Electronic Colloquium on Computational Complexity, Report No. 151 (2024)

Inapproximability of Sparsest Vector in a Real Subspace

Vijay Bhattiprolu *
vbhattip@uwaterloo.ca

Euiwoong Lee †
euiwoong@umich.edu

Abstract

We establish strong inapproximability for finding the sparsest nonzero vector in a real subspace (where sparsity refers to the number of nonzero entries). Formally we show that it is NP-Hard (under randomized reductions) to approximate the sparsest vector in a subspace within any constant factor. By simple tensoring the inapproximability factor can be made almost polynomial assuming NP cannot be solved in randomized quasipolynomial time. We recover as a corollary state of the art inapproximability factors for the shortest vector problem (SVP), a foundational problem in lattice based cryptography. Our proof is surprisingly simple, bypassing even the PCP theorem.

We are inspired by the homogenization framework developed over the course of several papers studying the inapproximability of minimum distance problems in integer lattices and error correcting codes. Our proof uses a combination of (a) *product testing via symmetric tensor codes* and (b) *encoding points of the hypercube as cosets of a random code in higher dimensional space* in order to embed an instance of non-homogeneous quadratic equations into the sparsest vector problem. (a) is inspired by Austrin and Khot's simplified proof of hardness of minimum distance of a code over finite fields, and (b) is inspired by Micciancio's partial derandomization of the hardness of SVP.

Our reduction involves the challenge of performing (a) over the reals. We prove that symmetric tensoring of the kernel of a +1/-1 random matrix furnishes an adequate product test (while still allowing (b)). The proof exposes a connection to Littlewood-Offord theory and relies on a powerful anticoncentration result of Rudelson and Vershynin.

Our main motivation in this work is the development of inapproximability techniques for problems over the reals. Analytic variants of sparsest vector have connections to small set expansion, quantum separability and polynomial maximization over convex sets, all of which cause similar barriers to inapproximability. The approach we develop could lead to progress on the hardness of some of these problems.

*Supported by NSERC Discovery Grant 2023-03599
†Supported in part by NSF grant CCF-2236669 and Google

ISSN 1433-8092

1 Introduction

Finding the sparsest nonzero vector in a real subspace is a classical and well studied computational task. It has many connections to machine learning and optimization, via topics like compressed sensing, robust subspace recovery, dictionary learning, and sparse blind deconvolution [HM13a, BKS14, BKS15], and has been actively studied in the nonconvex optimization literature [QSW14, QZL+20]. It is also closely related to central questions in complexity theory and quantum information [BBH+12, HM13b].

We establish strong inapproximability for this problem. Let $|x|_0$ denote the number of nonzero entries in x. Our main result is the following theorem.

Theorem 1.1. *Unless* $\mathbf{NP} \subseteq \mathbf{BPP}$*, for any constant* $c > 0$*, given a linear subspace* $V \subseteq \mathbb{R}^n$ *and* $d \in \mathbb{N}$*, no polynomial-time algorithm distinguishes apart the following cases.*

(YES) There exists a nonzero $x \in V \cap \{0,1\}^n$ *with* $|x|_0 \leq d$*.*

(NO) Every $x \in V \setminus \{0\}$ *satisfies* $|x|_0 \geq c \cdot d$*.*

Furthermore, unless $\mathbf{NP} \subseteq \cup_{k\in\mathbb{N}}\mathbf{BPTIME}(2^{\log^k n})$*, for any constant* $\varepsilon > 0$*, no polynomial-time algorithm distinguishes apart the following cases.*

(YES) There exists a nonzero $x \in V \cap \{0,1\}^n$ *with* $|x|_0 \leq d$*.*

(NO) Every $x \in V \setminus \{0\}$ *satisfies* $|x|_0 \geq 2^{\log^{1-\varepsilon} n} d$*.*

Previously, only NP-hardness of exact optimization was known [McC83, CP86]. The best known approximation algorithm achieves an $O(n/\log n)$-approximation [BK01].

The booleanity property of the (YES) case in our inapproximability result allows us to recover as an immediate corollary state of the art inapproximability factors for the Shortest Vector Problem (SVP) in lattices, a foundational problem in lattice-based cryptography [Kho05, HR07, Mic12].

Corollary 1.2. *Fix* $p \in (0, \infty)$*. Unless* $\mathbf{NP} \subseteq \mathbf{BPP}$*, for any constant* $c > 0$*, given a lattice* $L \subseteq \mathbb{Z}^n$ *and* $d \in \mathbb{N}$*, no polynomial-time algorithm distinguishes apart the following cases.*

(YES) There exists a nonzero vector $x \in L \cap \{0,1\}^n$ *with* $\|x\|_p^p = d$*.*

(NO) Every $x \in L \setminus \{0\}$ *satisfies* $\|x\|_p^p \geq cd$*.*

Furthermore, unless $\mathbf{NP} \subseteq \cup_{k\in\mathbb{N}}\mathbf{BPTIME}(2^{\log^k n})$*, for any constant* $\varepsilon > 0$*, no polynomial-time algorithm distinguishes apart the following cases.*

(YES) There exists a nonzero vector $x \in L \cap \{0,1\}^n$ *with* $\|x\|_p^p = d$*.*

(NO) Every $x \in L \setminus \{0\}$ *satisfies* $\|x\|_p^p \geq 2^{\log^{1-\varepsilon} n} d$*.*

Inapproximability in the $p \in (0, 1)$ regime above appears to be new. The proof of Theorem 1.1 is surprisingly simple, bypassing even the PCP theorem. It is inspired by the *homogenization framework* developed in [Ajt98, Mic01, DMS03, Kho05, CW09, Mic12, AK14] for lattices and error-correcting codes over finite fields. Inapproximability for these homogeneous problems (like SVP or minimum distance) may appear deceptively simple, however they have involved the overcoming of technical and conceptual barriers. In brief, the paradigm that has emerged from these works is to start with hardness of a *non-homogeneous problem* (whose hardness is easier to establish) and reduce it to the desired homogeneous problem using coding theoretic gadgets (e.g., locally dense codes/lattices). Pioneered in [Ajt98, Mic01] for lattices, this framework finally led to strong hardness for finite fields in [DMS03] and for lattices in [Kho05].

Motivation. Our main motivation in this work is the development of techniques to prove inapproximability results for problems over the reals. Sparsest vector in a subspace is one of a long list of problems (polynomial maximization over the sphere, quantum separability, maximize $\|\cdot\|_q/\|\cdot\|_2$ (an analytic notion of sparsity when $q > 2$) over a subspace (henceforth, $2 \to q$ sparsest vector), small set expansion, densest k-subgraph, sparse PCA, low rank matrix completion, tensor PCA/rank, etc.) that are bottlenecks for inapproximability results. A common theme in these problems is resistance to local gadget reductions from CSPs. Informally this is because extremely sparse solutions have high objective value, and it is unclear how to decode good assignments to the variables of a CSP from such solutions. We believe our findings in this work can be considered as progress towards this goal.

We further substantiate this by highlighting some concrete connections analytic variants of sparsest vector have. It was shown in [BBH+12] that the small set expansion of a graph G (an important problem at the heart of inapproximability) can be cast as finding the sparsest vector that is close (in ℓ_2 norm) to the top eigenspace of G. It is also shown [HM13b, BBH+12] that the hardness of approximately computing the $2 \to 4$ sparsity of a subspace is closely related to **QMA** $=$ **NEXP**, which is a long standing open problem in quantum information.

Quadratic maximization over convex sets is a rich and expressive family of continuous optimization problems. In [BLN21], it was shown that inapproximability of $2 \to q$ sparsest vector (for all $q > 2$) implies the inapproximability of quadratic maximization over any convex set with large type-2 value (informally type-2 is a quantitative notion of smoothness). All such optimization problems resist local gadget reductions from CSPs. It was also shown in [BLN21] that when the convex body has constant type-2 value and additionally admits a certain approximate separation oracle, the problem admits a polytime constant factor approximation. Thus NP-hardness of approximating $2 \to q$ sparsest vector would lead to a near-characterization of the approximability of this wide-ranging class under $\mathbf{P} \neq \mathbf{NP}$.

We believe our work provides a promising new line of attack on hardness of $2 \to q$ sparsest vector. The reduction we devise to prove Theorem 1.1 admits natural candidate analogues in the $2 \to q$ setting. Proving analytic versions of our product test using tensor codes and of our quotienting step would imply correctness of such a reduction.

