Restricted Isometries for Partial Random Circulant Matrices

Holger Rauhut, Justin Romberg, and Joel A. Tropp*

October 9, 2010

Abstract

In the theory of compressed sensing, *restricted isometry* analysis has become a standard tool for studying how efficiently a measurement matrix acquires information about sparse and compressible signals. Many recovery algorithms are known to succeed when the restricted isometry constants of the sampling matrix are small. Many potential applications of compressed sensing involve a data-acquisition process that proceeds by convolution with a random pulse followed by (nonrandom) subsampling. At present, the theoretical analysis of this measurement technique is lacking. This paper demonstrates that the sth order restricted isometry constant is small when the number m of samples satisfies $m \gtrsim (s \log n)^{3/2}$, where n is the length of the pulse. This bound improves on previous estimates, which exhibit quadratic scaling.

1 Introduction

The theory of compressed sensing [7, 9, 11, 16, 18, 38] predicts that a small number of linear samples suffice to capture all the information in a sparse vector and that, furthermore, we can recover the sparse vector from these samples using efficient algorithms. This discovery has a number of potential applications in signal processing, as well as other areas of science and technology.

The linear data acquisition process is described by a measurement matrix. The *restricted isometry property* (RIP) [10, 11, 18, 38] is a standard tool for studying how efficiently this matrix captures information about sparse signals. The RIP also streamlines the analysis of signal reconstruction algorithms. It is unknown whether any deterministic measurement matrix satisfies the RIP with the optimal scaling behavior. See, e.g., the discussion in [38, Sec. 2.5] or [18, Sec. 5.1]. In contrast, a variety of random measurement matrices exhibit the RIP with optimal scaling, including Gaussian matrices and Rademacher matrices [3, 11, 17, 39].

Although Gaussian random matrices are optimal for sparse recovery, they have limited use in practice because many measurement technologies impose structure on the matrix. Furthermore, recovery algorithms tend to be more efficient when the matrix admits a fast matrix–vector multiply. For example, random sets of rows from a Fourier transform matrix model the measurement process in MRI imaging, and these partial Fourier matrices lead to fast recovery algorithms because they can be applied using the FFT. It is known that a partial Fourier matrix satisfies a near-optimal RIP [11, 36, 38, 44]; the paper [38] contains some generalizations, and we refer to [40] for a variation related to recovery of sparse Legendre expansions.

Many potential applications of compressed sensing involve sampling processes that can be modeled by convolution with a random pulse. This measurement process can be modeled using a random circulant matrix. When we retain only a limited number of samples from the output of the convolution, the measurement process is described by a partial random circulant matrix. This situation has been studied in several works from the compressed sensing literature, including [2, 25, 37, 43, 48]. So far, the best available analysis of a partial random circulant matrix suggests that its restricted isometry constants do not exhibit optimal scaling. This work describes a new analysis that dramatically improves the previous estimates. Nevertheless, our results still fall short of the optimal scaling that one might hope for.

1.1 Compressed Sensing

The compressed sensing problem considers how to recover a vector $\boldsymbol{x} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ from the linear image

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x},$$

where the matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $m \ll n$. Clearly, it is impossible to reconstruct the vector $\boldsymbol{x}$ without additional prior information. Compressed sensing introduces the extra assumption that $\boldsymbol{x}$ is s-sparse, i.e., $\|\boldsymbol{x}\|_0 := \#\{\ell : x_\ell \neq 0\} \leq s$ for some $s \ll n$. More generally, we assume that $\boldsymbol{x}$ is well-approximated by a sparse vector.

The naïve approach of reconstructing $\boldsymbol{x}$ by solving the ℓ_0-minimization problem,

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

is NP-hard [32]. Therefore, several tractable heuristics have been proposed in the literature as alternatives to ℓ_0-minimization, most notably greedy algorithms [4, 20, 33, 34, 46] and ℓ_1-minimization [9, 13, 16]. The latter approach consists in solving the convex program

$$\min_{\boldsymbol{z}} \|\boldsymbol{z}\|_1 \quad \text{subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, \tag{1.1}$$

where $\|\cdot\|_p$ denotes the usual ℓ_p vector norm.

The restricted isometry property (RIP) offers a very elegant way to analyze ℓ_1-minimization and greedy algorithms. Define the restricted isometry constant δ_s of an $m \times n$ matrix $\boldsymbol{A}$ to be the smallest positive number that satisfies

$$(1 - \delta_s)\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leq (1 + \delta_s)\|\boldsymbol{x}\|_2^2 \quad \text{for all } \boldsymbol{x} \text{ with } \|\boldsymbol{x}\|_0 \leq s. \tag{1.2}$$

In words, the statement (1.2) requires that all column submatrices of $\boldsymbol{A}$ with at most s columns are well-conditioned. Informally, $\boldsymbol{A}$ is said to satisfy the RIP (with order s) when δ_s is small (for s close to m).

A number of recovery algorithms are provably effective for sparse recovery if the matrix $\boldsymbol{A}$ satisfies the RIP. More precisely, suppose that the matrix $\boldsymbol{A}$ obeys (1.2) with

$$\delta_{\kappa s} \;<\; \delta^* \tag{1.3}$$

for suitable constants $\kappa \geq 1$ and $\delta^* < 1$. Then these algorithms precisely recover all s-sparse vectors $\boldsymbol{x}$ from the measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. More generally, when the vector $\boldsymbol{x}$ is arbitrary and we acquire noisy observations

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

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

$$\|\boldsymbol{x} - \widetilde{\boldsymbol{x}}\|_2 \;\leq\; C_1 \frac{\sigma_s(\boldsymbol{x})_1}{\sqrt{s}} + C_2 \tau, \tag{1.4}$$

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

Algorithm	References	κ	δ^*
ℓ_1-minimization (1.1)	[8, 10, 19]	2	$\frac{3}{4+\sqrt{6}} \approx 0.4652$
CoSaMP	[21, 33]	4	$\sqrt{\frac{2}{5+\sqrt{73}}} \approx 0.3843$
Iterative Hard Thresholding	[4, 19]	3	$1/2$
Hard Thresholding Pursuit	[20]	3	$1/\sqrt{3} \approx 0.5774$

Table 1: Values of the constants κ and δ^* in (1.3) for various recovery algorithms.

A Gaussian random matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a matrix that has independent, normally distributed entries with mean zero and variance one. It is shown, e.g., in [3, 11, 31] that the restricted isometry constants of $\frac{1}{\sqrt{m}}\boldsymbol{A}$ satisfy $\delta_s \leq \delta$ with high probability provided that

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

It follows that the number m of Gaussian measurements required to reconstruct an s-sparse signal of length n is *linear* in the sparsity and *logarithmic* in the ambient dimension. See [3, 11, 18, 31, 38] for precise statements and extensions to Bernoulli and subgaussian matrices. It follows from lower estimates of Gelfand widths that this bound on the required samples is optimal [14, 22, 23]; that is, the log-factor must be present.

