
Construction of a Large Class of Deterministic Sensing Matrices that Satisfy a Statistical Isometry Property

Robert Calderbank
Mathematics and Electrical Engineering
Princeton University
NJ 08544, USA

Stephen Howard
DSTO
PO Box 1500
Edinburgh 5111, Australia

Sina Jafarpour
Computer Science
Princeton University
NJ 08544, USA

Abstract

Compressed Sensing aims to capture attributes of a signal using very few measurements. The *Restricted Isometry Property* is the condition that the sensing matrix acts as a near isometry on all k -sparse signals. Candès and Tao showed that this condition is sufficient for sparse reconstruction and that random matrices, where the entries are generated by an iid Gaussian or Bernoulli process, satisfy the RIP with high probability. This approach treats all k -sparse signals equally likely, in contrast to mainstream signal processing where the filtering is deterministic, and the signal is described probabilistically. In the mainstream framework the sensing matrix is deterministic and it is required to act as a near-isometry on k -sparse vectors with high probability. This paper provides weak conditions that are sufficient to show that a deterministic sensing matrix satisfies this *Statistical Restricted Isometry Property (STRIP)*. The proof is elementary and avoids intricate combinatorial arguments involving coherence of orthonormal bases. The new framework encompasses many families of deterministic sensing matrices, including those formed from discrete chirps, Delsarte-Goethals codes, and Extended BCH codes. It is resilient to noise, and generalizes to k -compressible signals, where only k entries are significant, and the magnitude of all remaining entries is close to zero.

1. Introduction

The central goal of *compressed sensing* is to capture attributes of a signal using very few measurements. In most work to date, this broader objective is exemplified by the important special case in which a k -sparse vector \mathbf{x} in $\mathbb{R}^{\mathcal{C}}$ with \mathcal{C} large is to be reconstructed from a small number N of linear measurements with $k < N \ll \mathcal{C}$. In this problem, the measurement data is a vector $\mathbf{y} = \Phi\mathbf{x}$, where Φ is an $N \times \mathcal{C}$ matrix called the *sensing matrix*. The two fundamental questions are construction of suitable sensing matrices Φ and efficient reconstruction of \mathbf{x} from \mathbf{y} . Note that work of Kashin [Kas] and Gluskin [Glu] on approximation theory implies that $\Omega(k \log \frac{\mathcal{C}}{k})$ measurements are required for sparse reconstruction. Also of interest are the introduction of noise into the measurement process, and the generalization of the reconstruction problem to encompass k -compressible signals, where only k entries are significant, and the magnitude of all remaining entries is close to zero.

The work of Donoho [Don] and of Candès, Romberg and Tao [CT], [CRTb], [CRTa] provides fundamental insight into the geometry of sensing matrices. The *Restricted Isometry Property (RIP)* formulated by Candès and Tao is that the sensing matrix acts as a near isometry on all k -sparse vectors, and this condition is sufficient for sparse reconstruction. We will use RIP-1 to indicate isometry with respect to the ℓ_1 metric, which provides performance guarantees on sparse

reconstruction algorithms that are based on linear programming. When $\frac{N}{C}$ and/or $\frac{k}{N}$ are small, deterministic sensing matrices with the RIP property have been constructed using methods from approximation theory [DeV] and coding theory [Ind]. More attention has been paid to probabilistic constructions where the entries of the sensing matrix are generated by an i.i.d Gaussian or Bernoulli process. These sensing matrices are known to satisfy the RIP with high probability [Don], [CT] and the number of measurements N is $k \log\left(\frac{C}{k}\right)$. Constructions of random sensing matrices of similar size that have the RIP, but require a smaller degree of randomness are given by several approaches including filtering [BHR⁺], [TWD⁺] and expander graphs [GLR], [BGI⁺], [IR], [JXHC].

Random sensing matrices are easy to construct and achieve the RIP with high probability but suffer from two important drawbacks. First, efficiency in sampling comes at the cost of complexity in reconstruction as is shown in Table 1. Second, all k -sparse signals are treated as equally likely, in contrast to many of the most valuable approaches in sensor signal processing, which capitalize on prior probability distributions or other side information about where the signal of interest resides within the signal space. The strength of algorithms such as Basis Pursuit [CRTb] or Matching Pursuit [GSTV] is sparse reconstruction that is resilient to noise. However performance guarantees are predicated on the RIP and there is no known algorithm for verifying whether a given sensing matrix has this property. Performance analysis of sparse reconstruction for sensing matrices constructed using expander graphs is asymptotic, the constants are large, and the algorithms are mostly not straightforward to implement.

The role of random measurement in compressive sensing is analogous to the role of random coding in Shannon theory. Here reliable communication is achieved by deterministic codes with fast encoding and decoding algorithms that are designed to improve typical rather than worst case performance. Coding theory provides an approach to the design of deterministic sensing matrices [SBB], [AT]. The strength of this approach is fast algorithms for sparse reconstruction and for Reed Solomon constructions [AT], the roots of the reconstruction algorithm go back to 1795 and the work of de Prony on interpolation in the complex domain [dP], [Wol]. Reed Solomon reconstruction uses the input data to construct an error-locator polynomial and the roots of this polynomial identify the signals appearing in the sparse superposition. However the correspondence between the coefficients of a polynomial and its roots is not well conditioned, making it very difficult to deal with compressible signals and noisy measurements.

This paper introduces a method of constructing deterministic sensing matrices that are guaranteed to act as a near-isometry on k -sparse vectors with high probability, and this geometric property will be referred to as the *Statistical Restricted Isometry Property* (STRIP). We suppose that the columns of the sensing matrix form a group under pointwise multiplication, that all row sums vanish, and we require only a simple upper bound on the expected absolute value of the sum of the entries in a column of the sensing matrix. Our framework is very general and includes sensing matrices for which the columns are *discrete chirps* either in the standard Fourier domain [AHSC] or the Walsh-Hadamard domain [HCS]. Numerical results in these papers had suggested that chirp sensing matrices satisfied the STRIP property; and we provide a very simple proof that avoids reasoning about coherence of collections of mutually unbiased bases (cf. [GH]). A point of comparison with random Gaussian sensing matrices is that the eigenvalues of the chirp matrix are, on average, closer to one by more than a standard deviation. The chirp sensing matrices are representative of a much broader class where columns in the sensing matrix are obtained by exponentiating codewords in a linear code. Note that binary chirps are defined by quadratic functions over the binary field but can be viewed as a linear code over \mathbb{Z}_4 , the ring of integers modulo 4 [JKC⁺]. Note also that in the binary case, the column sums take the form $N - 2w$ where w is the Hamming weight of the exponentiated codeword, and that a similar interpretation is possible for codes that are linear over \mathbb{Z}_4 . Hence the upper bound connects the geometry of the code domain as captured by the weight enumerator of a code with the geometry of the complex domain.

Tables 1 and 2 summarize properties of different approaches to reconstruction of k -sparse signals of length C via randomized and deterministic matrices respectively. The proof that chirp codes satisfy the STRIP property is given at the end of Section 2. Detailed analysis of the nonlinear decoding algorithm will appear elsewhere, but in this case it is possible to strengthen the STRIP property. The reconstruction algorithm involves pointwise multiplication of the k -sparse superposition with a shift of itself, followed by the Fourier / Walsh Hadamard transform, and the STRIP property

Table 1. Properties of k -sparse reconstruction algorithms that employ random sensing matrices with N Rows and C Columns. Note that explicit construction of the expander graphs requires a large number of measurements, and more practical alternatives are random sparse matrices which are expanders with high probability.

