

Compressibility of Infinite Sequences and its Interplay with Compressed Sensing Recovery

Jorge F. Silva and Eduardo Pavez
 Department of Electrical Engineering
 Information and Decision Systems Group
 Universidad de Chile

E-mails: josilva@ing.uchile.cl, epavez@ing.uchile.cl

Abstract—This work elaborates connections between notions of compressibility of infinite sequences, recently addressed by Amini *et al.* [1], and the performance of the compressed sensing (CS) type of recovery algorithms from linear measurements in the under-sample scenario. In particular, in the asymptotic regime when the signal dimension goes to infinity, we established a new set of compressibility definitions over infinite sequences that guarantees arbitrary good performance in an ℓ_1 -noise to signal ratio (ℓ_1 -NSR) sense with an arbitrary close to zero number of measurements per signal dimension.

I. INTRODUCTION

In the case of finite dimensions, compressed sensing (CS) has established sufficient conditions over a set of lineal measurements [2], [3], for perfect recovery and near-optimal reconstruction under a notion of signal sparsity and compressibility [4], [5], [6], [7]. Remarkably, these results cover the under-sampled regime where the number of measurements is strictly less than the signal dimension, and consequently perfect recovery over all possible signals cannot be achieved. In a nutshell, CS captures signal structure by means of a notion of sparsity (and compressibility), where restricted to these signal families, a way of doing perfect signal decoding (and near optimal best k -term approximation) is achieved by means of an ℓ_1 -minimization algorithm, which is implementable by linear programming [4]. The other critical ingredient in the CS theory is the role of random measurements that allow (with very large probability) perfect recovery under an $O(k \log(N/k))$ number of measurements [3], with k being the sparsity of the signal and N the signal dimension.

Moving on the countable infinity case (i.e., the space of sequences), Amini *et al.* [1] have introduced definitions of ℓ_p -compressibility for deterministic and random infinity sequences. Remarkably, these definitions allow the signals to have an infinite ℓ_p -norm, as they are based on a relative norm concept (the ratio between the norm of the best k -term approximation and the norm of the whole signal), reminiscent of an ℓ_p -signal to noise ratio (ℓ_p -SNR) fidelity measure. This relative fidelity measure allows capturing a wider spectrum of signal collections, which was particularly relevant when the authors addressed the random case generated from an independent and identically distributed (i.i.d.) process [1, Sec III, Th. 1]. In summary, based on their notion of ℓ_p -compressibility for deterministic and random sequences, they offered the basis

for categorizing sequences in a way that could be of interest for CS performance recovery.

In this work we move on in this direction, and provide a connection between ℓ_1 -compressibility notions for deterministic sequences and their respective performance in a CS (lineal measurement coding and ℓ_1 -minimization decoding) setting. In particular, we revisit and propose new compressibility definitions that provide a concrete categorization of sequences over which the CS signal representation setting achieves an arbitrary close to zero ℓ_1 -SNR distortion under arbitrary close to zero number of measurements per signal dimension. We show that these performances are obtained under different scenarios with compressibility notions that are stronger than original definitions of Amini *et al.* [1]. In summary, these results provide concrete connections between signal structure and guarantee performance for an infinitely countable CS setting, with focus on the zero measurement per signal dimension asymptotic regime.

The rest of the paper is organized as follows: Section II introduces the basic elements of the CS theory needed in the work. Section III is the main content section that presents the new definitions and CS recovery results. Section IV explores the important case of ℓ_1 -sequences, and finally Section V offers some directions for future work.

II. COMPRESSED SENSING

Let $x \in \mathbb{R}^N$ be a finite dimensional vector, where the usual ℓ_p -norm is given by

$$\|x\|_{\ell_p} = \left(\sum_{n=1}^N |x_n|^p \right)^{\frac{1}{p}}, \quad \forall p > 1. \quad (1)$$

We define the support of a x by $support(x) \equiv |\{i : |x_i| > 0\}| \leq N$. The signal x is said to be k -sparse if $support(x) \leq k$, and the collection of k -sparse signals is denoted by Σ_k . For an arbitrary signal $x \in \mathbb{R}^N$, x_k denotes its best k -term approximation, i.e., x_k retains the k -largest entries of x and zeroes out the rest of the coefficients¹. In this context the k -term approximation error of x is denoted by $\sigma_k(x)_p \equiv \inf_{\tilde{x} \in \Sigma_k} \|x - \tilde{x}\|_{\ell_p}$.

