From compression to compressed sensing

Shirin Jalali
New York Unversity
Department of Electrical and Computer Engineering
Brooklyn, NY, 11201

Arian Maleki
Columbia University
Department of Statistics
New York, NY, 10027

***Abstract*— Can compression algorithms be employed for recovering signals from their underdetermined set of linear measurements? Addressing this question is the first step towards applying compression algorithms for compressed sensing (CS). In this paper, we consider a family of compression algorithms $\mathcal{C}_R$, parametrized by rate R, for a compact class of signals $\mathcal{Q} \subset \mathbb{R}^n$. The set of natural images and JPEG2000 at different rates are examples of $\mathcal{Q}$ and $\mathcal{C}_R$, respectively. We establish a connection between the rate-distortion performance of $\mathcal{C}_R$, and the number of linear measurement required for successful recovery in CS. We then propose compressible signal pursuit (CSP) algorithm and prove that, with high probability, it accurately and robustly recovers signals from an underdetermined set of linear measurements.**

I. Introduction

The field of compressed sensing (CS) was established on a keen observation that if a signal is sparse in a certain basis it can be recovered from far fewer random linear measurements than its ambient dimension [1], [2]. In the last decade, CS recovery algorithms have evolved to capture more complicated signal structures such as group sparsity, atomic structure, and nuclear norm minimization [3]–[19]. In this paper, we consider a different type of structure based on compression algorithms. Suppose that a class of signals can be "efficiently" compressed by a compression algorithm. Intuitively speaking, such classes of signals have a certain "structure" that enables the compression algorithm to represent them with fewer bits. These structures are often much more complicated than sparsity, and employing them in CS can potentially reduce the number of measurements required for signal recovery.

In this paper, we aim to address the following problem. Is it possible to employ compression schemes in the CS problem and design algorithms that recover signals either exactly or with "small error", from random linear measurements? As we will prove in this paper, the answer to this question is affirmative. We propose a CS recovery algorithm based on exhaustive search over the set of "compressible" signals, that, under a certain condition on the rate-distortion function, recovers signals from fewer measurements than their ambient dimension. This result provides the first theoretical basis for using compression algorithms in CS.

The organization of the paper is as follows. Section II reviews the main concepts used in this paper and formally states the problem addressed in the paper. Section III summarizes our main contributions. Finally, Section V concludes the paper.

Due to space limitations, the proofs of all lemmas, theorems, etc. are presented in the full version of this paper which is available on arxiv [20].

II. Background and problem definition

In this section we first review the concept of compression and rate-distortion function. Then we state the problem we address in this paper more formally.

A. Notation

Boldfaced letters such as $\mathbf{x}$ and $\mathbf{X}$ represent vectors. Calligraphic letters denote sets. Given a finite set $\mathcal{A}$, $|\mathcal{A}|$ denotes its size. The ℓ_p-norm of $\mathbf{x} \in \mathbb{R}^n$ is defined as $\|x\|_p \triangleq (\sum_{i=1}^n |x_i|^p)^{1/p}$. The ℓ_0-norm is also defined as $\|\mathbf{x}\|_0 \triangleq |\{i \;:\; x_i \neq 0\}|$. Note that for $p < 1$, $\|\cdot\|_p$ is a semi-norm since it does not satisfy the triangle inequality.

B. Rate-distortion function

Let $\mathcal{Q}$ denote a compact subset of $\mathbb{R}^n$. Consider a compression algorithm for $\mathcal{Q}$ described by encoder and decoder mappings $(\mathcal{E}, \mathcal{D})$. Encoder

$$\mathcal{E} : \mathcal{Q} \rightarrow \{1, 2, \ldots, 2^R\},$$

maps signal $\mathbf{x} \in \mathcal{Q}$ to codeword $\mathcal{E}(\mathbf{x})$. Decoder

$$\mathcal{D} : \{1, 2, \ldots, 2^R\} \rightarrow \hat{\mathcal{Q}},$$

maps the coded signal $\mathcal{E}(\mathbf{x})$ back to the reconstruction domain $\hat{\mathcal{Q}}$. Let $\hat{\mathbf{x}} \triangleq \mathcal{D}(\mathcal{E}(\mathbf{x}))$ denote the reconstruction of signal $\mathbf{x} \in \mathcal{Q}$. The performance of the described coding scheme at rate R is measured in terms of its induced distortion defined as

$$D(R) \triangleq \sup_{\mathbf{x} \in \mathcal{Q}} \|\mathbf{x} - \mathcal{D}(\mathcal{E}(\mathbf{x}))\|_2.$$

Throughout the paper, we usually consider a family of fixed-rate compression algorithms $\{(\mathcal{E}_R, \mathcal{D}_R) \;:\; R > 0\}$ parametrized by the rate R, and its corresponding rate-distortion function $R(D)$ defined as $R(D) \triangleq \inf\{R : D(R) \leq D\}$.

