Improved bounds for sparse recovery from subsampled random convolutions

Shahar Mendelson*, Holger Rauhut†, Rachel Ward‡

October 17, 2016

Abstract

We study the recovery of sparse vectors from subsampled random convolutions via ℓ_1-minimization. We consider the setup in which both the subsampling locations as well as the generating vector are chosen at random. For a subgaussian generator with independent entries, we improve previously known estimates: if the sparsity s is small enough, i.e. $s \lesssim \sqrt{n/\log(n)}$, we show that $m \gtrsim s\log(en/s)$ measurements are sufficient to recover s-sparse vectors in dimension n with high probability, matching the well-known condition for recovery from standard Gaussian measurements. If s is larger, then essentially $m \geq s\log^2(s)\log(\log(s))\log(n)$ measurements are sufficient, again improving over previous estimates. Moreover, we also provide robustness estimates for measurement errors that are bounded in ℓ_q for $q > 2$ – in particular, allowing the case $q = \infty$ which is important for quantized compressive sensing. All these results are shown via ℓ_q-robust versions of the null space property and for $q > 2$ they represent the first non-trivial bounds for structured random matrices. As a crucial ingredient, our approach requires to lower bound expressions of the form $\inf_{v \in V_r} \|\Gamma_v \xi\|_q$, where Γ_v is a set of matrices indexed by unit norm r-sparse vectors and ξ is a subgaussian random vector. This involves the combination of small ball estimates with chaining techniques.

1 Introduction

Compressive sensing [6, 13, 17] considers the recovery of (approximately) sparse vectors from incomplete and possibly perturbed linear measurements via efficient algorithms such as ℓ_1-minimization. Provably optimal bounds for the minimal number of required measurements in terms of the sparsity have been shown for Gaussian and, more generally, subgaussian random matrices [4, 9, 13, 14, 17, 28, 29, 39].

Practical applications demand for structure in the measurement process which is clearly not present in Gaussian random matrices with independent entries. Several types of structured random matrices have been studied, including random partial Fourier matrices [6, 7, 39, 34, 37, 5], partial random circulant matrices (subsampled random convolutions) [24, 33, 34, 36, 38],

*Department of Mathematics, Technion, Haifa, 32000, Israel, shahar@tx.technion.ac.il & Centre for Mathematics and its Applications, The Australian National University, Canberra 0200 Australia

†Chair for Mathematics C (Analysis), RWTH Aachen University, rauhut@mathc.rwth-aachen.de

‡Mathematics Department, University of Texas at Austin, 2515 Speedway, Austin, Texas, 78712, rward@ma.utexas.edu

time-frequency structured random matrices [24, 35, 32], and more [2, 20]. In this article, we improve known recovery results for partial random circulant matrices.

In mathematical terms, linear measurements of a signal (vector) $x \in \mathbb{R}^n$ can be written as

$$y = Ax \quad \text{with } A \in \mathbb{R}^{m \times n},$$

and we are particularly interested in the case $m < n$. Compressive sensing predicts that this system can be solved for x using efficient algorithms if x is sparse enough, say in the sense that $\|x\|_0 = |\{\ell : x_\ell \neq 0\}|$ is small. While ℓ_0-minimization is NP-hard [17], several tractable algorithms have been introduced as alternatives, most notably ℓ_1-minimization (basis pursuit) [6, 10, 13, 17] which produces a minimizer of

$$\min_{z \in \mathbb{R}^n} \|z\|_1 \text{ subject to } Az = y.$$

If $A \in \mathbb{R}^{m \times n}$ is a random draw of a Gaussian matrix, i.e., all entries are standard normal random variables, then with probability at least $1 - e^{-cm}$, any s-sparse vector $x \in \mathbb{R}^n$, (i.e., $\|x\|_0 \leq s$), can be reconstructed in a stable way (see below) using ℓ_1-minimization from the given data $y = Ax$, provided that

$$m \geq Cs \ln(en/s) \tag{1.1}$$

for some absolute constant $C > 0$. This bound is optimal [13, 16, 17] in the sense that the combination of any recovery algorithm with any measurement matrix requires at least (1.1) many measurements in order to achieve stable reconstruction, i.e., for any $x \in \mathbb{R}^n$ (not necessarily s-sparse), the reconstruction $x^\sharp$ obtained from $y = Ax$, satisfies

$$\|x - x^\sharp\|_1 \leq C\sigma_s(x)_1 := C \inf\{\|x - z\|_1 : \|z\|_0 \leq s\}, \tag{1.2}$$

see [16, Theorem 2.7] for details. Moreover, *exact* s-sparse recovery via ℓ_1-minimization necessarily requires (1.1), see [16, Lemma 2.4].

Unfortunately, Gaussian random matrices are not suitable for many applications of compressive sensing – because of their lack of structure. In fact, structure is required in order to model realistic measurement scenarios and also to speed up the matrix-vector-multiplications that have to be applied many times in known ℓ_1-minimization algorithms.

An important example of structured random matrices are $m \times n$ matrices that are generated from the $n \times n$ discrete Fourier (more generally, from a Hadamard type matrix, see Definition 2.6 below), by randomly subsampling m rows. This corresponds to taking m random samples of the discrete Fourier transform of a vector. Again, ℓ_1-minimization successfully recovers s-sparse vectors with probability at least $1 - \varepsilon$ provided that

$$m \geq Cs \max\{\log^2(s) \log(n), \log(\varepsilon^{-1})\},$$

see, for example, [7, 39, 34, 5, 19].

In this article we will be concerned with subsampled random convolutions. The circular convolution on $\mathbb{R}^n$ is defined for two vectors $x, \xi \in \mathbb{R}^n$ as

$$(x * \xi)_k = \sum_{j=1}^{n} x_j \xi_{k-j \bmod n+1}, \quad k = 1, \dots, n.$$

For a subset $\Omega \subset \{1, \ldots, n\} =: [n]$ of cardinality m, let $P_\Omega : \mathbb{R}^n \to \mathbb{R}^m$ be the projection onto the coordinates indexed by Ω, i.e., $(P_\Omega x)_j = x_j$ for $j \in \Omega$. A subsampled convolution is defined as

$$Bx = P_\Omega(x * \xi) \tag{1.3}$$

and the corresponding matrix is a partial circulant matrix generated by ξ. Subsampled random convolutions find applications in radar and coded aperture imaging [3, 18, 38, 42], as well as in fast dimensionality reduction maps [25].

It was shown in [24] that if ξ is a (standard) Gaussian vector and Ω is an arbitrary (deterministic) subset of cardinality m, then with probability at least $1 - \varepsilon$, every s-sparse vector can be reconstructed from Bx via ℓ_1-minimization if

$$m \geq Cs \max\{\log^2(s) \log^2(n), \log(\varepsilon^{-1})\}. \tag{1.4}$$

Moreover, stability in the sense of (1.2) holds for such matrices, and the results are robust when the given measurements are corrupted by noise (see more details below). Moreover, the recovery result can be extended to circulant matrices generated by a subgaussian random vector – an object of central importance to our discussion which will be defined later.

Our focus is on sparse recovery via subsampled random convolutions, where the set Ω is chosen at random via independent selectors: let $(\delta_i)_{i=1}^n$ be independent, $\{0, 1\}$-valued random variables with mean $\delta = m/n \in (0, 1]$, and set $\Omega = \{i : \delta_i = 1\}$. Then the expected size of Ω is $\mathbb{E}|\Omega| = m$ and it follows from Bernstein's inequality that $m/2 \leq |\Omega| \leq 3m/2$ with probability at least $1 - 2\exp(-m/9)$.

For the sake of simplicity of this exposition, we shall first formulate our main theorem for a standard Gaussian generator, i.e., a random vector with independent, mean zero, variance one, normally distributed coordinates. However, the proof we present holds for more general L-subgaussian random vectors with independent coordinates and a more general type of random matrices (see Theorem 2.5).

Theorem 1.1 *Let $\xi \in \mathbb{R}^n$ be a random draw of a standard Gaussian random vector and let $\Omega \subset [n]$ be chosen at random, using independent selectors of mean $\delta = m/n$. Let B be the corresponding partial random circulant matrix defined in* (1.3)*. Let $s \leq c_1 \frac{n}{\log^4(n)}$ and assume that*

$$\begin{aligned} m &\geq c_3 s \log(en/s) && \textit{if } s \leq c_2\sqrt{\tfrac{n}{\log(n)}} \\ m &\geq c_3 s \log(en/s) \alpha_s^2 \log(\alpha_s) && \textit{if } c_2\sqrt{\tfrac{n}{\log(n)}} \leq s \leq c_1 \tfrac{n}{\log^4(n)}, \end{aligned} \tag{1.5}$$

where $\alpha_s = \log\left(\frac{s^2}{n} \max\{\log(en/s), \log(s)\}\right)$. Then with probability at least

$$1 - 2\exp\left(-c_0 \min\left\{\frac{n}{s}, s\log(en/s)\right\}\right)$$

the following holds. For all $x \in \mathbb{R}^n$, all $e \in \mathbb{R}^m$ with $\|e\|_2 \leq \eta$ and $y = Bx + e$, the minimizer $x^\sharp$ of

$$\min \|z\|_1 \quad \textit{subject to } \|Bz - y\|_2 \leq \eta \tag{1.6}$$

satisfies

$$\|x - x^\sharp\|_1 \le C\sigma_s(x)_1 + D\frac{\sqrt{s}\eta}{\sqrt{m}} \qquad \text{and} \tag{1.7}$$

$$\|x - x^\sharp\|_2 \le C\frac{\sigma_s(x)_1}{\sqrt{s}} + D\frac{\eta}{\sqrt{m}}. \tag{1.8}$$

Our estimates indicate a phase-transition that occurs when s is roughly of the order of $\sqrt{n}$. Below this level, the partial circulant matrix exhibits the same behavior as the Gaussian matrix (which is the optimal scaling of the number of measurements m as a function of the sparsity parameter s) – it requires $Cs\log(en/s)$ measurements to recover an s sparse vector. Above that level, more measurements are required; for example, if $s = n^\alpha$ for $1/2 < \alpha < 1$ then $c(L,\alpha)n^\alpha \log^3 n \cdot \log\log n$ measurements are needed.

As we will see later, the phase transition at $\sqrt{n/\log(n)}$ is not a coincidence – the analysis required in the low-sparsity case is truly different from the one needed to deal with the high-sparsity one. However, it is presently not clear whether the analysis for the high-sparsity case can be improved in order to remove the additional logarithmic factors.

In both cases (low and high sparsity) we improve the estimates from [24], though it should be noted that (1.4) applies to *any* set $\Omega \subset [n]$ of cardinality m, while (1.5) applies only to randomly chosen Ω. A random selection Ω has been considered in [38], but the estimates there require $m \ge Cs\log^6(n)$. On the other hand, [38] applies to vectors that are sparse in an arbitrary (fixed) orthonormal basis and not only in the canonical basis; our proof technique does not seem to extend to this case in a simple way.

Moreover, we stress that (1.1) provides a uniform recovery guarantee in the sense that a single random draw of the partial circulant matrix is able to recovery all s-sparse vectors simultaneously. This is in contrast to other previous so-called nonuniform results found in the literature [33, 34, 21] that only imply recovery of a fixed sparse vector from a random draw of the matrix. Moreover, these nonuniform results give no or weaker stability estimates than (1.7) and (1.8), see e.g. [17, Theorem 4.33] or [15].

Another improvement on known estimates is that our results hold for noisy measurements when the noise is bounded in ℓ_q for $q \ge 2$ and the ℓ_2-constraint in (1.6) is replaced by an ℓ_q-constraint (and the error estimates scale with the ℓ_q norm of the noise), see Theorem 6.1 for details. This allows us, for example, to explore quantized compressive sensing (see, e.g., [12]), when the quantization error has a natural ℓ_∞-bound. In contrast, all the recovery results mentioned above were derived via the restricted isometry property (RIP), and applying the RIP, knowing only that the noise is bounded in ℓ_q for $q > 2$, leads to a poor scaling of the number of required measurements in terms of the degree of sparsity (see [12] for a detailed discussion on this issue). The reason why we can handle bounded noise in ℓ_q for $q > 2$ is that the proof we present is not based on the (two-sided) RIP, but rather on suitable versions of the null-space property, which we define in the next section. Our new bound provides the first rigorous proof of the ℓ_q-robust null space for structured random matrices for $q > 2$ with an optimal scaling of the number of measurements in the sparsity up to possibly logarithmic factors, see Theorem 6.1 for details.

We note that for a few other constructions of structured random matrices (with fast matrix-vector multiplication), recovery results with the optimal number of measurements (1.1) have been shown under similar size restrictions on the sparsity as in our main theorem above [1, 2]. However, it seems that our construction is the simplest one and is arguably the only one among these which models a physically realizable measurement device. In contrast to these previous results, we are able to extend our bounds to the near-linear sparsity regime at the cost of some additional logarithmic factors.

Independently of our main results themselves, we believe that our proof techniques should be of interest as well. In fact, the crucial ingredient in our proof is a probabilistic lower bound on terms of the form $\inf_{v\in V_r}\|\Gamma_v\xi\|_q$, where Γ_v are matrices indexed by a set of unit norm r-sparse vectors and ξ is subgaussian random vector with independent coordinates. We use a combination small ball estimates, covering number bounds and chaining techniques. We are not aware that our proof technique was used before in a similar way and context. We remark that our main ingredient can be generalized to random vectors with heavier tails and less independence assumptions, for instance, log-concave random vectors. Details will be presented in a future contribution.

The article is structured as follows. Section 2 discusses preliminaries such as the null space property, subgaussian random vectors, states the main result and gives a brief explanation of its proof. Section 3 introduces the small ball estimates required for the proof as well as moment estimates for norms of subgaussian random vectors. It further provides some covering number estimates required in the sequel. Section 4 provides the main technical ingredient of the proof of our main results which consists in a lower bound for $\inf_{v\in V_r}\|\Gamma_v\xi\|_2$. Section 5 provides an upper bound over the one-sparse vectors, which is required for completing the proof of the null space property, see also Theorem 2.3. Finally, Section 6 provides the extension of our recovery result to robustness in ℓ_q for $q > 2$.

Acknowledgements

The authors would like to thank Grigoris Paouris for very valuable discussions on the subject of this article. S.M. and H.R. acknowledge funding by the German Israel Foundation (GIF) through the project *Analysis of structured random measurements in recovery problems* (G-1266-304.6/2014). R.W. was funded by NSF CAREER grant #1255631.

2 Preliminaries and main result

2.1 The null space property

Our analysis is based on a robust version of the null space property which is a sufficient and necessary condition for sparse recovery via ℓ_1-minimization. This version is stable when passing to approximately sparse vectors, and is robust when the measurements are noisy. We begin with several standard facts on this.

Given $v \in \mathbb{R}^n$ and $S \subset [n] = \{1, \dots, n\}$, let $v_S \in \mathbb{R}^n$ with entries $(v_S)_j = v_j$ for $j \in S$ and $(v_S)_j = 0$ for $j \notin S$. Further, $S^c = [n] \setminus S$ denotes the complement of S.

Definition 2.1 *For $1 \leq q \leq \infty$, a matrix A satisfies the ℓ_q-robust null-space property of order s with constants $\nu \in (0, 1)$ and $\tau > 0$ if*

$$\|v_S\|_2 \leq \frac{\nu}{\sqrt{s}}\|v_{S^c}\|_1 + \tau\|Av\|_q$$

for every $v \in \mathbb{R}^n$ and every $S \subset [n]$ of cardinality at most s.

The following result is standard by now (see, e.g., [17, Theorem 4.22]). It uses the notion of the error of best s-term approximation, defined as

$$\sigma_s(x)_1 = \min_{z:\|z\|_0 \le s} \|x - z\|_1;$$

that is, $\sigma_s(x)_1$ is the ℓ_1 distance between x and the set of s-sparse vectors.

Theorem 2.2 *Let $1 \le q \le \infty$ and let A satisfy the ℓ_q-robust null space property of order s with constants $\nu \in (0,1)$ and $\tau > 0$. Let $\|e\|_q \le \eta$, $x \in \mathbb{R}^n$ and put $y = Ax + e$. Then a minimizer $x^\sharp$ of*

$$\min_{z \in \mathbb{R}^n} \|z\|_1 \quad \text{subject to } \|Az - y\|_q \le \eta$$

satisfies

$$\|x - x^\sharp\|_1 \le C\sigma_s(x)_1 + D\sqrt{s}\eta \tag{2.1}$$

$$\|x - x^\sharp\|_2 \le \frac{C}{\sqrt{s}}\sigma_s(x)_1 + D\eta, \tag{2.2}$$

where the constants are given by $C = \frac{(1+\nu)^2}{1-\nu}$ and $D = \frac{(3+\nu)}{1-\nu}\tau$.