CS puts attention on the problem of recovering sparse signals through linear measurements. The CS setting can be

¹Alternatively, $x_k \equiv \arg \min_{\tilde{x} \in \Sigma_k} \|x - \tilde{x}\|_{\ell_1}$ for all $k \in \{1, \dots, N\}$.

seen as a coding-decoding framework over the operational constraint of a finite set of linear measurements [8], [3], [4]. The encoding function is a linear operator $\phi : \mathbb{R}^N \rightarrow \mathbb{R}^m$, that generates a measurement vector by $y = \phi x \in \mathbb{R}^m$. The main focus is on the under-sample regime, i.e., $m < N$, where the problem of recovering x from y is ill-posed. However, the main CS conjecture was that on constraining the signal domain to a collection of sparse signals, perfect reconstruction could be achieved when $m < N$. Furthermore, CS proposes a concrete and implementable (by linear programming) decoding function $\Delta^* : \mathbb{R}^m \rightarrow \mathbb{R}^N$ given by

$$\Delta^*(y) = \arg \min_{\{\tilde{x} \in \mathbb{R}^N : y = \phi \tilde{x}\}} \|\tilde{x}\|_{\ell_1}. \quad (2)$$

Remarkably, CS theory establishes sufficient conditions over ϕ (and implicitly over the number of measurements m) in order that $x = \Delta^*(\phi x)$ when $x \in \Sigma_k$ for some $k < N$. Here, we revisit the results derived from the celebrated *restricted isometry property* (RIP) [2], [4], which is the one we use in this work.

Definition 1: [2, Def. 1.1] Given a matrix ϕ and $k \in \{1, \dots, N\}$, the isometry constant δ_k of ϕ is the smaller non-negative number such that $\forall x \in \Sigma_k$

$$(1 - \delta_k) \cdot \|x\|_{\ell_2}^2 \leq \|\phi x\|_{\ell_2}^2 \leq (1 + \delta_k) \cdot \|x\|_{\ell_2}^2. \quad (3)$$

The isometry definition in (3) is the key to obtaining the following recovery result:

THEOREM 1: [2, Th. 1.2] If the measurement matrix ϕ has an isometry constant $\delta_{2k} < \sqrt{2} - 1$ for some $k > 0$, then

$$\|x - x^\ddagger\|_{\ell_1} \leq C_0 \|x - x_k\|_{\ell_1} = C_0 \cdot \sigma_k(x)_1 \quad (4)$$

where x^\ddagger is the solution of (2) and C_0 is a universal constant that is only a function of δ_k . In this case, we say that ϕ satisfies the RIP property for k .

Hence for a given level of approximation k , it is critical to characterize the minimum number of measurements m and the structure of the matrix ϕ that satisfy the RIP property of Theorem 1. The following result, in its original form stated in [5], shows that random measurements offer a solution to the problem of constructing matrixes that satisfy the RIP property with a near optimal relationship between m and k [8].

THEOREM 2: [3, Th. 5.2] Let $\phi(w)$ be a random matrix², $w \in \Omega^{mN}$, 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 $\{1/\sqrt{m}, -1/\sqrt{m}\}$. Then for any arbitrary number $\delta \in (0, 1)$ and $k \leq N$, if $m \geq C_1 k \log \frac{N}{k}$, $\phi(w)$ satisfies the condition in (3) with respect to δ and k with a probability over the space Ω^{mN} at least equal to $1 - 2^{-C_2 \cdot m}$, where C_1 and C_2 are universal constants that are only function of δ .

In summary, Theorems 1 and 2 provide sufficient conditions over the number of measurements m of concrete random matrixes (the Gaussian and Bernoulli random matrixes) to

²This result can be generalized to random matrixes satisfying a concentration inequality, which is not reported here for space considerations. See details on this observation in [3].

achieve perfect reconstruction for k -sparse signals, or near optimal k -term approximation errors for any arbitrary signal in \mathbb{R}^N . This can be obtained with high probability over the sample space of random matrixes, i.e., over Ω^{mN} .