Given compression algorithm $(\mathcal{E}_R, \mathcal{D}_R)$, let $\mathcal{C}_R$ denote its *codebook* defined as

$$\mathcal{C}_R \triangleq \{\mathcal{D}_R(\mathcal{E}_R(\mathbf{x})) : \mathbf{x} \in \mathcal{Q}\}.$$

Note that $|\mathcal{C}_R| \leq 2^R$.

In the remaining of this section we illustrate the concepts with two examples. Let $\mathcal{B}_p^n(\rho) \triangleq \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_p \leq \rho\}$ represent a ball of radius ρ in $\mathbb{R}^n$. Also, let Γ_k^n denote the set of all k-sparse signals in $\mathbb{R}^n$, i.e.,

$$\Gamma_k^n \triangleq \{\mathbf{x} \in \mathbb{R}^n \ : \ \|\mathbf{x}\|_0 \leq k\}.$$

Example 1: There exists a family of compression algorithms that achieves the following rate-distortion function on $\mathcal{B}_2^n(\rho)$:

$$R(D) = \frac{1}{2} n \log n + n \log\left(\frac{\rho}{D}\right) + cn,$$

for $D \leq \sqrt{n}$. Here, c is a constant less than 3.

Example 2: There exists a family of compression algorithms that achieves the following rate-distortion function on $\mathcal{B}_2^n(\rho) \cap \Gamma_k^n$:

$$R(D) = \log\binom{n}{k} + k \log\left(\frac{\sqrt{k}\rho}{D}\right) + ck,$$

where c is a constant less than 3.

Note that since $\mathcal{B}_2^n(\rho) \cap \Gamma_k^n \subset \mathcal{B}_2^n(\rho)$, we expect to compress $\mathcal{B}_2^n(\rho) \cap \Gamma_k^n$ more efficiently. This is specially clear as $D \downarrow 0$.

C. Problem statement

Consider the problem of recovering "structured" signal $\mathbf{x} \in \mathbb{R}^n$ from its undersampled set of linear measurements $\mathbf{y} = A\mathbf{x}$, where $\mathbf{y} \in \mathbb{R}^d$, and $A \in \mathbb{R}^{d \times n}$, $d < n$, denotes the measurement matrix. For various types of structure such as sparsity, it is well-known that $\mathbf{x}$ may be recovered from measurements $\mathbf{y}$ even though $d < n$. In this paper we explore a more elaborate type of structure based on compressibility.

Instead of being structured as sparse, smooth, etc., suppose that the signal belongs to a compact set $\mathcal{Q} \subset \mathbb{R}^n$ and there exists a family of compression algorithms $\{(\mathcal{E}_R, \mathcal{D}_R) : R > 0\}$ with rate-distortion function $R(D)$ for signals in $\mathcal{Q}$. For instance, we can consider the JPEG2000 compression algorithm [21] at different rates for the class of images. This family of compression algorithms might be exploiting the sparsity of the signal in a certain domain or any other type of structure. The actual mechanism by which the algorithm is compressing the signals in $\mathcal{Q}$ is not important for the purpose of this paper. Instead, we are interested in recovering vector $\mathbf{x} \in \mathcal{Q}$ from an undersampled set of linear equations $\mathbf{y} = A\mathbf{x}$ by employing the compression algorithms $\{(\mathcal{E}_R, \mathcal{D}_R) : R > 0\}$. The question is when this is possible, and what it implies on the rate-distortion function $R(D)$. Since for every compact set we can define a family of compression algorithms, it seems that existence of compression algorithms does not necessarily lead to a CS-recovery method. The following lemma confirms this intuition.

Lemma 1: Let $\mathcal{Q} = \mathcal{B}_2^n(1)$. If the number of measurements is less than the ambient dimension n, any CS-recovery algorithm will result in an ℓ_2 reconstruction error of at least 1, for any measurement matrix.