For a matrix consisting of m random rows from an $n \times n$ discrete Fourier transform matrix, slightly weaker estimates are available [11, 36, 38, 44]. The restricted isometry constants of this matrix satisfy $\delta_s \leq \delta$ with high probability provided that

$$m \;\geq\; C\delta^{-2} s \log^3(s) \log(n).$$

1.2 Partial Random Circulant Matrices

Given a vector $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_{n-1})^T \in \mathbb{R}^n$, we introduce the circulant matrix

$$\boldsymbol{\Phi}^\circ = \begin{bmatrix} \phi_0 & \phi_{n-1} & \cdots & \phi_1 \\ \phi_1 & \phi_0 & \cdots & \phi_2 \\ \vdots & \vdots & & \vdots \\ \phi_{n-1} & \phi_{n-2} & \cdots & \phi_0 \end{bmatrix} \in \mathbb{R}^{n \times n}. \tag{1.5}$$

Square matrices are not very interesting for compressed sensing, so we will restrict our attention to a row submatrix of $\boldsymbol{\Phi}^\circ$. Consider an *arbitrary* index set $\Omega \subset \{0, 1, \ldots, n-1\}$ whose cardinality $|\Omega| = m$. We define the operator $\boldsymbol{R}_\Omega \in \mathbb{R}^{m \times n}$ that restricts a vector to the entries listed in Ω. Then the corresponding partial circulant matrix generated with $\boldsymbol{\phi} \in \mathbb{R}^n$ is defined as

$$\boldsymbol{\Phi} = \frac{1}{\sqrt{m}} \boldsymbol{R}_\Omega \boldsymbol{\Phi}^\circ. \tag{1.6}$$

The action of $\boldsymbol{\Phi}$ can be interpreted as a circular convolution with the sequence $\frac{1}{\sqrt{m}} \boldsymbol{\phi}$ followed by a subsampling at locations indexed by Ω.

We will demonstrate that a partial circulant matrix with a random generator $\boldsymbol{\phi}$ has small restricted isometry constants. As a result, we can recover a sparse vector $\boldsymbol{x}$ robustly from measurements $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}$ using any of the algorithms mentioned above. Since $\boldsymbol{\Phi}^\circ$ can be diagonalized via the Fourier transform, the matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}^*$ both admit fast matrix–vector multiplication using the FFT algorithm. This fact allows us to accelerate recovery algorithms substantially.

A Rademacher sequence $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$ is a sequence of independent random variables, each taking the values $+1$ and -1 with equal probability. In the sequel, the matrix $\boldsymbol{\Phi}$ in (1.6) will always be generated by a Rademacher sequence $\boldsymbol{\phi} = \boldsymbol{\varepsilon}$, and we will refer to it as a *partial random circulant matrix*.

The main result of this paper is the following theorem.

Theorem 1.1 *Let Ω be an* arbitrary *subset of $\{0, 1, \ldots, n-1\}$ with cardinality $|\Omega| = m$. Let $\boldsymbol{\Phi}$ be the corresponding partial random circulant matrix (1.6) generated by a Rademacher sequence, and let δ_s denote the sth restricted isometry constant. Then*

$$\mathbb{E}[\delta_s] \leq C_1 \max \left\{ \frac{s^{3/2}}{m} \log^{3/2} n, \ \sqrt{\frac{s}{m}} \log s \log n \right\} \tag{1.7}$$

where $C_1 > 0$ is a universal constant.

In particular, (1.7) implies that for given $\delta \in (0, 1)$, we have $\mathbb{E}[\delta_s] \leq \delta$ provided

$$m \geq C_2 \max \left\{ \delta^{-1} s^{3/2} \log^{3/2} n, \ \delta^{-2} s \log^2 n \log^2 s \right\}, \tag{1.8}$$

where $C_2 > 0$ is another universal constant.

Theorem 1.1 tells us that partial random circulant matrices Φ obey (1.2) in expectation. The following theorem states that the random variable δ_s does not deviate much from its mean.

Theorem 1.2 *Let δ_s be as in Theorem 1.1. Then for $0 \leq \lambda \leq 1$*

$$\mathbb{P}\left(\delta_s \geq \mathbb{E}[\delta_s] + \lambda\right) \ \leq \ e^{-\lambda^2/\sigma^2} \quad \text{where} \quad \sigma^2 = C_3 \frac{s}{m} \log^2 s \log^2 n,$$

for a universal constant $C_3 > 0$.

The proof of Theorem 1.1 is connected with the approach for random partial Fourier matrices [38, 44]. We use a version of the classical Dudley inequality for Rademacher chaos that bounds the expectation of its supremum by the maximum of two entropy integrals that involve covering numbers with respect to two different metrics. We use elementary ideas from Fourier analysis to provide bounds for these metrics. This reduction allows us to exploit covering number estimates from the RIP analysis for partial Fourier matrices [38, 44] to complete the argument.

1.3 Discussion

In essence, the bound (1.8) exhibits the scaling behavior $m \gtrsim s^{3/2} \log^{3/2} n$. This result improves on the best available result for this type of matrix [25], but it falls short of the linear scaling in s that is typical in the compressive sensing literature. The bottleneck in our argument appears to be the bound on the "subexponential integral" (Section 2.4). It is not clear how to significantly improve (2.12) or the covering numbers from Proposition 2.3, so tightening this bound to $m \gtrsim s \log^p n$ for some constant p will probably require a different approach. Indeed, it is known that the central tool in this paper, the Dudley-type inequality for Rademacher chaos stated in Proposition 2.2, is not sharp for all examples [29, 45]. It might be that we are facing one of these cases.

The statement of Theorem 1.1 uses a very specific model for the measurement matrix based on a partial random convolution with a generator ϕ given by a Rademacher sequence. We have restricted our discussion to this example to simplify the exposition. Analogous results for other types of random generator sequences can be derived using the same type of analysis. In particular, one might consider the following variations.

Gaussian generating sequence. We can take the sequence ϕ to be iid Gaussian with zero mean and variance one. In this case, we can establish (1.8) by repeating the same steps because the central tool, Proposition 2.2, holds for Gaussian chaos processes as well as Rademacher chaos processes (perhaps with different constants). It is possible that, for this case, there is an extra factor of $\log n$ in the denominator of the variance σ^2 in the tail bound in Theorem 1.2.

Fourier-domain randomness. The generating sequence can also be iid Bernoulli in the Fourier domain. That is, we can take $\phi = \sqrt{n} \boldsymbol{F}^{-1} \boldsymbol{\varepsilon}$, where $\boldsymbol{F}$ is the Fourier matrix (see below) and $\boldsymbol{\varepsilon}$ is a Rademacher sequence. The analysis in this case is almost identical, except that we take the Fourier-domain expression (2.5) for the random process as our starting point.

This type of model was analyzed in [43] for the case where also Ω is chosen randomly; Theorem 1.1 gives us a result when Ω is *arbitrary*. Our model is also related to the *random demodulator* system analyzed in [47]. If we switch the roles of time and frequency, we can interpret the measurement system $\boldsymbol{\Phi}$ as taking a signal that is

sparse in the Fourier domain, multiplying it pointwise by a Rademacher sequence in the time domain (the random demodulation), and then recording the frequency components indexed by the set Ω. If Ω consists of a sequence of consecutive indices, then this operation is equivalent with random demodulation followed by bandpass filtering and, finally, acquiring m uniformly spaced samples. (In our model, we are observing the Fourier transform of the samples rather than the samples themselves.) This observation broadens the "randomly demodulate, integrate, then subsample" architecture of [47] to "randomly demodulate, bandpass filter, then subsample".

Complex generating sequences. It is also possible to take the generating sequence to be either a complex Steinhaus or complex Gaussian sequence. The proofs above remain essentially the same, the main difference would be establishing complex versions of Proposition 2.2 and Theorem 3.1 (some related results for Steinhaus sequences can be found in [38, Ch. 4]).

Toeplitz matrices. We can obtain analogous results for sections of a random Toeplitz matrix because a Toeplitz matrix can be embedded in a circulant matrix of twice the dimension.

Applications. From an engineering point of view, Theorems 1.1 and 1.2 tell us that we can identify a system with a sparse impulse response by probing it with a random input sequence and then taking a small number of samples of the output. This type of system identification, or *deconvolution*, problem is common task in signal processing, and the fact that it can be performed from a small number of samples allows for some interesting new design considerations.