Derandomization of SVP Hardness. A final point we highlight about our reduction is that it provides a new candidate approach towards derandomization of SVP, which is a long running pursuit in the homogenization literature (see [Mic12, BP22] and discussions therein). To derandomize our reduction, it suffices to construct a code with polynomial rate and minimum distance (in the hamming sense), support overlap property of codewords, and finally the quotienting property for embedding the hypercube. We raise the question of whether one can deterministically construct codes with all three properties above.

2 Techniques and Preliminaries

Product Testing via Symmetric Tensor Codes. Minimum Distance of Code (MDC) is the problem of finding the sparsest vector in a subspace over a finite field, whose hardness is shown via various instantiations of the homogenization framework [DMS03, CW09, AK14]. In particular, [AK14] show how over $\mathbb{F}_2$, MDC can be used to test if a matrix is rank-1 or not. They do this by starting with any code of good minimum distance, and considering the subspace of matrices given by the symmetric tensor product of the code with itself. [AK14] show that the minimum distance codewords in this code have rank one and any codewords of higher rank have higher distance by a multiplicative factor of 3/2. Using this product test, one can then embed non-homogeneous (and even nonlinear) optimization problems into the minimum distance problem.

We extend this idea to the reals, where it appears good minimum distance no longer suffices. Given a real code with the additional property that every pair of linearly independent codewords with small sparsity don't overlap much in support (henceforth referred to as the bounded codeword overlap property), we show that the sparsest codewords in the symmetric tensor product of the code with itself must have rank one. Furthermore we show that the kernel of a random Rademacher matrix satisfies the bounded overlap property along with adequate minimum distance, thereby furnishing a product test that embeds within the sparsest vector problem.

For a vector x let σ^x denote its support, and let $\mathbb{S}_d^N$ denote the set of N-dimensional unit vectors with support size at most d. All of the above random subspace properties are covered in the following theorem which we prove in Section 4.

Theorem 2.1. *(Minimum Distance and Codeword Overlap for a Random Rademacher Subspace)* *Fix any $\varepsilon \in (0, 1/2)$. For any $d \in \mathbb{N}$ sufficiently large, and any integer $N \in [d^{C_2/\varepsilon}, e^{d^{0.9}}]$, an $h \times N$ random matrix R with $h = \lceil 2(1+\varepsilon) d \log_d N \rceil$ and with i.i.d. ± 1-random entries satisfies with probability at least $1 - 1/N^{c_3 d/\log d}$,*

(a) $\mathbb{S}_d^N \cap \ker(R) = \emptyset$ *(Code Distance)*

(b) $\forall \ell \in \mathbb{N} \cap \left[1, \frac{\varepsilon}{C_2} \log_d N\right]$, $\text{Overlap}_\ell \cap \ker(R)^\ell = \emptyset$ *where,* *(Bounded Codeword Overlap)*

$$\text{Overlap}_\ell := \left\{(u_1, \ldots, u_\ell) \in \mathbb{R}^{N \times \ell} \;\middle|\; \text{rk}\{u_1, \ldots, u_\ell\} = \ell,\ \left|\textstyle\bigcup_{i \in [\ell]} \sigma^{u_i}\right| \leq \ell d\right\}$$

and where $C_2 > 1$, $c_3 \in (0, 1)$ are universal constants.

We prove the above theorem by making use of a celebrated result of [RV09] regarding the Littlewood-Offord problem (see also [RV08] for a one-dimensional precursor), which establishes a precise connection between approximate arithmetic structure of a collection of vectors $a_1, a_2 \ldots$ and the probability that a randomly signed sum of vectors $\sum_i \xi_i a_i$ lies in a ball of small radius. They give a beautiful proof of this result using harmonic analysis.

Embedding a Solution Set into a Quotient of a Random Subspace. In the full reduction of [AK14], the product test and non-homogenous hard problem (Max-NAND in their case) are coupled in an intricate manner. We are unable to extend this reduction to the reals.

Instead, we encode the solution set of our starting non-homogeneous optimization problem as cosets of a random code in a higher dimensional space. We need the additional property that the encodings of any two solution vectors have approximately equal weight, where weight of a coset refers to the sparsest vector in the coset. This component of our reduction is inspired by [Mic12], and it allows us to compose our product test with a non-homogenous optimization problem in a decoupled and modular fashion. In Section 3 we reduce from boolean quadratic equations whose solution set is the hypercube. Below we state formally what we require of the embedding of the hypercube.

We define the weight d slice of the N dimensional hypercube respectively as $H_d^N := \{x \in \{0,1\}^N \mid |x|_0 = d\}$.

Lemma 2.2. *(Embedding of $\{0,1\}^n$ in a Quotient of $\ker(R) \cap H_k^N$)* *Fix any $\varepsilon \in (0, 1/2)$. Let $n \in \mathbb{N}$, and set $k := 16^2 \lceil n^{2.2}/\varepsilon^2 \rceil$, $d := k(1-\varepsilon)/(1+\varepsilon)$, $h := \lceil 2(1+\varepsilon) d \log_d N \rceil$ and finally pick any integer $N \geq k^{5/\varepsilon}$. Let R be an $h \times N$ matrix with i.i.d. ± 1 random entries and let $T \in \{0,1\}^{n \times N}$ be such that each entry is set to 1 independently at random with probability $p = 1/(4kn)$. Then assuming k is sufficiently large, with probability at least $1 - 6/n^{0.1} - 1/N^{0.4}$, for every $y \in \{0,1\}^n$, there exists $x \in \ker(R) \cap H_k^N$ such that $T(x) = y$.*

3 PCP-Free Inapproximability of Sparsest Vector

We present our basic reduction. Our starting point is hardness of exactly solving a non-homogeneous system of quadratic equations. The proof uses an elementary reduction from the NP-hardness of 1-in-3-SAT [Sch78] and appears in Appendix A.

Proposition 3.1. *Given n variables $y_1, \ldots, y_n$ and m quadratic equations $(\sum_{i,j\in[n]} \alpha_{\ell,i,j} y_i y_j = b_\ell)_{\ell\in[m]}$, it is NP-hard to distinguish between the following two cases:*
(YES) There exists $y \in \{0,1\}^n$ satisfying all m equations.
(NO) There doesn't exist $y \in \mathbb{R}^n$ satisfying all m equations.

We next present our basic reduction that generates a constant multiplicative gap.

Reduction from non-homogeneous quadratic equations to sparsest vector in a subspace. Given the system of m_0 quadratic equations $(\sum_{i,j\in[n_0]} \alpha_{\ell,i,j} y_i y_j = b_\ell)_{\ell\in[m_0]}$ on n_0 variables and a parameter $\varepsilon \in (0, 1/2)$, let $k := 16^2 \lceil n_0^{2.2}/\varepsilon^2 \rceil$, $d := k(1-\varepsilon)/(1+\varepsilon)$, $N := \lceil k^{5/\varepsilon} \rceil$, and $h := \lceil 2(1+\varepsilon) d \log_d N \rceil$. Let $R \in \mathbb{R}^{h\times N}, T \in \mathbb{R}^{n_0\times N}$ be random matrices such that R is chosen to be an $h \times N$ matrix with i.i.d. ± 1 random entries and $T \in \{0,1\}^{n_0\times N}$ is chosen such that each entry is set to 1 independently at random with probability $p = 1/(4kn_0)$.

Remark. The above parameters are chosen precisely so that we may apply Theorem 2.1 and Lemma 2.2.

Definition of Subspace. Now we present the subspace $V \subseteq \mathbb{R}^{N\times N}$. For any $\mathbb{X} \in \mathbb{R}^{N\times N}$ and any $i \in [N]$, $\mathbb{X}_i$ denotes the i-th row of $\mathbb{X}$ (represented as a column vector) and $\mathbb{X}^i$ denotes the i-th column of $\mathbb{X}$. The linear subspace V is the set of matrices $\mathbb{X} \in \mathbb{R}^{N\times N}$ such that there exist $\mathbb{Y} \in \mathbb{R}^{n_0\times n_0}$ and $z \in \mathbb{R}$ satisfying the following system of homogeneous linear equations in $\mathbb{X}, \mathbb{Y}, z$.

$$\begin{aligned}
&R\mathbb{X}^i = 0, \;\; R\mathbb{X}_i = 0 \quad \forall i \in [n_0] \\
&\mathbb{X}^* = \mathbb{X} \\
&z = \sum_{i\in[n_0]} \mathbb{X}[i,i]/k \\
&\mathbb{Y} = T\mathbb{X}T^* \\
&\Big(\sum_{i,j\in[n_0]} \alpha_{\ell,i,j} \mathbb{Y}[i,j]\Big) - z b_\ell = 0 \quad \forall \ell \in [m_0]
\end{aligned} \tag{1}$$

Eq. (1) refers to the entire system rather than just the last line.

