Derandomized Compressed Sensing with Nonuniform Guarantees for $\ell_1$ Recovery

Charles Clum$^1$ · Dustin G. Mixon$^1$

Received: 11 February 2021 / Revised: 9 September 2021 / Accepted: 24 February 2022 / Published online: 29 March 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
In compressed sensing, the sensing matrices of minimal sample complexity are constructed with the help of randomness. Over 13 years ago, Tao (Open question: deterministic UUP matrices. https://terrytao.wordpress.com/2007/07/02/open-question-deterministic-uup-matrices/) posed the notoriously difficult problem of derandomizing these sensing matrices. While most work in this vein has been in the setting of explicit deterministic matrices with uniform guarantees, the present paper focuses on explicit random matrices of low entropy with non-uniform guarantees. Specifically, we extend the techniques of Hügel et al. (Found Comput Math 14:115–150, 2014) to show that for every $\delta \in (0, 1]$, there exists an explicit random $m \times N$ partial Fourier matrix $A$ with $m \leq C_1(\delta) s \log^{4/\delta}(N/\epsilon)$ and entropy at most $C_2(\delta) s^\delta \log^5(N/\epsilon)$ such that for every $s$-sparse signal $x \in \mathbb{C}^N$, there exists an event of probability at least $1 - \epsilon$ over which $x$ is the unique minimizer of $\|z\|_1$ subject to $Az = Ax$. The bulk of our analysis uses tools from decoupling to estimate the extreme singular values of the submatrix of $A$ whose columns correspond to the support of $x$.

Keywords Compressed sensing · Derandomization · $\ell_1$ minimization

1 Introduction
A vector $x \in \mathbb{C}^N$ is said to be $s$-sparse if it has at most $s$ nonzero entries. Natural images are well-approximated by sparse vectors in a wavelet domain, and this feature is exploited by JPEG2000 image compression [41]. In 2006, Candès, Romberg and Tao [13] and Donoho [18] discovered that sparsity could also be exploited for compressed sensing. One popular formulation of compressed sensing is to find a sensing matrix...
$A \in \mathbb{C}^{m \times N}$ such that every $s$-sparse vector $x \in \mathbb{C}^N$ with $s \leq m/\text{polylog}N$ can be efficiently reconstructed from the linear data $y = Ax$ by solving the convex program

$$\min \|z\|_1 \text{ subject to } Az = y. \quad (1)$$

To enjoy this $\ell_1$-recovery property, it suffices for $A$ to act as a near-isometry over the set of $s$-sparse vectors [11, 12]:

$$\frac{1}{2} \|x - y\|_2^2 \leq \|Ax - Ay\|_2^2 \leq \frac{3}{2} \|x - y\|_2^2 \quad \text{for every } s \text{-sparse } x, y \in \mathbb{C}^N. \quad (2)$$

We refer to such $A$ as $s$-restricted isometries. Equivalently, every submatrix $A_T$ that is comprised of $2s$ columns from $A$ has singular values $\sigma(A_T) \subseteq [\sqrt{1/2}, \sqrt{3/2}]$. Since random matrices exhibit predictable extreme singular values [40, 43, 46], it comes as no surprise that many distributions of random matrices $A \in \mathbb{C}^{m \times N}$ are known to be $s$-restricted isometries with high probability provided $m \geq s \text{polylog}N$, e.g., [8, 14, 32, 34]. Unfortunately, testing (2) is NP-hard [3, 42], and it is even hard for matrices with independent subgaussian entries, assuming the hardness of finding planted dense subgraphs [47].

In 2007, Tao [39] posed the problem of finding explicit $s$-restricted isometries $A \in \mathbb{C}^{m \times N}$ with $N^\epsilon \leq m \leq (1 - \epsilon)N$ and $m = s \text{polylog}N$. One may view this as an instance of Avi Wigderson’s hay in a haystack problem [4]. To be clear, we say a sequence $\{A_N\}$ of $m(N) \times N$ matrices with $N \to \infty$ is explicit if there exists an algorithm that on input $N$ produces $A_N$ in time that is polynomial in $N$. For example, we currently know of several explicit sequences of matrices $A$ with unit-norm columns $\{a_i\}_{i \in [N]}$ and minimum coherence:

$$\max_{i,j \in [N], i \neq j} |\langle a_i, a_j \rangle|. \quad (3)$$

See [21] for a survey. Since the columns of such matrices are nearly orthonormal, they are intuitively reasonable choices to ensure $\sigma(A_T) \subseteq [\sqrt{1/2}, \sqrt{3/2}]$. One may leverage the Gershgorin circle theorem to produce such an estimate [2, 5, 17], but this will only guarantee (2) for $s \leq m^{1/2}/\text{polylog}N$. In fact, this estimate is essentially tight since there exist $m \times N$ matrices of minimum coherence with $\Theta(\sqrt{m})$ linearly dependent columns [22, 30]. As an alternative to Gershgorin, Bourgain et al. [9, 10] introduced the so-called flat RIP estimate to demonstrate that certain explicit $m \times N$ matrices with $m = \Theta(N^{1-\epsilon})$ are $s$-restricted isometries for $s = O(m^{1/2+\epsilon})$, where $\epsilon = 10^{-16}$; see [35] for an expository treatment. It was conjectured in [5] that the Paley equiangular tight frames [37] are restricted isometries for even larger values of $\epsilon$, and the flat RIP estimate can be used to prove this, conditional on existing conjectures on cancellations in the Legendre symbol [6].

While it is difficult to obtain explicit $s$-restricted isometries for $s = m/\text{polylog}N$, there have been two approaches to make partial progress: random signals and derandomized matrices. The random signals approach explains a certain observation: While low-coherence $m \times N$ sensing matrices may not determine every $s$-sparse signal with...
\( s = m/\text{polylog}N \), they do determine most of these signals. In fact, even for the \( m \times N \) matrices \( A \) of minimum coherence with \( \Theta(\sqrt{m}) \) linearly dependent columns (indexed by \( T \), say), while \( y = Ax \) fails to uniquely determine any \( x \) with support containing \( T \), it empirically holds that random \( s \)-sparse vectors \( x \) can be reconstructed from \( y = Ax \) by solving (1). This behavior appears to exhibit a phase transition [36], and Tropp proved this behavior up to logarithmic factors in [44]; see also the precise asymptotic estimates conjectured by Haikin, Zamir and Gavish [25] and recent progress in [33].

As another approach, one may seek explicit random matrices that are \( s \)-restricted isometries for \( s = m/\text{polylog}N \) with high probability, but with as little entropy as possible; here, we say a sequence \( \{A_N\} \) of \( m(N) \times N \) random matrices is explicit if there exists an algorithm that on input \( N \) produces \( A_N \) in time that is polynomial in \( N \), assuming access to a random variable that is uniformly distributed over \( [k] := \{1, \ldots, k\} \) for any desired \( k \in \mathbb{N} \). Given a discrete random variable \( X \) that takes values in a finite set \( \mathcal{X} \) (e.g., a random matrix that resides in a finite set with probability 1), the entropy \( H(X) \) of \( X \) is defined by

\[
H(X) := -\sum_{x \in \mathcal{X}} \mathbb{P}\{X = x\} \log_2 \mathbb{P}\{X = x\}.
\]

Observe that a deterministic random variable has entropy 0. At another extreme, the uniform distribution over \( [2^H] \) has entropy \( H \), meaning it takes \( H \) independent tosses of a fair coin to simulate this distribution. This illustrates how \( H(X) \) measures the “amount of randomness” in a random variable \( X \), and so seeking random matrices of minimum entropy represents a natural relaxation of the derandomization problem. Considering it takes \( H \) bits to store the state of a random variable with entropy \( H \), this approach also minimizes the storage required to specify the sensing matrix. One popular random matrix in compressed sensing draws independent entries uniformly over \( \{\pm m^{-1/2}\} \) [8, 14, 20, 34], which has entropy \( H = mN = sN/\text{polylog}N \). One may use the Legendre symbol to derandomize this matrix to require only \( H = s/\text{polylog}N \) random bits [4]. Alternatively, one may draw \( m \) rows uniformly from the \( N \times N \) discrete Fourier transform to get \( H = m \log_2 N \approx s/\text{polylog}N \) [7, 14, 15, 26, 38].

Any choice of Johnson–Lindenstrauss projection [31] with \( m = \text{polylog}N \) is an \( s \)-restricted isometry with high probability [8], but these random matrices inherently require \( H = \Omega(m) \) [4]. To date, it is an open problem to find explicit random \( m \times N \) matrices with \( N^\epsilon \leq m \leq (1 - \epsilon)N \) and \( H \ll s \) that are \( s \)-restricted isometries for \( s = m/\text{polylog}N \) with high probability.

There is another meaningful way to treat the compressed sensing problem: Show that the distribution of a given random \( m \times N \) matrix \( A \) has the property that for every \( s \)-sparse \( x \in \mathbb{C}^N \), there exists a high-probability event \( \mathcal{E}(x) \) over which \( x \) can be reconstructed from \( y = Ax \) by solving (1). This nonuniform setting was originally studied by Candès, Romberg and Tao in [13], and later used to define the Donoho–Tanner phase transition [1, 19]. For applications, the nonuniform setting assumes that a fresh copy of \( A \) is drawn every time a signal \( x \) is to be sensed as \( y = Ax \), and then both \( A \) and \( y \) are passed to the optimizer to solve (1), which is guaranteed to recover \( x \) at least \( \inf_x \mathbb{P}(\mathcal{E}(x)) \) of the time. Notice that if an explicit random matrix is an \( s \)-restricted isometry in the high-probability event \( \mathcal{E} \), then it already enjoys a such
guarantee with $E(x) = \mathcal{E}$ for every $s$-sparse $x \in \mathbb{C}^N$. As such, it is natural to seek a nonuniform guarantee for an explicit random matrix with entropy $H \ll s$.

For any fixed $L \in \mathbb{N}$, we construct explicit $m \times N$ random matrices with $m = spolylogN$ and entropy $s^{1/L}polylogN$ that enjoy a nonuniform performance guarantee. The existence of such matrices is implied by the conjectural existence of pseudorandom number generators [24]. Indeed, consider any explicit $m \times N$ random matrices with $m = spolylogN$ and entropy $H = spolylogN$ that are $s$-restricted isometries with high probability. Simulate the randomness in these matrices with a pseudorandom number generator that stretches bit strings of length $n$ to bit strings of length $n^L$. Doing so results in explicit $m \times N$ random matrices with $m = spolylogN$ and entropy $H^{1/L} = s^{1/L}polylogN$. Furthermore, these low-entropy random matrices enjoy a nonuniform $\ell_1$-recovery guarantee, since otherwise any problematic $s$-sparse vector could be used to distinguish the pseudorandom number generator’s output from true randomness.

In 2014, Hügel, Rauhut and Strohmer [27] applied tools from decoupling to show that certain random matrices that arise in remote sensing applications enjoy a nonuniform $\ell_1$-recovery guarantee. One may directly apply their techniques to obtain an explicit random $m \times N$ partial Fourier matrix $A$ with entropy $\Theta(s^{1/2} \log N \log (N/\epsilon))$ such that each $s$-sparse vector $x \in \mathbb{C}^N$ can be recovered by $\ell_1$ minimization with probability at least $1 - \epsilon$. Explicitly, $A$ is the submatrix of the $N \times N$ discrete Fourier transform with rows indexed by $\{b_i + b_j : i, j \in [n]\}$, where $n = \Theta(s^{1/2} \log (N/\epsilon))$ and $b_1, \ldots, b_n$ are independent random variables with uniform distribution over $\mathbb{F}_N$. In this paper, we generalize this construction to allow for row indices of the form $\{b_1 + \cdots + b_{iL} : i_1, \ldots, i_L \in [n]\}$. Our main result, found in the next section, is that for each $L$, one may take $n = \Theta_L(s^{1/L} \log^4 (N/\epsilon))$ to obtain a nonuniform $\ell_1$-recovery guarantee. In a sense, this confirms our prediction from the previous paragraph, but it does not require the existence of pseudorandom number generators. Our proof hinges on a different decoupling result (Proposition 8) that reduces our key spectral norm estimate to an iterative application of the matrix Bernstein inequality (Proposition 10); see Sect. 3. In Sect. 4, we provide a simplified treatment of the moment method used in [27] to obtain an approximate dual certificate, though generalized for our purposes. Hopefully, similar ideas can be used to produce explicit random $m \times N$ matrices with $m = spolylogN$ and entropy $H \ll s$ that are $s$-restricted isometries with high probability. Also, we note that Iwen [29] identified explicit random $m \times N$ matrices with $m = \Theta(s \log^2 N)$ and entropy $H = \Theta(\log^2 s)$ for which a specialized algorithm enjoys a nonuniform recovery guarantee, and it would be interesting if this level of derandomization could also be achieved with $\ell_1$ recovery.