Therefore, the first step in employing compression algorithms for CS is to characterize the class of compression algorithms that can potentially lead to CS-recovery methods.

Definition 1: Compressed sensing is said to be *applicable* to a compact set $\mathcal{Q} \subset \mathbb{R}^n$ with $d < n$ measurements, if, for any $\epsilon > 0$, there exists a $d \times n$ matrix A_ϵ and a recovery algorithm $\mathcal{A}_\epsilon$,

$$\mathcal{A}_\epsilon : \mathbb{R}^d \to \mathcal{Q},$$

such that $\|\mathcal{A}_\epsilon(A_\epsilon \mathbf{x}) - \mathbf{x}\|_2 \leq \epsilon$, for all $\mathbf{x} \in \mathcal{Q}$.

According to Lemma 1, CS is not applicable to $\mathcal{B}_2^n(1)$ with d measurements, for any $d < n$. Next we define α-dimension for a rate-distortion function and establish its connection with CS-applicability.

Definition 2: Consider compact set $\mathcal{Q} \subset \mathbb{R}^n$, and a family of fixed-rate compression codes, $\{(\mathcal{E}_R, \mathcal{D}_R) : R > 0\}$, with rate-distortion function $R(D)$. Define the high resolution rate distortion dimension, or α-dimension, of a family of codes as

$$\alpha \triangleq \limsup_{D \to 0} \frac{R(D)}{\log(\frac{1}{D})}. \tag{1}$$

In Section IV we discuss the connection between α-dimension and other well-known concepts in information theory and functional analysis.

Consider a compact set $\mathcal{Q} \subset \mathbb{R}^n$, and without loss of generality, assume that $\mathbf{0} \in \mathcal{Q}$. Since the set is compact

$$\rho(\mathcal{Q}) \triangleq \sup_{\mathbf{x} \in \mathcal{Q}} \|\mathbf{x}\|_2$$

is finite. Therefore, $\mathcal{Q} \subset \mathcal{B}_2^n(\rho)$, and according to Example 1, there exists a family of compression algorithms with

$$R(D) \leq n \log\left(\frac{\rho}{D}\right) + c, \tag{2}$$

where $\rho = \rho(\mathcal{Q})$, $D < \rho$, and c is a constant independent of the distortion level D. Therefore, for any compact set $\mathcal{Q}$, there exists a family of compression codes with α-dimension upper-bounded by n. The interesting regime is when there exists a family of codes with α-dimension strictly smaller than n. For instance, the set of k-sparse signals in $\mathcal{B}_2^n(1)$, discussed in Example 2, is a set for which there exists a family of compression algorithms with α-dimension smaller than n. In the remaining of this section, we explore the connection between the CS-applicability of a compact set $\mathcal{Q}$ and the α-dimension of a family of codes for $\mathcal{Q}$. The question is whether CS is applicable to $\mathcal{Q}$ with number of measurements $d < n$, and if the answer is affirmative, what is the minimum number of measurements for this result to hold.

Given $\alpha > 0$, let $\mathcal{S}_\alpha^n$ denote the set of all subsets of $\mathcal{B}_2^n(1)$ for which there exists a family of compression algorithms with α-dimension upper-bounded by α. For each $\mathcal{Q} \in \mathcal{S}_\alpha^n$, define $d_{\min}(\mathcal{Q})$ as the minimum number of measurements for which CS is applicable to $\mathcal{Q}$. The following theorem provides a lower bound for the number of measurements.

Theorem 1: If CS is applicable to any element of $\mathcal{S}_{\alpha,n}$ with d measurements, then $d \geq \lfloor \alpha \rfloor$. In other words,

$$\sup_{\mathcal{Q} \in \mathcal{S}_\alpha^n} d_{\min}(\mathcal{Q}) \geq \lfloor \alpha \rfloor.$$

Note that the notion of CS-applicability with d measurements is the minimal requirement for the practical use of CS. In particular, it does not guarantee the robustness of the algorithm to measurement noise. Also, the measurement matrix can be adapted to the structure of data and the recovery algorithm can exploit any extra information about the set. In the rest of this paper, we show that considering random measurement matrices (nonadaptive measurements) and following Occam's principle results in an accurate and stable recovery algorithm.