For example, in radar imaging a transmitter sends out a pulse, which reflects off of a number of targets, and then a receiver observes this superposition of pulses (which can be modeled as the original pulse convolved with an unknown sparse range profile). The resolution to which we can resolve target locations is determined by the bandwidth of the pulse; to reconstruct the range profile digitally, then, the system requires an analog-to-digital converter (ADC) whose sample rate is on the same order as this bandwidth. Typical pulse bandwidths are in the gigahertz range, and ADCs that operate at this rate are expensive and low resolution. Indeed, lack of good high-speed ADCs has "historically slowed the introduction of digital techniques into radar signal processing" [41, Chap. 1.2]. Theorem 1.1 suggests that sample rate of the ADC depends primarily on the sparsity of the range profile, rather than the bandwidth of the pulse, which would allow us to achieve the same resolution with less expensive and more accurate hardware. See [26, 43] for more discussion of how convolution with a random waveform followed by subsampling can be applied to active imaging problems.

Another application of sparse recovery from a random convolution is increasing the field-of-view of a camera using a coded aperture [24, 30]. Here, we can imagine an optical architecture where a large image is convolved with a random code and then a small spatial portion of the result is sampled on a compact pixel array. If the image is sparse enough, Theorem 1.1 suggests that the entire field-of-view can be reconstructed to full resolution from this small set of observations.

Dimensionality reduction. The Johnson–Lindenstrauss lemma is an important tool for dimensionality reduction. It establishes that the pairwise distances between points in a

high-dimensional space are approximately preserved after we project the points into a significantly lower-dimensional space using a random linear map. While Gaussian or Bernoulli matrices were initially used for this task, more recent analyses show that structured random matrices also work. In particular, Hinrichs and Vybíral [27] have shown that one can perform dimension reduction using partial random circulant matrices with randomized column signs. These matrices are computationally efficient because they can be applied using the FFT algorithm. Krahmer and Ward subsequently showed that a matrix satisfying the RIP provides a Johnson–Lindenstrauss embedding if one randomizes the column signs [28]. Together with our result on the RIP of partial random circulant matrices, the work of Krahmer and Ward improves on a related result by Vybíral [49]. See [28] for a precise statement.

1.4 Relationship with previous work

Numerical results for compressive sampling by random convolution followed by subsampling appear in [48]. In this paper, the measurement process convolves with a pulse of length B and then extracts m equally spaced samples from the convolution. The effectiveness of this strategy is quantified empirically as a function of pulse length and the undersampling ratio: for long enough pulses, the number of samples required to reconstruct a signal is approximately linear in the sparsity.

Later, theoretical results for compressed sensing using random convolution were developed in [42, 43]. In these works, the measurement model is slightly different; the generating sequence ϕ is the discrete Fourier transform of an iid sequence of random signs. Convolution with this spectrally random sequence is followed by sampling at *random* locations (as opposed to the arbitrary set Ω we are considering in this work). This process is universally efficient, in that an s-sparse signal can be reconstructed from $m \gtrsim s \log^p n$ samples independent of the orthobasis in which it is sparse.

The first theoretical results for the measurement model we are using in this paper, convolution with a iid sequence followed by subsampling at fixed locations, can be traced to [2, 25]. They show that the matrix $\boldsymbol{\Phi}$ define in (1.6) has the RIP of order s with high probability when $m \gtrsim s^2 \log n$. These works are couched in the language of channel estimation, and so the results are stated explicitly for the case where Ω contains consecutive indices. Nevertheless, it appears that the same proof strategy extends to arbitrary Ω. Theorems 1.1 and 1.2 refine the sufficient condition for this model to $m \gtrsim s^{3/2} \log n$ using a completely different mathematical analysis.

A nonuniform recovery result for partial random circulant matrices has been established in [37, 38]. (See [38] for a discussion of the difference between nonuniform and uniform recovery guarantees.) Suppose that the number m of measurements satisfies $m \gtrsim s \log^2 n$. Let $\boldsymbol{x}_0$ be an s-sparse vector $\boldsymbol{x}_0$ whose nonzero components have random signs. With high probability, we can recover this vector exactly via ℓ_1-minimization using the measurements $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}_0$, where the partial random circulant matrix $\boldsymbol{\Phi}$ is drawn independently from $\boldsymbol{x}_0$. The proof involves duality for convex optimization, and it does not establish any type of RIP. As a consequence, this work does not offer any guarantees about stability in the presence of measurement noise or robustness when the signal $\boldsymbol{x}_0$ is not exactly s-sparse, in contrast with the RIP recovery bound (1.4).

2 Proof of Theorem 1.1 (Expectation)

We develop a method for estimating the restricted isometry constant δ_s for a fixed sparsity level s. Let T denote the set of all s-sparse signals in the Euclidean unit ball:

$$T := \{\boldsymbol{x} \in \mathbb{R}^n : \|\boldsymbol{x}\|_0 \leq s, \ \|\boldsymbol{x}\|_2^2 \leq 1\}. \tag{2.1}$$

Define a function $\|\cdot\|$ on Hermitian $n \times n$ matrices via the formula

$$\|\boldsymbol{A}\| := \sup_{\boldsymbol{x} \in T} |\boldsymbol{x}^* \boldsymbol{A} \boldsymbol{x}|.$$

This function can be extended to a norm on the set of all square matrices. We work with the quantity $\|\boldsymbol{\Phi}^* \boldsymbol{\Phi} - \mathbf{I}\|$ because

$$\|\boldsymbol{\Phi}^* \boldsymbol{\Phi} - \mathbf{I}\| \ = \ \sup_{\boldsymbol{x} \in T} |\langle (\boldsymbol{\Phi}^* \boldsymbol{\Phi} - \mathbf{I})\boldsymbol{x}, \boldsymbol{x} \rangle| \ = \ \sup_{\boldsymbol{x} \in T} \big| \, \|\boldsymbol{\Phi}\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_2^2 \, \big| \ = \ \delta_s. \tag{2.2}$$

Let $\boldsymbol{S}$ be the cyclic shift down operator on column vectors in $\mathbb{R}^n$. Applying the power $\boldsymbol{S}^k$ to $\boldsymbol{x}$ cycles $\boldsymbol{x}$ downward by k coordinates: $(\boldsymbol{S}^k \boldsymbol{x})_\ell = x_{\ell \ominus k}$, where $\ominus$ is subtraction modulo n. Note that $(\boldsymbol{S}^k)^* = \boldsymbol{S}^{-k} = \boldsymbol{S}^{n-k}$. We can now express $\boldsymbol{\Phi}$ as a random sum of shift operators,

$$\boldsymbol{\Phi} = \frac{1}{\sqrt{m}} \sum_{k=1}^{n} \varepsilon_k \boldsymbol{R}_\Omega \boldsymbol{S}^k.$$

It follows that

$$\boldsymbol{\Phi}^* \boldsymbol{\Phi} - \mathbf{I} \ = \ \frac{1}{m} \sum_{k \neq \ell} \varepsilon_k \varepsilon_\ell \, \boldsymbol{S}^{-k} \boldsymbol{R}_\Omega^* \boldsymbol{R}_\Omega \boldsymbol{S}^\ell \ = \ \frac{1}{m} \sum_{k \neq \ell} \varepsilon_k \varepsilon_\ell \, \boldsymbol{S}^{-k} \boldsymbol{P}_\Omega \boldsymbol{S}^\ell, \tag{2.3}$$

where $\boldsymbol{P}_\Omega = \boldsymbol{R}_\Omega^* \boldsymbol{R}_\Omega$ is the $n \times n$ diagonal projector onto the coordinates in Ω. Applying $\boldsymbol{P}_\Omega$ to $\boldsymbol{x}$ preserves the values of $\boldsymbol{x}$ on the set Ω while setting the values outside of Ω to zero.

Combining (2.2) and (2.3), we can view the restricted isometry constant as the supremum of a random process indexed by the set T:

$$\delta_s \ = \ \sup_{\boldsymbol{x} \in T} |G_{\boldsymbol{x}}| \quad \text{where} \quad G_{\boldsymbol{x}} \ = \ \frac{1}{m} \sum_{k \neq \ell} \varepsilon_k \varepsilon_\ell \, \boldsymbol{x}^* \boldsymbol{S}^{-k} \boldsymbol{P}_\Omega \boldsymbol{S}^\ell \boldsymbol{x} \tag{2.4}$$

We must bound the expected supremum of this process.

