Adaptive Robustness of Hypergrid Johnson-Lindenstrauss

Andrej Bogdanov[*] Alon Rosen[†] Neekon Vafa[‡] Vinod Vaikuntanathan[§]

Abstract

Johnson and Lindenstrauss (Contemporary Mathematics, 1984) showed that for $n > m$, a scaled random projection $\mathbf{A}$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ is an approximate isometry on any set S of size at most exponential in m. If S is larger, however, its points can contract arbitrarily under $\mathbf{A}$. In particular, the hypergrid $([-B, B] \cap \mathbb{Z})^n$ is expected to contain a point that is contracted by a factor of $\kappa_{\mathsf{stat}} = \Theta(B)^{-1/\alpha}$, where $\alpha = m/n$.

We give evidence that finding such a point exhibits a statistical-computational gap precisely up to $\kappa_{\mathsf{comp}} = \widetilde{\Theta}(\sqrt{\alpha}/B)$. On the algorithmic side, we design an online algorithm achieving κ_{comp}, inspired by a discrepancy minimization algorithm of Bansal and Spencer (Random Structures & Algorithms, 2020). On the hardness side, we show evidence via a multiple overlap gap property (mOGP), which in particular captures online algorithms; and a reduction-based lower bound, which shows hardness under standard worst-case lattice assumptions.

As a cryptographic application, we show that the rounded Johnson-Lindenstrauss embedding is a robust property-preserving hash function (Boyle, Lavigne and Vaikuntanathan, TCC 2019) on the hypergrid for the Euclidean metric in the computationally hard regime. Such hash functions compress data while preserving ℓ_2 distances between inputs up to some distortion factor, with the guarantee that even knowing the hash function, no computationally bounded adversary can find any pair of points that violates the distortion bound.

[*] University of Ottawa. abogdano@uottawa.ca. Work supported by an NSERC Discovery Grant.

[†] Bocconi University. alon.rosen@unibocconi.it. Work supported by European Research Council (ERC) under the EU's Horizon 2020 research and innovation programme (Grant agreement No. 101019547) and Cariplo CRYPTONOMEX grant.

[‡] MIT. nvafa@mit.edu. Work supported in part by NSF DGE-2141064 and the grants of the fourth author. Part of this work was done while visiting Bocconi, supported by European Research Council (ERC) under the EU's Horizon 2020 research and innovation programme (Grant agreement No. 101019547).

[§] MIT. vinodv@mit.edu. Research supported in part by DARPA under Agreement Number HR00112020023, NSF CNS-2154149, a Simons Investigator award and a Ford Foundation Chair at MIT.

Contents

1 Introduction

The celebrated Johnson-Lindenstrauss (henceforth JL) lemma [JL84, IM98] gives us a powerful dimension-reduction mechanism for data.[1] The JL lemma states that for all fixed (small) finite sets $S \subseteq \mathbb{R}^n$, for a random i.i.d. Gaussian matrix $\mathbf{A} \sim \mathcal{N}(0,1)^{m \times n}$, the linear map $\frac{1}{\sqrt{m}} \cdot \mathbf{A}$ embeds S into $\mathbb{R}^m$ in a way that approximately preserves all ℓ_2 norms. More concretely, the guarantee that

$$\Pr_{\mathbf{A} \sim \mathcal{N}(0,1)^{m \times n}} \left[\forall \mathbf{x} \in S,\ \|\mathbf{A}\mathbf{x}\|_2 \in (1 \pm \epsilon) \cdot \sqrt{m} \cdot \|\mathbf{x}\|_2 \right] \geq 2/3,$$

can be achieved when $m = \Omega(\log(|S|)/\epsilon^2)$. The JL lemma has seen a great deal of work in the mathematics and TCS literature, and has been extended in several directions including faster and more space-efficient versions [AC09, KN12, CJN18, JPS$^+$22], stronger guarantees [NN19], and a proof of optimality [LN16, LN17].

The statement of the JL lemma crucially relies on the fact the set S is defined independently of the matrix $\mathbf{A}$. For example, even considering only singleton sets $S = \{\mathbf{x}\}$, one can ask whether the order of quantifiers can be switched so that $\mathbf{x}$ can be chosen adaptively based on $\mathbf{A}$, namely:

$$\Pr_{\mathbf{A} \sim \mathcal{N}(0,1)^{m \times n}} \left[\forall \mathbf{x} \in \mathbb{R}^n,\ \|\mathbf{A}\mathbf{x}\|_2 \approx \sqrt{m} \cdot \|\mathbf{x}\|_2 \right] \overset{?}{\geq} 2/3.$$

However, one immediately observes that this is impossible in a very strong sense. For $n > m$, for any matrix $\mathbf{A}$, one can always find a non-zero vector $\mathbf{x} \in \ker(\mathbf{A})$, making $\|\mathbf{A}\mathbf{x}\|_2 = 0$ while $\|\mathbf{x}\|_2$ can be arbitrarily large. Thus, an adaptive choice of the set of points, one that depends on the dimensionality reduction matrix, kills all nice guarantees that JL gave us.

One does not have to look too far to see the plausibility of such a scenario. If the matrix $\mathbf{A}$ is chosen once and for all (such as in derandomized versions of JL [EIO02]), and made public, adversarial entities can choose a set of points that violates the correctness of the JL lemma. Less obviously, even if $\mathbf{A}$ is not public and only very limited access to it is available, an adversary can reconstruct it and use it to mount the above attack, as shown first by Hardt and Woodruff [HW13, GLW$^+$25]. More generally, such *adaptive* attacks have been extensively in the context of data structures, streaming, property testing, especially in the last decade [MNS08, CSS25, GLW$^+$24, ACSS24, CNS$^+$24, NST23, CLN$^+$22, BKM$^+$22, HKM$^+$20, BEKMR23, ABD$^+$21, BY20].

1.1 The Contracting Hypergrid Vector Problem

Faced with this pessimistic scenario, we ask whether we can recover the guarantees of the JL lemma if we constrain the set of points S, and restrict to *resource-bounded* adaptive adversaries.

One way in which one could constrain S, is to zero-out the "most significant bits" of the coordinate vectors $\mathbf{x}$, i.e., to limit them to some ℓ_∞ ball. However, the above kernel strategy still works by rescaling $\mathbf{x}$ down to live in the ball.

Another option would be to zero-out the "least significant bits" of $\mathbf{x}$ by requiring $\mathbf{x}$ to live in a discrete set, e.g. $\mathbf{x} \in \mathbb{Z}^n$. But then, if we have no upper-bound on $\mathbf{x}$, we can scale up kernel vectors arbitrarily high so that they in fact live on the integer grid.

[1] The original JL work used a matrix whose rows are unit norm and orthogonal to each other. Indyk and Motwani [IM98] observed that a Gaussian matrix suffices.

Computationally interesting phenomena occur when combining both these constraints. More precisely, we will require that for all $i \in [n]$, $|x_i| \leq B$ *and* $x_i \in \mathbb{Z}$, or more concisely,

$$\mathbf{x} \in ([-B, B] \cap \mathbb{Z})^n$$

for some bound $B \in \mathbb{N}$, which could be as small as 1 or polynomially large in n. Phrased another way, this puts a fixed bound on the precision of $\mathbf{x}$. Like the kernel examples above, one can ask whether contraction occurs for this *hypergrid variant of JL*. We can phrase the problem as follows.

Definition 1 (Contracting Hypergrid Vector)**.** *For $n, m, B \in \mathbb{N}$ and $\kappa \in \mathbb{R}_{>0}$, we define the contracting hypergrid vector (*CHV*) problem with parameters n, m, B, and κ as follows. Given as input $\mathbf{A} \sim \mathcal{N}(0,1)^{m \times n}$, a valid solution is some $\mathbf{x} \in ([-B, B] \cap \mathbb{Z})^n$ such that*

$$\|\mathbf{A}\mathbf{x}\|_2 < \kappa \cdot \|\mathbf{x}\|_2 \cdot \sqrt{m}.$$

Here, the parameter κ quantifies the quality of the solution. A direct computation shows that any non-zero choice of $\mathbf{x}$ achieves $\kappa = \Theta(1)$ in expectation. Statistically, the threshold is

$$\kappa_{\mathsf{stat}} = \Theta\left((2B+1)^{-n/m}\right).$$

That is, for $\kappa \gg \kappa_{\mathsf{stat}}$, solutions exist in expectation, and for $\kappa \ll \kappa_{\mathsf{stat}}$, they do not.

The story changes if we *computationally bound* the task of finding $\mathbf{x} \in ([-B, B] \cap \mathbb{Z})^n$ that violate JL. That is, we can consider CHV as a computational problem to be solved by polynomial time algorithms. It is then only natural to ask what the computational complexity of CHV is, and whether it exhibits a statistical-computational gap.

1.2 Our Results

If we let κ_{comp} denote the best-possible *efficiently achievable* κ, we show, in a sense to be made precise below, that

$$\kappa_{\mathsf{comp}} = \widetilde{\Theta}\left(\frac{1}{B}\sqrt{\frac{m}{n}}\right),$$

where $\widetilde{\Theta}$ hides logarithmic terms in all parameters. It is useful to look at these values in terms of the *aspect ratio* α, defined as $\alpha := m/n < 1$. In this language, we have

$$\kappa_{\mathsf{stat}} = \Theta\left((2B+1)^{-1/\alpha}\right), \quad \kappa_{\mathsf{comp}} = \widetilde{\Theta}\left(\frac{\sqrt{\alpha}}{B}\right).$$

To demonstrate how large the statistical-computational gap is, considering only $B = 1$ gives a statistical bound that decays exponentially in $1/\alpha = n/m$ as opposed to computationally, where it decays polynomially. We illustrate with a phase diagram in Figure 1.

To establish the above value of κ_{comp}, we give two algorithms for CHV. One algorithm is a simple variant of the kernel attack described above, while the other is an online algorithm inspired by a discrepancy minimization algorithm of Bansal and Spencer [BS20].

We then give matching lower bounds. One is via the multiple overlap gap property, which in particular captures online algorithms. The other is a reduction-based lower bound which shows hardness under standard worst-case lattice assumptions.

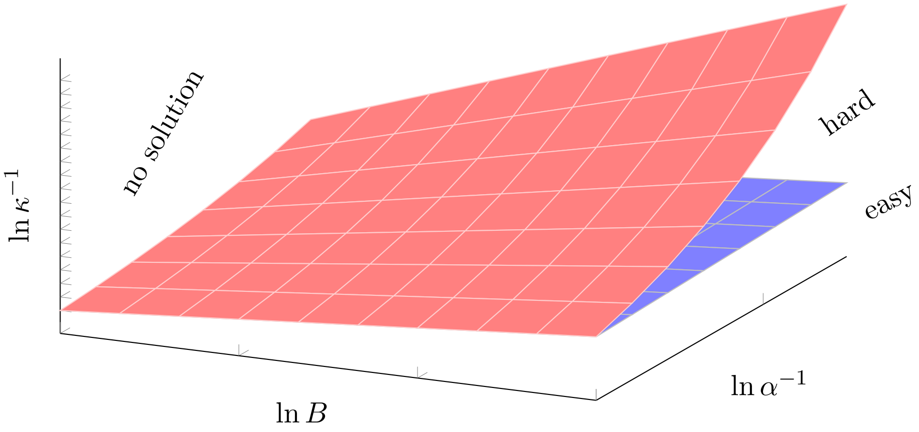


Figure 1: Phase diagram of CHV in the asymptotic regime (up to lower-order additive terms). The blue and red boundaries are $\ln \kappa^{-1} = \frac{1}{2}\ln \alpha^{-1} + \ln B$ and $\ln \kappa^{-1} = \alpha^{-1} \ln B$, respectively.