Our algorithm searches over the space of "compressible signals" and finds the one that matches the measurements the best. More formally, given a compression algorithm $(\mathcal{E}_R, \mathcal{D}_R)$ with codebook $\mathcal{C}_R$, consider *compressible signal pursuit* (CSP) algorithm for recovering $\mathbf{x}_o \in \mathcal{Q}$ from its measurements $\mathbf{y}_o = A\mathbf{x}_o$ defined as

$$\hat{\mathbf{x}}_o = \arg\min_{\mathbf{c}\in\mathcal{C}_R} \|\mathbf{y}_o - A\mathbf{c}\|_2^2. \tag{3}$$

Here the rate R can be considered as a free parameter, which, as we will see in Section III, plays a role in the tradeoff between the success probability of CSP, and its reconstruction error. Note that we still ignore one important aspect of practical algorithms and that is "computational complexity". CSP is based on an exhaustive search and hence is computationally very demanding. Practical implementation of such ideas is left for future research.

III. Main contributions

Consider the problem of recovering signal $\mathbf{x}_o \in \mathcal{Q} \subset \mathbb{R}^n$, from $d < n$ linear measurements $\mathbf{y}_o = A\mathbf{x}_o + \mathbf{z}$, where the entries of A are i.i.d. $\mathcal{N}(0,1)$, and $\mathbf{z} \in \mathbb{R}^d$ represents the measurements noise in the system. Furthermore, assume that there exists a family of compression algorithms, $\{(\mathcal{E}_R, \mathcal{D}_R) : R > 0\}$, for the signals of $\mathcal{Q}$, which has rate-distortion function $R(D)$. We employ the CSP algorithm described in (3) to recover $\mathbf{x}_o$ from $\mathbf{y}_o$.

A. Noiseless measurements

Our first result is on the performance of the CSP algorithm, when there is no noise in the system, i.e., $\mathbf{z} = \mathbf{0}$.

Theorem 2: Consider compression code $(\mathcal{E}, \mathcal{D})$ operating at rate R and distortion D for coding signals in set $\mathcal{Q}$. Let $A \in \mathbb{R}^{d\times n}$, where $A_{i,j}$ are i.i.d. $\mathcal{N}(0,1)$. For $\mathbf{x}_o \in \mathcal{Q}$, let $\hat{\mathbf{x}}_o$ denote the reconstruction of $\mathbf{x}_o$ from $\mathbf{y}_o = A\mathbf{x}_o$, $A \in \mathbb{R}^{d\times n}$, by the CSP algorithm employing $(\mathcal{E}, \mathcal{D})$. Then,

$$\|\hat{\mathbf{x}}_o - \mathbf{x}_o\|_2 \leq D\sqrt{\frac{1+\tau_1}{1-\tau_2}},$$

with probability at least

$$1 - 2^R \mathrm{e}^{\frac{d}{2}(\tau_2 + \log(1-\tau_2))} - \mathrm{e}^{-\frac{d}{2}(\tau_1 - \log(1+\tau_1))}.$$

Theorem 2 proves that in many cases CSP algorithm provides an "accurate estimate" of $\mathbf{x}_o$ with a number of measurements that is less than the ambient dimension n. Corollary 1 below describes one instance of such cases.

Corollary 1: Let $D < \mathrm{e}^{-1}$ and $d \geq \frac{4R(D)}{\log(1/\mathrm{e}D)}$. Then

$$\mathrm{P}(\|\hat{\mathbf{x}}_o - \mathbf{x}_o\|_2 \geq \sqrt{2D}) \leq \mathrm{e}^{-R(D)} + \mathrm{e}^{-\frac{R(D)}{8\log(1/\mathrm{e}D)}}.$$

Note that if $D \ll 1$ then $\frac{R(D)}{\log(1/eD)}$ is much smaller than $R(D)$ itself. In fact, according to (2) as $D \to 0$, $\limsup_{D\to 0} \frac{R(D)}{\log(1/eD)} \leq n$. The following corollary characterizes the number of measurements required and probability of correct recovery as a function of α-dimension.

Corollary 2: Consider a family of rate-distortion codes $\{(\mathcal{E}_R, \mathcal{D}_R) : R > 0\}$ with α-dimension α. Then for every $\epsilon > 0$, there exists $R > 0$ and a corresponding code $(\mathcal{E}_R, \mathcal{D}_R)$, such that if we employ this code in the CSP algorithm with $d > 4\alpha$ measurements, then

$$\mathrm{P}(\|\hat{\mathbf{x}}_o - \mathbf{x}_o\|_2 \geq \epsilon) \leq \mathrm{e}^{-0.1\alpha}.$$

Corollary 2 directly follows from Corollary 1, by taking the limit as $D \to 0$. Example 2 shows an application of this corollary.