2.1 Fourier representation of the random process

One of the key ideas in this work is to re-express the random process $G_{\boldsymbol{x}}$ in the Fourier domain. Let $\boldsymbol{F}$ be the $n \times n$ discrete Fourier transform matrix whose entries are given by the expression

$$F(\omega, \ell) \ := \ \mathrm{e}^{-\mathrm{i}2\pi \, \omega \ell / n}, \quad 0 \leq \omega, \ell \leq n - 1.$$

Note that we employ the electrical engineering convention that $\boldsymbol{F}$ is unnormalized. The hat symbol indicates the Fourier transform of a vector: $\hat{\boldsymbol{x}} := \boldsymbol{F}\boldsymbol{x}$. Recall that a shift in the

time domain followed by a Fourier transform can also be written as a Fourier transform followed by a frequency modulation:

$$\boldsymbol{F}\boldsymbol{S}^k \; = \; \boldsymbol{M}^k\boldsymbol{F},$$

where $\boldsymbol{M}$ is the diagonal matrix with entries $M(\omega,\omega) := \mathrm{e}^{-\mathrm{i}2\pi\,\omega/n}$ for $0 \le \omega \le n-1$.

The random process $G_{\boldsymbol{x}}$ has the Fourier-domain representation

$$G_{\boldsymbol{x}} \; = \; \frac{1}{m}\sum_{k\neq\ell}\varepsilon_k\varepsilon_\ell\,\hat{\boldsymbol{x}}^*\boldsymbol{M}^{-k}\hat{\boldsymbol{P}}_\Omega\boldsymbol{M}^\ell\hat{\boldsymbol{x}}, \tag{2.5}$$

where $\hat{\boldsymbol{P}}_\Omega = n^{-1}\boldsymbol{F}\boldsymbol{P}_\Omega\boldsymbol{F}^{-1}$. The matrix $\hat{\boldsymbol{P}}_\Omega$ has several nice properties that we use in the sequel.

Lemma 2.1 *The $n \times n$ matrix $\hat{\boldsymbol{P}}_\Omega = n^{-1}\boldsymbol{F}\boldsymbol{P}_\Omega\boldsymbol{F}^{-1}$ has the following properties:*

1. *$\hat{\boldsymbol{P}}_\Omega$ is circulant and conjugate symmetric.*

2. *Along the diagonal $\hat{P}_\Omega(\omega,\omega) = m/n^2$, and off the diagonal $|\hat{P}_\Omega(\omega,\xi)| \le m/n^2$.*

3. *Since the rows and columns of $\hat{\boldsymbol{P}}_\Omega$ are circular shifts of one another,*

$$\sum_\omega\left|\hat{P}_\Omega(\omega,\xi)\right|^2 \; = \; \sum_\xi\left|\hat{P}_\Omega(\omega,\xi)\right|^2 \; = \; \|\hat{\boldsymbol{P}}_\Omega\|_F^2/n \; = \; m/n^3.$$

4. *$\hat{\boldsymbol{P}}_\Omega$ has exactly m nonzero eigenvalues, each of which is equal to $1/n$. As such, $\hat{\boldsymbol{P}}_\Omega$ has spectral norm $\|\hat{\boldsymbol{P}}_\Omega\| = 1/n$ and Frobenius norm $\|\hat{\boldsymbol{P}}_\Omega\|_F^2 = m/n^2$.*

Proof These properties follow almost immediately from the fact that $\boldsymbol{P}_\Omega = \boldsymbol{R}_\Omega^*\boldsymbol{R}_\Omega$ is a diagonal matrix with 0–1 entries. The matrix $\hat{\boldsymbol{P}}_\Omega$ inherits conjugate symmetry from $\boldsymbol{P}_\Omega$. The matrix $\hat{\boldsymbol{P}}_\Omega$ is circulant because it is diagonalized by the Fourier transform. Since we form $\hat{\boldsymbol{P}}_\Omega$ by applying a similarity transform to $\boldsymbol{P}_\Omega$, they have the same eigenvalues modulo the scale factor n^{-1}. The remaining points follow from the simple calculations described in the statement of the lemma. ∎

2.2 Integrability of chaos processes

For the next step in the argument, we must rewrite the random process (2.5) again. Let $\boldsymbol{\varepsilon} = [\varepsilon_0,\ldots,\varepsilon_{n-1}]^*$. The process can now be expressed as a quadratic form:

$$G_{\boldsymbol{x}} \; = \; \langle\boldsymbol{\varepsilon},\boldsymbol{Z}_{\boldsymbol{x}}\,\boldsymbol{\varepsilon}\rangle \quad\text{where}\quad \boldsymbol{x}\in T. \tag{2.6}$$

The matrix $\boldsymbol{Z}_{\boldsymbol{x}}$ has entries

$$Z_{\boldsymbol{x}}(k,\ell) \; = \; \begin{cases} m^{-1}\hat{\boldsymbol{x}}^*\boldsymbol{M}^{-k}\hat{\boldsymbol{P}}_\Omega\boldsymbol{M}^\ell\hat{\boldsymbol{x}}, & k\neq\ell \\ 0, & k=\ell \end{cases}.$$

A short calculation verifies that this matrix can be written compactly.

$$\boldsymbol{Z_x} \;=\; \frac{1}{m}\left(\boldsymbol{F^*\hat{X}^*\hat{P}_\Omega\hat{X}F} - \operatorname{diag}(\boldsymbol{F^*\hat{X}^*\hat{P}_\Omega\hat{X}F})\right), \tag{2.7}$$

where $\boldsymbol{\hat{X}} := \operatorname{diag}(\boldsymbol{\hat{x}})$ is the diagonal matrix constructed from the vector $\boldsymbol{\hat{x}}$. The term *homogeneous second-order chaos* is used to refer to a random process $G_{\boldsymbol{x}}$ of the form (2.6) where each matrix $\boldsymbol{Z_x}$ is conjugate symmetric and hollow, i.e., has zeros on the diagonal.

To bound the expected supremum of the random process $G_{\boldsymbol{x}}$ over the set T, we apply a version of Dudley's inequality that is specialized to this setting. Define two pseudo-metrics on the index set T:

$$d_1(\boldsymbol{x},\boldsymbol{y}) \;:=\; \|\boldsymbol{Z_x} - \boldsymbol{Z_y}\| \quad\text{and}\quad d_2(\boldsymbol{x},\boldsymbol{y}) \;:=\; \|\boldsymbol{Z_x} - \boldsymbol{Z_y}\|_{\mathrm{F}}.$$

Let $N(T, d_i, u)$ denote the minimum number of balls of radius u in the metric d_i that we need to cover the set T.

Proposition 2.2 (Dudley's inequality for chaos) *Suppose that $G_{\boldsymbol{x}}$ is a homogeneous second-order chaos process indexed by a set T. Fix a point $\boldsymbol{x}_0 \in T$. There exists a universal constant* K *such that*

$$\mathbb{E}\sup_{\boldsymbol{x}\in T}|G_{\boldsymbol{x}} - G_{\boldsymbol{x}_0}| \;\leq\; \mathrm{K}\max\left\{\int_0^\infty \log N(T, d_1, u)\,\mathrm{d}u, \int_0^\infty \sqrt{\log N(T, d_2, u)}\,\mathrm{d}u\right\}. \tag{2.8}$$

Proposition 2.2 is based on the idea that the random process has a subexponential part, whose variation is controlled by the integral with respect to d_1, and a subgaussian part, whose variation is controlled by the integral with respect to d_2. This result appears in [29, Thm. 11.22] and [45, Thm. 2.5.2]. Our statement of the proposition looks different from the versions presented in the literature, so we sketch the derivation in Appendix A.

