Deterministic Compressive Sensing with Groups of Random Variables

Robert Calderbank
Mathematics and Electrical Engineering
Princeton University
NJ 08540, USA

Stephen Howard
DSTO
PO Box 1500
Edinburgh 5111, Australia

Sina Jafarpour
Computer Science
Princeton University
NJ, 08540, USA

Abstract

Compressed Sensing aims to capture attributes of a sparse signal using very few measurements. Candès and Tao showed that random matrices satisfy the Restricted Isometry Property or RIP property with high probability. They 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 as equally likely, in contrast to mainstream signal processing where the filtering is deterministic, and the signal is described probabilistically. This paper provides weak conditions that are sufficient to show that a deterministic sensing matrix satisfies the Statistical Restricted Isometry Property (StRIP), and to guarantee the uniqueness of k-sparse representations. The columns of the sensing matrix are required to form a group under pointwise multiplication, and it is this property that provides the concentration inequalities of McDiarmid and Bernstein with the leverage necessary to guarantee uniqueness of sparse representations. We provide a large class of deterministic sensing matrices for which Basis Pursuit is able to recover sparse Steinhaus signals. The new framework encompasses many families of deterministic sensing matrices, including those formed from discrete chirps, Delsarte-Goethals codes, and Extended BCH codes. In these cases it provides theoretical guarantees on the performance of nonlinear reconstruction algorithms with complexity that is only quadratic in the dimension of the measurement domain.

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 $\boldsymbol{\alpha}$ 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 $f = \boldsymbol{\Phi}\boldsymbol{\alpha}$, where $\boldsymbol{\Phi}$ is an $N \times \mathcal{C}$ matrix called the *sensing matrix*. The two fundamental questions are construction of suitable sensing matrices $\boldsymbol{\Phi}$ and efficient reconstruction of $\boldsymbol{\alpha}$ from f.

The work of Donoho [Don06] and of Candès, Romberg and Tao [CT06], [CRT06b], [CRT06a] 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. When $\frac{N}{\mathcal{C}}$ and/or $\frac{k}{N}$ are small, deterministic sensing matrices with the RIP property have been constructed using methods from approximation theory [DeV07] and coding theory [Ind08]. 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 [Don06], [CT06] and the number of measurements N is $k \log\left(\frac{\mathcal{C}}{k}\right)$. 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$^+$07], [TWD$^+$06] and expander graphs [GLR08], [BGI$^+$08], [IR08], [JXHC09].

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. 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 [CRT06b] or Matching Pursuit [GSTV07] 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. Similarly, there is no known algorithm tor verifying whether a random graph is an expander graph and explicit constructions of expander graphs are sub-optimal.

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 difference of the absolute values of the sum of the entries in any two columns 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 [AHSC08] or the Walsh-Hadamard domain [HCS08]. 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. [GH08]). Moreover, Finally, there are interesting connections between the techniques to prove the most general StRIP property and the techniques for approximating the frequency moments [AMS96].

Note that Tropp [Tro08a], [Tro08b] had previously derived conditions on the incoherence and spectral norm of a sensing matrix that are sufficient to prove that a random submatrix is highly likely to define a near isometry. He demonstrated feasibility of sparse recovery through $\ell 1$ minimization when the sensing matrix is incoherent, the sparse signal is supported on a well conditioned submatrix, and the nonzero entries in the sparse vector have random signs. Our approach uses the strong concentration inequality of McDiarmid to avoid the restrictive incoherency assumptions made by Tropp

Prior constructions of deterministic sensing matrices using approximation theory [DeV07] or number theory [Iwe09] require at least k^2 measurements. Those using unbalanced expander graphs require $O\left((k2^{(\log\log\mathcal{C})^{O(1)}}\right)$ measurements [IR08], and explicit constructions [GUV07] are complicated. Our deterministic sensing matrices on the other hand, at the presence of noise, only require $\Omega\left(k\log(\mathcal{C})\right)$ measurements, and are very easy to construct. The properties we require are satisfied by a large class of matrices constructed by exponentiating codewords from a linear code. Most important, we show that these relatively weak properties are sufficient to guarantee unique representation of sparse signals. It is the group property of columns in the sensing matrix that provides the concentration inequalities of McDiarmid with the necessary leverage.

Tables 1 and 2 summarize performance of different approaches to reconstruction of k-sparse signals of length $\mathcal{C}$ via randomized and deterministic matrices respectively. The quadratic reconstruction algorithm [AHSC08], [HCS08] involves pointwise multiplication of the k-sparse superposition with a shift of itself, followed by the Fourier / Walsh Hadamard transform, and the StRIP property holds for all offsets and every Fourier / Walsh-Hadamard coefficient. Experimental results in these papers suggested that these deterministic matrices satisfy the StRIP property, and the quadratic decoding algorithm with very high probability recovers the original sparse signal. We prove that these matrices satisfy StRIP, and we go further by proving that with overwhelming probability each sparse signal has a unique representation in the measurement domain. This guarantees that the reconstruction algorithm proposed in [AHSC08], [HCS08] which is quadratic in the dimension of the MEASUREMENT domain, will recover sparse signals with high probability. The quadratic decoding algorithm [AHSC08], [HCS08] applies to “chirp-like” sensing matrices. However, our construction of deterministic sensing matrices is much more general, and for this larger class of matrices, basis pursuit and ℓ_1 minimization will successfully recover a sparse signal with high probability.

Table 1. Properties of k-sparse reconstruction algorithms that employ random sensing matrices with N Rows and $\mathcal{C}$ Columns. Note that ℓ_p/ℓ_q means $\|\boldsymbol{\alpha}^* - \hat{\boldsymbol{\alpha}}\|_p \approx \|\boldsymbol{\alpha}^* - \boldsymbol{\alpha}_k^*\|_q$

Approach	Number of Measurements N	Complexity	Noise Resilience	Compressible Signals	RIP
Basis Pursuit (BP) [CRT06b]	$k\log\left(\frac{\mathcal{C}}{k}\right)$	$\mathcal{C}^3$	Yes	ℓ_2/ℓ_1	Yes
Orthogonal Matching Pursuit (OMP) [GSTV07]	$k\log^{\alpha}(\mathcal{C})$	$k^2\log^{\alpha}(\mathcal{C})$	No	ℓ_2/ℓ_1	Yes
Group Testing [CM06]	$k\log^{\alpha}(\mathcal{C})$	$k\log^{\alpha}(\mathcal{C})$	No	ℓ_1/ℓ_1	No
Expanders (Unique Neighborhood) [JXHC09]	$k\log\left(\frac{\mathcal{C}}{N}\right)$	$\mathcal{C}\log\left(\frac{\mathcal{C}}{k}\right)$	Yes	ℓ_1/ℓ_1	RIP-1
Expanders (BP) [BGI+08]	$k\log\left(\frac{\mathcal{C}}{k}\right)$	$\mathcal{C}^3$	Yes	ℓ_1/ℓ_1	RIP-1
Expander Matching Pursuit(EMP) [IR08]	$k\log\left(\frac{\mathcal{C}}{k}\right)$	$\mathcal{C}\log\left(\frac{\mathcal{C}}{k}\right)$	Yes	ℓ_1/ℓ_1	RIP-1
CoSaMP [NT08]	$k\log\left(\frac{\mathcal{C}}{k}\right)$	$\mathcal{C}k\log\left(\frac{\mathcal{C}}{k}\right)$	Yes	ℓ_2/ℓ_1	Yes
SSMP [DM08]	$k\log\left(\frac{\mathcal{C}}{k}\right)$	$\mathcal{C}k\log\left(\frac{\mathcal{C}}{k}\right)$	Yes	ℓ_2/ℓ_1	Yes