III. COMPRESSIBILITY AND CS PERFORMANCE GUARANTEE IN THE ASYMPTOTIC REGIME

Here we are interested in the role of CS coding-decoding for sequences $(x_n) \in \mathbb{R}^N$ and in the kind of compressibility notion in this domain that provides guaranteed performance in this context. We follow the approach proposed by Amini *et al.* [1], where the analysis is based on truncating the sequence to a finite dimension N and then taking the limit as $N \rightarrow \infty$. We begin revisiting the definitions introduced in [1] and include some new extensions.

Let $x = (x_n)$ be a sequence in \mathbb{R}^N . We denote $x^N \equiv (x_i)_{i=1}^N$ the N -truncated version of x and $(x_{ni})_{i=1}^N$ its N -order statistics, where $|x_{n1}| \geq |x_{n2}| \dots \geq |x_{nN}|$. Let

$$\varsigma_p(k, x^N) \equiv (|x_{n1}|^p + \dots + |x_{nk}|^p)^{\frac{1}{p}}, \quad \forall k \in \{1, \dots, N\} \quad (5)$$

be the ℓ_p -norm concentrated in the k -most significant entries of x^N , and let us define

$$\kappa_p(r, x^N) \equiv \min \left\{ k \in \{1, \dots, N\} : \frac{\varsigma_p(k, x^N)}{\|x^N\|_{\ell_p}} \geq r \right\}, \quad (6)$$

for all $r \in (0, 1]$.

Definition 2: (Amini *et al.* [1, Def. 4]) A sequence $(x_n) \in \mathbb{R}^N$ is said to be ℓ_p -compressible, if $\forall r \in (0, 1)$,

$$\lim_{N \rightarrow \infty} \frac{\kappa_p(r, x^N)}{N} = 0. \quad (7)$$

Definition 3: (x_n) is said to be strongly ℓ_p -compressible if $\forall r \in (0, 1)$,

$$\lim_{N \rightarrow \infty} \frac{\kappa_p(r, x^N) \cdot \log N}{N} = 0. \quad (8)$$

Definition 4: (x_n) is said to be τ -power dominated for ℓ_p for some $\tau \in (0, 1)$, if $\forall r \in (0, 1)$,

$$(\kappa_p(r, x^N))_{N>0} \text{ is } O(N^\tau). \quad (9)$$

Definition 5: (x_n) is said to be asymptotically sparse for ℓ_p , if $\forall r \in (0, 1)$,

$$\limsup_{N \rightarrow \infty} \kappa_p(r, x^N) < \infty. \quad (10)$$

Note that all these definitions of compressibility are hierarchical, from the weakest in Definition 2 to the strongest in Definition 5. Finally in this countable infinite setting, we can naturally say that (x_n) is k -sparse if $|\text{support}(x_n)| \leq k$.

Here we are interested in the recovery of x with a family of CS encoding-decoding pairs $\{(\phi_{m_N \times N}, \Delta_N^*) : N > 0\}$ of different signal-lengths, where for a fixed dimension $N > 0$ the encoder is given by a linear operator:

$$y^N \equiv \phi_{m_N \times N} x^N \quad (11)$$

from \mathbb{R}^N to \mathbb{R}^{m_N} , and the decoding process $x^{\ddagger N} \equiv \Delta_N^*(y^N) \in \mathbb{R}^N$ is the solution of the ℓ_1 -minimization in (2).

Sequences in \mathbb{R}^N have an infinite norm in general, then as a performance metric (or distortion measure) we consider the notion of *noise to signal ratio* (NSR) in the ℓ_p -sense given by:

$$D_{\ell_p}((\phi_{m_N \times N}, \Delta_N^*); x^N) \equiv \frac{\|x^{\dagger N} - x^N\|_{\ell_p}}{\|x^N\|_{\ell_p}} \in [0, 1]. \quad (12)$$

On the other hand, as a cost measure for the pair $(\phi_{m_N \times N}, \Delta_N^*)$ we consider the number of linear measurements per signal dimension given by:

$$R(\phi_{m_N \times N}, \Delta_N^*) \equiv m_N/N \in (0, 1]. \quad (13)$$

A. Asymptotic CS Recovery Results

From this point on, we will focus exclusively on ℓ_1 -norm compressibility as is seen in Theorem 1 (the right-hand side of (4)), it is the type of approximation condition over the signals that could provide guaranteed performance. We begin introducing an asymptotic definition of performance for the random CS scheme of Theorem 2.

