ℓ_p-Spread and Restricted Isometry Properties of Sparse Random Matrices

Venkatesan Guruswami
University of California, Berkeley, CA, USA

Peter Manohar
Carnegie Mellon University, Pittsburgh, PA, USA

Jonathan Mosheiff
Carnegie Mellon University, Pittsburgh, PA, USA

Abstract

Random subspaces X of $\mathbb{R}^n$ of dimension proportional to n are, with high probability, well-spread with respect to the ℓ_2-norm. Namely, every nonzero $x \in X$ is "robustly non-sparse" in the following sense: x is $\varepsilon \|x\|_2$-far in ℓ_2-distance from all δn-sparse vectors, for positive constants ε, δ bounded away from 0. This "ℓ_2-spread" property is the natural counterpart, for subspaces over the reals, of the minimum distance of linear codes over finite fields, and corresponds to X being a Euclidean section of the ℓ_1 unit ball. Explicit ℓ_2-spread subspaces of dimension $\Omega(n)$, however, are unknown, and the best known explicit constructions (which achieve weaker spread properties), are analogs of *low density parity check* (LDPC) codes over the reals, i.e., they are kernels of certain *sparse* matrices.

Motivated by this, we study the spread properties of the kernels of *sparse random matrices*. We prove that with high probability such subspaces contain vectors x that are $o(1) \cdot \|x\|_2$-close to $o(n)$-sparse with respect to the ℓ_2-norm, and in particular are *not* ℓ_2-spread. This is strikingly different from the case of random LDPC codes, whose distance is asymptotically almost as good as that of (dense) random linear codes.

On the other hand, for $p < 2$ we prove that such subspaces *are* ℓ_p-spread with high probability. The spread property of sparse random matrices thus exhibits a threshold behavior at $p = 2$. Our proof for $p < 2$ moreover shows that a random sparse matrix has the stronger restricted isometry property (RIP) with respect to the ℓ_p norm, and in fact this follows solely from the unique expansion of a random biregular graph, yielding a somewhat unexpected generalization of a similar result for the ℓ_1 norm [6]. Instantiating this with suitable explicit expanders, we obtain the first explicit constructions of ℓ_p-RIP matrices for $1 \le p < p_0$, where $1 < p_0 < 2$ is an absolute constant.

2012 ACM Subject Classification Theory of computation → Pseudorandomness and derandomization; Theory of computation → Random projections and metric embeddings

Keywords and phrases Spread Subspaces, Euclidean Sections, Restricted Isometry Property, Sparse Matrices

Digital Object Identifier 10.4230/LIPIcs.CCC.2022.7

Funding *Venkatesan Guruswami*: Supported in part by NSF grants CCF-1908125 and CCF-2210823, and a Simons Investigator Award.
Peter Manohar: Supported in part by an ARCS Scholarship, NSF Graduate Research Fellowship (under grant numbers DGE1745016 and DGE2140739), and NSF CCF-1814603. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Jonathan Mosheiff: Supported in part by NSF CCF-1814603.

Acknowledgements We thank Sidhanth Mohanty for helpful discussions about the works of [9, 7, 8, 26, 27, 29], and Yuval Peled for helpful discussions about convergence theorems for graph spectra. We thank Amir Shpilka for bringing the work of [23] to our attention, and Ioana Dumitriu for bringing the works of [9, 34] to our attention.

37th Computational Complexity Conference (CCC 2022).
Editor: Shachar Lovett; Article No. 7; pp. 7:1–7:17

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Classical results in asymptotic geometric analysis on the Gelfand/Kolmogorov widths of ℓ_2 balls [14, 24, 16] show that random subspaces X of $\mathbb{R}^n$ of dimension proportional to n (say, defined as the kernel of random $n/2 \times n$ matrices with i.i.d. Gaussian or ± 1 entries) are *good Euclidean sections* of ℓ_1^n: namely, $\|x\|_1 \geq \Omega(\sqrt{n})\,\|x\|_2$ for every $x \in X$. An elementary proof of this fact also follows from the Johnson-Lindenstrauss (JL) property of random matrices, its connection to the restricted isometry property (RIP) and compressed sensing, and their relationship to the Euclidean sections property [4].

The condition $\|x\|_1 \geq \Omega(\sqrt{n})\,\|x\|_2$ can equivalently be expressed as a "well-spreadness" criterion satisfied by every nonzero vector $x \in X$: the largest δn entries of x have at most $1-\varepsilon$ of its ℓ_2 mass, for some positive constants δ, ε bounded away from 0 as $n \to \infty$. Equivalently, this means that all nonzero vectors $x \in X$ are *incompressible* – there is no sparse vector that approximates x well in ℓ_2 norm (in other words, $\|x - y\|_2 \geq \varepsilon\,\|x\|_2$ for all δn-sparse vectors y). This can be naturally viewed as a robust analog, for subspaces of $\mathbb{R}^n$, of the distance property of linear error-correcting codes.

The above well-spreadness criterion can naturally be imposed with respect to any ℓ_p metric: a subspace X is said to be *ℓ_p-spread* if every nonzero vector $x \in X$ is $\varepsilon\,\|x\|_p$-far in ℓ_p-distance from all δn-sparse vectors. The ℓ_p-spread property is a more stringent requirement for larger p. For $p > 2$, the optimal asymptotic dimension of ℓ_p-spread subspaces is at most $O_p(n^{2/p})$ and thus $o(n)$ [17]. In this work, we therefore focus on $p \in [1, 2]$ where it is possible to have ℓ_p-spread subspaces of dimension proportional to n.

For a subspace X of $\mathbb{R}^n$, define its *ℓ_p-distortion* $\Delta_p(X)$ to be the following quantity:

$$\Delta_p(X) := \sup_{x \in X \setminus \{0^n\}} \frac{n^{1-\frac{1}{p}}\,\|x\|_p}{\|x\|_1} \ .$$

Note that $1 \leq \Delta_p(x) \leq n^{1-1/p}$. Good ℓ_p-spread of X can be captured by the condition that $\Delta_p(X)$ is bounded by a fixed constant independent of n; this generalizes the aforementioned equivalence of ℓ_2-spread and the Euclidean section property. The term *distortion* is used because the natural inclusion of X in $\mathbb{R}^n$ induces a bi-Lipschitz embedding of X, taken with the ℓ_p norm, into ℓ_1^n, with distortion $\Delta_p(X)$. The distortion/spread property of subspaces with respect to different ℓ_p norms has been extensively studied, owing to its connections to width properties in convex geometry [17, 25], embeddings between metric spaces [20], compressed sensing [13, 10, 25], error-correction over the reals [11, 18], and the restricted isometry (RIP) and dimensionality-reduction/Johnson-Lindenstrauss (JL) properties [25, 4, 1].

Despite a lot of interest and the abundance of probabilistic constructions, an outstanding question is to construct an *explicit* subspace $X \subseteq \mathbb{R}^n$ of dimension $\Omega(n)$ that is ℓ_2-spread, or equivalently has $\Delta_2(X) \leq O(1)$. By explicit, we mean deterministically constructing a basis for the subspace (or its dual) in $\mathrm{poly}(n)$ time.[1] This is a counterpart, for subspaces of $\mathbb{R}^n$, of the problem of constructing asymptotically good binary linear codes $C \subseteq \{0,1\}^n$: namely, codes whose dimension and minimum distance are both proportional to n. In addition to being a natural and basic challenge in pseudorandomness, explicit constructions are also valuable in applications of spread subspaces such as compressed sensing, as they provide a *guarantee* that the matrix will have the stipulated properties. This is particularly important since there are no known methods to efficiently certify the ℓ_2-spread (or even ℓ_p-spread) of random subspaces.

[1] Explicit constructions of ℓ_p-spread spaces of dimension $\Omega(n)$ are given in [6] ($p = 1$) and [23] ($1 \leq p < 2$).