The quadratic reconstruction algorithm [AHSC08], [HCS08] is described below as Algorithm 1 and is nonlinear. The first step is pointwise multiplication of a sparse superposition

$$f(x) = \frac{1}{\sqrt{N}}\sum_{i=1}^{k}\alpha_i\varphi^i(x)$$

with a shifted copy of itself. The sensing matrix is obtained by exponentiating quadratic functions so the first step produces a sparse superposition of pure frequencies (in the example below, these are Walsh functions in the binary domain) against a background of cross terms.

$$f(x+a)\overline{f(x)} = \frac{1}{N}\sum_{j=1}^{k}|\alpha_j|^2(-1)^{a^\top P_j x} + \frac{1}{N}\sum_{j\neq t}\alpha_j\overline{\alpha_t}\varphi^{P_j,b_j}(x+a)\overline{\varphi^{P_t,b_t}(x)}. \tag{1}$$

Then the fast Hadamard transform concentrates the energy of the first term $\frac{1}{N}\sum_{j=1}^{k}|\alpha_j|^2(-1)^{a^\top P_j x}$ at (at most) k Walsh-Hadamard tones, while the second term distributes energy uniformly across all N tones. The l^{th} Fourier coefficient is

$$\Gamma_a^l = \frac{1}{N^{3/2}}\sum_{j\neq t}\alpha_j\overline{\alpha_t}\sum_x(-1)^{l^\top x}\varphi^{P_j,b_j}(x+a)\overline{\varphi^{P_t,b_t}(x)}, \tag{2}$$

and it can be shown that the energy of the chirp-like cross terms is distributed uniformly in the Walsh-Hadamard domain. That is for any coefficient l

$$\lim_{N\to\infty}\mathbb{E}\left[N^2\left|\Gamma_a^l\right|^2\right] = \sum_{j\neq t}|\alpha_j|^2|\alpha_t|^2. \tag{3}$$

Table 2. Properties of k sparse reconstruction algorithms that employ deterministic sensing matrices with N Rows and $\mathcal{C}$ Columns. Note that for LDPC codes $k \ll \mathcal{C}$. 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. [GLR08] proved that this property also holds for deterministic sensing matrices constructed from expander codes. It follows from Theorem 2.3 in this paper that sensing matrices based on discrete chirps and Delsarte-Goethals codes satisfy the StRIP.

Approach	Number of Measurements N	Complexity	Noise Resilience	Compressible Signals	RIP
Low Density Parity Check Codes (LDPC) [SBB06]	$k \log \mathcal{C}$	$\mathcal{C} \log \mathcal{C}$	Yes	ℓ_2/ℓ_1	No
Reed-Solomon codes [AT07]	k	k^2	No	No	No
Embedding ℓ_2 into ℓ_1 (BP) [GLR08]	$k(\log \mathcal{C})^{\alpha \log \log \mathcal{C}}$	$\mathcal{C}^3$	No	ℓ_2/ℓ_1	No
Extractors [Ind08]	$k\mathcal{C}^{o(1)}$	$k\mathcal{C}^{o(1)} \log(\mathcal{C})$	No	No	No
Discrete chirps [AHSC08] (Quadratic Decoding)	$\sqrt{\mathcal{C}}$	$k\sqrt{\mathcal{C}} \log \mathcal{C}$ $(kN \log N)$	Yes	ℓ_2/ℓ_2 (This paper)	StRIP (This paper)
Delsarte-Goethals codes [HCS08] (Quadratic Decoding)	$2^{\sqrt{\log \mathcal{C}}}$	$k2^{\sqrt{\log \mathcal{C}}} \log \mathcal{C}$ $(kN \log^2 N)$	Yes	ℓ_2/ℓ_2 (This paper)	StRIP (This paper)

Equation (3) is related to the variance of f and may be viewed as a fine-grained concentration estimate. The Walsh-Hadamard tones appear as spikes above a uniform "noise" floor and the quadratic algorithm learns the terms in the sparse superposition by varying the offset a. These terms can be peeled off in decreasing order of signal strength or processed in a list (cf [GS99]), and it may also be possible to incorporate methods of finding locally significant Fourier coefficients (cf. [Aka08]). Experimental results show close approach to the information theoretic lower bound on the required number of measurements [HCS08]. Since our approach guarantees a good approximation $\hat{\boldsymbol{\alpha}}$ of the original signal $\boldsymbol{\alpha}$ among all signals with the same support T, the recovery is of the form ℓ_2/ℓ_2 that is

$$\|\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}\|_2 \approx \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_k\|_2, \tag{4}$$

which is tighter than ℓ_2/ℓ_1 of basis pursuit and matching pursuit methods and ℓ_1/ℓ_1 of expander-based methods. The quadratic algorithm is a recasting of the chirp detection algorithm commonly used in navigation radars which is known to work extremely well in the presence of noise.

2. Construction of Deterministic Sensing Matrices that satisfy the StRIP

Let Φ be a deterministic sensing matrix with N rows and $\mathcal{C}$ columns, and let $\varphi^i(x)$ denote the entry in row x and column i. We begin by imposing two very simple conditions.

1) the columns of Φ form a group $G_{\mathcal{C}}$ 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, that is for any row x and column i, we have $|\varphi^i(x)| = 1$. It also follows from conditions (1) and (2) that the normalized columns $\frac{1}{\sqrt{N}}\varphi$ form a tight frame with redundancy $\frac{\mathcal{C}}{N}$, and so $E_{\varphi \in G_{\mathcal{C}}}\left[\left|\sum_x \varphi(x)\right|^2\right] = N$ (Appendix E).

We now show that with high probability the sensing matrix $\frac{1}{\sqrt{N}}\Phi$ preserves the norm of any k-sparse input signal $\boldsymbol{\alpha}$ to within a small fraction. This is the Statistical Restricted Isometry Property or StRIP.

