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 signifi-
cantly improved estimates for the restricted isometry constants of partial random circulant
matrices and time-frequency structured random matrices. In both cases the required con-
dition 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 ma-
trices, chaos processes, $\gamma_2$-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 trig-
gered various applications in signal and image processing.

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

$$y = \Phi x.$$ 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 $x$. On the other hand, if it
is known a priori that $x$ is $s$-sparse then the situation changes. And, although the naive
approach for reconstruction, namely, $\ell_0$-minimization,

$$\min \|z\|_0 \quad \text{subject to } \Phi z = y$$
is NP-hard in general, there are several tractable alternatives – for instance, $\ell_1$-minimization
[11, 13, 7]

$$\min \|z\|_1 \quad \text{subject to } \Phi z = y,$$
(where $\|z\|_p$ denotes the usual $\ell_p$-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 $\delta_s$ is defined as the smallest number such that

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

One may show that under conditions of the form $\delta_{ks} \leq \delta^*$ for some $\delta^* < 1$ and some appropriate small integer $\kappa$, a variety of recovery algorithms reconstruct every $s$-sparse $x$ from $y = \Phi x$. Among these are $\ell_1$-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 $\varepsilon_{jk}$ 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 $\Phi$. Indeed, certain applications impose constraints on the matrix and recovery algorithms can be accelerated when fast matrix vector multiplication routines are available for $\Phi$. 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 $x, z \in \mathbb{C}^n$ by

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

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

We are interested in sparse recovery from subsampled convolutions with a random vector. Formally, let $\Omega \subset \{1, \ldots, n\}$ be an arbitrary (fixed) set of cardinality $m$, and denote by $R_\Omega : \mathbb{C}^n \to \mathbb{C}^m$ the operator that restricts a vector $x \in \mathbb{C}^n$ to its entries in $\Omega$. Let $\varepsilon = (\varepsilon_i)^{n}_{i=1}$ 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 $\Phi = m^{-1/2} R_\Omega H_\epsilon \in \mathbb{R}^{m \times n}$ and acts on vectors $x \in \mathbb{C}^n$ via

$$
\Phi x = \frac{1}{\sqrt{m}} R_\Omega (\epsilon \ast x).
$$

In other words, $\Phi$ is a circulant matrix generated by a Rademacher vector, where the rows outside $\Omega$ are removed. Our first main result establishes the restricted isometry property of $\Phi$ 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 $\epsilon$. If

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

then with probability at least $1 - n^{-{(\log n)}(\log^2 s)}$, the restricted isometry constant of $\Phi$ 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 $\ell_1$-minimization using a draw of a partial random circulant matrix. The condition derived there is $m \geq C \delta \log^2 n$, 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 $\ell_1$-minimization and other recovery methods for a single matrix $\Phi$.

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 $m \geq C \delta \log^6 n$. In contrast, our result holds for an arbitrary fixed selection of a set $\Omega \subset \{1, \ldots, n\}$, which is important in applications since in many practical problems, it is natural or desired to consider structured sampling sets such as $\Omega = \{L, 2L, 3L, \ldots, mL\}$ 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 $\eta$. Let $\Phi \in \mathbb{C}^{m \times n}$ be a partial circulant matrix generated by a Rademacher vector $\epsilon$. Furthermore, let $\epsilon' \in \mathbb{R}^n$ be a Rademacher vector independent of $\epsilon$ and set $D_{\epsilon'}$ to be the diagonal matrix with diagonal $\epsilon'$. Then with probability exceeding $1 - \eta$, for every $x \in E$,

$$
(1 - \delta)\|x\|_2^2 \leq \|\Phi D_{\epsilon'} x\|_2^2 \leq (1 + \delta)\|x\|_2^2.
$$
1.3 Time-frequency structured random matrices

The translation and modulation operators on $\mathbb{C}^m$ are defined by $(Th)_j = h_{j \oplus 1}$ and $(Mh)_j = e^{2\pi ij/m}h_j = \omega^j h_j$, where $\omega = e^{2\pi i/m}$ and $\oplus$ again denotes cyclic subtraction, this time modulo $m$. Observe that $(Th)_j = h_{j \oplus k}$ and $(Mh)_j = e^{2\pi ij/n}h_j = \omega^j h_j$.

The time-frequency shifts are given by $\pi(\lambda) = M^\ell T^k$, $\lambda = (k, \ell) \in \mathbb{Z}_m^2 = \{0, \ldots, m-1\}^2$.

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

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

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 $\varepsilon$ be a Rademacher vector and consider the Gabor synthesis matrix $\Psi_h \in \mathbb{C}^{m \times m^2}$ generated by $h = \frac{1}{\sqrt{m}} \varepsilon$. If

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

then with probability at least $1 - m^{-\left(\log m\right)\left(\log^2 s\right)}$, the restricted isometry constant of $\Psi_h$ 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 $\ell_1$-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 $\delta_s$ can be written as a random variable $X$ of the form

$$X = \sup_{A \in \mathcal{A}} \left\| A\varepsilon \right\|_2^2 - \mathbb{E}\left[\left\| A\varepsilon \right\|_2^2\right],$$

where $\mathcal{A}$ is a set of matrices and $\varepsilon$ 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 $d_{2\to 2}(\mathcal{A})$, is the radius of $\mathcal{A}$ in the Frobenius norm $\|A\|_F = \sqrt{tr(A^*A)}$ and the operator norm $\|A\|_{2\to 2} = \sup_{\|x\|_2 \leq 1} \|Ax\|_2$, respectively. That is, $d_F(\mathcal{A}) = \sup_{A \in \mathcal{A}} \|A\|_F$ and $d_{2\to 2}(\mathcal{A}) = \sup_{A \in \mathcal{A}} \|A\|_{2\to 2}$. For the second one, Talagrand’s functional $\gamma_2(A, \cdot \| \cdot \|_{2\to 2})$, 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 $\epsilon$ be a Rademacher vector of length $n$. Then

