

# ON THE LINEAR INDEPENDENCE OF SPIKES AND SINES

JOEL A. TROPP

ABSTRACT. The purpose of this work is to survey what is known about the linear independence of spikes and sines. The paper provides new results for the case where the locations of the spikes and the frequencies of the sines are chosen at random. This problem is equivalent to studying the spectral norm of a random submatrix drawn from the discrete Fourier transform matrix. The proof depends on an extrapolation argument of Bourgain and Tzafriri.

## 1. INTRODUCTION

An investigation central to sparse approximation is whether a given collection of impulses and complex exponentials is linearly independent. This inquiry appears in the early paper of Donoho and Stark on uncertainty principles [DS89], and it has been repeated and amplified in the work of subsequent authors. Indeed, researchers in sparse approximation have developed a much deeper understanding of general dictionaries by probing the structure of the unassuming dictionary that contains only spikes and sines.

The purpose of this work is to survey what is known about the linear independence of spikes and sines and to provide some new results on random subcollections chosen from this dictionary. The method is adapted from a paper of Bourgain–Tzafriri [BT91]. The advantage of this approach is that it avoids some of the complicated combinatorial arguments that are used in related works, e.g., [CRT06]. The proof also applies to other types of dictionaries, although we do not pursue this line of inquiry here.

**1.1. Spikes and Sines.** Let us shift to formal discussion. We work in the inner-product space  $\mathbb{C}^n$ , and we use the symbol  $*$  for the conjugate transpose. Define the Hermitian inner product  $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{x}$  and the  $\ell_2$  vector norm  $\|\mathbf{x}\| = |\langle \mathbf{x}, \mathbf{x} \rangle|^{1/2}$ . We also write  $\|\cdot\|$  for the spectral norm, i.e., the operator norm for linear maps from  $(\mathbb{C}^n, \ell_2)$  to itself.

We consider two orthonormal bases for  $\mathbb{C}^n$ . The standard basis  $\{\mathbf{e}_j : j = 1, 2, \dots, n\}$  is given by

$$\mathbf{e}_j(t) = \begin{cases} 1, & t = j \\ 0, & t \neq j \end{cases} \quad \text{for } t = 1, 2, \dots, n.$$

We often refer to the elements of the standard basis as *spikes* or *impulses*. The Fourier basis  $\{\mathbf{f}_j : j = 1, 2, \dots, n\}$  is given by

$$\mathbf{f}_j(t) = \frac{1}{\sqrt{n}} e^{2\pi i j t / n} \quad \text{for } t = 1, 2, \dots, n.$$

We often refer to the elements of the Fourier basis as *sines* or *complex exponentials*.

The *discrete Fourier transform* (DFT) is the  $n \times n$  matrix  $\mathbf{F}$  whose rows are  $\mathbf{f}_1^*, \mathbf{f}_2^*, \dots, \mathbf{f}_n^*$ . The matrix  $\mathbf{F}$  is unitary. In particular, its spectral norm  $\|\mathbf{F}\| = 1$ . Moreover, the entries of the DFT matrix are bounded in magnitude by  $n^{-1/2}$ . Let  $T$  and  $\Omega$  be subsets of  $\{1, 2, \dots, n\}$ . We write

---

*Date:* 4 September 2007. Revised 15 April 2008.

*2000 Mathematics Subject Classification.* Primary: 46B07, 47A11, 15A52. Secondary: 41A46.

*Key words and phrases.* Fourier analysis, local theory, random matrix, sparse approximation, uncertainty principle.

The author is with Applied & Computational Mathematics, MC 217-50, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125-5000. E-mail: [jtropp@acm.caltech.edu](mailto:jtropp@acm.caltech.edu). Supported by NSF 0503299.

$\mathbf{F}_{\Omega T}$  for the restriction of  $\mathbf{F}$  to the rows listed in  $\Omega$  and the columns listed in  $T$ . Since  $\mathbf{F}_{\Omega T}$  is a submatrix of the DFT matrix, its spectral norm does not exceed one.

We use the analysts' convention that upright letters represent universal constants. We reserve  $c$  for small constants and  $C$  for large constants. The value of a constant may change at each appearance.

**1.2. Linear Independence.** Let  $T$  and  $\Omega$  be subsets of  $\{1, 2, \dots, n\}$ . Consider the collection of spikes and sines listed in these sets:

$$\mathcal{X} = \mathcal{X}(T, \Omega) = \{\mathbf{e}_j : j \in T\} \cup \{\mathbf{f}_j : j \in \Omega\}.$$

Today, we will discuss methods for determining when  $\mathcal{X}$  is linearly independent. Since a linearly independent collection in  $\mathbb{C}^n$  contains at most  $n$  vectors, we obtain a simple necessary condition  $|T| + |\Omega| \leq n$ . Developing sufficient conditions, however, requires more sophistication.

We approach the problem by studying the Gram matrix  $\mathbf{G} = \mathbf{G}(\mathcal{X})$ , whose entries are the inner products between pairs of elements from  $\mathcal{X}$ . It is easy to check that the Gram matrix can be expressed as

$$\mathbf{G} = \begin{bmatrix} \mathbf{I}_{|\Omega|} & \mathbf{F}_{\Omega T} \\ (\mathbf{F}_{\Omega T})^* & \mathbf{I}_{|T|} \end{bmatrix}$$

where  $\mathbf{I}_m$  denotes an  $m \times m$  identity matrix and  $|\cdot|$  denotes the cardinality of a set.

It is well known that the collection  $\mathcal{X}$  is linearly independent if and only if its Gram matrix is nonsingular. The Gram matrix is nonsingular if and only if its eigenvalues are nonzero. A basic (and easily confirmed) fact of matrix analysis is that the extreme eigenvalues of  $\mathbf{G}$  are  $1 \pm \|\mathbf{F}_{\Omega T}\|$ . Therefore, *the collection  $\mathcal{X}$  is linearly independent if and only if  $\|\mathbf{F}_{\Omega T}\| < 1$ .*

One may also attempt to quantify the extent to which collection  $\mathcal{X}$  is linearly independent. To that end, define the *condition number*  $\kappa$  of the Gram matrix, which is the ratio of its largest eigenvalue to its smallest eigenvalue:

$$\kappa(\mathbf{G}) = \frac{1 + \|\mathbf{F}_{\Omega T}\|}{1 - \|\mathbf{F}_{\Omega T}\|}.$$

If  $\|\mathbf{F}_{\Omega T}\|$  is bounded away from one, then the condition number is constant. One may interpret this statement as evidence the collection  $\mathcal{X}$  is strongly linearly independent. The reason is that the condition number is the reciprocal of the relative spectral-norm distance between  $\mathbf{G}$  and the nearest singular matrix [Dem97, p. 33]. As we have mentioned,  $\mathbf{G}$  is singular if and only if  $\mathcal{X}$  is linearly dependent.

This article focuses on statements about linear independence, rather than conditioning. Nevertheless, many results can be adapted to obtain precise information about the size of  $\|\mathbf{F}_{\Omega T}\|$ .

**1.3. Summary of Results.** The major result of this paper to show that a random collection of spikes and sines is extremely likely to be strongly linearly independent, provided that the total number of spikes and sines does not exceed a constant proportion of the ambient dimension. We also provide a result which shows that the norm of a properly scaled random submatrix of the DFT is at most constant with high probability. For a more detailed statement of these theorems, turn to Section 2.3.

**1.4. Outline.** The next section provides a survey of bounds on the norm of a submatrix of the DFT matrix. It concludes with detailed new results for the case where the submatrix is random. Section 3 contains a proof of the new results. Numerical experiments are presented in Section 4, and Section 5 describes some additional research directions. Appendix A contains a proof of the key background result.

## 2. HISTORY AND RESULTS

The strange, eventful history of our problem can be viewed as a sequence of bounds on norm of the matrix  $\mathbf{F}_{\Omega T}$ . Results in the literature can be divided into two classes: the case where the sets  $\Omega$  and  $T$  are fixed and the case where one of the sets is random. In this work, we investigate what happens when both sets are chosen randomly.

