Compressibility Analysis of Asymptotically Mean Stationary Processes

Jorge F. Silva*

Information and Decision System Group

Department of Electrical Engineering, University of Chile

Av. Tupper 2007, Room 508, Santiago, Chile

Abstract

This work provides new results for the analysis of random sequences in terms of ℓ_p-compressibility. The results characterize the degree in which a random sequences can be approximated by its best k-sparse version under different rates of significant coefficients (compressibility analysis). In particular, the notion of strong ℓ_p-characterization is introduced to denote a random sequence that has a well-defined asymptotic limit (sample-wise) of its best k-term approximation error when a fixed rate of significant coefficients is considered (fixed-rate analysis). The main theorem of this work shows that the rich family of asymptotically mean stationary (AMS) processes has a strong ℓ_p-characterization and we present results for the characterization and analysis of its ℓ_p-approximation error function. Furthermore adding ergodicity in the analysis of AMS processes, we have a theorem that shows that its approximation error function is constant and determined in closed-form by the stationary mean of the process. The results and analysis presented in this paper offer a contribution to the theory and understanding of discrete-time sparse processes and, on the technical side, confirm how instrumental the point-wise ergodic theorem is to determine the compressibility expression of discrete-time processes even when stationarity and ergodicity assumptions are relaxed.

Keywords: Sparse models, discrete time random processes, best k-term approximation error analysis, compressible priors, processes with ergodic properties, AMS processes, the ergodic decomposition theorem.

1. Introduction

Quantifying sparsity and compressibility for random sequences has been a topic of active research largely motivated by the results on sparse signal recovery and compressed sensing (CS) [1, 2, 3, 4, 5, 6, 7]. Sparsity and compressibility can be understood, in general, as the degree in which one can represent a random sequence (perfectly and loosely, respectively) by its best k-sparse version in the non-trivial regime when k (the number of significant coefficients) is smaller than the signal or ambient dimension. Various forms of compressibility for a random sequence have been used in different signal processing problems, for instance in regression [8], signal reconstruction (in the classical random Gaussian linear measuring setting used in CS) [2, 3], and inference-decision [9, 10].

A process is an infinite dimensional random object and then the standard approach used to measure compressibility for finite dimensional signals (based on the rate of decay of the absolute approximation error) does not extend naturally for this infinite dimensional analysis. Addressing this issue, Amini et al. [1] and Gribonval et al. [2] proposed the use of a relative approximation error analysis to measure compressibility with the objective to quantify the rate of the best k-approximation error with respect to the energy of the signal, when the number of significant coefficients scales at a rate proportional to the dimension of the signal. This approach offered a meaningful way to determine the energy (and more generally the ℓ_p-norm) concentration signature of independent and identically distributed (i.i.d.) processes [1, 2]. In particular, they introduced the concept of ℓ_p-compressibility to name a random sequence that has the capacity to concentrate (with very high probability) almost all their ℓ_p-relative energy in an arbitrary small number of coordinates (relative to the ambient dimension) of the canonical or innovation domain. Two important results were presented for i.i.d. processes. [1, Theorem 3] showed that i.i.d. processes with heavy tail distribution (including the generalized Pareto, Students's t and log-logistic) are ℓ_p-compressible for some ℓ_p-norms. On the other hand, [1, Theorem 1] showed that i.i.d, processes with exponentially decaying tails (such as Gaussian, Laplacian and Generalized Gaussians) are not ℓ_p-compressible for any ℓ_p-norm. Completing this analysis, Silva et al. [3] stipulated a necessary and sufficient

Email address: josilva@ing.uchile.cl (Jorge F. Silva*)

condition over the process distribution to be ℓ_p-compressible (in the sense of Amini et al.[1, Def.6]) that reduces to look at the p-moment of the 1D marginal stationary distribution of the process. Importantly, the proof of this result was rooted on the almost sure convergence of two empirical distributions (random object function of the process) to their respective probabilities as the number of samples goes to infinity[1]. This argument offered the context to move from using the law of large numbers (to characterize i.i.d. processes) to the use of the point-wise ergodic theorem [11, 12]. Then a necessary and sufficient condition for ℓ_p-compressibility was obtained for the family of stationary and ergodic sources under the mild assumption that the process distribution projected on one coordinate, i.e., its 1D marginal on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, has a density [3, Theorem 1]. Furthermore, for non ℓ_p-compressible process Silva et al. [3] provided a closed-form expression for the so called ℓ_p-approximation error function, meaning that a stable asymptotic value of the relative ℓ_p-approximation error is obtained when the rate of significant coefficient is given (fixed-rate analysis).

Considering that the proof of the main result in [3] relies heavily on an almost sure (with probability one) convergence of empirical means to their respective expectations, the idea of relaxing some of the assumptions of the process, in particular stationarity, is an interesting direction in the pursuit of extending results for the analysis of ℓ_p-compressibility for general discrete time processes. In this work, we have two new results in this direction extending the compressibility analysis for a family of random sequences where stationarity or ergodicity is not assumed. In particular, this work studies the rich family of processes with ergodic properties and, in particular, the important family of asymptotically mean stationary (AMS) processes [11, 13]. This family of processes has been studied in the context of source coding and channel coding problems where its ergodic properties (with respect to the family of indicator functions) has been used to extend fundamental performance limits in source and channel coding problems. In our context, the reason for studying AMS processes in the first place is because the ℓ_p-characterization in [3] is fundamentally rooted on a form of ergodic property over a family of indicator functions, which is precisely the family of measurable functions where

[1]These almost sure convergences created a family of typical sets that was used to prove the main result in [3, Theorem 1].

AMS sources has (by definition) a stable almost-sure asymptotic behavior [11].

1.1. Contributions of this Work

Specifically, we consider a more refined and relevant (sample-wise) almost sure fixed-rate analysis of ℓ_p-approximation errors, first considered by Gribonval *et al* [2] for the analysis of i.i.d. processes, to determine the relationship between the rate of significant coefficients and ℓ_p approximation error of the process. Our first main result (Theorem 1) shows that this rate vs. approximation error has a well-defined expression function of the process distribution (in particular the stationary mean of the process) for the complete collection of AMS and ergodic processes. This result relaxes stationary as well as some of the regularity assumptions used in [3, Theorem 1] and, consequently, it is a significant extension of that result. As a corollary of this theorem, we extend the dichotomy of the ℓ_p-compressible process presented in [3, Theorem 1] to the family of AMS ergodic processes (see Corollaries 2 and 3). The second main result of this work (Theorem 2) uses the celebrated ergodic decomposition theorem (EDT) [11] to extend the strong ℓ_p-characterization to the family of AMS processes, where ergodicity and the stationarity assumptions on the process have been relaxed. Remarkably, we show that this family of processes do have a stable (almost sure) asymptotic ℓ_p-approximation error for any given rate of significant coefficients as the block of the analysis tends to infinity. Interestingly, this limiting value is in general a measurable (non-constant) function of the process, which is fully determined by the so-called ergodic decomposition (ED) function that maps elements of the sample space of the process to stationary and ergodic components [11].

1.2. Organization of the Paper

The rest of the paper is organized as follows. Section 2 introduces notations, preliminary results and some basic elements of the ℓ_p-compressibility analysis. In particular, Section 2.1 introduces the fixed-rate almost sure approximation error analysis that is the focus of this work. Section 3 presents the two main results of this paper for AMS processes. The summary and final discussion are presented in Section 4. To conclude, Section 5 provides some context for the construction AMS sources based on the basic principle of passing an innovation process through deterministic (coding) and random (channel) processing stages. The proofs of the two main results are presented in Section 6, while the proofs of supporting results are relegated to the Appendices.

2. Preliminaries

For any vector $x^n = (x_1, .., x_n)$ in $\mathbb{R}^n$, let $(x_{n,1}, .., x_{n,n}) \in \mathbb{R}^n$ denote the ordered vector such that $|x_{n,1}| \geq |x_{n,2}| \geq \ldots |x_{n,n}|$. For $p > 0$ and $k \in \{1, .., n\}$, let

$$\sigma_p(k, x^n) \equiv (|x_{n,k+1}|^p + \ldots + |x_{n,n}|^p)^{\frac{1}{p}}, \tag{1}$$

be the best k-term ℓ_p-approximation error of x^n, in the sense that if

$$\Sigma_k^n \equiv \{x^n \in \mathbb{R}^n : \sigma_p(k, x^n) = 0\}$$

is the collection of k-sparse signals, then $\sigma_p(k, x^n)$ is the solution of $\min_{\tilde{x}^n \in \Sigma_k^n} ||x^n - \tilde{x}^n||_{\ell_p}$.

Amini *et al.* [1] and Gribonval *et al.* [2] proposed the following relative best k-term ℓ_p-approximation error

$$\tilde{\sigma}_p(k, x^n) \equiv \frac{\sigma_p(k, x^n)}{||x^n||_{\ell_p}} \in [0, 1], \; k \in \{1, .., n\}, \tag{2}$$

for the analysis of infinite sequences, with the objective of extending notions of compressibility to sequences that have infinite ℓ_p-norms in $\mathbb{R}^{\mathbb{N}}$. More precisely, let $X_1, .., X_n, ...$ be a one-side random sequence with values in $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$. $(X_n)_{n \geq 1}$ is fully characterized by its consistent family of finite dimensional probabilities denoted by $\{\mu^n \in \mathcal{P}(\mathbb{R}^n) : n \geq 1\}$ [12], where $X^n = (X_1, .., X_n) \sim \mu^n$ for all $n \geq 1$ and $\mathcal{P}(\mathbb{R}^n)$ is the collection of probabilities on the space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ [12, 11].

For $d \in (0, 1)$, $n \geq 1$ and $k \in \{1, .., n\}$, let us define the following set

$$\mathcal{A}_d^{n,k} \equiv \{x^n \in \mathbb{R}^n : \tilde{\sigma}_p(k, x^n) \leq d\}. \tag{3}$$

Definition 1. *[1, Defs.5 and 6] Let us consider a process $(X_n)_{n \in \mathbb{N}}$ with distribution $\mu = \{\mu^n, n \geq 1\}$, the set $\mathcal{A}_d^{n,k}$ is said to be ϵ-typical for X^n (or μ^n) if*

$$\mu^n(\mathcal{A}_d^{n,k}) \geq 1 - \epsilon. \tag{4}$$

For $\epsilon > 0$ and $d \in (0,1)$,

$$\tilde{\kappa}_p(d, \epsilon, \mu^n) \equiv \min \left\{ k \in \{1, \ldots, n\} : \mu^n(\mathcal{A}_d^{n,k}) \geq 1 - \epsilon \right\}, \tag{5}$$

is the critical number of terms that makes $\mathcal{A}_d^{n,k}$ ϵ-typical for μ^n.

From the definition of critical dimension in (5), we can study the asymptotic rate of innovation of the process relative to an ℓ_p-approximation error $d \in (0,1)$ by

$$\tilde{r}_p^+(d, \epsilon, \mu) \equiv \lim \sup_{n \to \infty} \frac{\tilde{\kappa}_p(d, \epsilon, \mu^n)}{n}, \tag{6}$$

$$\tilde{r}_p^-(d, \epsilon, \mu) \equiv \lim \inf_{n \to \infty} \frac{\tilde{\kappa}_p(d, \epsilon, \mu^n)}{n}, \tag{7}$$

for all $\epsilon > 0$, where $\mu \in \mathcal{P}(\mathbb{R}^{\mathbb{N}})$ in the notation is a short-hand for the process distribution given by $\{\mu^n : n \geq 1\}$.

Alternatively, we can consider the following fixed-rate asymptotic analysis:

Definition 2. *[3, Defs.5 and 6] Let $(X_n)_{n \in \mathbb{N}}$ be a process and let us consider $\epsilon \in (0,1)$, $r \in (0,1)$ and $d \in (0,1)$. The rate-distortion pair (r,d) is said to be ℓ_p-achievable for (X_n) with probability ϵ, if there exists a sequence of positive integers (k_n) such that $\lim\sup_{n\to\infty} \frac{k_n}{n} \leq r$ and*

$$\lim \inf_{n \to \infty} \mu^n(\mathcal{A}_d^{n,k_n}) \geq 1 - \epsilon. \tag{8}$$

Then, the rate-approximation error function of $(X_n)_{n \in \mathbb{N}}$ with probability ϵ is given by

$$r_p(d, \epsilon, \mu) \equiv \inf \left\{ r \in [0,1], \textit{such that } (r,d) \textit{ is } \ell_p\textit{-achievable for } (X_n) \textit{ with probability } \epsilon \right\}. \tag{9}$$

In general, it follows that $r_p(d, \epsilon, \mu) \leq \tilde{r}_p^+(d, \epsilon, \mu)$ [3, Prop. 2]. Furthermore, for the important case of stationary and ergodic processes, it was shown in [3, Th. 1] that

$$r_p(d, \epsilon, \mu) = \tilde{r}_p^+(d, \epsilon, \mu) = \tilde{r}_p^-(d, \epsilon, \mu) \text{ for all } d \in (0,1). \tag{10}$$

2.1. Revisiting the ℓ_p-Approximation Error Analysis

The approximation properties of a process $(X_n)_{n \in \mathbb{N}}$ presented above relies on a weak convergence (in probability) of the event $\mathcal{A}_d^{n,k}$ (see Defs. 1 and 2). Here, we introduce a stronger (almost sure) convergence of the approximation error at a given rate of innovation to study a more essential asymptotic indicator of the best k-term ℓ_p-approximation attributes of $(X_n)_{n\in\mathbb{N}}$. This notion will be meaningful for a large collection of processes (details presented in Section 2.2) and it will imply specific approximation attributes for μ in terms of $r_p(d,\epsilon,\mu)$, $r_p^+(d,\epsilon,\mu)$, and $r_p^-(d,\epsilon,\mu)$.

Definition 3. *A process* $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$*, with distribution* $\mu = \{\mu^n, n \geq 1\}$*, is said to have a strong rate vs.* ℓ_p *best k-term approximation error characterization (in short, a strong* ℓ_p*-characterization) if for any* $r \in (0,1]$ *and for any sequence of non-negative integers* $(k_n)_{n\in\in\mathbb{N}}$ *satisfying that* $\frac{k_n}{n} \longrightarrow r$ *then*

$$\lim_{n\to\infty} \tilde{\sigma}_p(k_n, X^n) = f_{p,\mu}(\mathbf{X}, r), \tag{11}$$

μ*-almost surely, where* $f_{p,\mu}(\mathbf{X}, r)$ *is a well-defined (measurable) function of* $\mathbf{X}$*.*

A process with a strong ℓ_p-characterization has an almost everywhere asymptotic (with n) pattern for its ℓ_p-approximation error when a finite rate of significant coefficients is considered (i.e., a fixed-rate analysis). On top of this condition, an interesting scenario to consider is when the limiting function $f_{p,\mu}(\mathbf{X}, r)$, in (11), is constant μ-almost surely. This can be interpreted as an ergodic property of $\mathbf{X}$ with respect to its best-k term ℓ_p-approximation error, reflecting a typical (almost sure) approximation attribute that is constant for the entire process[2]. The following result offers a connection between $f_{p,\mu}(\mathbf{X}, r)$ and $r_p(d,\epsilon,\mu)$ in this very special case.