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

Furthermore, for $t > 0$,

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

where

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

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

The symmetry assumption $A = -\mathcal{A}$ 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 $\ell_2$-norms we can rewrite $X$ in (1.4) as

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

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

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

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

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

(see Section 2 for the definition of the $\gamma_\alpha$-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 $\gamma_1$-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 $\gamma_1$-functional but only the $\gamma_2$-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 $\pm 1$ 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 \(s \geq 1\), \(|T_s| \leq 2^{2s}\) and \(|T_0| = 1\). For \(\beta \geq 1\), define the \(\gamma_\beta\) 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 \(\gamma_\alpha\)-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 \(\gamma_2\)-functional of a set of matrices \(A\) equipped with the operator norm, the scenario which we will focus on in this article, is

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

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 \(T \subset L_2\), \(\gamma_2\) 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 $\theta \in S^{n-1}$, $E\langle X, \theta \rangle^2 = 1$. A random vector $X$ is called $L$-subgaussian if it is isotropic and $P(|\langle X, \theta \rangle| \geq t) \leq 2\exp(-t^2/2L^2)$ 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

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

Assume that a random vector $\xi$ has independent coordinates $\xi_i$, each of which is an $L$-subgaussian random variable of mean zero and variance one. One may verify by direct computation that $\xi$ 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 $x_1, \ldots, x_n \in \mathbb{C}^N$ and $T \subset \mathbb{C}^N$. If $\xi$ is an isotropic, $L$-subgaussian random vector and $Y = \sum_{j=1}^n \xi_j x_j$, then for every $p \geq 1$,

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

where $c$ is a constant which depends only on $L$ and $G = \sum_{j=1}^N g_j x_j$ for $g_1, \ldots, g_N$ 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

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

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 $\xi = (\xi_1, \ldots, \xi_n)$ be a sequence of independent, centered random variables, and let $F$ be a convex function. If $B$ is a collection of matrices and $\xi'$ is an independent copy of $\xi$, then

\[
\mathbb{E} \sup_{B \in B} F \left( \sum_{j,k=1}^{n} \xi_j \xi_k B_{j,k} \right) \leq \mathbb{E} \sup_{B \in B} F \left( 4 \sum_{j,k=1}^{n} \xi'_j \xi'_k B_{j,k} \right). \tag{2.5}
\]

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 $g = (g_1, \ldots, g_n)$ be a sequence of independent standard normal random variables. If $B$ is a collection of Hermitian matrices and $g'$ is an independent copy of $g$, then

\[
\mathbb{E} \sup_{B \in B} \left| \sum_{j,k=1}^{n} g_j g_k B_{j,k} + \sum_{j=1}^{n} (g_j^2 - 1) B_{j,j} \right|^p \leq C^p \mathbb{E} \sup_{B \in B} \left| \sum_{j,k=1}^{n} g'_j g'_k B_{j,k} \right|^p. \tag{2.6}
\]

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 $c_1, c_2, \ldots$: 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 $e_j$ and $B_2^n$ is the unit $\ell_2$-ball in $\mathbb{C}^n$.

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

\[
N_A(\xi) := \sup_{A \in A} \|A\xi\|_2, \quad B_A(\xi) := \sup_{A \in A} \left| \sum_{j,k=1}^{n} \xi_j \overline{\xi}_k \langle A^j, A^k \rangle \right|,
\]

\[
D_A(\xi) := \sup_{A \in A} \left| \sum_{j=1}^{n} (|\xi_j|^2 - \mathbb{E}|\xi_j|^2) \|A^j\|^2 \right|, \quad \text{and} \quad C_A(\xi) := \sup_{A \in A} \|A\xi\|_2^2 - \mathbb{E}\|A\xi\|_2^2.
\]
3 Chaos processes

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

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

\[
E = \gamma_2(A, \| \cdot \|_{2 \to 2}) (\gamma_2(A, \| \cdot \|_{2 \to 2}) + d_F(A)) + d_F(A) d_{2 \to 2}(A),
\]

\[
V = d_{2 \to 2}(A) (\gamma_2(A, \| \cdot \|_{2 \to 2}) + d_F(A)), \quad \text{and} \quad U = d^2_{2 \to 2}(A).
\]

Then, for $t > 0$,

\[
\mathbb{P}(C_A(\xi) \geq c_1 E + t) \leq 2 \exp \left( -c_2 \min \left\{ \frac{t^2}{V^2}, \frac{t}{U} \right\} \right).
\]

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 $A = -A$ ensures that $d_{2 \to 2}(A) \leq \gamma_2(A, \| \cdot \|_{2 \to 2})$. 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 $N_A$ and $C_A$, followed by applying Proposition 2.6. The first step is a bound on the moments of a decoupled version of $N_A$.

