# On the Characterization of $\ell_p$-Compressible Ergodic Sequences

Jorge F. Silva, *Member, IEEE*, and Milan S. Derpich, *Member, IEEE*

***Abstract*—This work offers a necessary and sufficient condition for a stationary and ergodic process to be $\ell_p$-compressible in the sense proposed by Amini, Unser and Marvasti ["Compressibility of deterministic and random infinity sequences," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5193–5201, 2011, Def. 6]. The condition reduces to check that the $p$-moment of the invariant distribution of the process is well defined, which contextualizes and extends the result presented by Gribonval, Cevher and Davies in ["Compressible distributions for high-dimensional statistics," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5016–5034, 2012, Prop. 1]. Furthermore, for the scenario of non-$\ell_p$-compressible ergodic sequences, we provide a closed-form expression for the best $k$-term relative approximation error (in the $\ell_p$-norm sense) when only a fraction (rate) of the most significant sequence coefficients are kept as the sequence-length tends to infinity. We analyze basic properties of this rate-approximation error curve, which is again a function of the invariant measure of the process. Revisiting the case of i.i.d. sequences, we completely identify the family of $\ell_p$-compressible processes, which reduces to look at a polynomial order decay (heavy-tail) property of the distribution.**

***Index Terms*—Asymptotic analysis, best $k$-term approximation error analysis, compressed sensing, compressibility of infinite sequences, compressible priors, ergodic processes, heavy-tail distributions.**

## I. Introduction

DEFINING notions of compressibility for a stochastic process, meaning that with high probability realizations of the process can be well-approximated in some sense by its best $k$-term sparse version [3], has been a recent topic of active research [1], [2], [4]–[6]. Quantifying compressibility for random sequences and the identification of compressible and sparse distributions (priors) are relevant problems considering the recent development of the compressed sensing theory [7]–[9] and its applications. These results can play an important role in regression [10], signal reconstruction (for instance in the classical compressed sensing setting [2, Th. 2]), inference, and decision-making problems [11], [12]. One important case is defining such a compressibility notion for i.i.d. processes where the probability measure is equipped with a density function[1] [1], [2]. In this context, realizations of the process are non-sparse (almost surely), and conventional ways of defining compressibility for finite dimensional signals, based on the power-law decay of the best $k$-term approximation error (or sequences that belong to the weak-$\ell_p$ ball), are not applicable either, as shown in [1], [2].

Motivated by this problem, Amini *et al.* [1] and Gribonval *et al.* [2] have introduced new definitions for compressible random sequences. These notions are not based on the typical absolute approximation error decay pattern of the signals, but on a relative $\ell_p$-best $k$-term approximation error behavior. In particular, Amini *et al.* [1] formally define the concept of $\ell_p$-compressible process (details in Section II below). This new definition provides a meaningful way of categorizing i.i.d. random sequences (and their distributions), in terms of the probability that almost all the $\ell_p$-relative energy of the process is concentrated in an arbitrarily small sub-dimension of the coordinate domain, as the block-length tends to infinity. Under this context, they provide two important results using the theory of order statistics [1]. First of all, [1, Theorem 3] shows that a concrete family of i.i.d. heavy-tail distributions is $\ell_p$-compressible (including the generalized Pareto, Students's $t$ and log-logistic), while on the other side, [1, Theorem 1] demonstrates that families with exponentially decaying tails (such as Gaussian, Laplace, generalized Gaussian) are not $\ell_p$-compressible. Therefore, it is interesting to ask about the compressibility of i.i.d processes not considered in that analysis. In this direction, we highlight the work of Gribonval *et al.* [2], which under an alternative notion of relative $\ell_p$-compressibility (involving almost sure convergences instead of convergence in measure, which was the criterion adopted in [1]) and a different analysis setting (fixed-rate instead of the variable rate used in [1]), elaborates an exact dichotomy between compressible and non-compressible i.i.d. sequences. This raises the question of whether it is possible to connect Amini *et al.* [1] $\ell_p$-compressibility with the more refined almost sure (a.s.) convergence analysis of the $\ell_p$ best $k$-term relative approximation error in [2, Prop. 1], with the idea of completing the analysis of [1, Ths. 1 and 3].

To address this question, we extend the analysis from i.i.d. sequences to stationary and ergodic processes. In this broader setting, the main result (Theorem 1) provides a necessary and sufficient condition for a stationary and ergodic process to be $\ell_p$-compressible (in the sense of Amini *et al.* [1, Def. 6]), for

[1]The probability is absolutely continuous with respect to the Lebesgue measure [13].

Manuscript received July 28, 2014; revised January 11, 2015; accepted March 13, 2015. Date of publication April 02, 2015; date of current version May 05, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Fancesco Verde. This material is based on work supported by grants of CONICYT-Chile, Fondecyt Grant 1140840 and the Advanced Center for Electrical and Electronic Engineering (AC3E), Basal Project FB0008. In addition, the work of M. S. Derpich was partially supported by CONICYT Fondecyt Grant 1140384. (*Corresponding Author: Jorge F. Silva.*)

J. F. Silva is with the Department of Electrical Engineering, Information and Decision Systems Group, University of Chile, 412-3 Santiago, Chile (e-mail: josilva@ing.uchile.cl).

M. S. Derpich is with the Department of Electronic Engineering, Universidad Técnica Federico Santa María, Valparaíso 2390123, Chile (e-mail: milan.derpich@usm.cl).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2015.2419183

any arbitrary $p > 0$. Furthermore, for the case of non $\ell_p$-compressible ergodic processes, we provide a closed-form expression for an achievable rate v/s $\ell_p$-approximation error function. The key element in the proof is the application of the *ergodic theorem* [14] and the derivation of intermediate almost sure convergence results (Lemma 2 and 3 in Section IV) that match and extend the approximation result presented by Gribonval *et al.* [2, Prop. 1] developed for the i.i.d. case. A corollary of Theorem 1 implies a necessary and sufficient condition to categorize i.i.d. random sequences in terms of $\ell_p$-compressibility, which completes the analysis presented in [1, Ths. 1 and 3]. In addition, for the class of non-$\ell_p$-compressible ergodic sequences, we provide an analysis of its rate-approximation error curve demonstrating that is continuous, differentiable (Theorem 2) and is convex under some conditions (Theorem 3). Finally as an application, we revisit the interplay between $\ell_1$-compressible ergodic sequences and the performance of the classical Gaussian compressed sensing (GCS) setting [15], in the asymptotic regime when the block-length tends to infinity. Using the well-known $\ell_1$-instance optimality performance guarantee of the GCS scheme [3], [15], [16], we show (Theorem 4) that an arbitrarily small number of linear measurements (zero-rate) is needed to achieve zero distortion, in an $\ell_1$-noise to signal ratio (NSR) sense. A preliminary version of this work was presented in [17]. The current version extends the presentation and analysis of the main result, provides further analysis of non-$\ell_p$ compressible ergodic sequences and explores connections with compressed sensing (CS).

The rest of the paper is organized as follows. Section II introduces some preliminary elements and definitions. Sections III and IV are devoted to the presentation of the main result on the characterization of compressible ergodic processes and its proof, respectively. Section V studies basic properties of the the rate-approximation error curve for non $\ell_p$-compressible processes. Finally, Section VI elaborates an interplay between $\ell_1$-compressibility and compressed sensing. Some of the proofs and derivations are presented in the Appendix sections.

## II. Preliminaries and Basic Definitions

For a finite dimensional 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 $|x_{n,1}| \geq |x_{n,2}| \geq \cdots \geq |x_{n,n}|$. For some $p > 0$ and $k \in \{1, .., n\}$, let

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

denote the $\ell_p$-norm of the best $k$-term approximation of $x^n$, where by definition $||x^n||_{\ell_p} = \zeta_p(n, x^n)$. In addition,

$$\sigma_p(k, x^n) \equiv (|x_{n,k+1}|^p + \cdots + |x_{n,n}|^p)^{\frac{1}{p}},\ \forall k \in \{1, .., n\}, \tag{2}$$

denotes the best $k$-term $\ell_p$-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 $\sigma_p(k, x^n) = \min_{\tilde{x}^n \in \Sigma_k^n} ||x^n - \tilde{x}^n||_{\ell_p}$. For the analysis of infinite sequences, Amini *et al.* [1] and Gribonval *et al.* [2] have proposed the following relative best $k$-term $\ell_p$-distortion indicator:

$$\tilde{\sigma}_p(k, x^n) \equiv \frac{\sigma_p(k, x^n)}{||x^n||_{\ell_p}} \in [0, 1],\ k \in \{1, .., n\}, \tag{3}$$

with the objective of extending notions of compressibility to sequences that have infinite $\ell_p$-norm.

### A. Rate of Innovation vs. Distortion for Infinite Sequences

*Definition 1:* For a sequence $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, the rate-distortion pair $(r, d) \in [0, 1]^2$ is $\ell_p$-achievable for $(x_n)$ if, there is a sequence of positive integers $(k_n)$ such that $\limsup_{n\to\infty} \frac{k_n}{n} = r$ and

$$\limsup_{n\to\infty} \tilde{\sigma}_p(k_n, x^n) \leq d, \tag{4}$$

where $x^n = (x_1, .., x_n) \in \mathbb{R}^n$ is the finite-block version of length $n$ of $(x_n)_{n\in\mathbb{N}}$.

Note that the use of the relative best $k$-term $\ell_p$-distortion in (4) allows the analysis of sequences with infinite $\ell_p$-norm.

*Definition 2:* For a sequence $(x_n)_{n\in\mathbb{N}}$ and $p > 0$, we define its rate-distortion $\ell_p$- *approximation* function by $r_p(d, (x_n)_{n\in\mathbb{N}}) \equiv$

$$\inf \{r \in [0, 1], (r, d) \text{ is } \ell_p\text{-achievable for } (x_n)_{n\in\mathbb{N}}\}, \tag{5}$$

for all $d \in [0, 1]$.

A simple consequence of these definitions is the following result:

*Proposition 1:* For all $(k_n)$ such that $\liminf_{n\to\infty} \frac{k_n}{n} > r_p(d, (x_n))$ then $\limsup_{n\to\infty} \tilde{\sigma}_p(k_n, x^n) \leq d$.

(The proof is presented in Appendix IV-A)

Hence, $r_p(d, (x_n)_{n\in\mathbb{N}})$ can be seen as the critical asymptotic rate of innovation of $(x_n)$ when a relative best $k$-term $\ell_p$-approximation error of magnitude $d$ is tolerated.

Alternatively, Amini, Unser and Marvasti [1] have introduced a notion of critical dimension for finite length signals, and from this, a notion of $\ell_p$-compressibility for infinite sequences. We revisit those notions here:

*Definition 3:* [1, Def. 4] For $x^n \in \mathbb{R}^n$ and $d \in (0, 1)$, let us define

$$\kappa_p(d, x^n) \equiv \min \{k \in \{1, .., n\} : \tilde{\sigma}_p(k, x^n) \leq d\}. \tag{6}$$

Then, a sequence $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ is called $\ell_p$-compressible if, $\forall d \in (0, 1)$

$$\lim_{n\to\infty} \frac{\kappa_p(d, x^n)}{n} = 0, \tag{7}$$

where $x^n = (x_1, .., x_n)$ is the truncated finite-block vector of $(x_n)_{n\in\mathbb{N}}$.

This notion of compressibility says that when the block-length tends to infinity, a negligible fraction of the coefficients is needed to represent $(x_n)_{n\in\mathbb{N}}$ with an arbitrary small $\ell_p$-distortion in the sense of (3). Note that $(\kappa_p(d, x^n))_{n\in\mathbb{N}}$ is signal dependent and a variable-rate sequence. In addition, it offers the critical number of terms needed to achieve a best $k$-term approximation error smaller or equal to $d$ in the sense of (3). From this, it should be related with the critical rate (from a fixed-rate analysis) described in Definition 2. That relationship is presented in the following result result:

*Lemma 1:* Let $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ and $d \in (0, 1)$, then

$$\liminf_{n\to\infty} \frac{\kappa_p(d, x^n)}{n} \leq r_p(d, (x_n)_{n\in\mathbb{N}}) \leq \limsup_{n\to\infty} \frac{\kappa_p(d, x^n)}{n}. \tag{8}$$

(The proof is presented in Appendix IV-B)

A corollary of this result implies that if the limit of $(\kappa_p(d,x^n)/n)_{n\in\mathbb{N}}$ exists, then $\lim_{n\to\infty}\frac{\kappa_p(d,x^n)}{n} = r_p(d,(x_n))$. In particular from Lemma 1, if $(x_n)$ is $\ell_p$-compressible, then $r_p(d,(x_n)) = 0$ for all $d \in (0,1)$. We refer the interested reader to Amini *et al.* [1] for further discussion and examples of $\ell_p$-compressible sequences.

### B. Rate of Innovation vs. Distortion for Random Sequences

Analogous notions of rate of innovation vs. best $k$-term $\ell_p$-distortion and compressibility can be stated for the case of random sequences (or processes). Let $X_1,..,X_n,..$ be a random sequence with values in $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ and characterized by its consistent family of finite-dimensional probabilities $\{\mu^n \in \mathcal{P}(\mathbb{R}^n) : n \geq 1\}$ [14], where $X^n = (X_1,..,X_n) \sim \mu^n$ and $\mathcal{P}(\mathbb{R}^n)$ denotes the space of probability measures for the Borel measurable space $(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$. As a short-hand, we denote by $\mathbb{P} = \{\mu^n : n \geq 1\}$ the process distribution of $(X_n)_{n\in\mathbb{N}}$.

Let us define the following measurable set:

$$\begin{aligned}\mathcal{A}_d^{n,k} &\equiv \{x^n \in \mathbb{R}^n : \tilde{\sigma}_p(k,x^n) \leq d\}\\ &= \{x^n \in \mathbb{R}^n : \kappa_p(d,x^n) \leq k\} \in \mathcal{B}(\mathbb{R}^n),\end{aligned} \tag{9}$$

where the equality is by (6). Then in analogy with Definition 3, Amini *et al.* [1] proposed the following:

*Definition 4:* [1, Defs.5 and 6] Let $(X_n)_{n\in\mathbb{N}}$ be a random sequence (equipped with $\mathbb{P}$). Then for any $\epsilon \in (0,1)$ and $d \in (0,1)$, $\tilde{\kappa}_p(d,\epsilon,\mu^n) \equiv$

$$\min\left\{k \in \{1,\ldots,n\} : \mu^n(\mathcal{A}_d^{n,k}) \geq 1-\epsilon\right\}, \tag{10}$$

is the critical number of terms that makes the set $\mathcal{A}_d^{n,k}$ $\epsilon$-typical with respect to $\mu^n$. The process $(X_n)_{n\in\mathbb{N}}$ (and $\mathbb{P}$, respectively) is said to be $\ell_p$-compressible, if $\forall\epsilon \in (0,1)$, $\forall d \in (0,1)$,

$$\lim_{n\to\infty}\frac{\tilde{\kappa}_p(d,\epsilon,\mu^n)}{n} = 0. \tag{11}$$

Alternatively, we can consider the following fixed-rate notions:

*Definition 5:* Let $(X_n)_{n\in\mathbb{N}}$ be a process characterized by $\mathbb{P}$, and let us consider $\epsilon \in (0,1)$, $r \in (0,1)$ and $d \in (0,1)$. We say that the rate-distortion pair (r,d) is $\ell_p$-achievable for $(X_n)$ with $\epsilon$ probability, if there exists a sequence of positive integers $(k_n)$ such that $\limsup_{n\to\infty}\frac{k_n}{n} = r$ and

$$\liminf_{n\to\infty}\mu^n(\mathcal{A}_d^{n,k_n}) \geq 1-\epsilon. \tag{12}$$

*Definition 6:* The rate vs. best $k$-term approximation error function of $(X_n)_{n\in\mathbb{N}}$ (in short the rate-approximation error function of $(X_n)$) with $\epsilon$ probability is given by[2]: $\tilde{r}_p(d,\epsilon,\mathbb{P}) \equiv$

$$\inf\{r \in [0,1], (r,d) \text{ is } \ell_p\text{-achievable for } (X_n) \text{ with } \epsilon \text{prob.}\}. \tag{13}$$

A simple relationship between $\tilde{r}_p(d,\epsilon,\mathbb{P})$ and the critical number of terms in (10) can be established in the asymptotic regime when $n$ goes to infinity, showing that our fixed-rate concept is a weaker one.

