The restricted isometry property for time-frequency structured random matrices

Götz E. Pfander[*], Holger Rauhut[†], Joel A. Tropp[‡]

June 17, 2011

Dedicated to Hans Georg Feichtinger on the occasion of his 60th birthday.

Abstract

We establish the restricted isometry property for finite dimensional Gabor systems, that is, for families of time–frequency shifts of a randomly chosen window function. We show that the s-th order restricted isometry constant of the associated $n \times n^2$ Gabor synthesis matrix is small provided $s \leq c\, n^{2/3} / \log^2 n$. This improves on previous estimates that exhibit quadratic scaling of n in s. Our proof develops bounds for a corresponding chaos process.

Key Words: compressive sensing, restricted isometry property, Gabor system, time-frequency analysis, random matrix, chaos process.

AMS Subject classification: 60B20, 42C40, 94A12

1 Introduction and statements of results

Sparsity has become a key concept in applied mathematics and engineering. This is largely due to the empirical observation that a large number of real-world signals can be represented well by a sparse expansion in an appropriately chosen system of basic signals. Compressive sensing [9, 11, 13, 19, 21, 44] predicts that a small number of linear samples suffices to capture all the information in a sparse vector and that, furthermore, we can recover the sparse vector 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.

Linear data acquisition is described by a measurement matrix. The *restricted isometry property* (RIP) [12, 13, 21, 44] is by-now a standard tool for studying how efficiently the measurement matrix captures information about sparse signals. The RIP also streamlines the analysis of signal reconstruction algorithms, including ℓ_1-minization, greedy and iterative algorithms. Up to date there are no deterministic constructions of measurement matrices available that satisfy the RIP with the optimal scaling behavior; see, for example, the discussions in [44, Sec. 2.5] and [21, Sec. 5.1]. In contrast, a variety of random measurement matrices exhibit the RIP with optimal scaling, including Gaussian matrices and Rademacher matrices [3, 20, 47, 13].

Although Gaussian random matrices are optimal for sparse recovery [19, 25], they have limited use in practice because many applications impose structure on the matrix. Furthermore, recovery algorithms are significantly more efficient when the matrix admits a fast matrix–vector multiplication. For example,

[*] GEP is with School of Engineering and Science, Jacobs University Bremen, 28759 Bremen, Germany, (e-mail: g.pfander@jacobs-university.de).

[†] HR is with 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).

[‡] JAT is with California Institute of Technology, Pasadena, CA 91125 USA (e-mail: jtropp@cms.caltech.edu).

random sets of rows from a discrete Fourier transform matrix model the measurement process in MRI imaging and other applications. These random partial Fourier matrices lead to fast recovery algorithms because they can utilize the FFT. It is known that a random partial Fourier matrix satisfies a near-optimal RIP [13, 49, 42, 44] with high probability; see also [44, 48] for some generalizations.

This paper studies another type of structured random matrix that arises from time-frequency analysis, and has potential applications for the channel identification problem [41] in wireless communications and sonar [35, 50], as well as in radar [30]. The columns of the considered $n \times n^2$ matrix consist of all discrete time-frequency shifts of a random vector. Previous analysis of this matrix has provided bounds for the coherence [41], as well as nonuniform sparse recovery guarantees using ℓ_1-minimization [45]. However, the so far best available bounds on the restricted isometry constants were derived from coherence bounds [41] and, therefore, exhibit highly non-optimal quadratic scaling of n in the sparsity s. This paper dramatically improves on these bounds. Such an improvement is important because the nonuniform recovery guarantees in [45] apply only for ℓ_1-minimization, they do not provide stability of reconstruction, and they do not show the existence of a single time-frequency structured measurement matrix that is able to recover all sufficiently sparse vectors. Also it is of theoretical interest whether Gabor systems, that is, the columns of our measurement matrix, can possess the restricted isometry property. Nevertheless, our results still fall short of the optimal scaling that one might hope for.

Our approach is similar to the recent restricted isometry analysis for partial random circulant matrices in [46]. Indeed, also here we bound a chaos process of order 2, by means of a Dudley type inequality for such processes due to Talagrand [53]. This requires to estimate covering numbers of the set of unit norm s-sparse vectors with respect to two different metrics induced by the process. In contrast to [46], the specific structure of our problem does not allow us to reduce to the Fourier case, and to apply covering number estimates shown in [49].

This paper is organized as follows. In Section 1.1 we recall central concepts in compressive sensing. Section 1.2 introduces the time-frequency structured measurement matrices that are considered in this paper, and we state our main result, Theorem 1. Remarks on applications in wireless communications and radar, as well as the relation of this paper to previous work are given in Sections 1.4 and 1.3, respectively. Sections 2, 3 and 4 provide the proof of Theorem 1.

1.1 Compressive Sensing

In general, reconstructing $\boldsymbol{x} = (x_1, \ldots, x_N)^T \in \mathbb{C}^N$ from

$$\boldsymbol{y} \;=\; \boldsymbol{A}\boldsymbol{x} \in \mathbb{C}^n, \tag{1}$$

where $\boldsymbol{A} \in \mathbb{C}^{n \times N}$ and $n \ll N$ (in this paper, we have $N = n^2$) is impossible without substantial *a-priori* information on $\boldsymbol{x}$. In compressive sensing the assumption that $\boldsymbol{x}$ is s-sparse, that is, $\|\boldsymbol{x}\|_0 := \#\{\ell : x_\ell \neq 0\} \leq s$ for some $s \ll N$ is introduced to ensure uniqueness and efficient recoverability of $\boldsymbol{x}$. More generally, under the assumption that $\boldsymbol{x}$ is well-approximated by a sparse vector, the question is posed whether an optimally sparse approximation to $\boldsymbol{x}$ can be found efficiently.

Reconstruction of a sparse vector $\boldsymbol{x}$ by means of the ℓ_0-minimization problem,

$$\min_{\boldsymbol{z}} \|\boldsymbol{z}\|_0 \quad \text{subject to} \quad \boldsymbol{y} \;=\; \boldsymbol{A}\boldsymbol{z},$$

is NP-hard [36] and therefore not tractable. Consequently, a number of alternatives to ℓ_0-minimization, for example, greedy algorithms [5, 23, 37, 54, 55], have been proposed in the literature. The most popular approach utilizes ℓ_1-minimization [11, 15, 19], that is, the convex program

$$\min_{\boldsymbol{z}} \|\boldsymbol{z}\|_1 \quad \text{subject to } \boldsymbol{y} \;=\; \boldsymbol{A}\boldsymbol{z}, \tag{2}$$

is solved, where $\|\boldsymbol{z}\|_1 = |z_1| + |z_2| + \ldots + |z_N|$ denotes the usual ℓ_1 vector norm.

To guarantee recoverability of the sparse vector $\boldsymbol{x}$ in (1) by means of ℓ_1-minimization and greedy algorithms, it suffices to establish the restricted isometry property (RIP) of the so-called measurement matrix $\boldsymbol{A}$: define the restricted isometry constant δ_s of an $n \times N$ matrix $\boldsymbol{A}$ to be the smallest positive number that satisfies

$$(1-\delta_s)\|\boldsymbol{x}\|_2^2 \;\leq\; \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \;\leq\; (1+\delta_s)\|\boldsymbol{x}\|_2^2 \quad \text{for all } \boldsymbol{x} \text{ with } \|\boldsymbol{x}\|_0 \leq s. \tag{3}$$

In words, the statement (3) requires that all column submatrices of $\boldsymbol{A}$ with at most s columns are well-conditioned. Informally, $\boldsymbol{A}$ is said to satisfy the RIP with order s when δ_s is “small”.