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

\[
\left\| \sup_{A \in A} \langle A\xi, A\xi' \rangle \right\|_{L_p} \lesssim_L \gamma_2(A, \| \cdot \|_{2 \to 2}) \| N_A(\xi) \|_{L_p} + \sup_{A \in A} \| \langle A\xi, A\xi' \rangle \|_{L_p}.
\]

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

Note that for every $A \in A$,

\[
\left| \langle A\xi, A\xi' \rangle \right| - \left| \langle (\pi_\ell A) \xi, (\pi_\ell A) \xi' \rangle \right| \leq \sum_{r=0}^{r_0-1} \left| \langle (\Delta_{r+1} A) \xi, (\pi_{r+1} A) \xi' \rangle \right| + \sum_{r=\ell}^{r_0-1} \left| \langle (\pi_r A) \xi, (\Delta_{r+1} A) \xi' \rangle \right|
\]

Furthermore, conditionally on $\xi'$,

\[
\langle (\Delta_{r+1} A) \xi, (\pi_{r+1} A) \xi' \rangle = \langle \xi, (\Delta_{r+1} A)^* (\pi_{r} A) \xi' \rangle
\]

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

\[
\mathbb{P}_\xi \left( \left| \langle (\xi, (\Delta_{r+1} A)^* (\pi_{r+1} A) \xi' \rangle \right| \geq u \| (\Delta_{r+1} A)^* (\pi_{r+1} A) \xi' \|_2 \right) \leq 2 \exp(-u^2/2). \tag{3.1}
\]

Recall that $|\{ \pi_r A : A \in A \}| = |T_r| \leq 2^r$, so there are at most $2^{r+2}$ possible values that $\Delta_r(A)^* \pi_{r+1}(A)$ 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 $\xi$-probability at least $1 - 2 \exp(-c_2 t^2)$ one has for every $\ell < r \leq r_0$ and every $A \in A$

$$|\langle (\Delta r+1)A, \xi, (\pi r+1)A\rangle| \leq t2^{r/2}||\Delta r+1 A||r ||(\Delta r+1)A||2. \tag{3.2}$$

Let $E_i(\xi')$ be the event for which (3.2) holds for all the possible choices of $r$ and $A$ as above. Since

$$\| (\Delta r+1 A)^* (\pi r+1 A)\xi' \|_2 \leq \|\Delta r+1 A\|_{2,2} \sup_{A \in A} \|A\xi'\|_2 = \|\Delta r+1 A\|_{2,2} N_A(\xi'),$$

one has on $E_i(\xi')$

$$S_1 := \sum_{r=\ell}^{r_0-1} |\langle (\Delta r+1)A, \xi, (\pi r+1)A\rangle| \leq \sum_{r=\ell}^{r_0-1} t2^{r/2} ||\Delta r+1 A||_{2,2} N_A(\xi') \leq t\gamma_2(A, \|\cdot\|_{2,2}) N_A(\xi').$$

We will now estimate

$$\|S_1\|_{L_p} = E \xi E \xi' S_{1}^{\xi} = E \xi \int_{0}^{\infty} pt^{p-1} P_\xi(S_1 > t|\xi') dt. \tag{3.3}$$

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

$$\int_{0}^{\infty} pt^{p-1} P_\xi(S_1 > t|\xi') dt \leq c_3^p W(\xi')^p + \int_{c_4^p W(\xi')}^{\infty} pt^{p-1} P_\xi(S_1 > t|\xi') dt \leq c_3^p W(\xi')^p + W(\xi')^p \int_{c_4^p}^{\infty} pu^{p-1} P_\xi(S_1 > u W(\xi')|\xi') du \leq c_4^p W(\xi')^p,$$

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

$$P_\xi(S_1 > u W(\xi')|\xi') \leq P_\xi(E_u(\xi')|\xi') \leq 2 \exp(-c_2 u^2 2^\ell) \leq 2 \exp(-c_2 u^2 p/2).$$

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

$$\|S_1 + S_2\|_{L_p} \leq c_5(L) \gamma_2(A, \|\cdot\|_{2,2}) N_A(\xi)||_p.$$

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

$$E \sup_{A \in A} |\langle (\pi r A, (\pi r A)\xi')| | \leq \sum_{A \in T_\ell} E|\langle A\xi, A\xi' | | \leq \exp(p) \sup_{A \in A} E|\langle A\xi, A\xi' | |.$$

Thus $\sup_{A \in A} |\langle (\pi r A, (\pi r A)\xi')| ||_{L_p} \leq c \sup_{A \in A} ||\langle A\xi, A\xi' | ||_{L_p}$, which completes the proof.

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

**Theorem 3.4** Let $L \geq 1$ and $\xi = (\xi_j)_{j=1}^n$, where $\xi_j$, $j = 1, \ldots, n$, are independent mean-zero, variance one, $L$-subgaussian random variables, and let $A$ be a class of matrices. Then for every $p \geq 1$,
(a) 
\[ \|N_\mathcal{A}(\xi)\|_{L_p} \lesssim L \gamma_2(\mathcal{A}, \| \cdot \|_{2 \to 2}) + d_F(\mathcal{A}) + \sqrt{pd}_{2 \to 2}(\mathcal{A}), \]

(b) 
\[ \|C_\mathcal{A}(\xi)\|_{L_p} \lesssim L \gamma_2(\mathcal{A}, \| \cdot \|_{2 \to 2}) (\gamma_2(\mathcal{A}, \| \cdot \|_{2 \to 2}) + d_F(\mathcal{A})) + \sqrt{pd}_{2 \to 2}(\mathcal{A}) (\gamma_2(\mathcal{A}, \| \cdot \|_{2 \to 2}) + d_F(\mathcal{A})) + pd^2_{2 \to 2}(\mathcal{A}). \]

**Proof.** We start by showing that if \( \xi' \) is an independent copy of \( \xi \), then
\[ \sup_{A \in \mathcal{A}} \| \langle A \xi, A \xi' \rangle \|_{L_p} \lesssim_L \sqrt{pd_F(A)}d_{2 \to 2}(\mathcal{A}) + pd^2_{2 \to 2}(\mathcal{A}). \] (3.4)

