Construction of binary matrices for near-optimal compressed sensing

Ivan Lau and Jonathan Jedwab
Department of Mathematics
Simon Fraser University
Burnaby, BC V5A 1S6, Canada
Email: {iplau, jed}@sfu.ca

***Abstract*—An efficient compressed sensing scheme requires a small number of measurements, a fast recovery algorithm, a small approximation error, and little or no randomness. In 2014, Iwen presented two compressed sensing schemes with near-optimal runtime, based on binary matrices. We combine ideas from these two schemes with a classical construction, used by Porat and Rothschild for near-optimal group testing, to produce a new compressed sensing scheme requiring significantly less randomness without compromising runtime. We give two variants of this compressed sensing scheme: the first is measurement-optimal, and the second is deterministic.**

I. Introduction

A. Context

Compressed sensing [1]–[3] concerns the problem of recovering an approximately sparse vector from a relatively small number of linear measurements. The unknown vector $\mathbf{x} \in \mathbb{R}^N$ we wish to recover is approximately k-sparse, where $k \ll N$. We take m non-adaptive linear measurements using a measurement matrix $\mathcal{M} \in \mathbb{C}^{m \times N}$ to yield a sketch $\mathbf{y} = \mathcal{M}\mathbf{x}$, and then use a recovery algorithm Δ to obtain an approximation $\widehat{\mathbf{x}} \in \mathbb{R}^N$ to $\mathbf{x}$ from $\mathbf{y}$. We would like the compressed sensing scheme $(\mathcal{M}, \Delta)$ to satisfy all the following properties:

(P1) the number of measurements m is small, ideally $O(k \ \mathrm{polylog} N)$.

(P2) the recovery algorithm Δ is fast, ideally having a runtime of $O(k \ \mathrm{polylog} N)$.

(P3) the error of the approximation $\mathbf{x} - \widehat{\mathbf{x}}$ is small, ideally achieving a mixed instance optimal error guarantee [4, §8] of the form

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_p \leq C k^{1/p-1/q} \min_{k\text{-sparse } \mathbf{x}_k} \|\mathbf{x} - \mathbf{x}_k\|_q \quad (1)$$

for some real constants p, q, C satisfying $1 \leq q \leq p$.

(P4) the number of random bits required to construct the measurement matrix $\mathcal{M}$ is small, ideally zero (in which case $\mathcal{M}$ is a deterministic matrix).

We will refer to (1) as an ℓ_p/ℓ_q error guarantee.

Uniform and nonuniform recovery models. When randomness is incorporated in accordance with property (P4), the measurement matrix $\mathcal{M}$ is randomly chosen from some specified distribution. There are two principal probabilistic models for the approximation error guarantee (1) to hold: the uniform and the nonuniform recovery model. In the uniform recovery model, for a single randomly-chosen matrix $\mathcal{M}$, with high probability the error guarantee is satisfied for all $\mathbf{x} \in \mathbb{R}^N$. For the nonuniform recovery model, for each fixed $\mathbf{x} \in \mathbb{R}^N$ and for a matrix $\mathcal{M}$ chosen randomly and independently for each $\mathbf{x}$, with high probability the error guarantee is satisfied.

In this paper, we provide a compressed sensing scheme which satisfies each of properties (P1) to (P4) under the nonuniform recovery model.

B. Related Work

A compressed sensing scheme which achieves an ℓ_1/ℓ_1, ℓ_2/ℓ_1, or ℓ_2/ℓ_2 error guarantee under the nonuniform recovery model must have a growth rate of $\Omega(k \log(N/k))$ for both the number of measurements and the runtime [5], [6]. There has been considerable work attempting to meet these bounds; we summarize the principal previous results in Table I. These schemes have either the best performance for at least one of the four properties (P1)–(P4) or a strong performance across all four. In addition to the compressed sensing schemes summarized in Table I and the references therein, we also refer the reader to [7]–[11] for previous work on deterministic schemes, to [12]–[14] for previous work on compressed sensing with derandomization, and to [15], [16] for background on combinatorial group testing.

C. Our Contributions

Our main contribution (Corollary 3) is a compressed sensing scheme with a nonuniform ℓ_2/ℓ_1 error guarantee. This scheme performs well across all four properties (P1)–(P4) simultaneously: few measurements, fast recovery algorithm, small approximation error, and few random bits. Both the number of measurements and the recovery algorithm runtime are $O(k \log k \cdot \log N)$. When $k = O\left(N^{1-\xi}\right)$ for an arbitrary positive constant ξ, these growth rates are within a factor of $O(\log k)$ of the lower bounds of Section I-B. Furthermore, the runtime is equal to that of the fastest known recovery algorithm [17, Theorem 5 (3)], and yet the number of random bits required is reduced from $O\left(Nk^2 \log N\right)$ to $O\left(\log k \cdot \log\left(k \log N\right)\right)$.

A first variant of Corollary 3 (Corollary 4) is a scheme using only $O(k \log N)$ measurements, at the cost of increased runtime and more random bits. When $k = O\left(N^{1-\xi}\right)$ for an arbitrary positive constant ξ, the number of measurements

is $O(k \log(N/k))$, which is order-optimal. To our knowledge, this scheme requires fewer random bits than all other measurement-optimal schemes satisfying an ℓ_p/ℓ_q guarantee.

A second variant of Corollary 3 (Corollary 5) is a deterministic scheme which, to our knowledge, achieves a faster runtime than all other deterministic schemes satisfying an ℓ_p/ℓ_q guarantee.

