Compressibility of Deterministic and Random Infinite Sequences

Arash Amini, Michael Unser, *Fellow, IEEE*, and Farokh Marvasti, *Senior Member, IEEE*

***Abstract*—We introduce a definition of the notion of compressibility for infinite deterministic and i.i.d. random sequences which is based on the asymptotic behavior of truncated subsequences. For this purpose, we use asymptotic results regarding the distribution of order statistics for heavy-tail distributions and their link with α-stable laws for $1 < \alpha < 2$. In many cases, our proposed definition of compressibility coincides with intuition. In particular, we prove that heavy-tail (polynomial decaying) distributions fulfill the requirements of compressibility. On the other hand, exponential decaying distributions like Laplace and Gaussian do not. The results are such that two compressible distributions can be compared with each other in terms of their degree of compressibility.**

***Index Terms*—Compressible prior, heavy-tail distribution, order statistics, stable law.**

I. Introduction

THE emerging field of compressed sensing investigates the problem of sampling and reconstructing signals (often finite-dimensional) that have sparse or compressible (almost sparse) representations in an orthonormal basis [1]–[3]. For instance, the so-called basis-pursuit method, which is based on the ℓ_1 minimization, can stably recover compressible vectors from their linear projections onto subspaces of much lower dimensions when the projection operator satisfies some constraints.

By definition, an $n \times 1$ vector (a finite sequence) is called "sparse", more precisely, k-sparse, when it contains at most k nonzero elements. In this case, a k-term representation of the vector exactly describes the vector, which is of special relevance when $k \ll n$. By extension, a vector is called "compressible" when its k-term representation is not exact, but only an approximation in some sense.

Manuscript received October 19, 2010; revised April 19, 2011; accepted July 06, 2011. Date of publication July 25, 2011; date of current version October 12, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Konstantinos I. Diamantaras. The research was supported in part by the Swiss National Science Foundation under Grant 200020-109415.

A. Amini and F. Marvasti are with the Advanced Communication Research Institute (ACRI), Electrical Engineering Department, Sharif University of Technology, Tehran 11365-9363, Iran (e-mail: arashsil@ee.sharif.edu; marvasti@sharif.edu).

M. Unser is with the Biomedical Imaging Group (BIG), EPFL, CH-1015 Lausanne, Switzerland (e-mail: michael.unser@epfl.ch).

This paper has supplementary downloadable multimedia material available at http://ieeexplore.ieee.org provided by the authors. This includes the MATLAB.m file, which shows how the figures are generated. This material is 0.005 MB in size.

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.2011.2162952

In the context of infinite sequences, finite-term approximations are classically studied for wavelet coefficients of the deterministic signals in Besov spaces [4] where the sequence of coefficients is in ℓ_p for some p and the energy (ℓ_p-norm) of the k-term approximation error decays like k^{-c} for some positive c. In more recent definitions, the sequence $\{x_i\}_{i\in\mathbb{N}}$ is called "compressible" in the sense of weak-ℓ_p norm $(0 < p)$ if it belongs to the weak-ℓ_p space [5], [6]. More precisely, for some R we should have

$$|x_i| \leq R i^{-\frac{1}{p}}. \tag{1}$$

Clearly, the above definition imposes a certain rate of decay on compressible sequences.

Introducing a stochastic model for a specific class of signals is a common approach in signal processing and it usually done by assuming an innovations domain. Typically, the signal is assumed to be independently and identically distributed in a transform domain. Now, to check the compressibility of stochastic signals, we should investigate the compressibility of their innovation sequence which is usually a sequence of i.i.d. random variables. Unfortunately, the definition of compressibility based on (1) is not applicable because sequences of i.i.d. random variables are excluded from ℓ_p. Moreover, they are not even decaying at all—except for some degenerate distributions. This fact suggests that a new definition for the compressibility of infinite sequences and i.i.d. sequences should be introduced. The main benefit of studying compressible priors is to provide the foundations for stochastic modeling of sparse/compressible processes. Most of the conventional stochastic signal models are based on Gaussian processes which are definitely not sparse. In the past five years, significant progress has been achieved in the relatively new field of sparse signal processing using a predominantly deterministic formulation. The investigation of sparse stochastic models may lead to new and efficient algorithms for reconstruction or denoising of the sparse signals.

The first attempts of identifying a compressible prior were based on the Bayesian interpretation of the successful basis-pursuit method. It was found that ℓ_1 minimization technique coincides with the maximum *a posteriori* (MAP) estimator [7], [8] when the observations of a Laplace-distributed signal having independent and identically distributed coefficients in the innovations domain are perturbed by Gaussian noise. However, several authors have pointed out that this choice does not meet the expectations from a compressible prior [9]–[11]. Employing quantile approximations of the expected order statistics, Cevher is the precursor for defining and identifying compressible distributions [9]. He proposed the following definition.

1053-587X/$26.00 © 2011 IEEE

Definition 1: For a random variable x with parameters $\boldsymbol{\theta}$, let $f_y(.)$ be the probability density function of $y = |x|$ and $F_y(.)$ denote the respective cumulative probability distribution. Also, for a given sequence length n, let $y_{n1} \geq y_{n2} \geq \dots y_{nn}$ denote the order statistics of the i.i.d. random variables $\{y_i\}_{i=1}^n$. The distribution of x is called compressible if there exist positive constants $R(n,\boldsymbol{\theta})$ and $p(n,\boldsymbol{\theta})$ such that

$$\mathcal{E}\{y_{ni}\} \approx F_y^{-1}\left(1 - \frac{i}{n+1}\right) \lesssim R(n,\boldsymbol{\theta}) i^{-\frac{1}{p(n,\boldsymbol{\theta})}} \qquad (2)$$

where $\mathcal{E}$ indicates expectation and $\lesssim$ means "less than or approximately equal to."

A few distributions, including the *Generalized Pareto* and *Student's* t, are introduced as compressible priors that fulfill Definition 1. Nonetheless, this definition is somewhat imprecise. Moreover, the distributions should be examined on a case-by-case basis. For each distribution, those samples of f^{-1} required in (2) should be evaluated first, and then, one should check if such $R(n,\boldsymbol{\theta})$ and $p(n,\boldsymbol{\theta})$ exist.

In this paper, we introduce a comprehensive definition of compressibility (strong compressibility or simply s-compressibility) for infinite sequences which ultimately leads to some simple practical classification criteria.[1] The main advantage is that this new definition is not based on the decay of the sequence. This enables us to generalize it to i.i.d. random sequences. Furthermore, we provide theorems to distinguish compressible priors for the given definition of compressibility. Using asymptotic results regarding the order statistics, we show that the s-compressibility of a prior closely depends on the decay of the probability density function. We demonstrate the s-compressibility of heavy-tail (polynomial decaying) distributions such as the Generalized Pareto and Student's t, while we prove that exponentially decaying distributions such as those from Laplace and Gauss are not s-compressible. Our definition appears to be compatible with Definition 1.

II. Compressible Infinite Sequences

The first idea that comes to mind for defining compressibility for an infinite sequence is to consider (1). Unfortunately, (1) is too restrictive since it prevents any i.i.d. random sequence to be called compressible, by lack of decay. To overcome this shortcoming, we define the compressibility of an infinite sequence using the asymptotic behavior of its truncated finite subsequences. Intuitively, an s-compressible infinite sequence is such that the energy of its truncated finite subsequences is concentrated in only a small fraction of the samples, and this fraction vanishes as the length of the subsequence increases. In order to put this insight into a rigorous mathematical form, we need to introduce some notations.

