Noname manuscript No.
(will be inserted by the editor)

The restricted isometry property for time–frequency structured random matrices

Götz E. Pfander · Holger Rauhut · Joel A. Tropp

Dedicated to Hans G. Feichtinger on occassion of his 60th birthday

the date of receipt and acceptance should be inserted later

Abstract This paper establishes the restricted isometry property for a Gabor system generated by n^2 time–frequency shifts of a random window function in n dimensions. The sth order restricted isometry constant of the associated $n \times n^2$ Gabor synthesis matrix is small provided that $s \leq cn^{2/3}/\log^2 n$. This bound provides a qualitative improvement over previous estimates, which achieve only quadratic scaling of the sparsity s with respect to n. The proof depends on an estimate for the expected supremum of a second-order chaos.

Keywords compressed sensing · restricted isometry property · Gabor system · time–frequency analysis · random matrix · chaos process

Mathematics Subject Classification (2000) 60B20 · 42C40 · 94A12

1 Introduction and statement of results

Sparsity has become a key concept in applied mathematics and engineering because of the empirical observation that, in many real-world settings, the signal of interest can be approximated accurately by means of a sparse expansion in an appropriately chosen system of basic signals. The theory of compressed sensing [7, 8, 10, 16, 18, 33] predicts that, to capture all the information in a sparse signal, it suffices to take a relatively small number of linear samples.

Götz E. Pfander
School of Engineering and Science, Jacobs University Bremen, 28759 Bremen, Germany, E-mail: g.pfander@jacobs-university.de.

Holger Rauhut
Hausdorff Center for Mathematics and Insitute for Numerical Simulation, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany E-mail: rauhut@hcm.uni-bonn.de.

Joel A. Tropp, California Institute of Technology, Pasadena, CA 91125, USA E-mail: jtropp@cms.caltech.edu.

Furthermore, we can identify the sparse signal from these samples using efficient algorithms. This discovery has a number of potential applications in signal processing, as well as other areas of science and technology.

In compressed sensing, a measurement matrix is used to model the process of linear data acquisition. The *restricted isometry property* (RIP) [9,10, 18,33] is a standard tool for studying how efficiently the measurement matrix captures information about sparse signals. The RIP also supports the analysis of various signal reconstruction algorithms, including ℓ_1 minimization, greedy pursuits, and other types of iterative algorithms. Many types of random matrices, including Gaussian and Rademacher matrices, obey the RIP with optimal scaling behavior [3,17,36,10]. In contrast, there are currently no deterministic constructions that satisfy an optimal RIP; see the discussion in [33, Sec. 2.5] or [18, Sec. 5.1].

In principle, Gaussian matrices are optimal for sparse recovery [16,19], but they have limited appeal in practice because most applications impose structure on the measurement matrix. Furthermore, most recovery algorithms are more efficient when the measurement matrix admits a fast matrix–vector multiply. For instance, we can model the signal acquisition process in MRI imaging by drawing a random set of rows from the discrete Fourier transform matrix. These matrices permit us to design fast recovery algorithms based on the FFT. With high probability, a random partial Fourier matrix satisfies the RIP with near-optimal scaling [10,38,31,33]. See [33,37] for some generalizations.

In this paper, we study a random Gabor system, which is a structured $n \times n^2$ matrix whose columns are obtained by taking all possible time–frequency shifts of a fixed random vector. The random Gabor system has many potential applications, including channel identification [30], underwater communications [27, 39], high-resolution radar [23], as well as the matrix probing problem [12].

The literature contains some work on the random Gabor system. The paper [30] uses coherence estimates to control the restricted isometry constants, and it results in suboptimal bounds. The paper [34] obtains nonuniform recovery bounds for ℓ_1 minimization. This analysis does not yield stable recovery results, it does not provide uniform recovery for all sparse signals, and it does not extend to other algorithms. The research in this paper makes some progress toward addressing these concerns.

Our approach is related to a recent restricted isometry analysis of the partial random circulant matrix [35]. Indeed, our argument requires us to bound the expected supremum of a second-order chaos, which we accomplish using a Dudley-type inequality due to Talagrand [42]. This approach involves an estimate the covering numbers of the set of unit-norm s-sparse vectors with respect to two metrics induced by the random process. In contrast to the situation in [35], we cannot exploit the covering number estimates from [38], and so we have been forced to perform a new analysis.

This paper is organized as follows. Section 1.1 introduces the random Gabor system, and it contains our main result on the restricted isometry constants. Section 1.2 includes some remarks that illustrate how time–frequency structured measurement matrices arise in applications such as in wireless communications and radar. We survey previous work in Section 1.3. Sections 2, 3 and 4 present the proof of the main result.

1.1 Time–frequency structured measurement matrices

This paper provides probabilistic estimates for the restricted isometry constants of a matrix whose columns consist of time–frequency shifts of a random vector. Let $\boldsymbol{T}$ denote the cyclic shift on $\mathbb{C}^n$, also known as the translation operator, and let $\boldsymbol{M}$ denote the frequency shift on $\mathbb{C}^n$, also known as the modulation operator. These operators are defined by the rules

$$(\boldsymbol{T}\boldsymbol{h})_q := h_{q\ominus 1} \quad \text{and} \quad (\boldsymbol{M}\boldsymbol{h})_q := e^{2\pi i q/n} h_q = \omega^q h_q, \tag{1}$$

where $\ominus$ is subtraction modulo n and $\omega := e^{2\pi i/n}$. Note that

$$(\boldsymbol{T}^k\boldsymbol{h})_q = h_{q\ominus k} \quad \text{and} \quad (\boldsymbol{M}^\ell\boldsymbol{h})_q = e^{2\pi i \ell q/n} h_q = \omega^{\ell q} h_q. \tag{2}$$

We introduce the time–frequency shift operators $\boldsymbol{\pi}(\lambda) = \boldsymbol{M}^\ell\boldsymbol{T}^k$, indexed by pairs $\lambda = (k,\ell)$, where k and ℓ range over $\mathbb{Z}_n := \{1,\dots,n\}$. The system $\{\boldsymbol{\pi}(\lambda) : \lambda \in \mathbb{Z}_n\times\mathbb{Z}_n\}$ of all time–frequency shifts forms a basis for the matrix space $\mathbb{C}^{n\times n}$ [25,24].

Next, we construct the random Gabor measurement matrix. Let $\boldsymbol{\epsilon} \in \mathbb{C}^n$ be a random vector that follows one of the following two distributions:

- Each entry of $\boldsymbol{\epsilon}$ is an independent Rademacher random variable, i.e., a variable that takes values ± 1 with equal probability.
- Each entry of $\boldsymbol{\epsilon}$ is an independent Steinhaus random variable, i.e., a variable that is uniformly distributed on the complex torus $S^1 = \{z \in \mathbb{C} : |z| = 1\}$.

Define a normalized window function

$$\boldsymbol{g} = \frac{1}{\sqrt{n}}\boldsymbol{\epsilon}.$$

The family

$$\{\boldsymbol{\pi}(\lambda)\boldsymbol{g} : \lambda \in \mathbb{Z}_n\times\mathbb{Z}_n\} \tag{3}$$

is called a full Gabor system with window $\boldsymbol{g}$ [21]. We can introduce a matrix $\boldsymbol{\Psi}_{\boldsymbol{g}} \in \mathbb{C}^{n\times n^2}$ whose columns range over the full Gabor system. The matrix $\boldsymbol{\Psi}_{\boldsymbol{g}}$ is referred to as the Gabor synthesis matrix [13,25,29]. Note that $\boldsymbol{\Psi}_{\boldsymbol{g}}$ admits a fast matrix–vector multiply by means of the FFT algorithm.

We say that a vector $\boldsymbol{x}$ is s-sparse when $\|\boldsymbol{x}\|_0 := \#\{\ell : x_\ell \neq 0\} \leq s$. Recall that, for an $n\times N$ matrix $\boldsymbol{A}$ and a positive integer $s \leq n$, the restricted isometry constant δ_s is defined as the smallest positive number that satisfies

$$(1-\delta_s)\|\boldsymbol{x}\|_2^2 \;\leq\; \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \;\leq\; (1+\delta_s)\|\boldsymbol{x}\|_2^2 \quad \text{for all } \boldsymbol{x} \text{ with } \|\boldsymbol{x}\|_0 \leq s. \tag{4}$$

When the matrix $\boldsymbol{A}$ has a sufficiently small restricted isometry constant δ_{2s}, then we can recover every s-sparse vector $\boldsymbol{x}$ from the measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{y}$

using a variety of algorithms, including ℓ_1 minimization and certain greedy pursuits. See [18] and the reference therein for more details.

The main result of this paper concerns the restricted isometry constants of the random Gabor system $\boldsymbol{\Psi_g}$. In the sequel, we write $\mathbb{E}$ for expectation and $\mathbb{P}$ for the probability of an event.

Theorem 1 *Let $\boldsymbol{\Psi_g} \in \mathbb{C}^{n\times n^2}$ be a draw of the random Gabor synthesis matrix with normalized Rademacher or Steinhaus generating vector.*

(a) For $s \le n$, the expectation of the restricted isometry constant δ_s of $\boldsymbol{\Psi_g}$ satisfies

$$\mathbb{E}\,\delta_s \le \max\Big\{C_1\sqrt{\frac{s^{3/2}}{n}}\log s\sqrt{\log n},\; C_2\frac{s^{3/2}\log^{3/2} n}{n}\Big\}, \tag{5}$$

where $C_1, C_2 > 0$ are universal constants.

(b) For $0 \le \lambda \le 1$, we have the probability bound

$$\mathbb{P}(\delta_s \ge \mathbb{E}[\delta_s] + \lambda) \le e^{-\lambda^2/\sigma^2}, \qquad \textit{where } \sigma^2 = \frac{C_3 s^{\frac{3}{2}}\log n\,\log^2 s}{n} \tag{6}$$

with $C_3 > 0$ being a universal constant.

In particular, the simplified condition

$$n \ge C s^{3/2}\log^3(n)\log(\varepsilon^{-1})$$

implies that the matrix $\boldsymbol{\Psi_g}$ satisfies the RIP of order s with probability exceeding $1-\varepsilon$. With slight variations of the proof one can show similar statements for normalized Gaussian or subgaussian random windows $\boldsymbol{g}$.

The paper [30] contains numerical tests that illustrate the performance of the random Gabor system $\boldsymbol{\Psi_g}$ for compressed sensing. This empirical work indicates that the behavior of the random Gabor system does not depend on the choice of random window, and the performance in all cases is similar to that of a fully Gaussian measurement matrix. Nevertheless, we must emphasize that numerical tests cannot verify the RIP.

We do not believe that Theorem 1 is optimal. We suspect that any significant improvement would demand more sophisticated techniques than the ones that we apply in this paper. Indeed, the literature contains examples [26, 42] where the central tool in this paper, the Dudley-type inequality for chaos processes (Theorem 3), is not sharp. We may well be facing one of these cases here.

1.2 Application in wireless communications and radar

In wireless communications, an important task is to identify the properties of the communication channel by probing it with a small number of known pilot signals. A common finite-dimensional model [4,14,20,28] for the channel operator is given by the formula

$$\boldsymbol{\Gamma} = \sum_{\lambda \in \mathbb{Z}_n \times \mathbb{Z}_n} x_\lambda \boldsymbol{\pi}(\lambda).$$

This model includes digital-to-analog-conversion at the transmitter, the action of the analog communications channel, and the analog-to-digital conversion at the receiver. Time shifts model delay due to multipath-propagation, while frequency shifts model Doppler effects due to motion of the transmitter, receiver, or scatterers. Physical considerations suggest that the vector $\boldsymbol{x}$ is rather sparse because the number of scatterers is typically quite small. Similar models appear in sonar [27,39] and radar [23] problems.

Our goal is to identify the coefficient vector $\boldsymbol{x}$ from a single input–output pair $(\boldsymbol{g}, \boldsymbol{\Gamma g})$. In other words, we need to reconstruct $\boldsymbol{\Gamma} \in \mathbb{C}^{n \times n}$ from its action $\boldsymbol{y} = \boldsymbol{\Gamma g}$ on a single vector $\boldsymbol{g}$. Write

$$\boldsymbol{y} = \boldsymbol{\Gamma g} = \sum_{\lambda \in \mathbb{Z}_n \times \mathbb{Z}_n} x_\lambda \boldsymbol{\pi}(\lambda)\boldsymbol{g} = \boldsymbol{\Psi_g x}, \tag{7}$$

where $\boldsymbol{x}$ is sparse but unknown. This is a compressed sensing problem. In this setting, the choice of pilot signal $\boldsymbol{g}$ remains at our discretion, so we may select $\boldsymbol{g}$ to be a random Rademacher or Steinhaus sequence. Theorem 1 demonstrates that $\boldsymbol{\Psi_g}$ has the RIP with high probability, so we can recover the coefficient vector $\boldsymbol{x}$, provided that it is sufficiently sparse.

1.3 Relation with previous work

Matrices with time–frequency structure have played a role in the sparsity literature for many years. Recall that the coherence of a matrix $\boldsymbol{A} = (\boldsymbol{a}_1| \dots |\boldsymbol{a}_N)$ with normalized columns $\|\boldsymbol{a}_\ell\|_2 = 1$ is defined as

$$\mu := \max_{\ell \neq k} |\langle \boldsymbol{a}_\ell, \boldsymbol{a}_k \rangle|.$$

Strohmer and Heath [40] considered a Gabor system $\boldsymbol{\Psi_g}$ based on the Alltop window $\boldsymbol{g} \in \mathbb{C}^n$, whose entries $g_\ell = n^{-1/2} e^{2\pi i \ell^3/n}$ where $n \geq 5$ is a prime [1]. The coherence of $\boldsymbol{\Psi_g}$ satisfies

$$\mu = \frac{1}{\sqrt{n}}.$$

For any $n \times N$ matrix, the coherence satisfies the bound $\mu \geq \sqrt{\frac{N-n}{n(N-1)}}$, so the Gabor–Alltop matrix has near-optimal coherence [40]. The coherence can be used to obtain a simple bound on the restricted isometry constant: $\delta_s \leq (s-1)\mu$. Therefore, for the Gabor–Alltop matrix, the restricted isometry constants satisfy

$$\delta_s \leq \frac{s-1}{\sqrt{n}}.$$

This bound requires that the sparsity $s \leq c\sqrt{n}$ for a nontrivial RIP to hold. Qualitatively, this estimate is somewhat worse than Theorem 1.