Example 3: Consider the class of k-sparse signals in $\mathbb{R}^n$ discussed in Example 2. It is straightforward to check that there exists a family of codes such that $R(D)/\log(1/D) \downarrow k$ as $D \downarrow 0$. Therefore, from Corollary 2, for every $\epsilon > 0$, the solution of the CSP algorithm satisfies

$$\mathrm{P}(\|\hat{\mathbf{x}}_o - \mathbf{x}_o\|_2 \geq \epsilon) \leq \mathrm{e}^{-0.1k},$$

if $d > 4k$.

By choosing different values for τ_1 and τ_2 we can derive different upper bounds for the recovery error. Here is another instance of such result.

Corollary 3: Taking the number of measurements d such that $d \geq 4R(D)/\log n$, then

$$\mathrm{P}\left(\frac{1}{\sqrt{n}}\|\mathbf{x}_o - \hat{\mathbf{x}}_o\|_2 \leq \sqrt{2}D\right) \leq \mathrm{e}^{-R(D)} + \mathrm{e}^{-\frac{R(D)}{8\log n}}.$$

As a final remark, we derive a lower bound for the number of measurements required according to Theorem 1 for the "success" of the CSP algorithm. Consider the success probability in Theorem 1. To keep success probability larger than zero we require

$$R\log(2) + \frac{d}{2}(\tau_1 - \log(1-\tau_1)) < 0 \tag{4}$$

Therefore, to reduce the number of measurement, we require τ_1 to be large. But τ_1 is always less than 1. Furthermore, if $\tau_1 > 1 - D^2$, the upper bound on the reconstruction error will be larger than 1, which is a trivial bound for any signal in $B_2^n(1)$. Hence, we consider $\tau_1 < 1 - D^2$. If we set $\tau_1 = 1 - D^2$ in (4), we obtain $d \geq \frac{2\log(2)R(D)}{2\log(D)+1-D^2}$.

B. Noisy measurements

In this section we analyze the performance of CSP in the presence of noise. Consider the case where the linear measurements are corrupted by i.i.d. noise, i.e., $\mathbf{y}_o = A\mathbf{x}_o + \mathbf{z}$, where $z_i \sim \mathcal{N}(0, \sigma^2)$, $i = 1, \ldots, d$. To recover signal $\mathbf{x}_o$ from $\mathbf{y}_o$, again we employ the CSP algorithm described in (3).

Theorem 3: Let $\hat{\mathbf{x}}_o$ denote the solution of CSP to input $\mathbf{y}_o$, using fixed-rate code $(\mathcal{E}, \mathcal{D})$ for compact set $\mathcal{Q}$, which operates at rate R and distortion $D \leq (5\mathrm{e})^{-1}$. For any $\mathbf{x}_o \in \mathcal{Q}$ and $\eta > 1$, choosing

$$d = \left\lceil \frac{4\eta R}{\log\frac{1}{\mathrm{e}D}} \right\rceil,$$

then

$$\|\hat{\mathbf{x}}_o - \mathbf{x}_o\|_2 \leq \frac{\sigma}{D\sqrt{\eta}} \log \frac{1}{\mathrm{e}D} + \sqrt{2D + \frac{2\sigma}{\sqrt{\eta}} + \frac{\sigma^2}{\eta D^2} \log^2 \left(\frac{1}{\mathrm{e}D}\right)},$$

with probability exceeding

$$1 - \mathrm{e}^{-\frac{R}{\log 1/(\mathrm{e}D)}} - 2\mathrm{e}^{-\frac{0.6\eta R}{\log 1/(\mathrm{e}D)}} - \mathrm{e}^{-(2\eta-1)R} - \mathrm{e}^{-(0.8\eta - \log 2)R} - \mathrm{e}^{-0.3R}. \quad (5)$$

Note that D (or equivalently R) acts as the free parameter of the CSP algorithm and control the bias and variance of the final estimate. Intuitively speaking small values of D lead to large variance since σ is divided by D in two terms. Large values of D make the variance small, but increase the bias (due to 2D term under the radical sign). The optimal choice of the free parameter R is dictated by the rate-distortion performance of the code, number of measurements, and the variance of the noise.