Algorithm 1 Quadratic Decoding Algorithm

Input: N dimensional vector $f = \Phi\boldsymbol{\alpha}$

Output: An approximation $\hat{\boldsymbol{\alpha}}$ of the signal $\boldsymbol{\alpha}$

for $t = 1, \cdots, k$ or while $\|f\|_2 \geq \epsilon$ **do**

  for each entry $x = 1$ to N **do**

    produces a Walsh function using poitwise multiplication of Equation (1)

  end for

  Compute the Fast Hadamard Transform of the Walsh vector: Equation (2)

  Find the position of the next peak l_t in the Hadamard domain: Equation (3) guarantees that chirp terms have no effect.

  Determine the value $\hat{\alpha}_l$ which minimizes $\|f - \hat{\alpha}_l \varphi^{P_l, b_l}\|_2$. StRIP property (This paper).

  Set $f \leftarrow f - \hat{\alpha}_l \varphi^{P_l, b_l}$.

end for

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 $T = \{T_1, \cdots, T_k\}$ be a random permutation of the columns of Φ. The image of α in the measurement domain is $f(x) = \frac{1}{\sqrt{N}} \sum_{j=1}^{k} \alpha_j \varphi^{T_j}(x)$ and we now show that Conditions 1) and 2) make it very easy to calculate the expected value and variance of $||f||^2$.

We have

$$||f||^2 = \sum_{x=1}^{N} |f(x)|^2 = \frac{1}{N} \sum_{x} \left(\sum_{t=1}^{k} |\alpha_t|^2 + \Psi(x) \right) \tag{5}$$

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

Note that the expectation $E[||f||^2]$ is over all admissible choices of columns φ^{T_j} based on the random permutation T.

Theorem 2.1. *Let T be a random column permutation, let Φ be a deterministic sensing matrix satisfying conditions 1) and 2) given above, and let $\boldsymbol{\alpha}$ be a k-sparse complex valued signal with non-zero entries in positions $T_1, \cdots, T_k$. Then*

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

Proof: See Appendix A. □

Next we show that Condition 1) greatly simplifies calculation of the variance of $||f||^2$. We have

$$\begin{aligned} ||f||^4 &= \sum_{x,x'} |f(x)|^2 |f(x')|^2 = \sum_{x,x'} f(x)\overline{f(x)} f(x')\overline{f(x')} \\ &= \frac{1}{N^2} \sum_{x,x'} \left(\sum_{t=1}^{k} |\alpha_t|^2 + \Psi(x) \right) \left(\sum_{t=1}^{k} |\alpha_t|^2 + \Psi(x') \right) \end{aligned} \tag{6}$$

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

In order to prove the Statistical Restricted Isometry Property (StRIP), we need to bound the expectation $E[||f||^4]$ taken over all admissible choices of columns φ^{T_j}. The first term in (6) 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 $V[||f||^2]$.

The second term in (6) is given by

$$\frac{2}{N^2}\left(\sum_{j=1}^{k}|\alpha_j|^2\right)\sum_{j\neq i}\alpha_j\overline{\alpha_i}\sum_x\varphi^{T_j}(x)\overline{\varphi^{T_i}(x)}$$

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

$$E_{j\neq i}\left[\sum_x\varphi^{T_j}(x)\overline{\varphi^{T_i}}\right]. \tag{7}$$

If C is the number of columns, then we can rewrite the second moment (7) as

$$\frac{1}{\mathcal{C}(\mathcal{C}-1)}\left[\mathbf{1}^T(\sum_{\substack{g,h\in\mathcal{G}\\ g\neq h}}gh^{-1})\right] \tag{8}$$

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. Condition 1) simplifies this second moment dramatically.

Lemma 2.2. *The map $\mathcal{G}\times\mathcal{G}\to\mathcal{G}$ given by $(g,h)\to gh^{-1}$ is a* homomorphism *and*

$$\sum_{\substack{g,h\in\mathcal{G}\\ g\neq h}}gh^{-1}=-\mathcal{C}\mathbf{1} \tag{9}$$

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. □

Using Lemma 2.2, the variance terms in (6) are bounded in Appendix B. Implications for the geometry of deterministic sensing matrices are found in the next theorem (cf. [CT06]).

Theorem 2.3 (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 $G_{\mathcal{C}}$ under pointwise multiplication.*
2) *the rows of Φ are orthogonal, and all row sums are equal to zero.*

Let $\boldsymbol{\alpha}$ 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 $1-\left({}^{2k}/_{N}+{}^{(2k+7)}/_{(\mathcal{C}-3)}\right)/\left(\delta-{}^{(k-1)}/_{(\mathcal{C}-1)}\right)^2$

$$(1-\delta)\|\boldsymbol{\alpha}\|^2\leq\|\frac{1}{\sqrt{N}}\Phi\boldsymbol{\alpha}\|^2\leq(1+\delta)\|\alpha\|^2.$$

Proof. See Appendix C. □

If any two column sums of the sensing matrix are also sufficiently close, then the rate of success of the StRIP property can be improved significantly.

Theorem 2.4. *Let Φ be a deterministic sensing matrix satisfying the conditions of Theorem 2.3, and also for any two columns i,j of Φ,*

$$\left|\sum_x\varphi^i(x)-\sum_x\varphi^j(x)\right|^2\leq N^{2-\gamma}, \tag{10}$$

for some positive γ. Then with probability $1-2\exp\left(\frac{-\delta^2 N^\gamma}{2k}\right)$ the following near-isometry holds:

$$(1-\delta)\|\boldsymbol{\alpha}\|^2 \leq \|\frac{1}{\sqrt{N}}\Phi\boldsymbol{\alpha}\|^2 \leq (1+\delta)\|\boldsymbol{\alpha}\|^2. \tag{11}$$

Proof: See Appendix D. □

3. The Null Space Property for Deterministic Sensing Matrices

Next we describe a third condition on deterministic sensing matrices that guarantees uniqueness of sparse representations with high probability (cf. the Null Space Property of [CDD08]]). This condition is satisfied by a large family of matrices constructed by exponentiating codewords from linear codes defined over the binary field and other rings. Uniqueness of sparse representation means that with high probability no other sparse signal will have the same representation in the measurement domain. This probability is only over the position of the non-zero entries of the sparse signal and the values can be any arbitrary complex numbers. A significant consequence is guarantees on the fidelity of the quadratic decoding algorithms proposed in [AHSC08] and [HCS08].