Lemma 1. *Let us consider a process* $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ *and its process distribution* μ*. Let us assume that* $\mathbf{X}$ *has a strong* ℓ_p*-characterization (Def. 3) and that its limiting function in (11) is constant* μ*-almost surely, denoted by* $(f_{p,\mu}(r))_{r\in(0,1]}$*. Then we have the following:*

i) If $d \in \{f_{p,\mu}(r), r \in (0,1]\}$ *and* $d > 0$*, then there exists a unique* $r_o \in (0,1)$ *such that* $f_{p,\mu}(r_o) = d$*, where for any* $r \in (0,1)$ *and* $(k_n)_{n\in\mathbb{N}}$ *such that* $\frac{k_n}{n} \longrightarrow r$ *it follows that*

$$\lim_{n\to\infty} \mu^n(\mathcal{A}_d^{n,k_n}) = \begin{cases} 1 & \text{if } r > r_o \\ 0 & \text{if } r < r_o. \end{cases} \tag{12}$$

[2] The next section shows that $f_{p,\mu}(\mathbf{X}, r)$ is a constant function for the family of AMS and ergodic processes [11]. However, it is not constant function for stationary and AMS processes in general as presented in Section 3.2.

ii) On the other hand if $f_{p,\mu}(r_o) = 0$ for some $r_o \in (0,1)$, then $\forall d \in (0,1)$, for any $r \geq r_o$ and for any $(k_n)_{n\in\mathbb{N}}$ such that $\frac{k_n}{n} \longrightarrow r$, it follows that

$$\lim_{n\to\infty} \mu^n(\mathcal{A}_d^{n,k_n}) = 1. \tag{13}$$

This result shows that for a process with a strong ℓ_p-characterization and a constant rate approximation error function, there is a 0-1 phase transition on the asymptotic probability of the events $\mathcal{A}_d^{n,k_n}$ when $k_n/n \longrightarrow r$, which is governed by $(f_{p,\mu}(r))_{r\in(0,1]}$ in (11). More precisely, we have the following direct implication:

Corollary 1. *Under the assumptions of Lemma 1, for any $d \in \{f_{p,\mu}(r), r \in (0,1]\} \setminus \{0\}$, and $\epsilon > 0$*

$$r_p(d,\epsilon,\mu) = \tilde{r}_p^+(d,\epsilon,\mu) = \tilde{r}_p^-(d,\epsilon,\mu) = f_{p,\mu}^{-1}(d). \tag{14}$$

On the other hand, if $f_{p,\mu}(r) = 0$ for some $r_o \in (0,1]$, then $\forall \epsilon \in (0,1)$, $\forall d \in (0,1)$,

$$r_p(d,\epsilon,\mu) \leq \tilde{r}_p^+(d,\epsilon,\mu) \leq r_o. \tag{15}$$

It is worth noting in (14) that the weak ℓ_p-approximation error function $r_p(d,\epsilon,\mu)$ is independent of ϵ and fully determined by $(f_{p,\mu}^{-1}(d))_{d\in\{f_{p,\mu}(r),r\in(0,1]\}}$. This is consistent with the result obtained for i.i.d. and stationary and ergodic processes in [3]. The proof of Lemma 1 and Corollary 1 are presented in Section 6.1.

2.2. AMS Processes

Let us briefly introduce the family of AMS processes that is the main object of study of this work[3].

Definition 4. *A process $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ represented by the probability space $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}), \mu)$ is said to have an ergodic property with respect to a measurable function $f : (\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ if the sample average*

$$< f >_n (\mathbf{X}) = \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(\mathbf{X})) \tag{16}$$

[3]A complete exposition of sources with ergodic properties viewed as a dynamical system is presented in [11, Chapts. 7, 8 and 10].

converges μ-almost surely as n tends to infinity to a measurable function of $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ *that is denoted by* $< f > (\mathbf{X})$*. In (16), T denotes the standard shift operator*[4] *[11].*

Definition 5. *A process $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ (or, equivalently, its underlying dynamical system representation $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}), \mu, T)$) is said to have an ergodic property with respect to a class of measurable functions $\mathcal{M}$ if $< f >_n (\mathbf{X})$ convergences μ-almost surely to a well-defined limit $< f > (\mathbf{X})$ for any $f \in \mathcal{M}$.*

For any $n > 0$, let us define the set of arithmetic mean probabilities by

$$\mu_n(F) = \frac{1}{n}\sum_{i=0}^{n-1} \mu(T^{-i}(F)), \tag{17}$$

for all $F \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$.

Definition 6. *A process $(X_n)_{n\geq 0}$ with distribution μ is said to be asymptotically mean stationary (AMS)*[5]*, if $\mu_n(F)$ in (17) convergences as n goes to infinity for all $F \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$.*

By the construction in (17), it is clear that $\{\mu_n : n \geq 1\} \subset \mathcal{P}(\mathbb{R}^{\mathbb{N}})$. Then if the limit of μ_n exists, in the sense that $(\mu_n(F))_{n\geq 1}$ convergences in $\mathbb{R}$ as n tends to infinity for any measurable event $F \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$, these values (indexed by $F \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$) induce a well defined probability in $\mathcal{P}(\mathbb{R}^{\mathbb{N}})$ [11, Lemma 7.4]. This object is called the stationary mean of μ and it is denoted by $\bar{\mu}$. It can be proved that $\bar{\mu}$ is a stationary measure with respect to T, in the sense that $\bar{\mu}(F) = \bar{\mu}(T^{-1}(F))$ for all $F \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$[6].

The following important result connects processes with ergodic properties and AMS processes:

Lemma 2. *[11, Ths. 7.1 and 8.1] A necessary and sufficient condition for a process $(X_n)_{n\geq 0}$ (and its distribution μ) to be AMS is that it has an ergodic property with respect to the family of indicator functions, i.e., $\left\{\mathbf{1}_F(x) : F \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})\right\}$.*

A stronger ergodic property to ask on $\mathbf{X}$ over a family $\mathcal{M}$ is that for all $f \in \mathcal{M}$ the sample average in (16) tends to a well-defined limit $< f > (\mathbf{X})$ that is constant μ-almost surely. For this analysis, it is relevant to introduce the following definition, which derives from the celebrated point-wise ergodic theorem for AMS sources [11, Th. 7.5]:

[4] For $\bar{x} \in \mathbb{R}^{\mathbb{N}}$, $\bar{z} = T(\bar{x})$ is given by the coordinate-wise relationship $z_i = x_{i+1}$ for all $i \geq 1$.

[5] By definition, if $(X_n)_{n\geq 0}$ is stationary then it is AMS.

[6] The process (X_n) is said to be stationary if its distribution μ is stationary with respect to T.

Definition 7. *A process* $\mathbf{X}$ *(or its equivalent dynamical system* $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}), \mu, T)$*) is said to be ergodic, if the collection of invariant events (the events* F *such that* $T^{-1}(F) = F$*) has* μ*-probability 1 or 0 [12, 11].*

The following result connects these concepts:

Lemma 3. *[11, Th. 7.5 and Lem. 7.14] A necessary and sufficient condition for an AMS process* $(X_n)_{n\geq 0}$ *to have a constant ergodic property for the family* $\left\{\mathbf{1}_F(x) : F \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})\right\}$ *is that* $(X_n)_{n\geq 0}$ *is ergodic.*

In general, AMS processes are not ergodic. In fact, the following result provides a condition for $\mathbf{X}$ to meet ergodicity that can be considered a form of a weak mixing (asymptotic independence) condition [11].

Lemma 4. *[11, Lem. 7.15] A necessary and sufficient condition for an AMS process* $(X_n)_{n\geq 0}$ *to be ergodic is that*

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} \mu(T^{-i}(F) \cup F) = \bar{\mu}(F)\mu(F) \tag{18}$$

for all $F \in \mathcal{F}$*, where* $\mathcal{F} \subset \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ *is the collection that generates* $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$*.*

For the case when the process is stationary, it follows that $\bar{\mu}(F) = \mu(F)$ for all $F \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$, then the condition in (18) can be interpreted as a mixing (asymptotic independence) property on $(X_n)_{n\geq 0}$.

3. Characterizing AMS Processes

Here we present the two main results of this paper beginning with the scenario of an AMS and ergodic process.

3.1. Strong ℓ_p-Characterization for AMS and Ergodic Processes

When an AMS process satisfies the mixing condition in (18) and, consequently, it is ergodic, we can state the following result:

Theorem 1. *Let* $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$*, or equivalently* $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}), \mu)$*), be an AMS and ergodic process (or dynamical system) with respect to* T *and let* $\bar{\mu} = \{\bar{\mu}_n : n \geq 1\}$ *be its stationary mean. Then for any* $p > 0$*,* $r \in (0,1]$*, and* $(k_n)_{n\geq 1}$ *with* $k_n \to r$*, it follows that*

$$\lim_{n\to\infty} \tilde{\sigma}_p(k_n, X_1^n) = f_{p,\bar{\mu}}(r), \ \mu - \textit{almost surely}, \tag{19}$$

where $f_{p,\bar{\mu}}(r)$ *is a well defined function of the stationary mean* $\bar{\mu}$ *of* $\mathbf{X}$ *projected over 1 dimensional cylinders in* $\mathcal{B}(\mathbb{R}^{\mathbb{N}})$ *(i.e., the marginal* $\bar{\mu}_1$ *in* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*). More precisely,* $(f_{p,\bar{\mu}}(r))_{r\in(0,1]}$ *is function of* $\bar{\mu}_1 \in \mathcal{P}(\mathbb{R})$ *with the following characterization:*

i) If the function $(x^p)_{x\in\mathbb{R}} \notin L_1(\bar{\mu}_1)$*, then it follows that*[7]

$$f_{p,\bar{\mu}}(r) = 0, \ \forall r \in (0,1].$$

ii) If $(x^p)_{x\in\mathbb{R}} \in L_1(\bar{\mu}_1)$ *and* $\bar{\mu}_1 \ll \lambda$*, we can introduce an induced probability* $v_p \in \mathcal{P}(\mathbb{R})$ *with* $v_p \ll \bar{\mu}_1$ *given by*

$$v_p(B) = \frac{\int_B |x|^p \, d\bar{\mu}_1(x)}{\int_{\mathbb{R}} |x|^p \, d\bar{\mu}_1(x)}, \forall B \in \mathcal{B}(\mathbb{R}).$$

Then for any $r \in (0,1]$*,*

$$f_{p,\bar{\mu}}(r) = \sqrt[p]{1 - v_p(B_{\tau(r)})},$$

where $B_\tau \equiv (-\infty, \tau] \cup [\tau, \infty)$ *and* $\tau(r) > 0$ *is the unique solution of:*

$$\bar{\mu}_1(B_\tau) = r.$$

ii) If $(x^p)_{x\in\mathbb{R}} \in L_1(\bar{\mu}_1)$ *and* $\bar{\mu}_1$ *is not absolutely continuous with respect to* λ*,*[8] *let us introduce the set*

$$\mathcal{R}^* \equiv \{\bar{\mu}_1(B_\tau), \tau \in [0,\infty)\} \ \textit{and} \ C_\tau \equiv (-\infty, \tau) \cup (\tau, \infty).$$

If $r \in \mathcal{R}^*$*, similarly to the previous case, it follows that*

$$f_{p,\bar{\mu}}(r) = \sqrt[p]{1 - v_p(B_{\tau(r)})},$$

[7] A real measurable function f is integrable with respect to $v \in \mathcal{P}(\mathbb{R})$ if $\int_{\mathbb{R}} |f(x)| \, dv(x) < \infty$. $L_1(v)$ denotes the collection of v-integrable functions [12, 14].

[8] In other words, $\bar{\mu}_1$ has atomic components.

where $\tau(r)$ *is the solution of* $\bar{\mu}_1(B_\tau) = r$.

On the other hand if $r \notin \mathcal{R}^*$, *there is* $\tau_o > 0$ *such that* $\bar{\mu}_1(\{-\tau_o, \tau_o\}) > 0$ *and* $r \in [\bar{\mu}_1(C_{\tau_o}), \bar{\mu}_1(B_{\tau_o}))$. *Then* $\exists \alpha_o \in [0,1)$ *such that*

$$r = \bar{\mu}_1(C_{\tau_o}) + \alpha_o(\bar{\mu}_1(B_{\tau_o}) - \bar{\mu}_1(C_{\tau_o})),$$

where

$$f_{p,\bar{\mu}}(r) = \sqrt[p]{1 - v_p(C_{\tau_o}) - \alpha_o(v_p(B_{\tau_o}) - v_p(C_{\tau_o}))}.$$

In the last expression,

$$v_p(B_{\tau_o}) - v_p(C_{\tau_o}) = v_p(\{-\tau_o, \tau_o\}) = |\tau_o|^p \cdot \bar{\mu}_1(\{-\tau_o, \tau_o\}) / \, ||(x^p)||_{L_1(\bar{\mu}_1)} > 0$$

if $\tau_o > 0$.

The proof of this result is presented in Section 6.2.

3.1.1. Analysis and Interpretations of Theorem 1

- The general result in (19) shows that any ergodic AMS process has a strong ℓ_p-characterization (Def. 3) where its point-wise (almost sure) approximation error function in (11) is completely determined by the 1D projection of its stationary mean, i.e., $\bar{\mu}_1 \in \mathcal{P}(\mathbb{R})$.

- Two important scenarios can be highlighted. The case $(x^p)_{x\in\mathbb{R}} \notin L_1(\bar{\mu}_1)$ in which $f_{p,\bar{\mu}}(r) = 0, \ \forall r \in (0,1]$ and the case $(x^p)_{x\in\mathbb{R}} \in L_1(\bar{\mu}_1)$ that has a non-trivial approximation error function expressed by the following collection of (rate, distortion) pairs:

$$\{(r, f_{p,\bar{\mu}}(r)), r \in (0,1]\} = \left\{ (\phi_{\bar{\mu}_1}(\tau), \sqrt[p]{1 - \phi_{v_p}(\tau)}), \tau \in [0, \infty) \right\}$$
$$\bigcup_{\tau_n \in \mathcal{Y}_{\bar{\mu}_1}} \bigcup_{\alpha \in [0,1)} \left\{ (\bar{\mu}_1(C_{\tau_n}) + \alpha \bar{\mu}_1(\{-\tau_n, \tau_n\}), \sqrt[p]{1 - v_p(C_{\tau_n}) - \alpha v_p(\{-\tau_n, \tau_n\})}) \right\} \tag{20}$$

  where $\mathcal{Y}_m = \{\tau \in [0,\infty), \bar{\mu}_1(\{-\tau, \tau\}) > 0\}$, which is shown to be at most a countable set. It is worth noting that the expression in (20) summarizes the continuous and non-continuous result stated in ii) and iii). The details of this analysis are presented in Section 6.2.

- Considering Lemma 1 and Corollary 1 (in particular the relationship expressed in Eq.(14) that connects the strong ℓ_p-caracterization in Def. 3 with the weak ℓ_p-characterization in (9)), Theorem 1 implies the following result:

Corollary 2. *If $(x^p)_{x\in\mathbb{R}} \in L_1(\bar{\mu}_1)$ then for any $d \in (0,1)$ and $\forall \epsilon > 0$,*

$$r_p(d,\epsilon,\mu) = \tilde{r}_p^+(d,\epsilon,\mu) = \tilde{r}_p^-(d,\epsilon,\bar{\mu}) = f_{p,\bar{\mu}}^{-1}(d).$$

Otherwise, if $(x^p)_{x\in\mathbb{R}} \notin L_1(\bar{\mu}_1)$, then for any $d \in (0,1)$ and $\epsilon > 0$,

$$r_p(d,\epsilon,\mu) = \tilde{r}_p^+(d,\epsilon,\mu) = \tilde{r}_p^-(d,\epsilon,\bar{\mu}) = 0.$$

- At this point, it is important to revisit the concept of ℓ_p-compressible processes introduced by Amini et al. [1] in light of Theorem 1 and Corollary 2.

Definition 8. *[1, Def.6] A process (X_n), with distribution μ, is said to be ℓ_p-compressible for $p > 0$, if for any $\epsilon \in (0,1)$ and $d \in (0,1)$, $\tilde{r}_p^+(d,\epsilon,\mu) = 0$.*

Then from Lemma 1 and Theorem 1, the following can be stated:

Corollary 3. *A necessary and sufficient condition for an AMS ergodic process (with stationary mean $\bar{\mu}$) to be ℓ_p-compressible is that $(x^p)_{x\in\mathbb{R}} \notin L_1(\bar{\mu}_1)$.*

Corollary 3 extends the dichotomy presented for the stationary and ergodic case in [3, Theorem 1].