The paper [30] contains an estimate for the coherence μ of a random Gabor system based on a Steinhaus window:

$$\mu \leq c\sqrt{\frac{\log(n/\varepsilon)}{n}},$$

with probability at least $1-\varepsilon$. As before, this bound only guarantees that the RIP constant δ_s is small when the s scales like $\sqrt{n}$.

The paper [34] develops a nonuniform recovery result for compressed sensing with a random Gabor system based on a Steinhaus window.

Theorem 2 *Let $\boldsymbol{x} \in \mathbb{C}^n$ be an s-sparse vector, and assume that*

$$s \leq \frac{cn}{\log(n/\varepsilon)}.$$

Draw a random Steinhaus sequence $\boldsymbol{g}$, and form the random Gabor system $\boldsymbol{\Psi_g}$. Then, with probability at least $1-\varepsilon$, the vector $\boldsymbol{x}$ can be recovered from the measurements $\boldsymbol{y} = \boldsymbol{\Psi_g x}$ using ℓ_1 minimization.

In this estimate, the sparsity s scales almost linearly with the dimension n, which is optimal. Clearly, this bound is better than the RIP estimate in our main result, Theorem 1. In many respects, the conclusion of Theorem 2 is weaker than what we obtain from a RIP bound. Indeed, Theorem 2 only guarantees that we can recover a single sparse vector with high probability on a random draw of the matrix $\boldsymbol{\Psi_g}$. In contrast, a RIP bound allows us to recover all sparse vectors with high probability on a single random draw of the matrix. Furthermore, Theorem 2 cannot guarantee that ℓ_1 minimization is stable for vectors that are not quite sparse or contain noise. The RIP allows us to assert that both these properties hold [18,9].

Finally, we mention a closely related measurement system based on the partial random circulant matrix [22,32,33,35]. This matrix models convolution by a random filter, followed by subsampling at an arbitrary (deterministic) set of outputs. At present, the best estimate for the restricted isometry constants of an $n \times N$ partial random circulant matrix require $n \geq c(s \log N)^{3/2}$ for a nontrivial bound [35]. This scaling is similar to what we achieve in this paper, in part because both results depend on the Dudley-type inequality (Theorem 3). We also mention that partial random circulant matrices satisfy nonuniform recovery guarantees similar to Theorem 2 [32,33]. For this measurement ensemble, the analysis is easier because we can use harmonic analysis to convert the time-domain problem to an easier problem in the Fourier domain. For Gabor synthesis matrices, this option is not available to us, so the arguments become more involved.

2 Expectation of the restricted isometry constants

We first estimate the expectation of the restricted isometry constants of the random Gabor synthesis matrix, that is, we shall prove Theorem 1(a). To this end, we first rewrite the restricted isometry constants δ_s. Let $T = T_s = \{\boldsymbol{x} \in \mathbb{C}^{n^2}, \|\boldsymbol{x}\|_2 = 1, \|\boldsymbol{x}\|_0 \leq s\}$. Introduce the following semi-norm on Hermitian matrices A,

$$|\!|\!|\boldsymbol{A}|\!|\!|_s = \sup_{\boldsymbol{x}\in T_s} |\boldsymbol{x}^*\boldsymbol{A}\boldsymbol{x}|.$$

Then the restricted isometry constants of $\boldsymbol{\Psi} = \boldsymbol{\Psi_g}$ can be written as

$$\delta_s = |\!|\!|\boldsymbol{\Psi}^*\boldsymbol{\Psi} - \boldsymbol{I}|\!|\!|_s,$$

where $\boldsymbol{I}$ denotes the identity matrix. Observe that the Gabor synthesis matrix $\boldsymbol{\Psi_g}$ takes the form

$$\boldsymbol{\Psi_g} = \left(\begin{array}{cccc|ccc|cc} g_0 & g_{n-1} & \cdots & g_1 & g_0 & \cdots & g_1 & \cdots & g_1 \\ g_1 & g_0 & \cdots & g_2 & \omega g_1 & \cdots & \omega g_2 & \cdots & \omega^{n-1} g_2 \\ g_2 & g_1 & \cdots & g_3 & \omega^2 g_2 & \cdots & \omega^2 g_3 & \cdots & \omega^{2(n-1)} g_3 \\ g_3 & g_2 & \cdots & g_4 & \omega^3 g_3 & \cdots & \omega^3 g_4 & \cdots & \omega^{3(n-1)} g_4 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ g_{n-1} & g_{n-2} & \cdots & g_0 & \omega^{n-1} g_{n-1} & \cdots & \omega^{n-1} g_0 & \cdots & \omega^{(n-1)^2} g_0 \end{array}\right).$$

Our analysis in this section employs the representation

$$\boldsymbol{\Psi_g} = \sum_{q=0}^{n-1} g_q \, \boldsymbol{A}_q$$

with

$$\boldsymbol{A}_0 = \left(\begin{array}{ccccc|ccccc|cc} 1 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & \omega & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \omega^2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & \omega^{n-1} & \cdots & \omega^{(n-1)^2} \end{array}\right)$$
$$= \big(\boldsymbol{I}\big|\boldsymbol{M}\big|\boldsymbol{M}^2\big|\cdots\big|\boldsymbol{M}^{n-1}\big),$$

$$\boldsymbol{A}_1 = \left(\begin{array}{ccccc|ccccc|cc} 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 1 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 & \omega & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & \omega^2 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \end{array}\right)$$
$$= \big(\boldsymbol{T}\big|\boldsymbol{M}\boldsymbol{T}\big|\boldsymbol{M}^2\boldsymbol{T}\big|\cdots\big|\boldsymbol{M}^{n-1}\boldsymbol{T}\big),$$

and so on. In short, for $q \in \mathbb{Z}_n$,

$$\boldsymbol{A}_q = \big(\boldsymbol{T}^q\big|\boldsymbol{M}\boldsymbol{T}^q\big|\boldsymbol{M}^2\boldsymbol{T}^q\big|\cdots\big|\boldsymbol{M}^{n-1}\boldsymbol{T}^q\big). \tag{8}$$

Observe that

$$\boldsymbol{H} := \boldsymbol{\Psi}^*\boldsymbol{\Psi} - \boldsymbol{I} = -\boldsymbol{I} + \frac{1}{n}\sum_{q,q'=0}^{n-1} \overline{\epsilon_{q'}}\epsilon_q \, \boldsymbol{A}_{q'}^*\boldsymbol{A}_q\,.$$

Using (26) below, it follows that

$$\boldsymbol{H} = \frac{1}{n} \sum_{q' \neq q} \overline{\epsilon_{q'}}\, \epsilon_q\, \boldsymbol{A}_{q'}^* \boldsymbol{A}_q = \frac{1}{n} \sum_{q',q} \overline{\epsilon_{q'}}\, \epsilon_q\, \boldsymbol{W}_{q',q}, \tag{9}$$

where, for notational simplicity, we use here and in the following

$$\boldsymbol{W}_{q',q} = \boldsymbol{A}_{q'}^* \boldsymbol{A}_q \text{ for } q \neq q' \text{ and } \boldsymbol{W}_{q',q} = 0 \text{ for } q = q'. \tag{10}$$

We shall use the matrix $\boldsymbol{B}(\boldsymbol{x}) \in \mathbb{C}^{n \times n}$, $\boldsymbol{x} \in T_s$, given by matrix entries

$$B(\boldsymbol{x})_{q',q} = \boldsymbol{x}^* \boldsymbol{W}_{q',q} \boldsymbol{x}. \tag{11}$$

Then we have

$$n\, \mathbb{E}\delta_s = \mathbb{E} \sup_{\boldsymbol{x} \in T_s} |Y_{\boldsymbol{x}}| = \mathbb{E} \sup_{\boldsymbol{x} \in T_s} |Y_{\boldsymbol{x}} - Y_{\boldsymbol{0}}|\,, \tag{12}$$

where

$$Y_{\boldsymbol{x}} \;=\; \boldsymbol{\epsilon}^* \boldsymbol{B}(\boldsymbol{x}) \boldsymbol{\epsilon} = \sum_{q' \neq q} \overline{\epsilon_{q'}}\, \epsilon_q\, \boldsymbol{x}^* \boldsymbol{A}_{q'}^* \boldsymbol{A}_q \boldsymbol{x} \tag{13}$$

and $\boldsymbol{x} \in T_s = \{\boldsymbol{x} \in \mathbb{C}^{n \times n}, \|\boldsymbol{x}\|_2 \leq 1, \|\boldsymbol{x}\|_0 \leq s\}$. A process of the type (13) is called Rademacher or Steinhaus chaos process of order 2. In order to bound such a process, we use the following Theorem, see for example, [26, Theorem 11.22] or [42, Theorem 2.5.2], where it is stated for Gaussian processes and in terms of majorizing measure (generic chaining) conditions. The formulation below requires the operator norm $\|\boldsymbol{A}\|_{2\to 2} = \max_{\|\boldsymbol{x}\|_2 = 1} \|\boldsymbol{A}\boldsymbol{x}\|_2$ and the Frobenius norm $\|\boldsymbol{A}\|_F = \mathrm{Tr}(\boldsymbol{A}^*\boldsymbol{A})^{1/2} = (\sum_{j,k} |A_{j,k}|^2)^{1/2}$, where $\mathrm{Tr}(\boldsymbol{A})$ denotes the trace of a matrix $\boldsymbol{A}$.

Theorem 3 *Let $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$ be a Rademacher or Steinhaus sequence, and let*

$$Y_{\boldsymbol{x}} := \boldsymbol{\epsilon}^* \boldsymbol{B}(\boldsymbol{x}) \boldsymbol{\epsilon} = \sum_{q',q=1}^{n} \overline{\epsilon_{q'}} \epsilon_q B(\boldsymbol{x})_{q',q}$$

be an associated chaos process of order 2, indexed by $x \in T$, where we additionally assume $\boldsymbol{B}(\boldsymbol{x})$ hermitian with zero diagonal, that is, $B(\boldsymbol{x})_{q,q} = 0$ and $B(\boldsymbol{x})_{q',q} = \overline{B(\boldsymbol{x})_{q,q'}}$. We define two (pseudo-)metrics on T,

$$d_1(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{B}(\boldsymbol{x}) - \boldsymbol{B}(\boldsymbol{y})\|_{2\to 2},$$
$$d_2(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{B}(\boldsymbol{x}) - \boldsymbol{B}(\boldsymbol{y})\|_F.$$

Let $N(T, d_i, u)$ be the minimum number of balls of radius u in the metric d_i needed to cover T. Then there exists a universal constant $K > 0$ such that, for an arbitrary $\boldsymbol{x_0} \in T$,

$$\mathbb{E} \sup_{\boldsymbol{x} \in T} |Y_{\boldsymbol{x}} - Y_{\boldsymbol{x_0}}| \;\leq\; K \max\Big\{ \int_0^\infty \log N(T, d_1, u)\, du\,,\; \int_0^\infty \sqrt{\log N(T, d_2, u)}\, du \Big\}. \tag{14}$$

Proof For a Rademacher sequence, the theorem is stated in [35, Proposition 2.2]. If $\boldsymbol{\epsilon}$ is a Steinhaus sequence and $\boldsymbol{B}$ a Hermitian matrix then

$$\boldsymbol{\epsilon}^*\boldsymbol{B}\boldsymbol{\epsilon} = \mathrm{Re}(\boldsymbol{\epsilon}^*\boldsymbol{B}\boldsymbol{\epsilon}) = \mathrm{Re}(\boldsymbol{\epsilon})^*\,\mathrm{Re}(\boldsymbol{B})\,\mathrm{Re}(\boldsymbol{\epsilon}) - \mathrm{Re}(\boldsymbol{\epsilon})^*\,\mathrm{Im}(\boldsymbol{B})\,\mathrm{Im}(\boldsymbol{\epsilon}) \\ + \mathrm{Im}(\boldsymbol{\epsilon})^*\,\mathrm{Im}(\boldsymbol{B})\,\mathrm{Re}(\boldsymbol{\epsilon}) + \mathrm{Im}(\boldsymbol{\epsilon})^*\,\mathrm{Re}(\boldsymbol{B})\,\mathrm{Im}(\boldsymbol{\epsilon}).$$

By decoupling, see, for example, [15, Theorem 3.1.1], we have with ϵ' denoting an independent copy of ϵ,

$$\mathbb{E}\sup_{x\in T}|\,\mathrm{Re}(\boldsymbol{\epsilon})^*\,\mathrm{Im}(\boldsymbol{B}(\boldsymbol{x}))\,\mathrm{Im}(\boldsymbol{\epsilon})| \le 8\,\mathbb{E}\sup_{x\in T}|\,\mathrm{Re}(\boldsymbol{\epsilon})^*\,\mathrm{Im}(\boldsymbol{B}(\boldsymbol{x}))\,\mathrm{Im}(\boldsymbol{\epsilon}')| \\ \le 8\,\mathbb{E}\sup_{x\in T}|\boldsymbol{\xi}^*\,\mathrm{Im}(\boldsymbol{B}(\boldsymbol{x}))\,\mathrm{Im}(\boldsymbol{\epsilon}')| \le 8\,\mathbb{E}\sup_{x\in T}|\boldsymbol{\xi}^*\,\mathrm{Im}(\boldsymbol{B}(\boldsymbol{x}))\boldsymbol{\xi}'|,$$

where $\boldsymbol{\xi},\boldsymbol{\xi}'$ denote independent Rademacher sequences. The second and third inequalities follow from the contraction principle [26, Theorem 4.4] (and symmetry of $\mathrm{Re}(\epsilon_\ell), \mathrm{Im}(\epsilon_\ell)$) first applied conditionally on $\boldsymbol{\epsilon}'$ and then conditionally on $\boldsymbol{\xi}$ (note that $|\,\mathrm{Re}(\epsilon_\ell)| \le 1$, $|\,\mathrm{Im}(\epsilon_\ell)| \le 1$ for all realizations of ϵ_ℓ). Using the triangle inequality we get

$$\mathbb{E}\sup_{x\in T}|Y_{\boldsymbol{x}} - Y_{\boldsymbol{x_0}}| \le 16\,\mathbb{E}\sup_{x\in T}|\boldsymbol{\xi}^*(\mathrm{Re}(\boldsymbol{B}(\boldsymbol{x})) - \mathrm{Re}(\boldsymbol{B}(x_0))\xi'| \\ + 16\,\mathbb{E}\sup_{x\in T}|\boldsymbol{\xi}^*(\mathrm{Im}(\boldsymbol{B}(\boldsymbol{x})) - \mathrm{Im}(\boldsymbol{B}(x_0)))\boldsymbol{\xi}'|. \quad (15)$$