**2.1. Bounds for fixed sets.** An early result, due to Donoho and Stark [DS89], asserts that an *arbitrary* collection of spikes and sines is linearly independent, provided that the collection is not too big.

**Theorem 1** (Donoho–Stark). *Suppose that  $|T| |\Omega| < n$ . Then  $\|\mathbf{F}_{\Omega T}\| < 1$ .*

The original argument relies on the fact that  $\mathbf{F}$  is a Vandermonde matrix. We present a short proof that is completely analytic. A similar argument using an inequality of Schur yields the more general result of Elad and Bruckstein [EB02, Thm. 1].

*Proof.* The entries of the  $|\Omega| \times |T|$  matrix  $\mathbf{F}_{\Omega T}$  are uniformly bounded by  $n^{-1/2}$ . Since the Frobenius norm dominates the spectral norm,  $\|\mathbf{F}_{\Omega T}\|^2 \leq \|\mathbf{F}_{\Omega T}\|_F^2 \leq |\Omega| |T| / n$ . Under the hypothesis of the theorem, this quantity does not exceed one.  $\square$

Theorem 1 has an elegant corollary that follows immediately from the basic inequality for geometric and arithmetic means.

**Corollary 2** (Donoho–Stark). *Suppose that  $|T| + |\Omega| < 2\sqrt{n}$ . Then  $\|\mathbf{F}_{\Omega T}\| < 1$ .*

The contrapositive of Theorem 1 is usually interpreted as an *discrete uncertainty principle*: a vector and its discrete Fourier transform cannot simultaneously be sparse. To express this claim quantitatively, we define the  $\ell_0$  “quasinorm” of a vector by  $\|\boldsymbol{\alpha}\|_0 = |\{j : \alpha_j \neq 0\}|$ .

**Corollary 3** (Donoho–Stark). *Fix a vector  $\mathbf{x} \in \mathbb{C}^n$ . Consider the representations of  $\mathbf{x}$  in the standard basis and the Fourier basis:*

$$\mathbf{x} = \sum_{j=1}^n \alpha_j \mathbf{e}_j \quad \text{and} \quad \mathbf{x} = \sum_{j=1}^n \beta_j \mathbf{f}_j.$$

*Then  $\|\boldsymbol{\alpha}\|_0 \|\boldsymbol{\beta}\|_0 \geq n$ .*

The example of the *Dirac comb* shows that Theorem 1 and its corollaries are sharp. Suppose that  $n$  is a square, and let  $T = \Omega = \{\sqrt{n}, 2\sqrt{n}, 3\sqrt{n}, \dots, n\}$ . On account of the Poisson summation formula,

$$\sum_{j \in T} \mathbf{e}_j = \sum_{j \in \Omega} \mathbf{f}_j.$$

Therefore, the set of vectors  $\mathcal{X}(T, \Omega)$  is linearly dependent and  $|T| |\Omega| = n$ .

The substance behind this example is that the abelian group  $\mathbb{Z}/\mathbb{Z}_n$  contains nontrivial subgroups when  $n$  is composite. The presence of these subgroups leads to arithmetic cancelations for properly chosen  $T$  and  $\Omega$ . See [DS89] for additional discussion.

One way to eradicate the cancelation phenomenon is to require that  $n$  be prime. In this case, the group  $\mathbb{Z}/\mathbb{Z}_n$  has no nontrivial subgroup. As a result, much larger collections of spikes and sines are linearly independent. Compare the following result with Corollary 2.

**Theorem 4** (Tao [Tao05, Thm. 1.1]). *Suppose that  $n$  is prime. If  $|T| + |\Omega| \leq n$ , then  $\|\mathbf{F}_{\Omega T}\| < 1$ .*

The proof of Theorem 4 is algebraic in nature, and it does not provide information about conditioning. Indeed, one expects that some submatrices have norms very near to one.

When  $n$  is composite, subgroups of  $\mathbb{Z}/\mathbb{Z}_n$  exist, but they have a very rigid structure. Consequently, one can also avoid cancelations by choosing  $T$  and  $\Omega$  with care. In particular, one may consider the situation where  $T$  is clustered and  $\Omega$  is spread out. Donoho and Logan [DL92] study

this case using the *analytic principle of the large sieve*, a powerful technique from number theory that can be traced back to the 1930s. See the lecture notes [Jam06] for an engaging introduction and references.

Here, we simply restate the (sharp) large sieve inequality [Jam06, LS1.1] in a manner that exposes its connection with our problem. The *spread* of a set is measured as the difference (modulo  $n$ ) between the closest pair of indices. Formally, define

$$\text{spread}(\Omega) = \min\{|j - k \bmod n| : j, k \in \Omega, j \neq k\}$$

with the convention that the modulus returns values in the symmetric range  $\{-\lceil n/2 \rceil + 1, \dots, \lfloor n/2 \rfloor\}$ . Observe that  $|\Omega| \cdot \text{spread}(\Omega) \leq n$ .

**Theorem 5** (Large Sieve Inequality). *Suppose that  $T$  is a block of adjacent indices:*

$$T = \{m + 1, m + 2, \dots, m + |T|\} \quad \text{for an integer } m. \quad (2.1)$$

For each set  $\Omega$ , we have

$$\|\mathbf{F}_{\Omega T}\|^2 \leq \frac{|T| + n/\text{spread}(\Omega) - 1}{n}.$$

In particular, when  $T$  has form (2.1), the bound  $|T| + n/\text{spread}(\Omega) < n + 1$  implies that  $\|\mathbf{F}_{\Omega T}\| < 1$ .

Of course, we can reverse the roles of  $T$  and  $\Omega$  in this theorem on account of duality. The same observation applies to other results where the two sets do not participate in the same way.

The discussion above shows that there are cases where delicately constructed sets  $T$  and  $\Omega$  lead to linearly dependent collections of spikes and sines. Explicit conditions that rule out the bad examples are unknown, but nevertheless the bad examples turn out to be quite rare. To quantify this intuition, we must introduce probability.

**2.2. Bounds when one set is random.** In their work [DS89, Sec. 7.3], Donoho and Stark discuss numerical experiments designed to study what happens when one of the sets of spikes or sines is drawn at random. They conjecture that the situation is vastly different from the case where the spikes and sines are chosen in an arbitrary fashion. Within the last few years, researchers have made substantial theoretical progress on this question. Indeed, we will see that the linearly dependent collections form a vanishing proportion of all collections, provided that the total number of spikes and sines is slightly smaller than the dimension  $n$  of the vector space.

First, we describe a probability model for random sets. Fix a number  $m \leq n$ , and consider the class  $\mathcal{S}_m$  of index sets that have cardinality  $m$ :

$$\mathcal{S}_m = \{S : S \subset \{1, 2, \dots, n\} \text{ and } |S| = m\}.$$

We may construct a random set  $\Omega$  by drawing an element from  $\mathcal{S}_m$  uniformly at random. That is,

$$\mathbb{P}\{\Omega = S\} = |\mathcal{S}_m|^{-1} \quad \text{for each } S \in \mathcal{S}_m.$$

In the sequel, we substitute the symbol  $|\Omega|$  for the letter  $m$ , and we say “ $\Omega$  is a random set with cardinality  $|\Omega|$ ” to describe this type of random variable. This phrase should cause no confusion, and it allows us to avoid extra notation for the cardinality.

In the sparse approximation literature, the first rigorous result on random sets is due to Candès and Romberg. They study the case where one of the sets is arbitrary and the other set is chosen at random. Their proof draws heavily on their prior work with Tao [CRT06].

**Theorem 6** (Candès–Romberg [CR06, Thm. 3.2]). *Fix a number  $s \geq 1$ . Suppose that*

$$|T| + |\Omega| \leq \frac{cn}{\sqrt{(s+1)\log n}}. \quad (2.2)$$

*If  $T$  is an arbitrary set with cardinality  $|T|$  and  $\Omega$  is a random set with cardinality  $|\Omega|$ , then*

$$\mathbb{P}\left\{\|\mathbf{F}_{\Omega T}\|^2 \geq 0.5\right\} \leq C((s+1)\log n)^{1/2}n^{-s}.$$

The numerical constant  $c \geq 0.2791$ , provided that  $n \geq 512$ .