1.1 Kernels of sparse matrices

In the case of $p = 2$, the best known explicit constructions of subspaces $X \subseteq \mathbb{R}^n$ with $\dim(X) \geq \Omega(n)$, in terms of their distortion $\Delta_2(X)$, are due to [19]. They give a construction analogous to Tanner codes from coding theory [32], combining appropriately chosen unbalanced bipartite expanders and local subspaces, to produce X with $\dim(X) \geq n - o(n)$ and $\Delta_2(X) \leq (\log n)^{O(\log\log\log n)}$ (so almost poly-logarithmic).[2] A simpler construction analogous to Sipser-Spielman codes [31], using s-regular spectral expanders and local well-spread subspaces of $\mathbb{R}^s$, was given in [18] and achieves[3] $\Delta_2(X) \leq n^{O(1/\log s)}$. An alternate probabilistic construction achieving similar parameters to [18] based on tensor products was given in [22]. The approach of [22] can further achieve distortion approaching 1 at the expense of making $\dim(X)$ smaller, but still $\Omega(n)$.

One notable attribute of the constructions above is that the subspace X can be expressed as the kernel of a matrix that is *sparse*. For instance, the construction of [18] picks a matrix where each row is s-sparse with ± 1 entries (that are chosen randomly for a probabilistic construction), and the construction in [22] defines the subspace $X \subseteq \mathbb{R}^n$ as the k-fold tensor product of another subspace, and so X can be defined as the kernel of an $n^{1/k}$-sparse matrix. Moreover, known explicit constructions of ℓ_p-spread subspaces for $1 \leq p < 2$ [6, 23] are also kernels of sparse matrices.

The sparsity of these constructions is inherited from the "underlying constructions" for codes; the constructions of [19, 18, 22] come from "lifting" constructions of linear codes (namely, Tanner codes [32], Sipser-Spielman codes [31], and tensor product codes, respectively) to this setting, and these constructions (for linear codes) are known to give good low density parity check (LDPC) codes: namely, codes that are the kernels of sparse matrices.

In light of these works, a natural question (and indeed one explicitly posed in [18]), is the following.

▶ **Question 1.** *Does there exist an $m \times n$ matrix A with $n - m \geq \Omega(n)$ whose rows are s-sparse for $s \leq O(1)$ (or even $s \leq \mathrm{polylog}(n)$) such that $\Delta_2(ker(A)) \leq O(1)$?*

The approaches of [18, 22] show that one can achieve $\Delta_2(\ker(A)) \leq \exp(O(1/\delta))$ when $s = n^\delta$. A positive answer to Question 1, even via random matrices, would likely yield good progress towards explicit constructions, as $O(1)$-sparse matrices are likely easier to derandomize than dense random ones. A negative answer to Question 1 would likely rule out explicit constructions based on the current state-of-the-art approaches of [19, 18, 22].

In addition to exploring the potential of the approaches behind the current best constructions, sparsity is desirable from a computational efficiency standpoint. Sparse matrices lead to faster algorithms, for example when used as measurement matrices in compressed sensing or to compute a sparse JL transform for dimensionality-reduction.

Motivated by these considerations, we study the ℓ_2-spread, and, more generally, ℓ_p-spread $(1 \leq p \leq 2)$ of subspaces defined as the kernel of *sparse random matrices*. Such subspaces are the continuous analogues of random low density parity check (LDPC) codes. Random LDPC codes have been studied in coding theory since Gallager's seminal work [15], with a renaissance since the mid 1990s [30] due to their fast iterative decoding algorithms and performance close to capacity.

[2] For sublinear dimension, an explicit construction of $X \subseteq \mathbb{R}^n$ with distortion $\Delta_2(X) \leq 1 + o(1)$ and $\dim(X) \geq n/2^{O((\log\log n)^2)}$ was given in [21].

[3] This construction is not explicit except for very small s, as the local subspace of $\mathbb{R}^s$ is either constructed by brute force or drawn at random.

Random LDPC codes are known to achieve rate vs. distance trade-offs approaching that of random (dense) linear codes [15]. Recently, even the list-decodability, and indeed any "local" property, of random LDPC codes was shown to be similar to that of random linear codes [28]. Given that random subspaces are well-spread and that random LDPC codes achieve similar properties to random (dense) codes, one might naturally expect, by analogy, that the kernels of sparse random matrices are also well-spread.

1.2 Our results

Our results paint a precise picture of the ℓ_p-spread of kernels X of sparse random matrices. Before stating our results, we first define ℓ_p-spread and state the random matrix model that we use.

▶ **Definition 2** (ℓ_p-spread). *Fix $p \in [1,\infty]$, $\varepsilon \in [0,1]$ and $k \le n \in \mathbb{N}$. A vector $y \in \mathbb{R}^n$ is* k-sparse *if $|\mathrm{supp}(y)| \le k$. A vector $x \in \mathbb{R}^n \setminus \{0^n\}$ is said to be* (k,ε)-ℓ_p-compressible *if there exists a k-sparse $y \in \mathbb{R}^n$ such that $\|x-y\|_p \le \varepsilon\|x\|_p$. Otherwise, we say that x is* (k,ε)-ℓ_p-spread.

A subspace $X \subseteq \mathbb{R}^n$ is (k,ε)-ℓ_p-spread *if every $x \in X \setminus \{0^n\}$ is (k,ε)-ℓ_p-spread.*

The random matrix model. A matrix $A \in \{0,1,-1\}^{m\times n}$ is said to be (s,t)-*biregular* if every row and column of A has exactly s and t nonzero entries, respectively. Let $\mathcal{M}_{m,n,s,t}$ denote the set of all (s,t)-biregular matrices in $\{0,1,-1\}^{m\times n}$.

All of our theorems for random matrices will be for a matrix A drawn uniformly at random from $\mathcal{M}_{m,n,s,t}$, where $\alpha = \frac{m}{n} = \frac{t}{s} \in (0,1)$ is a fixed constant and $n \to \infty$; for this exposition, we will use A to denote a random matrix from $\mathcal{M}_{m,n,s,t}$, and B to denote an arbitrary matrix in $\{0,1,-1\}^{m\times n}$. We additionally assume that $s := s(n) \le n^c$ for some absolute constant $0 < c < 1$, and $t = \alpha s \ge 3$. An event $\mathcal{E}$ is said to hold *with high probability* if $\lim_{n\to\infty} \Pr[\mathcal{E}] = 1$. All asymptotic notation refers to the regime of $n \to \infty$ and constant α. The constants implied by asymptotic notation are universal, unless stated otherwise. The symbols c, c', c_1 and c_2 always stand for positive universal constants, which may differ across different lemma and theorem statements.

1.2.1 Poor ℓ_2-spread of sparse random matrices

Our first theorem shows that, surprisingly, $\ker(A)$ is, with high probability, *not* ℓ_2-spread.

▶ **Theorem 3** (Poor ℓ_2-spread of $\ker(A)$). *With high probability over A, there exists an $(m^c, \frac{n^{-\Omega(\log(1/\alpha)/\log s)}}{1-\sqrt{\alpha}})$-$\ell_2$-compressible vector $x \in \ker(A)$, where $c < 1$ is an absolute constant. In particular,*

$$\Delta_2(\ker(A)) \ge (1-\sqrt{\alpha}) \cdot n^{\Omega(\log(1/\alpha)/\log s)} \ .$$

Moreover, there is a $\mathrm{poly}(n)$-time algorithm that, on input A, outputs such an x.

Choosing $s = O(1)$ in Theorem 3 (and letting α be bounded away from 1) implies[4] that $\Delta_2(\ker(A)) \ge n^{\Omega(1)}$ with high probability, and choosing $s = \mathrm{polylog}(n)$ implies $\Delta_2(\ker(A)) \ge n^{\Omega(\log(1/\alpha)/\log\log n)}$. We always trivially have $\Delta_2(\ker(A)) \le \sqrt{n}$, so not only does Theorem 3

[4] Note that since $\alpha s = t \ge 3$, we must have $\log s \ge \log \frac{1}{\alpha}$.

answer Question 1 in the negative for sparse random matrices, but it also does so in a very strong sense. For instance, when $s = O(1)$, Theorem 3 shows that $\Delta_2(\ker(A))$ is "maximally bad", up to a constant factor in the exponent.

Another point of interest is the choice $s = n^\delta$ for some fixed δ. This yields the tradeoff of $\Delta_2(\ker(A)) \geq (\frac{1}{\alpha})^{\Omega(\frac{1}{\delta})}$, which precisely matches the tradeoff (in terms of δ) achieved by both [18, 22]. While our matrix ensemble is "more random" compared to those in [18, 22], Theorem 3 can nonetheless be interpreted as giving evidence that this $\exp\left(O(\frac{1}{\delta})\right)$ tradeoff from [18, 22] is tight and inherent to sparse constructions.