1.1 Notation

Throughout, we take $e_N : \mathbb{R} \to \mathbb{C}$ defined by $e_N(x) := e^{2\pi ix/N}$. When $T \subseteq [N]$, let $\mathbb{C}^T \subseteq \mathbb{C}^N$ denote the $|T|$-dimensional space of vectors with support $T$. When $A \in \mathbb{C}^{m \times N}$, let $A_T$ be the $\mathbb{C}^{m \times |T|}$ matrix whose columns are indexed by $T$. Similarly, for $x \in \mathbb{C}^N$, let $x_T$ denote the coordinate projection of $x$ to $\mathbb{C}^T$. For $x \in \mathbb{C}^N$, $x_s$
denotes (any of) the best $s$-term approximation(s) of $x$, i.e., $x_s \in \mathbb{C}^N$ is an $s$-sparse vector such that
\[ \|x - x_s\|_1 \leq \|x - y\|_1 \]
for every $s$-sparse $y \in \mathbb{C}^N$. We write
\[ [n]_L := \{(i_1, \ldots, i_L) : i_k \in [n] \text{ for } k \in [L]\}. \]
$I_n^L$ will denote the subset of $[n]_L^L$ consisting of $L$-tuples with distinct entries. $e_t$ will denote the $t$-th Euclidean basis vector of $\mathbb{C}^N$. We will take $N$ to be prime and let $\mathbb{F}_N$ denote the finite field with $N$ elements.

### 2 Main Result

We will use the following random matrix as a compressed sensing matrix:

**Definition 1** Let $N$ be prime and put $m = nL$ for some $n, L \in \mathbb{N}$. Given independent, uniform random variables $b_1, \ldots, b_n$ over $[N]$, then the **Minkowski partial Fourier matrix** is the random $m \times N$ matrix $A$ with rows indexed by $[n]_L^L$, and whose entry at $(I, j) = (i_1, \ldots, i_L, j) \in [n]_L^L \times [N]$ is given by $A_{I,j} := n^{-L/2} \cdot e_N((b_{i_1} + \cdots + b_{i_L}) j)$. Note that the rows of $A$ are proportional to a subset of the rows of the $N \times N$ discrete Fourier transform matrix. In particular, the rows whose indices reside in the $L$-fold Minkowski sum of the set $\{b_{i_1}, \ldots, b_{i_L}\}$.

We take $N$ to be prime for convenience; we suspect that our results also hold when $N$ is not prime, but the proofs would be more complicated. To perform compressed sensing, we sense with the random matrix $A$ to obtain data $y$ and then we solve the following program:

\[
\begin{align*}
\text{minimize} & \quad \|z\|_1 \quad \text{subject to} \quad \|Az - y\|_2 \leq \eta. \\
\end{align*}
\]

Our main result states that if $x \in \mathbb{C}^N$ is nearly sparse, $n = s^{1/L} \log^4(N/\epsilon)$, and the noisy data $y = Ax + e$ satisfies $\|e\|_2 \leq \eta$, then the minimizer of (3) is a good approximation of $x$ with probability at least $1 - \epsilon$:

**Theorem 2** *(main result)* Fix $L \in \mathbb{N}$. There exists $C_L > 0$ depending only on $L$ such that the following holds. Given any prime $N$, positive integer $s \leq N$, and $\epsilon \in (0, 1)$, select any integer $n \geq C_L s^{1/L} \log^4(N/\epsilon)$. Then the Minkowski partial Fourier matrix $A$ with parameters $(N, n, L)$ has the property that for any fixed signal $x \in \mathbb{C}^N$ and noise $e \in \mathbb{C}^n$ with $\|e\|_2 \leq \eta$, then given random data $y = Ax + e$, the minimizer $\hat{x}$ of (3) satisfies the estimate

\[
\|\hat{x} - x\|_2 \leq 25 \cdot (\sqrt{s} \cdot \eta + \|x - x_s\|_1)
\]

with probability at least $1 - \epsilon$. 

© Birkhäuser
This result shows that an $m \times N$ Minkowski partial Fourier matrix with parameters $(N, n, L)$ will succeed at recovering $s$-sparse $x \in \mathbb{C}^N$ when $m = \lceil C_L s \log^{4L}(N/\epsilon) \rceil$ and with entropy at most $C''_L s^{1/L} \log^5(N/\epsilon)$, where $C'_L, C''_L$ are constants depending only on $L$. Therefore, this random matrix allows one to solve large compressed sensing instances with minimal sample complexity and low entropy, but note the trade-off between sample complexity and entropy. In particular, one may drive down the entropy’s sensitivity to $s$ by taking $L$ to be large, but this incurs a larger exponent in the sample complexity’s log factor. Furthermore, the $L$-dependent constant factors in both the sample complexity and entropy are sensitive functions of $L$. Indeed, while we have not made an effort to determine the precise growth rate of $C_L$, the constant appearing in Theorem 3.4.1 in [16] suggests that $C_L$ will grow super-exponentially in $L$. As such, one should only expect our random matrix with $L > 1$ to offer an improvement over the standard $L = 1$ setting in the regime where $s$ and $N$ are large.

To prove Theorem 2, we construct an approximate dual certificate that satisfies the hypotheses of the following proposition.

**Proposition 3** (Theorem 3.1 in [27], cf. Theorem 4.33 in [23]) Fix a signal $x \in \mathbb{C}^N$ and a measurement matrix $A \in \mathbb{C}^{m \times N}$ with unit-norm columns. Fix $s \in [N]$ and any $T \subseteq [N]$ of size $s$ with $\|x - x_T\|_1 = \|x - x_s\|_1$. Suppose

$$\|A_T^* A_T - I\|_2 \leq 1/2 \quad (4)$$

and that there exists $v \in \mathbb{C}^m$ such that $u := A^* v$ satisfies

$$u_T = \text{sgn}(x)_T, \quad \|u_T\|_\infty \leq 1/2, \quad \|v\|_2 \leq \sqrt{2s}. \quad (5)$$

Then for every $e \in \mathbb{C}^m$ with $\|e\|_2 \leq \eta$, given random data $y = Ax + e$, the minimizer $\hat{x}$ of (3) satisfies the estimate

$$\|\hat{x} - x\|_2 \leq 25 \cdot (\sqrt{s} \cdot \eta + \|x - x_s\|_1).$$

When applying Proposition 3, we will make use of a few intermediate lemmas, namely, Lemmas 4, 6, and 7. In this section, we prove Lemma 4 and sketch the proofs of Lemmas 6 and 7 before using these lemmas to prove Theorem 2. Sections 3 and 4 are then dedicated to the (lengthy) proofs of Lemmas 6 and 7. First, the following lemma gives that the columns of $A$ have low coherence with high probability:

**Lemma 4** Fix $L \in \mathbb{N}$. There exists $C_L^{(1)} > 0$ depending only on $L$ such that the following holds. Given any $N, n \in \mathbb{N}$, let $\{a_i\}_{i \in [N]}$ denote the column vectors of the Minkowski partial Fourier matrix $A$ with parameters $(N, n, L)$. Then

$$\max_{i, j \in [N]} |\langle a_i, a_j \rangle| \leq C_L^{(1)} \cdot n^{-L/2} \cdot \log^{L/2}(N/\epsilon)$$

with probability at least $1 - \frac{\xi}{3}$.
The proof of Lemma 4 follows from an application of the complex Hoeffding inequality:

**Proposition 5 (complex Hoeffding)** Suppose $X_1, \ldots, X_n$ are independent complex random variables with

$$\mathbb{E} X_i = 0 \quad \text{and} \quad |X_i| \leq K \quad \text{almost surely}.$$ 

Then for every $t \geq 0$, it holds that

$$\mathbb{P}\left\{ \left| \sum_{i=1}^{n} X_i \right| > t \right\} \leq 4 \exp \left( - \frac{t^2}{2nK^2} \right).$$

**Proof** Write $X_i = A_i + \sqrt{-1} \cdot B_i$. Then $|\sum_i X_i|^2 = |\sum_i A_i|^2 + |\sum_i B_i|^2$, and so

$$\mathbb{P}\left\{ \left| \sum_{i=1}^{n} X_i \right| > t \right\} \leq \mathbb{P}\left\{ \left| \sum_{i=1}^{n} A_i \right| > \frac{t}{\sqrt{2}} \right\} + \mathbb{P}\left\{ \left| \sum_{i=1}^{n} B_i \right| > \frac{t}{\sqrt{2}} \right\} \leq 4 \exp \left( - \frac{t^2}{2nK^2} \right),$$

where the last step applies Hoeffding’s inequality, namely, Theorem 2 in [28].

**Proof of Lemma 4** For each $j \in [N]$, consider $v_j \in \mathbb{C}^n$ whose $i$-th entry is $e_N(b_{ij})$, and observe that $a_j = n^{-L/2} \cdot v_j \otimes L$, where $v_j \otimes L$ denotes the Kronecker product $v_j \otimes \cdots \otimes v_j$ with $L$ factors. Apply the union bound and Proposition 5 to get

$$\mathbb{P}\left\{ \max_{j \neq k} |\langle v_j, v_k \rangle| > t \right\} \leq \sum_{j,k} \mathbb{P}\left\{ \left| \sum_{i=1}^{n} e_N(b_{ij}-k) \right| > t \right\} \leq 4N^2 e^{-t^2/(2n)} \leq \frac{\epsilon}{3},$$

where the last step takes $t := (C_L^{(1)} 1/L \cdot n^{1/2} \cdot \log^{1/2}(N/\epsilon))$. It follows that

$$\max_{j,k} |\langle a_j, a_k \rangle| = n^{-L} \max_{j,k} |\langle v_j, v_k \rangle|^L \leq C_L^{(1)} \cdot n^{-L/2} \cdot \log^{L/2}(N/\epsilon)$$

with probability at least $1 - \frac{\epsilon}{3}$, as desired.

Next, we show that the bound in (4) is satisfied with probability close to 1:

**Lemma 6** Fix $L \in \mathbb{N}$. There exists $C_L^{(2)} > 0$ depending only on $L$ such that the following holds. Given any prime $N$, positive integer $s \leq N$, and $\epsilon \in (0, 1)$, select any integer $n \geq C_L^{(2)} s^{1/L} \log^{3}(N/\epsilon)$. Then for every $T \subseteq [N]$ of size $|T| = s$, the Minkowski partial Fourier matrix $A$ with parameters $(N, n, L)$ satisfies $\|A_T^* A_T - I\|_{2 \to 2} \leq \frac{1}{\epsilon}$ with probability at least $1 - \frac{\epsilon}{3}$. 

\[ \text{Birkhäuser} \]
The proof of Lemma 6 appears in Sect. 3. Note that the failure rate in this lemma is too big to provide an RIP constant from the union bound over \( \binom{N}{s} \) choices for \( T \).