Roughly speaking, even if x is not s-sparse, but only approximated by an s-sparse vector, and if one receives linear measurements of x (i.e., Ax) that are corrupted by the "noise" e, then a solution to the minimization problem still yields a good approximation of x if A possesses the null space property. In particular, if x is s-sparse then $\sigma_s(x)_1 = 0$, and if $\eta = 0$ (no noise), then the reconstruction via equality constrained ℓ_1-minimization is exact.

In order to show the ℓ_q-robust null-space property, we will proceed in the following way. Let

$$\mathcal{T}_{\nu,s} := \left\{ v \in \mathbb{R}^n : \|v_S\|_2 \ge \frac{\nu}{\sqrt{s}} \|v_{S^c}\|_1 \right\}.$$

One may show (see, e.g., [12, 22]) that if

$$\inf_{x \in \mathcal{T}_{\nu,s} \cap S^{n-1}} \|Ax\|_q \ge \frac{1}{\tau},$$

then A satisfies the ℓ_q-robust null space property with constants ν and τ. Moreover, if we set

$$V_s = \{x \in \mathbb{R}^n : \|x\|_0 \le s, \|x\|_2 = 1\}$$

to be the set of s-sparse vectors in the unit sphere, then ([22, Lemma 3], see also [39])

$$\mathcal{T}_{\nu,s} \cap S^{n-1} \subset (2 + \nu^{-1}) \operatorname{conv} V_s, \tag{2.3}$$

allowing one to study $\operatorname{conv} V_s$ instead of $\mathcal{T}_{\nu,s}$, where $\operatorname{conv} S$ denotes the convex hull of the set S, that is, the set of all convex combinations of finite subsets of S.

It turns out that one may replace $\operatorname{conv} V_s$ with V_r for r sufficiently large by adding a condition on one-sparse vectors. This was observed for $q = 2$ in [31, Lemma 5.1] (see also [26, Theorem B]). We will extend this result to $q \ge 2$ in Theorem 6.2, which however comes with worse constants for $q = 2$ than the statement below.

Theorem 2.3 *Let $A \in \mathbb{R}^{m\times n}$ satisfy*

$$\inf_{x\in V_r} \|Ax\|_2 \geq \tau^{-1} \quad \text{and} \quad \max_{j\leq[n]} \|Ae_j\|_2 \leq M. \tag{2.4}$$

If $c(\nu) = \nu^2/(2\nu+1)^2$ and

$$s \leq c(\nu)\frac{r-1}{M^2\tau^2-1},$$

then

$$\inf_{x\in \mathcal{T}_{\nu,s}} \|Ax\|_2 \geq \frac{1}{\sqrt{2}\,\tau}.$$

Proof. By (2.3), it suffices to show that the conditions in (2.4) imply that

$$\inf_{x\in(2+\nu^{-1})\operatorname{conv} V_s\cap S^{n-1}} \|Ax\|_2 \geq 1/(\sqrt{2}\tau).$$

Applying [26, Lemma 2.6] (see also the proof of Lemma 6.2 specialized to $q=2$), it follows from (2.4) that for any $y \in \mathbb{R}^n$

$$\|Ay\|_2^2 \geq \tau^{-2}\|y\|_2^2 - \frac{1}{r-1}\left(\|y\|_1\sum_{j=1}^n \|Ae_j\|_2^2|y_j| - \tau^{-2}\|y\|_1^2\right).$$

Let B_1^n be the unit ball in ℓ_1^n and observe that $\operatorname{conv} V_s \subset \sqrt{s}B_1^n$. Thus, if $c_1(\nu) = 2+1/\nu$ and $y \in S^{n-1}\cap c_1(\nu)\operatorname{conv} V_s$,

$$\|y\|_2 = 1 \quad \text{and} \quad \|y\|_1 \leq c_1(\nu)\sqrt{s}.$$

Therefore,

$$\|Ay\|_2^2 \geq \tau^{-2}\left(1-\frac{\|y\|_1^2}{r-1}(\tau^2\max_{1\leq j\leq n}\|Ae_j\|_2-1)\right) \geq \tau^{-2}\left(1-\frac{c_1^2(\nu)s}{r-1}\left(\tau^2M^2-1\right)\right) \geq \frac{1}{2\tau^2}$$

by our choice of s. ■

With Theorem 2.3 at hand, we will take the following course of action: we will show that

$$\inf_{x\in V_r} \|P_\Omega Ax\|_2 \gtrsim \sqrt{m} \quad \text{and} \quad \max_{j\in[n]} \|P_\Omega Ae_j\|_2 \lesssim \sqrt{m} \tag{2.5}$$

for a partial circulant matrix $P_\Omega A$, whose rows are chosen using iid selectors.

2.2 Subgaussian random vectors

Just as in [24] we will focus on generators ξ that are isotropic, L-subgaussian and have independent coordinates.

Definition 2.4 *A centered random vector $\xi = (\xi_i)_{i=1}^n$ is L-subgaussian if for every $x \in \mathbb{R}^n$,*

$$\|\langle \xi, x\rangle\|_{L_p} \leq L\sqrt{p}\|\langle \xi, x\rangle\|_{L_2}.$$

Assume that $\xi_1, \ldots, \xi_n$ are independent, mean-zero, variance 1, L-subgaussian random variables. In other words, for every $p \geq 1$,

$$\|\xi_i\|_{L_p} \leq L\sqrt{p}\|\xi_i\|_{L_2} = L\sqrt{p}.$$

Then ξ is an *isotropic* random vector on $\mathbb{R}^n$: for every $x \in \mathbb{R}^n$, $\mathbb{E}\langle \xi, x\rangle^2 = \|x\|_2^2$, and it is standard to verify that it is L-subgaussian as well (see, e.g., [43]). Is it straightforward to show that if ξ is an L-subgaussian random vector, then for every $x \in \mathbb{R}^n$ and $u \geq 1$,

$$Pr\left(|\langle \xi, x\rangle| \geq uL\|\langle \xi, x\rangle\|_{L_2}\right) \leq 2\exp(-cu^2),$$

for a suitable absolute constant c. If, in addition, ξ is isotropic, then

$$Pr\left(|\langle \xi, x\rangle| \geq uL\|x\|_2\right) \leq 2\exp(-cu^2).$$

2.3 Main result

Our main result provides a bound for the null-space property of partial random circulant matrices. It implies Theorem 1.1 via Theorem 2.2.

Recall that

- ξ is a random vector whose coordinates are independent, mean-zero, variance 1, L-subgaussian random variables;
- A is the complete circulant matrix generated by ξ;
- $(\delta_i)_{i=1}^n$ are independent selectors of mean $\delta = m/n$, and $\Omega = \{i : \delta_i = 1\}$.

Theorem 2.5 *There exist constants $c_0, \ldots, c_6$ that depend only on L, ρ and τ for which the following holds. With probability at least*

$$1 - 2\exp\left(-c_0 \min\left\{\frac{n}{r}, r\log\left(\frac{en}{r}\right)\right\}\right),$$

the partial circulant matrix $P_\Omega A$ satisfies the ℓ_2-robust null-space property of order r with constants $\rho \in (0,1)$ and $\tau/\sqrt{m}$, where

$$\delta n = c_3 r\log\left(\frac{en}{r}\right) \quad \text{if} \quad r \leq c_4\sqrt{\frac{n}{\log n}},$$

and

$$\delta n = c_3 r\log\left(\frac{en}{r}\right) \cdot \alpha_r^2 \log\alpha_r \quad \text{if} \quad c_4\sqrt{\frac{n}{\log n}} < r \leq c_5\frac{n}{\log^4 n}.$$

The constant α_r satisfies

$$\alpha_r \leq \log\left(c_6\frac{r^2}{n}\max\left\{\log\left(\frac{en}{r}\right), \log(er)\right\}\right).$$

The result will actually be shown for a more general class of random matrices, see Definition (2.6) and the remarks following it.

2.4 The heart of the argument

The proof of Theorem 2.5 has two main components. We will begin by analyzing the way a complete circulant matrix $A : \mathbb{R}^n \to \mathbb{R}^n$ generated by ξ acts on V_r, and then apply a random "selector projection" P_Ω to the image AV_r. Our primary goal is to obtain a lower bound on

$$\inf_{v \in V_r} \|P_\Omega Av\|_2^2 = \inf_{v \in V_r} \sum_{i=1}^{n} \delta_i \big\langle Av, e_i \big\rangle^2. \tag{2.6}$$

Thanks to the nature of circulant matrices, there is a standard representation of $\{Av : v \in V_r\}$ via the Fourier transform. Let $\mathcal{F}$ be the (un-normalized) Fourier transform, i.e., $\mathcal{F}_{j,k} = e^{-2\pi ijk/n}$, (which we treat as a "real operator" from $\mathbb{R}^n$ onto the image of $\mathbb{R}^n$ in order to avoid working with $\mathbb{C}^n$) and set $\hat{v} = \mathcal{F}v$. If $D_x = \mathrm{diag}(x_1, \ldots, x_n)$ then

$$Av = \mathcal{F}^{-1} D_{\hat{v}} \mathcal{F} \xi.$$

We will consider a more general set of matrices:

Definition 2.6 *An orthogonal matrix O is of Hadamard type with constant β if for every $i, j \in [n]$, $|O_{i,j}| \le \beta/\sqrt{n}$.*

In what follows we will fix three matrices, U, W and O, all of which are of Hadamard type with constant β, and for $x \in \mathbb{R}^n$ we set

$$\Gamma_x = \sqrt{n} U D_{Wx} O, \tag{2.7}$$

where $D_{Wx} = \mathrm{diag}((\langle W_i, x\rangle)_{i=1}^n)$. Clearly, the representation of Av is precisely of this form: if A is the complete circulant matrix with the generator ξ then for every $v \in \mathbb{R}^n$, $Av = \sqrt{n} U D_{Wv} O \xi$ for the choice of $U = n^{-1/2}\mathcal{F}^{-1}$ and $W = O = n^{-1/2}\mathcal{F}$; in this case $\beta = 1$.

From here on, for $V \subset \mathbb{R}^n$ set

$$\Gamma_V = \{\sqrt{n} U D_{Wv} O : v \in V\};$$

naturally, the set of matrices we will be interested in is Γ_{V_r}.

Observe that if ξ is an isotropic random vector then for every $v \in S^{n-1}$, $\mathbb{E}\|\Gamma_v \xi\|_2^2 = n$, and at least on average, for a single vector $v \in V_r$, one expects to have $\|\Gamma_v \xi\|_2 \sim \sqrt{n}$.

Unfortunately, showing that $\inf_{v \in V_r} \|\Gamma_v \xi\|_2 \ge c\sqrt{n}$ does not lead to a nontrivial lower bound on (2.6). To see why, set

$$x^1 = (\sqrt{n}, 0, \ldots, 0) \quad \text{and} \quad x^2 = (1, \ldots, 1).$$

Both x^1 and x^2 have a Euclidean norm of $\sqrt{n}$, but any attempt of selecting a random subset of coordinates of cardinality $m \ll n$ fails miserably for x^1 and succeeds for x^2: typically, $P_\Omega x^1 = 0$ while $\|P_\Omega x^2\|_2 = \sqrt{m}$. We will be looking at this type of "good behavior", exhibiting a (one sided) *standard shrinking* phenomenon. The term "one-sided standard shrinking" used in this context usually refers to a random projection operator T of rank m, for which, with high probability,

$$\|Tv\|_2 \ge c\sqrt{m/n}\|v\|_2 \quad \text{for all vectors of interest.}$$

The operator we are interested in is indeed random, and of the form $T = P_\Omega$ – a random coordinate projection – but as the example of x^1 shows, P_Ω may map x to 0 even if x has a large norm – unless one imposes some additional condition on x.

The condition we will focus on here is that x has a regular coordinate structure, that is, for suitable constants α and θ,

$$\left|\left\{ i : |x_i| \geq \|x\|_2 \frac{\alpha}{\sqrt{n}} \right\}\right| \geq \theta n. \tag{2.8}$$

The notion of regularity in (2.8) implies that $|x_i|$ is at least $\sim \|x\|_2/\sqrt{n}$ for a large subset of coordinates – of cardinality that is proportional to the dimension n. That set of coordinates contributes at least a proportion of the Euclidean norm of x, and moreover, for a random choice of $\Omega \subset [n]$, $\|P_\Omega x\|_2 \geq c\|x\|_2 \cdot \sqrt{|\Omega|/n}$ with high probability.

Thus, in addition to showing that $\inf_{v \in V_r} \|\Gamma_v \xi\|_2 \geq c\sqrt{n}$, we will prove that each of the vectors $\Gamma_v \xi$ is regular in the sense of (2.8). We will do so by representing a typical realization of the set $\{\Gamma_v \xi : v \in V_r\}$ as a subset of the Minkowski sum of two (random) sets $T_1 + T_2$ defined in the following way: Let $H \subset V_r$ be a fine enough net with respect to the Euclidean distance and set $T_1 = \{\Gamma_x \xi : x \in H\}$. For each $v \in V_r$, choose $x = \pi(v) \in H$ minimizing $\|x - v\|_2$. Then $T_2 = \{\Gamma_{v-\pi(v)} \xi : v \in V_r\}$. We will show the following properties of the sets T_1, T_2:

- Every $t \in T_1$ satisfies $\|t\|_2 \gtrsim \sqrt{n}$ and has regular coordinate structure in the sense of (2.8). As a consequence, a random coordinate projection P_Ω will not shrink the ℓ_2-norm of elements in T_1 by more than a factor of $\sim \sqrt{\delta}$, and therefore, with high probability,

$$\inf_{x \in H} \|P_\Omega \Gamma_x \xi\|_2 \geq c\sqrt{\delta n}.$$

- The set of "random oscillations" T_2 has Euclidean diameter smaller than $(c/2)\sqrt{\delta n}$. Thus, its effect is negligible.

What may still appear mysterious is the claim that there is a phase transition in the choice of δ – and thus in the required number of measurements. The origin of the phase transition lies in a gap between the cardinality of the net H and the probability estimate one is likely to have for each Γ_v. Indeed, for reasons that will be clarified later, the probability that $\Gamma_v \xi$ is "well behaved" can be estimated by $\exp(-cn/r)$. In contrast, as a nontrivial Euclidean net in V_r, $|H| \geq \exp(c_1 r \log(en/r))$. In the low-sparsity case, when $n/r \gtrsim r \log(en/r)$, the individual probability estimate is strong enough to allow uniform control on all the vectors in the net H. In the high-sparsity case that is no longer true, and an additional argument is required to bridge the gap between n/r and $r \log(en/r)$. Specifically, we will show how one may "transfer information" from a set of cardinality $\exp(cn/r)$ to the net H whose cardinality is much larger – of the order of $\exp(cr \log(en/r))$.

2.5 Notation

Throughout this article, absolute constants are denoted by c, c_1, C, etc. The notation $c(L)$ refers to a constant that depends only on the parameter L; $a \sim b$ implies that there are absolute constants c and C for which $ca \leq b \leq Ca$; and $a \sim_L b$ means that the constants c and C depend only on L. The analogous one-sided notation is $a \lesssim b$ and $a \lesssim_L b$. Constants whose values remain unchanged throughout the article are denoted by κ_1, κ_2, etc. .

For $1 \le p \le \infty$ let ℓ_p^n be the normed space $(\mathbb{R}^n, \| \ \|_p)$ and set B_p^n to be its unit ball. S^{n-1} is the Euclidean unit sphere in $\mathbb{R}^n$. The expectation is denoted by $\mathbb{E}$ and Pr denotes the probability of an event. The L_p-norm of a random variable X is denoted $\|X\|_{L_p} = (\mathbb{E}|X|^p)^{1/p}$. We also recall that $[n] = \{1, \ldots, n\}$.

3 Small ball estimates and chaining

3.1 The random generator

Recall that the random vector ξ we are interested in has independent coordinates $(\xi_i)_{i=1}^n$ that are mean-zero, variance 1 and L-subgaussian. In particular, ξ is an isotropic, L-subgaussian random vector.