[2]Note that this rate-approximation error function (of $(X_n)_{n\in\mathbb{N}}$) is expressed as a function of $\mathbb{P}$.

*Proposition 2:* For any $\epsilon \in (0,1)$ and $d \in (0,1)$

$$\tilde{r}_p(d,\epsilon,\mathbb{P}) \leq \limsup_{n\to\infty}\frac{\tilde{\kappa}_p(d,\epsilon,\mu^n)}{n}. \tag{14}$$

(The proof is presented in Appendix IV-C)

In the next section, we will study the class of stationary and ergodic processes [14], where the best $k$-term approximation properties measured in terms of $\tilde{r}_p(d,\epsilon,\mathbb{P})$ will be characterized in closed-form. Furthermore, it will be shown for this class of random sequences that

$$\tilde{r}_p(d,\epsilon,\mathbb{P}) = \lim_{n\to\infty}\frac{\tilde{\kappa}_p(d,\epsilon,\mu^n)}{n}$$

for all $\epsilon > 0$ and $d \in (0,1)$, refining the basic relationship presented in Proposition 2.

## III. Analysis of Ergodic Processes

Let $(X_n)_{n\in\mathbb{N}}$ be a stationary and ergodic process with distribution $\mathbb{P} = \{\mu^n : n \geq 1\}$, where $\mu \in \mathcal{P}(\mathbb{R})$ denotes its marginal shift-invariant distribution [14]. For simplicity[3], we assume that $\mu \ll \lambda$ where $\lambda$ denotes the Lebesgue measure [14]. Then $\mu$ is equipped with a probability density function (pdf) and $d\mu(x) = \frac{d\mu}{d\lambda}(x)d\lambda(x)$.

For a measure $v$ on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, a measurable function $f : (\mathbb{R},\mathcal{B}(\mathbb{R})) \longrightarrow (\mathbb{R},\mathcal{B}(\mathbb{R}))$ is said to be integrable with respect to $v$ if [14]

$$\int_{\mathbb{R}}|f(x)|\,dv(x) < \infty, \tag{15}$$

where $L_1(v)$ denotes the collection of $v$-integrable functions.

We are in the position to state the main result:

*Theorem 1:* Let $(X_n)_{n\in\mathbb{N}}$ be a stationary and ergodic process with shift-invariant distribution $\mu \in \mathcal{P}(\mathbb{R})$ such that $\mu \ll \lambda$. Then for any $p > 0$, we have the following dichotomy:

i) If $(x^p)_{x\in\mathbb{R}} \notin L_1(\mu)$: then $(X_n)_{n\in\mathbb{N}}$ is $\ell_p$-compressible, i.e., $\forall\epsilon \in (0,1)$ and $\forall d \in (0,1)$,

$$\tilde{r}_p(d,\epsilon,\mathbb{P}) = \lim_{n\to\infty}\frac{\tilde{\kappa}_p(d,\epsilon,\mu^n)}{n} = 0. \tag{16}$$

ii) If $(x^p)_{x\in\mathbb{R}} \in L_1(\mu)$: then $(X_n)_{n\in\mathbb{N}}$ is not $\ell_p$-compressible. Furthermore, if we introduce the induced probability measure in $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ by:

$$v_p(B) \equiv \frac{\int_B |x^p|\,d\mu(x)}{\int_{\mathbb{R}}|x^p|\,d\mu(x)},\ \forall B \in \mathcal{B}(\mathbb{R}), \tag{17}$$

then $\forall d \in (0,1)$ and $\forall\epsilon \in (0,1)$

$$\tilde{r}_p(d,\epsilon,\mathbb{P}) = \lim_{n\to\infty}\frac{\tilde{\kappa}_p(d,\epsilon,\mu^n)}{n} = \mu(B_{\tau(d)}), \tag{18}$$

where $\forall\tau \geq 0$

$$B_\tau \equiv (-\infty,-\tau] \cup [\tau,\infty) \in \mathcal{B}(\mathbb{R}), \tag{19}$$

and $\tau(d) > 0$ is a solution of the identity:

$$v_p(B_{\tau(d)}) = 1-d^p. \tag{20}$$

[3]The general case when $\mu$ has atomic components can be extended from the result presented here. This extension does not offer new insights, while it requires the introduction of technicalities that make the statement and the proof of the result more involved.

The proof is presented in Section IV.

### A. Discussion and Interpretation of Theorem 1

1: Theorem 1 offers a necessary and sufficient condition for a stationary and ergodic process to be 1$p$-compressible in the sense elaborated in Definition 4.

2: In the case of non $\ell_p$-compressible processes, i.e., when $(x^p)_{x\in\mathbb{R}} \in L_1(\mu)$, Theorem 1 offers what we call the achievable rate-distortion region for the process, given by the set of critical rate-distortion pairs:

$$\mathcal{R}_\mu \equiv \left\{ \left( \mu(B_\tau), \sqrt[p]{1 - v_p(B_\tau)} \right) \in [0,1]^2 : \tau \geq 0 \right\}. \quad (21)$$

This region depends solely on the shift invariant measure $\mu \in \mathcal{P}(\mathbb{R})$ and its induced measure $v_p \in \mathcal{P}(\mathbb{R})$ in (17). More details on the characterization of this region will be presented in Section V.

3: Under the assumption that $(x^p)_{x\in\mathbb{R}} \in L_1(\mu)$ and $\mu \ll \lambda$, we have that any rate $r \in [0,1]$ and distortion $d \in [0,1]$ are achievable (see proof in Section IV and more details in Section V). This fact is used to derive a concrete analytical expression for $\tilde{r}_p(d,\epsilon,\mathbb{P})$ in (18). Furthermore, from the characterization in (18) and (20), it can be shown that $\tilde{r}_p(d,\epsilon,\mathbb{P})$ is a continuous and differentiable function with respect to $d \in (0,1)$ (see Theorem 2 in Section V).

4: In both scenarios i) and ii), the critical rate $\tilde{r}_p(d,\epsilon,\mathbb{P})$ for a stationary and ergodic process is independent of $\epsilon$. The reason is that asymptotically as $n$ goes to infinity, the characterization of $\tilde{r}_p(d,\epsilon,\mathbb{P})$ implies to compute probabilities on events that belong to the tail $\sigma$-field of the process, which is known to be trivial (i.e., their events have zero or one probability) for the case of ergodic processes [14], [18], [19]. Therefore, we obtain almost sure convergence results that make independent of $\epsilon$ the value of our object of interest $\tilde{r}_p(d,\epsilon,\mathbb{P})$ (see Section IV for details).

5: A natural order among stationary and ergodic process can be established from Theorem 1.

*Proposition 3:* If $(X_n)_{n\in\mathbb{N}}$ is $\ell_p$-compressible for some $p > 0$, then $(X_n)_{n\in\mathbb{N}}$ is $\ell_q$-compressible for all $q \geq p$.

*Proof:* If $(x^p)_{x\in\mathbb{R}} \notin L_1(\mu)$, then $(x^q)_{x\in\mathbb{R}} \notin L_1(\mu)$ for all $q \geq p$.

*Proposition 4:* If $(X_n)_{n\in\mathbb{N}}$ is not $\ell_p$-compressible for $p > 0$ then $(X_n)_{n\in\mathbb{N}}$ is not $\ell_q$-compressible for all $q \leq p$.

*Proof:* If $(x^p)_{x\in\mathbb{R}} \in L_1(\mu)$, then $(x^q)_{x\in\mathbb{R}} \in L_1(\mu)$ for all $q \leq p$.

6: Revisiting the i.i.d. scenario[4], we want to highlight the results by Amini *et al.* [1] related to $\ell_1$-compressibility in the sense of (11). In particular, [1, Theorem 1] says that if $\mu$ is such that for some $\gamma < 0$, $\mathbb{E}_{X\sim\mu}(e^{\gamma\cdot X}) < \infty$ then the i.i.d. process is not $\ell_1$-compressible. In contrast, [1, Theorem 3] says that if $\mu$ belongs to the domain of attraction of an $\alpha$-stable distribution [14, Chap. 9.11, pp. 207–213] with $\alpha \in (0,1)$, then the process is $\ell_1$-compressible. First for $p = 1$, Theorem 1 provides a refined result, revealing a richer (indeed, the complete) family of i.i.d. distributions that are not $\ell_1$-compressible. In fact, in addition to distributions that go to zero exponentially, and consequently $(x)_{x\in\mathbb{R}} \in L_1(\mu)$, (Gaussian, Laplacian, Gamma, etc.), heavy tail distributions whose density function are tail lower and upper dominated by a power law decay of the form $\frac{1}{|x|^{p+1}}$ with $p > 1$ are not $\ell_1$-compressible either[5]. On the other hand, concerning [1, Th. 3], it is simple to verify that any $\mu$ that is in the domain of attraction of an $\alpha$-stable law with $\alpha < 1$ [14, Ch.9] satisfies that $\mathbb{E}_{X\sim\mu}(|X|) = \infty$ (see Appendix IV-E for details), and consequently, part i) of Theorem 1 covers this family of $\ell_1$-compressible i.i.d. processes.

7: Complementing the previous point, from Theorem 1 we can state the following:

*Corollary 1:* Let $(X_n)_{n\in\mathbb{N}}$ be a stationary and ergodic process, if $\exists\gamma > 0$ where $\mathbb{E}_{X_1\sim\mu}(e^{-\gamma X_1}) < \infty$, then

$$(X_n)_{n\in\mathbb{N}} \text{ is not } \ell_p\text{-compressible for any } p > 0.$$

Therefore, stationary and ergodic processes equipped with a shift-invariant distribution that follows a Gaussian, generalized Gaussian, Laplacian and Gamma are not $\ell_p$-compressible in the sense of Definition 4, for any $p > 0$. In addition, if $\mu$ is finitely supported, i.e., $\exists C > 0$ where $\mu([-C,C]) = 1$, then its process is not $l_p$-compressible for any $p > 0$.

*Corollary 2:* Let $(X_n)_{n\in\mathbb{N}}$ be a ergodic process with invariant distribution $\mu \ll \lambda$ and density $f_\mu(x) \equiv \frac{d\mu}{d\lambda}(x), \forall x \in \mathbb{R}$. If $f_\mu(x)$ decays as $|x|^{-(\tau+1)}$ for some $\tau > 0$, then[6]

$$(X_n)_{n\in\mathbb{N}} \text{ is } l_p\text{-compressible, if and only if, } p \geq \tau.$$

Therefore, for shift invariant distributions characterized by a power-tail behavior, which belong to the category of heavy tail distributions, a complete picture of the range in which its ergodic process is $\ell_p$-compressible is obtained.

8: For the proof of Theorem 1, we derive almost sure convergence results (see Lemma 2 and 3 in Section IV). In the case when $(x^p)_{x\in\mathbb{R}} \in L_1(\mu)$: if $(k_n)$ is such that $\lim_{n\to\infty} \frac{k_n}{n} = \mu(B_\tau)$ for some $\tau > 0$, then $\lim_{n\to\infty} \tilde{\sigma}_p(k_n, X^n) = \sqrt[p]{1 - v_p(B_\tau)}$, $\mathbb{P} - a.s.$. Furthermore, for $d = \sqrt[p]{1 - v_p(B_\tau)}$ for some $\tau > 0$, it follows that $\lim_{n\to\infty} \frac{\kappa_p(d,X^n)}{n} = \mu(B_\tau)$, $\mathbb{P} - a.s.$. In the case when $(x^p)_{x\in\mathbb{R}} \notin L_1(\mu)$: if $\lim_{n\to\infty} \frac{k_n}{n} > 0$, then $\lim_{n\to\infty} \tilde{\sigma}_p(k_n, X^n) = 0$, $\mathbb{P} - a.s.$ These results are consistent and extend the result by Gribonval *et al.* [2, Prop. 1], which for the i.i.d. case shows the same almost-sure convergence limit for the object $\tilde{\sigma}_p(k_n, X^n)$. Their proof was based on the *Wald's lemma* of order statistics (see details in [2, Th. 6]). In contrast, our proof is based on the use of the tail events in (19), some induced empirical distributions on those events, and the convergence of those empirical measures through the application of the ergodic theory (see Section IV for details). The idea adopted in our proof was to look at the empirical distributions of $\mu$ and $v_p$ as the objects of interest, instead of the partial

[4]It is well-known that i.i.d. processes are stationary and ergodic [14].

[5]A measure $\mu \ll \lambda$ is tail lower and upper dominated by a no-negative function $g(x)$, if there exits $x_0 > 0$ and $0 < C_0 < C_1$ such that for any $x$ such that $|x| > x_o$ then $C_0 \cdot g(x) \leq f_\mu(x) \leq C_1 \cdot g(x)$. Here $f_\mu(x)$ denotes the pdf of $\mu$.

[6]We say that $f_\mu(x)$ decays as $|x|^{-(\tau+1)}$ if there exists $x_o > 0$ and $0 < K_1 < K_2 < \infty$ and $\lim_{x\to\infty} \frac{f_\mu(-x)}{f_\mu(x)} = \rho \in \mathbb{R}^+ \cup \{\infty\}$, where: if $\rho = 0$, then $\forall x > x_o$, $K_1 x^{-(\tau+1)} \leq f_\mu(x) \leq K_2 x^{-(\tau+1)}$; if $\rho = \infty$, then $\forall x < -x_o$, $K_1|x|^{-(\tau+1)} \leq f_\mu(x) \leq K_2|x|^{-(\tau+1)}$; and otherwise, $\forall |x| > x_o$ then $K_1|x|^{-(\tau+1)} \leq f_\mu(x) \leq K_2|x|^{-(\tau+1)}$.