Finally, we show a positive use of the statistical-computational gap by giving a cryptographic application: a construction of *robust property preserving hash functions* for the Euclidean metric. These hash functions compress data while preserving ℓ_2 distances between input points up to some distortion factor, with the guarantee that no computationally bounded adversary can find any points that violate the distortion bound (even though they must exist).

Robust property preserving hash functions imply collision resistance, demonstrating that the statistical-computational gap for CHV yields cryptography beyond the existence of one-way functions.

We elaborate on our algorithms, hardness, and applications results in the upcoming sections.

1.3 Algorithms

We show two algorithms for CHV. The first uses the same kernel strategy as in the attack against the non-discretized version but is analyzed carefully; and the second is a novel variant of the Bansal-Spencer online discrepancy algorithm [BS20]. Here, “online” means that the algorithm receives every column of $\mathbf{A}$ one at a time, and after seeing column $j \in [n]$, the algorithm must commit to some choice $x_j \in [-B, B] \cap \mathbb{Z}$.

Theorem 1 (Informal Version of Theorem 7)**.** *For all $n > m$ and $\mathbf{A} \sim \mathcal{N}(0,1)^{m\times n}$, scaling and rounding a random vector in $\ker(\mathbf{A})$ yields $\mathbf{x} \in ([-B, B] \cap \mathbb{Z})^n$ such that*

$$\frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = O\left(\frac{1}{B} \cdot \sqrt{m \log B}\right)$$

with probability $1 - o(1)$. This directly solves CHV where

$$\kappa = O\left(\frac{\sqrt{\log B}}{B}\right)$$

with probability $1 - o(1)$.

Theorem 2 (Informal Version of Theorem 6)**.** *There is an online algorithm for* CHV *and a universal constant K such that as long as $n \geq Km \log B$, the algorithm achieves*

$$\kappa = O\left(\frac{1}{B} \cdot \sqrt{\frac{m}{n}}\right)$$

except with probability at most $O(\log B) \cdot 2^{-\Omega(m)}$.

Taken together, we have the following corollary, which applies for all ranges of $n > m$.

Corollary 1. *There is an algorithm for* CHV *such that for all $n > m$ and all B, the algorithm achieves*

$$\kappa = O\left(\frac{\log B}{B} \cdot \sqrt{\frac{m}{n}}\right)$$

except with probability at most $O(\log B) \cdot 2^{-\Omega(m)} + o(1)$.

Proof of Corollary 1 given Theorems 1 and 2. If $n \geq Km \log B$ for the constant K given in Theorem 2, apply Theorem 2. Otherwise, apply Theorem 1. □

1.4 Hardness

We show two flavors of hardness. The first shows an overlap gap property [Gam21] for the problem, giving evidence that a family of algorithms including local algorithms will fail to solve the problem, and the second shows computational hardness under the worst-case hardness of lattice problems, using ideas from [VV25].

The multiple overlap gap property (rOGP) forbids the existence of certain configurations of multiple solutions $\mathbf{x}_1, \ldots, \mathbf{x}_r$ to a search problem. It is viewed as an impediment for "stable" average-case algorithms in which changing a single input cannot have a large effect on the output. In our setting, it posits that not all relative angles $\angle \mathbf{x}_i, \mathbf{x}_j$ can fall into some proscribed range (θ_-, θ_+) that does not vanish with n.

Theorem 3 (Informal Version of Theorem 8)**.** *Assuming $n \geq m$ and*

$$\kappa \ll \frac{1}{B} \cdot \frac{1}{\sqrt{\log(n/m)}} \cdot \sqrt{\frac{m}{n}},$$

most instances of CHV *satisfy rOGP for sufficiently large n.*

In Theorem 9, we derive hardness of online algorithms in the same parameter regime from OGP hardness. Our result is not fully general, in the sense that we assume the algorithm is committed to the approximate norm of the solution $\mathbf{x}$ (before seeing the input $\mathbf{A}$). Our online algorithm, as well as the ones of Bansal and Spencer [BS20], satisfy this assumption. We believe that some assumption of this type is necessary for an OGP-based argument to ensure stability. In general, we conjecture that no online algorithm can succeed when κ is at most $(\sqrt{\pi/8} - o(1))\sqrt{\alpha}/B$ (Conjecture 1).

The lattice-based lower bound applies to the parameter regime in which κ vanishes as n grows.

Theorem 4 (Informal Version of Theorem 10)**.** *Assuming the polynomial hardness of standard worst-case lattice problems, for all $B, n \leq \mathrm{poly}(m)$, there does not exist any polynomial time algorithm for* CHV *satisfying*

$$\kappa \ll \frac{1}{B} \cdot \frac{1}{\sqrt{n}}.$$

We emphasize that Theorem 4 shows hardness of an *average-case* problem (namely, CHV) assuming only the *worst-case* hardness of lattice problems. We note however that this lower bound is quantitatively weaker than the OGP analysis, and is meaningful for a somewhat more restricted range of parameters. For example, when $B = \Theta(1)$, we need $n \gg m \log m$ for a solution to exist for $\kappa \approx 1/(B\sqrt{n})$. Nonetheless, this lower bound is still much higher than κ_{stat} for a wide range of parameters. We leave it as a fascinating open question as to whether this lower bound can be improved to match that of Theorem 3 or similar.

1.5 Robust Locality Sensitive Hashing

Semantic embeddings, which compress data points into vectors whose (Euclidean, say) distance approximates some semantic distance between data points, are a recent and very popular paradigm in machine learning [MCCD13, RG19, RKH$^+$21, MTMR23, Hug23].

Several recent works [SRS20, TJS23, ZJBS24] point out the issue of adversarial robustness, under the term “adversarial semantic collisions”, namely, inputs (text, images, and so on) that are semantically different yet hash to the same (or close) outputs. In light of this real-world phenomenon, the question of whether one can design semantically collision-resistant hash functions, ones for which semantic collisions may exist but are computationally hard to find, is practically important.

While we do not solve this problem, we suggest a possible approach. First, we observe that the computational-statistical gap for CHV allows us to design robust locality-sensitive hash functions [BLV19] for the *Euclidean distance*. These compressing hash functions preserve the ℓ_2 distance between input points up to some distortion factor ξ, assuming the points are chosen by a computationally bounded adversary. (See Definition 3 for a formal definition.) Secondly, if one could design a *collision-free, possibly dimension-expanding* embedding from “semantic distance” to Euclidean distance, we can compose the two to get a compressing semantically collision-resistant hash function.

We do not pursue the second of these problems in this paper; rather, we restrict our attention to designing a robust locality-sensitive hash function for the Euclidean distance.

Theorem 5 (Informal Version of Theorem 11)**.** *Suppose that* CHV *is hard for parameters* n, m, B*, and* κ*. Then, there is a robust locality sensitive hash function for the Euclidean norm mapping the domain* $([0, B] \cap \mathbb{Z})^n$ *into* $\mathbb{R}^m$ *with distortion*

$$\xi = O\left(\frac{1}{\kappa}\sqrt{\frac{n}{m}}\right)$$

that is compressing as long as

$$(B+1)^n \gg \left(\frac{Bn}{\kappa\sqrt{m}}\right)^m .$$

The construction is a variant of the Johnson-Lindenstrauss embedding where the output gets rounded to some grid $\gamma\mathbb{Z}^m$. Phrased another way, we show that the rounded Johnson-Lindenstrauss embedding is itself a robust locality sensitive hash function for the Euclidean norm where the domain is a hypergrid. We give the explicit construction in Figure 5.

Corollary 2 (Informal)**.** *Suppose that the online algorithm in Theorem 2 is optimal for* CHV*. Then, there is a compressing robust locality sensitive hash function for the Euclidean norm mapping the domain* $([0, B] \cap \mathbb{Z})^n$ *into* $\mathbb{R}^m$ *with the following parameters:*

- *For $B = \Theta(1)$, as long as $n \gg m \log m$, the distortion ξ can be as low as $\approx n/m$.*
- *For $B = n^{\Theta(1)}$, as long as $n \gg m$, the distortion ξ can be as low as $\approx Bn/m$.*

We emphasize these distortion bounds are only meaningful up to $B\sqrt{n}$, as any distinct elements of $([0, B] \cap \mathbb{Z})^n$ have ℓ_2 distance between 1 and $B\sqrt{n}$.

For a concrete application of these hash functions, consider (e.g., grayscale) images with pixel values in $\{0, 1, \cdots, 255\}$, where ℓ_2 distance encodes some semantic information. Our hash function generically compresses such images (with $B = 255$) while preserving approximate ℓ_2 norms of images and approximate ℓ_2 distance between images (to a computationally bounded adversary).

As another simple corollary of the construction, we note that our definition of robust locality sensitive hash functions implies collision resistance, whose existence is known to be stronger than the existence of one-way functions [Sim98]. Therefore, in particular, the statistical-computational gap of CHV yields cryptographic utility beyond the existence of one-way functions.

1.6 Related Work

CHV, SBP and NBV. Two problems closely related to CHV are the Symmetric Binary Perceptron (SBP) problem, introduced by Aubin, Perkins, and Zdeborová [APZ19] and the Nearest Boolean Vector (NBV) problem, introduced by Mohanty, Raghavendra, and Xu [MRX20]. The differences from CHV are as follows:

- In SBP and NBV, the domain of $\mathbf{x}$ is fixed to $\{-1, 1\}^n$, as opposed to $([-B, B] \cap \mathbb{Z})^n$. In particular, the norm of $\mathbf{x}$ is fixed and $\mathbf{x}$ cannot have zero entries.
- SBP asks for a bound on $\|\mathbf{Ax}\|_\infty$, as opposed to $\|\mathbf{Ax}\|_2$. (Therefore there is no normalizing $\sqrt{m}$ term.)

The search variant of SBP has been studied extensively for its statistical and computational thresholds [APZ19, BDVLZ20, PX21, ALS21, ALS22, GKPX22, GKPX23, BEAKZ24]. Both the algorithm of Bansal and Spencer and the rOGP bound of Gamarnik, Kızıldağ, Perkins and Xu [GKPX22] match ours in the special case $B = 1$.

The study of NBV has focused on refuting proximity to a random subspace in the unsatisfiable regime [GJJ+20, PTVW22, BR23].

Adaptively Robust X. Adaptive robustness has been studied in many related contexts, perhaps stemming from the work of Mironov, Naor and Segev [MNS08] in the context of sketching algorithms. Many recent works have further explored this question in the context of sketching and streaming algorithms [CSS25, GLW+24, ACSS24, CNS+24, CLN+22, BKM+22, HKM+20], randomized data structures [NY19], property testing [BEKMR23], online algorithms [ABD+21], and sampling algorithms [BY20].

Gribelyuk, Lin, Woodruff, Yu, and Zhou [GLW+25] give an efficient algorithm that, for any sufficiently compressing unknown linear embedding $\mathbf{A}$, finds a hypergrid vector $\mathbf{x}$ that fails to embed almost-isometrically under $\mathbf{A}$ given only query access to $\mathbf{A}$. In particular, their result implies an easy regime for CHV when κ is close to one. In contrast, our algorithms apply to much smaller values of κ (and are thus stronger) but are specific for the Johnson-Lindenstrauss embedding and assume unrestricted access to $\mathbf{A}$.

