IMPROVED BOUNDS FOR SPARSE RECOVERY FROM SUBSAMPLED RANDOM CONVOLUTIONS

BY SHAHAR MENDELSON[*,†,1], HOLGER RAUHUT[‡,1] AND RACHEL WARD[§,2]

Technion[*], Australian National University[†], RWTH Aachen University[‡] and University of Texas at Austin[§]

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 sub-Gaussian generator with independent entries, we improve previously known estimates: if the sparsity s is small enough, that is, $s \lesssim \sqrt{n/\log(n)}$, 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. Our results are shown via the so-called robust null space property which is weaker than the standard restricted isometry property. Our method of proof involves a novel combination of small ball estimates with chaining techniques which should be of independent interest.

1. Introduction. Compressive sensing [6, 14, 18] 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, sub-Gaussian random matrices [4, 9, 14, 15, 18, 29, 30, 40].

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 [5–7, 36, 38, 40], partial random circulant matrices (subsampled random convolutions) [25, 35–37, 39], time-frequency structured random matrices [25, 33, 34] and more [2, 21]. 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

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

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 [18], several tractable algorithms have been introduced as alternatives, most notably ℓ_1-minimization (basis pursuit) [6, 10, 14, 18] which produces a minimizer of

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

If $A \in \mathbb{R}^{m \times n}$ is a random draw of a Gaussian matrix, that is, 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

$$(1.1) \qquad m \geq Cs \ln(en/s)$$

for some absolute constant $C > 0$. This bound is optimal [14, 17, 18] 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, that is, for any $x \in \mathbb{R}^n$ (not necessarily s-sparse), the reconstruction $x^\sharp$ obtained from $y = Ax$, satisfies

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

see [17], Theorem 2.7, for details. Moreover, *exact* s-sparse recovery via ℓ_1-minimization necessarily requires (1.1); see [17], 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, [5, 7, 20, 36, 40].

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

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

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, that is, $(P_\Omega x)_j = x_j$ for $j \in \Omega$. A subsampled convolution is defined as

$$(1.3) \qquad\qquad Bx = P_\Omega(x * \xi)$$

and the corresponding matrix is a partial circulant matrix generated by ξ. Subsampled random convolutions find applications in radar and coded aperture imaging [3, 19, 39, 43], as well as in fast dimensionality reduction maps [26].

It was shown in [25] 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

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

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 sub-Gaussian 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, that is, a random vector with independent, mean zero, variance one, normally distributed coordinates. However, the proof we present holds for more general L-sub-Gaussian random vectors with independent coordinates and a more general class of random matrices (see Theorem 2.5).

THEOREM 1.1.   *Let $\xi \in \mathbb{R}^n$ be a 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

$$m \geq c_3 s \log(en/s) \qquad \text{if } s \leq c_2 \sqrt{\frac{n}{\log(n)}},$$

(1.5)

$$m \geq c_3 s \log(en/s) \alpha_s^2 \log(e\alpha_s) \qquad \text{if } c_2 \sqrt{\frac{n}{\log(n)}} \leq s \leq c_1 \frac{n}{\log^4(n)},$$

where $\alpha_s = \log(\frac{s^2}{n} \max\{\log(en/s), \log(s)\})$. 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

(1.6) $$\min \|z\|_1 \qquad \text{subject to } \|Bz - y\|_2 \leq \eta$$

satisfies

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

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

The constants $c_1, c_2, c_3, C, D > 0$ are absolute.

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, our estimates require more measurements; for example, if $s = n^\alpha$ for $1/2 < \alpha < 1$ then $c(L, \alpha) s \log^3 n \cdot \log \log n$ 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 [25], 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 [39], but the estimates there require $m \geq Cs \log^6(n)$. On the other hand, [39] 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.

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 [22, 35, 36] 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, for example, [18], Theorem 4.33, or [16].

Another improvement on known estimates is that our results hold for noisy measurements when the noise is bounded in ℓ_q for $q \geq 2$ 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 2.5 for details. (Note that a standard RIP-based argument appeared after the first submission of this manuscript in [12], Theorem A.1.) This allows us, for example, to explore quantized compressive sensing (see, e.g., [13]), when the quantization error has a natural ℓ_∞-bound.

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.

Apart from our main results themselves, we believe that our proof techniques are of independent interest. In fact, the crucial ingredient is a probabilistic lower bound on terms of the form $\inf_{v \in V_r} \|\Gamma_v \xi\|_2$, where Γ_v are matrices indexed by a set of unit-norm r-sparse vectors, and ξ is a sub-Gaussian random vector with independent coordinates. It is based on a new approach that can be generalized to other generating random vectors with heavier tails and less independence assumptions, for instance, to log-concave random vectors.

The article is structured as follows. Section 2 discusses preliminaries such as the null space property, sub-Gaussian 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 sub-Gaussian 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: a lower bound for $\inf_{v \in V_r} \|\Gamma_v \xi\|_2$. Section 5 provides an upper bound for one-sparse vectors, which is the final ingredient in the proof of the null space property; see also Theorem 2.3.

2. Preliminaries and the 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 re-

covery via ℓ_1-minimization. This version is stable when passing to approximately sparse vectors and robust when the measurements are noisy.

Given $v \in \mathbb{R}^n$ and $S \subset [n] = \{1, \ldots, n\}$, 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 $1 \le q \le \infty$, a matrix A satisfies the ℓ_q-robust null-space property of order s with constants $\nu \in (0, 1)$ and $\tau > 0$ if

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

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., [18], 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

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

satisfies

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

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

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., [13, 23]) that if

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

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

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

to be the set of s-sparse vectors in the unit sphere, then ([23], Lemma 3, see also [40])

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

allowing one to study $\operatorname{conv} V_s$ instead of $\mathcal{T}_{v,s}$, where $\operatorname{conv} S$ denotes the convex hull of a 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 [32], Lemma 5.1 (see also [27], Theorem B), but extends also to $q > 2$ as we outline below.