A. Definitions and Notations

Definition 2: Let $\{a_i\}_{i=1}^n$ be a finite scalar sequence and let $\{a_{ni}\}_{i=1}^n$ be the ordered sequence according to nonincreasing modulus values ($|a_{n1}| \geq \dots \geq |a_{nn}|$). For $p > 0$ and $k \leq n$, define

$$\sigma_p\left(k; \{a_i\}_{i=1}^n\right) = \left(|a_{n1}|^p + \dots + |a_{nk}|^p\right)^{\frac{1}{p}}. \qquad (3)$$

In other words, $\sigma_p(k; \{a_i\}_{i=1}^n)$ is the ℓ_p-(semi)norm of the best k-term approximation of $\{a_i\}_{i=1}^n$. (Note that $\sigma_p(n; \{a_i\}_{i=1}^n)$ is the ℓ_p-norm of the sequence.)

Definition 3: For a finite scalar sequence $\{a_i\}_{i=1}^n$ (not all equal to zero) and real numbers $0 \leq r \leq 1$ and $p > 0$, define

$$\kappa_p\left(r; \{a_i\}_{i=1}^n\right) = \min\left\{k \,\middle|\, \frac{\sigma_p\left(k; \{a_i\}_{i=1}^n\right)}{\sigma_p\left(n; \{a_i\}_{i=1}^n\right)} \geq r\right\}. \qquad (4)$$

In fact, $\kappa_p(.)$ provides the minimum number of terms required to preserve the given fraction r of the total energy of the sequence in the ℓ_p sense. Finally, we define the s-compressibility of infinite sequences based on the asymptotic behavior of their truncated subsequences.

B. Definition 4

We call the sequence $\{a_i\}_{i\in\mathbb{N}}$ of real (or complex) numbers "ℓ_p-compressible" if

$$\forall\, 0 \leq r < 1 : \lim_{n\to\infty} \frac{\kappa_p\left(r; \{a_i\}_{i=1}^n\right)}{n} = 0. \qquad (5)$$

In addition, the sequence $\{a_i\}_{i\in\mathbb{N}}$ is called s-compressible if it is ℓ_p-compressible for some $0 < p$.

Definition 4 implies that, in order to capture almost all the energy of the truncated subsequences (with a large-enough number of terms) of a s-compressible infinite sequence, one only needs a negligible fraction of the terms. Note that we focus on the energy distribution among the elements, without imposing a decay on the sequence.

To illustrate Definition 4, let us consider the sequence $a_i = \frac{1}{2^i}$ for $i \in \mathbb{N}$. For any $0 \leq r < 1$ and $0 < p$, let $k = \left\lceil -\frac{\log_2(1-r^p)}{p} \right\rceil$. Then, for the k-term approximation of any subsequence of length $n \geq k$, we observe that

$$\frac{\sigma_p\left(k; \{a_i\}_{i=1}^n\right)}{\sigma_p\left(n; \{a_i\}_{i=1}^n\right)} = \left(\frac{\sum_{i=1}^k 2^{-ip}}{\sum_{i=1}^n 2^{-ip}}\right)^{\frac{1}{p}} \geq \left(\frac{\sum_{i=1}^k 2^{-ip}}{\sum_{i=1}^\infty 2^{-ip}}\right)^{\frac{1}{p}} = \left(1 - 2^{-kp}\right)^{\frac{1}{p}} \geq r \qquad (6)$$

which means

$$\kappa_p\left(r; \{a_i\}_{i=1}^n\right) \leq k. \qquad (7)$$

Therefore, since n can be increased independently of k, the limit of $n^{-1}\kappa_p(r; \{a_i\}_{i=1}^n)$ when $n \to \infty$ would be zero, and the sequence is ℓ_p-compressible for all p. The above arguments rely on the fact that the whole sequence has a finite energy and that one only needs a finite number of elements to preserve any given fraction of the total energy. This argument can be generalized to all ℓ_p sequences, so that a sequence in ℓ_p is also ℓ_p-compressible.

To show that the converse argument is not necessarily true, consider $a_i = \frac{1}{i}$. Although this sequence is not in ℓ_1, we show that it is ℓ_1-compressible. For a given $r < 1$, we choose $n >$

[1] While this paper was under review, we were informed of a paper submission with similar ideas by Gribonval, Cevher and Davis [12].

$e^{\frac{r}{1-r}}$ and we set $k = \lceil (ne)^r \rceil$. Using $\ln m < \sum_{i=1}^{m} \frac{1}{i} < 1 + \ln m$, we get

$$\frac{\sigma_p\left(k;\{a_i\}_{i=1}^n\right)}{\sigma_p\left(n;\{a_i\}_{i=1}^n\right)} = \frac{\sum_{i=1}^{k}\frac{1}{i}}{\sum_{i=1}^{n}\frac{1}{i}} \geq \frac{\ln k}{1+\ln n} \geq \frac{r\ln(ne)}{1+\ln n} = r. \quad (8)$$

Thus,

$$\kappa_1\left(r;\{a_i\}_{i=1}^n\right) \leq k \approx (ne)^r \quad (9)$$

which yields

$$\lim_{n\to\infty} \frac{\kappa_1\left(r;\{a_i\}_{i=1}^n\right)}{n} \leq \lim_{n\to\infty} \frac{(ne)^r}{n} = 0. \quad (10)$$

Hence, the sequence fulfills the requirements of ℓ_1-compressibility. The two examples above were decaying sequences. In Section IV, we shall consider some nondecaying examples as well.

C. Results

An interesting property of Definition 4 is the embedding of ℓ_p-compressible sequences with different ps. By using the following lemma, we show that an ℓ_p-compressible sequence is also ℓ_q-compressible for $q \geq p$:

Lemma 1: For a sequence $\{a_i\}_i$ and arbitrary integers $k < n$, the ratio $\left(\frac{\sigma_p(k;\{a_i\}_{i=1}^n)}{\sigma_p(n;\{a_i\}_{i=1}^n)}\right)^p$ is an increasing function of p.

Proof: Let $q \geq p$. For indices $1 \leq i \leq j \leq n$, we have that

$$\begin{aligned} |a_{ni}| \geq |a_{nj}| &\Rightarrow |a_{ni}|^{q-p} \geq |a_{nj}|^{q-p} \\ &\Rightarrow |a_{ni}|^q |a_{nj}|^p \geq |a_{ni}|^p |a_{nj}|^q. \end{aligned} \quad (11)$$

Now, by summing these inequalities for all $1 \leq i \leq k$ and $k+1 \leq j \leq n$, we get

$$\begin{aligned} &\sum_{i=1}^{k}\sum_{j=k+1}^{n} |a_{ni}|^q |a_{nj}|^p \geq \sum_{i=1}^{k}\sum_{j=k+1}^{n} |a_{ni}|^p |a_{nj}|^q \\ &\Rightarrow \frac{\sum_{i=1}^{k}|a_{ni}|^q}{\sum_{i=k+1}^{n}|a_{ni}|^q} \geq \frac{\sum_{i=1}^{k}|a_{ni}|^p}{\sum_{i=k+1}^{n}|a_{ni}|^p} \\ &\Rightarrow \frac{\sum_{i=1}^{k}|a_{ni}|^q}{\sum_{i=1}^{n}|a_{ni}|^q} \geq \frac{\sum_{i=1}^{k}|a_{ni}|^p}{\sum_{i=1}^{n}|a_{ni}|^p} \end{aligned} \quad (12)$$

which completes the proof. ∎

A direct consequence of Lemma 1 is that $\kappa_p(r^{\frac{1}{p}}, \{a_i\}_{i=1}^n)$ is a decreasing function of p. Hence, when the limit in Definition 4 vanishes for some $p > 0$, it will also vanish for all $q \geq p$, meaning that a sequence is either s-incompressible or ℓ_p-compressible for p larger than a threshold. The minimum (or infimum) of the positive values p for which a sequence is ℓ_p-compressible can be regarded as a measure of how compressible the sequence is: the lower the infimum value, the more compressible the sequence.

III. Compressible i.i.d. Random Sequences

In order to generalize the s-compressibility definition to i.i.d. random sequences, we need to modify Definition 3 to take statistic aspects into account. Since the sequence is now random, we can only guarantee that the best k-term approximation of a subsequence preserves the given fraction r of the energy with probability at least δ.