This is the first construction of deterministic sensing matrices, requiring only $\Omega\left(k\log\mathcal{C}\right)$ measurements for which uniqueness of sparse representation is guaranteed. The previous results are either based on group testing and approximation theory [DeV07] or number theory [Iwe09] requiring at least k^2 measurements, or require very strict incoherence assumptions, or are based on expander graphs [IR08], [JXHC09], [BGI+08] for which the best explicit construction requires $O\left((k2^{(\log\log\mathcal{C})^{O(1)}}\right)$ measurements [IR08], [GUV07], and has a more complicated explicit construction than the construction of deterministic matrices suggested in this paper.

The extra condition that we require is that any two column sums are sufficiently close. More precisely for any two columns i, j of the sensing matrix,

$$\forall\ i,j: \quad \left| |\sum_x \varphi^i(x)|^2 - |\sum_x \varphi^j(x)|^2 \right| \leq N^{2-\eta}, \tag{12}$$

for some $\eta < 1$[1]. For matrices constructed by exponentiating codewords from say a binary linear code it is satisfied if all the weights to lie in a range around $\frac{N}{2}$. Note that by unimodularity of the elements of Φ, for any column i we have $\left|\sum_x \varphi^i(x)\right|^2 \leq N^2$, and the tight frame property of Φ implies that $E_{\varphi\in U_\mathcal{C}}\left[|\sum_x \varphi(x)|^2\right] = N$.

For any set $S \subseteq \{1,\cdots,\mathcal{C}\}$ let Φ_S be Φ restricted to columns from S. Clearly if $|S| = 1$ then $\Phi_S = \varphi^S$. In order to prove the uniqueness of the sparse representation of a sparse signal $\boldsymbol{\alpha}$, we need a chain of lemmas to bound the inter-connection between the columns corresponding to support of $\boldsymbol{\alpha}$ and the remaining columns of the sensing matrix Φ.

Lemma 3.1. *Let w be a fixed column of Φ, and let T be a random k-subset of columns. Then*

$$\mathbb{E}_T\left[\left\|\frac{1}{\sqrt{N}}\Phi_T\frac{1}{\sqrt{N}}\varphi^w\right\|^2\right] = \frac{k}{N}. \tag{13}$$

Proof: Let $T = \{T_1,\cdots,T_k\}$. By linearity of expectation we have

$$\mathbb{E}_T\left[\left\|\frac{1}{\sqrt{N}}\Phi_T\frac{1}{\sqrt{N}}\varphi^w\right\|^2\right] = \frac{1}{N^2}\sum_{i=1}^k \mathbb{E}_{T_i}\left[\left|\left(\varphi^{T_i}\right)^\top\varphi^w\right|^2\right]. \tag{14}$$

1. Indeed, $\eta \approx 0.5$ is enough for our analysis.

Now since the columns of Φ form a group under the pointwise multiplication, we have

$$\left(\varphi^{T_i}\right)^\top \varphi^w \;=\; \sum_x \overline{\varphi^{T_i}(x)}\varphi^w(x) = \sum_x \varphi^{Z_i}(x),$$

for sum index Z_i, which has uniform distribution since w is fixed and T_i has uniform distribution and hence:

$$\sum_{i=1}^k \frac{1}{N^2}\mathbb{E}_{T_i}\left[\left|\left(\varphi^{T_i}\right)^\top \varphi^w\right|^2\right] = \sum_{i=1}^k \frac{1}{N^2}\mathbb{E}_{z_i}\left[\left|\sum_x \varphi^{Z_i}(x)\right|^2\right] = \frac{k}{N}.$$

□

Having bounded $\mathbb{E}_T\left[\|\Phi_T\varphi^w\|^2\right]$, we use McDiarmid's inequality [BBL05] to show that the random variable $\|\Phi_T\varphi^w\|^2$ has high concentration around its mean.

Proposition 1 (McDiarmid's inequality)**.** *Given a function f for which*

$$\forall\ x_1,\cdots,x_k,x_i' :\ \ \left|f(x_1,\cdots,x_i,\cdots,x_k) - f(x_1,\cdots,x_i',\cdots,x_k)\right| \le c_i,$$

and given $X_1,\cdots,X_k$ independent random variables. Then

$$\Pr\left[f(X_1,\cdots,X_k) \ge \mathbb{E}[f(X_1,\cdots,X_k)] + \eta\right] \le \exp\left(\frac{-2\eta^2}{\sum c_i^2}\right).$$

Using McDiarmid's inequality, and Condition (12) we now derive a uniform convergence bound for the random variable $\|\Phi_T\varphi^w\|^2$.

Theorem 3.2. *Let T be a set of k random columns of Φ. Then with probability $1-\delta$ for any column w of Φ we have*

$$\left\|\frac{1}{\sqrt{N}}\Phi_T\frac{1}{\sqrt{N}}\varphi^w\right\|^2 \le \frac{k}{N} + \frac{\sqrt{k\log \mathcal{C}/\delta}}{N^\eta}. \tag{15}$$

Proof: Let $f(t_1,\cdots,t_k) = \frac{1}{N^2}\sum_{i=1}^k\left|\left(\varphi^{t_i}\right)^\top\varphi^w\right|^2$. It follows from Condition (12) and the group property that for any i, i'

$$\frac{1}{N^2}\left|\left|\left(\varphi^{t_i}\right)^\top\varphi^w\right|^2 - \left|\left(\varphi^{t_i'}\right)^\top\varphi^w\right|^2\right| = \frac{1}{N^2}\left|\left|\sum_x \varphi^{z_i}(x)\right|^2 - \left|\sum_x \varphi^{z_i'}(x)\right|^2\right| = \frac{1}{N^\eta}.$$

Hence for any positive γ,

$$\Pr\left[\left\|\frac{1}{\sqrt{N}}\Phi_T\frac{1}{\sqrt{N}}\varphi^w\right\|^2 \ge \frac{k}{N} + \gamma\right] \le \exp\left(\frac{-2\gamma^2N^{2\eta}}{k}\right).$$

□

Now by applying union bounds over all possible columns of Φ we get:

$$\Pr_T\left[\exists\ w:\ \ \left\|\frac{1}{\sqrt{N}}\Phi_T\frac{1}{\sqrt{N}}\varphi^w\right\|^2 \ge \frac{k}{N} + \gamma\right] \le \mathcal{C}\exp\left(\frac{-2\gamma^2N^{2\eta}}{k}\right).$$

Writing γ in terms of δ completes the proof. Hence, one can choose the constant C_δ, and $N = C_\delta k\log\mathcal{C}$ so that if we choose k columns from the matrix uniformly at random, for any other column w we have

$$\left\|\frac{1}{\sqrt{N}}\Phi_T\frac{1}{\sqrt{N}}\varphi^w\right\|^2 \le (1-\delta)^2. \tag{16}$$

Before proving the uniqueness of sparse representation we need the following lemma (See also [Tro08b] for a similar lemma for incoherent dictionaries).