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

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

If $c(v) = v^2/(2v + 1)^2$ *and*

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

then

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

Consequently, A *also satisfies the* ℓ_q*-robust NSP, for* $q > 2$:

$$\inf_{x \in \mathcal{T}_{v,s}} \|Ax\|_q \geq \frac{1}{m^{1/2 - 1/q}} \inf_{x \in \mathcal{T}_{v,s}} \|Ax\|_2 \geq \frac{1}{m^{1/2 - 1/q} \sqrt{2}\tau}.$$

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

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

Applying [27], Lemma 2.6, it follows from (2.4) that for any $y \in \mathbb{R}^n$,

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

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

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

Therefore,

$$\|Ay\|_2^2 \ge \tau^{-2}\left(1 - \frac{\|y\|_1^2}{r-1}\left(\tau^2 \max_{1 \le j \le n}\|Ae_j\|_2^2 - 1\right)\right)$$

$$\ge \tau^{-2}\left(1 - \frac{c_1^2(v)s}{r-1}(\tau^2 M^2 - 1)\right) \ge \frac{1}{2\tau^2}$$

by our choice of s. The first inequality in (2.5) is due to Hölder. □

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

$$(2.6) \qquad \inf_{x \in V_r} \|P_\Omega Ax\|_2 \gtrsim \sqrt{m} \quad \text{and} \quad \max_{j \in [n]} \|P_\Omega Ae_j\|_2 \lesssim \sqrt{m}$$

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

2.2. *Sub-Gaussian random vectors.*   Just as in [25] we will focus on generators ξ that are isotropic, L-sub-Gaussian and have independent coordinates.

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

$$\left(\mathbb{E}|\langle \xi, x\rangle|^p\right)^{1/p} \le L\sqrt{p}\left(\mathbb{E}|\langle \xi, x\rangle|^2\right)^{1/2}.$$

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

$$\left(\mathbb{E}|\xi_i|^p\right)^{1/p} \le L\sqrt{p}\left(\mathbb{E}|\xi_i|^2\right)^{1/2} = L\sqrt{p}.$$

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-sub-Gaussian as well (see, e.g., [44]). Is it straightforward to show that if ξ is an L-sub-Gaussian random vector, then for every $x \in \mathbb{R}^n$ and $u \ge 1$,

$$\Pr\left(|\langle \xi, x\rangle| \ge uL\left(\mathbb{E}|\langle \xi, x\rangle|^2\right)^{1/2}\right) \le 2\exp(-cu^2),$$

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

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

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-sub-Gaussian 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 for all $q \geq 2$. 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_q-robust null-space property of order r with constants $\rho \in (0, 1)$ and $\frac{\tau}{m^{1/q}}$, where

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

and

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

The constant α_r satisfies

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

Theorem 2.5 is obtained from the analysis to follow by combining Theorem 5.1 and Corollary 4.2, and applying Theorem 2.3. Actually, our analysis holds 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

(2.7)
$$\inf_{v \in V_r} \|P_\Omega A v\|_2^2 = \inf_{v \in V_r} \sum_{i=1}^n \delta_i \langle Av, e_i \rangle^2.$$

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, that is, $\mathcal{F}_{j,k} = e^{-2\pi ijk/n}$ (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 $\beta \geq 1$ if for every $i, j \in [n]$, $|O_{i,j}| \leq \beta/\sqrt{n}$.

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

$$(2.8) \qquad \Gamma_x = \sqrt{n} U D_{Wx} O,$$

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 $\inf_{v \in V_r} \|\Gamma_v \xi\|_2 \geq c\sqrt{n}$ alone does not lead to a nontrivial lower bound on (2.7). 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,

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

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,

$$(2.9) \qquad \left| \left\{ i : |x_i| \geq \|x\|_2 \frac{\alpha}{\sqrt{n}} \right\} \right| \geq \theta n.$$

The notion of regularity in (2.9) 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 (2.9). 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 (2.9). 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 $1 \leq p \leq \infty$, 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-sub-Gaussian. In particular, ξ is an isotropic, L-sub-Gaussian 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

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

Indeed, for any $u \in \mathbb{R}$, $\|\xi_i - u\|_{L_2} \sim \max\{\|\xi_i\|_{L_2}, |u|\}$, and thus $\|\xi_i - u\|_{L_4} \leq c_3 L \|\xi_i - u\|_{L_2}$ 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., [18], 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 $\|\Gamma\|_{\mathrm{HS}}$ and $\|\Gamma\|_{2\to2}$ denote the Hilbert–Schmidt (Frobenius) and operator norms of a matrix Γ, respectively, and set

$$d_\Gamma = \left(\frac{\|\Gamma\|_{\mathrm{HS}}}{\|\Gamma\|_{2\to2}} \right)^2.$$

THEOREM 3.1 ([41]). *For $0 < p < 1$, there exists a constant $c = c(p) > 0$ for which the following holds. Let $X_1, \ldots, X_n$ be independent random variables that satisfy for some $t > 0$ and $0 < p < 1$,*

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

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

$$\Pr(\|\Gamma\mathbb{X}\|_2 \leq t\|\Gamma\|_{\mathrm{HS}}) \leq \left(\frac{1}{2}\right)^{cd_\Gamma}.$$

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 $\Gamma : \mathbb{R}^n \to \mathbb{R}^m$,*

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

It should be noted that a sub-Gaussian 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) sub-Gaussian random vector (see, e.g., [25], Theorem 2.3).

THEOREM 3.3. *There exists an absolute constant c for which the following holds. Let ξ be an isotropic, L-sub-Gaussian 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 $p \geq 1$,*

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

We will consider two families of norms associated with the nonincreasing rearrangement of the coordinates of a vector.

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

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

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,

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

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

LEMMA 3.5 ([24]). *Let $Z_1, \ldots, Z_n$ be mean-zero and L-sub-Gaussian random variables such that $\max_{i \in [n]} \|Z_i\|_{L_2} \leq M$. Then*

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

PROOF. Since the proof of this result is not provided in [24], we give it here for convenience. Since the Z_i are L-sub-Gaussian, there exist constants $c_0, c_1 > 0$ such that (see, e.g., [18], Proposition 7.23)

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

By Jensen's inequality and concavity of the logarithm,

$$\left[\mathbb{E}\left(\frac{1}{k}\sum_{i=1}^{k}(Z_i^*)^2\right)^{1/2}\right]^2 \le \mathbb{E}\frac{1}{k}\sum_{i=1}^{k}(Z_i^*)^2$$

$$\le c_0^{-1} M^2 L^2 \mathbb{E}\frac{1}{k}\sum_{i=1}^{k}\log(\exp(c_0(Z_i^*)^2/(L\|Z_i^*\|_{L_2})^2))$$

$$\le c_0^{-1} M^2 L^2 \log\left(\frac{1}{k}\sum_{i=1}^{k}\mathbb{E}\exp\left(c_0\frac{(Z_i^*)^2}{L^2\|Z_i^*\|_{L_2}^2}\right)\right)$$

$$\le c_0^{-1} M^2 L^2 \log\left(\frac{1}{k}\sum_{i=1}^{n}\mathbb{E}\exp\left(c_0\frac{Z_i^2}{L^2\|Z_i\|_{L_2}^2}\right)\right)$$

$$\le c_0^{-1} M^2 L^2 \log(c_1 n/k).$$

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 $p \ge 1$ and $k \in [n]$,

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

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

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

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-sub-Gaussian random vector. Then for every matrix Γ and any $p \ge 1$,*

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

PROOF. By Theorem 3.3, it suffices to estimate

$$\mathbb{E}\|G\|_{\Gamma,k} = \mathbb{E}\sup_{t\in\Gamma^*V_k}\langle G, t\rangle = \mathbb{E}\left(\sum_{i=1}^{k}(\langle G, \Gamma^*e_i\rangle^*)^2\right)^{1/2}$$

for a 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 follows immediately from Markov's inequality.

