The restricted isometry property for random convolutions

Felix Krahmer
Institute for Numerical and Applied Mathematics
University of Göttingen
Lotzestraße 16-18
37085 Göttingen, Germany
Email: f.krahmer@math.uni-goettingen.de

Shahar Mendelson
Department of Mathematics
Technion
Haifa 32000, Israel
Email: shahar@tx.technion.ac.il

Holger Rauhut
RWTH Aachen University
Lehrstuhl C für Mathematik (Analysis)
Templergraben 55
52056 Aachen, Germany
Email: rauhut@mathc.rwth-aachen.de

***Abstract*—We present significantly improved estimates for the restricted isometry constants of partial random circulant matrices as they arise in the matrix formulation of subsampled convolution with a random pulse. We show that the required condition on the number m of rows in terms of the sparsity s and the vector length n is $m \gtrsim s \log^2 s \log^2 n$.**

I. Introduction

The theory of *compressed sensing* is based on the observation that many natural signals are approximately sparse in appropriate representation systems, that is, only few entries are significant. The goal of the theory is to devise methods to recover such a signal $\boldsymbol{x}$ from linear measurements

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

For example, it has been shown [1] that under the assumption of a small *restricted isometry constant* on the matrix $\boldsymbol{\Phi}$, approximate recovery via ℓ_1-minimization

$$\min_{\boldsymbol{z}} \|\boldsymbol{z}\|_1 \quad \text{subject to } \boldsymbol{\Phi z} = \boldsymbol{y},$$

(where $\|\boldsymbol{z}\|_p$ denotes the usual ℓ_p-norm) is guaranteed even in the presence of noise.

Here, for a matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ and $s < n$, the restricted isometry constant $\delta_s = \delta_s(\boldsymbol{\Phi})$ is defined as the smallest number such that

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

If a matrix has a small restricted isometry constant, we also say that the matrix has the *restricted isometry property (RIP)*.

A class of measurement models that is of particular relevance for sensing applications is that of subsampled convolution with a random pulse. In such a model, the convolution of a signal $\boldsymbol{x} \in \mathbb{R}^n$ with a random vector $\boldsymbol{\epsilon} \in \mathbb{R}^n$ given by

$$\boldsymbol{x} \mapsto \boldsymbol{\epsilon} * \boldsymbol{x}, (\boldsymbol{\epsilon} * \boldsymbol{x})_k = \sum_{j=1}^{n} \epsilon_{(k-j) \bmod n}\, x_j.$$

is followed by a restriction $\boldsymbol{P}_\Omega$ to a deterministic subset of the coefficients $\Omega \subset \{1,\dots,n\}$ and normalization of the columns. The resulting measurement map is linear; its matrix representation $\boldsymbol{\Phi}$ given by

$$\boldsymbol{\Phi x} = \frac{1}{\sqrt{m}} \boldsymbol{\epsilon} * \boldsymbol{x}$$

is called a *partial random circulant matrix*. In this paper, we will focus on the case that the random vector $\boldsymbol{\epsilon}$ is a Rademacher random vector, that is, its entries are independent random variables with distribution $\mathbb{P}(\epsilon_i = \pm 1) = 1/2$. Note, however, that the corresponding results in [2] consider more general random vectors.

The problem of proving the RIP for subsampled convolutions has first been considered in [3]; these results have later been improved in [4]. In [5], a similar problem is considered. Both the sampling sets and the generators, however, are chosen at random. In contrast, our result below holds for an arbitrary fixed sampling sets $\Omega \subset \{1,\dots,n\}$, which is important in applications since in many practical problems, it is natural or desired to consider structured sampling sets such as $\Omega = \{L, 2L, 3L, \dots, mL\}$ for some $L \in \mathbb{N}$; these sets are clearly far from being random.

This paper is structured as follows. In Section II, we present our main result and compare it to the previously best known results. Section IV formulates the problem in terms of chaos processes and presents bounds for such processes in terms of complexity parameters, which are introduced before that in Section III. These bounds are then used to prove the main result in Section V.

II. Main result

Theorem II.1. *([2]) Let $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ be a draw of a partial random circulant matrix generated by a Rademacher vector $\boldsymbol{\epsilon}$. If*

$$m \ge c\delta^{-2} s\, (\log^2 s)(\log^2 n), \tag{1}$$

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

This result improves the best previously known estimates for a partial random circulant matrix [4], namely that $m \ge C_\delta (s \log n)^{3/2}$ is a sufficient condition for achieving $\delta_s \le \delta$ with high probability. In particular, Theorem II.1 removes the exponent $3/2$ of the sparsity s, which was already conjectured in [4] to be an artefact of the proof.