Further note that $\|\,\mathrm{Im}(\boldsymbol{B}(\boldsymbol{x})) - \mathrm{Im}(\boldsymbol{B}(\boldsymbol{y}))\|_F$, $\|\,\mathrm{Re}(\boldsymbol{B}(\boldsymbol{x})) - \mathrm{Re}(\boldsymbol{B}(\boldsymbol{y}))\|_F \le \|\boldsymbol{B}(\boldsymbol{x})-\boldsymbol{B}(\boldsymbol{y})\|_F$ and similarly, writing $\boldsymbol{B}(\boldsymbol{x})-\boldsymbol{B}(\boldsymbol{y})$ as a $2n\times 2n$ real block matrix acting on $\mathbb{R}^{2n}$ we see that also $\|\,\mathrm{Im}(\boldsymbol{B}(\boldsymbol{x}))-\mathrm{Im}(\boldsymbol{B}(\boldsymbol{y}))\|_{2\to 2}$, $\|\,\mathrm{Re}(\boldsymbol{B}(\boldsymbol{x}))-\mathrm{Re}(\boldsymbol{B}(\boldsymbol{y}))\|_{2\to 2} \le \|\boldsymbol{B}(\boldsymbol{x}) - \boldsymbol{B}(\boldsymbol{y})\|_{2\to 2}$. Furthermore, the statement for Rademacher chaos processes holds as well for decoupled chaos processes of the form above. (Indeed, its proof uses decoupling in a crucial way.) Therefore, the claim for Steinhaus sequences follows. □

Note that $\boldsymbol{B}(\boldsymbol{x})$ defined in (11) satisfies the hypotheses of Theorem 3 by definition. The pseudo-metrics are given by

$$d_2(\boldsymbol{x},\boldsymbol{y}) \;=\; \|\boldsymbol{B}(\boldsymbol{x}) - \boldsymbol{B}(\boldsymbol{y})\|_F = \Big(\sum_{q'\neq q}\big|\boldsymbol{x}^*\boldsymbol{A}_{q'}^*\boldsymbol{A}_q\boldsymbol{x} - \boldsymbol{y}^*\boldsymbol{A}_{q'}^*\boldsymbol{A}_q\boldsymbol{y}\big|^2\Big)^{1/2}, \quad (16)$$

and

$$d_1(\boldsymbol{x},\boldsymbol{y}) \;=\|\boldsymbol{B}(\boldsymbol{x}) - \boldsymbol{B}(\boldsymbol{y})\|_{2\to 2}.$$

The bound on the expected restricted isometry constant follows then from the following estimates on the covering numbers of T_s with respect to d_1 and d_2. Corresponding proofs will be detailed in Section 3. We start with $N(T_s, d_2, u)$.

Lemma 1 *For $u > 0$, it holds*

$$\log(N(T_s, d_2, u)) \le s\log(en^2/s) + s\log(1 + 4\sqrt{sn}u^{-1}).$$

The above estimate is useful only for small $u > 0$. For large u we require the following alternative bound.

Lemma 2 *The diameter of T_s with respect to d_2 is bounded by $4\sqrt{sn}$, and for $\sqrt{n} \le u \le 4\sqrt{sn}$, it holds*

$$\log(N(T_s, d_2, u)) \le cu^{-2} n s^{3/2} \log(n s^{5/2} u^{-1}),$$

where $c > 0$ is universal constant.

Covering number estimates with respect to d_1 are provided in the following lemma.

Lemma 3 *The diameter of T_s with respect to d_1 is bounded by $4s$, and for $u > 0$*

$$\log(N(T_s, d_1, u)) \le \min\left\{ s\log(en^2/s) + s\log(1 + 4su^{-1}), \right. \\ \left. cu^{-2} s^2 \log(2n) \log(n^2/u) \right\}, \quad (17)$$

where $c > 0$ is a universal constant.

Moreover, we require the following elementary estimate of an integral, see [33, Lemma 10.3].

Lemma 4 *For $\alpha > 0$, we have*

$$\int_0^\alpha \sqrt{\log(1 + u^{-1})} du \le \alpha \sqrt{\log(e(1 + \alpha^{-1}))}.$$

Based on these estimates and Theorem 3 we complete the proof of Theorem 1(a). By Lemmas 1 and 2, the subgaussian integral in (14) can be estimated as

$$\begin{aligned}
&\int_0^\infty \sqrt{\log(N(T_s, d_2, u))} du = \int_0^{4\sqrt{sn}} \sqrt{\log(N(T_s, d_2, u))} du \\
&= \int_0^{\sqrt{n}} \sqrt{\log(N(T_s, d_2, u))} du + \int_{\sqrt{n}}^{4\sqrt{sn}} \sqrt{\log(N(T_s, d_2, u))} du \\
&\le \int_0^{\sqrt{n}} \sqrt{s\log(en^2/s)} du + \int_0^{\sqrt{n}} \sqrt{s\log(1 + 4\sqrt{sn} u^{-1})} du \\
&+ c\sqrt{ns^{3/2}} \int_{\sqrt{n}}^{4\sqrt{sn}} u^{-1} \sqrt{\log(ns^{5/2}u^{-1})} du \\
&\le \sqrt{sn \log(en^2/s)} + 4s\sqrt{n} \int_0^{s^{-1/2}} \sqrt{\log(1 + u^{-1})} du \\
&+ c' \sqrt{s^{3/2} n} \sqrt{\log(n^{1/2} s^{5/2})} \log(\sqrt{s}) \\
&\le \sqrt{sn \log(en^2/s)} + 4\sqrt{sn} \sqrt{\log(e(1 + \sqrt{s}))} + c'' \sqrt{s^{3/2} n \log(n) \log^2(s)} \\
&\le \hat{C}_1 \sqrt{s^{3/2} n \log(n) \log^2(s)}.
\end{aligned} \quad (18)$$

In the second inequality, we have used Lemma 4. Due to Lemma 3 the subexponential integral obeys the estimate, for some $\kappa > 0$ to be chosen below,

$$\int_0^\infty \log(N(T_s, d_1, u))du = \int_0^{4s} \log(N(T_s, d_1, u))du$$
$$= \int_0^\kappa \log(N(T_s, d_1, u))du + \int_\kappa^{4s} \log(N(T_s, d_1, u))du$$
$$\leq \kappa s \log(en^2/s) + s \int_0^\kappa \log(1 + 4su^{-1})du + cs^2 \log(2n) \int_\kappa^{4s} u^{-2} \log(n^2/u)du$$
$$\leq \kappa s \log(en^2/s) + \kappa s \log(e(1 + 4s/\kappa)) + cs^2\kappa^{-1} \log(2n) \log(n^2/\kappa).$$

Choose $\kappa = \sqrt{s \log(n)}$ to reach

$$\int_0^\infty \log(N(T_s, d_1, u))du \leq \hat{C}_2 s^{3/2} \log^{3/2}(n). \tag{19}$$

Combining the above integral estimates with (12) and Theorem 3 yields

$$\mathbb{E}\delta_s = \frac{1}{n}\mathbb{E} \sup_{x\in T_s} |Y_{\boldsymbol{x}} - Y_0| \leq \frac{1}{n} \max\left\{ C_1\sqrt{s^{3/2} n \log(n) \log^2(s)}, C_2 s^{3/2} \log^{3/2}(n) \right\}. \tag{20}$$

This is the statement of Theorem 1(a).

Remark 1 In analogy to the estimate of a subgaussian entropy integral arising in the analysis of partial random circulant matrices in [35], we expect that the exponent 3/2 in (18) can be improved to 1. However, we doubt that for the subexponential integral (19) such improvement will be possible (indeed, the estimate of the subexponential integral in [35] also exhibits an exponent of 3/2 at the s-term), so that we did not pursue an improvement of (18) here as this would not provide a significant overall improvement of (20). We expect that an improvement of (20) would require more sophisticated tools than the Dudley type estimate for chaos processes of Theorem 3.

3 Proof of covering number estimates

In this section we provide the covering number estimates of Lemma 1, 2 and 3, which are crucial to the proof of our main result. We first introduce additional notation. Let $\delta(m, k) = \delta_{0,m-k}$ and $\delta(m) = \delta_{0,m}$ be the Kronecker symbol as usual. We denote by $\operatorname{supp} \boldsymbol{x} = \{\ell, x_\ell \neq 0\}$ the support of a vector $\boldsymbol{x}$. Let $\boldsymbol{A}$ be a matrix with vector of singular values $\boldsymbol{\sigma}(\boldsymbol{A})$. For $0 < q \leq \infty$, the Schatten S_q-norm is defined by

$$\|\boldsymbol{A}\|_{S_q} := \|\boldsymbol{\sigma}(\boldsymbol{A})\|_q, \tag{21}$$

where $\|\cdot\|_q$ is the usual vector ℓ_q norm. For an integer p, the S_{2p} norm can be expressed as

$$\|\boldsymbol{A}\|_{S_{2p}} = (\mathrm{Tr}((\boldsymbol{A}^*\boldsymbol{A})^p))^{1/(2p)}. \tag{22}$$

The S_∞-norm coincides with the operator norm, $\|\cdot\|_{S_\infty} = \|\cdot\|_{2\to 2}$. By the corresponding properties of ℓ_q-norms we have the inequalities

$$\|\boldsymbol{A}\|_{2\to 2} \le \|\boldsymbol{A}\|_{S_q} \le \operatorname{rank}(\boldsymbol{A})^{1/q}\|\boldsymbol{A}\|_{2\to 2}. \tag{23}$$

Moreover, we will require an extension of the quadratic form $\boldsymbol{B}(\boldsymbol{x})$ in (11) to a bilinear form,

$$(\boldsymbol{B}(\boldsymbol{x},\boldsymbol{z}))_{q',q} = \begin{cases} \boldsymbol{x}^*\boldsymbol{A}_{q'}^*\boldsymbol{A}_q\boldsymbol{z} & \text{if } q' \neq q, \\ 0 & \text{if } q' = q. \end{cases} \tag{24}$$

Then $\boldsymbol{B}(\boldsymbol{x}) = \boldsymbol{B}(\boldsymbol{x},\boldsymbol{x})$.

3.1 Time–frequency analysis on $\mathbb{C}^n$

Before passing to the actual covering number estimates we provide some facts and estimates related to time–frequency analysis on $\mathbb{C}^n$. Observe that the matrices $\boldsymbol{A}_q$ introduced in (8) satisfy

$$\boldsymbol{A}_q^* = \begin{pmatrix} (\boldsymbol{T}^q)^* \\ (\boldsymbol{M}\boldsymbol{T}^q)^* \\ (\boldsymbol{M}^2\boldsymbol{T}^q)^* \\ \vdots \\ (\boldsymbol{M}^{n-1}\boldsymbol{T}^q)^* \end{pmatrix} = \begin{pmatrix} \boldsymbol{T}^{-q} \\ \boldsymbol{T}^{-q}\boldsymbol{M}^{-1} \\ \boldsymbol{T}^{-q}\boldsymbol{M}^{-2} \\ \vdots \\ \boldsymbol{T}^{-q}\boldsymbol{M}^{1} \end{pmatrix},$$

and, hence,

$$(\boldsymbol{A}_q^*\boldsymbol{y})_{(k,\ell)} = y_{k+q}\ \omega^{-\ell(k+q)}.$$

Clearly,

$$\begin{aligned} \langle \boldsymbol{A}_q\boldsymbol{z},\boldsymbol{y}\rangle = \langle \boldsymbol{z},\boldsymbol{A}_q^*\boldsymbol{y}\rangle &= \textstyle\sum_{k,\ell} z_{(k,\ell)}\overline{y}_{k+q}\omega^{\ell(k+q)} = \sum_{k,\ell} z_{(k-q,\ell)}\overline{y}_k\omega^{\ell k} \\ &= \textstyle\sum_k \big(\sum_\ell z_{(k-q,\ell)}\omega^{\ell k}\big)\overline{y}_k \end{aligned}$$

and, hence,

$$(\boldsymbol{A}_q\boldsymbol{z})_k = \sum_\ell z_{(k-q,\ell)}\omega^{\ell k}.$$

In the following, $\boldsymbol{\mathcal{F}} : \mathbb{C}^n \mapsto \mathbb{C}^n$ denotes the normalized Fourier transform, that is,

$$(\boldsymbol{\mathcal{F}}\boldsymbol{v})_\ell = n^{-1/2}\sum_{q=0}^{n-1}\omega^{-q\ell}v_q.$$

For $\boldsymbol{v} \in \mathbb{C}^{n\times n}$, $\boldsymbol{\mathcal{F}}_2\boldsymbol{v}$ denotes the Fourier transform in the second variable of v.

Let $\{\boldsymbol{e}_\lambda\}_{\lambda\in\mathbb{Z}_n\times\mathbb{Z}_n}$ and $\{\boldsymbol{e}_q\}_{q\in\mathbb{Z}_n}$ denoting the Euclidean basis of $\mathbb{C}^{n\times n}$ respectively $\mathbb{C}^n$, and, let $\boldsymbol{P}_\lambda$ denote the orthogonal projection onto the one dimensional space $\operatorname{span}\{\boldsymbol{e}_\lambda\}$. The following relationships will be crucial for the covering number estimates below.

