

# Summary Based Structures with Improved Sublinear Recovery for Compressed Sensing

M. Amin Khajehnejad<sup>†</sup>, Juhwan Yoo<sup>†</sup>, Animashree Anandkumar<sup>\*</sup> and Babak Hassibi<sup>†</sup>

<sup>†</sup>California Institute of Technology, Pasadena CA 91125

<sup>\*</sup>University of California Irvine CA 92697 <sup>\*</sup>

March 8, 2011

## Abstract

We introduce a new class of measurement matrices for compressed sensing, using low order summaries over binary sequences of a given length. We prove recovery guarantees for three reconstruction algorithms using the proposed measurements, including  $\ell_1$  minimization and two combinatorial methods. In particular, one of the algorithms recovers  $k$ -sparse vectors of length  $N$  in sublinear time  $\text{poly}(k \log N)$ , and requires at most  $\Omega(k \log N \log \log N)$  measurements. The empirical oversampling constant of the algorithm is significantly better than existing sublinear recovery algorithms such as Chaining Pursuit and Sudocodes. In particular, for  $10^3 \leq N \leq 10^8$  and  $k = 100$ , the oversampling factor is between 3 to 8. We provide preliminary insight into how the proposed constructions, and the fast recovery scheme can be used in a number of practical applications such as market basket analysis, and real time compressed sensing implementation.

## 1 Introduction

Despite significant advances in the field of Compressed Sensing (CS), certain aspects of CS remain relatively immature. Thus far, CS has been viewed primarily as a data acquisition technique [1]. As a result, the applicability of CS to other computational applications has not enjoyed commensurate investigation. In addition, to the best of the authors' knowledge, there is no unified CS system that has been implemented for practical real-time applications. A few recent works have addressed the former by applying sparse reconstruction ideas to certain inference problems including learning and adaptive computational schemes ( e.g. [2, 3, 4]). Several other works have addressed the latter by designing hardware, which exploits the fact that CS enables the monitoring of a given bandwidth at a much lower sampling rate than traditional Nyquist-based methods (see e.g., [6]). The motivating factor behind these works is that

---

<sup>\*</sup>This work was supported in part by the National Science Foundation under grants CCF-0729203, CNS-0932428 and CCF-1018927, by the Office of Naval Research under the MURI grant N00014-08-1-0747, and by Caltech's Lee Center for Advanced Networking.

for a given maximum sampling rate (limited by the poor power consumption scaling with sampling rate) achievable by digitizing hardware, it is possible to either acquire signals over a much greater bandwidth, or with much less power for a given bandwidth. Recent work, inspired by this line of thought, has led to the development of hardware CS encoders (see e.g. [7, 8, 9, 10]). However, none of the previous works address the problem of real-time signal decoding, which is a critical requirement in many applications.

Although variant by the nature of the problem and physical constraints, perhaps two fundamental issues in the practical implementations of CS are the following: 1) construction of measurement matrices that are provably good, certifiable and inexpensive to implement (either as real time sketches or as pre-built constructions), 2) Time efficient and robust recovery algorithms. Our aim is to introduce and provide an analysis of a sparse reconstruction system that addresses the aforementioned problems and allude to the extensions of CS in the less explored directions.

We introduce a new class of measurement matrices for sparse recovery that are deterministic, structured and highly scalable. The constructions are based on labeling the ambient state space with binary sequences of length  $n = \log_2 N$ , and summing up entries of  $\mathbf{x}$  that share the same pattern (up to a fixed length) at various locations in their labeling sequences. The class of corresponding matrices are RIP-less matrices that are congruent with the Basis Pursuit algorithms, which are standard techniques for sparse reconstruction [11]. In addition, we provide two efficient combinatorial algorithms along with theoretical guarantees for the proposed measurement structures. The proposed algorithms are sub-linear in the ambient dimension of the signal. In particular, we propose a summarized support index inference (SSII) algorithm with a running time of  $\mathcal{O}(\text{poly}(k \log N))$  that requires  $\mathcal{O}(k \log N \log \log N)$  measurements to recover  $k$ -sparse vectors, and has a empirical required over-sampling factor significantly better than existing sublinear methods. Due to the particular structure of the measurements and decoding algorithms, we believe that the proposed compression/decompression framework is amenable to real time CS implementation, and offers significant simplification in the design of an existing CS encoder/decoder. Furthermore, observations collected based on the proposed constructions appear as low order statistics or “summaries” in a number of practical situations in which a similar intrinsic labeling of the state space exists. This includes certain inference and discrete optimization problems such as market basket (commodity bundle) analysis, advertising, online recommendation systems, genomic feature selection, social networks, etc.

It should be acknowledged that there are various results on sublinear sparse recovery in the literature, including [12, 13, 14, 15, 16]. Unlike most previous works, the constructions of this paper offer sublinear storage requirement and are compatible with the practical scenarios that we consider. The recovery time of the algorithm is sublinear in the signal dimension, and the empirical recovery bounds are significantly better than the existing sublinear algorithms, such as Chaining Pursuit and Sudocodes, especially for small and moderate sparsity levels and very large signal dimensions.

## 2 Proposed Measurement Structures

We define a class of structured binary measurement matrices, based on the following definition

**Definition 1.** Let  $m, n$  and  $d$  be integers. A  $(n, d)$  summary is a pair  $X = (S, \mathbf{c})$ , where  $S$  is a subset of  $\{1, 2, \dots, n\}$  of size  $d$ , and  $\mathbf{c}$  is a binary sequence of length  $d$ . A  $(m, n, d)$  summary codebook is a collection  $\mathcal{C} = \{(S_i, \mathbf{c}_j) \mid 1 \leq i \leq m, 0 \leq j \leq 2^d - 1\}$  of  $(n, d)$  summaries, where  $S_i$ 's are distinct subsets, and  $\mathbf{c}_j$  is the length  $d$  binary representation of the integer  $j$ . If  $m = \binom{n}{d}$ ,  $\mathcal{C}$  is called the complete  $(n, d)$  summary codebook.

