SPARSE RECONSTRUCTION FROM HADAMARD MATRICES: A LOWER BOUND

JAROSLAW BLASIOK, PATRICK LOPATTO, KYLE LUH, AND JAKE MARCINEK

Abstract. We give a short argument that yields a new lower bound on the number of subsampled rows from a bounded, orthonormal matrix necessary to form a matrix with the restricted isometry property. We show that for a \( N \times N \) Hadamard matrix, one cannot recover all \( k \)-sparse vectors unless the number of subsampled rows is \( \Omega(k \log k \log(N/k)) \), whenever \( \min(k, N/k) > \log^2(N) \).

1. Introduction

In their seminal work on sparse recovery [5], Candes and Tao were led to the notion of the restricted isometry property (RIP). A \( q \times N \) matrix \( M \) has the restricted isometry property of order \( k \) with constant \( \delta > 0 \) if for all \( k \)-sparse vectors \( x \in \mathbb{C}^N \),

\[
(1 - \delta)\|x\|^2_2 \leq \|Mx\|^2_2 \leq (1 + \delta)\|x\|^2_2.
\]

The importance of this property is that it guarantees that one can recover a \( k \)-sparse vector \( x \) from \( Mx \) via a convex program [5]. In applications, \( q \) is the number of measurements needed to recover a sparse signal (with \( \delta \) typically a small constant). Therefore, it is of interest to understand the minimal number of rows needed in a matrix with the RIP property.

It is known that for a properly normalized matrix of gaussian random variables, \( q = \Omega(k \log(N/k)) \) suffices to generate a RIP matrix with high probability (e.g. [8]). Yet, it is often beneficial to have more structure in the matrix \( M \) [13]. For example, if the matrix \( M \) is a submatrix of the discrete Fourier transform matrix, then the fast Fourier transform algorithm allows fast matrix–vector multiplication, speeding up the run time of the recovery algorithm [8, Chapter 12]. Additionally, generating a random submatrix requires fewer random bits and less storage space.

The first bound on the number of subsampled rows from a Fourier matrix necessary for recovery appeared in the groundbreaking work [5]. They show that if one randomly subsamples rows so that the expected number of rows is \( \Omega(k \cdot \log^6 N) \), then concatenating these rows forms a RIP matrix with high probability. Rudelson and Vershynin later improved this bound to \( \Omega(k \cdot \log^2 k \cdot \log(k \log N) \cdot \log N) \) via a gaussian process argument involving chaining techniques [14]. Their proof was then streamlined and their probability bounds strengthened [7, 13]. Cheraghchi, Gurusswami, and Velingker then proved a bound of \( \Omega(k \cdot \log^3 k \cdot \log N) \) [6], and Bourgain established the bound \( \Omega(k \cdot \log^2 k \cdot \log^2 N) \) [4]. The sharpest result in this direction is due to Haviv and Regev, who showed the upper bound \( O(k \cdot \log^2 k \cdot \log N) \) through a delicate application of the probabilistic method [10]. It is widely conjectured that for the discrete Fourier transform \( q = \Omega(k \log N) \) suffices. We note that most proofs in this line of work, including the strongest known upper bound [10], do not use the Fourier structure in an essential way and in fact apply to all bounded orthonormal matrices.

Date: March 29, 2019.
This paper addresses the problem of lower bounding \( q \), in other words determining a necessary number of samples for reconstruction. Our contribution is that surprisingly, for general bounded orthonormal matrices, and for a certain range of \( k \), \( \Omega(k \log^2 N) \) samples are needed. In particular, only a gap of \( \log k \) remains between our bound and the best known upper bound. We improve the previous best lower bound \( \Omega(k \cdot \log(N/k)) \) due to Bandeira, Lewis, and Mixon \cite{bandeira2016structured}, which in turn improved previous work establishing a \( \Omega(k \cdot \log(N/k)) \) lower bound \cite{blasiok2016generalized,blasiok2016improved,blasiok2016improved2,blasiok2016improved3}.