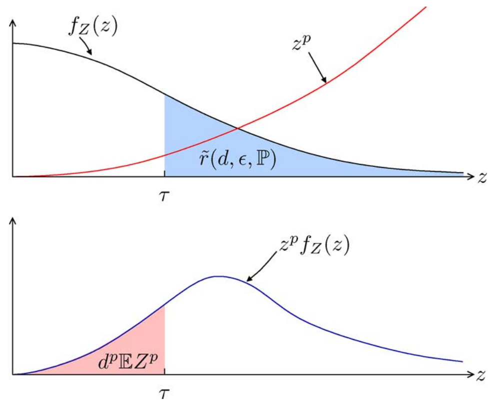


Fig. 1. Graphical representation of the relationship between $\tau$, $\tilde{r}(d,\epsilon,\mathbb{P})$ and $d^p$ in Theorem 1. Notice that $\mathbb{E}Z^p$ is the total area under the bottom curve.

sums of the ordered statistics considered in [1], [2]. That difference was essential to extend the mentioned almost sure convergence results (in Lemma 2 and 3) to the family of stationary and ergodic processes.

9: Under the assumption that $\mu \ll \lambda$, Theorem 1 shows another interesting dichotomy:

*Corollary 3:* If for some $\epsilon \in (0,1)$ and for some $d \in (0,1)$ it holds that

$$\tilde{r}_p(d,\epsilon,\mathbb{P}) = \lim_{n\to\infty} \frac{\tilde{\kappa}(d,\epsilon,\mu^n)}{n} = 0,$$

then the latter also holds for all $\epsilon \in (0,1)$ and for all $d \in (0,1)$. Likewise, if for some $\epsilon \in (0,1)$ and some $d \in (0,1)$ $\tilde{r}_p(d,\epsilon,\mathbb{P}) > 0$ then for all $\bar{d} \in (0,1)$ and all $\bar{\epsilon} \in (0,1)$ $\tilde{r}_p(\bar{d},\bar{\epsilon},\mathbb{P}) = \lim_{n\to\infty} \frac{\tilde{\kappa}(\bar{d},\bar{\epsilon},\mu^n)}{n} > 0$.

### B. Graphical Interpretation of $\mathcal{R}_\mu$

Note from (20) and (17) that $d^p = 1 - v_p(B_{\tau(d)})$ and $v_p(B_{\tau(d)}) = \int_{B_{\tau(d)}} |x^p|\, d\mu(x) / \int_{\mathbb{R}} |x^p|\, d\mu(x)$, respectively. The numerator in the last expression corresponds to the expected value of $Z^p \cdot \mathbf{1}_{B_{\tau(d)}}(Z)$, for a random variable $Z \triangleq |X|$ where $X \sim \mu$. Similarly, from (18), the optimal rate $\tilde{r}(d,\epsilon,\mathbb{P})$ equals the expected value of $\mathbf{1}_{B_{\tau(d)}}(Z)$. Thus, $\tilde{r}(d,\epsilon,\mathbb{P})$ corresponds to the area under the "tail" of $f_Z(z)$, which denotes the pdf of $Z$, depicted in Fig. 1 (top), while $d^p$ coincides with the area under the curve $z^p f_Z(z)$ to the left of $\tau(d)$, coloured in Fig. 1 (below). This graphical representation allows for an intuitive interpretation of the relationship between $p$ and the compressibility of a given stationary ergodic process $(X_n)_{n\in\mathbb{N}}$. In order for this process to be compressible, $\tilde{r}(d,\epsilon,\mathbb{P})$ must be zero for every $\epsilon \in (0,1)$ and for every $d \in (0,1)$. Equivalently, (and recalling from (16) that $\tilde{r}(d,\epsilon,\mathbb{P})$ is a limit), it must be possible to achieve any $d \in (0,1)$ while keeping an arbitrarily small fraction of the elements of the process, as $n \to \infty$. Such requirement is satisfied if and only if $\mathbb{E}Z^p = \infty$ (i.e., if $x^p \notin L_1(\mu)$), which implies that no matter how large $\tau$ is chosen, the shaded area in Fig. 1 (bottom), being infinite, will yield a zero $d$.

## IV. Proof of the Main Result

*Proof:* Let us first consider the case when $(x^p)_{x\in\mathbb{R}} \in L_1(\mu)$. For the rest, it is important to note that given that $\mu \ll \lambda$, then for all $r \in (0,1)$ there exists $\tau > 0$ such that $\mu(B_\tau) = r$, and for all $d \in (0,1)$ there exists $\tau > 0$ such that $\sqrt[p]{1 - v_p(B_\tau)} = d$.[7]

For $(X_1,..,X_n) \sim \mu^n$, we can define:

$$n(X^n, B_\tau) \equiv \sum_{i=1}^{n} 1_{B_\tau}(X_i), \tag{22a}$$

where from the *ergodic theorem* [14, Th. 6.28], $\forall \tau \geq 0$,

$$\lim_{n\to\infty} \frac{n(X^n,B_\tau)}{n} = \mathbb{E}_{X_1\sim\mu}(1_{B_\tau}(X_1)) = \mu(B_\tau),\ \mathbb{P}-a.s., \tag{22b}$$

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} |X_i|^p 1_{B_\tau}(X_i)}{\sum_{i=1}^{n} |X_i|^p} = \frac{\mathbb{E}_{X_1\sim\mu}(|X_i|^p 1_{B_\tau}(X_1))}{\mathbb{E}_{X_1\sim\mu}(|X_i|^p)} = \frac{\int_{B_\tau} |x|^p d\mu(x)}{\int_{\mathbb{R}} |x|^p d\mu(x)} = v_p(B_\tau),\ \mathbb{P}-a.s. \tag{22c}$$

The second almost sure convergence is from the assumption that $(x^p)_{x\in\mathbb{R}} \in L_1(\mu)$. Then, we can state the following:

*Lemma 2:* Let $(X_n)_{n\in\mathbb{N}}$ be a stationary and ergodic process with distribution $\mu \ll \lambda$ and $(x^p)_{x\in\mathbb{R}} \in L_1(\mu)$. Then for any $\tau \geq 0$ and sequence $(k_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} \frac{k_n}{n} = \mu(B_\tau)$, we have that

$$\lim_{n\to\infty} \tilde{\sigma}_p(k_n, X^n) = \sqrt[p]{1 - v_p(B_\tau)},\ \mathbb{P}-a.s. \tag{23}$$

In addition, $\forall d \in (0,1)$,

$$\lim_{n\to\infty} \frac{\kappa_p(d,X^n)}{n} = \mu(B_{\tau(d)}),\ \mathbb{P}-a.s. \tag{24}$$

where $\tau(d) \geq 0$ is a solution of $\sqrt[p]{1 - v_p(B_{\tau(d)})} = d$. (The proof is presented in Section IV-A)

In order to prove (18), let us fix $d \in (0,1)$. Then there exists $\tau(d)$, such that (20) holds and from Lemma 2 if $(k_n)_{n\in\mathbb{N}}$ is such that $\frac{k_n}{n} \to \mu(B_{\tau(d)})$, then $\lim_{n\to\infty} \tilde{\sigma}_p(k_n,X^n) = d$, $\mathbb{P}-a.s.$.

Let us consider an arbitrary $\tilde{\tau} < \tau(d)$ such that $\mu(B_{\tilde{\tau}}) > \mu(B_{\tau(d)})$, then again from Lemma 2, if a sequence $(\tilde{k}_n)$ is such that $\lim_{n\to\infty} \frac{\tilde{k}_n}{n} = \mu(B_{\tilde{\tau}})$, then $\lim_{n\to\infty} \tilde{\sigma}_p(\tilde{k}_n,X^n) = \sqrt[p]{1 - v_p(B_{\tilde{\tau}})} < d$, $\mathbb{P}-a.s.$. Consequently, $\tilde{\sigma}_p(\tilde{k}_n,X^n)$ convergences almost surely to a distortion strictly less than $d$, and then for all $\epsilon > 0$:

$$\lim_{n\to\infty} \inf \mu^n(\mathcal{A}_d^{n,\tilde{k}_n}) > 1 - \epsilon. \tag{25}$$

Hence from the definition of $\tilde{\kappa}_p(d,\epsilon,\mu^n)$ in (10), we have that $\tilde{\kappa}_p(d,\epsilon,\mu^n) \leq \tilde{k}_n$ eventually in $n$, which implies that

$$\lim_{n\to\infty} \sup \frac{\tilde{\kappa}_p(d,\epsilon,\mu^n)}{n} \leq \lim_{n\to\infty} \frac{\tilde{k}_n}{n} = \mu(B_{\tilde{\tau}}). \tag{26}$$

This upper bound is valid for any $\tilde{\tau} < \tau(d)$ such that $\mu(B_{\tilde{\tau}}) > \mu(B_{\tau(d)})$, then $\tilde{r}_p(d,\epsilon,\mathbb{P}) \leq$

$$\lim_{n\to\infty} \sup \frac{\tilde{\kappa}_p(d,\epsilon,\mu^n)}{n} = \inf_{\substack{\tilde{\tau}<\tau(d) \\ \mu(B_{\tilde{\tau}})>\mu(B_{\tau(d)})}} \mu(B_{\tilde{\tau}}) = \mu(B_{\tau(d)}). \tag{27}$$

[7]In general, the achievability condition on $\tau > 0$ for the rate ($\mu(B_\tau) = r$) and the distortion ($\sqrt[p]{1 - v_p(B_\tau)} = d$) are not unique.

The first inequality comes from Proposition 2 and the last equality from the fact that the function $\phi_\mu(\tau) \equiv \mu(B_\tau)$ is continuous with respect to $\tau$ as $\mu \ll \lambda$.

To derive a lower bound, let us consider an arbitrary $\tilde{r} \in (0, \mu(B_{\tau(d)}))$. We know that there exists $\tilde{\tau} > \tau(d)$ such that $\mu(B_{\tilde{\tau}}) = \tilde{r}$. Again from Lemma 2, for all $(\bar{k}_n)$ such that $\frac{\bar{k}_n}{n} \to \tilde{r}$, then $\lim_{n\to\infty} \tilde{\sigma}_p(\bar{k}_n, X^n) = \sqrt[p]{1 - v_p(B_{\tilde{\tau}})} = \tilde{d} > d$, $\mathbb{P} - a.s.$. Therefore,

$$\lim_{n\to\infty} \mu^n(\mathcal{A}_d^{n,\bar{k}_n}) = 0. \tag{28}$$

This result implies that eventually in $n$, $\bar{k}_n \le \tilde{\kappa}_p(d, \epsilon, \mu_n)$ and, consequently,

$$\lim_{n\to\infty} \frac{\bar{k}_n}{n} = \mu(B_{\tilde{\tau}}) = \tilde{r} \le \lim\inf_{n\to\infty} \frac{\tilde{\kappa}_p(d, \epsilon, \mu_n)}{n}. \tag{29}$$

On the other hand from (28), we have that it is necessary that $\tilde{r}_p(d, \epsilon, \mathbb{P}) \ge \tilde{r}$. This last inequality and (29) are valid for any $\tilde{r} \in (0, \mu(B_{\tau(d)}))$, then $\mu(B_{\tau(d)}) \le \tilde{r}_p(d, \epsilon, \mathbb{P})$ and $\mu(B_{\tau(d)}) \le \lim\inf_{n\to\infty} \frac{\tilde{\kappa}_p(d,\epsilon,\mu^n)}{n}$, which from (27) proves (18).

Moving to the case where $(x^p)_{x\in\mathbb{R}} \notin L_1(\mu)$, we have that $\mathbb{E}_{X\sim\mu}(|X|^p) = \infty$, then from the *ergodic theorem* [14] $\forall \tau > 0$,

$$\lim_{n\to\infty} \frac{n(X^n, B_\tau)}{n} = \mathbb{E}_{X_1\sim\mu}(1_{B_\tau}(X_1)) = \mu(B_\tau),\ \mathbb{P} - a.s., \tag{30}$$

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} |X_i|^p 1_{B_\tau}(X_i)}{\sum_{i=1}^{n} |X_i|^p} = 1,\ \mathbb{P} - a.s. \tag{31}$$

Equation (31) comes from $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n |X_i|^p = \mathbb{E}_{X\sim\mu}(|X|^p) = \infty$ a.s., and $\lim\sup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n |X_i|^p 1_{B_\tau^c}(X_i) \le \tau$. In other words, (31) means that $\lim_{n\to\infty} \tilde{\sigma}_p(n(X^n, B_\tau), X^n) = 0$, $\mathbb{P} - a.s.$ Furthermore, we have the following:

*Lemma 3:* Let $(X_n)_{n\in\mathbb{N}}$ be a stationary and ergodic process with distribution $\mu \ll \lambda$ and $(x^p)_{x\in\mathbb{R}} \notin L_1(\mu)$. Let us consider an arbitrary $r \in (0, 1]$ and $(k_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} \frac{k_n}{n} = r$, then

$$\lim_{n\to\infty} \tilde{\sigma}_p(k_n, X^n) = 0,\ \mathbb{P} - a.s. \tag{32}$$

(The proof is presented in Section IV-B)

Let fix an arbitrary $r > 0$ and $(k_n)$ where $\lim_{n\to\infty} \frac{k_n}{n} = r$, then from Lemma 3 we have that $\forall d > 0$:

$$\lim_{n\to\infty} \mu^n(\mathcal{A}_d^{n,k_n}) = 1. \tag{33}$$

Then for all $\epsilon \in (0, 1)$ and $d \in (0, 1)$, $\lim\inf_{n\to\infty} \mu^n(\mathcal{A}_d^{n,k_n}) > 1 - \epsilon$ and therefore $\tilde{\kappa}_p(d, \epsilon, \mu_n) \le k_n$ eventually in $n$. From this, $\lim\sup_{n\to\infty} \frac{\tilde{\kappa}_p(d,\epsilon,\mu_n)}{n} \le r$ for all $r > 0$, which concludes the result from Proposition 2. □

### A. Proof of Lemma 2

*Proof:* To begin let us prove the fixed-rate result in (23). It is first important to concentrate in the case when $\frac{k_n}{n} \to 0$ and to show that

$$\lim_{n\to\infty} \tilde{\sigma}_p(k_n, X^n) = 1.\ \mathbb{P} - a.s. \tag{34}$$

For any $\tau > 0$, let us define the sets

$$\mathcal{A}^\tau \equiv \left\{ (x_n)_n \in \mathbb{R}^{\mathbb{N}} : \lim_{n\to\infty} \frac{n(x^n, B_\tau)}{n} = \mu(B_\tau) \right\},$$
$$\mathcal{B}^\tau \equiv \left\{ (x_n)_n \in \mathbb{R}^{\mathbb{N}} : \lim_{n\to\infty} \tilde{\sigma}_p(n(x^n, B_\tau), x^n) = \sqrt[p]{1 - v_p(B_\tau)} \right\}, \tag{35}$$

where from (22b) and (22c), $\mathbb{P}(\mathcal{A}^\tau \cap \mathcal{B}^\tau) = 1$ because

$$\lim_{n\to\infty} \tilde{\sigma}_p(n(X^n, B_\tau), X^n) = \sqrt[p]{1 - v_p(B_\tau)},\ \mathbb{P} - a.s. \tag{36}$$

Let us fix an arbitrary $\delta > 0$ and let $d = 1-\delta$ and $\tau(d)$ be such that $\sqrt[p]{1 - v_p(B_{\tau(d)})} = d$, which implies that $\mu(B_{\tau(d)}) > 0$. Let us take an arbitrary $(x_n)_n \in \mathcal{A}^{\tau(d)} \cap \mathcal{B}^{\tau(d)}$. From the zero-rate assumption on $(k_n)$, the definition of $\mathcal{A}^\tau$ and the fact that $\mu(B_{\tau(d)}) > 0$, $\exists N > 0$ such that $\forall n \ge N$, $k_n < n(x^n, B_{\tau(d)})$, which implies that $\tilde{\sigma}_p(n(x^n, B_{\tau(d)}), x^n) < \tilde{\sigma}_p(k_n, x^n)$. Therefore considering that $(x_n)_n \in \mathcal{B}^{\tau(d)}$,

$$\lim\inf_{n\to\infty} \tilde{\sigma}_p(k_n, x^n) \ge d = 1 - \delta. \tag{37}$$

Performing the same steps for the sequence $d_m = 1 - 1/m$ and defining $\tau_m \equiv \tau(d_m)$ accordingly[8], we have from (37) that $\forall (x_n) \in \bigcap_{m\ge1}(\mathcal{A}^{\tau_m} \cap \mathcal{B}^{\tau_m})$,

$$\lim\inf_{n\to\infty} \tilde{\sigma}_p(k_n, x^n) \ge \sup_{m\ge1} 1 - 1/m = 1. \tag{38}$$

Then from the sigma additivity of $\mathbb{P}$, $\mathbb{P}(\bigcap_{m\ge1} \mathcal{A}^{\tau_m} \cap \mathcal{B}^{\tau_m}) = 1$, which proves the result in (34).