2.3 The subgaussian integral

In this section, we develop an estimate for the second integral in (2.8). To do so, we need a simpler bound for the metric d_2. First, note that

$$\begin{aligned}
d_2(\boldsymbol{x},\boldsymbol{y}) &\leq \frac{1}{m}\|\boldsymbol{F^*}(\boldsymbol{\hat{X}^*\hat{P}_\Omega\hat{X}} - \boldsymbol{\hat{Y}^*\hat{P}_\Omega\hat{Y}})\boldsymbol{F}\|_{\mathrm{F}} = \frac{n}{m}\|\boldsymbol{\hat{X}^*\hat{P}_\Omega\hat{X}} - \boldsymbol{\hat{Y}^*\hat{P}_\Omega\hat{Y}}\|_{\mathrm{F}} \\
&= \frac{n}{2m}\|(\boldsymbol{\hat{X}}-\boldsymbol{\hat{Y}})^*\boldsymbol{\hat{P}_\Omega}(\boldsymbol{\hat{X}}+\boldsymbol{\hat{Y}}) + (\boldsymbol{\hat{X}}+\boldsymbol{\hat{Y}})^*\boldsymbol{\hat{P}_\Omega}(\boldsymbol{\hat{X}}-\boldsymbol{\hat{Y}})\|_{\mathrm{F}} \\
&= \frac{n}{2m}\left[\sum\nolimits_{\omega,\xi}|\hat{P}_\Omega(\omega,\xi)|^2\cdot|\hat{Q}_{\boldsymbol{x},\boldsymbol{y}}(\omega,\xi)|^2\right]^{1/2}
\end{aligned}$$

where we have written

$$\hat{Q}_{\boldsymbol{x},\boldsymbol{y}}(\omega,\xi) \;:=\; (\hat{x}(\omega)+\hat{y}(\omega))(\hat{x}(\xi)-\hat{y}(\xi))^* + (\hat{x}(\omega)-\hat{y}(\omega))(\hat{x}(\xi)+\hat{y}(\xi))^*.$$

The first inequality arises when we re-introduce the diagonal entries of the hollow matrices. The next identity follows from the unitary invariance of the Frobenius norm. The second line is the polarization identity, and we obtain the last line by expressing the Frobenius norm in terms of coordinates.

Define the $\|\cdot\|_\infty$ norm to be the ℓ_∞ norm in the discrete Fourier domain

$$\|\boldsymbol{x}\|_\infty \ := \ \|\hat{\boldsymbol{x}}\|_\infty.$$

We can bound the entries of $\hat{\boldsymbol{Q}}_{\boldsymbol{x},\boldsymbol{y}}$ in terms of this norm. If we abbreviate $\boldsymbol{v} \ = \ \boldsymbol{x} + \boldsymbol{y}$,

$$|\hat{Q}_{\boldsymbol{x},\boldsymbol{y}}(\omega,\xi)| \ \leq \ \|\boldsymbol{x} - \boldsymbol{y}\|_\infty \cdot (|\hat{v}(\omega)| + |\hat{v}(\xi)|).$$

Introduce the latter bound into our estimate for the metric and apply the triangle inequality to reach

$$d_2(\boldsymbol{x},\boldsymbol{y}) \ \leq \ \frac{n\|\boldsymbol{x} - \boldsymbol{y}\|_\infty}{2m}\left[\left(\sum\nolimits_{\omega,\xi}|\hat{P}_\Omega(\omega,\xi)|^2\,|\hat{v}(\omega)|^2\right)^{1/2} + \left(\sum\nolimits_{\omega,\xi}|\hat{P}_\Omega(\omega,\xi)|^2\,|\hat{v}(\xi)|^2\right)^{1/2}\right].$$

Let us examine the first sum more closely.

$$\left(\sum\nolimits_{\omega,\xi}|\hat{P}_\Omega(\omega,\xi)|^2\,|\hat{v}(\omega)|^2\right)^{1/2} \ = \ \left(\frac{m}{n^3}\|\hat{\boldsymbol{v}}\|_2^2\right)^{1/2} \ \leq \ \sqrt{\frac{m}{n^3}}(\|\hat{\boldsymbol{x}}\|_2 + \|\hat{\boldsymbol{y}}\|_2) \ = \ \frac{2\sqrt{m}}{n}.$$

The first identity follows from Point 3 of Lemma 2.1. The second relation is the the triangle inequality. The last identity is a consequence of the fact that $\boldsymbol{x}$ and $\boldsymbol{y}$ have unit energy together with Parseval's identity. An analogous argument applies to the second sum. We conclude that

$$d_2(\boldsymbol{x},\boldsymbol{y}) \leq \frac{2\|\boldsymbol{x} - \boldsymbol{y}\|_\infty}{\sqrt{m}} = 2\sqrt{\frac{s}{m}} \cdot \frac{1}{\sqrt{s}}\|\boldsymbol{x} - \boldsymbol{y}\|_\infty.$$

This bound on d_2 allows us to estimate the subgaussian integral in terms of the covering numbers of T with respect to the norm $s^{-1/2}\|\cdot\|_\infty$. Abbreviating $\alpha = 2\sqrt{s/m}$, we compute that

$$
\begin{aligned}
I_2 \ &:= \ \int_0^\infty \sqrt{\log N(T, d_2, u)}\,\mathrm{d}u \ \leq \ \int_0^\infty \sqrt{\log N(T, \alpha s^{-1/2}\|\cdot\|_\infty, u)}\,\mathrm{d}u \\
&= \ \int_0^\infty \sqrt{\log N(T, s^{-1/2}\|\cdot\|_\infty, \alpha^{-1}u)}\,\mathrm{d}u \ = \ \alpha\int_0^\infty \sqrt{\log N(T, s^{-1/2}\|\cdot\|_\infty, u)}\,\mathrm{d}u \\
&= \ 2\sqrt{\frac{s}{m}}\int_0^1 \sqrt{\log N(T, s^{-1/2}\|\cdot\|_\infty, u)}\,\mathrm{d}u.
\end{aligned}
$$

The first inequality uses the fact that the metric balls in d_2 are *larger* than the balls in the norm $\alpha s^{-1/2}\|\cdot\|_\infty$ because the metric is *smaller* than the norm. The second line follows from an elementary scaling property of covering numbers along with a change of variables in the integral. Finally, we apply the fact that T is contained in the unit ball of $s^{-1/2}\|\cdot\|_\infty$ to see that the integrand vanishes for $u \geq 1$. We can now exploit some covering number estimates that appear in the literature [38, 44]. The first bound follows from a volume comparison argument; the second uses the empirical method invented by Maurey [35] and refined by Carl [12].

Proposition 2.3 (Covering Numbers) *For $u \in (0, 1]$, we have the following bound.*

$$N(T, s^{-1/2}\|\cdot\|_\infty, u) \ \leq \ \min\left\{\left(\frac{C_1 n}{s}\right)^s(1 + 2/u)^s,\ n^{C_2(\log n)/u^2}\right\}. \qquad (2.9)$$

The C_i are positive universal constants.

Explicit values of the constants C_1, C_2 can be found in [38, Lem. 8.3, eq. (8.14)]. We finish off the estimate for the first integral using Proposition 2.3. Splitting the integral at λ, we have

$$
\begin{aligned}
I_2 &\leq C\sqrt{\frac{s}{m}} \left[\int_0^\lambda \sqrt{s(\log(n/s) + \log(1 + 2/u))}\, du \;+\; \log(n) \int_\lambda^1 u^{-1}\, du \right] \\
&\leq C\sqrt{\frac{s}{m}} \left[\lambda\sqrt{s\log(n/s)} + \lambda\sqrt{s}\log(\lambda^{-1}) + \log(n)\log(\lambda^{-1}) \right] \\
&\leq C\sqrt{\frac{s\log^2(s)\log^2(n)}{m}},
\end{aligned}
\tag{2.10}
$$

where we have chosen $\lambda = s^{-1/2}$ in the last step.