Definition 6: The pair $(d, c) \in [0, 1) \times [0, 1)$ is achievable for $(x_n) \in \mathbb{R}^N$ adopting the standard random CS scheme $\{(\phi_{m_N \times N}(w), \Delta_N^*) : N > 0\}$ of Theorem 2, if there is a sequence (m_N) such that

$$\begin{aligned} \limsup_{N \rightarrow \infty} R(\phi_{m_N \times N}(w), \Delta_N^*) &\leq c \text{ and} \\ \lim_{N \rightarrow \infty} \mathbb{P}(\{w \in \Omega^{m_N \cdot N} : D_{\ell_1}((\phi_{m_N \times N}(w), \Delta_N^*); x^N) \leq d\}) &= 1, \end{aligned}$$

where \mathbb{P} denotes the process distribution of the random matrix process $\{\phi_{m_N \times N}(w) : N \geq 1\}$.

We can state the following result which is a corollary of Theorem 1:

LEMMA 1: Let (x_n) be an arbitrary sequence and $(\phi_{m_N \times N}, \Delta_N^*)$ be a CS encoding-decoding pair. If $\phi_{m_N \times N}$ satisfies the RIP of Theorem 1 for all $k \leq \kappa_1(r, x^N)$, then

$$D_{\ell_1}((\phi_{m_N \times N}, \Delta_N^*); x^N) \leq C_0 \cdot (1 - r). \quad (14)$$

Proof: In particular, $\phi_{m_N \times N}$ satisfies the RIP with respect to $k = \kappa_1(r, x^N)$, hence from (4),

$$\frac{\|x - x^\dagger\|_{\ell_1}}{\|x^N\|_{\ell_1}} \leq C_0 \frac{\|x - x_k\|_{\ell_1}}{\|x^N\|_{\ell_1}} \leq C_0(1 - r), \quad (15)$$

this last inequality by definition of $\kappa_1(r, x^N)$. \blacksquare

THEOREM 3: Let (x_n) be ℓ_1 -strongly compressible, then $\forall d \in (0, 1)$ there is $\{(\phi_{m_N \times N}(w), \Delta_N^*) : N > 0\}$ a random CS scheme³, where

$$\begin{aligned} \lim_{N \rightarrow \infty} \mathbb{P}(\{w \in \Omega^{m_N \cdot N} : D_{\ell_1}((\phi_{m_N \times N}(w), \Delta_N^*); x^N) \leq d\}) &= 1 \text{ and} \\ \lim_{N \rightarrow \infty} R(\phi_{m_N \times N}, \Delta_N^*) &= 0. \end{aligned}$$

In other words, for any strongly compressible signal and distortion level, there is a standard random CS scheme that achieves that distortion, under the asymptotic regime of zero

³ $\phi_{m_N \times N}(w)$ is either the Gaussian or Bernoulli random matrix of Theorem 2.

measurement per signal dimension, and with an arbitrary probability close to one. Hence, this result provides the operational interpretation of the notion of asymptotic compressibility stated in Definition 3.

Proof of Theorem 3: Let $d \in (0, 1)$ be an arbitrary number and $r \in (0, 1)$ such that $C_0 \cdot (1 - r) < d$. Let us consider the sequence $(\kappa_1(r, x^N))$ and take $(m_N) \sim (\kappa_1(r, x^N) \cdot \log N)$. In this context, Theorem 2 says that for an arbitrary number $N > 1$, $\phi_{m_N \times N}(w)$ satisfies the RIP property of Theorem 1 with respect to $k \leq \kappa_1(r, x^N)$ with probability at least equal to $1 - 2^{C_2 \cdot m_N}$. Therefore from Lemma 1,

$$\begin{aligned} \mathbb{P}(\{w : D_{\ell_1}((\phi_{m_N \times N}(w), \Delta_N^*); x^N) \leq C_0 \cdot (1 - r)\}) & \\ \geq 1 - 2^{C_2 \cdot m_N}. & \end{aligned} \quad (16)$$

Finally, the facts that $m_N \rightarrow \infty$ and $C_0 \cdot (1 - r) < d$ conclude the argument. \blacksquare

It is worth noting that the proof of this result uses a CS scheme that depends on the signal and the level of distortion $((m_N) \sim (\kappa_1(r, x^N) \cdot \log N)$ with $r > 1 - d/C_0$), which is a kind of oracle result. The following result takes advantage of the universal nature⁴ of the RIP property obtained for random matrices in Theorem 2, and offers an universal CS coding-decoding construction over a family of signals and distortion levels.