One should interpret this theorem as follows. Fix a set  $T$ , and consider all sets  $\Omega$  that satisfy (2.2). Of these, the proportion that are *not* strongly linearly independent is only about  $n^{-s}$ . One should be aware that the logarithmic factor in (2.2) is intrinsic when one of the sets is arbitrary. Indeed, one can construct examples related to the Dirac comb which show that the failure probability is constant unless the logarithmic factor is present. We omit the details.

The proof of Theorem 6 ultimately involves a variation of the moment method for studying random matrices, which was initiated by Wigner. The key point of the argument is a bound on the expected trace of a high power of the random matrix  $\sqrt{n/|\Omega|} \cdot \mathbf{F}_{\Omega T}^* \mathbf{F}_{\Omega T} - \mathbf{I}_{|T|}$ . The calculations involve delicate combinatorial techniques that depend heavily on the structure of the matrix  $\mathbf{F}$ .

This approach can also be used to establish that the smallest singular value of  $\mathbf{F}_{\Omega T}$  is bounded well away from zero [CRT06, Thm. 2.2]. This lower bound is essential in many applications, but we do not need it here. For extensions of these ideas, see also the work of Rauhut [Rau07].

Another result, similar to Theorem 6, suggests that the arbitrary set and the random set do not contribute equally to the spectral norm. We present one version, whose derivation is adapted from [Tro07, Thm. 10 et seq.].

**Theorem 7.** Fix a number  $s \geq 1$ . Suppose that

$$|T| \log n + |\Omega| \leq \frac{cn}{s}.$$

If  $T$  is an arbitrary set of cardinality  $|T|$  and  $\Omega$  is a random set of cardinality  $|\Omega|$ , then

$$\mathbb{P} \left\{ \|\mathbf{F}_{\Omega T}\|^2 \geq 0.5 \right\} \leq n^{-s}.$$

The proof of this theorem uses Rudelson's selection lemma [Rud99, Sec. 2] in an essential way. This lemma in turn hinges on the noncommutative Khintchine inequality [LP86, Buc01]. For a related application of this approach, see [CR07].

Theorems 6 and 7 are interesting, but they do not predict that a far more striking phenomenon occurs. A random collection of sines has the following property with high probability. To this collection, one can add an *arbitrary* set of spikes without sacrificing linear independence.

**Theorem 8.** Fix a number  $s \geq 1$ , and assume  $n \geq N(s)$ . Except with probability  $n^{-s}$ , a random set  $\Omega$  whose cardinality  $|\Omega| \leq n/3$  has the following property. For each set  $T$  whose cardinality

$$|T| \leq \frac{cn}{s \log^5 n},$$

it holds that  $\|\mathbf{F}_{\Omega T}\|^2 \leq 0.5$ .

This result follows from the (deep) fact that a random row-submatrix of the DFT matrix satisfies the *restricted isometry property* (RIP) with high probability. More precisely, a random set  $\Omega$  with cardinality  $|\Omega|$  verifies the following condition, except with probability  $n^{-s}$ .

$$\frac{|\Omega|}{2n} \leq \|\mathbf{F}_{\Omega T}\|^2 \leq \frac{3|\Omega|}{2n} \quad \text{when } |T| \leq \frac{c|\Omega|}{s \log^5 n}. \quad (2.3)$$

This result is adapted from [RV06, Thm. 2.2 et seq.].

The bound (2.3) was originally established by Candès and Tao [CT06] for sets  $T$  whose cardinality  $|T| \leq c|\Omega|/s \log^6 n$ . Rudelson and Vershynin developed a simpler proof and reduced the exponent on the logarithm [RV06]. Experts believe that the correct exponent is just one or two, but this conjecture is presently out of reach.

*Proof.* Let  $c$  be the constant in (2.3). Abbreviate  $m = c|\Omega|/s \log^5 n$ , and assume that  $m \geq 1$  for now. Draw a random set  $\Omega$  with cardinality  $|\Omega|$ , so relation (2.3) holds except with probability  $n^{-s}$ . Select an arbitrary set  $T$  whose cardinality  $|T| \leq cn/6s \log^5 n$ . We may assume that  $2|T|/m \geq 1$

because  $|\Omega| \leq n/3$ . Partition  $T$  into at most  $2|T|/m$  disjoint blocks, each containing no more than  $m$  indices:  $T = T_1 \cup T_2 \cup \dots \cup T_{2|T|/m}$ . Apply (2.3) to calculate that

$$\|\mathbf{F}_{\Omega T}\|^2 \leq \frac{2|T|}{m} \max_k \|\mathbf{F}_{\Omega T_k}\|^2 \leq |T| \cdot \frac{2s \log^5 n}{c|\Omega|} \cdot \frac{3|\Omega|}{2n} \leq \frac{1}{2}.$$

Adjusting constants, we obtain the result when  $|\Omega|$  is not too small.

In case  $m < 1$ , draw a random set  $\Omega$  and then draw additional random coordinates to form a larger set  $\Omega'$  for which  $c|\Omega'|/s \log^5 n \geq 1$  and  $|\Omega'| \leq n/3$ . This choice is possible because  $n \geq N(s)$ . Apply the foregoing argument to  $\Omega'$ . Since the spectral norm of a submatrix is not larger than the norm of the entire matrix, we have the bound  $\|\mathbf{F}_{\Omega T}\|^2 \leq \|\mathbf{F}_{\Omega' T}\|^2 \leq 0.5$  for each sufficiently small set  $T$ .  $\square$

**2.3. Bounds when both sets are random.** To move into the regime where the number of spikes and sines is proportional to the dimension  $n$ , we need to randomize both sets. The major goal of this article is to establish the following theorem.

**Theorem 9.** *Fix a number  $\varepsilon > 0$ , and assume that  $n \geq N(\varepsilon)$ . Suppose that*

$$|T| + |\Omega| \leq c(\varepsilon) \cdot n.$$

*Let  $T$  and  $\Omega$  be random sets with cardinalities  $|T|$  and  $|\Omega|$ . Then*

$$\mathbb{P} \left\{ \|\mathbf{F}_{\Omega T}\|^2 \geq 0.5 \right\} \leq \exp\{-n^{1/2-\varepsilon}\}.$$

*The constant  $c(\varepsilon) \geq e^{-C/\varepsilon}$ .*

Note that the probability bound here is superpolynomial, in contrast with the polynomial bounds of the previous section. The estimate is essentially optimal. Take  $\varepsilon > 0$ , and suppose it were possible to obtain a bound of the form

$$\mathbb{P} \{ \|\mathbf{F}_{\Omega T}\| = 1 \} \leq \exp\{-n^{1/2+\varepsilon}\} \quad \text{where} \quad |T| + |\Omega| \leq 2n^{1/2}.$$

According to Stirling's approximation, there are about  $\exp\{n^{1/2} \log n\}$  ways to select two sets satisfying the cardinality bound. At the same time, the proportion of sets that are linearly dependent is at most  $\exp\{-n^{1/2+\varepsilon}\}$ . Multiplying these two quantities, we find that no pair of sets meeting the cardinality bound is linearly dependent. This claim contradicts the fact that the Dirac comb yields a linearly dependent collection of size  $2n^{1/2}$ .

**Remark 10.** *As we will see, Theorem 9 holds for every  $n \times n$  matrix  $\mathbf{A}$  with constant spectral norm and uniformly bounded entries:*

$$\|\mathbf{A}\| \leq 1 \quad \text{and} \quad |a_{\omega t}| \leq n^{-1/2} \quad \text{for } \omega, t = 1, 2, \dots, n.$$

*The proof does not rely on any special properties of the discrete Fourier transform.*

**2.4. Random matrix theory.** Finally, we consider an application of this approach to random matrix theory. Note that each column of  $\mathbf{F}_{\Omega T}$  has  $\ell_2$  norm  $\sqrt{|\Omega|/n}$ . Therefore, it is appropriate to rescale the matrix by  $\sqrt{n/|\Omega|}$  so that its columns have unit norm. Under this scaling, it is possible that the norm of the matrix explodes when  $|\Omega|$  is small in comparison with  $n$ . The content of the next result is that this event is highly unlikely if the submatrix is drawn at random.