A simple observation is that the ξ_i's satisfy a small-ball property: there are positive constants c_1 and c_2 that depend only on L for which

$$\sup_{u \in \mathbb{R}} Pr(|\xi_i - u| \ge c_1) \ge c_2. \tag{3.1}$$

Indeed, for any $u \in \mathbb{R}$, $\|\xi_i - u\|_{L_2} \sim \max\{\|\xi_i\|_{L_2}, |u|\}$, and thus $\|\xi_i - u\|_{L_4} \le c_3 L \|\xi_i - u\|_{L_2}$ for a suitable absolute constant c_3. The small-ball property (3.1) is an immediate outcome of the Paley-Zygmund inequality (see, e.g., [17, Lemma 7.16]) applied to each $X_u = |\xi_i - u|$.

The small-ball property (3.1) tensorizes, leading to a vector small-ball property for $\xi = (\xi_i)_{i=1}^n$. To formulate this property, let $\|\Gamma\|_{HS}$ and $\|\Gamma\|_{2\to 2}$ denote the Hilbert-Schmidt (Frobenius) and operator norms of a matrix Γ, respectively, and set

$$d_\Gamma = \left(\frac{\|\Gamma\|_{HS}}{\|\Gamma\|_{2\to 2}} \right)^2.$$

Theorem 3.1 *[40] There exists an absolute constant c for which the following holds. Let $X_1, \ldots, X_n$ be independent random variables that satisfy for some $t > 0$ and $0 < p < 1$*

$$\sup_{u \in \mathbb{R}} Pr(|X_i - u| \le t) \le p.$$

Then, for $\mathbb{X} = (X_1, \ldots, X_n)$ and every matrix $\Gamma : \mathbb{R}^n \to \mathbb{R}^m$,

$$Pr(\|\Gamma \mathbb{X}\|_2 \le t \|\Gamma\|_{HS}) \le \left(\frac{1}{2} \right)^{c d_\Gamma}.$$

The small-ball property for individual ξ_i's from (3.1) and Theorem 3.1 imply that the random vector ξ satisfies a small-ball estimate.

Corollary 3.2 *There exist constants κ_1 and κ_2 that depend only on L such that, for any matrix $\Gamma : \mathbb{R}^n \to \mathbb{R}^m$,*

$$Pr(\|\Gamma \xi\|_2 \le \kappa_1 \|\Gamma\|_{HS}) \le \left(\frac{1}{2} \right)^{\kappa_2 d_\Gamma}. \tag{3.2}$$

It should be noted that a subgaussian vector with independent coordinates is not the only random vector that satisfies a small-ball estimate like (3.2). Moreover, such small-ball estimates can be used to extend our main result to a larger class of generators – a direction we will not explore further in this work.

The other type of bound we require deals with the way the moments of $\|\xi\|$ grow, for an arbitrary norm $\| \ \|$ on $\mathbb{R}^n$. Unlike Corollary 3.2, this feature does not require ξ to have independent coordinates, and it holds for any (isotropic) subgaussian random vector (see, for example, [24, Theorem 2.3]).

Theorem 3.3 *There exists an absolute constant c for which the following holds. Let ξ be an isotropic, L-subgaussian random vector in $\mathbb{R}^n$ and set $G = (g_1, \ldots, g_n)$ to be the standard gaussian vector in $\mathbb{R}^n$. Let $\| \ \|$ be a norm on $\mathbb{R}^n$ and set B° to be the unit ball of its dual norm. Then for every $p \geq 1$,*

$$(\mathbb{E}\|\xi\|^p)^{1/p} \leq cL\big(\mathbb{E}\|G\| + \sqrt{p} \sup_{t \in B^\circ} \|t\|_2\big).$$

We will consider two families of norms associated with the non-increasing rearrangement of the coordinates of a vector.

Definition 3.4 *Let $(x_i^*)_{i=1}^n$ denote the non-increasing rearrangement of $(|x_i|)_{i=1}^n$. For $k \in [n]$, set*

$$\|x\|_{[k]} = \max_{|I|=k} \big(\sum_{i \in I} x_i^2\big)^{1/2} = \big(\sum_{i=1}^k (x_i^*)^2\big)^{1/2}.$$

When $k = n$, the norm $\| \ \|_{[k]}$ is simply the Euclidean norm, and the unit ball of the dual norm is just the standard Euclidean unit ball. When $1 \leq k < n$, the dual unit ball consists of the set of unit-ℓ_2-norm k-sparse vectors, that is:

$$V_k = \{v \in S^{n-1} : \|v\|_0 \leq k\}.$$

To apply Theorem 3.3 to $\| \ \|_{[k]}$ one has to control $\mathbb{E}\|G\|_{[k]} = \mathbb{E}\big(\sum_{i=1}^k (g_i^*)^2\big)^{1/2}$. The following result for subgaussian random variables does not require independence.

Lemma 3.5 *[23] Let $Z_1, \ldots, Z_n$ be mean-zero and L-subgaussian random variables such that $\max_{i \in [n]} \|Z_i\|_{L_2} \leq M$. Then*

$$\mathbb{E}\big(\sum_{i=1}^k (Z_i^*)^2\big)^{1/2} \leq cLM\sqrt{k \log(en/k)}.$$

Proof. Since the proof of this result is not provided in [23], we give it here for convenience. Since the Z_i are L-subgaussian, there exist constants $c_0, c_1 > 0$ such that (see e.g. [17, Proposition 7.23])

$$\mathbb{E} \exp\left(c_0 \frac{Z_i^2}{L^2 \|Z_i\|_{L_2}^2}\right) \leq c_1.$$

By Jensen's inequality and concavity of the logarithm

$$\left[\mathbb{E}\big(\frac{1}{k}\sum_{i=1}^{k}(Z_i^*)^2\big)^{1/2}\right]^2 \leq \mathbb{E}\frac{1}{k}\sum_{i=1}^{k}(Z_i^*)^2 \leq c_0^{-1}M^2L^2\mathbb{E}\frac{1}{k}\sum_{i=1}^{k}\log(\exp(c_0(Z_i^*)^2/(L\|Z_i^*\|_{L_2})^2))$$
$$\leq c_0^{-1}M^2L^2\log\left(\frac{1}{k}\sum_{i=1}^{k}\mathbb{E}\exp\left(c_0\frac{(Z_i^*)^2}{L^2\|Z_i^*\|_{L_2}^2}\right)\right) \leq c_0^{-1}M^2L^2\log\left(\frac{1}{k}\sum_{i=1}^{n}\mathbb{E}\exp\left(c_0\frac{Z_i^2}{L^2\|Z_i\|_{L_2}^2}\right)\right)$$
$$\leq c_0^{-1}M^2L^2\log(c_1 n/k).$$

Rearranging this inequality and adjusting constants yields the claim. ∎

For the choice $Z_i = \xi_i$, it follows from Lemma 3.5 and Theorem 3.3 that for every $p \geq 1$ and $k \in [n]$,

$$(\mathbb{E}\|\xi\|_{[k]}^p)^{1/p} \leq cL\left(\sqrt{k\log(en/k)}+\sqrt{p}\right). \tag{3.3}$$

The second family of norms we require is a generalization of the first one. Let Γ be a matrix and set

$$\|x\| = \|\Gamma x\|_{[k]} = \big(\sum_{i=1}^{k}(\langle\Gamma x, e_i\rangle^*)^2\big)^{1/2} = \sup_{t\in\Gamma^* V_k}\langle x,t\rangle,$$

where the last equality follows from $\|z\|_{[k]} = \sup_{t\in V_k}\langle z,t\rangle$.

Lemma 3.6 *Let ξ be an isotropic, L-sugaussian random vector. Then for every matrix Γ and any $p \geq 1$,*

$$(\mathbb{E}\|\Gamma\xi\|_{[k]}^p)^{1/p} \leq cL\left(\sqrt{k\log(en/k)}\max_{1\leq i\leq n}\|\Gamma^* e_i\|_2 + \sqrt{p}\sup_{t\in V_k}\|\Gamma^* t\|_2\right)$$

Proof. By Theorem 3.3, it suffices to estimate $\mathbb{E}\sup_{t\in\Gamma^*V_k}\langle x,t\rangle = \mathbb{E}\big(\sum_{i=1}^{k}(\langle G,\Gamma^*e_i\rangle^*)^2\big)^{1/2}$ for the standard gaussian vector G. This expectation may be controlled using Lemma 3.5 for the choice of $Z_i = \langle G,\Gamma^* e_i\rangle$ and that fact that $M = \max_{i\in[n]}\|Z_i\|_{L_2} = \max_{i\in[n]}\|\Gamma^* e_i\|_2$. ∎

The following standard relation between moments and tails is recorded for convenience as it will be used frequently in the sequel. Its proof is follows immediately from Markov's inequality.

Lemma 3.7 *Assume that a random variable Z satisfies $(\mathbb{E}|Z|^p)^{1/p} \leq A$ for some $p > 0$ and $A > 0$. Then*

$$Pr(|Z| \geq eA) \leq e^{-p} \quad \text{and } Pr(|Z| \geq 2A) \leq 2^{-p}.$$

As noted above, the matrices we will be interested in are of the form

$$\Gamma_v = \sqrt{n}UD_{Wv}O,$$

where U, W and O are Hadamard type matrices with constant β and $v \in \mathbb{R}^n$. Thus,

$$\|\Gamma_v\|_{HS} = \sqrt{n}\|v\|_2, \quad \|\Gamma_v\|_{2\to 2} \leq \sqrt{n}\|Wv\|_\infty \leq \beta\|v\|_0^{1/2},$$

and for every $i \in [n]$,

$$\|\Gamma_v^* e_i\|_2 = \sqrt{n}\big(\sum_{\ell=1}^{n} \langle W_\ell, v\rangle^2 \cdot U_{i,\ell}^2\big)^{1/2} \leq \beta\|v\|_2.$$

Combining Corollary 3.2, Lemma 3.6 and Lemma 3.7, one has the following:

Corollary 3.8 *There exist constants κ_1, κ_2 and κ_3 that depend only on L and β for which the following holds. If $v \in V_r$ then*

$$Pr(\|\Gamma_v \xi\|_2 \leq \kappa_1 \sqrt{n}\|v\|_2) \leq \left(\frac{1}{2}\right)^{\kappa_2 n/\beta^2 r}, \tag{3.4}$$

and if $v \in \mathbb{R}^n$ and then with probability at least $1 - \exp(-p)$,

$$\|\Gamma_v \xi\|_{[k]} \leq \kappa_3(\|v\|_2\sqrt{k\log(en/k)} + \sqrt{p} \cdot \sqrt{n}\|Wv\|_\infty). \tag{3.5}$$

Remark 3.9 *In what follows,* (3.4) *and* (3.5) *are the key features of ξ that we will use. To establish those two facts we used rather special properties of ξ, but while those special properties (for example, that ξ is stochastically dominated by a gaussian vector) are highly restrictive,* (3.4) *and* (3.5)*, or even further relaxations of the two, actually hold for a wider variety of random vectors. We will pursue this direction in a future contribution.*

3.2 Definitions and basic facts

Let (T, d) be a metric space. A subset $T' \subset T$ is called ε-separated if for every $x, y \in T'$, $d(x, y) \geq \varepsilon$. By a standard comparison of packing and covering numbers, see e.g. [17, Lemma C.2], if T' is a maximal ε-separated subset of T, it is also an ε-cover: that is, for every $x \in T$ there is some $y \in T'$ for which $d(x, y) \leq \varepsilon$.

In what follows we denote by $N(T, d, \varepsilon)$ the cardinality of a minimal ε-cover of T. Note that if $T \subset \mathbb{R}^n$ and d is a norm on $\mathbb{R}^n$ whose unit ball is B, then $N(T, d, \varepsilon)$ is the minimal number of translates of εB needed to cover T. Therefore, we will sometimes abuse notation and write $N(T, \varepsilon B)$ instead of $N(T, d, \varepsilon)$.

We will also use the language of *Generic Chaining* [41] extensively.

Definition 3.10 *Given a metric space (T, d), an admissible sequence $(T_s)_{s\geq 0}$ is a sequence of subsets of T, with $|T_0| = 1$ and $|T_s| \leq 2^{2^s}$. Together with an admissible sequence one defines a collection of maps $\pi_s : T \to T_s$. Usually, $\pi_s t$ is chosen as a nearest point to t in T_s with respect to the metric d. For $s \geq 0$ set*

$$\Delta_s t = \pi_{s+1} t - \pi_s t.$$

Let us define several parameters that will be used throughout the proof of Theorem 2.5.

Definition 3.11 *For $r \in [n]$ set*

$$\rho = 10\log_2 e \cdot \max\left\{1, \frac{\log(er)}{\log(en/r)}\right\}. \tag{3.6}$$

Using the notation introduced earlier, put

$$\kappa_4 = \min\left\{\frac{\kappa_2}{2\beta^2}, \frac{\kappa_1^2}{64\kappa_3^2 L^2\beta^2}\right\}, \tag{3.7}$$

and observe that κ_4 depends only on L and β. Moreover, without loss of generality, $\kappa_1 \leq 1$, and $\kappa_3, L, \beta \geq 1$.

Set s_0 and s_1 to satisfy

$$2^{s_0} = \frac{\kappa_4 n}{r} \quad \text{and} \quad 2^{s_1} = \rho r \log(en/r)$$

and without loss of generality we will assume that s_0 and s_1 are integers. Finally, let

$$\alpha_r = \max\left\{1, \log\left(\frac{\rho r \log(en/r)}{\kappa_4(n/r)}\right)\right\} = \max\{1, \log(2^{s_1-s_0})\}.$$

A key part in the proof of Theorem 2.5 requires a different argument when $2^{s_0} \geq 2^{s_1}$ and when the reverse inequality holds. As we indicated earlier, we will call the former the "low-sparsity" case, and the latter the "high-sparsity" case. It is straightforward to verify that in the low-sparsity case ($2^{s_0} \geq 2^{s_1}$), this corresponds to

$$r \leq c\kappa_4^{1/2}\sqrt{\frac{n}{\log(cn/\kappa_4)}}, \quad \rho = 10\log_2 e, \quad \text{and} \quad \alpha_r = 1,$$

while in the "high-sparsity" case,

$$r \geq c\kappa_4^{1/2}\sqrt{\frac{n}{\log(cn/\kappa_4)}};$$

if $r \leq \sqrt{n}$ then $\rho = 10\log_2 e$, and otherwise, $\rho = 10\log_2 e \cdot \frac{\log(er)}{\log(en/r)}$. Thus,

$$\alpha_r = \log\left(\frac{cr^2}{\kappa_4 n} \cdot \log\left(\frac{en}{r}\right)\right) \quad \text{if} \quad r \leq \sqrt{n},$$

and

$$\alpha_r = \log\left(\frac{cr^2}{\kappa_4 n} \cdot \log(er)\right) \quad \text{otherwise.}$$

Let us mention that κ_1, κ_2, κ_3 and κ_4 are all constants that depend only on L and β – an observation that will be used throughout this article.

3.3 Covering of V_r

Let us begin by constructing (a part of) an admissible sequence for V_r.

Lemma 3.12 *Let $1 \leq r \leq n/2$ and s_1 as above. There exists an admissible sequence $(V_{r,s})_{s \geq s_1}$ for which*

$$\sup_{v \in V_r} \sum_{s \geq s_1} (\sqrt{n} + \sqrt{r}2^{s/2})\|\Delta_s v\|_2 \leq \frac{c}{n^{3/2}},$$

where $\pi_s v$ is the nearest point to v in $V_{r,s}$ with respect to the Euclidean norm, $\Delta_s v = \pi_{s+1}v - \pi_s v$ and c is an absolute constant.

We note that the exponent $3/2$ above is rather arbitrary. We could easily replace it by a larger one by adjusting constants.