Indeed, fix \( A \in \mathcal{A} \) and set \( S = \{ A^* Ax : x \in B^2_p \} \). Since the random vector \( \xi \) is \( L \)-subgaussian, the random variable \( \langle \xi, A^* A \xi' \rangle \) is subgaussian conditionally on \( \xi' \). Therefore, by (2.3),
\[ \| \langle A \xi, A \xi' \rangle \|_{L_p} = \left( \mathbb{E}_{\xi'} \left( \mathbb{E}_{\xi} \| \langle \xi, A^* A \xi' \rangle \|_p \right)^{1/p} \right)^{1/p} \lesssim (\mathbb{E}_{\xi'} L \sqrt{p} \| A^* A \xi' \|_2)^{1/p} \]
\[ = L \sqrt{p} \left( \mathbb{E}_{\xi'} \sup_{y \in S} |\langle y, \xi' \rangle| \right)^{1/p}. \]

Note that for a standard Gaussian vector \( g \),
\[ \mathbb{E} \sup_{y \in S} |\langle y, g \rangle| = \mathbb{E} \| A^* A g \|_2 \leq (\mathbb{E} \| A^* A g \|_2^2)^{1/2} = \| A^* A \|_F \leq \| A \|_F \| A \|_{2 \to 2}. \]

Also, if \( \xi' = \sum_{j=1}^n \xi_j e_j \) then, applying (2.3) again,
\[ \sup_{y \in S} (\mathbb{E} |\langle y, \xi' \rangle|^p)^{1/p} = \sup_{z \in B^2_p} (\mathbb{E} |\langle A^* A z, \xi' \rangle|^p)^{1/p} \lesssim_L L \sup_{z \in B^2_p} \sqrt{p} \| A^* A z \|_2 = L \sqrt{p} \| A \|_{2 \to 2}. \]

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

As a preliminary step, we will consider the special case of the Gaussian vector \( g \) and show that \( \mathbb{E} N_\mathcal{A}(g) \lesssim L \gamma_2(\mathcal{A}, \| \cdot \|_{2 \to 2}) + d_F(\mathcal{A}) \), which is (a) for \( p = 1 \). To that end observe that by Theorem 2.5 and Lemma 3.3,
\[ \|C_\mathcal{A}(g)\|_{L_p} = \left( \sup_{A \in \mathcal{A}} \left\| \sum_{j,k \neq k} g_j g_k^* \langle A^j, A^k \rangle + \sum_j (g_j^2 - 1) \| A^j \|^2 \right\|_{L_p} \right)^{1/p} \]
\[ \lesssim L \gamma_2(\mathcal{A}, \| \cdot \|_{2 \to 2}) \| N_\mathcal{A}(g) \|_{L_p} + \sup_{A \in \mathcal{A}} \| \langle A g, A g' \rangle \|_{L_p}. \] (3.5)

Combining (3.4) with (3.5), it follows that
\[ \|C_\mathcal{A}(g)\|_{L_p} \lesssim L \gamma_2(\mathcal{A}, \| \cdot \|_{2 \to 2}) \| N_\mathcal{A}(g) \|_{L_p} + \sqrt{pd_F(\mathcal{A})d_{2 \to 2}(\mathcal{A}) + pd^2_{2 \to 2}(\mathcal{A}).} \] (3.6)

Specifying \( p = 1 \) and using that \( d_F(\mathcal{A}) \geq d_{2 \to 2}(\mathcal{A}) \) as well as \( \mathbb{E} \| A g \|^2 \leq \| A \|^2_F \), we conclude
\[ \mathbb{E} N_\mathcal{A}^2(g) \leq \mathbb{E} C_\mathcal{A}(g) + d^2_F(\mathcal{A}) \lesssim L \gamma_2(\mathcal{A}, \| \cdot \|_{2 \to 2}) \mathbb{E} N_\mathcal{A}(g) + d^2_F(\mathcal{A}). \]
Therefore,
\[ \mathbb{E}N_{\mathcal{A}}(g) \leq (\mathbb{E}N_{\mathcal{A}}^2(g))^{1/2} \leq L \gamma_2(\mathcal{A}, \| \cdot \|_2) + d_F(\mathcal{A}), \]
as desired.

Finally, for general \( p \) and arbitrary subgaussian vectors, we apply Theorem 2.3 with the set \( S = \{ A^* x : x \in B_2^p, \ A \in \mathcal{A} \} \). Since \( \xi \) is \( L \)-subgaussian we obtain
\[
\| N_{\mathcal{A}}(\xi) \|_{L_p} = \mathbb{E} \sup_{A \in \mathcal{A}, x \in B_2^p} |\langle A \xi, x \rangle|^p \leq \mathbb{E} \sup_{u \in S} |\langle \xi, u \rangle|^p \leq \mathbb{E} N_{\mathcal{A}}(g) + \sqrt{p} \sup_{A \in \mathcal{A}, x \in B_2^p} \| A^* x \|_2
\]
\[
\leq L \mathbb{E} N_{\mathcal{A}}(g) + \sqrt{pd_{2 \to 2}(A)},
\]
which proves \( \text{(a)} \).