To a given  $(m, n, d)$  summary codebook  $\mathcal{C}$ , we associate a binary matrix  $A$  of size  $M \times N$  where  $M = 2^d \times m$ , and  $N = 2^n$ , in the following way. For every  $(S, \mathbf{c}) \in \mathcal{C}$ , there is a row  $\mathbf{a} = (a_1, \dots, a_N)$  in  $A$  that satisfies:

$$a_j = \mathbf{1} \{ \mathbf{b}_j(S) = \mathbf{c} \} \quad 1 \leq j \leq N \quad (1)$$

where  $\mathbf{b}_j$  is the  $n$ -bit binary representation of  $j$ , and  $\mathbf{b}_j(S)$  is the subsequence of the binary sequence  $\mathbf{b}_j$ , indexed by the entries of the set  $S^1$ . In other words,  $\mathbf{a}$  has a 1 in those columns  $\ell$  whose binary labeling conform to  $(S, \mathbf{c})$ . Every column of  $A$  has exactly  $m$  ones, and each row has exactly  $2^{n-d}$  ones. To clarify this definition, we consider the following example illustrated in Figure 1, in which  $n = 4$  and  $d = 2$ . Suppose that a summary  $(S, \mathbf{c})$  is given with  $S = \{1, 2\}$  and  $\mathbf{c} = 10$ . All possible binary sequences of length 4 that match  $(S, \mathbf{c})$  are listed in Figure 1. To find the corresponding indices of the listed labels, we should convert them to decimal values and increase by 1, which gives 9, 10, 11 and 12. The row  $\mathbf{a}$  of a measurement matrix that includes this summary is a vector of length  $2^4$  that has a 1 in those indices, as displayed.

$$\begin{array}{c} \begin{array}{cccc} \overline{1} & \overline{0} & x & x \\ \hline 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{array} \end{array} \quad \begin{array}{l} S = \{1, 2\}, \mathbf{c} = 10 \\ \mathbf{a} = (00000000111100000) \end{array}$$

Figure 1: An example  $(4, 2)$  summary and the corresponding row of the structured measurement matrix.

The defined matrices are very well motivated by some practical problems. In general, in a situation where the given signal space retains an intrinsic structured labeling similar to the one described, such constructions prove very useful. In particular, we consider the following two motivational examples.

**Resource Optimization.** Assume that a set  $\mathcal{F} = \{F_1, F_2, \dots, F_n\}$  of features (or parameters) is available, and assume that certain accumulations or collections of features form "lucrative" profiles (structures). In particular, a lucrative profile can be a subset of features which is representable by a binary sequence  $\mathbf{b} = b_1 b_2 \dots b_n$ , where  $b_i$  determines the presence of the  $i$ 'th feature. A practical assumption is that lucrative profiles are limited and weighted, meaning that their profitabilities are variable. The vector  $\mathbf{x} = (p_1, p_2, \dots, p_{2^n})^T$  formed by the respective profits of all feature collections is thus an approximately sparse vector. Furthermore, the available information about the profitability of profiles is often derived from a pool of observations or real world implementations, and are mostly given in the form of sum-

<sup>1</sup>Note that these structured matrices can be defined for any finite alphabets other than the binary field.

maries. More formally, what can be learned is the average profitability of a certain configuration of only  $d$  features. For example, it can be assessed that when  $F_1$  and  $F_2$  are present and  $F_3$  is absent, regardless of all other features, the average profit is some  $\bar{p}$ . The collection of summaries form an observation vector  $\mathbf{y}$ , that is related to  $\mathbf{x}$  through a set of linear equations  $\mathbf{y} = A\mathbf{x}$ , where  $A$  has a form similar to those obtained by summary codebooks. This setting arises in many practical applications such as market basket (commodity bundle) analysis, where the objective is to configure the structure of a market that complies the best with the needs and the behaviors of the customers. To that end, it is essential to understand which market configurations are winning and what packages of features (e.g. commodities, pricing options, interest rates, etc.) should be offered to customers, and with what percentages. Furthermore, the customers' behavioral information is often given in terms of high level summaries, e.g. in the lines of the statement "people who buy A and B, are likely to buy C".

**Compressed Sensing Hardware.** There are a few factors that severely limit the scalability of the existing CS hardware designs to larger problem dimensions. One of these factors is the generation of the measurement matrix  $A$ . In the simplest existing design,  $A$  is typically a pseudo-random matrix generated with a linear feedback shift register (LFSR) [8, 10]. The timing synchronization of a large number of measurements as well as the planar nature of physical implementations is very limiting. Using a more structured matrix may allow considerable simplification and reduction of the required hardware easing some of the previously mentioned limitations. The measurement structure defined in this work is potentially highly amenable to the implementation of practical CS hardware, due to the following two reasons. 1) There exist simple sublinear recovery algorithms for the proposed matrices, other than the linear programming method. This will be elaborated in the proceeding sections. 2) Due to the highly structured design, the integration matrix  $A$  can be implemented using one single LFSR seed, and a number of asynchronous digital circuits. Due to the lack of space and the irrelevance of the context, we avoid a detailed description of the latter, and postpone this to a future work.