Remark. Due to symmetry ($\mathbb{X}^* = \mathbb{X}$), adding both constraints $R\mathbb{X}^i = 0$, $R\mathbb{X}_i = 0$ is redundant. However we include these implied constraints to emphasize the importance of both the rows and the columns of $\mathbb{X}$ lying inside the kernel of R.

We next prove the validity of theabove reduction.

Proposition 3.2. *For any constant $\varepsilon \in (0, 1/2)$, given a linear subspace $V \subseteq \mathbb{R}^n$ and $k \in \mathbb{N}$, unless* **NP** $\subseteq$ **BPP***, no polynomial-time algorithm distinguishes between the following two cases.*
(YES) There exists a nonzero $x \in V \cap \{0,1\}^n$ with $|x|_0 \le k$.
(NO) Every $x \in V \setminus \{0\}$ has $|x|_0 \ge (2-\varepsilon)k$.

Proof. We use the above reduction from solving a system of quadratic equations whose hardness is proved in Proposition 3.1.

By Theorem 2.1 and Lemma 2.2, R and T satisfy the following properties with probability $1 - o(1)$, (that we will invoke shortly):

(Distance) : Every nonzero $x \in \ker(R)$ satisfies $|x|_0 \geq d$.
(Overlap) : For all $x, y \in \ker(R)$ that are linearly independent, $|\sigma^x \cup \sigma^y| \geq 2d$.
(Shattering) : $\{0,1\}^{n_0} \subseteq T(\ker(R) \cap H_k^N)$.

Completeness. Let $y \in \{0,1\}^{n_0}$ be a solution to the system of quadratic equations $(\sum_{i,j \in [n_0]} \alpha_{\ell,i,j} y_i y_j = b_\ell)_{\ell \in [m_0]}$. By property (Shattering), there exists $x \in \{0,1\}^N \cap \ker(R)$ with $|x|_0 = k$ such that $Tx = y$. Then it is easy to check that $\mathbb{Y} = yy^*$, $\mathbb{X} = xx^*$, $z = 1$ is a solution of Eq. (1) which further satisfies $|\mathbb{X}|_0 = k^2$. Therefore, $\mathbb{X}$ is a nonzero 0-1 vector in V whose sparsity is k^2.

Soundness. Let $\mathbb{X} \in V$ be nonzero with $|\mathbb{X}|_0 < 2d^2$. Then there exist $\mathbb{Y}$ and z such that $(\mathbb{X}, \mathbb{Y}, z)$ is a nonzero solution of Eq. (1). We will show that the system of quadratic equations admits a solution over reals.

If $\mathbb{X}$ has rank at least 2, then it has at least two linearly independent nonzero columns in $\ker(R)$. By (Overlap), these two columns certify that at least $2d$ rows of $\mathbb{X}$ are nonzero, and thus (Distance) implies that $|\mathbb{X}|_0 \geq 2d^2 = (2 - O(\varepsilon))k^2$.

Therefore, $\mathbb{X}$ has rank 1. Since $\mathbb{X}$ is symmetric, we conclude $\mathbb{X} = xx^*$ for some nonzero $x \in \mathbb{R}^N$. The constraint $z = \sum_i \mathbb{X}[i,i]/k$ implies that $z > 0$, and the constraint $\mathbb{Y} = T\mathbb{X}T^*$ implies that $\mathbb{Y} = yy^*$ where $y = Tx$. Then the final equation of Eq. (1) implies that for every $\ell \in [m_0]$,

$$\Big(\sum_{i,j \in [n_0]} \alpha_{\ell,i,j} y_i y_j\Big) = z b_\ell,$$

and thus $y/\sqrt{z}$ is a solution to the system of quadratic equations. ∎

Simple tensoring then implies our main result. Moreover the booleanity property of the completeness solution is preserved under tensoring. Given $V \subseteq \mathbb{R}^n$, let $V \otimes V \subseteq \mathbb{R}^{n \times n}$ denote the subspace such that $\mathbb{X} \in V \otimes V$ if and only if every row and column of $\mathbb{X}$ (viewed as an $n \times n$ matrix) lies in V.

Fact 3.3. *If there exists $x \in V \cap H^n_{d'}$, then $xx^T \in V \otimes V$ with $|x|_0 = (d')^2$. On the other hand, if any $x \in V \setminus \{0\}$ satisfies $|x|_0 \geq d'$, then for any $\mathbb{X} \in V \otimes V \setminus \{0\}$, $|\mathbb{X}|_0 \geq (d')^2$.*

Proof. For the second statement, if $\mathbb{X} \neq 0$, at least one row is nonzero. Since that row is in V, there are at least d' nonzero columns. Applying the same argument for each column implies that $|\mathbb{X}|_0 \geq (d')^2$. ∎

Using this we obtain

Theorem 1.1. *Unless* **NP** $\subseteq$ **BPP**, *for any constant $c > 0$, given a linear subspace $V \subseteq \mathbb{R}^n$ and $d \in \mathbb{N}$, no polynomial-time algorithm distinguishes apart the following cases.*

(YES) There exists a nonzero $x \in V \cap \{0,1\}^n$ with $|x|_0 \leq d$.
(NO) Every $x \in V \setminus \{0\}$ satisfies $|x|_0 \geq c \cdot d$.

Furthermore, unless **NP** $\subseteq \cup_{k \in \mathbb{N}}$**BPTIME**$(2^{\log^k n})$, *for any constant $\varepsilon > 0$, no polynomial-time algorithm distinguishes apart the following cases.*

(YES) There exists a nonzero $x \in V \cap \{0,1\}^n$ with $|x|_0 \leq d$.
(NO) Every $x \in V \setminus \{0\}$ satisfies $|x|_0 \geq 2^{\log^{1-\varepsilon} n} d$.

Proof. Let $V \subseteq \mathbb{R}^n$ and $d \in \mathbb{N}$ be an instance given by Proposition 3.2 with $\varepsilon = 0.1$. Let $\mathbb{V} = V^{\otimes t}$ and $d = 1.9^t$ for some $t \in \mathbb{N}$ fixed later. The new ambient dimension is $N = n^t$. Applying Fact 3.3, one can conclude that in the (YES) case of Proposition 3.2, $\mathbb{V} \cap H^N_{d^t} \neq \emptyset$, and in the (NO) case, every nonzero vector in $\mathbb{V}$ has sparsity at least $(1.9d)^t$.

Setting t to be an arbitrarily large constant implies the first statement, and setting $t = \log^{O(1/\varepsilon)} n$ implies the second statement as the gap is given by $1.9^t = 2^{\log^{1-\varepsilon} N}$. ∎

Our strong completeness and soundness conditions imply hardness of SVP as an immediate corollary, restated below.

Corollary 1.2. *Fix $p \in (0, \infty)$. Unless $\mathbf{NP} \subseteq \mathbf{BPP}$, for any constant $c > 0$, given a lattice $L \subseteq \mathbb{Z}^n$ and $d \in \mathbb{N}$, no polynomial-time algorithm distinguishes apart the following cases.*
(YES) There exists a nonzero vector $x \in L \cap \{0,1\}^n$ with $\|x\|_p^p = d$.
(NO) Every $x \in L \setminus \{0\}$ satisfies $\|x\|_p^p \geq cd$.

Furthermore, unless $\mathbf{NP} \subseteq \cup_{k \in \mathbb{N}} \mathbf{BPTIME}(2^{\log^k n})$, for any constant $\varepsilon > 0$, no polynomial-time algorithm distinguishes apart the following cases.
(YES) There exists a nonzero vector $x \in L \cap \{0,1\}^n$ with $\|x\|_p^p = d$.
(NO) Every $x \in L \setminus \{0\}$ satisfies $\|x\|_p^p \geq 2^{\log^{1-\varepsilon} n} d$.