- In the case where $(x^p)_{x\in\mathbb{R}} \in L_1(\bar{\mu}_1)$, the function $(f_{p,\bar{\mu}}(r))_{r\in(0,1]}$ given by (20) is proved to be continuous, strictly non-increasing in the domain $f_{p,\bar{\mu}}^{-1}(0,1)$ and achieving the range $[0,1)$ in the sense that for all $d \in [0,1)$ there exists $r \in (0,1]$ such that $f_{p,\bar{\mu}}(r)) = d$, where in addition $\lim_{r\to 0} f_{p,\bar{\mu}}(r) = 1$. The details of this analysis are presented in Lemma 8 (Section 6.2 and Appendix A).

- Furthermore when the process is not ℓ_p-compressible, i.e., $(x^p)_{x\in\mathbb{R}} \in L_1(\bar{\mu}_1)$, we highlight two important sub-scenarios: the sparse case, meaning that $\bar{\mu}_1(\{0\}) > 0$, and the non-sparse case meaning that $\bar{\mu}_1(\{0\}) = 0$. For the non ℓ_p-compressible and sparse case, it follows that zero approximation error is achieved at rates that are strictly smaller than 1. More precisely,

Corollary 4. *Let us consider a sparse AMS process, meaning that* $\bar{\mu}_1(\{0\}) > 0$*, with* $(x^p)_{x \in \mathbb{R}} \in L_1(\bar{\mu}_1)$ *then*

$$f_{p,\bar{\mu}}(r) = 0, \ \forall r \in [1 - \bar{\mu}_1(\{0\}), 1],$$

while if $r \in (0, 1 - \bar{\mu}_1(\{0\}))$ *then* $f_{p,\bar{\mu}}(r) > 0$*.*

On the other hand, for the ℓ_p-compressible and non-sparse case, zero distortion is exclusively achieved at a rate equal to 1, i.e., $f_{p,\bar{\mu}}(r) > 0$ if $r \in (0, 1)$. This result is elaborated in Section 6.2 and Appendix B.

3.2. Strong ℓ_p-Characterization for AMS Processes

Relaxing the ergodic assumptions for an AMS source is the focus of this part. It is worth noting that the ergodic result in Theorem 1 will be instrumental for this analysis in view of the ergodic decomposition (ED) theorem for AMS sources nicely presented in [11, Ths. 8.3 and 10.1] and references therein. In a nutshell, the ED theorem shows that the stationary mean of an AMS process (see Def. 6) can be decomposed as a convex combination of stationary and ergodic distributions (called the ergodic components) in $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$.

It is important to introduce one specific aspect of this result for the statement of the following theorem. Let us consider an arbitrary AMS process $(X_n)_{n \geq 1}$ equipped with its process distribution $\mu \in \mathcal{P}(\mathbb{R}^{\mathbb{N}})$ and its induced stationary mean $\bar{\mu} \in \mathcal{P}(\mathbb{R}^{\mathbb{N}})$. If we denote by $\tilde{\mathcal{P}} \subset \mathcal{P}(\mathbb{R}^{\mathbb{N}})$ the family of stationary and ergodic probabilities with respect to the shift operator, then one of the implications of the ergodic decomposition theorem [11, Ths. 8.3 and 10.1] is that there is a measurable space $(\Lambda, \mathcal{L})$ indexing this family, i.e., $\tilde{\mathcal{P}} = \{\mu_\lambda, \lambda \in \Lambda\}$. More importantly, there is a measurable function $\Psi : (\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}})) \to (\Lambda, \mathcal{L})$ that maps points in the sequence space $\mathbb{R}^{\mathbb{N}}$ to stationary and ergodic components (more details will be given in Section 6.3). Then using Ψ, there is a probability measure W_Ψ in $(\Lambda, \mathcal{L})$ induced by μ in the standard way, where $\forall A \in \mathcal{L}$, we have that $W_\Psi(A) = \mu(\Psi^{-1}(A))$. One of the implications of the ED theorem [11, Ths.

8.3 and 10.1] is that for all $F \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$[9]

$$\bar{\mu}(F) = \int \mu_\lambda(F) \partial W_\Psi(\lambda). \tag{21}$$

In other words, $\bar{\mu}$ can be expressed as the convex combination of stationary and ergodic components $\{\mu_\lambda, \lambda \in \Lambda\}$, where the mixture probability on $(\Lambda, \mathcal{L})$ is induced by the decomposition function Ψ, which is universal, meaning that the same function is valid to decompose any stationary distribution on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ in stationary and ergodic components in the sense presented in (21).

The following result uses the ED theorem for AMS sources [11, Ths. 8.3 and 10.1] and Theorem 1 to show that AMS sources have a strong ℓ_p-characterization as stated in Definition 3. Furthermore, the result offers an expression to specify the limit $(f_{p,\mu}(\mathbf{X}, r))_{(0,1]}$ in (11).

Theorem 2. *Let $\mathbf{X} = (X_n)_{n \in \mathbb{N}}$ be an AMS process with process distribution μ. Let us consider $\tilde{\mathcal{P}} = \{\mu_\lambda, \lambda \in \Lambda\}$ the collection of stationary and ergodic probabilities and the decomposition function $\Psi : (\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}})) \to (\Lambda, \mathcal{L})$ presented in the ED theorem [11, Th 10.1]. Then it follows that:*

i) The process $\mathbf{X} = (X_n)_{n \in \mathbb{N}}$ has a strong ℓ_p-characterization (Def. 3), where for any $r \in (0, 1]$ and $(k_n)_{n \geq 1}$ such that $k_n/n \to r$

$$\lim_{n \to \infty} \tilde{\sigma}_p(k_n, X_1^n) = f_{p,\mu}(\mathbf{X}, r) = f_{p,\mu_{\Psi(\mathbf{X})}}(r), \mu - \textit{almost surely}, \tag{22}$$

where $(f_{p,\mu_\lambda}(r))$ has been introduced and developed in Theorem 1[10].

ii) For any $r \in (0, 1]$, $d \in [0, 1)$, and (k_n) such that $k_n/n \to r$,

$$\begin{aligned} \lim_{n \to \infty} \mu^n(A_d^{n,k_n}) &= \int \lim_{n \to \infty} \mu_\lambda^n(A_d^{n,k_n}) \partial W_\Psi(\lambda) \\ &= \mu\left(\left\{\mathbf{x} \in \mathbb{R}^{\mathbb{N}} : \mu_{\Psi(\mathbf{x})} \textit{ is } \ell_p\textit{-compressible}\right\}\right) \\ &\quad + \mu\left(\left\{\mathbf{x} \in \mathbb{R}^{\mathbb{N}} : \mu_{\Psi(\mathbf{x})} \textit{ is not } \ell_p\textit{-compressible and } f_{p,\mu_{\Psi(\mathbf{x})}}(r) \leq d\right\}\right). \end{aligned} \tag{23}$$

The proof of this result is presented in Section 6.3.

[9]The assumption here is that for any F, $\mu_\lambda(F)$, as a function of λ, is measurable from $(\Lambda, \mathcal{L})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ [11].

[10]Note that $\forall \lambda \in \Lambda$, $\mu_\lambda \in \tilde{\mathcal{P}}$ is a stationary and ergodic process.

3.2.1. Analysis and Interpretations of Theorem 2:

- The first almost-sure point wise result in (22) provides a closed-form expression for the ℓ_p-characterization of the process $\mathbf{X}$ given by $f_{p,\mu}(\mathbf{X}, r)$, which is a function of $\mathbf{X}$ (not a constant function in general) through the ED function $\Psi(\cdot)$ that maps $\mathbf{X}$ to stationary and ergodic components in $\tilde{\mathcal{P}}$.

- An interesting interpretation of the result in (22), which is a consequence of the ED theorem, is that this limiting behaviour can be seen as if one selects at $t = 0$ an ergodic component $\mu_\lambda \in \tilde{\mathcal{P}}$ and then the process evolves with the statistic of μ_λ, which has a stable asymptotic characterized by Theorem 1. This is equivalent to state that there is one stationary ergodic component that is active all the time, but we do not know a priori which component. In fact to resolve the component that is active, we need to know the entire process $\mathbf{X}$, as the active component $\lambda \in \Lambda$ is given by $\Psi(\mathbf{X})$. This interpretation has a natural connection with the standard setting used in universal source coding as clearly argued by Gray and Kieffer in [13], where it is assumed that a process is fixed and belongs to a family of process distributions from beginning to end, but the observer (or the designer of the coding scheme) does not know which specific distribution is active. Therefore, when observing a realization of an AMS process, what we are really observing is a realization of one (unknown a priori) stationary and ergodic component in $\tilde{\mathcal{P}}$ and, consequently, its limiting behaviour is well defined as expressed in (22). The fact that this limit is expressed as a function of Ψ can be understood from the perspective that Ψ is the object that chooses the active component in $\tilde{\mathcal{P}}$ from $\mathbf{X}$.

- An intriguing aspect of this result, which is again a consequence of the ED theorem for AMS sources, is that if we look at the limit $f_{p,\mu}(\mathbf{X}, r)$ in (22), this is equal to $f_{p,\mu_{\Psi(\mathbf{X})}}(r)$, which does not depend on μ explicitly as long as μ is AMS. Then, we could say that the ED function Ψ characterizes the asymptotic limit for any AMS source universally.

- When we move to the weak ℓ_p-characterization result expressed in (23) (see Defs. 1 and 2), here we can observe explicitly the role of the distribution μ in the analysis, which is consistent with the almost sure result in (22). In the expression in the

LHS of (23), we note that the probability $\mu^n(A_d^{n,k_n})$ has a limit determined by the pair (r, d) and the distribution μ.

- Complementing the previous point, there are two clear terms in (23): the first is the probability (over μ) of the sequences that map through $\Psi(\cdot)$ to ℓ_p-compressible components in $\tilde{\mathcal{P}}$. The second term is the probability of the sequences that map through $\Psi(\cdot)$ to ergodic components that are not ℓ_p-compressible and satisfy that its ℓ_p-approximation error function (which is characterized in Theorem 1) evaluated at the rate r is smaller than the distortion d. Note that these two events on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ are distribution independent (universals) and therefore can be determined a priori (independent of μ) for this weak ℓ_p-compressibility analysis.

4. Summary and Discussion of the Results

In this work, we revisit the notion of ℓ_p-compressibility focusing on the study of the almost sure (with probability one) limit of the ℓ_p-relative best k-term approximation error when a fixed-rate of significant coefficients is considered for the analysis. We consider the study of processes with general ergodic properties relaxing the stationarity and ergodic assumptions considered in previous work. Interestingly, we found that the family of asymptotically mean stationary (AMS) processes has an (almost-sure) stable ℓ_p approximation error behavior (sample-wise) when considering any arbitrary rate of significant coefficients per dimension of the signal. In particular, our two main results offer expressions for this limit, which is a function of the entire process through the known ergodic decomposition (ED) mapping used in the proof the celebrated ED theorem. When ergodicity is added and we assume an AMS ergodic source, the ℓ_p-approximation error function reduces to a closed-form expression of the stationary mean of the process. As a corollary, we extend the dichotomy between ℓ_p-compressibility and non ℓ_p-compressibility observed in a previous result [3, Th.1].

In summary, the two main theorems of this paper significantly extend previous results in the literature of this problem (that were valid under the assumption of stationary, ergodicity and some extra regularity conditions on the process distributions) and on the technical side show the important role that the general point-wise ergodic theorem and, in particular, the ED theorem play for the extension of the ℓ_p-compressibility analysis to families of processes with general ergodic properties. Finally, from the proof of Theorem 1, we notice that being able to impose an ergodic property for the family of indicator functions is essential to obtain a stable (almost sure) result for the ℓ_p-approximation error function, in the way expressed in Def. 3, and, consequently, the AMS assumption for the process (see Lemma 2) seems to be crucial to achieve the desired strong (almost-sure) ℓ_p-approximation property declared in Definition 3.

5. On the Construction and Processing of AMS Processes

To conclude this paper, we provide some context to support the application of our results in Section 3. We consider a general generative scenario where a process is constructed as the output of an innovation source passing through a signal processor (or coding process) and a random corruption (or channel). In other words, we want to have an idea of the family of operations on a stationary and ergodic source (for example an i.i.d. source) that produces a process with a strong ℓ_p-characterization (Def. 3). For that we briefly revisit known results that guarantee that a process has stationarity and/or ergodic properties when it is produced (deterministically or randomly) from a stationary and ergodic source.[11]

A general way of representing a transformation of a process $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ into another process is using the concept of a channel. A channel is a collection of probabilities (or process distributions) in $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ indexed by elements in $\mathbb{R}^{\mathbb{N}}$, i.e., $\mathcal{C} = \left\{v_{\bar{x}}, \bar{x} \in \mathbb{R}^{\mathbb{N}}\right\} \subset \mathcal{P}(\mathbb{R}^{\mathbb{N}})$ such that for all $F \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ $v_{\bar{x}}(F)$ is a measurable function from $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then given μ the process distribution of $(X_n)_{n\in\mathbb{N}}$, the channel $\mathcal{C}$ induces a joint distribution in the product space $(\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}))$ by

$$\mu\mathcal{C}(F \times G) = \int_{\bar{x}\in F} v_{\bar{x}}(G) d\mu(\bar{x}),\ \forall F, G \in \mathcal{B}(\mathbb{R}^{\mathbb{N}}).$$

The joint process distribution is denoted by $\mu\mathcal{C}$. Then a new process $\mathbf{Y} = (Y_n)_{n\in\mathbb{N}}$ is obtained at the output of the channel when $(X_n)_{n\in\mathbb{N}}$ is its input. If we denote the distribution of $\mathbf{Y}$ by v, this is obtained by the marginalization of $\mu\mathcal{C}$, i.e., $v(G) \equiv \mathbb{P}(\mathbf{Y} \in G) = \mu\mathcal{C}(\mathbb{R}^{\mathbb{N}} \times G)$ for all $G \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$. Considering the shift operator T uses to characterize

[11] A complete exposition can be found in [15, Ch.2].

stationarity and ergodic properties for processes in $\mathbb{R}^{\mathbb{N}}$ (Sec. 2.2), the channel $\mathcal{C} = \left\{v_{\bar{x}}, \bar{x} \in \mathbb{R}^{\mathbb{N}}\right\}$ is said to be stationary with respect to T if [15, Sec. 2.3]

$$v_{T(\bar{x})}(G) = v_{\bar{x}}(T^{-1}(G)), \forall \bar{x} \in \mathbb{R}^{\mathbb{N}}, \forall G \in \mathcal{B}(\mathbb{R}^{\mathbb{N}}).$$

Then, the following result can be obtained:

Lemma 5. *[15, Lemma 2.2] Let us consider an AMS process* $\mathbf{X}$ *(with stationary mean given by* $\bar{\mu}$*) as the input of a stationary channel* $\mathcal{C} = \left\{v_{\bar{x}}, \bar{x} \in \mathbb{R}^{\mathbb{N}}\right\}$*. Then the output process* $\mathbf{Y}$ *is AMS and its stationary mean is given by*[12]

$$\bar{v}(G) = \lim_{n \longrightarrow \infty} 1/n \sum_{i=0}^{n-1} v(T^{-i}G) = \bar{\mu}\mathcal{C}(\mathbb{R}^{\mathbb{N}} \times G) \text{ for all } G \in \mathcal{B}(\mathbb{R}^{\mathbb{N}}).$$

Remarkably, Lemma 5 shows a general (stationary) random approach to produce AMS processes from another AMS process. Furthermore, the result provides a closed expression for the resulting stationary mean (function of the stationary mean of the input $\bar{\mu}$ and the channel $\mathcal{C}$), which is the object that determines its strong ℓ_p-compressibility signature from Theorems 1 and 2.

