AN IMPROVED LOWER BOUND FOR SPARSE RECONSTRUCTION FROM SUBSAMPLED HADAMARD MATRICES

JAROSŁAW BŁASIOK, PATRICK LOPATTO, KYLE LUH, JAKE MARCINEK, AND SHRAVAS RAO

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 a matrix formed by uniformly subsampling rows of an $N \times N$ Hadamard matrix contains a K-sparse vector in the kernel, unless the number of subsampled rows is $\Omega(K \log K \log(N/K))$ — our lower bound applies whenever $\min(K, N/K) > \log^C N$. Containing a sparse vector in the kernel precludes not only the restricted isometry property, but more generally the application of those matrices for uniform sparse recovery.

1. Introduction

In their seminal work on sparse recovery [5], Candés 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 $x \in \mathbb{C}^N$ (i.e. vectors with at most K nonzero entries) we have

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

The significance of this property is that it guarantees that one can recover an approximately K-sparse vector x^* from Mx^* via a convex program [5]. Specifically, they showed that if a matrix M satisfies $(2K, \sqrt{2}-1)$-RIP, then the minimizer

$$\tilde{x} := \arg\min_{x: Mx = Mx^*} \|x\|_1,$$

satisfies

$$\|\tilde{x} - x^*\|_2 \leq \frac{1}{\sqrt{k}} \|x^* - x_K^*\|_1,$$

where x_K^* is the best K-sparse approximation of x^* — in particular when x^* is exactly K-sparse, it can be efficiently recovered from Mx^* without any error.

In applications, q is the number of measurements needed to recover a sparse signal. 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 with independent gaussian entries, $q = \mathcal{O}(K \log(N/K))$ 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.

Date: April 8, 2019.

The first bound on the number of subsampled rows from a Fourier matrix necessary for recovery appeared in the groundbreaking work [5]. They showed that if one randomly subsamples rows so that the expected number of rows is $\mathcal{O}(K \cdot \log^6 N)$, then concatenating these rows forms a RIP matrix with high probability, after appropriate normalization. Rudelson and Verhsynin later improved this bound to $\mathcal{O}(K \cdot \log^2 K \cdot \log(K \log N) \cdot \log N)$ via a gaussian process argument involving chaining techniques [14]. Their proof was then streamlined and their probability bounds strengthened [7, 13]. Cheraghchi, Guruswami, and Velingker then proved a bound of $\mathcal{O}(K \cdot \log^3 K \cdot \log N)$ [6], and Bourgain established the bound $\mathcal{O}(K \cdot \log K \cdot \log^2 N)$ [4]. The sharpest result in this direction is due to Haviv and Regev, who showed the upper bound $\mathcal{O}(K \cdot \log^2 K \cdot \log N)$ through a delicate application of the probabilistic method [10]. It is widely conjectured that for the discrete Fourier transform $q = \mathcal{O}(K \log N)$ suffices [14].

It turns out that all proofs in this line of work, including the strongest known upper bound [10], apply in a more general setting where random matrix M is constructed by subsampling rows of any *bounded orthonormal matrix* — that is an orthonormal matrix with all entries bounded in magnitude by $\frac{B}{\sqrt{N}}$ for some constant B. The matrix of the Discrete Fourier Transform satisfy this property with $B = 1$.

This paper addresses the problem of 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 $\Omega(K \cdot \log N)$ due to Bandeira, Lewis, and Mixon [3] which applied to the DFT matrix. Those in turn improve upon more general lower bounds $\Omega(K \cdot \log(N/K))$ on the number of rows for *any* matrix that satisfies the RIP property [2, 9, 11, 12].

In the proof we consider an example of a bounded orthonormal matrix, the Hadamard matrix (i.e. the matrix of the Fourier transform on the additive group $\mathbb{Z}_2^n$), and we show that for this specific matrix at least $\Omega(K \log K \log N/K)$ samples is required. More concretely, by a second moment argument, we demonstrate that for fewer than $\mathcal{O}(K \log K \log N/K)$ subsampled rows, with high probability there exists a K-sparse vector in the kernel — ruling out both the RIP property, and in general any hope for sparse recovery algorithm with those matrices. The same proof can be applied more generally to show that for any prime r one needs to subsample at least $\Omega(K \log K \log(N/K)/\log(r))$ rows of a matrix corresponding to Fourier transform on the additive group $\mathbb{Z}_r^n$ — for the sake of simplicity of the argument we do not elaborate on this.

Acknowledgments. J.B. has been supported by Siebel Foundation. 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. S.R. has been supported by the Simons Collaboration on Algorithms and Geometry.