Now, if the matrix $\boldsymbol{A}$ obeys (3) with

$$\delta_{\kappa s} \;<\; \delta^* \tag{4}$$

for suitable constants $\kappa \geq 1$ and $\delta^* < 1$, then many algorithms precisely recover any s-sparse vectors $\boldsymbol{x}$ from the measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. Moreover, if $\boldsymbol{x}$ can be well approximated by an s sparse vector, then for noisy observations

$$\boldsymbol{y} \;=\; \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e} \quad \text{where} \quad \|\boldsymbol{e}\|_2 \leq \tau,$$

these algorithms return a reconstruction $\widetilde{\boldsymbol{x}}$ that satisfies an error bound of the form

$$\|\boldsymbol{x} - \widetilde{\boldsymbol{x}}\|_2 \;\leq\; C_1 \frac{\sigma_s(\boldsymbol{x})_1}{\sqrt{s}} + C_2\tau, \tag{5}$$

where $\sigma_s(\boldsymbol{x})_1 = \inf_{\|\boldsymbol{z}\|_0 \leq s} \|\boldsymbol{x} - \boldsymbol{z}\|_1$ denotes the error of best s-term approximation in ℓ_1 and C_1, C_2 are positive constants. For illustration, we include Table 1 which lists available values for the constants κ and δ^* in (4) that guarantee (5) for several algorithms along with respective references.

Algorithm	κ	δ^*	References
ℓ_1-minimization (2)	2	$\frac{3}{4+\sqrt{6}} \approx 0.4652$	[8, 10, 12, 22]
CoSaMP	4	$\sqrt{\frac{2}{5+\sqrt{73}}} \approx 0.3843$	[24, 54]
Iterative Hard Thresholding	3	$1/2$	[5, 22]
Hard Thresholding Pursuit	3	$1/\sqrt{3} \approx 0.5774$	[23]

Table 1: Values of the constants κ and δ^* in (4) that guarantee success for various recovery algorithms.

For example, Gaussian random matrices, that is, matrices that have independent, normally distributed entries with mean zero and variance one, have been shown [3, 13, 34] to have restricted isometry constants of $\frac{1}{\sqrt{n}}\boldsymbol{A}$ satisfy $\delta_s \leq \delta$ with high probability provided that

$$n \;\geq\; C\delta^{-2} s \log(N/s).$$

That is, the number n of Gaussian measurements required to reconstruct an s-sparse signal of length N is *linear* in the sparsity and *logarithmic* in the ambient dimension. See [3, 13, 34, 21, 44] for precise statements and extensions to Bernoulli and subgaussian matrices. It follows from lower estimates of Gelfand widths that this bound on the required samples is optimal [17, 25, 26], that is, the log-factor must be present.

As discussed above, no deterministic construction of a measurement matrix is known which provides RIP with optimal scaling of the recoverable sparsity s in the number of measurements n. In fact, all available proofs of the RIP with close to optimal scaling require the measurement matrix to contain some randomness. In Table 2 we list the Shannon entropy (in bits) of various random matrices along with the available RIP estimates. Compared to Gaussian random matrices, the Gabor synthesis measurement matrices constructed in this paper introduces only a small amount of randomness, that is, the presented measurement matrix depends only on the so-called Gabor window, a random vector of length n, which can be chosen to be a normalized copy of a Rademacher vector. Moreover, the random Gabor matrix provably provides scaling of s roughly in $n^{2/3}$, which significantly improves on known deterministic constructions. Clearly, such scaling falls short of the optimal one, but we expect that it is possible to establish linear scaling of s in n up to log-factors, similar to Gaussian matrices or partial random Fourier matrices. However, such improvement seems to require more powerful methods to estimate chaos processes than presently available.

$n \times N$ Measurement matrix	Shannon entropy	RIP regime	References
Gaussian	$nN\,\frac{1}{2}\log(2\pi e)$	$s \leq Cn/\log N$	[3, 20, 49]
Rademacher entries	nN	$s \leq Cn/\log N$	[3]
Partial Fourier matrix	$N\log_2 N - n\log_2 n$ $-(N-n)\log_2(N-n)$	$s \leq Cn/\log^4 N$	[46, 49]
Partial circulant Rademacher	N	$s \leq Cn^{2/3}/\log^{2/3} N$	[46]
Gabor, Rademacher window	n	$s \leq Cn^{2/3}/\log^2 n$	this paper
Gabor, Alltop window	0	$s \leq C\sqrt{n}$	[41]

Table 2: List of measurement matrices that have been proven to be RIP, scaling of sparsity s in the number of measurements n, and the respective Shannon entropy of the (random) matrix.

1.2 Time-frequency structured measurement matrices

In this paper, we provide probabilistic estimates of the restricted isometry constants for matrices whose columns are time–frequency shifts of a randomly chosen vector. To define these matrices, we let $\boldsymbol{T}$ denote the cyclic shift, also called translation operator, and $\boldsymbol{M}$ the modulation operator, or frequency shift operator, on $\mathbb{C}^n$. They are defined by

$$(\boldsymbol{T}\boldsymbol{h})_q = h_{q\ominus 1} \quad \text{and} \quad (\boldsymbol{M}\boldsymbol{h})_q = e^{2\pi i q/n} h_q = \omega^q h_q, \tag{6}$$

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

$$(\boldsymbol{T}^k\boldsymbol{h})_q = h_{q\ominus k} \quad \text{and} \quad (\boldsymbol{M}^\ell\boldsymbol{h})_q = e^{2\pi i\ell q/n} h_q = \omega^{\ell q} h_q. \tag{7}$$

The operators $\boldsymbol{\pi}(\lambda) = \boldsymbol{M}^\ell\boldsymbol{T}^k$, $\lambda = (k,\ell)$, are called time-frequency shifts and the system $\{\boldsymbol{\pi}(\lambda) : \lambda \in \mathbb{Z}_n{\times}\mathbb{Z}_n\}$, $\mathbb{Z}_n = \{0,1,\ldots,n-1\}$, of all time-frequency shifts forms a basis of the matrix space $\mathbb{C}^{n\times n}$ [32, 31].

We choose $\boldsymbol{\epsilon} \in \mathbb{C}^n$ to be a Rademacher or Steinhaus sequence, that is, a vector of independent random variables taking the values $+1$ and -1 with equal probability, respectively taking values uniformly distributed on the complex torus $S^1 = \{z \in \mathbb{C}, |z| = 1\}$. The normalized window is

$$\boldsymbol{g} = n^{-1/2}\boldsymbol{\epsilon},$$

and the set

$$\{\boldsymbol{\pi}(\lambda)\boldsymbol{g} : \lambda \in \mathbb{Z}_n{\times}\mathbb{Z}_n\} \tag{8}$$

is called a full Gabor system with window $\boldsymbol{g}$ [28]. The matrix $\boldsymbol{\Psi}_{\boldsymbol{g}} \in \mathbb{C}^{n\times n^2}$ whose columns list the members $\boldsymbol{\pi}(\lambda)\boldsymbol{g}$, $\lambda \in \mathbb{Z}_n{\times}\mathbb{Z}_n$, of the Gabor system is referred to as Gabor synthesis matrix [16, 32, 40]. Note that $\boldsymbol{\Psi}_{\boldsymbol{g}}$ allows for fast matrix vector multiplication algorithms based on the FFT. The main result of this paper addresses the restricted isometry constants of $\boldsymbol{\Psi}_{\boldsymbol{g}}$. Below $\mathbb{E}$ denotes expectation and $\mathbb{P}$ the probability of an event.

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

(a) The expectation of the restricted isometry constant δ_s of $\boldsymbol{\Psi}_{\boldsymbol{g}}$, $s \leq n$, satisfies

$$\mathbb{E}\,\delta_s \leq \max\Big\{C_1\sqrt{\frac{s^{3/2}}{n}}\log s\sqrt{\log n},\ C_2\frac{s^{3/2}\log^{3/2} n}{n}\Big\}, \tag{9}$$

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

(b) For $0 \leq \lambda \leq 1$, we have

$$\mathbb{P}(\delta_s \geq \mathbb{E}[\delta_s] + \lambda) \leq e^{-\lambda^2/\sigma^2}, \qquad \text{where } \sigma^2 = \frac{C_3 s^{\frac{3}{2}} \log n \log^2 s}{n} \tag{10}$$

with $C_3 > 0$ being a universal constant.