Lemma 3.3. *Let $T = \{T_1, \cdots, T_k\}$ be a set of k indices sampled uniformly from $\{1, \cdots, \mathcal{C}\}$. Assume that $\frac{1}{\sqrt{N}}\Phi$ satisfies the StRIP(δ) property, that is*

$$\sigma_{\min}(\Phi_T) \geq (1-\delta)\sqrt{N}, \tag{17}$$

for some positive constant δ and that the conditions of Theorem 3.2 are satisfied. Let S be any other subset of $\{1, \cdots, \mathcal{C}\}$ of size less than or equal to k. Then

$$\dim\left(\textit{range}(\Phi_T) \cap \textit{range}(\Phi_S)\right) < k. \tag{18}$$

Proof: First, note that we should assume that $\dim(\text{range}(\Phi_S)) = k$, otherwise the condition of the Equation (18) cannot be satisfied and we are done. Since $S \neq T$, S has at least one index not in T. Denote that index by s. Since the entries of the matrix are all unimodular we have

$$\|\varphi^s\|^2 = \sum_x |\varphi^s(x)|^2 = N. \tag{19}$$

Let $\mathbb{P}_T$ be the orthogonal projection operator on the range(Φ_T). To prove the claim of Equation (18), it is enough to show that $\|\mathbb{P}_T\varphi^s\|^2 < \|\varphi^s\|^2$, which implies that there exists a vector in the range Φ_S that is outside the range of Φ_T. Note that

$$\mathbb{P}_T = \left(\Phi_T^{\dagger}\right)^{\top}\Phi_T^{\top}, \tag{20}$$

where $\Phi_T^{\dagger}$ is the pseudo-inverse of Φ_T defined as $\left(\Phi_T^{\top}\Phi_T\right)^{-1}\Phi_T^{\top}$. Since $\frac{1}{\sqrt{N}}\Phi_T$ hast StRIP(δ) property, the usual norm estimate implies that

$$\|\mathbb{P}_T\varphi^s\|^2 \leq \left\|\Phi_T^{\dagger}\right\|_2^2 \left\|\Phi_T^{\top}\varphi^s\right\|^2 \leq \frac{\left\|\Phi_T^{\top}\varphi^s\right\|^2}{(\sigma_{\min}(\Phi_T))^2} \leq \frac{\left\|\Phi_T^{\top}\varphi^s\right\|^2}{N(1-\delta)^2}.$$

Equation (16) states that by choosing $N = C_\delta k \log \mathcal{C}$ where C_δ is some appropriate constant depending only on δ, we can guarantee that $\|\mathbb{P}_T\varphi^s\|_2 < N$. □

Theorem 3.4. *Let $f = \Phi\boldsymbol{\alpha}$, then with overwhelming probability, $\boldsymbol{\alpha}$ is the only unique vector with sparsity k that satisfies equations $\Phi\boldsymbol{x} = f$.*

Proof: Let T be the support of $\boldsymbol{\alpha}$ with size k and uniform distribution. Since Φ_T satisfies StRIP, it is non-singular and hence $\dim(\text{range}(\Phi_T)) = k$. The StRIP property implies that no two signals with support T can have the same value in the measurement domain. Lemma 3.3, on the other hand states that the set of signals in range(Φ_T), that have another sparse representation, must have at most dimension $k-1$, and hence this set has zero volume with respect to any nonatomic measure. Hence almost surely the signal cannot have any other sparse representation. □

Note that quadratic decoding algorithm finds the support of the sparse signal, and our result on the number of measurements $N = \Omega(k \log \mathcal{C})$ is aligned with the information theoretic lower bounds for the number of measurements required to recover the exact support of a sparse signal in the presence of noise [AT08], [Wai06].

4. Recovery of Complex Steinhaus Signals via $\ell 1$ Minimization

In this section, we show that for a particular family of (even complex) sparse signals, compressively sampled using a deterministic matrix satisfying StRIP, it is possible to recover the signal uniquely using the *basis pursuit* algorithm

$$\text{minimize } \|x\|_1 \text{ such that } \Phi x = y. \tag{21}$$

This section extends the work of Fuchs [Fuc04] and Tropp [Tro08b]. Their result require high incoherence among the columns of the sensing matrix. Here we show that the simple conditions required for StRIP property, and Equation (12) are enough. We assume that the sparse signal comes from the family of *Steinhaus* random variables. That is, the signal is in $\mathbb{C}^{\mathcal{C}}$, it has at most k non-zero entries, chosen uniformly at random from $\{1, \cdots, \mathcal{C}\}$, and each non-zero entry has a complex value whose phase is chosen uniformly at random from the unit complex circle, and its value is chosen arbitrarily or adversarially. Many signals of interest in spread-spectrum wireless communications. for example spreading sequences, can be modeled as Steinhaus random variables, and our work provides a new approach to classical intractable problems in that domain such as multi-user detection. We show that Steinhaus signals sensed using deterministic matrices with $\Omega\left(k \log \mathcal{C}\right)$ measurements can be recovered successfully using the basis pursuit algorithm.

Theorem 4.1. *Let $\boldsymbol{x}^*$ be a k-sparse complex Steinhaus signal compressively sampled using $\Omega\left(C_\delta k \log\ \mathcal{C}\right)$ measurements from a deterministic matrix with StRIP. Then with overwhelming probability, the basis pursuit algorithm of Equation (21) recovers $\boldsymbol{x}^*$ successfully.*

Proof: The proof is based on the propositions proved by Fuchs, and Tropp [Fuc04], [Tro08b], and the StRIP property, and is provided in Appendix F. □

Acknowledgment

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

References

[AHSC08] 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*, 2008.

[Aka08] A. Akavia. Finding Significant Fourier Transform Coefficients Deterministically and Locally. *Electronic Colloquium on Computational Complexity (ECCC) 15(102)*, 2008.

[AMS96] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *Proc. ACM STOC, 2029*, 1996.

[AT07] 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*, 2007.

[AT08] M. Akcakaya and V. Tarokh. Shannon Theoretic Limits on Noisy Compressive Sampling. *CISS*, 2008.

[BBL05] S. Boucheron, O. Bousquet, and G. Lugosi. Theory of Classification: a Survey of Recent Advances. *ESAIM: Probability and Statistics, 9:323–375*, 2005.

[BGI+08] 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+07] W. Bajwa, J. Haupt, G. Raz, S. Wright, and R. Nowak. Toeplitz-structured compressed sensing matrices. *in IEEE Workshop on Statistical Signal Processing*, 2007.

[CDD08] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc. 22 (2009), 211-231*, 2008.

[CHJ09] R. Calderbank, S. Howard, and S. Jafarpour. Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property. *Technical Report, Princeton University*, 2009.