2 Technical Overview

2.1 Algorithms

We focus on the online algorithm. The algorithm iterates over $t \in \{1, 2, \ldots, n\}$ in order, receives each column $\mathbf{a}_t \sim \mathcal{N}(0, 1)^m$ one at at time, and commits to $x_t \in [-B, B] \cap \mathbb{Z}$ before incrementing t.

At first, we describe a simpler variant of the algorithm that produces $x_t \in \{-1, 1\}$ (which is a stronger constraint, but will yield a weaker bound). This in particular ensures that $\|\mathbf{x}\|_2$ is fixed at $\sqrt{n}$, so the goal of the algorithm is simply to minimize $\|\mathbf{A}\mathbf{x}\|_2$. At any given point in time $t \in [n]$, there exists a current state

$$\mathbf{y}_t = \sum_{i=1}^{t-1} x_i \mathbf{a}_i \in \mathbb{R}^m,$$

corresponding to the result of the choices it has made so far. The algorithm must choose $\mathbf{y}_{t+1}$ as $\mathbf{y}_{t+1} = \mathbf{y}_t + b\mathbf{a}_t$ for some $b \in \{\pm 1\}$. By rotational symmetry of the Gaussian, it is clear that the optimal online choice is to choose

$$x_t = \operatorname*{argmin}_{b \in \{-1,1\}} \|\mathbf{y}_t + b\mathbf{a}_t\|_2,$$

as the only thing that matters about $\mathbf{y}_{t+1}$ is its ℓ_2 norm.

We proceed to analyze this algorithm. We can decompose $\mathbf{a}_t$ into its perpendicular and parallel components with respect to $\mathbf{y}_t$. Explicitly, by spherical symmetry of the Gaussian, we have

$$\mathbf{a}_t = a^{\|} \cdot \frac{\mathbf{y}_t}{\|\mathbf{y}_t\|_2} + \mathbf{a}^{\perp},$$

where $a^{\|} \sim \mathcal{N}(0, 1)$ and $\mathbf{a}^{\perp}$ is a spherical multivariate Gaussian on the $(m-1)$-dimensional subspace perpendicular to the line spanned by $\mathbf{y}_t$. By the Pythagorean theorem, it follows that

$$\|\mathbf{y}_t + b\mathbf{a}_t\|_2^2 = \left\| \mathbf{y}_t + ba^{\|} \cdot \frac{\mathbf{y}_t}{\|\mathbf{y}_t\|_2} + b\mathbf{a}^{\perp} \right\|_2^2 = \left\| \mathbf{y}_t + \frac{ba^{\|}}{\|\mathbf{y}_t\|_2} \cdot \mathbf{y}_t \right\|_2^2 + \|\mathbf{a}^{\perp}\|_2^2.$$

Choosing $b = -\operatorname{sign}(a^{\|})$ minimizes this quantity, which yields

$$\min_{b \in \{-1,1\}} \|\mathbf{y}_t + b\mathbf{a}_t\|_2^2 = \left(\|\mathbf{y}_t\|_2 - \left| a^{\|} \right| \right)^2 + \|\mathbf{a}^{\perp}\|_2^2.$$

Letting $R_t = \|\mathbf{y}_t\|_2$ and expanding, we get the stochastic recurrence

$$R_{t+1}^2 = R_t^2 - 2R_t|z_1| + \|\mathbf{z}\|_2^2,$$

where $\mathbf{z} = (z_1, \ldots, z_m) \sim \mathcal{N}(0, 1)^m$. As the typical value of $|z_1|$ is $\Theta(1)$ and the typical value of $\|\mathbf{z}\|_2^2$ is $\Theta(m)$, we observe that we get negative drift in R_t whenever $R_t \gg m$, in the sense that $R_{t+1} \ll R_t$ with good probability. One can show that the fixed point of this recurrence is $R_t = \Theta(m)$, independent of t. This results in $\|\mathbf{A}\mathbf{x}\|_2 \le O(m)$ and $\|\mathbf{x}\|_2 = \sqrt{n}$, giving

$$\kappa = \frac{\|\mathbf{A}\mathbf{x}\|_2}{\sqrt{m} \cdot \|\mathbf{x}\|_2} \le O\left(\sqrt{\frac{m}{n}} \right),$$

as desired.

We briefly describe how to reduce κ by a factor of B by allowing $x_t \in [-B, B] \cap \mathbb{Z}$. For the first half of the steps (i.e., $t \leq n/2$), we set the "temperature" all the way up to B, to enforce $x_t \in \{-B, B\}$. This ensures that $\|\mathbf{x}\|_2 \geq \Omega(B\sqrt{n})$, regardless of what happens in the second half of the steps. However, the fixed point of the recurrence becomes $R_{n/2} = \Theta(Bm)$, which defeats the purpose of increasing $\|\mathbf{x}\|_2$. For the second half of the steps, we carefully choose a "cooling" schedule (see Figure 2) to get back to temperature 1, quickly converging to $R_n = \Theta(m)$. This results in $\|\mathbf{Ax}\|_2 \leq O(m)$ and $\|\mathbf{x}\|_2 \geq \Omega(B\sqrt{n})$, giving

$$\kappa = \frac{\|\mathbf{Ax}\|_2}{\sqrt{m} \cdot \|\mathbf{x}\|_2} \leq O\left(\frac{1}{B}\sqrt{\frac{m}{n}}\right),$$

as desired, at the cost of requiring $n \geq Km \log B$ for some universal constant K. For more details, we defer to Section 4.1.

2.2 Hardness

The multi-OGP hardness (Theorem 8) is derived from a first moment (annealed) estimate of the expected number of forbidden configurations in a random instance. Our analysis extends that of Gamarnik et al. for the SBP [GKPX22], where the solution space is the Boolean cube.

The main conceptual difference is in the choice of distance metric. Unlike for the Boolean cube, the Euclidean and Hamming metrics are not equivalent over the hypergrid. We prove OGP with respect to the normalized inner product. The utility of this metric is demonstrated in our online lower bound (Theorem 9), which essentially shows that it captures the distance between the outputs of executions over correlated inputs that share the same prefix and independent suffixes.

For the lattice-based lower bound (Theorem 10), one approach would be to adapt the recent work of Vafa and Vaikuntanathan [VV25] that shows a reduction from worst-case lattice problems to SBP. Instead of opening up their proof in a white-box way, we choose to reduce from an intermediate problem called "Continuous Learning With Errors" (CLWE) [BRST21]. This average-case problem, which is known to be as hard as worst-case approximate lattice problems [BRST21, GVV22], asks to distinguish

$$(\mathbf{A}, \mathbf{s}^\top \mathbf{A} + \mathbf{e}^\top \pmod 1), \qquad \text{and} \qquad (\mathbf{A}, \mathbf{b}^\top),$$

where $\mathbf{A} \sim \mathcal{N}(0,1)^{m \times n}$, $\mathbf{s} \sim n^\varepsilon \cdot \mathbb{S}^{m-1}$, $\mathbf{e} \sim \mathcal{N}(0, 1/\text{poly}(n))^n$, and $\mathbf{b}$ is a uniformly random vector mod 1, where $\mathbb{S}^{m-1}$ denotes the unit sphere in $\mathbb{R}^m$. By using integrality of $\mathbf{x}$ and simultaneous smallness of $\mathbf{x}$ and $\mathbf{Ax}$, we can multiply the second element in the pair on the right by $\mathbf{x}$ and check if the result is small modulo 1. We defer the details to Section 5.4.

2.3 Robust Locality Sensitive Hashing

To design robust locality sensitive hash functions in the Euclidean norm, our starting-point is the JL lemma and the syntax of CHV. We set the hash function to be

$$\text{Hash}_{\mathbf{A}} : ([0, B] \cap \mathbb{Z})^n \to \mathbb{R}^m,$$
$$\mathbf{x} \mapsto \frac{1}{\sqrt{m}} \cdot \mathbf{Ax},$$

where the key of the hash function is the matrix $\mathbf{A} \sim \mathcal{N}(0,1)^{m \times n}$. By linearity and a direct reduction from (the hardness of) CHV, it is hard to find $\mathbf{x}_1, \mathbf{x}_2 \in ([0,B] \cap \mathbb{Z})^n$ such that

$$\|\mathrm{Hash}_{\mathbf{A}}(\mathbf{x}_1) - \mathrm{Hash}_{\mathbf{A}}(\mathbf{x}_2)\| < \kappa \|\mathbf{x}_1 - \mathbf{x}_2\|, \tag{1}$$

as otherwise, $\mathbf{x}_1 - \mathbf{x}_2 \in ([-B,B] \cap \mathbb{Z})^n$ would be a solution to CHV. Therefore, we have the guarantee for this function, it is computationally hard to find two points whose distance between the hashes is multiplicatively smaller by a factor of $1/\kappa$.

As is, this construction has two issues:

1. Eq. (1) is only a one-sided guarantee. How do we ensure that it is hard to find $\mathbf{x}_1, \mathbf{x}_2 \in ([-B,B] \cap \mathbb{Z})^n$ such that

   $$\|\mathrm{Hash}_{\mathbf{A}}(\mathbf{x}_1) - \mathrm{Hash}_{\mathbf{A}}(\mathbf{x}_2)\| > \eta \|\mathbf{x}_1 - \mathbf{x}_2\|,$$

   for some parameter $\eta \gg 1$?

2. Even if $m < n$, what does it mean for this function to be compressing in a bit-complexity sense if the codomain is $\mathbb{R}^m$?

Thankfully, the solutions to both Items 1 and 2 are relatively simple. For Item 1, we can use the spectral norm bound that $\|\mathbf{A}\|_2 \le O(\sqrt{n})$ with high probability. This implies that we can set $\eta = O(\sqrt{n/m})$ (since $\mathbf{A}$ is scaled down by $\sqrt{m}$). This results in overall distortion

$$\frac{\eta}{\kappa} = O\left(\frac{1}{\kappa}\sqrt{\frac{n}{m}}\right).$$

For Item 2, we can both discretize $\mathbb{R}^m$ into a grid $\gamma\mathbb{Z}^m$ and upper bound its ℓ_2 norm given that $\|\mathbf{A}\|_2 \le O(\sqrt{n})$ is bounded anyways. Then, we can set the codomain to $\gamma\mathbb{Z}^m \cap \mathrm{Ball}_m(r)$ for sufficiently large r. Showing compression amounts to counting the number of points in the discretized ball as compared to the the number of points in the domain, $(B+1)^n$. We defer to Section 6 for deatils.

3 Preliminaries

For a natural number $n \in \mathbb{N}$, we let $[n]$ denote the set $\{1, 2, \cdots, n\}$. For real numbers $a, b \in \mathbb{R}$ with $a \le b$, we let $[a,b]$ denote the continuous interval $\{x \in \mathbb{R} : a \le x \le b\}$. We say a function $f : \mathbb{N} \to \mathbb{R}_{>0}$ is negligible if for all $c > 0$, $\lim_{n\to\infty} f(n) \cdot n^c = 0$. We use the notation $\mathsf{negl}(n)$ to denote a function that is negligible (in its input n). We similarly use the notation $\mathrm{poly}(n)$ to denote a function that is at most $n^{O(1)}$. As shorthand, we say an algorithm is p.p.t. if it runs in probabilistic polynomial time.