PROPOSITION 1: Let us consider a random CS scheme $\{(\phi_{m_N \times N}(w), \Delta_N^*) : N > 0\}$ with $m_N \sim N^\rho \cdot \log N$ with $\rho \in (0, 1)$. Then for all (x_n) τ -power dominated with $\tau < \rho$ and for any $d \in (0, 1)$,

$$\lim_{N \rightarrow \infty} \mathbb{P}(\{w : D_{\ell_1}((\phi_{m_N \times N}(w), \Delta_N^*); x^N) \leq d\}) = 1, \quad (17)$$

furthermore, we have that

$$\lim_{N \rightarrow \infty} D_{\ell_1}((\phi_{m_N \times N}(w), \Delta_N^*); x^N) = 0, \quad \mathbb{P} - \text{almost surely.} \quad (18)$$

Note that the universal scheme of Proposition 1 is zero measurement per signal dimension by construction.

Proof of Proposition 1: Let $d \in (0, 1)$ be an arbitrary number and (x_n) an arbitrary signal τ -power dominated. By hypothesis as $m_N \sim N^\rho \cdot \log N$, considering any $r > (1 - d/C_0)$, we have that $\exists N_0$ such that for all $N \geq N_0$, $m_N \geq \kappa_1(r, x^N) \cdot \log N$. Therefore from Lemma 1, $\forall N \geq N_0$

$$\begin{aligned} \mathbb{P}(\{w \in \Omega^{m_N \cdot N} : D_{\ell_1}((\phi_{m_N \times N}(w), \Delta_N^*); x^N) \leq d\}) & \\ \geq 1 - 2^{C_2 \cdot m_N}, & \end{aligned} \quad (19)$$

then, taking the limit on N , we obtain (17). Note that (19) is valid for all $d > 0$, which is equivalent to saying that $\lim_{N \rightarrow \infty} D_{\ell_1}((\phi_{m_N \times N}(w), \Delta_N^*); x^N) = 0$ in probability (or measure) [9]. We can get the stronger almost-sure convergence in (18) from the *Borel-Cantelli lemma* [9] and from the fact that by the hypothesis we have that $\sum_{N \geq 1} 2^{-C_3 \cdot m_N} < \infty$. \blacksquare

Remark 1: i) Note that the construction of Proposition 1 shows that the distortion-measurement point $(d, 0)$, for d arbitrary close to zero, is achievable for the class of signals

⁴Over the signal collection Σ_k .

τ -power dominated K_τ . **ii)** Furthermore, there is a universal scheme that achieves this result over the family $\cup_{\tau < p} K_\tau$. **iii)** Finally, $D_{\ell_1}((\phi_{m_N \times N}(w), \Delta_N^*); x^N)$ converges to zero with probability one with respect to \mathbb{P} , which is stronger than the convergence in measure stipulated in Definition 6.

The universal zero distortion and zero measurement construction of Proposition 1 is also valid for the collection of asymptotically sparse signals of Definition 5, as they are τ -power dominated for any $\tau \in (0, 1)$. However, we can get a refined result if we restrict the analysis exclusively to this family of asymptotically sparse signals.