Remark II.2. *In certain application scenarios, the ambient dimension n as well as the number of measurements m may* *be given, while one is interested on the sparsity level that still guarantees recovery. To obtain such a bound, we estimate the logarithmic factors in s by $\log(n)$, so we obtain the condition $s \leq \frac{m}{\log^4(n)}$. Again, the dependence is linear up to logarithmic factors, which cannot be guaranteed using previous bounds.*

III. Important concepts and definitions

In the proof, two types of complexity parameters of a set of matrices $\mathcal{A}$ will play an important role. The first one, denoted by $d_F(\mathcal{A})$ and $d_{2\to2}(\mathcal{A})$, is the radius of $\mathcal{A}$ in the Frobenius norm $\|\boldsymbol{A}\|_F = \sqrt{\operatorname{tr}(\boldsymbol{A}^*\boldsymbol{A})}$ and the operator norm $\|\boldsymbol{A}\|_{2\to2} = \sup_{\|\boldsymbol{x}\|_2\leq1}\|\boldsymbol{A}\boldsymbol{x}\|_2$, respectively. That is, $d_F(\mathcal{A}) = \sup_{\boldsymbol{A}\in\mathcal{A}}\|\boldsymbol{A}\|_F$ and $d_{2\to2}(\mathcal{A}) = \sup_{\boldsymbol{A}\in\mathcal{A}}\|\boldsymbol{A}\|_{2\to2}$. The second one, Talagrand's functional $\gamma_2(\mathcal{A}, \|\cdot\|_{2\to2})$, is given by the following definition.

Definition III.1 ([6])**.** *For a metric space (T, d), an admissible sequence of T is a collection of subsets of T, $\{T_s : s \geq 0\}$, such that for every $s \geq 1$, $|T_s| \leq 2^{2^s}$ and $|T_0| = 1$. Then the γ_2 functional is given by*

$$\gamma_2(T, d) = \inf \sup_{t\in T} \sum_{s=0}^{\infty} 2^{s/2} d(t, T_s),$$

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

Recall that for a metric space (T, d) and $u > 0$, the covering number $N(T, d, u)$ is the minimal number of open balls of radius u in (T, d) needed to cover T. The γ_2-functionals can be bounded in terms of such covering numbers by the well-known Dudley integral (see, e.g., [6]). A formulation specific to a set of matrices $\mathcal{A}$ endowed with the operator norm is

$$\gamma_2(\mathcal{A}, \|\cdot\|_{2\to2}) \leq C \int_0^{d_{2\to2}(\mathcal{A})} \sqrt{\log N(\mathcal{A}, \|\cdot\|_{2\to2}, u)}\,du \qquad (2)$$

for some absolute constant C.

IV. Reformulation as a chaos process

Let $\boldsymbol{\Phi}$ be a partial circulant matrix based on a Rademacher vector, then

$$\begin{aligned}\delta_s(\boldsymbol{\Phi}) &= \sup_{\substack{\boldsymbol{x}\in S^{n-1}\\ |\operatorname{supp}\boldsymbol{x}|\leq s}} \left|\|\boldsymbol{\Phi}\boldsymbol{x}\|_2^2 - 1\right| \\ &= \sup_{\substack{\boldsymbol{x}\in S^{n-1}\\ |\operatorname{supp} x|\leq s}} \left|\left\|\frac{1}{\sqrt{m}}\boldsymbol{P}_\Omega \boldsymbol{x} * \boldsymbol{\epsilon}\right\|_2^2 - 1\right| \\ &= \sup_{\substack{\boldsymbol{x}\in S^{n-1}\\ |\operatorname{supp}\boldsymbol{x}|\leq s}} \left|\|\boldsymbol{V_x}\boldsymbol{\epsilon}\|_2^2 - \mathbb{E}\|\boldsymbol{V_x}\boldsymbol{\epsilon}\|_2^2\right|,\end{aligned}$$

where $\boldsymbol{V_x}$ is defined through $\boldsymbol{V_x}\boldsymbol{y} := \frac{1}{\sqrt{m}}\boldsymbol{P}_\Omega\boldsymbol{x} * \boldsymbol{y}$.

As it turns out, the expression $\|\boldsymbol{V_x}\boldsymbol{\epsilon}\|_2^2$ is a Rademacher chaos process, that is, it is of the form $\langle\boldsymbol{\epsilon}, \boldsymbol{M}\boldsymbol{\epsilon}\rangle$. This observation was already exploited in [4] to obtain their suboptimal bounds. Our result, however, incorporates the additional observation that the matrix M in the above scenario is $\boldsymbol{Vx}^*\boldsymbol{Vx}$, hence positive semidefinite.