Our constructions are similar to those of [17], in that we subsample rows of an incoherent binary matrix whose row count grows sublinearly. Indeed, certain aspects of our results could be inferred from [17]. However, our innovation lies in connecting in novel ways various ideas from the information theory literature involving compressed sensing, coding theory, and combinatorial group testing. In particular, we shall show that methods from Porat and Rothschild's construction of a non-adaptive combinatorial group testing scheme [18] can be modified to produce an incoherent binary matrix that can be seen to be optimal by means of the Johnson bound [19] of classical coding theory.

The compressed sensing schemes of Corollaries 3, 4, 5 each satisfy an ℓ_2/ℓ_1 error guarantee. We expect that these schemes can be modified in a straightforward way to account for post-measurement noise while still satisfying an ℓ_2/ℓ_1 error guarantee and without affecting performance in relation to properties (P1)–(P4).

II. Ingredients of the Construction

Throughout, we will write $[N]$ as $\{0, 1, \ldots, N-1\}$. Given a vector $\mathbf{x} \in \mathbb{R}^N$ and $k \in [N]$, we denote by $\mathbf{x}_k$ a best k-term approximation to $\mathbf{x}$ (namely $\mathbf{x}$ with all but k of the largest-magnitude terms set to zero). We consider rows and columns of matrices to be indexed from 0.

A. Incoherent Binary Matrices

We firstly introduce incoherent binary matrices, which will be used in the construction of the measurement matrix $\mathcal{M}$. See [27] for an overview of the relationship of these binary matrices to matrices with the restricted isometry property, error-correcting codes with large distance, incoherent spherical codes, list-decodable codes, disjunct matrices, and combinatorial designs.

Definition 1 ([28, Definition 4]). *Let N, K, α be integers which satisfy $1 \leq \alpha < K < N$. An $m \times N$ binary $\{0,1\}$ matrix $\mathcal{M}_{\mathrm{C}}$ is (K, α)-coherent if each column of $\mathcal{M}_{\mathrm{C}}$ contains at least K ones and each pair of distinct columns of $\mathcal{M}_{\mathrm{C}}$ has dot product at most α.*

[19, Theorem 3] introduces an auxiliary function associated with the well-known problem of determining the maximum number $A(n, w, d)$ of codewords in a length n binary code of distance d and constant weight w, and derives an upper bound on this function using elementary methods. By making a connection between this auxiliary function and a (K, α)-coherent matrix, we obtain the following lower bound on the row count of such a matrix.

Theorem 1. *Let $\mathcal{M}_{\mathrm{C}} \in \{0,1\}^{m \times N}$ be a (K, α)-coherent matrix. Then*

$$m \geq \frac{NK^2}{(N-1)\alpha + K} = \Omega\left(\frac{K^2}{\alpha}\right).$$

Kautz and Singleton, in a study of nonrandom binary superimposed codes, give a strongly explicit construction[1] of a certain disjunct matrix by transforming a Reed-Solomon code [29] into a binary code [30, Section V B]. DeVore [31, Theorem 3.1] uses the same construction to produce matrices with the restricted isometry property. By reformulating the Kautz-Singleton result, we obtain the following construction for a $(K, \log_K N - 1)$-coherent matrix.

Theorem 2. *Let N, K, α be positive integers satisfying $N > K > \alpha = \log_K N - 1$. Then we can construct a (K, α)-coherent matrix $\mathcal{M}_{\mathrm{KS}}$ of size $K^2 \times N$. Furthermore, the matrix comprises exactly K blocks of K rows, each column of each row block containing exactly one 1. The position of this 1 can be located in $O(\log_K N)$ using Horner's rule.*

Porat and Rothschild give a derandomized probabilistic construction of a linear code meeting the Gilbert-Varshamov bound [32], [33], and produce from this an explicit construction of a disjunct matrix suitable for group testing whose row count is close to optimal [18]. Cheraghchi [27, Section III] uses the same construction to produce matrices with the restricted isometry property. By interpreting the Porat-Rothschild result in the framework of (K, α)-coherent matrices, we obtain the following construction for a $(\Theta(K), \Theta(\log N))$-coherent matrix, having order-optimal row count by Theorem 1.

Theorem 3. *Let N, K, α be positive integers satisfying $N > K > \alpha = \Omega(\log N)$. Then we can construct a (K, α)-coherent matrix $\mathcal{M}_{\mathrm{PR}}$ of size $m \times N$ where $m = \Theta\left(K^2/\alpha\right)$. Furthermore, the matrix comprises exactly K blocks of $\Theta(K/\alpha)$ rows, each column of each row block containing exactly one 1.*

B. Columnwise Kronecker Product with Bit-Test Matrix

Definition 2. *The columnwise Kronecker product of matrices $\mathcal{B} \in \mathbb{R}^{b \times N}$ and $\mathcal{R} \in \mathbb{R}^{t \times N}$, denoted $\mathcal{B} \circledast \mathcal{R}$, is the $bt \times N$ matrix whose entries are given by $(\mathcal{B} \circledast \mathcal{R})_{ti+v,j} = \mathcal{B}_{i,j}\mathcal{R}_{v,j}$.*

Definition 3. *The N^{th} bit-test matrix $\mathcal{B}_N \in \{0,1\}^{(1+\lceil \log_2 N \rceil) \times N}$ is the binary $\{0,1\}$ matrix whose j^{th} column (read from bottom to top) equals the binary representation of j followed by 1.*