Proof. Let $V_{r,s}$ be a maximal ε_s separated subset of V_r with respect to the Euclidean norm and of cardinality 2^{2^s}. Thus it is also an ε_s-cover of V_r and

$$\|\Delta_s v\|_2 \le \|\pi_{s+1} v - v\|_2 + \|\pi_s v - v\|_2 \le 2\varepsilon_s.$$

To estimate ε_s, observe that by a standard volumetric estimate, see e.g. [17, Proposition C.3], and summing over all $\binom{n}{r}$ possible support subsets of $[n]$ of cardinality r, for any $0 < \varepsilon < 1/2$, the cardinality of a maximal ε-separated subset of V_r is at most

$$\binom{n}{r}\left(1+\frac{2}{\varepsilon}\right)^n \le \binom{n}{r}\left(\frac{3}{\varepsilon}\right)^r \le \left(\frac{3en}{r\varepsilon}\right)^r.$$

Hence,

$$\varepsilon_s \le 2^{-2^s/r}\left(\frac{3en}{r}\right),$$

and

$$\sup_{v\in V_r}\sum_{s\ge s_1} 2^{s/2}\|\Delta_s v\|_2 \le 2\sum_{s\ge s_1} 2^{s/2}\varepsilon_s \le c_0\left(\frac{n}{r}\right)\cdot\sum_{s\ge s_1} 2^{s/2-2^s/r}.$$

It is straightforward to verify that for every $s \ge s_1$,

$$2^s/r \ge 2(s/2). \tag{3.8}$$

This follows for s_1 because $2^{s_1} = \rho r\log(en/r)$ and

$$\frac{2^{s_1}}{r} = (10\log_2 e)\cdot\max\left\{\log\left(\frac{en}{r}\right), \log(er)\right\} \ge s_1,$$

and for $s > s_1$ because $s \mapsto 2^s/s$ is increasing. Therefore,

$$\sup_{v\in V_r}\sum_{s\ge s_1} 2^{s/2}\|\Delta_s v\|_2 \le c_0\frac{n}{r}\cdot\sum_{s\ge s_1} 2^{-2^s/(2r)},$$

which is dominated by a geometric series with power $2^{-2^{s_1}/2r} = 2^{-(\rho/2)\log(en/r)} \le 1/4$. Therefore,

$$\frac{n}{r}\sum_{s\ge s_1} 2^{-2^s/2r} \lesssim \frac{n}{r}\cdot 2^{-2^{s_1}/(2r)} \le e^{-1}\left(\frac{r}{en}\right)^{\frac{\rho}{2\log_2 e}-1}.$$

Note that

$$\left(\frac{r}{en}\right)^{\frac{\rho}{2\log_2 e}-1} \le \frac{1}{(en)^2}. \tag{3.9}$$

Indeed, if $r \le n/r$, i.e., if $r \le \sqrt{n}$, then $\rho/\log_2 e \ge 10$ and

$$\left(\frac{r}{en}\right)^{(\rho/2\log_2 e)-1} \le \left(\frac{1}{\sqrt{en}}\right)^4 \le \frac{1}{(en)^2};$$

otherwise, $\sqrt{n} \le r \le n/2$ and $\rho/\log_2 e = 10\log(er)/\log(en/r)$ so that

$$\log\left(\left(\frac{en}{r}\right)^{(\rho/2\log_2 e)-1}\right) = \log\left(\frac{en}{r}\right)\left[\frac{5\log(er)}{\log(en/r)} - 1\right] \ge \log\left((en)^4\right),$$

so that (3.9) holds also in this case. Therefore,

$$\sup_{v\in V_r}\sum_{s\ge s_1}(\sqrt{n}+\sqrt{r}2^{s/2})\|\Delta_s v\|_2 \le \frac{c_1}{n^{3/2}}$$

for an absolute constant c_1. ∎

4 The structure of a typical $\{\Gamma_v\xi : v \in V_r\}$

The main component in the proof of Theorem 2.5 is a structural result on a typical realization of the random set $\{\Gamma_v\xi : v \in V_r\}$.

Theorem 4.1 *There exist constants $c_0, \dots, c_3$ that depend only on L and β for which the following holds. With probability at least $1 - 2^{-c_0\min\{2^{s_0},2^{s_1}\}}$,*

$$\{\Gamma_v\xi : v \in V_r\} \subset T_1 + T_2,$$

where

$$|T_1| \le 2^{2^{s_1}} \quad \text{and} \quad T_2 \subset c_1 n^{-3/2} B_2^n.$$

Moreover, for every $t \in T_1$, there is a (random) set $I = I(t) \subset [n]$ of cardinality at least

$$|I| \ge c_2\frac{n}{\alpha_r^2\log\alpha_r},$$

for which

$$\sum_{i\in I}\left\langle\Gamma_t\xi, e_i\right\rangle^2 \ge (\kappa_1/4)^2 n \quad \text{and} \quad \max_{i\in I}|\left\langle\Gamma_t\xi, e_i\right\rangle| \le c_3\alpha_r\sqrt{\log(e\alpha_r)}.$$

Theorem 4.1 implies that a typical realization of $\{\Gamma_v\xi : v \in V_r\}$ is just a perturbation of the (random) set T_1, and that T_1 consists of vectors with a regular coordinate structure.

Let us examine Theorem 4.1 in the two cases: when $s_0 \ge s_1$ and when $s_0 < s_1$. In the former (the low-sparsity case), $\alpha_r = 1$, and the claim is that with probability at least $1-2^{-c_0\rho r\log(en/r)}$, for every $t \in T_1$, there is $I \subset [n]$ of cardinality $|I| \ge c_2 n$ for which

$$\|P_I t\|_2 \ge (\kappa_1/4)\sqrt{n} \quad \text{and} \quad \|P_I t\|_\infty \le c_3.$$

This forces I to contain at least $\sim c_2 n$ coordinates that are larger than a constant, and thus, each one of the vectors $t \in T_1$ has a regular coordinate structure in the sense of (2.8).

When $s_1 > s_0$, a similar type of claim holds, but with probability at least $1 - 2^{-c_0\kappa_4 n/r}$, and the regularity condition on the coordinates of $t \in T_1$ is slightly weaker: one no longer has a subset of cardinality that is proportional to n consisting of coordinates that are larger than a constant, but rather a (marginally) smaller set I; each $P_I t$ has a large ℓ_2 norm and a small ℓ_∞ norm.

Thanks to this information on the structure of a typical $\{\Gamma_v\xi : v \in V_r\}$, one may establish the required lower bound on $\inf_{v\in V_r}\|P_\Omega\Gamma_v\xi\|_2$.

Corollary 4.2 *There exist constants c_0, c_1 and c_2 that depend only on L and β for which the following holds. Let*

$$\delta n \geq c_0(\alpha_r^2 \log \alpha_r) \cdot \rho r \log(en/r),$$

with ρ defined in (3.6) and set $(\delta_i)_{i=1}$ to be independent selectors with mean δ. Then with probability at least $1 - 2^{-c_1 \min\{2^{s_0}, 2^{s_1}\}}$,

$$\inf_{v \in V_r} \sum_{i=1}^{n} \delta_i \left\langle \Gamma_v \xi, e_i \right\rangle^2 \geq c_2 \delta n.$$

Proof. Let ξ be a realization of the event from Theorem 4.1 – which holds with probability at least $1 - 2^{-c_0 \min\{2^{s_0}, 2^{s_1}\}}$ relative to ξ, and let $(\delta_i)_{i=1}^n$ be independent selectors with mean δ that are also independent of ξ.

Using the notation of Theorem 4.1, let $\Gamma_v \xi = t + y$ for $t \in T_1$ and $y \in T_2$; hence

$$\left(\sum_{i=1}^{n} \delta_i \left\langle \Gamma_v \xi, e_i \right\rangle^2\right)^{1/2} \geq \inf_{t \in T_1} \left(\sum_{i=1}^{n} \delta_i t_i^2\right)^{1/2} - \sup_{y \in T_2} \left(\sum_{i=1}^{n} \delta_i y_i^2\right)^{1/2} \geq \inf_{t \in T_1} \left(\sum_{i=1}^{n} \delta_i t_i^2\right)^{1/2} - \frac{c_1}{n^{3/2}}$$

for an absolute constant c_1. It suffices to show that

$$\inf_{t \in T_1} \sum_{i=1}^{n} \delta_i t_i^2 \geq c_2 \delta n,$$

the right hand being larger than $2c_1/n^3$, where c_2 is a suitable constant. Fix $t \in T_1$, let $I = I(t)$ be the set identified by Theorem 4.1 and put $x = P_I t$. Thus,

$$\|x\|_2 \geq (\kappa_1/4)\sqrt{n}, \quad \text{and} \quad \|x\|_\infty \leq c_3(L, \beta)\alpha_r \sqrt{\log(e\alpha_r)}. \tag{4.1}$$

By a straightforward application of Bernstein's inequality,

$$Pr\Big(\big|\sum_{i \in I} (\delta_i - \delta) x_i^2\big| \geq w\Big) \leq 2 \exp\Big(-c_4 \min\Big\{\frac{w^2}{\delta \sum_{i \in I} x_i^4}, \frac{w}{\max_{i \in I} x_i^2}\Big\}\Big).$$

Observe that $\sum_{i \in I} x_i^4 \leq \|x\|_\infty^2 \sum_{i \in I} x_i^2$, and thus, for $w = (\delta/2)\sum_{i \in I} x_i^2$, the probability estimate becomes

$$2 \exp(-c_4 \delta \|x\|_2^2 / \|x\|_\infty^2) \leq 2 \exp\Big(-c_5(L, \beta) \frac{\delta n}{\alpha_r^2 \log(e\alpha_r)}\Big).$$

By a union bound, with probability at least

$$1 - 2|T_1| 2 \exp\Big(-c_5 \frac{\delta n}{\alpha_r^2 \log(e\alpha_r)}\Big),$$

relative to $(\delta_i)_{i=1}^n$, this implies that, for every $t \in T_1$,

$$\sum_{i=1}^{n} \delta_i t_i^2 \geq \sum_{i \in I(t)} \delta_i t_i^2 \geq \frac{\delta}{2} \sum_{i \in I(t)} x_i^2 \geq \frac{\kappa_1^2}{32} \delta n.$$

The claim follows by setting

$$\delta n \geq c_5^{-1}(\alpha_r^2 \log(e\alpha_r)) \cdot 2^{s_1+1} = c_7(L, \beta) \cdot \alpha_r^2 \log(e\alpha_r) \cdot \rho r \log(en/r).$$

■

The proof of Theorem 4.1 is based on the following idea: $(V_{r,s})_{s\geq s_1}$ will be selected as an maximal ε_s-separated subset of V_r and $T_1 = \{\Gamma_v\xi : v \in V_{r,s_1}\}$. We will show that for every $v \in V_r$, $v - \pi_{s_1}v = \sum_{s\geq s_1} \Delta_s v$ is small enough to ensure that with high probability,

$$\sup_{v\in V_r} \|\Gamma_{v-\pi_{s_1}v}\xi\|_2 \leq \sum_{s\geq s_1} \|\Delta_s v\|_2 \leq \frac{c'}{n^{3/2}}$$

for c' that depends on L and β. Then, we will turn to the more difficult part of the argument – that with high probability, for every $v \in V_{r,s_1}$,

$$\|\Gamma_v\xi\|_2 \geq c\sqrt{n} \ \text{ and } \ \|\Gamma_v\xi\|_{[\theta n]} \leq (c/2)\sqrt{n} \tag{4.2}$$

for well-chosen c and

$$\theta = \begin{cases} c_1 & \text{for } r \leq c_2\sqrt{\frac{\kappa_4 n}{\log(c_2 n/\kappa_4)}} \\ \frac{c_3}{\alpha_r^2\log(\alpha_r)} & \text{for } c_2\sqrt{\frac{\kappa_4 n}{\log(c_2 n/\kappa_4)}} \leq r \leq \frac{c_4 n}{\log^4(n)} \end{cases} \tag{4.3}$$

where all constants c, c_1, c_2, c_3, c_4 only depend on L and β and κ_4 is defined in (3.7), see (4.8) and the following remarks as well as Lemma 4.10. Moreover, here and in the following, we assume for simplicity that θn is an integer. (The general case may need slightly different constants.) Then, for $t = \Gamma_v\xi$, one has

$$\sum_{i=\theta n+1}^{n} (t_i^*)^2 \geq (c/4)n \ \text{ and } \ t_{\theta n}^* \leq \Big(\frac{1}{\theta n}\sum_{i=1}^{\theta n}(t_i^*)^2\Big)^{1/2} \leq \frac{c}{2\sqrt{\theta}}.$$

The set $I = I(t)$ of indices corresponding to $t_{\theta n+1}^*, \ldots, t_n^*$ fulfills the properties claimed in Theorem 4.1, i.e., T_1 consists of vectors with a regular coordinate structure.

We begin the proof with its simpler part: a high probability estimate on $\sup_{v\in V_r} \|\Gamma_{v-\pi_{s_1}v}\xi\|_2$.

Lemma 4.3 *There exist constants c and c_1 that depend only on L and β for which the following holds. If $(V_{r,s})_{s\geq s_1}$ is an admissible sequence of V_r, then with probability at least $1 - 2^{-2^{s_1}}$, for every $v \in V_r$ and $s \geq s_1$,*

$$\|\Gamma_{\Delta_s v}\xi\|_2 \leq c\left(\sqrt{n} + 2^{s/2}\sqrt{r}\right)\|\Delta_s v\|_2. \tag{4.4}$$

In particular, if $(V_{r,s})_{s\geq s_1}$ is a maximal separated subset of V_r, then with probability at least $1 - 2^{-2^{s_1}}$,

$$\sup_{v\in V_r} \|\Gamma_{v-\pi_{s_1}v}\xi\|_2 \leq \frac{c_1}{n^{3/2}}.$$

Proof. The first part of the proof is a straightforward outcome of (3.5). Indeed, if we set $k = n$ and $p = 2^{s+3}$, then by Lemma 3.7, with probability at least $1 - \exp(-2^{-(s+3)})$,

$$\|\Gamma_{\Delta_s v}\xi\|_2 \leq 4\kappa_3(\sqrt{n}\|\Delta_s v\|_2 + 2^{s/2} \cdot \sqrt{n}\|W\Delta_s v\|_\infty).$$

Since v is r-sparse, $\Delta_s v = \pi_{s+1}v - \pi_s v$ is $2r$-sparse, and since W is a Hadamard type matrix with constant β,

$$\sqrt{n}\|W\Delta_s v\|_\infty = \sqrt{n}\max_{i\in[n]} |\langle W_i, \Delta_s v\rangle| \le \beta\|\Delta_s v\|_1 \le \beta\|\Delta_s v\|_0^{1/2} \cdot \|\Delta_s v\|_2.$$

Therefore,

$$\|\Gamma_{\Delta_s v}\xi\|_2 \le 8\kappa_3(\sqrt{n} + \beta\sqrt{r}2^{s/2})\|\Delta_s v\|_2. \tag{4.5}$$

There are at most $2^{2^s} \cdot 2^{2^{s+1}} \le 2^{2^{s+2}}$ vectors of the form $\Delta_s v$, and thus, by a union bound, with probability at least $1 - 2^{-2^{s+2}}$, (4.5) holds for every $v \in V_r$ for that choice of s. Summing the probabilities for $s \ge s_1$, one has that with probability at least $1 - 2^{-2^{s_1+1}}$, for every $v \in V_r$,

$$\|\Gamma_{v-\pi_{s_1}v}\xi\|_2 = \|\sum_{s\ge s_1} \Gamma_{\Delta_s v}\xi\|_2 \le \sum_{s\ge s_1} \|\Gamma_{\Delta_s v}\xi\|_2 \le 8\kappa_3\beta \sum_{s\ge s_1} \left(\sqrt{n} + \sqrt{r}2^{s/2}\right) \|\Delta_s v\|_2 \le \frac{c(L,\beta)}{n^{3/2}},$$

where the last inequality is just Lemma 3.12. ∎

Lemma 4.3 shows that if we set $T_2 = \{\Gamma_{v-\pi_{s_1}v}\xi : v \in V_r\}$, then with probability at least $1 - 2^{-2^{s_1}}$, $\sup_{t\in T_2} \|t\|_2 = \sup_{v\in V_r} \|\Gamma_{v-\pi_{s_1}v}\xi\|_2 \le c(L,\beta)/n^{3/2}$, as required in Theorem 4.1. Therefore, all that is left is to study the structure of $T_1 = \{\Gamma_{\pi_{s_1}v}\xi : v \in V_r\}$, and to show that with high probability, it consists of vectors with a regular coordinate structure. To that end we shall split the argument into two cases: the low-sparsity case, when $s_0 \ge s_1$ and the high-sparsity one, in which the reverse inequality holds. The analysis in both cases is based on Corollary 3.8.