With slight variations of the proof one can show similar statements for normalized Gaussian or subgaussian random windows $\boldsymbol{g}$.

Roughly speaking $\boldsymbol{\Psi_g}$ satisfies the RIP of order s with high probability if $n \geq Cs^{3/2} \log^3(n)$, or equivalently if,

$$s \leq cn^{2/3}/\log^2 n.$$

We expect that this is not the optimal estimate, but improving on this seems to require more sophisticated techniques than pursued in this paper. There are known examples [33, 53] for which the central tool in this paper, the Dudley type inequality for chaos processes stated in Theorem 3, is not sharp. We may well be facing one of these cases here.

Numerical tests illustrating the use of $\boldsymbol{\Psi_g}$ for compressive sensing are presented in [41]. They illustrate that empirically $\boldsymbol{\Psi_g}$ performs very similarly to a Gaussian matrix.

1.3 Application in wireless communications and radar

An important task in wireless communications is to identify the communication channel at hand, that is, the channel opperator, by probing it with a small number of known transmit signals; ideally a single probing signal. A common finite-dimensional model for the channel operator, that combines digital (discrete) to analog conversion, the analog channel, and analog to digital conversion. It is given by [4, 18, 27, 38]

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

Time-shifts model delay due to multipath-propagation, while frequency-shifts model the Doppler effect due to moving transmitter, receiver, and/or scatterers. Physical considerations often suggest that $\boldsymbol{x}$ is rather sparse as, indeed, the number of present scatterers can be assumed to be small in most cases. The same model is used as well in sonar [35, 50] and radar [30].

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

$$\boldsymbol{y} = \boldsymbol{\Gamma g} \; = \sum_{\lambda \in \mathbb{Z}_n \times \mathbb{Z}_n} x_\lambda \boldsymbol{\pi}(\lambda) \boldsymbol{g} = \boldsymbol{\Psi_g x} \tag{11}$$

with unknown but sparse $\boldsymbol{x}$, we arrive at a compressive sensing problem. In this setup, we clearly have the freedom to choose $\boldsymbol{g}$, and we may choose it as a random Rademacher or Steinhaus sequence. Then the restricted isometry property of $\boldsymbol{\Psi_g}$, as shown in Theorem 1, ensures recovery of sufficiently sparse $\boldsymbol{x}$, and hence, of the associated operator $\boldsymbol{\Gamma}$.

Recovery of the sparse $\boldsymbol{x}$ in (11) can also be interpreted as finding a sparse time-frequency representation of a given $\boldsymbol{y}$ with respect to the window $\boldsymbol{g}$. From an application point of view though, the vectors considered here are not well suited to describe meaningful sparse time-frequency representations of $\boldsymbol{x}$ as all $\boldsymbol{g}$ that are known to guarantee RIP of $\boldsymbol{\Psi_g}$ are very poorly localized both in time and in frequency.

1.4 Relation with previous work

Time-frequency structured matrices $\boldsymbol{\Psi_g}$ appeared in the study of frames with (near-)optimal coherence. 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

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

Choosing the Alltop window [1, 51] $\boldsymbol{g} \in \mathbb{C}^n$ with entries $g_\ell = n^{-1/2} e^{2\pi i \ell^3/n}$ for $n \geq 5$ prime yields $\boldsymbol{\Psi_g}$ with coherence

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

Due to the general lower bound $\mu \geq \sqrt{\frac{N-n}{n(N-1)}}$ for an $n \times N$ matrix [51], this coherence is almost optimal. Together with the bound $\delta_s \leq (s-1)\mu$ we obtain

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

This requires a scaling $s \leq c\sqrt{n}$ to achieve sufficiently small RIP and sparse recovery, which clearly is worse than the main result of this paper.

The coherence of $\boldsymbol{\Psi_g}$ with Steinhaus sequence $\boldsymbol{g}$ is estimated in [41] by

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

holding with probability at least $1-\varepsilon$. As before, this does not give better than quadratic scaling of n in s in order to have small RIP constants δ_s.

The following nonuniform recovery results for ℓ_1-minimization with $\boldsymbol{\Psi_g}$ and Steinhaus sequence $\boldsymbol{g}$ was derived in [45].

theorem 2 *Let $\boldsymbol{x} \in \mathbb{C}^n$ be s-sparse. Choose a Steinhaus sequence $\boldsymbol{g}$ at random. Then with probability at least $1-\varepsilon$, the vector $\boldsymbol{x}$ can be recovered from $\boldsymbol{y} = \boldsymbol{\Psi_g x}$ via ℓ_1-minimization provided*

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

Clearly, the (optimal) almost linear scaling of n in s of this estimate is better than the RIP estimate of the main Theorem 1. However, the conclusion is weaker than what can be derived using the restricted isometry property: recovery in Theorem 2 is nonuniform in the sense that a given s-sparse vector can be recovered with high probability from a random draw of the matrix $\boldsymbol{\Psi_g}$. It is not stated that a single matrix $\boldsymbol{\Psi_g}$ can recover all s-sparse vectors simultaneously. Moreover, nothing is said about the stability of recovery, while in contrast, small RIP constants imply (5). Therefore, our main Theorem 1 is of high interest and importance, despite the better scaling in Theorem 2. Moreover, we expect that an improvement of the RIP estimate is possible, although it is presently not clear how this can be achieved.

Partial random circulant matrices are a different, but closely related measurement matrix, studied in [29, 43, 44, 46]. They model convolution with a random vector followed by subsampling on an arbitrary (deterministic) set. The so far best estimate of the restricted isometry constants δ_s of such an $n \times N$ matrix in [46] requires $n \geq c(s \log N)^{3/2}$, similarly to the main result of this paper. The corresponding analysis requires to bound as well a chaos process, which is also achieved by the Dudley type bound of Theorem 3 below. Nonuniform recovery guarantees for partial random circulant matrices similarly to Theorem 2 are contained in [43, 44]. The analysis of circulant matrices benefits from a simplified arithmetic in the Fourier domain, a tool not available to us in the case of Gabor synthesis matrices. Hence, the analysis presented here is 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

$$\begin{aligned} \boldsymbol{A}_0 &= \left(\begin{array}{ccccc|ccccc|cc} 1 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & \omega & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 & \omega^2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & \omega^{n-1} & \cdots & \omega^{(n-1)^2} \end{array}\right) \\ &= \big(\boldsymbol{I}\big|\boldsymbol{M}\big|\boldsymbol{M}^2\big|\cdots\big|\boldsymbol{M}^{n-1}\big), \\ \boldsymbol{A}_1 &= \left(\begin{array}{ccccc|ccccc|cc} 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 1 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 & \omega & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & \omega^2 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \end{array}\right) \\ &= \big(\boldsymbol{T}\big|\boldsymbol{MT}\big|\boldsymbol{M}^2\boldsymbol{T}\big|\cdots\big|\boldsymbol{M}^{n-1}\boldsymbol{T}\big), \end{aligned}$$

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

$$\boldsymbol{A}_q = \big(\boldsymbol{T}^q\big|\boldsymbol{M}\boldsymbol{T}^q\big|\boldsymbol{M}^2\boldsymbol{T}^q\big|\cdots\big|\boldsymbol{M}^{n-1}\boldsymbol{T}^q\big). \tag{12}$$