[CM06] G. Cormode and S. Muthukrishnan. Combinatorial algorithms for Compressed Sensing. *In Proc. 40th Ann. Conf. Information Sciences and Systems, Princeton*, 2006.

[CRT06a] 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.

[CRT06b] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math., 59(8):12081223*, 2006.

[CT06] 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.

[DeV07] R. DeVore. Derministic constructions of compressed sensing matrices. *J. of Complexity, 23, pp. 918 - 925*, 2007.

[DM08] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing: Closing the gap between performance and complexity. *Arxiv:0803.0811*, 2008.

[Don06] D. Donoho. Compressed Sensing. *IEEE Trans. on Information Theory, 52(4), pp. 1289 - 1306*, 2006.

[Fuc04] J-J Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Trans. Info. Theory, 50(6):13411344*, 2004.

[GH08] Sh. Gurevich and R. Hadani. Incoherent dictionaries and the statistical restricted isometry property. *Preprint*, 2008.

[GLR08] V. Guruswami, J. Lee, and A. Razborov. Almost euclidean subspaces of ℓ_1 via expander codes. *Soda*, 2008.

[GS99] V. Guruswami and M. Sudan. Improved Decoding of ReedSolomon Codes and Algebraic Geometry Codes. *IEEE Trans. Inform. Theory, vol. 45, no. 6, pp. 17571767*, 1999.

[GSTV07] 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.

[GUV07] V. Guruswami, C. Umans, and S. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes. *IEEE Conference on Computational Complexity (CCC)*, 2007.

[HCS08] 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*, 2008.

[Ind08] P. Indyk. Explicit constructions for compressed sensing of sparse signals. *SODA*, 2008.

[IR08] P. Indyk and M. Ruzic. Near-optimal sparse recovery in the $\ell 1$ norm. *FOCS*, 2008.

[Iwe09] M. Iwen. Simple Deterministically Constructible RIP Matrices with Sublinear Fourier Sampling Requirements. *Submitted to CISS 2009*, 2009.

[JXHC09] S. Jafarpour, W. Xu, B. Hassibi, and R. Calderbank. Efficient compressed sensing using high-quality expander graphs. *to appear in IEEE Transactions on Information Theory*, 2009.

[NT08] D. Needell and J. A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comp. Harmonic Anal*, 2008.

[SBB06] Sh. Sarvotham, D. Baron, and R. Baraniuk. Compressed sensing reconstruction via belief propagation. *Rice ECE Department Technical Report TREE 0601*, 2006.

[Tro08a] J. Tropp. Norms of Random Submatrices and Sparse Approximation. *Comptes Rendus Mathematique*, 2008.

[Tro08b] J. Tropp. On the conditioning of random subdictionaries. *Appl. Comput. Harmonic Anal., vol. 25, pp. 124*, 2008.

[TWD+06] 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*, 2006.

[Wai06] M. Wainwright. Sharp thresholds for noisy and high-dimensional recovery of sparsity using $\ell 1$-constrained quadratic programming. *Technical report, UC Berkeley, Department of Statistics*, 2006.

Appendix A: Proof of Lemma 4.2

The first term in (5) 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 $\frac{k}{\mathcal{C}}||\boldsymbol{\alpha}||^2$.

Lemma 4.2. *Let T be a random column permutation, let Φ be a deterministic sensing matrix satisfying conditions 1) and 2) given above, and let $\boldsymbol{\alpha}$ be a k-sparse complex valued signal with non-zero entries in positions $T_1, \cdots, T_k$. Then*

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

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

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

$$E_T\left[\sum_x \sum_{j\neq i} \alpha_j \overline{\alpha_i} \varphi^{T_j}(x)\overline{\varphi^{T_i}(x)}\right] = \sum_{j\neq i} \alpha_j \overline{\alpha_i} E_{T_i,T_j}\left[\sum_x \varphi^{T_j}(x)\overline{\varphi^{T_i}(x)}\right].$$

Hence, we should calculate

$$E_{T_i,T_j}\left[\sum_x \varphi^{T_j}(x)\overline{\varphi^{T_i}(x)}\right]. \tag{22}$$

Since the columns of the matrix form a group under poinwise multiplication, equation (22) can be rewritten as

$$\underset{j'\neq 0}{E_{T_{j'}}}\left[\sum_x \varphi^{T_{j'}}(x)\right]. \tag{23}$$

We can simply use the method of double counting for calculating equation (23). 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,

$$E_T\left[\sum_x\sum_{j\neq i}\alpha_j\overline{\alpha_i}\varphi^{T_j}(x)\overline{\varphi^{T_i}(x)}\right]=\frac{-N}{\mathcal{C}-1}\sum_{j\neq i}\alpha_j\overline{\alpha_i}.$$

Applying the Cauchy-Schwartz inequality, we obtain

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

which immediately implies

$$-\frac{k-1}{\mathcal{C}-1}||\boldsymbol{\alpha}||^2\leq E_T\left[\frac{1}{N}\sum_x\sum_{j\neq i}\alpha_j\overline{\alpha_i}\varphi^{T_j}(x)\overline{\varphi^{T_i}(x)}\right]\leq\frac{1}{\mathcal{C}-1}||\boldsymbol{\alpha}||^2. \tag{24}$$

□

Appendix B: Bounding the variance terms of Equation (6)

It follows from the Lemma 2.2 that

$$\frac{1}{\mathcal{C}(\mathcal{C}-1)}[\mathbf{1}^T(\sum_{\substack{g,h\in\mathcal{G}\\ g\neq h}} gh^{-1})]=\frac{N}{\mathcal{C}-1}, \tag{25}$$

and hence

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

The third term in (6) is given by

$$\frac{1}{N^2}\sum_{x,x'}\sum_{\substack{j\neq i\\ s\neq t}}\alpha_j\overline{\alpha_i}\alpha_s\overline{\alpha_t}\varphi^{T_j}(x)\overline{\varphi^{T_i}}(x)\varphi^{T_s}(x')\overline{\varphi^{T_t}}(x')$$

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

$$E_{j\neq i\neq s\neq t}\left[\sum_{x,x'}\varphi^{T_j}(x)\overline{\varphi^{T_i}}(x)\varphi^{T_s}(x')\overline{\varphi^{T_t}}(x')\right]$$

by rewriting as

$$\frac{1}{\mathcal{C}(\mathcal{C}-1)(\mathcal{C}-2)(\mathcal{C}-3)}\left[\mathbf{1}^T(\sum_{\substack{g,h\in\mathcal{G}\\ g\neq h}}gh^{-1})(\sum_{\substack{v,w\in\mathcal{G}\\ v\neq w,g,h\\ w\neq g,h}}(vw^{-1})^T\mathbf{1}\right] \tag{26}$$