**Theorem 11.** *Fix a number  $\delta \in (0, c)$ . Suppose that  $n \geq N(\delta)$  and that*

$$|T| \leq |\Omega| = \delta n.$$

*If  $T$  and  $\Omega$  are random sets with cardinalities  $|T|$  and  $|\Omega|$ , then*

$$\mathbb{P} \left\{ \sqrt{\frac{n}{|\Omega|}} \|\mathbf{F}_{\Omega T}\| \geq 9 \right\} \leq n^{-C}.$$

For  $\delta$  in the range  $[c, 1]$ , it is evident that

$$\sqrt{\frac{n}{|\Omega|}} \|\mathbf{F}_{\Omega T}\| \leq c^{-1}.$$

Therefore, we obtain a constant bound for the norm of a normalized random submatrix throughout the entire parameter range.

**Remark 12.** *Theorem 11 also holds for the class of matrices described in Remark 10.*

### 3. NORMS OF RANDOM SUBMATRICES

In this section, we prove Theorem 9 and Theorem 11. First, we describe some problem simplifications. Then we provide a moment estimate for the norm of a very small random submatrix, and we present a device for extrapolating a moment estimate for the norm of a much larger random submatrix. This moment estimate is used to prove a tail bound, which quickly leads to the two major results of the paper.

**3.1. Reductions.** Denote by  $\mathbf{P}_\delta$  a random  $n \times n$  diagonal matrix where exactly  $m = \lfloor \delta n \rfloor$  entries equal one and the rest equal zero. This matrix can be seen as a projector onto a random set of  $m$  coordinates. With this notation, the restriction of a matrix  $\mathbf{A}$  to  $m$  random rows and  $m$  random columns can be expressed as  $\mathbf{P}_\delta \mathbf{A} \mathbf{P}'_\delta$ , where the two projectors are statistically independent from each other.

**Lemma 13** (Square case). *Let  $\mathbf{A}$  be an  $n \times n$  matrix. Suppose that  $T$  and  $\Omega$  are random sets with cardinalities  $|T|$  and  $|\Omega|$ . If  $\delta \geq \max\{|T|, |\Omega|\}/n$ , then*

$$\mathbb{P} \{\|\mathbf{A}_{\Omega T}\| \geq u\} \leq \mathbb{P} \{\|\mathbf{P}_\delta \mathbf{A} \mathbf{P}'_\delta\| \geq u\} \quad \text{for } u \geq 0.$$

*Proof.* It suffices to show that the probability is weakly increasing as the cardinality of one set increases. Therefore, we focus on  $\Omega$  and remove  $T$  from the notation for clarity. Let  $\Omega$  be a random subset of cardinality  $|\Omega|$ . Conditional on  $\Omega$ , we may draw a uniformly random element  $\omega$  from  $\Omega^c$ , and put  $\Omega' = \Omega \cup \{\omega\}$ . This  $\Omega'$  is a uniformly random subset with cardinality  $|\Omega| + 1$ . We have

$$\begin{aligned} \mathbb{P} \{\|\mathbf{A}_\Omega\| \geq u\} &= \mathbb{E} I(\|\mathbf{A}_\Omega\| \geq u) \\ &\leq \mathbb{E} I(\|\mathbf{A}_{\Omega \cup \{\omega\}}\| \geq u) \\ &= \mathbb{E} I(\|\mathbf{A}_{\Omega'}\| \geq u) \\ &= \mathbb{P} \{\|\mathbf{A}_{\Omega'}\| \geq u\} \end{aligned}$$

where we have written  $I(E)$  for the indicator variable of an event. The inequality follows because the spectral norm is weakly increasing when we pass to a larger matrix, and so we have the inclusion of events  $\{\Omega' : \|\mathbf{A}_\Omega\| \geq u\} \subset \{\Omega' : \|\mathbf{A}_{\Omega \cup \{\omega\}}\| \geq u\}$ .  $\square$

It can be inconvenient to work with projectors of the form  $\mathbf{P}_\delta$  because their entries are dependent. We would prefer a model where coordinates are selected independently. To that end, denote by  $\mathbf{R}_\delta$  a random  $n \times n$  diagonal matrix whose entries are independent 0–1 random variables of mean  $\delta$ . This matrix can be seen as a projector onto a random set of coordinates with *average* cardinality  $\delta n$ . The following lemma establishes a relationship between the two types of coordinate projectors. The argument is drawn from [CR06, Sec. 3].

**Lemma 14** (Random coordinate models). *Fix a number  $\delta$  in  $[0, 1]$ . For every  $n \times n$  matrix  $\mathbf{A}$ ,*

$$\mathbb{P} \{\|\mathbf{P}_\delta \mathbf{A}\| \geq u\} \leq 2\mathbb{P} \{\|\mathbf{R}_\delta \mathbf{A}\| \geq u\} \quad \text{for } u \geq 0.$$

*In particular,*

$$\mathbb{P} \{\|\mathbf{P}_\delta \mathbf{A} \mathbf{P}'_\delta\| \geq u\} \leq 4\mathbb{P} \{\|\mathbf{R}_\delta \mathbf{A} \mathbf{R}'_\delta\| \geq u\} \quad \text{for } u \geq 0.$$

*Proof.* Given a coordinate projector  $\mathbf{R}$ , denote by  $\sigma(\mathbf{R})$  the set of coordinates onto which it projects. For typographical felicity, we use  $\#\sigma(\mathbf{R})$  to indicate the cardinality of this set.

First, suppose that  $\delta n$  is an integer. For every  $u \geq 0$ , we may calculate that

$$\begin{aligned} \mathbb{P}\{\|\mathbf{R}_\delta \mathbf{A}\| \geq u\} &\geq \sum_{j=\delta n}^n \mathbb{P}\{\|\mathbf{R}_\delta \mathbf{A}\| \geq u \mid \#\sigma(\mathbf{R}_\delta) = j\} \cdot \mathbb{P}\{\#\sigma(\mathbf{R}_\delta) = j\} \\ &\geq \mathbb{P}\{\|\mathbf{R}_\delta \mathbf{A}\| \geq u \mid \#\sigma(\mathbf{R}_\delta) = \delta n\} \cdot \sum_{j=\delta n}^n \mathbb{P}\{\#\sigma(\mathbf{R}_\delta) = j\} \\ &\geq \frac{1}{2} \mathbb{P}\{\|\mathbf{P}_\delta \mathbf{A}\| \geq u\}. \end{aligned}$$

The second inequality holds because the spectral norm of a submatrix is smaller than the spectral norm of the matrix. The third inequality relies on the fact [JS68, Thm. 3.2] that the medians of the binomial distribution  $\text{BINOMIAL}(\delta, n)$  lie between  $\delta n - 1$  and  $\delta n$ .

In case  $\delta n$  is not integral, the monotonicity of the spectral norm yields that

$$\mathbb{P}\{\|\mathbf{R}_\delta \mathbf{A}\| \geq u\} \geq \mathbb{P}\{\|\mathbf{R}_{\lfloor \delta n \rfloor / n} \mathbf{A}\| \geq u\}.$$

Since  $\mathbf{P}_{\lfloor \delta n \rfloor / n} = \mathbf{P}_\delta$ , this point completes the argument.  $\square$

**3.2. Small submatrices.** We focus on matrices with uniformly bounded entries. The first step in the argument is an elementary estimate on the norm of a random submatrix with expected order one. In this regime, the bound on the matrix entries determines the norm of the submatrix; the signs of the entries do not play a role. The proof shows that most of the variation in the norm actually derives from the fluctuation in the order of the submatrix.