Equipped with this result, for an arbitrary $r \in (0, 1)$ let us consider $(k_n)$ such that $\lim_{n\to\infty} k_n/n = r$. We know that there exists $\tau_o > 0$ such that $r = \mu(B_{\tau_o})$, and for this $\tau_o$ we consider the sets $\mathcal{A}^{\tau_o}$ and $\mathcal{B}^{\tau_o}$ as defined in (35). Then for any $(x_n) \in \mathcal{A}^{\tau_o} \cap \mathcal{B}^{\tau_o}$ it follows that

$$\lim_{n\to\infty} \left| \frac{k_n}{n} - \frac{n(x^n, B_{\tau_o})}{n} \right| = 0. \tag{39}$$

Furthermore, from definition of ordered sequences, it is simple to verify that (see (1)):

$$\begin{aligned} &\left|\tilde{\sigma}_p(k_n, x^n)^p - \tilde{\sigma}_p(n(x^n, B_{\tau_o}), x^n)^p\right| \\ &\quad \le \frac{\zeta_p(|k_n - n(x^n, B_{\tau_o})|, x^n)^p}{||x^n||_{\ell_p}^p} \\ &\quad = 1 - \tilde{\sigma}_p(|k_n - n(x^n, B_{\tau_o})|, x^n)^p \end{aligned} \tag{40}$$

This is where the zero-rate result in (34) is used. In particular, if we consider $(x_n)_{n\in\mathbb{N}} \in (\mathcal{A}^{\tau_o} \cap \mathcal{B}^{\tau_o}) \cap \bigcap_{m\ge1} \mathcal{A}^{\tau_m} \cap \mathcal{B}^{\tau_m}$, from (40), (38) and the fact that $|k_n - n(x^n, B_{\tau_o})|$ is $o(n)$ in (39), we have that $\lim_{n\to\infty} \tilde{\sigma}_p(k_n, x^n) =$

$$\lim_{n\to\infty} \tilde{\sigma}_p(n(x^n, B_{\tau_o}), x^n) = \sqrt[p]{1 - v_p(B_{\tau_o})}, \tag{41}$$

the last equality in (41) from definition of $\mathcal{B}^{\tau_o}$ in (35). Finally from (22b) and (22c), $\mathbb{P}\left( (\mathcal{A}^{\tau_o} \cap \mathcal{B}^{\tau_o}) \cap \bigcap_{m\ge1} \mathcal{A}^{\tau_m} \cap \mathcal{B}^{\tau_m} \right) = 1$ which proves (23).

Concerning the fixed-distortion result in (24), for $d \in (0, 1)$ let $\tau(d) > 0$ be such that $\sqrt[p]{1 - v_p(B_{\tau(d)})} = d$. Let us consider

[8] $\tau_m$ is such that $\sqrt[p]{1 - v_p(B_{\tau_m})} = d_m$, $\forall m \in \mathbb{N}$.

an arbitrary $\bar{\tau} > \tau(d)$ where $\mu(B_{\bar{\tau}}) < \mu(B_{\tau(d)})$ and consequently $v_p(B_{\bar{\tau}}) < v_p(B_{\tau(d)})$ (note that $\mu$ and $v_p$ are mutually absolutely continuous). Again for this $\bar{\tau}$, we use the sets in (35), where for all $(x_n) \in \mathcal{A}^{\bar{\tau}} \cap \mathcal{B}^{\bar{\tau}}$, it follows that $\forall \epsilon > 0$, $\exists N_\epsilon$ such that $\forall n \geq N_\epsilon$, $\left|\tilde{\sigma}_p(n(x^n, B_{\bar{\tau}}), x^n) - \sqrt[p]{1 - v_p(B_{\bar{\tau}})}\right| < \epsilon$. Considering $\epsilon < \frac{\sqrt[p]{1-v_p(B_{\bar{\tau}})}-d}{2}$, it follows by definition in (6) that:

$$n(x^n, B_{\bar{\tau}}) < \kappa_p(d, x^n), \quad \forall n \geq N_\epsilon \tag{42}$$

then $\lim_{n\to\infty} n(x^n, B_{\bar{\tau}})/n = \mu(B_{\bar{\tau}}) \leq \liminf_{n\to\infty} \kappa_p(d, x^n)/n$. We can do the same for the countable collection:

$$\mathcal{C} = \left\{\tilde{\tau}_m \equiv \tau(d) + \frac{1}{m} : m \in \mathbb{N} \text{ such that } \mu(B_{\tilde{\tau}_m}) < \mu(B_{\tau(d)})\right\} \tag{43}$$

where for any $(x_n) \in \bigcap_{\tilde{\tau}_m \in \mathcal{C}} \mathcal{A}^{\tilde{\tau}_m} \cap \mathcal{B}^{\tilde{\tau}_m}$,

$$\mu(B_{\tau(d)}) = \sup_{\tilde{\tau}_m \in \mathcal{C}} \mu(B_{\tilde{\tau}_m}) \leq \lim \inf_{n\to\infty} \kappa_p(d, x^n)/n. \tag{44}$$

The first equality follows from the continuity of the function $f_\mu(\tau) \equiv \mu(B_\tau)$ with respect to $\tau$ as $\mu \ll \lambda$. Then from the fact that $\mathbb{P}(\bigcap_{\tilde{\tau}_m \in \mathcal{C}} \mathcal{A}^{\tilde{\tau}_m} \cap \mathcal{B}^{\tilde{\tau}_m}) = 1$, we have that, $\mu(B_{\tau(d)}) \leq \liminf_{n\to\infty} \kappa_p(d, X^n)/n$, $\mathbb{P}$-a.s. Finally, proving that $\limsup_{n\to\infty} \kappa_p(d, X^n)/n \leq \mu(B_{\tau(d)})$ $\mathbb{P}$-a.s. follows an equivalent symmetric argument and we omit it. □

### B. Proof of Lemma 3

*Proof:* For $r > 0$ let us consider $\bar{r} \in (0, r)$ and $\tau > 0$, such that $\mu(B_\tau) = \bar{r}$ and $(k_n)$ such that $\lim_{n\to\infty} k_n/n = r$. Considering the sets $\mathcal{A}^\tau \equiv \left\{(x_n)_n \in \mathbb{R}^{\mathbb{N}} : \frac{n(x^n, B_\tau)}{n} \to \mu(B_\tau)\right\}$ and $\mathcal{B}^\tau \equiv \left\{(x_n)_n \in \mathbb{R}^{\mathbb{N}} : \tilde{\sigma}_p(n(x^n, B_\tau), x^n) \to 0\right\}$, we have that $\mathbb{P}(\mathcal{A}^\tau \cap \mathcal{B}^\tau) = 1$ from (30) and (31). Let us fix an arbitrary $(x_n)_{n\in\mathbb{N}} \in \mathcal{A}^\tau \cap \mathcal{B}^\tau$. Considering that $\bar{r} < r$, then $n(x^n, B_\tau) < k_n$ eventually in $n$, and therefore $\tilde{\sigma}_p(n(x^n, B_\tau), X^n) \geq \tilde{\sigma}_p(k_n, X^n)$ eventually, which implies from definition of $\mathcal{B}^\tau$ that $\lim_{n\to\infty} \tilde{\sigma}_p(k_n, x^n) = 0$. Finally, the fact the event $\mathcal{A}^\tau \cap \mathcal{B}^\tau$ happens $\mathbb{P}$-almost surely concludes the result. □

## V. Properties of the Rate-Approximation Error Curve for Non $\ell_p$-Compressible Processes

For the family of non $\ell_p$-compressible ergodic processes, in this section we study two tail functions that characterize the achievable rate-distortion region in (21). Let $(X_n)_{n\in\mathbb{N}}$ be stationary and ergodic with $\mu \in \mathcal{P}(\mathbb{R})$ its invariant probability measure. Here we focus on the case where $(x^p)_{x\in\mathbb{R}} \in L_1(\mu)$, then the measure $v_p \in \mathcal{P}(\mathbb{R})$ in (17) is well-defined and by construction $v_p \ll \mu$, where the *Radon-Nikodym* (RN) derivative (or density) of $v_p$ with respect to $\mu$ is given by $\frac{dv_p}{d\mu}(x) = \frac{|x|^p}{\int_{\mathbb{R}} |x|^p d\mu(x)}$ for all $x$. Furthermore, from the strict positivity of $x^p$ on $\mathbb{R}\setminus\{0\}$, it is clear that $\mu \ll v_p$, where $\frac{d\mu}{dv_p}(x) = \frac{dv_p}{d\mu}(x)^{-1}$ and then these two measures a mutually absolutely continuous [14], i.e.,

$$\forall p > 0, \forall B \in \mathcal{B}(\mathbb{R}),\ \mu(B) = 0, \text{ if an only if, } v_p(B) = 0. \tag{45}$$

This implies a close interplay between the tail-probability functions

$$\phi_\mu(\tau) \equiv \mu(B_\tau) \text{ and } \phi_{v_p}(\tau) \equiv v_p(B_\tau), \tag{46}$$

that characterize $\mathcal{R}_\mu = \left\{\left(\phi_\mu(\tau), \sqrt[p]{1 - \phi_{v_p}(\tau)}\right) : \tau \geq 0\right\}$ in (21). The following basic properties can be stated:

*Proposition 5:* [9]

a) $\phi_\mu(\tau)$ and $\phi_{v_p}(\tau)$ are left-continuous.
b) $\lim_{n\to\infty} \phi_\mu(\tau + 1/n) = \phi_\mu(\tau) - \mu(\{\tau\} \cup \{-\tau\})$. Then, $\phi_\mu(\tau)$ is continuous at $\tau \geq 0$, if and only if, $\mu(\{\tau\} \cup \{-\tau\}) = 0$.
c) For any $\tau \geq 0$, $\phi_\mu(\tau)$ is continuous at $\tau$, if and only if, $\phi_{v_p}(\tau)$ is continuous at $\tau$.
d) For any pair $\tau > \tilde{\tau} \geq 0$, $\phi_\mu(\tau) < \phi_\mu(\tilde{\tau})$, if and only if, $\phi_{v_p}(\tau) < \phi_{v_p}(\tilde{\tau})$.
e) $\phi_\mu(0) = \phi_{v_p}(0) = 1$ and $\lim_{\tau\to\infty} \phi_\mu(\tau) = \lim_{\tau\to\infty} \phi_{v_p}(\tau) = 0$.

From these properties we can state the following:

*Theorem 2:* Assuming that $\mu \ll \lambda$, then for any $\epsilon > 0$:

1) $\tilde{r}_p(d, \epsilon, \mathbb{P})$ is a continuous function with respect to $d \in [0, 1]$.
2) $\forall d \in [0, 1]$ there is $\tau \geq 0$ such that $\sqrt[p]{1 - \phi_{v_p}(\tau)} = d$ and $\forall r \in [0, 1]$ there exists $\tau \geq 0$ such that $\phi_{v_p}(\tau) = r$. Then, the collection of critical rates $\{\tilde{r}_p(d, \epsilon, \mathbb{P}) : d \in [0, 1]\}$ achieves all the values in $[0, 1]$, and any distortion $d \in [0, 1]$ is achieved in $\mathcal{R}_\mu$ with a given rate.
3) For $0 \leq d < \bar{d} \leq 1$, then $\tilde{r}_p(d, \epsilon, \mathbb{P}) > \tilde{r}_p(\bar{d}, \epsilon, \mathbb{P})$.
4) $\tilde{r}_p(d, \epsilon, \mathbb{P})$ is a differentiable function in $(0, 1)$, and

$$\frac{\partial \tilde{r}_p(d, \epsilon, \mathbb{P})}{\partial d} = -\frac{p\, ||x^p||_{L_1(\mu)}}{d} \cdot \left(\frac{d}{\tilde{\phi}_{v_p}^{-1}(d)}\right)^p, \ \forall d \in (0, 1), \tag{47}$$

where $||x^p||_{L_1(\mu)} \equiv \int_{\mathbb{R}} |x|^p \, d\mu(x)$ and $\tilde{\phi}_{v_p}^{-1}(d)$ denotes the inverse of the auxiliary function $\tilde{\phi}_{v_p}(\tau) \equiv \sqrt[p]{1 - \phi_{v_p}(\tau)}$ for $\tau \geq 0$.

The proof is presented in Appendix I.

In summary, the rate-distortion approximation function $(\tilde{r}_p(d, \epsilon, \mathbb{P}))_{d\in[0,1]}$ is continuous, injective and achieves all the rates, in the sense that for any $r \in [0, 1]$ there is only one $d \in [0, 1]$ such that $\tilde{r}_p(d, \epsilon, \mathbb{P}) = r$. In addition, it is strictly decreasing and satisfies the following boundary conditions: $\tilde{r}_p(0, \epsilon, \mathbb{P}) = 1$ and $\tilde{r}_p(1, \epsilon, \mathbb{P}) = 0$. Furthermore, it is simple to verify that $\lim_{d\to 0} \tilde{\phi}_{v_p}^{-1}(d) = 0$ and $\lim_{d\to 1} \tilde{\phi}_{v_p}^{-1}(d) = \infty$. Then from (47), it seems that $\frac{\partial \tilde{r}_p(d,\epsilon,\mathbb{P})}{\partial d}$ should be a non-increasing function as $d$ progress to 1 (considering that $\lim_{d\to 1} \frac{\partial \tilde{r}_p(d,\epsilon,\mathbb{P})}{\partial d} = 0$), and consequently, $\tilde{r}_p(d, \epsilon, \mathbb{P})$ should present a convex dominating behavior[10] eventually as $d$ progresses to 1. The next section analyzes the convexity of $\tilde{r}_p(d, \epsilon, \mathbb{P})$ more formally.