For \( \text{(b)} \), observe that as the \( \xi_j \) are unit variance, we have \( \mathbb{E} \| A \xi \|_2^2 = \| A \|_F^2 = \sum_{j=1}^n \| A^j \|_2^2 \)
and, consequently, \( C_{\mathcal{A}}(\xi) \) can be split up into the diagonal and the off-diagonal contributions as follows.
\[
C_{\mathcal{A}}(\xi) = \sup_{A \in \mathcal{A}} \left| \sum_{j,k=1}^n \xi_j \xi_k \langle A^j, A^k \rangle + \sum_{j=1}^n (|\xi_j|^2 - 1) \| A^j \|_2^2 \right| \leq B_{\mathcal{A}}(\xi) + D_{\mathcal{A}}(\xi).
\]
Hence it suffices to estimate the moments of \( B_{\mathcal{A}}(\xi) \) and \( D_{\mathcal{A}}(\xi) \); one concludes using the triangle inequality.

For the off-diagonal term, we use Theorem 2.4 and Lemma 3.3 to bound
\[
\| B_{\mathcal{A}}(\xi) \|_{L_p} \leq 4 \mathbb{E} \left\| \sum_{j,k} \xi_j \xi_k \langle A^j, A^k \rangle \right\|_{L_p} = 4 \sup_{A \in \mathcal{A}} \left\| \langle A \xi, A^j \rangle \right\|_{L_p}
\]
\[
\leq L \gamma_2(\mathcal{A}, \| \cdot \|_2)(\| N_{\mathcal{A}}(\xi) \|_{L_p} + \sup_{A \in \mathcal{A}} \| \langle A \xi, A^j \rangle \|_{L_p}).
\]
Combining this estimate with (3.4) and part \( \text{(a)} \), we obtain that
\[
\| B_{\mathcal{A}}(\xi) \|_{L_p} \leq L \gamma_2(\mathcal{A}, \| \cdot \|_2) + \gamma_2(\mathcal{A}, \| \cdot \|_2) + d_F(\mathcal{A} + \sqrt{pd_{2 \to 2}(A)}
\]
\[
+ \sqrt{pd_F(\mathcal{A} + \sqrt{pd_{2 \to 2}(A)} + pd_{2 \to 2}(\mathcal{A}).
\]
(3.7)

For the diagonal term, observe that by a standard symmetrization argument (see, e.g., [27, Lemma 6.3]),
\[
\| D_{\mathcal{A}}(\xi) \|_{L_p} = \sup_{A \in \mathcal{A}} \left\| \sum_{j=1}^n (b_j^2 - \mathbb{E} b_j^2) \| A^j \|_2^2 \right\|_{L_p} \leq \sup_{A \in \mathcal{A}} \left\| \sum_{j=1}^n \epsilon_j b_j^2 \| A^j \|_2^2 \right\|_{L_p}
\]
where \( \epsilon = (\epsilon_1, \ldots, \epsilon_n) \) is a Rademacher vector independent of \( \xi \). Furthermore, let \( g = (g_1, \ldots, g_n) \) be a sequence of independent standard normal variables. Then, as \( \xi_j \) is \( L \)-subgaussian, there is an absolute constant \( c \) for which \( \mathbb{P}(b_j^2 \geq tL^2) \leq \mathbb{P}(g_j^2 \geq t) \) for every \( t > 0 \). Moreover, \( \epsilon_j b_j^2 \) and \( \epsilon_j g_j^2 \) are symmetric, so by the contraction principle (see
Remark 3.5, which, together with (3.7), proves (b).

Therefore, by Theorem 2.3 and a standard chaining argument, and de-symmetrization [27, Lemma 6.3],

\[
\|D_A(\xi)\|_{L^p} \lesssim \left\| \sup_{A \in \mathcal{A}} \sum_j \epsilon_j g_j^2 \|A^j\|_2^2 \right\|_{L^p}
\leq 2 \left\| \sup_{A \in \mathcal{A}} \sum_j (|g_j|^2 - 1)\|A^j\|_2^2 \right\|_{L^p} + \left\| \sup_{A \in \mathcal{A}} \sum_j \epsilon_j \|A^j\|_2^2 \right\|_{L^p}
= 2 \|D_A(g)\|_{L^p} + \left\| \sup_{A \in \mathcal{A}} \sum_j \epsilon_j \|A^j\|_2^2 \right\|_{L^p}.
\]

Now observe that \(D_A(g) \leq C_A(g) + B_A(g)\), and thus, by (3.6) and (3.7),

\[
\|D_A(g)\|_{L^p} \leq \|C_A(g)\|_{L^p} + \|B_A(g)\|_{L^p}
\lesssim \gamma_2(\mathcal{A}, \| \cdot \|_{2 \rightarrow 2}) (\gamma_2(\mathcal{A}, \| \cdot \|_{2 \rightarrow 2}) + d_F(\mathcal{A}))
+ \sqrt{pd_F(\mathcal{A}) (d_{2 \rightarrow 2}(\mathcal{A}) + \gamma_2(\mathcal{A}, \| \cdot \|_{2 \rightarrow 2}))} + pd_{2 \rightarrow 2}(\mathcal{A}).
\]

Finally, note that \(A \rightarrow \sum_j \epsilon_j \|A^j\|_2^2\) is a subgaussian process relative to the metric

\[
d(A, B) = \left( \sum_{j=1}^n (\|A^j\|_2^2 - \|B^j\|_2^2)^2 \right)^{1/2}
\leq \left( \sum_{j=1}^n \|A^j - B^j\|_2^2 \cdot (\|A^j\|_2 + \|B^j\|_2)^2 \right)^{1/2}
\leq 2d_F(\mathcal{A})\|A - B\|_{2 \rightarrow 2}.
\]