In what follows, we briefly sketch the approach and technical obstacles in Lemma 6. \( A_T^*A_T - I \) may be decomposed as as a sum of \( n^L \) rank-1 outer products of the rows of \( A_T \). The row \((i_1, \ldots, i_L) \in [n]^L\) of \( A_T \) relies only on the random variables \( (b_{i_1}, \ldots, b_{i_L}) \), and so we may write the corresponding rank-1 outer product as \( h(b_{i_1}, \ldots, b_{i_L}) \), which has \((i, j)\) entry \( e_N((b_{i_1} + \cdots + b_{i_L})(j - i)) \). Exploiting the multiplicative nature of \( e_N \), we may write each term as a product of diagonal matrices \( D_{b_k} \) depending only on \( b_k \). In particular, we write

\[
A_T^*A_T - I = \sum_{(i_1, \ldots, i_L) \in [n]^L} D_{b_{i_L}} \cdots D_{b_{i_1}} (11^* - I) D_{b_{i_1}}^* \cdots D_{b_{i_L}}^*.
\]

The principal difficulty is the dependence between different terms in the sum. Proposition 8 (i.e., Theorem 3.4.1 in [16]) is a decoupling inequality that allows us to relate the tail of the spectral norm of this sum to the tail of the spectral norm of a sum with less dependence between terms. Explicitly, it enables us to pass to

\[
\sum_{(i_1, \ldots, i_L) \in [n]^L} D_{b_i^{(L)}} \cdots D_{b_i^{(1)}} (11^* - I) D_{b_i^{(1)}}^* \cdots D_{b_i^{(L)}}^*,
\]

where \( b_i^{(j)} \) is now an array of \( nL \) independent random variables all uniform over \([N]\). This additional independence permits an iterative application of the matrix Bernstein inequality, where first the tail of \( \sum_{i_1} D_{b_i^{(1)}} (11^* - I) D_{b_i^{(1)}}^* \) is bounded, then the tail of

\[
\sum_{i_1} \sum_{i_2} D_{b_i^{(2)}} D_{b_i^{(1)}} (11^* - I) D_{b_i^{(1)}}^* D_{b_i^{(2)}}^*,
\]

and so on. This iterative argument is given in Lemma 9.

Finally, we describe how we use the hypothesis in Lemma 6 that \( N \) is prime. Proposition 8 does not apply to sums indexed over \([n]^L\), as we desire, but to sums over the tuples \((i_1, \ldots, i_L)\) with distinct elements. We handle this by grouping all tuples \((i_1, \ldots, i_L)\) with repeated elements and applying Proposition 8 many times. As an illustrative example, consider all tuples \((i_1, \ldots, i_L)\) with \( i_1 = i_2 \) and no other equalities. Then \( h(b_{i_1}, \ldots, b_{i_L}) \) will have entries \( e_N((2bi_1 + bi_3 + \cdots + bi_L)(j - i)) \).

Since \( kb_{i_1} \) is also uniform on \([N]\) for any \( k \in \{1, \ldots, N - 1\} \), we may handle all such \( L\)-tuples as \((L - 1)\)-tuples with distinct elements, which allows analysis via Proposition 8.

We will soon prove Theorem 2 by applying Proposition 3 with the choice \( v = (A_T^*)^\dagger \text{sgn}(x)_T \). Most of the difficulty arises from verifying that \( \|u_{T^c}\|_\infty \leq \frac{1}{2} \), which is equivalent to showing that for each \( u \in T^c \), the complex scalar

\( Birkhäuser \)
is shown that the constraints on the integers \( \ell \) on any \( z \in \text{Fix } N \) are controlled by our final technical lemma:

**Lemma 7** Fix \( N, n, L, s \in \mathbb{N} \) such that \( n \geq 2s^{1/L} \), and take any \( T \subseteq [N] \) of size \( s \) and any \( z \in \mathbb{C}^T \) with \( \|z\|_\infty \leq 1 \). Let \( \{a_i\}_{i \in [N]} \) denote the column vectors of the Minkowski partial Fourier matrix \( A \) with parameters \( (N, n, L) \). Take any \( u \in T^c, k \in \mathbb{N} \) and \( p \geq 2 \), and put \( \eta := (k + 1)Lp \). Then

\[
\mathbb{E}|a_u^* A_T (I - A_T^* A_T)^k z|^p \leq 2\eta^{2\eta} (s^{1/L} n^{-1})^{\eta/2}.
\]

In what follows, we sketch the main ideas of the proof of Lemma 7, leaving the formal proof for Sect. 4. The expression \( |a_u^* A_T (I - A_T^* A_T)^k z|^p \), when \( p = 2M \) is an even integer, is a matrix–vector product involving only entries of \( A, z \), and their conjugates. Further, each entry of \( A \) is a product of complex exponentials \( e_N(kb_i) \) for \( i \in [n] \) and \( k \in \mathbb{Z} \). Therefore, we claim that a straightforward calculation using linearity of expectation and independence of \( b_i, i \in [n] \), yields that

\[
\mathbb{E}|a_u^* A_T (I - A_T^* A_T)^k z|^{2M} = \frac{1}{n^D} \sum_{a} \prod_{i=1}^{n} \mathbb{E}(e_N(c_{a,i} b_i)),
\]

for some \( c_{a,i} \in \mathbb{Z} \) and \( D \in \mathbb{N} \). These expectations are 1 when \( c_{a,i} = 0 \mod N \) and 0 otherwise, and so we may compute this moment by counting the number of nonzero terms of this sum.

It is natural to write this sum with \((k+1)(2M)\) integer indices in \( T \), corresponding to the \((k+1)(2M)\) matrix products involving \( A_T^* A_T \). Additionally, each entry of \( A^* A \) is itself a sum over \([n]^L \), and so an additional \( n^L \) integer indices are needed for each of the \((k+1)(2M)\) factors of \( A_T^* A_T \). It is most convenient to organize these indices as \( \ell \in T^{(k+1)M} \) and \( j \) an arbitrary function \([k+1] \times [2M] \times [L] \rightarrow [n] \), where it may be helpful to recall that such a function is, formally, a list of integer tuples. Then we may rewrite the previous expression as

\[
\frac{1}{n^D} \sum_{\ell \in \mathcal{L}} \sum_{j \in \mathcal{J}} \prod_{i=1}^{n} \mathbb{E}(e_N(c_{\ell,i} b_i)) = \left\{ (\ell, j) \in \mathcal{L} \times \mathcal{J} : c_{\ell,i} = 0 \mod N, i \in [n] \right\}.
\]

See the discussion preceding Eq. (22) for precise definitions of \( \mathcal{L} \) and \( \mathcal{J} \). In Eq. (23), it is shown that the constraints on the integers \( c_{\ell,j,i} \) amount to \( n \) affine linear constraints on \( \ell \) as a member of the vector space \( \mathbb{P}_N^{(k+1)(2M)} \), where each constraint \( i \in [n] \) is determined by \( j^{-1}(i) \). We bound the size of this set using the following ideas:

1. We show that the number of linearly independent affine constraints on \( \ell \) can be estimated from the size of the image of \( j \). This is the subject of Lemma 11.
2. When there are \( r \) linearly independent constraints on \( \ell \), there are at most \(|T|^{(k+1)(2M) - r}\) indices \( \ell \) that can satisfy those constraints.
3. For technical reasons discussed in the proof of Lemma 7, the constraints cannot be satisfied when \( j \in \mathcal{J} \) has any preimage of size 1. Although there are \( n^L \) functions \([L] \rightarrow [n]\), there are less than \( L^{L+1} n^{L/2} \) having no preimage of size 1.

These ideas lead to the bounds in Lemma 7 in the case where \( p \) is an even integer. A brief interpolation argument then gives the result for arbitrary \( p \geq 2 \).

**Proof of Theorem 2** We will restrict to an event of the form \( \mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3 \) such that the hypotheses of Proposition 3 are satisfied over \( \mathcal{E} \) and each \( \mathcal{E}_i \) has probability at least \( 1 - \frac{\epsilon}{3} \). By Lemma 6, we may take

\[
\mathcal{E}_1 := \{ \| A_T^* A_T - I \|_{2 \rightarrow 2} \leq 1/e \}.
\]

Then (4) is satisfied over \( \mathcal{E}_1 \supseteq \mathcal{E} \). For (5), we take \( v := (A_T^*)^\dagger \text{sgn}(x)_T \). Since \( A_T \) has trivial kernel over \( \mathcal{E}_1 \), we may write \((A_T^*)^\dagger = A_T (A_T^* A_T)^{-1} \), and so

\[
u_T = (A_T^* v)_T = A_T^* v = A_T^* (A_T^*)^\dagger \text{sgn}(x)_T = A_T^* A_T (A_T^* A_T)^{-1} \text{sgn}(x)_T = \text{sgn}(x)_T.
\]

Next, the nonzero singular values of any \( B \) and \( B^\dagger \) are reciprocal to each other, and so

\[
\|v\|_2 = \|(A_T^*)^\dagger \text{sgn}(x)_T\|_2 \leq \|(A_T^*)^\dagger\|_2 \cdot \|\text{sgn}(x)_T\|_2 \leq (1 - 1/e)^{-1/2} \cdot s^{1/2} \leq \sqrt{2s}.
\]

As such, for (5), it remains to verify that \( \|u_Tc\|_\infty \leq 1/2 \). We will define \( \mathcal{E}_2 \) and \( \mathcal{E}_3 \) in such a way that Lemmas 4 and 7 imply this bound. Since \( \|I - A_T^* A_T\|_{2 \rightarrow 2} < 1 \), we may write \((A_T^* A_T)^{-1} = \sum_{k=0}^{\infty} (I - A_T^* A_T)^k \). For any \( \omega \in \mathbb{N} \), the triangle inequality then gives

\[
\|u_T c\|_\infty = \|A_T^* v\|_\infty
= \|A_T^* A_T (A_T^* A_T)^{-1} \text{sgn}(x)_T\|_\infty
\leq \left\| A_T^* A_T \sum_{k=0}^{\omega-1} (I - A_T^* A_T)^k \text{sgn}(x)_T \right\|_\infty
+ \left\| A_T^* A_T \sum_{k=\omega}^{\infty} (I - A_T^* A_T)^k \text{sgn}(x)_T \right\|_\infty.
\]

Given \( B \in \mathbb{C}^{a \times b} \), denote \( \|B\|_\infty := \max_{i \in [a], j \in [b]} |B_{ij}| \). Observe that \( \|B\|_{\infty \rightarrow \infty} \leq b \|B\|_\infty \), and recall that \( \|B\|_{\infty \rightarrow \infty} \leq \sqrt{b} \|B\|_{2 \rightarrow 2} \). We apply these estimates to obtain

\[
(7) \leq s \|A_T^* A_T\|_{\infty \rightarrow \infty} \cdot \left( \sum_{k=\omega}^{\infty} (I - A_T^* A_T)^k \right)_{\infty \rightarrow \infty}
\leq s \|A_T^* A_T\|_{\infty \rightarrow \infty} \cdot \sqrt{s} \left( \sum_{k=\omega}^{\infty} \|I - A_T^* A_T\|_{2 \rightarrow 2} \right) \leq s^{3/2} \cdot \max_{i,j \in [N]} |\langle a_i, a_j \rangle| \cdot \sum_{k=\omega}^{\infty} e^{-k}.
\]
Next, by Lemma 4, we may take

$$\mathcal{E}_2 := \left\{ \max_{i \neq j} \sum_{i,j \in [N]} |\langle a_i, a_j \rangle| \leq C_L^{(1)} \cdot n^{-L/2} \cdot \log^{L/2}(N/\epsilon) \right\}.$$ 

Then since $n \geq (C_L^{(1)})^{2/L} \cdot s^{1/L} \cdot \log(N/\epsilon)$ by assumption, we have

$$\mathcal{E}_2 \subseteq \mathcal{E}, \quad \text{where the last step follows from selecting } \omega := \lceil 2 \log N \rceil, \text{ say.}$$