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 ± 1 entries, the matrix H is also known as a Hadamard matrix. We refer the reader to [15] 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 $\mathbb{1}_V \in \mathbb{R}^N$ be the vector corresponding to the indicator function for V with the normalization $\|\mathbb{1}_V\|_2 = 1$. Then*

$$H\mathbb{1}_V = \mathbb{1}_{V^\perp}.$$

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*

$$2^{d(n-d)} < |\mathbb{G}_{n,d}| < 2^{d(n-d+1)}. \tag{2.1}$$

Proof. A standard counting argument gives the explicit formula

$$|\mathbb{G}_{n,d}| = \prod_{k=0}^{d-1} \frac{2^n - 2^k}{2^d - 2^k}. \tag{2.2}$$

Using the inequalities

$$2^{n-d} < \frac{2^n - 2^k}{2^d - 2^k} < 2^{n-d+1} \tag{2.3}$$

on each factor individually gives the result. □

We also make use of the following trivial counting lemma.

Lemma 2.3. *For $U, V \in \mathbb{G}_{n,k}$,*

$$\max(n - 2k, 0) \leq \dim(U^\perp \cap V^\perp) \leq n - k.$$

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 $\delta_1, \dots, \delta_N$ be a set of independent Bernoulli random variables which take the value 1 with probability $\hat{p}$. Random variables δ_i 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 $N\hat{p}$ and standard concentration arguments can be used to obtain sharp bounds on its size. We say that 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 $\min(k, n-k) \geq 12 \log n$, where $N = 2^n$ and $K = 2^k$, there exists a positive constant $c > 0$ such that for $\hat{p} \leq \frac{cK}{N} \log K \log(N/K)$, there exists a K-sparse vector in the kernel of H_Q with probability $1 - o(1)$.*[1]

Proof. We will define $p := -\ln(1-\hat{p})$ for future convenience, and note that $\hat{p} \leq p \leq 2\hat{p}$, for small enough $\hat{p}$.

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

$$X = \sum_{V \in \mathbb{G}_{n,k}} X_V. \tag{3.1}$$

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]),

$$\mathbb{P}(X = 0) \leq \frac{\operatorname{Var} X}{(\mathbb{E}X)^2}. \tag{3.2}$$

We can easily obtain an expression for the first moment:

$$\begin{aligned}
\mathbb{E}X &= |\mathbb{G}_{n,k}| \cdot \mathbb{E}X_V \\
&= |\mathbb{G}_{n,k}| \, (1-\hat{p})^{|V^\perp|} \\
&= |\mathbb{G}_{n,k}| \exp(-p\frac{N}{K}) \\
&\geq \exp((\ln 2 - 2c)k(n-k)).
\end{aligned}$$

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

$$\begin{aligned}
\frac{\operatorname{Var} X}{(\mathbb{E}X)^2} &= \frac{\sum_{U,V \in \mathbb{G}_{n,k}} \operatorname{Cov}(X_U, X_V)}{|\mathbb{G}_{n,k}|^2 \, (\mathbb{E}X_U)^2} \\
&= \frac{\sum_{d=d_0}^{n-k} \sum_{U,V : \dim(U^\perp \cap V^\perp) = d} \operatorname{Cov}(X_U X_V)}{|\mathbb{G}_{n,k}|^2 \, (\mathbb{E}X_U)^2}.
\end{aligned} \tag{3.3}$$

We can explicitly compute each term in the sum above as follows.

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

Claim 3.2. *For $U, V \in \mathbb{G}_{n,k}$ such that $\dim(U^\perp \cap V^\perp) = d$, we have*

$$\frac{Cov(X_U, X_V)}{(\mathbb{E}X_U)^2} = \exp(p2^d) - 1.$$

Proof. Observe that

$$\begin{aligned}\mathbb{E}X_U X_V &= \mathbb{P}(U^\perp \cap Q = \emptyset \wedge V^\perp \cap Q = \emptyset)\\ &= \exp(-p|U^\perp \cup V^\perp|)\\ &= \exp(-2p|U^\perp| + p|U^\perp \cap V^\perp|)\\ &= (\mathbb{E}X_U)^2 \exp(p2^d).\end{aligned}$$

□

We plug this expression back to the sum (3.3), in order to arrive at

$$\frac{\operatorname{Var} X}{(\mathbb{E}X)^2} = \sum_{d=d_0}^{n-k} \sum_{U,V:\dim(U^\perp \cap V^\perp)=d} \frac{1}{|\mathbb{G}_{n,k}|^2} \left(\exp(p2^d) - 1\right).$$

Let us use $T(n,k,d)$ to denote number of pairs $U, V \in \mathbb{G}_{n,k}$ such that $\dim(U^\perp \cap V^\perp) = d$. With this notation, the entire sum simplifies to

$$\sum_{d=d_0}^{n-k} \frac{T(n,k,d)}{|\mathbb{G}_{n,k}|^2} \left(\exp(p2^d) - 1\right).$$