Lemma 5 *Let $\boldsymbol{A}_q$ be as given in* (8)*. Then, for $\lambda \in \mathbb{Z}_n \times \mathbb{Z}_n$, $q \in \mathbb{Z}_n$,*

$$\boldsymbol{A}_q \boldsymbol{e}_\lambda = \boldsymbol{\pi}(\lambda) \boldsymbol{e}_q \,, \tag{25}$$

$$\sum_{q=0}^{n-1} \boldsymbol{A}_q^* \boldsymbol{A}_q = n \, \boldsymbol{I} \,, \tag{26}$$

$$\sum_{q=0}^{n-1} \boldsymbol{A}_q \boldsymbol{P}_\lambda \boldsymbol{A}_q^* = \boldsymbol{I} \ , \tag{27}$$

$$\sum_{q=0}^{n-1} \sum_{q'=0}^{n-1} \left| \boldsymbol{x}^* \boldsymbol{A}_{q'}^* \boldsymbol{A}_q \boldsymbol{y} \right|^2 \leq n \, \|\boldsymbol{x}\|_0 \, \|\boldsymbol{x}\|_2^2 \, \|\boldsymbol{y}\|_2^2. \tag{28}$$

Proof For (25), observe that

$$\begin{aligned}
(\boldsymbol{A}_q \boldsymbol{e}_{(k_0, \ell_0)})_k &= \sum_\ell \delta(k - q - k_0, \ell - \ell_0) \omega^{\ell k} = \delta(q - (k - k_0)) \omega^{\ell_0 k} \\
&= (\boldsymbol{\pi}(k_0, \ell_0) \boldsymbol{e}_q)_k \,.
\end{aligned}$$

To see (26), choose $\boldsymbol{z} \in \mathbb{C}^{n \times n}$ and compute

$$\begin{aligned}
\left(\boldsymbol{A}_{q'}^* \boldsymbol{A}_q \boldsymbol{z} \right)_{(k', \ell')} &= \sum_\ell z_{(k'+q'-q, \ell)} \omega^{\ell(k'+q')} \omega^{-\ell'(k'+q')} \\
&= \sum_\ell z_{(k'+q'-q, \ell)} \omega^{(\ell - \ell')(k'+q')} \,.
\end{aligned}$$

Hence,

$$\begin{aligned}
\sum_q \left(\boldsymbol{A}_q^* \boldsymbol{A}_q \boldsymbol{z} \right)_{(k', \ell')} &= \sum_q \sum_\ell z_{(k', \ell)} \omega^{(\ell - \ell')(k'+q)} = \sum_\ell z_{(k', \ell)} \sum_q \omega^{(\ell - \ell')(k'+q)} \\
&= \sum_\ell z_{(k', \ell)} n \, \delta(\ell - \ell') = n \, z_{(k', \ell')} \,.
\end{aligned}$$

Finally, observe that all but one column of $\boldsymbol{A}_q \boldsymbol{P}_{\{(\ell_0, k_0)\}}$ are 0, the nonzero column being column (ℓ_0, k_0), and only its $(k_0 + q)$th entry is nonzero, namely, it is $\omega^{\ell_0 (k_0 + q)}$. We have

$$\boldsymbol{A}_q \boldsymbol{P}_{\{(\ell_0, k_0)\}} \boldsymbol{A}_q^* = \boldsymbol{A}_q \boldsymbol{P}_{\{(\ell_0, k_0)\}} \boldsymbol{P}_{\{(\ell_0, k_0)\}} \boldsymbol{A}_q^* = \boldsymbol{A}_q \boldsymbol{P}_{\{(\ell_0, k_0)\}} (\boldsymbol{A}_q \boldsymbol{P}_{\{(\ell_0, k_0)\}})^*,$$

and hence, $\boldsymbol{A}_q \boldsymbol{P}_{\{(\ell_0, k_0)\}} \boldsymbol{A}_q^* = \boldsymbol{P}_{\{k_0 + q\}}$ and $\sum_q \boldsymbol{A}_q \boldsymbol{P}_{\{(\ell_0, k_0)\}} \boldsymbol{A}_q^* = \boldsymbol{I}$.

Let $\boldsymbol{x} \in \mathbb{C}^{n\times n}$ and $\Lambda = \operatorname{supp} \boldsymbol{x}$, then

$$\begin{aligned}
\sum_q \sum_{q'} \left| \boldsymbol{x}^* \boldsymbol{A}_{q'}^* \boldsymbol{A}_q \boldsymbol{y} \right|^2 &= \sum_q \sum_{q'} \Big| \sum_{(k',\ell')\in\Lambda} x_{(k',\ell')} \overline{\left(\boldsymbol{A}_{q'}^* \boldsymbol{A}_q \boldsymbol{y}\right)_{k',\ell'}} \Big|^2 \\
&\le \|\boldsymbol{x}\|_2^2 \sum_q \sum_{q'} \sum_{(k',\ell')\in\Lambda} \left| \left(\boldsymbol{A}_{q'}^* \boldsymbol{A}_q \boldsymbol{y}\right)_{k',\ell'} \right|^2 \\
&= \|\boldsymbol{x}\|_2^2 \sum_q \sum_{q'} \sum_{(k',\ell')\in\Lambda} \Big| \omega^{-\ell'(k'+q')} \sum_\ell \omega^{\ell(k'+q')} y_{(k'-(q-q'),\ell)} \Big|^2 \\
&= \|\boldsymbol{x}\|_2^2 \sum_q \sum_{q'} \sum_{(k',\ell')\in\Lambda} \Big| \sum_\ell \omega^{\ell(k'+q')} y_{(k'-(q-q'),\ell)} \Big|^2 \\
&= n \, \|\boldsymbol{x}\|_2^2 \sum_{(k',\ell')\in\Lambda} \sum_q \sum_{q'} \left| \left(\boldsymbol{\mathcal{F}}_2 \boldsymbol{y}\right)_{(k'-(q-q'),k'+q')} \right|^2 \\
&= n \, \|\boldsymbol{x}\|_2^2 \sum_{(k',\ell')\in\Lambda} \left\| \boldsymbol{\mathcal{F}}_2 \boldsymbol{y} \right\|_2^2 = n \, |\Lambda| \, \|\boldsymbol{x}\|_2^2 \, \|\boldsymbol{y}\|_2^2 = n \, \|\boldsymbol{x}\|_0 \|\boldsymbol{x}\|_2^2 \|\boldsymbol{y}\|_2^2
\end{aligned}$$

by unitarity of $\boldsymbol{\mathcal{F}}_2$. □

3.2 Proof of Lemma 1

For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^{n^2}$,

$$d_2(\boldsymbol{x}, \boldsymbol{y}) \;\le\; \Big(\sum_{q'\neq q} \Big| \boldsymbol{x}^* \boldsymbol{A}_{q'}^* \boldsymbol{A}_q (\boldsymbol{x} - \boldsymbol{y}) \Big|^2 \Big)^{1/2} + \Big(\sum_{q'\neq q} \Big| (\boldsymbol{x} - \boldsymbol{y})^* \boldsymbol{A}_{q'}^* \boldsymbol{A}_q \boldsymbol{y} \Big|^2 \Big)^{1/2}.$$

Inequality (28) implies that for $\boldsymbol{x}, \boldsymbol{y} \in T_s$,

$$\Big(\sum_{q'\neq q} \Big| \boldsymbol{x}^* \boldsymbol{A}_{q'}^* \boldsymbol{A}_q (\boldsymbol{x} - \boldsymbol{y}) \Big|^2 \Big)^{1/2}, \; \Big(\sum_{q'\neq q} \Big| (\boldsymbol{x} - \boldsymbol{y})^* \boldsymbol{A}_{q'}^* \boldsymbol{A}_q \boldsymbol{y} \Big|^2 \Big)^{1/2} \le \sqrt{sn} \, \|\boldsymbol{x} - \boldsymbol{y}\|_2$$

and, hence,

$$d_2(\boldsymbol{x}, \boldsymbol{y}) \;\le\; 2\sqrt{sn} \, \|\boldsymbol{x} - \boldsymbol{y}\|_2. \tag{29}$$

Using the volumetric argument, see, for example, [33, Proposition 10.1], we obtain

$$N(T_s, \|\cdot\|_2, u) \le \binom{n^2}{s} (1 + 2/u)^s \le (en^2/s)^s (1 + 2/u)^s.$$

By a rescaling argument

$$\begin{aligned}
N(T_s, d_2, u) &\le N(T_s, 2\sqrt{sn}\|\cdot\|_2, u) = N(T_s, \|\cdot\|_2, u/(2\sqrt{sn})) \\
&\le (en^2/s)^s (1 + 4\sqrt{sn} u^{-1})^s.
\end{aligned}$$

Taking the logarithm completes the proof. □

3.3 Proof of Lemma 2

Now, we seek a suitable estimate of the covering numbers $N(T_s, d_1, u)$ for $u \geq \sqrt{n}$. We use Maurey's empirical method [11], similarly as done in [38]: For a fixed vector $x \in T_s$, we introduce a discrete random vector $\boldsymbol{Z}$ with expectation x. We form the empirical mean over m copies of $\boldsymbol{Z}$ to estimate the deviation to x in the metric d_1. This allows us to find a value of m which leads to a prescribed distance u to x; since $\boldsymbol{Z}$ takes only a finite number of values we are finally able to derive a bound on the covering numbers. Hereby we use the fact that, by construction, the values attained by $\boldsymbol{Z}$ are independent of the choice of x.

Since $d_1(x, y) \leq d_2(x, y)$, inequality (29) implies that the diameter of T_s with respect to d_1 is at most $4\sqrt{sn}$. Hence, it suffices to consider $N(T_s, d_1, u)$ for

$$\sqrt{n} \leq u \leq 4\sqrt{sn}, \tag{30}$$

as stated in the lemma. We define the norm $\|\cdot\|_*$ on $\mathbb{C}^{n\times n}$ by

$$\|\boldsymbol{x}\|_* = \sum_\lambda |\operatorname{Re} x_\lambda| + |\operatorname{Im} x_\lambda|\,. \tag{31}$$

For $\boldsymbol{x} \in T_s$ we define a random vector $\boldsymbol{Z}$, which takes $\|\boldsymbol{x}\|_* \operatorname{sgn}(\operatorname{Re} x_\lambda)\boldsymbol{e}_\lambda$ with probability $\frac{|\operatorname{Re} \boldsymbol{x}_\lambda|}{\|\boldsymbol{x}\|_*}$, and the value $i\|\boldsymbol{x}\|_* \operatorname{sgn}(\operatorname{Im} x_\lambda)\boldsymbol{e}_\lambda$ with probability $\frac{|\operatorname{Im} x_\lambda|}{\|\boldsymbol{x}\|_*}$.

Now, let $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m, \boldsymbol{Z}'_1, \ldots, \boldsymbol{Z}'_m$ be independent copies of $\boldsymbol{Z}$. We set $\boldsymbol{y} = \frac{1}{m}\sum_{j=1}^m \boldsymbol{Z}_j$ and $\boldsymbol{y}' = \frac{1}{m}\sum_{j=1}^m \boldsymbol{Z}'_j$ and attempt to approximate $\boldsymbol{B}(\boldsymbol{x})$ by

$$\boldsymbol{B} := \boldsymbol{B}(\boldsymbol{y}, \boldsymbol{y}') = \frac{1}{m^2}\sum_{j,j'=1}^m \boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}'_{j'})\,. \tag{32}$$

First, compute

$$\begin{aligned}
\mathbb{E}\|\boldsymbol{B} - \boldsymbol{B}(\boldsymbol{x})\|_F^2 &= \mathbb{E}\sum_{q,q'} \Big|\boldsymbol{x}^*\boldsymbol{W}_{q',q}\boldsymbol{x} - \frac{1}{m^2}\sum_{j,j'=1}^m \boldsymbol{Z}_j^*\boldsymbol{W}_{q',q}\boldsymbol{Z}'_{j'}\Big|^2 \\
&= \sum_{q,q'} \Big(|\boldsymbol{x}^*\boldsymbol{W}_{q',q}\boldsymbol{x}|^2 - 2\operatorname{Re}\Big(\overline{\boldsymbol{x}^*\boldsymbol{W}_{q',q}\boldsymbol{x}}\,\mathbb{E}\Big[\frac{1}{m^2}\sum_{j,j'=1}^m \boldsymbol{Z}_j^*\boldsymbol{W}_{q,q'}\boldsymbol{Z}'_{j'}\Big]\Big) \\
&\quad + \mathbb{E}\Big[\Big|\frac{1}{m^2}\sum_{j,j'=1}^m \boldsymbol{Z}_j^*\boldsymbol{W}_{q,q'}\boldsymbol{Z}'_{j'}\Big|^2\Big]\Big) \\
&= \sum_{q,q'} \Big(- |\boldsymbol{x}^*\boldsymbol{W}_{q',q}\boldsymbol{x}|^2 + \frac{1}{m^4}\sum_{j,j',j'',j'''=1}^m \mathbb{E}\Big[\boldsymbol{Z}_j^*\boldsymbol{W}_{q,q'}\boldsymbol{Z}'_{j'}(\boldsymbol{Z}'_{j''})^*\boldsymbol{W}_{q,q'}^*\boldsymbol{Z}_{j'''}\Big]\Big)\,,
\end{aligned}$$

where we used that $\mathbb{E}[\boldsymbol{Z}_j^*\boldsymbol{W}_{q,q'}\boldsymbol{Z}'_{j'}] = \boldsymbol{x}^*\boldsymbol{W}_{q,q'}\boldsymbol{x}$, $j, j' = 1, \ldots m$, by independence. Moreover, for $j \neq j'''$ and $j' \neq j''$, independence implies

$$\mathbb{E}\Big[\boldsymbol{Z}_j^*\boldsymbol{W}_{q,q'}\boldsymbol{Z}'_{j'}(\boldsymbol{Z}'_{j''})^*\boldsymbol{W}_{q,q'}^*\boldsymbol{Z}_{j'''}\Big] = |\boldsymbol{x}^*\boldsymbol{W}_{q,q'}\boldsymbol{x}|^2.$$

To estimate summands with $j' = j''$, note that

$$\boldsymbol{Z}_j^* \boldsymbol{W}_{q',q} \boldsymbol{Z}'_{j'} (\boldsymbol{Z}'_{j'})^* \boldsymbol{W}_{q,q'} \boldsymbol{Z}_{j'''} = \|\boldsymbol{x}\|_*^2 \boldsymbol{Z}_j^* \boldsymbol{A}_{q'}^* \boldsymbol{A}_q \boldsymbol{P}_{\{\lambda\}} \boldsymbol{A}_q^* \boldsymbol{A}_{q'} \boldsymbol{Z}_{j'''},$$

where $\{\lambda\} = \operatorname{supp} \boldsymbol{Z}_{j'}$ is random. Hence, in this case, we compute using (27) in Lemma 5

$$\begin{aligned}
&\sum_{q' \neq q} \mathbb{E}\Big[\boldsymbol{Z}_j^* \boldsymbol{A}_{q'}^* \boldsymbol{A}_q \boldsymbol{Z}'_{j'} (\boldsymbol{Z}'_{j'})^* \boldsymbol{A}_q^* \boldsymbol{A}_{q'} \boldsymbol{Z}_{j'''}\Big] \\
&\leq \|\boldsymbol{x}\|_*^2 \sum_{q',q} \mathbb{E}\Big[\boldsymbol{Z}_j^* \boldsymbol{A}_{q'}^* \boldsymbol{A}_q \boldsymbol{P}_{\{\lambda\}} \boldsymbol{A}_q^* \boldsymbol{A}_{q'} \boldsymbol{Z}_{j'''}\Big] \\
&= \|\boldsymbol{x}\|_*^2 \mathbb{E}\Big[\boldsymbol{Z}_j^* \sum_{q'} \Big(\boldsymbol{A}_{q'}^* \Big(\sum_q \boldsymbol{A}_q \boldsymbol{P}_{\{\lambda\}} \boldsymbol{A}_q^*\Big) \boldsymbol{A}_{q'}\Big) \boldsymbol{Z}_{j'''}\Big] \\
&= \|\boldsymbol{x}\|_*^2 \mathbb{E}\Big[\boldsymbol{Z}_j^* \sum_{q'} \Big(\boldsymbol{A}_{q'}^* \boldsymbol{A}_{q'}\Big) \boldsymbol{Z}_{j'''}\Big] = n \|\boldsymbol{x}\|_*^2 \mathbb{E}[\boldsymbol{Z}_j^* \boldsymbol{Z}_{j'''}] \\
&= \begin{cases} n\|\boldsymbol{x}\|_*^4, & \text{if } j = j''', \\ n\|\boldsymbol{x}\|_*^2 \mathbb{E}[\boldsymbol{Z}_j^*]\mathbb{E}[\boldsymbol{Z}_{j'''}] = n\|\boldsymbol{x}\|_*^2 \|\boldsymbol{x}\|_2^2 \leq n\|\boldsymbol{x}\|_*^2, & \text{else.} \end{cases}
\end{aligned}$$