**Lemma 15** (Small Submatrices). *Let  $\mathbf{A}$  be an  $n \times n$  matrix whose entries are bounded in magnitude by  $n^{-1/2}$ . Abbreviate  $\varrho = 1/n$ . When  $q \geq 2 \log n \geq e$ ,*

$$\left( \mathbb{E} \|\mathbf{R}_\varrho \mathbf{A} \mathbf{R}'_\varrho\|^{2q} \right)^{1/2q} \leq 2qn^{-1/2}.$$

*Proof.* By homogeneity, we may rescale  $\mathbf{A}$  so that its entries are bounded in magnitude by one. Define the event  $\Sigma_{jk}$  where the random submatrix has order  $j \times k$ .

$$\Sigma_{jk} = \{\#\sigma(\mathbf{R}_\varrho) = j \text{ and } \#\sigma(\mathbf{R}'_\varrho) = k\}.$$

On this event, the norm of the submatrix can be bounded as

$$\|\mathbf{R}_\varrho \mathbf{A} \mathbf{R}'_\varrho\| \leq \|\mathbf{R}_\varrho \mathbf{A} \mathbf{R}'_\varrho\|_{\text{F}} \leq \sqrt{jk}.$$

Using elementary inequalities, we may estimate the probability that this event occurs.

$$\mathbb{P}(\Sigma_{jk}) = \binom{n}{j} \binom{n}{k} \varrho^{j+k} (1 - \varrho)^{2n - (j+k)} \leq \left(\frac{en}{j}\right)^j \left(\frac{en}{k}\right)^k n^{-(j+k)} = (e/j)^j \cdot (e/k)^k.$$

With this information at hand, the rest of the proof follows from some easy calculations:

$$\begin{aligned} \mathbb{E} \|\mathbf{R}_\varrho \mathbf{A} \mathbf{R}'_\varrho\|^{2q} &= \sum_{j,k=1}^n \mathbb{E} \left[ \|\mathbf{R}_\varrho \mathbf{A} \mathbf{R}'_\varrho\|^{2q} \mid \Sigma_{jk} \right] \cdot \mathbb{P}(\Sigma_{jk}) \\ &\leq \sum_{j,k=1}^n (jk)^q \cdot (e/j)^j \cdot (e/k)^k \\ &= \left[ \sum_{k=1}^n k^q \cdot (e/k)^k \right]^2. \end{aligned}$$

A short exercise in differential calculus shows that the maximum term in the sum occurs when  $k \log k = q$ . Write  $k_\star$  for the solution to this equation, and note that  $k_\star \leq q$ . Bounding all the terms by the maximum, we find

$$\sum_{k=1}^n k^q \cdot (e/k)^k \leq n \cdot \exp\{q \log k_\star - k_\star \log k_\star + k_\star\} \leq n \cdot \exp\{q \log k_\star\} \leq n \cdot q^q.$$

Combining the last two inequalities, we reach

$$\left(\mathbb{E} \|\mathbf{R}_\varrho \mathbf{A} \mathbf{R}'_\varrho\|^{2q}\right)^{1/2q} \leq (n^2 \cdot q^{2q})^{1/2q} = n^{1/q} \cdot q.$$

When  $q \geq 2 \log n$ , the first term is less than two.  $\square$

**Remark 16.** *This argument delivers a moment estimate that is roughly a factor of  $\log q$  smaller than the one stated. This fact can be used to sharpen the major results slightly at a cost we prefer to avoid.*

**3.3. Extrapolation.** The key technique in the proof is an extrapolation of the moments of the norm of a large random submatrix from the moments of a smaller random submatrix. Without additional information, extrapolation must be fruitless because the signs of matrix entries play a critical role in determining the spectral norm. It turns out that we can fold in information about the signs by incorporating a bound on the spectral norm of the matrix. The proof, which we provide in Appendix A, ultimately depends on the minimax property of the Chebyshev polynomials. The method is essentially the same as the one Bourgain and Tzafriri develop to prove Proposition 2.7 in [BT91]. See also [Tro08, Sec. 7].

**Proposition 17.** *Suppose that  $\mathbf{A}$  is an  $n \times n$  matrix with  $\|\mathbf{A}\| \leq 1$ . Let  $q$  be an integer that satisfies  $13 \log n \leq q \leq n/2$ . Write  $\varrho = 1/n$ , and choose  $\delta$  in the range  $[1/n, 1]$ . For each  $\lambda \in (0, 1)$ , it holds that*

$$\left(\mathbb{E} \|\mathbf{R}_\delta \mathbf{A} \mathbf{R}'_\delta\|^{2q}\right)^{1/2q} \leq 8\delta^\lambda \max \left\{ 1, n^\lambda \left(\mathbb{E} \|\mathbf{R}_\varrho \mathbf{A} \mathbf{R}'_\varrho\|^{2q}\right)^{1/2q} \right\}.$$

Although the statement is a little complicated, we require the full power of this estimate. As usual, the parameter  $q$  is the moment that we seek. The proposition extrapolates from a matrix of expected order 1 up to a matrix of expected order  $\delta n$ . The parameter  $\lambda$  is a tuning knob that controls how much of the estimate is determined by the spectral norm of the full matrix and how much is determined by the norm bound for small submatrices. Indeed, the first member of the maximum reflects the spectral norm bound  $\|\mathbf{A}\| \leq 1$ .

**3.4. A tail bound.** We are now prepared to develop a tail bound for the random norm  $\|\mathbf{R}_\delta \mathbf{A} \mathbf{R}'_\delta\|$ .

**Lemma 18 (Tail Bound).** *Let  $\mathbf{A}$  be an  $n \times n$  matrix for which*

$$\|\mathbf{A}\| \leq 1 \quad \text{and} \quad |a_{jk}| \leq n^{-1/2} \quad \text{for } j, k = 1, 2, \dots, n.$$

*Choose  $\delta$  from  $[1/n, 1]$  and an integer  $q$  that satisfies  $13 \log n \leq q \leq n/2$ . For each  $\lambda \in (0, 1)$ , it holds that*

$$\mathbb{P} \left\{ \|\mathbf{R}_\delta \mathbf{A} \mathbf{R}'_\delta\| \geq 8\delta^\lambda \max \{1, 2qn^{\lambda-1/2}\} \cdot u \right\} \leq u^{-2q} \quad \text{for } u \geq 1.$$

*Proof of Lemma 18.* Choose an integer  $q$  in the range  $[13 \log n, n/2]$ . Markov's inequality allows that

$$\mathbb{P} \left\{ \|\mathbf{R}_\delta \mathbf{A} \mathbf{R}'_\delta\| \geq \left(\mathbb{E} \|\mathbf{R}_\delta \mathbf{A} \mathbf{R}'_\delta\|^{2q}\right)^{1/2q} \cdot u \right\} \leq u^{-2q}.$$

Therefore, we may establish the result by obtaining a moment estimate. This estimate is a direct consequence of Lemma 15 and Proposition 17:

$$\left(\mathbb{E} \|\mathbf{R}_\delta \mathbf{A} \mathbf{R}'_\delta\|^{2q}\right)^{1/2q} \leq 8\delta^\lambda \max \left\{ 1, n^\lambda \cdot 2qn^{-1/2} \right\}.$$

Combine the two bounds to complete the argument.  $\square$

The two major results of this paper, Theorem 9 and Theorem 11, both follow from a simple corollary of Lemma 18.

**Corollary 19.** *Suppose that  $T$  and  $\Omega$  are random sets with cardinalities  $|T|$  and  $|\Omega|$ . Assume  $\delta \geq \max\{|T|, |\Omega|\}/n$ . For each integer  $q$  that satisfies  $13 \log n \leq q \leq n/2$  and for  $\lambda \in [0, 1]$ , it holds that*

$$\mathbb{P} \left\{ \|\mathbf{F}_{\Omega T}\| \geq 8\delta^\lambda \max\{1, 2qn^{\lambda-1/2}\} \cdot u \right\} \leq 4u^{-2q} \quad \text{for } u \geq 1.$$

*Proof.* Consider the matrix  $\mathbf{A} = \mathbf{F}$ . Perform the reductions from Section 3.1, Lemma 13 and Lemma 14. Then apply the tail bound, Lemma 18.  $\square$

**3.5. Proof of Theorem 9.** The content of Theorem 9 is to provide a bound on  $\delta$  which ensures that  $\|\mathbf{F}_{\Omega T}\|$  is somewhat less than one with extremely high probability. To that end, we want to make  $\lambda$  close to zero and  $q$  large. The following selections accomplish this goal:

$$\lambda = \frac{\log 16}{\log(1/\delta)} \quad \text{and} \quad q = \lfloor 0.5n^{1/2-\lambda} \rfloor.$$

Note that we can make  $\lambda$  as small as we like by taking  $\delta$  sufficiently small. For any value of  $\lambda < 0.5$ , the number  $q$  satisfies the requirements of Corollary 19 as soon as  $n$  is sufficiently large.