A. Definitions and Notations

Definition 5: Let $\{x_i\}_{i\in\mathbb{N}}$ be an i.i.d. sequence of random variables[2] with probability density f_x. For given $0 \leq r, \delta \leq 1$ and $p > 0$, we define

$$\tilde{\kappa}_p(r,\delta,n;f_x) = \min\left\{k \middle| P\left(\frac{\sigma_p\left(k;\{x_i\}_{i=1}^n\right)}{\sigma_p\left(n;\{x_i\}_{i=1}^n\right)} \geq r\right) \geq \delta\right\}. \quad (13)$$

Thus, $\tilde{\kappa}_p(.)$ is almost equivalent to $\kappa_p(.)$ with the addition of the probability measure. Note that there are two lower-bounds in Definition 5: one for the energy fraction r and one for the probability δ. Now, similarly to Definition 4, we can define the s-compressibility of random sequences. Because the s-compressibility only depends on the distribution f_x instead of the sequence, we define the s-compressibility in terms of the distribution.

Definition 6: The probability density f_x is called "ℓ_p-compressible" if the sequence of i.i.d. realizations of this distribution is almost surely ℓ_p-compressible, in the sense that

$$\forall\, 0 \leq r, \quad \delta < 1, \quad \lim_{n\to\infty} \frac{\tilde{\kappa}_p(r,\delta,n;f_x)}{n} = 0. \quad (14)$$

Similarly, the probability density f_x is called s-compressible if it is ℓ_p-compressible for some $0 < p$.

Although we focus on the above definition in this paper, we also propose the following distinction between sparsity (exact representation) and compressibility (approximation):

Definition 7: The probability density f_x is called "ℓ_p-sparse" if it is ℓ_p-compressible and if, for finite but long-enough sequences of i.i.d. realizations of f_x, the whole energy (in any sense) is almost surely concentrated in a fraction of the terms

$$\forall p, \quad 0 \leq \delta < 1, \quad \lim_{n\to\infty} \frac{\tilde{\kappa}_p(1,\delta,n;f_x)}{n} < 1. \quad (15)$$

The only difference between s-compressible and sparse distributions lies in the fact that the choice of $r = 1$ is excluded for s-compressible distributions. One can easily verify that none of our mentioned s-compressible examples (deterministic) is sparse. The additional condition in the definition of sparsity implies that some of the elements of a sparse i.i.d. realization necessarily vanish. This in turn implies that ℓ_p-sparse distributions always contain a mass probability at $x = 0$. Conversely, when $P(x = 0) = \pi_0 > 0$, the whole energy of an i.i.d. sequence of length n is contained with high probability in only $n(1-p)$ elements. Therefore, the limit in Definition 7 is less than one. This means that sparse distributions are those s-compressible distributions that contain a mass probability at $x = 0$. To generate a sparse i.i.d. sequence, one can select with probability $1 - \pi_0$ values drawn from a s-compressible distribution, and select with probability π_0 the value 0.

[2]In this paper, we use the lower-case letter f exclusively to represent the probability density function of a random variable.

Compared to [9], Definition 6 explains the concept of compressibility more intuitively. Moreover, we are going to show that Definition 6 leads naturally to the derivation of simple tools for examining the s-compressibility of distributions.

In the rest of the paper, we study distributions from the point of view of their ℓ_p-compressibility. There are two main contributions: We first exclude a large class of distributions, namely, those with exponential decay. The well-studied Laplace distribution is a member of this non ℓ_P-compressible class. Then, we introduce heavy-tail distributions (polynomial decay) as possible s-compressible candidates.

Lemma 1 allowed us to identify an embedding property of sequences, likewise, there is a similar embedding for s-compressible (sparse) distributions: An ℓ_p-compressible (ℓ_p-sparse) distribution is also ℓ_q-compressible (ℓ_q-sparse) for all $q \geq p$. Moreover, it is noteworthy to mention the following link between the ℓ_p-compressible distributions with different ps.

Lemma 2: The distribution of the random variable x is ℓ_p-compressible if and only if the distribution of the random variable $y = |x|^p$ is ℓ_1-compressible.

Proof: The statement follows from the fact that ℓ_p-norm of a sequence of x_is to the power p is equal to ℓ_1-norm of the corresponding sequence of y_is.

B. Main Results

The statement in Lemma 2 shows that we only need to identify ℓ_1-compressible distributions in order to identify the whole class. Hence, from now on, we only study the ℓ_1-compressibility condition.

Theorem 1: If the probability density f_x is such that, for some $0 < \gamma$, the expectation $\mathcal{E}_x\{e^{\gamma|x|}\}$ exists, then the distribution is not ℓ_1-compressible.

Proof: Let $\{x_i\}$ be an i.i.d. sequence of random variables with distribution f_x and let $y_i = |x_i|$. Obviously, the y_is form an i.i.d. sequence with distribution $f_y(y) = (f_x(y)+f_x(-y))u(y)$, where $u(.)$ is the Heaviside (unit) step function. Also, let μ_y and σ_y^2 be the mean and variance of a random variable with distribution f_y. The existence of $\mathcal{E}_x\{e^{\gamma|x|}\}$ implies the existence of μ_y and σ_y. Moreover, let m be a large-enough positive integer such that $\mathcal{E}_y\{e^{\gamma y}\} < e^{m\gamma r\mu_y}(3m)^{-1}$, where γ is the positive value for which $\mathcal{E}\{e^{\gamma y}\}$ is finite, and where $0 < r < 1$ is an arbitrary ratio. To simplify the notations, let us define the following events:

$$\begin{cases} EV_1: & \sigma_1\left(k;\{y_i\}_{i=1}^{mk}\right) \geq r\sum_{i=1}^{mk} y_i \\ EV_2: & \sum_{i=1}^{mk} y_i \geq mk\mu_y - (mk)^{0.75}\sigma_y \\ EV_3: & \sum_{i=1}^{mk} y_i < mk\mu_y - (mk)^{0.75}\sigma_y \end{cases} \tag{16}$$

which yields

$$\begin{aligned} P(EV_1) &= P(EV_1|EV_2)P(EV_2) + P(EV_1|EV_3)P(EV_3) \\ &\leq P(EV_1|EV_2) + P(EV_3) \\ &\leq P\left(\sigma_1\left(k;\{y_i\}_{i=1}^{mk}\right) \geq r\left(mk\mu_y - (mk)^{0.75}\sigma_y\right)\right) \\ &\quad + P\left(\sum_{i=1}^{mk} y_i < mk\mu_y - (mk)^{0.75}\sigma_y\right). \end{aligned} \tag{17}$$

The independence of the y_is results in

$$\begin{cases} \mathcal{E}_y\left\{\sum_{i=1}^{mk} y_i\right\} = mk\mu_y \\ \mathrm{Var}\left\{\sum_{i=1}^{mk} y_i\right\} = mk\sigma_y^2. \end{cases} \tag{18}$$

Therefore, by applying Chebychev's inequality we obtain

$$P\left(\sum_{i=1}^{mk} y_i < mk\mu_y - (mk)^{0.75}\sigma_y\right) \leq \frac{1}{\sqrt{mk}}. \tag{19}$$

Furthermore, $\sigma_1(k;\{y_i\}_{i=1}^{mk})$ is generated by one of the $\binom{mk}{k}$ possible ways of selecting k elements. Therefore, we get

$$\begin{aligned} &P\left(\sigma_1\left(k;\{y_i\}_{i=1}^{mk}\right) \geq r\left(mk\mu_y - (mk)^{0.75}\sigma_y\right)\right) \\ &\quad \leq \binom{mk}{k} P\left(\sum_{i=1}^{k} y_i \geq r\left(mk\mu_y - (mk)^{0.75}\sigma_y\right)\right). \end{aligned} \tag{20}$$