### 3 Proposed Recovery Algorithms

For the measurement matrices described in the previous section we propose three reconstruction algorithms and provide success guarantees. These algorithms include the Basis Pursuit algorithm (a.k.a  $\ell_1$  minimization), as well as two fast algorithms that can recover sparse vectors from a sublinear number of measurements and in a sublinear amount of time. The detailed specifications will be given in the sequel. For the sake of the theoretical arguments that appear in the remainder of this section, we need to define the following notions:

**Definition 2.** Let  $n$  and  $l$  be integers with  $l < n$ . We define  $f_S(n, l)$ ,  $f_W(n, l, p, \epsilon)$  and  $f'_W(n, l, p, \epsilon)$  to be the largest integer  $k$  such that when  $k$  binary sequences of length  $n$  are selected at random, the following happens respectively:

1. With probability 1, there exists a  $(n, d)$  summary that appears in exactly one of the sequences.

2. With probability at least  $p$ , for each of the binary sequences, at least a fraction  $\epsilon$  of its  $(n, d)$  summaries are unique.
3. With probability at least  $p$ , for each of the binary sequences, at least a fraction  $\epsilon$  of its  $(n, d)$  summaries that include the first bit are unique.

It is important to note that the recovery guarantees of the presented combinatorial algorithms are only valid for a class of vectors in which no two disjoint subsets of nonzero coefficients have the exact same sum. For simplicity, we refer to these vectors as “distinguishable” signals. This is not the case for Basis Pursuit.

### 3.1 Basis Pursuit

The success of the basis pursuit algorithm for recovering sparse signals is certified by several conditions. Two major classes of conditions are the Restricted Isometry Property (RIP) and the null space property [11, 5]. It is provable that the measurement structures defined in this paper do not maintain the RIP properties, due to the existence of columns with fairly large coherence. This however does not discard the suitability of these constructions for  $\ell_1$  minimization, since RIP is known to provide a sufficient condition (see e.g., [17]). Instead, we prove that certain null space conditions hold for the considered class of matrices, and therefore provide a sparse signal recovery bound for  $\ell_1$  minimization. We restrict our attention to nonnegative vectors in this case. The reconstruction method is the following program with the additional nonnegativity constraint.

$$\begin{aligned} & \text{minimize} \quad \|\mathbf{x}\|_1 \\ & \text{subject to} \quad A\mathbf{x} = \mathbf{y}, \mathbf{x} \geq 0 \end{aligned} \tag{2}$$

The performance of the above program was studied for 0-1 matrices in [18]. In particular, it was shown that a nonnegative vector  $\mathbf{x}$  can be recovered from (2), if and only if it is the unique nonnegative solution of the linear system of equations, which is stated formally in the following lemma.

**Lemma 3.1** (from [18]). *Suppose  $A \in \mathbb{R}^{m \times n}$  is a matrix with constant column sum, and  $\mathbf{x}_0 \in \mathbb{R}^{n \times 1}$  is a nonnegative vector.  $\mathbf{x}_0$  is the unique solution to (2), if and only if  $\mathbf{x}_0$  is the unique nonnegative solution to  $A\mathbf{x} = A\mathbf{x}_0$ .*

Using the above lemma, we can evaluate the performance of the Basis Pursuit algorithm when used with the presented measurement matrices. The following theorem is fundamental to this analysis.

**Theorem 3.1** (Strong Recovery for Basis Pursuit). *Let  $k \leq f_S(n, d - 1)$  be an integer, and let  $A$  correspond to a complete  $(n, d)$  summary codebook. Then every  $k$ -sparse nonnegative vector  $\mathbf{x}$  is perfectly recovered by (2).*

*Proof.* Let  $k \leq f_S(n, d - 1)$  and let  $\mathbf{x}_0$  be a nonnegative  $k$ -sparse vector. Also, let the  $n$ -bit binary labels associated to the support set of  $\mathbf{x}_0$  be  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$ . We show that if  $A$  corresponds to a complete  $(n, d)$

summary codebook, then  $\mathbf{x}_0$  is the unique nonnegative solution to  $A\mathbf{x} = A\mathbf{x}_0$ . Therefore, by Lemma 3.1 it follows that  $\mathbf{x}_0$  can be recovered via (2). We prove this by contradiction. Suppose that there is another nonnegative vector  $\mathbf{x} \neq \mathbf{x}_0$  with  $A\mathbf{x} = A\mathbf{x}_0$ . Due to the nonnegativity assumption, we may assume that the support sets of  $\mathbf{x}$  and  $\mathbf{x}_0$  do not overlap. Let the  $n$ -bit labels of the support set of  $\mathbf{x}$  be the binary sequences  $\mathbf{b}'_1, \mathbf{b}'_2, \dots, \mathbf{b}'_\ell$ . From the definition of  $f_S(\cdot)$ , we can assert that there is a  $(n, d-1)$  summary that appears in exactly one of the sequences  $\mathbf{b}_1, \dots, \mathbf{b}_k$ . Let us assume without the loss of generality that the first  $d-1$  bits of  $\mathbf{b}_1$  are unique, and that  $\mathbf{b}_1$  is the all zero binary sequence. Therefore, there are at least  $n-d+1$  measurements in  $\mathbf{y} = A\mathbf{x}_0$  that are equal to the entry of  $\mathbf{x}_0$  that corresponds to the label  $\mathbf{b}_1$ . These measurements are those that correspond to the summaries