Symmetry implies an identical estimate for $j = j'''$, $j' \neq j''$. As $\boldsymbol{x} \in T_s$ is s-sparse we have $\|\boldsymbol{x}\|_* \leq \sqrt{2}\|\boldsymbol{x}\|_1 \leq \sqrt{2s}\|\boldsymbol{x}\|_2 \leq \sqrt{2s}$. Using (10) we conclude

$$\begin{aligned}
&\sum_{q',q} \sum_{j,j',j'',j'''=1}^{m} \mathbb{E}\Big[\boldsymbol{Z}_j^* \boldsymbol{W}_{q,q'} \boldsymbol{Z}'_{j'} (\boldsymbol{Z}'_{j''})^* \boldsymbol{W}_{q,q'}^* \boldsymbol{Z}_{j'''}\Big] \\
&\leq m^2(m-1)^2 \sum_{q',q} |\boldsymbol{x}^* \boldsymbol{W}_{q,q'} \boldsymbol{x}|^2 + m^2 n 4 s^2 + 2m^2(m-1)n \cdot 2s.
\end{aligned}$$

For $m \geq \frac{11ns^{\frac{3}{2}}}{u^2}$ and $u \leq 4\sqrt{sn}$, we finally obtain,

$$\begin{aligned}
\mathbb{E}\|\boldsymbol{B} - \boldsymbol{B}(\boldsymbol{x})\|_F^2 &\leq \sum_{q',q} -|\boldsymbol{x}^* \boldsymbol{W}_{q',q} \boldsymbol{x}|^2 + \frac{m^2(m^2-1)}{m^4} \sum_{q',q} |\boldsymbol{x}^* \boldsymbol{W}_{q,q'} \boldsymbol{x}|^2 \\
&\quad + \frac{m^2 n 4 s^2}{m^4} + \frac{4m^2(m-1)ns}{m^4} \\
&\leq \frac{4ns^2}{m^2} + \frac{4ns}{m} \leq \frac{4ns^2}{121 n^2 s^3} u^4 + \frac{4ns}{11ns^{\frac{3}{2}}} u^2 \leq \frac{64ns}{121ns} u^2 + \frac{44}{121\sqrt{s}} u^2 \leq u^2 \,.
\end{aligned} \tag{33}$$

Since $\|\boldsymbol{x}\|_*$ can take any value in $[1, \sqrt{2s}]$, we still have to discretize this factor in the definition of the random variable $\boldsymbol{Z}$. To this end, set

$$\boldsymbol{B}_\alpha := \frac{1}{m^2} \sum_{j=1,j'=1}^{m} \boldsymbol{B}(\alpha \operatorname{sgn}(x_{\lambda_j}) \boldsymbol{e}_{\lambda_j}, \alpha \operatorname{sgn}(x_{\lambda'_{j'}}) \boldsymbol{e}_{\lambda'_{j'}}) \,.$$

Next, we observe that, for $\lambda = (k,\ell)$ and $\lambda' = (k',\ell')$,

$$\boldsymbol{B}(\boldsymbol{e}_{\lambda'},\boldsymbol{e}_\lambda)_{q',q} = (\boldsymbol{A}_{q'}e_{\lambda'})^*\boldsymbol{A}_q\boldsymbol{e}_\lambda = \langle \boldsymbol{\pi}(\lambda)\boldsymbol{e}_q, \boldsymbol{\pi}(\lambda')\boldsymbol{e}_{q'}\rangle$$
$$= \begin{cases} \omega^{(\ell-\ell')(k+q)}, & \text{if } k'+q' = k+q\,; \\ 0, & \text{else,} \end{cases} \tag{34}$$

and, hence, $\|\boldsymbol{B}(\boldsymbol{e}_{\lambda'},\boldsymbol{e}_\lambda)\|_F^2 = n$. Now, assume α is chosen such that $|\|\boldsymbol{x}\|_*^2 - \alpha^2| \leq \frac{u}{\sqrt{n}}$. Then

$$\begin{aligned}
&\|\boldsymbol{B}_\alpha - \boldsymbol{B}_{\|\boldsymbol{x}\|_*}\|_F \\
&= \Big\| \frac{1}{m^2} \sum_{j=1,j'=1}^{m} \boldsymbol{B}(\alpha\,\mathrm{sgn}(x_{\lambda_j})\boldsymbol{e}_{\lambda_j}, \alpha\,\mathrm{sgn}(x_{\lambda'_{j'}})\boldsymbol{e}_{\lambda'_{j'}}) \\
&\qquad - \frac{1}{m^2} \sum_{j,j'=1}^{m} \boldsymbol{B}(\|\boldsymbol{x}\|_*\,\mathrm{sgn}(x_{\lambda_j})\boldsymbol{e}_{\lambda_j}, \|\boldsymbol{x}\|_*\,\mathrm{sgn}(x_{\lambda'_{j'}})\boldsymbol{e}_{\lambda'_{j'}}) \Big\|_F \\
&= |\|\boldsymbol{x}\|_*^2 - \alpha^2| \|\frac{1}{m^2} \sum_{j,j'=1}^{m} \boldsymbol{B}(\mathrm{sgn}(x_{\lambda_j})\boldsymbol{e}_{\lambda_j}, \mathrm{sgn}(x_{\lambda'_{j'}})\boldsymbol{e}_{k'_{j'}})\|_F \\
&\leq \frac{u}{m^2\sqrt{n}} \sum_{j,j'=1}^{m} \|\boldsymbol{B}(\boldsymbol{e}_{\lambda_j}, \boldsymbol{e}_{\lambda_{j'}})\|_F \\
&= u.
\end{aligned} \tag{35}$$

We conclude that it suffices to choose

$$K := \left\lceil \frac{2s-1}{\frac{u}{\sqrt{n}}} \right\rceil \leq \lceil 2s\sqrt{n}/u \rceil$$

values $\alpha_k \in J_s := [1,2s]$, $k = 1,\ldots,K$, such that for each $\beta \in J_s$ there exists k satisfying $|\beta - \alpha_k| \leq u/\sqrt{n}$.

Now, given $\boldsymbol{x}$ we can find $\boldsymbol{z}_1,\ldots,\boldsymbol{z}_m,\boldsymbol{z}'_1,\ldots,\boldsymbol{z}'_m$ of the form $\|\boldsymbol{x}\|_* p_\lambda \boldsymbol{e}_\lambda$, $p_\lambda \in \{1,-1,i,-i\}$ such that $\|\boldsymbol{B} - \boldsymbol{B}(\boldsymbol{x})\|_F \leq u$. Further, we can find k such that $|\|\boldsymbol{x}\|_*^2 - \alpha_k^2| \leq u/\sqrt{n}$. We replace the $\boldsymbol{z}_1,\ldots,\boldsymbol{z}_m,\boldsymbol{z}'_1,\ldots,\boldsymbol{z}'_m$ by the respective $\tilde{\boldsymbol{z}}_1,\ldots,\tilde{\boldsymbol{z}}_m,\tilde{\boldsymbol{z}}'_1,\ldots,\tilde{\boldsymbol{z}}'_m$ of the form $\alpha_j p_\lambda \boldsymbol{e}_\lambda$.

Then, using (33), (35) and the triangle inequality, we obtain

$$\|\boldsymbol{B}(\boldsymbol{x}) - \frac{1}{m^2}\sum_{j,j'=1}^{m} \boldsymbol{B}(\tilde{\boldsymbol{z}}_j, \tilde{\boldsymbol{z}}'_{j'})\|_F \leq 2u.$$

Now, each $\tilde{\boldsymbol{z}}_j$, $\tilde{\boldsymbol{z}}'_j$ can take at most $\lceil 2s\sqrt{n}/u\rceil \cdot 4 \cdot n^2$ values, so that

$$\frac{1}{m^2}\sum_{j,j'=1}^{m} \boldsymbol{B}(\tilde{\boldsymbol{z}}_j, \tilde{\boldsymbol{z}}'_{j'})$$

can take at most $(4\lceil\frac{2s\sqrt{n}}{u}\rceil n^2)^{2m} \leq (Csn^{\frac{5}{2}}/u)^{2m}$ values. Hence, we found a $2u$-covering of the set of matrices $\boldsymbol{B}(\boldsymbol{x})$ with $\boldsymbol{x} \in T_s$ of cardinality at most

$(Csn^{\frac{5}{2}}/u)^{2m}$. Unfortunately, the matrices of the covering are not necessarily of the form $\boldsymbol{B}(\boldsymbol{x})$. Nevertheless, we may replace each matrix of the form $\frac{1}{m^2}\sum_{j,j'=1}^{m}\boldsymbol{B}(\tilde{\boldsymbol{z}}_j,\tilde{\boldsymbol{z}}'_{j'})$ which is used to cover some $\boldsymbol{B}(\boldsymbol{x})$ by a matrix $\boldsymbol{B}(\tilde{\boldsymbol{x}})$ with

$$\|\boldsymbol{B}(\tilde{\boldsymbol{x}}) - \frac{1}{m^2}\sum_{j,j'=1}^{m}\boldsymbol{B}(\tilde{\boldsymbol{z}}_j,\tilde{\boldsymbol{z}}'_{j'})\|_F \leq 2u.$$

Again, the set of such chosen $\tilde{\boldsymbol{x}}$ has cardinality at most $(Csn^{\frac{5}{2}}/u)^{2m}$ and, by the triangle inequality, for each $\boldsymbol{x}$ we can find $\tilde{\boldsymbol{x}}$ of the covering such that

$$d_2(\boldsymbol{x},\tilde{\boldsymbol{x}}) \leq 4u.$$

For $m \geq 11u^{-2}ns^{\frac{3}{2}}$, we consequently get

$$\log(N(T_s,d_2,4u)) \leq \log((Csn^{\frac{5}{2}}/u)^{2m}) = 2m\log(Cns^{5/2}/u).$$

The choice $m = \lceil 11u^{-2}ns^{\frac{3}{2}}\rceil \leq 27u^{-2}ns^{\frac{3}{2}}$ and rescaling gives

$$\log(N(T_s,d_2,u)) \leq 27u^{-2}ns^{\frac{3}{2}}\log(4Cns^{5/2}/u) \leq cu^{-2}ns^{\frac{3}{2}}\log(ns^{5/2}/u).$$

The proof of Lemma 2 is completed. □

3.4 Proof of Lemma 3, Part I

Now we show the estimate

$$\log(N(T_s,d_1,u)) \leq s\log(en^2/s) + s\log(1+4su^{-1}),$$

which will establish one part of (17). Before doing so, we note that one can quickly obtain an estimate for $N(T_s,d_1,u)$ for small u using that the Frobenius norm dominates the operator norm, and, hence $d_1(\boldsymbol{x},\boldsymbol{y}) \leq d_2(\boldsymbol{x},\boldsymbol{y}) \leq 2\sqrt{sn}\|\boldsymbol{x}-\boldsymbol{y}\|_2$. In fact, this estimate would not deteriorate the estimate in Theorem 1(a). But in the proof of Theorem 1(b), the more involved estimate $d_1(\boldsymbol{x},\boldsymbol{y}) \leq 2s\|\boldsymbol{x}-\boldsymbol{y}\|_2$ developed below is useful.

Let us first rewrite d_1. Recall (25) in Lemma 5, namely, $\boldsymbol{A}_q\boldsymbol{e}_\lambda = \boldsymbol{\pi}(\lambda)\boldsymbol{e}_q$, and, with $\lambda=(k,\ell)$ and $\lambda'=(k',\ell')$, we obtain

$$\boldsymbol{\pi}(\lambda')^*\boldsymbol{\pi}(\lambda) = \omega^{k'(\ell-\ell'')}\boldsymbol{\pi}(\lambda-\lambda') \equiv \omega(\lambda,\lambda')\boldsymbol{\pi}(\lambda-\lambda').$$

Writing now $\boldsymbol{x} = \sum_{\lambda\in\mathbb{Z}_n\times\mathbb{Z}_n} x_\lambda \boldsymbol{e}_\lambda$, the entries of the matrix $\boldsymbol{B}(\boldsymbol{x})$ in (24) for $q' \neq q$ are given by

$$\begin{aligned}\boldsymbol{B}(\boldsymbol{x})_{q'q} &= \sum_{\lambda,\lambda'} x_\lambda \overline{x}_{\lambda'}\boldsymbol{e}^*_{\lambda'}\boldsymbol{A}^*_{q'}\boldsymbol{A}_q\boldsymbol{e}_\lambda = \sum_{\lambda,\lambda'} x_\lambda\overline{x}_{\lambda'}\boldsymbol{e}^*_{q'}\boldsymbol{\pi}(\lambda')^*\boldsymbol{\pi}(\lambda)\boldsymbol{e}_q \\ &= \sum_{\lambda,\lambda'} x_\lambda\overline{x}_{\lambda'}\omega(\lambda,\lambda')\ \boldsymbol{e}^*_{q'}\boldsymbol{\pi}(\lambda-\lambda')\boldsymbol{e}_q = \sum_{\lambda\neq\lambda'} x_\lambda\overline{x}_{\lambda'}\omega(\lambda,\lambda')\ \boldsymbol{e}^*_{q'}\pi(\lambda-\lambda')\boldsymbol{e}_q \\ &= \boldsymbol{e}^*_{q'}\Big(\sum_{\lambda\neq\lambda'} x_\lambda\overline{x}_{\lambda'}\omega(\lambda,\lambda')\ \boldsymbol{\pi}(\lambda-\lambda')\Big)\boldsymbol{e}_q.\end{aligned}$$