C. Discussion

The results we have discussed so far are the same for finite and infinite dimensional classes [20]. Such results may mislead us to a conclusion that as long as the performance of CSP is concerned the ambient dimension is not important. However, the finiteness of ambient dimension may help derive stronger results. Next theorem is an instance of such results.

Theorem 4: Let $A \in \mathbb{R}^{d\times n}$ be a measurement matrix with $A_{ij} \overset{iid}{\sim} N(0,1)$. For any $\mathbf{x}_o \in \mathbb{Q}$, we denote the reconstruction of the CSP algorithm at rate R with $\hat{\mathbf{x}}_o$. We have

$$\mathrm{P}\left(\forall\, \mathbf{x}_o \in \mathcal{Q}:\ \|\mathbf{x}_o - \hat{\mathbf{x}}_o\|_2 \geq \frac{2(\sqrt{nd^{-1}} + (1+t))}{1-\tau} D\right) \leq \mathrm{e}^{-dt^2/2} + 2^{2R}\mathrm{e}^{\frac{d}{2}(\tau + \log(1-\tau))}.$$

Note that there is a major difference between Theorem 4 and Theorem 2. Theorem 4 claims that once we draw a random matrix from Gaussian distribution, this matrix with high probability works for any signal in $\mathcal{Q}$. However, Theorem 2 considers individual sequences. Note that the strength of Theorem 4 has come at a price of larger reconstruction error and lower success probability.

IV. Related work

A. Connection of compression and compressed sensing

In this paper we consider the problem of using a family of compression algorithms for compressed sensing. The other direction, i.e., using CS for compression have also been extensively studied in the literature [22]–[30]. In this line of work the rate-distortion that is achieved by scaler (or in a few cases adaptive) quantization of random linear measurements has been derived. However, such results are different from our work since they only consider either sparse or approximately sparse signals. Furthermore, we consider a different direction, that is, the direction of deriving CS recovery algorithms based on compression schemes.

B. Kolmogorov's ϵ-entropy and compressed sensing

The ϵ-entropy of a compact set $\mathcal{Q}$ is defined as

$$H_\epsilon(\mathcal{Q}) = \log_2 N_\epsilon(\mathcal{Q}),$$

where $N_\epsilon(\mathcal{Q})$ is the minimum number of elements in an ϵ-covering of $\mathcal{Q}$ [31]. The ϵ-entropy, H_ϵ, provides a lower bound on the rate distortion of any family of compression algorithms. In other words, if $R(D)$ is the rate distortion function of a family of compression algorithms on $\mathcal{Q}$, then

$$R(D) \geq H_D(\mathcal{Q}).$$

It is clear that our results can be stated in terms of Kolmogorov's ϵ-entropy by considering it as the optimal compression scheme from the perspective of rate-distortion tradeoff.

C. Stochastic settings

While in this paper we only focus on deterministic signal models, stochastic settings have also been extensively studied in CS. (See for instance [32]–[41].) In such models the data is assumed to follow a certain distribution (often i.i.d.) and the probability of correct recovery is measured as the ambient dimension tends to infinity. In many cases the algorithms exhibit certain phase transitions in the probability of correct recovery.

In [38], [40], the authors characterize the asymptotic performance of "information-theoretically" optimal algorithms in the stochastic setting. There are several major differences between our work and the work of [38], [40]. First our framework is concerned with the deterministic signal models. Second, our results are for finite-dimensional signals, and hence are non-asymptotic. Third, we consider arbitrary family of compression algorithms and characterize when such schemes are suitable for signal recovery from random linear measurements.

D. Kolmogorov complexity

Our work is mainly inspired by series of work on the connection between Kolmogorov complexity of sequences and CS [42]–[47]. In particular, [42] defines the *Kolmogorov information dimension* of $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in [0,1]^n$ at resolution m as

$$\kappa_{m,n}(\mathbf{x}) \triangleq \frac{K^{[\cdot]_m}(x_1, x_2, \ldots, x_n)}{m},$$

where intuitively speaking, $K^{[\cdot]_m}(x_1, x_2, \ldots, x_n)$ denotes the Kolmogorov complexity of vector $\mathbf{x}$ where each component is quantized by m bits, and proves that if the Kolmogorov information dimension of a sequence is small compared to its ambient dimension one can recover it from an undersampled set of linear measurements. Our results have several connections with [42] and is the first step towards practical implementation of [42]. While the CSP algorithm is computationally demanding at this point, it provides an approach to designing suboptimal algorithms such as greedy methods. Furthermore, CSP algorithm may enable us to employ universal compression algorithms to develop universal compressed sensing methods.