Furthermore adding ergodicity, we highlight the following result:

Lemma 6. *[15, Lemma 2.7] If the channel* $\mathcal{C} = \left\{v_{\bar{x}}, \bar{x} \in \mathbb{R}^{\mathbb{N}}\right\}$ *is weakly mixing in the sense that for all* $\bar{x} \in \mathbb{R}^{\mathbb{N}}$ *and measurable rectangles* F*,* $G \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$

$$\lim_{n \longrightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \left| v_{\bar{x}}(T^{-i}(F) \cap G) - v_{\bar{x}}(T^{-i}(F)) v_{\bar{x}}(G) \right| = 0,$$

then if the input process is AMS and ergodic then the output of the channel is also AMS and ergodic.

We will cover two important families of channels below.

5.1. Deterministic Channels: Stationary Codes and LTI Systems

A deterministic transformation (or measurable function) of an AMS process $\mathbf{X} = (X_n)_{n \in \mathbb{N}}$ can be seen as an important example of the channel framework presented above.

[12] The result shows more generally that the joint process $(\mathbf{X}, \mathbf{Y})$ is AMS with respect to $T \times T$ ($T \times T(\bar{x}, \bar{y}) = (T(\bar{x}), T(\bar{y}))$), where its stationary mean is $\bar{\mu}\mathcal{C}$.

Let us consider a measurable function $f : (\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}})) \longrightarrow (\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ and the induced process $\mathbf{Y} = f(\mathbf{X})$, where the process distribution is $v(G) = \mu(f^{-1}(G))$. This form of encoding $\mathbf{X}$ is a special case of a channel, where $v^f_{\bar{x}}(G) = \mathbf{1}_{f^{-1}(G)}(\bar{x})$. Importantly, the deterministic channel $\mathcal{C}^f \equiv \left\{v^f_{\bar{x}}, \bar{x} \in \mathbb{R}^{\mathbb{N}}\right\}$ induced by f is stationary if, and only if, $f(T(\bar{x})) = T(f(\bar{x}))$ [15]. In this context, we say that f produces a stationary coding of $\mathbf{X}$.

Corollary 5. *Any stationary coding of an AMS process produces an AMS process, where the stationary mean of* $\mathbf{Y}$ *is given by* $\bar{v}(G) = \bar{\mu}(f^{-1}(G))$ *for all* $G \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$*.*

The proof of this result follows directly from Lemma 5.

There is a stronger result for deterministic and stationary channels:

Lemma 7. *[15, Lemma 2.4] Let us consider a deterministic and stationary channel* $\mathcal{C}^f$*. If the input process to* $\mathcal{C}^f$ *is AMS and ergodic then the output process is AMS and ergodic.*

It is worth noting that a direct way of constructing stationary coding is by a scalar measurable function $\phi : (\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}})) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where given $\bar{x} \in \mathbb{R}^{\mathbb{N}}$ the output is produced by $y_n = \phi(T^n(\bar{x}))$ for all $n \geq 0$.[13] Then, there is an infinity collection of stationary coding that preserves the AMS and ergodic characteristics of an input process. Two emblematic cases to consider are the finite length sliding block code where $y_n = \phi(X_{n+M}, ..., X_{n+D})$ with $D > M \geq 0$ and $\phi : \mathbb{R}^{D-M+1} \longrightarrow \mathbb{R}$, and the case when ϕ is linear function, i.e., $\phi(\bar{x}) = \sum_{i\geq 0} a_i \cdot x_i$, and, consequently, f produces a linear and time invariant (LTI) coding of $\mathbf{X}$.[14]

5.2. Memoryless Channels

A channel $\mathcal{C} = \left\{v_{\bar{x}}, \bar{x} \in \mathbb{R}^{\mathbb{N}}\right\}$ is said to be memoryless if for any finite dimensional cylinder $\times_{i\in J} F_i \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$ and for any $\bar{x} \in \mathbb{R}^{\mathbb{N}}$ it follows that $v_x(\times_{i\in J} F_i) = \prod_{i\in J} p_{x_i}(F_i)$ where $\{p_x, x \in \mathbb{R}\} \subset \mathcal{P}(\mathbb{R})$. Basically we have that probabilities of the channel decompose as the multiplication of its marginals (memoryless). The classical example is the

[13] Conversely, for any stationary code f there is a function $\phi(\bar{x}) = \pi_0(f(\bar{x}))$ that induces f, where $\pi_0()$ denotes the first coordinate projection of the sequence.

[14] Stationary codings play an important role in ergodic theory for the analysis of isomorphic processes [11].

additive white Gaussian noise (AWGN) channel used widely in signal processing and communications where $p_x = \mathcal{N}(\mu_x, \sigma)$ is a normal distribution with the mean depending on x and $\sigma > 0$. It is simple to check that memoryless channels are stationary. Consequently, Lemma 5 tells us that a memoryless corruption of an AMS process produces an AMS process. In addition, the mixing condition of Lemma 6 is simple to verify for memoryless channels. Consequently, a memoryless corruption of an AMS and ergodic process preserves the ergodicity of the input at the output of the channel.

Finally, using Lemmas 5, 6 and 7, we can have a rich collection of processing steps where AMS as well as AMS and ergodicity are preserved from the input to the output and, consequently, Theorems 1 and 2 can be adopted for compressibility analysis of these processes.

6. Proofs of the Main Results

6.1. Lemma 1 (and Corollary 1)

Proof: First, some properties of $(f_{p,\mu}(r))_{r\in(0,1]}$ will be needed.

Proposition 1. *It follows that:*

- *If* $0 < r_1 < r_2 \leq 1$, *then* $f_{p,\mu}(r_2) \leq f_{p,\mu}(r_1)$.
- *If* $0 < r_1 < r_2 \leq 1$ *and* $f_{p,\mu}(r_2) = f_{p,\mu}(r_1)$, *then* $f_{p,\mu}(r_2) = f_{p,\mu}(r_1) = 0$.

The proof of this result derives directly from the definition of $\tilde{\sigma}_p(k_n, X^n)$ and some basic inequalities.[15] From Proposition 1, $f_{p,\mu}(\cdot)$ is strictly monotonic and injective in the domain $f^{-1}_{p,\mu}(0,1)$. Therefore, $f^{-1}_{p,\mu}(d)$ is well defined for any $d \in \{f_{p,\mu}(r), r \in (0,1]\} \setminus \{0\}$.

Let us first consider the case $d \in \{f_{p,\mu}(r), r \in (0,1]\} \setminus \{0\}$ assuming for a moment that this set is non-empty. Then, there exists $r_o \in (0,1)$ such that $d = f_{p,\mu}(r_o)$, where by the strict monotonicity of $f_{p,\mu}(\cdot)$, we have that $f_{p,\mu}(r_2) < d < f_{p,\mu}(r_1)$ for any $r_1 < r_o < r_2 \leq 1$. On the other hand, using the convergence of the approximation error to the function $f_{p,\mu}(r)$ in (11) and the definition of $\mathcal{A}^{n,k}_d$ in (3), it follows that for any $r \in (0,1)$, k_n with $k_n/n \longrightarrow r$, and $\epsilon > 0$

$$\lim_{n\to\infty} \mu^n(\mathcal{A}^{n,k_n}_{f_{p,\mu}(r)+\epsilon}) = 1 \text{ and} \tag{24}$$

[15] This result is revisited and proved (including additional properties) in Lemma 8, Section 6.2.

$$\lim_{n\to\infty} \mu^n(\mathcal{A}^{n,k_n}_{f_{p,\mu}(r)-\epsilon}) = 0. \tag{25}$$

Then assuming that $k_n/n \longrightarrow r_o$, if $r > r_o$ then $f_{p,\mu}(r) < d$ and from (24) we obtain that $\lim_{n\to\infty} \mu^n(\mathcal{A}^{n,k_n}_d) = 1$. On the other hand, if $r < r_o$ then $f_{p,\mu}(r) > d$ and from (25) we obtain that $\lim_{n\to\infty} \mu^n(\mathcal{A}^{n,k_n}_d) = 0$. This proves (12).

Remark 1. *Adopting the definition in (9) and setting $\epsilon > 0$, it follows from (24) and (25) that for any arbitrary small $\delta > 0$, $r_o - \delta \leq r_p(d,\epsilon,\mu) \leq r_o + \delta$, and, consequently, $r_p(d,\epsilon,\mu) = r_o = f^{-1}_{p,\mu}(d)$. Furthermore, adopting the definition of $\tilde{\kappa}_p(d,\epsilon,\mu^n)$ in (5) with a fixed $\epsilon > 0$, and its asymptotic limits (with n) in (6) and (7), it follows from (24) and (25) that for any arbitrary small $\delta > 0$, $r_o - \delta \leq \tilde{r}^-_p(d,\epsilon,\mu) \leq \tilde{r}^+_p(d,\epsilon,\mu) \leq r_o + \delta$, and, consequently, $\tilde{r}^-_p(d,\epsilon,\mu) = \tilde{r}^+_p(d,\epsilon,\mu) = f^{-1}_{p,\mu}(d)$.*

Concerning the second part of the result, let us assume $r_o \in (0,1)$ such that $f_{p,\mu}(r_o) = 0$. From the convergence in (11) assumed in this result, we have that if $(k_n)_{n\geq 1}$ is such that $k_n/n \longrightarrow r_o$ then

$$\lim_{n\to\infty} \tilde{\sigma}_p(k_n, X^n) = 0, \mu - a.s. \tag{26}$$

Then adopting $\mathcal{A}^{n,k}_d$ in (3), it follows from (26) that for any $d > 0$

$$\lim_{n\to\infty} \mu^n(\mathcal{A}^{n,k_n}_d) = 1, \tag{27}$$

which proves (13).

Remark 2. *Using the definition of $\tilde{\kappa}_p(d,\epsilon,\mu^n)$ in (5), from (27) it is clear that for any $\epsilon > 0$, $\tilde{\kappa}_p(d,\epsilon,\mu^n) \leq k_n$ eventually (in n). From (6), this last inequality implies that $\tilde{r}^+_p(d,\epsilon,\mu) \leq r_o$.*

□

6.2. Theorem 1

Proof: First, we introduce some preliminary results, definitions, and properties that will be essential to elaborate the main argument.

6.2.1. Preliminaries

First, for the case of AMS and ergodic sources (see Lemmas 2 and 4), the ergodic theorem [11, Th. 7.5] tells us that for any ℓ_1-integrable function with respect to $\bar{\mu}_1$, $f : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, the sampling mean (computed with a realization of $\mathbf{X}$) converges with probability one (with respect to μ) to the expectation of f with respect to $\bar{\mu}_1$, i.e.,

$$\lim_{n \longrightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) = \mathbb{E}_{X \sim \bar{\mu}_1}(f(X)) < \infty, \mu - a.s. \tag{28}$$

Therefore, we have that for any $B \in \mathcal{B}(\mathbb{R})$,

$$\lim_{n \longrightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_B(X_i) = \bar{\mu}_1(B), \mu - a.s. \tag{29}$$

In addition, if $(x^p)_{x \in \mathbb{R}} \in L_1(\bar{\mu}_1)$ then for any $B \in \mathcal{B}(\mathbb{R})$,

$$\lim_{n \longrightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_B(X_i) \cdot |X_i|^p = \int_B |x|^p \, d\bar{\mu}_1(x) = ||(x^p)||_{L_1(\bar{\mu}_1)}, \mu - a.s. \tag{30}$$

and, consequently,

$$\lim_{n \longrightarrow \infty} \frac{\sum_{i=0}^{n-1} \mathbf{1}_B(X_i) \cdot |X_i|^p}{\sum_{i=0}^{n-1} |X_i|^p} = v_p(B) = \frac{\int_B |x|^p \, d\bar{\mu}_1(x)}{||(x^p)||_{L_1(\bar{\mu}_1)}}, \mu - a.s. \tag{31}$$

Let us define the tail distribution function of $m \in \mathcal{P}(\mathbb{R})$ by $\phi_m(\tau) \equiv m(B_\tau)$ for all $\tau \in [0, \infty)$. It is simple to verify that:

Proposition 2. *For any $m \in \mathcal{P}(\mathbb{R})$*

i) if $\tau_1 > \tau_2$ then $\phi_m(\tau_1) \leq \phi_m(\tau_2)$ and $\phi_m(\tau_1) = \phi_m(\tau_2)$ if, and only if, $m([\tau_2, \tau_1) \cup (-\tau_1, \tau_2]) = 0$.

ii) $\phi_m(0) = 1$ and $\lim_{\tau \longrightarrow \infty} \phi_m(\tau) = 0$.

iii) $(\phi_m(\tau))_{\tau \geq 0}$ is left continuous and $\phi_m^+(\tau) \equiv \lim_{t_n \longrightarrow \tau, t_n > \tau} \phi_m(t_n) = \phi_m(\tau) - m(\{\tau\} \cup \{-\tau\})$.

Therefore, $(\phi_m(\tau))$ is a continuous function except on the points where m has atomic mass (see Fig. 1). From a well-known result on real analysis [16], using the fact that $(\phi_m(\tau))$ is non-decreasing then this function has at most a countable number of discontinuities.

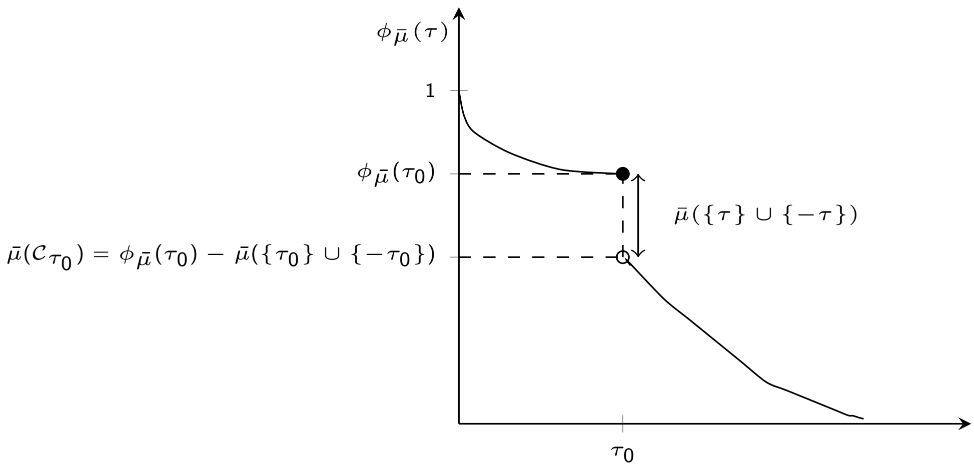

Figure 1: Illustration of the tail distribution function $\phi_{\bar{\mu}}(\tau)$ of $\bar{\mu} \in \mathcal{P}(\mathbb{R})$ with a single discontinuous point at $\tau_o > 0$ in $[0, \infty)$.

This means that m has at most a countable number of non-zero probability events on the collection $\{\{\tau\} \cup \{-\tau\}, \tau \in [0, \infty)\} \subset \mathcal{B}(\mathbb{R})$ that we index and denote by $\mathcal{Y}_m \subset [0, \infty)$.

By definition of v_p, the discontinuity points of $(\phi_{\bar{\mu}_1}(\tau))$ and $(\phi_{v_p}(\tau))$ agree[16] from the fact that if $\tau > 0$ then $\bar{\mu}_1(\{\tau\} \cup \{-\tau\}) = 0 \Leftrightarrow v_p(\{\tau\} \cup \{-\tau\}) = 0$. Therefore, we have that $\mathcal{Y}_{v_p} = \mathcal{Y}_{\bar{\mu}_1} \setminus \{0\}$. For the rest of the proof, it will be relevant to consider the range of these tail functions. These can be characterized as follows (see Figure 1):

$$\mathcal{R}^*_{\bar{\mu}_1} \equiv \{\phi_{\bar{\mu}_1}(\tau), \tau \geq 0\} = (0, 1] \setminus \bigcup_{\tau_n \in \mathcal{Y}_{\bar{\mu}_1}} [\bar{\mu}_1(C_{\tau_n}), \bar{\mu}_1(B_{\tau_n})), \tag{32}$$

$$\mathcal{R}^*_{v_p} \equiv \left\{\phi_{v_p}(\tau), \tau \geq 0\right\} = (0, 1] \setminus \bigcup_{\tau_n \in \mathcal{Y}_{\bar{\mu}_1} \setminus \{0\}} [v_p(C_{\tau_n}), v_p(B_{\tau_n})), \tag{33}$$

where $\mathcal{Y}_{\bar{\mu}_1}$ is either the empty set, or a finite, or a countable set.