Note that, since $\mathcal{E}\{\sum_{i=1}^{k} y_i\} = k\mu_y$, the probability of the event $\sum_{i=1}^{k} y_i \geq r(mk\mu_y - (mk)^{0.75}\sigma_y)$ is in fact a tail probability for large m. The exponential decay of this tail probability is given by

$$\begin{aligned} P\left(\sum_{i=1}^{k} y_i \geq T\right) &= P\left(e^{\gamma\sum_{i=1}^{k} y_i} \geq e^{\gamma T}\right) \\ &\leq \frac{\mathcal{E}\left\{e^{\gamma\sum_{i=1}^{k} y_i}\right\}}{e^{\gamma T}} = \left(\mathcal{E}\{e^{\gamma y}\}e^{-\gamma\frac{T}{k}}\right)^k \\ &< \left(\frac{e^{\gamma\left(rm\mu_y - \frac{T}{k}\right)}}{3m}\right)^k \end{aligned} \tag{21}$$

where the first inequality is obtained by Markov's inequality. Thus, using $\binom{mk}{k} < (em)^k$ and the results in (20) and (21), we have

$$\begin{aligned} &P\left(\sigma_1\left(k;\{y_i\}_{i=1}^{mk}\right) \geq r\left(mk\mu_y - (mk)^{0.75}\sigma_y\right)\right) \\ &\qquad < \left(\frac{e}{3}e^{\gamma r\sigma_y m^{0.75}k^{-0.25}}\right)^k \end{aligned} \tag{22}$$

which, in combination with (17) and (19), yields

$$\begin{aligned} P(EV_1) &= P\left(\frac{\sigma_1\left(k;\{x_i\}_{i=1}^{mk}\right)}{\sigma_1\left(mk;\{x_i\}_{i=1}^{mk}\right)} \geq r\right) \\ &< \frac{1}{\sqrt{mk}} + \left(\frac{e}{3}e^{\gamma r\sigma_y m^{0.75}k^{-0.25}}\right)^k. \end{aligned} \tag{23}$$

It is easy to check that, for fixed m, the latter upperbound vanishes as $k \to \infty$. Hence, to keep the probability greater than a pre-specified value δ, we require to keep more than k terms among the total of mk, so that

$$\lim_{k\to\infty} \frac{\tilde{\kappa}_p(r,\delta,km;f_x)}{km} \geq \frac{1}{m} \tag{24}$$

where m depends only on r and not k. Consequently, the probability density f_x is not ℓ_1-compressible. ■

Theorem 1 reveals that distributions with exponential decay such as Laplace, Gamma, and Gaussian are not ℓ_1-compressible. Our next step is to show that a class of heavy-tail distributions

fulfills the requirements of Definition 6. However, we need to review a few preliminaries first.

According to the Central Limit Theorem, if $\{x_i\}_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables such that the distribution has finite mean μ_x and variance σ_x^2, then the distribution of the normalized sum $n^{-0.5}\sum_{i=1}^{n}(x_i-\mu_x)$ is a zero-mean Gaussian with variance σ_x^2 in the limit. In addition, the sum of two independent Gaussian random variables is again a Gaussian random variable: the Gaussian distribution is *stable* under the summation operation.

In fact, the class of stable distributions is not limited to the Gaussian case; there is a class of distributions indexed by the parameter $\alpha\in(0,2]$ with the name α-stable, which includes the Gaussian distribution for the special case $\alpha=2$. Except for Gaussians, an α-stable distribution decays like $|t|^{-(\alpha+1)}$ and has finite moments only for orders less than α. None of them has finite variance [13]. Similar to the Gaussian case, the distribution of the normalized sum of i.i.d. random variables (for some distributions) tends to an α-stable law in the limit. For this to happen, the normalization factor must be set to $n^{-\frac{1}{\alpha}}$. These distributions are said to be in the attraction domain of a stable law with index α. For instance, the attraction domain for $\alpha=2$ contains all the distributions with finite variance.

To classify the attraction domain for $0<\alpha<2$, let x be a random variable and define $y=|x|$ and $G(t)=P(y>t)$. It is known that x is in the domain of attraction of a stable law with index $0<\alpha<2$ if and only if [14]

1) the function $t^{\alpha}G(t)$ varies slowly[3] as $t\to\infty$; and
2) the limit $\lim_{t\to\infty}\frac{P(x>t)}{G(t)}$ exists.

Because of their polynomial decay, the distributions that are in the attraction domain of stable laws are suitable candidates for s-compressible distributions. As discussed earlier, we need to study order statistics to inspect the k-term approximation of i.i.d. sequences. The following theorem by Lepage, Woodroofe, and Zinn demonstrates the asymptotic order statistics of the distributions in the attraction domain of stable laws [15].

Theorem 2: Let $\{\eta_i\}$ be an i.i.d. sequence with standard exponential distribution, $f_\eta(t)=e^{-t}u(t)$ ($u(.)$ is the Heaviside step function), and define $\Gamma_i=\sum_{j=1}^{i}\eta_j$. For the i.i.d. sequence $\{x_i\}$ for which the distribution is in the attraction domain of an α-stable law, we have that

$$\lim_{n\to\infty} a_n^{-1}(y_{n1},\ldots,y_{nn},0,0,\ldots)=_d\left(\Gamma_1^{-\frac{1}{\alpha}},\Gamma_2^{-\frac{1}{\alpha}},\ldots\right) \quad (25)$$

where $=_d$ denotes the convergence in distribution (convergence of all finite-dimensional distributions), $\{y_{ni}\}_{i=1}^{n}$ represents the ordered sequence of $\{|x_i|\}_{i=1}^{n}$, and

$$a_n=\lim_{t\to\infty}\frac{G^{-1}\left(\frac{1}{nt^{\alpha}}\right)}{t}. \quad (26)$$

The importance of Theorem 2 is to enable us to study the asymptotic order statistics of a large class of distributions by investigating the i.i.d. sequence of standard exponential distributions. We have now the required tools to establish a link between the s-compressible distributions and the attraction domain of stable laws.

Theorem 3: If the random variable x with the distribution f_x is in the domain of attraction of a stable law with index $\alpha<1$, then f_x is ℓ_1-compressible.

Proof: To prove the s-compressibility, we start by considering the probability involved in Definition 5. Since $\sigma_1(k;\{x_i\}_{i=1}^{n})=\sigma_1(k;\{y_i\}_{i=1}^{n})$, where $y_i=|x_i|$, we have that

$$P\left(\frac{\sigma_1\left(k;\{x_i\}_{i=1}^{n}\right)}{\sigma_1\left(n;\{x_i\}_{i=1}^{n}\right)}\geq r\right) = P\left(\frac{a_n^{-1}\left(y_{n(k+1)}+\cdots+y_{nn}\right)}{a_n^{-1}(y_{n1}+\cdots+y_{nn})}\leq 1-r\right). \quad (27)$$

The s-compressibility definition deals with the asymptotic behavior of the above probability; therefore, based on Theorem 2, replacing $a_n^{-1}y_{ni}$ with $\Gamma_i^{-\frac{1}{\alpha}}$ does not change the probability in the asymptotic case $n\to\infty$. Thus, we can write

$$P\left(\frac{\sigma_1\left(k;\{x_i\}_{i=1}^{n}\right)}{\sigma_1\left(n;\{x_i\}_{i=1}^{n}\right)}\geq r\right)\sim P\left(\frac{\Gamma_{k+1}^{-\frac{1}{\alpha}}+\cdots+\Gamma_n^{-\frac{1}{\alpha}}}{\Gamma_1^{-\frac{1}{\alpha}}+\cdots+\Gamma_n^{-\frac{1}{\alpha}}}\leq 1-r\right). \quad (28)$$