Proof. Given an instance (V, d) of Theorem 1.1, define the lattice $L := V \cap \mathbb{Z}^n$. In the (YES) case, there exists $x \in L \cap H^N_d$, and so $\|x\|_p^p = d$. In the (NO) case, if $x \in L \setminus \{0\}$ satisfies $|x|_0 \geq \alpha d$ for some gap parameter α, since every nonzero coordinate has absolute value ≥ 1, we must have $\|x\|_p^p \geq \alpha d$. ∎

4 Properties of Rademacher Random Subspaces

4.1 Minimum Distance and Bounded Codeword Overlap

Throughout this section, let $\xi_1, \ldots, \xi_d$ denote i.i.d. Rademacher (± 1) random variables. For a vector $u \in \mathbb{R}^N$, σ^u denotes its support. In this section we borrow heavily from the exposition, ideas and notions in [RV09, RV08].

Small Ball Probability and LCD.

Definition 4.1. *The* Lévy concentration function *of a random vector Ξ in $\mathbb{R}^\ell$ is defined for $t > 0$ as*

$$\mathcal{L}(\Xi, t) := \sup_{v \in \mathbb{R}^\ell} \mathbb{P}\left[\|\Xi - v\|_2 \leq t\right].$$

Following [RV09], LCD is defined as

$$\mathrm{LCD}_{\alpha,\gamma}(a) := \inf\left\{\theta > 0 \mid \mathrm{dist}(\theta \cdot a, \mathbb{Z}^N) < \min(\gamma \|\theta \cdot a\|_2, \alpha)\right\}.$$

It can be thought of as a measure of arithmetic structure. Let $E \subset \mathbb{R}^N$ be a subspace. We define

$$\mathrm{LCD}_{\alpha,\gamma}(E) := \inf_{a \in \mathbb{S}(E)} \mathrm{LCD}_{\alpha,\gamma}(a).$$

where $\mathbb{S}(E)$ denotes the euclidean sphere restricted to the subspace E.

The following theorem which connects multidimensional small ball probability of a signed sum of vectors to the LCD of their Rowspace, is the main workhorse of our proof of the overlap property.

Theorem 4.2 (Multidimensional Small ball probability, Theorem 3.3 of [RV09])**.**
Consider a sequence of orthonormal vectors $u_1, \dots, u_\ell \in \mathbb{S}^{N-1}$. Let $u^1, \dots u^N \in \mathbb{R}^\ell$ be the columns of the $\ell \times N$ matrix whose rows are formed by $u_1, \dots u_\ell$. Let $\xi_1, \dots, \xi_N$ be i.i.d. Rademacher random variables and let $\Xi := \sum_{j=1}^N \xi_j \cdot u^j$ be a sum of randomly signed vectors. Then for any $\alpha > 0$, $\gamma \in (0,1)$, and $t \geq \sqrt{\ell}/\mathrm{LCD}_{\alpha,\gamma}(\mathrm{span}\{u_1, \dots, u_\ell\})$, we have

$$\mathcal{L}(\Xi, t\sqrt{\ell}) \leq \left(\frac{Ct}{\gamma}\right)^\ell + C^\ell e^{-\alpha^2}.$$

where C is a universal constant.

Remark 4.3. *Above, we specialized their theorem to the case of orthonormal vectors. For a reader interested in the details of the specialization, see the proof of Theorem 4.2 in [RV09].*

Restricted Isometry. The following restricted isometry result is well known in related parameter regimes. In order to make parameter choices that are compatible with the proof of Theorem 2.1, we derive it here again using off-the-shelf concentration estimates for the maximum and minimum singular values of a random matrix. For a matrix $A \in \mathbb{R}^{h \times N}$ and $k \in \mathbb{N}$, let

$$\sigma_{min}^k(A) := \min_{|T| \leq k, |T| \subseteq [N]} \sigma_{min}(A_T^* A_T)^{1/2}$$
$$\sigma_{max}^k(A) := \max_{|T| \leq k, |T| \subseteq [N]} \sigma_{max}(A_T),$$

where A_T denotes the column submatrix of A restricted to T.

Proposition 4.4 (R is an Isometry when restricted to any subset of $\Theta(d/\log d)$ Columns)**.**
Let $d, \ell, N \in \mathbb{N}$, $\eta \in [1, 3/2]$ and let $h := 2\eta d \log N$. There are universal constants $C_0 > 1, c_0 \in (0,1)$, such that an $h \times N$ random matrix R with i.i.d. ± 1 entries satisfies with probability at least $1 - N^{-\ell d}$ that $\sigma_{max}^{\ell d}(R) \leq C_0 \sqrt{\ell d \log N}$. Moreover $\sigma_{min}^{\overline{d}}(R) \geq c_0 \sqrt{h}$ with probability at least $1 - N^{-\overline{d}}$, where $\overline{d} := \lceil c_0 d / \log d \rceil$.

Proof. By Theorem 5.39 in [Ver10], there is a universal constant $C_0 > 1$ such that for any fixed $T \in \binom{[N]}{\ell d}$,

$$\mathbb{P}\left[\sigma_{max}(R_T) > C_0 \sqrt{\ell d \log N}\right] \leq N^{-2\ell d}.$$

Taking union bound over all choices of T implies the first claim.

It is shown in [BDJ+77], that there is a universal constant $c_0 \in (0,1)$ such that for any fixed $\overline{T} \in \binom{[N]}{\overline{d}}$,

$$\mathbb{P}\left[\sigma_{min}(R_{\overline{T}}) < c\sqrt{h}\right] \leq N^{-2\overline{d}}.$$

where $\overline{d} := \lceil c_0 d / \log d \rceil$. Taking union bound over $\binom{N}{\overline{d}}$ choices of $\overline{T}$ implies the second claim. ∎

Standard Net Size Estimates. We use the following standard facts about ε-nets. Recall that when T is a metric space with distance d and let $E \subset T$, $\mathcal{N} \subseteq T$ is called an ε-net of E if for every $x \in E$ there exists $y \in \mathcal{N}$ such that $d(x,y) \le \varepsilon$.

Fact 4.5. *Let T be a metric space and let $E \subset T$. Let $\mathcal{N} \subset T$ be a t-net of the set E. Then there exists a $(2t)$-net $\mathcal{N}'$ of E whose cardinality does not exceed that of $\mathcal{N}$, and such that $\mathcal{N}' \subset E$.*

Fact 4.6. *There is a t-net (in ℓ_2 norm) within $\mathbb{S}^{d-1}$, of size at most $(\frac{6}{t})^d$.*

Partitioning Space for Chaining according to Small ball Probability. We define $\mathbb{S}^N_d$ to be the union of all subspheres of $\mathbb{R}^N$ supported on d coordinates. We then partition $\mathbb{S}^N_d$ into compressible and incompressible vectors. This paritioning is required since compressible vectors can have relatively high small ball probability. The saving grace here is that portion of $\mathbb{S}^N_d$ that is compressible is much smaller than the incompressible portion.

Define for any $\ell \in \mathbb{N}$ and any $\delta, \rho \in (0,1)$ the following sets:

$$\mathbb{S}^N_d := \{u \in \mathbb{R}^N : \|u\|_2 = 1,\ |u|_0 \le d\}$$

$$\mathrm{Comp}^{(\ell)}_{\rho,\delta} := \left\{u \in \mathbb{S}^N_{\ell d} \mid \exists \overline{T} \subseteq \sigma^u, |\overline{T}| \le \delta \ell d, \text{ s.t. } \|u_{\sigma^u \setminus \overline{T}}\|_2 \le \rho \right\}$$

i.e. the set of vectors of sparsity ℓd that are ρ-close to a $\delta \ell d$-sparse vector

$$\mathrm{Incomp}^{(\ell)}_{\rho,\delta} := \mathbb{S}^N_{\ell d} \setminus \mathrm{Comp}^{(\ell)}_{\rho,\delta}$$

$$\rho_0(\ell) := c_0/(2C_0\sqrt{\ell \log d}),\ \ \delta_0(\ell) := c_0/(\ell \log d), \quad t_0(\ell) := 2\sqrt{\ell \log d/(c_0 d)}$$

$$\mathrm{Comp}^{(\ell)} := \mathrm{Comp}^{(\ell)}_{\rho_0(\ell),\delta_0(\ell)}, \quad \mathrm{Incomp}^{(\ell)} := \mathrm{Incomp}^{(\ell)}_{\rho_0(\ell),\delta_0(\ell)}$$

$$\mathrm{Bases}_\ell := \{(u_1,\ldots,u_\ell) \in \left(\mathbb{S}^{N-1}\right)^\ell \ \mid\ u_1,\ldots,u_\ell \text{ orthonormal}, \\ |\bigcup_{i\in[\ell]} \sigma^{u_i}| \le \ell d,\ \mathrm{Span}(\{u_1,\ldots,u_\ell\}) \cap \mathrm{Comp}^{(\ell)} = \emptyset\}$$

where c_0, C_0 are universal constants determined by Proposition 4.4.