It remains to establish (6) $\leq 1/4$. To this end, we define

$$\mathcal{E}_3 := \{(6) \leq 1/4\},$$

and we will use Lemma 7 to prove $\mathbb{P}(\mathcal{E}_3^c) \leq \epsilon/3$ by the moment method. For every choice of $\beta_0, \ldots, \beta_\omega > 0$ satisfying $\sum_{k=0}^{\omega-1} \beta_k \leq 1/4$ and $p_0, \ldots, p_{\omega-1} \geq 2$, the union bound gives

$$\mathbb{P}(\mathcal{E}_3^c) \leq \sum_{u \in \mathcal{T}} \mathbb{P} \left\{ \sum_{k=0}^{\omega-1} |a_u^* A_T (I - A_T^* A_T)^k \text{sgn}(x)_T| \geq \beta_k \right\} \leq \sum_{u \in \mathcal{T}} \mathbb{P} \left\{ \sum_{k=0}^{\omega-1} |a_u^* A_T (I - A_T^* A_T)^k \text{sgn}(x)_T| > \beta_k \right\} \leq \sum_{u \in \mathcal{T}} \mathbb{P} \left\{ \sum_{k=0}^{\omega-1} \beta_k^{-p_k} \cdot \mathbb{E} |a_u^* A_T (I - A_T^* A_T)^k \text{sgn}(x)_T|^{p_k} \right\},$$

where the last step follows from Markov’s inequality. For simplicity, we select $\beta_k := 5^{-(k+1)}$. Put $\eta_k := (k+1)Lp_k$. Then Lemma 7 gives that the $k$-th term of (8) satisfies

$$\beta_k^{-p_k} \cdot \mathbb{E} |a_u^* A_T (I - A_T^* A_T)^k \text{sgn}(x)_T|^{p_k} \leq 5^{(k+1)p_k} \cdot 2^{2\eta_k} \cdot (s^{1/L} n^{-1})^{\eta_k/2} = 2 \exp \left( 2\eta_k \cdot \log \left( s^{1/(2L)} \cdot (s^{1/L} n^{-1})^{1/4} \right) \right),$$

and the right-hand side is minimized when the inner logarithm equals $-1$, i.e., when

$$\eta_k = e^{-1} 5^{-1/(2L)} n^{1/4} s^{-1/(4L)}.$$
In particular, the optimal choice of $\eta_k$ does not depend on $k$. It remains to verify that $p_k \geq 2$ for every $k$ (so that Lemma 7 applies) and that $(8) \leq \epsilon/3$. Since $\epsilon < 1$, we have

$$n \geq C_L s^{1/L} \log^4 (N/\epsilon) \geq C_L s^{1/L} \log^4 N,$$

which combined with (9) and $k + 1 \leq \omega \leq 3 \log N$ implies that $p_k \geq 2$. Also,

$$(8) \leq N \cdot \omega \cdot 2e^{-2\eta_0} = \exp \left( \log N + \log \omega + \log 2 - 2e^{-1.5^{1/(2L)} n^{1/4}s^{-1/(4L)}} \right),$$

and the right-hand side is less than $\epsilon/3$ since $n \geq C_L s^{1/L} \log^4 (N/\epsilon)$. \qed

3 Proof of Lemma 6

Recall that the rows in the submatrix $A_T$ are not statistically independent. To overcome this deficiency, our proof of Lemma 6 makes use of the following decoupling estimate:

**Proposition 8** (Theorem 3.4.1 in [16]) Fix $k \in \mathbb{N}$. There exists a constant $C_k > 0$ depending only on $k$ such that the following holds. Given independent random variables $X_1, \ldots, X_n$ in a measurable space $S$, a separable Banach space $B$, and a measurable function $h : S^k \to B$, then for every $t > 0$, it holds that

$$P \left\{ \left\| \sum_{(i_1, \ldots, i_k) \in I_n^k} b(X_{i_1}, \ldots, X_{i_k}) \right\|_B > t \right\} \leq C_k \cdot P \left\{ \left\| \sum_{(i_1, \ldots, i_k) \in I_n^k} b(\tilde{X}_{i_1}^{(1)}, \ldots, \tilde{X}_{i_k}^{(k)}) \right\|_B > t \right\},$$

where for each $i \in [n]$ and $\ell \in [k]$, $X_i^{(\ell)}$ is an independent copy of $X_i$.

Proposition 8 will allow us to reduce Lemma 6 to the following simpler result, which uses some notation that will be convenient throughout this section: For a fixed $T \subseteq [N]$, let $X_0$ denote the matrix $11^* - I \in \mathbb{C}^{T \times T}$, and for each $x \in [N]$, let $D_x \in \mathbb{C}^{T \times T}$ denote the diagonal matrix whose $t$-th diagonal entry is given by $e^N(-xt)$.

**Lemma 9** Fix $L \in \mathbb{N}$. There exists a constant $\tilde{C}_L^{(2)}$ depending only on $L$ such that the following holds. Select any $s, n, N \in \mathbb{N}$ such that $N > nL \geq 1$, and any $T \subseteq [N]$ with $|T| = s$. Draw $\{b_i^{(j)}\}_{i \in [n], j \in [L]}$ independently with uniform distribution over $[N]$. Then for each $\ell \in [L]$ and $\alpha \in (0, L^{-1})$, it holds that

$$P \left\{ \left\| \sum_{i_1, \ldots, i_L \in [n]} D_{b_{i_L}^{(\ell)}} \cdots D_{b_{i_1}^{(1)}} X_0 D_{b_{i_L}^{(1)}}^* \cdots D_{b_{i_1}^{(1)}}^* \right\|_{2 \to 2} > \tilde{C}_L^{(2)} n^{L/2} s^{1/2} \log^{3L/2} (N/\alpha) \right\} \leq L\alpha.$$

We prove Lemma 9 by an iterative application of both the complex Hoeffding and matrix Bernstein inequalities:
Proposition 10 (matrix Bernstein, Theorem 1.4 in [45]) Suppose $X_1, \ldots, X_n$ are independent random $d \times d$ self-adjoint matrices with

$$\mathbb{E}X_i = 0 \quad \text{and} \quad \|X_i\|_2 \leq K \quad \text{almost surely.}$$

Then for every $t \geq 0$, it holds that

$$\mathbb{P}\left\{ \left\| \sum_{i=1}^{n} X_i \right\|_2 > t \right\} \leq 2d\exp\left(-\frac{t^2}{2\sigma^2 + \frac{2}{3}Kt}\right), \quad \sigma^2 := \left\| \sum_{i=1}^{n} \mathbb{E}X_i^2 \right\|_2.$$

Proof of Lemma 9 Iteratively define $X_\ell := \sum_{j=1}^{n} D_{b_j}^{(\ell)}X_{\ell-1}D_{b_j}^{*}$. Then our task is to prove

$$\mathbb{P}\{\|X_\ell\|_2 > C^{(2)}_L n^{L/2}s^{1/2} \log^{3L/2}(N/\alpha)\} \leq L\alpha$$

for every $\ell \in [L]$ and $\alpha \in (0, L^{-1})$. To accomplish this, first put

$$\sigma^2_\ell(X_\ell) := \left\| \sum_{j=1}^{n} \mathbb{E}\left[ (D_{b_j}^{(\ell+1)}X_\ell D_{b_j}^{*})^2 \right| X_\ell \right\|_2 \leq \left\| n \cdot \text{diag}(\text{diag}(X_\ell^2)) \right\|_2 = n \max_{t \in T} \|X_\ell e_t\|^2,$$

where the last step uses the fact that $X_\ell$ is self-adjoint. Next, fix thresholds $u_0, \ldots, u_L > 0$ and $v_0, \ldots, v_{L-1} > 0$ (to be determined later), and define events

$$A_\ell := \{\|X_\ell\|_2 \leq u_\ell\}, \quad B_\ell := \{\sigma^2_\ell(X_\ell) \leq v_\ell\}, \quad E_\ell := \bigcap_{i=0}^{\ell} B_i, \quad E := E_{L-1}.$$

In this notation, our task is to bound $\mathbb{P}(A_\ell^c)$ for each $\ell \in [L]$.

First, we take $v_0 = ns$ so that $E_0^c = B_0^c$ is empty, i.e., $\mathbb{P}(E_0^c) = 0$. Now fix $\ell \geq 0$ and suppose $\mathbb{P}(E_\ell^c) \leq \ell\alpha/2$. Then we can condition on $E_\ell$, and

$$\mathbb{P}(E_{\ell+1}^c) = \mathbb{P}(E_{\ell+1}^c \cap E_\ell) + \mathbb{P}(E_{\ell+1}^c \cap E_\ell^c) \leq \mathbb{P}(E_{\ell+1}^c \cap E_\ell) + \mathbb{P}(E_\ell^c) = \mathbb{P}(B_{\ell+1}^c|E_\ell) + \mathbb{P}(E_\ell^c).$$

Later, we will apply Proposition 5 to obtain the bound

$$\mathbb{P}(B_{\ell+1}^c|E_\ell) \leq \alpha/2,$$

which combined with (11) implies that $\mathbb{P}(E_{\ell+1}^c) \leq (\ell + 1)\alpha/2$. By induction, we have

$$\mathbb{P}(E^c) \leq L\alpha/2,$$
and so we can condition on \( E \). Next, we take \( u_0 = s \) so that \( A_0^c \) is empty, i.e., \( \mathbb{P}(A_0^c | E) = 0 \). Now fix \( \ell \geq 0 \) and suppose \( \mathbb{P}(A_\ell^c | E) \leq \ell \alpha/2 \). Then we can condition on \( A_\ell \cap E \), and

\[
\mathbb{P}(A_{\ell+1}^c | E) = \mathbb{P}(A_{\ell+1}^c | A_\ell \cap E) \mathbb{P}(A_\ell | E) + \mathbb{P}(A_{\ell+1}^c \cap A_\ell^c | E) \\
\leq \mathbb{P}(A_{\ell+1}^c | A_\ell \cap E) + \mathbb{P}(A_\ell^c | E).
\]

(14)

Later, we will apply Proposition 10 to obtain the bound

\[
\mathbb{P}(A_{\ell+1}^c | A_\ell \cap E) \leq \alpha/2,
\]

(15)

which combined with (14) implies

\[
\mathbb{P}(A_{\ell+1}^c | E) \leq (\ell + 1)\alpha/2.
\]

(16)

Combining (13) and (16) then gives

\[
\mathbb{P}(A_{\ell+1}^c) = \mathbb{P}(A_{\ell+1}^c | E) \mathbb{P}(E) + \mathbb{P}(A_{\ell+1}^c \cap E^c) \\
\leq \mathbb{P}(A_{\ell+1}^c | E) + \mathbb{P}(E^c) \leq L\alpha,
\]

as desired. Overall, to prove (10), it suffices to select thresholds

\[
0 < u_1 \leq u_2 \leq \cdots \leq u_L \leq \tilde{C}_L^{(2)} n^{L/2} s^{1/2} \log^{3L/2} (N/\alpha)
\]

(17)

and \( v_1, \ldots, v_{L-1} > 0 \) in such a way that (12) and (15) hold simultaneously.