Therefore, by Theorem 2.3 and a standard chaining argument,

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

This shows that

\[
\|D_A(\xi)\|_{L^p} \lesssim \gamma_2(\mathcal{A}, \| \cdot \|_{2 \rightarrow 2}) (\gamma_2(\mathcal{A}, \| \cdot \|_{2 \rightarrow 2}) + d_F(\mathcal{A}))
+ \sqrt{pd_{2 \rightarrow 2}(\mathcal{A}) (\gamma_2(\mathcal{A}, \| \cdot \|_{2 \rightarrow 2}) + d_F(\mathcal{A})) + pd_{2 \rightarrow 2}(\mathcal{A}),
\]

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 \(D_A \equiv 0\), 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 \(\xi\) 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 \( \xi_i \)'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 \( V_xz = \frac{1}{\sqrt{m}}P_\Omega(x \ast z) \), where the projection operator \( P_\Omega : \mathbb{C}^n \to \mathbb{C}^n \) is given by \( P_\Omega = R_{\Omega}^*R_\Omega \), that is, \( (P_\Omega x)_\ell = x_\ell \) for all \( \ell \in \Omega \) and \( (P_\Omega x)_\ell = 0 \) for \( \ell \notin \Omega \).

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

\[
\delta_s = \sup_{x \in T_s} \| R_\Omega(\xi \ast x) \|_2^2 - \| x \|_2^2 = \sup_{x \in T_s} \| P_\Omega(\xi \ast x) \|_2^2 - \| x \|_2^2 = \sup_{x \in T_s} \| V_x \xi \|_2^2 - \| x \|_2^2.
\]

Since \( |\Omega| = m \), it follows that

\[
\mathbb{E}\| V_x \xi \|_2^2 = \frac{1}{m} \sum_{\ell \in \Omega} \sum_{k,j=1}^n \xi_k \xi_j x_{\ell \oplus j} x_{\ell \oplus k} = \frac{1}{m} \sum_{\ell \in \Omega} \sum_{k=1}^n |x_{\ell \oplus k}|^2 = \| x \|_2^2
\]

and hence

\[
\delta_s = \sup_{x \in T_s} \| V_x \xi \|_2^2 - \mathbb{E}\| V_x \xi \|_2^2,
\]

which shows that \( \delta_s \) is the process \( C_4 \) studied in the previous section for \( A = \{ V_x : x \in T_s \} \). 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 \( \eta, \delta \in (0,1) \),

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

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 \( \xi \) 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 \( \Phi \). Let \( F \) by the unnormalized Fourier transform with elements \( F_{jk} = e^{2\pi ijk/n} \). By the convolution theorem, for every \( 1 \leq j \leq n \), \( F(x \ast y)_j = (Fx)_j \cdot (Fy)_j \). Therefore,

\[
V_x \xi = \frac{1}{\sqrt{m}}P_\Omega F^{-1} \hat{X} F \xi,
\]

where \( \hat{X} = \text{diag}(Fx) \) is the diagonal matrix, whose diagonal is the Fourier transform \( Fx \).

In short,

\[
V_x = \frac{1}{\sqrt{m}} \hat{P}_\Omega X F,
\]

where \( \hat{P}_\Omega = \frac{1}{\sqrt{m}} P_\Omega F^{-1} \).

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

Since the matrices \( V_x \) consist of shifted copies of \( x \) in all of their \( m \) nonzero rows, the \( \ell_2 \)-norm of each nonzero row is \( m^{-1/2}\|x\|_2 \); thus \( \| V_x \|_F = \| x \|_2 \leq 1 \) for all \( x \in T_s \) and

\[
d_F(A) = 1.
\]
Also, observe that for every \( x \in T \) with associated diagonal matrix \( \hat{X} \),
\[
\|V_x\|_{2 \rightarrow 2} = \frac{1}{\sqrt{m}}\|P_\Omega \hat{X} F\|_{2 \rightarrow 2} \leq \sqrt{\frac{n}{m}}\|P_\Omega F^{-1}\|_{2 \rightarrow 2} \|\hat{X}\|_{2 \rightarrow 2} \leq \frac{1}{\sqrt{m}}\|\hat{X}\|_{2 \rightarrow 2} = \frac{1}{\sqrt{m}}\|Fx\|_\infty.
\] (4.2)

Setting \( \|x\|_\infty := \|Fx\|_\infty \) it is evident that \( \|Fx\|_\infty \leq \|x\|_1 \leq \sqrt{s}\|x\|_2 \leq \sqrt{s} \) for every \( x \in T \), and thus
\[
d_{2 \rightarrow 2}(A) \leq \sqrt{s/m}.
\]

Next, to estimate the \( \gamma_2 \) functional, recall from (2.1) that
\[
\gamma_2(A, \cdot \|_{2 \rightarrow 2}) \lesssim \int_0^{d_{2 \rightarrow 2}(A)} \log^{1/2} N(A, \| \cdot \|_{2 \rightarrow 2}, u) du.
\]

By (4.2),
\[
\|V_x - V_y\| = \|V_{x-y}\| \leq m^{-1/2}\|x - y\|_\infty,
\]
and hence for every \( u > 0 \), \( N(A, \| \cdot \|_{2 \rightarrow 2}, u) \leq N(T, m^{-1/2}\| \cdot \|_\infty, u) \). 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
\[
\log N(T, m^{-1/2}\| \cdot \|_\infty, u) \leq \log N(s^{1/2}B_1^n, m^{-1/2}\| \cdot \|_\infty, u)
\]
\[
\lesssim \left( \frac{s}{u} \right)^2 \log^2 (nu^2/\Delta^2).
\]

