

Suprema of Chaos Processes and the Restricted Isometry Property

Felix Krahmer*, Shahar Mendelson† and Holger Rauhut‡

July 1, 2012

Abstract

We present a new bound for suprema of a special type of chaos processes indexed by a set of matrices, which is based on a chaining method. As applications we show significantly improved estimates for the restricted isometry constants of partial random circulant matrices and time-frequency structured random matrices. In both cases the required condition on the number m of rows in terms of the sparsity s and the vector length n is $m \gtrsim s \log^2 s \log^2 n$.

Key words. Compressive sensing, restricted isometry property, structured random matrices, chaos processes, γ_2 -functionals, generic chaining, partial random circulant matrices, random Gabor synthesis matrices.

1 Introduction and Main Results

1.1 Compressive Sensing

Compressive sensing [7, 13, 18, 37] is a method aimed at recovering sparse vectors from highly incomplete information using efficient algorithms. This discovery has recently triggered various applications in signal and image processing.

To formulate the procedure, a vector $\mathbf{x} \in \mathbb{C}^n$ is called s -sparse if $\|\mathbf{x}\|_0 := |\{\ell : x_\ell \neq 0\}| \leq s$. Given a matrix $\Phi \in \mathbb{C}^{m \times n}$, called the measurement matrix, the task is to reconstruct \mathbf{x} from the linear measurements

$$\mathbf{y} = \Phi \mathbf{x}.$$

We are interested in the case $m \ll n$, so that this system is under-determined, and thus, without additional information it is impossible to reconstruct \mathbf{x} . On the other hand, if it is known a priori that \mathbf{x} is s -sparse then the situation changes. And, although the naive approach for reconstruction, namely, ℓ_0 -minimization,

$$\min \|\mathbf{z}\|_0 \quad \text{subject to } \Phi \mathbf{z} = \mathbf{y}$$

is NP-hard in general, there are several tractable alternatives – for instance, ℓ_1 -minimization [11, 13, 7]

$$\min \|\mathbf{z}\|_1 \quad \text{subject to } \Phi \mathbf{z} = \mathbf{y},$$

*Institute for Numerical and Applied Mathematics, University of Göttingen, Lotzestraße 16-18, 37083 Göttingen, Germany, f.krahmer@math.uni-goettingen.de

†Department of Mathematics, Technion, Haifa, 32000, Israel, shahar@tx.technion.ac.il

‡Hausdorff Center for Mathematics and Institute for Numerical Simulation, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany, rauhut@hcm.uni-bonn.de

(where $\|\mathbf{z}\|_p$ denotes the usual ℓ_p -norm) which is a convex optimization problem and may be solved efficiently.

The restricted isometry property streamlines the analysis of recovery algorithms. For a matrix $\Phi \in \mathbb{C}^{m \times n}$ and $s < n$, the restricted isometry constant δ_s is defined as the smallest number such that

$$(1 - \delta_s)\|\mathbf{x}\|_2^2 \leq \|\Phi\mathbf{x}\|_2^2 \leq (1 + \delta_s)\|\mathbf{x}\|_2^2 \quad \text{for all } s\text{-sparse } \mathbf{x}.$$

One may show that under conditions of the form $\delta_{\kappa s} \leq \delta^*$ for some $\delta^* < 1$ and some appropriate small integer κ , a variety of recovery algorithms reconstruct every s -sparse \mathbf{x} from $\mathbf{y} = \Phi\mathbf{x}$. Among these are ℓ_1 -minimization as mentioned above [8, 19, 31], orthogonal matching pursuit [47], CoSaMP [45, 19], iterative hard thresholding [4] and hard thresholding pursuit [20].

Remarkably, all optimal measurement matrices known so far are random matrices. For example, a Bernoulli random matrix $\Phi \in \mathbb{R}^{m \times n}$ has entries $\Phi_{jk} = \varepsilon_{jk}/\sqrt{m}$, where the ε_{jk} are independent, symmetric $\{-1, 1\}$ -valued random variables. Its restricted isometry constant satisfies $\delta_s \leq \delta$ with probability at least $1 - \eta$ provided that

$$m \geq C\delta^{-2}(s \ln(en/s) + \ln(\eta^{-1})),$$

where C is an absolute constant [9, 30, 3].

In practice, structure is an additional requirement on the measurement matrix Φ . Indeed, certain applications impose constraints on the matrix and recovery algorithms can be accelerated when fast matrix vector multiplication routines are available for Φ . Unfortunately, a Bernoulli random matrix does not possess any structure. This motivates the study of random matrices with more structure. Also, structured random matrix constructions usually involve a reduced degree of randomness. For example, partial random Fourier matrices $\Phi \in \mathbb{C}^{m \times n}$ arise as random row submatrices of the discrete Fourier matrix and their restricted isometry constants satisfy $\delta_s \leq \delta$ with high probability provided that

$$m \geq C\delta^{-2}s \log^3 s \log n,$$

see [9, 41].

This article provides a similar estimate for two further types of structured random matrices, namely partial random circulant matrices and time-frequency structured random matrices. The key proof ingredients will be new estimates for suprema of chaos processes of a certain type.

1.2 Partial random circulant matrices

Circulant matrices are connected to circular convolution, defined for two vectors $\mathbf{x}, \mathbf{z} \in \mathbb{C}^n$ by

$$(\mathbf{z} * \mathbf{x})_j := \sum_{k=1}^n z_{j \ominus k} x_k, \quad j = 1, \dots, n,$$

where $j \ominus k = j - k \pmod n$ is the cyclic subtraction. The circulant matrix $\mathbf{H} = \mathbf{H}_{\mathbf{z}} \in \mathbb{C}^{n \times n}$ associated with \mathbf{z} is given by $\mathbf{H}\mathbf{x} = \mathbf{z} * \mathbf{x}$ and has entries $H_{jk} = z_{j \ominus k}$.

We are interested in sparse recovery from subsampled convolutions with a random vector. Formally, let $\Omega \subset \{1, \dots, n\}$ be an arbitrary (fixed) set of cardinality m , and denote by $\mathbf{R}_{\Omega} : \mathbb{C}^n \rightarrow \mathbb{C}^m$ the operator that restricts a vector $\mathbf{x} \in \mathbb{C}^n$ to its entries in Ω . Let $\boldsymbol{\varepsilon} = (\varepsilon_i)_{i=1}^n$ be a Rademacher vector of length n , i.e., a random vector with independent entries distributed according to $\mathbb{P}(\varepsilon_i = \pm 1) = \frac{1}{2}$. Then the associated partial

random circulant matrix is given by $\Phi = m^{-1/2} \mathbf{R}_\Omega \mathbf{H}_\epsilon \in \mathbb{R}^{m \times n}$ and acts on vectors $\mathbf{x} \in \mathbb{C}^n$ via

$$\Phi \mathbf{x} = \frac{1}{\sqrt{m}} \mathbf{R}_\Omega (\epsilon * \mathbf{x}).$$

In other words, Φ is a circulant matrix generated by a Rademacher vector, where the rows outside Ω are removed. Our first main result establishes the restricted isometry property of Φ in a near-optimal parameter regime:

Theorem 1.1 *Let $\Phi \in \mathbb{R}^{m \times n}$ be a draw of a partial random circulant matrix generated by a Rademacher vector ϵ . If*

$$m \geq c\delta^{-2} s (\log^2 s) (\log^2 n), \quad (1.1)$$

then with probability at least $1 - n^{-(\log n)(\log^2 s)}$, the restricted isometry constant of Φ satisfies $\delta_s \leq \delta$. The constant $c > 0$ is universal.

In Section 4, we will prove a more general version of this theorem, just requiring that the generating random variable is mean-zero, variance one, and subgaussian. These results improve the best previously known estimates for a partial random circulant matrix [38], namely that $m \geq C_\delta (s \log n)^{3/2}$ is a sufficient condition for achieving $\delta_s \leq \delta$ with high probability (see also [22] for an earlier work on this problem). In particular, Theorem 1.1 removes the exponent 3/2 of the sparsity s , which was already conjectured in [38] to be an artefact of the proof.

A related non-uniform recovery result is contained in [36, 37] where one considers the probability that a fixed s -sparse vector is reconstructed via ℓ_1 -minimization using a draw of a partial random circulant matrix. The condition derived there is $m \geq Cs \log^2 n$, which is slightly better than (1.1). However, the statement of Theorem 1.1 is considerably stronger because it implies uniform and stable recovery of all s -sparse vectors via ℓ_1 -minimization and other recovery methods for a single matrix Φ .

Note that in [39], the restricted isometry property has been established for partial random circulant matrices with *random* sampling sets and random generators under the condition $m \geq Cs \log^6 n$. In contrast, our result holds for an arbitrary fixed selection of a set $\Omega \subset \{1, \dots, n\}$, which is important in applications since in many practical problems, it is natural or desired to consider structured sampling sets such as $\Omega = \{L, 2L, 3L, \dots, mL\}$ for some $L \in \mathbb{N}$; these sets are clearly far from being random.

Potential applications of compressive sensing with subsampled random convolutions include system identification, radar and cameras with coded aperture. We refer to [22, 39, 38] for a discussion on these applications.

Combining our result with the work [26] on the relation between the restricted isometry property and the Johnson-Lindenstrauss lemma we also obtain an improved estimate for Johnson-Lindenstrauss embeddings arising from partial random circulant matrices, see also [24, 46] for earlier work in this direction.

Theorem 1.2 *Fix $\eta, \delta \in (0, 1)$, and consider a finite set $E \subset \mathbb{R}^n$ of cardinality $|E| = p$. Choose*

$$m \geq C_1 \delta^{-2} \log(C_2 p) (\log \log(C_2 p))^2 (\log n)^2,$$

where the constants C_1, C_2 depend only on η . Let $\Phi \in \mathbb{C}^{m \times n}$ be a partial circulant matrix generated by a Rademacher vector ϵ . Furthermore, let $\epsilon' \in \mathbb{R}^n$ be a Rademacher vector independent of ϵ and set $\mathbf{D}_{\epsilon'}$ to be the diagonal matrix with diagonal ϵ' . Then with probability exceeding $1 - \eta$, for every $\mathbf{x} \in E$,

$$(1 - \delta) \|\mathbf{x}\|_2^2 \leq \|\Phi \mathbf{D}_{\epsilon'} \mathbf{x}\|_2^2 \leq (1 + \delta) \|\mathbf{x}\|_2^2.$$

1.3 Time-frequency structured random matrices

The translation and modulation operators on \mathbb{C}^m are defined by $(\mathbf{T}\mathbf{h})_j = h_{j\ominus 1}$ and $(\mathbf{M}\mathbf{h})_j = e^{2\pi i j/m} h_j = \omega^j h_j$, where $\omega = e^{2\pi i/m}$ and \ominus again denotes cyclic subtraction, this time modulo m . Observe that

$$(\mathbf{T}^k \mathbf{h})_j = h_{j\ominus k} \quad \text{and} \quad (\mathbf{M}^\ell \mathbf{h})_j = e^{2\pi i \ell j/n} h_j = \omega^{\ell j} h_j. \quad (1.2)$$

The time-frequency shifts are given by

$$\boldsymbol{\pi}(\lambda) = \mathbf{M}^\ell \mathbf{T}^k, \quad \lambda = (k, \ell) \in \mathbb{Z}_m^2 = \{0, \dots, m-1\}^2.$$

For $\mathbf{h} \in \mathbb{C}^m \setminus \{0\}$ the system $\{\boldsymbol{\pi}(\lambda)\mathbf{h} : \lambda \in \mathbb{Z}_m^2\}$, is called a Gabor system [16, 21, 25], and the $m \times m^2$ matrix $\boldsymbol{\Psi}_{\mathbf{h}}$ whose columns are the vectors $\boldsymbol{\pi}(\lambda)\mathbf{h}$, $\lambda \in \mathbb{Z}_m^2$ is called a Gabor synthesis matrix,

$$\boldsymbol{\Psi}_{\mathbf{h}} = [\boldsymbol{\pi}(\lambda)\mathbf{h}]_{\lambda \in \mathbb{Z}_m^2} \in \mathbb{C}^{m \times m^2}.$$

Note that here the signal length n is coupled to the embedding dimension m via $n = m^2$ (so that $\log n = 2 \log m$ below).