2.4   The subexponential integral

We can also bound the d_1 metric in terms of the norm $\|\cdot\|_\infty$, which allows us to re-use the estimates for the covering numbers given in Proposition 2.3 to control the first integral in (2.8). To begin,

$$
d_1(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{A}_{\boldsymbol{x},\boldsymbol{y}} - \boldsymbol{D}_{\boldsymbol{x},\boldsymbol{y}}\| \leq \|\boldsymbol{A}_{\boldsymbol{x},\boldsymbol{y}}\| + \|\boldsymbol{D}_{\boldsymbol{x},\boldsymbol{y}}\|
$$

where the matrix $\boldsymbol{A}_{\boldsymbol{x},\boldsymbol{y}}$ is given by the expression

$$
\begin{aligned}
\boldsymbol{A}_{\boldsymbol{x},\boldsymbol{y}} &:= \frac{1}{m}\boldsymbol{F}^*(\hat{\boldsymbol{X}}^*\hat{\boldsymbol{P}}_\Omega\hat{\boldsymbol{X}} - \hat{\boldsymbol{Y}}^*\hat{\boldsymbol{P}}_\Omega\hat{\boldsymbol{Y}})\boldsymbol{F} \\
&= \frac{1}{2m}\boldsymbol{F}^*((\hat{\boldsymbol{X}} + \hat{\boldsymbol{Y}})^*\hat{\boldsymbol{P}}_\Omega(\hat{\boldsymbol{X}} - \hat{\boldsymbol{Y}}) + (\hat{\boldsymbol{X}} - \hat{\boldsymbol{Y}})^*\hat{\boldsymbol{P}}_\Omega(\hat{\boldsymbol{X}} + \hat{\boldsymbol{Y}}))\boldsymbol{F},
\end{aligned}
\tag{2.11}
$$

and $\boldsymbol{D}_{\boldsymbol{x},\boldsymbol{y}}$ denotes the diagonal of the matrix $\boldsymbol{A}_{\boldsymbol{x},\boldsymbol{y}}$.

We bound the diagonal term first. Let $\boldsymbol{f}_k$ be the kth column of $\boldsymbol{F}$, and note that $\|\boldsymbol{f}_k\|_2^2 = n$. Owing to Lemma 2.1, we have

$$
\begin{aligned}
\|\boldsymbol{D}_{\boldsymbol{x},\boldsymbol{y}}\| &= \frac{1}{m}\max_k \left| \boldsymbol{f}_k^*(\hat{\boldsymbol{X}} + \hat{\boldsymbol{Y}})^*\hat{\boldsymbol{P}}_\Omega(\hat{\boldsymbol{X}} - \hat{\boldsymbol{Y}})\boldsymbol{f}_k \right| \\
&\leq \frac{1}{m}\max_k \|\hat{\boldsymbol{P}}_\Omega(\hat{\boldsymbol{X}} + \hat{\boldsymbol{Y}})\boldsymbol{f}_k\|_2 \cdot \|(\hat{\boldsymbol{X}} - \hat{\boldsymbol{Y}})\boldsymbol{f}_k\|_2 \\
&\leq \frac{1}{m}\max_k \|\hat{\boldsymbol{P}}_\Omega\| \cdot \|\hat{\boldsymbol{X}} + \hat{\boldsymbol{Y}}\| \cdot \|\hat{\boldsymbol{X}} - \hat{\boldsymbol{Y}}\| \cdot \|\boldsymbol{f}_k\|_2^2 \;=\; \frac{1}{m}\|\boldsymbol{x} + \boldsymbol{y}\|_\infty \cdot \|\boldsymbol{x} - \boldsymbol{y}\|_\infty \\
&\leq \frac{2s}{m} \cdot \frac{1}{\sqrt{s}}\|\boldsymbol{x} - \boldsymbol{y}\|_\infty.
\end{aligned}
$$

In the last inequality, we have used the fact that $\|\boldsymbol{x} + \boldsymbol{y}\|_\infty \leq \|\boldsymbol{x} + \boldsymbol{y}\|_1 \leq 2\sqrt{s}$ for $\boldsymbol{x}, \boldsymbol{y} \in T$. For the off-diagonal term, we use Lemma 2.1 to compute

$$
\begin{aligned}
\|\boldsymbol{A}_{\boldsymbol{x},\boldsymbol{y}}\| &= \frac{1}{m}\|\boldsymbol{F}^*(\hat{\boldsymbol{X}} + \hat{\boldsymbol{Y}})^*\hat{\boldsymbol{P}}_\Omega(\hat{\boldsymbol{X}} - \hat{\boldsymbol{Y}})\boldsymbol{F}\| \;\leq\; \frac{n}{m}\|(\hat{\boldsymbol{X}} + \hat{\boldsymbol{Y}})^*\hat{\boldsymbol{P}}_\Omega(\hat{\boldsymbol{X}} - \hat{\boldsymbol{Y}})\| \\
&\leq \frac{n}{m}\|(\hat{\boldsymbol{X}} + \hat{\boldsymbol{Y}})\| \cdot \|\hat{\boldsymbol{P}}_\Omega\| \cdot \|(\hat{\boldsymbol{X}} - \hat{\boldsymbol{Y}})\| \;=\; \frac{1}{m}\|\boldsymbol{x} + \boldsymbol{y}\|_\infty \cdot \|\boldsymbol{x} - \boldsymbol{y}\|_\infty \\
&\leq \frac{2s}{m} \cdot \frac{1}{\sqrt{s}}\|\boldsymbol{x} - \boldsymbol{y}\|_\infty.
\end{aligned}
$$

In summary,

$$d_1(\boldsymbol{x}, \boldsymbol{y}) \le \frac{4s}{m} \cdot \frac{1}{\sqrt{s}} \|\boldsymbol{x} - \boldsymbol{y}\|_\infty. \tag{2.12}$$

The covering number estimates of Proposition 2.3 allow us to bound the subexponential integral.

$$
\begin{aligned}
I_1 &:= \int_0^\infty \log N(T, d_1, u) \; \mathrm{d}u \le \frac{4s}{m} \int_0^1 N(T, s^{-1/2}\|\cdot\|_\infty, u) \; \mathrm{d}u \\
&\le \frac{\mathrm{C}s}{m} \left(\int_0^\lambda (s \log(n/s) + s \log(1 + 2/u)) \, \mathrm{d}u \; + \; \log^2(n) \int_\lambda^1 u^{-2} \, \mathrm{d}u \right) \\
&\le \frac{\mathrm{C}s}{m} \left(\lambda s \log(n/s) + \lambda s \log(1 + 2/\lambda) + \lambda^{-1} \log^2(n) \right) \\
&\le \frac{\mathrm{C}s^{3/2} \log^{3/2}(n)}{m}.
\end{aligned}
\tag{2.13}
$$

We have taken $\lambda = s^{-1/2} \log^{1/2}(n)$ in the last step.

2.5   Denouement

As we noted in (2.4), the restricted isometry constant δ_s is given by the supremum of the random process $G_{\boldsymbol{x}}$. To compute the expectation of this supremum, we simply apply Proposition 2.2. Select $\boldsymbol{x}_0 = \boldsymbol{0}$ so that $G_{\boldsymbol{x}_0} = 0$. Introduce the estimate (2.13) for the subexponential integral and (2.10) for the subgaussian integral into Dudley's inequality (2.8).

$$\mathbb{E}\,\delta_s = \mathbb{E} \sup_{\boldsymbol{x} \in T} |G_{\boldsymbol{x}}| \le \mathrm{K} \left[\frac{s^{3/2} \log^{3/2}(n)}{m} + \sqrt{\frac{s \log^2(s) \log^2(n)}{m}} \right].$$

This point completes the proof.