With the tail functions $(\phi_{\bar{\mu}_1}(\tau))$ and $(\phi_{v_p}(\tau))$ we can introduce the collection

$$\left\{(\phi_{\bar{\mu}_1}(\tau), \sqrt[p]{1 - \phi_{v_p}(\tau)}), \tau \in [0, \infty)\right\} \tag{34}$$

that the first coordinate covers the range $\mathcal{R}^*_{\bar{\mu}_1}$. For the non-continuous case, i.e., $|\mathcal{Y}_{\bar{\mu}_1}| > 0$, we can complete the range on the first coordinate to cover the non-achievable values

[16]There is only one possible exception when $\tau = 0$.

$\bigcup_{\tau_n \in \mathcal{Y}_{\bar{\mu}_1}} [\bar{\mu}_1(C_{\tau_n}), \bar{\mu}_1(B_{\tau_n}))$ in (32) (Fig. 1 illustrates this range when $\mathcal{Y}_{\bar{\mu}_1} = \{\tau_0\}$) by the following simple extension:

$$\mathcal{F}_{\bar{\mu}_1} \equiv \left\{ (\phi_{\bar{\mu}_1}(\tau), \sqrt[p]{1-\phi_{v_p}(\tau)}), \tau \in [0,\infty) \right\} \\ \bigcup_{\tau_n \in \mathcal{Y}_{\bar{\mu}_1}} \left\{ (\bar{\mu}_1(C_{\tau_n}) + \alpha\bar{\mu}_1(\{-\tau_n, \tau_n\}), \sqrt[p]{1 - v_p(C_{\tau_n}) - \alpha v_p(\{-\tau_n,\tau_n\})}), \alpha \in [0,1) \right\}. \tag{35}$$

The collection of pairs in $\mathcal{F}_{\bar{\mu}_1}$ defines a function from $(0,1]$ to $[0,1)$. In fact for any $r \in (0,1]$ we have that either: $r \in \mathcal{R}^*_{\bar{\mu}_1}$ in (32) for which there is a unique $\tau^* \geq 0$ such that $r = \phi_{\bar{\mu}_1}(\tau^*)$ and, consequently, there is only one $d = \sqrt[p]{1-\phi_{v_p}(\tau^*)} \in [0,1)$ such that $(r,d) \in \mathcal{F}_{\bar{\mu}_1}$, or $r \in \bigcup_{\tau_n \in \mathcal{Y}_{\bar{\mu}_1}} [\bar{\mu}_1(C_{\tau_n}), \bar{\mu}_1(B_{\tau_n}))$, for which there is a unique pair $(\tau^*, \alpha^*) \in \mathcal{Y}_{\bar{\mu}_1} \times [0,1)$ such that $r = \bar{\mu}_1(C_{\tau^*}) + \alpha^* \bar{\mu}_1(\{-\tau^*\} \cup \{\tau^*\})$ and, consequently, a unique $d = \sqrt[p]{1 - \phi_{v_p}(\tau^*) - \alpha^* v_p(\{-\tau^*\} \cup \{\tau^*\})}$ such that again $(r,d) \in \mathcal{F}_{\bar{\mu}_1}$. This means that the set $\mathcal{F}_{\bar{\mu}_1}$ induces a function that we denote by $(f_{\bar{\mu}_1}(r))_{r\in(0,1]}$. In addition, from the properties of $(\phi_{\bar{\mu}_1}(\tau))$ and $(\phi_{v_p}(\tau))$ the following can be shown:

Lemma 8. *The function induced by the set $\mathcal{F}_{\bar{\mu}_1}$ in (35) has the following properties:*

i) $(f_{\bar{\mu}_1}(r))$ *is continuous in* $(0,1]$.

ii) $(f_{\bar{\mu}_1}(r))$ *is strictly decreasing in the domain* $f^{-1}_{\bar{\mu}_1}((0,1)) \subset (0,1]$.[17] *More precisely, if* $0 < r_1 < r_2 \leq 1$ *then either* $f_{\bar{\mu}_1}(r_2) < f_{\bar{\mu}_1}(r_1)$, *or* $f_{\bar{\mu}_1}(r_1) = f_{\bar{\mu}_1}(r_2) = 0$. *Furthermore,* $(f_{\bar{\mu}_1}(r))$ *is strictly monotonic in* $(0,1]$ *if, and only if,* $0 \notin \mathcal{Y}_{\bar{\mu}_1}$.

iii) The range of $(f_{\bar{\mu}_1}(r))_{r\in(0,1]}$ *is* $[0,1)$.

The proof of this result is presented in Appendix A.

We are in a position to prove the main result:

6.2.2. Main Argument — Case $(x^p)_{x\in\mathbb{R}} \in L_1(\bar{\mu}_1)$

Let us assume that $(x^p)_{x\in\mathbb{R}} \in L_1(\bar{\mu}_1)$. Let us consider an arbitrary $r \in [1,0)$ and a sequence $(k_n)_{n\geq 1}$ such $k_n/n \longrightarrow r$ as n tends to infinity.

[17] $f^{-1}_{\bar{\mu}_1}((0,1)) = (0,1)$ if, and only if, $0 \notin \mathcal{Y}_{\bar{\mu}_1}$.

2.1) *Continuous Scenario:* Let us first consider the case where $r \in int(\mathcal{R}^*_{\bar{\mu}_1})$, i.e., $r \in \mathcal{R}^*_{\bar{\mu}_1} \setminus \{\bar{\mu}_1(B_{\tau_n}), \tau_n \in \mathcal{Y}_{\bar{\mu}_1}\}$, and, consequently, there is τ_o such that $r = \phi_{\bar{\mu}_1}(\tau_o)$ being τ_o a continuous point of the tail function $\phi_{\bar{\mu}_1}(\cdot)$ (see iii) in Proposition 2).

Let us define $n_\tau(x^n) \equiv \sum_{i=1}^n \mathbf{1}_{B_\tau}(x_i)$, then using the (point-wise) ergodic result in (30), it follows that for all $\tau \geq 0$

$$\lim_{n\to\infty} \frac{n_\tau(X^n)}{n} = \phi_{\bar{\mu}_1}(\tau), \mu - a.s., \tag{36}$$

and from (31) and (2)

$$\lim_{n\to\infty} \tilde{\sigma}_p(n_\tau(X^n), X^n) = \sqrt[p]{1 - \phi_{v_p}(\tau)}, \mu - a.s. \tag{37}$$

In other words, we have the following family of (typical) sets:

$$\mathcal{A}^\tau \equiv \left\{ (x_n)_{n\geq 0}, \lim_{n\to\infty} \frac{n_\tau(x^n)}{n} = \phi_{\bar{\mu}_1}(\tau) \right\} \tag{38}$$

$$\mathcal{B}^\tau \equiv \left\{ (x_n)_{n\geq 0}, \lim_{n\to\infty} \tilde{\sigma}_p(n_\tau(x^n), x^n) = \sqrt[p]{1 - \phi_{v_p}(\tau)} \right\}, \tag{39}$$

satisfying $\mu(\mathcal{A}^\tau \cap \mathcal{B}^\tau) = 1$ for all $\tau \geq 0$.

Using the fact that $\phi_{\bar{\mu}_1}(\cdot)$ is continuous at τ_o and the observation that $(\phi_{\bar{\mu}_1}(\cdot))$ has at most a countable number of discontinuities, there is $\delta \in (0, r)$ where the interval $(r - \delta, r + \delta)$ defines an open domain containing τ_o, given by $(\tau_1, \tau_2) = \phi^{-1}_{\bar{\mu}_1}((r - \delta, r + \delta))$ where the function $(\phi_{\bar{\mu}_1}(\cdot))$ is continuous (see Figure 2). Associated with this domain, we can consider $\left\{\phi_{v_p}(\tau), \tau \in (\tau_1, \tau_2)\right\} = (v_2, v_1)$ where by monotonicity $v_1 = \phi_{v_p}(\tau_1)$ and $v_2 = \phi_{v_p}(\tau_2)$ (see Figure 2). It is simple to show (by the construction of v_p from $\bar{\mu}_1$)[18] that for any $\tau > 0$ and $\epsilon > 0$

$$\phi_{\bar{\mu}_1}(\tau + \epsilon) < \phi_{\bar{\mu}_1}(\tau) \text{ if, and only if, } \phi_{v_p}(\tau + \epsilon) < \phi_{v_p}(\tau). \tag{40}$$

Therefore, this mutual absolutely continuity property between $\bar{\mu}_1$ and v_p implies that

$$\phi_{\bar{\mu}_1}(\tau_1) > \phi_{\bar{\mu}_1}(\tau_o) = r \Leftrightarrow \phi_{v_p}(\tau_1) = v_1 > \phi_{v_p}(\tau_o), \text{ and} \tag{41}$$

$$\phi_{\bar{\mu}_1}(\tau_2) < \phi_{\bar{\mu}_1}(\tau_o) = r \Leftrightarrow \phi_{v_p}(\tau_2) = v_2 < \phi_{v_p}(\tau_o). \tag{42}$$

[18]Note that for any $B \in \mathcal{B}(\mathbb{R})$ where $0 \notin B$, $\bar{\mu}_1(B) = 0$ if, and only if, $v_p(B) = 0$.

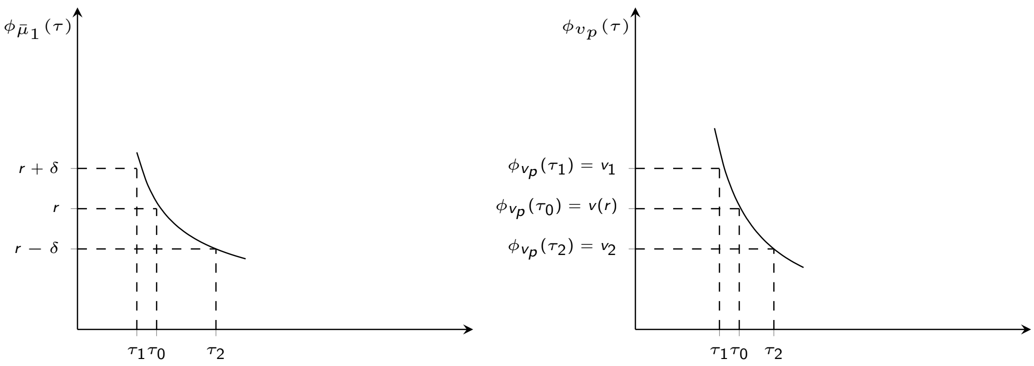

Figure 2: Illustration of the tail distribution functions of $\bar{\mu}_1$ and v_p at a continuous point τ_0, where $r = \phi_{\bar{\mu}_1}(\tau_0)$.

We can then find $M > 0$ sufficiently large such that for all $m \geq M$, $\phi_{v_p}(\tau_o) + 1/m < v_1$. For any of these $m \geq M$, there is $\tau_m \in (\tau_1, \tau_2)$ (from the continuity of $\phi_{v_p}(\cdot)$ in (τ_1, τ_2)) such that $\phi_{v_p}(\tau_m) = \phi_{v_p}(\tau_o) + 1/m$, where again by (40) $\phi_{\bar{\mu}_1}(\tau_m) > \phi_{\bar{\mu}_1}(\tau_o) = r$. Therefore, for any $m \geq M$ and $\forall (x_n)_{n\geq 0} \in \mathcal{A}^{\tau_m} \cap \mathcal{B}^{\tau_m}$, $n_{\tau_m}(x^n) > k_n$ eventually in n (as n tends to infinity). This comes from the assumption that $k_n/n \longrightarrow r < \phi_{\bar{\mu}_1}(\tau_m)$ and the definition of $\mathcal{A}^{\tau_m}$ given in (38). Consequently, under this context, it follows that

$$\tilde{\sigma}_p(n_{\tau_m}(x^n), x^n) \leq \tilde{\sigma}_p(k_n, x^n), \tag{43}$$

eventually in n. Finally, using explicitly that $(x_n)_{n\geq 1} \in \mathcal{B}^{\tau_m}$ (see Eq.(39)), we have that

$$\sqrt[p]{1 - (\phi_{v_p}(\tau_o) + 1/m)} \leq \lim \inf_{n\to\infty} \tilde{\sigma}_p(k_n, x^n). \tag{44}$$

Repeating this argument, if $(x_n)_{n\geq 1} \in \bigcap_{m\geq M}(\mathcal{A}^{\tau_m} \cap \mathcal{B}^{\tau_m})$ it follows from (44) that[19]

$$\sqrt[p]{1 - \phi_{v_p}(\tau_o)} \leq \lim \inf_{n\to\infty} \tilde{\sigma}_p(k_n, x^n). \tag{45}$$

By the sigma additivity [12] and the fact that from the ergodic theorem $\mu(\mathcal{A}^{\tau_m} \cap \mathcal{B}^{\tau_m}) = 1$ for any $m \geq 1$, it follows that

$$\sqrt[p]{1 - \phi_{v_p}(\tau_o)} \leq \lim \inf_{n\to\infty} \tilde{\sigma}_p(k_n, X^n), \mu - a.s. \tag{46}$$

[19]This is obtained by taking the supremum ($m \geq M$) in the LHS of (44) and using the continuity of the function $\sqrt[p]{1-x}$ in $x \in (0, 1)$.

The exact argument can be used to prove that

$$\limsup_{n\to\infty} \tilde{\sigma}_p(k_n, X^n) \leq \sqrt[p]{1-\phi_{v_p}(\tau_o)}, \mu - a.s.. \quad (47)$$

by using the sequences $\tilde{\tau}_m$ such that $\phi_{v_p}(\tilde{\tau}_m) = \phi_{v_p}(\tau_o) - 1/m$ for $m \geq \tilde{M}$ and $\tilde{M}$ sufficiently large. This is omitted for the sake of space. Finally, (46) and (47) prove the result in the continuous case.[20]

2.2) <u>*Discontinuous scenario:*</u> Let us consider the case where $r \notin \mathcal{R}^*_{\bar{\mu}_1}$ (see Eq.(32)), which means that $\exists \tau_i \in \mathcal{Y}_{\bar{\mu}_1}$ such that

$$r \in [\bar{\mu}_1(C_{\tau_i}), \bar{\mu}_1(B_{\tau_i})), \quad (48)$$

(see the illustration in Fig. 3). For the moment let us assume that $r \in (\bar{\mu}_1(C_{\tau_i}), \bar{\mu}_1(B_{\tau_i}))$,[21] then there is a unique $\alpha_o \in (0,1)$ such that

$$r = \bar{\mu}_1(C_{\tau_i}) + \alpha_o \cdot \bar{\mu}_1(\{-\tau_i, \tau_i\}). \quad (49)$$

Here we need to use an extended version of the point-wise ergodic result in (28). For that, let us introduce an i.i.d. Bernoulli process $\mathbf{Y} = (Y_i)_{i\geq 0}$ of parameter $\rho \in [0,1]$, where $\mathbb{P}(Y_i = 1) = \rho$ for all $i \geq 0$, that is independent of $\mathbf{X} = (X_n)_{n\geq 0}$. Let us denote by η its (i.i.d) process distribution in $\{0,1\}^{\mathbb{N}}$. Then, from the ergodic result for AMS process in (28) it follows, as a natural extension of (29), that for all $\tau \geq 0$

$$\lim_{n\longrightarrow\infty} \frac{1}{n}\sum_{i=0}^{n-1} \mathbf{1}_{C_\tau}(X_i) + \frac{1}{n}\sum_{i=0}^{n-1} \mathbf{1}_{\{-\tau,\tau\}}(X_i)\cdot Y_i = \bar{\mu}_1(C_\tau) + \bar{\mu}_1(\{-\tau,\tau\})\rho, \quad (50)$$

$$\lim_{n\longrightarrow\infty} \frac{\sum_{i=0}^{n-1}(\mathbf{1}_{C_\tau}(X_i) + \mathbf{1}_{\{-\tau,\tau\}}(X_i)Y_i)\,|X_i|^p}{\sum_{i=0}^{n-1}|X_i|^p} = \frac{\int_{C_\tau}|x|^p\, d\bar{\mu}_1(x) + \bar{\mu}_1(\{-\tau,\tau\})\tau^p\rho}{||(x^p)||_{L_1(\bar{\mu}_1)}}, \quad (51)$$

[20]The proof assumes that $r < 1$. The proof for the asymmetric case when $r = 1 \in int(\mathcal{R}^*_{\bar{\mu}_1})$, i.e., $r = 1$ is a continuous point of $\phi_{\bar{\mu}_1}(\cdot)$ follows from the argument above. On the one hand, $\tilde{\sigma}_p(k_n, x^n) \geq 0$ for any k_n by definition. On the other hand, the argument used to obtain (47) follows without any problem in this context, implying that $\limsup_{n\to\infty}\tilde{\sigma}_p(k_n, X^n) \leq \sqrt[p]{1-\phi_{v_p}(\tau_o)} = 0$ $\mu - a.s.$, considering that $\tau_o = 0$ in this case.