Approach	Number of Measurements N	Complexity	Compressible Signals	Noise Resilience	RIP
Basis Pursuit (BP) [CRTb]	$k \log \left(\frac{C}{k} \right)$	C^3	Yes	Yes	Yes
Orthogonal Matching Pursuit (OMP) [GSTV]	$k \log^\alpha(C)$	$k^2 \log^\alpha(C)$	Yes	No	Yes
Group Testing [CM06]	$k \log^\alpha(C)$	$k \log^\alpha(C)$	Yes	No	No
Expanders (BP) [BGI ⁺]	$k \log \left(\frac{C}{k} \right)$	C^3	Yes	Yes	RIP-1
Expander Matching Pursuit(EMP) [IR]	$k \log \left(\frac{C}{k} \right)$	$C \log \left(\frac{C}{k} \right)$	Yes	Yes	RIP-1
CoSaMP [NT]	$k \log \left(\frac{C}{k} \right)$	$Ck \log \left(\frac{C}{k} \right)$	Yes	Yes	Yes
SSMP [DM]	$k \log \left(\frac{C}{k} \right)$	$Ck \log \left(\frac{C}{k} \right)$	Yes	Yes	Yes

Table 2. Properties of k sparse reconstruction algorithms that employ deterministic sensing matrices with N Rows and C Columns. Note that for LDPC codes $k \ll C$. Note also that RIP holds for random matrices where it implies existence of a low-distortion embedding from ℓ_2 into ℓ_1 . Guruswami et al. [GLR] proved that this property also holds for deterministic sensing matrices constructed from expander codes. It follows from Theorem 2.4 in this paper that sensing matrices based on discrete chirps and Delsarte-Goethals codes satisfy the STRIP.

Approach	Number of Measurements N	Complexity	Compressible Signals	Noise Resilience	RIP
Low Density Parity Check Codes (LDPC) [SBB]	$k \log C$	$C \log C$	Yes	Yes	No
Reed-Solomon codes [AT]	k	k^2	No	No	No
Embedding ℓ_2 into ℓ_1 (BP) [GLR]	$k(\log C)^{\alpha \log \log C}$	C^3	Yes	No	No
Extractors [Ind]	$kC^{o(1)}$	$kC^{o(1)} \log(C)$	No	No	No
Discrete chirps [AHSC]	\sqrt{C}	$kN \log N$	Yes	Yes	STRIP
Delsarte-Goethals codes [HCS]	$2^{\sqrt{\log C}}$	$kN \log^2 N$	Yes	Yes	STRIP

holds for all offsets and every Fourier / Walsh-Hadamard coefficient. Note that the number of measurements N is 2^m and that the number of codewords in the second order Reed Muller code is $2^{\frac{m(m-1)}{2}}$. This is exponentially large, but there is a chain of Delsarte-Goethals codes, each linear over the ring of integers modulo 4, that makes it possible to match the dimension of the input data to the size of the code. It is of course always possible to randomly pad input data with zeros to match the number of columns in the sensing matrix.

2. The STRIP and Simple Sufficient Conditions on Deterministic Sensing Matrices

Let Φ be a deterministic sensing matrix with N rows and C columns, and let $\varphi^i(x)$ denote the entry in row x and column i . We make two simple assumptions :

- 1) the columns of Φ form a group U_C under pointwise multiplication,
- 2) the rows of Φ are orthogonal, and all row sums are equal to zero.

It follows from condition (1) that all entries of Φ are unimodular, i.e for any row x and column i , $|\varphi^i(x)| = 1$. Also, it follows from conditions (1) and (2) that the normalized columns $\frac{1}{\sqrt{N}}\varphi$ form a tight frame with redundancy $\frac{\mathcal{C}}{N}$, that is $\Phi\Phi^\dagger = \mathcal{C}I_{N \times N}$. Since, if $\Phi\Phi^\dagger = \mathcal{C}I_{N \times N}$, then

$$\sum_{j=1}^{\mathcal{C}} \varphi^j(i) \overline{\varphi^j(k)} = \mathcal{C} \delta_{ik}.$$

Hence for any vector v

$$\sum_j |\langle v | \varphi^j \rangle|^2 = \sum_j \langle v | \varphi^j \rangle \langle \varphi^j | v \rangle = \langle v | \left(\sum_j \varphi^j \right) \langle \varphi^j | v \rangle = \mathcal{C} \langle v | v \rangle = \mathcal{C} \|v\|^2.$$

As a result, if the normalized columns $\frac{1}{\sqrt{N}}\varphi$ form a tight frame with redundancy $\frac{\mathcal{C}}{N}$ then

$$\begin{aligned} E_{\varphi \in U_{\mathcal{C}}} \left[\left| \sum_x \varphi(x) \right|^2 \right] &= \sum_{x, x'} \frac{1}{\mathcal{C}} \sum_{\varphi \in U_{\mathcal{C}}} \varphi(x) \overline{\varphi(x')} \\ &= \sum_{x, x'} \delta(x - x') = N. \end{aligned}$$

We now show that the sensing matrix Φ preserves the norm of any k -sparse input signal α to within a small fraction. This is the Statistical Restricted Isometry Property or STRIP.

Our model for the signal is that the positions of the entries are chosen randomly, and the values of the entries are chosen adversarially or arbitrarily. Let $\pi = (\pi_j)$ be a random permutation of the columns of Φ . Setting $f(x) = \frac{1}{\sqrt{N}} \sum_{j=1}^k \alpha_j \varphi^{\pi_j}(x)$ we calculate the expected value and variance of $\|f\|^2$.

We have

$$\|f\|^2 = \sum_{x=1}^N |f(x)|^2 = \frac{1}{N} \sum_x \left(\sum_{t=1}^k |\alpha_t|^2 + \Psi(x) \right) \quad (1)$$

where $\Psi(x) = \sum_{j \neq i} \alpha_j \overline{\alpha_i} \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)}$.

Note that the expectation $E[\|f\|^2]$ is over all admissible choices of columns φ^{π_j} based on the random permutation π . The first term in (1) is independent of the choice of columns and is just $\sum_{j=1}^k |\alpha_j|^2$; The following lemma shows that the second term is bounded in absolute value by $\frac{k}{\mathcal{C}} \|\alpha\|^2$.

Lemma 2.1. *Let π be a random permutation of $\{1, \dots, \mathcal{C}\}$, Φ be the matrix satisfying the two conditions given above, and let α be a k -sparse signal in $\mathbb{R}^{\mathcal{C}}$ such that the positions of the k non-zero entries are chosen according to the random permutation π . Then*

$$-\frac{N(k-1)}{\mathcal{C}-1} \|\alpha\|^2 \leq E_{\pi} \left[\sum_x \sum_{j \neq i} \alpha_j \overline{\alpha_i} \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)} \right] \leq \frac{N}{\mathcal{C}-1} \|\alpha\|^2,$$

where the expectation is taken over all possible random permutations π .