We used for the fourth inequality that $\boldsymbol{e}_{q'}^*\boldsymbol{\pi}(\ell_0,k_0)\boldsymbol{e}_q = 0$ if $q' \neq q$ and $k_0 = 0$. This shows that

$$\boldsymbol{B}(\boldsymbol{x}) = \sum_{\lambda\neq\lambda'} x_\lambda \overline{x}_{\lambda'}\omega(\lambda,\lambda')\ \boldsymbol{\pi}(\lambda-\lambda').$$

The estimate (23) for the Schatten norms shows

$$\begin{aligned}
d_1^{2p}(\boldsymbol{x},\boldsymbol{y}) &= \|\sum_{\lambda\neq\lambda'}(x_\lambda\overline{x}_{\lambda'} - y_\lambda\overline{y}_{\lambda'})\omega(\lambda,\lambda')\ \boldsymbol{\pi}(\lambda-\lambda')\|_{2\to2}^{2p}\\
&\leq \|\sum_{\lambda\neq\lambda'}(x_\lambda\overline{x}_{\lambda'} - y_\lambda\overline{y}_{\lambda'})\omega(\lambda,\lambda')\ \boldsymbol{\pi}(\lambda-\lambda')\|_{S_{2p}}^{2p}\\
&= \sum_{\lambda_1\neq\lambda_1',\lambda_2\neq\lambda_2',\dots,\lambda_{2p}\neq\lambda_{2p}'} (x_{\lambda_1}\overline{x}_{\lambda_1'} - y_{\lambda_1}\overline{y}_{\lambda_1'})\cdots(x_{\lambda_{2p}}\overline{x}_{\lambda_{2p}'} - y_{\lambda_{2p}}\overline{y}_{\lambda_{2p}'})\times\\
&\quad\times\omega(\lambda_1,\lambda_1')\cdots\omega(\lambda_{2p},\lambda_{2p}')\ \mathrm{Tr}\Big(\boldsymbol{\pi}(\lambda_1-\lambda_1')\cdots\boldsymbol{\pi}(\lambda_{2p}-\lambda_{2p}')\Big).
\end{aligned}$$

Setting $(\ell_0,k_0) = \lambda_1-\lambda_1'+\lambda_2-\lambda_2'+\cdots+\lambda_{2p}-\lambda_{2p}'$ we observe that the trace in the last expression sums over zero entries if $k_0\neq 0$ and sums over roots of unity to zero if $\ell_0 \neq 0$. We conclude that

$$\Big|\mathrm{Tr}\Big(\boldsymbol{\pi}(\lambda_1-\lambda_1')\cdots\boldsymbol{\pi}(\lambda_{2p}-\lambda_{2p}')\Big)\Big| \leq n\,\delta_{0,\lambda_1-\lambda_1'+\lambda_2-\lambda_2'+\cdots+\lambda_{2p}-\lambda_{2p}'}\,.$$

Hence,

$$\begin{aligned}
d_1(\boldsymbol{x},\boldsymbol{y})^{2p} &\leq n\sum_{\lambda_1\neq\lambda_1'}\big|x_{\lambda_1}\overline{x}_{\lambda_1'} - y_{\lambda_1}\overline{y}_{\lambda_1'}\big|\sum_{\lambda_2\neq\lambda_2'}\big|x_{\lambda_2}\overline{x}_{\lambda_2'} - y_{\lambda_2}y_{\lambda_2'}\big|\cdots\\
&\cdots\sum_{\lambda_{2p-1}\neq\lambda_{2p-1}'}\big|x_{\lambda_{2p-1}}\overline{x}_{\lambda_{2p-1}'} - y_{\lambda_{2p-1}}\overline{y}_{\lambda_{2p-1}'}\big|\sum_{\lambda_{2p}}\big|x_{\lambda_{2p}}\overline{x}_{\lambda_1-\lambda_1'+\cdots+\lambda_{2p}} - y_{\lambda_{2p}}\overline{y}_{\lambda_1-\lambda_1'+\cdots+\lambda_{4p}}\big|.
\end{aligned}$$

Now observe that, setting $t = \lambda_1-\lambda_1'+\cdots+\lambda_{2p-1}-\lambda_{2p-1}'$, and using the Cauchy-Schwarz inequality

$$\begin{aligned}
&\sum_\lambda |x_\lambda\overline{x}_{t+\lambda} - y_\lambda\overline{y}_{t+\lambda}| \leq \sum_\lambda |x_\lambda||x_{t+\lambda}-y_{t+\lambda}| + \sum_\lambda |x_\lambda - y_\lambda||y_{\lambda+t}|\\
&\leq \|\boldsymbol{x}\|_2\|\boldsymbol{x}-\boldsymbol{y}\|_2 + \|\boldsymbol{x}-\boldsymbol{y}\|_2\|\boldsymbol{y}\|_2 = (\|\boldsymbol{x}\|_2+\|\boldsymbol{y}\|_2)\|\boldsymbol{x}-\boldsymbol{y}\|_2.
\end{aligned}$$

We obtain similarly

$$\sum_{\lambda,\lambda'}|x_\lambda\overline{x}_{\lambda'} - y_\lambda\overline{y}_{\lambda'}| = \sum_{\lambda,\lambda'}|x_\lambda|\,|x_{\lambda'}-y_{\lambda'}| + |y_{\lambda'}|\,|x_\lambda - y_\lambda| \leq (\|\boldsymbol{x}\|_1+\|\boldsymbol{y}\|_1)\|\boldsymbol{x}-\boldsymbol{y}\|_1.$$

For $\boldsymbol{x},\boldsymbol{y}$ with $\operatorname{supp}\boldsymbol{x} = \operatorname{supp}\boldsymbol{y} = \Lambda$ for $|\Lambda|\leq s$ and $\|\boldsymbol{x}\|_2 = \|\boldsymbol{y}\|_2 = 1$ we have $\|\boldsymbol{x}\|_1 \leq \sqrt{s}\|\boldsymbol{x}\|_2 = \sqrt{s}$ (and similarly for $\boldsymbol{y}$) as well as $\|\boldsymbol{x}-\boldsymbol{y}\|_1 \leq \sqrt{s}\|\boldsymbol{x}-\boldsymbol{y}\|_2$. Hence,

$$(\|\boldsymbol{x}\|_1+\|\boldsymbol{y}\|_1)\|\boldsymbol{x}-\boldsymbol{y}\|_1 \leq 2s\|\boldsymbol{x}-\boldsymbol{y}\|_2.$$

This finally yields

$$d_1(\boldsymbol{x},\boldsymbol{y})^{2p} \le 2^{2p} n s^{2p-1} \|\boldsymbol{x}-\boldsymbol{y}\|_2^{2p}$$

for such $\boldsymbol{x},\boldsymbol{y}$. As this holds for all $p \in \mathbb{N}$ we conclude that

$$d_1(\boldsymbol{x},\boldsymbol{y}) \le 2s\|\boldsymbol{x}-\boldsymbol{y}\|_2. \tag{36}$$

With the volumetric argument, see for example [33, Proposition 10.1], we obtain the bound

$$\log(N(T_s,\|\cdot\|_2,u)) \le s\log(en^2/s) + s\log(1+2/u).$$

Rescaling yields

$$\begin{aligned}\log(N(T_s,d_1,u)) &\le \log(N(T_s,2s\|\cdot\|_2,u)) = \log(N(T_s,\|\cdot\|_2,u/(2s)))\\ &\le s\log(en^2/s) + s\log(1+4su^{-1}),\end{aligned}$$

which is the claimed inequality. □

3.5 Proof of Lemma 3, Part II

Next we establish the remaining estimate of (17),

$$\log(N(T_s,d_1,u)) \le cu^{-2}s^2\log(2n)\log(n^2/u).$$

To this end, we use again Maurey's empirical method as in Section 3.3.

For $\boldsymbol{x} \in T_s$, we define $\boldsymbol{Z}_1,\dots,\boldsymbol{Z}_m$ and $\boldsymbol{Z}'_1,\dots,\boldsymbol{Z}'_m$ as in Section 3.3, that is, each takes independently the value $\|\boldsymbol{x}\|_* \operatorname{sgn}(\operatorname{Re} x_\lambda)\boldsymbol{e}_\lambda$ with probability $\frac{|\operatorname{Re} x_\lambda|}{\|\boldsymbol{x}\|_*}$, and the value $i\|\boldsymbol{x}\|_* \operatorname{sgn}(\operatorname{Im} x_\lambda)\boldsymbol{e}_\lambda$ with probability $\frac{|\operatorname{Im} x_\lambda|}{\|\boldsymbol{x}\|_*}$.

As before, we set

$$B(\boldsymbol{Z},\boldsymbol{Z}') = (\boldsymbol{Z}^*\boldsymbol{W}_{q'q}\boldsymbol{Z}')_{q',q}, \tag{37}$$

where $\boldsymbol{W}_{q'q} = \boldsymbol{A}_{q'}^*\boldsymbol{A}_q$ for $q' \ne q$ and $\boldsymbol{W}_{q,q} = 0$, $j = 1,\dots,N$, and attempt to approximate $\boldsymbol{B}(\boldsymbol{x})$ with

$$\boldsymbol{B} := \frac{1}{m}\sum_{j=1}^{m} \boldsymbol{B}(\boldsymbol{Z}_j,\boldsymbol{Z}'_j). \tag{38}$$

That is, we will estimate $\mathbb{E}\|\boldsymbol{B}-\boldsymbol{B}(\boldsymbol{x})\|_{2\to 2}^2$.

We will use symmetrization as formulated in the following lemma [33, Lemma 6.7], see also [26, Lemma 6.3], [15, Lemma 1.2.6]. Note that we will use this result with $\boldsymbol{Y}_j = \boldsymbol{B}(\boldsymbol{Z}_j,\boldsymbol{Z}'_j)$.

Lemma 6 *(Symmetrization) Assume that* $(\boldsymbol{Y}_j)_{j=1}^m$ *is a sequence of independent random vectors in* $\mathbb{C}^r$ *equipped with a (semi-)norm* $\|\cdot\|$*, having expectations* $\beta_j = \mathbb{E}\boldsymbol{Y}_j$*. Then for* $1 \leq p < \infty$

$$\Big(\mathbb{E}\|\sum_{j=1}^m(\boldsymbol{Y}_j - \beta_j)\|^p\Big)^{1/p} \leq 2\Big(\mathbb{E}\|\sum_{j=1}^m \epsilon_j \boldsymbol{Y}_j\|^p\Big)^{1/p}, \tag{39}$$

where $(\epsilon_j)_{j=1}^N$ *is a Rademacher series independent of* $(\boldsymbol{Y}_j)_{j=1}^m$*.*

To estimate the $2p$-th moment of $\|\boldsymbol{B}(\boldsymbol{x}) - \boldsymbol{B}\|_{2\to 2}$, we will use the noncommutative Khintchine inequality [6, 33] which makes use of the Schatten p-norms introduced in (21).

Theorem 4 *(Noncommutative Khintchine inequality) Let* $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_m)$ *be a Rademacher sequence, and let* $\boldsymbol{A}_j$*,* $j = 1, \ldots, m$*, be complex matrices of the same dimension. Choose* $p \in \mathbb{N}$*. Then*

$$\mathbb{E}\|\sum_{j=1}^m \epsilon_j \boldsymbol{A}_j\|_{S_{2p}}^{2p} \leq \frac{(2p)!}{2^p p!} \max\Big\{\Big\|\Big(\sum_{j=1}^m \boldsymbol{A}_j\boldsymbol{A}_j^*\Big)^{1/2}\Big\|_{S_{2p}}^{2p}, \Big\|\Big(\sum_{j=1}^m \boldsymbol{A}_j^*\boldsymbol{A}_j\Big)^{1/2}\Big\|_{S_{2p}}^{2p}\Big\}. \tag{40}$$

Let $p \in \mathbb{N}$. We apply symmetrization with $\boldsymbol{Y}_j = \boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}'_j)$, estimate the operator norm by the Schatten-$2p$-norm and apply the noncommutative Khintchine inequality (after using Fubini's theorem), to obtain

$$\begin{aligned}
&\Big(\mathbb{E}\|\boldsymbol{B} - \boldsymbol{B}(\boldsymbol{x})\|_{2\to 2}^{2p}\Big)^{\frac{1}{2p}} = \Big(\mathbb{E}\|\frac{1}{m}\sum_{j=1}^m(\boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}'_j) - \mathbb{E}\boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}'_j))\|_{2\to 2}^{2p}\Big)^{\frac{1}{2p}} \\
&\leq \frac{2}{m}\Big(\mathbb{E}\|\sum_{j=1}^m \epsilon_j \boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}'_j)\|_{2\to 2}^{2p}\Big)^{\frac{1}{2p}} \leq \frac{2}{m}\Big(\mathbb{E}\|\sum_{j=1}^m \epsilon_j \boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}'_j)\|_{S_{2p}}^{2p}\Big)^{\frac{1}{2p}} \\
&\leq \frac{2}{m}\Big(\frac{(2p)!}{2^p p!}\Big)^{\frac{1}{2p}}\Big(\mathbb{E}\max\Big\{\Big\|\Big(\sum_{j=1}^m \boldsymbol{B}(Z_j, Z'_j)^*\boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}'_j)\Big)^{1/2}\Big\|_{S_{2p}}^{2p}, \\
&\qquad\qquad \Big\|\Big(\sum_{j=1}^m \boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}'_j)\boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}'_j)^*\Big)^{1/2}\Big\|_{S_{2p}}^{2p}\Big\}\Big)^{\frac{1}{2p}}.
\end{aligned} \tag{41}$$