$$(\{1, 2, \dots, d-1, i\}, \mathbf{0}), \quad d \leq i \leq n \quad (3)$$

Since,  $A\mathbf{x} = A\mathbf{x}_0$ , there must be a nonzero entries in  $\mathbf{x}$  with labeling indices that satisfy the above summaries. In particular, without loss of generality assume that the first  $d$  bits of  $\mathbf{b}'_1$  are all zero. However, since the support sets of  $\mathbf{x}$  and  $\mathbf{x}_0$  do not overlap,  $\mathbf{b}'_1$  is different from  $\mathbf{b}_1$  in at least one bit, say  $\mathbf{b}_1(j) \neq \mathbf{b}'_1(j)$  for some  $j > d$ . Now consider the summary  $(S, \mathbf{c}) = (\{1, 2, \dots, d-1, j\}, 00\dots 01)$ , which represent the set of all binary sequences that are zero on the first  $d-1$  bits and one on the  $j$ th bit. Because  $A$  corresponds a complete  $(n, d)$  codebook, there is a row of  $A$  that is based on  $(S, \mathbf{c})$ , and moreover the corresponding value of  $\mathbf{y}$  is nonzero, because  $\mathbf{b}'_1$  conforms to  $(S, \mathbf{c})$ . On the other hand this cannot be true when considering the equations  $\mathbf{y} = A\mathbf{x}$ , because it requires that one of the labels  $\mathbf{b}_1, \dots, \mathbf{b}_k$  conform to  $(S, \mathbf{c})$ , which cannot be  $\mathbf{b}_1$  (recall that  $\mathbf{b}_1$  is the all zero codeword, whereas  $\mathbf{c}$  includes a 1). The existence of such a label contradicts the assumption that  $\mathbf{b}_1$  is the only label whose  $d-1$  first bits are all zero.

■

The complexity of Basis Pursuit is generally polynomial in the ambient dimension of the signal. Specifically, one can implement (2) in  $\mathcal{O}(N^3)$  operations, without exploiting any of the available structural information of the measurement matrix. Although there are some advantages to Basis Pursuit, such as robustness to noise, its complexity is impractical for problems where  $N$  scales exponentially. In these situations, sublinear time algorithms are preferred.

### 3.2 Summarized Support Index Inference

The first sublinear algorithm discussed in this subsection is called the summarized support index inference (SSII). The algorithm is based on iteratively inferring the nonzero entries of the signal based on one of the distinct values of  $\mathbf{y}$  and its various occurrences. The method is described below.

At the beginning of the algorithm, distinct nonzero values of the observations  $\mathbf{y}$  are identified, and are separated from the zero values. Due to the distinguishability assumption on  $\mathbf{x}$ , each distinct nonzero value of  $\mathbf{y}$  is a sum of a unique subset of nonzeros of  $\mathbf{x}$ , and can thus be used to infer the position

---

**Algorithm 1** SSII

---

- 1: Repeat until all nonzeros of  $\mathbf{x}$  are identified.
  - 2: Identify distinct nonzeros of  $\mathbf{y}$ , exhaust the following:
  - 3: Consider all occurrences of a value  $y_{\pi(1)} = \dots = y_{\pi(t)}$ .
  - 4: Construct a binary sequence  $\mathbf{b}$  by setting  $\mathbf{b}(S_{\pi(i)}) := \mathbf{c}_{\pi(i)}$ ,  $\forall 1 \leq i \leq t$ , where  $(S_j, \mathbf{c}_j)$  is the summary corresponding to measurement  $\mathbf{y}_j$ .
  - 5: If  $\mathbf{b}$  is fully characterized without conflict from previous step, then a nonzero entry of  $\mathbf{x}$  has been determined. subtract it, update  $\mathbf{y}$  and go to step 2. Otherwise, exhaust the following step.
  - 6: Find a subset  $S'$ , such that among summaries  $(S', \mathbf{c})$  that do not contradict with  $\mathbf{b}$ , exactly one corresponds to a nonzero of  $\mathbf{y}$ , say  $(S', \mathbf{c}')$ , and set  $\mathbf{b}(S') := S'$ .
- 

of at least one nonzero entry. The index of a nonzero entry of  $\mathbf{x}$  is determined by its unique labeling, which is a binary sequence of length  $n$ . Therefore, the algorithm attempts to infer all relevant binary sequences. Suppose that a nonzero value of  $\mathbf{y}$  is chosen that has  $t$  occurrences, say without loss of generality,  $y_1 = y_2 = \dots = y_t$ . Also, let the  $(n, d)$  summary which corresponds to the  $i$ th row of  $A$  be denoted by  $(S_i, \mathbf{c}_i)$  (see equation (1)). The algorithm explores the possibility that  $y_1, y_2, \dots, y_t$  are all equal to a single nonzero entry of  $\mathbf{x}$ , by trying to build a binary sequence  $\mathbf{b}$  that conforms to the summaries  $\{(S_i, \mathbf{c}_i)\}_{i=1}^t$ , i.e., by setting:

$$\mathbf{b}(S_i) := \mathbf{c}_i, \forall 1 \leq i \leq t \quad (4)$$

If there is a conflict in the set of equations in (4), then that value of  $\mathbf{y}$  is discarded in the current iteration, and the search is continued for other values. Otherwise, two events may occur. If (4) uniquely identifies  $\mathbf{b}$ , then one nonzero position and value of  $\mathbf{x}$  has been determined. It is subtracted, measurements are updated and the algorithm is continued. However, there might be a case where only  $n_1 < n$  bits of  $\mathbf{b}$  are determined by (4). In this case, we use the zero values of  $\mathbf{y}$  to infer the remaining  $n - n_1$  bits in the following way. Let the set of known and unknown bits of  $\mathbf{b}$  be denoted by  $S_1$  and  $S_2$ , respectively. We consider the summaries  $(S, \mathbf{c})$  which contribute to  $A$ , and among all, consider all distinct subsets  $S$ . If there is a subset  $S'$  such that among all the measurements corresponding to  $(S', \mathbf{c})$  where  $\mathbf{c}$  does not conflict with  $\mathbf{b}(S')$ , exactly one of them are nonzero, say  $(S', \mathbf{c}')$ , then the bits of  $\mathbf{b}$  over  $S' \cap S_2$  can be uniquely determined by setting  $\mathbf{b}(S') = \mathbf{c}'$ . This procedure is repeated until either  $\mathbf{b}$  is completely identified, or all possibilities are exhausted. A high level description of the presented method is given in Alg. 1, for which we can assert the following weak and strong recovery guarantees.