We require a lower bound on the LCD of vectors that are incompressible.

Lemma 4.7 (LCD of incompressible vectors [RV09])**.**
Consider any $\rho, \delta \in (0,1)$, and any $a \in \mathrm{Incomp}^{(\ell)}_{\rho,\delta}$. Then, for every $\gamma \in (0,\ \rho^2\sqrt{\delta}/2)$ and every $\alpha > 0$, one has $\mathrm{LCD}_{\alpha,\gamma}(a) > \sqrt{\delta \ell d}/2$.

We are finally ready to use Theorem 4.2 to derive an estimate on the joint small ball probability of an orthogonal basis whose span does not contain compressible vectors.

Corollary 4.8 (Joint Small Ball Probability for Basis Elements of an Incompressible Subspace)**.**
There is a universal constant $C_1 > 1$ such that for any $\ell \in \mathbb{N}$, any $(u_1,\ldots,u_\ell) \in \mathrm{Bases}_\ell$, and any d sufficiently large,

$$\mathbb{P}\left[\max_{i\in[\ell]} \|Ru_i\|_\infty \le t_0(\ell)\right] \le (C_1 \log^2 d)^{h\ell} \cdot \ell^{2h\ell}/\sqrt{d}^{h\ell}$$

Proof. Assume $\gamma_0(\ell) := \rho_0(\ell)^2\sqrt{\delta_0(\ell)}/3$. Lemma 4.7 gives a universal lower bound of $\sqrt{c_0 d/\log d}/2$, on the LCD of vectors in $\mathrm{Incomp}^{(\ell)}$. Indeed for any $\alpha > 0$ we have

$$\mathrm{LCD}_{\alpha,\gamma_0(\ell)}(\mathrm{span}\{u_1,\ldots,u_\ell\}) = \inf_{a \in \mathbb{S}^{N-1}\cap \mathrm{span}\{u_1,\ldots,u_\ell\}} \mathrm{LCD}_{\alpha,\gamma_0(\ell)}(a) \ge \inf_{a\in \mathrm{Incomp}^{(\ell)}} \mathrm{LCD}_{\alpha,\gamma_0(\ell)}(a)$$

$$\geq \sqrt{c_0 d/\log d}/2\,,$$

where we used the fact that $\mathbb{S}^{N-1}\cap \operatorname{span}\{u_1,\ldots,u_\ell\} \subseteq \mathrm{Incomp}^{(\eta)}$ by definition of Bases.

We then apply Theorem 4.2 with $\alpha \leftarrow \sqrt{\ell}\log d$, $\gamma \leftarrow \gamma_0(\ell)$, $t \leftarrow t_0(\ell)$, to obtain that

$$\mathbb{P}\left[\sum_{i\in[\ell]}\langle \xi, u_i\rangle^2/\ell \leq t_0(\ell)^2\right] \leq C^\ell \frac{t_0(\ell)^\ell}{\gamma_0(\ell)^\ell} + C^\ell e^{-\ell\log^2 d} = O(\log^2 d)^\ell \cdot \ell^{2\ell}/\sqrt{d}^{\,\ell}\,.$$

Since $\max_{i\in[\ell]}\{|\langle \xi, u_i\rangle|\} \geq \left(\sum_{i\in[\ell]}\langle \xi, u_i\rangle^2/\ell\right)^{1/2}$, we obtain

$$\mathbb{P}\left[\max_{i\in[\ell]}|\langle \xi, u_i\rangle| \leq t_0(\ell)\right] = O(\log^2 d)^\ell \cdot \ell^{2\ell}/\sqrt{d}^{\,\ell}\,.$$

The claim then follows by observing that by independence of the rows of R,

$$\mathbb{P}\left[\max_{i\in[\ell]}\|Ru_i\|_\infty \leq t_0(\ell)\right] = \left(\mathbb{P}\left[\max_{i\in[\ell]}|\langle \xi, u_i\rangle| \leq t_0(\ell)\right]\right)^h.$$

■

Equipped with our joint small ball estimate, we are now ready to prove the bounded overlap property.

Theorem 2.1. *(Minimum Distance and Codeword Overlap for a Random Rademacher Subspace) Fix any $\varepsilon \in (0,1/2)$. For any $d \in \mathbb{N}$ sufficiently large, and any integer $N \in [d^{C_2/\varepsilon}, e^{d^{0.9}}]$, an $h \times N$ random matrix R with $h = \lceil 2(1+\varepsilon)d\log_d N\rceil$ and with i.i.d. ± 1-random entries satisfies with probability at least $1 - 1/N^{c_3 d/\log d}$,*

(a) $\mathbb{S}_d^N \cap \ker(R) = \emptyset$ *(Code Distance)*

(b) $\forall \ell \in \mathbb{N}\cap \left[1, \frac{\varepsilon}{C_2}\log_d N\right]$, $\mathrm{Overlap}_\ell \cap \ker(R)^\ell = \emptyset$ *where,* *(Bounded Codeword Overlap)*

$$\mathrm{Overlap}_\ell := \left\{(u_1,\ldots,u_\ell) \in \mathbb{R}^{N\times\ell} \;\middle|\; \mathrm{rk}\{u_1,\ldots,u_\ell\} = \ell,\ |\textstyle\bigcup_{i\in[\ell]}\sigma^{u_i}| \leq \ell d\right\}$$

and where $C_2 > 1$, $c_3 \in (0,1)$ are universal constants.

Proof. Claim (a) is a special case of (b) for $\ell \leftarrow 1$. So we need only prove (b). We first show that for any $\ell \in \mathbb{N}$, $\mathrm{Comp}^{(\ell)} \cap \ker(R) = \emptyset$ with high probability, which will then allow us to focus on incompresible subspaces.

Compressible vectors:

For any $u \in \mathrm{Comp}^{(\ell)}$, let $\overline{T}$ be the set of $\delta_0(\ell)\ell d$ coordinates of largest magnitude in u. Let y be the vector such that $y_j = u_j$ for $j \in \overline{T}$ and $y_j = 0$ otherwise, and $z = u - y$. Then $\|y\|_2 \geq \sqrt{1-\rho_0(\ell)^2}$, and $\|z\|_2 \leq \rho_0(\ell)$, $|y|_0 \leq \delta_0(\ell)\ell d$, and $|z|_0 \leq \ell d$. Then we have

$$\begin{aligned}\|Ru\|_2 &\geq \sigma_{min}^{\delta_0(\ell)\ell d}(R)\|y\|_2 - \sigma_{max}^{\ell d}(R)\|z\|_2\\ &\geq \sigma_{min}^{c_0 d/\log d}(R)\sqrt{1-\rho_0(\ell)^2} - \sigma_{max}^{\ell d}(R)\cdot\rho_0(\ell)\,.\end{aligned}$$

We have by Proposition 4.4, that with probability at least $1 - N^{-c_0 d/\log d}$,

$$\inf_{u\in\mathrm{Comp}^{(\ell)}}\|Ru\|_2 \;\geq\; c_0\sqrt{h}\sqrt{1 - 1/(4C_0^2\ell\log d)} - c_0\sqrt{d\log N/\log d}/2 \;\geq\; 3c_0\sqrt{h}/4\,. \tag{2}$$

which implies that for any $\ell \in \mathbb{N}$, $\mathrm{Comp}^{(\ell)} \cap \ker(R) = \emptyset$ with probability $1 - N^{-c_0 d/\log d}$, for d sufficiently large.