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

$$\Pr(|Z| \ge \alpha A) \le \alpha^{-p}.$$

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

$$\Gamma_v = \sqrt{n}U D_{Wv}O,$$

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

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

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

$$(3.4) \qquad \|\Gamma_v^*e_i\|_2 = \sqrt{n}\left(\sum_{\ell=1}^{n}\langle W_\ell, v\rangle^2 \cdot U_{i,\ell}^2\right)^{1/2} \le \beta\|v\|_2.$$

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*

$$(3.5) \qquad \Pr\left(\|\Gamma_v\xi\|_2 \le \kappa_1\sqrt{n}\|v\|_2\right) \le \left(\frac{1}{2}\right)^{\kappa_2 n/r},$$

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

$$(3.6) \qquad \|\Gamma_v\xi\|_{[k]} \le \kappa_3\left(\|v\|_2\sqrt{k\log(en/k)} + \sqrt{p}\cdot\sqrt{n}\|Wv\|_\infty\right).$$

REMARK 3.9. In what follows, (3.5) and (3.6) 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 (e.g., that ξ is stochastically dominated by a Gaussian vector) are highly restrictive, (3.5) and (3.6), 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., [18], 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* [42] 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.

1. For $r \in [n]$ set

$$(3.7) \qquad \rho = 10 \log_2 e \cdot \max\left\{1, \frac{\log(er)}{\log(en/r)}\right\}.$$

2. Using the notation introduced earlier, put

$$(3.8) \qquad \kappa_4 = \min\left\{\frac{\kappa_2}{2}, \frac{\kappa_1^2}{256\kappa_3^2 L^2 \beta^2}\right\},$$

and observe that κ_4 depends only on L and β. Moreover, without loss of generality, $\kappa_4 \leq 1$, and $\kappa_3, L, \beta \geq 1$.

3. Set s_0 and s_1 to satisfy

$$(3.9) \qquad 2^{s_0} = \frac{\kappa_4 n}{r} \quad \text{and} \quad 2^{s_1} = \rho r \log(en/r)$$

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

4. Finally, let

$$(3.10) \qquad \alpha_r = \max\left\{1, \log\left(\frac{\rho r \log(en/r)}{\kappa_4(n/r)}\right)\right\} = \max\{1, \log(2^{s_1 - s_0})\}.$$

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

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

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,

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

and

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

Let us again emphasize 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 3.11. *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

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

To estimate ε_s, observe that by a standard volumetric estimate (see, e.g., [18], 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

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

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 $s \ge s_1$,

(3.11) $$\qquad\qquad\qquad\qquad 2^s/r \ge 2(s/2).$$

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

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

Indeed, if $r \le n/r$, that is, 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

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

so that (3.12) 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. $\qquad\square$

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.   *Recall the definitions of ρ, s_0, s_1, and α_r from (3.7), (3.9) and (3.10), respectively. There exist constants $c_0, \ldots, c_3$ 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

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

Moreover, for each $t \in T_1$:

1. *$\|t\|_2 \geq \kappa_1 \sqrt{n}$, and*
2. *$\|t_I\|_2 \leq \frac{\kappa_1 \sqrt{n}}{4}$ on the set $I = I(t) \subset [n]$ of largest-magnitude coordinates of t of cardinality $|I| \geq c_2 \frac{n}{\alpha_r^2 \log(e\alpha_r)}$.*

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: for each $t \in T_1$ consider the set $J = I(t)^c$ and observe that it follows from $\|t\|_2 \geq \kappa_1 \sqrt{n}$ and $\|t_I\|_2 \leq \frac{\kappa_1 \sqrt{n}}{4}$ that $\|t_J\|_2 \geq \frac{15\kappa_1}{16} \sqrt{n}$ and

$$(4.1) \qquad \|t_J\|_\infty \leq |t^*_{|I(t)|}| \leq \left(\frac{1}{|I(t)|} \sum_{j \in I(t)} |t^*_j|^2 \right)^{1/2} \leq \frac{\kappa_1}{4\sqrt{c_2}} \alpha_r \sqrt{\log(e\alpha_r)}.$$

This implies that at least $\theta = c/(\alpha_r^2 \log(e\alpha_r))$ coordinates of t_J are larger than a constant. When $s_0 \geq s_1$ (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 a subset $I \subset [n]$ of cardinality $|I| \geq c_2 n$ of coordinates which 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 (2.9).

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.

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*

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

with ρ defined in (3.7) *and set* $(\delta_i)_{i=1}$ *to be independent selectors with mean* δ. *Then with probability at least* $1 - 2^{-c_1 \min\{2^{s_0}, 2^{s_1}\}}$,

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

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

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

(4.2)

$$\geq \inf_{t \in T_1} \left(\sum_{i=1}^{n} \delta_i t_i^2 \right)^{1/2} - \frac{c_1}{n^{3/2}}$$

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 side being larger than $2c_1/n^3$, where c_2 is a suitable constant.

Fix $t \in T_1$, let $J = J(t) = I^c(t)$ be the *complement* of the set $I(t)$ identified by Theorem 4.1 and put $x = P_J t$. Then as argued in (4.1) and the line above

(4.3) $$\|x\|_2 \geq \frac{15c_1}{16} \sqrt{n} \quad \text{and} \quad \|x\|_\infty \leq c_3 \alpha_r \sqrt{\log(e\alpha_r)}.$$

By Bernstein's inequality,

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

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

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

By a union bound, with probability at least

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

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

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

The claim follows by setting

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

REMARK 4.3. The following alternative argument starting from (4.3) leads to a lower bound in ℓ_q for any $1 \leq q < \infty$—at least in the small sparsity regime $s_0 \leq s_1$, that is, $s \leq c\sqrt{n/\log(n)}$. For fixed $t \in T_1$, choose $I = I(t)$ as the θn largest absolute coefficients of t where θ is defined in (4.5) below (with suitable constants). We know from the theorem that $\|t\|_2 \geq \kappa_1 \sqrt{n}$ and $\|t\|_{[\theta n]} \leq \frac{\kappa}{4}\sqrt{n}$. This implies that