The low-sparsity case

Assume that $2^{s_0} \ge 2^{s_1}$. Then $\frac{\kappa_4 n}{r} \ge \rho r\log(en/r)$ and in particular, by the choice of κ_4,

$$\frac{\kappa_2}{2\beta^2} \cdot \frac{n}{r} \ge \rho r \log\left(\frac{en}{r}\right). \tag{4.6}$$

Fix $v \in V_{r,s_1}$. It follows from (3.4) that with probability at least $1 - 2^{-\kappa_2 n/\beta^2 r}$,

$$\|\Gamma_v\xi\|_2 \ge \kappa_1\sqrt{n}. \tag{4.7}$$

Let $0 < \theta < 1$ to be named later. Observe that by (3.5) for $k = \theta n$ and $p = 2\rho r\log(en/r) = 2^{s_1+1}$, with probability at least $1 - 2^{-2^{s_1+1}}$,

$$\begin{aligned}\|\Gamma_v\xi\|_{[k]} \le& 2\kappa_3(\|v\|_2\sqrt{k\log(en/k)} + \sqrt{\rho r\log(en/r)}\sqrt{n}\|Wv\|_\infty) \\ \le& 2\kappa_3(\sqrt{n}\sqrt{\theta\log(e/\theta)} + \beta\sqrt{\rho r\log(en/r)} \cdot \sqrt{r}),\end{aligned} \tag{4.8}$$

because $\|v\|_2 = 1$ and $\sqrt{n}\|Wv\|_\infty \le \beta\sqrt{r}$.

Recall that $|V_{r,s_1}| \le 2^{2^{s_1}}$ and thus, by (4.6) and the union bound, with probability at least $1 - 2^{-2^{s_1}}$, (4.7) and (4.8) hold for every $v \in V_{r,s_1}$. All that remains is ensure that

$$2\kappa_3\beta\sqrt{\theta\log(e/\theta)}\sqrt{n} \le \frac{\kappa_1}{4}\sqrt{n} \quad \text{and} \quad 2\kappa_3\beta\sqrt{r} \cdot \sqrt{\rho r\log(en/r)} \le \frac{\kappa_1}{4}\sqrt{n}.$$

The first condition holds for the right choice of the constant $\theta = \theta(L, \beta)$. For the second, note that by the definition of κ_4 in (3.7)

$$\rho r \log(en/r) = 2^{s_1} \leq 2^{s_0} \leq \frac{\kappa_1^2}{64\kappa_3^2 L^2 \beta^2} \cdot \frac{n}{r}.$$

Therefore,

$$2\kappa_3 \beta \sqrt{r} \cdot \sqrt{\rho r \log(en/r)} \leq 2\kappa_3 \beta \sqrt{r} \cdot \frac{\kappa_1}{8\kappa_3 L \beta} \cdot \sqrt{\frac{n}{r}} \leq \frac{\kappa_1}{4}\sqrt{n},$$

because $L \geq 1$.

This concludes the proof of Theorem 4.1 in the low sparsity case. ■

The high-sparsity case

Now consider the case $2^{s_0} \leq 2^{s_1}$; that is $\kappa_4 n/r \leq \rho r \log(en/r)$, and there is a substantial gap between the individual probability estimate (3.4) and the cardinality of V_{r,s_1}, so that a simple union bound does not give a non-trivial probability estimate. The difficulty one faces here is bridging this gap, and the key to the proof in the high-sparsity case is the following result.

Theorem 4.4 *There exist constants c_1 and c_2 that depend only on L and β for which the following holds. If $r \leq c_1 n/\log^4 n$ then with probability at least $1 - 2\exp(-c_2 n/r)$,*

$$\inf_{v \in V_{r,s_1}} \|\Gamma_v \xi\|_2 \geq \frac{\kappa_1}{2}.$$

For the proof of Theorem 4.4, one has to 'transfer' the lower bound on $\inf_{v \in V_{r,s_0}} \|\Gamma_v \xi\|_2$, which may be obtained directly from (3.4) and the union bound, to the much larger set V_{r,s_1}. Thus, it suffices to show that with high probability,

$$\sup_{v \in V_{r,s_1}} \left| \|\Gamma_v \xi\|_2^2 - \|\Gamma_{\pi v}\|_2^2 \right| \leq \frac{\kappa_1^2}{4} n, \tag{4.9}$$

for an approximating $\pi v \in V_{r,s_0}$.

To address (4.9), we proceed along the lines of [24] and consider the following, more general situation: let $\mathcal{A}$ be a class of matrices, $|\mathcal{A}| \leq 2^{2^{s_1}}$ and set $(\mathcal{A}_s)_{s \geq s_0}$ to be an admissible sequence of $\mathcal{A}$; that is, $\mathcal{A}_s$ is of cardinality at most 2^{2^s}. Let $\pi_s A$ to be the nearest point to A in $\mathcal{A}_s$ with respect to the $\| \ \|_{2\to2}$ norm, set $\Delta_s A = \pi_{s+1} A - \pi_s A$, and put

$$\gamma_{s_0,s_1}(\mathcal{A}) = \sup_{A \in \mathcal{A}} \sum_{s=s_0}^{s_1-1} 2^{s/2} \|\Delta_s A\|_{2\to2}.$$

Lemma 4.5 *Let ξ be an isotropic, L-subgaussian random vector and set ξ' to be an independent copy of ξ. Let $N_{\mathcal{A}}(\xi) = \sup_{A \in \mathcal{A}} \|A\xi\|_2$ and put $\mathcal{Z} = \sup_{A \in \mathcal{A}} \left| \langle A\xi, A\xi' \rangle - \langle (\pi_{s_0} A)\xi, (\pi_{s_0} A)\xi' \rangle \right|$. Then, for every $p \geq 1$,*

$$\|\mathcal{Z}\|_{L_p} \leq cL \sup_{A \in \mathcal{A}} \gamma_{s_0,s_1}(\mathcal{A}) \|N_{\mathcal{A}}(\xi)\|_{L_p}.$$

for an absolute constant c.

Since the proof of Lemma 4.5 is contained in the proof of Lemma 3.3 in [24], it will not be presented here.

We will be interested in the specific class of matrices $\mathcal{A} = \{\Gamma_v : v \in V_{r,s_1}\}$:

Theorem 4.6 *There exists constants c_1 and c_2 that depend only on L and β for which the following holds. Let $\mathcal{A} = \{\Gamma_v : v \in V_{r,s_1}\}$ and set $(\mathcal{A}_s)_{s \geq s_0}$ to be an admissible sequence of $\mathcal{A}$. Then, with probability at least $1 - 2^{-c_1 n/r}$,*

$$\sup_{A \in \mathcal{A}} \left| \|A\xi\|_2^2 - \|\pi_{s_0} A\xi\|_2^2 \right| \leq c_2 \gamma_{s_0,s_1}(\mathcal{A}) \left(\gamma_{s_0,s_1}(\mathcal{A}) + \sqrt{n} \right). \tag{4.10}$$

Before proving Theorem 4.6, let us recall a standard fact that can be established using tail integration.

Lemma 4.7 *Let Z be a nonnegative random variable and assume that*

$$Pr(Z \geq A_1 + uA_2) \leq 2\exp(-u^2/2) \quad \text{for every } u \geq A_3.$$

Then, for every $p \geq 1$, $\|Z\|_{L_p} \leq c(A_1 + A_2 \cdot \max\{A_3, \sqrt{p}\})$, and c is an absolute constant.

Proof of Theorem 4.6. Denote by A^j, $j \in [n]$, the columns of the matrix A and observe that

$$\|A\xi\|_2^2 - \|\pi_{s_0} A\xi\|_2^2 = \left\langle A\xi, A\xi \right\rangle^2 - \left\langle \pi_{s_0} A\xi, \pi_{s_0} A\xi \right\rangle^2 = \sum_{j,k} \xi_j \xi_k \left(\left\langle A^j, A^k \right\rangle - \left\langle (\pi_{s_0} A)^j, (\pi_{s_0} A)^k \right\rangle \right).$$

Since each A is of the form Γ_v for some $v \in S^{n-1}$, $\left\langle A^j, A^j \right\rangle = \|v\|_2^2 = 1$, and the same holds for $\pi_{s_0} A$. Therefore, $\left\langle A^j, A^j \right\rangle - \left\langle (\pi_{s_0} A)^j, (\pi_{s_0} A)^j \right\rangle = 0$ and all that remains is to control the "off-diagonal" terms,

$$\sum_{j \neq k} \xi_j \xi_k \left(\left\langle A^j, A^k \right\rangle - \left\langle (\pi_{s_0} A)^j, (\pi_{s_0} A)^k \right\rangle \right) =: F_A.$$

Applying a standard decoupling argument (see, for example, [24, Theorem 2.4] or [11]) and Lemma 4.5

$$\begin{aligned}
(\mathbb{E} \sup_{A \in \mathcal{A}} |F_A|^p)^{1/p} =& \| \sup_{A \in \mathcal{A}} (*)_A \|_{L_p} \leq c_1 \left\| \sup_{A \in \mathcal{A}} \left| \sum_{j,k} \xi_j \xi_k' \left(\left\langle A^j, A^k \right\rangle - \left\langle (\pi_{s_0} A)^j, (\pi_{s_0} A)^k \right\rangle \right) \right| \right\|_{L_p} \\
=& c_1 \left\| \sup_{A \in \mathcal{A}} \left| \left\langle A\xi, A\xi' \right\rangle - \left\langle \pi_{s_0} A\xi, \pi_{s_0} A\xi' \right\rangle \right| \right\|_{L_p} \\
\leq& c_2 L \Big(\sup_{A \in \mathcal{A}} \sum_{s=s_0}^{s_1-1} 2^{s/2} \|\Delta_s A\|_{2\to 2} \Big) \cdot \|N_{\mathcal{A}}(\xi)\|_{L_p} = c_2 L \gamma_{s_0,s_1}(\mathcal{A}) \cdot \|N_{\mathcal{A}}(\xi)\|_{L_p}.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\|(N_{\mathcal{A}}(\xi))^2\|_{L_p} \leq & \left\| \sup_{A \in \mathcal{A}} \left| \|A\xi\|_2^2 - \|\pi_{s_0} A\xi\|_2^2 \right| \right\|_{L_p} + \left\| \sup_{A \in \mathcal{A}} \|\pi_{s_0} A\xi\|_2^2 \right\|_{L_p} \\
\leq & c_2 L \gamma_{s_0,s_1}(\mathcal{A}) \cdot \|N_{\mathcal{A}}(\xi)\|_{L_p} + \left\| \sup_{A \in \mathcal{A}} \|\pi_{s_0} A\xi\|_2^2 \right\|_{L_p}.
\end{aligned}$$

By (3.5) for $k=n$ and $p=u^2 \geq 2^{s_0+3}$, it follows that with probability at least $1-2\exp(-u^2/2)$,

$$\sup_{A \in \mathcal{A}} \|\pi_{s_0} A\xi\|_2 = \sup_{v \in V_{r,s_0}} \|\Gamma_v \xi\|_2 \leq \kappa_3(\sqrt{n} + u\sqrt{n}\|Wv\|_\infty) \leq c_3(L,\beta)(\sqrt{n}+u\sqrt{r}),$$

where we used that $\|Wv\|_\infty \leq \beta\sqrt{r/n}$. Thus, the random variable $Z = \sup_{A \in \mathcal{A}} \|\pi_{s_0} A\xi\|_2$ satisfies the conditions of Lemma 4.7 for $A_1 = c_3\sqrt{n}$, $A_2 = c_3\sqrt{r}$ and $A_3 = 4 \cdot 2^{s_0/2}$, implying that for every $p \geq 1$,

$$\left\| \sup_{A \in \mathcal{A}} \|\pi_{s_0} A\xi\|_2^2 \right\|_{L_p} \leq (\mathbb{E}Z^{2p})^{1/p} \leq c_4(L,\beta)(\sqrt{n} + \sqrt{r}\max\{2^{s_0/2}, \sqrt{p}\})^2.$$

Setting $p = 2^{s_0} = \kappa_4 n/r$ we obtain

$$\left\| \sup_{A \in \mathcal{A}} \|\pi_{s_0} A\xi\|_2^2 \right\|_{L_p} \leq c_5(L,\beta)n;$$

therefore,

$$\|(N_{\mathcal{A}}(\xi))^2\|_{L_p} \leq c_6(L,\beta)\left(\gamma_{s_0,s_1}(\mathcal{A})\|N_{\mathcal{A}}(\xi)\|_{L_p} + n\right),$$

implying that

$$\|N_{\mathcal{A}}(\xi)\|_{L_p} \leq c_7(L,\beta)\max\{\gamma_{s_0,s_1}(\mathcal{A}), \sqrt{n}\}$$

and

$$\left(\mathbb{E}\sup_{A \in \mathcal{A}} \left|\|A\xi\|_2^2 - \|\pi_{s_0} A\xi\|_2^2\right|^p\right)^{1/p} \leq c_8(L,\beta)\gamma_{s_0,s_1}(\mathcal{A})\left(\gamma_{s_0,s_1}(\mathcal{A}) + \sqrt{n}\right).$$

The claim now follows from Lemma 3.7 and the definition of $p = \kappa_4 n/r$. ■

The next step is to estimate $\gamma_{s_0,s_1}(\mathcal{A})$ for our choice $\mathcal{A} = \{\Gamma_v : v \in V_{r,s_1}\}$. We will construct the admissible sequence $\mathcal{A}_s = \{\Gamma_v : v \in V_{r,s}\}$ for $s_0 \leq s < s_1$ based on the fact that

$$\|\Delta_s A\|_{2\to 2} = \|\Gamma_{\pi_{s+1}v} - \Gamma_{\pi_{s+1}v}\|_{2\to 2} = \sqrt{n}\|W\Delta_s v\|_\infty,$$

for $\Delta_s v = \pi_{s+1}v - \pi_s v$. Hence,

$$\gamma_{s_0,s_1}(\mathcal{A}) = \sup_{v \in V_{r,s_1}} \sum_{s=s_0}^{s_1-1} 2^{s/2} \cdot \sqrt{n}\|W\Delta_s v\|_\infty,$$

and the admissible sequence will be constructed as maximal separated subsets of V_{r,s_1} with respect to the norm $\sqrt{n}\|W(\cdot)\|_\infty$.

We will require a well-known fact, due to Carl [8].

Lemma 4.8 *There is an absolute constant c for which the following holds. Let $w_1, \ldots, w_n \in \mathbb{R}^n$ that satisfy $\|w_i\|_\infty \leq K$, put $\|z\| = \max_{i \in [n]} |\langle z, w_i\rangle|$ and set B to be the unit ball with respect to that norm. Then for every $t > 0$,*

$$\log N(\sqrt{r}B_1^n, tB) \leq C\frac{K^2 r}{t^2}\log^2(nt^2/r).$$

In the case we are interested in, $w_i = \sqrt{n} W_i$ and $\|w_i\|_\infty \leq \beta$. Moreover, $V_{r,s_1} \subset V_r \subset \sqrt{r} B_1^n$; therefore,

$$\log N(V_{r,s_1}, tB) \leq \log N(\sqrt{r} B_1^n, tB) \leq C \frac{\beta^2 r}{t^2} \log^2(nt^2/r). \tag{4.11}$$

Corollary 4.9 *There is an admissible sequence of V_{r,s_1}, for which, for every $v \in V_{r,s_1}$,*

$$\sqrt{n} \|W \Delta_s v\|_\infty \leq c\beta 2^{-s/2} \sqrt{r} \log(en/2^s)$$

for an absolute constant c. Therefore,

$$\sup_{v \in V_{r,s_1}} \sum_{s=s_0}^{s_1-1} 2^{s/2} \cdot \sqrt{n} \|W \Delta_s v\|_\infty \leq c_1 \beta \sqrt{r} \alpha_r \log(er/\kappa_4). \tag{4.12}$$

Corollary 4.9 follows from (4.11), a straightforward computation and the definition of s_0 and s_1. We omit the details.