V. CONCLUSION

In this paper, we studied the problem of employing a family of compression algorithms for compressed sensing, i.e., recovering structured signals from their undersampled set of random linear measurements. Addressing this problem enables CS schemes to exploit complicated structures integrated in compression algorithms. We proposed compressible signal pursuit (CSP) algorithm that outputs the codeword that best matches the measurements. We proved that employing a family of compression algorithms whose rate-distortion function satisfies $\limsup_{D\to 0} R(D)/\log(1/D) \leq \alpha$, with α smaller than the ambient dimension, with high probability, CSP recovers signals from 4α measurements. The CSP algorithm is still computationally demanding and requires approximation or simplification for practical applications. This important direction is left for future research.

REFERENCES

[1] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, Apr. 2006.

[2] E. Candès, J. Romberg, , and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, Feb. 2006.

[3] S. Bakin. Adaptive regression and model selection in data mining problems. *Ph.D. Thesis, Australian National University*, 1999.

[4] Y. C. Eldar, P. Kuppinger, and H. Bolcskei. Block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Trans. Signal Processing*, 58(6):3042–3054, Jun. 2010.

[5] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *J. Roy. Statist. Soc. Ser. B*, 68(1):49–67, 2006.

[6] S. Ji, D. Dunson, and L. Carin. Multi-task compressive sensing. *IEEE Trans. Signal Processing*, 57(1):92–106, 2009.

[7] A. Maleki, L. Anitori, Z. Yang, and R. G. Baraniuk. Asymptotic analysis of complex lasso via complex approximate message passing (CAMP). *arXiv preprin arXiv:1108.0477v1*, 2011.

[8] M. Stojnic. Block-length dependent thresholds in block-sparse compressed sensing. *Preprint arXiv:0907.3679*, 2009.

[9] M. Stojnic, F. Parvaresh, and B. Hassibi. On the reconstruction of block-sparse signals with an optimal number of measurements. *IEEE Trans. Signal Processing*, 57(8):3075–3085, 2009.

[10] M. Stojnic. ℓ_2/ℓ_1-optimization in block-sparse compressed sensing and its strong thresholds. *IEEE J. Select. Top. Signal Processing*, 4(2):350–357, 2010.

[11] L. Meier, S. Van De Geer, and P. Buhlmann. The group Lasso for logistic regression. *J. Roy. Statist. Soc. Ser. B*, 70(1):53–71, 2008.

[12] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. Willsky. The convex geometry of linear inverse problems. *Preprint*, 2010.

[13] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Trans. Inform. Theory*, 56(4):1982 –2001, Apr. 2010.

[14] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum rank solutions to linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, Apr. 2010.

[15] M. Vetterli, P. Marziliano, and T. Blu. Sampling signals with finite rate of innovation. *IEEE Trans. Signal Processing*, 50(6):1417–1428, Jun. 2002.

[16] S. Som and P. Schniter. Compressive imaging using approximate message passing and a markov-tree prior. *IEEE Trans. Signal Processing*, 60(7):3439–3448, 2012.

[17] E.J. Candés, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *arXiv preprint arXiv:0912.3599*, 2009.

[18] V. Chandrasekaran, S. Sanghavi, P.A. Parrilo, and A.S. Willsky. Rank-sparsity incoherence for matrix decomposition. *SIAM J. Optimization*, 21(2):572–596, 2011.

[19] M. F. Duarte, W. U. Bajwa, and R. Calderbank. The performance of group Lasso for linear regression of grouped variables. Technical Report TR-2010-10, Duke University, Dept. Comp. Science, Durham, NC, 2011.

[20] S. Jalali and A. Maleki. From compression to compressed sensing. *arXiv:1212.4210*, 2012.

[21] D. S. Taubman and M. W. Marcellin. *JPEG2000: Image Compression Fundamentals, Standards and Practice*. Kluwer Academic Publishers, 2002.

[22] S. Sarvotham, D. Baron, and R.G. Baraniuk. Measurements vs. bits: Compressed sensing meets information theory. In *Proc. Allerton Conf. Comm., Control, and Comp.*, 2006.

[23] V.K. Goyal, A.K. Fletcher, and S. Rangan. Compressive sampling and lossy compression. *Signal Processing Magazine*, 25(2):48–56, 2008.