Our second main result establishes the restricted isometry property for Gabor synthesis matrices generated by a random vector. The following formulation again focuses on normalized Rademacher vectors, postponing a more general version of our results until Section 5.

Theorem 1.3 *Let $\boldsymbol{\varepsilon}$ be a Rademacher vector and consider the Gabor synthesis matrix $\boldsymbol{\Psi}_{\mathbf{h}} \in \mathbb{C}^{m \times m^2}$ generated by $\mathbf{h} = \frac{1}{\sqrt{m}} \boldsymbol{\varepsilon}$. If*

$$m \geq c\delta^{-2} s (\log s)^2 (\log m)^2, \quad (1.3)$$

then with probability at least $1 - m^{-(\log m) \cdot (\log^2 s)}$, the restricted isometry constant of $\boldsymbol{\Psi}_{\mathbf{h}}$ satisfies $\delta_s \leq \delta$.

Again, Theorem 1.3 improves the best previously known estimate from [34], in which the sufficient condition of $m \geq Cs^{3/2} \log^3 m$ was derived. In particular, it implies the first uniform sparse recovery result with a linear scaling of the number of samples m in the sparsity s (up to log-factors).

A non-uniform recovery result for Gabor synthesis matrices with Steinhaus generator (see Section 2 for the definition) appears in [32], where it was shown that a fixed s -sparse vector is recovered from its image under a random draw of the $m \times m^2$ Gabor synthesis matrix via ℓ_1 -minimization with high probability provided that $m \geq Cs \log m$. Again, the conclusion of Theorem 1.3 is stronger than this previous result in the sense that it implies uniform and stable s -sparse recovery. Further related material may be found in [33, 2].

Applications of random Gabor synthesis matrices include operator identification (channel estimation in wireless communications), radar and sonar [2, 23, 33].

1.4 Suprema of chaos processes

Both for partial random circulant matrices and for time-frequency structured random matrices generated by Rademacher vectors, the restricted isometry constants δ_s can be written as a random variable X of the form

$$X = \sup_{\mathbf{A} \in \mathcal{A}} \left| \|\mathbf{A}\boldsymbol{\varepsilon}\|_2^2 - \mathbb{E}\|\mathbf{A}\boldsymbol{\varepsilon}\|_2^2 \right|, \quad (1.4)$$

where \mathcal{A} is a set of matrices and $\boldsymbol{\varepsilon}$ is a Rademacher vector. Due to the identity (1.7) below, X is the supremum of a chaos process.

Our third main result – the main ingredient of the proofs of Theorems 1.1 and 1.3, but also of independent interest – provides expectation and deviation bounds for random vectors X of this form in terms of two types of complexity parameters of the set of matrices \mathcal{A} . The first one, denoted by $d_F(\mathcal{A})$ and $d_{2 \rightarrow 2}(\mathcal{A})$, is the radius of \mathcal{A} in the Frobenius norm $\|\mathbf{A}\|_F = \sqrt{\text{tr}(\mathbf{A}^* \mathbf{A})}$ and the operator norm $\|\mathbf{A}\|_{2 \rightarrow 2} = \sup_{\|\mathbf{x}\|_2 \leq 1} \|\mathbf{A}\mathbf{x}\|_2$, respectively. That is, $d_F(\mathcal{A}) = \sup_{\mathbf{A} \in \mathcal{A}} \|\mathbf{A}\|_F$ and $d_{2 \rightarrow 2}(\mathcal{A}) = \sup_{\mathbf{A} \in \mathcal{A}} \|\mathbf{A}\|_{2 \rightarrow 2}$. For the second one, Talagrand’s functional $\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2})$, we refer to Definition 2.1 below for a precise description.

With these notions, our result reads as follows.

Theorem 1.4 *Let $\mathcal{A} \subset \mathbb{C}^{m \times n}$ be a symmetric set of matrices, $\mathcal{A} = -\mathcal{A}$. Let $\boldsymbol{\epsilon}$ be a Rademacher vector of length n . Then*

$$\mathbb{E} \sup_{\mathbf{A} \in \mathcal{A}} \left| \|\mathbf{A}\boldsymbol{\epsilon}\|_2^2 - \mathbb{E} \|\mathbf{A}\boldsymbol{\epsilon}\|_2^2 \right| \leq C_1 (d_F(\mathcal{A})\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2})^2) =: C_1 E. \quad (1.5)$$

Furthermore, for $t > 0$,

$$\mathbb{P} \left(\sup_{\mathbf{A} \in \mathcal{A}} \left| \|\mathbf{A}\boldsymbol{\epsilon}\|_2^2 - \mathbb{E} \|\mathbf{A}\boldsymbol{\epsilon}\|_2^2 \right| \geq C_2 E + t \right) \leq 2 \exp \left(-C_3 \min \left\{ \frac{t^2}{V^2}, \frac{t}{U} \right\} \right), \quad (1.6)$$

where

$$V = d_{2 \rightarrow 2}(\mathcal{A})(\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A})) \quad \text{and} \quad U = d_{2 \rightarrow 2}^2(\mathcal{A}).$$

The constants $C_1, C_2, C_3 > 0$ are universal.

The symmetry assumption $\mathcal{A} = -\mathcal{A}$ was made for the sake of simplicity. The more general Theorem 3.1 below does not use this assumption but requires an additional term on the right hand side of the estimate. Furthermore, Theorem 3.1 will actually be stated under more general conditions on the generating random vector.

Let us relate our new bound to previous estimates. By expanding the ℓ_2 -norms we can rewrite X in (1.4) as

$$X = \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_{j \neq k} \epsilon_j \epsilon_k (\mathbf{A}^* \mathbf{A})_{j,k} \right|, \quad (1.7)$$

which is a homogeneous chaos processes of order 2 indexed by the positive semidefinite matrices $\mathbf{A}^* \mathbf{A}$. Talagrand [44] considers general homogeneous chaos process of the form

$$Y = \sup_{\mathbf{B} \in \mathcal{B}} \left| \sum_{j \neq k} \epsilon_j \epsilon_k B_{j,k} \right|,$$

where $\mathcal{B} \subset \mathbb{C}^{n \times n}$ is a set of (not necessarily positive semidefinite) matrices. He derives the bound

$$\mathbb{E} Y \leq C_1 \gamma_2(\mathcal{B}, \|\cdot\|_F) + C_2 \gamma_1(\mathcal{B}, \|\cdot\|_{2 \rightarrow 2}) \quad (1.8)$$

(see Section 2 for the definition of the γ_α -functional). This estimate was an essential component in the proofs of the previous bounds for the restricted isometry constants of partial random circulant matrices [38] and of random Gabor synthesis matrices [34]. In fact, the appearance of the γ_1 -functional leads to the non-optimal exponent 3/2 in the sparsity s in the estimate of the required number m of samples. In contrast, as our bound for the chaos at hand does not involve the γ_1 -functional but only the γ_2 -functional, this issue does not arise here.

Remark. The benchmark problems of estimating the singular values and the restricted isometry constant of a Bernoulli matrix (with independent ± 1 entries) can also be recast as a supremum of chaos processes of the form (1.4). The bounds resulting from Theorem 1.4 are then optimal up to a constant factor. Again, we are not aware of a way to deduce such bounds from (1.8).

Acknowledgements

H.R. and F.K. acknowledge support by the Hausdorff Center for Mathematics. H.R. is funded by the European Research Council by the Starting Grant StG 258926. Parts of this research were developed during a stay of H.R. and F.K. at the Institute for Mathematics and Its Applications, University of Minnesota. Both are grateful for the support and the stimulating research environment. S.M. acknowledges the support of the Centre for Mathematics and its Applications, The Australian National University, Canberra, Australia. Additional support to S.M. was given by an Australian Research Council Discovery grant DP0986563, the European Community's Seventh Framework Programme (FP7/2007-2013) under ERC grant agreement 203134, and the Israel Science Foundation grant 900/10.

2 Preliminaries

2.1 Chaining

The following definition is due to Talagrand [44] and forms the core of the generic chaining methodology.

Definition 2.1 *For a metric space (T, d) , an admissible sequence of T is a collection of subsets of T , $\{T_s : s \geq 0\}$, such that for every $s \geq 1$, $|T_s| \leq 2^{2^s}$ and $|T_0| = 1$. For $\beta \geq 1$, define the γ_β functional by*

$$\gamma_\beta(T, d) = \inf \sup_{t \in T} \sum_{s=0}^{\infty} 2^{s/\beta} d(t, T_s),$$

where the infimum is taken with respect to all admissible sequences of T .

Recall that for a metric space (T, d) and $u > 0$, the covering number $N(T, d, u)$ is the minimal number of open balls of radius u in (T, d) needed to cover T . The γ_α -functionals can be bounded in terms of such covering numbers by the well-known Dudley integral (see, e.g., [44]). A more specific formulation for the γ_2 -functional of a set of matrices \mathcal{A} equipped with the operator norm, the scenario which we will focus on in this article, is

$$\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) \leq c \int_0^{d_{2 \rightarrow 2}(\mathcal{A})} \sqrt{\log N(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u)} du \quad (2.1)$$

This type of entropy integral was introduced by Dudley [15] to bound the supremum of Gaussian processes, and was extended by Pisier [35] as a way of bounding processes that satisfy different decay properties.

When considered for a set $T \subset L_2$, γ_2 has close connections with properties of the canonical Gaussian process indexed by T ; we refer the reader to [14, 44] for detailed expositions on these connections. One can show that under mild measurability assumptions, if $\{G_t : t \in T\}$ is a centered Gaussian process indexed by a set T , then

$$c_1 \gamma_2(T, d) \leq \mathbb{E} \sup_{t \in T} G_t \leq c_2 \gamma_2(T, d), \quad (2.2)$$

where c_1 and c_2 are absolute constants, and for every $s, t \in T$, $d^2(s, t) = \mathbb{E}|G_s - G_t|^2$. The upper bound is due to Fernique [17] and the lower bound is Talagrand's majorizing measures theorem [42, 44].

2.2 Subgaussian random vectors

In this section, we will discuss different classes of random vectors that are needed in the formulation of the main results in a more general framework. In the following definition, S^{n-1} denotes the unit sphere in \mathbb{R}^n (resp. in \mathbb{C}^n).