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 (29) below, it follows that

$$\boldsymbol{H} = \frac{1}{n}\sum_{q' \neq q} \overline{\epsilon_{q'}} \, \epsilon_q \, \boldsymbol{A}_{q'}^* \boldsymbol{A}_q = \frac{1}{n}\sum_{q',q} \overline{\epsilon_{q'}} \, \epsilon_q \, \boldsymbol{W}_{q',q}, \tag{13}$$

where, for notational simplicity, we use here and in the following $\boldsymbol{W}_{q',q} = \boldsymbol{A}_{q'}^*\boldsymbol{A}_q$ for $q \neq q'$ and $\boldsymbol{W}_{q',q} = 0$ for $q = q'$. We employ the matrix $\boldsymbol{B}(\boldsymbol{x}) \in \mathbb{C}^{n\times n}$, $\boldsymbol{x} \in T_s$, given by matrix entries

$$B(\boldsymbol{x})_{q',q} = \boldsymbol{x}^*\boldsymbol{W}_{q',q}\boldsymbol{x}. \tag{14}$$

Then we have

$$n\,\mathbb{E}\delta_s = \mathbb{E}\sup_{\boldsymbol{x}\in T_s} |Y_{\boldsymbol{x}}| = \mathbb{E}\sup_{\boldsymbol{x}\in T_s} |Y_{\boldsymbol{x}} - Y_{\boldsymbol{0}}| \, , \tag{15}$$

where

$$Y_{\boldsymbol{x}} \; = \boldsymbol{\epsilon}^*\boldsymbol{B}(\boldsymbol{x})\boldsymbol{\epsilon} = \sum_{q' \neq q} \overline{\epsilon_{q'}} \, \epsilon_q \, \boldsymbol{x}^*\boldsymbol{A}_{q'}^*\boldsymbol{A}_q\boldsymbol{x} \tag{16}$$

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 (16) 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, [33, Theorem 11.22] or [53, 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 $\|\boldsymbol{A}\|_F = \operatorname{Tr}(\boldsymbol{A}^*\boldsymbol{A})^{1/2} = (\sum_{j,k} |A_{j,k}|^2)^{1/2}$, where $\operatorname{Tr}(\boldsymbol{A})$ denotes the trace of a matrix $\boldsymbol{A}$.

theorem 3 *Let $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$ 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,

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

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 $\boldsymbol{x_0} \in T$,

$$\mathbb{E} \sup_{\boldsymbol{x} \in T} |Y_{\boldsymbol{x}} - Y_{\boldsymbol{x_0}}| \; \leq \; K \max \left\{ \int_0^\infty \log N(T, d_1, u) \, du \int_0^\infty \sqrt{\log N(T, d_2, u)} \, du, \; \right\}. \tag{17}$$

Proof: For a Rademacher sequence, the theorem is stated in [46, Proposition 2.2]. If $\boldsymbol{\epsilon}$ is a Steinhaus sequence and $\boldsymbol{B}$ a Hermitian matrix then

$$\begin{aligned} \boldsymbol{\epsilon}^* \boldsymbol{B} \boldsymbol{\epsilon} = \operatorname{Re}(\boldsymbol{\epsilon}^* \boldsymbol{B} \boldsymbol{\epsilon}) &= \operatorname{Re}(\boldsymbol{\epsilon})^* \operatorname{Re}(\boldsymbol{B}) \operatorname{Re}(\boldsymbol{\epsilon}) - \operatorname{Re}(\boldsymbol{\epsilon})^* \operatorname{Im}(\boldsymbol{B}) \operatorname{Im}(\boldsymbol{\epsilon}) \\ &\quad + \operatorname{Im}(\boldsymbol{\epsilon})^* \operatorname{Im}(\boldsymbol{B}) \operatorname{Re}(\boldsymbol{\epsilon}) + \operatorname{Im}(\boldsymbol{\epsilon})^* \operatorname{Re}(\boldsymbol{B}) \operatorname{Im}(\boldsymbol{\epsilon}). \end{aligned}$$

By decoupling, see, for example, [39, Theorem 3.1.1], we have with ϵ' denoting an independent copy of ϵ,

$$\begin{aligned} \mathbb{E} \sup_{x \in T} |\operatorname{Re}(\boldsymbol{\epsilon})^* \operatorname{Im}(\boldsymbol{B}(\boldsymbol{x})) \operatorname{Im}(\boldsymbol{\epsilon})| &\leq 8 \, \mathbb{E} \sup_{x \in T} |\operatorname{Re}(\boldsymbol{\epsilon})^* \operatorname{Im}(\boldsymbol{B}(\boldsymbol{x})) \operatorname{Im}(\boldsymbol{\epsilon'})| \\ \leq 8 \, \mathbb{E} \sup_{x \in T} |\boldsymbol{\xi}^* \operatorname{Im}(\boldsymbol{B}(\boldsymbol{x})) \operatorname{Im}(\boldsymbol{\epsilon}')| &\leq 8 \, \mathbb{E} \sup_{x \in T} |\boldsymbol{\xi}^* \operatorname{Im}(\boldsymbol{B}(\boldsymbol{x})) \boldsymbol{\xi}'|, \end{aligned}$$

where $\boldsymbol{\xi}, \boldsymbol{\xi}'$ denote independent Rademacher sequences. The second and third inequalities follow from the contraction principle [33, Theorem 4.4] (and symmetry of $\operatorname{Re}(\epsilon_\ell), \operatorname{Im}(\epsilon_\ell)$) first applied conditionally on $\boldsymbol{\epsilon}'$ and then conditionally on $\boldsymbol{\xi}$ (note that $|\operatorname{Re}(\epsilon_\ell)| \leq 1$, $|\operatorname{Im}(\epsilon_\ell)| \leq 1$ for all realizations of ϵ_ℓ). Using the triangle inequality we get

$$\begin{aligned} \mathbb{E} \sup_{x \in T} |Y_{\boldsymbol{x}} - Y_{\boldsymbol{x_0}}| &\leq 16 \, \mathbb{E} \sup_{x \in T} |\boldsymbol{\xi}^* (\operatorname{Re}(\boldsymbol{B}(\boldsymbol{x})) - \operatorname{Re}(\boldsymbol{B}(x_0)) \xi'| \\ &\quad + 16 \, \mathbb{E} \sup_{x \in T} |\boldsymbol{\xi}^* (\operatorname{Im}(\boldsymbol{B}(\boldsymbol{x})) - \operatorname{Im}(\boldsymbol{B}(x_0))) \boldsymbol{\xi}'|. \end{aligned} \tag{18}$$

Further note that $\|\operatorname{Im}(\boldsymbol{B}(\boldsymbol{x})) - \operatorname{Im}(\boldsymbol{B}(\boldsymbol{y}))\|_F$, $\|\operatorname{Re}(\boldsymbol{B}(\boldsymbol{x})) - \operatorname{Re}(\boldsymbol{B}(\boldsymbol{y}))\|_F \leq \|\boldsymbol{B}(\boldsymbol{x}) - \boldsymbol{B}(\boldsymbol{y})\|_F$ 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 $\|\operatorname{Im}(\boldsymbol{B}(\boldsymbol{x})) - \operatorname{Im}(\boldsymbol{B}(\boldsymbol{y}))\|_{2\to 2}$, $\|\operatorname{Re}(\boldsymbol{B}(\boldsymbol{x})) - \operatorname{Re}(\boldsymbol{B}(\boldsymbol{y}))\|_{2\to 2} \leq \|\boldsymbol{B}(\boldsymbol{x}) - \boldsymbol{B}(\boldsymbol{y})\|_{2\to 2}$. 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 (14) satisfies the hypotheses of Theorem 3 by definition. The pseudometrics are given by

$$d_2(\boldsymbol{x},\boldsymbol{y}) \;=\; \|\boldsymbol{B}(\boldsymbol{x})-\boldsymbol{B}(\boldsymbol{y})\|_F = \Big(\sum_{q'\neq q}\left|\boldsymbol{x}^*\boldsymbol{A}_{q'}^*\boldsymbol{A}_q\boldsymbol{x}-\boldsymbol{y}^*\boldsymbol{A}_{q'}^*\boldsymbol{A}_q\boldsymbol{y}\right|^2\Big)^{1/2}, \tag{19}$$

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 4 *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 5 *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*

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

where $c > 0$ is universal constant.