Our proof of Theorem 3 is *constructive*, in the sense that we give a very simple, efficient algorithm to find such an $x \in \ker(A)$. This moreover shows that for sparse random matrices, one can efficiently *refute* the claim that $\Delta_2(\ker(A)) = O(1)$, as the vector x is a refutation witness. Our algorithm provides an interesting counterpoint to the work of [3], who gave an algorithm based on the sum-of-squares SDP hierarchy to certify that $\Delta_2(\ker(A)) \leq O(1)$ with high probability for *dense* matrices A where $\dim(\ker(A)) \leq O(\sqrt{n})$. In contrast, our algorithm succeeds when $\dim(\ker(A)) = \Omega(n)$ and the matrix A is *sparse*. The two results taken together suggest an interesting relationship between the density and $\dim(\ker(A))$ of matrices A for which we can efficiently certify or refute bounds on $\Delta_2(\ker(A))$.

We also note that, by the well-known duality formula relating Kolmorogov and Gelfand widths (see [KT07] and the references therein), Theorem 3 implies that the row span of A is far from approximating the ℓ_2-sphere in ℓ_∞ distance. Concretely, with high probability over A there exists $x \in \mathbb{R}^n$ with $\|x\|_2 = 1$ that is $(1 - \sqrt{\alpha}) \cdot n^{\Omega(\log(1/\alpha)/\log s)}/\sqrt{n}$-far in ℓ_∞ norm from all vectors of the form $A^\top y$, where $y \in \mathbb{R}^m$.

The proof of Theorem 3 requires the following strong bound that we show on the singular values of A.

▶ **Theorem 4** (Singular value bound). *With high probability, the set of singular values $\sigma(A)$ of A satisfy*

$$\sigma(A) \subseteq \left[\sqrt{s-1} - (1+o(1)) \cdot \sqrt{t-1}, \sqrt{s-1} + (1+o(1)) \cdot \sqrt{t-1}\right] .$$

Moreover, the above bound holds even without our (otherwise global) assumption that $s \leq n^c$ for some absolute constant $c < 1$.

Theorem 4 should not be surprising, especially given the recent works of [9, 7, 8, 26, 27, 29, 34], and indeed our proof follows the same overall blueprint of these works. Most of these papers, however, only handle the case when the degree of the graph is *constant* as $n \to \infty$;[5] this corresponds to the case of $s = O(1)$ in Theorem 4. Theorem 4 thus differs as it allows for $s = \omega(1)$, and indeed we can even take $s = n^c$ for some absolute constant $c < 1$. On the other hand, most of the aforementioned works deal with the case of *unsigned* adjacency matrices, whereas we only prove Theorem 4 for *randomly signed* adjacency matrices. We note that proving the analogue of Theorem 4 for unsigned adjacency matrices (where $\sigma(A)$ now denotes the set of singular values, excluding the trivial value of $\sqrt{st}$) and for all (non-constant) s, t remains open.

The singular value bound in Theorem 4 is challenging to prove because it is so sharp. Indeed, it is not too difficult to show that $\sigma(A) \subseteq [\sqrt{s} - O(\sqrt{t}), \sqrt{s} + O(\sqrt{t})]$ with high probability via black-box applications of known results, e.g., [2]. However, this does not

[5] The exceptions are [26], which handles polylog(n) degree, and [34], which handles n^c degree but does not obtain as sharp bounds.

suffice for our use in the proof of Theorem 3, as the aforementioned weaker bound would only suffice to prove Theorem 3 provided that $\alpha \leq c$ for some absolute constant c, where c depends on the absolute constant c' hidden in the "$O(\sqrt{t})$". We need the sharp bound of Theorem 4 in order to allow for α to be an arbitrary constant in $(0, 1)$.

As a counterpart to Theorem 3, we give the following partial converse, which shows that $\ker(A)$ *is* (k, ε)-ℓ_2-spread for a weak choice of parameters k and ε.

▶ **Theorem 5** (Converse to Theorem 3)**.** *Assume that $t \geq 9$. Then, with high probability over A, the space* $\ker(A)$ *is* $\left(\Omega(\alpha^2 n/t^4), \alpha^{O(\log n/\log t)}\right)$*-$\ell_2$-spread.*

We note that in Theorem 5, the parameter k is $\Omega(\alpha^2 n/t^4) = m^{\Omega(1)}$ and the parameter ε is $\alpha^{O(\log n/\log s)}$.[6] Theorem 5 thus shows that the parameters in Theorem 3 are tight up to the universal constants in the exponent. Our proof of Theorem 5 is an adaptation of the proof of [5, Lemma 3.4].

1.2.2 ℓ_p-RIP and ℓ_p-spread for $p < 2$

We next focus on the ℓ_p norm for $p < 2$. For $p < 2$, there are known explicit constructions of ℓ_p-spread subspaces [23]. Because of this, for $p < 2$ we focus on the stronger, well-studied *Restricted Isometry Property (RIP)*. We also note that the constructions of [23] are highly structured, and so even though they also come from sparse matrices, they do not tell us anything about the ℓ_p-spread of sparse *random* matrices.

We prove that sparse random matrices are not only ℓ_p-spread, but are also ℓ_p-RIP, and, moreover, this follows merely from the expansion of the underlying bipartite graph of the random matrix A. In particular, we prove that *any* signed adjacency matrix B of a left-regular bipartite expander graph G is ℓ_p-RIP, provided that the maximum right degree $s_{\max}$ is above a small threshold independent of n.

The RIP is a well-studied property of matrices from the compressed sensing literature, defined as follows.

▶ **Definition 6** (ℓ_p-RIP)**.** *Let $B \in \mathbb{R}^{m \times n}$ be a matrix. We say that B is* (k, ε)-ℓ_p-RIP *if there exists $K > 0$ such that for every k-sparse $x \in \mathbb{R}^n$, it holds that*[7]

$$K(1-\varepsilon)\|x\|_p \leq \|Bx\|_p \leq K(1+\varepsilon)\|x\|_p \quad .$$

We note that ℓ_p-RIP implies ℓ_p-spread, and in fact it is a strictly stronger property [25].

RIP matrices have been studied extensively in the context of compressed sensing, as they yield a polynomial-time algorithm based on linear programming for the *robust sparse recovery problem*. Namely, given a "noisy measurement sketch" $y = Bx + e$ of a vector x, where B is (k, ε)-ℓ_p-RIP and $\|e\|_p \leq \eta$, there is a polynomial-time algorithm to recover an estimate $\hat{x}$ for x with the so-called "ℓ_p-ℓ_1 guarantee," namely the estimate $\hat{x}$ satisfies $\|\hat{x} - x\|_p \leq O\left(k^{-(1-\frac{1}{p})}\|x - x^*\|_1 + \eta\right)$, where x^* is a k-sparse vector minimizing $\|x - x^*\|_1$ (see Appendix A in [1] for details). We note that if B is merely ℓ_p-spread, then B suffices for the (non-robust) sparse recovery problem, i.e., when there is no noise e.

We now turn to formally stating our results. We first recall the definition of a (unique) bipartite expander.

---

[6] This follows since $t = \alpha s \leq s$, $m = \alpha n$, and $s \leq n^c$ for some absolute constant c.

[7] We note that the standard definition of RIP typically appears without the normalization factor K above. We include the parameter K for convenience, as the random sparse matrices we consider are not normalized.

▶ **Definition 7** (Unique expanders). *A bipartite graph $G = (V_L = [n], V_R = [m], E)$ is a* t-left-regular (γ, μ)-unique expander *if (1)* $\deg(u) = t$ *for all* $u \in V_L$*, and (2) for all* $S \subseteq V_L$*,* $|S| \leq \gamma n$*, there are at least* $t(1-\mu)|S|$ *vertices* $v \in V_R$ *which each have exactly one neighbor in* S*.*

A matrix $B \in \{0, 1, -1\}^{m \times n}$ is a *signed adjacency matrix* of a bipartite graph $G = (V_L = [n], V_R = [m], E)$ if

$$B_{r,u} \neq 0 \iff (u, r) \in E$$

for all $u \in V_L$, $r \in V_R$.