Proof. Since π is a random permutation, the choice of coefficients α_j is independent of the choice of columns φ^{π_j} so by linearity of expectation,

$$E_{\pi} \left[\sum_x \sum_{j \neq i} \alpha_j \overline{\alpha_i} \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)} \right] = \sum_{j \neq i} \alpha_j \overline{\alpha_i} E_{\pi} \left[\sum_x \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)} \right].$$

Hence, we should calculate

$$E_{\pi} \left[\sum_x \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)} \right]. \quad (2)$$

Since the columns of the matrix form a group under pointwise multiplication, equation (2) can be rewritten as

$$E_{j' \sim \pi} \left[\sum_x \varphi^{\pi_{j'}}(x) \right]. \quad (3)$$

We can simply use the method of double counting for calculating equation (3). Since the row sums of Φ are all equal to zero, the sum of all entries of Φ is zero, hence the average of the column sums over all except the identity column is $\frac{-N}{\mathcal{C}-1}$. Hence,

$$E_{\pi} \left[\sum_x \sum_{j \neq i} \alpha_j \bar{\alpha}_i \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)} \right] = \frac{-N}{\mathcal{C}-1} \sum_{j \neq i} \alpha_j \bar{\alpha}_i.$$

Applying the Cauchy-Schwartz inequality, we obtain

$$0 \leq \sum_{\substack{j,i=1 \\ j \neq i}}^k \alpha_j \bar{\alpha}_i + \sum_{j=1}^k |\alpha_j|^2 = \left| \sum_{j=1}^k \alpha_j \right|^2 \leq k \sum_{j=1}^k |\alpha_j|^2,$$

which immediately implies

$$-\frac{k-1}{\mathcal{C}-1} \|\alpha\|^2 \leq E_{\pi} \left[\frac{1}{N} \sum_x \sum_{j \neq i} \alpha_j \bar{\alpha}_i \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)} \right] \leq \frac{1}{\mathcal{C}-1} \|\alpha\|^2. \quad (4)$$

□

The following Theorem is a direct consequence of Lemma 2.1.

Theorem 2.2. *Let π be a random permutation of $\{1, \dots, \mathcal{C}\}$, Φ be the matrix satisfying the two conditions given above, and let α be a k -sparse signal in $\mathbb{R}^{\mathcal{C}}$ such that the positions of the k non-zero entries are chosen according to the random permutation π . Then*

$$\left(1 - \frac{k-1}{\mathcal{C}-1}\right) \|\alpha\|^2 \leq E_{\pi} [\|\mathbf{f}\|^2] \leq \left(1 + \frac{1}{\mathcal{C}-1}\right) \|\alpha\|^2.$$

Now we compute the variance of $\|\mathbf{f}\|^2$. We have

$$\begin{aligned} \|\mathbf{f}\|^4 &= \sum_{x, x'} |f(x)|^2 |f(x')|^2 \\ &= \sum_{x, x'} f(x) \overline{f(x)} f(x') \overline{f(x')} \\ &= \frac{1}{N^2} \sum_{x, x'} \left(\sum_{t=1}^k |\alpha_t|^2 + \Psi(x) \right) \left(\sum_{t=1}^k |\alpha_t|^2 + \Psi(x') \right) \end{aligned} \quad (5)$$

where $\Psi(x) = \sum_{j \neq i} \alpha_j \bar{\alpha}_i \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)}$.

In order to prove the statistical restricted isometry property (STRIP), we need to bound the expectation $E[\|\mathbf{f}\|^4]$ taken over all admissible choices of columns φ^{π_j} . The first term in (5) is independent of the choice of columns and is just $\left(\sum_{j=1}^k |\alpha_j|^2\right)^2$. Hence the remaining terms constitute the variance $V[\|\mathbf{f}\|^2]$.

The second term in (5) is given by

$$\frac{2}{N^2} \left(\sum_{j=1}^k |\alpha_j|^2 \right) \sum_{j \neq i} \alpha_j \bar{\alpha}_i \sum_x \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)}$$

The choice of coefficients α_j is independent of the choice of columns φ^{P_j, b_j} so by linearity of expectation, we should calculate

$$E_{j \neq i} \left[\sum_x \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}} \right]. \quad (6)$$

Let \mathcal{C} be the number of column vectors in Φ . Then we can rewrite (6) as

$$\frac{1}{\mathcal{C}(\mathcal{C}-1)} \left[\mathbf{1}^T \left(\sum_{\substack{g, h \in \mathcal{G} \\ g \neq h}} gh^{-1} \right) \right] \quad (7)$$

where $\mathbf{1}^T$ is the row vector of length 2^m with entries indexed by index binary m -tuples \underline{x} and every entry equal to 1. The initial factor is just the frequency with which any admissible pair (g, h) is chosen. The next Lemma employs elementary group theory to bound this summation.

Lemma 2.3. *The map $\mathcal{G} \times \mathcal{G} \rightarrow \mathcal{G}$ given by $(g, h) \rightarrow gh^{-1}$ is a homomorphism and*

$$\sum_{\substack{g, h \in \mathcal{G} \\ g \neq h}} gh^{-1} = -\mathcal{C}\mathbf{1} \quad (8)$$

Proof. Every element of \mathcal{G} except the identity appears exactly \mathcal{C} times in the left hand sum. Since all row sums of the sensing matrix vanish, the sum of the all element of the group is zero. This means that sum of all except the identity element should be $-\mathbf{1}$ which completes the proof. \square

It follows from the Lemma that

$$\left| \frac{1}{\mathcal{C}(\mathcal{C}-1)} [\mathbf{1}^T \left(\sum_{\substack{g, h \in \mathcal{G} \\ g \neq h}} gh^{-1} \right)] \right| \leq \frac{N}{\mathcal{C}-1}, \quad (9)$$

and hence

$$E_\pi \left[\left| \frac{2}{N^2} \left(\sum_{j=1}^k |\alpha_j|^2 \right) \sum_{j \neq i} \alpha_j \overline{\alpha_i} \sum_x \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}}(x) \right| \right] \leq \frac{2(k-1)}{N(\mathcal{C}-1)} \|\boldsymbol{\alpha}\|^4.$$

The third term in (5) is given by

$$\frac{1}{N^2} \sum_{\substack{x, x' \\ s \neq t}} \sum_{\substack{j \neq i}} \alpha_j \overline{\alpha_i} \alpha_s \overline{\alpha_t} \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}}(x) \varphi^{\pi_s}(x') \overline{\varphi^{\pi_t}}(x')$$

and again we calculate the expectation over choices of admissible columns. There are several cases.

Case 1: The indices j, i, s, t are all distinct.

We calculate

$$E_{j \neq i \neq s \neq t} \left[\sum_{x, x'} \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}}(x) \varphi^{\pi_s}(x') \overline{\varphi^{\pi_t}}(x') \right]$$

by rewriting as

$$\frac{1}{\mathcal{C}(\mathcal{C}-1)(\mathcal{C}-2)(\mathcal{C}-3)} \left[\mathbf{1}^T \left(\sum_{\substack{g, h \in \mathcal{G} \\ g \neq h}} gh^{-1} \right) \left(\sum_{\substack{v, w \in \mathcal{G} \\ v \neq w, g, h \\ w \neq g, h}} (vw^{-1})^T \mathbf{1} \right) \right] \quad (10)$$

Now since

$$\sum_{\substack{v,w \\ v \neq w}} vw^{-1} = -\mathcal{C}\mathbf{1}, \quad (11)$$

removing all terms $(g, *)$, $(h, *)$, $(*, g^{-1})$, $(*, h^{-1})$ from ((11)) and adding back in the terms (g, h^{-1}) , (h, g^{-1}) that have been removed twice, we count each element other than gh^{-1} and hg^{-1} exactly $\mathcal{C} - 4$ times and obtain

$$\sum_{\substack{v,w \\ v \neq w, g, h \\ w \neq g, h}} vw^{-1} = -(\mathcal{C} - 4)\mathbf{1} + gh^{-1} + hg^{-1}.$$

Hence Equation. (10) becomes

$$\frac{1}{\mathcal{C}(\mathcal{C} - 1)(\mathcal{C} - 2)(\mathcal{C} - 3)} \sum_{\substack{g,h \\ g \neq h}} \mathbf{1}^T (gh^{-1}) (-(\mathcal{C} - 4)\mathbf{1} + gh^{-1} + hg^{-1})^T \mathbf{1}. \quad (12)$$