[9]The arguments are simple to verify and they are omitted for the space constraint.

[10]Note that for $p \in (0, 1]$ the fact that $\frac{\partial \tilde{r}_p(d,\epsilon,\mathbb{P})}{\partial d}$ is non-increasing is guaranteed from (47). On the other hand, from this analysis, when $p > 1$, it is not absolutely clear that $\left(\frac{d}{\tilde{\phi}_{v_p}^{-1}(d)}\right)^p$ is non-decreasing in the whole range $d \in (0, 1)$, however it goes to zero as $d$ approaches 1 from below. This is formally analyzed in Theorem 3.

### A. Convexity of $\tilde{r}_p(d,\epsilon,\mathbb{P})$

First in this section we show that $\tilde{r}_p(d,\epsilon,\mathbb{P})$ is a convex function of $D \equiv d^p$. Then, we provide a necessary and sufficient condition for $\tilde{r}_p(d,\epsilon,\mathbb{P})$ to be a convex function of $d$.

*Proposition 6:* For every $p>0$, $\tilde{r}_p(d,\epsilon,\mathbb{P})$ is a convex function of $D \equiv d^p$ over $(0,1)$.

(The proof is presented in Appendix IV-D)

*Theorem 3:* Let $(X_n)$ by a non $\ell_p$-compressible stationary and ergodic sequence equipped with $\mu \ll \lambda$. Then

1) if $p \le 1$, $\tilde{r}_p(d,\epsilon,\mathbb{P})$ is a convex function for all $d \in (0,1)$.
2) Otherwise, $\tilde{r}_p(d,\epsilon,\mathbb{P})$ is convex, if and only if,

$$\frac{\tau^{p+1} f_Z(\tau)}{\mathcal{D}(\tau)} \le \frac{p^2}{p-1}, \quad \forall \tau > 0, \tag{48}$$

where $f_Z$ is the pdf of $Z=|X|$, with $X \sim \mu$, and $\mathcal{D}(\tau) \equiv \int_0^\tau z^p f(z)dz$.

The proof is presented in Appendix II.

*Remark 1:* The condition in (48) is implicit and may be difficult to verify. A simple to check sufficient condition for the convexity of $\tilde{r}_p(d,\epsilon,\mathbb{P})$ is the following:[11]

$$\int_0^\tau f'_Z(z) z^{p+1} dz \le 0, \quad \forall \tau > 0. \tag{49}$$

where $f'(z) = \frac{d}{dz} f_Z(z)$. In particular, a non increasing pdf (i.e., $f'_Z(z) \le 0$ almost everywhere in $\mathbb{R}^+$) characterizes a convex rate-approximation error function.

### B. Examples of Heavy Tail and Exponentially Decaying Tail Distributions

We present few examples of rate-approximation error curves of non $\ell_p$-compressible i.i.d. processes. In particular following [1], we consider the Gaussian (exponentially decaying distribution), which is non $\ell_p$-compressible for any $p>0$, from Corollary 1, and the family of Student's $t$-distribution with parameter $q>0$[12], whose pdf goes to cero as $O(|x|^{-(q+1)})$. From Corollary 2, the i.i.d. process with a Student's t-distribution ($q>0$) is $\ell_p$-compressible for any $p \ge q$ and non-$\ell_p$-compressible for $p<q$.

To compute the rate-distortion function $(\tilde{r}_p(d,\epsilon,\mathbb{P}))_{d\in[0,1]}$, we use the fact that $\mathcal{R}_\mu = \{(\tilde{r}_p(d,\epsilon,\mathbb{P}),d) : d \in [0,1]\} =$

$$\left\{ \left( \mu(B_\tau), \sqrt[p]{1-v_p(B_\tau)} \right) \in [0,1]^2 : \tau \ge 0 \right\}.$$

Then the problem reduces to compute $\phi_\mu(\tau) = \mu(B_\tau)$ and $\phi_{v_p}(\tau) = v_p(B_\tau)$, for all $\tau \ge 0$. For this we consider an estimation approach. Considering a sufficiently large set of i.i.d. realizations of $\mu$ (let say $X_1,..,X_n$), the law of large numbers [14] tells us that for any $\tau \ge 0$ and $p>0$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{B_\tau}(X_i) = \phi_\mu(\tau) \tag{50}$$

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} \mathbf{1}_{B_\tau}(X_i)\cdot|X_i|^p}{\sum_{i=1}^{n}(X_i)\cdot|X_i|^p} = \frac{\int_{B_\tau}|x|^p\, d\mu(x)}{\int_{\mathbb{R}}|x|^p\, d\mu(x)} = \phi_{v_p}(\tau) \tag{51}$$

[11]The proof is omitted for the sake of space.

[12]The pdf of a Student's $t$-distribution with $q$ degrees of freedom is given by $f(x) = \frac{\Gamma((q+2)/2)}{\sqrt{q\pi}}(1+x^2/q)^{-\frac{q+1}{2}}$, where $\Gamma(\cdot)$ is the gamma function.

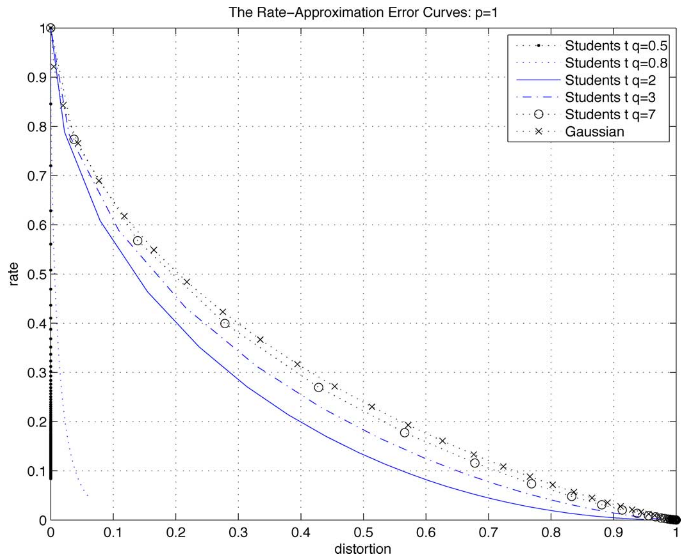


Fig. 2. Numerically estimated rate approximation error curves for several non $\ell_1$-compressible i.i.d. processes and one $\ell_1$-compressible i.i.d. process.

with probability one, assuming that $\int |x|^p\, d\mu(x) < \infty$ and the use of (17). Then, by sampling the space of thresholds $\tau \in [0,\infty)$ and considering a sufficiently large $n$, we can estimate with an arbitrary good precision the rate distortion region $\mathcal{R}_\mu$. Following this path, Fig. 2 shows the estimated rate-approximation error curves for the Gaussian, and several Student's $t$-distribution for $p=1$. We verify that some Student's $t$-distribution are $\ell_1$-compressible (cases $q=0.5$ and $q=0.8$) and others are non $\ell_1$-compressible (cases $q=2$, $q=3$ and $q=7$) as the Theorem 1 predicts. More interesting is to validate in all the cases of non compressible priors, that the curves have a convex behavior, which is justified from (49). Furthermore, the density with the exponentially decaying tail is less compressible than any prior with a power law decay, in the sense that for achieving a distortion $d \in (0,1)$ the Gaussian i.i.d. process needs a higher rate. From these curves, as $q$ goes to infinity the i.i.d. process with a heavy-tail distribution approaches the approximation error behavior of the Gaussian law.

Fig. 3 shows the rate-approximation error curves for the Gaussian prior for different values of $p>0$. It is interesting to observe the increasing monotonic behavior of $\tilde{r}_p(d,\epsilon,\mathbb{P})$ as $p$ increases, for any fixed value of $d>0$. Again all curves have a convex behavior. To contrast, Fig. 4 shows a set of curves for the Cauchy distribution (i.e., Student's $t$ distribution with $q=1$), where no clear monotonic pattern is observed as a function of $p$.

## VI. $\ell_1$-Compressibility and Compressed Sensing

We conclude this work analyzing compressible stationary and ergodic sequences, as characterized in Theorem 1, in terms of their ability to be represented with an arbitrary small proportion of linear measurements adopting for that the classical compressed sensing (CS) measurement and reconstruction setting. In particular, the focus is on $\ell_1$-compressible processes, as the standard Gaussian i.i.d. linear acquisition and $\ell_1$-minimization (sparsity promoting) decoder of CS [15], [16] offer a well-known $\ell_1$-instance optimality guarantee [3] (stated below in Lemma 4) that matches the modeling assumption of $\ell_1$-compressible processes.

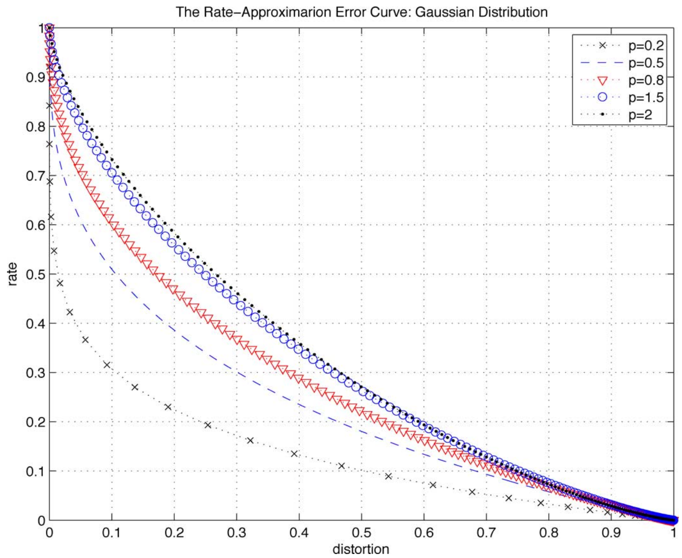


Fig. 3. Numerically estimated rate approximation error curves for the Gaussian i.i.d. process considering different $\ell_p$-norms.

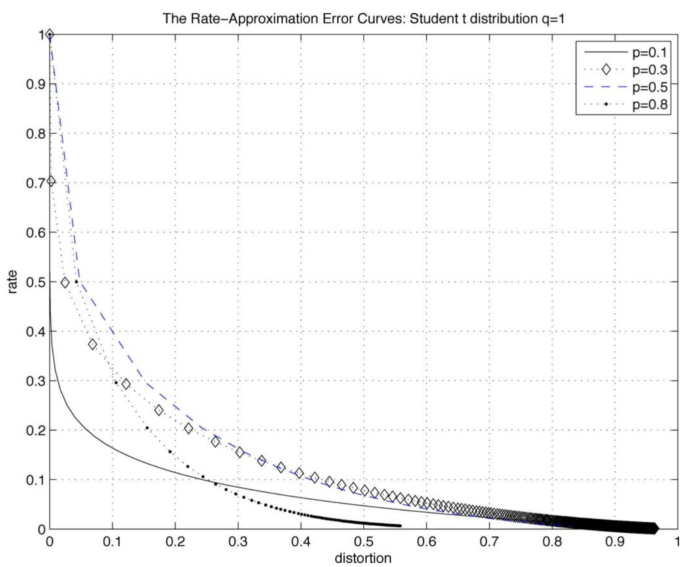


Fig. 4. Numerically estimated rate approximation error curves for the Cauchy (Student's $t$-distribution with $q = 1$) i.i.d. process considering different $\ell_p$-norms in the regime where the process is non $\ell_p$-compressible (i.e., $p < 1$ from Corollary 2).

### A. Compressed Sensing in a Nutshell

In the finite dimensional setting, the analysis phase of the CS is a linear operator $\phi(w) : \mathbb{R}^n \to \mathbb{R}^m$, that given a signal $x \in \mathbb{R}^n$ generates a measurement vector $y = \phi x \in \mathbb{R}^m$. The case of interest is on the under-sampled regime, i.e., $m < n$, where under sparse or compressible assumptions on $x$, CS can offer perfect or near-optimal reconstruction by the solution of the following (linear programing) problem [16]:

$$\Delta^*(y) = \arg \min_{\{\tilde{x} \in \mathbb{R}^n : y = \phi \tilde{x}\}} ||\tilde{x}||_{\ell_1}. \tag{52}$$

Notably, the CS theory, based on the restricted isometry property (RIP), establishes sufficient conditions over $\phi$ (and implicitly over the number of measurements $m$) in order that $x = \Delta^*(\phi x)$, when $x \in \Sigma_k$ for some $k < n$. The next result, in its original form stated in [9], shows that random measurements offer a solution to that problem with a near optimal relationship between $m$ and $k$ [3][13].

*Lemma 4:* ([15, Th. 5.2] and [16, Th. 1.2]) Let $\phi(w)$ be a random matrix[14], $w \in \Omega^{mn}$, whose entries are driven by i.i.d realizations of a Gaussian distribution $\mathcal{N}(0, 1/m)$ or a binary variable with uniform distribution over $\{1/\sqrt{m}, -1/\sqrt{m}\}$. For any arbitrary $k \leq n$ and $x \in \mathbb{R}^n$, we have that:

$$\left|\left|x - x^\ddagger\right|\right|_{\ell_1} \leq C_0 \cdot \sigma_1(k, x) \tag{53}$$

if

$$m \geq C_1 k \log \frac{n}{k}, \tag{54}$$

with a probability, over the sensing sampling space $\Omega^{mn}$, at least equal to $1 - 2^{-C_2 \cdot m}$. Here $x^\ddagger = \Delta^*(\phi(w)x)$ is the solution of (52) and, $C_0$, $C_1$ and $C_2$ are positive universal constants independent of $n$ and $k$.[15]

(The proof of this result derives directly from [15, Th. 5.2] and [16, Th. 1.2])

### B. Zero-Rate Reconstruction for $\ell_1$-Compressible Processes

Here we formalize the reconstruction of infinite sequences $(x_n)_{n \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ using CS. For this, we consider a finite-length (or fixed-rate) approach, where the idea is to analyze consecutive finite-block versions of the sequence, i.e., to sense and reconstruct $x^n = (x_1, .., x_n)$ for any $n \geq 1$, and study reconstruction performances in the limit when the block-length tends to infinity.