In what follows, we demonstrate (12). Let \( \mathbb{P}_\ell \) denote the probability measure obtained by conditioning on the event \( E_\ell \), and let \( \mathbb{E}_\ell \) denote expectation with respect to this measure. As we will see, Proposition 5 implies that \( \mathbb{P}_\ell(B_{\ell+1}^c | X_\ell) \leq \alpha/2 \) holds \( \mathbb{P}_\ell \)-almost surely, which in turn implies

\[
\mathbb{P}(B_{\ell+1}^c | E_\ell) = \mathbb{E}_\ell[\mathbb{P}_\ell(B_{\ell+1}^c | X_\ell)] \leq \alpha/2
\]

by the law of total probability. For each \( t \in T \) and \( x \in [N] \), let \( D'_{t,x} \in \mathbb{C}^{T \times T} \) denote the diagonal matrix whose \( t \)-th diagonal entry is 0, and whose \( u \)-th diagonal entry is \( e_N(-xu) \) for \( u \in T \setminus \{t\} \). In particular, \( D'_{t,x} \) is identical to \( D_x \), save the \( t \)-th diagonal entry. Then since the \( t \)-th entry of \( X_\ell e_t \) is 0, we may write

\[
X_{\ell+1} e_t = \sum_{j=1}^n D_{b_j^{(\ell+1)} t} X_t d_{b_j^{(\ell+1)} t} e_t = \sum_{j=1}^n e_N(b_j^{(\ell+1)} t \cdot D_{b_j^{(\ell+1)} t} X_\ell e_t = \sum_{j=1}^n e_N(b_j^{(\ell+1)} t \cdot D'_{t,b_j^{(\ell+1)} t} X_\ell e_t.
\]

We then take norms to get

\[
\|X_{\ell+1} e_t\|_2^2 \leq \sum_{j=1}^n e_N(b_j^{(\ell+1)} t \cdot D'_{t,b_j^{(\ell+1)} t}) \|X_\ell e_t\|_2 = \max_{u \in T \setminus \{t\}} \|X_\ell e_t\|_2 \left( \sum_{j=1}^n e_N(b_j^{(\ell+1)} (t-u)) \right) \|X_\ell e_t\|_2.
\]

Birkhäuser
Recalling the definition of \( \sigma^2(X_{\ell}) \), it follows that

\[
\sigma^2_{\ell+1}(X_{\ell+1}) = n \cdot \max_{t \in T} \|X_{\ell+1} e_t\|^2_F \\
\leq \sigma^2(X_{\ell}) \max_{t \neq u} \left| \sum_{j=1}^{n} e_N(b^{(\ell+1)}_j (t - u)) \right|^2 \leq \frac{v_{\ell+1}}{v_{\ell}} \max_{t \neq u} \left| \sum_{j=1}^{n} e_N(b^{(\ell+1)}_j (t - u)) \right|^2,
\]

where the last step holds \( \mathbb{P}_\ell \)-almost surely since \( B_{\ell} \supseteq E_{\ell} \). As such, the following bound holds \( \mathbb{P}_\ell \)-almost surely:

\[
\mathbb{P}_\ell(B^c_{\ell+1} | X_{\ell}) = \mathbb{P}_\ell((\sigma^2_{\ell+1}(X_{\ell+1}) > v_{\ell+1}) | X_{\ell}) \\
\leq \mathbb{P}_\ell\left\{ \max_{t,u \in T, t \neq u} \left| \sum_{j=1}^{n} e_N(b^{(\ell+1)}_j (t - u)) \right|^2 > \frac{v_{\ell+1}}{v_{\ell}} \right\} \\
\leq \sum_{t,u \in T, t \neq u} \mathbb{P}_\ell\left\{ \left| \sum_{j=1}^{n} e_N(b^{(\ell+1)}_j (t - u)) \right|^2 > \frac{v_{\ell+1}}{v_{\ell}} \right\},
\]

where the last step applies the union bound. Next, for each \( t, u \in T \) with \( t \neq u \), it holds that \( \{e_N(b^{(\ell+1)}_j (t - u))\}_{j \in [n]} \) are \( \mathbb{P}_\ell \)-independent complex random variables with zero mean and unit modulus \( \mathbb{P}_\ell \)-almost surely. We may therefore apply Proposition 5 to continue:

\[
\mathbb{P}_\ell(B^c_{\ell+1} | X_{\ell}) \leq 4s^2 \exp \left( -\frac{v_{\ell+1}}{2nv_{\ell}} \right) \quad \mathbb{P}_\ell \text{-almost surely.}
\]

As such, selecting \( v_{\ell+1} := 2n \log(8s^2 / \alpha) \cdot v_{\ell} \) ensures that \( \mathbb{P}_\ell(B^c_{\ell+1} | X_{\ell}) \leq \alpha/2 \) holds \( \mathbb{P}_\ell \)-almost surely, as desired.

Next, we demonstrate (15). For convenience, we change the meanings of \( \mathbb{P}_\ell \) and \( \mathbb{E}_\ell \): Let \( \mathbb{P}_\ell \) denote the probability measure obtained by conditioning on the event \( A_{\ell} \cap E \), and let \( \mathbb{E}_\ell \) denote expectation with respect to this measure. As we will see, Proposition 10 implies that \( \mathbb{P}_\ell(A^c_{\ell+1} | X_{\ell}) \leq \alpha/2 \) holds \( \mathbb{P}_\ell \)-almost surely, which in turn implies

\[
\mathbb{P}(A^c_{\ell+1} | A_{\ell} \cap E) = \mathbb{E}_\ell[\mathbb{P}_\ell(A^c_{\ell+1} | X_{\ell})] \leq \alpha/2
\]

by the law of total probability. Let \( \mathcal{X} \subseteq \mathbb{C}^T \times T \) denote the support of the discrete distribution of \( X_{\ell} \) under \( \mathbb{P}_\ell \). For each \( x \in \mathcal{X} \), let \( \mathbb{P}_{\ell,x} \) denote the probability measure obtained by conditioning \( \mathbb{P}_\ell \) on the event \( \{X_{\ell} = x\} \). Then \( \{D_{b^{(\ell+1)}_j} X_{\ell} D_{b^{(\ell+1)}_j}^*\}_{j \in [n]} \) are \( \mathbb{P}_{\ell,x} \)-independent random \( s \times s \) self-adjoint matrices with

\[
\mathbb{E}_{\ell,x} D_{b^{(\ell+1)}_j} X_{\ell} D_{b^{(\ell+1)}_j}^* = 0, \quad \|D_{b^{(\ell+1)}_j} X_{\ell} D_{b^{(\ell+1)}_j}^*\|_{2 \to 2} \leq \|x\|_{2 \to 2} \quad \mathbb{P}_{\ell,x} \text{-almost surely},
\]

Birkhäuser
and so Proposition 10 gives
\[
\mathbb{P}_\ell (A^c_{\ell+1} | \{X_\ell = x\}) \\
= \mathbb{P}_\ell (\{\|X_{\ell+1}\|_{2\to2} > u_{\ell+1}\} | \{X_\ell = x\}) \\
= \mathbb{P}_{\ell,x} \left\{ \sum_{j=1}^n D_{b_j^{(\ell+1)} X_\ell D_{b_j}^{*}} \right\}_{2\to2} > u_{\ell+1} \right\} \\
\leq 2\exp\left( -\frac{u_{\ell+1}^2}{2\sigma^2_\ell(x) + \frac{2}{3} \|x\|_{2\to2} u_{\ell+1}} \right) \leq 2\exp\left( -\frac{u_{\ell+1}^2}{2v_\ell + \frac{2}{3} u_\ell u_{\ell+1}} \right),
\]
where the last step follows from the fact that \( x \) resides in the support of the \( \mathbb{P}_\ell \)-distribution of \( X_\ell \), and furthermore \( \sigma^2_\ell (X_\ell) \leq v_\ell \) holds over \( B_\ell \supseteq A_\ell \cap E \) and \( \|X_\ell\|_{2\to2} \leq u_\ell \) holds over \( A_\ell \supseteq A_\ell \cap E \). Thus,
\[
\mathbb{P}_\ell (A^c_{\ell+1} | X_\ell) \leq 2\exp\left( -\frac{u_{\ell+1}^2}{2v_\ell + \frac{2}{3} u_\ell u_{\ell+1}} \right) \leq 2\exp\left( -\frac{1}{4} \min \left\{ \frac{u_{\ell+1}^2}{v_\ell}, \frac{3u_{\ell+1}^2}{u_\ell} \right\} \right)
\]
holds \( \mathbb{P}_\ell \)-almost surely, and so we select \( u_{\ell+1} \) so that the right-hand side is at most \( \alpha / 2 \). One may show that it suffices to take \( u_1 := v_{L-1}^{1/2} \cdot 4 \log(4s/\alpha) \) and \( u_{\ell+1} = u_\ell \cdot (4/3) \log(4s/\alpha) \) for \( \ell \geq 1 \). This choice satisfies (17), from which the result follows.

\[\square\]

**Proof of Lemma 6** For indices \( i_1, \ldots, i_\ell \in [n] \) we define
\[
H(b_{i_1}, \ldots, b_{i_\ell}) := H_{i_1 \ldots i_\ell} := D_{b_{i_1}} \cdots D_{b_{i_\ell}} X_0 D_{b_{i_\ell}}^* \cdots D_{b_{i_1}}^*,
\]
Notice that \( H(\cdot) \) is a deterministic function, but \( H_{i_1 \ldots i_\ell} \) is random since \( b_{i_1}, \ldots, b_{i_\ell} \) are random. Observe that we may decompose \( A_\ell^* A_T - I \) as
\[
A_\ell^* A_T - I = \frac{1}{n^L} \sum_{i_1, \ldots, i_L \in [n]} H_{i_1 \ldots i_L}.
\]
For each \( \ell \in [L] \), it will be convenient to partition \([n]^\ell = \mathcal{I}_n^\ell \sqcup \mathcal{J}_n^\ell \), where \( \mathcal{I}_n^\ell \) denotes the elements with distinct entries and \( \mathcal{J}_n^\ell \) denotes the elements that do not have distinct entries. In addition, we let \( S([L], \ell) \) denote the set of partitions of \([L] \) into \( \ell \) nonempty sets, and we put \( S(L, \ell) := |S([L], \ell)| \). To analyze the above sum, we will relate \( A_\ell^* A_T - I \) to a sum indexed over \( \mathcal{I}_n^\ell \). Specifically, every tuple \((i_1, \ldots, i_L) \in [n]^\ell\) is uniquely determined by three features:

- The number \( \ell \in [L] \) of distinct entries in the tuple,
- A partition \( S \in S([L], \ell) \) of the tuple indices into \( \ell \) nonempty sets, and
- A tuple \((j_1, \ldots, j_\ell) \in \mathcal{I}_n^\ell \) of distinct elements of \([n]\).

To be clear, we order the members of \( S = \{S_1, \ldots, S_\ell\} \) lexicographically so that \( i_k = j_l \) for every \( k \in S_l \). For every \( S \in S([L], \ell) \), we may therefore abuse notation by

\[\copyright\] Birkhäuser
considering the function $S : [n] \overset{\ell}{\rightarrow} [n]^L$ defined by $S(j_1, \ldots, j_\ell) = (i_1, \ldots, i_L)$. It will also be helpful to denote $\sigma : [L] \rightarrow [\ell]$ such that for every $i \in [L]$, it holds that $i \in S_{\sigma(i)}$, that is, $i$ resides in the $\sigma(i)$-th member of the partition $S$. This gives

$$A_T^* A_T - I = \frac{1}{n^L} \sum_{i_1, \ldots, i_L \in [n]} H_{i_1 \cdots i_L} = \frac{1}{n^L} \sum_{\ell=1}^L \sum_{S \in \mathcal{S}([L], \ell)} \sum_{(j_1, \ldots, j_\ell) \in \mathcal{T}_n^\ell} H_{S(j_1, \ldots, j_\ell)}.$$  