Remark 1. *Let $\mathcal{B}_N = (\mathbf{b}_i)_{i=0}^{\lceil \log_2 N \rceil}$ be the N^{th} bit-test matrix, and let $\mathcal{R} = (\mathbf{r}_j)_{j=0}^{t-1}$ be a $t \times N$ matrix. The vector $\mathbf{y}_{\mathrm{id}} = (\mathcal{B}_N \circledast \mathcal{R})\mathbf{x}$ comprises exactly $1 + \lceil \log_2 N \rceil$ blocks of t entries, the i^{th} block being $(\mathbf{b}_i \circledast \mathcal{R})\mathbf{x} \in \mathbb{R}^t$. Furthermore, $(\mathbf{y}_{\mathrm{id}})_{ti+j} = \langle \mathbf{b}_i \circledast \mathbf{r}_j, \mathbf{x} \rangle = ((\mathbf{b}_i \circledast \mathcal{R})\,\mathbf{x})_j$ for each i and each j. This will be used in Line 4 of Algorithm 1.*

[1] A construction of an $m \times N$ matrix is strongly explicit if each column of the matrix can be constructed in time poly(m), and explicit if each column can be constructed in time poly(m, N).

Paper	D/U/N	# m of measurements	Runtime of Δ	# random bits	Error Guarantee
[9], [20]	D	$k^{2-\varepsilon}$	LP	Deterministic	ℓ_2/ℓ_1
Herein, Corollary 5	D	$k^2 \log^2 N$	$k^2 \log^2 N$	Deterministic	ℓ_2/ℓ_1
[21]	U	$k \log(N/k)$	LP	$O\left(k \log(N/k) \cdot \log\left(k \log\left(N/k\right)\right)\right)$	ℓ_2/ℓ_1
[22]	U	$k \log^{\geq 2} N$	$k^2 \log^{\geq 2} N$	$\Omega(N)$	ℓ_2/ℓ_1
[23]	U	$k \log N$	$k^{>1} \log^{\geq 2} N$	$\Omega(N)$	ℓ_1/ℓ_1
[24, Theorem 4]	N	$k \log^3 N$	$k \log^3 N$	$\Omega(N)$	ℓ_2/ℓ_2
[25, Theorem 1.1]	N	$k \log(N/k)$	$k \log^{\geq 2} N$	$\Omega(N)$	ℓ_2/ℓ_2
[26, Theorem 1.2]	N	$k \log(N/k)$	$k \log^2(N/k)$	$\Omega(N)$	ℓ_2/ℓ_2
[17, Theorem 5 (2)]	N	$k \log^2 N$	$k \log^2 N$	$O\left(\log k \cdot \log\left(k \log N\right)\right) =^* O(\log^2 k)$	ℓ_2/ℓ_1
[17, Theorem 5 (3)]	N	$k \log k \cdot \log N$	$k \log k \cdot \log N$	$O\left(Nk^2 \log N\right)$	ℓ_2/ℓ_1
		$*$ follows from the assumption of [17] that $k = \Omega(\log N)$			
Herein, Corollary 3	N	$k \log k \cdot \log N$	$k \log k \cdot \log N$	$O\left(\log k \cdot \log\left(k \log N\right)\right)$	ℓ_2/ℓ_1
Herein, Corollary 4	N	$k \log N$	$N \log N$	$O\left(\log N \cdot \log\left(k \log N\right)\right)$	ℓ_2/ℓ_1

TABLE I

SUMMARY OF THE PRINCIPAL PREVIOUS RESULTS AND THE RESULTS OBTAINED IN THIS PAPER. THE COLUMN "D/U/N" INDICATES WHETHER THE COMPRESSED SENSING SCHEME APPLIES TO THE DETERMINISTIC (D) MODEL, THE UNIFORM (U) RECOVERY MODEL OR THE NONUNIFORM (N) RECOVERY MODEL. THE MEASUREMENT AND RUNTIME COMPLEXITIES ARE EACH SUBJECT TO O-NOTATION, SUPPRESSED FOR BREVITY. "LP" DENOTES THE TIME COMPLEXITY OF SOLVING A LINEAR PROGRAM OF N VARIABLES.

III. MAIN RESULTS

A. Overview of Techniques

The compressed sensing scheme that we will present in Corollary 3 combines the advantages of two schemes S_1, S_2 proposed by Iwen [17, Theorem 5 (2) and Theorem 5 (3)].

Both schemes S_1, S_2 use the recovery algorithm Δ given in Algorithm 1 below. Their measurement matrix takes the form $\mathcal{M} = \left[\begin{array}{c} \mathcal{M}_{\text{id}} \\ \hline \mathcal{M}_{\text{est}} \end{array}\right]$, where $\mathcal{M}_{\text{id}}$ is an identification matrix and $\mathcal{M}_{\text{est}}$ is an estimation matrix.

The estimation matrix $\mathcal{M}_{\text{est}}$ is obtained by randomly subsampling (with replacement) blocks of rows from a Kautz-Singleton (K, α)-coherent matrix $\mathcal{M}_{\text{KS}}$. The identification matrix $\mathcal{M}_{\text{id}}$ is the columnwise Kronecker product of the N^{th} bit-test matrix with a binary matrix obtained by randomly subsampling (with replacement) rows from some (K, α)-coherent matrix $\mathcal{M}_{\text{C}}$.

In scheme S_1, the matrix $\mathcal{M}_{\text{C}}$ is fixed to be $\mathcal{M}_{\text{KS}}$, and the required amount of randomness is reduced by subsampling on row blocks rather than on individual rows. In contrast, in scheme S_2, the matrix $\mathcal{M}_{\text{C}}$ is itself randomly generated [17, Theorem 2], and random subsampling is then carried out on individual rows. The row count of this randomly generated matrix $\mathcal{M}_{\text{C}}$ meets the lower bound in Theorem 1, and in comparison with S_1 requires fewer measurements and gives faster runtime, but the process of random subsampling on individual rows requires more random bits.