More precisely, let $(m_n)_{n \geq 1}$ be a sequence of positive integers such that $1 \leq m_n \leq n$. From this sequence, we consider the family of Gaussian CS encoding-decoding pairs $\{(\phi_{m_n \times n}(w), \Delta_n^*(\cdot)) : n > 0\}$ where for any $n > 0$, $\phi_{m_n \times n}(w)$ is the random sensing matrix of $m_n \times n$ generated by i.i.d. entries as mentioned in Lemma 4, and $\Delta_n^*(\cdot)$ is the function from $\mathbb{R}^{m_n}$ to $\mathbb{R}^n$ that solves the $\ell_1$-minimization problem in (52). Given a sequence $(x_n)_{n \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ and any finite block-length $n > 0$, we can apply the CS approach over $x^n$ to recover $\hat{x}^n(w) = \Delta_n^*(\phi_{m_n \times n}(w) x^n)$, which is a random reconstruction function of the matrix $\phi_{m_n \times n}(w)$. In relation with the $\ell_1$-relative approximation error introduced in (3), we consider as a fidelity indicator the $\ell_1$-noise to signal ratio (NSR) given by: $D_{\ell_1}((\phi_{m_n \times n}(w), \Delta_n^*); x^n) \equiv$

$$\frac{||\hat{x}^n(w) - x^n||_{\ell_1}}{||x^n||_{\ell_1}} \in [0, 1], \;\; \forall n \geq 1. \tag{55}$$

More generally, if we have a process $(X_n)_{n \in \mathbb{N}}$ with distribution $\mathbb{P} = \{\mu^n : n \geq 1\}$ and a sequence of lengths $(m_n)_{n \geq 1}$ with its associated Gaussian CS finite-block scheme $\{(\phi_{m_n \times n}(w), \Delta_n^*(\cdot)) : n > 0\}$, we can also analyze the finite-block performance of the scheme by the object $D_{\ell_1}((\phi_{m_n \times n}(w), \Delta_n^*); X^n)$, which is a random variable function of two independent random objects: the vector $X^n \sim \mu^n$ and the random matrix $\phi_{m_n \times n}(w)$. Therefore, it is important

[13]A systematic and lucid exposition of this CS theory can be found in the work of Candes [16], Baraniuk *et al.* [15] and Cohen *et al.* [3].

[14]This result can be generalized to random matrices satisfying a concentration inequality, which is not reported here for space considerations. See more details in [15].

[15]Refined results can be found in [3], [16], [20].

to consider the average NSR with respect to the statistics of the source (i.e., $\mu^n$) by, $D_{\ell_1}\left((\phi_{m_n \times n}(w), \Delta_n^*); \mu^n\right) \equiv$

$$\mathbb{E}_{X^n \sim \mu^n}\left\{D_{\ell_1}\left((\phi_{m_n \times n}(w), \Delta_n^*); X^n\right)\right\} \in [0,1]. \quad (56)$$

Then the question we focus here is: for an $\ell_1$-compressible process that satisfies (16), what is the minimum rate of measurements $r \in (0,1)$ (i.e., $m_n = rn$, or more generally $\lim_{n\to\infty} \frac{m_n}{n} = r$) of the classical Gaussian CS scheme that ensures that:

$$\lim_{n\to\infty} D_{\ell_1}\left((\phi_{m_n \times n}(w), \Delta_n^*); \mu^n\right) = 0, \quad (57)$$

with probability one with respect to the statistics of the sequence of random matrices $\{\phi_{m_n \times n}(w) : n \geq 1\}$?

From Theorem 1 and the RIP-based $\ell_1$-instance optimality result of the Gaussian CS setting in Lemma 4, we can state the following result:

*Theorem 4:* Let $(X_n)_{n\in\mathbb{N}}$ be a stationary ergodic process. If $(X_n)_{n\in\mathbb{N}}$ is $\ell_1$-compressible, then for any sequence $(m_n)$ such that $\lim_{n\to\infty} \frac{m_n}{n} > 0$, it follows that

$$\lim_{n\to\infty} D_{\ell_1}\left((\phi_{m_n \times n}(w), \Delta_n^*); \mu^n\right) = 0 \text{ and}$$
$$\lim_{n\to\infty} D_{\ell_1}\left((\phi_{m_n \times n}(w), \Delta_n^*); X^n\right) = 0, \quad (58)$$

almost-surely with respect to the statistics of $\{\phi_{m_n \times n}(w) : n \geq 1\}$ and the joint statistics of $(X_n)_{n\in\mathbb{N}}$ and $\{\phi_{m_n \times n}(w) : n \geq 1\}$, respectively.

The proof is presented in Appendix III.

### C. Discussion and Interpretation of Theorem 4

1: This result states that in order to achieve zero distortion in the reconstruction for an $\ell_1$-compressible process, almost-surely in the NSR sense of (56), the CS scheme needs an arbitrary small number of measurements per sample. In other words, under the $\ell_1$-compressibility model assumption for the process, the minimum rate to achieve zero distortion is zero for the Gaussian CS scheme. Then, it is remarkable to validate that CS is able to achieve the same zero critical rate that it is obtained by the analysis of the pure *oracle* best-$k$-term approximation error of $\ell_1$-compressible process (see the result in Lemma 3).

2: This result shows that the lucid notion of $\ell_1$-compressibility proposed by Amini *et al.* [1][16] really translates in a meaningful performance result for the classical Gaussian CS (GCS) setting in the asymptotic regime when the block-length goes to infinity. In other words, we can say that $\ell_1$-compressibility, meaning a sort of zero-rate of innovation in the process, implies zero-rate of measurements (per signal dimension) for perfect recovery (in the sense of NSR distortion) for the CS scheme. This result closes a gap not explored in [1] between their notion of $\ell_1$-compressibility and CS performance guarantee in the asymptotic regime.

3: Concerning compressibility of random sequences and CS performance guarantee, we want to highlight the work of Gribonval *et al.* [2] for the case of i.i.d. processes. They show in [2, Theorem 2] that if $\mathbb{E}_{X\sim\mu}(|X|^2) = \infty$ then $\lim_{n\to\infty} D_{\ell_2}\left((\phi_{m_n \times n}(w), \Delta_n^*); X^n\right) = 0$ (for distribution equipped with a pdf) and they enunciate a version of Theorem 4 for the i.i.d. case [2, Remark 1]. Then, we want to give credit to this contribution to be the first result that offers a connection between notions of compressibility for i.i.d. processes (based on relative approximation errors) and the performance (in the asymptotic regime) of the classical GCS scheme. In this context, Theorem 4 can be seen as an extension of these results to the case of stationary and ergodic sequences and, in the technical side, an extension on the use of the $\ell_1$-instance optimality property of the $\ell_1$-minimization decoder[17]. On the other hand, focusing on the i.i.d. context, Theorems 4 and 1 offer a way to verify that Amini *et al.* [1] $\ell_1$-compressible notion (variable rate in nature) has a connection with the results in Gribonval *et al.* [2, Th. 2 and Rem. 1] in terms of what GCS can achieves for the case of $\ell_1$-compressible processes.

## VII. Discussion and Final Remarks

The main result of this work (Theorem 1) provides a connection between Gribonval *et al.* [2, Prop. 1] almost sure convergence result of relative approximation errors, and Amini *et al.* [1, Def. 6] notion of $\ell_p$-compressibility for random sequences. More importantly, Theorem 1 offers new techniques to extend that connection (and, consequently, a dichotomy between being and non-being $\ell_p$-compressible random sequences) to the family of stationary and ergodic processes. This extension is constructed over the almost sure convergence of the empirical distributions of $\mu$ and $v_p$, respectively (see definitions in the statement of Theorem 1) to the true probabilities on the family of tail events $\{B_\tau : \tau \geq 0\}$ (details in Section IV). The idea of looking at specific empirical measures as the basic object of interest, in (22b) and (22c), instead of the statistics of the sum of the ordered sequence as considered in [2, Prop. 1] and [1], was essential to extend the analysis from the i.i.d. case to the case of stationary ergodic processes.

Finally, one can notice from the proof of Theorem 1 that this result does not rely on a stationary property, as it is essentially based on an almost sure convergence (asymptotic in nature) over the family of indicator functions of the tail events $\{B_\tau : \tau \geq 0\}$ in (19). Then, we conjecture that the analysis of compressible priors can be extended over a family of random sequences with a specific ergodic property over the tail events, which is an interesting direction for future work. This observation leads us to put the attention on the general theory of (non-stationary) processes with ergodic properties [14], [18], [22]–[24].

## Appendix I
## Proof of Theorem 2

*Proof:* We first verify the second point to then move to the rest of the points.

*Point 2):* (Achievability of all rates and distortions in $\mathcal{R}_\mu$): Using Proposition 5 b) and c) and the hypothesis that $\mu \ll \lambda$, it follows that $\phi_\mu(\tau)$ and $\phi_{v_p}(\tau)$ are continuous functions in $[0,\infty)$. Then, adopting Proposition 5 e), we have that $\mathcal{R}_\mu$ achieves all the rates and distortions in the range $[0,1]$.

*Point 1):* (Continuity of $\tilde{r}_p(d,\epsilon,\mathbb{P})$):

First, it is important to verify that the implicit characterization of $\tilde{r}_p(d,\epsilon,\mathbb{P})$ presented in (18) and (20) offers a well-defined

[16]This notion was motivated by the performance guarantees result of CS for finite dimensional sparse and compressible signals.

[17]The argument in the proof of [2, Theorem 2] uses the $\ell_2$-instance optimality in probability of the $\ell_1$-minimization decoder [21].

function. By contradiction, let assume that such characterization is not a function in the sense that for a given distortion $d \in [0,1]$ there are two values $\tau_1 > \tau_2 \geq 0$ solution of (20) associated to two different rates $\phi_\mu(\tau_1) < \phi_\mu(\tau_1)$. This last condition implies that $\phi_{v_p}(\tau_1) < \phi_{v_p}(\tau_1)$ from (45), which contradicts the fact that $\tau_1$ and $\tau_2$ are solutions of (20).

Moving to the continuity, let us fix an arbitrary $d \in (0,1)$ and $\epsilon > 0$. Then by the achievability of the distortions, $\exists \tau_o > 0$ such that $d = \sqrt[p]{1-\phi_{v_p}(\tau_o)}$. On the other hand, by the achievability of the rates, there exists $\tau_1 > \tau_o > \tau_2$, where $\phi_\mu(\tau_1) = \phi_\mu(\tau_o) - \epsilon$ and $\phi_\mu(\tau_2) = \phi_\mu(\tau_o) + \epsilon$ (without loss of generality we assume that $\epsilon < \phi_\mu(\tau_o) = r_o$). Then from monotonicity of $\phi_\mu(\cdot)$ we have that for any $\tau \in (\tau_2, \tau_1)$, $\phi_\mu(\tau) \in B_\epsilon(\phi_\mu(\tau_o))$.[18] At this point, we can obtain the distortions $d_1 = \sqrt[p]{1-\phi_{v_p}(\tau_1)} > d > d_2 = \sqrt[p]{1-\phi_{v_p}(\tau_2)}$, where the strict inequalities that relate them follow from the fact that by construction $\phi_\mu(\tau_1) < \phi_\mu(\tau_o) < \phi_\mu(\tau_2)$ and (45). Then, we can define $\delta = \min\{d - d_1, d_2 - d\}$, where for any $\tilde{d} \in B_\delta(d) \subset (d_2, d_1)$ we have (by the monotonicity of the function $\phi_{v_p}(\cdot)$) that there exists $\tilde{\tau} \in (\tau_2, \tau_1)$ such that $\tilde{d} = \sqrt[p]{1-\phi_{v_p}(\tilde{\tau})}$, and consequently $\tilde{r}_p(\tilde{d}, \epsilon, \mathbb{P}) = \phi_\mu(\tilde{\tau}) \in B_\epsilon(\phi_\mu(\tau_o)) = B_\epsilon(\tilde{r}_p(d, \epsilon, \mathbb{P}))$, the last set of equalities from (18) and (20). As $d \in (0,1)$ and $\epsilon > 0$ are arbitrary numbers, this proves the continuity of $(\tilde{r}_p(d, \epsilon, \mathbb{P}))_{d \in [0,1]}$.

*Point 3):* (Strict monotonicity of $\tilde{r}_p(d, \epsilon, \mathbb{P})$):

Let fix $d_1 < d_2 \in [0,1]$. By definition (13), we have that $\tilde{r}_p(d_1, \epsilon, \mathbb{P}) \geq \tilde{r}_p(d_2, \epsilon, \mathbb{P})$. Furthermore, from the characterization given in (18) and (20), we have that there exists $\tau_1 > \tau_2$ such that $d_1 = \sqrt[p]{1-\phi_{v_p}(\tau_1)}$ and $d_2 = \sqrt[p]{1-\phi_{v_p}(\tau_2)}$. This implies that $\phi_{v_p}(\tau_1) > \phi_{v_p}(\tau_2)$, and consequently from (45) we have that $\tilde{r}_p(d_1, \epsilon, \mathbb{P}) = \phi_\mu(\tau_1) > \phi_\mu(\tau_2) = \tilde{r}_p(d_2, \epsilon, \mathbb{P})$.

*Point 4:* (Differentiability of $\tilde{r}_p(d, \epsilon, \mathbb{P})$):

First, it is clear that both functions $\phi_\mu(\tau)$ and $\phi_{v_p}(\tau)$ are differentiable by construction. In fact from (46), $\forall \tau \geq 0$, $\frac{\partial \phi_\mu}{\partial \tau}(\tau) =$

$$-(f_\mu(\tau) + f_\mu(-\tau)) \text{ and } \frac{\partial \phi_{v_p}}{\partial \tau}(\tau) = -(f_{v_p}(\tau) + f_{v_p}(-\tau)), \tag{59}$$

where $f_\mu(\tau) \equiv \frac{d\mu}{d\lambda}(\tau)$ and $f_{v_p}(\tau) \equiv \frac{dv_p}{d\lambda}(\tau) = \frac{|\tau|^p}{\int_{\mathbb{R}} |x|^p d\mu(x)} f_\mu(\tau)$ denote the pdf of $\mu$ and $v_p$, respectively. Furthermore, we can introduce the auxiliary function $\tilde{\phi}_{v_p}(\tau) = \sqrt[p]{1-\phi_{v_p}(\tau)}$ that is differentiable, and

$$\frac{\partial \tilde{\phi}_{v_p}}{\partial \tau}(\tau) = \frac{|\tau|^p \left(1-\phi_{v_p}(\tau)\right)^{1/p-1}}{p} \cdot \frac{f_\mu(\tau) + f_\mu(-\tau)}{\int_{\mathbb{R}} |x|^p \, d\mu(x)}, \ \forall \tau > 0. \tag{60}$$

Finally, for a fix $d \in (0,1)$ there exists $\tau_o > 0$ such that $d = \sqrt[p]{1-\phi_{v_p}(\tau_o)}$ and, consequently,[19]

$$\begin{aligned} \frac{\partial \tilde{r}_p(d, \epsilon, \mathbb{P})}{\partial d} &= \frac{\partial \phi_\mu}{\partial \tau}(\tau_o) \cdot \left( \frac{\partial \tilde{\phi}_{v_p}}{\partial \tau}(\tau_o) \right)^{-1} \\ &= -\frac{p \, ||x^p||_{L_1(\mu)}}{|\tau_o|^p \left(1-\phi_{v_p}(\tau_o)\right)^{1/p-1}} \\ &= -\frac{p \, ||x^p||_{L_1(\mu)}}{d} \cdot \left( \frac{d}{\tilde{\phi}_{v_p}^{-1}(d)} \right)^p, \end{aligned} \tag{61}$$

the first equality by the characterization of $\tilde{r}_p(d, \epsilon, \mathbb{P})$ in (18) and (20), and the third using that $(1-\phi_{v_p}(\tau_o))^{1/p-1} = d^{1-p}$ and $\tau_o = \tilde{\phi}_{v_p}^{-1}(d)$. □