Incompressible Subspaces. Assume $\ell \le (\varepsilon/C_2)\log_d N$ for a sufficiently large constant C_2 chosen later. We will show with high probability that for any $(u_1, \dots, u_\ell) \in \mathrm{Overlap}_\ell$, at least one of $u_1, \dots, u_\ell$ doesn't lie inside $\ker(R)$. Observe that if $(u_1, \dots, u_\ell) \in \mathrm{Overlap}_\ell \cap \ker(R)^\ell$, then for any orthonormal basis $\widetilde{u}_1, \dots, \widetilde{u}_\ell$ of $\mathrm{span}\{u_1, \dots u_\ell\}$, it holds that $(\widetilde{u}_1, \dots, \widetilde{u}_\ell) \in \mathrm{Overlap}_\ell \cap \ker(R)^\ell$, and so it suffices to show that for any orthonormal basis $(u_1, \dots, u_\ell)$ within $\mathrm{Overlap}_\ell$, at least one of $u_1, \dots, u_\ell$ doesn't lie inside $\ker(R)$. Since we showed above that $\ker(R)$ avoids compressible vectors with high probability, we may assume that $\mathrm{span}\{u_1, \dots u_\ell\}$ is incompressible. Thus we need only show that $\mathrm{Bases}_\ell \cap \ker(R)^\ell = \emptyset$ with high probability, which we do by combining Corollary 4.8 with a union bound over a sufficiently fine net.

Orthogonal Bases of Incompressible Subspaces. Let $\mathcal{O} \subseteq \mathrm{Bases}_\ell$ be a minimum size $1/d^2$-net of Bases_ℓ according to the norm $\|(u_1, \dots u_\ell)\| := \max_{i \in [\ell]} \|u_i\|_2$. We have

$$
\begin{aligned}
&\inf_{(\widetilde{u}_1,\dots,\widetilde{u}_\ell)\in \mathrm{Bases}_\ell} \max_{i\in[\ell]} \|R\widetilde{u}_i\|_\infty \\
&\underset{\text{(triangle inequality)}}{\ge} \min_{(u_1,\dots,u_\ell)\in\mathcal{O}} \max_{i\in[\ell]} \|Ru_i\|_\infty - \max_{\substack{(u'_1,\dots,u'_\ell)\in \mathrm{Bases}_\ell,\ s.t.\\ \forall i\in[\ell],\ \|u'_i\|_2 \le 1/d^2}} \max_{i\in[\ell]} \|Ru'_i\|_\infty \\
&\underset{\text{(Cauchy-Schwarz)}}{\ge} \min_{(u_1,\dots,u_\ell)\in\mathcal{O}} \max_{i\in[\ell]} \|Ru_i\|_\infty - \sigma_{max}^{\ell d}(R)/d^2 \\
&\underset{\text{(Proposition 4.4)}}{\ge} \min_{(u_1,\dots,u_\ell)\in\mathcal{O}} \max_{i\in[\ell]} \|Ru_i\|_\infty - O\big(\sqrt{\ell \log N}/d^{1.5}\big) \quad \text{with probability } 1 - N^{-c_0 d/\log d} \\
&\underset{\text{(since } N \le e^{d^{0.1}},\ \ell = O(\log_d N))}{\ge} \min_{(u_1,\dots,u_\ell)\in\mathcal{O}} \max_{i\in[\ell]} \|Ru_i\|_\infty - o(t_0(\ell))
\end{aligned} \tag{3}
$$

It suffices now to show that $\min_{(u_1,\dots,u_\ell)\in\mathcal{O}} \max_{i\in[\ell]} \|Ru_i\|_\infty \ge t_0(\ell)$ with high probability, to obtain that $\mathrm{Bases}_\ell \cap \ker(R)^\ell = \emptyset$. We again proceed by anticoncentration+union bound. For any fixed $(u_1, \dots, u_\ell) \in \mathrm{Bases}_\ell$, we have by Corollary 4.8 that

$$
\mathbb{P}\left[\max_{i\in[\ell]} \|Ru_i\|_\infty \le t_0(\ell)\right] \le (C_1 \log^2 d)^{h\ell} \cdot \ell^{2h\ell} / \sqrt{d}^{h\ell} = \frac{1}{N^{\ell d(1+\varepsilon/3) - o(\ell d)}} .
$$

where the final step assumes that $\ell = O(\log_d N)$. As for net size, observe that Bases_ℓ has a $1/(2d^2)$-net of size $d^{O(\ell^2 d)} \cdot N^{\ell d}$ since the ℓd-dimensional sphere has a $1/(2d^2)$-net of size $d^{O(\ell d)}$, and any basis in Bases_ℓ can be generated by choosing a subset of size ℓd and then choosing ℓ vectors in the sphere supported on those coordinates. We then apply Fact 4.5 to obtain a $1/d^2$-net for Bases_ℓ that is also a subset. So we have $|\mathcal{O}| = d^{O(\ell^2 d)} \cdot N^{\ell d}$. Thus the probability that $\mathrm{Bases}_\ell \cap \ker(R)^\ell \ne \emptyset$ is at most $d^{O(\ell^2 d)}/N^{\varepsilon \ell d/2} \le 1/N^{\ell d(\varepsilon/3 - O(\ell/\log_d N))}$. The claim then follows assuming by taking a union bound over all choices of positive integers $\ell \le (\varepsilon/C_2)\log_d N$ for a sufficiently large universal constant $C_2 > 1$. ■

4.2 Embedding a Discrete Solution Set as a Quotient of a Random Rademacher Subspace

One of the key ingredients in Section 3 is the idea (inspired by precedents in the homogenization literature [Mic12]) of encoding (linearly) the solutions of a hard non-homogeneous optimization problem as nearly minimial weight cosets of a subspace with good minimum distance, where weight of a coset refers to the sparsity of the sparsest vector in the coset. In Section 3 we reduce from boolean quadratic equations (resp. affine sparsest vector) whose solution space is the hypercube. In this section we prove this embedding result for the case of a random Rademacher subspace.

The proof is not difficult and follows from a routine albeit technical second moment calculation. Potentially useful intuition here is that random Rademacher subspaces contain many nearly-minimum weight boolean vectors that are quite uncorrelated with one another. With respect to most projections to low enough dimensional space, this set of boolean vectors will behave like a high dimensional set in an appropriate sense and after projection will completely cover the low dimensional hypercube.

We prove the following technical lemma that will be used to deduce our embedding result discussed above.

For any $y \in \{0,1\}^n$, let $\mathcal{C}_y := \{x \in \{0,1\}^N \mid \mathrm{proj}(x) = y\}$.

Lemma 4.9 (Random Subspace of Distance $\geq d$ has Large Intersection with $H^N_{d(1+\varepsilon)}$**).**
Fix any $\varepsilon \in (0, 1/2)$. Let $k \in \mathbb{N}$ be even, set $d := k(1-\varepsilon)/(1+\varepsilon)$, $h := \lceil 2(1+\varepsilon) \cdot d \log_d N \rceil$, and let R be an $h \times N$ matrix with i.i.d. ± 1-random entries.

(a) If $N \geq k^{2/\varepsilon}$, then for k sufficiently large we have $\mathbb{P}[|H^N_k \cap \ker(R)| \geq N^{\varepsilon k/2}] \geq 1 - 1/N^{0.4}$.

(b) Consider any $n \geq k$ and set the column duplication parameter $\overline{N} := N/n$. If $N \geq n^{2/\varepsilon}$, then for k sufficiently large and any fixed $y \in H^n_k$, we have $\mathbb{P}[|\mathcal{C}_y \cap \ker(R)| \geq N^{\varepsilon k/2}] \geq 1 - 1/N^{0.4}$.

Proof. Since for any $y \in H^n_k$ we have $\mathcal{C}_y \subseteq H^N_k$, claim (a) follows as an immediate consequence of (b) by setting $n \leftarrow k$. We now prove the stronger statement (b).