Now, the bound of Corollary 19 results in

$$\mathbb{P} \left\{ \|\mathbf{F}_{\Omega T}\| \geq 0.5u \right\} \leq 4u^{-2q}.$$

For  $u = \sqrt{2}$ , we see that

$$\mathbb{P} \left\{ \|\mathbf{F}_{\Omega T}\|^2 \geq 0.5 \right\} \leq 4 \cdot 2^{-q}.$$

It follows that, for any assignable  $\varepsilon > 0$ , we can make

$$\mathbb{P} \left\{ \|\mathbf{F}_{\Omega T}\|^2 \geq 0.5 \right\} \leq \exp\{-n^{1/2-\varepsilon}\}$$

provided that  $\delta \leq e^{-C/\varepsilon} = c(\varepsilon)$  and that  $n \geq N(\varepsilon)$ .

**3.6. Proof of Theorem 11.** To establish Theorem 11, we must make the parameter  $\lambda$  as close to 0.5 as possible. Choose

$$\lambda = \frac{1}{2} - \frac{0.1}{\log(1/\delta)} \quad \text{and} \quad q = \lfloor C \log n \rfloor.$$

where  $C$  is a large constant. These choices are acceptable once  $\delta$  is sufficiently small and  $n$  is sufficiently large.

Corollary 19 delivers

$$\mathbb{P} \left\{ \|\mathbf{F}_{\Omega T}\| \geq 8.9\delta^{1/2}u \right\} \leq 4u^{-C \log n}.$$

For  $u = 90/89$ , we reach

$$\mathbb{P} \left\{ \|\mathbf{F}_{\Omega T}\| \geq 9\delta^{1/2} \right\} \leq n^{-C},$$

adjusting constants as necessary. Finally, we transfer the factor  $\delta^{1/2}$  to the other side of the inequality and set  $\delta = |\Omega|/n$  to complete the proof.

#### 4. NUMERICAL EXPERIMENTS

The theorems of this paper provide gross information about the norm of a random submatrix of the DFT. To complement these results, we performed some numerical experiments to give a more detailed empirical view.

The first set of experiments concerns random square submatrices of a DFT matrix of size  $n$ , where we varied the parameter  $n$  over several orders of magnitude. Given a value of  $\delta \in (0, 0.5)$ , we formed one hundred random submatrices with dimensions  $\delta n \times \delta n$  and computed the average spectral norm of these matrices. We did not plot data when  $\delta \in (0.5, 1)$  because the norm of a random submatrix equals one.

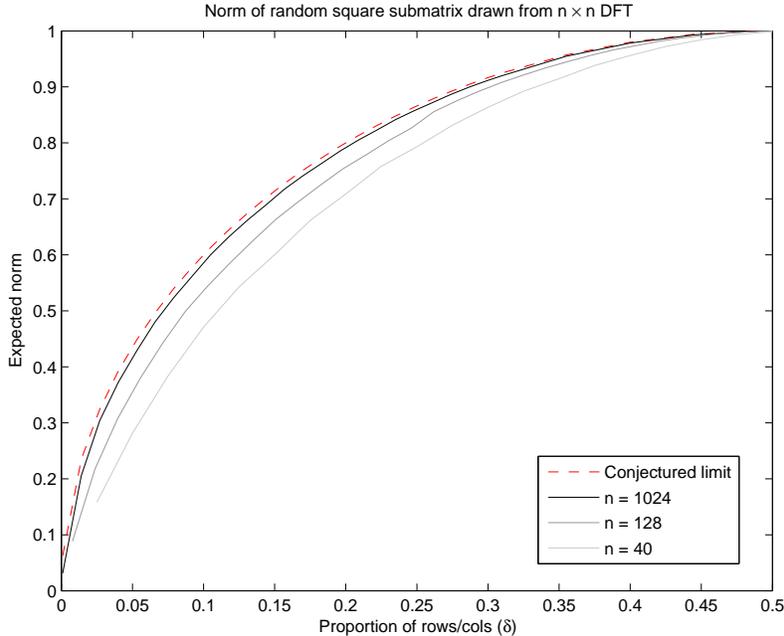


FIGURE 1. Sample average of the norm of a random  $\delta n \times \delta n$  submatrix drawn from the  $n \times n$  DFT.

Figure 1 shows the raw data for this first experiment. As  $n$  grows, one can see that the norm tends toward an apparent limit:  $2\sqrt{\delta(1-\delta)}$ . In Figure 2, we re-scale each matrix by  $\delta^{-1/2}$  so its columns have unit norm and then compute the average spectral norm. More elaborate behavior is visible in this plot:

- For  $\delta = 1/n$ , the norm of a random submatrix is identically equal to one.
- For  $\delta = 2/n$ , the norm tends toward  $1 + 2^{-1/2} = 1.7071\dots$ , which can be verified by a relatively simple analytic computation.
- The maximum value of the norm appears to occur at  $\delta = 2/\sqrt{n}$ .
- The apparent limit of the scaled norm is  $2\sqrt{1-\delta}$ , in agreement with the first figure.

These phenomena are intriguing, and it would be valuable to understand them in more detail. Unfortunately, the methods of this paper are not refined enough to provide an explanation.

In the second set of experiments, we studied the norm of a random rectangular submatrix of the  $128 \times 128$  DFT matrix. We varied the proportion  $\delta_T$  of columns and the proportion  $\delta_\Omega$  of rows in the range  $(0, 1)$ . For each pair  $(\delta_T, \delta_\Omega)$ , we drew 100 random submatrices and computed the average norm. Figure 3 shows the raw data. The apparent trend is that

$$\mathbb{E} \|\mathbf{P}_{\delta_\Omega} \mathbf{F} \mathbf{P}'_{\delta_T}\| = 2\sqrt{\delta(1-\delta)} \quad \text{where} \quad \delta = \frac{|T| + |\Omega|}{2}.$$

Figure 4 shows the same data, rescaled by  $\max\{|T|, |\Omega|\}^{-1/2}$ . As in the square case, this plot reveals a variety of interesting phenomena that are worth attention.

## 5. FURTHER RESEARCH DIRECTIONS

The present research suggests several directions for future exploration.

- (1) It may be possible to improve the constants in Proposition 17 using a variation of the current approach. Instead of using the Chebyshev polynomial to estimate the coefficients of the polynomial that arises in the proof, one might use the nonnegative polynomial of least

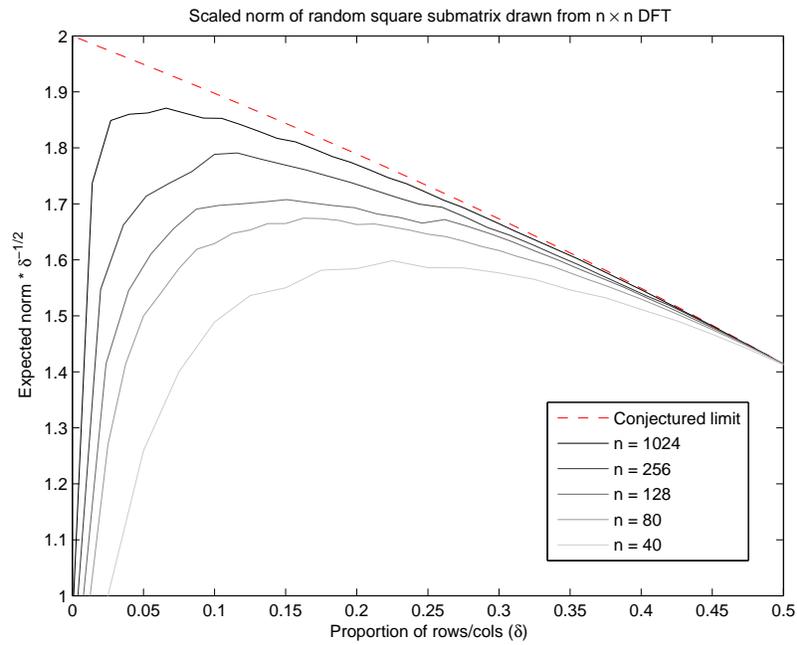


FIGURE 2. Sample average of the norm of a random  $\delta n \times \delta n$  submatrix drawn from the  $n \times n$  DFT and re-scaled by  $\delta^{-1/2}$ .