Proof of Theorem 4.4. Combining the individual small ball estimate in (3.4), Theorem 4.6 and Corollary 4.9, one has that with probability at least $1 - 2^{-c_1(L,\beta)n/r}$, for every $v \in V_{r,s_1}$,

$$\|\Gamma_{\pi_{s_0} v} \xi\|_2 \geq \kappa_1 \sqrt{n},$$

and

$$\left| \|\Gamma_v \xi\|_2^2 - \|\Gamma_{\pi_{s_0} v} \xi\|_2^2 \right| \leq c_2(L,\beta) \sqrt{n} \cdot \sqrt{r} \alpha_r \log(c_2 r).$$

It is straightforward verify that the latter term is bounded by $(\kappa_1^2/4)n$ provided that $r \leq c_3(L,\beta) n / \log^4 n$. ■

The more difficult step consists in exposing the regular coordinate structure of vectors in a typical $\{\Gamma_v \xi : v \in V_{r,s_1}\}$, and we will do that by finding a suitable upper bound on

$$\sup_{v \in V_{r,s_1}} \|\Gamma_v \xi\|_{[k]} = \sup_{v \in V_{r,s_1}} \Big(\sum_{i \leq k} (\langle \Gamma_v \xi, e_i \rangle^*)^2 \Big)^{1/2}. \tag{4.13}$$

Specifically, the next result identifies the largest possible k for which (4.13) is smaller than $(\kappa_1/4)\sqrt{n}$.

Lemma 4.10 *Let $2^{s_0} \leq 2^{s_1}$ and $r \leq c_1 n / \log^4(n)$. Assume that*

$$k \leq c_2 \frac{n}{\alpha_r^2 \log(\alpha_r)}.$$

Then with probability at least $1 - 2^{-c_3 2^{s_0}}$

$$\sup_{v \in V_{r,s_1}} \|\Gamma_v \xi\|_{[k]} \leq \frac{\kappa_1}{4} \sqrt{n}. \tag{4.14}$$

The constants c_1, c_2, c_3 only depend on L and β.

Remark 4.11 *One should note that an upper bound on the supremum in* (4.14) *cannot be obtained via an individual estimate and the union bound. The subgaussian property of ξ is enough to ensure that for a fixed v,*

$$(\mathbb{E}\|\Gamma_v\xi\|_{[k]}^p)^{1/p} \le cL\left(\|v\|_2\sqrt{k\log(en/k)} + \sqrt{p}\sqrt{n}\|Wv\|_\infty\right).$$

However, because one only knows that for $v \in V_r$, $\sqrt{n}\|Wv\|_\infty \le \beta\sqrt{r}\|v\|_2$, individual tail estimates suffice for a uniform bound in V_{r,s_0}, but not in the much larger set V_{r,s_1}.

Proof. Observe that

$$\sup_{v\in V_{r,s_1}} \|\Gamma_v\xi\|_{[k]} = \sup_{x\in V_k}\sup_{v\in V_{r,s_1}} \langle \xi, \Gamma_v^* x\rangle.$$

We will study the supremum of the linear process $w \mapsto \langle \xi, w\rangle$ indexed by the set

$$\{\Gamma_v^* x : v \in V_{r,s_1},\ x \in V_k\}.$$

Let $(V_{k,s})_{s\ge s_0}$ be an admissible sequence of V_k which will be specified later on. For $x \in V_k$, we consider $\pi_{s_1}x \in V_{k,s_1}$, whose cardinality is $2^{2^{s_1}}$, and $\pi_{s_0}x \in V_{k,s_0}$, whose cardinality is $2^{2^{s_0}}$ and write, for $v \in V_{r,s_1}$,

$$\Gamma_v^* x = \underbrace{\Gamma_v^*(x-\pi_{s_1}x)}_{=:H_1} + \underbrace{(\Gamma_v^*\pi_{s_1}x - \Gamma_{\pi_{s_0}v}^*\pi_{s_0}x)}_{=:H_2} + \underbrace{\Gamma_{\pi_{s_0}v}^*\pi_{s_0}x}_{=:H_3}. \tag{4.15}$$

While H_3 corresponds to the "starting points" of every chain, the difference between H_1 and H_2 lies in the "balance" between the contribution of V_k and V_{r,s_1} to each one of them. For H_1, there are $\sim 2^{2^{s_1}}$ points $v \in V_{r,s_1}$, but for an admissible sequence for V_k one has

$$\Gamma_v^*(x-\pi_{s_1}x) = \sum_{s\ge s_1}\Gamma_v^*(\pi_{s+1}x - \pi_s x) = \sum_{s\ge s_1}(\Gamma_v^*\Delta_s x),$$

and for $s \ge s_1$, $|\{\Delta_s x : x \in V_k\}| \ge 2^{2^{s_1}} = |V_{r,s_1}|$. Hence, it is possible to treat H_2 for each $v \in V_{r,s_1}$ separately. In contrast, the situation for H_2 is "more balanced", and requires a different argument.

To deal with H_1 in (4.15) let us fix $v \in V_{r,s_1}$ and recall that by Lemma 3.5,

$$\begin{aligned}\mathbb{E}\sup_{t\in\Gamma_v^*V_k}\langle G,t\rangle =&\mathbb{E}\|\Gamma_v G\|_{[k]} = \mathbb{E}\Big(\sum_{i=1}^k (\langle G, \Gamma_v^* e_i\rangle^*)^2\Big)^{1/2}\\ \le& c\sqrt{k\log(en/k)}\max_{1\in[n]}\|\Gamma_v^* e_i\|_2 \le c\sqrt{k\log(en/k)}.\end{aligned}$$

By the Majorizing Measures Theorem [41], there exists an admissible sequence of V_k for which

$$\sup_{x\in V_k}\sum_{s=0}^\infty 2^{s/2}\|\Gamma_v^*\Delta_s x\|_2 \sim \mathbb{E}\sup_{t\in\Gamma_v^* V_k}\langle G,t\rangle \le c\sqrt{k\log(en/k)}.$$

Let us consider a part of that admissible sequence, namely, $(V_{k,s})_{s \geq s_1}$. As ξ is an isotropic, L-subgaussian random vector, it follows that for every $s \geq s_1$ and every $x \in V_k$, with probability at least $1 - 2^{-2^{s+3}}$,

$$|\langle \xi, \Gamma_v^* \Delta_s x \rangle| \leq cL 2^{s/2} \|\Gamma_v^* \Delta_s x\|_2.$$

Summing for $s \geq s_1$, one has that with probability at least $1 - 2^{-2^{s_1+2}}$, for every $x \in V_k$,

$$|\langle \xi, \Gamma_v^*(x - \pi_{s_1} x) \rangle| \leq \sum_{s \geq s_1} |\langle \xi, \Gamma_v^* \Delta_s x \rangle| \leq cL \sum_{s \geq s_1} 2^{s/2} \|\Gamma_v^* \Delta_s x\|_2 \leq cL\sqrt{k \log(en/k)}. \tag{4.16}$$

Repeating this argument for every $v \in V_{r,s_1}$ and applying the union bound, one has that with probability at least $1 - 2^{-2^{s_1+1}}$, (4.16) holds for every $v \in V_{r,s_1}$.

Next, let us turn to H_2 in (4.15). We will construct approximating subsets in the following way: let $(V_{r,s})_{s=s_0}^{s_1}$ be the admissible sequence of V_{r,s_1} used earlier, consisting of maximal separated subsets with respect to the norm $\sqrt{n}\|W(\cdot)\|_\infty$, and put ν_s to be the mesh width of the net $V_{r,s}$. For every $s_0 \leq s \leq s_1$ and $z \in V_{r,s}$, let $\| \ \|_z$ be the ellipsoid norm

$$\|x\|_z^2 = \|D_{Wz} U^* x\|_2^2 = n \sum_{\ell=1}^n \langle W_\ell, z \rangle^2 \langle U^\ell, x \rangle^2,$$

and set $T_s(z)$ to be a maximal separated subset in V_k with respect to $\| \ \|_z$, of cardinality 2^{2^s}. Denote its mesh width by $\varepsilon_s(z)$. Thus, for $s_0 \leq s < s_1$, $v \in V_{r,s+1}$ and $x \in V_k$,

$$\Gamma_v x = \Gamma_{v-v'}^* x + \Gamma_{v'}^*(x - x') + \Gamma_{v'}^* x', \tag{4.17}$$

where $v' \in V_{s,r}$ satisfies $\sqrt{n}\|W(v - v')\|_\infty \leq \nu_s$, and $x' \in V_k$ belongs to $T_s(v')$ – the net of V_k with respect to the norm $\| \ \|_{v'}$, and fulfills

$$\|\Gamma_{v'}^*(x - x')\|_2 = \|x - x'\|_{v'} \leq \varepsilon_s(v'). \tag{4.18}$$

Moreover, the cardinality of $\bigcup_{v' \in V_{r,s}} T_s(v')$ is at most $2^{2^s} \cdot 2^{2^s} \leq 2^{2^{s+1}}$.

The required estimate on H_2 follows once we bound ν_s – that is, establish a covering number estimate for V_{r,s_1} with respect to the norm $\sqrt{n}\|W(\cdot)\|_\infty$ – and control $\varepsilon_s(z)$ with respect to each one of the $\| \ \|_z$ norms. Indeed,

$$|\langle \xi, \Gamma_v^* x \rangle| \leq |\langle \xi, \Gamma_{v-v'}^* x \rangle| + |\langle \xi, \Gamma_{v'}^*(x - x') \rangle| + |\langle \xi, \Gamma_{v'}^* x' \rangle|$$

for $v \in V_{r,s+1}$, $v' \in V_{r,s}$, $x \in T_{s+1}(v)$ and $x' \in T_s(v')$. Since ξ is isotropic and L-subgaussian, we have with probability at least $1 - 2^{2^{s+2}}$, for every $v \in V_{r,s+1}$ and $x \in T_{s+1}(v)$,

$$\begin{aligned}
|\langle \xi, \Gamma_v^* x \rangle| \leq & cL\big(2^{s/2}\|\Gamma_{v-v'}^* x\|_2 + 2^{s/2}\|\Gamma_{v'}^*(x - x')\|_2\big) + |\langle \xi, \Gamma_{v'}^* x' \rangle| \\
\leq & cL\big(2^{s/2} \cdot \sqrt{n}\|W(v - v')\|_2 + 2^{s/2}\|x - x'\|_{v'}\big) + |\langle \xi, \Gamma_{v'}^* x' \rangle| \\
\leq & cL\big(2^{s/2}\nu_s + 2^{s/2} \sup_{z \in V_r} \varepsilon_s(z)\big) + |\langle \xi, \Gamma_{v'}^* x' \rangle|,
\end{aligned}$$

for $v' \in V_{r,s}$ and $x' \in \bigcup_{v' \in V_{r,s}} T_s(v')$, hence,

$$|\langle \xi, \Gamma_v^* x - \Gamma_{v'}^* x' \rangle| \leq cL\big(2^{s/2}\nu_s + 2^{s/2} \sup_{z \in V_r} \varepsilon_s(z)\big). \tag{4.19}$$

Iterating (4.19), summing for $s_0 \le s < s_1$ and recalling that $2^{s_0} = \kappa_4 n/r$, it follows that with probability at least $1 - 2^{-c_1(L,\beta)n/r}$, for every $v \in V_{r,s_1}$ and every $x \in V_k$,

$$|\langle \xi, \Gamma_v^* \pi_{s_1} x - \Gamma_{\pi_{s_0} v}^* \pi_{s_0} x \rangle| \le c_2(L,\beta)\Big(\sum_{s=s_0}^{s_1-1} 2^{s/2}\nu_s + \sum_{s=s_0}^{s_1-1} 2^{s/2} \sup_{z\in V_r} \varepsilon_s(z)\Big), \tag{4.20}$$

and $\pi_{s_0} v \in V_{r,s_0}$, $\pi_{s_0} x \in \bigcup_{v\in V_{r,s_0}} T_s(v)$.

The first sum in (4.20) has been estimated earlier, in (4.12). In particular,

$$c_2 \sum_{s=s_0}^{s_1-1} 2^{s/2}\nu_s \le c_3\sqrt{r}\alpha_r \cdot \log(c_3 r) \le (\kappa_1/16)\sqrt{n}$$

for $c_3 = c_3(L,\beta)$ and as long as $r \le c_4(L,\beta) n/\log^4 n$.

To bound the second sum in (4.20) we require another covering estimate.

Theorem 4.12 *There exists an absolute constant c for which, for every $z \in S^{n-1}$,*

$$\log N(V_k, \|\ \|_z, \varepsilon) \le c\beta\varepsilon^{-2} k \log(en/k).$$

Proof. The proof of Theorem 4.12 is an outcome of Sudakov's inequality. Fix $z \in S^{n-1}$ and define the linear operator $S : \mathbb{R}^n \to \mathbb{R}^n$ by $Se_i = U^i$. Observe that for $i \in [n]$ and $t \in \mathbb{R}^n$

$$\langle \sqrt{n} D_{Wz} S^* t, e_i \rangle = \sqrt{n}\langle S^* t, D_{Wz} e_i\rangle = \sqrt{n}\langle W_i, z\rangle\langle S^* t, e_i\rangle = \sqrt{n}\langle W_i, z\rangle\langle U^i, t\rangle.$$

Therefore, $\|\sqrt{n} D_{Wz} S^* t\|_2 = \|t\|_z$ and $\log N(V_k, \|\ \|_z, \varepsilon) = \log N(\sqrt{n} D_{Wz} S^* V_k, \varepsilon B_2^n)$.

Set $T = \sqrt{n} D_{Wz} S^* V_k$ and let G be the standard Gaussian vector in $\mathbb{R}^n$. By Sudakov's inequality (see, e.g., [27]), there is an absolute constant c for which

$$c\varepsilon^2 \log N(T, \varepsilon B_2^n) \le \mathbb{E} \sup_{t\in T} \langle G, t\rangle,$$

where G is a standard Gaussian vector in $\mathbb{R}^n$, and

$$\mathbb{E}\sup_{t\in T}\langle G, t\rangle = \mathbb{E}\sup_{v\in V_k}\langle G, \sqrt{n} D_{Wz} S^* v\rangle = \sqrt{n}\mathbb{E}\sup_{v\in V_k}\sum_{i=1}^n g_i \langle W_i, z\rangle\langle U^i, v\rangle \equiv \mathbb{E}\Big(\sum_{j\le k}(Z_j^*)^2\Big)^{1/2},$$

for $Z_j = \sqrt{n}\sum_{i=1}^n g_i\langle W_i, z\rangle\langle U^i, e_j\rangle$. Each one of the Z_j's is a Gaussian variable, and since U is a Hadamard type matrix with constant β,

$$\mathbb{E}Z_j^2 = n\sum_{i=1}^n \langle W_i, z\rangle^2 \langle U^i, e_j\rangle^2 \le \beta^2.$$

Finally, by Lemma 3.5,

$$\mathbb{E}\Big(\sum_{j=1}^k (Z_j^*)^2\Big)^{1/2} \le c\beta\sqrt{k\log(en/k)}.$$

■

Invoking Theorem 4.12, it follows that for every $z \in V_r$, $\varepsilon_s(z) \leq \beta 2^{-s/2}\sqrt{k\log(en/k)}$. Hence,

$$\sup_{z\in V_r} \varepsilon_s(z) \leq \beta 2^{-s/2}\sqrt{k\log(en/k)},$$

and the second sum in (4.20) is bounded by

$$c(L,\beta)(s_1-s_0)\sqrt{k\log(en/k)} \leq c_1(L,\beta)\alpha_r\sqrt{k\log(en/k)} \leq (\kappa_1/16)\sqrt{n}$$

provided that

$$k \leq c_2(L,\beta)\frac{n}{\alpha_r^2\log\alpha_r}. \tag{4.21}$$

This concludes the required estimate on H_2 and leads to the condition on k.