3   Proof of Theorem 1.2 (Tail Bound)

In this section, we develop a tail bound on the supremum of the process G_x. We require the following result, which is Theorem 17 from [5]. Let $\mathcal{F}$ be a collection of $n \times n$ symmetric real matrices, and assume that $Z(k, k) = 0$ for each $\boldsymbol{Z} \in \mathcal{F}$. We are concerned with the tail behavior of the real-valued random variable

$$Y := \sup_{\boldsymbol{Z} \in \mathcal{F}} \sum_{k,\ell=1}^n \varepsilon_k \varepsilon_\ell \, Z(k, \ell).$$

Define two variance parameters:

$$U := \sup_{\boldsymbol{Z} \in \mathcal{F}} \|\boldsymbol{Z}\|$$

and

$$V^2 := \mathbb{E} \sup_{\boldsymbol{Z} \in \mathcal{F}} \sum_{k=1}^n \left| \sum_{\ell=1}^n \varepsilon_\ell Z(k, \ell) \right|^2.$$

The parameter V^2 describes the variance of X near its mean, while the second parameter U is the scale on which large deviations occur.

Proposition 3.1 (Tail Bound for Chaos) *Under the preceding assumptions,*

$$\mathbb{P}\{Y \geq \mathbb{E}[Y] + \lambda\} \leq \exp\left(-\frac{\lambda^2}{32V^2 + 65U\lambda/3}\right) \tag{3.1}$$

for all $\lambda \geq 0$.

Recall from (2.4), (2.6), and (2.7) that the restricted isometry constant can be written as

$$\delta_s = \sup_{\boldsymbol{x} \in T} |G_{\boldsymbol{x}}| = \sup_{\boldsymbol{x} \in T} \sum_{k,\ell} \varepsilon_k \varepsilon_\ell \, Z_{\boldsymbol{x}}(k,\ell)$$

where the matrix $\boldsymbol{Z_x}$ has the expression

$$\boldsymbol{Z_x} = \boldsymbol{A_x} - \operatorname{diag}(\boldsymbol{A_x}) \quad \text{for} \quad \boldsymbol{A_x} = \frac{1}{m}\boldsymbol{F}^*\hat{\boldsymbol{X}}^*\hat{\boldsymbol{P}}_\Omega\hat{\boldsymbol{X}}\boldsymbol{F}.$$

As a consequence, Theorem 3.1 applies to the random variable δ_s.

To bound the first parameter U, we first apply the triangle inequality to obtain

$$\|\boldsymbol{Z_x}\| \leq \|\boldsymbol{A_x}\| + \|\operatorname{diag}(\boldsymbol{A_x})\|.$$

Emulating the arguments in Section 2.4, we can bound each of the two terms.

$$\|\boldsymbol{A_x}\| \leq \frac{n}{m}\|\hat{\boldsymbol{P}}_\Omega\| \cdot \|\boldsymbol{x}\|_\infty^2 \leq \frac{s}{m},$$

Similarly,

$$\|\operatorname{diag}(\boldsymbol{A_x})\| = \frac{1}{m}\max_k \left|\boldsymbol{f}_k^*\hat{\boldsymbol{X}}^*\hat{\boldsymbol{P}}_\Omega\hat{\boldsymbol{X}}\boldsymbol{f}_k\right| \leq \frac{s}{m}.$$

In total, $U \leq 2s/m$.

To bound the other parameter V^2, we use the following "vector version" of the Dudley inequality, which we prove in the Appendix.

Proposition 3.2 *Consider the vector-valued random process*

$$\boldsymbol{h_x} = \boldsymbol{Z_x}\,\boldsymbol{\varepsilon} \quad \text{for } \boldsymbol{x} \in T.$$

Recall the definition of the pseudo-metric

$$d_2(\boldsymbol{x}, \boldsymbol{y}) := \|\boldsymbol{Z_x} - \boldsymbol{Z_y}\|_{\mathrm{F}}$$

Fix a point $\boldsymbol{x}_0 \in T$. There exists a universal constant $\mathrm{K} > 0$ such that

$$\left(\mathbb{E}\sup_{\boldsymbol{x} \in T}\|\boldsymbol{h_x} - \boldsymbol{h_{x_0}}\|_2^2\right)^{1/2} \leq \mathrm{K}\int_0^\infty \sqrt{N(T, d_2, u)}\,\mathrm{d}u. \tag{3.2}$$

14

With $\boldsymbol{x}_0 = \boldsymbol{0}$, the left-hand side of (3.2) is precisely V. We have already studied the integral on the right-hand side of (3.2) in Section 2.3. We import (2.10) to reach

$$V^2 \leq \frac{\mathrm{C}s}{m} \log^2(s) \log^2 n.$$

We are prepared to complete the tail bound for δ_s. For $\lambda \leq 1$,

$$\frac{\lambda^2}{32C(s/m)\log^2(s)\log^2(n) + (130/3)(s/m)\lambda} \leq \frac{1}{\mathrm{C}'} \min\left(\frac{\lambda^2}{(s/m)\log^2(s)\log^2(n)}, \frac{\lambda}{s/m} \right)$$

$$\leq \frac{\lambda^2}{\mathrm{C}'(s/m)\log^2(s)\log^2(n)}.$$

Applying (3.1), we reach

$$\mathbb{P}\{\delta_s > \mathbb{E}[\delta_s] + \lambda\} \leq \mathrm{e}^{-\lambda^2/\mathrm{C}'\sigma^2},$$

with $\sigma^2 = (s/m)\log^2(s)\log^2(n)$.

A A Dudley-type inequality for chaos processes

We provide a proof sketch for Proposition 2.2. Let $\{\varepsilon'_k\}$ be a Rademacher sequence independent of $\{\varepsilon_k\}$. The decoupling method (see, for example, [15, Th. 3.1.2]) yields

$$\mathbb{E}\sup_{\boldsymbol{x}\in T} |G_{\boldsymbol{x}_0} - G_{\boldsymbol{x}}| = \mathbb{E}\sup_{\boldsymbol{x}\in T} |\sum_{k,\ell} \varepsilon_k \varepsilon_\ell (Z_{\boldsymbol{x}_0}(k,\ell) - Z_{\boldsymbol{x}}(k,\ell))|$$

$$\leq 8\,\mathbb{E}\sup_{\boldsymbol{x}\in T} |\sum_{k,\ell} \varepsilon_k \varepsilon'_\ell (Z_{\boldsymbol{x}_0}(k,\ell) - Z_{\boldsymbol{x}}(k,\ell))|.$$

Now we introduce two independent standard Gaussian sequences $\{g_k\}$ and $\{g'_k\}$. Applying the contraction principle [29, eq. (4.8)] twice, first conditioned on ε_ℓ and then on g'_ℓ, leads to

$$\mathbb{E}\sup_{\boldsymbol{x}\in T} |G_{\boldsymbol{x}_0} - G_{\boldsymbol{x}}| \leq 8\sqrt{\frac{\pi}{2}}\,\mathbb{E}\sup_{\boldsymbol{x}\in T} |\sum_{k,\ell} \varepsilon_k g'_\ell (Z_{\boldsymbol{x}_0}(k,\ell) - Z_{\boldsymbol{x}}(k,\ell))|$$

$$\leq 4\pi\,\mathbb{E}\sup_{\boldsymbol{x}\in T} |\sum_{k,\ell} g_k g'_\ell (Z_{\boldsymbol{x}_0}(k,\ell) - Z_{\boldsymbol{x}}(k,\ell))|. \tag{A.1}$$

Thus our task is to bound the expected supremum of a decoupled Gaussian chaos process. Using [45, Th. 1.2.7], we see that

$$\mathbb{E}\sup_{\boldsymbol{x}\in T} |G_{\boldsymbol{x}} - G_{\boldsymbol{x}_0}| \leq \mathrm{C}(\gamma_1(T,d_2) + \gamma_2(T,d_1)),$$

where $\gamma_\alpha(T,d)$ is the γ_α-functional of the metric space (T,d); see [45, Def. 1.2.5]. It is known that

$$\gamma_\alpha(T,d) \leq \mathrm{C}\int_0^\infty \log^{1/\alpha}(N(T,d,u))\,\mathrm{d}u.$$

This is established carefully in [45, p. 13] for the case $\alpha = 2$ and the general case is analogous. The statement of the theorem follows.