Let p^t_ℓ denote the probability that the sum of t Rademacher variables equals ℓ. Fix any $y \in H^n_k$ and let $\mathcal{F} := \{\sigma^x \mid x \in \mathcal{C}_y\}$. We first check that $\mu := \mathbb{E}[|\mathcal{C}_y \cap \ker(R)|]$ is large. Let X_S denote the indicator random variable for the event that $R\mathbf{1}_S = 0$. We have $\mathbb{E}[X_S] = \mathbb{P}[R\mathbf{1}_S = 0] = (p^k_0)^h = \Theta(1/\sqrt{k})^h$ and so

$$\mu = \mathbb{E}\left[\sum_{S \in \mathcal{F}} X_S\right] \geq \overline{N}^k \cdot (p^k_0)^h = \overline{N}^k / N^{(1-\varepsilon)k} \geq N^{\varepsilon k/2} \,. \tag{4}$$

In order to use Chebyshev's inequality, we begin with a second moment calculation. Let $F_t := |\{(S_1, S_2) \;:\; S_1, S_2 \in \mathcal{F},\ |S_1 \cap S_2| = t\}|$, and observe that

$$\mathbb{E}\left[(\sum_{S \in \mathcal{F}} X_S)^2\right] = \sum_{t=0}^{k} F_t \cdot \mathbb{P}\left[R\mathbf{1}_{S_1} = 0,\ R\mathbf{1}_{S_2} = 0\right] \tag{5}$$

where S_1, S_2 above are any fixed sets satisfying $S_1, S_2 \in \binom{[N]}{k}$, $|S_1 \cap S_2| = t$. Above we used the fact that the joint distribution of $R\mathbf{1}_{S_1}, R\mathbf{1}_{S_2}$ is the same for any pair S_1, S_2 of k-sets that intersect in t elements.

We next give upper estimates for the terms in the aforementioned sum. Clearly, $F_t \leq \binom{k}{t} \cdot \overline{N}^t \cdot \overline{N}^{2k-2t} \leq \overline{N}^{2k-t} \cdot (ke/t)^t$. Assume $k - t$ is even and $t \leq k(1 - \varepsilon/10)$. We have

$$\begin{aligned}
&\mathbb{P}\left[R\mathbf{1}_{S_1}, R\mathbf{1}_{S_2} = 0\right] \\
&= \mathbb{P}\left[R\mathbf{1}_{S_1 \setminus S_2} = -R\mathbf{1}_{S_1 \cap S_2},\ R\mathbf{1}_{S_2 \setminus S_1} = -R\mathbf{1}_{S_1 \cap S_2}\right] \\
&= \sum_{z \in \mathbb{Z}^h} \mathbb{P}\left[R\mathbf{1}_{S_1 \setminus S_2} = -z,\ R\mathbf{1}_{S_2 \setminus S_1} = -z\right] \cdot \mathbb{P}\left[R\mathbf{1}_{S_1 \cap S_2} = z\right]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{z\in\mathbb{Z}^h} \mathbb{P}\left[R\mathbf{1}_{S_1\setminus S_2} = -z\right]\cdot \mathbb{P}\left[R\mathbf{1}_{S_2\setminus S_1} = -z\right]\cdot \mathbb{P}\left[R\mathbf{1}_{S_1\cap S_2} = z\right] \\
&= \Big(\sum_{\ell\in\mathbb{Z}} (p^{k-t}_{-\ell})^2\cdot(p^t_\ell)\Big)^h \\
&\le (p^{k-t}_0)^{2h} \\
&= (p^k_0)^{2h}\cdot\left(2^{-k}\cdot\binom{k}{k/2}\cdot 2^{k-t}\cdot\binom{k-t}{(k-t)/2}\right)^{2h} \\
&\le (p^k_0)^{2h}\cdot\left(e^{1/4(k-t)}\sqrt{k/k-t}\right)^{2h} \\
&\left(\text{Stirling's Approximation: } \frac{\ell!}{\sqrt{2\pi\ell}\cdot(\ell/e)^\ell}\in\left(e^{1/12\ell+1}, e^{1/12\ell}\right)\right) \\
&\le (p^k_0)^{2h}\cdot\left(e^{1/4(k-t)}\, e^{t/2(k-t)}\right)^{2h} && (\theta>0\Rightarrow\sqrt{1+\theta}<e^{\theta/2}) \\
&= (p^k_0)^{2h}\cdot e^{O(th/k)} && (\text{since } k-t\ge \varepsilon k/10)
\end{aligned}
$$

When $k-t$ is odd, we replace p^{k-t}_0 with p^{k-t}_1 above and follow the same approach, observing that $p^k_0/p^{k-t}_1 = 1+O(t/(k-t))$, to get the same bound. For the case when $t\ge k-\varepsilon k/10$, we will use a different upper estimate:

$$\mathbb{P}\left[R\mathbf{1}_{S_1}, R\mathbf{1}_{S_2}=0\right]\le \mathbb{P}\left[R\mathbf{1}_{S_1}=0\right]\le (p^k_0)^h .$$

Combining the above three estimates with Eq. (5), we obtain

$$
\begin{aligned}
&\mathbb{E}\left[\left(\textstyle\sum_{S\in\mathcal{F}} X_S\right)^2\right] \\
&\le F_0(p^k_0)^{2h} + \sum_{t=1}^{k(1-\varepsilon/10)} \left(e^{O(h/k)}ke/(t\overline{N})\right)^t\cdot\overline{N}^{2k}\cdot(p^k_0)^{2h} + \sum_{t=k(1-\varepsilon/10)}^{k} \overline{N}^{2k-t}\cdot(ke/t)^t\cdot(p^k_0)^h \\
&\le \mu^2+\mu^2\cdot\sum_{t=1}^{k(1-\varepsilon/10)}\left(e^{O(h/k)}ke/(t\overline{N})\right)^t + \mu\cdot k\cdot\overline{N}^{\varepsilon k/9} \\
&\le \mu^2+\mu^2\cdot\sum_{t=1}^{k(1-\varepsilon/10)}\left(\overline{N}^{O(1/\log k)}ke/(t\overline{N})\right)^t + O(\mu^{1+1/8}) \\
&\le \mu^2\left(1+k/\overline{N}^{1-O(1/\log k)}\right)\le \mu^2\left(1+1/N^{1-\varepsilon-O(1/\log k)}\right)
\end{aligned}
$$

By Chebyshev's inequality, we conclude

$$\mathbb{P}\left[|\mathcal{C}_y\cap\ker(R)|\ge\mu/2\right] = \mathbb{P}\left[\textstyle\sum_{S\in\mathcal{F}} X_S\ge\mu/2\right]\ge 1-4/N^{1-\varepsilon-O(1/\log k)}$$

which proves (b). ∎

We are now ready to conclude our desired embedding result.

Lemma 2.2. *(Embedding of $\{0,1\}^n$ in a Quotient of $\ker(R)\cap H^N_k$)*
Fix any $\varepsilon\in(0,1/2)$. Let $n\in\mathbb{N}$, and set $k := 16^2\lceil n^{2.2}/\varepsilon^2\rceil$, $d := k(1-\varepsilon)/(1+\varepsilon)$, $h := \lceil 2(1+\varepsilon)d\log_d N\rceil$ and finally pick any integer $N\ge k^{5/\varepsilon}$. Let R be an $h\times N$ matrix with i.i.d. ±1 random entries and let $T\in\{0,1\}^{n\times N}$ be such that each entry is set to 1 independently at random with probability $p=1/(4kn)$. Then assuming k is sufficiently large, with probability at least $1-6/n^{0.1}-1/N^{0.4}$, for every $y\in\{0,1\}^n$, there exists $x\in\ker(R)\cap H^N_k$ such that $T(x)=y$.

Proof. We combine the fact that $\ker(R)$ contains many vectors in H_k^N (by Lemma 4.9 (a)) with the following result which shows that the random projection of a sufficeintly large subset of a hypercube slice, to a sufficeintly low dimension, must cover the entire hypercube.

Theorem 4.10 (Theorem 5.9 of [Mic01], Random Projections Shatter Large Subsets of H_k^N)**.** *For any $k, n, N \in \mathbb{N}$ and any $t > 0$, let $\mathcal{F} \subseteq \{0,1\}^N$ be a set of at least $k! N^{4\sqrt{k}n/t}$ vectors, each with k non-zero entries. If $T \in \{0,1\}^{n \times N}$ is chosen by setting each entry to 1 independently at random with probability $p = 1/(4kn)$, then the probability that all of $\{0,1\}^n$ is contained in $T(\mathcal{F}) = \{Tx \mid x \in \mathcal{F}\}$ is at least $1 - 6t$.*

We apply the above theorem with the substitution $t \leftarrow 1/n^{0.1}$, $n \leftarrow n$, $k \leftarrow k$, $N \leftarrow N$. It is easily checked that the assumptions of our claim imply that $k! N^{4\sqrt{k}n/t} < N^{\epsilon k/2}$, and so the application of the above theorem is valid. ∎

Acknowledgements

The authors would like to thank Venkatesan Guruswami and Madhur Tulsiani for many useful discussions on these topics over the years. We would also like to thank an anonymous referee on a previous version of this manuscript for pointing us to the reference [Mic12]. The form of quotienting used in [Mic12] eventually played a role in making our results unconditional.