Definition 2.2 *A mean-zero random vector X on \mathbb{C}^n is called **isotropic** if for every $\theta \in S^{n-1}$, $\mathbb{E}|\langle X, \theta \rangle|^2 = 1$. A random vector X is called **L -subgaussian** if it is isotropic and $\mathbb{P}(|\langle X, \theta \rangle| \geq t) \leq 2 \exp(-t^2/2L^2)$ for every $\theta \in S^{n-1}$, and any $t > 0$.*

It is well known that, up to an absolute constant, the tail estimates in the definition of a subgaussian random vector are equivalent to the moment characterization

$$\sup_{\theta \in S^{n-1}} (\mathbb{E}|\langle X, \theta \rangle|^p)^{1/p} \leq \sqrt{p}L. \quad (2.3)$$

Assume that a random vector $\boldsymbol{\xi}$ has independent coordinates ξ_i , each of which is an L -subgaussian random variable of mean zero and variance one. One may verify by direct computation that $\boldsymbol{\xi}$ is L -subgaussian. Rademacher vectors, standard Gaussian vectors, (that is, random vectors with independent normally distributed entries of mean zero and variance one), as well as Steinhaus vectors (that is, random vectors with independent entries that are uniformly distributed on $\{z \in \mathbb{C} : |z| = 1\}$), are examples of isotropic, L -subgaussian random vectors for an absolute constant L .

We will require the following well-known bound relating strong and weak moments. For convenience, a proof based on chaining and the majorizing measures theorem is provided in the appendix.

Theorem 2.3 *Let $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{C}^N$ and $T \subset \mathbb{C}^N$. If $\boldsymbol{\xi}$ is an isotropic, L -subgaussian random vector and $\mathbf{Y} = \sum_{j=1}^n \xi_j \mathbf{x}_j$, then for every $p \geq 1$,*

$$\left(\mathbb{E} \sup_{\mathbf{t} \in T} |\langle \mathbf{t}, \mathbf{Y} \rangle|^p \right)^{1/p} \leq c \left(\mathbb{E} \sup_{\mathbf{t} \in T} |\langle \mathbf{t}, \mathbf{G} \rangle| + \sup_{\mathbf{t} \in T} (\mathbb{E} |\langle \mathbf{t}, \mathbf{Y} \rangle|^p)^{1/p} \right), \quad (2.4)$$

where c is a constant which depends only on L and $\mathbf{G} = \sum_{j=1}^N g_j \mathbf{x}_j$ for g_1, \dots, g_N independent standard normal random variables.

Note that if $\|\cdot\|$ is some norm on \mathbb{C}^N and B_* is the unit ball in the dual norm of $\|\cdot\|$ then the above theorem implies that

$$(\mathbb{E} \|\mathbf{Y}\|^p)^{1/p} \leq c \left(\mathbb{E} \|\mathbf{G}\| + \sup_{\mathbf{t} \in B_*} (\mathbb{E} |\langle \mathbf{t}, \mathbf{Y} \rangle|^p)^{1/p} \right).$$

In the remainder of this article, we will state and prove generalizations of our main results Theorem 1.1, Theorem 1.3, and Theorem 1.4 to arbitrary isotropic vectors, whose coordinates are independent L -subgaussian random variables. Since a Rademacher vector has all these properties, the above formulations of our results will directly follow.

2.3 Further probabilistic tools

The following decoupling inequality is a slight variation of a result found for instance in [12], see also [6, 37].

Theorem 2.4 Let $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)$ be a sequence of independent, centered random variables, and let F be a convex function. If \mathcal{B} is a collection of matrices and $\boldsymbol{\xi}'$ is an independent copy of $\boldsymbol{\xi}$, then

$$\mathbb{E} \sup_{\mathbf{B} \in \mathcal{B}} F \left(\sum_{\substack{j,k=1 \\ j \neq k}}^n \xi_j \xi_k \mathbf{B}_{j,k} \right) \leq \mathbb{E} \sup_{\mathbf{B} \in \mathcal{B}} F \left(4 \sum_{j,k=1}^n \xi_j \xi'_k \mathbf{B}_{j,k} \right). \quad (2.5)$$

We require also a slightly stronger decoupling inequality which is valid in the Gaussian case and follows from specifying results from [1, Section 2] to an order 2 Gaussian chaos.

Theorem 2.5 There exists an absolute constant C such that the following holds for all $p \geq 1$. Let $\mathbf{g} = (g_1, \dots, g_n)$ be a sequence of independent standard normal random variables. If \mathcal{B} is a collection of Hermitian matrices and \mathbf{g}' is an independent copy of \mathbf{g} , then

$$\mathbb{E} \sup_{\mathbf{B} \in \mathcal{B}} \left| \sum_{\substack{j,k=1 \\ j \neq k}}^n g_j g_k \mathbf{B}_{j,k} + \sum_{j=1}^n (g_j^2 - 1) \mathbf{B}_{j,j} \right|^p \leq C^p \mathbb{E} \sup_{\mathbf{B} \in \mathcal{B}} \left| \sum_{j,k=1}^n g_j g'_k \mathbf{B}_{j,k} \right|^p. \quad (2.6)$$

Since some steps in our estimates are formulated in terms of moments, the transition to a tail bound can be established by the following standard estimate, which easily follows from Markov's inequality.

Proposition 2.6 Suppose Z is a random variable satisfying

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

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

$$\mathbb{P}(|Z| \geq e(\alpha + \beta\sqrt{u} + \gamma u)) \leq e^{-u}.$$

2.4 Notation

Absolute constants will be denoted by c_1, c_2, \dots ; their value may change from line to line. We write $A \lesssim B$ if there is an absolute constant c_1 for which $A \leq c_1 B$. $A \sim B$ means that $c_1 A \leq B \leq c_2 A$ for absolute constants c_1 and c_2 . If the constants depend on some parameter r we will write $A \lesssim_r B$ or $A \sim_r B$.

The L_p -norm of a random variable, or its p -th moment, is given by $\|X\|_{L_p} = (\mathbb{E}|X|^p)^{1/p}$. For a random variable X independent from all other random variables which appear, we denote the expectation and probability conditional on all variables except X by \mathbb{E}_X and \mathbb{P}_X , respectively. The canonical unit vectors in \mathbb{C}^n are denoted \mathbf{e}_j and B_2^n is the unit ℓ_2 -ball in \mathbb{C}^n .

Finally, we introduce shorthand notations for some quantities that we will study. To that end, let \mathcal{A} be a set of matrices on \mathbb{R}^n or on \mathbb{C}^n and set a random vector $\boldsymbol{\xi} = (\xi_i)_{i=1}^n$. For a given matrix \mathbf{A} , denote its j -th column by \mathbf{A}^j and set

$$\begin{aligned} N_{\mathcal{A}}(\boldsymbol{\xi}) &:= \sup_{\mathbf{A} \in \mathcal{A}} \|\mathbf{A}\boldsymbol{\xi}\|_2, & B_{\mathcal{A}}(\boldsymbol{\xi}) &:= \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_{\substack{j,k=1 \\ j \neq k}}^n \xi_j \overline{\xi_k} \langle \mathbf{A}^j, \mathbf{A}^k \rangle \right|, \\ D_{\mathcal{A}}(\boldsymbol{\xi}) &:= \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_{j=1}^n (|\xi_j|^2 - \mathbb{E}|\xi_j|^2) \|\mathbf{A}^j\|^2 \right|, & \text{and} & C_{\mathcal{A}}(\boldsymbol{\xi}) &:= \sup_{\mathbf{A} \in \mathcal{A}} \left| \|\mathbf{A}\boldsymbol{\xi}\|_2^2 - \mathbb{E}\|\mathbf{A}\boldsymbol{\xi}\|_2^2 \right|. \end{aligned}$$

3 Chaos processes

We are now well-equipped to prove the following generalized version of Theorem 1.4.

Theorem 3.1 *Let \mathcal{A} be a set of matrices, and let $\boldsymbol{\xi}$ be a random vector whose entries ξ_j are independent, mean-zero, variance 1, and L -subgaussian random variables. Set*

$$\begin{aligned} E &= \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) (\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A})) + d_F(\mathcal{A})d_{2 \rightarrow 2}(\mathcal{A}), \\ V &= d_{2 \rightarrow 2}(\mathcal{A})(\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A})), \quad \text{and} \quad U = d_{2 \rightarrow 2}^2(\mathcal{A}). \end{aligned}$$

Then, for $t > 0$,

$$\mathbb{P}(C_{\mathcal{A}}(\boldsymbol{\xi}) \geq c_1 E + t) \leq 2 \exp\left(-c_2 \min\left\{\frac{t^2}{V^2}, \frac{t}{U}\right\}\right).$$

The constants c_1, c_2 depend only on L .

Remark 3.2 *Theorem 3.1 directly implies the tail estimate of Theorem 1.4. Indeed, the symmetry assumption $\mathcal{A} = -\mathcal{A}$ ensures that $d_{2 \rightarrow 2}(\mathcal{A}) \leq \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2})$. The estimate for the expectation in Theorem 1.4 follows from Theorem 3.4 below by choosing $p = 1$.*

The proof is based on estimating the moments of the random variables $N_{\mathcal{A}}$ and $C_{\mathcal{A}}$, followed by applying Proposition 2.6. The first step is a bound on the moments of a decoupled version of $N_{\mathcal{A}}$.

Lemma 3.3 *Let \mathcal{A} be a set of matrices, let $\boldsymbol{\xi} = (\xi_i)_{i=1}^n$ be an L -subgaussian random vector, and let $\boldsymbol{\xi}'$ be an independent copy of $\boldsymbol{\xi}$. Then for every $p \geq 1$,*

$$\left\| \sup_{\mathbf{A} \in \mathcal{A}} \langle \mathbf{A}\boldsymbol{\xi}, \mathbf{A}\boldsymbol{\xi}' \rangle \right\|_{L_p} \lesssim_L \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) \|N_{\mathcal{A}}(\boldsymbol{\xi})\|_{L_p} + \sup_{\mathbf{A} \in \mathcal{A}} \|\langle \mathbf{A}\boldsymbol{\xi}, \mathbf{A}\boldsymbol{\xi}' \rangle\|_{L_p}.$$

Proof. The proof is based on a chaining argument. Since the space involved is finite dimensional, one may assume without loss of generality that \mathcal{A} is finite. Fix an admissible sequence (T_r) of \mathcal{A} , let $\pi_r \mathbf{A} = \operatorname{argmin}_{\mathbf{B} \in T_r} \|\mathbf{B} - \mathbf{A}\|_{2 \rightarrow 2}$ and set $\Delta_r \mathbf{A} = \pi_r \mathbf{A} - \pi_{r-1} \mathbf{A}$. Since \mathcal{A} is finite, there is some r_0 for which $|\mathcal{A}| \leq 2^{2^{r_0}}$. Given $p \geq 1$, let ℓ be the largest integer for which $2^\ell \leq p$, and we may assume that $\ell < r_0$ as the modifications needed when $\ell \geq r_0$ are minimal.