B   A Dudley type inequality for vector-valued Rademacher processes

In this section we give a sketch of the proof of Proposition 3.2, the vector-valued version of Dudley's inequality. It is a consequence of the following proposition.

Proposition B.1 *Let $\boldsymbol{A}$ be an $m \times n$ matrix with columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. For each $u \geq 0$,*

$$\mathbb{P}\left(\| \sum_{j=1}^{n} \varepsilon_j a_j \|_2 \geq \|\boldsymbol{A}\|_{\mathrm{F}} \cdot u \right) \leq 2\mathrm{e}^{-cu^2},$$

where c *is a universal constant.*

Proof It is easily seen that $\mathbb{E} \| \sum_{j=1}^{n} \varepsilon_j \boldsymbol{a}_j \|_2^2 = \sum_j^n \|\boldsymbol{a}_j\|_2^2 = \|\boldsymbol{A}\|_{\mathrm{F}}^2$. The vector version of Khintchine's inequality, given in [15, Th. 1.3.1] implies that, for $p \geq 2$,

$$\left(\mathbb{E} \| \sum_{j=1}^{n} \varepsilon_j \boldsymbol{a}_j \|_2^p \right)^{1/p} \leq \sqrt{p} \left(\mathbb{E} \| \sum_{j=1}^{n} \varepsilon_j \boldsymbol{a}_j \|_2^2 \right)^{1/2} = \sqrt{p}\|\boldsymbol{A}\|_F.$$

This moment growth implies the tail estimate, see e.g. [38, Proposition 6.5]

$$\mathbb{P}(\| \sum_{j} \varepsilon_j \boldsymbol{a}_j \|_2 \geq \mathrm{e}^{1/2}\|\boldsymbol{A}\|_{\mathrm{F}} \cdot u) \leq \mathrm{e}^{-u^2}, \quad u \geq \sqrt{2},$$

which yields the conclusion.   ∎

An explicit value of c $= 1/2$ for the constant above can be found using non-commutative Khintchine inequalities [6, 38].

With this proposition in place, we can prove Proposition 3.2 as follows. From Proposition (B.1),

$$\mathbb{P}(\|\boldsymbol{h_x} - \boldsymbol{h_y}\|_2 \geq \|\boldsymbol{Z_x} - \boldsymbol{Z_y}\|_{\mathrm{F}} \cdot u) \leq 2\mathrm{e}^{-cu^2} \quad \text{ for all } \boldsymbol{x}, \boldsymbol{y} \in T.$$

This sets us into the position to follow the standard proof of Dudley's inequality for scalar-valued subgaussian processes; see [38, Theorem 6.23] or [1, 29, 45]. One only has to replace the triangle inequality for the absolute value by the one for $\|\cdot\|_2$ in $\mathbb{C}^m$. This finally yields the stated conclusion.

References

[1] J.-M. Azais and M. Wschebor. *Level Sets and Extrema of Random Processes and Fields.* John Wiley & Sons Inc., Hoboken, NJ, 2009.

[2] W. U. Bajwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak. Toeplitz-structured compressed sensing matrices. In *Proc. IEEE Stat. Sig. Proc. Workshop*, pages 294–298, Madison, WI, August 2007.

[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. and Comp. Harmonic Analysis*, 27(3):265–274, November 2009.

[5] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, 1(3):1583–1614, 2003.

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

[7] E. J. Candès. Compressive sampling. In *Proceedings of the International Congress of Mathematicians*, Madrid, Spain, 2006.

[8] E. J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Acad. Sci. Paris S'er. I Math.*, 346:589–592, 2008.

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

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

[11] 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.

[12] 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.

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

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

[15] V. H. de la Peña and E. Giné. *Decoupling: From Dependence to Independence*. Springer, 1999.

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

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

[18] M. Fornasier and H. Rauhut. Compressive sensing. In O. Scherzer, editor, *Handbook of Mathematical Methods in Imaging*. Springer, to appear.

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

[20] S. Foucart. Hard thresholding pursuit: an algorithm for compressive sensing. *preprint*, 2010.

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

[22] S. Foucart, A. Pajor, H. Rauhut, and T. Ullrich. The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$. *J. Complexity*, to appear.

[23] A. Y. Garnaev and E. D. Gluskin. On widths of the Euclidean ball. *Sov. Math., Dokl.*, 30:200–204, 1984.

[24] S. R. Gottesman and E. E. Fenimore. New family of binary arrays for coded aperture imaging. *Appl. Opt.*, 28:4344–4352, 1989.

[25] J. Haupt, W. Bajwa, G. Raz, and R. Nowak. Toeplitz compressed sensing matrices with applications to sparse channel estimation. *To appear in IEEE Trans. Inform. Theory*, 2010.

[26] M. A. Herman and T. Strohmer. High-resolution radar via compressed sensing. *IEEE Trans. Sig. Proc.*, 57(6):2275–2284, June 2009.

[27] A. Hinrichs and J. Vybíral. Johnson-Lindenstrauss lemma for circulant matrices. *Random Struct. Algorithms*, to appear.

[28] F. Krahmer and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *preprint*, 2010.

[29] M. Ledoux and M. Talagrand. *Probability in Banach Spaces.* Springer-Verlag, Berlin, Heidelberg, NewYork, 1991.

[30] R. F. Marcia, C. Kim, J. Kim, D. Brday, and R. M. Willett. Fast disambiguation of superimposed images for increased field of view. In *Proc. IEEE Int. Conf. Image Proc.*, pages 2620–2623, 2008.

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

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

[33] D. Needell and J. Tropp. COSAMP: Iterative signal recovery from incomplete and inaccurate measurements. *Appl. and Comp. Harmonic Analysis*, 26(3):301–321, May 2009.

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

[35] G. Pisier. Remarques sur un résultat non publié de B. Maurey. In *Seminar on Functional Analysis, 1980–1981*, pages Exp. No. V, 13. École Polytech., Palaiseau, 1981.

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

[37] H. Rauhut. Circulant and Toeplitz matrices in compressed sensing. In Rémi Gribonval, editor, *SPARS'09 - Signal Processing with Adaptive Sparse Structured Representations*, Saint Malo France, 2009. Inria Rennes - Bretagne Atlantique.

[38] 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.

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

[40] H. Rauhut and R. Ward. Sparse Legendre expansions via l_1-minimization. *preprint*, 2010.

[41] M. Richards. *Fundamentals of Radar Signal Processing.* McGraw-Hill, 2005.

[42] J. Romberg. Compressive sampling via random convolution. In *Proc. Int'l Cong. Indus. and Appl. Math. (ICIAM)*, Zurich, Switzerland, July 2007.

[43] J. Romberg. Compressive sensing by random convolution. *SIAM J. Imaging Sci.*, 2(4):1098–1128, 2009.

[44] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61(8):1025–1045, 2008.

[45] M. Talagrand. *The Generic Chaining.* Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005.

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

[47] J. A. Tropp, J. N. Laska, M. F. Duarte, J. Romberg, and R. G. Baraniuk. Beyond Nyquist: efficient sampling of sparse bandlimited signals. *IEEE Trans. Inform. Theory*, 56(1), January 2010.

[48] J. A. Tropp, M. B. Wakin, M. F. Duarte, D. Baron, and R. G. Baraniuk. Random filters for compressive sampling and reconstruction. In *Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc.*, volume 3, pages III–872–875, Toulouse, France, May 2006.

[49] J. Vybíral. A variant of the Johnson-Lindenstrauss lemma for circulant matrices. *preprint*, 2010.