References

[Ajt98] Miklós Ajtai. The shortest vector problem in l2 is np-hard for randomized reductions. In *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, pages 10–19, 1998. 1

[AK14] Per Austrin and Subhash Khot. A simple deterministic reduction for the gap minimum distance of code problem. *IEEE Transactions on Information Theory*, 60(10):6636–6645, 2014. 1, 2, 3

[BBH+12] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou. Hypercontractivity, sum-of-squares proofs, and their applications. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 307–326. ACM, 2012. 1, 2

[BDJ+77] G Bennett, LE Dor, V GOODMAN WB JOHNSON, V Goodman, and CM Newman. On uncomplemented subspaces. *Israel Journal of Mathematics*, 26(2), 1977. 7

[BK01] Piotr Berman and Marek Karpinski. Approximating minimum unsatis ability of linear equations. In *Electron. Colloquium Comput. Complexity (ECCC)*, volume 25, 2001. 1

[BKS14] Boaz Barak, Jonathan A Kelner, and David Steurer. Rounding sum-of-squares relaxations. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 31–40. ACM, 2014. 1

[BKS15] Boaz Barak, Jonathan A Kelner, and David Steurer. Dictionary learning and tensor decomposition via the sum-of-squares method. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pages 143–151. ACM, 2015. 1

[BLN21] Vijay Bhattiprolu, Euiwoong Lee, and Assaf Naor. A framework for quadratic form maximization over convex sets through nonconvex relaxations. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 870–881, 2021. 2

[BP22] Huck Bennett and Chris Peikert. Hardness of the (approximate) shortest vector problem: A simple proof via reed-solomon codes. *arXiv preprint arXiv:2202.07736*, 2022. 2

[CP86] Thomas F Coleman and Alex Pothen. The null space problem i. complexity. *SIAM Journal on Algebraic Discrete Methods*, 7(4):527–537, 1986. 1

[CW09] Qi Cheng and Daqing Wan. A deterministic reduction for the gap minimum distance problem. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 33–38, 2009. 1, 2

[DMS03] Ilya Dumer, Daniele Micciancio, and Madhu Sudan. Hardness of approximating the minimum distance of a linear code. *IEEE Transactions on Information Theory*, 49(1):22–37, 2003. 1, 2

[HM13a] Moritz Hardt and Ankur Moitra. Algorithms and hardness for robust subspace recovery. In *Conference on Learning Theory*, pages 354–375. PMLR, 2013. 1

[HM13b] Aram W Harrow and Ashley Montanaro. Testing product states, quantum Merlin-Arthur games and tensor optimization. *Journal of the ACM (JACM)*, 60(1):3, 2013. 1, 2

[HR07] Ishay Haviv and Oded Regev. Tensor-based hardness of the shortest vector problem to within almost polynomial factors. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 469–477, 2007. 1

[Kho05] Subhash Khot. Hardness of approximating the shortest vector problem in lattices. *Journal of the ACM (JACM)*, 52(5):789–808, 2005. 1

[McC83] Seth Thomas McCormick. *A combinatorial approach to some sparse matrix problems*, volume 83. Stanford University Stanford, CA, 1983. 1

[Mic01] Daniele Micciancio. The shortest vector in a lattice is hard to approximate to within some constant. *SIAM journal on Computing*, 30(6):2008–2035, 2001. 1, 13

[Mic12] Daniele Micciancio. Inapproximability of the shortest vector problem: Toward a deterministic reduction. *Theory of Computing*, 8(1):487–512, 2012. 1, 2, 3, 10, 13

[QSW14] Qing Qu, Ju Sun, and John Wright. Finding a sparse vector in a subspace: Linear sparsity using alternating directions. *Advances in Neural Information Processing Systems*, 27, 2014. 1

[QZL+20] Qing Qu, Zhihui Zhu, Xiao Li, Manolis C Tsakiris, John Wright, and René Vidal. Finding the sparsest vectors in a subspace: Theory, algorithms, and applications. *arXiv preprint arXiv:2001.06970*, 2020. 1

[RV08] Mark Rudelson and Roman Vershynin. The littlewood–offord problem and invertibility of random matrices. *Advances in Mathematics*, 218(2):600–633, 2008. 3, 6

[RV09] Mark Rudelson and Roman Vershynin. Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 62(12):1707–1739, 2009. 3, 6, 7, 8

[Sch78] Thomas J Schaefer. The complexity of satisfiability problems. In *Proceedings of the tenth annual ACM symposium on Theory of computing*, pages 216–226, 1978. 4, 15

[Ver10] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010. 7

A NP-Hardness of Exactly Solving non-Homogeneous Quadratic Equations

Proposition 3.1. *Given n variables $y_1, \ldots, y_n$ and m quadratic equations $(\sum_{i,j \in [n]} \alpha_{\ell,i,j} y_i y_j = b_\ell)_{\ell \in [m]}$, it is NP-hard to distinguish between the following two cases:*
(YES) There exists $y \in \{0,1\}^n$ satisfying all m equations.
(NO) There doesn't exist $y \in \mathbb{R}^n$ satisfying all m equations.

Proof. We reduce from 1-in-3-SAT, where the input consists of n Boolean variables $x_1, \ldots, x_n$ and a 3-CNF formula $\varphi = C_1 \wedge \cdots \wedge C_m$ on where each C_i contains exactly three literals, and the goal is to find a Boolean assignment so that each C_i has at least one true literal and at least one false literal. It is one of the earliest problems proved to be NP-Complete [Sch78].

Given an instance of 1-in-3-SAT, we create $n+1$ variables $y_0, \ldots, y_n$, and add the following constraints. Let C_j contains three literals $s_{j,1}x_{i_{j,1}}$, $s_{j,2}x_{i_{j,2}}$, $s_{j,3}x_{i_{j,3}}$ where $i_{j,1}, i_{j,2}, i_{j,3} \in [n]$ denote the indices of the variables appearing in C_j and $s_{j,1}, s_{j,2}, s_{j,3} \in \{\pm 1\}$ indicate their signs. (I.e., $-x_1$ means $\neg x_1$.) The quadratic constraints are the following.

$$
\begin{aligned}
& y_0^2 = 1 && \\
& y_i^2 - y_0 y_i = 0 && \forall i \in [n] \\
& s_{j,1}(2y_{i_{j,1}} - y_0)s_{j,2}(2y_{i_{j,2}} - y_0) + s_{j,2}(2y_{i_{j,2}} - y_0)s_{j,3}(2y_{i_{j,3}} - y_0) && \\
& \quad + s_{j,3}(2y_{i_{j,3}} - y_0)s_{j,1}(2y_{i_{j,1}} - y_0) = -1 && \forall j \in [m]
\end{aligned}
$$

Completeness. Let $x_1, \ldots, x_n \in \{0,1\}$ be an assignment satisfying every 1-in-3-SAT constraint, where 1 denotes True and 0 denotes False. Then $y_0 = 1$, $y_i = x_i$ satisfies all the quadratic constraints; in particular, for each $j \in [m]$ and $\ell \in [3]$, $s_{j,\ell}(2y_{i_{j,\ell}} - y_0)$ becomes 1 if the corresponding literal $s_{j,\ell}x_{i_{j,\ell}}$ is True and -1 otherwise. So the LHS of the final constraint for $j \in [m]$ is 3 if all literals of C_j have the same value and -1 otherwise.

Soundness. Let $y_0, \ldots, y_n \in \mathbb{R}$ be an assignment satisfying all the quadratic equations. Then $y_0 = 1$ or -1. If $y_0 = 1$, then each $y_i \in \{0,1\}$ and $x_i = y_i$ for every $i \in [n]$ is a satisfying assignment for φ. The same argument as the completeness case shows that (x_i) is a solution for 1-in-3-SAT.

If $y_0 = -1$, $y_i \in \{0,-1\}$. Let us consider the assignment $x_i = -y_i$ for every $i \in [n]$. Then $s_{j,\ell}(2y_{i_{j,\ell}} - y_0)$ becomes -1 if the corresponding literal $s_{j,\ell}x_{i_{j,\ell}}$ is True and 1 otherwise. Again following the same argument, $(x_i)_{i \in [n]}$ is a satisfying assignment for φ. ■

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il