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

$$\geq \left(\frac{1}{n}(\kappa_1^2 n - \kappa_1^2 n/16)\right)^{1/2} = \frac{\sqrt{15}\kappa_1}{4} = c_6.$$

For $\Omega = \{i : \delta_i = 1\}$, let $K = K(t) = I(t) \cap \Omega$. Then, by Chernov's inequality the cardinality of K satisfies $|K| = \sum_{j \in I(t)} \delta_i \geq \frac{1}{2}\delta|I| = \frac{1}{2}\delta\theta n$ probability at least $1 - e^{-\delta\theta n/8}$. On this event,

$$\|P_\Omega t\|_q \geq \left(\sum_{j \in K} |t_j|^q\right)^{1/q} \geq c_6(\theta\delta n/2)^{1/q}$$

By the union bound, this holds for all $t \in T_1$ with probability at least

$$1 - 2|T_1|\exp(-\delta\theta n) \geq 1 - 2 \cdot 2^{2^{s_1}} \exp(-\delta\theta n/8)$$

$$\geq 1 - 2\exp\left(-\delta\theta n/8 + \ln(2)\rho r \log(en/r)\right).$$

Therefore, in the small sparsity regime $s_0 \leq s_1$, that is, $\theta = c_3$, a combination with a similar estimate as in (4.2) gives

$$\inf_{v \in V_r} \sum_{i=1}^{n} \delta_i \langle \Gamma_v \xi, e_i \rangle^q \geq c_8^q \delta n$$

with probability at least $1 - 2\exp(-c_9 \delta n)$ provided that

$$\delta n \geq c_{10} r \log(en/r).$$

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 $v \in V_{r,s_1}$,

(4.4) $$\|\Gamma_v \xi\|_2 \geq c\sqrt{n} \quad \text{and} \quad \|\Gamma_v \xi\|_{[\theta n]} \leq (c/4)\sqrt{n}$$

for well chosen c and

(4.5) $$\theta = \begin{cases} c_1 & \text{for } r \leq c_2 \sqrt{\dfrac{\kappa_4 n}{\log(c_2 n/\kappa_4)}}, \\[4mm] \dfrac{c_3}{\alpha_r^2 \log(e\alpha_r)} & \text{for } c_2 \sqrt{\dfrac{\kappa_4 n}{\log(c_2 n/\kappa_4)}} \leq r \leq \dfrac{c_4 n}{\log^4(n)}, \end{cases}$$

where all constants c, c_1, c_2, c_3, c_4 only depend on L and β and κ_4 is defined in (3.8); see (4.10) and the following remarks as well as Lemma 4.11. Moreover, here and in the following, we assume for simplicity that θn is an integer. (The general case may need slightly different constants.)

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 4.4. *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 $s \geq s_1$,*

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

In particular, if V_{r,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 (3.6). Indeed, if we set $k = n$ and $p = 2^{s+3}$, then by Corollary 3.8, 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,

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

Therefore,

(4.7) $$\|\Gamma_{\Delta_s v} \xi\|_2 \leq 4\kappa_3(\sqrt{n} + \beta\sqrt{2r}2^{s/2})\|\Delta_s v\|_2.$$

There are at most $2^{2^s} \cdot 2^{2^{s+1}} \leq 2^{2^{s+2}}$ vectors of the form $\Delta_s v$, and thus, by a union bound, with probability at least $1 - 2^{-2^{s+2}}$, (4.7) holds for every $v \in V_r$ for that

choice of s. Summing the probabilities for $s \geq s_1$, one has that with probability at least $1 - 2^{-2^{s_1+1}}$, for every $v \in V_r$,

$$\|\Gamma_{v-\pi_{s_1}v}\xi\|_2 = \left\|\sum_{s \geq s_1} \Gamma_{\Delta_s v}\xi\right\|_2 \leq \sum_{s \geq s_1} \|\Gamma_{\Delta_s v}\xi\|_2$$

$$\leq 4\sqrt{2}\kappa_3\beta \sum_{s \geq s_1}(\sqrt{n} + \sqrt{r}2^{s/2})\|\Delta_s v\|_2 \leq \frac{c(L,\beta)}{n^{3/2}},$$

where the last inequality is just Lemma 3.11. $\square$

Lemma 4.4 shows that if we set $T_2 = \{\Gamma_{v-\pi_{s_1}v}\xi : v \in V_r\}$, then with probability at least $1 - 2^{-2^{s_1}}$, $\sup_{t \in T_2} \|t\|_2 = \sup_{v \in V_r} \|\Gamma_{v-\pi_{s_1}v}\xi\|_2 \leq c(L,\beta)/n^{3/2}$, 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 $s_0 \geq s_1$ and the high-sparsity one, in which the reverse inequality holds. The analysis is based on Corollary 3.8 in both cases.

The low-sparsity case. Assume that $2^{s_0} \geq 2^{s_1}$, that is, $\frac{\kappa_4 n}{r} \geq \rho r \log(en/r)$ and in particular, by the choice of κ_4 in (3.8),

$$(4.8) \qquad \frac{\kappa_2}{2} \cdot \frac{n}{r} \geq \rho r \log\left(\frac{en}{r}\right).$$

Fix $v \in V_{r,s_1}$. It follows from (3.5) that with probability at least $1 - 2^{-\kappa_2 n/r}$,

$$(4.9) \qquad \|\Gamma_v \xi\|_2 \geq \kappa_1\sqrt{n}.$$

Let $0 < \theta < 1$ to be named later. Observe that by (3.6) for $k = \theta n$ and $p = 2\rho r \log(en/r) = 2^{s_1+1}$, with probability at least $1 - e^{-2^{s_1+1}}$,

$$(4.10) \quad \begin{aligned} \|\Gamma_v \xi\|_{[k]} &\leq 2\kappa_3\left(\|v\|_2\sqrt{k\log(en/k)} + \sqrt{\rho r \log(en/r)}\sqrt{n}\|Wv\|_\infty\right) \\ &\leq 2\kappa_3\left(\sqrt{n}\sqrt{\theta\log(e/\theta)} + \beta\sqrt{\rho r \log(en/r)} \cdot \sqrt{r}\right), \end{aligned}$$

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

Recall that $|V_{r,s_1}| \leq 2^{2^{s_1}}$, and thus, by (4.8) and the union bound, with probability at least $1 - 2^{-2^{s_1}}$, (4.9) and (4.10) hold for every $v \in V_{r,s_1}$. All that remains is to ensure that

$$2\kappa_3\beta\sqrt{\theta\log(e/\theta)}\sqrt{n} \leq \frac{\kappa_1}{8}\sqrt{n} \quad \text{and} \quad 2\kappa_3\beta\sqrt{r} \cdot \sqrt{\rho r \log(en/r)} \leq \frac{\kappa_1}{8}\sqrt{n}.$$