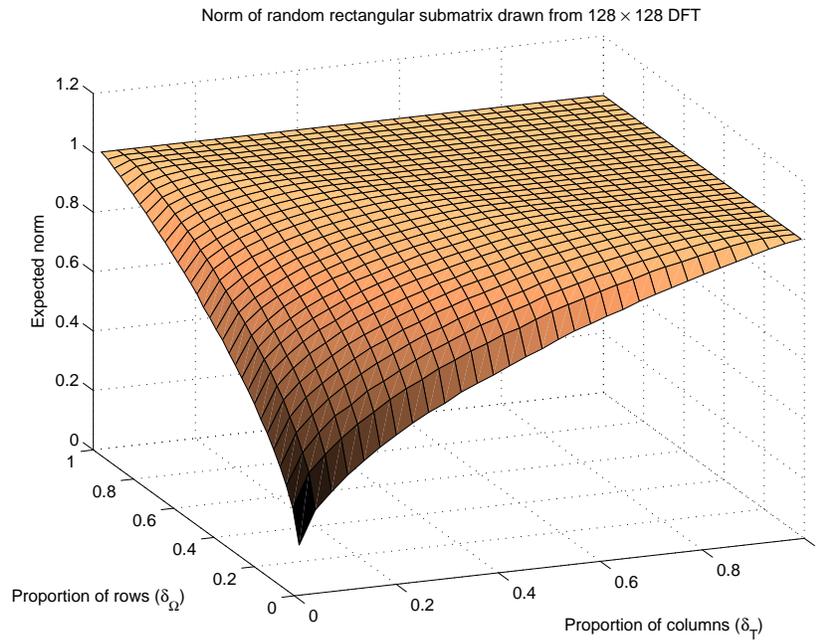


FIGURE 3. Sample average of the norm of a random  $\delta_\Omega n \times \delta_\Gamma n$  submatrix drawn from the  $128 \times 128$  DFT matrix.

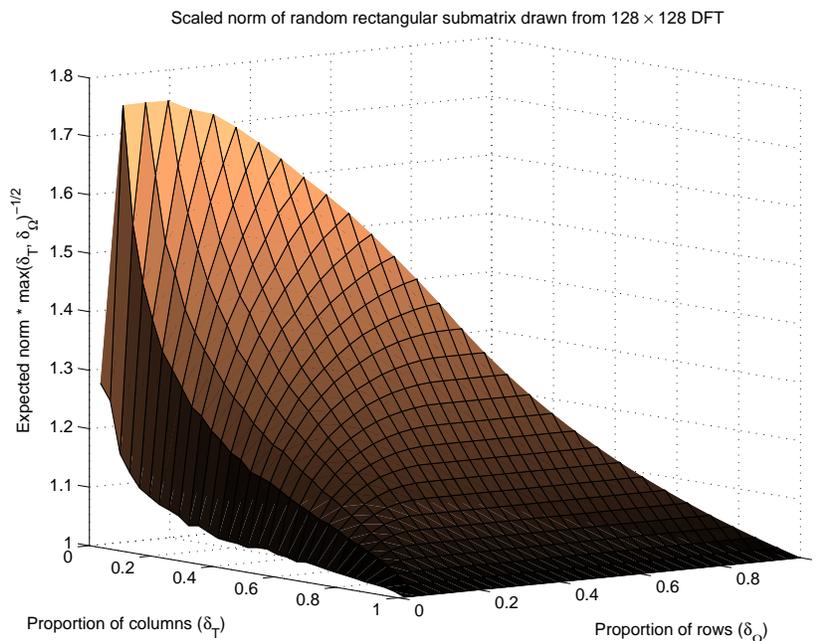


FIGURE 4. Sample average of the norm of a random  $\delta_\Omega n \times \delta_T n$  submatrix drawn from the  $128 \times 128$  DFT matrix and rescaled by  $\max\{|T|, |\Omega|\}^{-1/2}$ .

deviation from zero on the interval  $[0, 1]$ . The paper [BK85] is relevant in this connection: its authors identify the nonnegative polynomials with least deviation from zero with respect to  $L_p$  norms for  $p < \infty$ . The  $p = \infty$  case appears to be open, and uniqueness may be an issue.

- (2) Instead of reducing the problem to the square case, it would be valuable to understand the rectangular case directly. Again, it may be possible to adapt Proposition 17 to handle this situation. This approach would probably require the bivariate polynomials of least deviation from zero identified by Sloss [Slo65].
- (3) A harder problem is to determine the limiting behavior of the expected norm of a random submatrix as the dimension grows and the proportion of rows and columns remains fixed. We frame the following conjecture.

**Conjecture 20** (Quartercircle Law). *A random square submatrix of the  $n \times n$  DFT satisfies*

$$\mathbb{E} \|P_\delta \mathbf{F} P'_\delta\| \leq 2\sqrt{\delta(1-\delta)}.$$

*The inequality becomes an equality as  $n \rightarrow \infty$ .*

One can develop a similar statement about random rectangular submatrices. At present, however, these conjectures are out of reach.

- (4) Finally, one might study the behavior of the lower singular value of a (suitably normalized) random submatrix drawn from the DFT. There are some results available when one set, say  $T$ , is fixed [CRT06]. It is possible that the behavior will be better when both sets are random. The present methods do not seem to provide much information about this problem.

## ACKNOWLEDGMENTS

One of the anonymous referees provided a wealth of useful advice that substantially improved the quality of this work. In particular, the referee described a version of Lemma 15 and demonstrated that it offers a simpler route to the main results than the argument in earlier drafts of this paper.

## APPENDIX A. CHEBYSHEV EXTRAPOLATION

One of the major tools in the proof of Theorem 9 is Proposition 17. This result extrapolates the moments of the norm of a large random submatrix drawn from a fixed matrix, given information about a small random submatrix. An important idea behind the result is to fold information about the spectral norm of the matrix into the estimate. The extrapolation technique is due to Bourgain and Tzafriri [BT91]. We require a variant of their result, so we repeat the argument in its entirety. The complete statement of the result follows.

**Proposition 21.** *Suppose that  $\mathbf{A}$  is an  $n \times n$  matrix with  $\|\mathbf{A}\| \leq 1$ . Let  $q$  be an integer that satisfies  $13 \log n \leq q \leq n/2$ . Choose parameters  $\varrho \in (0, 1)$  and  $\delta \in [\varrho, 1]$ . For each  $\lambda \in [0, 1]$ , it holds that*

$$\left( \mathbb{E} \|\mathbf{R}_\delta \mathbf{A} \mathbf{R}'_\delta\|^{2q} \right)^{1/2q} \leq 8\delta^\lambda \max \left\{ 1, \varrho^{-\lambda} \left( \mathbb{E} \|\mathbf{R}_\varrho \mathbf{A} \mathbf{R}'_\varrho\|^{2q} \right)^{1/2q} \right\}.$$

*The same result holds if we replace  $\mathbf{R}'_\delta$  by  $\mathbf{R}_\delta$  and replace  $\mathbf{R}'_\varrho$  by  $\mathbf{R}_\varrho$ .*

V. A. Markov observed that the coefficients of an arbitrary polynomial can be bounded in terms of the coefficients of a Chebyshev polynomial because Chebyshev polynomials are the unique polynomials of least deviation from zero on the unit interval. See [Tim63, Sec. 2.9] for more details.