We will split this sum into two parts and bound them separately

$$\begin{aligned}&\sum_{d=d_0}^{n-k-3\log n} \frac{T(n,k,d)}{|\mathbb{G}_{n,k}|^2} \left(\exp(p2^d) - 1\right)\\ &\quad + \sum_{d=n-k-3\log n}^{n-k} \frac{T(n,k,d)}{|\mathbb{G}_{n,k}|^2} \left(\exp(p2^d) - 1\right)\\ &=: (I) + (II).\end{aligned}$$

The first part of the summation is easy to control: for $d < n-k-3\log n$ we have $p2^d \le \frac{2c}{n}$, which implies $\exp(p2^d) - 1 \le \frac{4c}{n}$, and

$$\begin{aligned}(I) &\le \sum_{d=d_0}^{n-k-3\log n} \frac{T(n,k,d)}{|\mathbb{G}_{n,k}|^2} \frac{4c}{n}\\ &\le \frac{4c}{n} \sum_{d} \frac{T(n,k,d)}{|\mathbb{G}_{n,k}|^2}\\ &= \frac{4c}{n} = o(1).\end{aligned} \tag{3.4}$$

We can now turn our attention to bounding (II).

Claim 3.3. *For $d \ge n-k-3\log n$, we have*

$$\frac{T(n,k,d)}{|\mathbb{G}_{n,k}|^2} \le \exp\left(-\frac{\ln(2)}{2} k(n-k)\right).$$

Proof. First, we have the bound $T(n,k,d) \le |\mathbb{G}_{n,d}||\mathbb{G}_{n-d,n-k-d}|^2$. Indeed, to choose two subspaces $U^\perp, V^\perp$ of dimension k with $\dim(U^\perp \cap V^\perp) = d$, we can first choose $T = U^\perp \cap V^\perp$ as a subspace of $\mathbb{F}_2^n$ (there are $|\mathbb{G}_{n,d}|$ ways of doing this), and then we can consider the quotient space $\mathbb{F}_2^n/T$ and count the number of disjoint subspaces $U/T, V/T \subset \mathbb{F}_2^n/T$ — the number of such choices is upper bounded by $|\mathbb{G}_{n-d,n-k-d}|^2$ — the number of all pairs of subspaces $U/T, V/T \in \mathbb{F}_2^n/T$.

Applying Lemma 2.2 to $|\mathbb{G}_{n,d}|$ and $|\mathbb{G}_{n-d,n-k-d}|$, we obtain

$$T(n,k,d) \le \exp\left(\ln(2)\left[d(n-d+1) + 2(n-k-d+1)(k+1)\right]\right).$$

The quadratic in the exponent is maximized for $d = \frac{n-2k-1}{2}$, hence in the range $d \ge n-k-3\log n$, the maximum is attained exactly at $d = n-k-3\log n$. This yields

$$\begin{aligned} T(n,k,d) &\le \exp\left(\ln(2)\left[(n-k-3\log n)(k+3\log n+1) + 2(3\log n+1)(k+1)\right]\right) \\ &\le \exp\left(\ln(2)\left[(n-k)(\frac{5}{4}k) + \frac{1}{4}(n-k)k\right]\right) \\ &\le \exp\left(\ln(2)\left[\frac{3}{2}(n-k)k\right]\right), \end{aligned}$$

where the second inequality follows from the fact that $\min(k, n-k) \ge 12\log n$.

On the other hand, using Lemma 2.2 again, we have $\frac{1}{|\mathbb{G}_{n,k}|^2} \le 2^{-2k(n-k)}$ and the statement of the claim follows by combining these two inequalities. □

Now we can introduce a simple upper bound of the sum (II)

$$\begin{aligned} \sum_{d=n-k-3\log n}^{n-k} \frac{T(n,k,d)}{|\mathbb{G}_{n,k}|^2}\left(\exp(p2^d)-1\right) &\le 3(\log n)2^{-\frac{1}{2}k(n-k)}\exp\left(p2^{n-k}\right) \\ &\le 3(\log n)2^{(\frac{2c}{\ln(2)}-\frac{1}{2})k(n-k)} \\ &\le o(1), \end{aligned} \tag{3.5}$$

where the first inequality follows from Claim 3.3, the second one follows from $p2^{n-k} < 2ck(n-k)$, and the third can one be applied as soon as $\frac{2c}{\ln(2)} < \frac{1}{2}$. The statement of the Theorem now follows by combining (3.2), (3.4) and (3.5). □

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

Theorem 3.4. *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 $\min(K, N/K) \ge \log^C N$, 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 $o(K\log K\log(N/K))$. 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 H_Q is not injective when restricted to the set of all K-sparse vectors. The statement about the RIP property follows directly from the definition. □

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

(Jarosław Błasiok) John A. Paulson School of Engineering and Applied Sciences, Harvard University
E-mail address: jblasiok@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

(Shravas Rao) Courant Institute, New York University
E-mail address: rao@cims.nyu.edu