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

Lemma 6 *The diameter of T_s with respect to d_1 is bounded by $4s$, and for $u > 0$*

$$\begin{aligned}\log(N(T_s,d_1,u)) \le \min\big\{&s\log(en^2/s)+s\log(1+4su^{-1}),\\ &cu^{-2}s^2\log(2n)\log(n^2/u)\big\}\,,\end{aligned} \tag{20}$$

where $c > 0$ is a universal constant.

Based on these estimates and Theorem 3 we complete the proof of Theorem 1(a). By Lemmas 4 and 5, the subgaussian integral in (17) can be estimated as

$$\begin{aligned}
&\int_0^\infty \sqrt{\log(N(T_s,d_2,u))}du = \int_0^{4\sqrt{sn}} \sqrt{\log(N(T_s,d_2,u))}du\\
&= \int_0^{\sqrt{n}} \sqrt{\log(N(T_s,d_2,u))}du + \int_{\sqrt{n}}^{\sqrt{sn}} \sqrt{\log(N(T_s,d_2,u))}du\\
&\le \int_0^{\sqrt{n}} \sqrt{s\log(en^2/s)}du + \int_0^{\sqrt{n}} \sqrt{s\log(1+4\sqrt{sn}u^{-1})}du\\
&+ c\sqrt{ns^{3/2}}\int_{\sqrt{n}}^{4\sqrt{sn}} u^{-1}\sqrt{\log(ns^{5/2}u^{-1})}du\\
&\le \sqrt{sn\log(en^2/s)} + 4s\sqrt{n}\int_0^{s^{-1/2}} \sqrt{\log(1+u^{-1})}du\\
&+ c\sqrt{s^{3/2}n}\sqrt{\log(n^{1/2}s^{5/2})}\log(\sqrt{s})\\
&\le \sqrt{sn\log(en^2/s)} + 4\sqrt{sn}\sqrt{\log(e(1+\sqrt{s}))} + c'\sqrt{s^{3/2}n\log(n)\log^2(s)}\\
&\le \hat{C}_1\sqrt{s^{3/2}n\log(n)\log^2(s)}.
\end{aligned} \tag{21}$$

Hereby, we have used [44, Lemma 10.3], and that $s \leq n$. Due to Lemma 6 the subexponential integral obeys the estimate, for some $\kappa > 0$ to be chosen below,

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

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

$$\int_0^\infty \log(N(T_s, d_1, u))du \leq \hat{C}_2 s^{3/2} \log^{3/2}(n). \tag{22}$$

Combining the above integral estimates with (15) and Theorem 3 yields

$$\mathbb{E}\delta_s = \frac{1}{n}\mathbb{E} \sup_{x \in T_s} |Y_{\boldsymbol{x}} - Y_0| \leq \frac{1}{n} \max\left\{C_1\sqrt{s^{3/2} n \log(n) \log^2(s)}, C_2 s^{3/2} \log^{3/2}(n)\right\}. \tag{23}$$

This is the statement of Theorem 1(a).

Remark 7 In analogy to the estimate of a subgaussian entropy integral arising in the analysis of partial random circulant matrices in [46], we expect that the exponent 3/2 in (21) can be improved to 1. However, we doubt that for the subexponential integral (22) such improvement will be possible (indeed, the estimate of the subexponential integral in [46] also exhibits an exponent of 3/2 at the s-term), so that we did not pursue an improvement of (21) here as this would not provide a significant overall improvement of (23). We expect that an improvement of (23) 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 4, 5 and 6, 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

$$\|\boldsymbol{A}\|_{S_q} := \|\boldsymbol{\sigma}(\boldsymbol{A})\|_q, \tag{24}$$

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

$$\|\boldsymbol{A}\|_{S_{2p}} = (\operatorname{Tr}((\boldsymbol{A}^*\boldsymbol{A})^p))^{1/(2p)}. \tag{25}$$

The S_∞-norm coincides with the operator norm, $\|\cdot\|_{S_\infty} = \|\cdot\|_{2\to2}$. By the corresponding properties of ℓ_q-norms we have the inequalities

$$\|\boldsymbol{A}\|_{2\to2} \leq \|\boldsymbol{A}\|_{S_q} \leq \operatorname{rank}(\boldsymbol{A})^{1/q}\|\boldsymbol{A}\|_{2\to2}. \tag{26}$$

Moreover, we will require an extension of the quadratic form $\boldsymbol{B}(\boldsymbol{x})$ in (14) to a bilinear form,

$$(\boldsymbol{B}(\boldsymbol{x}, \boldsymbol{z}))_{q',q} = \begin{cases} \boldsymbol{x}^*\boldsymbol{A}_{q'}^*\boldsymbol{A}_q\boldsymbol{z} & \text{if } q' \neq q, \\ 0 & \text{if } q' = q. \end{cases} \tag{27}$$

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 (12) 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 span $\{\boldsymbol{e}_\lambda\}$. The following bounds will be crucial for the covering number estimates below.

Lemma 8 *Let $\boldsymbol{A}_q$ be as given in (12). Then, for $\lambda \in \mathbb{Z}_n\times\mathbb{Z}_n$, $q \in \mathbb{Z}_n$,*

$$\boldsymbol{A}_q\boldsymbol{e}_\lambda = \boldsymbol{\pi}(\lambda)\boldsymbol{e}_q\,, \tag{28}$$

$$\sum_{q=0}^{n-1}\boldsymbol{A}_q^*\boldsymbol{A}_q = n\,\boldsymbol{I}\,, \tag{29}$$

$$\sum_{q=0}^{n-1}\boldsymbol{A}_q\boldsymbol{P}_\lambda\boldsymbol{A}_q^* = \boldsymbol{I}\ , \tag{30}$$

$$\sum_{q=0}^{n-1}\sum_{q'=0}^{n-1}\left|\boldsymbol{x}^*\boldsymbol{A}_{q'}^*\boldsymbol{A}_q\boldsymbol{y}\right|^2 \le n\,\|\boldsymbol{x}\|_0\,\|\boldsymbol{x}\|_2^2\,\|\boldsymbol{y}\|_2^2. \tag{31}$$

Proof: For (28), 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 (29), choose $\boldsymbol{z} \in \mathbb{C}^{n\times n}$ and compute

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

Hence,

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

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

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

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

3.2 Proof of Lemma 4

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 (31) 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,

$$d_2(\boldsymbol{x},\boldsymbol{y})\ \le\ 2\sqrt{sn}\,\|\boldsymbol{x}-\boldsymbol{y}\|_2. \tag{32}$$

Using the volumetric argument, see, for example, [44, Proposition 10.1], we obtain

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

By a rescaling argument

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

Taking the logarithm completes the proof.