The proof constructs an example of a bounded orthonormal matrix, the Hadamard matrix, that requires \( \Omega(k \log k \log N/k) \) samples. We interpret the Hadamard matrix as the Fourier transform on the additive group \( \mathbb{Z}_N^2 \). By a second moment argument, we demonstrate that for fewer than \( O(k \log k \log N/k) \) subsampled rows, there exists a \( k \)-sparse vector in the kernel.

**Remark 1.1.** Shravas Rao has simultaneously and independently proved a similar result and we refer the reader to his forthcoming preprint for the details.

**Acknowledgments.** P.L. has been partially supported by the NSF Graduate Research Fellowship Program under grant DGE-1144152. K.L. has been partially supported by NSF postdoctoral fellowship DMS-1702533.

### 2. Preliminaries

Throughout this note, we use \( \log \) to denote the base 2 logarithm. For an integer \( n \geq 1 \), we set \( N = 2^n \) and fix a bijection between \([N]\) and \( \mathbb{Z}_2^n \); this identification remains in force for the rest of the paper.

We say a function \( \chi : \mathbb{Z}_2^n \to \{\pm 1\} \) is a character if it is a group homomorphism. To an element \( a \in \mathbb{Z}_2^n \), we associate the character

\[
\chi_a(x) = (-1)^{\langle a, x \rangle}
\]

for all \( x \in \mathbb{Z}_2^n \). The Fourier transform of a function \( f : \mathbb{Z}_2^n \to \mathbb{C} \) is defined to be

\[
\hat{f}(a) = \frac{1}{\sqrt{N}} \sum_{x \in \mathbb{Z}_2^n} f(x) \chi_a(x)
\]

for all \( a \in \mathbb{Z}_2^n \). Let \( H \) be the \( N \times N \) matrix representing the Fourier transform on the group \( \mathbb{Z}_2^n \). In other words,

\[
H_{ij} = \frac{1}{\sqrt{N}}(-1)^{\sum_{k=1}^{n} i_k j_k}.
\]

When normalized to have \( \pm 1 \) entries, the matrix \( H \) is also known as a Hadamard matrix. We refer the reader to \cite{du2000hadamard} for a thorough discussion of Fourier analysis on finite groups.

The Grassmannian \( \mathbb{G}_{n,d} = \mathbb{G}_{n,d}(\mathbb{Z}_2) \) is defined as the collection of vector subspaces of \( \mathbb{Z}_2^n \) of dimension \( d \). Our proof uses the following well-known result about the Fourier transform.

**Lemma 2.1.** For a subspace \( V \in \mathbb{G}_{n,d} \), we let \( 1_V \in \mathbb{R}^N \) be the vector corresponding to the indicator function for \( V \) with the normalization \( \|1_V\|_2 = 1 \). Then

\[
H 1_V = 1_{V^\perp}.
\]

where \( V^\perp \) is the orthogonal complement of \( V \).

In this way, \( H \) implements a bijection between \( \mathbb{G}_{n,d} \) and \( \mathbb{G}_{n,n-d} \). We also make use of the following bounds on the size of \( \mathbb{G}_{n,d} \).
Lemma 2.2. The size of $\mathbb{G}_{n,d}$ is bounded by
\[ 2^{d(n-d-1)} < |\mathbb{G}_{n,d}| < 2^{d(n-d)}. \] (2.1)

Proof. A standard counting argument gives the explicit formula
\[ |\mathbb{G}_{n,d}| = \prod_{k=0}^{d-1} \frac{2^n - 2^k}{2^d - 2^k}. \] (2.2)

Using the inequalities
\[ 2^{n-d-1} < \frac{2^n - 2^k}{2^d - 2^k} < 2^{n-d} \] (2.3)
on each factor individually gives the result. \qed

We also make use of the following trivial counting lemma.