Note that for every $\mathbf{A} \in \mathcal{A}$,

$$\begin{aligned} & |\langle \mathbf{A}\boldsymbol{\xi}, \mathbf{A}\boldsymbol{\xi}' \rangle - \langle (\pi_\ell \mathbf{A})\boldsymbol{\xi}, (\pi_\ell \mathbf{A})\boldsymbol{\xi}' \rangle| \\ & \leq \sum_{r=\ell}^{r_0-1} |\langle (\Delta_{r+1} \mathbf{A})\boldsymbol{\xi}, (\pi_{r+1} \mathbf{A})\boldsymbol{\xi}' \rangle| + \sum_{r=\ell}^{r_0-1} |\langle (\pi_r \mathbf{A})\boldsymbol{\xi}, (\Delta_{r+1} \mathbf{A})\boldsymbol{\xi}' \rangle| \end{aligned}$$

Furthermore, conditionally on $\boldsymbol{\xi}'$,

$$\langle (\Delta_{r+1} \mathbf{A})\boldsymbol{\xi}, (\pi_{r+1} \mathbf{A})\boldsymbol{\xi}' \rangle = \langle \boldsymbol{\xi}, (\Delta_{r+1} \mathbf{A})^* (\pi_{r+1} \mathbf{A})\boldsymbol{\xi}' \rangle$$

is a subgaussian random variable, as for every $u > 0$,

$$\mathbb{P}_{\boldsymbol{\xi}}(|\langle \boldsymbol{\xi}, (\Delta_{r+1} \mathbf{A})^* (\pi_{r+1} \mathbf{A})\boldsymbol{\xi}' \rangle| \geq uL \|(\Delta_{r+1} \mathbf{A})^* (\pi_{r+1} \mathbf{A})\boldsymbol{\xi}'\|_2) \leq 2 \exp(-u^2/2). \quad (3.1)$$

Recall that $|\{\pi_r \mathbf{A} : \mathbf{A} \in \mathcal{A}\}| = |T_r| \leq 2^{2^r}$, so there are at most $2^{2^{r+2}}$ possible values that $\Delta_r(\mathbf{A})^* \pi_{r+1}(\mathbf{A})$ can assume in (3.1). Therefore, via a union bound over all these choices

and on all the levels $\ell < r \leq r_0$ (cf. (A.2) below), it is evident that there are constants $c_1, c_2 > 0$ for which if $t \geq c_1$ then with $\boldsymbol{\xi}$ -probability at least $1 - 2 \exp(-c_2 2^\ell t^2)$ one has for every $\ell < r \leq r_0$ and every $\mathbf{A} \in \mathcal{A}$

$$|\langle (\Delta_{r+1} \mathbf{A}) \boldsymbol{\xi}, (\pi_{r+1} \mathbf{A}) \boldsymbol{\xi}' \rangle| \leq t 2^{r/2} \|(\Delta_{r+1} \mathbf{A})^* (\pi_{r+1} \mathbf{A}) \boldsymbol{\xi}'\|_2. \quad (3.2)$$

Let $\mathcal{E}_t(\boldsymbol{\xi}')$ be the event for which (3.2) holds for all the possible choices of r and \mathbf{A} as above. Since

$$\|(\Delta_{r+1} \mathbf{A})^* (\pi_{r+1} \mathbf{A}) \boldsymbol{\xi}'\|_2 \leq \|\Delta_{r+1} \mathbf{A}\|_{2 \rightarrow 2} \sup_{\mathbf{A} \in \mathcal{A}} \|\mathbf{A} \boldsymbol{\xi}'\|_2 = \|\Delta_{r+1} \mathbf{A}\|_{2 \rightarrow 2} N_{\mathcal{A}}(\boldsymbol{\xi}'),$$

one has on $\mathcal{E}_t(\boldsymbol{\xi}')$

$$\begin{aligned} S_1 &:= \sum_{r=\ell}^{r_0-1} |\langle (\Delta_{r+1} \mathbf{A}) \boldsymbol{\xi}, (\pi_{r+1} \mathbf{A}) \boldsymbol{\xi}' \rangle| \leq \sum_{r=\ell}^{r_0-1} t 2^{r/2} \|\Delta_{r+1} \mathbf{A}\|_{2 \rightarrow 2} N_{\mathcal{A}}(\boldsymbol{\xi}') \\ &\leq t \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) N_{\mathcal{A}}(\boldsymbol{\xi}'). \end{aligned}$$

We will now estimate

$$\|S_1\|_{L_p}^p = \mathbb{E}_{\boldsymbol{\xi}} \mathbb{E}_{\boldsymbol{\xi}'} S_1^p = \mathbb{E}_{\boldsymbol{\xi}'} \int_0^\infty p t^{p-1} \mathbb{P}_{\boldsymbol{\xi}}(S_1 > t | \boldsymbol{\xi}') dt. \quad (3.3)$$

Setting $W(\boldsymbol{\xi}') = \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) N_{\mathcal{A}}(\boldsymbol{\xi}')$, observe that

$$\begin{aligned} \int_0^\infty p t^{p-1} \mathbb{P}_{\boldsymbol{\xi}}(S_1 > t | \boldsymbol{\xi}') dt &\leq c_3^p W(\boldsymbol{\xi}')^p + \int_{c_3 W(\boldsymbol{\xi}')}^\infty p t^{p-1} \mathbb{P}_{\boldsymbol{\xi}}(S_1 > t | \boldsymbol{\xi}') dt \\ &\leq c_3^p W(\boldsymbol{\xi}')^p + W(\boldsymbol{\xi}')^p \int_{c_3}^\infty p u^{p-1} \mathbb{P}_{\boldsymbol{\xi}}(S_1 > u W(\boldsymbol{\xi}') | \boldsymbol{\xi}') du \leq c_4^p W(\boldsymbol{\xi}')^p, \end{aligned}$$

where $c_3 \geq c_1$ and c_4 are constants that depends only on L . Indeed, for $u \geq c_1$,

$$\mathbb{P}_{\boldsymbol{\xi}}(S_1 > u W(\boldsymbol{\xi}') | \boldsymbol{\xi}') \leq \mathbb{P}_{\boldsymbol{\xi}}(\mathcal{E}_u(\boldsymbol{\xi}') | \boldsymbol{\xi}') \leq 2 \exp(-c_2 u^2 2^\ell) \leq 2 \exp(-c_2 u^2 p/2).$$

Repeating this argument for $S_2 = \sum_{r=\ell}^{r_0-1} |\langle (\pi_r \mathbf{A}) \boldsymbol{\xi}, (\Delta_{r+1} \mathbf{A}) \boldsymbol{\xi}' \rangle|$, it follows that

$$\|S_1 + S_2\|_{L_p} \leq c_5(L) \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) \|N_{\mathcal{A}}(\boldsymbol{\xi})\|_{L_p}.$$

Finally, since $|\{\pi_\ell \mathbf{A} : \mathbf{A} \in \mathcal{A}\}| \leq 2^{2^\ell} \leq \exp(p)$, we conclude

$$\begin{aligned} \mathbb{E} \sup_{\mathbf{A} \in \mathcal{A}} |\langle (\pi_\ell \mathbf{A}) \boldsymbol{\xi}, (\pi_\ell \mathbf{A}) \boldsymbol{\xi}' \rangle|^p &\leq \sum_{\tilde{\mathbf{A}} \in \mathcal{T}_\ell} \mathbb{E} |\langle \tilde{\mathbf{A}} \boldsymbol{\xi}, \tilde{\mathbf{A}} \boldsymbol{\xi}' \rangle|^p \\ &\leq \exp(p) \sup_{\mathbf{A} \in \mathcal{A}} \mathbb{E} |\langle \mathbf{A} \boldsymbol{\xi}, \mathbf{A} \boldsymbol{\xi}' \rangle|^p. \end{aligned}$$

Thus $\|\sup_{\mathbf{A} \in \mathcal{A}} |\langle (\pi_\ell \mathbf{A}) \boldsymbol{\xi}, (\pi_\ell \mathbf{A}) \boldsymbol{\xi}' \rangle|\|_{L_p} \leq e \sup_{\mathbf{A} \in \mathcal{A}} \|\langle \mathbf{A} \boldsymbol{\xi}, \mathbf{A} \boldsymbol{\xi}' \rangle\|_{L_p}$, which completes the proof. \blacksquare

With these preliminary results at hand, we can now proceed to establish moment bounds for the quantities in questions.

Theorem 3.4 *Let $L \geq 1$ and $\boldsymbol{\xi} = (\xi_j)_{j=1}^n$, where $\xi_j, j = 1, \dots, n$, are independent mean-zero, variance one, L -subgaussian random variables, and let \mathcal{A} be a class of matrices. Then for every $p \geq 1$,*

(a)

$$\|N_{\mathcal{A}}(\boldsymbol{\xi})\|_{L_p} \lesssim_L \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A}) + \sqrt{p}d_{2 \rightarrow 2}(\mathcal{A}),$$

(b)

$$\begin{aligned} \|C_{\mathcal{A}}(\boldsymbol{\xi})\|_{L_p} &\lesssim_L \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) (\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A})) \\ &\quad + \sqrt{p}d_{2 \rightarrow 2}(\mathcal{A}) (\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A})) + pd_{2 \rightarrow 2}^2(\mathcal{A}). \end{aligned}$$

Proof. We start by showing that if $\boldsymbol{\xi}'$ is an independent copy of $\boldsymbol{\xi}$, then

$$\sup_{\mathbf{A} \in \mathcal{A}} \|\langle \mathbf{A}\boldsymbol{\xi}, \mathbf{A}\boldsymbol{\xi}' \rangle\|_{L_p} \lesssim_L \sqrt{p}d_F(\mathcal{A})d_{2 \rightarrow 2}(\mathcal{A}) + pd_{2 \rightarrow 2}^2(\mathcal{A}). \quad (3.4)$$

Indeed, fix $\mathbf{A} \in \mathcal{A}$ and set $S = \{\mathbf{A}^* \mathbf{A} \mathbf{x} : \mathbf{x} \in B_2^n\}$. Since the random vector $\boldsymbol{\xi}$ is L -subgaussian, the random variable $\langle \boldsymbol{\xi}, \mathbf{A}^* \mathbf{A} \boldsymbol{\xi}' \rangle$ is subgaussian conditionally on $\boldsymbol{\xi}'$. Therefore, by (2.3),

$$\begin{aligned} \|\langle \mathbf{A}\boldsymbol{\xi}, \mathbf{A}\boldsymbol{\xi}' \rangle\|_{L_p} &= \left(\mathbb{E}_{\boldsymbol{\xi}'} \left(\left(\mathbb{E}_{\boldsymbol{\xi}} |\langle \boldsymbol{\xi}, \mathbf{A}^* \mathbf{A} \boldsymbol{\xi}' \rangle|^p \right)^{1/p} \right)^p \right)^{1/p} \lesssim (\mathbb{E}_{\boldsymbol{\xi}'} L \sqrt{p} \|\mathbf{A}^* \mathbf{A} \boldsymbol{\xi}'\|_2^p)^{1/p} \\ &= L \sqrt{p} \left(\mathbb{E}_{\boldsymbol{\xi}'} \sup_{\mathbf{y} \in S} |\langle \mathbf{y}, \boldsymbol{\xi}' \rangle|^p \right)^{1/p}. \end{aligned}$$

Note that for a standard Gaussian vector \mathbf{g} ,

$$\mathbb{E} \sup_{\mathbf{y} \in S} |\langle \mathbf{y}, \mathbf{g} \rangle| = \mathbb{E} \|\mathbf{A}^* \mathbf{A} \mathbf{g}\|_2 \leq (\mathbb{E} \|\mathbf{A}^* \mathbf{A} \mathbf{g}\|_2^2)^{1/2} = \|\mathbf{A}^* \mathbf{A}\|_F \leq \|\mathbf{A}\|_F \|\mathbf{A}\|_{2 \rightarrow 2}.$$

Also, if $\boldsymbol{\xi}' = \sum_{j=1}^n \xi'_j \mathbf{e}_j$ then, applying (2.3) again,

$$\sup_{\mathbf{y} \in S} (\mathbb{E} |\langle \mathbf{y}, \boldsymbol{\xi}' \rangle|^p)^{1/p} = \sup_{\mathbf{z} \in B_2^n} (\mathbb{E} |\langle \mathbf{A}^* \mathbf{A} \mathbf{z}, \boldsymbol{\xi}' \rangle|^p)^{1/p} \lesssim L \sup_{\mathbf{z} \in B_2^n} \sqrt{p} \|\mathbf{A}^* \mathbf{A} \mathbf{z}\|_2 = L \sqrt{p} \|\mathbf{A}\|_{2 \rightarrow 2}^2.$$

Hence, Equation (3.4) follows by applying Theorem 2.3 and taking the supremum over $\mathbf{A} \in \mathcal{A}$.