3.3 Proof of Lemma 5

Now, we seek a suitable estimate of the covering numbers $N(T_s, d_1, u)$ for $u \geq \sqrt{n}$. Observe that by (32) 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

$$\sqrt{n} \leq u \leq 4\sqrt{sn}, \tag{33}$$

as stated in the lemma. We use the empirical method [14], similarly as in [49]. We define the norm $\|\cdot\|_*$ on $\mathbb{C}^{n\times n}$ by

$$\|\boldsymbol{x}\|_* = \sum_\lambda |\operatorname{Re} x_\lambda| + |\operatorname{Im} x_\lambda|\,. \tag{34}$$

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

$$\boldsymbol{B} := \boldsymbol{B}(\boldsymbol{y}, \boldsymbol{y}') = \frac{1}{m^2}\sum_{j,j'=1}^m \boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}_{j'}')\,. \tag{35}$$

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 (30) in Lemma 8

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

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

$$
\begin{aligned}
\mathbb{E}\|\boldsymbol{B} - \boldsymbol{B}(\boldsymbol{x})\|_F^2 &\leq \sum_{q',q} -|\boldsymbol{x}^* \boldsymbol{W}_{q',q} \boldsymbol{x}|^2 + \frac{m^2(m^2-1)}{m^4} \sum_{q',q} |\boldsymbol{x}^* \boldsymbol{W}_{q,q'} \boldsymbol{x}|^2 \\
&\quad + \frac{m^2 n 4 s^2}{m^4} + \frac{4m^2(m-1)ns}{m^4} \\
&\leq \frac{4ns^2}{m^2} + \frac{4ns}{m} \leq \frac{4ns^2}{121 n^2 s^3} u^4 + \frac{4ns}{11ns^{\frac{3}{2}}} u^2 \leq \frac{64ns}{121ns} u^2 + \frac{44}{121\sqrt{s}} u^2 \leq u^2 \, . \qquad (36)
\end{aligned}
$$

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 $\lambda' = (k', \ell')$,

$$
\begin{aligned}
\boldsymbol{B}(\boldsymbol{e}_{\lambda'}, \boldsymbol{e}_\lambda)_{q',q} &= (\boldsymbol{A}_{q'} e_{\lambda'})^* \boldsymbol{A}_q \boldsymbol{e}_\lambda = \langle \boldsymbol{\pi}(\lambda) \boldsymbol{e}_q, \boldsymbol{\pi}(\lambda') \boldsymbol{e}_{q'} \rangle \\
&= \begin{cases} \omega^{(\ell - \ell')(k+q)}, & \text{if } k' + q' = k + q\,; \\ 0, & \text{else,} \end{cases} \qquad (37)
\end{aligned}
$$

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

$$\begin{aligned}
&\|\boldsymbol{B}_\alpha - \boldsymbol{B}_{\|\boldsymbol{x}\|_*}\|_F \\
&= \Big\| \frac{1}{m^2} \sum_{j=1,j'=1}^{m} \boldsymbol{B}(\alpha \operatorname{sgn}(x_{\lambda_j})\boldsymbol{e}_{\lambda_j}, \alpha \operatorname{sgn}(x_{\lambda'_{j'}})\boldsymbol{e}_{\lambda'_{j'}}) \\
&\qquad - \frac{1}{m^2} \sum_{j,j'=1}^{m} \boldsymbol{B}(\|\boldsymbol{x}\|_* \operatorname{sgn}(x_{\lambda_j})\boldsymbol{e}_{\lambda_j}, \|\boldsymbol{x}\|_* \operatorname{sgn}(x_{\lambda'_{j'}})\boldsymbol{e}_{\lambda'_{j'}}) \Big\|_F \\
&= |\|\boldsymbol{x}\|_*^2 - \alpha^2| \|\frac{1}{m^2} \sum_{j,j'=1}^{m} \boldsymbol{B}(\operatorname{sgn}(x_{\lambda_j})\boldsymbol{e}_{\lambda_j}, \operatorname{sgn}(x_{\lambda'_{j'}})\boldsymbol{e}_{k'_{j'}})\|_F \\
&\leq \frac{u}{m^2\sqrt{n}} \sum_{j,j'=1}^{m} \|\boldsymbol{B}(\boldsymbol{e}_{\lambda_j}, \boldsymbol{e}_{\lambda_{j'}})\|_F \\
&= u. \qquad (38)
\end{aligned}$$

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 $\alpha_k \in J_s := [1, 2s]$, $k = 1, \dots, K$, 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 $\boldsymbol{z}_1, \dots, \boldsymbol{z}_m, \boldsymbol{z}'_1, \dots, \boldsymbol{z}'_m$ 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 $\boldsymbol{z}_1, \dots, \boldsymbol{z}_m, \boldsymbol{z}'_1, \dots, \boldsymbol{z}'_m$ by the respective $\tilde{\boldsymbol{z}}_1, \dots, \tilde{\boldsymbol{z}}_m, \tilde{\boldsymbol{z}}'_1, \dots, \tilde{\boldsymbol{z}}'_m$ of the form $\alpha_j p_\lambda \boldsymbol{e}_\lambda$.

Then, using (36), (38) 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 relevant matrix. (Clearly, if for a matrix $\frac{1}{m^2} \sum_{j,j'=1}^{m} \boldsymbol{B}(\tilde{\boldsymbol{z}}_j, \tilde{\boldsymbol{z}}'_{j'})$ there is no such $\tilde{\boldsymbol{x}}$, then we can discard that matrix.) $\frac{1}{m^2} \sum_{j,j'=1}^{m} \boldsymbol{B}(\tilde{\boldsymbol{z}}_j, \tilde{\boldsymbol{z}}'_{j'})$ 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 5 is completed.

3.4 Proof of Lemma 6, 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 (20). 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 (28) in Lemma 8, 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 (27) 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 (26) for the Schatten norms shows

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

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,

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

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

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

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

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

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

$$d_1(\boldsymbol{x},\boldsymbol{y}) \leq 2s\|\boldsymbol{x}-\boldsymbol{y}\|_2. \tag{39}$$

With the volumetric argument, see for example [44, Proposition 10.1], we obtain the bound

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

Rescaling yields

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

which is the claimed inequality.

3.5 Proof of Lemma 6, Part II

Next we establish the remaining estimate of (20),

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

To this end, we use again the empirical method as in Section 3.3.

For $\boldsymbol{x} \in T_s$, we define $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m$ and $\boldsymbol{Z}_1', \ldots, \boldsymbol{Z}_m'$ 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

$$B(\boldsymbol{Z},\boldsymbol{Z}') = (\boldsymbol{Z}^*\boldsymbol{W}_{q'q}\boldsymbol{Z}')_{q',q}, \tag{40}$$

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

$$\boldsymbol{B} := \frac{1}{m}\sum_{j=1}^{m} \boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}_j'). \tag{41}$$

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 [44, Lemma 6.7], see also [33, Lemma 6.3], [39, Lemma 1.2.6]. Note that we will use this result with $\boldsymbol{B}_j = \boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}_j')$.

Lemma 9 *(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 $1 \leq p < \infty$*

$$\Big(\mathbb{E}\|\sum_{j=1}^{m}(\boldsymbol{Y}_j - \beta_j)\|^p\Big)^{1/p} \leq 2\Big(\mathbb{E}\|\sum_{j=1}^{m}\epsilon_j\boldsymbol{Y}_j\|^p\Big)^{1/p}, \tag{42}$$

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 [7, 44] which makes use of the Schatten p-norms introduced in (24).

theorem 10 *(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*

$$\mathbb{E}\|\sum_{j=1}^{m}\epsilon_j\boldsymbol{A}_j\|_{S_{2p}}^{2p} \leq \frac{(2p)!}{2^p p!}\max\Big\{\Big\|\Big(\sum_{j=1}^{m}\boldsymbol{A}_j\boldsymbol{A}_j^*\Big)^{1/2}\Big\|_{S_{2p}}^{2p}, \Big\|\Big(\sum_{j=1}^{m}\boldsymbol{A}_j^*\boldsymbol{A}_j\Big)^{1/2}\Big\|_{S_{2p}}^{2p}\Big\}. \tag{43}$$