[21]We left the case $r \in \{\bar{\mu}_1(C_{\tau_n}) : \tau_n \in \mathcal{Y}_{\bar{\mu}_1}\} \cup \{\bar{\mu}_1(B_{\tau_n}) : \tau_n \in \mathcal{Y}_{\bar{\mu}_1}\}$ for the mixed scenario below.

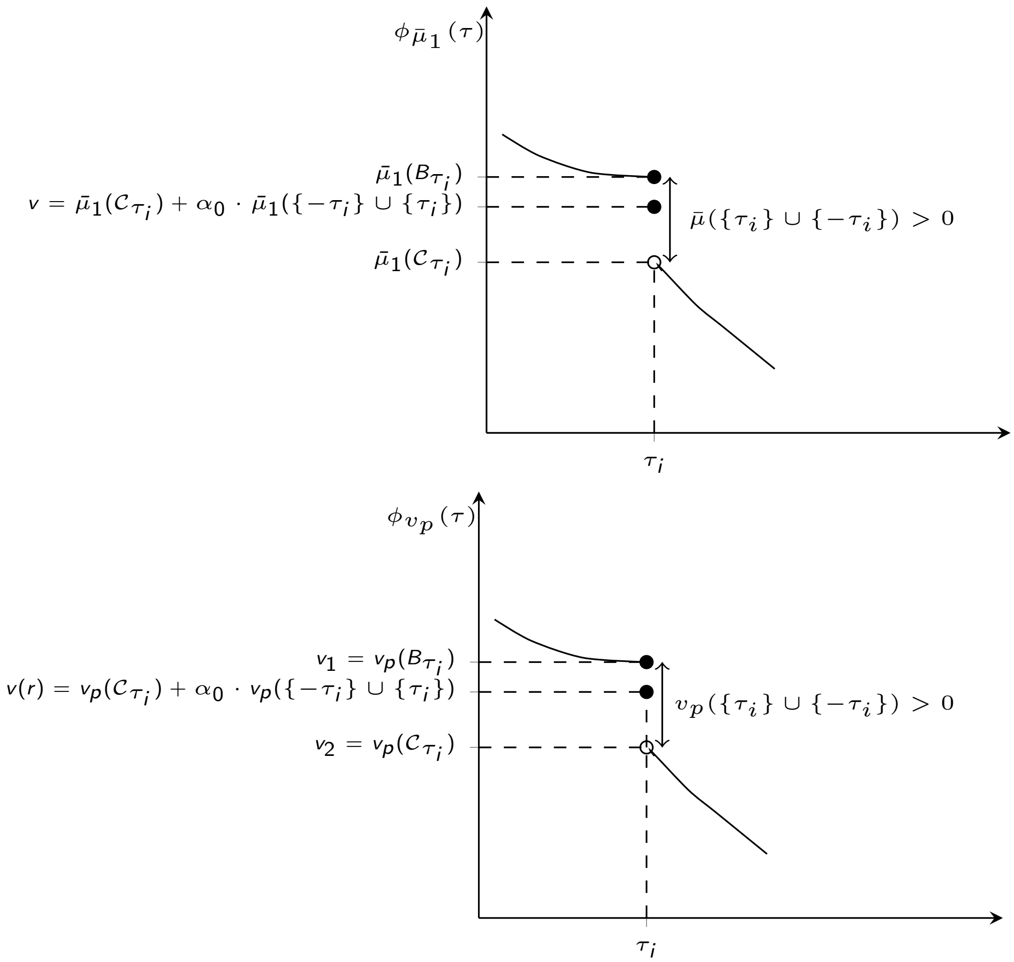

Figure 3: Illustration of the tail distribution functions of $\bar{\mu}_1$ and v_p at a discontinuous point $\tau_i > 0$.

with probability one with respect to joint process distribution of $(\mathbf{X}, \mathbf{Y})$ denoted by $\mu \times \eta$.

Returning to the argument, let us consider an arbitrary $(k_n)_{\geq 1}$ such that $k_n/n \longrightarrow r$ as n goes to infinity. Let us consider $\alpha_m \equiv \alpha_o + 1/m$ and M sufficiently large to make $\alpha_M < 1$. For any $m \geq M$, let us construct an auxiliary i.i.d. Bernoulli process $\mathbf{Y}(\alpha_\mathbf{m}) \equiv (Y_i)_{i\geq 0}$, where $\mathbb{P}(Y_i = 1) = \alpha_m$. The process distribution of $\mathbf{Y}(\alpha_\mathbf{m})$ is denoted by η_m. In this context, if we define the a joint count function $n_\tau(x^n, y^n) \equiv \sum_{i=0}^{n-1} \mathbf{1}_{C_\tau}(x_i) +$

$\sum_{i=0}^{n-1} \mathbf{1}_{\{\tau\}\cup\{-\tau\}}(x_i) \cdot y_i$ it follows from (50) and (51) that for τ_i introduced in (48),

$$\lim_{n\longrightarrow\infty} \frac{n_{\tau_i}(X^n, Y^n)}{n} = \bar{\mu}_1(C_{\tau_i}) + (\alpha_o + 1/m) \cdot \bar{\mu}_1(\{-\tau_i, \tau_i\}), \tag{52}$$

$$\lim_{n\longrightarrow\infty} \tilde{\sigma}_p(n_{\tau_i}(X^n, Y^n), X^n) = \sqrt[p]{1 - (v_p(C_{\tau_i}) - (\alpha_o + 1/m) \cdot v_p(\{-\tau_i, \tau_i\}))} \tag{53}$$

$\mu \times \eta_m$ -almost surely. Importantly in (53), $v_p(\{-\tau_i, \tau_i\}) > 0$ from the fact that $\bar{\mu}_1(\{-\tau_i, \tau_i\}) > 0$.[22] Let us consider an arbitrary (typical) sequence $((x_n)_{n\geq 1}, (y_n)_{n\geq 1})$ satisfying the limiting conditions in (52) and (53). From (52), it follows that $n_\tau(x^n, y^n) > k_n$ eventually in n as $k_n/n \longrightarrow r = \bar{\mu}_1(C_{\tau_i}) + \alpha_o \cdot \bar{\mu}_1(\{-\tau_i, \tau_i\}) < \bar{\mu}_1(C_{\tau_i}) + (\alpha_o + 1/m) \cdot \bar{\mu}_1(\{-\tau_i, \tau_i\})$ by construction. Therefore,

$$\tilde{\sigma}_p(n_\tau(x^n, y^n), x^n) \leq \tilde{\sigma}_p(k_n, x^n), \text{ eventually in } n. \tag{54}$$

But the left hand side of (54) converges to $\sqrt[p]{1 - (v_p(C_{\tau_i}) - (\alpha_o + 1/m) \cdot v_p(\{-\tau_i, \tau_i\}))}$ as n tends to infinity by the construction of $((x_n)_{n\geq 1}, (y_n)_{n\geq 1})$. Finally, by the almost sure convergence in (52) and (53), it follows that

$$\lim \inf_{n\longrightarrow\infty} \tilde{\sigma}_p(k_n, X^n) \geq \sqrt[p]{1 - (v_p(C_{\tau_i}) - (\alpha_o + 1/m) \cdot v_p(\{-\tau_i, \tau_i\}))}, \tag{55}$$

μ- almost surely.[23]

Let us denote by $\mathcal{D}^{\tau_m} \equiv \{(x_n)_{n\geq 0}$, where (55) holds$\}$. From (55), $\mu(\mathcal{D}^{\tau_m}) = 1$ and by sigma-additivity [12] it follows that $\mu(\cap_{m\geq M}\mathcal{D}^{\tau_m}) = 1$, which implies that

$$\lim \inf_{n\longrightarrow\infty} \tilde{\sigma}_p(k_n, X^n) \geq \sqrt[p]{1 - (v_p(C_{\tau_i}) - \alpha_o \cdot v_p(\{-\tau_i, \tau_i\}))}, \mu - a.s. \tag{56}$$

To conclude, an equivalent (symmetric) argument can be used to prove that

$$\lim \sup_{n\longrightarrow\infty} \tilde{\sigma}_p(k_n, X^n) \leq \sqrt[p]{1 - (v_p(C_{\tau_i}) - \alpha_o \cdot v_p(\{-\tau_i, \tau_i\}))}, \mu - a.s., \tag{57}$$

using $\tilde{\alpha}_m \equiv \alpha_o - 1/m$ and $\tilde{M}$ sufficiently large to make $\tilde{\alpha}_{\tilde{M}} > 0$. For sake of space the proof is omitted. This concludes the result in this case.

2.3) *Mixed scenario:* Here we consider the scenario where $r \in \{\bar{\mu}_1(B_{\tau_n}), \bar{\mu}_1(C_{\tau_n}) : \tau_n \in \mathcal{Y}_{\bar{\mu}_1}\}$. The proof reduces to the same procedure presented above in the continuous and discontinues scenarios, but adopted in a mixed form. A sketch with the steps will be provided as no new technical elements are introduced here.

[22] Here we assume that $\tau_i > 0$. The important sparse case when $\tau_i = 0$ will be treated below.

[23] We remove the dependency on η_m, as both terms in (55) (in the limit) turn out to be independent of the auxiliary process $(Y_i)_{i\geq 0}$.

For $r = \bar{\mu}_1(B_{\tau_i})$ for $\tau_i \in \mathcal{Y}_{\bar{\mu}_1}$ and $\tau_i \neq 0$, the same argument adopted in the continuos case (to obtain (46)) can be adopted here to obtain that

$$\liminf_{n\to\infty} \tilde{\sigma}_p(k_n, X^n) \geq \sqrt[p]{1 - \phi_{v_p}(\tau_i)}, \mu - a.s., \tag{58}$$

for any sequence $(k_n)_{n\geq 1}$ such that $k_n/n \longrightarrow \bar{\mu}_1(B_{\tau_i})$. For the other inequality, the strategy with the auxiliary Bernoulli process presented in the proof of the discontinuous case can be adopted considering $\alpha_o = 1$ and $\tilde{\alpha}_m = 1 - 1/m$ for m sufficiently large. Then, a result equivalent to (57) is obtained, meaning in this specific context that

$$\limsup_{n\to\infty} \tilde{\sigma}_p(k_n, X^n) \leq \sqrt[p]{1 - \phi_{v_p}(\tau_i)}, \mu - a.s. \tag{59}$$

For $r = \bar{\mu}_1(C_{\tau_i})$ for $\tau_i \in \mathcal{Y}_{\bar{\mu}_1}$ and $\tau_i \neq 0$, the same argument with the auxiliary Bernoulli process used to obtain (56) can be adopted here, considering $\alpha_o = 0$ and $\alpha_m = 1/m$ for m sufficiently large, to obtain that

$$\liminf_{n\longrightarrow\infty} \tilde{\sigma}_p(k_n, X^n) \geq \sqrt[p]{1 - v_p(C_{\tau_i})}, \mu - a.s., \tag{60}$$

for any sequence $(k_n)_{n\geq 1}$ such that $k_n/n \longrightarrow \bar{\mu}_1(C_{\tau_i})$. For the other inequality, the argument of the continuous case proposed to obtain (47) can be adopted here (with no differences) to obtain that

$$\limsup_{n\longrightarrow\infty} \tilde{\sigma}_p(k_n, X^n) \leq \sqrt[p]{1 - v_p(C_{\tau_i})}, \mu - a.s. \tag{61}$$

2.4) *Sparse scenario:* The sparse scenario, meaning that $0 \in \mathcal{Y}_{\bar{\mu}_1}$, deserves a special treatment because this analysis offers insights about an important property of the function $(f_{p,\bar{\mu}}(r))_{r\in(0,1]}$. Let us consider the case that $\bar{\mu}_1(\{0\}) = \rho_o > 0$, then $\phi^+_{\bar{\mu}_1}(0) = \lim_{\tau\to 0} \phi_{\bar{\mu}_1}(\tau) = \bar{\mu}_1(C_0) = 1 - \rho_o \in (0, 1)$ (see, the illustration in Fig. 4). On the other hand, we have that $v_p(\{0\}) = \frac{0^p \cdot \bar{\mu}_1(\{0\})}{||(x^p)||_{L_1(\bar{\mu}_1)}} = 0$. Therefore, $(\phi_{v_p}(\tau))$ is continuous at $\tau = 0$ (Fig. 4). From the fact that $\mathcal{Y}_{\bar{\mu}_1}$ is at most a countable set, there is $\tau_1 > 0$ with $\phi_{\bar{\mu}_1}(\tau_1) < 1 - \rho_o$ where $\phi_{\bar{\mu}_1}(\cdot)$ is continuous in $(0, \tau_1)$ and, consequently, so is $\phi_{v_p}(\cdot)$ in $(0, \tau_1)$ (from Proposition 3 in Appendix 8). If we consider the range of $\phi_{v_p}(\cdot)$ in this continuous domain, we have that $\left\{\phi_{v_p}(\tau), \tau \in (0, \tau_1)\right\} = (v_1, 1)$ where $v_1 = \phi_{v_p}(\tau_1) < 1$.