Observe that if $S_{g,h} = \sum_x g(x)h^{-1}(x)$ then we can write $S_{g,h} = \mathbf{1}^T (gh^{-1})$ then

$$S_{g,h}^2 = \mathbf{1}^T (gh^{-1})(gh^{-1})^T \mathbf{1}$$

and

$$|S_{g,h}|^2 = \mathbf{1}^T (gh^{-1})(hg^{-1})^T \mathbf{1}$$

and it is obvious that $|S_{g,h}|^2 = |S_{g,h}^2|$. So equation (12) can be written as

$$\frac{1}{\mathcal{C}(\mathcal{C} - 1)(\mathcal{C} - 2)(\mathcal{C} - 3)} \left[N\mathcal{C}(\mathcal{C} - 4) + \sum_{\substack{g,h \\ g \neq h}} (S_{g,h}^2 + |S_{g,h}|^2) \right]. \quad (13)$$

We can apply the hypothesis on column sums and obtain

$$E_{g,h \sim U_{\mathcal{C}}} \left[|S_{g,h}|^2 \right] = E_{f \sim U_{\mathcal{C}}} \left[|\mathbf{1}^T f|^2 \right] = E_{f \sim U_{\mathcal{C}}} \left[\left| \sum_x \varphi^f(x) \right|^2 \right] \leq N.$$

Hence (13) is bounded above in absolute value by

$$\frac{N\mathcal{C}(\mathcal{C} - 4) + 2\mathcal{C}(\mathcal{C} - 1)N}{\mathcal{C}(\mathcal{C} - 1)(\mathcal{C} - 2)(\mathcal{C} - 3)}.$$

As a result

$$E_{\pi} \left[\left| \frac{1}{N^2} \sum_{x,x'} \sum_{\substack{j \neq i \\ \neq s \neq t}} \alpha_j \bar{\alpha}_i \alpha_s \bar{\alpha}_t \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)} \varphi^{\pi_s}(x') \overline{\varphi^{\pi_t}(x')} \right| \right] \leq \frac{3(k-1)^2}{N(\mathcal{C}-2)(\mathcal{C}-3)} \|\alpha\|^4.$$

Case 2: The indices j, i, s, t take on 3 distinct values.

There are two subcases. The first subcase is

$$E_{j,i,t \text{ distinct}} \left[\left(\sum_x \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)} \right) \left(\sum_{x'} \varphi^{\pi_i}(x') \overline{\varphi^{\pi_t}(x')} \right) \right] \quad (14)$$

which we rewrite as

$$\frac{1}{\mathcal{C}(\mathcal{C} - 1)(\mathcal{C} - 2)} \sum_{h \in \mathcal{G}} \mathbf{1}^T \left(\sum_{g \neq h} gh^{-1} \right) \left(\sum_{w \neq g, h} hw^{-1} \right)^T \mathbf{1}. \quad (15)$$

Now

$$\sum_{w \neq g, h} hw^{-1} = -(\mathbf{1} + hg^{-1})$$

and so (15) is bounded above in absolute value by

$$\begin{aligned} & \frac{1}{\mathcal{C}(\mathcal{C}-1)(\mathcal{C}-2)} \left[\mathbf{1}^T \left(\sum_{\substack{g,h \\ g \neq h}} gh^{-1} \right) \mathbf{1}^T \mathbf{1} + \sum_{\substack{g,h \\ g \neq h}} |\mathbf{1}^T (gh^{-1})(hg^{-1})^T \mathbf{1}| \right] \\ & \leq \frac{\mathcal{C}N^2 + \mathcal{C}(\mathcal{C}-1)N}{\mathcal{C}(\mathcal{C}-1)(\mathcal{C}-2)}, \end{aligned}$$

which gives

$$E_\pi \left[\frac{1}{N^2} \sum_{x,x'} \sum_{j \neq i \neq t} \alpha_j \overline{\alpha_i} \alpha_i \overline{\alpha_t} \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)} \varphi^{\pi_i}(x') \overline{\varphi^{\pi_t}(x')} \right] \leq \frac{2(k-1)}{N(\mathcal{C}-2)} \|\boldsymbol{\alpha}\|^4.$$

The analysis of the second subcase

$$E_{j,i,t \text{ distinct}} \left[\left(\sum_x \varphi^{\pi_i}(x) \overline{\varphi^{\pi_j}(x)} \right) \left(\sum_{x'} \varphi^{\pi_i}(x') \overline{\varphi^{\pi_t}(x')} \right) \right]$$

is very similar. All that changes is that the terms gh^{-1} are replaced by hg^{-1} , so that the terms $|\mathbf{1}(gh^{-1})(hg^{-1})^T \mathbf{1}|$ are replaced by $|\mathbf{1}(hg^{-1})(hg^{-1})^T \mathbf{1}|$. But since $|S_{g,h}|^2 = |S_{g,h}^2|$ the upper bound on absolute value stays the same.

Case 3 The indices j, i, s, t take 2 distinct values.

We calculate

$$E_{j \neq i} \left[\left(\sum_x \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)} \right) \left(\sum_{x'} \varphi^{\pi_i}(x') \overline{\varphi^{\pi_j}(x')} \right) \right] \quad (16)$$

which we rewrite as

$$\frac{1}{\mathcal{C}(\mathcal{C}-1)} \sum_{\substack{g,h \\ g \neq h}} \mathbf{1}^T (hg^{-1})(gh^{-1})^T \mathbf{1}.$$

This quantity is bounded above in absolute value by

$$\frac{\mathcal{C}(\mathcal{C}-1)}{\mathcal{C}(\mathcal{C}-1)} N. \quad (17)$$

Note that

$$\sum_{\substack{j,i=1 \\ j \neq i}}^k |\alpha_j|^2 |\alpha_i|^2 = \left(\sum_{j=1}^k |\alpha_j|^2 \right)^2 - \sum_{j=1}^k |\alpha_j|^4 \leq (k-1) \sum_{j=1}^k |\alpha_j|^4 \leq (k-1) \|\boldsymbol{\alpha}\|^4.$$

As a result

$$E_\pi \left[\frac{1}{N^2} \sum_{x,x'} \sum_{j \neq i} \alpha_j \overline{\alpha_i} \alpha_i \overline{\alpha_j} \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)} \varphi^{\pi_i}(x') \overline{\varphi^{\pi_j}(x')} \right] \leq \frac{(k-1)}{N} \|\boldsymbol{\alpha}\|^4. \quad (18)$$

The second subcase is

$$E_{j \neq i} \left[\left(\sum_x \varphi^{\pi_j}(x) \overline{\varphi^{\pi_i}(x)} \right) \left(\sum_{x'} \varphi^{\pi_j}(x') \overline{\varphi^{\pi_i}(x')} \right) \right]$$

which we rewrite as

$$\frac{1}{\mathcal{C}(\mathcal{C}-1)} \sum_{\substack{g,h \\ g \neq h}} \mathbf{1}^T (hg^{-1})(hg^{-1})^T \mathbf{1}. \quad (19)$$

and since

$$\sum_{\substack{j,i=1 \\ j \neq i}} \alpha_j^2 \bar{\alpha}_i^2 = \left| \sum_{j=1}^k \alpha_j^2 \right|^2 - \sum_{j=1}^k |\alpha_j|^4 \leq (k-1) \sum_{j=1}^k |\alpha_j|^4 \leq (k-1) \|\alpha\|^4,$$

the upper bound (18) stays the same. Now we are ready to prove the following theorem.