As a preliminary step, we will consider the special case of the Gaussian vector \mathbf{g} and show that $\mathbb{E} N_{\mathcal{A}}(\mathbf{g}) \lesssim_L \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A})$, which is (a) for $p = 1$. To that end observe that by Theorem 2.5 and Lemma 3.3,

$$\begin{aligned} \|C_{\mathcal{A}}(\mathbf{g})\|_{L_p} &= \left\| \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_{\substack{j,k \\ j \neq k}} g_j g_k \langle \mathbf{A}^j, \mathbf{A}^k \rangle + \sum_j (g_j^2 - 1) \|\mathbf{A}^j\|^2 \right| \right\|_{L_p} \\ &\lesssim \left\| \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_{j,k} g_j g'_k \langle \mathbf{A}^j, \mathbf{A}^k \rangle \right| \right\|_{L_p} = \left\| \sup_{\mathbf{A} \in \mathcal{A}} |\langle \mathbf{A}\mathbf{g}, \mathbf{A}\mathbf{g}' \rangle| \right\|_{L_p} \\ &\lesssim_L \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) \|N_{\mathcal{A}}(\mathbf{g})\|_{L_p} + \sup_{\mathbf{A} \in \mathcal{A}} \|\langle \mathbf{A}\mathbf{g}, \mathbf{A}\mathbf{g}' \rangle\|_{L_p}. \quad (3.5) \end{aligned}$$

Combining (3.4) with (3.5), it follows that

$$\|C_{\mathcal{A}}(\mathbf{g})\|_{L_p} \lesssim_L \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) \|N_{\mathcal{A}}(\mathbf{g})\|_{L_p} + \sqrt{p}d_F(\mathcal{A})d_{2 \rightarrow 2}(\mathcal{A}) + pd_{2 \rightarrow 2}^2(\mathcal{A}). \quad (3.6)$$

Specifying $p = 1$ and using that $d_F(\mathcal{A}) \geq d_{2 \rightarrow 2}(\mathcal{A})$ as well as $\mathbb{E} \|\mathbf{A}\mathbf{g}\|_2^2 = \|\mathbf{A}\|_F^2$, we conclude

$$\mathbb{E} N_{\mathcal{A}}^2(\mathbf{g}) \leq \mathbb{E} C_{\mathcal{A}}(\mathbf{g}) + d_F^2(\mathcal{A}) \lesssim_L \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) \mathbb{E} N_{\mathcal{A}}(\mathbf{g}) + d_F^2(\mathcal{A}).$$

Therefore,

$$\mathbb{E}N_{\mathcal{A}}(\mathbf{g}) \leq (\mathbb{E}N_{\mathcal{A}}^2(\mathbf{g}))^{1/2} \lesssim_L \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A}),$$

as desired.

Finally, for general p and arbitrary subgaussian vectors, we apply Theorem 2.3 with the set $S = \{\mathbf{A}^* \mathbf{x} : \mathbf{x} \in B_2^n, \mathbf{A} \in \mathcal{A}\}$. Since $\boldsymbol{\xi}$ is L -subgaussian we obtain

$$\begin{aligned} \|N_{\mathcal{A}}(\boldsymbol{\xi})\|_{L_p} &= (\mathbb{E} \sup_{\mathbf{A} \in \mathcal{A}, \mathbf{x} \in B_2^n} |\langle \mathbf{A}\boldsymbol{\xi}, \mathbf{x} \rangle|^p)^{1/p} = (\mathbb{E} \sup_{\mathbf{u} \in S} |\langle \boldsymbol{\xi}, \mathbf{u} \rangle|^p)^{1/p} \\ &\lesssim_L \mathbb{E}N_{\mathcal{A}}(\mathbf{g}) + \sup_{\mathbf{u} \in S} (\mathbb{E} |\langle \boldsymbol{\xi}, \mathbf{u} \rangle|^p)^{1/p} \lesssim_L \mathbb{E}N_{\mathcal{A}}(\mathbf{g}) + \sqrt{p} \sup_{\mathbf{A} \in \mathcal{A}, \mathbf{x} \in B_2^n} \|\mathbf{A}^* \mathbf{x}\|_2 \\ &\lesssim_L \mathbb{E}N_{\mathcal{A}}(\mathbf{g}) + \sqrt{p} d_{2 \rightarrow 2}(\mathcal{A}), \\ &\lesssim_L \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A}) + \sqrt{p} d_{2 \rightarrow 2}(\mathcal{A}), \end{aligned}$$

which proves (a).

For (b), observe that as the ξ_j are unit variance, we have $\mathbb{E}\|\mathbf{A}\boldsymbol{\xi}\|_2^2 = \|\mathbf{A}\|_F^2 = \sum_{j=1}^n \|\mathbf{A}^j\|_2^2$ and, consequently, $C_{\mathcal{A}}(\boldsymbol{\xi})$ can be split up into the diagonal and the off-diagonal contributions as follows.

$$C_{\mathcal{A}}(\boldsymbol{\xi}) = \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_{\substack{j,k=1 \\ j \neq k}}^n \xi_j \bar{\xi}_k \langle \mathbf{A}^j, \mathbf{A}^k \rangle + \sum_{j=1}^n (|\xi_j|^2 - 1) \|\mathbf{A}^j\|_2^2 \right| \leq B_{\mathcal{A}}(\boldsymbol{\xi}) + D_{\mathcal{A}}(\boldsymbol{\xi}).$$

Hence it suffices to estimate the moments of $B_{\mathcal{A}}(\boldsymbol{\xi})$ and $D_{\mathcal{A}}(\boldsymbol{\xi})$; one concludes using the triangle inequality.

For the off-diagonal term, we use Theorem 2.4 and Lemma 3.3 to bound

$$\begin{aligned} \|B_{\mathcal{A}}(\boldsymbol{\xi})\|_{L_p} &\leq 4 \left\| \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_{j,k} \xi_j \bar{\xi}_k \langle \mathbf{A}^j, \mathbf{A}^k \rangle \right| \right\|_{L_p} = 4 \left\| \sup_{\mathbf{A} \in \mathcal{A}} |\langle \mathbf{A}\boldsymbol{\xi}, \mathbf{A}\boldsymbol{\xi}' \rangle| \right\|_{L_p} \\ &\lesssim_L \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) \|N_{\mathcal{A}}(\boldsymbol{\xi})\|_{L_p} + \sup_{\mathbf{A} \in \mathcal{A}} \|\langle \mathbf{A}\boldsymbol{\xi}, \mathbf{A}\boldsymbol{\xi}' \rangle\|_{L_p}. \end{aligned}$$

Combining this estimate with (3.4) and part (a), we obtain that

$$\begin{aligned} \|B_{\mathcal{A}}(\boldsymbol{\xi})\|_{L_p} &\lesssim_L \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) (\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A}) + \sqrt{p} d_{2 \rightarrow 2}(\mathcal{A})) \\ &\quad + \sqrt{p} d_F(\mathcal{A}) d_{2 \rightarrow 2}(\mathcal{A}) + p d_{2 \rightarrow 2}^2(\mathcal{A}). \end{aligned} \quad (3.7)$$

For the diagonal term, observe that by a standard symmetrization argument (see, e.g., [27, Lemma 6.3]),

$$\|D_{\mathcal{A}}(\boldsymbol{\xi})\|_{L_p} = \left\| \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_{j=1}^n (|\xi_j|^2 - \mathbb{E}|\xi_j|^2) \|\mathbf{A}^j\|_2^2 \right| \right\|_{L_p} \leq 2 \left\| \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_j \epsilon_j |\xi_j|^2 \|\mathbf{A}^j\|_2^2 \right| \right\|_{L_p}$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)$ is a Rademacher vector independent of $\boldsymbol{\xi}$. Furthermore, let $\mathbf{g} = (g_1, \dots, g_n)$ be a sequence of independent standard normal variables. Then, as ξ_j is L -subgaussian, there is an absolute constant c for which $\mathbb{P}(|\xi_j|^2 \geq tL^2) \leq c\mathbb{P}(g_j^2 \geq t)$ for every $t > 0$. Moreover, $\epsilon_j |\xi_j|^2$ and $\epsilon_j g_j^2$ are symmetric, so by the contraction principle (see

[27, Lemma 4.6]), a rescaling argument, and de-symmetrization [27, Lemma 6.3],

$$\begin{aligned}
\|D_{\mathcal{A}}(\boldsymbol{\xi})\|_{L_p} &\lesssim_L \left\| \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_j \epsilon_j g_j^2 \|\mathbf{A}^j\|_2^2 \right| \right\|_{L_p} \\
&\leq 2 \left\| \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_j (|g_j|^2 - 1) \|\mathbf{A}^j\|_2^2 \right| \right\|_{L_p} + \left\| \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_j \epsilon_j \|\mathbf{A}^j\|_2^2 \right| \right\|_{L_p} \\
&= 2 \|D_{\mathcal{A}}(\mathbf{g})\|_{L_p} + \left\| \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_j \epsilon_j \|\mathbf{A}^j\|_2^2 \right| \right\|_{L_p}.
\end{aligned}$$

Now observe that $D_{\mathcal{A}}(\mathbf{g}) \leq C_{\mathcal{A}}(\mathbf{g}) + B_{\mathcal{A}}(\mathbf{g})$, and thus, by (3.6) and (3.7),

$$\begin{aligned}
\|D_{\mathcal{A}}(\mathbf{g})\|_{L_p} &\leq \|C_{\mathcal{A}}(\mathbf{g})\|_{L_p} + \|B_{\mathcal{A}}(\mathbf{g})\|_{L_p} \\
&\lesssim \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) (\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A})) \\
&\quad + \sqrt{p} d_F(\mathcal{A}) (d_{2 \rightarrow 2}(\mathcal{A}) + \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2})) + p d_{2 \rightarrow 2}^2(\mathcal{A}).
\end{aligned}$$

Finally, note that $\mathbf{A} \rightarrow \sum_j \epsilon_j \|\mathbf{A}^j\|_2^2$ is a subgaussian process relative to the metric

$$\begin{aligned}
d(\mathbf{A}, \mathbf{B}) &= \left(\sum_{j=1}^n (\|\mathbf{A}^j\|_2^2 - \|\mathbf{B}^j\|_2^2)^2 \right)^{1/2} \\
&\leq \left(\sum_{j=1}^n \|\mathbf{A}^j - \mathbf{B}^j\|_2^2 \cdot (\|\mathbf{A}^j\|_2 + \|\mathbf{B}^j\|_2)^2 \right)^{1/2} \leq 2 d_F(\mathcal{A}) \|\mathbf{A} - \mathbf{B}\|_{2 \rightarrow 2}.
\end{aligned}$$

Therefore, by Theorem 2.3 and a standard chaining argument,

$$\left\| \sup_{\mathbf{A} \in \mathcal{A}} \left| \sum_j \epsilon_j \|\mathbf{A}^j\|_2^2 \right| \right\|_{L_p} \lesssim d_F(\mathcal{A}) \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + \sqrt{p} d_F(\mathcal{A}) d_{2 \rightarrow 2}(\mathcal{A}).$$

This shows that

$$\begin{aligned}
\|D_{\mathcal{A}}(\boldsymbol{\xi})\|_{L_p} &\lesssim_L \gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) (\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A})) \\
&\quad + \sqrt{p} d_{2 \rightarrow 2}(\mathcal{A}) (\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) + d_F(\mathcal{A})) + p d_{2 \rightarrow 2}^2(\mathcal{A}),
\end{aligned}$$

which, together with (3.7), proves (b). \blacksquare

Remark 3.5 (a) Observe that Theorem 3.1 can be deduced from Theorem 3.4 using Proposition 2.6.