In the following, we will provide a bound for suprema of chaos processes under such structural assumptions. That is, we study expressions of the form

$$\sup_{\boldsymbol{A}\in\mathcal{A}} \left|\|\boldsymbol{A}\boldsymbol{\epsilon}\|_2^2 - \mathbb{E}\|\boldsymbol{A}\boldsymbol{\epsilon}\|_2^2\right|.$$

Here $\mathcal{A}$ is an arbitrary set of matrices, which is assumed to be symmetric, i.e., $\mathcal{A} = -\mathcal{A}$.

Theorem IV.1 ([2])**.** *Let $\mathcal{A} \subset \mathbb{R}^{m\times n}$ be a symmetric set of matrices and let $\boldsymbol{\epsilon}$ be a Rademacher vector of length n. Then*

$$\begin{aligned}&\mathbb{E}\sup_{\boldsymbol{A}\in\mathcal{A}} \left|\|\boldsymbol{A}\boldsymbol{\epsilon}\|_2^2 - \mathbb{E}\|\boldsymbol{A}\boldsymbol{\epsilon}\|_2^2\right| \\ &\quad\leq C_1\left(d_F(\mathcal{A})\gamma_2(\mathcal{A}, \|\cdot\|_{2\to2}) + \gamma_2(\mathcal{A}, \|\cdot\|_{2\to2})^2\right) \\ &\quad=: C_1 E.\end{aligned}$$

Furthermore, for $t > 0$,

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

where $U = d_{2\to2}^2(\mathcal{A})$ and

$$V = d_{2\to2}(\mathcal{A})(\gamma_2(\mathcal{A}, \|\cdot\|_{2\to2}) + d_F(\mathcal{A})).$$

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

The proof of this theorem is based on decoupling and a chaining argument, see [2].

V. Proof of Theorem II.1

The proof will be mainly based on Theorem IV.1. Thus we need to control the parameters $d_{2\to2}(\mathcal{A})$, $d_F(\mathcal{A})$, as well as $\gamma_2(\mathcal{A}, \|\cdot\|_{2\to2})$ for the set

$$\mathcal{A} = \{\boldsymbol{V_x} : \boldsymbol{x} \in D_{s,N}\},$$

where

$$D_{s,N} = \{\boldsymbol{x} \in \mathbb{R}^N : |\operatorname{supp}\boldsymbol{x}| \leq s\}.$$

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

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

To bound $d_{2\to2}(\mathcal{A})$, we will use a Fourier domain description of $\boldsymbol{\Phi}$. Let $\boldsymbol{F}$ by the unnormalized Fourier transform with elements $F_{jk} = e^{2\pi ijk/n}$. As the Fourier transform diagonalizes the convolution operator, for every $1 \leq j \leq n$, $\boldsymbol{F}(\boldsymbol{x} * \boldsymbol{y})_j = (\boldsymbol{F}\boldsymbol{x})_j \cdot (\boldsymbol{F}\boldsymbol{y})_j$. Therefore,

$$\boldsymbol{V_x}\boldsymbol{\xi} = \frac{1}{\sqrt{m}}\boldsymbol{P}_\Omega\boldsymbol{F}^{-1}\widehat{\boldsymbol{X}}\boldsymbol{F}\boldsymbol{\xi},$$

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

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

where $\widehat{\boldsymbol{P}}_\Omega = \boldsymbol{P}_\Omega \boldsymbol{F}^{-1}$. Now observe that for every $\boldsymbol{x} \in D_{s,N}$ with the associated diagonal matrix $\widehat{\boldsymbol{X}}$,

$$\begin{aligned} \|\boldsymbol{V}_{\boldsymbol{x}}\|_{2\to 2} &= \frac{1}{\sqrt{m}} \|\widehat{\boldsymbol{P}}_\Omega \widehat{\boldsymbol{X}} \boldsymbol{F}\|_{2\to 2} \\ &\le \sqrt{\frac{n}{m}} \|\boldsymbol{P}_\Omega \boldsymbol{F}^{-1}\|_{2\to 2} \|\widehat{\boldsymbol{X}}\|_{2\to 2} \\ &\le \frac{1}{\sqrt{m}} \|\widehat{\boldsymbol{X}}\|_{2\to 2} \\ &= \frac{1}{\sqrt{m}} \|\boldsymbol{F}\boldsymbol{x}\|_\infty. \end{aligned} \tag{3}$$