We will show in Corollary 3 that, by instead fixing $\mathcal{M}_{\text{C}}$ to be a Porat-Rothschild matrix $\mathcal{M}_{\text{PR}}$ and randomly subsampling blocks of rows, we can retain the advantages of S_2 over S_1 (fewer measurements and faster runtime) without incurring the penalty of more random bits.

B. Recovery Algorithm: Identification, Estimation, Pruning

We use the recovery algorithm of Iwen [17, Algorithm 1], which we present here as Algorithm 1. For each $n \in [N]$, we write $\mathcal{M}_{\text{est}}(n)$ to represent the submatrix of $\mathcal{M}_{\text{est}}$ comprising the rows of $\mathcal{M}_{\text{est}}$ whose entry in the n^{th} column is 1.

This algorithm has three phases: identification, estimation, and pruning. The identification phase uses the identification measurement $\mathbf{y}_{\text{id}} = \mathcal{M}_{\text{id}}\,\mathbf{x}$ to identify a set $T \subseteq [N]$ containing the indices of the entries of $\mathbf{x}$ having largest magnitude, potentially with some false positives. Then the estimation phase uses the estimation matrix $\mathcal{M}_{\text{est}}$, the estimation measurement $\mathbf{y}_{\text{est}} = \mathcal{M}_{\text{est}}\,\mathbf{x}$, and the index set T to construct a vector $\widehat{\mathbf{x}} \in \mathbb{R}^N$ by estimating the entries x_n at only the indices $n \in T$. Finally, the pruning phase sets to 0 all but $2k$ entries of $\widehat{\mathbf{x}}$ with the largest magnitude so that this modified approximation $\widehat{\mathbf{x}}$ to $\mathbf{x}$ satisfies an ℓ_2/ℓ_1 guarantee.

In the next three subsections, we describe the identification, estimation, and pruning phases in more detail.

C. Identification

We randomly subsample (with replacement) blocks of rows from a Porat-Rothschild (K, α)-coherent matrix $\mathcal{M}_{\text{PR}}$ (see Theorem 3) to generate a random binary matrix $\mathcal{R}$, and then construct the identification matrix $\mathcal{M}_{\text{id}}$ as the columnwise Kronecker product $\mathcal{B}_N \circledast \mathcal{R}$. The properties of the identification matrix and the performance of the identification phase of Algorithm 1 are described in Corollary 1 below, using the following lemma derived by combining results given in [17, Section 4.1].

Lemma 1. *Let $\mathcal{M}_{\text{C}}$ be an $m \times N$ (K, α)-coherent matrix comprising K blocks of $\frac{m}{K}$ rows, each column of each block containing exactly one 1. Let $\epsilon \in (0, 1]$, $\sigma \in \left[\frac{2}{3}, 1\right)$, $\mathbf{x} \in \mathbb{R}^N$ and $k \in [K\frac{\epsilon}{14\alpha}]$. Construct a matrix $\mathcal{R}$ by randomly choosing with replacement at least $\frac{7}{6} \ln\left(\frac{2k/\epsilon}{1-\sigma}\right)$ of the K blocks of $\mathcal{M}_{\text{C}}$, and let $\mathcal{M}_{\text{id}} = \mathcal{B}_N \circledast \mathcal{R}$. Then, with probability at least σ, the identification phase of Algorithm 1 recovers from the identification measurement $\mathbf{y}_{\text{id}} = \mathcal{M}_{\text{id}}\,\mathbf{x}$ the set T of indices $n \in [N]$ for which*

$$|x_n| \geq \frac{4\epsilon}{k} \left\|\mathbf{x} - \mathbf{x}_{k/\epsilon}\right\|_1.$$

Furthermore, the number of random bits used to construct $\mathcal{R}$ is $O\left(\log\left(\frac{2k/\epsilon}{1-\sigma}\right) \cdot \log K\right)$, and both the number of measure-

Algorithm 1 Recovery algorithm for approximating $\mathbf{x}$ [17]

Input: $\mathcal{M} = \left[\begin{array}{c} \mathcal{B}_N \circledast \mathcal{R} \\ \hline \mathcal{M}_{\text{est}} \end{array}\right]$, $\mathbf{y} = \left[\begin{array}{c} \mathbf{y}_{\text{id}} \\ \hline \mathbf{y}_{\text{est}} \end{array}\right] = \mathcal{M}\mathbf{x}$, and k, where $\mathcal{R}$ has size $t \times N$

Output: an approximation to $\mathbf{x}$ containing at most $2k$ nonzero terms

1: Initialize $T \leftarrow \emptyset$, $\widehat{\mathbf{x}} \leftarrow \mathbf{0} \in \mathbb{R}^N$, $\mathbf{v} \leftarrow \mathbf{0} \in \mathbb{R}^{\lceil \log_2 N \rceil}$

IDENTIFICATION PHASE

2: **for** j from 0 to $t-1$ **do**
3: **for** i from 1 to $\lceil \log_2 N \rceil$ **do**
4: **if** $|\mathbf{y}_{ti+j}| > |\mathbf{y}_j - \mathbf{y}_{ti+j}|$ **then**
5: $v_i \leftarrow 1$
6: **else**
7: $v_i \leftarrow 0$
8: **end if**
9: **end for**
10: $n \leftarrow \sum_{i=1}^{\lceil \log_2 N \rceil} v_i 2^i$
11: $T \leftarrow T \cup \{n\}$
12: **end for**

ESTIMATION PHASE

13: **for each** n in T **do**
14: $\widehat{x}_n \leftarrow$ median of the entries of $\mathcal{M}_{\text{est}}(n)\,\mathbf{x}$
15: **end for**

PRUNING PHASE

16: Sort by magnitude the w nonzero entries of $\widehat{\mathbf{x}}$ (where $w \leq |T|$) so that $|\widehat{x}_{n_1}| \geq |\widehat{x}_{n_2}| \geq \cdots \geq |\widehat{x}_{n_w}|$
17: **for** j from $\min(2k+1, w)$ to w **do**
18: $\widehat{x}_{n_j} \leftarrow 0$
19: **end for**
20: Output: $\widehat{\mathbf{x}}$

ments made by $\mathcal{M}_{\text{id}}$ and the runtime of the identification phase are $O\left(\frac{m}{K}\log\left(\frac{2k/\epsilon}{1-\sigma}\right)\cdot\log N\right)$.