The triangle inequality and union bound then give

$$E := \mathbb{P} \left\{ \left\| A_T^* A_T - I \right\|_2 > \frac{1}{e} \right\}$$

$$\leq \mathbb{P} \left\{ \left\| \sum_{\ell=1}^L \sum_{S \in \mathcal{S}([L], \ell)} \sum_{(j_1, \ldots, j_\ell) \in \mathcal{T}_n^\ell} H_{S(j_1, \ldots, j_\ell)} \right\|_2 > \frac{n^L}{e} \right\}$$

$$\leq \sum_{\ell=1}^L \sum_{S \in \mathcal{S}([L], \ell)} \mathbb{P} \left\{ \left\| \sum_{(j_1, \ldots, j_\ell) \in \mathcal{T}_n^\ell} H_{S(j_1, \ldots, j_\ell)} \right\|_2 > \frac{n^L}{eL^2} \right\},$$

where the last inequality uses the simple bound $S(L, \ell) \leq 2^L$. To proceed, we will apply Proposition 8 with $k = \ell$, random variables $X_i = b_i$ for $i \in [n]$, measurable space $\mathcal{S} = [N]$, separable Banach space $B = (\mathbb{C}^{T \times T}, \| \cdot \|_2)$, and measurable function $h_S : [N]^{\ell} \rightarrow \mathbb{C}^{T \times T}$ defined by

$$h_S(x_1, \ldots, x_\ell) := H(x_{\sigma(1)}, \ldots, x_{\sigma(\ell)}) = H(s_1 x_1, \ldots, s_\ell x_\ell),$$

where $s_k := |S_k|$ for $k \in [\ell]$. Observe that $H_{S(j_1, \ldots, j_\ell)} = h_S(b_{j_1}, \ldots, b_{j_\ell})$. As such, we may continue our bound:

$$E \leq \sum_{\ell=1}^L \sum_{S \in \mathcal{S}([L], \ell)} \mathbb{P} \left\{ \left\| \sum_{(j_1, \ldots, j_\ell) \in \mathcal{T}_n^\ell} H_{S(j_1, \ldots, j_\ell)} \right\|_2 > \frac{n^L}{eL^2} \right\}$$

$$= \sum_{\ell=1}^L \sum_{S \in \mathcal{S}([L], \ell)} \mathbb{P} \left\{ \left\| \sum_{(j_1, \ldots, j_\ell) \in \mathcal{T}_n^\ell} h_S(b_{j_1}, \ldots, b_{j_\ell}) \right\|_2 > \frac{n^L}{eL^2} \right\}$$

$$\leq \sum_{\ell=1}^L \sum_{S \in \mathcal{S}([L], \ell)} C_\ell \cdot \mathbb{P} \left\{ C_\ell \left\| \sum_{(j_1, \ldots, j_\ell) \in \mathcal{T}_n^\ell} h_S(b_{j_1}^{(1)}, \ldots, b_{j_\ell}^{(\ell)}) \right\|_2 > \frac{n^L}{eL^2} \right\}$$

$$= \sum_{\ell=1}^L \sum_{S \in \mathcal{S}([L], \ell)} C_\ell \cdot \mathbb{P} \left\{ C_\ell \left\| \sum_{(j_1, \ldots, j_\ell) \in \mathcal{T}_n^\ell} H(s_{j_1} b_{j_1}^{(1)}, \ldots, s_{j_\ell} b_{j_\ell}^{(\ell)}) \right\|_2 > \frac{n^L}{eL^2} \right\}.$$


Since $L < N$, then for each $S \in \mathcal{S}([L], \ell)$, we have $s_i \in (0, N)$ for every $i \in [\ell]$. Also, $N$ is prime, and so the mapping $M : \mathbb{F}^{\ell \times n}_N \rightarrow \mathbb{F}^{\ell \times n}_N$ defined by $M(X) = \text{diag}(s_1, \ldots, s_{\ell}) \cdot X$ is invertible. Thus, $(s_i b_j^{(i)})_{i \in [\ell], j \in [n]}$ has the same (uniform) distribution as $(b_j^{(i)})_{i \in [\ell], j \in [n]}$. Put

$$H'_{j_1 \cdots j_\ell} := D_{b_{j_\ell}} \cdots D_{b_{j_1}} X_0 D_{b_j}^* \cdots D_{b_1}^*.$$  

Denoting $C := \max_{\ell \leq L} C_\ell$ in terms of the absolute constants of Proposition 8, then

$$E \leq 2^{L^2} C \cdot \sum_{\ell=1}^L \mathbb{P}\left\{ \left\| \sum_{j_1, \ldots, j_\ell \in [n]} H'_{j_1 \cdots j_\ell} \right\|_{2 \rightarrow 2} > \frac{n^L}{CeL2L^2} \right\}.$$  

We will use Lemma 9 to bound each of the probabilities in the above sum. However, the sums in Lemma 9 are indexed by $[n]^\ell$ instead of $\mathcal{J}_n^\ell$. To close this gap, we perform another sequence of union bounds. First,

$$E \leq 2^{L^2} C \cdot \sum_{\ell=1}^L \mathbb{P}\left\{ \left\| \sum_{j_1, \ldots, j_\ell \in [n]} H'_{j_1 \cdots j_\ell} - \sum_{(j_1, \ldots, j_\ell) \in \mathcal{J}_n^\ell} H'_{j_1 \cdots j_\ell} \right\|_{2 \rightarrow 2} > \frac{n^L}{CeL2L^2} \right\}$$

$$\leq 2^{L^2} C \cdot \sum_{\ell=1}^L \mathbb{P}\left\{ \left\| \sum_{j_1, \ldots, j_\ell \in [n]} H'_{j_1 \cdots j_\ell} \right\|_{2 \rightarrow 2} > \frac{n^L}{2CeL2L^2} \right\} \quad (18)$$

$$+ 2^{L^2} C \cdot \sum_{\ell=1}^L \mathbb{P}\left\{ \left\| \sum_{(j_1, \ldots, j_\ell) \in \mathcal{J}_n^\ell} H'_{j_1 \cdots j_\ell} \right\|_{2 \rightarrow 2} > \frac{n^L}{2CeL2L^2} \right\}. \quad (19)$$

Next, we will decompose the sum over $\mathcal{J}_n^\ell$ into multiples of sums of over $[n]^1, \ldots, [n]^\ell-1$. Each $(j_1, \ldots, j_\ell) \in \mathcal{J}_n^\ell$ induces a partition of $[\ell]$ into at most $\ell - 1$ parts. We may rewrite this sum by first summing over each partition $S$ into $\ell - 1$ parts and then over each assignment of elements to this partition, represented by a member of $[n]^\ell-1$. That is, we may consider

$$\sum_{S \in \mathcal{S}([\ell], \ell-1)} \sum_{i_1, \ldots, i_{\ell-1} \in [n]} H'_{S(i_1, \ldots, i_{\ell-1})}.$$  

The above expression will include $H'_{j_1 \cdots j_\ell}$ exactly once for each $(j_1, \ldots, j_\ell)$ that induces a partition of $[\ell]$ into exactly $\ell - 1$ parts. However, since the indices $(i_1, \ldots, i_{\ell-1}) \in [n]^\ell-1$ need not be distinct, for each partition into more than $\ell - 1$ parts, the above sum will overcount the corresponding terms $H'_{j_1 \cdots j_\ell}$. To remedy this, we subtract the appropriate multiple of the sum over all partitions into $\ell - 2$ parts. Then, having subtracted too many terms of the form $H_{j_1 \cdots j_\ell}$ corresponding to parti-
tions into \( \ell - 3 \) parts, we may add back the appropriate number of such terms. Iterating this process leads to a decomposition of the form

\[
\sum_{(j_1, \ldots, j_t) \in \mathcal{J}_n^t} H'_{j_1 \ldots j_t} = (-1)^{\ell-1} \sum_{j=1}^{\ell-1} \sum_{1 \leq s_1, \ldots, s_j \leq \ell \atop s_1 + \cdots + s_j = \ell \atop |S_1| = s_1, \ldots, |S_j| = s_j} c_{s_1, \ldots, s_j} \sum_{S \in \mathcal{S}([\ell], j)} \sum_{i_1, \ldots, i_j \in [n]} \sum_{i_1, \ldots, i_j \in [n]} H'_{S(i_1, \ldots, i_j)}.
\]

For our purposes, it suffices to observe that \( c_{s_1, \ldots, s_j} \) is some positive integer crudely bounded by \( 2^{L^3} \); indeed, \( 2^{L^2} \) is a bound on the number of partitions of \([\ell]\), and there are at most \( L - 1 \) steps in the iterative process above, so

\[
c_{s_1, \ldots, s_j} \leq 1 + 2L^2 + 2^{2L^2} + \cdots + 2^{(L-1)L^2} \leq 2L^3.
\]

The triangle inequality then gives

\[
\left\| \sum_{(j_1, \ldots, j_t) \in \mathcal{J}_n^t} H'_{j_1 \ldots j_t} \right\|_{2 \to 2} \leq \sum_{j=1}^{\ell-1} \sum_{1 \leq s_1, \ldots, s_j \leq \ell \atop s_1 + \cdots + s_j = \ell \atop |S_1| = s_1, \ldots, |S_j| = s_j} c_{s_1, \ldots, s_j} \left\| \sum_{i_1, \ldots, i_j \in [n]} H'_{S(i_1, \ldots, i_j)} \right\|_{2 \to 2} \leq 2L^3 \cdot \sum_{j=1}^{\ell-1} \sum_{i_1, \ldots, i_j \in [n]} \left\| H'_{S(i_1, \ldots, i_j)} \right\|_{2 \to 2}.
\]

Next, we apply a union bound:

\[
(19) = 2L^2 C \cdot \sum_{\ell=1}^{L} \left\{ \left\| \sum_{(j_1, \ldots, j_t) \in \mathcal{J}_n^t} H'_{j_1 \ldots j_t} \right\|_{2 \to 2} > \frac{n^L}{2CeL2L^2} \right\} \leq 2L^2 C \cdot \sum_{\ell=1}^{L} \left\{ \sum_{j=1}^{\ell-1} \sum_{S \in \mathcal{S}([\ell], j)} \left\| \sum_{i_1, \ldots, i_j \in [n]} H'_{S(i_1, \ldots, i_j)} \right\|_{2 \to 2} > \frac{n^L}{2CeL2L^2} \right\} \leq 2L^2 C \cdot \sum_{\ell=1}^{L} \sum_{j=1}^{\ell-1} \sum_{S \in \mathcal{S}([\ell], j)} \left\{ \sum_{i_1, \ldots, i_j \in [n]} H'_{S(i_1, \ldots, i_j)} \right\}_{2 \to 2} > \frac{n^L}{2CeL2L^2}.
\]

Let \( F \in \mathbb{F}_N^{j \times \ell} \) denote the matrix whose \( i \)th row is the indicator vector of \( S_i \). Since the \( S_i \)'s are nonempty and partition \([\ell]\), it follows that \( F \) has rank \( j \). Define the mapping \( M : \mathbb{F}_N^{\ell \times n} \to \mathbb{F}_N^{j \times n} \) by \( M(X) = FX \). Then \( M \) is surjective, and so \( M((b_k^{(i)})_{i \in [\ell], k \in [n]}) \) has the same (uniform) distribution as \( (b_k^{(i)})_{i \in [\ell], k \in [n]} \). Next, we take \( g : \mathbb{F}_N^{\ell \times n} \to (\mathbb{C}^T \times T)^{n^j} \) such that \( g((x_k^{(i)})_{i \in [\ell], k \in [n]})_{i_1, \ldots, i_j} = H(x_1^{(1)}, \ldots, x_j^{(j)}) \).
Then \( (H'_{S(i_1,\ldots,i_j)})_{i_1,\ldots,i_j \in [n]} = g(M((b_k^{(i)})_{i \in [\ell], k \in [n]})) \) has the same distribution as \( (H'_{i_1,\ldots,i_j})_{i_1,\ldots,i_j \in [n]} = g((b_k^{(i)})_{i \in [j], k \in [n]}). \) With this, we continue:

\[
(19) \leq 2^{L^2} C \cdot \sum_{\ell=1}^{L} \sum_{j=1}^{\ell-1} \sum_{S \in \mathcal{S}([\ell],j)} \mathbb{P}\left\{ \left\| \sum_{i_1,\ldots,i_j \in [n]} H'_{S(i_1,\ldots,i_j)} \right\|_{2 \to 2} > \frac{nL}{2CeL^2 23 L^3} \right\}
\]

\[
= 2^{L^2} C \cdot \sum_{\ell=1}^{L} \sum_{j=1}^{\ell-1} \sum_{S \in \mathcal{S}([\ell],j)} \mathbb{P}\left\{ \left\| \sum_{i_1,\ldots,i_j \in [n]} H'_{i_1,\ldots,i_j} \right\|_{2 \to 2} > \frac{nL}{2CeL^2 23 L^3} \right\}. \tag{20}
\]

Finally, pick \( \alpha := \epsilon / (6CL^3 2L^2) \) and take \( C_L^{(2)} \) large enough so that

\[
n^{L/2} \geq (C_L^{(2)})^{L/2} s^{1/2} \log^{3L/2} (N/\epsilon) \geq 2CeL^2 23L^3 \cdot C_L^{(2)} s^{1/2} \log^{3L/2} (N/\alpha).
\]

Then by Lemma 9, each of the terms in (18) and (20) are at most \( L \alpha \), and so

\[
E \leq (18) + (20) \leq 2^{L^2} C \cdot L \cdot L \alpha + 2^{L^2} C \cdot L \cdot L \cdot 2L^2 \cdot L \alpha \leq 2CL^3 2L^2 \alpha = \epsilon / 3.
\]

\( \square \)

4 Proof of Lemma 7

We will use the following lemma, whose proof we save for later.