## APPENDIX II
## PROOF OF THEOREM 3

*Proof:* Considering that $Z = |X|$, let $f(\cdot)$ be its pdf. In view of Theorem 1, for a fixed $d \geq 0$, there exists $\tau(d) > 0$ such that $\tilde{r}_p(d, \epsilon, \mu^n)$ can be expressed as:

$$\tilde{r}_p(d, \epsilon, \mathbb{P}) = \phi_\mu(\tau(d)) \tag{62}$$

where $d = \tilde{\phi}_{v_p}(\tau(d)) = \frac{\mathcal{D}(\tau(d))^{1/p}}{||x^p||_{L_1(\mu)}}$, for which we introduce the short-hands $\tilde{\phi}_{v_p}(\tau) \equiv \sqrt[p]{1-\phi_{v_p}(\tau)}$ and $\mathcal{D}(\tau) \equiv \int_0^\tau z^p f(z) dz$. Then using (61),

$$\begin{aligned} \frac{\partial \tilde{r}_p(d, \epsilon, \mathbb{P})}{\partial d} &= -\,||x^p||_{L_1(\mu)} \cdot p \frac{\mathcal{D}(\tau(d))^{1-\frac{1}{p}}}{\tau(d)^p} \\ &= -||x^p||_{L_1(\mu)} \cdot p \left[ \left( \frac{\mathcal{D}(\tau(d))}{\tau(d)^{p+1}} \right)^{p-1} \tau(d)^{-1} \right]^{1/p} \end{aligned} \tag{63}$$

By construction $\mathcal{D}(\tau)$ is non-decreasing with $\tau$. Hence, when $p \leq 1$, the middle term in (63) is negative and increases with $\tau$, proving the convexity for that case. For the case $p > 1$, from the right hand side of (63) convexity will hold, if and only if, $\frac{d}{d\tau} \left[ \left( \frac{\mathcal{D}(\tau)}{\tau^{p+1}} \right)^{p-1} \tau^{-1} \right] < 0$ for all $\tau > 0$. To check that, it is useful first to note that

$$\begin{aligned} \frac{d}{d\tau} \left( \frac{\mathcal{D}(\tau)}{\tau^{p+1}} \right) &= \tau^p f(\tau) \tau^{-(p+1)} - (p+1) \tau^{-(p+2)} \mathcal{D}(\tau) \\ &= (p+1) \tau^{-(p+2)} \left( \frac{\tau^{p+1}}{p+1} f(\tau) - \mathcal{D}(\tau) \right). \end{aligned} \tag{64}$$

Then $\frac{d}{d\tau} \left[ \left( \frac{\mathcal{D}(\tau)}{\tau^{p+1}} \right)^{p-1} \tau^{-1} \right]$

$$\begin{aligned} &= (p-1) \left( \frac{\mathcal{D}(\tau)}{\tau^{p+1}} \right)^{p-2} \tau^{-1} \frac{d}{d\tau} \left( \frac{\mathcal{D}(\tau)}{\tau^{p+1}} \right) - \left( \frac{\mathcal{D}(\tau)}{\tau^{p+1}} \right)^{p-1} \tau^{-2} \\ &= \tau^{-1} \left( \frac{\mathcal{D}(\tau)}{\tau^{p+1}} \right)^{p-1} \left[ (p-1) \frac{\tau^{p+1}}{\mathcal{D}(\tau)} \cdot \frac{d}{d\tau} \left( \frac{\mathcal{D}(\tau)}{\tau^{p+1}} \right) - \tau^{-1} \right] \\ &\overset{(a)}{=} \tau^{-1} \left( \frac{\mathcal{D}(\tau)}{\tau^{p+1}} \right)^{p-1} \cdot \Bigg[ (p-1) \frac{\tau^{p+1}}{\mathcal{D}(\tau)} \\ &\qquad \cdot (p+1) \tau^{-(p+2)} \left( \frac{\tau^{p+1}}{p+1} f(\tau) - \mathcal{D}(\tau) \right) - \tau^{-1} \Bigg] \\ &= \tau^{-2} \left( \frac{\mathcal{D}(\tau)}{\tau^{p+1}} \right)^{p-1} \left[ (p-1) \frac{p+1}{\mathcal{D}(\tau)} \cdot \left( \frac{\tau^{p+1}}{p+1} f(\tau) - \mathcal{D}(\tau) \right) - 1 \right] \\ &= \tau^{-2} \left( \frac{\mathcal{D}(\tau)}{\tau^{p+1}} \right)^{p-1} \left[ (p-1) \frac{\tau^{p+1}}{\mathcal{D}(\tau)} f(\tau) - p^2 \right], \end{aligned} \tag{65}$$

where (a) follows from (64). Hence, convexity will hold, if and only if, $(p-1) \frac{\tau^{p+1}}{\mathcal{D}(\tau)} f(\tau) - p^2 < 0$ for all $\tau > 0$. □

## APPENDIX III
## PROOF OF THEOREM 4

*Proof:* Let fix $d \in (0,1)$ and $\epsilon \in (0,1)$. Assuming that $(X_n)_{n \in \mathbb{N}}$ is $\ell_1$-compressible, from Theorem 1 we have that $\lim_{n \to \infty} \frac{\tilde{\kappa}_1(d, \epsilon, \mu_n)}{n} = 0$ and, consequently, it follows that:

$$\lim_{n \to \infty} \frac{\tilde{\kappa}_1(d, \epsilon, \mu_n)}{n} \cdot \log \frac{n}{\tilde{\kappa}_1(d, \epsilon, \mu_n)} = 0. \tag{66}$$

[18] $B_\epsilon(x) \equiv \{y : |x - y| < \epsilon\}$ denotes the open ball of radius $\epsilon$ centered at $x$.

[19] Without loss of generality we assume that $f_\mu(\tau_o) + f_\mu(-\tau_o) > 0$.

Then from the fact that $\lim_{n\to\infty}\frac{m_n}{n}>0$, $\exists N(d,\epsilon)>0$, such that $\forall n>N(d,\epsilon)$, $m_n>C_1\cdot\tilde{\kappa}_1(d,\epsilon,\mu_n)\log\frac{n}{\tilde{\kappa}_1(d,\epsilon,\mu_n)}$, where $C_1$ is the universal constant in (54). Then by the application of the $\ell_1$-instante optimality bound in (53), given the condition in (54) (Lemma 4), we have that $\forall n>N(d,\epsilon)$:

$$\begin{aligned} D_{\ell_1}&((\phi_{m_n\times n}(w),\Delta_n^*);X^n)\\ &=\frac{||X^n-\Delta_n^*(\phi_{m_n\times n}(w)X^n)||_{\ell_1}}{||X^n||_{\ell_1}}\\ &\le C_0\cdot\frac{\sigma_1(\tilde{\kappa}_1(d,\epsilon,\mu_n),X^n)}{||X^n||_{\ell_1}}=C_0\cdot\tilde{\sigma}_1(\tilde{\kappa}_1(d,\epsilon,\mu_n),X^n),\end{aligned} \tag{67}$$

with probability $1-2^{-C_2\cdot m_n}$ with respect to the probability of $\phi_{m_n\times n}(w)$ in $\Omega^{m_n n}$. We can take the expected value with respect to $X^n\sim\mu^n$ in (67) to obtain that $\forall n>N(d,\epsilon)$:

$$\begin{aligned} D_{\ell_1}&((\phi_{m_n\times n}(w),\Delta_n^*);\mu^n)\\ &=\mathbb{E}_{X^n\sim\mu^n}D_{\ell_1}((\phi_{m_n\times n}(w),\Delta_n^*);X^n)\\ &\le C_0\cdot\mathbb{E}_{X^n\sim\mu^n}\tilde{\sigma}_1(\tilde{\kappa}_1(d,\epsilon,\mu_n),X^n)\\ &\le C_0\cdot\left[d\cdot\mu^n(\mathcal{A}_d^{n,\tilde{\kappa}_1(d,\epsilon,\mu_n)})+1-\mu^n(\mathcal{A}_d^{n,\tilde{\kappa}_1(d,\epsilon,\mu_n)})\right]\\ &\le C_0\,[d+\epsilon],\end{aligned} \tag{68}$$

where the first inequality is from the definition of the typical set $\mathcal{A}_k^{n,k}$ in (9) and the fact that $\tilde{\sigma}_1(\tilde{\kappa}_1(d,\epsilon,\mu_n),X^n)$ is bounded by 1, and the second from definition of $\tilde{\kappa}_1(d,\epsilon,\mu_n)$ in (10). Again this inequality is valid with probability $1-2^{-C_2\cdot m_n}$ over sampling space $w\in\Omega^{m_n n}$ of the random object $\phi_{m_n\times n}(w)$. Therefore taking the limit when $n$ goes to infinity, we have that

$$\lim_{n\to\infty}D_{\ell_1}((\phi_{m_n\times n}(w),\Delta_n^*);\mu^n)\le C_0\,[d+\epsilon], \tag{69}$$

with probability one with respect to distribution of $\{\phi_{m_n\times n}(w),n\ge1\}$[20], which is valid for any arbitrary small $d\in(0,1)$ and $\epsilon\in(0,1)$. In other words, if we define the set $\Omega_{\epsilon,d}=\{w:\lim_{n\to\infty}D_{\ell_1}((\phi_{m_n\times n}(w),\Delta_n^*);\mu^n)\le C_0\,[d+\epsilon]\}$, then $\mathbb{P}(\Omega_{\epsilon,d})=1$, $\forall\epsilon\in(0,1)$ and $d\in(0,1)$. Finally, $\mathbb{P}(\cap_{i\in\mathbb{N}}\cap_{j\in\mathbb{N}}\Omega_{1/i,1/j})=1$ from sigma additivity of $\mathbb{P}$, which implies that

$$\lim_{n\to\infty}D_{\ell_1}((\phi_{m_n\times n}(w),\Delta_n^*);\mu^n)=0, \tag{70}$$

with probability one with respect to distribution of $\{\phi_{m_n\times n}(w),n\ge1\}$.

For the almost-sure convergence result on the sequence $D_{\ell_1}((\phi_{m_n\times n}(w),\Delta_n^*);X^n)$, we need a slightly different argument. Under the assumption that $\lim_{n\to\infty}m_n/n>0$, it is simple to verify that there exists a sequence $(k_n)_{n\in\mathbb{N}}$ of positive integers such that $\lim_{n\to\infty}k_n/n=r>0$ and, more importantly, it satisfies that

$$\lim_{n\to\infty}\frac{m_n}{n}>\lim_{n\to\infty}C_1\frac{k_n}{n}\log\frac{n}{k_n}. \tag{71}$$

[20]The almost sure convergence with respect to the statistics of $\{\phi_{m_n\times n}(w),n\ge1\}$ derives from the Borel-Cantelli Lemma [14] and the fact that $\sum_{n\ge}2^{-C_2\cdot m_n}<\infty$ as by hypothesis $\lim_{n\to\infty}m_n/n>0$.

Following similar steps than before, there exists $N>0$ such that $\forall n>N$ we have that $m_n>C_1\frac{k_n}{n}\log\frac{n}{k_n}$, and therefore, from Lemma 4,

$$D_{\ell_1}((\phi_{m_n\times n}(w),\Delta_n^*);X^n)\le C_0\cdot\tilde{\sigma}_1(k_n,X^n), \tag{72}$$

with probability $1-2^{-C_2\cdot m_n}$ in $\Omega^{m_n n}$. It is important to define the set $\mathcal{C}_n\equiv\{w\in\Omega^{m_n n}:\text{where (72) is satisfied}\}$ where Lemma 4 tells us that $\mathbb{P}(\mathcal{C}_n)\ge1-2^{-C_2\cdot m_n}$. Furthermore, from Lemma 3, we have that

$$\lim_{n\to\infty}\tilde{\sigma}_1(k_n,X^n)=0, \tag{73}$$

with probability one with respect to the process distribution of $\{X_n,n\ge1\}$. In other words, if we define the set $\mathcal{A}\equiv\{(x_n)_{n\in\mathbb{N}}:\lim_{n\to\infty}\tilde{\sigma}_1(k_n,X^n)=0\}$, we have that $\mathbb{P}(\mathcal{A})=1$ from (73). Finally, we are interested in the set

$$\mathcal{B}\equiv\left\{(w,(x_n)_{n\in\mathbb{N}}):\lim_{n\to\infty}D_{\ell_1}((\phi_{m_n\times n}(w),\Delta_n^*);x^n)=0\right\},$$

where it is simple to show that $(\cup_{l\ge1}\cap_{p\ge l}\mathcal{C}_p)\bigcap\mathcal{A}\subset\mathcal{B}$, by (72) and the definitions of $\mathcal{A}$ and $\mathcal{C}_p$. Hence, the problem reduces to evaluate,

$$\begin{aligned}\mathbb{P}(\underbrace{\cup_{l\ge1}\cap_{p\ge l}\mathcal{C}_p}_{=\liminf_{n\to\infty}\mathcal{C}_n})&=1-\mathbb{P}\left(\cap_{l\ge1}\cup_{p\ge l}\mathcal{C}_p^c\right)\\ \ge1-\lim_{l\to\infty}\sum_{p\ge l}\mathbb{P}\left(\mathcal{C}_p^c\right)&\ge1-\lim_{l\to\infty}\sum_{p\ge l}2^{-C_2\cdot m_p}=1,\end{aligned} \tag{74}$$

the last equality from the fact that $\lim_{n\to\infty}m_n/n>0$. Then from the additivity and monotony of the measure $\mathbb{P}(\mathcal{B})=1$, which concludes the result by the definition of $\mathcal{B}$. □

## Appendix IV
## Complementary Results

### A. Proposition 1

*Proof:* From the definition of $r_p(d,(x_n)_{n\in\mathbb{N}})$ and the hypothesis on $(k_n)$, it follows that $\exists(\tilde{k}_n)$ where $\limsup_{n\to\infty}\frac{\tilde{k}_n}{n}<\liminf_{n\to\infty}\frac{k_n}{n}$ and $\limsup_{n\to\infty}\tilde{\sigma}_p(\tilde{k}_n,x^n)\le d$. Consequently eventually in $n$, $\tilde{k}_n<k_n$, which implies that eventually $\tilde{\sigma}_p(\tilde{k}_n,x^n)\ge\tilde{\sigma}_p(k_n,x^n)$. This concludes the result. □

### B. Lemma 1

*Proof:* As a short-hand, let $k_n^*\equiv\kappa_p(d,x^n)$ for all $n$. By definition $\tilde{\sigma}_p(k_n^*,x^n)\le d$ for all $n$, and consequently, $\limsup_{n\to\infty}\tilde{\sigma}_p(k_n^*,x^n)\le d$. Then from (5) and Def. 1, it follows that $\limsup_{n\to\infty}\frac{k_n^*}{n}\ge r_p(d,(x_n)_{n\in\mathbb{N}})$.