PROPOSITION 2: Let $\{(\phi_{m_N \times N}(w), \Delta_N^*) : N > 0\}$ be the usual random CS scheme of Theorem 3, and let $K_\infty \subset \mathbb{R}^N$ denote the set of asymptotically sparse signals. If we assume that $(1/m_N)$ is $o(1/\log N)$ and (m_N) is $o(N)$, then for any $d > 0$ and for all $(x_n) \in K_\infty$,

$$\lim_{N \rightarrow \infty} \mathbb{P}(\{w : D_{\ell_1}((\phi_{m_N \times N}(w), \Delta_N^*); x^N) \leq d\}) = 1. \quad (20)$$

Thus this result shows sufficient conditions for a zero measurement universal CS scheme to achieve close to zero arbitrary distortion as N tends to infinity. It is important to note that the condition $(\frac{1}{m_N})$ is $o(\frac{1}{\log N})$ is weaker than its universal counterpart in Proposition 1, obtained for the family of τ -power dominated signals. However, we pay the price that the almost-sure convergence in (18) is not necessarily holding, where from (20) $D_{\ell_1}((\phi_{m_N \times N}(w), \Delta_N^*); x^N)$ converges to zero in probability.

Proof of Proposition 2: Let us fix an arbitrary $r \in (0, 1)$. As $\limsup_{N \rightarrow \infty} \kappa_1(r, x^N) < \infty$, then by hypothesis $\exists N_0$ such that $\forall N \geq N_0$, $m_N \geq \kappa_1(r, x^N) \log N$. The rest of the argument follows the same steps as those in the proof of Proposition 1. ■

Remark 2: To conclude this section, it is worth emphasizing that the compressibility in Definition 2, although it provides a nice interpretation of asymptotic energy compaction of the signal in the sense of best k -term approximation [1], does not grant any performance in the sense of NSR versus number of measurements per dimension for the operational context of CS analysis and synthesis studied in this work. As a result, stringent compressibility notions were needed to fill this gap, which motivated the inclusion of Definitions 3-5.

IV. THE $\ell_1(\mathbb{N})$ CASE

Here we study the implications of the type of results presented in the previous section, for the special case of sequences in $\ell_1(\mathbb{N})$. We begin by mentioning the following fact:

LEMMA 2: If (x_n) is $\ell_1(\mathbb{N})$ then it is asymptotically sparse (Definition 5).

The proof is presented in the Appendix.

As $\ell_1(\mathbb{N}) \subset K_\infty$, then, from Proposition 2, we have a universal CS scheme achieving zero distortion with zero measurement per dimension for this class of signals. Furthermore considering that $\|(x_n)\|_{\ell_1} < \infty$, Theorems 1 and 2 offer the

following result concerning ℓ_1 -absolute error in the asymptotic regime:

THEOREM 4: Let $\{(\phi_{m_N \times N}(w), \Delta_N^*) : N > 0\}$ be the standard CS random scheme. If $(1/m_N)$ is $o(1/\log N)$ and (m_N) is $o(N)$, then $\forall (x_n) \in \ell_1(\mathbb{N})$ and for any $d > 0$,

$$\lim_{N \rightarrow \infty} \mathbb{P}(\{w : \|x^N - \Delta_N^*(\phi_{m_N \times N}(w) \cdot x^N)\|_{\ell_1} \leq d\}) = 1. \quad (21)$$

Furthermore, if $(m_N) \sim (N^\rho \cdot \log N)$ for some $\rho \in (0, 1)$, then $\forall (x_n) \in \ell_1(\mathbb{N})$

$$\lim_{N \rightarrow \infty} \|x^N - \Delta_N^*(\phi_{m_N \times N}(w) \cdot x^N)\|_{\ell_1} = 0, \quad \mathbb{P}\text{-almost surely.} \quad (22)$$

The proof follows the same arguments as the proof of Propositions 1 and 2. For completeness it is presented here.

Proof: We show first the result in (21). Let us consider $r \in (0, 1)$. As $\limsup_{N \rightarrow \infty} \kappa_1(r, x^N) < \infty$ by Lemma 2, then by the hypothesis $\exists N_0(r)$ such that $\forall N \geq N_0(r)$, $m_N \geq \kappa_1(r, x^N) \log N$. Then from (19) and the fact that $\|x^N\|_{\ell_1} \leq \|(x_n)\|_{\ell_1} < \infty$, $\forall N \geq N_0(r)$

$$\mathbb{P}\left\{\|x^N - \Delta_N^*(\phi_{m_N \times N}(w) \cdot x^N)\|_{\ell_1} \leq C_0(1-r)\|(x_n)\|_{\ell_1}\right\} \geq 1 - 2^{C_2 \cdot m_N}, \quad (23)$$

At this point for any arbitrary $d > 0$, we can choose $r > 0$ sufficiently small such that $C_0(1-r)\|(x_n)\|_{\ell_1} < d$, and, consequently, we obtain the result in Eq. (21) from (23) since by hypothesis m_N goes to infinity.