(b) In the Rademacher case, once the bound for the expectation is derived, one may alternatively deduce the tail bound from the concentration inequality in [5, Theorem 17], see also [43].

(c) In the Rademacher case, one has $D_{\mathcal{A}} \equiv 0$, so the contraction principle and the more sophisticated decoupling inequality for Gaussian random variables are not needed in the proof.

(d) Note that the assumption that $\boldsymbol{\xi}$ has independent coordinates has only been used in the decoupling steps of the proof of Theorem 3.4.

4 The Restricted Isometry Property of Partial Random Circulant Matrices

In this section we study the restricted isometry constants of a partial random circulant matrix $\Phi \in \mathbb{R}^{m \times n}$ generated by a random vector $\xi = (\xi_i)_{i=1}^n$, where the ξ_i 's are independent mean-zero, L -subgaussian random variables of variance one. Arguably the most important case is when $\xi = \epsilon$ is a Rademacher vector, as introduced in Section 1.2.

Throughout this section and following the notation of the introduction, let $\mathbf{V}_x \mathbf{z} = \frac{1}{\sqrt{m}} \mathbf{P}_\Omega(\mathbf{x} * \mathbf{z})$, where the projection operator $\mathbf{P}_\Omega : \mathbb{C}^n \rightarrow \mathbb{C}^n$ is given by $\mathbf{P}_\Omega = \mathbf{R}_\Omega^* \mathbf{R}_\Omega$, that is, $(\mathbf{P}_\Omega \mathbf{x})_\ell = x_\ell$ for $\ell \in \Omega$ and $(\mathbf{P}_\Omega \mathbf{x})_\ell = 0$ for $\ell \notin \Omega$.

Recalling that $T_s = \{\mathbf{x} \in \mathbb{C}^n : \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x}\|_0 \leq s\}$, the restricted isometry constant of Φ is

$$\delta_s = \sup_{\mathbf{x} \in T_s} \left| \|\mathbf{R}_\Omega(\xi * \mathbf{x})\|^2 - \|\mathbf{x}\|_2^2 \right| = \sup_{\mathbf{x} \in T_s} \left| \|\mathbf{P}_\Omega(\mathbf{x} * \xi)\|^2 - \|\mathbf{x}\|_2^2 \right| = \sup_{\mathbf{x} \in T_s} \left| \|\mathbf{V}_x \xi\|_2^2 - \|\mathbf{x}\|_2^2 \right|.$$

Since $|\Omega| = m$, it follows that

$$\mathbb{E} \|\mathbf{V}_x \xi\|_2^2 = \frac{1}{m} \sum_{\ell \in \Omega} \mathbb{E} \sum_{k,j=1}^n \xi_j \bar{\xi}_k x_{\ell \ominus j} \overline{x_{\ell \ominus k}} = \frac{1}{m} \sum_{\ell \in \Omega} \sum_{k=1}^n |x_{\ell \ominus k}|^2 = \|\mathbf{x}\|_2^2$$

and hence

$$\delta_s = \sup_{\mathbf{x} \in T_s} \left| \|\mathbf{V}_x \xi\|_2^2 - \mathbb{E} \|\mathbf{V}_x \xi\|_2^2 \right|,$$

which shows that δ_s is the process $C_{\mathcal{A}}$ studied in the previous section for $\mathcal{A} = \{\mathbf{V}_x : \mathbf{x} \in T_s\}$. Hence the tail decay can be analyzed using Theorem 3.1.

Theorem 4.1 *Let $\xi = (\xi_j)_{j=1}^n$ be a random vector with independent mean-zero, variance one, L -subgaussian entries. If, for $s \leq n$ and $\eta, \delta \in (0, 1)$,*

$$m \geq c\delta^{-2}s \max\{(\log s)^2(\log n)^2, \log(\eta^{-1})\} \quad (4.1)$$

then with probability at least $1 - \eta$, the restricted isometry constant of the partial random circulant matrix $\Phi \in \mathbb{R}^{m \times n}$ generated by ξ satisfies $\delta_s \leq \delta$. The constant $c > 0$ depends only on L .

The proof of Theorem 4.1 requires a Fourier domain description of Φ . Let \mathbf{F} by the unnormalized Fourier transform with elements $F_{jk} = e^{2\pi ijk/n}$. By the convolution theorem, for every $1 \leq j \leq n$, $\mathbf{F}(\mathbf{x} * \mathbf{y})_j = (\mathbf{F}\mathbf{x})_j \cdot (\mathbf{F}\mathbf{y})_j$. Therefore,

$$\mathbf{V}_x \xi = \frac{1}{\sqrt{m}} \mathbf{P}_\Omega \mathbf{F}^{-1} \widehat{\mathbf{X}} \mathbf{F} \xi,$$

where $\widehat{\mathbf{X}} = \text{diag}(\mathbf{F}\mathbf{x})$ is the diagonal matrix, whose diagonal is the Fourier transform $\mathbf{F}\mathbf{x}$. In short,

$$\mathbf{V}_x = \frac{1}{\sqrt{m}} \widehat{\mathbf{P}}_\Omega \widehat{\mathbf{X}} \mathbf{F},$$

where $\widehat{\mathbf{P}}_\Omega = \mathbf{P}_\Omega \mathbf{F}^{-1}$.

Proof of Theorem 4.1. In light of Theorem 3.4 and Theorem 3.1, it suffices to control the parameters $d_{2 \rightarrow 2}(\mathcal{A})$, $d_F(\mathcal{A})$, and $\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2})$ for the set $\mathcal{A} = \{\mathbf{V}_x : \mathbf{x} \in T_s\}$.

Since the matrices \mathbf{V}_x consist of shifted copies of \mathbf{x} in all of their m nonzero rows, the ℓ_2 -norm of each nonzero row is $m^{-1/2} \|\mathbf{x}\|_2$; thus $\|\mathbf{V}_x\|_F = \|\mathbf{x}\|_2 \leq 1$ for all $\mathbf{x} \in T_s$ and

$$d_F(\mathcal{A}) = 1.$$

Also, observe that for every $\mathbf{x} \in T_s$ with associated diagonal matrix $\widehat{\mathbf{X}}$,

$$\begin{aligned}\|\mathbf{V}_{\mathbf{x}}\|_{2 \rightarrow 2} &= \frac{1}{\sqrt{m}} \|\widehat{\mathbf{P}}_{\Omega} \widehat{\mathbf{X}} \mathbf{F}\|_{2 \rightarrow 2} \leq \sqrt{\frac{n}{m}} \|\mathbf{P}_{\Omega} \mathbf{F}^{-1}\|_{2 \rightarrow 2} \|\widehat{\mathbf{X}}\|_{2 \rightarrow 2} \leq \frac{1}{\sqrt{m}} \|\widehat{\mathbf{X}}\|_{2 \rightarrow 2} \\ &= \frac{1}{\sqrt{m}} \|\mathbf{F}\mathbf{x}\|_{\infty}.\end{aligned}\tag{4.2}$$

Setting $\|\mathbf{x}\|_{\infty} := \|\mathbf{F}\mathbf{x}\|_{\infty}$ it is evident that $\|\mathbf{F}\mathbf{x}\|_{\infty} \leq \|\mathbf{x}\|_1 \leq \sqrt{s}\|\mathbf{x}\|_2 \leq \sqrt{s}$ for every $x \in T_s$, and thus

$$d_{2 \rightarrow 2}(\mathcal{A}) \leq \sqrt{s/m}.$$

Next, to estimate the γ_2 functional, recall from (2.1) that

$$\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) \lesssim \int_0^{d_{2 \rightarrow 2}(\mathcal{A})} \log^{1/2} N(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u) du.$$

By (4.2),

$$\|\mathbf{V}_{\mathbf{x}} - \mathbf{V}_{\mathbf{y}}\| = \|\mathbf{V}_{\mathbf{x} - \mathbf{y}}\| \leq m^{-1/2} \|\mathbf{x} - \mathbf{y}\|_{\infty},$$

and hence for every $u > 0$, $N(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u) \leq N(T_s, m^{-1/2} \|\cdot\|_{\infty}, u)$. Using an argument due to Carl [10, Prop. 3], (see also [41] or [37, Lemma 8.3]), and setting $\Delta = \sqrt{s/m}$, it is evident that

$$\begin{aligned}\log N(T_s, m^{-1/2} \|\cdot\|_{\infty}, u) &\leq \log N(s^{1/2} B_1^n, m^{-1/2} \|\cdot\|_{\infty}, u) \\ &\lesssim \left(\frac{\Delta}{u}\right)^2 \log^2(nu^2/\Delta^2).\end{aligned}$$

Since T_s is the union of s -dimensional Euclidean balls, a standard volumetric argument (see, e.g., [41] or [37, Chapter 8.4]) yields

$$\log N(T_s, m^{-1/2} \|\cdot\|_{\infty}, u) \lesssim s \log(en/su)$$

(which is stronger than the bound above for $u \lesssim 1/\sqrt{m}$).

Combining the two covering number estimates, a straightforward computation of the entropy integral, (see also [41] or [37, eq. (8.15)]), reveals that

$$\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) \lesssim \sqrt{\frac{s}{m}} (\log s) (\log n),$$

which implies that $\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) \lesssim \delta$ for the given choice of m .

Now, by choosing the constant c in (4.1) appropriately (depending only on L), one obtains

$$E \leq \frac{\delta}{2c_1},$$

where E and c_1 are chosen as in Theorem 3.1. Then Theorem 3.1 yields

$$\mathbb{P}(\delta_s \geq \delta) \leq \mathbb{P}(\delta_s \geq c_1 E + \delta/2) \leq \exp(-c_2(m/s)\delta^2) \leq \eta,$$

which, after possibly increasing the value of c enough to compensate c_2 , completes the proof. \blacksquare

5 Time-Frequency Structured Random Matrices

In this section, we will treat the restricted isometry property of random Gabor synthesis matrices, as described in Section 1.3.