Take $\alpha = \Theta(\log N)$ and $K = \alpha\frac{k}{\epsilon} = \Theta\left(\frac{k}{\epsilon}\log N\right)$ and $\mathcal{M}_{\text{C}} = \mathcal{M}_{\text{PR}}$ and $\sigma = 0.99$ in Lemma 1 to obtain the following properties of the identification matrix and the performance of the identification phase of our scheme. Recall from Theorem 3 that the matrix $\mathcal{M}_{\text{PR}}$ comprises exactly K blocks of $\Theta\left(K/\alpha\right)$ rows, each column of each row block containing exactly one 1.

Corollary 1. *Let $\epsilon \in (0,1]$, $\mathbf{x} \in \mathbb{R}^N$ and $k \in [N]$. Let $\mathcal{M}_{\text{PR}}$ be a Porat-Rothschild $\left(\Theta\left(\frac{k}{\epsilon}\log N\right), \Theta(\log N)\right)$-coherent matrix. Construct a matrix $\mathcal{R}$ by randomly choosing with replacement at least $\frac{7}{6}\ln\left(200\frac{k}{\epsilon}\right)$ of the $\Theta(\frac{k}{\epsilon}\log N)$ blocks of $\Theta\left(\frac{k}{\epsilon}\right)$ rows of $\mathcal{M}_{\text{PR}}$, and let $\mathcal{M}_{\text{id}} = \mathcal{B}_N \circledast \mathcal{R}$. Then, with probability at least 0.99, the identification phase of Algorithm 1 recovers from the identification measurement $\mathbf{y}_{\text{id}} = \mathcal{M}_{\text{id}}\,\mathbf{x}$ the set T of indices $n \in [N]$ for which*

$$|x_n| \geq \frac{4\epsilon}{k}\left\|\mathbf{x} - \mathbf{x}_{k/\epsilon}\right\|_1.$$

Furthermore, the number of random bits used to construct $\mathcal{R}$ (and hence $\mathcal{M}_{\text{id}}$) is $O\left(\log\left(\frac{k}{\epsilon}\right)\cdot\log\left(\frac{k}{\epsilon}\log N\right)\right)$, and both the number of measurements made by the measurement matrix $\mathcal{M}_{\text{id}}$ and the runtime of the identification phase are $O\left(\frac{k}{\epsilon}\log\left(\frac{k}{\epsilon}\right)\cdot\log N\right)$.

D. Estimation

The estimation matrix $\mathcal{M}_{\text{est}}$ is generated by randomly subsampling (with replacement) blocks of rows from a Kautz-Singleton (K,α)-coherent matrix $\mathcal{M}_{\text{KS}}$ (see Theorem 2). The properties of the estimation matrix and the performance of the estimation phase of Algorithm 1 are described in Corollary 2 below, using the following lemma derived by combining results given in [17, Section 4.2].

Lemma 2. *Let $\mathcal{M}_{\text{C}}$ be an $m \times N$ (K,α)-coherent matrix comprising K blocks of $\frac{m}{K}$ rows, each column of each block containing exactly one 1. Let $\epsilon \in (0,1]$, $\sigma \in \left[\frac{2}{3}, 1\right)$, $\mathbf{x} \in \mathbb{R}^N$, $k \in [K\frac{\epsilon}{14\alpha}]$, and let T be a subset of $[N]$. Construct a matrix $\mathcal{M}_{est}$ by randomly choosing with replacement at least $28.56\ln\left(\frac{2|T|}{1-\sigma}\right)$ of the K blocks of $\mathcal{M}_{\text{C}}$. Then, with probability at least σ, the estimation phase of Algorithm 1 estimates a vector $\widehat{\mathbf{x}}$ satisfying*

$$|\widehat{x}_n - x_n| \leq \frac{\epsilon}{k}\left\|\mathbf{x} - \mathbf{x}_{(k/\epsilon)}\right\|_1 \quad \textit{for all } n \in T.$$

Furthermore, the number of random bits used to construct $\mathcal{M}_{\text{est}}$ is $O\left(\log\left(\frac{2|T|}{1-\sigma}\right)\log K\right)$, and the number of rows of the estimation matrix $\mathcal{M}_{\text{est}}$ is $O\left(\frac{m}{K}\log\left(\frac{2|T|}{1-\sigma}\right)\right)$. The runtime of the estimation phase is $O\left(|T|\log\left(\frac{2|T|}{1-\sigma}\right)\log_K N\right)$ in the case that $\mathcal{M}_{\text{C}}$ is a Kautz-Singleton $(\Theta(K), \Theta(\log_K N))$-coherent matrix $\mathcal{M}_{\text{KS}}$; the runtime of the estimation phase is $O\left(|T|\log|T|\right)$ in the case that $\mathcal{M}_{\text{C}}$ is a Porat-Rothschild $(\Theta(K), \Theta(\log N))$-coherent matrix[2] $\mathcal{M}_{\text{PR}}$.