Now since

$$\sum_{\substack{v,w\\ v\neq w}}vw^{-1}=-\mathcal{C}\mathbf{1}, \tag{27}$$

removing all terms $(g,*),(h,*),(*,g^{-1}),(*,h^{-1})$ from ((27)) 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. (26) becomes

$$\frac{1}{\mathcal{C}(\mathcal{C}-1)(\mathcal{C}-2)(\mathcal{C}-3)}\sum_{\substack{g,h\\ g\neq h}}\mathbf{1}^T(gh^{-1})\left(-(\mathcal{C}-4)\mathbf{1}+gh^{-1}+hg^{-1}\right)^T\mathbf{1}. \tag{28}$$

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 (28) can be written as

$$\frac{1}{\mathcal{C}(\mathcal{C}-1)(\mathcal{C}-2)(\mathcal{C}-3)}\left[N\mathcal{C}(\mathcal{C}-4)+\sum_{\substack{g,h\\ g\neq h}}\left(S_{g,h}^2+|S_{g,h}|^2\right)\right]. \tag{29}$$

We can apply the hypothesis on column sums and obtain

$$E_{\substack{g,h\sim G_{\mathcal{C}}\\ g\neq h}}\left[|S_{g,h}|^2\right]=E_{\substack{f\sim G_{\mathcal{C}}\\ f\neq 0}}\left[\left|\mathbf{1}^Tf\right|^2\right]=E_{\substack{f\sim G_{\mathcal{C}}\\ i\neq 0}}\left[\left|\sum_x\varphi^f(x)\right|^2\right]\leq N.$$

Hence (29) 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

$$E_T\left[\left|\frac{1}{N^2}\sum_{x,x'}\sum_{\substack{j\neq i\\ \neq s\neq t}}\alpha_j\overline{\alpha_i}\alpha_s\overline{\alpha_t}\varphi^{T_j}(x)\overline{\varphi^{T_i}}(x)\varphi^{T_s}(x')\overline{\varphi^{T_t}}(x')\right|\right]\leq\frac{3(k-1)^2}{N(\mathcal{C}-2)(\mathcal{C}-3)}||\boldsymbol{\alpha}||^4.$$

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

There are two subcases. The first subcase is

$$E_{\text{j,i,t distinct}}\left[\left(\sum_x\varphi^{T_j}(x)\overline{\varphi^{T_i}(x)}\right)\left(\sum_{x'}\varphi^{T_i}(x')\overline{\varphi^{T_t}(x')}\right)\right] \tag{30}$$

which we rewrite as

$$\frac{1}{\mathcal{C}(\mathcal{C}-1)(\mathcal{C}-2)}\sum_{h\in\mathcal{G}}\mathbf{1}^T\left(\sum_{g\neq h}gh^{-1}\right)\left(\sum_{w\neq g,h}hw^{-1}\right)^T\mathbf{1}. \tag{31}$$

Now

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

and so (31) is bounded above in absolute value by

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

which gives

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

The analysis of the second subcase

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

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

$$E_{j\neq i}\left[\left(\sum_x\varphi^{T_j}(x)\overline{\varphi^{T_i}(x)}\right)\left(\sum_{x'}\varphi^{T_i}(x')\overline{\varphi^{T_j}(x')}\right)\right] \tag{32}$$

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

$$\frac{\mathcal{C}(\mathcal{C}-1)}{\mathcal{C}(\mathcal{C}-1)}N. \tag{33}$$

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

$$E_T\left[\left|\frac{1}{N^2}\sum_{x,x'}\sum_{j\neq i}\alpha_j\overline{\alpha_i}\alpha_i\overline{\alpha_j}\varphi^{T_j}(x)\overline{\varphi^{T_i}}(x)\varphi^{T_i}(x')\overline{\varphi^{T_j}}(x')\right|\right] \leq \frac{(k-1)}{N}||\boldsymbol{\alpha}||^4. \tag{34}$$

The second subcase is

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

which we rewrite as

$$\frac{1}{\mathcal{C}(\mathcal{C}-1)}\sum_{\substack{g,h\\ g\neq h}}\mathbf{1}^T(hg^{-1})(hg^{-1})^T\mathbf{1}. \tag{35}$$

and since

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

the upper bound (34) stays the same.

Appendix C: Proof of Theorem 2.3

Let $f = \frac{1}{\sqrt{N}}\Phi\boldsymbol{\alpha}$. By Theorem 2.1 we have

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

We have also shown that

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

which implies

$$\begin{aligned} Var[||f||^2] &\leq ||\boldsymbol{\alpha}||^4 + \left(\frac{2k}{N}+\frac{9}{\mathcal{C}-3}\right)||\boldsymbol{\alpha}||^4 - ||\boldsymbol{\alpha}||^4 + \frac{2(k-1)}{\mathcal{C}-1}||\boldsymbol{\alpha}||^4 - \left(\frac{k-1}{(\mathcal{C}-1)}\right)^2||\boldsymbol{\alpha}||^4 \\ &\leq \left(\frac{2k}{N}+\frac{2k+7}{\mathcal{C}-3}\right)||\boldsymbol{\alpha}||^4. \end{aligned} \tag{36}$$

Now, using Chebyshev's inequality we have

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

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

$$\begin{aligned}\Pr\left[\left|||f||^2 - ||\boldsymbol{\alpha}||^2\right| \geq \delta||\boldsymbol{\alpha}||^2\right] &\leq \frac{\frac{2c_1}{\kappa+1} + \frac{2c_1N - 7\kappa + 7}{(\kappa+1)(c_2N^{\beta}-3)}}{\left(\delta - \frac{2c_1}{\kappa+1}\right)^2} \\ &\sim \frac{2c_1}{\delta^2(\beta-1)\log N} \quad \text{as } N \to \infty,\end{aligned}$$

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

Appendix D: Proof of Theorem 2.4

Suppose that in addition to the two simple conditions of Section 2, any two column sums of the sensing matrix are also sufficiently close. More precisely for any two columns i, j of the sensing matrix and some positive γ,

$$\left|\sum_x \varphi^i(x) - \sum_x \varphi^j(x)\right|^2 \leq N^{2-\gamma}. \tag{37}$$

We show that if $N = \Omega\left(k \log\left(\frac{\mathcal{C}}{k}\right)\right)$ and $\gamma = 1$ then the probability of failure of the StRIP property decays at a rate proportional to ${}^{k}/{}_{\mathcal{C}}$. Let $f(t_1, \cdots, t_k) = \frac{1}{\sqrt{N}} \sum_{i=1}^{k} \alpha_i \varphi^{t_i}$, and $h(t_1, \cdots, t_k) = \|f(t_1, \cdots, t_k)\|^2$. Clearly

$$h(t_1, \cdots, t_k) = \frac{1}{N} \sum_{i,j=1}^{k} \alpha_i \overline{\alpha_j} \left(\varphi^{t_i}\right)^{\top} \overline{\varphi^{t_j}}. \tag{38}$$