Theorem 5.1 *Let $\boldsymbol{\xi} = (\xi_j)_{j=1}^m$ be a random vector with independent mean-zero, variance one, L -subgaussian entries. Let $\Psi_{\mathbf{h}} \in \mathbb{C}^{m \times m^2}$ be the random Gabor synthesis matrix generated by $\mathbf{h} = \frac{1}{\sqrt{m}}\boldsymbol{\xi}$. If, for $s \in \mathbb{N}$ and $\delta, \eta \in (0, 1)$,*

$$m \geq c\delta^{-2}s \max\{(\log^2 s)(\log^2 m), \log(\eta^{-1})\}$$

then with probability at least $1 - \eta$ the restricted isometry constant of $\Psi_{\mathbf{h}}$ satisfies that $\delta_s \leq \delta$. The constant $c > 0$ depends only on L .

Before presenting the proof we will need several observations. First, note that for $\mathbf{x} \in \mathbb{C}^m$, $\Psi_{\mathbf{h}}\mathbf{x} = \mathbf{V}_{\mathbf{x}}\boldsymbol{\xi}$, where the $m \times m$ matrix $\mathbf{V}_{\mathbf{x}}$ is given by

$$\mathbf{V}_{\mathbf{x}} = \frac{1}{\sqrt{m}} \sum_{\lambda \in \mathbb{Z}_m^2} x_{\lambda} \boldsymbol{\pi}(\lambda).$$

It is straightforward to check that $\{\frac{1}{\sqrt{m}}\boldsymbol{\pi}(\lambda) : \lambda \in \mathbb{Z}_m^2\}$ is an orthonormal system in the space of complex $m \times m$ matrices endowed with the Frobenius norm. Therefore,

$$\mathbb{E}\|\mathbf{V}_{\mathbf{x}}\boldsymbol{\xi}\|_2^2 = \|\mathbf{V}_{\mathbf{x}}\|_F^2 = \|m^{-1/2} \sum_{\lambda \in \mathbb{Z}_m^2} x_{\lambda} \boldsymbol{\pi}(\lambda)\|_F^2 = \|\mathbf{x}\|_2^2.$$

Hence, if $T_s = \{\mathbf{x} \in \mathbb{C}^{m^2} : \|\mathbf{x}\|_2 \leq 1, \|\mathbf{x}\|_0 \leq s\}$ and $\mathcal{A} = \{\mathbf{V}_{\mathbf{x}} : \mathbf{x} \in T_s\}$, the restricted isometry constant is

$$\delta_s = \sup_{\mathbf{x} \in T_s} \left| \|\Psi_{\mathbf{h}}\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 \right| = \sup_{\mathbf{V}_{\mathbf{x}} \in \mathcal{A}} \left| \|\mathbf{V}_{\mathbf{x}}\boldsymbol{\xi}\|_2^2 - \mathbb{E}\|\mathbf{V}_{\mathbf{x}}\boldsymbol{\xi}\|_2^2 \right|.$$

Thus Theorem 3.1 applies again and we need to estimate the associated Dudley integral. Note that, as $\boldsymbol{\pi}(\lambda)$ is unitary, one has for $\mathbf{x} \in T_s$

$$\|\mathbf{V}_{\mathbf{x}}\|_{2 \rightarrow 2} \leq \frac{1}{\sqrt{m}} \sum_{\lambda \in \mathbb{Z}_m^2} |x_{\lambda}| \|\boldsymbol{\pi}(\lambda)\|_{2 \rightarrow 2} \leq \|\mathbf{x}\|_1 / \sqrt{m} \leq \sqrt{s/m} \|\mathbf{x}\|_2, \quad (5.1)$$

so the upper integration limit will be $d_{2 \rightarrow 2}(\mathcal{A}) \leq \sqrt{\frac{s}{m}}$.

Lemma 5.2 *There exists an absolute constant c such that for every $0 < u \leq \sqrt{\frac{s}{m}}$,*

$$\log N(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u) \leq cs \left(\log(em^2/s) + \log(3\sqrt{s/m}/u) \right),$$

and

$$\log N(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}, u) \leq c \frac{s \log^2 m}{mu^2}.$$

Before proving the lemma, let us recall the following simple modification of the Maurey Lemma, essentially due to Carl [10]. For convenience, a proof is provided in the appendix. Below, for a set \mathcal{U} in a vector space, $\text{conv}(\mathcal{U})$ denotes its convex hull.

Lemma 5.3 *There exists an absolute constant c for which the following holds. Let X be a normed space, consider a finite set $\mathcal{U} \subset X$ of cardinality N , and assume that for every $L \in \mathbb{N}$ and $(\mathbf{u}_1, \dots, \mathbf{u}_L) \in \mathcal{U}^L$, $\mathbb{E}_{\varepsilon} \|\sum_{j=1}^L \varepsilon_j \mathbf{u}_j\|_X \leq A\sqrt{L}$, where $(\varepsilon_j)_{j=1}^L$ denotes a Rademacher vector. Then for every $u > 0$,*

$$\log N(\text{conv}(\mathcal{U}), \|\cdot\|_X, u) \leq c(A/u)^2 \log N.$$

Proof of Lemma 5.2. Define the norm $\|\cdot\|$ on \mathbb{R}^n by $\|\mathbf{x}\| = \|\mathbf{V}_x\|_{2 \rightarrow 2}$, fix $S \subset \mathbb{Z}_m^2$ of cardinality s and put $B_S = \{\mathbf{x} \in \mathbb{C}^{m^2} : \|\mathbf{x}\|_2 \leq 1, \text{supp}(\mathbf{x}) \subset S\}$. Then, by (5.1) and a volumetric estimate,

$$N(B_S, \|\cdot\|, u) \leq N(B_S, \sqrt{s/m}\|\cdot\|_2, u) \leq \left(1 + 2\frac{\sqrt{s/m}}{u}\right)^{2s} \leq \left(3\frac{\sqrt{s/m}}{u}\right)^{2s},$$

where the last step uses that $u \leq \sqrt{s/m}$. Since there are at most $\binom{m^2}{s} \leq (em^2/s)^s$ such subsets S of \mathbb{Z}_m^2 , the first part of the claim follows.

To prove the second part, note that $T_s \subset \sqrt{2s}(\text{conv}(e_\lambda, ie_\lambda, -e_\lambda, -ie_\lambda)_{\lambda \in \mathbb{Z}_m^2})$. Consider $(\mathbf{u}_j)_{j=1}^L$ selected from the extreme points (with possible repetitions). Then, by the non-commutative Khintchine inequality, due to Lust-Picard, [28, 29, 40],

$$\mathbb{E}_\varepsilon \left\| \sum_{j=1}^L \varepsilon_j \mathbf{V}_{\mathbf{u}_j} \right\|_{2 \rightarrow 2} \lesssim \sqrt{\log m} \max \left\{ \left\| \sum_{j=1}^L \mathbf{V}_{\mathbf{u}_j} \mathbf{V}_{\mathbf{u}_j}^* \right\|_{2 \rightarrow 2}, \left\| \sum_{j=1}^L \mathbf{V}_{\mathbf{u}_j}^* \mathbf{V}_{\mathbf{u}_j} \right\|_{2 \rightarrow 2} \right\}^{1/2}.$$

Recall that for every such \mathbf{u}_j , $\mathbf{V}_{\mathbf{u}_j} = \alpha \boldsymbol{\pi}(\lambda)$ with $|\alpha| = \sqrt{2s/m}$. Therefore, $\mathbf{V}_{\mathbf{u}_j}^* \mathbf{V}_{\mathbf{u}_j} = \mathbf{V}_{\mathbf{u}_j} \mathbf{V}_{\mathbf{u}_j}^* = (2s/m)\mathbf{I}$ and thus

$$\mathbb{E}_\varepsilon \left\| \sum_{j=1}^L \varepsilon_j \mathbf{V}_{\mathbf{u}_j} \right\|_{2 \rightarrow 2} \lesssim \sqrt{\log m} \sqrt{s/m} \sqrt{L}.$$

Applying Lemma 5.3 for $A \sim \sqrt{s/m} \sqrt{\log m}$, it follows that

$$\log N(T_s, \|\cdot\|, u) \lesssim (A/u)^2 \log(m^2) \lesssim \frac{s \log^2 m}{mu^2}.$$

■

Proof of Theorem 5.1. The proof follows an identical path to that of Theorem 4.1. First, as was noted above, $d_F(\mathcal{A}) \leq 1$ and $d_{2 \rightarrow 2}(\mathcal{A}) \leq \sqrt{s/m}$. Also, using the bound (2.1) by the Dudley type integral and by a direct application of Lemma 5.2,

$$\gamma_2(\mathcal{A}, \|\cdot\|_{2 \rightarrow 2}) \lesssim \int_0^{d_{2 \rightarrow 2}(\mathcal{A})} \sqrt{\log N(\mathcal{A}, \|\cdot\|, u)} du \lesssim \sqrt{s/m} (\log s) (\log m).$$

Here we used the first bound of Lemma 5.2 for $u \lesssim m^{-1/2}$, the second bound for $u \gtrsim m^{-1/2}$. The claim is now a direct application of Theorem 3.1. ■

Remark 5.4 *The only properties of the system $\{\boldsymbol{\pi}(\lambda) : \lambda \in \mathbb{Z}_m^2\}$ that have been used in the proof are the facts that all $\boldsymbol{\pi}(\lambda)$ are unitary and that $\{m^{-1/2}\boldsymbol{\pi}(\lambda) : \lambda \in \mathbb{Z}_m^2\}$ is an orthonormal system with respect to the Frobenius inner product. Therefore, Theorem 5.1 also holds true for general systems of operators with these two properties.*

A Appendix

A.1 Proof of Theorem 2.3

Without loss of generality assume that T is finite. Fix an admissible sequence (T_r) of T , let $\pi_r(\mathbf{t}) \in T_r$, for $\mathbf{t} \in T$, be an element in T_r with the smallest ℓ_2 -distance to \mathbf{t} , and choose

ℓ for which $2^{\ell-1} \leq 2p \leq 2^\ell$. Since one may assume that $\pi_r(\mathbf{t}) = \mathbf{t}$ for a sufficiently large r , one has

$$\sup_{\mathbf{t} \in T} |\langle \pi_\ell(\mathbf{t}), Y \rangle| \leq \sup_{\mathbf{t} \in T} |\langle \pi_\ell(\mathbf{t}), Y \rangle| + \sup_{\mathbf{t} \in T} \sum_{r=\ell}^{\infty} |\langle \pi_{r+1}(\mathbf{t}) - \pi_r(\mathbf{t}), Y \rangle|. \quad (\text{A.1})$$

The p -th moment of the first term can be estimated as

$$\begin{aligned} \left(\mathbb{E} \sup_{\mathbf{t} \in T} |\langle \pi_\ell(\mathbf{t}), Y \rangle|^p \right)^{1/p} &\leq \left(\mathbb{E} \sum_{\mathbf{t} \in T_\ell} |\langle \mathbf{t}, Y \rangle|^p \right)^{1/p} \leq (|T_\ell|)^{1/p} \sup_{\mathbf{t} \in T_\ell} (\mathbb{E} |\langle \mathbf{t}, Y \rangle|^p)^{1/p} \\ &\leq (2^{2^\ell})^{1/p} \sup_{\mathbf{t} \in T} (\mathbb{E} |\langle \mathbf{t}, Y \rangle|^p)^{1/p} \leq 16 \sup_{\mathbf{t} \in T} (\mathbb{E} |\langle \mathbf{t}, Y \rangle|^p)^{1/p}, \end{aligned}$$

where the last inequality follows from the choice of p .