Now recall that the $\boldsymbol{Z}_j, \boldsymbol{Z}'_j$ may take the values $\|\boldsymbol{x}\|_* p_\lambda \boldsymbol{e}_\lambda$, with $p_\lambda \in \{1, -1, i, -i\}$. Further, observe that $\boldsymbol{B}(\boldsymbol{e}_{\lambda'}, \boldsymbol{e}_\lambda)^* = \boldsymbol{B}(\boldsymbol{e}_\lambda, \boldsymbol{e}_{\lambda'})$, and, for $q \neq q'$,

$$\begin{aligned}
&(\boldsymbol{B}(\boldsymbol{e}_{\lambda'}, \boldsymbol{e}_\lambda)^*\boldsymbol{B}(\boldsymbol{e}_{\lambda'}, \boldsymbol{e}_\lambda))_{q,q''} = \sum_{q'} \boldsymbol{e}_\lambda^* \boldsymbol{A}_q^* \boldsymbol{A}_{q'} e_{\lambda'}\, \boldsymbol{e}_{\lambda'}^* \boldsymbol{A}_{q'}^* \boldsymbol{A}_{q''} \boldsymbol{e}_\lambda \\
&= \sum_{q'} \boldsymbol{e}_\lambda^* \boldsymbol{A}_q^* \boldsymbol{A}_{q'} \boldsymbol{P}_{\lambda'} \boldsymbol{A}_{q'}^* \boldsymbol{A}_{q''} \boldsymbol{e}_\lambda = \boldsymbol{e}_\lambda^* \boldsymbol{A}_q^* \Big(\sum_{q'} \boldsymbol{A}_{q'} \boldsymbol{P}_{\lambda'} \boldsymbol{A}_{q'}^*\Big)\ \boldsymbol{A}_{q''} \boldsymbol{e}_\lambda \\
&= \boldsymbol{e}_\lambda^* \boldsymbol{A}_q^* \boldsymbol{A}_{q''} \boldsymbol{e}_\lambda = \langle \boldsymbol{\pi}(\lambda) \boldsymbol{e}_{q''}, \boldsymbol{\pi}(\lambda) \boldsymbol{e}_q \rangle = \langle \boldsymbol{e}_{q''}, \boldsymbol{e}_q \rangle = \delta(q'' - q).
\end{aligned}$$

Therefore, $\boldsymbol{B}(\boldsymbol{e}_{\lambda'}, \boldsymbol{e}_\lambda)^* \boldsymbol{B}(\boldsymbol{e}_{\lambda'}, \boldsymbol{e}_\lambda) = \boldsymbol{I}$ and

$$\boldsymbol{B}(\boldsymbol{Z}_\ell, \boldsymbol{Z}'_\ell)^* \boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}'_j) = \|\boldsymbol{x}\|_*^4 \boldsymbol{I}. \tag{42}$$

Since $\|\boldsymbol{I}\|_{S_{2p}}^{2p} = n$, $\|\boldsymbol{x}\|_* \leq 2s\|\boldsymbol{x}\|_2 = 2s$, we obtain

$$\begin{aligned} &\|\Big(\sum_{j=1}^m \boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}'_j)^* \boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}'_j)\Big)^{1/2}\|_{S_{2p}}^{2p} = \|\Big(\sum_{j=1}^m \|\boldsymbol{x}\|_*^4 \boldsymbol{I}\Big)^{1/2}\|_{S_{2p}}^{2p} = \|\boldsymbol{x}\|_*^{4p} m^p n \\ &\leq (2s)^{2p} m^p n\,. \end{aligned} \tag{43}$$

By symmetry this inequality applies also to the second term in the maximum in (41). This yields

$$\Big(\mathbb{E}\|\boldsymbol{B} - \boldsymbol{B}(\boldsymbol{x})\|_{2\to 2}^{2p}\Big)^{\frac{1}{2p}} \leq \frac{2}{m}\Big(\frac{(2p)!}{2^q q!}\Big)^{\frac{1}{2p}} \; 2sm^{\frac{1}{2}} n^{\frac{1}{2p}} \leq \frac{4s}{\sqrt{m}} n^{1/(2p)} \Big(\frac{(2p)!}{2^p p!}\Big)^{\frac{1}{2p}}.$$

Using Hölder's inequality, we can interpolate between $2p$ and $2p+2$, and an application of Stirling's formula yields for arbitrary moments $p \geq 2$, see also [33],

$$\Big(\mathbb{E}\|\boldsymbol{B} - \boldsymbol{B}(\boldsymbol{x})\|_{2\to 2}^{p}\Big)^{1/p} \leq 2^{3/(4p)} n^{1/p} e^{-1/2} \sqrt{p} \frac{4s}{\sqrt{m}}. \tag{44}$$

Now we use the following lemma relating moments and tails [32,33].

Proposition 1 *Suppose Ξ is a random variable satisfying*

$$(\mathbb{E}|\Xi|^p)^{1/p} \leq \alpha \beta^{1/p} p^{1/\gamma} \quad \text{for all } p \geq p_0$$

for some constants $\alpha, \beta, \gamma, p_0 > 0$. Then

$$\mathbb{P}(|\Xi| \geq e^{1/\gamma} \alpha v) \leq \beta e^{-v^\gamma/\gamma}$$

for all $v \geq p_0^{1/\gamma}$.

Applying the lemma with $p_0 = 2$, $\gamma = 2$, $\beta = 2^{3/4} n$, $\alpha = e^{-1/2} \frac{4s}{\sqrt{m}}$, and

$$v = u \frac{e^{-1/\gamma}}{\alpha} = u \frac{e^{-1/2}\sqrt{m}}{e^{-1/2} 4s} = u \frac{\sqrt{m}}{4s} \geq \sqrt{2}$$

gives

$$\mathbb{P}\Big(\|\boldsymbol{B} - \boldsymbol{B}(\boldsymbol{x})\|_{2\to 2} \geq u\Big) \leq 2^{3/4} n e^{-\frac{mu^2}{32s^2}}, \quad u \geq 4s\sqrt{2/m}.$$

In particular, if

$$m > \frac{32s^2}{u^2} \log(2^{3/4} n) \tag{45}$$

then there exists a matrix of the form $\frac{1}{m}\sum_{j=1}^m \boldsymbol{B}(\boldsymbol{z}_j, \boldsymbol{z}_j')$ with $\boldsymbol{z}_j, \boldsymbol{z}_j'$ of the given form $\|\boldsymbol{x}\|_* p_\lambda \boldsymbol{e}_\lambda$ for some k such that

$$\Big\|\frac{1}{m}\sum_{j=1}^m \boldsymbol{B}(\boldsymbol{z}_j, \boldsymbol{z}_j') - \boldsymbol{B}(\boldsymbol{x})\Big\| \leq u.$$

As before, we still have to discretize the prefactor $\|\boldsymbol{x}\|_*$. Assume that α is chosen such that $|\|\boldsymbol{x}\|_*^2 - \alpha^2| \leq u$. Then, similarly as in (35),

$$\begin{aligned}
&\Big\|\frac{1}{m}\sum_{j=1}^m \boldsymbol{B}(\alpha\,\mathrm{sgn}(x_{\lambda_j})\boldsymbol{e}_{\lambda_j}, \alpha\,\mathrm{sgn}(x_{\lambda_{j'}})\boldsymbol{e}_{\lambda_{j'}}) \\
&\quad - \frac{1}{m}\sum_{j=1}^m \boldsymbol{B}(\|\boldsymbol{x}\|_1\,\mathrm{sgn}(x_{\lambda_j})\boldsymbol{e}_{\lambda_j}, \|\boldsymbol{x}\|_1\,\mathrm{sgn}(x_{\lambda_{j'}})\boldsymbol{e}_{\lambda_{j'}})\Big\|_{2\to 2} \\
&= |\|\boldsymbol{x}\|_1^2 - \alpha^2|\|\frac{1}{m}\sum_{j=1}^m \boldsymbol{B}(\mathrm{sgn}(x_{\lambda_j})\boldsymbol{e}_{\lambda_j}, \mathrm{sgn}(x_{\lambda_{j'}})\boldsymbol{e}_{\lambda_{j'}})\|_{2\to 2} \\
&\leq \frac{u}{m}\sum_{j=1}^m \|\boldsymbol{B}(\mathrm{sgn}(x_{\lambda_j})\boldsymbol{e}_{\lambda_j}, \mathrm{sgn}(x_{\lambda_{j'}})\boldsymbol{e}_{\lambda_{j'}})\|_{2\to 2} = u.
\end{aligned}$$

Hereby, we used $\|\boldsymbol{B}(\mathrm{sgn}(x_{\lambda_j})\boldsymbol{e}_{\lambda_j}, \mathrm{sgn}(x_{\lambda_{j'}})\boldsymbol{e}_{\lambda_{j'}})\|_{2\to 2} = 1$.

As in Section 3.3, we use a discretization of $J_s = [1, 2s]$ with about $K = \lceil \frac{2s}{u} \rceil$ elements, $\alpha_1, \ldots, \alpha_K$ such that for any β in J_s there exists k such $|\beta - \alpha_k^2| \leq u$. Now, provided (45) holds, for given $\boldsymbol{x}$ we can find $\tilde{\boldsymbol{z}}_1, \ldots, \tilde{\boldsymbol{z}}_m, \tilde{\boldsymbol{z}}_1', \ldots, \tilde{\boldsymbol{z}}_m'$ of the form $\alpha_k\,\mathrm{sgn}(x_\lambda)\boldsymbol{e}_\lambda$, $p(\lambda) \in \{1, -1, i, -i\}$, with

$$\|\boldsymbol{B}(\boldsymbol{x}) - \frac{1}{m}\sum_{j=1}^m \boldsymbol{B}(\tilde{\boldsymbol{z}}_j, \tilde{\boldsymbol{z}}_j')\|_{2\to 2} \leq 2u.$$

Observe as in Section 3.3 that each $\tilde{\boldsymbol{z}}_j$ can take $4\lceil \frac{2s}{u} \rceil n^2$ values, so that $\frac{1}{m}\sum_{j=1}^m B(\tilde{z}_j, \tilde{z}_j')$ can take at most $(4\lceil \frac{2s}{u} \rceil n^2)^{2m} \leq (Cn^2 s/u)^{2m}$ values. As seen before, this establishes a $4u$ covering of the set of matrices $\boldsymbol{B}(\boldsymbol{x})$ with $\boldsymbol{x} \in T_s$ of cardinality at most $(Cn^2 s/u)^{2m}$, and we conclude

$$\begin{aligned}
\log(N(T_s, d_1, u)) &\leq \log((Cn^2 s/u)^{2m}) \leq C' \frac{s^2}{u^2} \log(2^{3/4} n) \log(Cn^2 s/u) \\
&\leq \tilde{C} \frac{s^2}{u^2} \log(2n) \log(n^2/u).
\end{aligned}$$

This completes the proof of Lemma 3. □

4 Probability estimate

To prove Theorem 1(b) will use the following concentration inequality, which is a slight variant of Theorem 17 in [5], which in turn is an improved version of a striking result due to Talagrand [41]. Note that with $\boldsymbol{B}(\boldsymbol{x})$ as defined above, Y below satisfies $\mathbb{E}Y = n\,\mathbb{E}\delta_s$.

Theorem 5 *Let $\mathscr{B} = \{\boldsymbol{B}(\boldsymbol{x})\}_{\boldsymbol{x}\in T}$ be a countable collection of $n \times n$ complex Hermitian matrices, and let $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^T$ be a sequence of i.i.d. Rademacher or Steinhaus random variables. Assume that $B(\boldsymbol{x})_{q,q} = 0$ for all $\boldsymbol{x} \in T$. Let Y be the random variable*

$$Y = \sup_{\boldsymbol{x}\in T} \Big|\boldsymbol{\epsilon}^* \boldsymbol{B}(\boldsymbol{x})\boldsymbol{\epsilon}\Big| = \sup_{\boldsymbol{x}\in T} \Big| \sum_{q,q'=1}^{n} \overline{\epsilon_{q'}} \epsilon_q B(\boldsymbol{x})_{q',q} \Big|.$$

Define U and V to be

$$U = \sup_{\boldsymbol{x}\in T} \|\boldsymbol{B}(\boldsymbol{x})\|_{2\to 2}$$

and

$$V = \mathbb{E} \sup_{\boldsymbol{x}\in T} \|\boldsymbol{B}(\boldsymbol{x})\boldsymbol{\epsilon}\|_2^2 = \mathbb{E} \sup_{\boldsymbol{x}\in T} \sum_{q'=1}^{n} \Big| \sum_{q=1}^{n} \epsilon_q B(\boldsymbol{x})_{q',q} \Big|^2. \tag{46}$$

Then, for $\lambda \geq 0$,

$$\mathbb{P}\Big(Y \geq \mathbb{E}[Y] + \lambda\Big) \;\leq\; \exp\Big(- \frac{\lambda^2}{32V + 65U\lambda/3}\Big). \tag{47}$$

Proof For Rademacher variables, the statement is exactly Theorem 17 in [5]. For Steinhaus sequences, we provide a variation of its proof. For $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)$, let $g_{\boldsymbol{M}}(\boldsymbol{\epsilon}) = \sum_{j,k=1}^{n} \overline{\epsilon_j} \epsilon_k M_{j,k}$ and set

$$Y = f(\boldsymbol{\epsilon}) = \sup_{\boldsymbol{M}\in\mathscr{B}} \Big| g_{\boldsymbol{M}}(\boldsymbol{\epsilon}) \Big|.$$

Further, for an independent copy $\widetilde{\epsilon}_\ell$ of ϵ_ℓ, set $\boldsymbol{\epsilon}^{(\boldsymbol{\ell})} = (\epsilon_1, \ldots, \epsilon_{\ell-1}, \widetilde{\epsilon}_\ell, \epsilon_{\ell+1}, \ldots, \epsilon_n)$ and $Y^{(\ell)} = f(\boldsymbol{\epsilon}^{(\boldsymbol{\ell})})$. Conditional on $(\epsilon_1, \ldots, \epsilon_n)$, let $\widehat{\boldsymbol{M}} = \widehat{\boldsymbol{M}}(\boldsymbol{\epsilon})$ be the matrix giving the maximum in the definition of Y. (If the supremum is not attained, then one has to consider finite subsets $T \subset \mathscr{B}$. The derived estimate will not depend on T, so that one can afterwards pass over to the possibly infinite, but countable, set $\mathscr{B}$.) Then we obtain, using $\widehat{\boldsymbol{M}}^* = \widehat{\boldsymbol{M}}$ and $\widehat{M}_{kk} = 0$ in the last step,

$$\begin{aligned}
&\mathbb{E}\Big[(Y - Y^{(\ell)})^2 \mathbf{1}_{Z > Z^{(\ell)}} |\epsilon\Big] \leq \mathbb{E}\Big[|g_{\widehat{M}}(\boldsymbol{\epsilon}) - g_{\widehat{M}}(\boldsymbol{\epsilon}^{(\boldsymbol{\ell})})|^2 \mathbf{1}_{Z>Z^{(\ell)}} |\epsilon\Big] \\
&= \mathbb{E}\Big[|(\overline{\epsilon_\ell - \widetilde{\epsilon}_\ell}) \sum_{j=1, j\neq \ell}^{n} \epsilon_j \widehat{M}_{j,\ell} + (\epsilon_\ell - \widetilde{\epsilon}_\ell) \sum_{k=1, k\neq\ell}^{n} \overline{\epsilon_k} \widehat{M}_{\ell,k}|^2 \mathbf{1}_{Z>Z^{(\ell)}} |\epsilon\Big] \\
&\leq 4 \mathbb{E}_{\widetilde{\epsilon}_\ell} |\epsilon_\ell - \widetilde{\epsilon}_\ell|^2 \Big| \sum_{j=1, j\neq\ell}^{n} \epsilon_j \widehat{M}_{j,\ell} \Big|^2 = 8 \Big| \sum_{j=1}^{n} \epsilon_j \widehat{M}_{j,\ell} \Big|^2.
\end{aligned}$$

The remainder of the proof is analogous to the one in [5] and therefore omitted. □

We first note that we may pass from T_s to a dense countable subset T_s° without changing the supremum, hence Theorem 5 is applicable. Now, it remains to estimate U and V. To this end, note that (36) implies

$$U = \sup_{\boldsymbol{x}\in T_s} \|\boldsymbol{B}(\boldsymbol{x})\|_{2\to 2} \le \sup_{\boldsymbol{x}\in T_s} 2s\|\boldsymbol{x}\|_2 = 2s\,.$$

The remainder of this section develops an estimate of the quantity V in (46). Hereby, we rely on a Dudley type inequality for Rademacher or Steinhaus processes with values in ℓ_2, see below. First we note the following Hoeffding type inequality.