Lemma 2.3. For $U, V \in \mathbb{G}_{n,m}$,
\[ \max(n - 2m, 0) \leq \dim(U^\perp \cap V^\perp) \leq n - m. \]

3. Main Result

For a subset $Q \subset [N]$, we let $H_Q$ denote the matrix generated from the rows of $H$ indexed by $Q$. Let $\delta_1, \ldots, \delta_N$ be a set of independent Bernoulli random variables which take the value 1 with probability $p$. These random variables will indicate which rows to include in our measurement matrix, $H_Q$, meaning
\[ Q = \{ j \in [N] : \delta_j = 1 \}. \]

Note that $Q$ has average cardinality $Np$ and standard concentration arguments can be used to obtain sharp bounds on its size. We say a vector $v \in \mathbb{R}^N$ is $k$-sparse if it has at most $k$ nonzero entries. The following theorem is our main technical result.

Theorem 3.1. For $\min(m, n-m) \geq 16 \log n$, where $N = 2^n$ and $k = 2^m$, there exists a positive constant $C > 0$ such that for $p \leq \frac{2^k \log N}{N} \log N/k$, there exists a $k$-sparse vector in the kernel of $H_Q$ with probability $1 - o(1)$.\(^1\)

Proof. For convenience, we assume that $k = 2^m$ is a power of 2 in what follows, although this is not essential to the argument.

We restrict our attention to the $k$-sparse vectors that correspond to $1_V$ for $V \in \mathbb{G}_{n,m}$, the indicator functions of subspaces of dimension $m$. For such $V$, set $X_V$ to be the indicator function for the event that $Q \cap V^\perp = \emptyset$. Define
\[ X = \sum_{V \in \mathbb{G}_{n,m}} X_V. \] (3.1)

Observe that by Lemma 2.1, if $X$ is non-zero then there exists a $k$-sparse vector in the kernel of $H_Q$. We proceed by the second moment method to show that $X$ is nonzero with high probability. By the second moment method (e.g. [1]),
\[ \mathbb{P}(X = 0) \leq \frac{\text{Var} X}{(\mathbb{E} X)^2}. \] (3.2)

\(^1\) $o(1)$ indicates a quantity that tends to zero as $N \to \infty$. All asymptotic notation is applied under the assumption that $N \to \infty$. 

We can easily obtain an expression for the first moment:

\[ EX = |G_{n,m}| \cdot EX \]
\[ = |G_{n,m}|(1-p)^{|V^\perp|} \]
\[ \geq |G_{n,m}|(1-p)^{N/k}. \quad (3.3) \]

The second moment requires a more delicate calculation. We partition the sum into pairs of orthogonal complements with the same dimension of intersection. By Lemma 2.3, and letting \( d_0 \) denote \( \max(n-2m,0) \), we have

\[
\frac{\text{Var} X}{(EX)^2} = \frac{\sum_{d=d_0}^{n-m} \sum_{U,V:\dim(U \cap V^\perp) = d} \text{Cov}(X_U X_V)}{|G_{n,m}|^2 (1-p)^{2N/k}}
\]
\[
= \frac{\sum_{d=d_0}^{n-m} \sum_{U,V:\dim(U \cap V^\perp) = d} E(X_U X_V) - EX \cdot EX_{X_V}}{|G_{n,m}|^2 (1-p)^{2N/k}}
\]
\[
= \frac{\sum_{d=d_0}^{n-m} \sum_{U,V:\dim(U \cap V^\perp) = d} E(X_U X_V) - EX \cdot EX_{X_V}}{|G_{n,m}|^2 (1-p)^{2N/k}}
\]
\[
= \frac{\sum_{d=d_0}^{n-m} \sum_{U,V:\dim(U \cap V^\perp) = d} (1-p)^{2N/k - 2d} - (1-p)^{2N/k}}{|G_{n,m}|^2 (1-p)^{2N/k}}
\]
\[
\leq \sum_{d=d_0}^{n-m} \min\{\frac{|G_{n,d}|}{|G_{n,m}|^2}, |G_{n,m}| \} \left( \exp(p \cdot 2^d) - 1 \right).
\]