**Theorem 3.2** (Strong Recovery for SSII). *Let  $k \leq f_S(n, d - 1)$  be an integer, and let  $A$  correspond to a complete  $(n, d)$  summary codebook. Then every  $k$ -sparse distinguishable vector  $\mathbf{x}$  is perfectly recovered by Alg. 1.*

*Proof.* Let  $k \leq f_S(n, d - 1)$  and let  $\mathbf{x}$  be a  $k$ -sparse vector. Also, let the  $n$ -bit binary labels associated to the support set of  $\mathbf{x}$  be  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$ . We show that at least one of these labels can be inferred from one of the nonzero values of the vector  $\mathbf{y} = A\mathbf{x}$ , by solving (4). From the definition, there is a  $(n, d - 1)$  summary that appears in exactly one of the labels  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$ . Without loss of generality, let's assume

that the first  $d-1$  bits of  $\mathbf{b}_1$  are unique, and that  $\mathbf{b}_1$  is the all zero binary sequence. Also, let the nonzero value of  $\mathbf{x}$  in the position given by  $\mathbf{b}_1$  be  $\gamma$ . Now consider all summaries  $(S, \mathbf{c})$  for which the value of the corresponding entry in  $\mathbf{y}$  is equal to  $\gamma$ . Let these summaries be denoted by  $\{(S_i, \mathbf{c}_i)\}_{i=1}^t$ , where  $t$  is the number of occurrences of  $\gamma$  in  $\mathbf{y}$ . We show that there is a unique binary sequence  $\mathbf{b}'$  that conforms to all of these summaries. In other words, we prove that equation (4) has a unique solution which is equal to  $\mathbf{b}' = \mathbf{b}_1$ .

Due to the distinguishability assumption on the nonzero values of  $\mathbf{x}$ , The set  $\{(S_i, \mathbf{c}_i)\}_{i=1}^t$  should include the following summaries:

$$(\{1, 2, \dots, d-1, i\}, \mathbf{0}), \quad d \leq i \leq n \quad (5)$$

Where  $\mathbf{0}$  indicates the all zero bit sequence of length  $d$ . Clearly the only length  $n$  binary sequence that conforms to all of the above summaries is the all zero binary sequence, namely  $\mathbf{b}_1$ . Thus, we only need to show that  $\mathbf{b}_1(S_i) = \mathbf{c}_i$  for all other summaries  $(S_i, \mathbf{c}_i)$ ,  $1 \leq i \leq t$ . This also follows immediately from the distinguishability assumption on  $\mathbf{x}$ , and the fact that every instance of  $\gamma$  in the vector  $\mathbf{y}$  is only the result of the nonzero value in  $\mathbf{x}$  labeled by  $\mathbf{b}_1$  (i.e. it is not the direct sum of another subset of the entries of  $\mathbf{x}$ ). ■

**Theorem 3.3** (Weak Recovery for SSII). *Let  $k \leq f'_W(n, d, p, \epsilon)$  be an integer, and let  $A$  correspond to a random  $(n, m, d)$  summary codebook. Then, a random  $k$ -sparse distinguishable vector  $\mathbf{x}$  is recovered by Alg. 1 with probability at least  $1 - kn(1 - p + p(1 - \frac{\epsilon d}{n})^m)$ .*

*Proof.* We define an event  $\mathcal{E}$  which is stronger than the success event of Algorithm 1, namely a sufficient condition for the success of SSII. Let the  $n$ -bit binary labels associated to the support set of  $\mathbf{x}$  be  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$ , and let  $\mathcal{C}$  be the  $(n, m, d)$  summary codebook based on which  $A$  is constructed. The sufficient condition for success of SSII is that for every  $1 \leq i \leq k$ , and every bit  $1 \leq j \leq n$ , there exists a summary  $(S, \mathbf{c})$  in  $\mathcal{C}$  such that  $j \in S$  and in addition  $\mathbf{b}_i(S) = \mathbf{c}$  and  $\mathbf{b}_\ell \neq \mathbf{c} \forall \ell \neq i$ . In other words, for each of the  $k$  labels corresponding to the support of  $\mathbf{x}$  and each of the  $n$  bits, there is a summary in the codebook that includes the considered bit and only conforms to that particular label.

We find a lower bound on the probability of the complementary event  $\mathcal{E}^c$  by using union bounds. Note that there are  $m$  distinct subsets in the summaries of the codebook  $\mathcal{C}$ , which are chosen randomly. We assume that the subsets are chosen independently at random, and allow repetition. In case of repetition, the repeated subset is excluded, which only makes things worst. Consider a label  $\mathbf{b}_1$  and the first bit. The probability that a randomly chosen subset of bits of length  $d$  includes the first bit is  $\frac{d}{n}$ . Furthermore, let us say that at least a fraction  $\epsilon'$  of the summaries that conform to  $\mathbf{b}_1$  and include the first bit, does not conform to the remaining  $\mathbf{b}_i$ 's (i.e. only appear in  $\mathbf{b}_1$ ). Then, when a random subset  $S$  is chosen, with probability at least  $\frac{\epsilon' d}{n}$ , the following happens:

$$1 \in S \text{ and } \mathbf{b}_i(S) \neq \mathbf{b}_1(S) \forall 1 < i \leq k \quad (6)$$