Note that the relation

$$\begin{aligned}\frac{\Gamma_{k+1}^{-\frac{1}{\alpha}}+\cdots+\Gamma_n^{-\frac{1}{\alpha}}}{\Gamma_1^{-\frac{1}{\alpha}}+\cdots+\Gamma_n^{-\frac{1}{\alpha}}} &= \frac{k^{2-\alpha}\left(\frac{\Gamma_1}{\Gamma_{k+1}}\right)^{\frac{1}{\alpha}}}{1+\left(\frac{\Gamma_1}{\Gamma_2}\right)^{\frac{1}{\alpha}}+\cdots+\left(\frac{\Gamma_1}{\Gamma_n}\right)^{\frac{1}{\alpha}}} \times\frac{1+\left(\frac{\Gamma_{k+1}}{\Gamma_{k+2}}\right)^{\frac{1}{\alpha}}+\cdots+\left(\frac{\Gamma_{k+1}}{\Gamma_n}\right)^{\frac{1}{\alpha}}}{k^{2-\alpha}}\\ &\leq \underbrace{k^{2-\alpha}\left(\frac{\Gamma_1}{\Gamma_{k+1}}\right)^{\frac{1}{\alpha}}}_{A_k}\underbrace{\frac{1+\left(\frac{\Gamma_{k+1}}{\Gamma_{k+2}}\right)^{\frac{1}{\alpha}}+\cdots+\left(\frac{\Gamma_{k+1}}{\Gamma_n}\right)^{\frac{1}{\alpha}}}{k^{2-\alpha}}}_{B_{k,n}} \end{aligned} \quad (29)$$

implies that

$$\begin{aligned} P&\left(\frac{\Gamma_{k+1}^{-\frac{1}{\alpha}}+\cdots+\Gamma_n^{-\frac{1}{\alpha}}}{\Gamma_1^{-\frac{1}{\alpha}}+\cdots+\Gamma_n^{-\frac{1}{\alpha}}}\leq 1-r\right)\\ &\geq 1-P(A_kB_{k,n}\geq 1-r)\\ &\geq 1-P(A_k>1)-P(B_{k,n}>1-r)\\ &\geq 1-\mathcal{E}\{A_k\}-\frac{1}{1-r}\mathcal{E}\{B_{k,n}\}. \end{aligned} \quad (30)$$

The last result is obtained by applying Markov's inequality. In the remainder of the proof we show that both $\mathcal{E}\{A_k\}$ and $\mathcal{E}\{B_{k,n}\}$ vanish as $k\to\infty (n>k+2)$.

To show the vanishing property of $\mathcal{E}\{A_k\}$, we first state a lemma which is proved in Appendix A.

[3]The function $f(t)$ is said to vary slowly at $t\to\infty$ if $\lim_{t\to\infty}\frac{f(ct)}{f(t)}=1$ for all $c>0$.

Lemma 3: Let $\{\eta_i\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables with standard exponential distribution and let θ, β be arbitrary positive real numbers. We have that

$$0 \leq \mathcal{E}_{\{\eta_i\}_i}\left\{\frac{1}{\left(1+\frac{1}{\theta}\sum_{i=1}^{k}\eta_i\right)^{\beta}}\right\} < 0.83^k + \left(\frac{2\theta}{k}\right)^{\beta}. \quad (31)$$

Having stated this Lemma B, we now recall the definition of $\mathcal{E}\{A_k\}$, i.e.,

$$\begin{aligned}\mathcal{E}\{A_k\} &= k^{2-\alpha}\mathcal{E}_{\{\eta_i\}_i}\left\{\left(\frac{\Gamma_1}{\Gamma_{k+1}}\right)^{\frac{1}{\alpha}}\right\} \\ &= k^{2-\alpha}\mathcal{E}_{\eta_1}\left\{\mathcal{E}_{\{\eta_i\}_{i>1}}\left\{\left(1+\frac{1}{\eta_1}\sum_{i=2}^{k+1}\eta_i\right)^{-\frac{1}{\alpha}}\right\}\right\}. \quad (32)\end{aligned}$$

Now, the following upper-bound is derived from Lemma 3:

$$\mathcal{E}_{\{\eta_i\}_{i>1}}\left\{\left(1+\frac{1}{\eta_1}\sum_{i=2}^{k+1}\eta_i\right)^{-\frac{1}{\alpha}}\right\} < 0.83^k + \left(\frac{2\eta_1}{k}\right)^{\frac{1}{\alpha}}. \quad (33)$$

Therefore, it stands that

$$\mathcal{E}\{A_k\} \leq k^{2-\alpha}0.83^k + k^{2-\alpha-\frac{1}{\alpha}}\mathcal{E}_{\eta}\left\{(2\eta)^{\frac{1}{\alpha}}\right\}. \quad (34)$$

Since $2 < \alpha + \frac{1}{\alpha}$, it is clear that $\lim_{k\to\infty}\mathcal{E}\{A_k\} = 0$.

To show the vanishing property of $\mathcal{E}\{B_{k,n}\}$, we employ the following lemma (see Appendix B for the proof):

Lemma 4: Let $\{\eta_i\}_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables with standard exponential distribution and let $0 < \theta$ and $1 < \beta$ be arbitrary numbers. We have the following identity:

$$\mathcal{E}_{\{\eta_i\}_i}\left\{1+\sum_{k=1}^{\infty}\frac{1}{\left(1+\frac{1}{\theta}\sum_{i=1}^{k}\eta_i\right)^{\beta}}\right\} = 1+\frac{\theta}{\beta-1}. \quad (35)$$

Before we prove the vanishing property of $\mathcal{E}\{B_{k,n}\}$ with the help of Lemma 4, it is interesting to point out the link between this lemma and the Riemann zeta function: In Lemma 4, for the simple case $\theta = 1$, we are interested in finding the expected value of $1+\sum_{k=1}^{\infty}\Lambda_k^{-\beta}$ where $\{\Lambda_k\}_k$ are random variables such that $\Lambda_k \leq \Lambda_{k+1}$ and $\mathcal{E}\{\Lambda_k\} = k+1$. In the extreme case where the variance of all Λ_ks are zero (removing the randomness), we have that $\Lambda_k = k+1$, and, therefore, the expected value of the summation is $1+\sum_{k=2}^{\infty}k^{-\beta}$. This can be identified with the value of the zeta function at β. Hence, the summation in (35), can be regarded as the randomized version of the zeta function. Similar to the zeta function, this randomized version is also converges when $\beta > 1$ and contains a pole at $\beta = 1$.

Now we get back to $B_{k,n}$, which has

$$\begin{aligned}\mathcal{E}\{B_{k,n}\} &= k^{\alpha-2}\mathcal{E}\left\{1+\sum_{t=k+2}^{n}\left(\frac{\Gamma_{k+1}}{\Gamma_t}\right)^{\frac{1}{\alpha}}\right\} \\ &\leq k^{\alpha-2}\mathcal{E}\left\{1+\sum_{t=k+2}^{\infty}\left(\frac{\Gamma_{k+1}}{\Gamma_t}\right)^{\frac{1}{\alpha}}\right\} \quad (36)\end{aligned}$$

as expectation. Using Lemma 4 we can write

$$\begin{aligned}&\mathcal{E}_{\{\eta_i\}_{i>k+1}}\left\{1+\sum_{t=k+2}^{\infty}\left(\frac{\Gamma_{k+1}}{\Gamma_t}\right)^{\frac{1}{\alpha}}\right\} \\ &\quad= \mathcal{E}_{\{\eta_i\}_{i>k+1}}\left\{1+\sum_{t=k+2}^{\infty}\frac{1}{\left(1+\frac{\sum_{i=k+2}^{t}\eta_i}{\Gamma_{k+1}}\right)^{\frac{1}{\alpha}}}\right\} \\ &\quad= 1+\frac{\alpha}{1-\alpha}\Gamma_{k+1} \quad (37)\end{aligned}$$

which leads to

$$\begin{aligned}\mathcal{E}\{B_{k,n}\} &\leq k^{\alpha-2}\mathcal{E}_{\Gamma_{k+1}}\left\{1+\frac{\alpha}{1-\alpha}\Gamma_{k+1}\right\} \\ &= k^{\alpha-2}\left(1+\frac{\alpha(k+1)}{1-\alpha}\right). \quad (38)\end{aligned}$$