The final and easiest component is to control H_3 in (4.15). Indeed,

$$\|\Gamma^*_{\pi_{s_0}v}\pi_{s_0}x\|_2 \leq \sqrt{n}\|W_{\pi_{s_0}}v\|_\infty \leq \beta\sqrt{r},$$

and there are at most $2^{2^{s_0}} \cdot 2^{2^{s_0}} = 2^{2^{s_0+1}}$ pairs $(\pi_{s_0}v, \pi_{s_0}x)$. Since ξ is isotropic and L-subgaussian, one has that with probability at least $1-2\exp(-u^2/2)$,

$$|\langle\xi, \Gamma^*_{\pi_{s_0}v}\pi_{s_0}x\rangle| \leq Lu\|\Gamma^*_{\pi_{s_0}v}\pi_{s_0}x\|_2 \leq Lu\beta\sqrt{r}.$$

Let $u = 2^{(s_0+2)/2}$, and by the union bound, with probability at least $1-2\exp(-c2^{s_0})$,

$$\sup_{v\in V_{r,s_1}, x\in V_k} |\langle\xi, \Gamma^*_{\pi_{s_0}v}\pi_{s_0}x\rangle| \leq 2^{(s_0+2)/2}L\beta\sqrt{r} \leq 2\cdot\frac{\kappa_1}{8\kappa_3 L\beta}\sqrt{\frac{n}{r}}\cdot L\beta\sqrt{r} \leq \frac{\kappa_1}{4}\sqrt{n},$$

because $\kappa_3, L, \beta \geq 1$. Taking the union bound over the events in which the bounds for H_1, H_2 and H_3 apply, we deduce that

$$\sup_{v\in V_{r,s_1}} \|\Gamma_v\xi\|_{[k]} \leq \kappa_1\sqrt{n}$$

with probability at least $1-2^{-c_1 n/r} - 2e^{-c_2 2^{s_0}} - 2^{-2^{s_1+1}} \geq 1-2^{-c' 2^{s_0}}$ for a suitable constant c', under condition (4.21) on k. This completes the proof of Lemma 4.10 after relabelling constants. ■

5 The upper bound on one-sparse vectors

To complete the proof of Theorem 2.5 one has to show that with high probability,

$$\max_{i\in[n]} \|P_\Omega\Gamma_{e_i}\xi\|_2 \leq c\sqrt{\delta n},$$

for a suitable constant c and $\Omega = \{i : \delta_i = 1\}$, see also Theorem 2.3.

Theorem 5.1 *There exist constants c_0, c_1 and c_2 that depend only on L and β for which the following holds. If $\delta \geq c_0\frac{\log n}{n}$, then with probability at least $1-2\exp(-c_1\delta n)$,*

$$\max_{i\in[n]}\left(\sum_{j=1}^n \delta_j\langle\Gamma_{e_i}\xi, e_j\rangle^2\right)^{1/2} \leq c_2\sqrt{\delta n}.$$

The proof of Theorem 5.1 is based on the fact that for every $i \in [n]$, $\Gamma_{e_i}\xi$ has a regular coordinate structure, in the sense that is clarified in the following lemma.

Lemma 5.2 *There exist an absolute constant c_1 and a constant c_2 that depends on β and L for which the following holds. Let ξ be an isotropic, L-subgaussian random vector and let $m \in [n]$. Then for every $i \in [n]$, with probability at least $1-\exp(-c_1 m \log(en/m))$, $\Gamma_{e_i}\xi = x+y$, where*

$$\|x\|_2 \le c_2\sqrt{m\log(en/m)} \ \text{ and } \ \max_{i\in[n]} \frac{y_i^*}{\sqrt{\log(en/i)}} \le c_2.$$

Proof. Let $i \in [n]$ and set $z = \Gamma_{e_i}\xi$. Let I be the (random) set of the m largest coordinates of z and define

$$x = \sum_{j\in I} z_j e_j \ \text{ and } \ y = \sum_{j\in I^c} z_j e_j. \tag{5.1}$$

Observe that $\|x\|_2 = \|\Gamma_{e_i}\xi\|_{[m]}$. By (3.5) for $p = m\log(en/m)$, and noting that $\|We_i\|_\infty \le \beta/\sqrt{n}$, one has

$$Pr(\|x\|_2 \ge c(L,\beta)\sqrt{m\log(en/m)}) \le \exp(-m\log(en/m)).$$

Repeating this argument for $m \le \ell \le n$, it follows that with probability at least

$$1 - \sum_{\ell=m}^{n} \exp(-\ell\log(en/\ell)) \ge 1 - \exp(-c_1 m\log(en/m)),$$

for every $m \le \ell \le n$,

$$z_\ell^* \le \frac{1}{\sqrt{\ell}}\|z\|_{[\ell]} \le 2c(\beta,L)\sqrt{\log(en/\ell)},$$

and the claim follows. ∎

The coordinate structure of $\Gamma_{e_i}\xi$ comes into play thanks to a fact from [30].

Lemma 5.3 *Let $a \in \mathbb{R}^n$, set $\|a\|_{\psi_1^n} = \max_{1\le j\le n} a_j^*/\log(en/j)$ and put $0 < t < \|a\|_{\psi_1^n}/2$. Then*

$$Pr\Big(\big|\sum_{j=1}^{n}(\delta_j-\delta)a_j\big| > t\delta n\Big) \le 2\exp(-ct^2\delta n/\|a\|_{\psi_1^n}^2)$$

where c is an absolute constant and $(\delta_i)_{j=1}^n$ are independent selectors with mean δ.

In particular, if $a_j = \log(en/j)$ then with probability at least $1-2\exp(-c_1\delta n)$,

$$\sum_{i=1}^{n} \delta_j \log(en/j) \le 5\delta n.$$

Proof of Theorem 5.1. Let $m = \delta n/\log(e/\delta)$ and consider the decomposition of $\Gamma_{e_i}\xi = x+y$ established in Lemma 5.2. Conditioned on the event from that lemma which holds with probability at least $1-2\exp(-c_0\delta n)$,

$$\Big(\sum_{j=1}^{n}\delta_j x_j^2\Big)^{1/2} \le \Big(\sum_{j=1}^{n} x_j^2\Big)^{1/2} \le c_1(L,\beta)\sqrt{m\log(en/m)} \le c_2(L,\beta)\sqrt{\delta n}.$$

Also, by Lemma 5.3, with probability at least $1-2\exp(-c_3\delta n)$,

$$\sum_{j=1}^{n}\delta_j y_j^2 \le c_4(L,\beta)\sum_{i=1}^{n}\delta_j\log(en/j)\le c_5(L,\beta)\delta n.$$

Hence, with probability at least $1-2\exp(-c_6\delta n)$ with respect to both ξ and $(\delta_j)_{j=1}^n$, one has that

$$\Big(\sum_{j=1}^{n}\delta_j\big\langle\Gamma_{e_i}\xi,e_j\big\rangle^2\Big)^{1/2}\le c_7(L,\beta)\sqrt{\delta n}.$$

Recalling that $\delta n\ge c_8\log n$ for a well chosen c_8, it follows from the union bound that with probability at least $1-2n\exp(-c_6\delta n)\ge 1-2\exp(-c_9\delta n)$,

$$\max_{i\in[n]}\Big(\sum_{j=1}^{n}\delta_j\big\langle\Gamma_{e_i}\xi,e_j\big\rangle^2\Big)^{1/2}\le c_7(L,\beta)\sqrt{\delta n}.$$

■

6 The ℓ_q-robust null space property for $q>2$

We will now extend from the ℓ_2-robustness to ℓ_q-robustness for general $2\le q\le\infty$. We state our main result for the general type of matrices as used before, i.e., for three Hadamard type matrices U,W,O with constant β,

$$Av=\Gamma_v\xi=\sqrt{n}UD_{Wv}O\xi,$$

where ξ is a random vector with independent mean-zero, variance one, L-subgaussian coordinates. Clearly, random circulant matrices fall into this class of random matrices A.

Theorem 6.1 *Let $2\le q\le\infty$ and A be a draw of an $n\times n$ random matrix as described above. Let $(\delta_i)_{i\in[n]}$ be a sequence of independent selectors with mean δ, set $\Omega=\{i\in[n]:\delta_i=1\}$ and put $B=P_\Omega A$. For $\nu\in(0,1)$, set $\alpha_{s,\nu}=\log\left(\frac{s^2}{\nu^2 n}\max\left\{\log\left(\frac{e\nu^2 n}{s}\right),\log(\nu^{-2}s)\right\}\right)$,*

$$\zeta(s)=\begin{cases} c_1 & \text{if } s\le c_2\sqrt{\frac{n}{\log n}},\\ c_3\alpha_{s,\nu}^2\log(\alpha_{s,\nu}) & \text{if } c_2\sqrt{\frac{n}{\log n}}\le s\le c_3\frac{n}{\log^4(n)}.\end{cases}$$

and $\rho(s)=\max\{1,\log(e\nu^{-2}r)/\log(en/s)\}$. Assume, for $2\le q<\infty$,

$$\delta n\ge c_4\nu^{-2}s\log(en/s)\rho(s)\zeta(s)\max\{q^2,q\log^{1-2/q}(en)\}\quad\text{and}\quad\delta n\ge c_5(q/2)^{-q/2}\log^{q-2}(en)\tag{6.1}$$

and, for $q=\infty$,

$$\delta n\ge c_7\nu^{-2}s\log(en/s)\rho(s)\zeta(s)\log^2(en).\tag{6.2}$$

Then with probability at least $1-n^{-c_6}$ the following holds: For all $x\in\mathbb{R}^n$, all $e\in\mathbb{R}^m$ with $\|e\|_q\le\eta$ and $y=Bx+e$, the minimizer $x^\sharp$ of

$$\min\|z\|_1\quad\textit{subject to}\ \|Bz-y\|_q\le\eta$$

satisfies

$$\|x - x^\sharp\|_1 \leq \frac{C}{1-\nu}\sigma_s(x)_1 + \frac{D}{1-\nu}\frac{\sqrt{s}\eta}{(\delta n/\zeta(s))^{1/q}} \quad \text{and}$$
$$\|x - x^\sharp\|_2 \leq \frac{C}{1-\nu}\frac{\sigma_s(x)_1}{\sqrt{s}} + \frac{D}{1-\nu}\frac{\eta}{(\delta n/\zeta(s))^{1/q}},$$

where we set $1/q = 0$ for $q = \infty$. All constants $c_0, \ldots, c_8$ only depend on L and β.

Remark.

a) In the theorem above, it suffices to restrict q to the range $[2, \log(en)]$ because for $q \geq \log(en) = q_0$, $\|\cdot\|_q$ and $\|\cdot\|_{q_0}$ are equivalent up to absolute constants.

b) For $q > 2$, the conditions (6.1), (6.2) on the required number of measurements δn have more logarithmic factors than the one for $q = 2$. It is presently not clear whether this is an artefact of the proof and one can work with only one logarithmic factor as in the case of matrices with i.i.d. entries [12].

c) The case $q = \infty$ is important for quantized compressive sensing, where one would like to work with consistent reconstruction methods such as ℓ_∞-constraint ℓ_1-minimization,

$$\min \|z\|_1 \quad \text{subject to } \|Bz - y\|_\infty \leq \eta.$$

Our result represents the first near-optimal bound for structured random matrices in this context. We refer to [12] for a detailed discussion of connections to quantized compressive sensing. In the low sparsity case $s \leq c_2\sqrt{\frac{n}{\log n}}$ the condition of the required number of samples reads

$$\delta n \geq c'\nu^{-2} s \log(en/s) \log^2(en).$$

d) The logarithmic factors of $\log(en)$ may actually be slightly improved to $\log(e\delta n)$, but for convenience we stated the theorem in the above way.

We first provide an extension of Theorem 2.3 in order to reduce the ℓ_q-null space property to a lower bound of $\|Bx\|_q$ over $x \in V_r$.

Theorem 6.2 *Let $2 \leq q < \infty$. Assume that, for $B \in \mathbb{R}^{m\times n}$,*

$$\inf_{x\in V_r} \|Bx\|_q \geq \tau_q^{-1} \quad \text{and} \quad \max_{j\in[n]} \|Be_j\|_q \leq M_q.$$

If $r \geq 10(2+\nu^{-1})^2\tau_q^2 M_q^2 q\, s$, then

$$\inf_{x\in\mathcal{T}_{\nu,s}\cap S^{n-1}} \|Bx\|_q \geq \frac{1}{2\,\tau_q}.$$

Proof. For a vector $z \in \mathbb{R}^n \setminus \{0\}$, we introduce the (discrete) random vector W on $\mathbb{R}^n$ defined via

$$\mathbb{P}(W = \operatorname{sign}(z_j)\|z\|_1 e_j) = \frac{|z_j|}{\|z\|_1}.$$

Then $\mathbb{E}W = z$. We take r independent copies $W_1, \ldots, W_r$ of W and define $Z = \frac{1}{r}\sum_{j=1}^r W_j$, so that $\mathbb{E}Z = z$ and $Z/\|Z\|_2 \in V_r$ for every realization of Z so that by assumption $\|BZ\|_q \geq \tau_q^{-1}\|Z\|_2$ and

$$\mathbb{E}\|BZ\|_q^2 \geq \tau_q^{-2}\mathbb{E}\|Z\|_2^2.$$

By the triangle inequality,

$$\|Bz\|_q \geq (\mathbb{E}\|BZ\|_q^2)^{1/2} - (\mathbb{E}\|BZ - Bz\|_q^2)^{1/2} \geq \tau_q^{-1}\left(\mathbb{E}\|Z\|_2^2\right)^{1/2} - \left(\mathbb{E}\|BZ - Bz\|_q^q\right)^{1/q}.$$

The expectation in the first term can be computed as

$$\mathbb{E}\|Z\|_2^2 = \frac{1}{r^2}\sum_{i,j=1}^r \mathbb{E}\langle W_i, W_j\rangle = \frac{1}{r}\mathbb{E}\langle W, W\rangle + \frac{1}{r^2}\sum_{i\neq j}\mathbb{E}\langle W_i, W_j\rangle = \frac{\|z\|_1^2}{r} + \frac{r-1}{r}\|z\|_2^2.$$

Denoting by $(\epsilon_j)_{j=1}^r$ a Rademacher vector independent of (W_j), symmetrization and Khintchine's inequality yields

$$\begin{aligned}
\left(\mathbb{E}\|BZ - Bz\|_q^q\right)^{1/q} &= \left(\mathbb{E}\|r^{-1}\sum_{j=1}^r (BZ - \mathbb{E}BZ)\|_q^q\right)^{1/q} \leq \frac{2}{r}\left(\mathbb{E}\|\sum_{j=1}^r \epsilon_j BW_j\|_q^q\right)^{1/q} \\
&= \frac{2}{r}\left(\sum_{k=1}^m \mathbb{E}|\sum_{j=1}^r \epsilon_j (BW_j)_k|^q\right)^{1/q} \leq \frac{2C\sqrt{q}}{r}\left(\mathbb{E}\sum_{k=1}^m\left(\sum_{j=1}^r |(BW_j)_k|^2\right)^{q/2}\right)^{1/q} \\
&\leq \frac{2C\sqrt{q}}{r}\left(\sum_{j=1}^r \mathbb{E}\|BW_j\|_q^2\right)^{1/2} \leq \frac{2C\sqrt{q}}{r}M_p r^{1/2}\|z\|_1 = \frac{2C\sqrt{q}M_p}{\sqrt{r}}\|z\|_1,
\end{aligned}$$

where $C = 2^{3/8}e^{-1/2}$ (see e.g. [17, Corollary 8.7]). Hereby, we have also used that $q \geq 2$ and that $\|Be_j\|_q \leq M_q$. Altogether, we obtain

$$\|Bz\|_q \geq \tau_q^{-1}\sqrt{\frac{\|z\|_1^2}{r} + \frac{r-1}{r}\|z\|_2^2} - \frac{2C\sqrt{q}M_q}{\sqrt{r}}\|z\|_1 \geq \frac{\|z\|_2}{\tau_q} - \frac{2C\sqrt{q}M_q\|z\|_1}{\sqrt{r}}.$$

Now for $z \in \mathcal{T}_{\nu,s} \cap S^{n-1} \subset (2+\nu^{-1})\operatorname{conv} V_s$ we have $\|z\|_1 \leq (2+\nu^{-1})\sqrt{s}$, so that

$$\|Bz\|_q \geq \tau_q^{-1} - \frac{2C\sqrt{q}M_q(2+\nu^{-1})\sqrt{s}}{\sqrt{r}} \geq \frac{1}{2\tau_q}$$

provided that

$$r \geq \widetilde{C}\tau_q^2(2+\nu^{-1})^2 M_q^2 qs$$

with $\widetilde{C} = 16C^2 = 16 \cdot 2^{3/4}e^{-1} \approx 9.899 < 10.$ ∎

Remark. The same proof strategy can also be applied for $1 \le q < 2$. However, since ℓ_q is of type q for $1 \le q < 2$, see e.g. [27, Chapter 9], one only obtains $\mathbb{E}\|BZ - Bz\|_q \le \frac{C'}{r} M_q r^{1/q} \|z\|_1$ in this case, leading to the condition $r^{1-1/q} \ge C\tau_q \sqrt{s}$ or

$$r \gtrsim C_q \, s^{\frac{q}{2(q-1)}}$$

for $1 < q < 2$, hence, a polynomial scaling of r in s and as a result also polynomial scaling of the number of measurements in s in the end. It is presently not clear whether a different proof strategy may mend this problem.