Here, we adopt the same argument used in the continuous scenario to obtain the upper bound in (47). Let us consider an arbitrary sequence $(k_n)_{n\geq 1}$ such that $k_n/n \longrightarrow 1 - \rho_o$

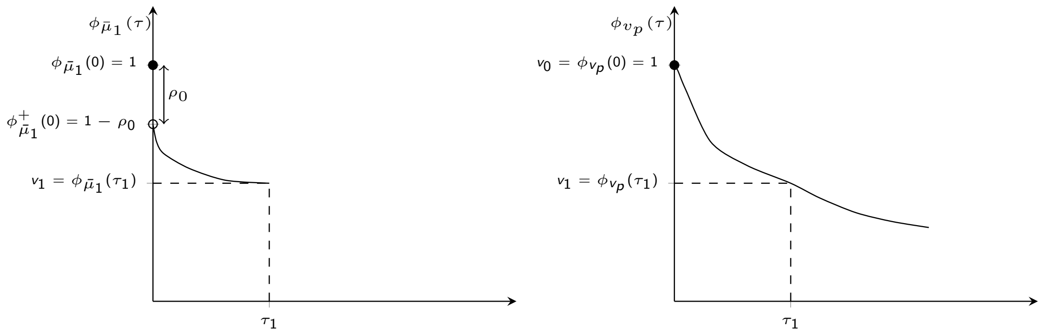

Figure 4: Illustration of the tail distribution functions of $\bar{\mu}_1$ and v_p in the sparse case where $\bar{\mu}_1(\{0\}) = \rho_0 > 0$.

with n. By the continuity of $\phi_{v_p}(\tau)$ in $(0, \tau_1)$ for any m sufficiently large such that $1 - \frac{1}{m} < v_1$, there is $\tau_m > 0$ such that $\phi_{v_p}(\tau_m) = 1 - \frac{1}{m}$. For any of these τ_m, it follows that $\phi_{\bar{\mu}_1}(\tau_m) < 1 - \rho_o$.[24] Then, we can consider the set of typical sequences defined in (38) and (39), where if $(x_n)_{n \geq 1} \in \mathcal{A}^{\tau_m} \cap \mathcal{B}^{\tau_m}$ then eventually in n it follows that $k_n > n_{\tau_m}(x^n)$ (from the fact that $\phi_{\bar{\mu}_1}(\tau_m) < 1 - \rho_o$) and, consequently,

$$\limsup_{n \longrightarrow \infty} \tilde{\sigma}_p(k_n, x^n) \leq \sqrt[p]{1/m}, \tag{62}$$

this last result from the definition of $\mathcal{B}^{\tau_m}$ and the construction of τ_m (i.e., $\phi_{v_p}(\tau_m) = 1 - \frac{1}{m}$). Then if $(x_n)_{n \geq 1} \in \bigcap_{m \geq M} (\mathcal{A}^{\tau_m} \cap \mathcal{B}^{\tau_m})$, where $M > 0$ is set such that $1 - \frac{1}{M} < v_1$, then

$$\limsup_{n \longrightarrow \infty} \tilde{\sigma}_p(k_n, x^n) \leq 0. \tag{63}$$

Finally, from the (point-wise) ergodic result for AMS sources in (28), it follows that $\mu(\bigcap_{m \geq M} \mathcal{A}^{\tau_m} \cap \mathcal{B}^{\tau_m}) = 1$, meaning from (63) that $\lim_{n \longrightarrow \infty} \tilde{\sigma}_p(k_n, X^n) = 0$, μ-almost surely.

The last observation to conclude this part is that if $(\tilde{k}_n)$ dominates (k_n), in the sense that $\tilde{k}_n \geq k_n$ eventually, then from definition $\tilde{\sigma}_p(\tilde{k}_n, x^n) \leq \tilde{\sigma}_p(k_n, x^n)$ for all x^n. Therefore from (63), for any $r \in [1 - \rho_o, 1]$ and for any (k_n) such that $k_n/n \longrightarrow r$, it

[24] This from the fact that if $\phi_{v_p}(\tau) < \phi_{v_p}(\tilde{\tau})$ then $\phi_{\bar{\mu}_1}(\tau) < \phi_{\bar{\mu}_1}(\tilde{\tau})$ from the definition of v_p.

follows that

$$\lim_{n\longrightarrow\infty} \tilde{\sigma}_p(k_n, X^n) = 0, \ \mu - a.s. \tag{64}$$

Then we obtain in this case that $\forall r \in [1 - \bar{\mu}_1(\{0\}), 1]$

$$f_{p,\bar{\mu}}(r) = 0, \tag{65}$$

while $f_{p,\bar{\mu}}(r) > 0$ if $r \in (0, 1 - \bar{\mu}_1(\{0\}))$.

Remark 3. *The result in (64) is consistent with the statement of Theorem 1, because if $r \in [1-\rho_o, 1]$ then it can be written as $r = \bar{\mu}_1((0,\infty)) + \alpha \cdot \bar{\mu}_1(\{0\})$ for some $\alpha \in [0,1]$ where $f_{p,\bar{\mu}}(r) = \sqrt[p]{1 - v_p((0,\infty)) - \alpha \cdot v_p(\{0\})} = 0$.*

6.2.3. Main Argument — Case $(x^p)_{x\in\mathbb{R}} \notin L_1(\bar{\mu}_1)$

When $(x^p)_{x\in\mathbb{R}} \notin L_1(\bar{\mu}_1)$, it follows that $\forall \tau \geq 0$,

$$\lim_{n\longrightarrow\infty} \frac{\sum_{i=0}^{n-1}(1 - \mathbf{1}_{B_\tau}(X_i)) \cdot |X_i|^p}{\sum_{i=0}^{n-1} |X_i|^p} = 0, \mu - a.s.. \tag{66}$$

this from the (point-wise) ergodic result in (28) and the fact that $\int_{\mathbb{R}} |x|^p \, d\bar{\mu}_1(x) = \infty$. Then from (29) and (66) it follows in this case that

$$\lim_{n\to\infty} \frac{n_\tau(X^n)}{n} = \phi_{\bar{\mu}_1}(\tau), \mu - a.s., \tag{67}$$

$$\lim_{n\to\infty} \tilde{\sigma}_p(n_\tau(X^n), X^n) = 0, \mu - a.s. \tag{68}$$

for all $\tau \geq 0$. Again we can consider $\mathcal{A}^\tau = \left\{(x_n)_{n\geq 0}, \lim_{n\to\infty} \frac{n_\tau(x^n)}{n} = \phi_{\bar{\mu}_1}(\tau)\right\}$ and

$$\mathcal{B}^\tau = \left\{(x_n)_{n\geq 0}, \lim_{n\to\infty} \tilde{\sigma}_p(n_\tau(x^n), x^n) = 0\right\},$$

where $\mu(\mathcal{A}^\tau \cap \mathcal{B}^\tau) = 1$ for all τ.

Let us fix $r \in (0,1]$ and $(k_n)_{n\geq 1}$ such that $k_n/n \longrightarrow r$. We can consider $\bar{r} < r$, and τ_o such that $\phi_{\bar{\mu}_1}(\tau_o) = \bar{r}$. Then for any $(x_n) \in \mathcal{A}^{\tau_o} \cap \mathcal{B}^{\tau_o}$ it follows that $k_n > n_{\tau_o}(x^n)$ eventually in n (from the fact that $r > \bar{r}$ and the definition of $\mathcal{A}^{\tau_o}$), therefore eventually $\tilde{\sigma}_p(n_\tau(x^n), x^n) \geq \tilde{\sigma}_p(k_n, x^n)$. Finally from the definition of $\mathcal{B}^{\tau_o}$, $\lim_{n\to\infty} \tilde{\sigma}_p(k_n, x^n) = 0$. The proof concludes noting that $\mu(\mathcal{A}^{\tau_o} \cap \mathcal{B}^{\tau_o}) = 1$. □

6.3. Theorem 2

Proof: First, we introduce formally the ergodic decomposition (ED) theorem:

Theorem 3. *[11, Th. 10.1] Let $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ be an AMS process characterized by $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}), \mu)$. Then there is a measurable space given by $(\Lambda, \mathcal{L})$ that parametrizes the family of stationary and ergodic distribution, i.e., $\tilde{\mathcal{P}} = \{\mu_\lambda, \lambda \in \Lambda\}$, and a measurable function $\Psi : (\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}})) \to (\Lambda, \mathcal{L})$ such that:*

i) Ψ is invariant with respect to T, i.e., $\Psi(\mathbf{x}) = \Psi(T(\mathbf{x}))$ for all $\mathbf{x} \in \mathbb{R}^{\mathbb{N}}$.

ii) Using the stationary mean $\bar{\mu}$ of $\mathbf{X}$, and its induced probability in $(\Lambda, \mathcal{L})$, denoted by W_Ψ, it follows that $\forall F \in \mathcal{B}(\mathbb{R}^{\mathbb{N}})$

$$\bar{\mu}(F) = \int \mu_\lambda(F) \partial W_\Psi(\lambda). \tag{69}$$

iii) Finally[25], for any $L_1(\bar{\mu})$-integrable and measurable function $f : (\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(T^i(\mathbf{X})) = \mathbb{E}_{\mathbf{Z}\sim\mu_{\Psi(\mathbf{X})}}\left(f(\mathbf{Z})\right), \ \mu - \textit{almost surely}, \tag{70}$$

where $\mathbf{Z}$ in (70) denotes a stationary and ergodic process in $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ with process distribution given by $\Psi(\mathbf{X}) \in \Lambda$.

Let us first prove the almost sure sample-wise convergence in (22). For $r \in (0, 1])$ and $(k_n)_{n\geq 1}$ such that $k_n/n \to r$, we need to study the limit of the following random object $Y_n = \tilde{\sigma}_p(k_n, X_1^n)$. As in the proof of Theorem 1, we consider the tail events

$$B_\tau = (-\infty, \tau] \cup [\tau, \infty) \text{ and } C_\tau = (-\infty, \tau) \cup (\tau, \infty) \tag{71}$$

for $\tau \geq 0$. From Theorem 3 it follows that for any $B \in \mathcal{B}(\mathbb{R})$,

$$\lim_{n\longrightarrow\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{1}_B(X_i) = \mathbb{E}_{\mathbf{Z}\sim\mu_{\Psi(\mathbf{X})}}(\mathbf{1}_B(Z_1)) = \mu_{1,\Psi(\mathbf{X})}(B), \mu - a.s. \tag{72}$$

[25]This result can be interpreted as a more sophisticated re-statement of the point-wise ergodic theorem for AMS sources under the assumption of a standard space, which is the case for $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$. Details and the interpretations of this result are presented in [11, Chs. 8 and 10].

where $\mathbf{Z} = (Z_i)_{i\geq 1}$ and $\mu_{1,\Psi(\mathbf{X})}$ denotes the probability of Z_1 (the 1D marginalization of the process distribution $\mu_{\Psi(\mathbf{X})}$) in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. In addition, from Theorem 3 we have that

$$\lim_{n\longrightarrow\infty} \frac{\sum_{i=0}^{n-1} \mathbf{1}_{B_\tau}(X_i)\cdot |X_i|^p}{\sum_{i=0}^{n-1} |X_i|^p} = \xi_p(\mathbf{X}, B_\tau),\ \mu - a.s., \tag{73}$$

where

$$\xi_p(\mathbf{X}, B_\tau) \equiv \begin{cases} \frac{\int_{B_\tau} |x|^p d\mu_{1,\Psi(\mathbf{X})}(x)}{||(x^p)||_{L_1(\mu_{1,\Psi(\mathbf{X})})}} & \text{if } (x^p)_{x\in\mathbb{R}} \in L_1(\mu_{1,\Psi(\mathbf{X})}) \\ 1 & \text{if } (x^p)_{x\in\mathbb{R}} \notin L_1(\mu_{1,\Psi(\mathbf{X})}) \end{cases}. \tag{74}$$

From the results in (72) and (73), we can proceed with the same arguments used in the proof of Theorem 1 to obtain that[26]

$$\lim_{n\longrightarrow\infty} \tilde{\sigma}_p(k_n, X_1^n) = f_{p,\mu_{\Psi(\mathbf{X})}}(r),\ \mu - \text{almost surely}, \tag{75}$$

where $f_{p,\mu_{\Psi(\mathbf{X})}}(r)$ is the almost-sure asymptotic limit of the stationary and ergodic component $\mu_{\Psi(\mathbf{X})} \in \tilde{\mathcal{P}}$ stated in (19) and elaborated in the statement of Theorem 1. This proves the first part of the result.

For the second part, we consider again $r \in (0, 1]$ and $(k_n)_{n\geq 1}$ such that $k_n/n \to r$. Let us denote the almost sure limit in (75) by $f_p(\mathbf{X}, r)$[27], which is in general a random variable from $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}(\mathbb{R}^{\mathbb{N}}))$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. For an arbitrary $d \in [0, 1)$, we need to analyze the asymtotic limit of $\mu^n(A_d^{n,k_n})$. By additivity, we decompose this probability in two terms:

$$\begin{aligned} \mu^n(A_d^{n,k_n}) =& \mu\left(\left\{\bar{x} = (x_i)_{i\geq 1} \in \mathbb{R}^{\mathbb{N}} : \tilde{\sigma}_p(k_n, x_1^n) \leq d, f_p(\bar{x}, r) \leq d\right\}\right) \\ &+\mu\left(\left\{\bar{x} \in \mathbb{R}^{\mathbb{N}} : \tilde{\sigma}_p(k_n, x_1^n) \leq d, f_p(\bar{x}, r) > d\right\}\right), \end{aligned} \tag{76}$$

For the first term (from left to right) in the RHS of (76), we can consider the following bounds

$$\begin{aligned} \mu\left(\{\bar{x} : f_p(\bar{x}, r) \leq d\}\right) \geq \mu\left(\{\bar{x} : \tilde{\sigma}_p(k_n, x_1^n) \leq d, f_p(\bar{x}, r) \leq d\}\right) \geq \\ \mu\left(\{\bar{x} : \tilde{\sigma}_p(k_n, x_1^n) \leq f_{p,\mu}(\bar{x}, r), f_p(\bar{x}, r) \leq d\}\right). \end{aligned} \tag{77}$$

[26] We omit the argument here as they are redundant, following directly the structure presented in Section 6.2.

[27] We omit the dependency on μ in the notation, because this limit (as a random variable of $\mathbf{X}$) is independent of μ.

The lower and upper bounds in (77) has the same asymptotic limit, i.e.,

$$\lim_{n\longrightarrow\infty} \mu\left(\{\bar{x} : \tilde{\sigma}_p(k_n, x_1^n) \leq f_p(\bar{x}, r), f_p(\bar{x}, r) \leq d\}\right) = \lim_{n\longrightarrow\infty} \mu\left(\{\bar{x} : f_p(\bar{x}, r) \leq d\}\right). \quad (78)$$

This can be shown by the following equality

$$\begin{aligned}\mu\left(\{\bar{x} : f_p(\bar{x}, r) \leq d\}\right) &= \mu\left(\{\bar{x} : \tilde{\sigma}_p(k_n, x_1^n) \leq f_p(\bar{x}, r), f_p(\bar{x}, r) \leq d\}\right) \\ &\quad + \mu\left(\{\bar{x} : \tilde{\sigma}_p(k_n, x_1^n) > f_p(\bar{x}, r), f_p(\bar{x}, r) \leq d\}\right), \qquad (79)\end{aligned}$$

and $\mu\left(\{\bar{x} : \tilde{\sigma}_p(k_n, x_1^n) > f_p(\bar{x}, r), f_p(\bar{x}, r) \leq d\}\right) \leq \mu\left(\{\bar{x} : \tilde{\sigma}_p(k_n, x_1^n) > f_p(\bar{x}, r)\}\right)$, where the almost sure convergence of $\tilde{\sigma}_p(k_n, X_1^n)$ to $f_p(\mathbf{X}, r)$ in (75) implies that

$$\lim_{n\longrightarrow\infty} \mu\left(\{\bar{x} : \tilde{\sigma}_p(k_n, x_1^n) > f_p(\bar{x}, r), f_p(\bar{x}, r) \leq d\}\right) = 0$$

obtaining the result in (78). Consequently, we have from (77) that

$$\lim_{n\longrightarrow\infty} \mu\left(\left\{\bar{x} \in \mathbb{R}^{\mathbb{N}} : \tilde{\sigma}_p(k_n, x_1^n) \leq d, f_p(\bar{x}, r) \leq d\right\}\right) = \lim_{n\longrightarrow\infty} \mu\left(\left\{\bar{x} \in \mathbb{R}^{\mathbb{N}} : f_p(\bar{x}, r) \leq d\right\}\right). \quad (80)$$

For the second term in the RHS of (76), it is simple to verify that

$$\mu\left(\left\{\bar{x} \in \mathbb{R}^{\mathbb{N}} : \tilde{\sigma}_p(k_n, x_1^n) \leq d, f_p(\bar{x}, r) > d\right\}\right) \leq \mu\left(\left\{\bar{x} \in \mathbb{R}^{\mathbb{N}} : \tilde{\sigma}_p(k_n, x_1^n) < f_p(\bar{x}, r)\right\}\right),$$

then the almost sure convergence in (75) implies that

$$\lim_{n\longrightarrow\infty} \mu\left(\left\{\bar{x} \in \mathbb{R}^{\mathbb{N}} : \tilde{\sigma}_p(k_n, x_1^n) \leq d, f_p(\bar{x}, r) > d\right\}\right) = 0.$$

Putting this result in (76) and using (80), it follows that

$$\lim_{n\longrightarrow\infty} \mu^n(A_d^{n,k_n}) = \lim_{n\longrightarrow\infty} \mu\left(\left\{\bar{x} \in \mathbb{R}^{\mathbb{N}} : f_p(\bar{x}, r) \leq d\right\}\right), \quad (81)$$

which concludes the argument. Finally to obtain the specific statement presented in (23), we first note that $f_p(\bar{x}, r) = f_{p,\mu_{\Psi(\bar{x})}}(r)$ for all $\bar{x} \in \mathbb{R}^{\mathbb{N}}$, where $(f_{p,\mu_\lambda}(r))_{r\in(0,1]}$ is the expression that has been fully characterized in Theorem 1 for any $\mu_\lambda \in \tilde{\mathcal{P}}$. In addition, we can use Theorem 1 i) stating that when μ_λ is ℓ_p-compressible, meaning that $(x^p)_{x\in\mathbb{R}} \notin L_1(\mu_{\lambda 1})$, then $f_{p,\mu_\lambda}(r) = 0$ for all $r \in (0, 1]$. Therefore, all the stationary and ergodic components $\mu_{\Psi(\bar{x})}$ that are ℓ_p-compressible satisfies that $f_p(\bar{x}, r) = f_{p,\mu_{\Psi(\bar{x})}}(r) \leq d$ independent of the pair (r, d), which explains the first term in the expression presented in (23). □

7. Acknowledgment

This material is based on work supported by grants of CONICYT-Chile, Fondecyt 1170854 and the Advanced Center for Electrical and Electronic Engineering, Basal Project FB0008. The author thanks Professor Martin Adams for proofreading this material and providing valuable comments about the organization and presentation of this paper. The author thanks Sebastian Espinosa for preparing the figures of this paper.