We let $\mathcal{N}(\mu, \sigma^2)$ denote the univariate Gaussian (or normal) distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma \in \mathbb{R}_{>0}$. For any distribution $\mathcal{D}$, we let $\mathcal{D}^m$ denote the product distribution of m i.i.d. copies of $\mathcal{D}$. We let χ^2_m denote the chi-squared distribution with m degrees of freedom, which is defined as the distribution of a random variable Z such that

$$Z = \sum_{i \in [m]} X_i^2$$

for i.i.d. $X_i \sim \mathcal{N}(0,1)$. Equivalently, for $\mathbf{v} \sim \mathcal{N}(0,1)^m$, the distribution of $\|\mathbf{v}\|_2^2$ is identically χ^2_m.

Definition 2 (Contracting Hypergrid Vector Problem (CHV))**.** *For $n, m, B \in \mathbb{N}$ and $\kappa \in \mathbb{R}_{>0}$ with $m < n$, we define the* CHV *problem with parameters n, m, B, and κ as follows. Given as input $\mathbf{A} \sim \mathcal{N}(0,1)^{m \times n}$, a valid solution is some $\mathbf{x} \in ([-B, B] \cap \mathbb{Z})^n$ such that*

$$\|\mathbf{A}\mathbf{x}\|_2 < \kappa \|\mathbf{x}\|_2 \sqrt{m}. \tag{2}$$

To match notation in the literature, we use α to denote the aspect ratio $m/n < 1$.

We say that CHV is (computationally) *hard* for parameters n, m, B, κ if for all p.p.t. algorithms $\mathcal{A}$, the probability that $\mathcal{A}$ outputs a valid solution to CHV for parameters n, m, B, κ is $\mathsf{negl}(n)$.

4 Algorithms

We analyze two efficient algorithms for solving CHV. Taken together, they give algorithms for CHV whenever $\kappa \gg \sqrt{\alpha} \cdot \log B / B$, provided $m < n$.

4.1 Online Norm Minimization

Our algorithm processes the columns $\mathbf{a}$ of $\mathbf{A}$ in sequence producing the corresponding entries of $\mathbf{x}$ one by one. It is inspired by one of the algorithms of Bansal and Spencer [BS20] where both $\mathbf{A}$ and $\mathbf{x}$ are restricted to ± 1 values and, less importantly, their objective is to minimize the infinity norm of $\mathbf{A}\mathbf{x}$.

We assume B is a power of two. If not, use the largest available. K is a sufficiently large absolute constant.

```
Online Algorithm Cool

State: $\mathbf{y} \in \mathbb{R}^m$, initialized with $\mathbf{0}$.
Parameter: The temperature $b \in [-B, B] \cap \mathbb{Z}$.

Step: On sample $\mathbf{a} \in \mathbb{R}^m$,
    Update $\mathbf{y}$ to $\mathbf{y} - b\mathbf{a}$ or $\mathbf{y} + b\mathbf{a}$, whichever is smaller in 2-norm.
    Output the minimizer $-b$ or $b$.

Algorithm Cool: Run $n$ steps with the temperature $b$ set to:
    $B$ in the first $n - Km(\log(B) - 1)$ steps,
    $B/2$ in the next $Km$ steps,
    $B/4$ in the next $Km$ steps,
        $\vdots$
    1 in the last $Km$ steps.
```

Figure 2: Online Norm Minimization Algorithm Cool, as analyzed in Theorem 6.

The only difference between Bansal and Spencer's transition rule and ours is the choice of norm to be minimized.[2] The 2-norm is more natural in the Gaussian setting. Our innovation is the cooling schedule which is responsible for the factor B scaling of the discrepancy.

Theorem 6. *Assuming the samples are independent normals and $n \geq 2Km\log B$, the output $\mathbf{x}$ of* Cool *satisfies $\|\mathbf{Ax}\|/\|\mathbf{x}\| = O(m/B\sqrt{n})$ except with probability at most $O(\log B)\cdot 2^{-\Omega(m)}$.*

The norm of $\mathbf{x}$ is dominated by the temperature B part of the schedule so it is at least $(B/2)\sqrt{n}$. It remains to show that the final state has norm at most $O(m)$.

The state update $\mathbf{y}' = \mathbf{y} + x\mathbf{a}$ can be decomposed as $\mathbf{y} + x\mathbf{a}^{\|} + x\mathbf{a}^{\perp}$, where $\mathbf{a}^{\|}$ is the component of $\mathbf{a}$ in the direction of $\mathbf{y}$, and $\mathbf{a}^{\perp}$ is its orthogonal complement. By Pythagoras' theorem,

$$\|\mathbf{y}'\|^2 = \|\mathbf{y} + x\mathbf{a}^{\|}\|^2 + \|x\mathbf{a}^{\perp}\|^2 = \|\mathbf{y} + x\mathbf{a}^{\|}\|^2 + b^2\|\mathbf{a}^{\perp}\|^2.$$

As $\mathbf{y}$ and $\mathbf{a}^{\|}$ are aligned, $\|\mathbf{y} + x\mathbf{a}^{\|}\|$ is either $\|\mathbf{y}\| - b\|\mathbf{a}\|$ or $\|\mathbf{y}\| + b\|\mathbf{a}\|$. The first choice is clearly the minimizing one and

$$\|\mathbf{y}'\|^2 = \left(\|\mathbf{y}\| - b\|\mathbf{a}^{\|}\|\right)^2 + b^2\|\mathbf{a}^{\perp}\|^2.$$

As the entries of $\mathbf{a}$ are independent normals, by spherical symmetry, $\|\mathbf{a}^{\|}\|$ and $\|\mathbf{a}^{\perp}\|^2$ are the absolute value $|N|$ of a normal random variable and an independent chi-squared random variable χ^2_{m-1} with $m-1$ degrees of freedom, respectively. The length $L = \|\mathbf{y}\|_2$ satisfies the stochastic recurrence

$$L' = \sqrt{(L - b|N|)^2 + b^2\chi^2_{m-1}}, \tag{3}$$

where L' denotes the updated length after one step.

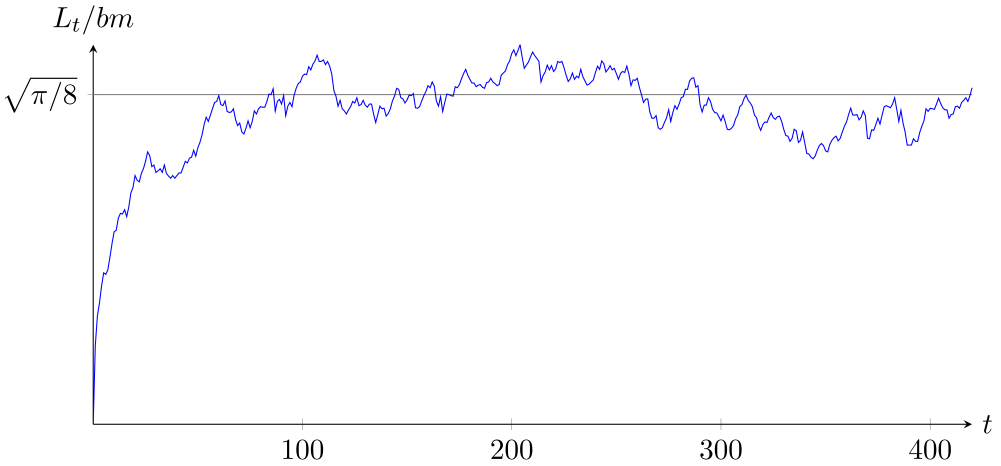


Figure 3: A sample realization of the stochastic process (3) with $m = 50$ with fixed temperature b. Since b is fixed, L_t is homogeneously linear in b, so the stochastic process $L_t/(bm)$ is independent of b.

Applying the inequality $\sqrt{1+x} \leq 1 + x/2$ valid for $x \geq 0$ yields

$$L' \leq |L - b|N|| + \frac{b^2\chi^2_{m-1}}{2|L - b|N||}. \tag{4}$$

[2] Specifically, their "majority rule" strategy (Strategy 2) can be interpreted as minimizing the ℓ_1 norm in the online step [BS20].

The typical magnitude of $|N|$ is constant and the typical magnitude of χ^2_{m-1} is on the order of m. If L is larger than about bm, the drift is typically negative.

Claim 1. *Assume $m \geq 4$. Conditioned on $L \geq 2bm$, L' is stochastically dominated by $L - b\mathsf{B}$ where B is independent of L, has mean at least 0.25, and has constant subexponential norm, in the sense that $\Pr[|\mathsf{B}| \geq t] \leq O\left(\exp\left(-\Omega(t)\right)\right)$ for all $t > 0$.*

For now, we defer the proof of Claim 1.

To handle the cases when L is small or $|N|$ is atypically large, we can use the simple bound that $\|\mathbf{y}'\|$ is still at most $\|\mathbf{y}\| + b\|\mathbf{a}\|$ by the triangle inequality so

$$L' \leq L + b\mathsf{A}, \qquad \text{where } \mathsf{A} \text{ is of type } \sqrt{\chi^2_m}. \tag{5}$$

We will refer to the corresponding random variables as in (5) and Claim 1 as being of type A and type B, respectively. By standard properties of the χ^2_m distribution and Jensen's inequality, type A is of mean at most $\sqrt{m}$ and has subgaussian norm $\Omega(\sqrt{m})$.

Let $L_0, L_1, \dots$ be a stochastic process that evolves according to (3) (with fixed b).

Claim 2. *If L_0 is at most $8bm$ then for every $t^\star \geq Km$, $L_{t^\star}$ is at most $4bm$ except with probability $O(2^{-\Omega(m)})$.*

Proof. Let $A_1, A_2, \dots$ and $B_1, B_2, \dots$ be sequences of independent type-A and type-B random variables, respectively. The drift $L_t - L_{t-1}$ is stochastically dominated by bA_t if $L_{t-1} < 2bm$ and by $-bB_t$ otherwise.

Let T be the last time $T \leq Km$ at which $L_T < 2bm$, if such a time exists. By Claim 1 and (5), $L_{t^\star}$ is then stochastically dominated by $2bm + b(A_T - B_{T+1} - \cdots - B_{t^\star})$. Otherwise, L_t is stochastically dominated by $L_0 - bB_1 - \cdots - bB_{t^\star}$. By a union bound over the possible choices of T,

$$\begin{aligned}
\Pr[L_{t^\star} > 4bm] &\leq \Pr[L_0 - bB_1 - \cdots - bB_{t^\star} > 4bm] + \sum_{t=1}^{t^\star} \Pr\left[bA_t - bB_{t+1} - \cdots - bB_{t^\star} > 2bm\right] \\
&\leq \Pr\left[B_1 + \cdots + B_{t^\star} \leq 4m\right] + \sum_{t=0}^{\infty} \Pr\left[A_1 - B_1 - \cdots - B_t > 2m\right].
\end{aligned}$$

As the type-B random variables have mean at least 0.25 and are subexponential, the first probability is at most 2^{-m} provided $t^* > Km$. As for the tth term in the sum, by union and tail bounds,

$$\begin{aligned}
\Pr\left[A_1 - B_1 - \cdots - B_t > 2m\right] &\leq \Pr\left[A_1 > m + t/8\right] + \Pr\left[B_1 + \cdots + B_t < -m + t/8\right] \\
&\leq \exp\left(-\Omega\left(\sqrt{m} + t/\sqrt{m}\right)^2\right) + \exp\left(-\Omega(m+t)\right) \\
&\leq 2\exp\left(-\Omega(m+t)\right).
\end{aligned}$$

By convergence of the geometric series, the sum over t is also bounded by $O\left(\exp\left(-\Omega(m)\right)\right)$. □

Proof of Theorem 6. Apply Claim 2 to $L = \|\mathbf{y}\|$. At the end of the first stage L is at most $4Bm = 8(B/2)m$ except with probability $O(2^{-\Omega(m)})$. Assuming it is, at the end of the second stage L is at most $4(B/2)m = 8(B/4)m$ except with probability $O(2^{-\Omega(m)})$, and so on. At the very end L is at most $4m$ as desired. The cumulative failure probability is at most $O(\log B)2^{-\Omega(m)}$. □

We now proceed to prove Claim 1.