To obtain more manageable notation, we define

\[ S(d) = \max\{2^{d(n-d)+2m(n-m-d)}, |G_{n,m}|^2\}. \]

Therefore, we have

\[
\frac{\text{Var} X}{(EX)^2} \leq \sum_{d=d_0}^{n-m} \frac{S(d)}{|G_{n,m}|^2} \left( \exp(p \cdot 2^d) - 1 \right)
\]
\[
= \sum_{d=d_0}^{n-m-4\log n} \frac{S(d)}{|G_{n,m}|^2} \left( \exp(p \cdot 2^d) - 1 \right)
\]
\[
+ \sum_{d=n-m-4\log n+1}^{n-m} \frac{S(d)}{|G_{n,m}|^2} \left( \exp(p \cdot 2^d) - 1 \right)
\]
\[ := (I) + (II). \]

We handle each sum separately.

For \( d_0 \leq d \leq n-m-4\log n \),

\[ p \cdot 2^d \leq \frac{Ck \log^2 N}{N} 2^{n-m-4\log n} = o(1/n) \]

and we have the simple bound \( S(d)/|G_{n,m}|^2 \leq 1 \). Therefore,

\[ (I) \leq \sum_{d=d_0}^{n-m-4\log n} \left( \exp(p \cdot 2^d) - 1 \right) = o(1). \]
In the range \( d > n - m - 4 \log n \),
\[
\frac{S(d)}{|\mathbb{G}_{n,m}|^2} \leq \frac{S(d)}{2^{2m(n-m-1)}}.
\]
In the definition of \( S(d) \), exponent \( d(n-d) + 2m(n-m-d) \) is a quadratic function in \( d \) with negative leading term. Optimizing in \( d \), we find that \( S(d) \) is maximized at \( d = n - 2m/2 = n - m - 4 \log n \). Therefore, when \( d > n - m - 4 \log n \), using \( \min(m, n-m) \geq 16 \log n \) we have
\[
\log S(d) \leq (n - m)(m + 4 \log n) + 4m \log n
\]
\[
\leq \frac{3}{2}(n - m)m.
\]
This implies
\[
\frac{S(d)}{|\mathbb{G}_{n,m}|^2} \leq \frac{2^{2(n-m)(3m/2)}}{2^{2m(n-m-1)}} = 2^{-cm(n-m-1)}.
\]
Thus, we can conclude that
\[
(II) \leq \sum_{d=n-m-4 \log n+1}^{n-m} 2^{-cm(n-m-1)} \left( \exp(p \cdot 2^d) - 1 \right)
\]
\[
\leq 4(\log n)2^{-cm(n-m-1)} \exp(p \cdot 2^{n-m}) = o(1)
\]
where the last line follows from \( p \leq \frac{Ck \log k \log N/k}{N} \), i.e. \( p \cdot 2^{n-m} \leq Cm(n-m) \), where \( C \) is small enough constant depending on \( c \) (we can set \( C = c/2 \)).

We can now state our main result in terms of sparse recovery.

**Theorem 3.2.** Let \( N \) and \( k \) be as in Theorem 3.1. For there to exist a method to recover every \( k \)-sparse vector from \( H_Q \), for any \( k \) such that \( \min(k, N/k) \geq \log C'N \), the expected cardinality of the number of rows of \( H_Q \) must be \( \Omega(k \log k \log (N/k)) \). Further, for any constant \( \delta > 0 \), the expected number of rows of \( H_Q \) must be \( \Omega(k \log k \log (N/k)) \) for \( H_Q \) to have the RIP property.