For the other inequality, we consider the nontrivial case when $\liminf_{n\to\infty}\frac{k_n^*}{n}=r_1>0$. We prove it by contradiction assuming that $r_p(d,(x_n)_{n\in\mathbb{N}})<r_1$. Then from (5), there exists $r_0<r_1$ and a sequence $\tilde{k}_n=r_0 n$ such that $\limsup_{n\to\infty}\tilde{\sigma}_p(\tilde{k}_n,x^n)\le d$. Under the fact that $r_0<r_1$, there exists $N>0$ such that $\forall n>N$, $\tilde{k}_n<k_n^*$. Using this and the definition of $k_n^*$ in (6), $\forall n>N$,

$$d<\tilde{\sigma}_p(k_n^*-1,x^n)\le\tilde{\sigma}_p(\tilde{k}_n,x^n) \tag{75}$$

where, consequently, $d\le\liminf_{n\to\infty}\tilde{\sigma}_p(\tilde{k}_n,x^n)\le\limsup_{n\to\infty}\tilde{\sigma}_p(\tilde{k}_n,x^n)\le d$. Therefore,

$$\lim_{n\to\infty}\tilde{\sigma}_p(\tilde{k}_n,x^n)=\lim_{n\to\infty}\tilde{\sigma}_p(k_n^*-1,x^n)=d. \tag{76}$$

*Definition 7:* For $0 \leq k < \bar{k} \leq n$, we define $\tilde{\zeta}_p(\bar{k}, k, x^n) \equiv \frac{(|x_{n,k+1}|^p + \cdots + |x_{n,\bar{k}}|^p)^{\frac{1}{p}}}{||x^n||_{\ell_p}}$.

Note that $\tilde{\sigma}_p(\tilde{k}_n, x^n) - \tilde{\sigma}_p(k_n^* - 1, x^n) = \tilde{\zeta}_p(k_n^* - 1, \tilde{k}_n, x^n)$, therefore (76) implies that $\lim_{n\to\infty} \tilde{\zeta}_p(k_n^* - 1, \tilde{k}_n, x^n) = 0$. Then, considering that $r_1 > r_0$, and the fact that $\forall ko > 0$, $l > 0$ and $s > 0$, $\tilde{\zeta}_p(k_o + l, k_o, x^n) \geq \tilde{\zeta}_p(k_o + l + s, k_o + s, x^n)$, we have that:

$$\lim_{n\to\infty} \tilde{\zeta}_p((k_n^* - 1 - \tilde{k}_n) + k_n^* - 1, k_n^* - 1, x^n) = 0,$$
$$\lim_{n\to\infty} \tilde{\zeta}_p(2(k_n^* - 1 - \tilde{k}_n) + k_n^* - 1, (k_n^* - 1 - \tilde{k}_n) + k_n^* - 1, x^n) = 0, \ldots. \quad (77)$$

This approach can be iterated a finite number of times (independent of the length $n$ as $r_0 < r_1$) to obtain that

$$\lim_{n\to\infty} \tilde{\zeta}_p(n, \tilde{k}_n, x^n) = 0 \Leftrightarrow \lim_{n\to\infty} \tilde{\sigma}_p(\tilde{k}_n, x^n) = 0, \quad (78)$$

which contradicts (75). □

### *C. Proposition 2*

*Proof:* Fixing $\epsilon \in (0,1)$ and $d > 0$, let us define $k_n^* = \tilde{\kappa}_p(d, \epsilon, \mu_n)$, $\forall n \geq 1$. Then by (10), $\liminf_{n\to\infty} \mu^n(\mathcal{A}_d^{n,k_n^*}) \geq 1 - \epsilon$, where from (13) it follows that $\limsup_{n\to\infty} \frac{k_n^*}{n} \geq \tilde{r}_p(d, \epsilon, \mathbb{P})$. □

### *D. Proof of Proposition 6*

*Proof:* Using the arguments to prove Theorem 2 (Point 4), if we consider $D = d^p$, then for any $D \in (0,1)$ there exists $\tau_o > 0$ such that $D = \bar{\phi}_{v_p}(\tau_o) \equiv (1 - \phi_{v_p}(\tau_o))$. This last auxiliary function is diferentiable (with respect to $\tau$) and we have that:

$$\frac{\partial \tilde{r}_p(D^{1/p}, \epsilon, \mathbb{P})}{\partial D} = \frac{\partial \phi_\mu}{\partial \tau}(\tau_o) \cdot \left(\frac{\partial \bar{\phi}_{v_p}}{\partial \tau}(\tau_o)\right)^{-1}$$
$$= -\,||x^p||_{L_1(\mu)} \cdot \left(\frac{1}{\bar{\phi}_{v_p}^{-1}(D)}\right)^p, \quad (79)$$

the last equality by considering that $\frac{\partial \bar{\phi}_{v_p}}{\partial \tau}(\tau) = \frac{|\tau|^p}{\int_{\mathbb{R}} |x|^p d\mu(x)} \cdot (f_\mu(\tau) + f_\mu(-\tau))$, $\forall \tau > 0$. Since $\bar{\phi}_{v_p}(\tau)$ is non-decreasing with $\tau$, it follows that $\partial \tilde{r}/\partial D$ is negative and increasing with $D$, i.e., $\tilde{r}$ is a convex function of $D$. □

### *E. Analysis of the Domain of Attraction of $\alpha$-Stable Distributions*

The family of stable laws is the class of non-degenerate probabilities that are limit (in distribution) of sequences of random objects of the form [14, Ch.9]:

$$\frac{X_1 + .. + X_n}{A_n} - B_n \quad (80)$$

where $X_1, .., X_n$ are i.i.d realizations of a random variable, and $(A_n)$ and $(B_n)$ are a sequences of real numbers. For the well-known scenario when $\mathbb{E}|X_1|^2 < \infty$, the *Central Limit Theorem* tells us that the limit is a normal law. For the case $\mathbb{E}|X_1|^2 = \infty$, we have the less known family of $\alpha$-stable laws, whose characteristic function $f(u)$ is given by [14, Th.9.27]:

$$\log f(u) = ju\beta + m_1 \int_0^\infty \left(e^{jux} - 1 - \frac{jux}{1+x^2}\right)\frac{dx}{x^{1+\alpha}}$$
$$+ m_2 \int_{-\infty}^0 \left(e^{jux} - 1 - \frac{jux}{1+x^2}\right)\frac{dx}{|x|^{1+\alpha}} \quad (81)$$

being $\alpha \in (0,2)$ the exponent of the law, and $m_1 \geq 0$, $m_2 \geq 0$ and $\beta$ constants.

*Definition 8:* [14, Def. 9.33] The distribution $\mu \in \mathcal{P}(\mathbb{R})$ is said to be in the domain of attraction of an $\alpha$-stable law with $0 < \alpha < 2$, which we denote by $D(\alpha) \subset \mathcal{P}(\mathbb{R})$, if there exists $(A_n)$ and $(B_n)$ such that: $\frac{X_1 + .. + X_n}{A_n} - B_n \to X$ (in distribution) and $X$ follows the $\alpha$-stable distribution in (81).

The collection $\bigcup_{\alpha\in(0,2)} D(\alpha)$ is non-empty and is characterized by the following result:

*Theorem 5:* [14, Th.9.34 and Prop. 9.39] Let $\mu \in \mathcal{P}(\mathbb{R})$ and let us define[21]:

$$H_\alpha^+(x) \equiv x^\alpha(1 - F_\mu(x)) \text{ and } H_\alpha^-(x) \equiv x^\alpha F_\mu(-x) \quad (82)$$

on $(0,\infty)$. Then $\mu$ belongs to $D(\alpha)$, if there exists $M^+ \geq 0$ and $M^- \geq 0$ with $M^+ + M^- > 0$ such that:

i) $\lim_{x\to\infty} \frac{F_\mu(-x)}{1 - F_\mu(x)} = \frac{M^-}{M^+} \in \mathbb{R}^+ \cup \{\infty\}$,
ii) $M^+ > 0$ implies that $H_\alpha^+(x)$ is slowly changing[22], and
iii) $M^- > 0$ implies that $H_\alpha^-(x)$ is slowly changing.

Then we can state the following:

*Proposition 7:* If $\mu \in D(\alpha)$ for some $\alpha \in (0,1)$, then $\mathbb{E}_{X\sim\mu}(|X|) = \infty$.

*Proof:* Let $\mu \in D(\alpha)$ and $\mu \ll \lambda$. Without loss of generality let us assume that $H_\alpha^+(x)$ is slowly changing. Then it is simple to verify that $\exists \bar{\alpha} \in (\alpha, 1)$, $\exists x_o \geq 0$ and $C_o > 0$ such that $\forall x \geq x_o$ $f_\mu(x) \geq C_o \frac{1}{x^{(1+\bar{\alpha})}}$. Therefore $\mathbb{E}_{X\sim\mu}(|X|) \geq \int_{x_o}^\infty x f_\mu(x) dx \geq C_o \int_{x_o}^\infty \frac{1}{x^{\bar{\alpha}}} dx = \infty$. □

## ACKNOWLEDGMENT

The authors thank the anonymous reviewers for their valuable comments, and their colleagues R. Lobos, C. Azurdia and M. Orchard for proofreading part of this material.

## REFERENCES

[1] A. Amini, M. Unser, and F. Marvasti, "Compressibility of deterministic and random infinity sequences," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5193–5201, Nov. 2011.
[2] R. Gribonval, V. Cevher, and M. E. Davies, "Compressible distributions for high-dimensional statistics," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5016–5034, 2012.
[3] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best $k$-term approximation," *J. Amer. Math. Soc.*, vol. 22, no. 1, pp. 211–231, Jul. 2009.
[4] V. Cevher, "Learning with compressible priors," presented at the Neural Inf. Process. Syst.(NIPS), Vancouver, Canada, 2008.
[5] A. Amini and M. Unser, "Sparsity and infinity divisibility," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2346–2358, Apr. 2014.
[6] M. Unser, P. Tafti, A. Amini, and H. Kirshner, "A unified formulation of Gaussian versus sparse stochastic processes—Part II: Discrete domain theory," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 3036–3050, May 2014.
[7] D. L. Donoho, "Compressed sensing," *IEEE Transact. on Informat. Theory*, vol. 52, pp. 1289–1306, 2006.

[21] $F_\mu(x) = \mu((\infty, x))$ $\forall x$ is the distribution function of $\mu$.

[22] A function $G(x)$ is slowly changing if $\forall \rho > 0$, $\lim_{x\to\infty} \frac{G(\rho \cdot x)}{G(x)} = 1$.

[8] E. J. Candes and T. Tao, "Decoding by linear programing," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[9] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertanty principle: Exact signal reconstruction from highly imcomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[10] R. Gribonval, "Should penalized least squares regression be interpreted as maximum a posteriori estimation?," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2405–2410, May 2011.

[11] A. Amini, U. Kamilov, E. Bostan, and M. Unser, "Bayesian estimation for continuous-time sparse stochastic processes," *IEEE Trans. Signal Process.*, vol. 61, no. 4, pp. 907–929, 2013.

[12] R. Prasad and C. Murthy, "Cramer-Rao-type of bounds for sparse Bayesian learning," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 622–632, 2013.

[13] S. Varadhan, *Probability Theory*. Providence, RI, USA: Amer. Math. Soc., 2001.

[14] L. Breiman, *Probability*. Reading, MA, USA: Addison-Wesley, 1968.

[15] R. G. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, 2008, DOI: 10.1007/500365-007-9003-x.

[16] E. J. Candes, "The restricted isometry property and its applications for compressed sensing," *C. R. Acad. Sci. Paris*, vol. I 346, pp. 589–592, 2008.

[17] J. F. Silva and M. S. Derpich, "Precise best $k$-term approximation error analysis of ergodic sequences," presented at the IEEE Int. Symp. Inf. Theory, Honolulu, HI, USA, 2014.

[18] R. M. Gray, *Probability, Random Processes, and Ergodic Properties*. New York, NY, USA: Springer, 2009.

[19] K. Petersen, *Ergodic Theory*. Cambridge, U.K.: Cambridge Univ. Press, 1983.

[20] S. Founcart and M. Lai, "Sparsest solutions of underdemined linear systems via $\ell_p$-minimization," *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 395–407, 2009.

[21] R. DeVore, G. Petrova, and P. Wojtaszczyk, "Instance-optimality in probability with an $\ell_1$-minimization decoder," *Appl. Comput. Harmon. Anal.*, vol. 27, pp. 275–288, 2009.

[22] R. M. Gray, *Entropy and Information Theory*. New York, NY, USA: Springer-Verlag, 1990.

[23] P. Billingsley, *Ergodic Theory and Information*. New York, NY, USA: Wiley, 1965.

[24] G. Masi and K. Stettner, "Ergodicity of hidden markov models," *Math. Control, Signals, Syst.*, vol. 17, no. 4, pp. 269–296, 2005.



**Jorge F. Silva** (S'08–M'09) is Associate Professor at the Electrical Engineering Department and director of the Information and Decision Systems (IDS) Group at the University of Chile, Santiago, Chile. He received the Master of Science (2005) and Ph.D. (2008) in Electrical Engineering from the University of Southern California (USC). He is IEEE member of the Signal Processing and Information Theory Societies and he has participated as a reviewer in various IEEE journals on Signal Processing. Mr. Silva is recipient of the Viterbi Doctoral Fellowship 2007–2008 and Simon Ramo Scholarship 2007–2008 at USC.

Jorge F. Silva was research assistant at the Signal Analysis and Interpretation Laboratory (SAIL) at USC (2003–2008) and was also research intern at the Speech Research Group, Microsoft Corporation, Redmond (Summer 2005). His current research interests include: detection and estimation, universal source coding, statistical learning, information theory and statistics, compressible models for information sources, compressed sensing and its applications.



**Milan S. Derpich** (S'08–M'09) received the Ingeniero Civil Electrónico degree from the Universidad Técnica Federico Santa María (UTFSM), Valparaíso, Chile in 1999. During his time at the university he was supported by a full scholarship from the alumni association and upon graduating received several university-wide prizes. Mr. Derpich also worked by the electronic circuit design and manufacturing company Protonic Chile S.A. between 2000 and 2004. In 2009 he received the Ph.D. degree in electrical engineering from the University of Newcastle, Australia. He received the Guan Zhao-Zhi Award at the Chinese Control Conference 2006, and the Research Higher Degrees Award from The Faculty of Engineering and Built Environment, University of Newcastle, Australia, for his Ph.D. thesis. Since 2009 he has been with the Department of Electronic Engineering at Universidad Técnica Federico Santa María, Chile. His main research interests include rate-distortion theory, communications, networked control systems, sampling and quantization.