We shall apply Lemma 2, taking T to be the set of indices identified in the identification phase of our scheme. Take $\alpha = \Theta\left(\log_{k/\epsilon} N\right)$ and $K = \alpha\frac{k}{\epsilon} = \Theta\left(\frac{k}{\epsilon}\log_{k/\epsilon} N\right)$ and $\mathcal{M}_{\text{C}} = \mathcal{M}_{\text{KS}}$ and $\sigma = 0.99$, so that $|T| = \Theta\left(\frac{k}{\epsilon}\log\left(\frac{k}{\epsilon}\right)\right)$ by Corollary 1 and Line 2 of Algorithm 1. We then obtain the following properties of the estimation matrix and performance of the estimation phase of our scheme.

Corollary 2. *Let $\epsilon \in (0,1]$, $\mathbf{x} \in \mathbb{R}^N$, $k \in [N]$, and let T be the set of indices identified in the identification phase, as described in Corollary 1. Let $\mathcal{M}_{\text{KS}}$ be a Kautz-Singleton $\left(\Theta\left(\frac{k}{\epsilon}\log_{k/\epsilon} N\right), \Theta(\log_{k/\epsilon} N)\right)$-coherent matrix. Construct a matrix $\mathcal{M}_{\text{est}}$ by randomly choosing with replacement at least $28.56\ln\left(200|T|\right)$ of the $\Theta\left(\frac{k}{\epsilon}\log N\right)$ blocks of $\Theta\left(\frac{k}{\epsilon}\log N\right)$ rows of $\mathcal{M}_{\text{KS}}$. Then, with probability at least 0.99, the estimation phase of Algorithm 1 estimates a vector $\widehat{\mathbf{x}}$ satisfying*

$$|\widehat{x}_n - x_n| \leq \frac{\epsilon}{k}\left\|\mathbf{x} - \mathbf{x}_{(k/\epsilon)}\right\|_1 \quad \textit{for all } n \in T.$$

Furthermore, the number of random bits used to construct $\mathcal{M}_{\text{est}}$ is $O\left(\log\left(\frac{k}{\epsilon}\right)\cdot\log\left(\frac{k}{\epsilon}\log N\right)\right)$, and the runtime of the estimation phase is $O\left(\frac{k}{\epsilon}\log\left(\frac{k}{\epsilon}\right)\cdot\log N\right)$.

[2]This assumes that the locations of the nonzero entries of each column of $\mathcal{M}_{\text{PR}}$ are precomputed. A similar assumption is not needed when $\mathcal{M}_{\text{C}} = \mathcal{M}_{\text{KS}}$ because of the strongly explicit construction of $\mathcal{M}_{\text{KS}}$: see discussion prior to Theorem 2.

E. Pruning

We summarize the overall performance of our scheme in Corollary 3, using the following lemma given in [17, Theorem 5]. We then show that this scheme satisfies an ℓ_2/ℓ_1 guarantee.

Lemma 3. *Let $\epsilon \in (0,1]$, $\sigma \in \left[\frac{2}{3}, 1\right)$, $\mathbf{x} \in \mathbb{R}^N$, and $k \in [N]$. Construct a matrix $\mathcal{R}$ randomly according to Lemma 1, and an estimation matrix $\mathcal{M}_{\text{est}}$ randomly according to Lemma 2 (not necessarily from the same (K,α)-coherent matrix $\mathcal{M}_{\text{C}}$). Then, with probability at least σ^2, Algorithm 1 returns a vector $\widehat{\mathbf{x}}$ whose approximation error satisfies*

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \leq \|\mathbf{x} - \mathbf{x}_k\|_2 + \frac{22\epsilon}{\sqrt{k}} \left\|\mathbf{x} - \mathbf{x}_{k/\epsilon}\right\|_1. \tag{2}$$

Remark 2. *The following modification of Lemma 3 gives a compressed sensing scheme satisfying a nonuniform ℓ_2/ℓ_1 guarantee. Take $\epsilon = 1$ in Lemma 3, and replace the parameter k by $2k$. The resulting approximation error guarantee satisfies*

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \leq \|\mathbf{x} - \mathbf{x}_{2k}\|_2 + \frac{22}{\sqrt{2}\sqrt{k}} \|\mathbf{x} - \mathbf{x}_{2k}\|_1$$

for each fixed $\mathbf{x} \in \mathbb{R}^N$. The first term on the right side can be bounded using $\|\mathbf{x} - \mathbf{x}_{2k}\|_2 = \|(\mathbf{x} - \mathbf{x}_k) - (\mathbf{x} - \mathbf{x}_k)_k\|_2$ $\leq \frac{1}{\sqrt{k}} \|\mathbf{x} - \mathbf{x}_k\|_1$ from [3, Proposition 2.3], and the second term can be bounded using $\|\mathbf{x} - \mathbf{x}_{2k}\|_1 \leq \|\mathbf{x} - \mathbf{x}_k\|_1$. Therefore this scheme satisfies a nonuniform ℓ_2/ℓ_1 error guarantee.

Apply the modification of Lemma 3 described in Remark 2 with $\sigma = 0.99$, taking the matrix $\mathcal{R}$ to be constructed randomly according to Corollary 1, and the estimation matrix $\mathcal{M}_{\text{est}}$ to be constructed randomly according to Corollary 2. We then obtain the following overall performance of our scheme.