Proof of Claim 1. Set

$$\mathsf{B} = \begin{cases} |N| - \chi^2_{m-1}/(2m), & \text{if } |N| \leq m \\ -\sqrt{m^2 + \chi^2_{m-1}}, & \text{otherwise.} \end{cases}$$

Stochastic domination follows from (4) and (5), as we can decompose the type A random variable into its component in the direction of $\mathbf{y}$ (i.e., N) and its $m-1$ independent other components. By standard facts and Cauchy-Schwarz, the mean of B is at least

$$\begin{aligned} \mathbb{E}[\mathsf{B}] &= \mathbb{E}[|N|] - \frac{\mathbb{E}\left[\chi^2_{m-1}\right]}{2m} - \mathbb{E}\left[\left(\sqrt{m^2 + \chi^2_{m-1}} + |N| - \frac{\chi^2_{m-1}}{2m}\right) \mathbb{1}\left[|N| > m\right]\right] \\ &\geq \sqrt{\frac{2}{\pi}} - \frac{m-1}{2m} - \left(\sqrt{m^2 + \mathbb{E}\left[\chi^2_{m-1}\right]} + \sqrt{\mathbb{E}\left[N^2\right]} + \sqrt{\frac{\mathbb{E}\left[\chi^4_{m-1}\right]}{4m^2}}\right) \sqrt{\Pr\left[|N| > m\right]} \\ &\geq 0.29 - (2m+2) \cdot 2\exp\left(-m^2/4\right) \\ &\geq 0.25. \end{aligned}$$

For the subexponential norm, by union and large deviation bounds,

$$\begin{aligned} \Pr\left[|\mathsf{B}| \geq t\right] &\leq \Pr\left[|N| \geq t\right] + \Pr\left[\frac{\chi^2_{m-1}}{2m} \geq t\right] + \Pr\left[\sqrt{m^2 + \chi^2_{m-1}} \geq t \text{ and } |N| > m\right] \\ &\leq 2\exp\left(-t^2/2\right) + \exp\left(-t/2\right) + \Pr\left[\sqrt{m^2 + \chi^2_{m-1}} \geq t \text{ and } |N| > m\right] \\ &= O(\exp(-t/2)) + \Pr\left[\sqrt{m^2 + \chi^2_{m-1}} \geq t \text{ and } |N| > m\right]. \end{aligned}$$

To bound the right-hand term, we consider two cases. If $t < m^2$, then

$$\Pr\left[\sqrt{m^2 + \chi^2_{m-1}} \geq t \text{ and } |N| > m\right] \leq \Pr[|N| > m] \leq O\left(\exp\left(-m^2/4\right)\right) \leq O(\exp(-\Omega(t))).$$

If $t \geq m^2$,

$$\begin{aligned} \Pr\left[\sqrt{m^2 + \chi^2_{m-1}} \geq t \text{ and } |N| > m\right] &\leq \Pr\left[\chi^2_{m-1} \geq t^2 - m^2\right] \\ &\leq \exp\left(-\Omega\left(t^2/m - m\right)\right) \\ &\leq \exp\left(-\Omega\left(tm - m\right)\right) \\ &\leq \exp\left(-\Omega\left(t\right)\right). \end{aligned}$$

Therefore, in all cases, $\Pr[|\mathsf{B}| \geq t] \leq O(\exp(-\Omega(t)))$. □

Limiting behavior In the limit $m \to \infty$ under the scaling $U = L/bm$, $dt = 1/m$, (4) is approximated by the stochastic differential equation

$$dU = \left(-\mu + \frac{1}{2U}\right)dt + \sigma dW,$$

where $\mu = \sqrt{2/\pi}$ and $\sigma = \sqrt{1-2/\pi}$ are the statistics of $|N|$, and W is the Wiener process. The drift pushes U towards the fixed point

$$U = \frac{1}{2\mu} = \sqrt{\frac{\pi}{8}} \approx 0.627.$$

Conjecture 1 (Online threshold conjecture)**.** *For every δ and B there exists a sufficiently small α and ϵ so that every online algorithm fails to find $\mathbf{x} \in ([-B,B] \cap \mathbb{Z})^n$ such that $\|\mathbf{A}\mathbf{x}\|/(\sqrt{m}\|\mathbf{x}\|) < (\sqrt{\pi/8}-\delta)\sqrt{\alpha}/B$ for at least an ϵ fraction of αn by n matrices A for all sufficiently large n.*

4.2 Kernel Rounding

```
Algorithm KernelRound

Input: $\mathbf{A} \sim \mathcal{N}(0,1)^{m\times n}$ where $m < n$.

  1. Sample a random $\mathbf{x}$ such that $\mathbf{A}\mathbf{x} = \mathbf{0}$ according to the Haar measure.
  2. Scale $\mathbf{x}$ to have length $\sqrt{\chi_n^2} \cdot \frac{B}{\sqrt{4K \ln^+ B}}$.
  3. Define the rounded vector $\mathbf{z} = \lceil \mathbf{x} \rfloor_B$.

Output: The vector $\mathbf{z} \in ([-B,B] \cap \mathbb{Z})^n$.
```

Figure 4: Kernel Rounding Algorithm for CHV, as analyzed in Theorem 7.

The rounding function is applied entrywise as

$$\lceil x \rfloor_B = \begin{cases} \lceil x \rfloor, & \text{if } |x| \le B, \\ B \operatorname{sign} x, & \text{otherwise,} \end{cases}$$

where $\lceil x \rfloor$ denotes rounding $x \in \mathbb{R}$ to the nearest integer (and tie-breaking arbitrarily). We will let $\ln^+$ denote the function $\max\{\ln, 1\}$. $K \ge 2$ is a (constant) parameter that controls the tradeoff between approximation quality and failure probability in the regime $B = \Omega(n^{1/4})$.

Theorem 7. *For $\mathbf{A} \sim \mathcal{N}(0,1)^{m\times n}$, the algorithm* KernelRound *outputs a vector $\mathbf{z} \in ([-B,B] \cap \mathbb{Z})^n$ such that*

$$\frac{\|\mathbf{A}\mathbf{z}\|_2}{\|\mathbf{z}\|_2} = O\left(\frac{\sqrt{mK \ln^+ B}}{B}\right),$$

except with probability $2^{-\Omega(m)} + \min\left\{O\left(2^{-\Omega\left(n(K\ln B)^2/B^4\right)}\right), 2nB^{-K}\right\}$.

As the space spanned by the rows of $\mathbf{A}$ is a random m-dimensional subspace of $\mathbb{R}^n$, the marginal distribution of $\mathbf{x}$ is identically $\mathcal{N}(0, B^2/(4K \ln^+ B))^n$. We analyze the conditional distribution of $\langle \mathbf{a}, \mathbf{x} \rangle$ for a single row $\mathbf{a}$ of $\mathbf{A}$.

Claim 3. *Suppose $\mathbf{a} \sim \mathcal{N}(0,1)^n$. Conditioned on $\mathbf{x}$ and $\langle \mathbf{a}, \mathbf{x} \rangle = 0$, the random variable $\langle \mathbf{a}, \lceil \mathbf{x} \rfloor_B \rangle$ is a centered normal of variance at most $\|\{\mathbf{x}\}_B\|^2$, where $\{x\}_B = x - \lceil x \rfloor_B$ is applied coordinate-wise.*

We emphasize that $\{x\}_B$ can be arbitrarily large when considering $|x| > B$.

Claim 4. *If $N \sim \mathcal{N}(0,1)$, then the random variable*

$$\left\{ \frac{B}{\sqrt{4K \ln^+ B}} \cdot N \right\}_B^2$$

has mean at most $1/4 + O(B^{-K+2})$ and subexponential norm $O(B^2/(K \ln^+ B))$.

Proof of Theorem 7. The quantity $\|\{\mathbf{x}\}_B\|^2$ is a sum of n independent

$$\left\{ \frac{B}{\sqrt{4K \ln^+ B}} \cdot N \right\}_B^2$$

random variables where $N \sim \mathcal{N}(0,1)$. By Claim 4 and Bernstein's inequality [Ver18, Theorem 2.8.1], $\|\{\mathbf{x}\}_B\|^2$ is $O(n)$ except with probability $O(2^{-\Omega(n(K \ln B)^2/B^4)})$, which is meaningful whenever $B = O(n^{1/4})$. When B is large relative to n, by a Gaussian tail bound and a union bound, none of the n entries of $\mathbf{x}$ exceed B in absolute value and so $\|\{\mathbf{x}\}_B\|^2 \leq n/4$ except with probability $2n \cdot B^{-K}$.

By Claim 3, conditioned on $\mathbf{x}$, the entries of $\mathbf{A}\lceil \mathbf{x} \rfloor_B$ are m independent normals of variance $O(n)$ with the same exceptional probability. By Hoeffding's inequality $\|\mathbf{A}\lceil \mathbf{x} \rfloor_B\|^2$ itself is bounded by $O(mn)$ except with additional probability $2^{-\Omega(m)}$.

Lastly, we lower bound $\|\lceil \mathbf{x} \rfloor_B\|$. Each entry of $\lceil \mathbf{x} \rfloor_B$ is of magnitude at least $B/\left(4\sqrt{K \ln^+ B}\right)$ with constant probability. By a large deviation bound,

$$\|\lceil \mathbf{x} \rfloor_B\| \geq \Omega\left(\frac{B\sqrt{n}}{\sqrt{K \ln^+ B}}\right),$$

except with probability $2^{-\Omega(n)}$. By a union bound, the target ratio becomes

$$\frac{\|\mathbf{A}\lceil \mathbf{x} \rfloor_B\|}{\|\lceil \mathbf{x} \rfloor_B\|} \leq O\left(\frac{\sqrt{mn}}{B\sqrt{n}/\left(\sqrt{K \ln^+ B}\right)}\right) = O\left(\frac{\sqrt{mK \ln^+ B}}{B}\right).$$

□

Proof of Claim 3. $\langle \mathbf{a}, \lceil \mathbf{x} \rfloor_B \rangle = \langle \mathbf{a}, \mathbf{x} \rangle - \langle \mathbf{a}, \{\mathbf{x}\}_B \rangle$. Conditioned on $\mathbf{x}$ and $\langle \mathbf{a}, \mathbf{x} \rangle = 0$, $\mathbf{a}$ is jointly normal and centered (but not independent), so $\langle \mathbf{a}, \{\mathbf{x}\}_B \rangle$ is also centered normal. Conditioning on $\langle \mathbf{a}, \mathbf{x} \rangle = 0$ implies that $\langle \mathbf{a}, \lceil \mathbf{x} \rfloor_B \rangle$ is centered normal too.