Since $\boldsymbol{\xi}$ is an L -subgaussian vector, one obtains for the second term in (A.1)

$$\begin{aligned} &\mathbb{P} \left(\sup_{\mathbf{t} \in T} \sum_{r=\ell}^{\infty} |\langle \pi_{r+1}(\mathbf{t}) - \pi_r(\mathbf{t}), Y \rangle| \geq uL \sum_{r=\ell}^{\infty} 2^{r/2} \|(\langle \pi_{r+1}(\mathbf{t}) - \pi_r(\mathbf{t}), \mathbf{x}_j \rangle)_{j=1}^n\|_2 \right) \\ &\leq \sum_{r=\ell}^{\infty} \sum_{\mathbf{t} \in T_{r+1}} \sum_{\mathbf{t}' \in T_r} \mathbb{P} \left(\left| \sum_{j=1}^m \xi_j(\mathbf{t} - \mathbf{t}', \mathbf{x}_j) \right| \geq uL 2^{r/2} \|(\langle \mathbf{t} - \mathbf{t}', \mathbf{x}_j \rangle)_{j=1}^n\|_2 \right) \\ &\leq \sum_{r=\ell}^{\infty} 2^{2^{r+1}} \cdot 2^{2^r} \exp(-2^r u^2/2) \leq 2 \exp(-2^\ell u^2/4) \leq 2 \exp(-pu^2/2), \end{aligned} \quad (\text{A.2})$$

when $u \geq c$ for an appropriate choice of c (independent of ℓ). Therefore, by integration,

$$\begin{aligned} \left(\mathbb{E} \sup_{\mathbf{t} \in T} \sum_{r=\ell}^{\infty} |\langle \pi_{r+1}(\mathbf{t}) - \pi_r(\mathbf{t}), Y \rangle|^p \right)^{1/p} &\lesssim_L \sum_{r=\ell}^{\infty} 2^{r/2} \|(\langle \pi_{r+1}(\mathbf{t}) - \pi_r(\mathbf{t}), \mathbf{x}_j \rangle)_{j=1}^m\|_2 \\ &\lesssim_L \gamma_2(T', \|\cdot\|_2) \end{aligned}$$

where $T' = \{(\langle \mathbf{t}, \mathbf{x}_j \rangle)_{j=1}^n : \mathbf{t} \in T\}$. By the majorizing measures theorem,

$$\gamma_2(T', \|\cdot\|_2) \lesssim \mathbb{E} \sup_{\mathbf{z} \in T'} \left| \sum_{j=1}^m z_j g_j \right| = \mathbb{E} \sup_{\mathbf{t} \in T} \left| \sum_{j=1}^m g_j \langle \mathbf{t}, \mathbf{x}_j \rangle \right| = \mathbb{E} \sup_{\mathbf{t} \in T} |\langle \mathbf{t}, G \rangle|,$$

which yields the claim.

A.2 Proof of Lemma 5.3

If $\mathbf{x} \in \text{conv}(\mathcal{U})$ then $\mathbf{x} = \sum_{j=1}^N \theta_j \mathbf{u}_j$ with $\theta_j \geq 0$, $\sum_{j=1}^N \theta_j = 1$. Let $\mathbf{Z} \in X$ be a random vector which takes the value \mathbf{u}_j with probability θ_j for $j = 1, \dots, N$ and thus satisfies $\mathbb{E}\mathbf{Z} = \mathbf{x}$. Let L be a number to be determined later, set $\mathbf{Z}_1, \dots, \mathbf{Z}_L$ be independent copies of \mathbf{Z} , and put

$$\mathbf{Y} = \frac{1}{L} \sum_{j=1}^L \mathbf{Z}_j.$$

If $(\epsilon_j)_{j=1}^m$ is a Rademacher sequence independent of (\mathbf{Z}_j) then by a standard symmetrization argument (see, e.g., [27]) and because $(\mathbf{Z}_j)_{j=1}^L$ ranges over \mathcal{U}^L

$$\mathbb{E} \|\mathbf{x} - \mathbf{Y}\|_X = \frac{1}{L} \mathbb{E} \left\| \sum_{j=1}^L (\mathbf{x} - \mathbf{Z}_j) \right\|_X \leq \frac{2}{L} \mathbb{E} \left\| \sum_{j=1}^L \epsilon_j \mathbf{Z}_j \right\|_X \leq 2A/\sqrt{L}.$$

Thus, for $L \sim (A/u)^2$ there exists a realization $\mathbf{y} = \frac{1}{L} \sum_{j=1}^L \mathbf{z}_\ell$, for some $\mathbf{z}_\ell \in \mathcal{U}$, of \mathbf{Y} for which

$$\|\mathbf{x} - \mathbf{y}\|_X \leq u.$$

As this argument applies for any $\mathbf{x} \in \text{conv}(\mathcal{U})$, any such \mathbf{x} can be approximated by some \mathbf{y} of this form. Since \mathbf{y} can assume at most N^L different values, this yields

$$\log N(\text{conv}(\mathcal{U}), \|\cdot\|_X, u) \leq L \log N \leq c(A/u)^2 \log N,$$

as claimed.

References

- [1] M. Arcones and E. Giné. On decoupling, series expansions, and tail behavior of chaos processes. *J. Theoret. Probab.*, 6(1):101–122, 1993.
- [2] W. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak. Compressed channel sensing: A new approach to estimating sparse multipath channels. *Proc. IEEE*, 98(6):1058–1076, June 2010.
- [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] T. Blumensath and M. Davies. Iterative hard thresholding for compressed sensing. *Appl. Comput. Harmon. Anal.*, 27(3):265–274, 2009.
- [5] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, 31(3):1583–1614, 2003.
- [6] J. Bourgain and L. Tzafriri. Invertibility of ‘large’ submatrices with applications to the geometry of Banach spaces and harmonic analysis. *Israel J. Math.*, 57(2):137–224, 1987.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] S. S. Chen, David L. Donoho, and M. A. Saunders. Atomic decomposition by Basis Pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1999.
- [12] V. de la Peña and Evarist Giné. *Decoupling. From Dependence to Independence*. Probability and its Applications (New York). Springer-Verlag, New York, 1999.
- [13] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
- [14] R. Dudley. *Uniform central limit theorems*, volume 63 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 1999.
- [15] R. M. Dudley. The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *J. Functional Analysis*, 1:290–330, 1967.

- [16] H. G. Feichtinger, F. Luef, and T. Werther. A guided tour from linear algebra to the foundations of Gabor analysis. In *Gabor and Wavelet Frames*, volume 10 of *Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap.*, pages 1–49. World Sci. Publ., Hackensack, 2007.
- [17] X. Fernique. Régularité des trajectoires des fonctions aléatoires gaussiennes. In *École d'Été de Probabilités de Saint-Flour, IV-1974*, pages 1–96. Lecture Notes in Math., Vol. 480. Springer, 1975.
- [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. Sparse recovery algorithms: sufficient conditions in terms of restricted isometry constants. In *Proceedings of the 13th International Conference on Approximation Theory*, 2010.
- [20] S. Foucart. Hard thresholding pursuit: an algorithm for compressive sensing. *SIAM J. Numer. Anal.*, 49(6):2543–2563, 2011.
- [21] K. Gröchenig. *Foundations of Time-Frequency Analysis*. Appl. Numer. Harmon. Anal. Birkhäuser Boston, 2001.
- [22] J. Haupt, W. Bajwa, G. Raz, and R. D. 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] A. Hinrichs and J. Vybíral. Johnson-Lindenstrauss lemma for circulant matrices. *Random Struct. Algorithms*, 39(3):391–398, 2011.
- [25] F. Kraher, G. E. Pfander, and P. Rashkov. Uncertainty principles for time–frequency representations on finite abelian groups. *Appl. Comput. Harmon. Anal.*, 25(2):209–225, 2008.
- [26] F. Kraher and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.*, 43(3):1269–1281, 2011.
- [27] M. Ledoux and M. Talagrand. *Probability in Banach Spaces*. Springer-Verlag, Berlin, Heidelberg, New York, 1991.
- [28] F. Lust-Piquard. Inégalités de Khintchine dans C_p ($1 < p < \infty$). *C. R. Acad. Sci. Paris Sér. I Math.*, 303:289–292, 1986.
- [29] F. Lust Piquard and G. Pisier. Noncommutative Khintchine and Paley inequalities. *Ark. Mat.*, 29(2):241–260, 1991.
- [30] S. Mendelson, A. Pajor, and N. Tomczak Jaegermann. Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geom. Funct. Anal.*, 17(4):1248–1282, 2007.
- [31] Q. Mo and S. Li. New bounds on the restricted isometry constant δ_{2k} . *Appl. Comput. Harmon. Anal.*, 31(3):460–468, 2011.
- [32] G. E. Pfander and H. Rauhut. Sparsity in time-frequency representations. *J. Fourier Anal. Appl.*, 16(2):233–260, 2010.
- [33] G. E. Pfander, H. Rauhut, and J. Tanner. Identification of matrices having a sparse representation. *IEEE Trans. Signal Process.*, 56(11):5376–5388, 2008.
- [34] G. E. Pfander, H. Rauhut, and J. A. Tropp. The restricted isometry property for time-frequency structured random matrices. *Probab. Theory Relat. Fields*, to appear.

- [35] G. Pisier. Some applications of the metric entropy condition to harmonic analysis. In *Banach spaces, harmonic analysis, and probability theory (Storrs, Conn., 1980/1981)*, volume 995 of *Lecture Notes in Math.*, pages 123–154. Springer, 1983.
- [36] H. Rauhut. Circulant and Toeplitz matrices in compressed sensing. In *Proc. SPARS'09*, Saint-Malo, France, 2009.
- [37] 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.
- [38] H. Rauhut, J. K. Romberg, and J. A. Tropp. Restricted isometries for partial random circulant matrices. *Appl. Comput. Harmon. Anal.*, 32(2):242–254, 2012.
- [39] J. K. Romberg. Compressive sensing by random convolution. *SIAM J. Imaging Sci.*, 2(4):1098–1128, 2009.
- [40] M. Rudelson. Random vectors in the isotropic position. *J. Funct. Anal.*, 164(1):60–72, 1999.
- [41] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.
- [42] M. Talagrand. Regularity of Gaussian processes. *Acta Math.*, 159(1-2):99–149, 1987.
- [43] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.
- [44] M. Talagrand. *The Generic Chaining*. Springer Monographs in Mathematics. Springer-Verlag, 2005.
- [45] J. Tropp and D. Needell. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.*, 26(3):301–321, 2008.
- [46] J. Vybíral. A variant of the Johnson-Lindenstrauss lemma for circulant matrices. *J. Funct. Anal.*, 260(4):1096–1105, 2011.
- [47] T. Zhang. Sparse recovery with orthogonal matching pursuit under RIP. *IEEE Trans. Inform. Theory*, 57(9):6215 – 6221, 2011.