**Proof.** By Theorem 3.1, there exists a \( 2k \)-sparse vector \( x \) in the kernel of \( H_Q \) with high probability if the expected number of rows of \( H_Q \) is \( o(k \log^2 N) \). Let us write \( x = y - z \) where \( y \) and \( z \) are both \( k \)-sparse vectors. Then \( H_Q y = H_Q z \), which proves that one cannot distinguish all \( k \)-sparse vectors. The statement about the RIP property follows directly from the definition — existence of a \( k \) sparse vector in the kernel precludes \((k, \delta)\)-RIP property for any \( \delta \).

**References**

[1] Noga Alon and Joel H. Spencer, *The probabilistic method*, fourth ed., Wiley Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2016. MR 3524748

[2] Khanh Do Ba, Piotr Indyk, Eric Price, and David P Woodruff, *Lower bounds for sparse recovery*, Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms, SIAM, 2010, pp. 1190–1197.

[3] Afonso S. Bandeira, Megan E. Lewis, and Dustin G. Mixon, *Discrete uncertainty principles and sparse signal processing*, Journal of Fourier Analysis and Applications (2017), 1–22.
[4] Jean Bourgain, *An improved estimate in the restricted isometry problem*, Geometric aspects of functional analysis, Lecture Notes in Math., vol. 2116, Springer, Cham, 2014, pp. 65–70. MR 3364679

[5] Emmanuel J. Candes and Terence Tao, *Near-optimal signal recovery from random projections: Universal encoding strategies?*, IEEE transactions on information theory 52 (2006), no. 12, 5406–5425.

[6] Mahdi Cheraghchi, Venkatesan Guruswami, and Ameya Velingker, *Restricted isometry of Fourier matrices and list decodability of random linear codes*, SIAM J. Comput. 42 (2013), no. 5, 1888–1914. MR 3108113

[7] Sjoerd Dirksen, *Tail bounds via generic chaining*, Electronic Journal of Probability 20 (2015).

[8] Simon Foucart and Holger Rauhut, *A mathematical introduction to compressive sensing*, Applied and Numerical Harmonic Analysis, Birkhäuser/Springer, New York, 2013. MR 3100033

[9] Andrej Y. Garnaev and Efim D. Gluskin, *On widths of the euclidean ball*, Soviet Mathematics Doklady, vol. 277, 1984, pp. 1048–1052.

[10] Ishay Haviv and Oded Regev, *The restricted isometry property of subsampled fourier matrices*, Geometric Aspects of Functional Analysis, Springer, 2017, pp. 163–179.

[11] Boris Sergeevich Kashin, *Diameters of some finite-dimensional sets and classes of smooth functions*, Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya 41 (1977), no. 2, 334–351.

[12] Jelani Nelson and Huy L. Nguyen, *Sparsity lower bounds for dimensionality reducing maps*, Proceedings of the forty-fifth annual ACM symposium on Theory of computing, ACM, 2013, pp. 101–110.

[13] Holger Rauhut, *Compressive sensing and structured random matrices*, Theoretical foundations and numerical methods for sparse recovery 9 (2010), 1–92.

[14] Mark Rudelson and Roman Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Communications on Pure and Applied Mathematics 61 (2008), no. 8, 1025–1045.

[15] Audrey Terras, *Fourier analysis on finite groups and applications*, London Mathematical Society Student Texts, vol. 43, Cambridge University Press, Cambridge, 1999. MR 1695775

(Jaroslaw Blasiok) JOHN A. PAULSON SCHOOL OF ENGINEERING AND APPLIED SCIENCES, HARVARD UNIVERSITY

E-mail address: jblasio@g.harvard.edu

(Patrick Lopatto) DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY

E-mail address: lopatto@math.harvard.edu

(Kyle Luh) CENTER OF MATHEMATICAL SCIENCES AND APPLICATIONS, HARVARD UNIVERSITY

E-mail address: kluh@cmsa.fas.harvard.edu

(Jake Marcinek) DEPARTMENT OF MATHEMATICS, HARVARD UNIVERSITY

E-mail address: marcinek@math.harvard.edu