To derive its variance, we first calculate the conditional covariances of the entries of $\mathbf{a}$ given $\mathbf{x}$. Unconditionally, the entries a_i of $\mathbf{a}$ are independent standard normal. They decompose as

$$a_i = \frac{x_i}{\|\mathbf{x}\|^2} \langle \mathbf{a}, \mathbf{x} \rangle + a_i^{\perp},$$

where $a_i^{\perp}$ is a centered normal independent of $\langle \mathbf{a}, \mathbf{x} \rangle$ given $\mathbf{x}$. By independence and expanding out the expression,

$$
\begin{aligned}
\mathrm{Cov}(a_i, a_j \mid \mathbf{x}, \langle \mathbf{a}, \mathbf{x} \rangle = 0) &= \mathrm{Cov}\left(a_i^{\perp}, a_j^{\perp} \mid \mathbf{x}, \langle \mathbf{a}, \mathbf{x} \rangle = 0\right) \\
&= \mathrm{Cov}\left(a_i^{\perp}, a_j^{\perp} \mid \mathbf{x}\right) \\
&= \mathbb{E}\left[\left(a_i - \frac{x_i}{\|\mathbf{x}\|^2}\langle \mathbf{a}, \mathbf{x} \rangle\right)\left(a_j - \frac{x_j}{\|\mathbf{x}\|^2}\langle \mathbf{a}, \mathbf{x} \rangle\right) \;\Big|\; \mathbf{x}\right] \\
&= \mathbb{1}[i = j] - \frac{x_i x_j}{\|\mathbf{x}\|^2}.
\end{aligned}
$$

By the variance of sum formula,

$$
\begin{aligned}
\mathrm{Var}\left(\langle \mathbf{a}, \{\mathbf{x}\}_B \rangle \mid \mathbf{x}, \langle \mathbf{a}, \mathbf{x} \rangle = 0\right) &= \sum_i \{x_i\}_B^2 - \sum_{i,j} \{x_i\}_B \{x_j\}_B \cdot \frac{x_i x_j}{\|\mathbf{x}\|^2} \\
&= \|\{\mathbf{x}\}_B\|^2 - \left(\frac{\langle \mathbf{x}, \{\mathbf{x}\}_B \rangle}{\|\mathbf{x}\|}\right)^2 \\
&\leq \|\{\mathbf{x}\}_B\|^2.
\end{aligned}
$$

Once again, conditioned on $\mathbf{x}$ and $\langle \mathbf{a}, \mathbf{x} \rangle = 0$, this implies that the variance of $\langle \mathbf{a}, \lceil \mathbf{x} \rfloor_B \rangle$ is also at most $\|\{\mathbf{x}\}_B\|^2$. □

Proof of Claim 4. Let $L = B/\sqrt{4K \ln^+ B}$. Then by Cauchy-Schwarz and standard tail bounds,

$$
\begin{aligned}
\mathbb{E}\left[\{LN\}_B^2\right] &\leq \mathbb{E}\left[(LN - \lceil LN \rfloor)^2 \cdot \mathbb{1}[LN \leq B]\right] + \mathbb{E}\left[(LN - B)^2 \cdot \mathbb{1}[LN > B]\right] \\
&\leq \frac{1}{4} + \sqrt{\mathbb{E}\left[(LN - B)^4\right]}\sqrt{\Pr\left[LN > B\right]} \\
&\leq \frac{1}{4} + O\left(B^2 \exp\left(-B^2/4L^2\right)\right) \\
&\leq \frac{1}{4} + O\left(B^{2-K}\right).
\end{aligned}
$$

For every $t > 1/2$,

$$
\Pr\left[\{LN\}_B^2 \geq t\right] = \Pr\left[|LN| \geq B + \sqrt{t}\right] \leq 2\exp\left(-\frac{(B + \sqrt{t})^2}{2L^2}\right) = 2\exp\left(-\left(\frac{B/\sqrt{t} + 1}{L\sqrt{2}}\right)^2 \cdot t\right).
$$

The parenthesized expression squared is at least $1/(2L^2)$, so $\{LN\}_B^2$ is subexponential of norm $O(L^2)$. □

5 Hardness

5.1 Overlap Gap Property

We study the typical structure of the solution space of (2) under the parametrization $m = \alpha n$ for fixed α, κ, B and large n. Gamarnik et al. [GKPX22] showed that when the solution $\mathbf{x}$ is restricted to the Boolean cube $\{\pm 1\}^n$ the multi-overlap gap property holds in the regime $\alpha \gg \kappa^2 \log 1/\kappa$. We establish the following extension.

Theorem 8. *For all α, κ, B with $\kappa \ll 1/B$ and $\alpha \gg (B\kappa)^2 \log 1/B\kappa$ there exists β and r such that for all sufficiently large n, for all replicas $\mathbf{x}_1, \dots, \mathbf{x}_r$ such that*

$$1 - \beta \le \frac{\langle \mathbf{x}_i, \mathbf{x}_j \rangle}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|} \le 1 - \beta + \beta/2r \qquad \textit{for all } i \neq j, \tag{6}$$

at least one $\mathbf{x}_i$ fails to satisfy (2) *except with probability* $\exp -\Omega(\alpha r n)$.

This theorem can be interpreted as evidence for CHV being hard in the regime $\alpha \gg B^2\kappa^2 \log 1/\kappa$.

In the algorithmic study of random disordered systems, algorithmic efficiency is predicted by "replica symmetry" of the solution space. Under the replica symmetric model, the uniform distribution over solutions of $\|\mathbf{A}\mathbf{x}\| \le \kappa m$ is approximated by a product distribution P over $\{\pm 1\}^n$ with different biases across coordinates.

Assuming P has sufficient entropy, the overlap $\langle \mathbf{x}, \mathbf{x}' \rangle / \|\mathbf{x}\| \|\mathbf{x}'\|$ between two random solutions ought to be bounded away from 1. Let $B(t)$ be the "bouquet" of vectors $\mathbf{x}_1, \dots, \mathbf{x}_r$ in which the first t coordinates are sampled identically from P and the rest are sampled independently from P. By the law of large numbers all pairwise overlaps in $B(t)$ will typically concentrate around some value $1 - \beta(t)$. As $\beta(0)$ is bounded away from zero, $\beta(n)$ is zero, and β is $O(1/n)$-Lipschitz, by the intermediate value theorem $\beta(t)$ is bound to hit the interval $(\beta - \beta/2r, \beta)$ for some t, contradicting Theorem 8.

While is difficult to justify the accuracy of the replica symmetric model, and algorithmic easiness may persist even under replica symmetry breaking, the underlying logic can be used to rigorously rule out certain natural classes of algorithms. In Section 5.3 we prove that a natural extension of Theorem 8, known as the ensemble (multi) overlap gap property, rules out online algorithms under a certain stability restriction in the claimed regime. Before stating this extension it is instructive to see the proof of Theorem 8.

5.2 Proof of the Overlap Gap Property

The proof goes by a first moment (annealed) estimate. Claim 7 is used to count the number of forbidden configurations $\mathbf{x}_1, \dots, \mathbf{x}_r$. Claim 8, together with the explicit formula for the multivariate Gaussian PDF, is used to bound the probability of any such configuration simultaneously solving (2). For the parameter regime of interest the expected number of configuration, which is the product of these two numbers, is close to zero.

Claim 5. *Assuming $t \le 1/12B$ the number of nonzero B-bounded points $\mathbf{x} \in \mathbb{Z}^n$ that are within angle* $\arctan t$ *of some fixed $\mathbf{x}_0$ is at most* $O((1 + \ln B)/t) \cdot \exp O(nt^2B^2 \ln 1/tB)$.

Claim 6. *For $\rho \le 1/2$ an n-dimensional ball of radius $\sqrt{\rho n}$ contains at most* $\exp O(n\rho \log 1/\rho)$ *integer points.*

Proof. The number of integer points within distance $\sqrt{\rho n}$ of $\mathbf{y}$ is at most [OM90, Lemma 1]

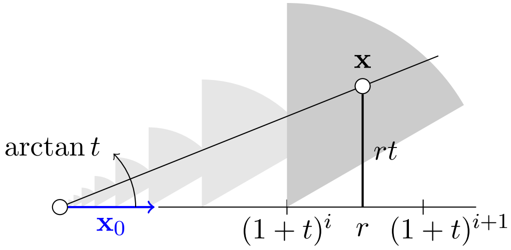
$$\exp(s\rho n) \prod_i \sum_k \exp(-s(k - y_i)^2)$$

for every $s > 0$. The summation is minimized at $y_i = 0$ giving an upper bound of

$$\exp(s\rho n) \Big(\sum_k \exp(-sk^2) \Big)^n \le \exp(s\rho n) \Big(1 + \frac{2e^{-s}}{1 - e^{-s}} \Big)^n$$

Setting $s = \ln 1/\rho$ produces the desired bound. □

Claim 7. *Assume $t \le 1$ and $\|\mathbf{x}_0\| = 1$. The union of balls with centers $(1+t)^i\mathbf{x}_0$ and radii $\sqrt{5}t(1+t)^i$ where i ranges over the integers covers all nonzero $\mathbf{x}$ within angle $\arctan t$ of $\mathbf{x}_0$.*



Proof. Let $r\mathbf{x}_0$, $r > 0$ be the projection of $\mathbf{x}$ in the direction of $\mathbf{x}_0$. Take the largest i such that $(1+t)^i \le r$. The distance between $\mathbf{x}$ and its projection $r\mathbf{x}_0$ is at most rt. By Pythagoras' theorem the distance from $\mathbf{x}$ to the center of ball i is at most

$$\begin{aligned}\sqrt{r^2t^2 + \left((1+t)^{i+1} - (1+t)^i\right)^2} &\le \sqrt{r^2t^2 + (1+t)^{2i}t^2} \\ &< (1+t)^i t\sqrt{(1+t)^2 + 1}\end{aligned}$$

because $r < (1+t)^{i+1}$. As $t \le 1$ the square root is at most 5. □

Proof of Claim 5. Those points are covered by the balls in Claim 7. We can discard the balls indexed by (negative) i such that $(1+\sqrt{5})(1+t)^i < 1$ as they fit into the unit ball and do not cover any integer points. We can also discard the balls indexed by i with $(1+t)^i > (1+\sqrt{5})B\sqrt{n}$ as they only cover points of magnitude larger than $B\sqrt{n}$. The largest of the remaining balls has radius at most $(5+\sqrt{5})Bt/\sqrt{n} \le \sqrt{n/2}$ so by Claim 6 it contains at most $\exp O(n(Bt)^2 \log 1/Bt)$ points. The number of such balls is $O((1+\ln B)/t)$. □

Claim 8. *An r by r matrix with diagonal 1 and off-diagonal entries between $1-\beta$ and $1-\beta+\beta/2r$ is positive semidefinite and has determinant at least $(\beta/2)^{r-1}(\beta/2 + (1-\beta)r)$.*

Proof. This matrix has the form $(1-\beta)I + \beta J + (\beta/2r)E$ for some E of infinity-norm at most 1 so spectral norm at most r. Here I and J are the identity and the all-ones matrices. The eigenvalues of $(1-\beta)I + \beta J$ are β of multiplicity $r-1$ and $\beta + (1-\beta)r$. The $(\beta/2r)E$ term cannot change them by more than $\beta/2$ each giving the desired bounds. □

Proof of Theorem 8. Let C be a sufficiently large absolute constant. For fixed $\mathbf{x}_1$, the number of $\mathbf{x}_j$ with the required correlation is at most $O_n(1) \cdot \exp O(n\beta B^2 \ln 1/\beta B^2)$ by Claim 5, provided $\beta < 1/CB^2$. Thus the number of requisite configurations $\mathbf{x}_1, \dots, \mathbf{x}_r$ is at most $\exp O(n \ln B + n(r-1)\beta B^2 \ln(1/\beta B^2))$.