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 (3.8)

$$\rho r \log(en/r) = 2^{s_1} \leq 2^{s_0} \leq \frac{\kappa_1^2}{256\kappa_3^2 L^2\beta^2} \cdot \frac{n}{r}.$$

Therefore,

$$2\kappa_3\beta\sqrt{r}\cdot\sqrt{\rho r \log(en/r)} \leq 2\kappa_3\beta\sqrt{r}\cdot\frac{\kappa_1}{16\kappa_3 L\beta}\cdot\sqrt{\frac{n}{r}} \leq \frac{\kappa_1}{8}\sqrt{n},$$

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, $\kappa_4 n/r \leq \rho r \log(en/r)$. Then there is a substantial gap between the individual probability estimate (3.5) and the cardinality of V_{r,s_1}, so that a simple union bound does not give a nontrivial 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 4.5. *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 $1 - 2\exp(-c_2 n/r)$,*

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

For the proof of Theorem 4.5, 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 (3.5) and the union bound, to the much larger set V_{r,s_1}. Thus, it suffices to show that with high probability,

$$(4.11) \qquad \sup_{v \in V_{r,s_1}} \left| \|\Gamma_v \xi\|_2^2 - \|\Gamma_{\pi_{s_0}v}\|_2^2 \right| \leq \frac{\kappa_1^2}{4}n,$$

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

To address (4.11), we proceed along the lines of [25] 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 4.6. *Assume $\mathcal{A} = \mathcal{A}_{s_1}$ to be finite. Let ξ be an isotropic, L-sub-Gaussian 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 4.6 is contained in the proof of Lemma 3.3 in [25], 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 4.7. *There exist 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 $1 - 2^{-c_1 n/r}$,*

$$(4.12) \qquad \sup_{A \in \mathcal{A}} \left| \|A\xi\|_2^2 - \|\pi_{s_0} A\xi\|_2^2 \right| \leq c_2 \gamma_{s_0,s_1}(\mathcal{A})\left(\gamma_{s_0,s_1}(\mathcal{A}) + \sqrt{n}\right).$$

Before proving Theorem 4.7, let us recall a standard fact that can be established using tail integration.

LEMMA 4.8. *Let Z be a nonnegative random variable and assume that*

$$\Pr(Z \geq A_1 + u A_2) \leq 2 \exp(-u^2/2) \qquad \text{for every } u \geq A_3.$$

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

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

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

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

Denoting by ξ' an independent copy of ξ, applying a standard decoupling argument (see, e.g., [25], Theorem 2.4, or [11]) and Lemma 4.6,

$$\left(\mathbb{E} \sup_{A \in \mathcal{A}} |F_A|^p \right)^{1/p} = \left\| \sup_{A \in \mathcal{A}} (*)_A \right\|_{L_p}$$

$$\leq 4 \left\| \sup_{A \in \mathcal{A}} \left| \sum_{j,k} \xi_j \xi_k' \left(\langle A^j, A^k \rangle - \langle (\pi_{s_0} A)^j, (\pi_{s_0} A)^k \rangle \right) \right| \right\|_{L_p}$$

$$= 4 \left\| \sup_{A \in \mathcal{A}} \left| \langle A\xi, A\xi' \rangle - \langle \pi_{s_0} A\xi, \pi_{s_0} A\xi' \rangle \right| \right\|_{L_p}$$

$$\leq c_1 L \gamma_{s_0,s_1}(\mathcal{A}) \cdot \left\| N_{\mathcal{A}}(\xi) \right\|_{L_p}.$$

Moreover,

$$\left\|(N_{\mathcal{A}}(\xi))^2\right\|_{L_p} \leq \left\|\sup_{A\in\mathcal{A}}\left|\|A\xi\|_2^2 - \|\pi_{s_0}A\xi\|_2^2\right|\right\|_{L_p} + \left\|\sup_{A\in\mathcal{A}}\|\pi_{s_0}A\xi\|_2^2\right\|_{L_p}$$

$$\leq c_1 L\gamma_{s_0,s_1}(\mathcal{A})\cdot\left\|N_{\mathcal{A}}(\xi)\right\|_{L_p} + \left\|\sup_{A\in\mathcal{A}}\|\pi_{s_0}A\xi\|_2^2\right\|_{L_p}.$$

By (3.6) for $k = n$ and $p = u^2 \geq 2^{s_0+3}$, it follows that with probability at least $1 - 2\exp(-u^2/2)$,

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

$$\leq c_2(L,\beta)(\sqrt{n} + u\sqrt{r}),$$

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 4.8 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 $p \geq 1$,

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

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,

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

implying that

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

and

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

The claim now follows from Lemma 3.7 with $\alpha = 2$ and the definition of $p = \kappa_4 n/r$. $\square$

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 4.9. *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,

$$(4.13) \qquad \log N(V_{r,s_1}, tB) \leq \log N(\sqrt{r}B_1^n, tB) \leq C\frac{\beta^2 r}{t^2}\log^2(nt^2/r).$$

COROLLARY 4.10. *There is an admissible sequence of* V_{r,s_1}, *for which, for every* $v \in V_{r,s_1}$,

$$\sqrt{n}\|W\Delta_s v\|_\infty \leq c\beta 2^{-s/2}\sqrt{r}\log(en/2^s) \qquad \text{for } s \leq s_1,$$

for an absolute constant c. Therefore,

$$(4.14) \qquad \sup_{v \in V_{r,s_1}} \sum_{s=s_0}^{s_1-1} 2^{s/2} \cdot \sqrt{n}\|W\Delta_s v\|_\infty \leq c_1\beta\sqrt{r}\alpha_r\log(er/\kappa_4).$$

Corollary 4.10 follows from (4.13), a straightforward computation and the definition of s_0 and s_1. We omit the details.

PROOF OF THEOREM 4.5. Combining the individual small ball estimate in (3.5), Theorem 4.7 and Corollary 4.10, 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 $r \leq c_3(L,\beta)n/\log^4 n$. $\square$

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

$$(4.15) \qquad \sup_{v \in V_{r,s_1}} \|\Gamma_v\xi\|_{[k]} = \sup_{v \in V_{r,s_1}} \left(\sum_{i \leq k}(\langle\Gamma_v\xi, e_i\rangle^*)^2\right)^{1/2}.$$

Specifically, the next result bounds the largest possible k for which (4.15) is smaller than $(\kappa_1/4)\sqrt{n}$.