We have shown in Appendix A that

$$\mathbb{E}_T\left[h(T_1, \cdots, T_k)\right] \approx \|\boldsymbol{\alpha}\|^2. \tag{39}$$

Note that since Φ is unimodular, for each index i:

$$h(t_1, \cdots, t_i, \cdots, t_k) - h(t_1, \cdots, t_i', \cdots, t_k) = \frac{1}{N} \sum_{j \neq i} \left(\alpha_i \overline{\alpha_j} \left(\overline{\varphi^{t_i}} - \overline{\varphi^{t_i'}}\right)^{\top} \varphi^{t_j} - \alpha_j \overline{\alpha_i} \overline{\varphi^{t_j}}^{\top} \left(\varphi^{t_i} - \varphi^{t_i'}\right)\right) \tag{40}$$

Now since the columns form a group, the inner product between any two columns is another column sum. Hence using Condition (37) we get

$$\left|h(t_1, \cdots, t_i, \cdots, t_k) - h(t_1, \cdots, t_i', \cdots, t_k)\right| \leq \frac{2}{N}|\alpha_i| \sum_{j \neq i} |\alpha_j| \left|\sum_x \varphi^{z_i}(x) - \varphi^{z_i'}(x)\right| \leq \frac{2}{N^{\gamma/2}}|\alpha_i| \sum_{j \neq i} |\alpha_j|. \tag{41}$$

Hence using McDiarmid's inequality we get

$$\Pr_T\left[\left|\|f\|^2 - \mathbb{E}[\|f\|^2\right| \geq \epsilon]\right] \leq 2\exp\left(\frac{-2\epsilon^2 N^{\gamma}}{4\sum_i |\alpha_i|^2 \left(\sum_{j \neq i} |\alpha_j|\right)^2}\right). \tag{42}$$

Note that using Cauchy-Schwartz's inequality

$$\begin{aligned}\sum_{i=1}^{k}|\alpha_i|^2\left(\sum_{j\neq i}|\alpha_j|\right)^2 &= \sum_{i=1}^{k}|\alpha_i|^4+\sum_{i=1}^{k}|\alpha_i|^2\left(\sum_{j=1}^{k}|\alpha_j|\right)^2-2\sum_{i=1}^{k}|\alpha_i|^3\left(\sum_{j=1}^{k}|\alpha_j|\right) \\ &\leq \sum_{i=1}^{k}|\alpha_i|^4+k\left(\sum_{i=1}^{k}|\alpha_i|^2\right)^2-2\left(\sum_{i=1}^{k}|\alpha_i|^2\right)^2\leq k\|\boldsymbol{\alpha}\|_2^4.\end{aligned} \tag{43}$$

Hence for each positive δ, with probability at least $1-2\exp\left(\frac{-\delta^2 N^{\gamma}}{2k}\right)$ the following near-isometry holds:

$$(1-\delta)\|\boldsymbol{\alpha}\|^2\leq\|f\|^2\leq(1+\delta)\|\boldsymbol{\alpha}\|^2, \tag{44}$$

and if $N=k\log\left(\frac{\mathcal{C}}{k}\right)$ then the probability of failure decays at a rate proportional to $\frac{k}{\mathcal{C}}$.

Appendix E: Tight Frames

It follows from conditions (1) and (2) that the normalized columns $\frac{1}{\sqrt{N}}$ 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

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

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

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

Appendix F: Proof of Theorem 4.1

We use the following proposition proved by Fuchs, and Tropp [Fuc04], [Tro08b]

Proposition 2. *Let $\boldsymbol{x}^*$ be a complex sparse signal from the Steinhaus family, with support T. If for any column $w\notin T$ of the normalized sensing matrix Φ*

$$\left|\left(\Phi_T^{\dagger}\varphi^w\right)^{\top}sgn(\boldsymbol{x}^*)\right|<1, \tag{45}$$

then the basis pursuit algorithm of Equation (21) recovers the signal $\boldsymbol{x}^*$ successfully.

An analysis very similar to analysis of Theorem 3.2 shows that for a fixed column w, if the elements of T are chosen uniformly at random, if the matrix satisfies StRIP(δ) property, and also condition of Equation (12) holds then there exists a constant κ_δ such that except with exponentially small probability

$$\|\Phi_T\varphi^w\|^2<\kappa_\delta\sqrt{\frac{k}{N}}. \tag{46}$$

We use the Bernstein inequality to show that the sufficient conditions of Equation (45) holds and hence basis pursuit algorithm can recover the Steinhaus signal (cf. [Tro08b]).

Proposition 3 (Complex Bernstein Inequality)**.** *Let* $\boldsymbol{X} = [X_1, \cdots, X_k]^\top$ *be a vector of* k *i.i.d Steinhaus random variables, and* $\boldsymbol{a}$ *be any complex vector. Then for any positive* ρ

$$\Pr\left[\left|\boldsymbol{a}^\top \boldsymbol{X}\right| \geq \rho \|\boldsymbol{a}\|_2\right] \leq 2\exp\left(\frac{-\rho^2}{2}\right). \tag{47}$$

Theorem 4.3. *Let* $\boldsymbol{x}^*$ *be a* k*-sparse complex Steinhaus signal compressively sampled using* $\Omega\left(C_\delta k \log \mathcal{C}\right)$ *measurements from a deterministic matrix with StRIP(*δ*). Then with overwhelming probability, the basis pursuit algorithm of Equation (21) recovers* $\boldsymbol{x}^*$ *successfully.*

Proof: We show that (45) hold with high probability. First fix a column w. Since $\boldsymbol{x}^*$ is Steinhaus, and considering Equation (46), we can apply the Bernstein inequality with $\boldsymbol{a} = \Phi_T^\dagger \varphi^w$, and $\boldsymbol{X} = sgn(\boldsymbol{x}^*)$. Hence

$$\Pr\left[\left|\left(\Phi_T^\dagger \varphi^w\right)^\top sgn(\boldsymbol{x}^*)\right| \geq 1\right] \leq 2\exp\left(\frac{-N}{\kappa^2 k}\right).$$

Now, adding the total probability of failure of Equation (46) for a fixed w and applying the union bounds to all $\mathcal{C}$ possible choices of w we get

$$\Pr\left[\exists\ w: \ \ \left|\left(\Phi_T^\dagger \varphi^w\right)^\top sgn(\boldsymbol{x}^*)\right| \geq 1\right] \leq 3\mathcal{C}\exp\left(\frac{-N}{\kappa^2 k}\right). \tag{48}$$

This probability drops to zero exponentially fast, provided that $N = \Omega\left(k \log \mathcal{C}\right)$. □