Again, since $\alpha < 1$, the upperbound in the above inequality vanishes as $k \to \infty$. Thus, we have that $\lim_{k\to\infty}\mathcal{E}\{B_{k,n}\} = 0$. Along with (30), this implies that

$$\lim_{k\to\infty} P\left(\frac{\Gamma_{k+1}^{-\frac{1}{\alpha}}+\cdots+\Gamma_n^{-\frac{1}{\alpha}}}{\Gamma_1^{-\frac{1}{\alpha}}+\cdots+\Gamma_n^{-\frac{1}{\alpha}}} \leq 1-r\right) = 1. \quad (39)$$

Note that the upper bounds for $\mathcal{E}\{A_k\}$ in (34) and $\mathcal{E}\{B_{k,n}\}$ in (38), respectively, do not depend on n and vanish as k grows. Hence, for a lower bound $1-\frac{\epsilon}{2}$ (arbitrary $0 < \epsilon < 1$) on the probability in (39), it is sufficient to choose $k \geq k_0$. On the other hand, Theorem 2 states that for this fixed k_0 (which depends on ϵ) and for large-enough values of n (above a threshold), the difference between the two probabilities in (28) would be less than $\frac{\epsilon}{2}$. Hence, for this fixed k_0 and for large-enough values of n, we conclude that

$$P\left(\frac{\sigma_1(k_0;\{x_i\}_{i=1}^n)}{\sigma_1(n;\{x_i\}_{i=1}^n)} \geq r\right) \geq 1-\epsilon. \quad (40)$$

Obviously, when n grows we have that $\frac{k_0}{n} \to 0$, which completes the proof. ∎

Theorem 3 roughly states that a heavy-tail distribution $f_x(t)$ that decays as $|t|^{-(\alpha+1)}$ for $\alpha < 1$ is ℓ_1-compressible. At first glance, it seems that this theorem is of no use when the decay is faster than $|t|^{-2}$, particularly for those distributions in the attraction domain of stable laws with $1 \leq \alpha < 2$, or even for those heavy-tail distributions in the attraction domain of Gaussian distribution that decay faster than $|t|^{-3}$ and have a finite variance. However, since the distribution of $y = |x|^p$ is $f_y(t) = \frac{f_x\left(t^{\frac{1}{p}}\right)+f_x\left(-t^{\frac{1}{p}}\right)}{pt^{1-\frac{1}{p}}}$, if f_x decays as $|t|^{-(q+1)}$, then f_y will decay as $|t|^{-(\frac{q}{p}+1)}$; hence, according to Lemma 2, f_x is ℓ_p-compressible for $p > q$.

In summary, the main results in this section are as follows:

1) if the distribution of a random variable x is such that $\mathcal{E}_x\{e^{\gamma x}\}$ is finite for some $\gamma > 0$, then the distribution is not ℓ_p-compressible for $p \leq 1$; and
2) if the distribution of a random variable x decays as $|t|^{-(q+1)}$, then it is ℓ_p-compressible for $p > q$.

TABLE I
SOME EXAMPLES OF HEAVY-TAIL DISTRIBUTIONS

Distribution	$f_x(t)$
Generalized Pareto	$\frac{q}{2\lambda}\left(1+\frac{\lvert t\rvert}{\lambda}\right)^{-(q+1)}$
Student's t	$\frac{\Gamma((q+1)/2)}{\Gamma(q/2)\sqrt{2\pi\lambda}}\left(1+\frac{t^2}{2\lambda}\right)^{-(q+1)/2}$
Extreme Value	$\frac{q}{2\lambda}e^{-(1+\frac{\lvert t\rvert}{\lambda})^{-q}}\left(1+\frac{\lvert t\rvert}{\lambda}\right)^{-(q+1)}$
Log-Logistic	$\frac{(q/\lambda)(t/\lambda)^{q-1}}{\left(1+(t/\lambda)^q\right)^2}$

IV. DISCUSSION

We proposed in Section II a definition of the compressibility of infinite sequences (called s-compressibility) that was not based on the decay of the sequence; however, all of our examples so far were decaying sequences. Here, we analyze our definition in more details and introduce nondecaying s-compressible sequences.

Assume that we have a monotonically decreasing sequence of positive numbers which is s-compressible. Whenever we truncate the sequence, the best k-term approximation of this finite sequence is, in fact, the first k elements. Note that only the energy of the best k terms plays a role in Definition 4 and not their location: The s-compressibility status of the sequence would not have changed if these k significant elements would have been located at the end of the truncated subsequence. Consequently, a monotonically increasing sequence of numbers where the increasing rate is similar to the decreasing rate of a decaying s-compressible sequence is also compressible.

A. An Exponentially Increasing S-Compressible Sequence

We show that the exponentially increasing sequence of $a_i = 2^i$ is ℓ_p-compressible for any p. The approach is similar to the one presented for $a_i = 2^{-i}$. Let $0 \le r < 1$ and $0 < p$ be arbitrary numbers and set $k = \left\lceil -\frac{\log_2(1-r^p)}{p} \right\rceil$. For $n \ge k$ we have that

$$\frac{\sigma_p\left(k;\{a_i\}_{i=1}^n\right)}{\sigma_p\left(n;\{a_i\}_{i=1}^n\right)} = \left(\frac{\sum_{i=n-k+1}^{n} 2^{ip}}{\sum_{i=1}^{n} 2^{ip}}\right)^{\frac{1}{p}} = \left(\frac{\sum_{i=1}^{k} 2^{-ip}}{\sum_{i=1}^{n} 2^{-ip}}\right)^{\frac{1}{p}} > (1-2^{-kp})^{\frac{1}{p}} \ge r. \quad (41)$$

Similarly to the arguments used for $a_i = 2^{-i}$, this sequence is also ℓ_p-compressible for all $0 < p$.

B. S-Compressible Distributions

In Section III we showed that heavy-tail distributions are in some sense compressible. To show some examples, we have listed a few polynomial-decaying distributions in Table I. In fact, those are the compressible distributions introduced in [9]. It is easy to check that these distributions decay like $t^{-(q+1)}$ and therefore, they are ℓ_p-compressible for $p > q$. To verify the compressibility of these distributions based on Definition 1, note that for $y = |x|$, we have

$$1 - F_y(t) = \int_t^\infty f_y(\tau)d\tau \propto t^{-q} \Rightarrow F_y^{-1}\left(1-\frac{i}{n+1}\right) \propto \left(\frac{i}{n+1}\right)^{-\frac{1}{q}} \quad (42)$$

which suggests $R(n,\theta) \propto (n+1)^{\frac{1}{q}}$ and $p(n,\theta) = q$, as required in (2).

In addition to the examples in Table I, all α-stable distributions for $0 < \alpha < 2$ (which include Cauchy and Lévy distributions) are ℓ_p-compressible for $p > \alpha$.

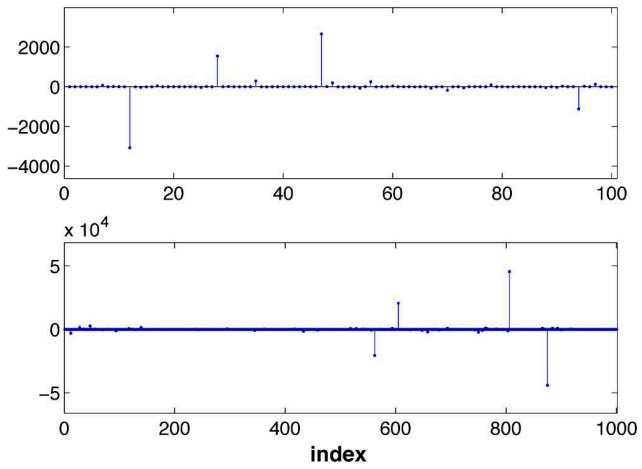


Fig. 1. A realization of an i.i.d. sequence of random variables with Student's t distribution ($q = 0.5$).