▶ **Theorem 8** (ℓ_p-RIP of expander graphs). *Let G be a bipartite t-left-regular (γ, μ)-unique expander with maximum right degree $s_{\max}$, and let B be any signed adjacency matrix of G. Let $0 < \varepsilon \leq 1$ and $1 \leq p < 2$ such that $\varepsilon^2 \geq 9\mu s_{\max}^{p-1}$. Then, B is $(\gamma n, \varepsilon)$-ℓ_p-RIP, i.e., for every γn-sparse $x \in \mathbb{R}^n$,*

$$t^{\frac{1}{p}}(1-\varepsilon)\|x\|_p \leq \|Bx\|_p \leq t^{\frac{1}{p}}(1+\varepsilon)\|x\|_p \quad .$$

Theorem 8 generalizes a result of [6], which shows that any signed adjacency matrix B of G is ℓ_1-RIP, provided that G is an expander. This is somewhat surprising, as the proof in [6] makes heavy use of properties specific to the ℓ_1 norm.[8]

The ℓ_p-RIP of matrices for general p has been studied in other contexts, most notably in [1]. As is typical when studying RIP matrices, they view the sparsity parameter k as a fixed function of n, and determine m as a function of k, n. However, the results in [1] are incomparable to ours, as they hold only for the low-sparsity case of $k = O(n^{1/p})$ (so $k = o(n)$ if $p > 1$), but we are concerned with the case of $k = \Omega(n)$, when the sparsity is a small constant fraction of n.

As a random t-left-regular bipartite graph is a good expander with high probability, we obtain the following corollary of Theorem 8, which shows that $\ker(A)$ for $A \leftarrow \mathcal{M}_{m,n,s,t}$ achieves very good ℓ_p-spread for every $p \in [1, 2)$. Thus, the poor ℓ_2-spread of $\ker(A)$ is in fact specific to the case of $p = 2$.

▶ **Corollary 9** (Good ℓ_p-RIP and ℓ_p-spread of A). *Fix $p \in [1, 2)$, $0 < \varepsilon < \frac{1}{2}$, and suppose that $s \geq \left(\frac{18}{\alpha\varepsilon^2}\right)^{\frac{1}{2-p}}$. Then, with high probability over A, the matrix A is $(\Omega(\gamma n), \varepsilon)$-$\ell_p$-RIP for $\gamma = \frac{\alpha^2}{t^4}$: for every $\Omega(\gamma n)$-sparse $x \in \mathbb{R}^n$, it holds that*

$$t^{\frac{1}{p}}(1-\varepsilon)\|x\|_p \leq \|Ax\|_p \leq t^{\frac{1}{p}}(1+\varepsilon)\|x\|_p \quad .$$

In particular, the subspace $\ker(A)$ is $\left(\Omega(\gamma n), \Omega\left(\gamma^{1-\frac{1}{p}}\right)\right)$-$\ell_p$-spread and $\Delta_p(\ker(A)) \leq O\left(1/\gamma^{2-\frac{2}{p}}\right)$.

Fixing p, α, ε to be constants and taking s to be a large enough constant, this shows that $\ker(A)$ is $(\Omega(n), \Omega(1))$-ℓ_p-spread with high probability, and therefore $\Delta_p(\ker(A)) = O(1)$. Together with Theorem 3, this shows that the ℓ_p-spread property of $\ker(A)$ exhibits an interesting *threshold phenomenon* at $p = 2$.

[8] They also show that their proof for ℓ_1-RIP extends to ℓ_p-RIP for $p \leq 1 + O(\frac{1}{\log n})$, because the "Hölder factor" of $n^{1-\frac{1}{p}}$ is $O(1)$, but it does not extend to ℓ_p for any constant $p > 1$.

We also combine Theorem 8 with the explicit constructions of expander graphs of [12] to obtain the following corollary, which gives an explicit construction of ℓ_p-RIP matrices for all $p \in [1, p_0)$, where $1 < p_0 < 2$ is an absolute constant. We thus obtain the first explicit construction of a matrix B achieving the "ℓ_p/ℓ_1 guarantee" for the robust sparse recovery problem, and our matrices are for the regime $k = \Theta(n)$ and any $p \in [1, p_0)$. Previously, such constructions were only known for $p \le 1 + O\left(\frac{1}{\log n}\right)$ [6]. Unlike Corollary 9, our explicit constructions only extend up to some threshold $p_0 < 2$. This is because the expanders of [12] achieve weaker expansion than random graphs. Concretely, the "expansion error" μ of the [12] expanders is $\mu = O(1/t)^\tau$ for some constant $\tau < 1$, which yields the threshold of $p_0 = 1 + \tau$, whereas random graphs achieve $\mu = O(1/t)$, allowing for $p_0 = 2$.

▶ **Corollary 10** (Explicit construction of ℓ_p-RIP matrices). *Let $0 < \varepsilon < \frac{1}{2}$, $\alpha \in (0,1)$, and let $n \in \mathbb{N}$ be sufficiently large. For some universal constant $1 < p_0 < 2$, there exists a deterministic algorithm which, given $p \in [1, p_0)$, ε, α and n, outputs in time $\mathrm{poly}(n/\delta) + 2^{O(1/\delta)}$ a matrix $B \in \{0,1\}^{m \times n}$, for some $m \le \alpha n$, such that B is $(\gamma n, \varepsilon)$-ℓ_p-RIP, for some $\delta, \gamma = \mathrm{poly}(\varepsilon, \alpha)^{\frac{1}{p_0 - p}}$. In particular, $\ker(B)$ is $(\gamma n, \gamma^{1-\frac{1}{p}})$-$\ell_p$-spread and $\Delta_p(\ker(B)) \le 1/\gamma^{2-\frac{2}{p}}$.*

Note that as ε, α and p are constants, the matrix B in Corollary 10 is $(\Omega(n), O(1))$-ℓ_p-RIP, $\ker(B)$ is $(\Omega(n), \Omega(1))$-ℓ_p-spread, and $\Delta_p(\ker(B)) \le O(1)$.

As noted earlier, [23] gives explicit constructions of $(\Omega(n), \Omega(1))$-ℓ_p-spread subspaces, for *all* $1 \le p < 2$. This is incomparable to Corollary 10: on one hand, [23] obtains the full range of $1 \le p < 2$, but on the other hand, his matrices only are ℓ_p-spread and do not satisfy the (strictly stronger) ℓ_p-RIP. Our construction is moreover the "simplest" black-box reduction to expansion: we show that the mere adjacency matrix of a bipartite expander is ℓ_p-RIP. While the constructions in [23] are themselves not too complicated, we think that this is nonetheless an interesting conceptual contribution of our work.

Finally, we also prove the following partial converse to Corollary 9, which shows that when $s^{2-p} \lessapprox \frac{1}{\alpha}$ (i.e., s^{2-p} is a constant factor below the threshold in Corollary 9), then A is *not* ℓ_p-RIP.

▶ **Theorem 11** (Partial converse to Corollary 9). *Let $p \in [1,2)$, $\varepsilon > 0$. If $s - 1 \le \left(\frac{1}{(1+\varepsilon)\alpha}\right)^{\frac{1}{2-p}}$, then with high probability over A, there exists an n^c-sparse vector $x \in \mathbb{R}^n \setminus \{0^n\}$ such that*

$$\frac{\|Ax\|_p}{\|x\|_p} \le t^{\frac{1}{p}} \cdot m^{-\Omega\left(\frac{\varepsilon}{\log s}\right)} .$$

Note that $\|Ae_1\|_p / \|e_1\|_p = t^{\frac{1}{p}}$ always holds, so Corollary 9 demonstrates that, given small enough s, the ratio $\frac{\|Ax\|_p}{\|x\|_p}$ has a large range over different choices of n^c-sparse x.

2 Proof overview

We outline the proofs of our results. For the purposes of this exposition, we will adopt the same convention as in Section 1.2 and use A and B to denote a uniformly sampled matrix from $\mathcal{M}_{m,n,s,t}$ and arbitrary matrix from $\{0,1,-1\}^{m \times n}$, respectively. Recall that $\mathcal{M}_{m,n,s,t}$ denotes the set of (s,t)-biregular matrices with entries in $\{0,1,-1\}$, and that $\frac{m}{n} = \frac{t}{s} = \alpha$ for some constant α, and $n \to \infty$. For simplicity of this exposition, in this section we restrict ourselves to the regime $s = O(1)$ unless stated otherwise.