Let us now consider the lower bound over V_r. The following result generalizes Corollary 4.2 and holds for any $1 \le q < \infty$.

Theorem 6.3 *Let $1 \le q < \infty$. There exist constants $c_0, \ldots, c_5$ that depend only on L and β for which the following holds. Let*

$$\zeta_r = \begin{cases} c_0, & \text{if } r \le c_1 \sqrt{\frac{\kappa_4 n}{\log(c_1 n/\kappa_4)}} \\ \alpha_r^2 \log(\alpha_r), & \text{if } c_1 \sqrt{\frac{\kappa_4 n}{\log(c_1 n/\kappa_4)}} \le r \le c_2 \frac{n}{\log^4(n)} \end{cases} \tag{6.3}$$

and assume that

$$\delta n \ge c_3 \rho \zeta_r r \log(en/r),$$

with ρ defined in (3.6). Set $(\delta_i)_{i=1}$ to be independent selectors with mean δ and $\Omega = \{i : \delta_i = 1\}$. Then with probability at least $1 - 2^{-c_4 \min\{2^{s_0}, 2^{s_1}\}}$,

$$\inf_{v \in V_r} \|P_\Omega \Gamma_v \xi\|_q \ge c_5 (\delta n/\zeta_r)^{1/q}. \tag{6.4}$$

Proof. According to Theorem 4.1, with probability at least $1 - 2^{-c_2 \min\{2^{s_0}, 2^{s_1}\}}$ we have

$$\{\Gamma_v \xi : v \in V_r\} \subset T_1 + T_2$$

with $|T_1| \le 2^{2^{s_1}}$ and $T_2 \subset c_1 n^{-3/2} B_2^n$. We assume in the following to be in that event. For $v \in V_r$ we write $\Gamma_v \xi = t + y$ with $t \in T_1$ and $y \in T_2$ and conclude that

$$\begin{aligned} \left(\sum_{i=1}^n \delta_i |\langle \Gamma_v \xi, e_i \rangle|^q\right)^{1/q} &\ge \inf_{t \in T_1} \left(\sum_{i=1}^n \delta_i |t_i|^q\right)^{1/q} - \sup_{y \in T_2} \left(\sum_{i=1}^n \delta_i |y_i|^q\right)^{1/q} \\ \ge \inf_{t \in T_1} \left(\sum_{i=1}^n \delta_i |y_i|^q\right)^{1/q} &- \sup_{y \in T_2} \sum_{i=1}^n |y_i| \ge \inf_{t \in T_1} \left(\sum_{i=1}^n \delta_i |y_i|^q\right)^{1/q} - n^{1/2} \sup_{y \in T_2} \|y\|_2 \\ \ge \inf_{t \in T_1} \left(\sum_{i=1}^n \delta_i |y_i|^q\right)^{1/p} &- c_1 n^{-1}. \end{aligned}$$

It remains to estimate $\inf_{t \in T_1} (\sum_{i=1}^n \delta_i |y_i|^q)^{1/q}$ from below. Recall from the proof of Theorem 4.1 that $T_1 = V_{r,s_1}$ and for a suitable constant c, on the event of probability at least $1 - 2^{-c_1 \min\{2^{s_0}, 2^{s_1}\}}$, for every $t = \Gamma_v \xi \in T_1$,

$$\|t\|_2 \ge c\sqrt{n} \quad \text{ and } \|t\|_{[\theta n]} \le c\sqrt{n}/2, \tag{6.5}$$

see (4.2) as well as (4.8) and Lemma 4.10, where θ is defined in (4.3) and the constant c only depends on L, β. Now, for $t \in T_1$, we choose $I = I(t)$ to be the index set corresponding to the θn largest absolute coefficients of t, and consider $x = P_I t$. Then

$$t^*_{\theta n} \geq \left(\frac{1}{(1-\theta)n} \sum_{i=\theta n}^{n} (t_i^*)^2 \right)^{1/2} = \left(\frac{1}{(1-\theta)n} \left(\|t\|_2^2 - \|t\|_{[\theta n]}^2 \right) \right)^{1/2}$$
$$\geq \left(\frac{1}{n} (c^2 n - c^2 n/4) \right)^{1/2} = \frac{\sqrt{3}c}{2} = c_1.$$

For $\Omega = \{i : \delta_i = 1\}$ let $K = K(t) = I(t) \cap \Omega$. Then the cardinality of K satisfies $|K| \geq \frac{1}{2}\delta|I| \geq c_2 \theta \delta n$ with probability at least $1 - e^{-c_3 \delta \theta n}$. On this event,

$$\|P_\Omega t\|_q \geq \left(\sum_{j \in K} |t_j|^q \right)^{1/q} \geq c_1 (c_2 \theta \delta n)^{1/q}$$

By the union over all $t \in T_1$ we conclude that $\inf_{t \in T_1} \|P_\Omega t\|_q \geq c_3 (\theta \delta n)^{1/q}$ with probability at least

$$1 - 2|T_1| \exp(-c_3 \delta \theta n) = 1 - 2 \cdot 2^{2^{s_1}} \exp(-c_3 \delta \theta n).$$

Hence, by definition of s_1, see Definition 3.11, if

$$\delta n \geq C \theta^{-1} \rho r \log(en/r)$$

then (6.4) holds with probability at least $1 - 2^{-c_0 2^{s_1}}$, completing the proof by setting $\zeta_r = \theta^{-1}$. ∎

As the next step, we provide an upper bound for $\|P_\Omega \Gamma_{e_i} \xi\|_q$.

Theorem 6.4 *Let $2 \leq q < \infty$. There are constants $c_0, \ldots, c_4 > 0$ such that if $\delta n \geq c_0 \log(n) \max\{1, \log^{q-2}(e/\delta), (q/2)^{-q/2}\}$ then with probability at least $1 - n^{-c_1}$, it holds*

$$\max_{i \in [n]} \|P_\Omega \Gamma_{e_i} \xi\|_q \leq c_2 (\delta n)^{1/q} \sqrt{\max\{q, \log(n)^{1-2/q}\}},$$

where $\Omega = \{i \in [n] : \delta_i = 1\}$. Moreover, if $\delta n \geq c_0 \log(en)$, then

$$\max_{i \in [n]} \|P_\Omega \Gamma_{e_i} \xi\|_\infty \leq c_3 \sqrt{\log(en)}, \tag{6.6}$$

with probability at least $1 - n^{-c_4}$.

Proof. We proceed similarly to Section 5. For fixed $i \in [n]$, let $z = \Gamma_{e_i} \xi$ and, for $m \in [n]$ let I be the random set of m largest absolute coordinates of z. We define x and y as in (5.1). As in the proof of Lemma (5.2) we conclude that

$$Pr(\|x\|_{[m]} \geq c(L, \beta) \sqrt{m \log(en/m)}) \leq \exp(-m \log(en/m))$$

and

$$\max_{i\in[n]} \frac{z_i^*}{\sqrt{\log(en/i)}} \le c_2$$

with probability at least $1-\exp(-c_1 m\log(en/m))$. In conclusion

$$y_i^* \le c_2\sqrt{\log\left(\frac{en}{m+i-1}\right)}.$$

Moreover, by (3.5) for $k=1$ and $p=(\alpha+1)\log(n)$,

$$\|x\|_\infty = \|z\|_\infty = \|\Gamma_{e_i}\xi\|_\infty \le c(\alpha+1)\sqrt{\log(en)}, \tag{6.7}$$

with probability at least $1-e^{-(\alpha+1)\log(n)} = 1-n^{-\alpha-1}$. It follows that

$$\|x\|_q \le \|x\|_2^{2/q}\|x\|_\infty^{1-2/q} \le c_3(\alpha+1)^{1-2/q} m^{1/q}\log^{1/(2q)}\left(\frac{en}{m}\right)\log^{1/2-1/(2q)}(en)$$

with probability at least $1-n^{-\alpha}-\exp(-c_4 m\log(en/m))$. In order to apply Lemma 5.3 for $\sum_{j=1}^n \delta_j|y_j|^q$ we note that the sequence $a_j = \log^{q/2}\left(\frac{en}{m+j-1}\right)$ satisfies

$$\sum_{j=1}^n a_j \le c_5\left(\frac{q}{2}\right)^{q/2} n$$

and $\|a\|_{\psi_1^n} \le \log^{q/2-1}\left(\frac{en}{m}\right)$. Lemma 5.3 implies that

$$Pr\left(\sum_{j=1}^n \delta_j|y_j|^q \ge 2c_5(q/2)^{q/2}\delta n\right) \le 2\exp\left(-c_6\frac{\delta n(q/2)^{q/2}}{\log^{q-2}(en/m)}\right).$$

Now choose $m=\delta n/\log(e/\delta)$. Then

$$\left(\sum_{j=1}^n \delta_j|x_j|^q\right)^{1/q} \le \|x\|_q \le c_3(\alpha+1)^{1-2/q}(\delta n)^{1/q}\log^{1/2-1/q)}(en)$$

with probability at least $1-n^{-\alpha+1}-\exp(-c_7\delta n)$ and

$$\|P_\Omega y\|_q \le c_8\sqrt{q/2}(\delta n)^{1/q}$$

with probability at least $1-\exp(-c_9\delta n(q/2)^{q/2}/\log^{q-2}(e/\delta))$. Taking the union bound over the events for x and y and then over $i\in[n]$ while using that

$$\delta n \ge c_0\max\{\log(n), \log^{q-2}(e/\delta)(q/2)^{-q/2}\}$$

we obtain

$$\max_{i\in I}\|P_\Omega\Gamma_{e_i}\xi\|_q \le c_8\sqrt{\max\{q,\log^{1-2/q}(en)\}}(\delta n)^{1/q}$$

with probability at least $1-n^{-\alpha}-n^{-c_7c_0+1}-n^{-c_9c_0+1}$. For suitable c_0 and α this proves the claim for $2\le q<\infty$. For $q=\infty$, the claim follows by taking the union over the events for $i\in[n]$ which ensure (6.7). ∎

We are now in the position to conclude the proof of the main result of this section.

Proof of Theorem 6.1. The constants M_q and τ_q of Theorem 6.2 can be chosen according Theorems 6.3 and 6.4 as

$$\tau_q^{-1} = c_1(\delta n/\tau)^{1/q} \quad \text{and} \quad M_q = c_2(\delta n)^{1/q}\sqrt{\max\{q, \log(en)^{1-2/q}\}}$$

under the condition

$$\delta n \geq c_3 \rho \zeta_r r \log(en/r),$$

where ζ_r is defined in (6.3). With $r = 10(2+\nu^{-1})^2 \tau_q^2 M_q^2 q\, s$, Theorems 6.3–6.4 yield

$$\inf_{x \in \mathcal{T}_{\nu,s} \cap S^{n-1}} \|P_\Omega \Gamma_x \xi\|_q \geq c_4(\delta n/\zeta_r)^{1/q}$$

with probability at least $1 - n^{-c_5}$ provided that

$$\delta n \geq c_6 \nu^{-2} \rho \zeta_r q s \max\{q, \log^{1-2/q}(en)\} \log(en/s).$$

Since it suffices to consider $2 \leq q \leq \log(en) = q_0$ by equivalence of $\|\cdot\|_{q_0}$ and $\|\cdot\|_q$ for $q > q_0$, the quantity ζ_r is equivalent to $\zeta(s)$ defined in Theorem 6.1 and ρ turns into $\rho(s)$. By Theorem 2.2, this concludes the proof for the case $q < \infty$.

The case $q = \infty$ is proven in the same way by using (6.6) and observing that for $q_0 = \log(en)$, the norms $\|\cdot\|_{q_0}$ and $\|\cdot\|_\infty$ are equivalent. ∎

References

[1] N. Ailon and E. Liberty. Fast Dimension Reduction Using Rademacher Series on Dual BCH Codes. *Discrete & Computational Geometry*, 42(4):615–630, 2009.

[2] N. Ailon and H. Rauhut. Fast and RIP-optimal transforms. *Discrete & Computational Geometry*, 52(4):780–798, 2014.

[3] W. U. Bajwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak. Toeplitz-structured compressed sensing matrices. In *2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, pages 294–298. IEEE, 2007.

[4] 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.

[5] J. Bourgain. An improved estimate in the restricted isometry problem. In *Geometric Aspects of Functional Analysis*, pages 65–70. Springer, 2014.

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

[7] 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.

[8] 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.

[9] V. Chandrasekaran, B. Recht, P. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6):805–849, 2012.

[10] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1998.

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

[12] S. Dirksen, G. Lecué, and H. Rauhut. On the gap between restricted isometry properties and sparse recovery conditions. *IEEE Trans. Inform. Theory*, to appear.

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

[14] 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.

[15] D. Dorsch and H. Rauhut. Refined analysis of sparse MIMO radar. *Journal of Fourier Analysis and Applications*, to appear. DOI:10.1007/s00041-016-9477-7.

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

[17] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing.* Applied and Numerical Harmonic Analysis. Birkhäuser, 2013.

[18] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak. Toeplitz compressed sensing matrices with applications to sparse channel estimation. *IEEE Transactions on Information Theory*, 56(11):5862–5875, 2010.

[19] I. Haviv and O. Regev. The restricted isometry property of subsampled fourier matrices. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '16, pages 288–297, 2016.

[20] M. Hügel, H. Rauhut, and T. Strohmer. Remote sensing via l1-minimization. *Found. Comput. Math.*, 14:115–150, 2014.

[21] D. James and H. Rauhut. Nonuniform sparse recovery with random convolutions. In *Proc. SampTA 2015, Washington*, 2015.

[22] M. Kabanava and H. Rauhut. Analysis ℓ_1-recovery with frames and Gaussian measurements. *Acta Appl. Math.*, 140(1):173–195, 2015.

[23] B. Klartag. $5n$ Minkowski symmetrizations suffice to arrive at an approximate Euclidean ball. *Ann. of Math. (2)*, 156(3):947–960, 2002.

[24] F. Krahmer, S. Mendelson, and H. Rauhut. Suprema of chaos processes and the restricted isometry property. *Comm. Pure Appl. Math.*, 67(11):1877–1904, 2014.

[25] F. Krahmer and R. Ward. New and improved johnson-lindenstrauss embeddings via the restricted isometry property. *SIAM Journal on Mathematical Analysis*, 43(3):1269–1281, 2011.

[26] G. Lecué and S. Mendelson. Sparse recovery under weak moment assumptions. *J. Eur. Math. Soc. (JEMS)*, to appear.

[27] M. Ledoux and M. Talagrand. *Probability in Banach Spaces.* Springer-Verlag, 1991.

[28] S. Mendelson, A. Pajor, and N. Tomczak Jaegermann. Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geom. Funct. Anal.*, 17(4):1248–1282, 2007.

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

[30] S. Mendelson and R. Vershynin. Remarks on the geometry of coordinate projections in $\mathbb{R}^n$. *Israel J. Math.*, 140:203–220, 2004.

[31] R. Oliveira. The lower tail of random quadratic forms, with applications to ordinary least squares and restricted eigenvalue properties. *ArXiv:1312.2903*, 2013.

[32] G. E. Pfander, H. Rauhut, and J. A. Tropp. The restricted isometry property for time-frequency structured random matrices. *Probab. Theory Relat. Fields*, (156):707–737, 2013.

[33] H. Rauhut. Circulant and Toeplitz matrices in compressed sensing. In *Proc. SPARS'09*, 2009.

[34] 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.

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

[36] 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.

[37] H. Rauhut and R. Ward. Interpolation via weighted ℓ_1 minimization. *Appl. Comput. Harm. Anal.*, 40(2):321–351, 2016.

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

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

[40] M. Rudelson and R. Vershynin. Small ball probabilities for linear images of high-dimensional distributions. *Internat. Math. Res. Notices*, 2015(19):9594–9617, 2015.

[41] M. Talagrand. *Upper and Lower Bounds For Stochastic Processes*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]*. Springer, Heidelberg, 2014.

[42] J. A. Tropp, M. B. Wakin, M. F. Duarte, D. Baron, and R. G. Baraniuk. Random filters for compressive sampling and reconstruction. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 3, pages III–III. IEEE, 2006.

[43] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Y. Eldar and G. Kutyniok, editors, *Compressed Sensing: Theory and Applications*, pages 210–268. Cambridge Univ Press, 2012.