Appendix A. Proof of Lemma 8

Proof: For the proof, the following properties of the tail functions (that defines $\mathcal{F}_{\bar{\mu}_1}$ in (35)) will be used:

Proposition 3.
- $\mathcal{Y}_{v_p} = \mathcal{Y}_{\bar{\mu}_1} \setminus \{0\}$, *meaning that for all* $\tau > 0$, $(\phi_{\bar{\mu}_1}(\cdot))$ *is continuous at* τ *if, and only if,* $(\phi_{v_p}(\cdot))$ *is continuous at* τ.
- $\forall \tau_1 > \tau_2 > 0$, $\phi_{\bar{\mu}_1}(\tau_1) = \phi_{\bar{\mu}_1}(\tau_2)$ *if, and only if,* $\phi_{v_p}(\tau_1) = \phi_{v_p}(\tau_2)$.

The proof is presented in Appendix C.

Proof of i): Let us first show that $(f_{\bar{\mu}_1}(\cdot))$ is continuous in $(0, 1]$. It is sufficient to prove continuity on the function $\tilde{f}_{\bar{\mu}_1}(r)) \equiv 1 - (f_{\bar{\mu}_1}(r))^p$, which is induced by the following more simple relationship:[28]

$$\tilde{\mathcal{F}}_{\bar{\mu}_1} \equiv \left\{(\phi_{\bar{\mu}_1}(\tau), \phi_{v_p}(\tau)), \tau \in [0, \infty)\right\} \tag{A.1}$$

$$\bigcup_{\tau_n \in \mathcal{Y}_{\bar{\mu}_1}} \left\{(\bar{\mu}_1(C_{\tau_n}) + \alpha \bar{\mu}_1(\{-\tau_n, \tau_n\}), v_p(C_{\tau_n}) + \alpha v_p(\{-\tau_n, \tau_n\})), \alpha \in [0, 1)\right\}. \tag{A.2}$$

There are three distinct scenarios to consider:

- Let us first focus on the case where $r \in \mathcal{R}^*_{\bar{\mu}_1} \setminus \{\bar{\mu}_1(B_{\tau_n}), \tau_n \in \mathcal{Y}_{\bar{\mu}_1}\}$ (see, Eq.(32)). Under this assumption there exists $\tau_o \in [0, \infty) \setminus \mathcal{Y}_{\bar{\mu}_1}$ (in the domain where $\phi_{\bar{\mu}_1}(\cdot)$ is continuous) where $r = \phi_{\bar{\mu}_1}(\tau_o)$. From Proposition 3, $\phi_{v_p}(\cdot)$ is also continuos at τ_o where by construction in (A.1) $\tilde{f}_{\bar{\mu}_1}(r) = \phi_{v_p}(\tau_o)$. Let us consider an arbitrary $\epsilon > 0$. From the continuity of $\phi_{v_p}(\cdot)$ at τ_o there exists $\delta > 0$ such that $\left\{\phi_{v_p}(\tau), \tau \in B_\delta(\tau_o)\right\} \subset B_\epsilon(\tilde{f}_{\bar{\mu}_1}(r))$.[29] Without loss of generality, we can assume

[28] This from the continuity of the function $g(x) = \sqrt[p]{1-x}$ in $x \in [0, 1]$.

[29] $B_\epsilon(x) \equiv (x - \epsilon, x + \epsilon) \subset \mathbb{R}$ denotes the open ball of radius $\epsilon > 0$ centered at $x \in \mathbb{R}$.

that $\phi_{v_p}(\tau_o - \delta) > \tilde{f}_{\bar{\mu}_1}(r) = \phi_{v_p}(\tau_o) > \phi_{v_p}(\tau_o + \delta)$. Then from Proposition 3, it follows that $\phi_{\bar{\mu}_1}(\tau_o - \delta) > r = \phi_{\bar{\mu}_1}(\tau_o) > \phi_{\bar{\mu}_1}(\tau_o + \delta)$. Then, there exists $\bar{\delta} > 0$ such that $B_{\bar{\delta}}(r) \subset \{\phi_{\bar{\mu}_1}(\tau), \tau \in B_\delta(\tau_o)\}$. Therefore from (A.1), we have that for any $\bar{r} \in B_{\bar{\delta}}(r)$ there exists $\tau_{\bar{r}} \in B_\delta(\tau_o)$ where $\bar{r} = \phi_{\bar{\mu}_1}(\tau_{\bar{r}})$ and, consequently, $\tilde{f}_{\bar{\mu}_1}(\bar{r}) = \phi_{v_p}(\tau_{\bar{r}}) \in B_\epsilon(\tilde{f}_{\mu_1}(r))$, which concludes the argument in this case.

- Let us assume that $r \in \bigcup_{\tau_n \in \mathcal{Y}_{\bar{\mu}_1}} (\bar{\mu}_1(C_{\tau_n}), \bar{\mu}_1(B_{\tau_n}))$ (see, Eq.(32)). Then there is $\tau_n \in \mathcal{Y}_{\bar{\mu}_1}$ and a unique $\alpha_o \in (0,1)$ such that $r = \bar{\mu}_1(C_{\tau_n}) + \alpha_o \cdot \bar{\mu}_1(B_{\tau_n} \setminus C_{\tau_n})$ and, consequently, $\tilde{f}_{\bar{\mu}_1}(r) = v_p(C_{\tau_n}) + \alpha_o \cdot v_p(B_{\tau_n} \setminus C_{\tau_n})$ from (A.2). Without loss of generality, let us consider $\epsilon > 0$ small enough such that $B_\epsilon(\tilde{f}_{\bar{\mu}_1}(r)) \subset (v_p(C_{\tau_n}), v_p(B_{\tau_n}))$. Then from the continuity of the affine function $g(\alpha) \equiv v_p(C_{\tau_n}) + \alpha \cdot v_p(B_{\tau_n} \setminus C_{\tau_n})$ in $(0,1)$, there exists $\delta > 0$ (function of ϵ) such that $\{g(\alpha), \alpha \in B_\delta(\alpha_o)\} \subset B_\epsilon(\tilde{f}_{\bar{\mu}_1}(r))$. Therefore for any $\bar{r} \in \{\bar{\mu}_1(C_{\tau_n}) + \alpha \cdot \bar{\mu}_1(B_{\tau_n}), \alpha \in (\alpha_o - \delta, \alpha_o + \delta)\}$, $\tilde{f}_{\bar{\mu}_1}(\bar{r}) \in B_\epsilon(\tilde{f}_{\bar{\mu}_1}(r))$ from the construction in (A.2). Finally fixing $\bar{\delta} = \delta \cdot \bar{\mu}_1(B_{\tau_n} \setminus C_{\tau_n})$, we have that $\left\{\tilde{f}_{\bar{\mu}_1}(\bar{r}), \bar{r} \in B_{\bar{\delta}}(r)\right\} \subset B_\epsilon(\tilde{f}_{\bar{\mu}_1}(r))$, which concludes the argument in this case.

- Finally, we need to consider the case where $r \in \{\bar{\mu}_1(C_{\tau_n}), \tau_n \in \mathcal{Y}_{\bar{\mu}_1}\} \cup \{\bar{\mu}_1(B_{\tau_n}), \tau_n \in \mathcal{Y}_{\bar{\mu}_1}\}$. The argument mixed the steps already presented in the two previous scenarios and for sake of space it is omitted here as no new technical elements are needed.

Proof of ii): Let us consider $r_2 > r_1$ and assume that both belong to $\mathcal{R}^*_{\bar{\mu}_1}$. This means that there exist $\tau_1 > \tau_2 \geq 0$ such that $r_1 = \phi_{\bar{\mu}_1}(\tau_1)$ and $r_2 = \phi_{\bar{\mu}_1}(\tau_1)$. Then $\phi_{v_p}(\tau_1) < \phi_{v_p}(\tau_2)$ from Proposition 3, which implies the result by the construction of $f_{\bar{\mu}_1}(\cdot)$ in (35). Another important scenario to cover is the case when $r_1 = \bar{\mu}_1(C_{\tau_n}) + \alpha_1 \bar{\mu}_1(B_{\tau_n} \setminus C_{\tau_n})$ and $r_2 = \bar{\mu}_1(C_{\tau_n}) + \alpha_2 \bar{\mu}_1(B_{\tau_n} \setminus C_{\tau_n})$ with $\alpha_2 > \alpha_1$, $\tau_n \in \mathcal{Y}_{\bar{\mu}_1}$ and $\tau_n > 0$. Then in this case $v_p(C_{\tau_n}) + \alpha_2 v_p(B_{\tau_n} \setminus C_{\tau_n}) > v_p(C_{\tau_n}) + \alpha_1 v_p(B_{\tau_n} \setminus C_{\tau_n})$ because from Proposition 3 it follows that $v_p(B_{\tau_n} \setminus C_{\tau_n}) > 0$ if $\tau_n \in \mathcal{Y}_{\bar{\mu}_1} \setminus \{0\}$. Again the result in this case follows from (35). Mixing these two scenarios and using the monotonic property of the tails functions $(\phi_{\bar{\mu}_1}(\cdot), \phi_{v_p}(\cdot))$ proves the strict monotonic property of $(f_{\bar{\mu}_1}(\cdot))$ in $(0,1]$ if $0 \notin \mathcal{Y}_{\bar{\mu}_1}$ and, the strict monotonic property of $(f_{\bar{\mu}_1}(\cdot))$ in $(0,1] \setminus [1 - \bar{\mu}_1(\{0\}), 1]$ if $0 \in \mathcal{Y}_{\bar{\mu}_1}$. Then the remaining scenario to consider is $r_2 > r_1$ in $[1 - \bar{\mu}_1(\{0\}), 1]$ when $0 \in \mathcal{Y}_{\bar{\mu}_1}$ (the sparse case). In this context, from definition $v_p(\{0\}) = 0$, which implies from (35) that $f_{\bar{\mu}_1}(r_2) = f_{\bar{\mu}_1}(r_1) = 0$. Finally, it is worth noting that this is the only context ($0 \in \mathcal{Y}_{\bar{\mu}_1}$) and regime ($r \in [1-\bar{\mu}_1(\{0\}), 1]$) where $f_{\bar{\mu}_1}(\cdot)$ is not strictly monotonic.

Proof of iii): This part comes directly from the continuity of $f_{\bar{\mu}_1}(\cdot)$ in $(0,1)$ and the limiting values of the function (i.e., $f_{\bar{\mu}_1}(1) = 0$ and $\lim_{r \longrightarrow 0} f_{\bar{\mu}_1}(r) = 1$ from (35)). □

Appendix B. Proof of Proposition 2

Proof: i) follows from the definition of $\phi(\cdot)$ and ii) comes from the continuity of a measure under a monotone sequence of events converging to a limit [14]. The left continuous property of $\phi(\cdot)$ and the fact that $\phi^+_m(\tau) = \phi_m(\tau) - m(\{\tau\} \cup \{-\tau\})$ (in iii)) follows mainly from the continuity of a measure [14]. □

Appendix C. Proof Proposition 3

Proof: The proof of these two points derives directly from the definition of the tail function and the construction of v_p from $\bar{\mu}_1$. More precisely, both results derive from the observation that these two measures are almost mutually absolutely continuous in the sense that for all $B \in \mathcal{B}(\mathbb{R})$ such that $0 \notin B$, $\bar{\mu}_1(B) = 0$ if, and only if, $v_p(B) = 0$. In fact, for all $B \in \mathcal{B}(\mathbb{R})$ such that $0 \notin B$, $v_p(B) = \int_B \frac{|x|^p}{||(x^p)||_{L_1(\bar{\mu}_1)}} d\bar{\mu}_1(x)$ and, conversely, $\bar{\mu}_1(B) = \int_B \frac{||(x^p)||_{L_1(\bar{\mu}_1)}}{|x|^p} dv_p(x)$. □

References

[1] A. Amini, M. Unser, F. Marvasti, Compressibility of deterministic and random infinity sequences, IEEE Transactions on Signal Processing 59 (11) (2011) 5193–5201.

[2] R. Gribonval, V. Cevher, M. E. Davies, Compressible distributions for hight-dimensional statistics, IEEE Transactions on Information Theory 58 (8) (2012) 5016–5034.

[3] J. F. Silva, M. S. Derpich, On the characterization of ℓ_p-compressible ergodic sequences, IEEE Transactions on Signal Processing 63 (11) (2015) 2915–2928.

[4] V. Cevher, Learning with compressible priors, in: Neural Inf. Process. Syst.(NIPS), Canada, 2008.

[5] A. Amini, M. Unser, Sparsity and infinity divisibility, IEEE Transactions on Information Theory 60 (4) (2014) 2346–2358.

[6] M. Unser, P. Tafti, A. Amini, H. Kirshner, A unified formualtion of gaussian versus sparse stochastic processes — part ii: Discrete domain theory, IEEE Transactions on Information Theory 60 (5) (2014) 3036–3050.

[7] M. Unser, P. D. Tafti, An Introduction to Sparse Stochastic Processes, Cambridge Univ Press, 2014.

[8] R. Gribonval, Should penalized least squares regression be interpreted as maximum a posetriori estimation?, IEEE Transactions on Signal Processing 59 (5) (2011) 2405–2410.

[9] A. Amini, U. Kamilov, E. Bostan, M. Unser, Bayesian estimation for continuous-time sparse stocastic processes, IEEE Transactions on Signal Processing 61 (4) (2013) 907–929.

[10] R. Prasad, C. Murthy, Cramer-Rao-type of bounds for sparse bayesian learning, IEEE Transactions on Signal Processing 61 (3) (2013) 622–632.

[11] R. M. Gray, Probability, Random Processes, and Ergodic Properties, 2nd Edition, Springer, 2009.

[12] L. Breiman, Probability, Addison-Wesley, 1968.

[13] R. M. Gray, J. Kieffer, Asymptotically mean stationary measures, The Annals of Probability 8 (5) (1980) 962–973.

[14] S. Varadhan, Probability Theory, American Mathematical Society, 2001.

[15] R. M. Gray, Entropy and Information Theory, Springer - Verlag, New York, 1990.

[16] H. L. Royden, P. Fitzpatrick, Real Analysis, Pearson Education, 2010.