We naturally associate with B the bipartite graph $G = G_B = (V_L, V_R, E)$ with $n = |V_L|$ left vertices, $m = |V_R|$ right vertices, and an edge between $u \in V_L$ and $r \in V_R$ if $B_{r,u} \neq 0$. We view the rows and columns of B as indexed by V_R and V_L, respectively, and identify $\mathbb{R}^n$ with $\mathbb{R}^{V_L}$, and $\mathbb{R}^m$ with $\mathbb{R}^{V_R}$. In addition, we define the function $\mathsf{sign} = \mathsf{sign}_B : E \to \{1, -1\}$, which maps an edge $\{u, r\}$ as above to $B_{r,u}$. We note that the combination of G_B and sign_B completely describes B.

2.1 Theorem 3: $\ker(A)$ in not ℓ_2-spread

For simplicity, we only sketch here why $\ker(A)$ is likely to contain an $(o(n), o(1))$-compressible vector.

The proof of Theorem 3 consists of two steps. In the first step we find an $o(n)$-sparse vector $x \in \mathbb{R}^n$ with $\|x\|_2 \geq 1$ and $\|Ax\|_2 \leq o(1)$. In the second step we find a vector $y \in \ker(A)$ with $\|y - x\|_2 \leq o(1) \cdot \|y\|_2$. In particular, y is $(o(n), o(1))$-compressible, so $\ker(A)$ cannot be ℓ_2-spread.

Below, we outline these two steps. It is straightforward to see, given the construction described below, that both x and y can be computed in polynomial time given A.

We also note that an ℓ_p analog of Step 1 is the main technical component in the proof of Theorem 11.

Step 1: constructing a sparse x with small $\|Ax\|_2$. To obtain the vector x, we first prove that G is highly likely to contain a vertex $v^* \in V_L$ such that the ball of radius $2\ell + 1$ about v^*, for some $\ell \leq O(\log n)$, contains no cycles. That is, the radius-ℓ neighborhood of v^* is a complete (t, s)-biregular tree T rooted at v^*. Recall that a rooted tree is (t, s)-biregular if the even depth (resp. odd depth) inner vertices have degree t (resp. s). The existence of such a vertex v^* is the only random property of A needed in this step of the proof. In particular, assuming that G has the aforementioned property, our construction of x is always possible, regardless of the sign function.

To describe the construction of x itself, we assume for simplicity that $\mathsf{sign}(e) = 1$ for all $e \in E$. Namely, all the non-zero entries of A are 1. In this setting, let $v \in V_L \cap T$ be a vertex of depth $2k$ in the tree for some $k \geq 0$ (note that a vertex in V_L must have even depth), and set $x_v = (-(s-1))^{-k}$. For any $x \in V_L \setminus T$, set $x_v = 0$. Note that $\mathrm{supp}(x) \subseteq T$. We choose ℓ above to be as large as possible, i.e., $O(\log n)$, so that the size of T is roughly n^c for some $c < 1$. In particular, x is $\approx n^c$-sparse. Also, note that $\|x\|_2^2 \geq x_{v^*}^2 = 1$. We informally refer to the vector x produced by this construction as a *tree vector*.

Our construction guarantees that $(Ax)_r = 0$ for every internal node $r \in T \cap V_R$. Indeed, suppose that r is of depth $2k + 1$. Then, it has one neighbor of depth $2k$, and $s - 1$ neighbors of depth $2k + 2$. As $(Ax)_r$ is the sum of x_v over neighbors v of r, we have $(Ax)_r = (-(s-1))^{-k} + (s-1) \cdot (-(s-1))^{-(k+1)} = 0$.

To compute $\|Ax\|_2$, it thus suffices to compute $|(Ax)_r|$ when r is one of the $t(t-1)^\ell(s-1)^\ell$ leaves of T. It is not hard to see that in this case $|(Ax)_r| = (s-1)^\ell$, and so

$$\|Ax\|_2^2 = t(t-1)^\ell(s-1)^\ell \cdot (s-1)^{-2\ell} = e^{-\Omega(\ell)} = o(1) \ .$$

We note that our tree vector construction is similar in spirit to a construction by Noga Alon [18, Theorem 8], which demonstrates the limitations of expander-based analysis of the spread property. In [18], however, they *choose* their graph G so that (their analog of) the tree vector x will lie in (their analog of) $\ker(A)$ *by design*. Our graph is random and not

up to our choice, so we cannot simply orchestrate the graph so that our tree vector x to belong to $\ker(A)$. This necessitates that we perform the nontrivial step of rounding x to some $y \in \ker(A)$, which we discuss next.

Step 2: finding $y \in \ker(A)$ close to x. Our main goal in this step is to establish the following lemma:

▶ **Lemma 12** (Informal). *With high probability over A, it holds that every $x \in \mathbb{R}^n$ with $\|Ax\|_2 \le o(1)$ is $o(1)$-close to some vector $y \in \ker(A)$.*

Indeed, let x be the tree vector constructed in Step 1. Then x is $o(n)$-sparse with $\|x\|_2 \ge 1$ and $\|Ax\|_2 \le o(1)$. By Lemma 12, there exists a vector $y \in \ker(A)$, which is $o(1)$-close to x. This vector y is $(o(n), o(1))$-compressible, which yields Theorem 3 in the present parameter setting.

One may naively try to prove Lemma 12 by locally perturbing x to try to make $Ax = 0^m$, e.g. by designing a greedy algorithm for this task. This approach, however, seems difficult to execute, especially given that Lemma 12 is in fact not true in general. For example, it could be the case that x is a (unit norm) right singular vector of A with singular value $o(1)$. Then, $\|Ax\|_2 = o(1)$, but $\|x - y\|_2 \ge 1$ for all $y \in \ker(A)$, and in fact the closest vector in $\ker(A)$ to x is 0^n.

Instead, we set y to be the orthogonal projection of x onto $\ker(A)$. In hindsight, this is the obvious choice for y, as then $y \in \ker(A)$ is the vector that minimizes $\|x - y\|_2$. How large can $\|x - y\|_2$ be? Intuitively, we would like to say that $\|Ax\|_2$ being small implies that $\|x - y\|_2$ is small as well. As the earlier example shows, this is not true for a general matrix A, as A could have small singular values. However, the implication *does* hold provided that all singular values of A are $\Omega(1)$.[9] Indeed, the singular value decomposition of A implies that

$$\|Ax\|_2 = \|A(x-y)\|_2 \ge \sigma_{\min}(A)\,\|x-y\|_2 \quad ,$$

where $\sigma_{\min}(A)$ is the minimum singular value of A and the inequality holds as $x - y$ is orthogonal to $\ker(A)$. Hence, $\|x - y\|_2 \le \frac{\|Ax\|_2}{\sigma_{\min}(A)}$.

The main technical component of Step 2 is therefore the lower bound on $\sigma_{\min}(A)$, given by Theorem 4. As we have argued above, the crude lower bound of $\sigma_{\min}(A) \ge \Omega(1)$ suffices to yield Lemma 12. Indeed, if $\sigma_{\min}(A) \ge \Omega(1)$, then $\|x - y\|_2 \le o(1)$, and so $\ker(A)$ contains an $(o(n), o(1))$-compressible vector. The precise high-probability lower bound on $\sigma_{\min}(A)$ established in Theorem 4 implies a finer quantitative version of Lemma 12, which yields the full Theorem 3. The latter gives a much sharper bound on the $o(1)$ term, and also applies to sparsity all the way up to $O(n^c)$ for some $c > 0$.

We remark that one can easily show that $\sigma_{\min}(A) \ge \sqrt{s} - O(\sqrt{t})$ via "off-the-shelf" methods, such as [2]. However, this would only allow us to prove Theorem 3 provided that $\alpha \le c$ for some absolute constant $c < 1$ (related to the $O(1)$ factor in front of $\sqrt{t}$ above), and thus would not allow us to take α to be *any* constant in $(0, 1)$, e.g., $\alpha = 0.999$. Our sharper bound also highlights the difficulty in lower bounding the minimum singular value when $\alpha = m/n$ is close to 1.

We postpone our discussion of the proof of Theorem 4 to Section 2.3, and turn next to our positive result for ℓ_p-spread for $p < 2$.