C. Realizations of an i.i.d. Sequence

We illustrate in Fig. 1, a realization of an i.i.d. sequence with Student's t distribution ($q = 0.5$; decaying like $t^{-1.5}$). In the upper graph, we focus on the first $n = 100$ terms, while we show a broader view of the same sequence in the lower graph. We observe that the significant samples in the first 100 terms become insignificant when 1000 terms are considered. In other words, as n increases, we observe larger peaks in the sequence, which leads to dramatic changes in the energy of the truncated sequence depending on whether or not these peaks are considered.

D. Energy Fraction in Practice

To visualize the concept of s-compressibility in i.i.d. sequences, we have plotted the average curves of $\frac{\sigma_1(k;\{a_i\}_{i=1}^n)}{\sigma_1(n;\{a_i\}_{i=1}^n)}$ for Gaussian, Laplace, Cauchy, and Student's $t(q = 0.5)$ distributions in Fig. 2. We set $n = 10^4$ and averaged the curves over 500 different realizations. Theorem 3 implies that Student's t distribution with $q = 0.5$ is ℓ_1 compressible. In Fig. 2, we observe that its energy fraction curve saturates at very small values of $\frac{k}{n}$. According to Theorem 3, the curve for Student's t distribution will be a step function in the limiting case of $n \to \infty$. Conversely, Gaussian and Laplace distributions are not compressible in any sense and we expect their asymptotic curves ($n \to \infty$) to be close to the ones depicted

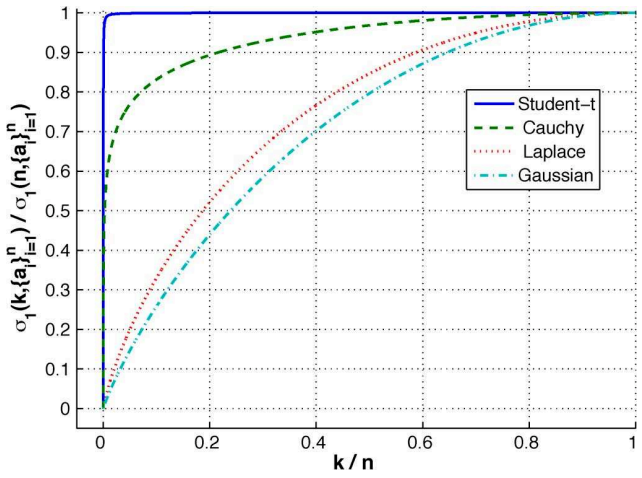


Fig. 2. Average curve of $\frac{\sigma_p(k;\{a_i\}_{i=1}^n)}{\sigma_p(n;\{a_i\}_{i=1}^n)}$ with respect to $\frac{k}{n}$ when $p=1$ and $n=10^4$ for i.i.d. realizations of the Gaussian, Laplace, Cauchy, and Student's $t(q=0.5)$ distributions.

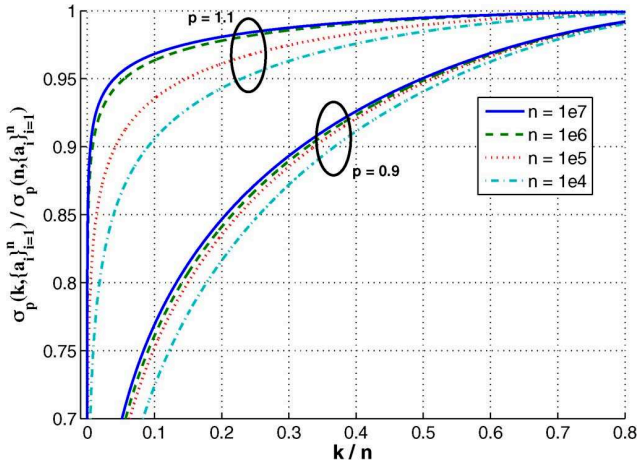


Fig. 3. Average curve of $\frac{\sigma_p(k;\{a_i\}_{i=1}^n)}{\sigma_p(n;\{a_i\}_{i=1}^n)}$ with respect to $\frac{k}{n}$ for $p=0.9$ and $p=1.1$ with different sequence lengths n when i.i.d. realizations of the Cauchy distribution are considered.

in Fig. 2. Meanwhile, the fate of the Cauchy distribution is as yet undecided; since it decays like t^{-2}, we know that it is ℓ_p-compressible for $p>1$; however, our results are not enough to decide for $p=1$.

E. Effect of p

To investigate the role of p in s-compressibility of distributions, we have plotted the average energy fraction curves for Cauchy distribution with $p=0.9$ and $p=1.1$ and $n=10^i$, where $i\in\{4,5,6,7\}$. As predicted by Theorem 3, the curves for $p=1.1$ approach the step function when n increases. The curves for $p=0.9$ seem to have a continuous limit.

F. Linear Transformations

It is interesting to mention that, since α-stable distributions satisfy a generalized form of the central-limit theorem, they are closed under linear transformations. In other words, the stability index of a stable law does not change whatever the linear-transformation may be. This suggests that the s-compressibility status of a sequence can be determined in any domain. However, due to the requirement of independence, the k-term approximation of the sequence should be carried out preferentially in the innovation domain.

V. Conclusion

In this paper, we first proposed a definition for compressibility of infinite sequences (s-compressibility) that is consistent with the intuition and is mathematically exploitable. We then defined compressibility and sparsity for i.i.d. random sequences. Our definitions are based on the asymptotic distribution of the energy among the elements in truncated subsequences. We showed that some exponentially decaying probability density such as the Laplace distribution are not proper candidates for a s-compressible prior. This arises because they distribute the energy in the sequence rather uniformly. On the other hand, by using order statistics, we demonstrated the connection between the s-compressibility of a heavy-tail distribution and the stability index to which it is attracted.

Appendix A
Proof of Lemma 3

Since η_i and θ are non-negative, the left inequality is trivial. Defining $\Gamma_k=\sum_{i=1}^k\eta_i$, we can write that

$$\begin{aligned}\mathcal{E}_{\{\eta_i\}}\left\{\frac{1}{\left(1+\frac{\Gamma_k}{\theta}\right)^\beta}\right\}&=\int_0^\infty\frac{1}{\left(1+\frac{\gamma}{\theta}\right)^\beta}f_{\Gamma_k}(\gamma)d\gamma\\&=\int_0^{\frac{k}{2}}\frac{1}{\left(1+\frac{\gamma}{\theta}\right)^\beta}f_{\Gamma_k}(\gamma)d\gamma\\&\quad+\int_{\frac{k}{2}}^\infty\frac{1}{\left(1+\frac{\gamma}{\theta}\right)^\beta}f_{\Gamma_k}(\gamma)d\gamma\\&\le\int_0^{\frac{k}{2}}f_{\Gamma_k}(\gamma)d\gamma+\int_{\frac{k}{2}}^\infty\left(\frac{2\theta}{k}\right)^\beta f_{\Gamma_k}(\gamma)d\gamma\\&\le P\left(\Gamma_k\le\frac{k}{2}\right)+\left(\frac{2\theta}{k}\right)^\beta.\end{aligned}\tag{43}$$