Theorem 2.4 (Statistical restricted isometry property (STRIP)). *Let Φ be a deterministic $N \times \mathcal{C}$ sensing matrix such that*

- 1) *The columns of Φ form a group $U_{\mathcal{C}}$ under pointwise multiplication.*
- 2) *the rows of Φ are orthogonal, and all row sums are equal to zero.*

Let α be a k -sparse signal where the positions of the k non-zero entries are equiprobable. Given a δ with $1 > \delta > (k-1)/(\mathcal{C}-1)$, then with probability $1 - [2^k/N + (2k+7)/(\mathcal{C}-3)]/[\delta - (k-1)/(\mathcal{C}-1)]^2$

$$(1 - \delta) \|\alpha\|^2 \leq \|\Phi \alpha\|^2 \leq (1 + \delta) \|\alpha\|^2.$$

Proof. Let $y = \Phi \alpha$. By Theorem 2.2 we have

$$\left(1 - \frac{k-1}{\mathcal{C}-1}\right) \|\alpha\|^2 \leq E[|y|^2] \leq \left(1 + \frac{1}{\mathcal{C}-1}\right) \|\alpha\|^2.$$

We have also shown that

$$\begin{aligned} E[|y|^4] - \|\alpha\|^4 &\leq \frac{2(k-1)}{N(\mathcal{C}-1)} \|\alpha\|^4 + \frac{3(k-1)^2}{N(\mathcal{C}-2)(\mathcal{C}-3)} \|\alpha\|^4 + \frac{4(k-1)}{N(\mathcal{C}-2)} \|\alpha\|^4 + \frac{2k}{N} \|\alpha\|^4 \\ &\leq \left(\frac{2k}{N} + \frac{9}{\mathcal{C}-3}\right) \|\alpha\|^4, \end{aligned}$$

which implies

$$\begin{aligned} \text{Var}[|y|^2] &\leq \|\alpha\|^4 + \left(\frac{2k}{N} + \frac{9}{\mathcal{C}-3}\right) \|\alpha\|^4 - \|\alpha\|^4 + \frac{2(k-1)}{\mathcal{C}-1} \|\alpha\|^4 - \left(\frac{k-1}{\mathcal{C}-1}\right)^2 \|\alpha\|^4 \\ &\leq \left(\frac{2k}{N} + \frac{2k+7}{\mathcal{C}-3}\right) \|\alpha\|^4. \end{aligned} \quad (20)$$

Now, using Chebyshev's inequality we have

$$\begin{aligned} \Pr[||y|^2 - \|\alpha\|^2| \geq \delta \|\alpha\|^2] &\leq \Pr\left[| |y|^2 - E[|y|^2] | \geq \left(\delta - \frac{k-1}{\mathcal{C}-1}\right) \|\alpha\|^2\right] \\ &\leq \frac{\left(\frac{2k}{N} + \frac{2k+7}{\mathcal{C}-3}\right) \|\alpha\|^4}{\left(\delta - \frac{k-1}{\mathcal{C}-1}\right)^2 \|\alpha\|^4}. \end{aligned}$$

If $k \leq c_1 N / (\log \mathcal{C} / N + 1)$ and $\mathcal{C} \geq c_2 N^\beta$ for some $c_1, c_2 > 0$ and $\beta > 1$, then

$$\begin{aligned} \Pr[||y|^2 - \|\alpha\|^2| \geq \delta \|\alpha\|^2] &\leq \frac{\frac{2c_1}{\kappa+1} + \frac{2c_1 N - 7\kappa + 7}{(\kappa+1)(c_2 N^\beta - 3)}}{\left(\delta - \frac{2c_1}{\kappa+1}\right)^2} \\ &\sim \frac{2c_1}{\delta^2 (\beta - 1) \log N} \quad \text{as } N \rightarrow \infty, \end{aligned}$$

where $\kappa = (\beta - 1)(\log N + \log c_2)$.

□

Remark 1 If the tight frame condition of Theorem (2.4) is weakened to

$$E_{\varphi \in U_{\mathcal{C}}} \left[\left| \sum_x \varphi(x) \right|^2 \right] \leq N^{2\epsilon} \text{ for some } 0 < \epsilon < 1,$$

then the proof given above leads to the success probability $1 - \frac{\left(\frac{2k}{N^2(1-\epsilon)} + \frac{2k+7}{C-3}\right)}{\left(\delta - \frac{k-1}{C-1}\right)^2}$.

Remark 2 It is shown in [BDDW07] that with high probability, $N \times C$ compressive sensing matrices with entries that are independent random Bernoulli and Gaussian random variables obtain k -RIP for $k \leq c_1 n / (\log C / N + 1)$, for some $c_1 > 0$. Taking k as such and further assuming that $C \geq c_2 N^\beta$, for some $c_2 > 0$ and $\beta > 1$, in theorem 2.4 we proved that

$$\Pr \left[\left| \|y\|^2 - \|\alpha\|^2 \right| \geq \delta \|\alpha\|^2 \right] \sim \frac{2c_1}{\delta^2(\beta-1) \log N} \quad \text{as } N \rightarrow \infty,$$

We have found in practice that for compressive sensing matrices with columns selected from second order Reed-Muller functions $\Pr \left[\left| \|y\|^2 - \|\alpha\|^2 \right| \geq \delta \|\alpha\|^2 \right]$ has much faster decay with N than suggested by this bound.

3. Resilience to Noise

3.1. Noisy Measurements

In this Section, we consider deterministic sensing matrices satisfying the hypothesis of Theorem 2.4, and show resilience to independent identically distributed (iid) Gaussian noise that is uncorrelated with the measured signal.

Theorem 3.1. *Let Φ and α be such that*

$$(1 - \delta)^2 \|\alpha\|^2 \leq \|\Phi\alpha\|^2 \leq (1 + \delta)^2 \|\alpha\|^2. \quad (21)$$

for some $0 < \delta \leq 1$ and let $f = \Phi\alpha + \nu$, where the noise samples $\nu(x)$ are iid complex Gaussian random variables with zero mean and variance $2\sigma^2$. Then, for $\epsilon \geq 0$,

$$(1 - \delta - \epsilon)^2 \|\alpha\|^2 \leq \|f\|^2 \leq (1 + \delta + \epsilon)^2 \|\alpha\|^2, \quad (22)$$

with probability greater than

$$1 - \frac{2S^N}{(N-1)!} \left(\int_\epsilon^{1-\delta} e^{-Sy^2} (1 - \delta - y)^{2N-1} dy + \int_\epsilon^\infty e^{-Sy^2} (1 + \delta + y)^{2N-1} dy \right),$$

where $S = \|\alpha\|^2 / 2\sigma^2$.

Note we have introduced the square of $(1 \pm \delta)$ in (21) which is inconsequential for the bound but leads to a neater result.