Setting $\|\boldsymbol{x}\|_{\widehat{\infty}} := \|\boldsymbol{F}\boldsymbol{x}\|_\infty$ we observe that

$$\|\boldsymbol{F}\boldsymbol{x}\|_\infty \le \|\boldsymbol{x}\|_1 \le \sqrt{s}\|\boldsymbol{x}\|_2 \le \sqrt{s}$$

for every $x \in D_{s,N}$, and hence

$$d_{2\to 2}(\mathcal{A}) \le \sqrt{s/m}.$$

Next, to estimate the γ_2 functional, recall from (2) that

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

where C is an absolute constant. By (3),

$$\|\boldsymbol{V}_{\boldsymbol{x}} - \boldsymbol{V}_{\boldsymbol{y}}\|_{2\to 2} = \|\boldsymbol{V}_{\boldsymbol{x}-\boldsymbol{y}}\|_{2\to 2} \le m^{-1/2} \|\boldsymbol{x} - \boldsymbol{y}\|_{\widehat{\infty}},$$

and hence for every $u > 0$,

$$N(\mathcal{A}, \|\cdot\|_{2\to 2}, u) \le N(D_{s,N}, m^{-1/2}\|\cdot\|_{\widehat{\infty}}, u).$$

Such covering numbers and the corresponding Dudley integral have been bounded before, e.g., in the context of proving the restricted isometry property for partial random Fourier matrices [7]. The resulting bound for the γ_2-functional is

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

where C is an absolute constant. This implies that

$$\gamma_2(\mathcal{A}, \|\cdot\|_{2\to 2}) \le \frac{C}{c}\delta$$

for the given choice of m.

Now, by choosing the constant c in (1) appropriately, one obtains

$$E \le \frac{\delta}{2C_2},$$

where E and C_2 are chosen as in Theorem IV.1. Then Theorem IV.1 yields

$$\mathbb{P}(\delta_s \ge \delta) \le \mathbb{P}(\delta_s \ge C_2 E + \delta/2) \le \exp(-C_3(m/s)\delta^2),$$

which, after possibly increasing the value of c enough to compensate C_3, exactly amounts to the probability bound given in the theorem. □

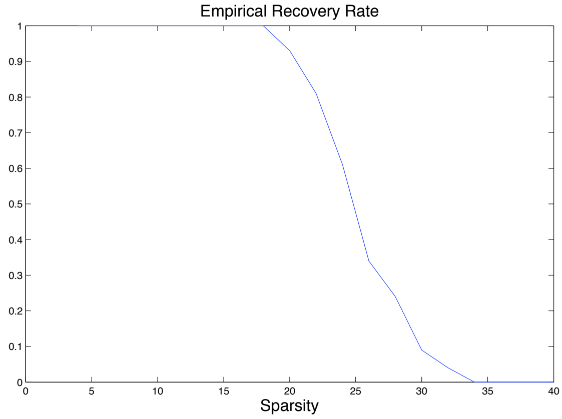

Fig. 1. Empirical recovery rate from partial random circulant measurements for $n = 500$, $m = 100$, and different sparsity levels

VI. Numerical Illustration

We illustrate our results by a numerical example, considering signals of length $n = 500$ and $m = 100$ measurements, letting the sparsity vary. We used a partial random circulant matrix based on a Bernoulli vector, where the rows are selected at random. The plot shows the empirical success rate, that is, in which fraction of the trials the correct signal was recovered (see Figure 1). One should note that our rather simple tests depict the non-uniform success rate: Given a signal, what is the probability that it can be recovered from randomly generated measurements? What we proved above are uniform recovery guarantees: With high probability, a single randomly chosen matrix allows for the recovery of all sparse vectors. This property is much harder to check, as one needs to find the worst vector. While we leave such tests in the context of partial random circulant matrices for future work, we note that strategies to check for this property have been investigated recently in [8].

VII. Conclusion

In this paper we derive bounds on the embedding dimension necessary for a partial random circulant matrix, which is linear in the sparsity. This improves on previous results, in which the sparsity appears with an exponent of $\frac{3}{2}$.

References

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

[2] F. Krahmer, S. Mendelson, and H. Rauhut. Suprema of chaos processes and the Restricted Isometry Property. *Comm. Pure Appl. Math.*, to appear.

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

[4] H. Rauhut, J. K. Romberg, and J. A. Tropp. Restricted isometries for partial random circulant matrices. *Appl. Comput. Harmon. Anal.*, 32(2):242–254, 2012.

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

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

[7] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.
[8] J. Blanchard, C. Cartis, and J. Tanner. Compressed Sensing: How sharp is the Restricted Isometry Property? *SIAM Review*, 53(1):105–125, 2011.