Since \( T \) is the union of \( s \)-dimensional Euclidean balls, a standard volumetric argument (see, e.g., [41] or [37, Chapter 8.4]) yields
\[
\log N(T, m^{-1/2}\| \cdot \|_\infty, u) \lesssim s \log(en/su)
\]
(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
\[
\gamma_2(A, \| \cdot \|_{2 \rightarrow 2}) \lesssim \sqrt{\frac{s}{m}} (\log s)(\log n),
\]
which implies that \( \gamma_2(A, \| \cdot \|_{2 \rightarrow 2}) \lesssim \delta \) 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
\[
\mathbb{P}(\delta_x \geq \delta) \leq \mathbb{P}(\delta_x \geq c_1 E + \delta/2) \leq \exp(-c_2 (m/s) \delta^2) \leq \eta,
\]
which, after possibly increasing the value of \( c \) enough to compensate \( c_2 \), completes the proof. ■
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 ξ = (ξ_j)_{j=1}^m be a random vector with independent mean-zero, variance one, L-subgaussian entries. Let Ψ_h ∈ ℂ^{m×m^2} be the random Gabor synthesis matrix generated by h = 1/√m ξ. If, for s ∈ ℤ and δ, η ∈ (0, 1),

\[ m ≥ cδ^{-2}s \max\{(log^2 m), log(η^{-1})\} \]

then with probability at least 1 − η the restricted isometry constant of Ψ_h satisfies that δ_s ≤ δ. The constant c > 0 depends only on L.

Before presenting the proof we will need several observations. First, note that for x ∈ ℂ^m, Ψ_hx = V_xξ, where the m × m matrix V_x is given by

\[ V_x = \frac{1}{\sqrt{m}} \sum_{λ ∈ ℤ_m^2} x_λ \pi(λ). \]

It is straightforward to check that \( \{ \frac{1}{\sqrt{m}} \pi(λ) : λ ∈ ℤ_m^2 \} \) is an orthonormal system in the space of complex m × m matrices endowed with the Frobenius norm. Therefore,

\[ \mathbb{E}\|V_x ξ\|_2^2 = \|V_x\|_F^2 = m^{-1/2} \sum_{λ ∈ ℤ_m^2} \|x_λ \pi(λ)\|_F^2 = \|x\|_2^2. \]

Hence, if \( T_s = \{ x ∈ ℂ^{m^2} : \|x\|_2 ≤ 1, \|x\|_0 ≤ s \} \) and \( A = \{ V_x : x ∈ T_s \} \), the restricted isometry constant is

\[ δ_s = \sup_{x ∈ T_s} \|Ψ_h x\|_2 - \|x\|_2 = \sup_{V_x ∈ A} \|V_x ξ\|_2 - \mathbb{E}\|V_x ξ\|_2. \]

Thus Theorem 3.1 applies again and we need to estimate the associated Dudley integral. Note that, as \( \pi(λ) \) is unitary, one has for \( x ∈ T_s \)

\[ \|V_x\|_{2→2} ≤ \frac{1}{\sqrt{m}} \sum_{λ ∈ ℤ_m^2} |x_λ| \|\pi(λ)\|_{2→2} ≤ \|x\|_1 / \sqrt{m} ≤ \sqrt{s/m} \|x\|_2, \quad (5.1) \]

so the upper integration limit will be \( d_{2→2}(A) ≤ \sqrt{s/m}. \)

**Lemma 5.2** There exists an absolute constant c such that for every 0 < u ≤ \( \sqrt{s/m} \),

\[ \log N(A, \|\cdot\|_{2→2}, u) ≤ cs \left( \log(em^2/s) + \log(3\sqrt{s/m}/u) \right), \]

and

\[ \log N(A, \|\cdot\|_{2→2}, u) ≤ c\frac{s \log^2 m}{mu^2}. \]

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 \( U \) in a vector space, \( \text{conv}(U) \) 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 \( U ⊂ X \) of cardinality N, and assume that for every \( L ∈ ℤ \) and \( (u_1, \ldots, u_L) ∈ U^L \),

\[ \varepsilon_x \sum_{j=1}^L ε_j u_j \|x\|_X ≤ A\sqrt{L}, \]

where \( (ε_j)_{j=1}^L \) denotes a Rademacher vector. Then for every \( u > 0 \),

\[ \log N(\text{conv}(U), \|\cdot\|_X, u) ≤ c(A/u)^2 \log N. \]
Proof of Lemma 5.2. Define the norm $\| \cdot \|$ on $\mathbb{R}^n$ by $\| x \| = \| V_r \|_{2\to 2}$, fix $S \subset \mathbb{Z}_m^2$ of cardinality $s$ and put $B_S = \{ x \in \mathbb{C}^{m^2} : \| x \|_2 \leq 1, \text{supp}(x) \subset S \}$. 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 $(m^2/s) \leq (em^2/s)^s$ such subsets 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 $(u_j)_{j=1}^L$ selected from the extreme points (with possible repetitions). Then, by the non-commutative Khintchine inequality, due to Lust-Picard, [28, 29, 40],

$$\mathbb{E}_\pi \left[ \sum_{j=1}^L \varepsilon_j V_{u,j} \right]_{2\to 2} \lesssim \sqrt{\log m} \max \left\{ \| \sum_{j=1}^L V_{u,j} V_{u,j}^* \|_{2\to 2}, \| \sum_{j=1}^L V_{u,j}^* V_{u,j} \|_{2\to 2} \right\}^{1/2}.$$ 