[9] Technically, what matters is the minimum *nonzero* singular value. However, with high probability the matrix A will be full rank (i.e., rank m), so that $\sigma_{\min}(A) > 0$. Indeed, this is trivially implied by Theorem 4.

2.2 Theorem 8: ℓ_p-RIP for $p < 2$ from vertex expansion

We sketch the proof of Theorem 8. For simplicity, we will assume that $B \in \mathcal{M}_{m,n,s,t}$, i.e., that the bipartite graph G_B is t-left-regular and s-right-regular (and hence $s_{\max} = s$) and also that G_B is a (γ, μ)-unique expander. For this exposition, we only discuss the claimed lower bound on $\|Bx\|_p$ stated in the theorem, namely,

$$\|Bx\|_p \geq t^{\frac{1}{p}}(1-\varepsilon)\|x\|_p \tag{1}$$

for all γn-sparse $x \in \mathbb{R}^n$, as the upper bound is obtained via a variation on the same method.

Theorem 8 for tree vectors. As a warm-up for the proof of Theorem 8, we show why the $o(n)$-sparse tree vector x constructed in Section 2.1 does not yield a counterexample to Equation (1). Let $x^{(k)}$ $(0 \leq k \leq \ell)$ denote the restriction of x to the vertices in the $2k$-th level of the tree T. Then,

$$\left\|x^{(k)}\right\|_p^p = t(t-1)^{k-1}(s-1)^{(1-p)k} \ .$$

For $p = 2$, this expression decreases exponentially in k, and thus the ℓ_2-mass of x is concentrated at the top of the tree. For $p < 2$, however, $\left\|x^{(k)}\right\|_p^p$ actually *grows* exponentially in k provided that s is large enough (concretely, one needs $s^{2-p} \gtrapprox \frac{1}{\alpha}$).[10] In this case, the ℓ_p-mass is concentrated towards the bottom of the tree. Moreover, one can take s large enough so that all but an $\frac{\varepsilon}{2}$-fraction of the mass lies in the bottom layer. Then,

$$\frac{\|Bx\|_p^p}{\|x\|_p^p} \geq \left(1-\frac{\varepsilon}{2}\right)\cdot\frac{\|Bx\|_p^p}{\left\|x^{(\ell)}\right\|_p^p} = \left(1-\frac{\varepsilon}{2}\right)\cdot\frac{t(t-1)^{\ell}(s-1)^{(1-p)\ell}}{t(t-1)^{\ell-1}(s-1)^{(1-p)\ell}} = \left(1-\frac{\varepsilon}{2}\right)\cdot(t-1) \geq (1-\varepsilon)t \ .$$

Hence, x is not a counterexample to Equation (1).

Theorem 8 for general vectors with tree-shaped support. Fix a set $S \subseteq V_L$ such that the subgraph induced by $T := S \cup N(S)$ is a (t,s)-biregular tree. We generalize the above discussion of tree vectors by explaining why Equation (1) holds for any vector $x \in \mathbb{R}^n$ supported on S.

Given $r \in N(S)$, let v_r denote the parent of r in the tree T. In an overly optimistic scenario, if we could show that $|(Bx)_r| \approx |v_r|$ for all $r \in N(S)$, then we would be done, as each vertex $v \in S$ has $t-1$ children.[11] Each of these children then contributes $\approx |x_v|^p$ mass to $\|Bx\|_p^p$, so that $\|Bx\|_p^p \approx (t-1)\cdot\sum_{v\in S}|x_v|^p = (t-1)\cdot\|x\|_p^p$, implying Equation (1). As the tree vector case shows, one cannot, in fact, hope to guarantee $|(Bx)_r| \approx |v_r|$ for all $r \in N(S)$. Indeed, for a tree vector x we have $(Bx)_r = 0$ for any non-leaf $r \in N(S)$. Thus, a more delicate analysis is required.

For intuition, let us consider the viewpoint of an adversary seeking to construct an x supported on S such that $\|Bx\|_p^p$ is small. We shall think of the adversary as assigning values to $\{x_v\}_{v\in S}$ starting from the root, and then moving down the tree.

For each non-leaf $r \in N(S)$, let W_r denote the set of $s-1$ children of r. Recall that v_r is the parent of r. Note that $|(Bx)_r| \geq |x_{v_r}| - \sum_{u\in W_r}|x_u|$ due to the triangle inequality. Hence, when assigning values to the vertices in W_r, the adversary morally has two choices: (1) either

[10] And, indeed, if instead $s^{2-p} \lessapprox \frac{1}{\alpha}$, then we have $\|Bx\|_p = o(1)$, and this gives us Theorem 11.
[11] Except for the root, which has t children.

make $\sum_{u\in W_r}|x_u| \ll |x_{v_r}|$, in which case $|(Bx)_r| \approx |x_{v_r}|$, or (2) make $|(Bx)_r| \approx 0$, in which case $\sum_{u\in W_r}|x_u| \approx |x_{v_r}|$.[12] Let us fix some $\beta < 1$, and suppose that the adversary chooses values for $\{x_u\}_{u\in W_r}$ such that $\sum_{u\in W_r}|x_u| \le \beta|x_v|$, i.e., the adversary chooses Case (1). We then have that

$$|(Bx)_r| \ge \left(|x_{v_r}| - \sum_{u\in W_r}|x_u|\right) \ge (1-\beta)|x_{v_r}| \ ,$$

so $|(Bx)_r| \approx |x_{v_r}|$, which is what we wanted. Next, suppose that the adversary makes $\sum_{u\in W_r}|x_u| \ge \beta|x_v|$, i.e., the adversary chooses Case (2). Then, $|(Bx)_r|$ can be small, but applying Hölder's inequality, we have

$$\sum_{u\in W_r}|x_u|^p \ge \frac{\beta^p}{(s-1)^{p-1}}\cdot|x_v|^p \ .$$

Now, suppose that all children r of v_r have this property. Then, the total ℓ_p^p mass of all of the grandchildren of v_r must be at least $\frac{(t-1)\beta^p}{(s-1)^{p-1}}\cdot|x_v|^p \gg |x_v|^p$. We thus see that, intuitively, the adversary has merely pushed its task down to the grandchildren of v_r, and in doing so has not made any progress towards its overall goal. Indeed, this is precisely what happens in the case of a tree vector!

The above informal argument shows that the adversary does not "win" in either case. We can concretely capture this intuition via the following potential function:

$$a_r(x) := |(Bx)_r|^p + \Theta(1)\cdot\frac{(s-1)^{p-1}}{\varepsilon^{p-1}}\sum_{u\in W_r}|x_u|^p \ .$$

In the actual proof, this choice of $a_r(x)$ allows us to cleanly express the intuition that either $|(Bx)_r|$ is large or $\sum_{u\in W_r}|x_u|$ is large, and further extends beyond the "toy case" of tree-supported vectors.

Using expansion when $S\cup N(S)$ is not a tree. We now turn to the general case, where the subgraph induced by $S\cup N(S)$ (where $S = \mathrm{supp}(x)$) is not necessarily a tree. We observe that above, we are only using the tree structure to show that the rooted tree $S\cup N(S)$ trivially has a 1-to-$(t-1)$ "matching" with the following properties: (1) every vertex $r\in N(S)$ is matched with exactly one vertex $v\in S$, and (2) every vertex $v\in S$ is matched with at least $t-1$ vertices in $N(S)$. Indeed, when $S\cup N(S)$ is a tree, such a matching exists by matching each vertex $r\in N(S)$ with its parent v_r.

To generalize the above, we use the (unique) expansion of G to construct a similar matching that suffices for the proof. Recall that G is a (γ,μ)-unique expander, meaning that every set $S\subseteq V_L$ of size $\le \gamma n$ has at least $t(1-\mu)|S|$ unique neighbors, i.e., neighbors of a unique element of S. We construct the matching by "peeling off" vertices one at a time from S, each time matching a vertex with $\ge t(1-\mu)$ vertices in $N(S)$, namely its neighbors that are not neighbors of any of the remaining "unpeeled" vertices in S.