**Lemma 11** Fix \( u \in \mathbb{F}_N \) and \( j : [k + 1] \times [2M] \times [L] \rightarrow [n] \) with image\(^1\) of size \( m \). Then at least \( \lceil m/L \rceil \) of the following \( n \) linear constraints on \( \ell \in \mathbb{F}_N^{(k+1) \times 2M} \) are linearly independent:

\[
\sum_{(h, p, q) \in j^{-1}(i)} (-1)^p (\ell_{h, p} - \ell_{h-1, p}) = 0, \quad i \in [n],
\]

where \( \ell_{0, p} := u \) for every \( p \in [2M] \).

**Proof of Lemma 7** For now, we assume that \( p \) takes the form \( p = 2M \) for some \( M \in \mathbb{N} \), and later we interpolate with Littlewood’s inequality. As such, we will bound the expectation of

\[
|a_n^* A_T (I - A_T^* A_T)^k z|^{2M}.
\]

\( \tag{21} \)

\(^1\) To be clear, \( j \) is a function, and we demand that the range of this function has exactly \( m \) elements.
In what follows, we write (21) as a large sum of products, and then we will take the expectation of each product in the sum. First, we denote \( \ell_0 := u \) and observe

\[
a_u^* A_T (I - A_T^* A_T)^k z = - \sum_{\ell_1, \ldots, \ell_{k+1} \in T_h} \prod_{h=1}^{k+1} (I - A^* A)_{\ell_{h-1}, \ell_h z \ell_{k+1}}.
\]

Since \( A \) has unit-norm columns, the diagonal entries of \( I - A^* A \) are all zero, and so we can impose the constraint that \( \ell_h \neq \ell_{h+1} \) for every \( h \in [k] \). For the moment, we let \( T_{k+1}^* \) denote the set of all \((k+1)\)-tuples \( \ell \) of members of \( T \) that satisfy this constraint. Then

\[
a_u^* A_T (I - A_T^* A_T)^k z = - \sum_{\ell \in T_{k+1}^*} \prod_{h=1}^{k+1} (I - A^* A)_{\ell_{h-1}, \ell_h z \ell_{k+1}}.
\]

Next, we take the squared modulus of this quantity and raise it to the \( M \)-th power. To do so, it will be convenient to adopt the following notation to keep track of complex conjugation: Write \( \bar{x}^{(p)} \) to denote \( x \) if \( p \) is odd and \( \bar{x} \) if \( p \) is even. Then

\[
|a_u^* A_T (I - A_T^* A_T)^k z|^{2M} = - \sum_{\ell \in T_{k+1}^*} \prod_{h=1}^{k+1} (I - A^* A)_{\ell_{h-1}, \ell_h z \ell_{k+1}}^{2M}.
\]

where \( \ell_{0, p} = u \) and \( \ell_{h, p} \) denotes the \( h \)-th entry of \( \ell^{(p)} \) for every \( p \in [2M] \). It is helpful to think of each tuple \((\ell^{(1)}, \ldots, \ell^{(2M)}) \in (T_{k+1}^*)^{2M}\) as a matrix \( \ell \in T(\times 1 \times 2M) \) satisfying constraints of the form \( \ell_{h-1, p} \neq \ell_{h, p} \). Let \( \mathcal{L} \) denote the set of these matrices. Next, we write out the entries of \( I - A^* A \) in terms of the entries of \( A \):

\[
(I - A^* A)_{\ell_{h-1, p}, \ell_{h, p}} = -(A^* A)_{\ell_{h-1, p}, \ell_{h, p}}
\]

\[
= -\frac{1}{nL} \sum_{j_1, \ldots, j_L \in [n]} e_N (((\ell_{h, p} - \ell_{h-1, p}) (b_{j_1} + \cdots + b_{j_L})).
\]

After applying the distributive law, it will be useful to index with \( j_{h, p, 1}, \ldots, j_{h, p, L} \in [n] \) in place of \( j_1, \ldots, j_L \in [n] \) above, and in order to accomplish this, we let \( \mathcal{J} \) denote the set of functions \( j : [k+1] \times [2M] \times [L] \rightarrow [n] \). We also collect the entries of \( z \) by writing \( z_{\ell} := \sum_{p=1}^{2M} \prod_{h=1}^{k+1} \ell^{(p)}_{h, p} \). Then the distributive law gives

\[
|a_u^* A_T (I - A_T^* A_T)^k z|^{2M} = \sum_{\ell \in \mathcal{L}} \prod_{p=1}^{2M} \frac{1}{nL} \sum_{j_1, \ldots, j_L \in [n]} e_N((-1)^p ((\ell_{h, p} - \ell_{h-1, p}) (b_{j_1} + \cdots + b_{j_L})).
\]
\[
\frac{1}{n^{2ML(k+1)}} \sum_{\ell \in \mathcal{L}} z_{\ell} \prod_{j \in \mathcal{J}} \prod_{p=1}^{2M} \prod_{h=1}^{k+1} e_N \left( \sum_{q=1}^{L} (-1)^{p} (\ell_{h,p} - \ell_{h-1,p}) b_{j_{h,p,q}} \right) \\
= \frac{1}{n^{2ML(k+1)}} \sum_{\ell \in \mathcal{L}} z_{\ell} \sum_{j \in \mathcal{J}} e_N \left( \sum_{p=1}^{2M} \sum_{h=1}^{k+1} \sum_{q=1}^{L} (-1)^{p} (\ell_{h,p} - \ell_{h-1,p}) b_{j_{h,p,q}} \right) \\
= \frac{1}{n^{2ML(k+1)}} \sum_{\ell \in \mathcal{L}} z_{\ell} \sum_{j \in \mathcal{J}} e_N \left( \sum_{i \in [n]} (h,p,q) \in j^{-1}(i) (-1)^{p} (\ell_{h,p} - \ell_{h-1,p}) b_{i} \right) \\
= \frac{1}{n^{2ML(k+1)}} \sum_{\ell \in \mathcal{L}} z_{\ell} \sum_{j \in \mathcal{J}} \prod_{i \in [n]} e_N \left( \sum_{(h,p,q) \in j^{-1}(i)} (-1)^{p} (\ell_{h,p} - \ell_{h-1,p}) b_{i} \right). \tag{22}
\]

When bounding the expectation, we can remove the \(z_{\ell}\)’s. To be explicit, we apply the linearity of expectation and the independence of the \(b_{i}\)’s, the triangle inequality and the fact that \(\|z\|_{\infty} \leq 1\), and then the fact that each expectation \(\mathbb{E}e_N\) is either 0 or 1:

\[
\mathbb{E} \sum_{\ell} z_{\ell} \prod_{j} \prod_{i} e_N = \sum_{\ell} z_{\ell} \prod_{j} \prod_{i} \mathbb{E}e_N \leq \sum_{\ell} \left| \sum_{j} \prod_{i} \mathbb{E}e_N \right| = \sum_{\ell} \sum_{j} \prod_{i} \mathbb{E}e_N.
\]

This produces the bound

\[
\mathbb{E}\|a^*_n A_T (I - A_T^2 A_T)^k z\|^{2M} \leq \frac{1}{n^{2ML(k+1)}} \sum_{\ell \in \mathcal{L}} \sum_{j \in \mathcal{J}} \prod_{i \in [n]} \mathbb{E}e_N \left( \sum_{(h,p,q) \in j^{-1}(i)} (-1)^{p} (\ell_{h,p} - \ell_{h-1,p}) b_{i} \right) \\
= \frac{1}{n^{2ML(k+1)}} \left\{ (\ell, j) \in \mathcal{L} \times \mathcal{J} : (\ell, j) \in S \}
\]

\[
\mathbb{E} \left( \sum_{i \in [n]} (h,p,q) \in j^{-1}(i) (-1)^{p} (\ell_{h,p} - \ell_{h-1,p}) b_{i} \right) = 0 \mod N \ \forall i \in [n] \right\}. \tag{23}
\]

We estimate the size of this set \(S\) by first bounding the number \(N_1(j)\) of \(\ell \in \mathcal{L}\) such that \((\ell, j) \in S\), and then bounding the sum \(|S| = \sum_{j \in \mathcal{J}} N_1(j)\). Let \(m(j)\) denote the size of the image of \(j \in \mathcal{J}\) and put \(r := \lceil m(j)/L \rceil\). By Lemma 11, there exist linearly independent \(\{w_{i}\}_{i \in [r]}\) in \(\mathbb{F}_N^{(k+1) \times 2M}\) and \(\{b_{i}\}_{i \in [r]}\) in \(\mathbb{F}_N\) such that \((\ell, j) \in S\) only if \(\ell \in \mathbb{F}_N^{(k+1) \times 2M}\) satisfies \(\ell^T w_i = b_i\) for every \(i \in [r]\). Of course, \((\ell, j) \in S\) also requires \(\ell \in T^{(k+1) \times 2M}\). As such, we use identity basis elements to complete \(\{w_{i}\}_{i \in [r]}\) to a basis \(\{w_{i}\}_{i \in [2M(k+1)]}\) for \(\mathbb{F}_N^{(k+1) \times 2M}\). Then \((\ell, j) \in S\) only if \(\ell \in \mathbb{F}_N^{(k+1) \times 2M}\) satisfies

\[
w_{i}^T \ell = b_{i} \ \forall i \in \{1, \ldots, r\} \quad \text{and} \quad w_{i}^T \ell \in T \ \forall i \in \{r + 1, \ldots, 2M(k+1)\}.
\]

It follows that \(N_1(j) \leq |T|^{2M(k+1) - r}\). Next, observe that \(\eta\) is the size of the domain of \(j \in \mathcal{J}\). If \(m(j) > \eta/2\), then there exists \(i \in [n]\) for which the preimage \(j^{-1}(i)\) is
a singleton set \{(h, p, q)\}, in which case there is no \((\ell, j) \in S\) since \(\ell_{h,p} \neq \ell_{h-1,p}\). Overall, 

\[
N_1(j) \leq \begin{cases} 
2^{M(k+1) - [m(j)/L]} & \text{if } m(j) \leq \eta/2 \\
0 & \text{else.}
\end{cases}
\]

Next, let \(N_2(m)\) denote the number of \(j \in J\) with image of size \(m\). Then \(N_2(m) \leq n^m m^n\), since there are \(\binom{n}{m}\) choices for the image, and for each image, there are at most \(m^n\) choices for \(j\) (we say “at most” here since the image of \(j\) needs to have size \(m\)). Then

\[
|S| = \sum_{j \in J} N_1(j) \\
\leq \sum_{m=1}^{\eta/2} N_2(m) 2^{M(k+1) - [m/L]} \\
\leq \sum_{m=1}^{\eta/2} n^m m^n (s^{1/L})^{n-m} \\
\leq (\eta/2)^n (s^{1/L})^n \sum_{m=1}^{\eta/2} (ns^{-1/L})^m \leq (\eta/2)^n (s^{1/L})^n \cdot 2(ns^{-1/L})^{\eta/2},
\]

(24)

where the last step uses the fact that \(x := ns^{-1/L} \geq 2\), or equivalently \(x - 1 \geq x/2\), which implies \(\sum_{i=1}^{k} x^i = \frac{x^{k+1} - x}{x-1} \leq 2(x^k - 1) \leq 2x^k\). We now combine (23) and (24):

\[
\mathbb{E}[a_u^* A_T (I - A_T^* A_T)^k z] \leq n^{-\eta} \cdot (\eta/2)^n (s^{1/L})^n \cdot 2(ns^{-1/L})^{\eta/2} = 2(\eta/2)^n (s^{1/L} n^{-1})^{\eta/2}.
\]

(25)