Concerning the result in Eq.(22), note that this is obtained from (23), which is valid in this scenario as $(m_N) \sim (N^\rho \cdot \log N)$, and the Borel-Cantelli lemma, because $\sum_{N \geq 1} 2^{-C_3 \cdot m_N} < \infty$. ■

Finally, given that $\|(x_n)\|_{\ell_2} \leq \|(x_n)\|_{\ell_1} < \infty$, we can obtain a version of Theorem 4 considering ℓ_2 -norm as a fidelity measure.

V. SUMMARY AND FUTURE WORK

This work introduces new definitions of compressibility of infinite sequences. In this context, we elaborate a compressed sensing analysis-synthesis scenario, in which results are established that guarantee recovery performances in the sense of achieving a given number of measurements per signal dimension and an ℓ_1 noise to signal ratio (ℓ_1 -NSR) distortion measure. In particular, we explore the zero-measurement and zero distortion regime, and their interplay with the previously mentioned notions of compressibility in the asymptotic regime when the signal dimension goes to infinity.

There are a number of interesting directions to pursue in further research on this topic. Just to name a few key ones: **i)** the characterization of concrete sequences that satisfy the proposed notions of compressibility; **ii)** the extension of performance recovery for the important case of an ℓ_2 -based distortion measure; **iii)** the extension of these results for random sequences, where the focus would be on average performances instead of the universal (minimax) [8] performance criterion explored in this work.

ACKNOWLEDGMENT

This material is based on work partially supported by Program U-Apoya, University of Chile and grants of CONICYT-Chile, Fondecyt 1110145. We thank Sandra Beckman for proofreading this work.

VI. APPENDICES

A. Proof of Lemma 2

As (x_n) is $\ell_1(\mathbb{N})$, then $\forall \epsilon > 0$, $\exists N(\epsilon) \geq 1$ such that $\sum_{n \geq N(\epsilon)} |x_n| < \epsilon$. Let us define

$$k_1(r, N, (x_n)) \equiv \min \left\{ k \in \{1, \dots, N\} : \frac{\varsigma_1(k, x^N)}{\|(x_n)\|_{l_1}} \geq r \right\}, \quad (24)$$

where $k_1(r, N, (x_n)) \geq \kappa_p(r, x^N)$ as $\|(x_n)\|_{l_1} \geq \|x^N\|_{l_1}$. We have that for all $N > 0$, and for any $k \in \{1, \dots, N\}$,

$$\frac{|x_{N1}| + \dots + |x_{Nk}|}{\|(x_n)\|_{l_1}} \geq \frac{|x_1| + \dots + |x_k|}{\|(x_n)\|_{l_1}} = 1 - \frac{\sum_{n \geq k} |x_n|}{\|(x_n)\|_{l_1}}, \quad (25)$$

where $\{x_{N1}, \dots, x_{NN}\}$ is the order statistics of $\{x_1, \dots, x_N\}$. Hence, for all $\epsilon > 0$, if $N \geq k > N(\epsilon)$, $\frac{|x_{N1}| + \dots + |x_{Nk}|}{\|(x_n)\|_{l_1}} \geq 1 - \epsilon / \|(x_n)\|_{l_1}$. Therefore, if $\epsilon < (1 - r) \|(x_n)\|_{l_1}$, $\limsup_{N \rightarrow \infty} k_1(r, N, (x_n)) \leq N(\epsilon) < \infty$ for all $r \in (0, 1)$. ■

REFERENCES

- [1] A. Amini, M. Unser, and F. Marvasti, "Compressibility of deterministic and random infinity sequences," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5193–5201, November 2011.
- [2] 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.
- [3] R. G. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, 2008.
- [4] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203–4215, December 2005.
- [5] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principle: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [6] —, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. on Pure and Appl. Math.*, vol. 59, pp. 1207–1223, 2006.
- [7] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.
- [8] A. Cohen, W. Dahmen, and R. DeVore, "Compressed sensing and best k -term approximation," *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 211–231, July 2009.
- [9] L. Breiman, *Probability*. Addison-Wesley, 1968.