Proof: First consider the probability that $\|f\|$ exceeds the upper bound in (22). Writing $B_u = (1 + \delta)\|\alpha\|$ and $g = \Phi\alpha$, this is

$$\begin{aligned} \Pr(\|f\| \geq B_u + \epsilon\|\alpha\|) &= \frac{1}{(2\pi\sigma^2)^N} \int_{\|f\| \geq B_u + \epsilon\|\alpha\|} \exp\left(-\frac{1}{2\sigma^2} \|f - g\|^2\right) df \\ &< \frac{1}{(2\pi\sigma^2)^N} \int_{\|f\| \geq B_u + \epsilon\|\alpha\|} \exp\left(-\frac{1}{2\sigma^2} (\|f\| - \|g\|)^2\right) df \\ &\leq \frac{1}{(2\pi\sigma^2)^N} \int_{\|f\| \geq B_u + \epsilon\|\alpha\|} \exp\left(-\frac{1}{2\sigma^2} (\|f\| - B_u)^2\right) df. \end{aligned}$$

Then writing $f = (B_u + y\|\alpha\|)\hat{n}$, for $y \geq 0$, and carrying out the resulting surface integration over unit vectors \hat{n} , we obtain

$$\Pr(\|f\| \geq B_u + \epsilon\|\alpha\|) < \frac{\Omega_{2N-1} \|\alpha\|^{2N}}{(2\pi\sigma^2)^N} \int_\epsilon^\infty \exp\left(-\frac{\|\alpha\|^2}{2\sigma^2} y^2\right) (1 + \delta + y)^{2N-1} dy,$$

where $\Omega_{2N-1} = 2\pi^N / (N-1)!$ is the surface area of the unit $(2N-1)$ -sphere. In a similiary way, with $B_\ell = (1 - \delta)\|\alpha\|$,

$$\begin{aligned} \Pr(\|f\| \leq B_\ell - \epsilon\|\alpha\|) &< \frac{1}{(2\pi\sigma^2)^N} \int_{\|f\| \leq B_\ell - \epsilon\|\alpha\|} \exp\left(-\frac{1}{2\sigma^2} (\|f\| - B_\ell)^2\right) df \\ &= \frac{\Omega_{2N-1} \|\alpha\|^{2N}}{(2\pi\sigma^2)^N} \int_\epsilon^{1-\delta} \exp\left(-\frac{\|\alpha\|^2}{2\sigma^2} y^2\right) (1 - \delta - y)^{2N-1} dy, \end{aligned}$$

where we have made the change of variable $f = (B_\ell - y\|\alpha\|)\hat{n}$, for $0 \leq y \leq 1 - \delta$.

□

3.2. Noisy Signals

If the signal α is contaminated by noise then the measurements are given by

$$\mathbf{y} = \Phi\alpha + \Phi\boldsymbol{\mu}, \quad (23)$$

where $\boldsymbol{\mu}$ is complex multivariate Gaussian distributed, with zero mean and covariance

$$E(\boldsymbol{\mu}\boldsymbol{\mu}^\dagger) = 2\sigma^2 I_{\mathcal{C} \times \mathcal{C}}. \quad (24)$$

The reconstruction algorithm needs to recover the signal from the noisy measurements

$$\mathbf{y} = \mathbf{f} + \boldsymbol{\nu}, \quad (25)$$

where $\boldsymbol{\nu} = \Phi\boldsymbol{\mu}$ is complex multivariate Gaussian distributed with mean zero and covariance

$$E(\boldsymbol{\nu}\boldsymbol{\nu}^\dagger) = 2\sigma^2 \Phi\Phi^\dagger. \quad (26)$$

The deterministic compressive sensing schemes considered in this paper have some advantage over random compressive sensing schemes in that $\Phi\Phi^\dagger = \frac{\mathcal{C}}{N} I_{N \times N}$ and consequently $E(\nu(x)\nu(x')) = \frac{2\sigma^2\mathcal{C}}{N} \delta(x - x')$, i.e., the noise samples on distinct measurements are independent.

If we had unlimited computational power there would be little to distinguish compressive sensing schemes which satisfy RIP or STRIP in terms of statistical performance of reconstruction. However computational power is always limited and consequently one would like to develop reconstruction algorithms which are both fast and statistically robust. The advantage of the deterministic compressive sensing schemes described here is that \mathbf{f} is a linear combination of functions which have a great deal of structure and this can be exploited in developing such algorithms (for examples see [HCS], [AHSC]). Random schemes, by their very nature, do not have this property.

4. Families of Deterministic Sensing Matrices

The hypotheses of Theorem 2.4 are relatively weak and there are many families of deterministic sensing matrices that satisfy the STRIP.

Discrete Chirps: Let k be a prime and let ω be a primitive (complex) k^{th} root of unity. A length k chirp signal takes the form

$$\varphi^{m,r}(x) = \omega^r \omega^{mx+rx^2} \quad \text{where } x = 0, 1, \dots, k-1$$

where m is the base frequency and r is the chirp rate. It is clear that the chirp signals form the columns of a sensing matrix that satisfies the hypotheses of Theorem 2.4. Note that the purpose of the initial phase is to make all row sums vanish. Applebaum et al. [AHSC] have analyzed an algorithm for sparse reconstruction that exploits the efficiency of the FFT in each of two steps: the first to recover the chirp rates and second to recover the chirp frequency. Their paper uses the Gershgorin Circle Theorem to prove that the RIP holds for sets of \sqrt{k} columns.

Kerdock, Delsarte-Goethals and Second Order Reed-Muller Codes: These codes were originally constructed as nonlinear binary codes defined by collections of quadratic forms (see [Ker], [DG], also [MS], chapter 15). Hammons et al. [JKC⁺] then showed that each of these codes arises as the binary image under the Gray map of a linear code over \mathbb{Z}_4 , the ring of integers modulo 4. Codewords are determined by triples $(\mathbf{P}, \mathbf{b}, \epsilon)$ where \mathbf{P} is an $m \times m$ binary symmetric matrix, \mathbf{b} is a binary m -tuple, and $\epsilon = 0, 1, 2$ or 3 . The entries of the codeword $c(\mathbf{P}, \mathbf{b}, \epsilon)$ are indexed by binary m -tuples \mathbf{x} according to the rule

$$c(\mathbf{P}, \mathbf{b}, \epsilon) = \mathbf{x}\mathbf{P}\mathbf{x}^T + 2\mathbf{b}\mathbf{x}^T + \epsilon \quad (27)$$

where all arithmetic takes place in \mathbb{Z}_4 , the ring of integers modulo 4. There are $(m+1)/2$ Delsarte-Goethals codes $DG(m+1, r)$ of length $N = 2^m$ over \mathbb{Z}_4 . The Kerdock code is $DG(m+1, 0)$, the second order Reed Muller code is $DG(m+1, \frac{m-1}{2})$, and the codes are all nested, so that $DG(m+1, r)$ is contained in $DG(m+1, r+1)$ for $r = 0, 1, \dots, \frac{m+1}{2}$. The Delsarte-Goethals codes determine nested sets of binary symmetric matrices $(m+1, r)$, each closed under addition modulo 2, with the property that for any distinct pair of matrices \mathbf{P}, \mathbf{Q} in $(m+1, r)$, the rank

of the binary sum $P + Q$ is at least $m - 2r$. There are $2^{2m+2+mr}$ codewords in $DG(m+1, r)$ determined in part by the $2^{m(r+1)}$ matrices in $\Omega(m+1, r)$, and since the first $2r+1$ rows of any pair of binary symmetric matrices in $\Omega(m+1, r)$ are distinct, this is best possible. Sets of matrices that are extremal with respect to rank distance appear in early work of Gabidulin [Gab] and more recently in the construction of space-time codes for wireless communication (see [LK] and references therein) where they provide a mechanism for translating constraints in the binary domain into lower bounds on diversity protection in the complex domain.