LEMMA 4.11. *Assume that $2^{s_0} \leq 2^{s_1}$, $r \leq c_1 n/\log^4(n)$ and*

$$(4.16) \qquad k \leq c_2 \frac{n}{\alpha_r^2 \log(e\alpha_r)}.$$

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

$$(4.17) \qquad \sup_{v \in V_{r,s_1}} \|\Gamma_v \xi\|_{[k]} \leq \frac{\kappa_1}{4}\sqrt{n}.$$

The constants $c_1, c_2, c_3 > 0$ only depend on L and β.

REMARK 4.12. One should note that an upper bound on the supremum in (4.17) cannot be obtained via an individual estimate and the union bound. The sub-Gaussian property of ξ is enough to ensure that for a fixed v,

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

However, because one only knows that for $v \in V_r$, $\sqrt{n}\|Wv\|_\infty \leq \beta\sqrt{r}\|v\|_2$, 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

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

Let $(V_{k,s})_{s \geq s_0}$ 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 $v \in V_{r,s_1}$,

$$(4.18) \qquad \Gamma_v^* x = \underbrace{\Gamma_v^*(x - \pi_{s_1} x)}_{=:H_1} + \underbrace{\left(\Gamma_v^* \pi_{s_1} x - \Gamma_{\pi_{s_0}v}^* \pi_{s_0} x\right)}_{=:H_2} + \underbrace{\Gamma_{\pi_{s_0}v}^* \pi_{s_0} x}_{=:H_3}.$$

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

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

and for $s \geq s_1$, $|\{\Delta_s x : x \in V_k\}| \geq 2^{2^{s_1}} = |V_{r,s_1}|$. Hence, it is possible to treat H_1 for each $v \in V_{r,s_1}$ separately and apply a simple union bound. In contrast, the situation for H_2 is "more balanced," and requires a different argument.

To deal with H_1 in (4.18), fix $v \in V_{r,s_1}$ and apply Lemma 3.5 for a standard Gaussian vector G in $\mathbb{R}^n$ together with (3.4) to obtain

$$\mathbb{E} \sup_{t \in \Gamma_v^* V_k} \langle G, t \rangle = \mathbb{E} \|\Gamma_v G\|_{[k]} = \mathbb{E} \left(\sum_{i=1}^{k} (\langle G, \Gamma_v^* e_i \rangle^*)^2 \right)^{1/2}$$

$$\leq c \sqrt{k \log(en/k)} \max_{i \in [n]} \|\Gamma_v^* e_i\|_2 \leq c\beta \sqrt{k \log(en/k)}.$$

By the majorizing measures theorem [42], there exists an admissible sequence of V_k for which

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

Let us consider a part of that admissible sequence, namely, $(V_{k,s})_{s \geq s_1}$. As ξ is an isotropic, L-sub-Gaussian 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 $x \in V_k$,

$$|\langle \xi, \Gamma_v^*(x - \pi_{s_1} x) \rangle| \leq \sum_{s \geq s_1} |\langle \xi, \Gamma_v^* \Delta_s x \rangle| \leq cL \sum_{s \geq s_1} 2^{s/2} \|\Gamma_v^* \Delta_s x\|_2$$

(4.19)

$$\leq c\beta L \sqrt{k \log(en/k)} \leq \frac{\kappa_1}{16} \sqrt{n}$$

for a suitable choice of c_2 in (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 $1 - 2^{-2^{s_1+1}}$, (4.19) holds for every $v \in V_{r,s_1}$.

Next, let us turn to H_2 in (4.18). 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 v_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

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

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 $x \in V_k$,

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

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

$$(4.21) \qquad \|\Gamma_{v'}^*(x - x')\|_2 = \|x - x'\|_{v'} \le \varepsilon_s(v').$$

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

The required estimate on H_2 follows from (4.14) and once we control $\varepsilon_s(z)$ with respect to each one of the $\| \|_z$ norms. Indeed, using the above notation

$$\left|\langle \xi, \Gamma_v^* x\rangle - \langle \xi, \Gamma_{\pi_s v}^* \tilde{\pi}_s x\rangle\right| \le \left|\langle \xi, \Gamma_{v - \pi_s v}^* x\rangle\right| + \left|\langle \xi, \Gamma_{\pi_s v}^*(x - \tilde{\pi}_s x)\rangle\right|,$$

where, for $v \in V_{r,s+1}$, $\pi_s v \in V_{r,s}$ is closest to v with respect to $\sqrt{n}\|W(\cdot)\|_\infty$, $x \in T_{s+1}(v)$ and $\tilde{\pi}_s x \in T_s(\pi_s v)$ is closest to x with respect to $\| \|_{\pi_s v}$. Since ξ is isotropic and L-sub-Gaussian, we have with probability at least $1 - 2^{2^{s+2}}$, for every $v \in V_{r,s+1}$ and $x \in T_{s+1}(v)$,

$$(4.22) \qquad \begin{aligned} &\left|\langle \xi, \Gamma_v^* x\rangle - \langle \xi, \Gamma_{\pi_s v}^* \tilde{\pi}_s x\rangle\right| \\ &\le cL\left(2^{s/2}\|\Gamma_{v - \pi_s v}^* x\|_2 + 2^{s/2}\|\Gamma_{\pi_s v}^*(x - \tilde{\pi}_s x)\|_2\right) \\ &= cL\left(2^{s/2} \cdot \sqrt{n}\|W(v - \pi_s v)\|_\infty + 2^{s/2}\|x - \tilde{\pi}_s x\|_{\pi_s v}\right) \\ &\le cL\left(2^{s/2}v_s + 2^{s/2}\sup_{z \in V_r}\varepsilon_s(z)\right). \end{aligned}$$

Iterating (4.22), 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 $x \in V_k$,

$$(4.23) \quad \left|\langle \xi, \Gamma_v^* \pi_{s_1} x - \Gamma_{\pi_{s_0} v}^* \tilde{\pi}_{s_0} x\rangle\right| \le c_2(L,\beta)\left(\sum_{s=s_0}^{s_1-1} 2^{s/2}v_s + \sum_{s=s_0}^{s_1-1} 2^{s/2}\sup_{z \in V_r}\varepsilon_s(z)\right).$$

The first sum in (4.23) has been estimated earlier, in (4.14). In particular,

$$c_2\sum_{s=s_0}^{s_1-1} 2^{s/2}v_s \le c_3\sqrt{r}\alpha_r \cdot \log(c_3 r) \le \frac{\kappa_1}{32}\sqrt{n}$$

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

In order to bound the second sum in (4.23), we require another covering estimate.