**Proposition 22** (Markov). *Let  $p(t) = \sum_{k=0}^r c_k t^k$ . The coefficients of the polynomial  $p$  satisfy the inequality*

$$|c_k| \leq \frac{r^k}{k!} \max_{|t| \leq 1} |p(t)| \leq e^r \max_{|t| \leq 1} |p(t)|.$$

*for each  $k = 0, 1, \dots, r$ .*

With Markov's result at hand, we can prove Proposition 21.

*Proof of Proposition 21.* We establish the result when the two diagonal projectors are independent; the other case is almost identical because this independence is never exploited. Define the function

$$F(s) = \mathbb{E} \|\mathbf{R}_s \mathbf{A} \mathbf{R}'_s\|^{2q} \quad \text{for } s \in [0, 1].$$

Note that  $F(s) \leq 1$  because  $\|\mathbf{R}_s \mathbf{A} \mathbf{R}'_s\| \leq \|\mathbf{A}\| \leq 1$ . Furthermore,  $F$  does not decrease.

The function  $F$  is comparable with a polynomial. Use the facts that  $2q$  is even and that  $\mathbf{A}$  has dimension  $n$  to check the inequalities

$$F(s) \leq \mathbb{E} \text{trace}[(\mathbf{R}_s \mathbf{A} \mathbf{R}'_s)^* (\mathbf{R}_s \mathbf{A} \mathbf{R}'_s)]^q \leq nF(s). \quad (\text{A.1})$$

Define a second function

$$p(s) = \mathbb{E} \text{trace}[(\mathbf{R}_s \mathbf{A} \mathbf{R}'_s)^* (\mathbf{R}_s \mathbf{A} \mathbf{R}'_s)]^q = \mathbb{E} \text{trace}(\mathbf{A}^* \mathbf{R}_s \mathbf{A} \mathbf{R}'_s)^q,$$

where we used the cyclicity of the trace and the fact that  $\mathbf{R}_s$  and  $\mathbf{R}'_s$  are diagonal matrices with 0–1 entries. Expand the product and compute the expectation using the additional fact that the entries of the diagonal matrices are independent random variables of mean  $s$ . We discover that  $p$  is a polynomial of maximum degree  $2q$  in the variable  $s$ :

$$p(s) = \sum_{k=1}^{2q} c_k s^k$$

The polynomial has no constant term because  $\mathbf{R}_0 = \mathbf{0}$ .

We can use Markov's technique to bound the coefficients of the polynomial. First, make the change of variables  $s = \varrho t^2$  to see that

$$\left| \sum_{k=1}^{2q} c_k \varrho^k t^{2k} \right| = |p(\varrho t^2)| \leq nF(\varrho t^2) \leq nF(\varrho) \quad \text{for } |t| \leq 1.$$

The first inequality follows from (A.1) and the second follows from the monotonicity of  $F$ . The polynomial  $p(\varrho t^2)$  has degree  $4q$  in the variable  $t$ , so Proposition 22 yields

$$|c_k| \varrho^k \leq n e^{4q} F(\varrho) \quad \text{for } k = 1, 2, \dots, 2q. \quad (\text{A.2})$$

Evaluate this expression at  $\varrho = 1$  and recall that  $F \leq 1$  to obtain a second bound,

$$|c_k| \leq n e^{4q} \quad \text{for } k = 1, 2, \dots, 2q. \quad (\text{A.3})$$

To complete the proof, we evaluate the polynomial at a point  $\delta$  in the range  $[\varrho, 1]$ . Fix a value of  $\lambda$  in  $[0, 1]$ , and set  $K = \lfloor 2\lambda q \rfloor$ . In view of (A.2) and (A.3), we obtain

$$\begin{aligned} F(\delta) &\leq \sum_{k=1}^K |c_k| \delta^k + \sum_{k=K+1}^{2q} |c_k| \delta^k \\ &\leq \sum_{k=1}^K n e^{4q} F(\varrho) (\delta/\varrho)^k + \sum_{k=K+1}^{2q} n e^{4q} \delta^k \\ &\leq n e^{4q} [K (\delta/\varrho)^K F(\varrho) + (2q - K) \delta^{K+1}] \\ &\leq n e^{4q} \delta^{2\lambda q} [K \varrho^{-2\lambda q} F(\varrho) + (2q - K)] \\ &\leq n e^{4q} \delta^{2\lambda q} \cdot 2q \max\{1, \varrho^{-2\lambda q} F(\varrho)\} \end{aligned}$$

The third and fourth inequalities use the conditions  $\delta/\varrho \geq 1$  and  $\delta \leq 1$ , and the last bound is an application of Jensen's inequality. Taking the  $(2q)$ th root, we reach

$$F(\delta)^{1/2q} \leq (2qn)^{1/2q} e^2 \delta^\lambda \max\{1, \varrho^{-\lambda} F(\varrho)^{1/2q}\}.$$

The leading constant is less than 8, provided that  $13 \log n \leq q \leq n/2$ . □

## REFERENCES

- [BK85] V. F. Babenko and V. A. Kofanov. Polynomials of fixed sign that deviate least from zero in the spaces  $L_p$ . *Math. Notes*, 37(2):99–105, Feb. 1985.
- [BT91] J. Bourgain and L. Tzafriri. On a problem of Kadison and Singer. *J. reine angew. Math.*, 420:1–43, 1991.
- [Buc01] A. Buchholz. Operator Khintchine inequality in non-commutative probability. *Math. Annalen*, 319:1–16, 2001.
- [CR06] E. J. Candès and J. Romberg. Quantitative robust uncertainty principles and optimally sparse decompositions. *Foundations of Comput. Math*, 6:227–254, 2006.
- [CR07] E. J. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23:969–985, June 2007.
- [CRT06] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete Fourier information. *IEEE Trans. Info. Theory*, 52(2):489–509, Feb. 2006.
- [CT06] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Info. Theory*, 52(12):5406–5425, Dec. 2006.
- [Dem97] J. Demmel. *Applied Numerical Linear Algebra*. SIAM, Philadelphia, 1997.
- [DL92] D. L. Donoho and P. Logan. Signal recovery and the large sieve. *SIAM J. Appl. Math.*, 52(2):577–591, 1992.
- [DS89] D. L. Donoho and P. B. Stark. Uncertainty principles and signal recovery. *SIAM J. Appl. Math.*, 49(3):906–931, June 1989.
- [EB02] M. Elad and A. M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Trans. Info. Theory*, 48(9):2558–2567, 2002.
- [Jam06] G. J. O. Jameson. Notes on the large sieve. Available online: <http://www.maths.lancs.ac.uk/~jameson/lsv.pdf>, 2006.
- [JS68] K. Jogdeo and S. M. Samuels. Monotone convergence of binomial probabilities and generalization of Ramanujan's equation. *Ann. Math. Stat.*, 39:1191–1195, 1968.
- [LP86] F. Lust-Picquard. Inégalités de Khintchine dans  $C_p$  ( $1 < p < \infty$ ). *Comptes Rendus Acad. Sci. Paris, Série I*, 303(7):289–292, 1986.

- [Rau07] H. Rauhut. Random sampling of trigonometric polynomials. *Appl. Comp. Harmonic Anal.*, 22(1):16–42, 2007.
- [Rud99] M. Rudelson. Random vectors in the isotropic position. *J. Functional Anal.*, 164:60–72, 1999.
- [RV06] M. Rudelson and R. Vershynin. Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements. In *Proc. 40th Annual Conference on Information Sciences and Systems*, Princeton, Mar. 2006.
- [Slo65] J. M. Sloss. Chebyshev approximation to zero. *Pacific J. Math*, 15(1):305–313, 1965.
- [Tao05] T. Tao. An uncertainty principle for cyclic groups of prime order. *Math. Res. Lett.*, 12(1):121–127, 2005.
- [Tim63] A. F. Timan. *Theory of approximation of functions of a real variable*. Pergamon, 1963.
- [Tro07] J. A. Tropp. On the conditioning of random subdictionaries. *Appl. Comp. Harmonic Anal.*, 2007. To appear.
- [Tro08] J. A. Tropp. The random paving property for uniformly bounded matrices. *Studia Math.*, 185(1):67–82, 2008.