The above step can be viewed as extracting a "tree-like" subgraph from $S\cup N(S)$, where each vertex $v\in S$ has at least $t(1-\mu)$ "children" (the vertices it was matched with), and at most μt "parents" (its neighbors that it was *not* matched with). Each vertex $r\in N(S)$

[12] Note that the adversary has the third choice of setting $\sum_{u\in W_r}|x_u| \gg |x_{v_r}|$, but this is worse for the adversary.

still has exactly one "parent" and $\leq s-1$ "children". Once we have the above "tree-like" subgraph, the argument for trees goes through with only minor modifications, so this finishes the proof.

We note that the existence of this "tree-like" subgraph for any set S with $|S| \leq \gamma n$ immediately implies that G is a (γ, μ)-vertex expander, and hence a $(\gamma, 2\mu)$-unique expander. Thus, the existence of such a subgraph for every S of size at most γn is equivalent to unique expansion, up to a factor of 2 loss in the parameter μ.

Comparison with [6]. We briefly summarize the proof in [6] for the case of $p = 1$, and explain why their proof does not extend to the case of $p > 1$.

The proof in [6] proceeds as follows. For a vector x supported on S, let E_0 denote the set of edges between S and $N(S)$. First, they match each $r \in N(S)$ to its neighbor $v \in S$ with $|x_v|$ maximized. Let E_1 be the set of edges in this matching, and let $E_2 = E_0 \setminus E_1$. For any $r \in N(S)$, it then follows that $|(Bx)_r| \geq |x_v| - \sum_{u \in W_r} |x_u|$, where $(v, r) \in E_1$ and $W_r = \{u : (u, r) \in E_2\}$. Hence, $\|Bx\|_1 \geq \sum_{(v,r)\in E_1} |x_v| - \sum_{(u,r)\in E_2} |x_u|$. We observe that this step of the proof is specific to the ℓ_1 norm, and does not generalize to larger ℓ_p norms.

The main step in the proof is to argue that $\sum_{(v,r)\in E_2} |x_v| \leq t\varepsilon \|x\|_1$ using expansion. With this in hand, it immediately follows that $\sum_{(v,r)\in E_1} |x_v| \geq t(1-\varepsilon)\|x\|_1$, because $\sum_{(v,r)\in E_0} |x_v| = t\|x\|_1$ by regularity. It then follows that $\|Bx\|_1 \geq t(1-2\varepsilon)\|x\|_1$. Note that the upper bound $\|Bx\|_1 \leq t\|x\|_1$ is trivial, so this shows that B is ℓ_1-RIP.

One may attempt to generalize this proof to $p > 1$ by replacing $|x_v|$ with $|x_v|^p$. For example, using expansion it follows that $\sum_{(v,r)\in E_2} |x_v|^p \leq t\varepsilon \|x\|_p^p$, and as $\sum_{(v,r)\in E_0} |x_v|^p = t\|x\|_p^p$, we then have $\sum_{(v,r)\in E_1} |x_v|^p \geq t(1-\varepsilon)\|x\|_p^p$. But this is not enough to complete the proof, as it does *not* follow that $\|Bx\|_p^p \geq \sum_{(v,r)\in E_1} |x_v|^p - \sum_{(v,r)\in E_2} |x_v|^p$. Indeed, this is a fundamental barrier, and is the reason why our analysis for $p \geq 1$ proceeds by analyzing the "local" potential function $a_r(x)$, rather than the two "global" sums over E_1 and E_2 above.

2.3 Theorem 4: bounds on the singular values of A

We give a brief overview of the proof of Theorem 4. First, we observe that in order to bound the singular values of A, it suffices to bound the spectrum of $M := AA^\top - s \cdot I$, as each singular value of A is the square root of an eigenvalue of $AA^\top$. Note that M is a square matrix with an all-0 diagonal, by regularity of A.

Step 1: reducing to the nomadic walk matrix via a modified Ihara–Bass formula. The first step in the proof is to relate bounds on the spectrum of M to the spectral radius (i.e., maximum eigenvalue in absolute value) $\rho(B)$ of B, the *nomadic walk matrix* introduced in [27].[13] The nomadic walk matrix B is indexed by pairs of edges[14] (e_1, e_2) in G that form a length 2 non-backtracking walk in G, and its $((e_1, e_2), (e_1', e_2'))$-th entry is $\mathsf{sign}(e_1')\mathsf{sign}(e_2')$ if $e_1 \to e_2 \to e_1' \to e_2'$ forms a non-backtracking walk of length 4 in G, and is 0 otherwise. Note that B is *not* symmetric.

[13] We note that one could most likely also prove Theorem 4 using the standard nonbacktracking walk matrix and Ihara–Bass formula, e.g., with similar methods as in [9].

[14] In [27], the nomadic walk matrix is indexed by *directed* edges. In our context, this is equivalent to a length 2 oriented walk $e_1 \to e_2$, which is equivalent to a pair (e_1, e_2) of *undirected* edges, as the ordering in the pair gives the unique orientation $e_1 \to e_2$ in the walk.

▶ **Theorem 13** (Modified Ihara–Bass formula, Theorem 3.1 of [27], informal). *If* $\rho(B) \leq (1+o(1))\sqrt{(s-1)(t-1)}$*, then the spectrum* $\mathrm{Spec}(M)$ *of* M *satisfies:*

$$\mathrm{Spec}(M) \subseteq [t-2-2(1+o(1))\sqrt{(s-1)(t-1)}, t-2+2(1+o(1))\sqrt{(s-1)(t-1)}] \ .$$

The above theorem thus shows that it suffices to prove that $\rho(B) \leq (1+o(1))\sqrt{(s-1)(t-1)}$ with high probability.

We remark that bounds on the spectra of matrices of the form of M were studied in [27] for the case of $s = O(1)$. Unfortunately, this is insufficient to prove Theorem 4, as we wish to allow s to be any function of n (provided that $s \leq n^c$ for some absolute constant c). However, [27, Theorem 3.1] is a general statement that holds regardless of s, so we can make use of it in our setting.

Step 2: bounding $\rho(B)$ via the trace method by counting hikes. The natural approach to bound $\rho(B)$ is by applying the trace method. As the matrix B is not symmetric, we compute:

$$T := \mathbb{E}_{\mathsf{sign}}[\mathrm{tr}(B^\ell (B^\top)^\ell)] \ ,$$

where the expectation is taken over the function sign that determines the signs of the entries of A. By carefully expanding this expectation, one can show that the nonzero contributions to T roughly come from length $4(\ell-1)$ closed walks in G where (1) each edge in the walk appears an even number of times, and (2) the walk is non-backtracking, except possibly at the middle step in the walk. Such walks (of length 4ℓ) are commonly referred to as (2ℓ)-*hikes* [26].

To finish the proof of Theorem 4, we thus turn to obtaining a careful bound on the number of such walks.

Counting these walks requires extra care in our setting as our graph is bipartite, and so the bound needs to be sensitive to the difference in right/left degree. The counting of such hikes also differs greatly depending on whether $s \leq \mathrm{polylog}(n)$ or $s = \omega(\mathrm{polylog}(n))$.

Step 3: counting the number of hikes when $s \leq \mathrm{polylog}(n)$. [26, Section 3] counts the number of such hikes when $s = O(1)$, provided that G is bicycle-free at radius $O(\log n)$. Namely, any vertex v participates in at most one cycle of length $O(\log n)$. By repeating their proof, one can show that their bounds can be extended to the case when $s \leq \mathrm{polylog}(n)$. However, we still cannot use their bound on the number of such hikes naively, as their counting is for non-bipartite graphs and thus yields a bound of $m(1+o(1))^\ell (s-1)^{2\ell}$, simply because it treats left and right vertices the same, and the maximum degree of a vertex is s. We refine their approach to ensure that right and left vertices contribute roughly equally, which will yield the desired bound. One may, at first glance, be tempted to assume that this is trivial because a closed walk in a bipartite graph has an equal number of left and right vertices, but this is not the case, as we shall see.

We adopt the bookkeeping approach of [26]. We think of a hike as discovering the graph G as one traverses the hike. A step in a hike is *fresh* if uses an edge for the first time and ends at a previously undiscovered vertex; it is *boundary* if it uses an edge for the first time but ends up at an old vertex; finally, it is *stale* if it uses an old edge.