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

$$\begin{aligned}
\Big(\mathbb{E}\|\boldsymbol{B} - \boldsymbol{B}(\boldsymbol{x})\|_{2\to 2}^{2p}\Big)^{\frac{1}{2p}} &= \Big(\mathbb{E}\|\frac{1}{m}\sum_{j=1}^{m}(\boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}_j') - \mathbb{E}\boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}_j'))\|_{2\to 2}^{2p}\Big)^{\frac{1}{2p}} \\
\leq \frac{2}{m}\Big(\mathbb{E}\|\sum_{j=1}^{m}\epsilon_j\boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}_j')\|_{2\to 2}^{2p}\Big)^{\frac{1}{2p}} &\leq \frac{2}{m}\Big(\mathbb{E}\|\sum_{j=1}^{m}\epsilon_j\boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}_j')\|_{S_{2p}}^{2p}\Big)^{\frac{1}{2p}} \\
\leq \frac{2}{m}\Big(\frac{(2p)!}{2^p p!}\Big)^{\frac{1}{2p}}\Big(\mathbb{E}\max\Big\{&\Big\|\Big(\sum_{j=1}^{m}\boldsymbol{B}(Z_j, Z_j')^*\boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}_j')\Big)^{1/2}\Big\|_{S_{2p}}^{2p}, \\
&\Big\|\Big(\sum_{j=1}^{m}\boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}_j')\boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}_j')^*\Big)^{1/2}\Big\|_{S_{2p}}^{2p}\Big\}\Big)^{\frac{1}{2p}}.
\end{aligned} \tag{44}$$

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

$$\boldsymbol{B}(\boldsymbol{Z}_\ell, \boldsymbol{Z}_\ell')^*\boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}_j') = \|\boldsymbol{x}\|_*^4\boldsymbol{I}. \tag{45}$$

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

$$\begin{aligned}
&\|\Big(\sum_{j=1}^{m} \boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}_j')^* \boldsymbol{B}(\boldsymbol{Z}_j, \boldsymbol{Z}_j')\Big)^{1/2}\|_{S_{2p}}^{2p} = \|\Big(\sum_{j=1}^{m} \|\boldsymbol{x}\|_*^4 \boldsymbol{I}\Big)^{1/2}\|_{S_{2p}}^{2p} = \|\boldsymbol{x}\|_*^{4p} m^p n \\
&\leq (2s)^{2p} m^p n\,. \qquad (46)
\end{aligned}$$

By symmetry this inequality applies also to the second term in the maximum in (44). 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 [44],

$$\Big(\mathbb{E}\|\boldsymbol{B} - \boldsymbol{B}(\boldsymbol{x})\|_{2\to 2}^{p}\Big)^{1/p} \leq 2^{3/(4p)} n^{1/p} e^{-1/2} \sqrt{p} \frac{4s}{\sqrt{m}}. \qquad (47)$$

Now we use the following lemma relating moments and tails [43, 44].

Proposition 11 *Suppose Ξ is a random variable satisfying*

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

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

$$m > \frac{32s^2}{u^2}\log(2^{3/4}n) \qquad (48)$$

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 (38),

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

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

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 $|\beta - \alpha_k^2| \le u$. Now, provided (48) 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 $\alpha_k \operatorname{sgn}(x_\lambda)\boldsymbol{e}_\lambda$, $p(\lambda) \in \{1, -1, i, -i\}$, with

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

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 $(4\lceil \frac{2s}{u} \rceil n^2)^{2m} \le (Cn^2 s/u)^{2m}$ 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

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

This completes the proof of Lemma 6.

4 Probability estimate

To prove Theorem 1(b) will use the following concentration inequality, which is a slight variant of Theorem 17 in [6], which in turn is an improved version of a striking result due to Talagrand [52]. Note that with $\boldsymbol{B}(\boldsymbol{x})$ as defined above, Y below satisfies $\mathbb{E}Y = n\,\mathbb{E}\delta_s$.

theorem 12 *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*

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

Define U and V to be

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

and

$$V = \mathbb{E} \sup_{\boldsymbol{x}\in T} \|\boldsymbol{B}(\boldsymbol{x})\boldsymbol{\epsilon}\|_2^2 = \mathbb{E} \sup_{\boldsymbol{x}\in T} \sum_{q'=1}^{n} \Big| \sum_{q=1}^{n} \epsilon_q B(\boldsymbol{x})_{q',q} \Big|^2. \tag{49}$$

Then, for $\lambda \ge 0$,

$$\mathbb{P}\Big(Y \ge \mathbb{E}[Y] + \lambda\Big) \;\le\; \exp\Big(-\frac{\lambda^2}{32V + 65U\lambda/3}\Big). \tag{50}$$

<u>**Proof:**</u> For Rademacher variables, the statement is exactly Theorem 17 in [6]. 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 $\boldsymbol{\epsilon}^{(\boldsymbol{\ell})} = (\epsilon_1, \ldots, \epsilon_\ell, \widetilde{\epsilon_\ell}, \epsilon_{\ell+1}, \ldots, \epsilon_n)$ 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,

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

The remainder of the proof is analogous to the one in [6] 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 12 is applicable. Now, it remains to estimate U and V. To this end, note that (39) implies

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

The remainder of this section develops an estimate of the quantity V in (49). 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 13 *Let $\boldsymbol{\epsilon} = (\epsilon_q)_{q=1}^n$ be a Steinhaus sequence and let $\boldsymbol{B} \in \mathbb{C}^{m\times n}$. Then, for $u \geq 0$,*

$$\mathbb{P}\Big(\|\boldsymbol{B\epsilon}\|_2 \geq u\|\boldsymbol{B}\|_F\Big) \leq 8e^{-u^2/16}. \tag{51}$$

<u>**Proof:**</u> In [46, Proposition B.1], it is shown that

$$\mathbb{P}\Big(\|\boldsymbol{B\epsilon}\|_2 \geq u\|\boldsymbol{B}\|_F\Big) \leq 2e^{-u^2/2}. \tag{52}$$

for Rademacher sequences. We extend this result using the contraction principle [33, Theorem 4.4], as in the proof of Theorem 3.

In fact, [33, 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

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

Hence,

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

■

With more effort, one may also derive (51) 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 13 and (52) that the increments of the process satisfy

$$\mathbb{P}(\|\boldsymbol{B}(\boldsymbol{x})\boldsymbol{\epsilon} - \boldsymbol{B}(\boldsymbol{x}')\boldsymbol{\epsilon}\|_2 \geq u\|\boldsymbol{B}(\boldsymbol{x}) - \boldsymbol{B}(\boldsymbol{x}')\|_F) \leq 8e^{-u^2/16}. \tag{53}$$

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

theorem 14 *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 $K > 0$,

$$\Big(\mathbb{E} \sup_{\boldsymbol{x}\in T} \|\boldsymbol{R}_{\boldsymbol{x}} - \boldsymbol{R}_{\boldsymbol{x_0}}\|_2^2\Big)^{1/2} \leq K \int_0^\infty \sqrt{\log(N(T, d, u))}du, \tag{54}$$

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 [44, Theorem 6.23] or [2, 33, 53]. 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, (21) provides us with the right hand side of (54). 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 (50) and simplifying leads to our result, compare with [46]. In particular, Theorem 1(b) follows.

acknoeldegements

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] Alltop, W.O.: Complex sequences with low periodic correlations. IEEE Trans. Inform. Theory **26**(3), 350–354 (1980)

[2] Azais, J.M., Wschebor, M.: Level Sets and Extrema of Random Processes and Fields. John Wiley & Sons Inc. (2009)

[3] Baraniuk, R.G., Davenport, M., DeVore, R.A., Wakin, M.: A simple proof of the restricted isometry property for random matrices. Constr. Approx. **28**(3), 253–263 (2008)

[4] Bello, P.A.: Characterization of Randomly Time-Variant Linear Channels. IEEE Trans. Comm. **11**, 360–393 (1963)

[5] Blumensath, T., Davies, M.: Iterative hard thresholding for compressed sensing. Appl. Comput. Harmon. Anal. **27**(3), 265–274 (2009)

[6] Boucheron, S., Lugosi, G., Massart, P.: Concentration inequalities using the entropy method. Ann. Probab. **31**(3), 1583–1614 (2003)

[7] Buchholz, A.: Operator Khintchine inequality in non-commutative probability. Math. Ann. **319**, 1–16 (2001)