THEOREM 4.13.   *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 4.13 is an outcome of Sudakov's inequality. Fix $z \in S^{n-1}$ and define a linear operator $S : \mathbb{R}^n \to \mathbb{R}^n$ by $Se_i = U^i$. Observe that for $i \in [n]$ and $t \in \mathbb{R}^n$

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

Therefore, $\| \sqrt{n} D_{W_z} S^* t \|_2 = \| t \|_z$ and

$$\log N(V_k, \| \ \|_z, \varepsilon) = \log N(\sqrt{n} D_{W_z} S^* V_k, \varepsilon B_2^n).$$

Set $T = \sqrt{n} D_{W_z} S^* V_k$ and let G be a standard Gaussian vector in $\mathbb{R}^n$. By Sudakov's inequality (see, e.g., [28]), there is an absolute constant c for which

(4.24)
$$c \varepsilon^2 \log N(T, \varepsilon B_2^n) \le \mathbb{E} \sup_{t \in T} \langle G, t \rangle,$$

and

$$\mathbb{E} \sup_{t \in T} \langle G, t \rangle = \mathbb{E} \sup_{v \in V_k} \langle G, \sqrt{n} D_{W_z} S^* v \rangle$$

$$= \sqrt{n} \mathbb{E} \sup_{v \in V_k} \sum_{i=1}^{n} g_i \langle W_i, z \rangle \langle U^i, v \rangle \equiv \mathbb{E} \left(\sum_{j \le k} (Z_j^*)^2 \right)^{1/2},$$

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, Lemma 3.5 yields $\mathbb{E}(\sum_{j=1}^{k} (Z_j^*)^2)^{1/2} \le c\beta \sqrt{k \log(en/k)}$, which completes the proof by (4.24). $\square$

Invoking Theorem 4.13, it follows that

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

and due to the definition of α_r in (3.10) the second sum in (4.23) is bounded by

$$\sum_{s=s_0}^{s_1} 2^{s/2} \sup_{z \in V_r} \epsilon_s(z) \le c(L, \beta)(s_1 - s_0) \sqrt{k \log(en/k)}$$

$$\le c_1(L, \beta) \alpha_r \sqrt{k \log(en/k)} \le \frac{\kappa_1}{32} \sqrt{n}$$

provided that

(4.25)
$$k \le c_2(L, \beta) \frac{n}{\alpha_r^2 \log(e\alpha_r)}.$$

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 (4.18). Indeed,

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

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-sub-Gaussian, one has that with probability at least $1 - 2\exp(-u^2/2)$,

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

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

$$\sup_{v\in V_{r,s_1}, x\in V_k}|\langle \xi, \Gamma^*_{\pi_{s_0}v}\pi_{s_0}x\rangle| \le 2^{(s_0+2)/2}L\beta\sqrt{r} \le \frac{2\kappa_1}{16\kappa_3 L\beta}\sqrt{\frac{n}{r}} \cdot L\beta\sqrt{r} \le \frac{\kappa_1}{8}\sqrt{n},$$

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

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

with probability at least $1 - 2^{-c_1 n/r} - 2e^{-c_2 2^{s_0}} - 2^{-2^{s_1+1}} \ge 1 - 2^{-c'2^{s_0}}$ for a suitable constant c', under condition (4.25) on k. This completes the proof of Lemma 4.11 after relabeling constants. $\square$

5. The upper bound on one-sparse vectors.

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

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

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 $\delta \ge c_0 \frac{\log n}{n}$, then with probability at least $1 - 2\exp(-c_1\delta n)$,*

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

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-sub-Gaussian 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*

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

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

$$(5.1) \qquad x = \sum_{j \in I} z_j e_j \quad \text{and} \quad y = \sum_{j \in I^c} z_j e_j.$$

Observe that $\|x\|_2 = \|\Gamma_{e_i}\xi\|_{[m]}$. By (3.6) for $p = m\log(en/m)$, and noting that $\|We_i\|_\infty \le \beta/\sqrt{n}$, one has

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

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

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

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. $\quad\square$

The coordinate structure of $\Gamma_{e_i}\xi$ comes into play thanks to a fact from [31].

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*

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

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 $1 - 2\exp(-c_0\delta n)$,

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

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

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

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 $1 - 2n\exp(-c_6\delta n) \ge 1 - 2\exp(-c_9\delta n)$,

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

REFERENCES

[1] AILON, N. and LIBERTY, E. (2009). Fast dimension reduction using Rademacher series on dual BCH codes. *Discrete Comput. Geom.* **42** 615–630. MR2556458

[2] AILON, N. and RAUHUT, H. (2014). Fast and RIP-optimal transforms. *Discrete Comput. Geom.* **52** 780–798. MR3279549

[3] BAJWA, W. U., HAUPT, J. D., RAZ, G. M., WRIGHT, S. J. and NOWAK, R. D. (2007). Toeplitz-structured compressed sensing matrices. In 2007 *IEEE/SP* 14*th Workshop on Statistical Signal Processing* 294–298. IEEE, New York.

[4] BARANIUK, R., DAVENPORT, M., DEVORE, R. and WAKIN, M. (2008). A simple proof of the restricted isometry property for random matrices. *Constr. Approx.* **28** 253–263. MR2453366

[5] BOURGAIN, J. (2014). An improved estimate in the restricted isometry problem. In *Geometric Aspects of Functional Analysis. Lecture Notes in Math.* **2116** 65–70. Springer, Cham. MR3364679

[6] CANDÈS, E. J., ROMBERG, J. and TAO, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory* **52** 489–509. MR2236170

[7] CANDES, E. J. and TAO, T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inform. Theory* **52** 5406–5425. MR2300700

[8] CARL, B. (1985). Inequalities of Bernstein–Jackson-type and the degree of compactness of operators in Banach spaces. *Ann. Inst. Fourier* (*Grenoble*) **35** 79–118. MR0810669

[9] CHANDRASEKARAN, V., RECHT, B., PARRILO, P. A. and WILLSKY, A. S. (2012). The convex geometry of linear inverse problems. *Found. Comput. Math.* **12** 805–849. MR2989474

[10] CHEN, S. S., DONOHO, D. L. and SAUNDERS, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.* **20** 33–61. MR1639094

[11] DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling. From Dependence to Independence, Randomly Stopped Processes. U-Statistics and Processes. Martingales and Beyond.* Springer, New York. MR1666908

[12] DIRKSEN, S., JUNG, H. and RAUHUT, H. (2017). One-bit compressed sensing with partial Gaussian circulant matrices. Preprint. Available at arXiv:1710.03287.

[13] DIRKSEN, S., LECUÉ, G. and RAUHUT, H. (2016). On the gap between restricted isometry properties and sparse recovery conditions. *IEEE Trans. Inform. Theory*.