Therefore, the probability that the above event does not happen for any of  $m$  randomly chosen subsets  $S$  is at most  $(1 - \frac{\epsilon'd}{n})^m$ . From the definition of  $f'_W(\cdot)$  and the fact that  $k \leq f'_W(n, d, p, \epsilon)$ , we know that with probability at least  $p$ ,  $\epsilon' \geq \epsilon$ , and therefore, the probability that (6) does not happen for any set  $S$  in the codebook  $\mathcal{C}$  is at most  $1 - p + p(1 - \frac{\epsilon d}{n})^m$ . If we union bound the probability of such event for all possible  $k$  labels and all possible  $n$  bits, we conclude that the probability of the undesirable event  $\mathcal{E}^c$  is bounded by:

$$\mathbb{P}(\mathcal{E}^c) \leq nk(1 - p + p(1 - \frac{\epsilon d}{n})^m) \quad (7)$$

Which concludes the proof of the theorem.  $\blacksquare$

The explicit recovery bounds given by above theorem are calculated in Section 4. Alg. 1 can be implemented very efficiently, with  $\mathcal{O}(\max(\text{poly}(M), k \log N))$  operations, which is sublinear in the dimension of the problem. The computational advantage is owed to the most part to the structural definition of the measurement matrices which facilitates sublinear search over the column space of the matrix. In addition, we do not require an exponential memory for decoding, since the information about  $A$  and the current inferred indices of the unknown vector at each stage can be retained by only storing the corresponding binary indices.

### 3.3 Mix and Match Algorithm

We describe a third recovery method, which is on the lines of the algorithm proposed in [2] with slight modifications. The algorithm consists of two subroutines: a value identification phase in which the nonzero values of the unknown signal is determined, and a second phase for identifying the support set of  $\mathbf{x}$ . The method is based on measurements given by  $\mathbf{y} = (\mathbf{y}^{(1)T}, \mathbf{y}^{(2)T})^T = (A_1^T, A_2^T)^T \mathbf{x}$ , where only  $\mathbf{y}^{(1)}$  is used for the first phase, and  $\mathbf{y}^{(2)}$  and  $A_2$  are used in the second phase. For details of this method please refer to [2]. We analyze this algorithm for the proposed measurement structures of this paper, which is different from the analysis of [2].

---

#### Algorithm 2 M&M

---

Find the set  $Y$  of nonzero entries of  $\mathbf{y}_1$  and set  $X = \emptyset$ .  $X$  will determine the set of nonzeros of  $\mathbf{x}$ .  
Repeat steps 2,3 until  $S(X) = Y$ .  
Update the set  $S(X)$  of sums of subsets of  $X$ .  
Find the smallest entry of  $Y$  that is not in  $S(X)$ , and add it to  $X$ .  
Initiate zero binary sequences  $\{\mathbf{b}_x | x \in X\}$ , which will determine the labeling of the support indices of  $\mathbf{x}$ .  
For every nonzero entry of  $\mathbf{y}_2$ , find the corresponding summary  $(S, \mathbf{c})$ , and a subset  $X' \subset X$  that sum up to that value of  $\mathbf{y}_2$ . Set  $\mathbf{b}_x(S) = \mathbf{c}, \forall x \in X'$ .

---

**Theorem 3.4** (Weak Recovery for M&M). *Let  $k \leq f_W(n, d, p, \epsilon)$  be an integer, and  $A = (A_1^T A_2^T)^T$  where  $A_1$  and  $A_2$  correspond to a random  $(m, n, d)$  summary codebook, and a complete  $(n, 1)$  summary codebook, respectively. Then, a random nonnegative  $k$ -sparse distinguishable vector  $\mathbf{x}$  is recovered by Alg. 2 with probability at least  $p(1 - k(1 - \epsilon)^m)$ .*

*Proof.* Let the  $n$ -bit binary labels associated to the support set of  $\mathbf{x}$  be  $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_k$ , and let  $\mathcal{C}_1$  be the  $(n, m, d)$  summary codebook based on which  $A_1$  is constructed. It can be shown that the value identification subroutine of Alg. 2 identifies all nonzero values of the nonnegative vector  $\mathbf{x}$  correctly, if in the observation vector  $\mathbf{y}$ , all every nonzero value of  $\mathbf{x}$  appear at least once. We find the probability that this condition holds, when the  $m$  subsets of the random codebook  $\mathcal{C}_1$  are chosen at random. For every  $1 \leq i \leq k$ , we define the following set of subsets of  $\{1, 2, \dots, n\}$ :

$$\mathcal{U}_i = \{S \mid |S| = d, \mathbf{b}_j(S) \neq \mathbf{b}_i(S) \forall j \neq i\} \quad (8)$$

If a subset  $S$  in the codebook  $\mathcal{C}_1$  belongs to  $\mathcal{U}_i$ , then the nonzero entry  $\gamma_i$  that corresponds to the label  $\mathbf{b}_i$  appears in the observation vector  $\mathbf{y}$ . Therefore, we are interested in finding the probability that the set of  $m$  subsets of  $\mathcal{C}_1$  has a nonempty overlap with all  $\mathcal{U}_i$ 's. Let us assume that for some  $\epsilon' > 0$ , the following holds:

$$|\mathcal{U}_i| \geq \epsilon' \binom{n}{d}, \forall 1 \leq i \leq k \quad (9)$$

When a subset  $S$  is chosen at random, the probability that it belongs to  $\mathcal{U}_i$  is at least  $\epsilon'$ . Therefore the probability that  $\mathcal{U}_i$  does not overlap with the set of all subsets  $S$  appearing in  $\mathcal{C}_1$  is at most  $(1 - \epsilon')^m$ . Using a union bound over all  $1 \leq i \leq k$ , we conclude that the probability that this undesirable event happens for at least one of the sets  $\mathcal{U}_i$  is at most  $k(1 - \epsilon')^m$ , which means that the probability of success is at least  $1 - k(1 - \epsilon')^m$ . However, we know from the definition of  $f_W(\cdot)$ , and the fact that  $k \leq f_W(n, d, p, \epsilon)$ , that with probability at least  $p$ , we have  $\epsilon' > \epsilon$ . Therefore the overall probability of success is at least:

$$1 - p + p(1 - k(1 - \epsilon)^m) \geq p(1 - k(1 - \epsilon)^m) \quad (10)$$

■

The complexity of Alg. 2 is  $\mathcal{O}(\max(\text{poly}(M), 2^k))$ , and thus explodes when  $k$  grows.