[8] Cai, T., Wang, L., Xu, G.: Shifting inequality and recovery of sparse vectors. IEEE Trans. Signal Process. **58**(3), 1300–1308 (2010)

[9] Candès, E.J.: Compressive sampling. In: Proceedings of the International Congress of Mathematicians. Madrid, Spain (2006)

[10] Candès, E.J.: The restricted isometry property and its implications for compressed sensing. preprint (2008)

[11] Candès, E.J., J., Tao, T., Romberg, J.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory **52**(2), 489–509 (2006)

[12] Candès, E.J., Romberg, J., Tao, T.: Stable signal recovery from incomplete and inaccurate measurements. Comm. Pure Appl. Math. **59**(8), 1207–1223 (2006)

[13] Candès, E.J., Tao, T.: Near optimal signal recovery from random projections: universal encoding strategies? IEEE Trans. Inform. Theory **52**(12), 5406–5425 (2006)

[14] Carl, B.: Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in Banach spaces. Ann. Inst. Fourier (Grenoble) **35**(3), 79–118 (1985)

[15] Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by Basis Pursuit. SIAM J. Sci. Comput. **20**(1), 33–61 (1999)

[16] Christensen, O.: An Introduction to Frames and Riesz Bases. Applied and Numerical Harmonic Analysis. Birkhäuser Boston Inc., Boston, MA (2003)

[17] Cohen, A., Dahmen, W., DeVore, R.A.: Compressed sensing and best k-term approximation. J. Amer. Math. Soc. **22**(1), 211–231 (2009)

[18] Correia, L.M.: Wireless Flexible Personalized Communications. John Wiley & Sons, Inc., New York, NY, USA (2001)

[19] Donoho, D.L.: Compressed sensing. IEEE Trans. Inform. Theory **52**(4), 1289–1306 (2006)

[20] Donoho, D.L., Tanner, J.: Counting faces of randomly-projected polytopes when the projection radically lowers dimension. J. Amer. Math. Soc. **22**(1), 1–53 (2009)

[21] Fornasier, M., Rauhut, H.: Compressive sensing. In: O. Scherzer (ed.) Handbook of Mathematical Methods in Imaging, pp. 187–228. Springer (2011)

[22] Foucart, S.: A note on guaranteed sparse recovery via ℓ_1-minimization. Appl. Comput. Harmon. Anal. **29**(1), 97–103 (2010)

[23] Foucart, S.: Hard thresholding pursuit: an algorithm for compressive sensing. preprint (2010)

[24] Foucart, S.: Sparse recovery algorithms: sufficient conditions in terms of restricted isometry constants. In: Proceedings of the 13th International Conference on Approximation Theory (2010)

[25] Foucart, S., Pajor, A., Rauhut, H., Ullrich, T.: The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$. J. Complexity **26**(6), 629–640 (2010)

[26] Garnaev, A., Gluskin, E.: On widths of the Euclidean ball. Sov. Math., Dokl. **30**, 200–204 (1984)

[27] Grip, N., Pfander, G.: A discrete model for the efficient analysis of time-varying narrowband communication channels. Multidim. Syst. Signal Processing **19**(1), 3–40 (2008)

[28] Gröchenig, K.: Foundations of Time-Frequency Analysis. Applied and Numerical Harmonic Analysis. Birkhäuser, Boston, MA (2001)

[29] Haupt, J., Bajwa, W., Raz, G., Nowak, R.: Toeplitz compressed sensing matrices with applications to sparse channel estimation. IEEE Trans. Inform. Theory **56**(11), 5862–5875 (2010)

[30] Herman, M., Strohmer, T.: High-resolution radar via compressed sensing. IEEE Trans. Signal Process. **57**(6), 2275–2284 (2009)

[31] Krahmer, F., Pfander, G.E., Rashkov, P.: Uncertainty in time-frequency representations on finite abelian groups and applications. Appl. Comput. Harmon. Anal. **25**(2), 209–225 (2008)

[32] Lawrence, J., Pfander, G., Walnut, D.: Linear independence of Gabor systems in finite dimensional vector spaces. J. Fourier Anal. Appl. **11**(6), 715–726 (2005)

[33] Ledoux, M., Talagrand, M.: Probability in Banach spaces. Springer-Verlag, Berlin, Heidelberg, NewYork (1991)

[34] Mendelson, S., Pajor, A., Tomczak Jaegermann, N.: Uniform uncertainty principle for Bernoulli and subgaussian ensembles. Constr. Approx. **28**(3), 277–289 (2009)

[35] Middleton, D.: Channel modeling and threshold signal processing in underwater acoustics: An analytical overview. IEEE J. Oceanic Eng. **12**(1), 4–28 (1987)

[36] Natarajan, B.K.: Sparse approximate solutions to linear systems. SIAM J. Comput. **24**, 227–234 (1995)

[37] Needell, D., Vershynin, R.: Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. Found. Comput. Math. **9**(3), 317–334 (2009)

[38] Pätzold, M.: Mobile Fading Channels: Modelling, Analysis and Simulation. John Wiley & Sons, Inc. (2001)

[39] de la Peña, V., Giné, E.: Decoupling. From Dependence to Independence. Probability and its Applications (New York). Springer-Verlag (1999)

[40] Pfander, G., Rauhut, H.: Sparsity in time–frequency representations. J. Fourier Anal. Appl. **16**(2), 233–260 (2010)

[41] Pfander, G.E., Rauhut, H., Tanner, J.: Identification of matrices having a sparse representation. IEEE Trans. Signal Process. **56**(11), 5376–5388 (2008)

[42] Rauhut, H.: Stability results for random sampling of sparse trigonometric polynomials. IEEE Trans. Information Theory **54**(12), 5661–5670 (2008)

[43] Rauhut, H.: Circulant and Toeplitz matrices in compressed sensing. In: Proc. SPARS'09 (2009)

[44] Rauhut, H.: Compressive Sensing and Structured Random Matrices. In: M. Fornasier (ed.) Theoretical Foundations and Numerical Methods for Sparse Recovery, *Radon Series Comp. Appl. Math.*, vol. 9, pp. 1–92. deGruyter (2010)

[45] Rauhut, H., Pfander, G.E.: Sparsity in time-frequency representations. J. Fourier Anal. Appl. **16**(2), 233–260 (2010)

[46] Rauhut, H., Romberg, J., Tropp, J.: Restricted isometries for partial random circulant matrices. Appl. Comput. Harmonic Anal. (to appear). DOI:10.1016/j.acha.2011.05.001

[47] Rauhut, H., Schnass, K., Vandergheynst, P.: Compressed sensing and redundant dictionaries. IEEE Trans. Inform. Theory **54**(5), 2210 – 2219 (2008)

[48] Rauhut, H., Ward, R.: Sparse Legendre expansions via l_1-minimization. preprint (2010)

[49] Rudelson, M., Vershynin, R.: On sparse reconstruction from Fourier and Gaussian measurements. Comm. Pure Appl. Math. **61**, 1025–1045 (2008)

[50] Stojanovic, M.: Underwater acoustic communications. In: J.G. Webster (ed.) Encyclopedia of Electrical and Electronics Engineering, vol. 22, pp. 688–698. John Wiley & Sons (1999)

[51] Strohmer, T., Heath, R.W.j.: Grassmannian frames with applications to coding and communication. Appl. Comput. Harmon. Anal. **14**(3), 257–275 (2003)

[52] Talagrand, M.: New concentration inequalities in product spaces. Invent. Math. **126**(3), 505–563 (1996)

[53] Talagrand, M.: The Generic Chaining. Springer Monographs in Mathematics. Springer-Verlag (2005)

[54] Tropp, J., Needell, D.: CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. Appl. Comput. Harmon. Anal. **26**(3), 301–321 (2008)

[55] Tropp, J.A.: Greed is good: Algorithmic results for sparse approximation. IEEE Trans. Inform. Theory **50**(10), 2231–2242 (2004)