[14] DONOHO, D. L. (2006). Compressed sensing. *IEEE Trans. Inform. Theory* **52** 1289–1306. MR2241189

[15] DONOHO, D. L. and TANNER, J. (2009). Counting faces of randomly projected polytopes when the projection radically lowers dimension. *J. Amer. Math. Soc.* **22** 1–53. MR2449053

[16] DORSCH, D. and RAUHUT, H. (2017). Refined analysis of sparse MIMO radar. *J. Fourier Anal. Appl.* **23** 485–529. MR3649470

[17] FOUCART, S., PAJOR, A., RAUHUT, H. and ULLRICH, T. (2010). The Gelfand widths of ℓ_p-balls for $0 < p \le 1$. *J. Complexity* **26** 629–640. MR2735423

[18] FOUCART, S. and RAUHUT, H. (2013). *A Mathematical Introduction to Compressive Sensing.* Birkhäuser, Basel. MR3100033

[19] HAUPT, J., BAJWA, W. U., RAZ, G. and NOWAK, R. (2010). Toeplitz compressed sensing matrices with applications to sparse channel estimation. *IEEE Trans. Inform. Theory* **56** 5862–5875. MR2808938

[20] HAVIV, I. and REGEV, O. (2016). The restricted isometry property of subsampled Fourier matrices. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms* 288–297. ACM, New York. MR3478398

[21] HÜGEL, M., RAUHUT, H. and STROHMER, T. (2014). Remote sensing via ℓ_1-minimization. *Found. Comput. Math.* **14** 115–150. MR3160709

[22] JAMES, D. and RAUHUT, H. (2015). Nonuniform sparse recovery with random convolutions. In *Proc. SampTA* 2015, Washington.

[23] KABANAVA, M. and RAUHUT, H. (2015). Analysis ℓ_1-recovery with frames and Gaussian measurements. *Acta Appl. Math.* **140** 173–195. MR3424652

[24] KLARTAG, B. (2002). 5n Minkowski symmetrizations suffice to arrive at an approximate Euclidean ball. *Ann. of Math.* (2) **156** 947–960. MR1954241

[25] KRAHMER, F., MENDELSON, S. and RAUHUT, H. (2014). Suprema of chaos processes and the restricted isometry property. *Comm. Pure Appl. Math.* **67** 1877–1904. MR3263672

[26] KRAHMER, F. and WARD, R. (2011). New and improved Johnson–Lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.* **43** 1269–1281. MR2821584

[27] LECUÉ, G. and MENDELSON, S. (2017). Sparse recovery under weak moment assumptions. *J. Eur. Math. Soc.* (*JEMS*) **19** 881–904. MR3612870

[28] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes. Ergebnisse der Mathematik und Ihrer Grenzgebiete* (3) [*Results in Mathematics and Related Areas* (3)] **23**. Springer, Berlin. MR1102015

[29] MENDELSON, S., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2007). Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geom. Funct. Anal.* **17** 1248–1282. MR2373017

[30] MENDELSON, S., PAJOR, A. and TOMCZAK-JAEGERMANN, N. (2008). Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constr. Approx.* **28** 277–289. MR2453368

[31] MENDELSON, S. and VERSHYNIN, R. (2004). Remarks on the geometry of coordinate projections in $\mathbb{R}^n$. *Israel J. Math.* **140** 203–220. MR2054844

[32] OLIVEIRA, R. I. (2013). The lower tail of random quadratic forms with applications to ordinary least squares and restricted eigenvalue properties. Preprint. Available at arXiv:1312.2903.

[33] PFANDER, G. E. and RAUHUT, H. (2010). Sparsity in time-frequency representations. *J. Fourier Anal. Appl.* **16** 233–260. MR2600959

[34] PFANDER, G. E., RAUHUT, H. and TROPP, J. A. (2013). The restricted isometry property for time-frequency structured random matrices. *Probab. Theory Related Fields* **156** 707–737. MR3078284

[35] RAUHUT, H. (2009). Circulant and Toeplitz matrices in compressed sensing. In *Proc. SPARS '09*.

[36] RAUHUT, H. (2010). Compressive sensing and structured random matrices. In *Theoretical Foundations and Numerical Methods for Sparse Recovery* (M. Fornasier, ed.). *Radon Ser. Comput. Appl. Math.* **9** 1–92. de Gruyter, Berlin. MR2731597

[37] RAUHUT, H., ROMBERG, J. and TROPP, J. A. (2012). Restricted isometries for partial random circulant matrices. *Appl. Comput. Harmon. Anal.* **32** 242–254. MR2880281

[38] RAUHUT, H. and WARD, R. (2016). Interpolation via weighted ℓ_1 minimization. *Appl. Comput. Harmon. Anal.* **40** 321–351. MR3440176

[39] ROMBERG, J. (2009). Compressive sensing by random convolution. *SIAM J. Imaging Sci.* **2** 1098–1128. MR2559161

[40] RUDELSON, M. and VERSHYNIN, R. (2008). On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.* **61** 1025–1045. MR2417886

[41] RUDELSON, M. and VERSHYNIN, R. (2015). Small ball probabilities for linear images of high-dimensional distributions. *Int. Math. Res. Not. IMRN* **2015** 9594–9617. MR3431603

[42] TALAGRAND, M. (2014). *Upper and Lower Bounds for Stochastic Processes*. Springer, Heidelberg. MR3184689

[43] TROPP, J. A., WAKIN, M. B., DUARTE, M. F., BARON, D. and BARANIUK, R. G. (2006). Random filters for compressive sampling and reconstruction. In 2006 *IEEE International Conference on Acoustics Speech and Signal Processing Proceedings* **3** III–III. IEEE, New York.

[44] VERSHYNIN, R. (2012). Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing* 210–268. Cambridge Univ. Press, Cambridge. MR2963170

S. MENDELSON
DEPARTMENT OF MATHEMATICS
TECHNION
HAIFA, 32000
ISRAEL
AND
MATHEMATICAL SCIENCES INSTITUTE
AUSTRALIAN NATIONAL UNIVERSITY
CANBERRA 0200
AUSTRALIA
E-MAIL: shahar@tx.technion.ac.il

H. RAUHUT
CHAIR FOR MATHEMATICS C (ANALYSIS)
RWTH AACHEN UNIVERSITY
PONTDRIESCH 10
52056 AACHEN
GERMANY
E-MAIL: rauhut@mathc.rwth-aachen.de

R. WARD
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TEXAS AT AUSTIN
2515 SPEEDWAY
AUSTIN, TEXAS 78712
USA
E-MAIL: rward@ma.utexas.edu