If (2) holds for all $\mathbf{x}_i$ then all but a 1/4 fraction of the normalized dot products $\langle \mathbf{a}, \mathbf{x}_i\rangle / \|\mathbf{a}\|\|\mathbf{x}_i\|$ are bounded by 4κ as a ranges over the rows of $\mathbf{A}$. We bound the probability that some fixed subset S of $3rm/4$ correlations are simultaneously at most 4κ. At least half the $\mathbf{a}$s then have normalized dot product at most 4κ with some $r/2$ of the $\mathbf{x}_i$s. The normalized dot products between $\mathbf{a}$ and

$\mathbf{x}_1, \dots, \mathbf{x}_r$ are jointly normal with pairwise correlations between $1-\beta$ and $1-\beta+\beta/2r$. By a union bound the probability that all $r'=r/2$ of them are at most 4κ is at most

$$\frac{1}{(2\pi)^{r'/2}D^{1/2}}(8\kappa)^{r'} \tag{7}$$

where D is the determinant of the correlation matrix restricted to these r' entries. By Claim 8 D is at least $(\beta/2)^{r'}$ which gives an overall probability of at most $O(\kappa^2/\beta)^{r/4}$ for a given $\mathbf{a}$. As the rows of $\mathbf{A}$ are independent the overall probability is at most $O(\kappa^2/\beta)^{mr/8}$. By a union bound, the probability that such an S exists is of the same order as there are at most 2^{mr} choices for S.

The expected number of configurations that solve (2) is therefore at most

$$Z = \exp O(n \ln B + n(r-1)\beta B^2 \ln(1/\beta B^2)) - (\alpha n r/8)\ln(\beta/\kappa^2) + O(\alpha n r),$$

where we replaced m by αn. Setting $\beta = C\kappa^2$, $\log_n Z$ becomes negative in the claimed regime $\alpha \geq C^2(B\kappa)^2 \log C/B\kappa$ when r is sufficiently large. By Markov's inequality $\mathbf{x}_1, \dots, \mathbf{x}_r$ cannot simultaneously solve (2) with high probability in the limit $n \to \infty$. □

5.3 Hardness for Online Algorithms

The ensemble OGP refers to a correlated ensemble of instances $\mathbf{A}_1, \dots, \mathbf{A}_r$. The condition posits that among all $\mathbf{x}_1, \dots, \mathbf{x}_r$ satisfying (6), at least one pair $\mathbf{x}_i$ fails to solve instance $\mathbf{A}_i$ of (2).

The proof of Theorem 8 extends readily to establish ensemble OGP for any collection of instances $\mathcal{A}(t) = (\mathbf{A}_1, \dots, \mathbf{A}_r)$ in which the first t columns are sampled identically and the other $n-t$ are sampled independently across instances, for any t. The reason is that the value of the determinant D in (7) may only increase and all other quantities are preserved.

In contrast, we argue that any "norm-concentrated" online algorithm admits a choice of t for which the resulting solutions do satisfy (6). An algorithm is (ϵ, δ)-norm concentrated if the norm of the output $\mathbf{x}$ is within some $\epsilon\sqrt{n}$-interval of the median except with probability at most δ. Our algorithm Cool is $(1, 0)$ concentrated.

The analysis relies on another mild assumption that can be enforced. We say the algorithm is nice if the (unconditional) distribution of the entries of $\mathbf{x}$ is symmetric and the entries have independent signs.

Claim 9. *Every online algorithm that on input $\mathbf{A}$ produces a solution $\mathbf{x}$ to (2) can be made nice without affecting its success probability and the norm of the output.*

Proof. Choose a random $\mathbf{r} \in \{\pm 1\}^m$. Multiply the columns of $\mathbf{A}$ pointwise by $\mathbf{r}$, run the algorithm then multiply the solution $\mathbf{x}$ pointwise by $\mathbf{r}$. The pointwise multiplications cancel out, preserving the solution. The solution is symmetrized and its entries become sign-independent. □

Claim 10. *Assume $\epsilon \leq \min\{\beta\rho/16r, \beta\rho^{3/2}/16B^2r\}$. For every nice online algorithm that is (ϵ, δ)-concentrated around norm at least $\sqrt{\rho n}$, there exists a (random) t such that its outputs on inputs $\mathcal{A}(t)$ satisfy (6) except with probability $rn\delta + r^2 n 2^{-\Omega(\epsilon^2 n/B^2)}$ for sufficiently large n.*

Proof. $\mathcal{A}(t)$ can be sampled from a common "stem" $\mathbf{A}$ by picking the first t columns in $\mathbf{A}_i$ as in $\mathbf{A}$ and the rest independently. Let $\mathbf{x}$ be the output of the algorithm on input $\mathbf{A}$ and $\mathbf{x}^{\leq t}$ (resp., $\mathbf{A}^{\leq t}$) be its first t entries (resp., columns). In particular, $\mathbf{x}_1(n) = \cdots = \mathbf{x}_n(n) = \mathbf{x}$.

The sequence of random variables $\langle \mathbf{x}_i(t), \mathbf{x}_j(t)\rangle - \|\mathbf{x}^{\leq t}\|^2$ is a martingale with respect to the filtration $\mathbf{A}^{\leq t}$. The martingale property is a consequence of niceness. As its increments are $2B^2$-bounded, by Azuma's inequality $\langle \mathbf{x}_i(t), \mathbf{x}_j(t)\rangle$ is within ϵn of $\|\mathbf{x}^{\leq t}\|^2$ given $\mathbf{A}^{\leq t}$ except with probability $2\exp -\Omega(\epsilon^2 n/B^2)$. By a union bound, this holds simultaneously for all times and all replica pairs except with $\binom{r}{2}n$ times this probability.

By another union bound, the 2-norms of all replicas $\mathbf{x}_i(t)$ at all times t are $\epsilon\sqrt{n}$-concentrated around at least $\sqrt{\rho n}$ except with probability $rn\delta$.

Assume that the exceptional events do not occur. Let $f(t) = \|\mathbf{x}^{\leq t}\|^2/\|\mathbf{x}\|^2$. As $f(0) = 0$, $f(1) = 1$, and f is $B^2/(4\epsilon^2 n)$-Lipschitz, by the intermediate value theorem there must be a (random) time t at which it must be within $B^2/(4\epsilon^2 n)$ of $\beta - \beta/4r$.

We argue that $\mathbf{x}_1(t), \dots, \mathbf{x}_r(t)$ must satisfy (6). For every pair $i \neq j$,

$$\begin{aligned}
\left| f(t) - \frac{\langle \mathbf{x}_i(t), \mathbf{x}_j(t)\rangle}{\|\mathbf{x}_i(t)\|\|\mathbf{x}_j(t)\|} \right| &\leq \|\mathbf{x}^{\leq t}\|^2 \left| \frac{1}{\|\mathbf{x}_i(t)\|\|\mathbf{x}_j(t)\|} - \frac{1}{\|\mathbf{x}\|^2} \right| + \frac{|\|\mathbf{x}^{\leq t}\|^2 - \langle \mathbf{x}_i(t), \mathbf{x}_j(t)\rangle|}{\|\mathbf{x}\|^2} \\
&\leq B^2 n \cdot \frac{2\epsilon}{\rho^{3/2} n} + \frac{\epsilon n}{\rho n} \\
&= \frac{2\epsilon B^2}{\rho^{3/2}} + \frac{\epsilon}{\rho}.
\end{aligned}$$

By the assumption on ϵ this is at most $\beta/6r$, so for sufficiently large n all overlaps are within $\beta/4r$ of $\beta - \beta/4r$ as desired. □

Theorem 9. *For all κ, α, B such that $\kappa \ll 1/B$ and $\alpha \gg (B\kappa)^2 \log 1/B\kappa$ there exists an ϵ such that an online $(\epsilon, o(1/n))$-norm concentrated algorithm cannot solve* (2) *with probability $1 - o(1/n)$.*

Proof. Assume it does. By Claim 9 the algorithm can be assumed nice. Let r and β be the as promised by Theorem 8. Let $\rho = (c\alpha \log 1/\kappa)/\log(\alpha \log 1/\kappa)^{-1}$ for a sufficiently small constant c. By Claim 6, Lemma 3, and a union bound, no solutions to (2) of norm at most $\sqrt{\rho n}$ exist except with probability $O(\kappa)^{\alpha n} = o(1/n)$. By a union bound, all rn instances $\mathcal{A}(1) \cup \cdots \cup \mathcal{A}(n)$ are solved by the algorithm and their solutions have norm at least $\sqrt{\rho n}$ with at least constant probability.

By Claim 10, the outputs of the algorithm on input $\mathcal{A}(t)$ satisfy (6) and (2). Thus (6) and (2) simultaneously hold with constant probability. This contradicts the ensemble OGP extension of Theorem 8. □

Can the norm concentration assumption be removed from Theorem 9? We believe that some stability restriction on the output is needed for the outputs to satisfy a condition like (6). A general online algorithm may be unstable in the sense that changing even a single input can induce an arbitrarily large change in the norm of its output.

5.4 Reduction from Continuous Learning with Errors

Here, we adapt the proof of [VV25] to give a computational lower bound assuming the worst-case hardness of approximate lattice problems. To make the proof simpler (albeit less direct), we reduce from an intermediate problem called Continuous Learning With Errors (CLWE) [BRST21], which is an average-case problem that is known to be as hard as worst-case approximate lattice problems [BRST21, GVV22].

Let $\mathbb{S}^{m-1}$ denote the unit sphere (according to the usual ℓ_2 metric) in $\mathbb{R}^m$. For this section, we let the $\pmod 1$ notation denote taking fractional representatives in $[-1/2, 1/2)$ coordinate-wise. We choose these representatives to ensure that $|x \pmod 1| \leq |x|$ for all $x \in \mathbb{R}$.

Assumption 1 (Continuous Learning With Errors (CLWE))**.** *For all choices of parameters $n(m) \leq \mathrm{poly}(m)$, $\beta(m) \geq 1/\mathrm{poly}(m)$, and $\gamma(m) = m^{\Omega(1)}$, the following holds. For all p.p.t. adversaries, the following two distributions cannot be distinguished with advantage $\Omega(1)$:*

$$\left(\mathbf{A}, \mathbf{s}^\top \mathbf{A} + \mathbf{e}^\top \pmod 1\right), \quad \left(\mathbf{A}, \mathbf{b}^\top\right),$$

where $\mathbf{A} \sim \mathcal{N}(0,1)^{m \times n}$, $\mathbf{s} \sim \gamma \cdot \mathbb{S}^{m-1}$, $\mathbf{e} \sim \mathcal{N}(0, \beta^2)^n$, and $\mathbf{b} \sim [-1/2, 1/2)^n$.

Previous works [Reg09, BLP$^+$13, BRST21, GVV22] show the following facts:

- Assumption 1 holds under the *quantum* worst-case polynomial hardness of the approximate shortest independent vectors problem on lattices (SIVP) and [GVV22, Corollary 2]. For $\gamma \geq 2\sqrt{m}$, this follows from [BRST21, Corollary 3.2], and to get the full range of parameters, from combining [Reg09, Theorem 1.1] and [GVV22, Corollary 2].