Recall that for every such $u_j$, $V_{u_j} = \alpha \pi(\lambda)$ with $|\alpha| = \sqrt{2s/m}$. Therefore, $V_{u,j}^* V_{u,j} = V_{u,j} V_{u,j}^* = (2s/m)I$ and thus

$$\mathbb{E}_\pi \left[ \sum_{j=1}^L \varepsilon_j V_{u,j} \right]_{2\to 2} \lesssim \sqrt{\log m \sqrt{s/m} L}.$$ 

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

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

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

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

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 $\{ \pi(\lambda) : \lambda \in \mathbb{Z}_m^2 \}$ that have been used in the proof are the facts that all $\pi(\lambda)$ are unitary and that $\{ m^{-1/2} \pi(\lambda) : \lambda \in \mathbb{Z}_m^2 \}$ 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 $\pi_r(t) \in T_r$, for $t \in T$, be an element in $T_r$ with the smallest $\ell_2$-distance to $t$, and choose
\( \ell \) for which \( 2^{\ell-1} \leq 2p \leq 2^\ell \). Since one may assume that \( \pi_r(t) = t \) for a sufficiently large \( r \), one has
\[
\sup_{t \in T} |\langle \pi_{r+1}(t) - \pi_r(t), Y \rangle| \leq \sup_{t \in T} |\langle \pi_r(t), Y \rangle| + \sup_{t \in T} \sum_{r=\ell}^{\infty} |\langle \pi_{r+1}(t) - \pi_r(t), Y \rangle|.
\tag{A.1}
\]
The \( p \)-th moment of the first term can be estimated as
\[
\left( E \sup_{t \in T} |\langle \pi_r(t), Y \rangle|^p \right)^{1/p} \leq \left( E \sum_{t \in T} |\langle t, Y \rangle|^p \right)^{1/p} \leq (|T|)^{1/p} \sup_{t \in T} (E |\langle t, Y \rangle|^p)^{1/p} \leq (2^{2\ell})^{1/p} \sup_{t \in T} (E |\langle t, Y \rangle|^p)^{1/p} \leq 16 \sup_{t \in T} (E |\langle t, Y \rangle|^p)^{1/p},
\]
where the last inequality follows from the choice of \( p \).
Since \( \xi \) is an \( L \)-subgaussian vector, one obtains for the second term in (A.1)
\[
P \left( \sup_{t \in T} \sum_{r=\ell}^{\infty} |\langle \pi_{r+1}(t) - \pi_r(t), Y \rangle| \geq uL \sum_{r=\ell}^{\infty} 2^{r/2} |\langle \pi_{r+1}(t) - \pi_r(t), x_j \rangle| \right)_{j=1}^{n} \|_2 \right)
\leq uL 2^{r/2} |\langle t - t', x_j \rangle| \right)_{j=1}^{n} \|_2 \right)
\leq \sum_{r=\ell}^{\infty} 2^{2^{r+1}} \cdot 2^{2^{r}} \exp(-2^{r}u^{2}/2) \leq 2 \exp(-2^{r}u^{2}/4) \leq 2 \exp(-pu^{2}/2),
\tag{A.2}
\]
when \( u \geq c \) for an appropriate choice of \( c \) (independent of \( \ell \)). Therefore, by integration,
\[
\left( E \sup_{t \in T} \sum_{r=\ell}^{\infty} |\langle \pi_{r+1}(t) - \pi_r(t), Y \rangle|^p \right)^{1/p} \lesssim L \sum_{r=\ell}^{\infty} 2^{r/2} |\langle \pi_{r+1}(t) - \pi_r(t), x_j \rangle| \right)_{j=1}^{n} \|_2 \leq L \gamma_2(T', \| \cdot \|_2)
\]
where \( T' = \{(t, x_j) \}_{j=1}^{n} : t \in T \} \). By the majorizing measures theorem,
\[
\gamma_2(T', \| \cdot \|_2) \lesssim \sup_{z \in T'} |\sum_{j=1}^{m} \epsilon_j g_j| = \sup_{t \in T} |\sum_{j=1}^{m} g_j(t, x_j)| = \sup_{t \in T} |\langle t, G \rangle|,
\]
which yields the claim.

### A.2 Proof of Lemma 5.3

If \( x \in \text{conv}(U) \) then \( x = \sum_{j=1}^{N} \theta_j u_j \) with \( \theta_j \geq 0 \), \( \sum_{j=1}^{N} \theta_j = 1 \). Let \( Z \in X \) be a random vector which takes the value \( u_j \) with probability \( \theta_j \) for \( j = 1, \ldots, N \) and thus satisfies \( E Z = x \). Let \( L \) be a number to be determined later, set \( Z_1, \ldots, Z_L \) be independent copies of \( Z \), and put
\[
Y = \frac{1}{L} \sum_{j=1}^{L} Z_j.
\]
If \( (\epsilon_j)_{j=1}^{m} \) is a Rademacher sequence independent of \( (Z_j) \) then by a standard symmetrization argument (see, e.g., [27]) and because \( (Z_j)_{j=1}^{L} \) ranges over \( U^L \)
\[
E \| x - Y \|_X = \frac{1}{L} E \| \sum_{j=1}^{L} (x - Z_j) \|_X \leq \frac{2}{L} E \| \sum_{j=1}^{L} \epsilon_j Z_j \|_X \leq 2A/\sqrt{L}.
\]
\[18\]
Thus, for $L \sim (A/u)^2$ there exists a realization $y = \frac{1}{L} \sum_{j=1}^{L} z_j$, for some $z_j \in U$, of $Y$ for which

$$\|x - y\|_X \leq u.$$  

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

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

as claimed.

References