Finally, we interpolate using Littlewood’s inequality. Put \(X := a_u^* A_T (I - A_T^* A_T)^k z\), given any \(p \geq 2\), let \(M\) denote the largest integer for which \(2M \leq p\), and put \(\theta := p/2 - M\). Consider the function defined by \(\eta(x) := (k+1)Lx\), and put \(\eta := \eta(p), \eta_1 := \eta(2M), \) and \(\eta_2 := \eta(2M + 2)\). Then (25) implies

\[
\mathbb{E}[|X|^p] \leq (\mathbb{E}[|X|^{2M}]^{1-\theta} (\mathbb{E}[|X|^{2M+2}]^\theta \\
\leq 2((\eta_1/2)^{\eta_1})^{1-\theta} ((\eta_2/2)^{\eta_2})^\theta (s^{1/L} n^{-1})^{\eta/2} \leq 2\eta^{2\eta} (s^{1/L} n^{-1})^{\eta/2},
\]

where the last step applies the fact that \(\eta_1, \eta_2 \leq 2\eta\). \(\Box\)

**Proof of Lemma 11** First, we isolate the constant terms in the left-hand side:

\[
\sum_{(h, p, q) \in j^{-1}(i)} (-1)^p (\ell_{h,p} - \ell_{h-1,p}) = \sum_{(1, p, q) \in j^{-1}(i)} (-1)^p (\ell_{1,p} - \ell_{0,p}) + \sum_{(h, p, q) \in j^{-1}(i) : h > 1} (-1)^p (\ell_{h,p} - \ell_{h-1,p}).
\]
Since $\ell_{0,p} = u$ for every $p$, we have

$$\sum_{(1,p,q)\in j^{-1}(i)} (-1)^p \ell_{1,p} + \sum_{(h,p,q)\in j^{-1}(i)} (-1)^p (\ell_{h,p} - \ell_{h-1,p}) = \left( \sum_{(1,p,q)\in j^{-1}(i)} (-1)^{p+1} \right) u.$$ (26)

Let $e_{h,p} \in \mathbb{F}_N^{(k+1)\times 2M}$ denote the matrix that is 1 at entry $(h, p)$ and 0 otherwise, and consider the basis $\{b_{h,p}\}_{h\in[k+1], p\in[2M]}$ of $\mathbb{F}_N^{(k+1)\times [2M]}$ defined by

$$b_{h,p} := \begin{cases} (-1)^p e_{1,p} & \text{if } h = 1 \\ (-1)^p (e_{h,p} - e_{h-1,p}) & \text{else.} \end{cases}$$

Then the left-hand side of (26) may be rewritten as $\left( \sum_{(h,p,q)\in j^{-1}(i)} b_{h,p} \right)^\top \ell$. It remains to find a subset $S \subseteq [n]$ of size $\lceil m/L \rceil$ for which $\{\sum_{(h,p,q)\in j^{-1}(i)} b_{h,p} \}_i \in S$ is linearly independent. Initialize $S = \emptyset$, $R := \text{im}(j)$, and $t = 1$, and then do the following until $R$ is empty:

- Select any $(h_t, p_t)$ for which there exists $q_t$ such that $i_t := j(h_t, p_t, q_t) \in R$, and
- Update $S \leftarrow S \cup \{i_t\}$, $R \leftarrow R \setminus \{j(h_t, p_t, q) : q \in [L]\}$ and $t \leftarrow t + 1$.

Since each iteration removes at most $L$ members from $R$, the resulting $S$ has size at least $\lceil \text{im}(j)/L \rceil = \lceil m/L \rceil$. By construction, every $t$ has the property that there is no $q \in [L]$ or $u < t$ for which $(h_t, p_t, q) \in j^{-1}(i_t)$, and it follows that $\sum_{(h,p,q)\in j^{-1}(i_t)} b_{h,p}$ is the first member of the sequence to exhibit a contribution from $b_{h_t, p_t}$. Thanks to this triangularization, we may conclude that $\{\sum_{(h,p,q)\in j^{-1}(i)} b_{h,p} \}_i \in S$ is linearly independent. \qed

Acknowledgements  The authors thank Holger Rauhut for pointing out [27] as a potential avenue for derandomized compressed sensing, and the anonymous reviewer for helpful suggestions that improved the presentation of our results. DGM was partially supported by AFOSR FA9550-18-1-0107 and NSF DMS 1829955.

References

1. Amelunxen, D., Lotz, M., McCoy, M.B., Tropp, J.A.: Living on the edge: phase transitions in convex programs with random data. Inf. Inference 3, 224–294 (2014)

2. Applebaum, L., Howard, S.D., Searle, S., Calderbank, R.: Chirp sensing codes: deterministic compressed sensing measurements for fast recovery. Appl. Comput. Harmon. Anal. 26, 283–290 (2009)

3. Bandeira, A.S., Dobriban, E., Mixon, D.G., Sawin, W.F.: Certifying the restricted isometry property is hard. IEEE Trans. Inf. Theory 59, 3448–3450 (2013)

4. Bandeira, A.S., Fickus, M., Mixon, D.G., Moreira, J.: Derandomizing restricted isometries via the Legendre symbol. Constr. Approx. 43, 409–424 (2016)

5. Bandeira, A.S., Fickus, M., Mixon, D.G., Moreira, J.: The road to deterministic matrices with the restricted isometry property. J. Fourier Anal. Appl. 19, 1123–1149 (2013)

6. Bandeira, A.S., Mixon, D.G., Moreira, J.: A conditional construction of restricted isometries. Int. Math. Res. Not. 2017, 372–381 (2017)

7. Bourgain, J.: An improved estimate in the restricted isometry problem. In: Geometric Aspects of Functional Analysis, pp. 65–70. Springer, Berlin (2014)

8. Baraniuk, R., Davenport, M., DeVore, R., Wakin, M.: A simple proof of the restricted isometry property for random matrices. Constr. Approx. 28, 253–263 (2008)
9. Bourgain, J., Dilworth, S., Ford, K., Konyagin, S., Kutzarova, D.: Explicit constructions of RIP matrices and related problems. Duke Math. J. 159, 145–185 (2011)
10. Bourgain, J., Dilworth, S., Ford, K., Konyagin, S., Kutzarova, D.: Breaking the $k^2$ barrier for explicit RIP matrices, pp. 637–644. STOC (2011)
11. Cai, T.T., Zhang, A.: Sharp RIP bound for sparse signal and low-rank matrix recovery. Appl. Comput. Harmon. Anal. 35, 74–93 (2013)
12. Candès, E.J.: The restricted isometry property and its implications for compressed sensing. C. R. Acad. Sci. Paris Ser. I 346, 589–592 (2008)
13. Candès, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Inf. Theory 52, 489–509 (2006)
14. Candès, E.J., Tao, T.: Near-optimal signal recovery from random projections: universal encoding strategies? IEEE Inf. Theory 52, 5406–5425 (2006)
15. Cheraghchi, M., Guruswami, V., Velingker, A.: Restricted isometry of Fourier matrices and list decodability of random linear codes. SIAM J. Comput. 42, 1888–1914 (2013)
16. de la Peña, V., Giné, E.: Decoupling: From Dependence to Independence. Springer, Berlin (2012)
17. DeVore, R.A.: Deterministic constructions of compressed sensing matrices. J. Complex. 23, 918–925 (2007)
18. Donoho, D., Elad, M.: Optimally sparse representation in general (non-linear) dictionaries. IEEE Trans. Inf. Theory 52, 603–621 (2006)
19. Donoho, D., Tanner, J.: Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. Philos. Trans. R. Soc. A 367, 4273–4293 (2009)
20. Duarte, M.F., Davenport, M.A., Takhar, D., Laska, J.N., Sun, T., Kelly, K.F., Baraniuk, R.G.: Single-pixel imaging via compressive sampling. IEEE Signal Process. Mag. 25, 83–91 (2008)
21. Fickus, M., Mixon, D.G.: Tables of the existence of equiangular tight frames. arXiv:1504.00253
22. Fickus, M., Mixon, D.G., Treiman, J.C.: Steiner equiangular tight frames. Linear Algebra Appl. 436, 1014–1027 (2012)
23. Foucart, S., Rauhut, H.: A Mathematical Introduction to Compressive Sensing. Springer, Berlin (2013)
24. Goldreich, O.: Foundations of Cryptography I: Basic Tools. Cambridge University Press, Cambridge (2001)
25. Haikin, M., Zamir, R., Gavish, M.: Random subsets of structured deterministic frames have MANOVA spectra. Proc. Natl. Acad. Sci. USA 114, E5024–E5033 (2017)
26. Haviv, I., Regev, O.: The restricted isometry property of subsampled Fourier matrices. In: Geometric Aspects of Functional Analysis, pp. 163–179. Springer, Berlin (2017)
27. Hütig, M., Rauhut, H., Strohmer, T.: Remote sensing via $\ell_1$-minimization. Found. Comput. Math. 14, 115–150 (2014)
28. Hoeffding, W.: Probability inequalities for sums of bounded random variables. J. Am. Stat. Assoc. 58, 13–30 (1963)
29. Iwen, M.A.: Compressed sensing with sparse binary matrices: instance optimal error guarantees in near-optimal time. J. Complex. 30, 1–15 (2014)
30. Jasper, J., Mixon, D.G., Fickus, M.: Kirkman equiangular tight frames and codes. IEEE Trans. Inf. Theory 60, 170–181 (2013)
31. Johnson, W.B., Lindenstrauss, J.: Extensions of Lipschitz mappings into a Hilbert space. Contemp. Math. 26, 189–206 (1984)
32. Krahmer, F., Mendelson, S., Rauhut, H.: Suprema of chaos processes and the restricted isometry property. Commun. Pure Appl. Math. 67, 1877–1904 (2014)
33. Maggioni, M., Mixon, D.G., Parshall, H.: Kesten–McKay law for random subensembles of Paley equiangular tight frames. arXiv:1905.04360
34. Mendelson, S., Pajor, A., Tomczak-Jaegermann, N.: Uniform uncertainty principle for Bernoulli and subgaussian ensembles. Constr. Approx. 28, 277–289 (2009)
35. Mixon, D.G.: Explicit matrices with the restricted isometry property: breaking the square-root bottleneck. In: Compressed sensing and its applications, pp. 389–417. Birkhäuser (2015)
36. Monajemi, H., Jafarpour, S., Gavish, M., Donoho, D.L.: Stat 330/CME 362 collaboration, deterministic matrices matching the compressed sensing phase transitions of Gaussian random matrices. Proc. Natl. Acad. Sci. USA 110, 1181–1186 (2013)
37. Renes, J.M.: Equiangular tight frames from Paley tournaments. Linear Algebra Appl. 426, 497–501 (2007)
38. Rudelson, M., Vershynin, R.: On sparse reconstruction from Fourier and Gaussian measurements. Commun. Pure Appl. Math. 61, 1025–1045 (2008)
39. Tao, T.: Open question: deterministic UUP matrices. https://terrytao.wordpress.com/2007/07/02/open-question-deterministic-uup-matrices/
40. Tao, T.: Topics in random matrix theory. Am. Math. Soc. (2012)
41. Taubman, D., Marcellin, M.: JPEG2000 Image Compression Fundamentals, Standards and Practice. Springer, Berlin (2012)
42. Tillmann, A.M., Pfetsch, M.E.: The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. IEEE Trans. Inf. Theory 60, 1248–1259 (2013)
43. Tropp, J.A.: An introduction to matrix concentration inequalities, Found. Trends. Mach. Learn. 8, 1–230 (2015)
44. Tropp, J.A.: On the conditioning of random subdictionaries. Appl. Comput. Harmon. Anal. 25, 1–24 (2008)
45. Tropp, J.A.: User-friendly tail bounds for sums of random matrices. Found. Comput. Math. 12, 389–434 (2012)
46. Vershynin, R.: High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge University Press, Cambridge (2018)
47. Wang, T., Berthet, Q., Plan, Y.: Average-case hardness of RIP certification. NeurIPS 29, 3819–3827 (2016)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.