[24] P. Boufounos and R. Baraniuk. Quantization of sparse representations. Technical report, Rice University Technical Report, 2007.

[25] E. Candés and J. Romberg. Encoding the ℓ_p ball from limited measurements. In *Proc. Data Compression Conference (DCC)*, pages 33–42. IEEE, 2006.

[26] W. Dai, H.V. Pham, and O. Milenkovic. Distortion-rate functions for quantized compressive sensing. In *Proc. Information Theory Workshop (ITW)*, pages 171–175, 2009.

[27] J.Z. Sun and V.K. Goyal. Optimal quantization of random measurements in compressed sensing. In *Proc. IEEE Int. Symp. Inform. Theory*, pages 6–10, 2009.

[28] C. Deng, W. Lin, B. Lee, and C.T. Lau. Robust image compression based on compressive sensing. In *Proc. Int. Conference Multimedia and Expo (ICME)*, pages 462–467, 2010.

[29] A. Schulz, L. Velho, and E.A.B. Da Silva. On the empirical rate-distortion performance of compressive sensing. In *Proc. Int. Conference Image Processing (ICIP)*, pages 3049–3052, 2009.

[30] P.T. Boufounos. Universal rate-efficient scalar quantization. *IEEE Trans. Info. Theory*, 58(3):1861–1872, 2012.

[31] A.N. Kolmogorov and V.M. Tikhomirov. ε-entropy and ε-capacity of sets in function spaces. *Uspekhi Matematicheskikh Nauk*, 14(2):3–86, 1959.

[32] J. Shihao, X. Ya, and L. Carin. Bayesian compressive sensing. *IEEE Trans. Signal Processing*, 56(6):2346 –2356, Jun. 2008.

[33] P. Schniter. Turbo reconstruction of structured sparse signals. In *Proc. IEEE Conf. Inform. Science and Systems (CISS)*, Mar. 2010.

[34] D. Guo, D. Baron, and S. Shamai. A single-letter characterization of optimal noisy compressed sensing. In *47th Annual Allerton Conf. on Comm. Control, and Comp.*, pages 52–59, 2009.

[35] S. Rangan, A. K. Fletcher, and V. K. Goyal. Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing. *arXiv preprint arXiv:0906.3234v3*, 2009.

[36] D. L. Donoho, A. Maleki, and A. Montanari. Noise sensitivity phase transition. *IEEE Trans. Inform. Theory*, 57(10), Oct. 2011.

[37] D. L. Donoho and J. Tanner. Precise undersampling theorems. *Proc. IEEE*, 98(6):913–924, Jun. 2010.

[38] Y. Wu and S. Verdú. Rényi information dimension: Fundamental limits of almost lossless analog compression. *IEEE Trans. Inform. Theory*, 56(8):3721–3748, Aug. 2010.

[39] A. Maleki and D. L. Donoho. Optimally tuned iterative thresholding algorithm for compressed sensing. *IEEE J. Select. Top. Signal Processing*, Apr. 2010.

[40] Y. Wu and S. Verdu. Optimal phase transitions in compressed sensing. *IEEE Trans. Info. Theory*, 58(10):6241 –6263, oct. 2012.

[41] D.L. Donoho, A. Javanmard, and A. Montanari. Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. In *Proc. Int. Symposium Info. Theory (ISIT)*, pages 1231–1235, 2012.

[42] S. Jalali, A. Maleki, and R. Baraniuk. Minimum complexity pursuit for universal compressed sensing. *arXiv preprint arXiv:1208.5814*, 2012.

[43] D. L. Donoho, H. Kakavand, and J. Mammen. The simplest solution to an underdetermined system of linear equations. In *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, pages 1924–1928, Jul. 2006.

[44] D. Baron and M. F. Duarte. Signal recovery in compressed sensing via universal priors. *Arxiv preprint arXiv:1204.2611*, 2012.

[45] D. Baron and M. F. Duarte. Universal map estimation in compressed sensing. In *49th Annual Allerton Conference on Comm. Control, and Comp.*, pages 768–775, 2011.

[46] S. Jalali and A. Maleki. Minimum complexity pursuit. In *Proc. Allerton Conf. Communication, Control, and Computing*, pages 1764–1770, 2011.

[47] S. Jalali, A. Maleki, and R. Baraniuk. Minimum complexity pursuit: Stability analysis. In *Proc. Int. Symposium Info. Theory (ISIT)*, pages 1857–1861, 2012.