Since Γ_k is a sum of k i.i.d. random variables with unit mean, we have $\mathcal{E}\{\Gamma_k\}=k$. Thus, $P(\Gamma_k\le\frac{k}{2})$ is a tail probability. The following approach is a classical method to show the exponential decay of the tail probability:

$$\begin{aligned}P\left(\Gamma_k\le\frac{k}{2}\right)&=P\left(-\sum_{i=1}^k\eta_i\ge-\frac{k}{2}\right)\\&=P\left(e^{-\sum_{i=1}^k\eta_i}\ge e^{-\frac{k}{2}}\right)\\&\le\frac{\mathcal{E}_{\{\eta_i\}_i}\left\{e^{-\sum_{i=1}^k\eta_i}\right\}}{e^{-\frac{k}{2}}}\\&=e^{\frac{k}{2}}\left(\mathcal{E}_\eta\{e^{-\eta}\}\right)^k=\left(\frac{\sqrt{e}}{2}\right)^k<0.83^k\end{aligned}\tag{44}$$

where the first inequality is obtained by applying Markov's inequality. ■

APPENDIX B
PROOF OF LEMMA 4

Similarly to the proof of Lemma 3, let $\Gamma_k = \sum_{i=1}^{k} \eta_i$. Since η_is are i.i.d., the probability density function of Γ_k is found by convolving k times the standard exponential distribution with itself:

$$f_{\Gamma_k}(\gamma) = \left(\underbrace{f_\eta * \cdots * f_\eta}_{k-\text{times}} \right)(\gamma) = \mathcal{F}_\omega^{-1}\left\{(1+j\omega)^{-k}\right\}(\gamma). \quad (45)$$

Thus, we have that

$$\mathcal{E}_{\{\eta_i\}}\left\{\frac{1}{\left(1+\frac{\Gamma_k}{\theta}\right)^\beta}\right\} = \int_0^\infty \int_{\mathbb{R}} \frac{1}{\left(1+\frac{\gamma}{\theta}\right)^\beta} \frac{e^{j\omega\gamma}}{(1+j\omega)^k} \frac{d\omega}{2\pi} d\gamma \quad (46)$$

which yields

$$\begin{aligned}
&\mathcal{E}_{\{\eta_i\}_i}\left\{1+\sum_{k=1}^{\infty}\frac{1}{\left(1+\frac{\Gamma_k}{\theta}\right)^\beta}\right\} \\
&= \int_0^\infty \int_{\mathbb{R}} \frac{e^{j\omega\gamma}}{\left(1+\frac{\gamma}{\theta}\right)^\beta} \sum_{k=0}^{\infty} \frac{1}{(1+j\omega)^k} \frac{d\omega}{2\pi} d\gamma \\
&= \int_0^\infty \int_{\mathbb{R}} \frac{e^{j\omega\gamma}}{\left(1+\frac{\gamma}{\theta}\right)^\beta} \left(1+\frac{1}{j\omega}\right) \frac{d\omega}{2\pi} d\gamma \\
&= \int_{\mathbb{R}} \frac{\delta(\gamma)+1}{\left(1+\frac{\gamma}{\theta}\right)^\beta} d\gamma = 1 + \frac{\theta}{\beta-1} < \infty. \qquad (47)
\end{aligned}$$

■

ACKNOWLEDGMENT

The authors would like to thank P. Thévenaz and P. Tafti for their fruitful comments.

REFERENCES

[1] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
[2] E. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
[3] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
[4] R. A. DeVore, "Nonlinear approximation," *Acta Numer.*, vol. 7, pp. 51–150, 1998.
[5] E. J. Candes, "Compressive sampling," in *Proc. Int. Congr. Math.*, 2006, vol. 3, pp. 1433–1452.
[6] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k-term approximation," *J. Amer. Math. Soc.*, vol. 22, pp. 211–231, 2009.
[7] D. P. Wipf and B. D. Rao, "Sparse Baysian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
[8] T. Park and G. Casella, "The Bayesian lasso," *J. Amer. Stat. Assoc.*, vol. 103, pp. 681–686, Jun. 2008.
[9] V. Cevher, "Learning with compressible priors," presented at the Neural Inf. Process. Syst. (NIPS), Vancouver, B.C., Canada, 2008.
[10] V. Cevher, "Approximate distributions for compressible signals," in *Proc. IEEE Inf. Theory Workshop*, Oct. 2009, p. 353.
[11] Y. Zhang, J. Schneider, and A. Dubrawski, "Learning compressible models," in *SIAM Int. Conf. Data Mining*, Apr. 2010, pp. 872–881.
[12] R. Gribonval, V. Cevher, and M. Davies, "Compressible priors for high-dimensional statistics," Feb. 2011 [Online]. Available: http://arxiv.org/abs/1102.1249
[13] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes*. London, U.K.: Chapman & Hall/CRC, 1994.
[14] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York: Wiley, 1991, vol. 2.
[15] R. Lepage, M. Woodroofe, and J. Zinn, "Convergence to a stable distribution via order statistic," *Ann. Probability*, vol. 9, no. 4, pp. 624–632, 1981.

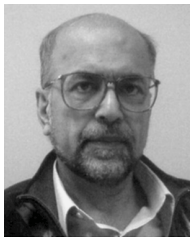
Arash Amini received B.Sc., M.Sc., and Ph.D. degrees in electrical engineering (communications and signal processing) in 2005, 2007, and 2011, respectively, and the B.Sc. degree in petroleum engineering (Reservoir) in 2005, all from Sharif University of Technology (SUT), Tehran, Iran.

Since April 2011, he has been a Postdoctoral Researcher in the Biomedical Imaging Group (BIG), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. His research interests include different aspects of sampling, specially compressed sensing.

Michael Unser (M'89–SM'94–F'99) received the M.S. (*summa cum laude*) and Ph.D. degrees in electrical engineering from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 1981 and 1984, respectively.

From 1985 to 1997, he worked as a Scientist with the National Institutes of Health, Bethesda, MD. He is currently a Full Professor and Director of the Biomedical Imaging Group at the EPFL. His main research area is biomedical image processing. He has a strong interest in sampling theories, multiresolution algorithms, wavelets, and the use of splines for image processing. He has published 200 journal papers on those topics.

Dr. Unser has held the position of Associate Editor-in-Chief for the IEEE TRANSACTIONS ON MEDICAL IMAGING from 2003 to 2005 and has served as Associate Editor for the same journal from 1999 to 2002 and from 2006 to 2007), the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1992 to 1995, and the IEEE SIGNAL PROCESSING LETTERS from 1994 to 1998. He is currently member of the editorial boards of Foundations and Trends in Signal Processing, and Sampling Theory in Signal and Image Processing. He co-organized the first IEEE International Symposium on Biomedical Imaging (ISBI 2002) and was the founding chair of the technical committee of the IEEE-SP Society on Bio Imaging and Signal Processing (BISP). He received the 1995 and 2003 Best Paper Awards, the 2000 Magazine Award, and two IEEE Technical Achievement Awards (2008 SPS and 2010 EMBS). He is an EURASIP Fellow and a member of the Swiss Academy of Engineering Sciences. He is one of ISIs Highly Cited authors in Engineering (http://isihighlycited.com).



Farokh Marvasti (S'72–M'74–SM'83) received the B.S., M.S., and Ph.D. degrees all from Rensselaer Polytechnic Institute, Troy, NY, in 1970, 1971, and 1973, respectively.

He has worked, consulted, and taught in various industries and academic institutions since 1972, among which are Bell Labs, the University of California Davis, the Illinois Institute of Technology, the University of London, and King's College. He is currently a Professor at Sharif University of Technology, Tehran, Iran, and the Director of the Advanced Communications Research Institute (ACRI) and head of the Center for Multi-Access Communications Systems. He has about 100 journal publications and has written several reference books, his last book being *Nonuniform Sampling: Theory and Practice* (Kluwer, 2001). He also has several international patents.

Dr. Marvasti was one of the editors and associate editors of the IEEE TRANSACTIONS ON COMMUNICATIONS and the TRANSACTIONS ON SIGNAL PROCESSING from 1990 to 1997. He was also a Guest Editor on the May 2008 Special Issue on Nonuniform Sampling for the *Sampling Theory & Signal and Image Processing* journal. Besides being the co-founder of two international conferences (ICTs and SampTAs), he has been the organizer and special session chair of many IEEEE conferences, including ICASSP conferences.