## 4 Recovery Bounds

We derive recovery bounds for (2) and Alg.'s 1 and 2 by obtaining explicit bounds on the terms of definition 2 and replacing them in the recovery guarantees of Section 3, namely Theorems 3.1-3.4. The proof of the following lemma is based on some combinatorial techniques and Chernoff concentration bounds.

**Lemma 4.1.** *Let  $n, l$  and  $k$  be integers and  $0 < \alpha < 1/2$ . Also, let  $\epsilon = 1 - k \binom{\frac{n}{2}(1+\sqrt{2\alpha})}{l} / \binom{n}{l}$ , and  $p = 1 - k^2 e^{-\alpha n}$ . Then,*

1.  $f_S(n, l) \geq 2^l$ .
2.  $f_W(n, l, p, \epsilon) \geq k$ .

$$3. f'_W(n, l, p, \epsilon) \geq f_W(n-1, l-1, p, \epsilon).$$

By exploiting the expressions of the above lemma in Theorems 3.1-3.4, we obtain the following bounds for different methods:

**Basis Pursuit.** If a complete  $(n, d)$  summary codebook is used to build  $A$ , then the number of measurements is  $M = 2^d \binom{n}{d}$ , and every sparsity  $k \leq 2^{d-1}$  is guaranteed to be recovered. When put together (recall that  $n = \log_2 N$ ), an upper bound on the the required number of measurements for recoverable sparsity  $k$  is given by:

$$M = 2k \binom{\log N}{\log k} \quad (11)$$

In particular, for small values of  $k$ , the above bound is comparable with the  $M = 2k \log N$  bound of  $\ell_1$  minimization for random Gaussian matrices [19].

**SSII Algorithm.** We focus on the weak bound, namely the one obtained from Theorem 3.3. The general strategy is to take the values of  $p$  and  $\epsilon$  according to Lemma 4.1 with  $l = d - 1$ , and choose  $k$  and  $m$  in such a way that firstly,  $\epsilon$  is bounded away from zero, and secondly, the probability of recovery failure approaches zero as  $n \rightarrow \infty$ . Taking  $k = \lambda 2^{-d \log_2(\sqrt{\alpha/2+1/2})}$  for some  $0 < \lambda < 1$ , a few basic algebraic steps lead to the following:

$$\mathbb{P}(\text{failure}) \leq k^3 n e^{-\alpha n} + kn (1 - (1 - \lambda)d/n)^m, \quad (12)$$

It follows that the above expression approaches zero if  $m = \Omega(n \log n)$ . Furthermore,  $\alpha$  can be chosen arbitrarily close to zero. Therefore, it follows that an upper bound on the required number of measurements for successful recovery with high probability is given by:

$$M = \Omega(k \log N \log \log N). \quad (13)$$

**M&M Algorithm.** We take  $k = \lambda 2^{-d \log_2(\sqrt{\alpha/2+1/2})}$ , and  $\epsilon, p$  according to Lemma 4.1, it follows that:

$$\mathbb{P}(\text{failure}) \leq k^2 e^{-\alpha n} + k \lambda^m, \quad (14)$$

Which asymptotically vanishes if  $m = \Omega(\log k)$ . Recall that the number of measurements in this case is determined by the matrix  $A = [A_1^T, A_2^T]^T$  described in Theorem 3.4, which is equal to  $M = 2 \log N + m \times 2^d$ . Therefore, it follows that an upper bound on the required number of measurements for successful recovery with high probability is given by:

$$M = 2 \log N + \Omega(k \log k). \quad (15)$$

In particular, when  $k = o(\log \log N)$ , this means only  $\mathcal{O}(\log N)$  measurements are required, and the running time of the algorithm is  $\mathcal{O}(\log N)$  (see Section 3), both of which are almost optimal.

## 5 Simulations

Since Alg. 2 is only efficient for very small values of  $k$ , we present the empirical performance of Alg.

1. Due to the efficiency of the method, it is possible to perform simulations for very large values of

$N$ . In Figure 2, the empirical required over-sampling rate for Alg. 2 and the proposed constructions is plotted versus the signal dimension  $N$ , for various sparsity levels  $k$ . The required criteria here is that the probability of successful recovery be larger than 90%. Note that when  $N$  is increased by 3 orders of magnitude, the required number of measurements is increased by a factor of 3, which is an indication of the logarithmic dependence of  $M$  to  $N$ . Furthermore, as the signal becomes less sparse (i.e.  $k$  increases), the required oversampling factor decreases. For  $k = 100$ , this ratio is only about 3 for  $N = 1024$ , and about 8 for  $N = 3.3 \times 10^7$ . This is significantly better than existing sublinear recovery algorithms. Note that the optimal value of  $d$  for constructing the measurement matrices for every  $k, N$  is found empirically.