Proposition 2 *Let $\boldsymbol{\epsilon} = (\epsilon_q)_{q=1}^n$ be a Steinhaus sequence and let $\boldsymbol{B} \in \mathbb{C}^{m\times n}$. Then, for $u \ge 0$,*

$$\mathbb{P}\Big(\|\boldsymbol{B}\boldsymbol{\epsilon}\|_2 \ge u\|\boldsymbol{B}\|_F\Big) \le 8e^{-u^2/16}. \tag{48}$$

Proof In [35, Proposition B.1], it is shown that

$$\mathbb{P}\Big(\|\boldsymbol{B}\boldsymbol{\epsilon}\|_2 \ge u\|\boldsymbol{B}\|_F\Big) \le 2e^{-u^2/2}. \tag{49}$$

for Rademacher sequences. We extend this result using the contraction principle [26, Theorem 4.4], as in the proof of Theorem 3.

In fact, [26, Theorem 4.4] implies that for $\boldsymbol{B} \in \mathbb{C}^{n\times n}$ and $\boldsymbol{\epsilon}$ being a Steinhaus sequence and $\boldsymbol{\xi}$ a Rademacher sequence, we have, for example

$$\mathbb{P}(\|\operatorname{Re}(\boldsymbol{B})\operatorname{Re}(\boldsymbol{\epsilon})\|_2 \ge u\|\boldsymbol{B}\|_F) \le 2\mathbb{P}(\|\operatorname{Re}\boldsymbol{B}\boldsymbol{\xi}\|_2 \ge u\|\boldsymbol{B}\|_F) \le 4e^{-u^2/2}.$$

Hence,

$$\begin{aligned}
&\mathbb{P}(\|\boldsymbol{B}\boldsymbol{\epsilon}\|_2 \ge u\|\boldsymbol{B}\|_F) = \mathbb{P}(\|\operatorname{Re}(\boldsymbol{B}\boldsymbol{\epsilon})\|_2^2 + \|\operatorname{Im}(\boldsymbol{B}\boldsymbol{\epsilon})\|_2^2 \ge u^2\|\boldsymbol{B}\|_F^2)\\
&\le \mathbb{P}(\|\operatorname{Re}(\boldsymbol{B}\boldsymbol{\epsilon})\|_2^2 \ge \frac{u^2}{\sqrt{2}}) + \mathbb{P}(\|\operatorname{Im}(\boldsymbol{B}\boldsymbol{\epsilon})\|_2^2 \ge \frac{u}{\sqrt{2}}\|\boldsymbol{B}\|_F^2)\\
&\le \mathbb{P}(\|\operatorname{Re}\boldsymbol{B}\operatorname{Re}\boldsymbol{\epsilon})\|_2 \ge \frac{u}{\sqrt{8}}\|\boldsymbol{B}\|_F^2) + \mathbb{P}(\|\operatorname{Im}\boldsymbol{B}\operatorname{Im}\boldsymbol{\epsilon})\|_2 \ge \frac{u}{\sqrt{8}}\|\boldsymbol{B}\|_F^2)\\
&+ \mathbb{P}(\|\operatorname{Re}\boldsymbol{B}\operatorname{Im}\boldsymbol{\epsilon})\|_2 \ge \frac{u}{\sqrt{8}}\|\boldsymbol{B}\|_F^2) + \mathbb{P}(\|\operatorname{Im}\boldsymbol{B}\operatorname{Re}\boldsymbol{\epsilon})\|_2 \ge \frac{u}{\sqrt{8}}\|\boldsymbol{B}\|_F^2)\\
&\le 8e^{-u^2/16}.
\end{aligned}$$

□

With more effort, one may also derive (48) with better constants. Let us now estimate the quantity

$$V = \mathbb{E}\sup_{\boldsymbol{x}\in T_s} \|\boldsymbol{B}(\boldsymbol{x})\boldsymbol{\epsilon}\|_2^2 = \mathbb{E}\sup_{\boldsymbol{x}\in T_s} \sum_{q'=1} |\sum_{q=1} \epsilon_q B(\boldsymbol{x})_{q',q}|^2.$$

It follows immediately from Proposition 2 and (49) that the increments of the process satisfy

$$\mathbb{P}(\|\boldsymbol{B}(\boldsymbol{x})\boldsymbol{\epsilon} - \boldsymbol{B}(\boldsymbol{x}')\boldsymbol{\epsilon}\|_2 \geq u\|\boldsymbol{B}(\boldsymbol{x}) - \boldsymbol{B}(\boldsymbol{x}')\|_F) \leq 8e^{-u^2/16}. \tag{50}$$

This allows to apply the following variant of Dudley's inequality for vector-valued processes in ℓ_2.

Theorem 6 *Let $\boldsymbol{R}_x$, $\boldsymbol{x} \in T$, be a process with values in $\mathbb{C}^m$ indexed by a metric space (T, d), with increments that satisfy the subgaussian tail estimate*

$$\mathbb{P}(\|\boldsymbol{R}_{\boldsymbol{x}} - \boldsymbol{R}_{\boldsymbol{x}'}\|_2 \geq ud(\boldsymbol{x}, \boldsymbol{x}')) \leq 8e^{-u^2/16}.$$

Then, for an arbitrary $\boldsymbol{x_0} \in T$ and a universal constant $K > 0$,

$$\Big(\mathbb{E} \sup_{\boldsymbol{x}\in T} \|\boldsymbol{R}_{\boldsymbol{x}} - \boldsymbol{R}_{\boldsymbol{x_0}}\|_2^2\Big)^{1/2} \leq K \int_0^\infty \sqrt{\log(N(T, d, u))}du, \tag{51}$$

where $N(T, d, u)$ denote the covering numbers of T with respect to d and radius $u > 0$.

Proof The proof follows literally the lines of the standard proof of Dudley's inequalities for scalar-valued subgaussian processes, see for instance [33, Theorem 6.23] or [2,26,42]. One only has to replace the triangle inequality for the absolute value by the one for $\|\cdot\|_2$ in $\mathbb{C}^m$. □

We have $d = d_2$ defined above, and, hence, (18) provides us with the right hand side of (51). Using the fact that here, $\boldsymbol{R}_x = \boldsymbol{B}(\boldsymbol{x})\boldsymbol{\epsilon}$, we conclude that

$$\begin{aligned} V &= \mathbb{E} \sup_{\boldsymbol{x}\in T_s} \|\boldsymbol{B}(\boldsymbol{x})\boldsymbol{\epsilon}\|_2^2 = \mathbb{E} \sup_{\boldsymbol{x}\in T_s} \|\boldsymbol{B}(\boldsymbol{x})\boldsymbol{\epsilon} - \boldsymbol{B}(\boldsymbol{0})\boldsymbol{\epsilon}\|_2^2 \\ &\leq \big(KC\sqrt{ns^{3/2}}\sqrt{\log(n)}\log(s)\big)^2 \leq C'ns^{3/2}\log(n)\log^2(s). \end{aligned}$$

Plugging these estimates into (47) and simplifying leads to our result, compare with [35]. In particular, Theorem 1(b) follows.

Acknowledgements Götz E. Pfander appreciates the support by the *Deutsche Forschungsgemeinschaft* (DFG) under grant 50292 DFG PF-4 Sampling Operators. Holger Rauhut acknowledges generous support by the Hausdorff Center for Mathematics, and funding by the Starting Independent Researcher Grant StG-2010 258926-SPALORA from the European Research Council (ERC). Joel A. Tropp was supported in part by the Defense Advanced Research Projects Agency (DARPA) and the Office of Naval Research (ONR) under Grants N66001-06-1-2011 and N66001-08-1-2065.

References

1. W. O. Alltop. Complex sequences with low periodic correlations. *IEEE Trans. Inform. Theory*, 26(3):350–354, 1980.
2. J.-M. Azais and M. Wschebor. *Level Sets and Extrema of Random Processes and Fields.* John Wiley & Sons Inc., 2009.
3. R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.
4. P. A. Bello. Characterization of Randomly Time-Variant Linear Channels. *IEEE Trans. Comm.*, 11:360–393, 1963.
5. S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, 31(3):1583–1614, 2003.
6. A. Buchholz. Operator Khintchine inequality in non-commutative probability. *Math. Ann.*, 319:1–16, 2001.
7. E. J. Candès. Compressive sampling. In *Proceedings of the International Congress of Mathematicians*, Madrid, Spain, 2006.
8. E. J. Candès, J., T. Tao, and J. Romberg. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
9. E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.
10. E. J. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.
11. B. Carl. Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in Banach spaces. *Ann. Inst. Fourier (Grenoble)*, 35(3):79–118, 1985.
12. J. Chiu and L. Demanet. Matrix probing and its conditioning. *SIAM J. Numer. Anal.*, 50(1):171–193, 2012.
13. O. Christensen. *An Introduction to Frames and Riesz Bases.* Applied and Numerical Harmonic Analysis. Birkhäuser Boston Inc., Boston, MA, 2003.
14. L. M. Correia. *Wireless Flexible Personalized Communications.* John Wiley & Sons, Inc., New York, NY, USA, 2001.
15. V. de la Peña and E. Giné. *Decoupling. From Dependence to Independence.* Probability and its Applications (New York). Springer-Verlag, 1999.
16. D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
17. D. L. Donoho and J. Tanner. Counting faces of randomly-projected polytopes when the projection radically lowers dimension. *J. Amer. Math. Soc.*, 22(1):1–53, 2009.
18. M. Fornasier and H. Rauhut. Compressive sensing. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*, pages 187–228. Springer, 2011.
19. S. Foucart, A. Pajor, H. Rauhut, and T. Ullrich. The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$. *J. Complexity*, 26(6):629–640, 2010.
20. N. Grip and G. Pfander. A discrete model for the efficient analysis of time-varying narrowband communication channels. *Multidim. Syst. Signal Processing*, 19(1):3–40, 2008.
21. K. Gröchenig. *Foundations of Time-Frequency Analysis.* Applied and Numerical Harmonic Analysis. Birkhäuser, Boston, MA, 2001.
22. J. Haupt, W. Bajwa, G. Raz, and R. Nowak. Toeplitz compressed sensing matrices with applications to sparse channel estimation. *IEEE Trans. Inform. Theory*, 56(11):5862–5875, 2010.
23. M. Herman and T. Strohmer. High-resolution radar via compressed sensing. *IEEE Trans. Signal Process.*, 57(6):2275–2284, 2009.
24. F. Krahmer, G. E. Pfander, and P. Rashkov. Uncertainty in time-frequency representations on finite abelian groups and applications. *Appl. Comput. Harmon. Anal.*, 25(2):209–225, 2008.
25. J. Lawrence, G. Pfander, and D. Walnut. Linear independence of Gabor systems in finite dimensional vector spaces. *J. Fourier Anal. Appl.*, 11(6):715–726, 2005.
26. M. Ledoux and M. Talagrand. *Probability in Banach spaces.* Springer-Verlag, Berlin, Heidelberg, NewYork, 1991.

27. D. Middleton. Channel modeling and threshold signal processing in underwater acoustics: An analytical overview. *IEEE J. Oceanic Eng.*, 12(1):4–28, 1987.
28. M. Pätzold. *Mobile Fading Channels: Modelling, Analysis and Simulation.* John Wiley & Sons, Inc., 2001.
29. G. Pfander and H. Rauhut. Sparsity in time–frequency representations. *J. Fourier Anal. Appl.*, 16(2):233–260, 2010.
30. G. E. Pfander, H. Rauhut, and J. Tanner. Identification of matrices having a sparse representation. *IEEE Trans. Signal Process.*, 56(11):5376–5388, 2008.
31. H. Rauhut. Stability results for random sampling of sparse trigonometric polynomials. *IEEE Trans. Information Theory*, 54(12):5661–5670, 2008.
32. H. Rauhut. Circulant and Toeplitz matrices in compressed sensing. In *Proc. SPARS'09*, 2009.
33. H. Rauhut. Compressive Sensing and Structured Random Matrices. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1–92. deGruyter, 2010.
34. H. Rauhut and G. E. Pfander. Sparsity in time-frequency representations. *J. Fourier Anal. Appl.*, 16(2):233–260, 2010.
35. H. Rauhut, J. Romberg, and J. Tropp. Restricted isometries for partial random circulant matrices. *Appl. Comput. Harmonic Anal.*, to appear. DOI:10.1016/j.acha.2011.05.001.
36. H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *IEEE Trans. Inform. Theory*, 54(5):2210 – 2219, 2008.
37. H. Rauhut and R. Ward. Sparse Legendre expansions via l_1-minimization. *preprint*, 2010.
38. M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.
39. M. Stojanovic. Underwater acoustic communications. In J. G. Webster, editor, *Encyclopedia of Electrical and Electronics Engineering*, volume 22, pages 688–698. John Wiley & Sons, 1999.
40. T. Strohmer and R. W. j. Heath. Grassmannian frames with applications to coding and communication. *Appl. Comput. Harmon. Anal.*, 14(3):257–275, 2003.
41. M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.
42. M. Talagrand. *The Generic Chaining.* Springer Monographs in Mathematics. Springer-Verlag, 2005.