Corollary 3. *Let $\mathbf{x} \in \mathbb{R}^N$ and $k \in [N]$. Construct a matrix $\mathcal{R}$ randomly according to Corollary 1, and a matrix $\mathcal{M}_{\text{est}}$ randomly according to Corollary 2. Then, with probability at least 0.99^2, Algorithm 1 (with the parameter k replaced by $2k$ as in Remark 2) returns a vector $\widehat{\mathbf{x}}$ whose approximation error satisfies a nonuniform ℓ_2/ℓ_1 guarantee. Furthermore, the number of random bits used to construct the measurement matrix $\mathcal{M} = \left[\frac{\mathcal{B}_N \circledast \mathcal{R}}{\mathcal{M}_{\text{est}}}\right]$ is $O(\log k \cdot \log(k \log N))$, and both the number of measurements made by $\mathcal{M}$ and the runtime of Algorithm 1 are $O(k \log k \cdot \log N)$.*

IV. Measurement-Optimal Variant

If we omit the identification measurement and the identification phase of Algorithm 1, and apply the estimation phase directly with $T = [N]$, then we obtain the measurement-optimal variant of Corollary 3 stated in Corollary 4. The measurement and runtime performance of this scheme matches that of [17, Theorem 5 (1)] but uses fewer random bits. To obtain this, apply the modification of Lemma 3 described in Remark 2, where Lemma 2 is called with $\alpha = \Theta(\log N)$ and $\epsilon = 1$ and $K = \Theta(k \log N)$ and $\mathcal{M}_{\text{C}} = \mathcal{M}_{\text{PR}}$ and $\sigma = 0.99$ and $T = [N]$.

Corollary 4. *Let $\mathbf{x} \in \mathbb{R}^N$ and $k \in [N]$. Let $\mathcal{M}_{\text{PR}}$ be a Porat-Rothschild $(\Theta(k \log N), \Theta(\log N))$-coherent matrix. Construct a measurement matrix $\mathcal{M}$ by randomly choosing with replacement at least $28.56 \ln(200N)$ of the $\Theta(k \log N)$ blocks of $\Theta(k)$ rows. Then, with probability at least 0.99, Algorithm 1 (with the parameter k replaced by $2k$ as in Remark 2) returns a vector $\widehat{\mathbf{x}}$ whose approximation error satisfies a nonuniform ℓ_2/ℓ_1 guarantee. Furthermore, the number of random bits used to construct $\mathcal{M}$ is $O(\log N \log(k \log N))$, the number of measurements made by the measurement matrix $\mathcal{M}$ is $O(k \log N)$, and the runtime [3] is $O(N \log N)$.*

V. Deterministic Variant

Finally, we provide a deterministic variant of Corollary 3, using a variant [4] of Algorithm 1 and [28, Theorem 4]. Essentially, we take $\mathcal{R}$ to be a Porat-Rothschild matrix $\mathcal{M}_{\text{PR}}$ (instead of generated by randomly subsampling rows from $\mathcal{M}_{\text{PR}}$), and $\mathcal{M}_{\text{est}}$ to be a Kautz-Singleton matrix $\mathcal{M}_{\text{KS}}$. This gives a deterministic guarantee at the cost of more measurements and slower runtime.

Corollary 5. *Let $\mathbf{x} \in \mathbb{R}^N$ and $k \in [N]$. Let $\mathcal{R} = \mathcal{M}_{\text{PR}}$ be a Porat-Rothschild $(\Theta(k \log N), \Theta(\log N))$-coherent matrix, and $\mathcal{M}_{\text{est}} = \mathcal{M}_{\text{KS}}$ be a Kautz-Singleton $(\Theta(k \log_k N), \Theta(\log_k N))$-coherent matrix. Then the measurement matrix $\mathcal{M} = \left[\frac{\mathcal{B}_N \circledast \mathcal{M}_{\text{PR}}}{\mathcal{M}_{\text{KS}}}\right]$ and the recovery algorithm of Algorithm 1 (with the parameter k replaced by $2k$ as in Remark 2, and certain lines modified as in Footnote 4) produces a deterministic compressed sensing scheme satisfying an ℓ_2/ℓ_1 error guarantee. Both the number of measurements made by the measurement matrix $\mathcal{M}$ and the runtime of Algorithm 1 are $O(k^2 \log^2 N)$.*

VI. Conclusion and Further Research

We have presented a nonuniform compressed sensing scheme (Corollary 3) for which both the number of measurements and the runtime are within a factor of $O(\log k)$ of the known lower bound $\Omega(k \log N/k)$.

As shown in Table I, the technique of [26, Theorem 1.2] achieves an order-optimal number of measurements, but its runtime is only within a factor of $O(\log N/k)$ of the lower bound. It remains as an important open question whether there are nonuniform compressed sensing schemes with an ℓ_1/ℓ_1, ℓ_2/ℓ_1, or ℓ_2/ℓ_2 error guarantee for which both the number of measurements and the runtime are order-optimal.

Acknowledgment

The authors thank Mark Iwen for suggesting this problem and for helpful feedback on a preliminary description of our results.

[3] assuming the locations of the nonzero entries of each column of $\mathcal{M}_{\text{PR}}$ are precomputed.

[4] T is initialized as a multiset in Line 1; Line 11 is changed from union to multiset union: "$T \leftarrow T \uplus \{n\}$"; Line 13 is changed to "**for each** n in T with multiplicity greater than $K/2$ **do**". See also [28, Algorithm 1]

REFERENCES

[1] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Transactions on Information Theory, vol. 52, no. 2, pp. 489–509, 2006.

[2] D. L. Donoho, "Compressed sensing," IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] S. Foucart and H. Rauhut, A Mathematical Introduction to Compressive Sensing. Birkhäuser Basel, 2013.