Because each edge must appear an even number of times, a hike can have at most 2ℓ fresh steps. Each fresh step "pays" a factor of $(s-1)$ (if we move from a right to a left vertex) or $(t-1)$ (if we move from a left to a right vertex) in our bound in the number of hikes, as this is the number of choices for the next vertex that the hike moves to. [26, Theorem 2.13]

implies that, since G is bicycle-free, the number of boundary steps is $\ll \ell$; they also show that one can bound the "number of choices" for the stale steps by some small factor, which we ignore here.

We need to augment the argument of [26] with the following addition: if a hike has c fresh steps, then the number of fresh steps c_R that start at a right vertex is $\approx \frac{c}{2}$, and similarly for c_L for left vertices. Note that by definition, $c = c_R + c_L$. The key observation is that a fresh step from a right (resp. left) vertex must be followed by either a fresh step from a left (resp. right) vertex, or by a boundary step. Indeed, after we take a fresh step we are at a previously unvisited vertex, so the next step must use a new edge; in particular, it cannot be *stale*. This implies that the deviation of each of c_L, c_R from $\frac{c}{2}$ is bounded by the number of boundary steps, which is $\ll \ell$.

This implies a bound of $m(1+o(1))^\ell (s-1)^\ell (t-1)^\ell$ on the number of hikes, provided that $s \leq \text{polylog}(n)$. The m comes from the number of start vertices in the hike, and the $(1+o(1))^\ell$ comes from the stale and boundary steps, as well as our new deviation term. This yields the desired bound for sparse s.

Step 4: counting the number of hikes when $s = \omega(\text{polylog}(n))$. For s this large, the graph G is "dense", and so it will not be bicycle-free at radius $O(\log n)$. This rules out the approach of [26], which relies on G being bicycle-free. Instead, we adapt a standard counting approach (for bounding the operator norm of a random $n \times n$ Gaussian matrix) given in [33, Section 2.3.6] to our bipartite setting. Our crucial observation here is to note that any hike can have at most ℓ distinct left vertices. As we pay a factor of $(s-1)$ every time we move to a left vertex, it then follows that the "power" of $(s-1)$ in our bound can only be at most ℓ. The standard counting argument of [33, Section 2.3.6] for Gaussian matrices then goes through, yielding the desired bound.

References

1 Zeyuan Allen-Zhu, Rati Gelashvili, and Ilya P. Razenshteyn. Restricted isometry property for general p-norms. In *31st International Symposium on Computational Geometry, SoCG 2015, June 22-25, 2015, Eindhoven, The Netherlands*, volume 34 of *LIPIcs*, pages 451–460. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2015.

2 Afonso S Bandeira and Ramon Van Handel. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Annals of Probability*, 44(4):2479–2506, 2016.

3 Boaz Barak, Fernando G. S. L. Brandão, Aram Wettroth Harrow, Jonathan A. Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 307–326. ACM, 2012.

4 Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the Restricted Isometry property for random matrices. Available at `https://www.math.tamu.edu/~rdevore/publications/131.pdf`, 2007.

5 Anirban Basak and Mark Rudelson. Invertibility of sparse non-hermitian matrices. *Advances in Mathematics*, 310:426–483, 2017.

6 R. Berinde, A. C. Gilbert, P. Indyk, H. Karloff, and M. J. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. In *2008 46th Annual Allerton Conference on Communication, Control, and Computing*, pages 798–805, 2008. `doi:10.1109/ALLERTON.2008.4797639`.

7 Charles Bordenave. A new proof of friedman's second eigenvalue theorem and its extension to random lifts. In *Annales scientifiques de l'Ecole normale supérieure*, 2019.

8 Charles Bordenave and Benoît Collins. Eigenvalues of random lifts and polynomials of random permutation matrices. *Annals of Mathematics*, 190(3):811–875, 2019.

9 Gerandy Brito, Ioana Dumitriu, and Kameron Decker Harris. Spectral gap in random bipartite biregular graphs and applications. *arXiv preprint*, 2018. arXiv:1804.07808.

10 Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59:1207–1223, 2006.

11 Emmanuel J. Candès and Terence Tao. Decoding by linear programming. *IEEE Trans. Inf. Theory*, 51(12):4203–4215, 2005.

12 Michael R. Capalbo, Omer Reingold, Salil P. Vadhan, and Avi Wigderson. Randomness conductors and constant-degree lossless expanders. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 659–668. ACM, 2002.

13 David L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006. doi:10.1109/TIT.2006.871582.

14 T. Figiel, J. Lindenstrauss, and V. D. Milman. The dimension of almost spherical sections of convex bodies. *Acta Math.*, 139(1-2):53–94, 1977.

15 R. G. Gallager. *Low-Density Parity-Check Codes*. MIT Press, 1963.

16 A. Garnaev and E. D. Gluskin. The widths of Euclidean balls. *Doklady An. SSSR.*, 277:1048–1052, 1984.

17 E. D. Gluskin. Norms of random matrices and diameters of finite-dimensional sets. *Mat. Sb. (N.S.)*, 120(162)(2):180–189, 286, 1983.

18 V. Guruswami, J. Lee, and A. Wigderson. Euclidean sections of with sublinear randomness and error-correction over the reals. In *12th International Workshop on Randomization and Combinatorial Optimization: Algorithms and Techniques (RANDOM)*, pages 444–454, 2008.

19 Venkatesan Guruswami, James R. Lee, and Alexander A. Razborov. Almost Euclidean subspaces of ℓ_1^N via expander codes. *Combinatorica*, 30(1):47–68, 2010. doi:10.1007/s00493-010-2463-9.

20 Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of the ACM*, 53(3):307–323, 2006.

21 Piotr Indyk. Uncertainty principles, extractors, and explicit embeddings of L_1 into L_2. In *Proceedings of the 39th Annual ACM Symposium on the Theory of Computing*, pages 615–620, 2007.

22 Piotr Indyk and Stanislaw J. Szarek. Almost-euclidean subspaces of ℓ_1^n via tensor products: A simple approach to randomness reduction. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 13th International Workshop, APPROX 2010, and 14th International Workshop, RANDOM 2010, Barcelona, Spain, September 1-3, 2010. Proceedings*, volume 6302 of *Lecture Notes in Computer Science*, pages 632–641. Springer, 2010.

23 Zohar Shay Karnin. Deterministic construction of a high dimensional l_p section in l_1^n for any $p<2$. In Lance Fortnow and Salil P. Vadhan, editors, *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 645–654. ACM, 2011. doi:10.1145/1993636.1993722.

24 B. S. Kashin. The widths of certain finite-dimensional sets and classes of smooth functions. *Izv. Akad. Nauk SSSR Ser. Mat.*, 41(2):334–351, 478, 1977.

25 Boris S Kashin and Vladimir N Temlyakov. A remark on compressed sensing. *Mathematical notes*, 82(5):748–755, 2007.

26 Sidhanth Mohanty, Ryan O'Donnell, and Pedro Paredes. Explicit near-ramanujan graphs of every degree. In *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 510–523. ACM, 2020.

27 Sidhanth Mohanty, Ryan O'Donnell, and Pedro Paredes. The SDP value for random two-eigenvalue csps. In *37th International Symposium on Theoretical Aspects of Computer Science, STACS 2020, March 10-13, 2020, Montpellier, France*, volume 154 of *LIPIcs*, pages 50:1–50:45. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

28 Jonathan Mosheiff, Nicolas Resch, Noga Ron-Zewi, Shashwat Silas, and Mary Wootters. LDPC codes achieve list decoding capacity. In *Proceedings of the 61st IEEE Annual Symposium on Foundations of Computer Science*, pages 458–469, 2020.

29 Ryan O'Donnell and Xinyu Wu. Explicit near-fully X-ramanujan graphs. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 1045–1056. IEEE, 2020. doi:10.1109/FOCS46700.2020.00101.

30 Tom Richardson and Ruediger Urbanke. *Modern Coding Theory*. Cambridge University Press, USA, 2008.

31 Michael Sipser and Daniel A. Spielman. Expander codes. *IEEE Trans. Inform. Theory*, 42(6, part 1):1710–1722, 1996. Codes and complexity.

32 Robert M. Tanner. A recursive approach to low complexity codes. *IEEE Transactions on Information Theory*, 27(5):533–547, 1981.

33 Terence Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

34 Yizhe Zhu. On the second eigenvalue of random bipartite biregular graphs. *arXiv preprint*, 2020. arXiv:2005.08103.