Two columns of a sensing matrix should not differ by a scalar factor, but the constant $\epsilon = 0, 1, 2,$ or 3 appearing in (27) produces sets of four exponentiated codewords where any pair differ by a 4th root of unity. Therefore we define a collection of exponentiated codewords $\varphi^{P,b}$, with exactly one representative from each set as follows:

$$\varphi^{P,b}(\mathbf{x}) = i^{wt(d_P)+2wt(\mathbf{b})} i^{\mathbf{x}P\mathbf{x}^T+2\mathbf{b}\mathbf{x}^T}, \text{ where } \mathbf{x} \in \mathbb{Z}_2^m$$

The index P ranges over binary symmetric matrices in the Delsarte-Goethals set $\Omega(m+1, r)$ and d_P denotes the main diagonal of P . If we fix a matrix P , then the 2^m exponentiated codewords $\varphi^{P,b}$ that are obtained by varying \mathbf{b} form an orthonormal basis Γ_P . This basis is obtained by postmultiplying the Walsh-Hadamard basis by the diagonal matrix with entry $i^{\mathbf{x}P\mathbf{x}}$ in position \mathbf{x} . These bases are in one to one correspondence with cosets of the first order Reed-Muller code in $DG(m+1, r)$. Coherence between bases Γ_P and Γ_Q indexed by the matrices P and Q respectively depends on the rank R of the binary difference $P + Q$ (see Lemmas 4.2 and 4.3 below and [CCKS] for an alternative proof that uses properties of the symplectic group $\text{Sp}(m, 2)$). Any vector in one of the orthonormal bases has inner product $2^{-R/2}$ with 2^R vectors in the other basis and is orthogonal to the remaining vectors in that basis. The next lemma is well known, but the proof is non-standard and makes a useful connection to exponential sums.

Lemma 4.1. *The diagonal d_P of a binary symmetric matrix P is contained in the row space of P .*

Proof. We suppose that there is no solution to the equation $\mathbf{z}P = d_P$ and show that the vector $(i^{\mathbf{x}P\mathbf{x}^T})$ is orthogonal to every vector in the Walsh-Hadamard orthonormal basis. Set

$$S = \left((i^{\mathbf{x}P\mathbf{x}^T}), (i^{2\mathbf{b}\mathbf{x}^T}) \right) = \sum_{\mathbf{x}} i^{\mathbf{x}P\mathbf{x}^T+2\mathbf{b}\mathbf{x}^T},$$

then

$$\begin{aligned} S^2 &= \sum_{\mathbf{x}, \mathbf{y}} i^{\mathbf{x}P\mathbf{x}^T+\mathbf{y}P\mathbf{y}^T+2\mathbf{b}(\mathbf{x}+\mathbf{y})^T} \\ &= \sum_{\mathbf{x}, \mathbf{y}} i^{(\mathbf{x}+\mathbf{y})P(\mathbf{x}+\mathbf{y})^T+\mathbf{x}P\mathbf{y}^T+2\mathbf{b}(\mathbf{x}+\mathbf{y})^T}. \end{aligned}$$

Changing variables to $\mathbf{z} = \mathbf{x} \oplus \mathbf{y}$ and \mathbf{y} gives

$$S^2 = \sum_{\mathbf{z}} i^{\mathbf{z}P\mathbf{z}^T} \sum_{\mathbf{y}} (-1)^{(d_P+\mathbf{z}P)\mathbf{y}^T}$$

and since there is no solution for $\mathbf{z}P = d_P$ we have $S = 0$. □

Lemma 4.2. *If $S = \sum_{\mathbf{x}} i^{\mathbf{x}P\mathbf{x}^T+2\mathbf{b}\mathbf{x}^T}$, then either $S = 0$ or*

$$S^2 = i^{\mathbf{z}_1 P \mathbf{z}_1^T + 2\mathbf{b} \mathbf{z}_1^T} 2^{2m-\mathcal{R}},$$

where $\mathbf{z}_1 P = d_P$ and \mathcal{R} is the rank of the binary symmetric matrix P .

Proof. As in the proof of Lemma 4.1

$$S^2 = \sum_{\mathbf{z}} i^{\mathbf{z}P\mathbf{z}^T} \sum_{\mathbf{y}} (-1)^{(d_P+\mathbf{z}P)\mathbf{y}^T}.$$

The solution to the equation $\mathbf{z}P = 0$ form a vector space E and for all $e, \mathbf{f} \in E$

$$ePe^T + \mathbf{f}P\mathbf{f}^T = (e + \mathbf{f})P(e + \mathbf{f})^T \pmod{4.}$$

Hence

$$\begin{aligned} S^2 &= 2^m \sum_{e \in E} i^{(z_1+e)P(z_1+e)^T+2(z_1+e)b^T} \\ &= 2^m i^{z_1 P z_1^T + 2z_1 b^T} \sum_{e \in E} i^{e P e^T + 2e b^T}. \end{aligned}$$

The map $e \rightarrow e P e^T$ is a linear map from E to \mathbb{Z}_2 so that the numerator $e P e^T + 2e b^T$ also defines a linear map from E to \mathbb{Z}_2 . If this linear map is the zero map then

$$S^2 = 2^{2m-\mathcal{R}} i^{z_1 P z_1^T + 2z_1 b^T}$$

and if it is not zero then $S = 0$. □

Lemma 4.3. *If $S = \sum_x i^{\mathbf{x} P \mathbf{x}^T + 2\mathbf{b} \mathbf{x}^T}$ then the expected value*

$$E_{\mathbf{b} \in \mathbb{Z}_2^m} [|S|] = 2^m.$$

Proof. In the proof of Lemma 4.2, there are $2^{\mathcal{R}}$ choices for \mathbf{b} such that $e \rightarrow e P e^T + 2e b^T$ is the zero map. Hence

$$E_{\mathbf{b} \in \mathbb{Z}_2^m} [|S|] = \frac{2^{\mathcal{R}}}{2^m} 2^{2m-\mathcal{R}} = 2^m$$

regardless of the rank of P . □

Lemma 4.4. *Let $\mathcal{G} = \mathcal{G}(m, i)$ be the set of column vectors $\varphi^{\mathbf{P}, \mathbf{b}}$ where*

$$\varphi^{\mathbf{P}, \mathbf{b}}(\mathbf{x}) = i^{wt(d_P) + 2wt(\mathbf{b})} i^{\mathbf{x} P \mathbf{x}^T + 2\mathbf{b} \mathbf{x}^T}, \text{ for } \mathbf{x} \in \mathbb{Z}_2^m,$$

where $\mathbf{b} \in \mathbb{Z}_2^m$ and where the binary symmetric matrix P is drawn from a Delsarte-Goethals ensemble $DG(m, i)$. Then \mathcal{G} is a group under pointwise multiplication.

Proof. We have

$$\varphi^{\mathbf{P}, \mathbf{b}}(x) \varphi^{\mathbf{P}', \mathbf{b}'}(x) = i^{wt(d_P) + wt(d_{P'}) + 2wt(\mathbf{b} \oplus \mathbf{b}')} i^{\mathbf{x}(\mathbf{P} + \mathbf{P}') \mathbf{x}^T + 2(\mathbf{b} \oplus \mathbf{b}') \mathbf{x}^T}.$$