[4] 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, 2009.

[5] K. D. Ba, P. Indyk, E. Price, and D. P. Woodruff, "Lower bounds for sparse recovery," in Proceedings of the 2010 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), pp. 1190–1197, 2010.

[6] E. Price and D. P. Woodruff, "(1+ ε)-approximate sparse recovery," in 2011 IEEE 52nd Annual Symposium on Foundations of Computer Science, pp. 295–304, IEEE, 2011.

[7] M. A. Herman and T. Strohmer, "High-resolution radar via compressed sensing," IEEE Transactions on Signal Processing, vol. 57, no. 6, pp. 2275–2284, 2009.

[8] L. Applebaum, S. D. Howard, S. Searle, and R. Calderbank, "Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery," Applied and Computational Harmonic Analysis, vol. 26, no. 2, pp. 283–290, 2009.

[9] J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, and D. Kutzarova, "Explicit constructions of RIP matrices and related problems," Duke Mathematical Journal, vol. 159, no. 1, pp. 145–185, 2011.

[10] S. Li and G. Ge, "Deterministic sensing matrices arising from near orthogonal systems," IEEE Transactions on Information Theory, vol. 60, pp. 2291–2302, April 2014.

[11] R. R. Naidu, P. Jampana, and C. S. Sastry, "Deterministic compressed sensing matrices: Construction via Euler squares and applications," IEEE Transactions on Signal Processing, vol. 64, no. 14, pp. 3566–3575, 2016.

[12] A. S. Bandeira, M. Fickus, D. G. Mixon, and J. Moreira, "Derandomizing Restricted Isometries via the Legendre Symbol," Constructive Approximation, vol. 43, no. 3, pp. 409–424, 2016.

[13] P. Jung, R. Kueng, and D. G. Mixon, "Derandomizing compressed sensing with combinatorial design," Frontiers in Applied Mathematics and Statistics, vol. 5, p. 26, 2019.

[14] C. Clum and D. G. Mixon, "Derandomized compressed sensing with nonuniform guarantees for ℓ_1 recovery," arXiv preprint arXiv:1912.12045, 2019.

[15] D. Du and F. Hwang, Combinatorial group testing and its applications, vol. 12. World Scientific, 2000.

[16] M. Aldridge, O. Johnson, J. Scarlett, et al., "Group testing: an information theory perspective," Foundations and Trends in Communications and Information Theory, vol. 15, no. 3-4, pp. 196–392, 2019.

[17] M. Iwen, "Compressed sensing with sparse binary matrices: Instance optimal error guarantees in near-optimal time," Journal of Complexity, vol. 30, no. 1, pp. 1 – 15, 2014.

[18] E. Porat and A. Rothschild, "Explicit nonadaptive combinatorial group testing schemes," IEEE Transactions on Information Theory, vol. 57, no. 12, pp. 7982–7989, 2011.

[19] S. Johnson, "A new upper bound for error-correcting codes," IRE Transactions on Information Theory, vol. 8, no. 3, pp. 203–207, 1962.

[20] D. G. Mixon, "Explicit Matrices with the Restricted Isometry Property: Breaking the Square-Root Bottleneck," in Compressed Sensing and its Applications, pp. 389–417, Springer, 2015.

[21] D. M. Kane and J. Nelson, "A Derandomized Sparse Johnson-Lindenstrauss Transform," arXiv preprint arXiv:1006.3585, 2010.

[22] A. C. Gilbert, M. J. Strauss, J. A. Tropp, and R. Vershynin, "One sketch for all: Fast algorithms for compressed sensing," in Proceedings of the 39th annual ACM Symposium on Theory of Computing, pp. 237–246, 2007.

[23] A. C. Gilbert, Y. Li, E. Porat, and M. J. Strauss, "For-all sparse recovery in near-optimal time," ACM Transactions on Algorithms, vol. 13, no. 3, pp. 1–26, 2017.

[24] G. Cormode and S. Muthukrishnan, "Combinatorial Algorithms for Compressed Sensing," in International Colloquium on Structural Information and Communication Complexity, pp. 280–294, Springer, 2006.

[25] A. C. Gilbert, Y. Li, E. Porat, and M. J. Strauss, "Approximate Sparse Recovery: Optimizing Time and Measurements," SIAM Journal on Computing, vol. 41, no. 2, pp. 436–453, 2012.

[26] V. Nakos and Z. Song, "Stronger L2/L2 compressed sensing; without iterating," in Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, pp. 289–297, 2019.

[27] M. Cheraghchi, "Coding-theoretic methods for sparse recovery," in 2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 909–916, IEEE, 2011.

[28] J. Bailey, M. A. Iwen, and C. V. Spencer, "On the design of deterministic matrices for fast recovery of Fourier compressible functions," SIAM Journal on Matrix Analysis and Applications, vol. 33, no. 1, pp. 263–289, 2012.

[29] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," Journal of the Society for Industrial and Applied Mathematics, vol. 8, no. 2, pp. 300–304, 1960.

[30] W. Kautz and R. Singleton, "Nonrandom binary superimposed codes," IEEE Transactions on Information Theory, vol. 10, no. 4, pp. 363–377, 1964.

[31] R. A. DeVore, "Deterministic constructions of compressed sensing matrices," Journal of Complexity, vol. 23, no. 4, pp. 918–925, 2007.

[32] E. Gilbert, "A Comparison of Signalling Alphabets," The Bell System Technical Journal, vol. 31, no. 3, pp. 504–522, 1952.

[33] R. Varshamov, "Estimate of the number of signals in error correcting codes," Doklady Akademii Nauk SSSR, vol. 117, pp. 739–741, 1957.