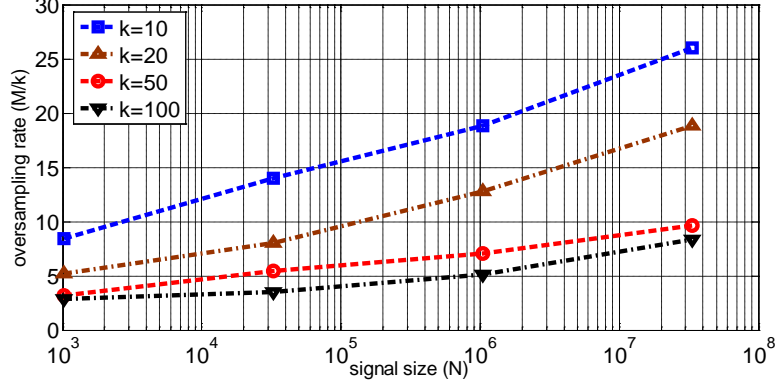


Figure 2: Required oversampling rate for successful recovery of Alg. 1 on proposed constructions versus signal dimension for various sparsity levels.

In Figure 3, the probability of successful recovery is plotted against the sparsity level  $k$  for  $N = 32768$ , and  $M = 140$  and  $240$ . We can see that although the number of measurements has only increased by a factor of 1.7, the recoverable sparsity (given a fixed probability of success) has improved in some cases by a factor of 5. These curves are comparable with the performance of  $\ell_1$  minimization over dense matrices, with  $N = 900$ , as displayed, which is an indication of the strong performance of the proposed scheme.

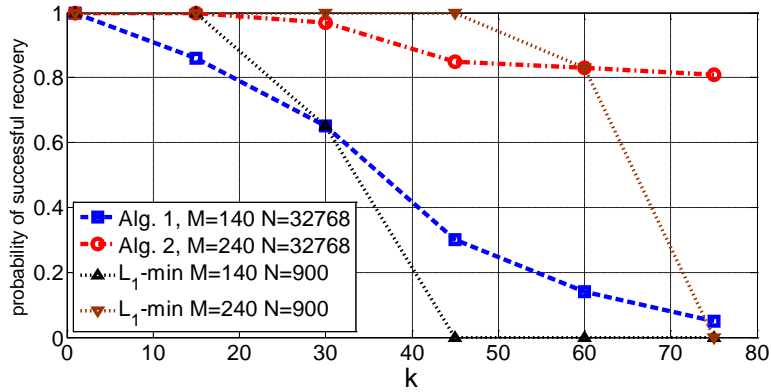


Figure 3: Probability of successful recovery of Alg. 1 versus sparsity level  $k$ , for  $N = 32768$  and  $M = 140, 240$ , and the same curves for  $\ell_1$ -minimization over i.i.d Gaussian matrices with  $N = 900$ .

## References

- [1] Compressed Sensing Online Resources, <http://dsp.rice.edu/cs>.
- [2] S. Jagabathula and D. Shah, *Inferring popular rankings under constrained sensing*, NIPS, 2008.
- [3] X. Jiang, Y. Yao and L. Guibas, *Stable Identification of Cliques with Restricted Sensing*, NIPS 2009.
- [4] V. Cevher, *Learning with Compressible Priors*, NIPS 2009.
- [5] Mihailo Stojnic, Weiyu Xu and Babak Hassibi, *Compressed Sensing - Probabilistic Analysis of a Null-space Characterization*, ICASSP 2008.
- [6] Y. Eldar, *Compressed Sensing of Analog Signals in Shift-Invariant Spaces*, IEEE Tran. on Sig. Proc., 57(8), 2986-2997.
- [7] J. Laska, S. Kirolos, M. Duarte, T. Ragheb and R. Baraniuk and Y. Massoud, *Theory and implementation of an analog-to-information converter using random demodulation*, ISCAS 2007.
- [8] M. Mishali and Y. C. Eldar, *From Theory to Practice: Sub-Nyquist Sampling of Sparse Wideband Analog Signals*, IEEE J. of Sel. Top. on Sig. Proc., 4(2):375391, 2010.
- [9] J. Luo, Yi Lu and B. Prabhaka, *Prototyping Counter Braids on NetFPGA*, Tech. Rep. 2008.
- [10] J. Yoo, S. Becker, E. Candès and Azita Emami, *A Random Modulation Pre-Integration Receiver for Sub-Nyquist Rate Signal Acquisition*, Preprint 2010.
- [11] E. J. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 4203-4215.
- [12] A. Gilbert, M. Strauss, J. Tropp, and R. Vershynin, *One Sketch for All: Fast Algorithms for Compressed Sensing*, STOC 2007.
- [13] D. Sarvotham, D. Baron, and R. Baraniuk, *Sudocodes - Fast Measurement and Reconstruction of Sparse Signals*, ISIT 2006.
- [14] W. Dai, O. Milenkovic and H. Pham, *Structured sublinear compressive sensing via dense belief propagation*, Preprint 2011.
- [15] G. Cormode and S. Muthukrishnan, *Combinatorial algorithms for compressed sensing*, CISS 2006.
- [16] R. Berinde, A. Gilbert, P. Indyk, M. Karloff, and M. Strauss, *Combining Geometry and Combinatorics: a Unified Approach to Sparse Signal Recovery*, Allerton 2008.
- [17] J. Blanchard, C. Cartis, and J. Tanner, *Compressed Sensing: How Sharp Is the Restricted Isometry Property?*, SIAM Rev. 53 105-125, 2010.

- [18] A. Khajehnejad, W. Xu, A. G. Dimakis, B. Hassibi, *Sparse Recovery of Positive Signals with Minimal Expansion*, IEEE Tran. on Signal Proc., 59(1),196-208, 2010.
- [19] V. Chandrasekaran, B. Recht, P. Parrilo and A. Willsky, *The Convex Geometry of Linear Inverse Problems*, available at arXiv:1012.0621v1.