Write $\mathbf{P} + \mathbf{P}' = (\mathbf{P} \oplus \mathbf{P}') + 2\mathbf{Q}$ (mod 4) where \mathbf{Q} is a binary symmetric matrix. Observe that $2\mathbf{x} \mathbf{Q} \mathbf{x}^T = 2d_{\mathbf{Q}} \mathbf{x}^T$ (mod 4), where the diagonal $d_{\mathbf{Q}} = d_P * d_{P'}$ is the pointwise product of d_P and $d_{P'}$. Thus

$$\begin{aligned} &\varphi^{\mathbf{P}, \mathbf{b}}(x) \varphi^{\mathbf{P}', \mathbf{b}'}(x) \\ &= i^{([wt(d_P) + wt(d_{P'}) + 2wt(d_P * d_{P'}]) + 2wt(\mathbf{b} \oplus \mathbf{b}' \oplus d_P * d_{P'}))} i^{\mathbf{x}(\mathbf{P} \oplus \mathbf{P}') \mathbf{x}^T + 2(\mathbf{b} \oplus \mathbf{b}' \oplus d_P * d_{P'}) \mathbf{x}^T} \\ &= \varphi^{\mathbf{P} \oplus \mathbf{P}', \mathbf{b} \oplus \mathbf{b}' \oplus d_P * d_{P'}}(x) \end{aligned}$$

as required. □

Duals of Extended Binary BCH Codes: If $e = 2t + 1$, then all nonzero weights in the extended binary BCH code of length 2^m with designed distance e , except for $wt(\mathbf{1}) = 2^m$, are contained in the interval $[2^{m-1} - (t-1)2^{\frac{m}{2}}, 2^{m-1} + (t-1)2^{\frac{m}{2}}]$. This follows from the Carlitz-Uchiyama bound (see [MS], Chapter 9). Hence we may construct sensing matrices that satisfy the hypotheses of Theorem 2.4 by exponentiating codewords from these codes. Note that for odd m , the weight distribution of these codes coincides with that of certain Gray images of linear codes defined over \mathbb{Z}_4 , but the codes are inequivalent (see [CM95]). The column φ^c of the sensing matrix indexed by the codeword c is given by

$$\varphi^c(j) = (-1)^{\mathbf{b} \cdot c} (-1)^{c_j} \text{ where } j = 0, 1, \dots, 2^m - 1$$

where \mathbf{b} is a binary vector that is not orthogonal to the code.

Acknowledgment

The authors would like to thank Lorne Applebaum, Doug Cochran, Ingrid Daubechies, Anna Gilbert, Shamgar Gurevich, Ronnie Hadani, and Rachel Ward for helpful suggestions.

References

- [AHSC] L. Applebaum, S. Howard, S. Searle, and R. Calderbank. Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery. in *Applied and Computational Harmonic Analysis*, September 2008.
- [AT] M. Akcakaya and V. Tarokh. A frame construction and a universal distortion bound for sparse representations. *IEEE Int. Symposium on Information Theory (ISIT), Nice, France, June 2007*.
- [BDDW07] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 2007.
- [BGI⁺] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss. Combining geometry and combinatorics: a unified approach to sparse signal recovery. *Allerton*, 2008.
- [BHR⁺] W. Bajwa, J. Haupt, G. Raz, S. Wright, and R. Nowak. Toeplitz-structured compressed sensing matrices. in *IEEE Workshop on Statistical Signal Processing*.
- [CCKS] A.R. Calderbank, P.J. Cameron, W.M. Kantor, and J.J. Seidel. \mathbb{Z}_4 -Kerdock Codes, orthogonal spreads and extremal euclidean line sets. *Proceedings of London Math. Society*, vol. 75, pp. 436-480, 1997.
- [CM95] A.R. Calderbank and G. McGuire. \mathbb{Z}_4 -linear codes obtained as projections of Kerdock and Delsarte-Goethals Codes. *Linear Algebra and its Applications*, vol. 226-228, pp. 647-665, 1995.
- [CM06] G. Cormode and S. Muthukrishnan. Combinatorial algorithms for Compressed Sensing. In *Proc. 40th Ann. Conf. Information Sciences and Systems, Princeton*, 2006.
- [CRTa] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Inf. Theory*, 52(2):489509, 2006.
- [CRTb] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):12081223, 2006.
- [CT] E. Candès and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies. *IEEE Transactions on Information Theory*, vol. 52, pp.5406-5425, 2006.
- [DeV] R. A. DeVore. Deterministic constructions of compressed sensing matrices. *Journal of Complexity*, 23, pp. 918 - 925, August 2007.
- [DG] P. Delsarte and J. M. Goethals. Alternating bilinear forms over $\text{GF}(q)$. *Journal of Combinatorial Theory*, vol. 19, pp. 26-50, 1975.
- [DM] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing: Closing the gap between performance and complexity. *Arxiv:0803.0811*, 2008.
- [Don] D. Donoho. Compressed Sensing. *IEEE Trans. on Information Theory*, 52(4), pp. 1289 - 1306, April 2006.
- [dP] M. R. de Prony. Essai experimentalle et analytique. *J. Ecole Polytech.Paris*, vol. 1, pp. 2476, 1795.
- [Gab] E. Gabidulin. Theory of codes with maximum rank distance, Problems of Information Transmission. vol. 21, no. 1, pp. 3-14, 1985.
- [GH] Sh. Gurevich and R. Hadani. Incoherent dictionaries and the statistical restricted isometry property. *Preprint 2008*.
- [GLR] V. Guruswami, J. Lee, and A. Razborov. Almost euclidean subspaces of ℓ_1 via expander codes. *Soda 2008*.

- [Glu] E. D. Gluskin. Norms of random matrices and widths of finite dimensional sets. *Math USSR Sbornik*, vol. 48, pp. 173182, 1984.
- [GSTV] A. Gilbert, M. Strauss, J. Tropp, and R. Vershynin. One sketch for all: fast algorithms for compressed sensing. In *ACM STOC 2007*, pages 237246, 2007.
- [HCS] S. Howard, R. Calderbank, and S. Searle. A fast reconstruction algorithm for deterministic compressive sensing using second order Reed-Muller codes. *Conf. on Info. Sciences and Systems (CISS)*, Princeton, New Jersey, March 2008.
- [Ind] P. Indyk. Explicit constructions for compressed sensing of sparse signals. *SODA 2008*.
- [IR] P. Indyk and M. Ruzic. Near-optimal sparse recovery in the ℓ_1 norm. *FOCS 2008*.
- [JKC⁺] A. R. Hammons Jr, P. V. Kumar, A. R. Calderbank, N. J. A. Sloane, and P. Sole. The \mathbb{Z}_4 -linearity of Kerdock Codes, Preparata, Goethals, and related codes. *IEEE Transactions on Information Theory*, vol. 40, pp. 301-319, 1994.
- [JXHC] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank. Efficient compressed sensing using high-quality expander graphs. *submitted to the IEEE transactions on Information Theory*.
- [Kas] B. Kashin. The widths of certain finite dimensional sets and classes of smooth functions. *Izvestia*, vol. 41, pp. 334351, 1977.
- [Ker] A.M. Kerdock. A class of low-rate nonlinear binary codes. *Information and Control*, vol. 20, pp.182-187, 1972.
- [LK] H.F. Lu and P.V. Kumar. A unified construction of space-time codes with optimal rate diversity tradeoff. *IEEE Transactions on Information Theory*, vol. 51, pp. 1709-1730, 2005.
- [MS] F.J. MacWilliams and N.J.A. Sloane. *The Theory of Error-Correcting Codes*, chapter 9, 15. North-Holland: Amsterdam, 1977.
- [NT] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comp. Harmonic Anal.*
- [SBB] Sh. Sarvotham, D. Baron, and R. Baraniuk. Compressed sensing reconstruction via belief propagation. *Rice ECE Department Technical Report TREE 0601*, 2006.
- [TWD⁺] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk. Random filters for compressive sampling and reconstruction. in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2006.
- [Wol] J. Wolf. Decoding of Bose-Chaudhuri-Hocquenghem Codes and Prony's method for curve fitting. *IEEE Trans. Inform. Theory*, vol.IT-3, p. 608, 1967.