- Assumption 1 holds under the *classical* worst-case polynomial hardness of the gap shortest vector problem on lattices (GapSVP). This follows from combining [BLP$^+$13, Theorem 1.1] and [GVV22, Corollary 2].

In both cases, Assumption 1 holds under the worst-case polynomial hardness of approximately solving lattice problems.

Theorem 10. *Under Assumption 1, for all $\varepsilon > 0$ and $B, n \leq \mathrm{poly}(m)$, there does not exist a p.p.t. algorithm for CHV that succeeds with probability at least $2/3$ for*

$$\kappa = O\left(\frac{1}{Bn^{1/2+\varepsilon}}\right).$$

Proof. Suppose for contradiction that there exists $\varepsilon > 0$ and $B, n \leq \mathrm{poly}(m)$ with a p.p.t. algorithm $\mathcal{A}$ succeeding with probability at least 2/3. We then violate Assumption 1 with $\beta = 1/(B \cdot n)$ and $\gamma = n^{\varepsilon/2}$ by the algorithm $\mathcal{A}'$ described as follows. On input $(\mathbf{A}, \mathbf{b}^\top)$, $\mathcal{A}'$ runs $\mathbf{x} \leftarrow \mathcal{A}(\mathbf{A})$, ensures that $\mathbf{x} \in ([-B, B] \cap \mathbb{Z})^n$ and $\|\mathbf{A}\mathbf{x}\|_2 < \kappa \|\mathbf{x}\|_2 \sqrt{m}$, and then outputs 1 if and only if

$$\left|\mathbf{b}^\top \mathbf{x} \pmod 1\right| < \frac{1}{4}. \tag{8}$$

If $\mathbf{x}$ is not a solution to CHV, then the algorithm outputs 0.

We now analyze the performance of $\mathcal{A}'$. For the "null" case of CLWE, where $\mathbf{b} \sim [-1/2, 1/2)^n$, since $\mathbf{x} \neq 0$ and $\mathbf{x} \in \mathbb{Z}^n$, it follows that $\mathbf{b}^\top \mathbf{x} \pmod 1$ is distributed uniformly randomly in $[-1/2, 1/2)$. Therefore, (8) holds with probability 1/2 (conditioned on $\mathbf{x}$ being a valid solution).

For the planted case, we know $\mathbf{b}^\top = \mathbf{s}^\top \mathbf{A} + \mathbf{e}^\top \pmod 1$ for $\mathbf{s} \sim \gamma \cdot \mathbb{S}^{m-1}$, $\mathbf{e} \sim \mathcal{N}(0, \beta^2)^n$. By spherical symmetry of the Gaussian, we can write $\mathbf{s}$ as $\gamma \cdot \mathbf{s}'/\|\mathbf{s}'\|_2$ for $\mathbf{s}' \sim \mathcal{N}(0,1)^n$. Since $\mathbf{x} \in \mathbb{Z}^n$, we then have

$$\begin{aligned}\left|\mathbf{b}^\top \mathbf{x} \pmod 1\right| &= \left|\mathbf{s}^\top \mathbf{A}\mathbf{x} + \mathbf{e}^\top \mathbf{x} \pmod 1\right| \\ &\le \left|\mathbf{s}^\top \mathbf{A}\mathbf{x}\right| + \left|\mathbf{e}^\top \mathbf{x}\right| \\ &= \frac{\gamma}{\|\mathbf{s}'\|_2}\left|(\mathbf{s}')^\top \mathbf{A}\mathbf{x}\right| + \left|\mathbf{e}^\top \mathbf{x}\right| \\ &= \frac{\gamma}{\|\mathbf{s}'\|_2}\left|v_1\right| + \left|v_2\right|,\end{aligned}$$

where $v_1 \sim \mathcal{N}(0, \|\mathbf{A}\mathbf{x}\|_2^2)$ and $v_2 \sim \mathcal{N}(0, \beta^2\|\mathbf{x}\|_2^2)$, by Gaussianity and independence of $\mathbf{s}'$ and $\mathbf{e}$ from $\mathbf{A}$ and $\mathbf{x}$. By standard tail bounds on the (univariate) Gaussian distribution, with probability at least 99/100, we know that $|v_1| \le O\left(\|\mathbf{A}\mathbf{x}\|_2\right)$ and $|v_2| \le O(\beta\|\mathbf{x}\|_2)$. Also, by Lemma 3, we know that $\|\mathbf{s}'\|_2 \ge \Omega(\sqrt{m})$ with probability $1 - o(1)$. Putting these all together and continuing the inequality, with probability at least $99/100 - o(1)$ (conditioned on $\mathbf{x}$ being a valid solution), we have

$$\begin{aligned}\left|\mathbf{b}^\top \mathbf{x} \pmod 1\right| &\le \frac{\gamma}{\|\mathbf{s}'\|_2}\left|v_1\right| + \left|v_2\right| \\ &\le O\left(\frac{\gamma\|\mathbf{A}\mathbf{x}\|_2}{\sqrt{m}} + \beta\|\mathbf{x}\|_2\right) \\ &\le O\left(\frac{\gamma\kappa\|\mathbf{x}\|_2\sqrt{m}}{\sqrt{m}} + \beta\|\mathbf{x}\|_2\right) \\ &= O\left(\gamma\kappa + \beta\right)\cdot\|\mathbf{x}\|_2. \\ &\le O\left(\gamma\kappa + \beta\right)\cdot B\sqrt{n}.\end{aligned}$$

For $\mathcal{A}'$ to get advantage $\Omega(1)$, it suffices to show that $(\gamma\kappa + \beta)B\sqrt{n} = o(1)$. We directly have $\beta = o(1/(B\sqrt{n}))$ by the way we have set β. For the last remaining term, we can plug in our settings of γ and κ to see that

$$\gamma\kappa B\sqrt{n} = n^{\varepsilon/2}\cdot O\left(\frac{1}{Bn^{1/2+\varepsilon}}\right)\cdot B\sqrt{n} = O\left(\frac{1}{n^{\varepsilon/2}}\right) = o(1),$$

as desired. □

6 Robust Locality Sensitive Hash Functions

6.1 Definition

Here, we define robust locality-sensitive hash functions, specialized to the Euclidean norm, following [BLV19].

Definition 3 (Robust Locality Sensitive Hash Functions)**.** *For natural numbers n and $m = m(n) < n$, let $\mathcal{X}_n \subseteq \mathbb{R}^n$ and $\mathcal{Y}_n \subseteq \mathbb{R}^m$ be finite sets. A* robust locality sensitive hash function *with approximation factors $\alpha = \alpha(n), \beta = \beta(n)$ consists of p.p.t. algorithms* (KeyGen, Hash) *with the following syntax:*

- KeyGen$(1^n) \to \mathsf{k}$. *This algorithm is randomized and outputs some key $\mathsf{k} \in \{0,1\}^{\mathrm{poly}(n)}$.*

- Hash : $\{0,1\}^{\mathrm{poly}(n)} \times \mathcal{X}_n \to \mathcal{Y}_n$. *This algorithm is deterministic. As shorthand, we will write* $\mathrm{Hash}_{\mathsf{k}} : \mathcal{X}_n \to \mathcal{Y}_n$ *to denote the hash function* $\mathrm{Hash}(\mathsf{k}, -)$ *for fixed key* $\mathsf{k} \in \{0,1\}^{\mathrm{poly}(n)}$.

Moreover, we require the following three properties:

1. ***Compression**: We have* $|\mathcal{Y}_n| \le \frac{1}{2}|\mathcal{X}_n|$. *That is, the function* $\mathrm{Hash}_{\mathsf{k}} : \mathcal{X}_n \to \mathcal{Y}_n$ *is compressing by at least a factor of* 2 *(typically, significantly more).*
2. ***Statistical Non-Expansion**:*
$$\Pr_{\mathsf{k} \sim \mathrm{KeyGen}(1^n)} \left[\exists \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}_n : \|\mathrm{Hash}_{\mathsf{k}}(\mathbf{x}_1) - \mathrm{Hash}_{\mathsf{k}}(\mathbf{x}_2)\|_2 > \alpha \cdot \|\mathbf{x}_1 - \mathbf{x}_2\|_2\right] = \mathsf{negl}(n).$$
3. ***Computational Non-Contraction**: For all p.p.t. adversaries* $\mathcal{A}$,
$$\Pr_{\mathsf{k} \sim \mathrm{KeyGen}(1^n)} \left[(\mathbf{x}_1, \mathbf{x}_2) \leftarrow \mathcal{A}(1^n, \mathsf{k}) : \begin{array}{c} \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}_n \wedge \\ \|\mathrm{Hash}_{\mathsf{k}}(\mathbf{x}_1) - \mathrm{Hash}_{\mathsf{k}}(\mathbf{x}_2)\|_2 < \beta \|\mathbf{x}_1 - \mathbf{x}_2\|_2 \end{array}\right] = \mathsf{negl}(n),$$
where the probability is also taken over the internal randomness of $\mathcal{A}$.

We refer to the quantity $\xi = \alpha/\beta$ *as the* distortion *of the hash function.*

We note that this definition is, in particular, stronger than a collision-resistant hash function.[3] To see this, note that if $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}_n$ form a collision, then $\mathrm{Hash}_{\mathsf{k}}(\mathbf{x}_1) = \mathrm{Hash}_{\mathsf{k}}(\mathbf{x}_2)$ while $\|\mathbf{x}_1 - \mathbf{x}_2\|_2 > 0$, which violates computational non-contraction. Therefore, computational assumptions are necessary to achieve the definition.

6.2 Preliminaries

Let $\mathrm{Ball}_m(r, \mathbf{y}) \subseteq \mathbb{R}^m$ denote the ball of radius r (according to the usual ℓ_2 norm) in dimension m centered at $\mathbf{y} \in \mathbb{R}^m$. If $\mathbf{y}$ is ommitted, it is taken to be the all 0s vector.

Lemma 1 (E.g., Lemma 16 in [KF15]). *For all* $m \in \mathbb{N}$ *and* $r \in \mathbb{R}_{>0}$, *we have the bound*
$$|\mathrm{Ball}_m(r) \cap \mathbb{Z}^m| \le \mathrm{Vol}\left(\mathrm{Ball}_m\left(r + \frac{\sqrt{m}}{2}\right)\right)$$

Proof. For all $\mathbf{y} \in \mathrm{Ball}_m(r) \cap \mathbb{Z}^m$, consider the (open) cube $\mathbf{y} + (-1/2, 1/2)^m$. Note that since $\mathbf{y} \in \mathbb{Z}^m$, all such cubes are disjoint. Since all cubes have volume 1, we have
$$\begin{aligned} |\mathrm{Ball}_m(r) \cap \mathbb{Z}^m| &= \mathrm{Vol}\left(\bigsqcup_{\mathbf{y} \in \mathrm{Ball}_m(r) \cap \mathbb{Z}^m} (\mathbf{y} + (-1/2, 1/2))^m\right) \\ &\le \mathrm{Vol}\left(\bigcup_{\mathbf{y} \in \mathrm{Ball}_m(r) \cap \mathbb{Z}^m} \mathrm{Ball}_m\left(\frac{\sqrt{m}}{2}, \mathbf{y}\right)\right) \\ &\le \mathrm{Vol}\left(\mathrm{Ball}_m\left(r + \frac{\sqrt{m}}{2}\right)\right), \end{aligned}$$
as desired. □

[3] The only syntactic difference is that the codomain here is not expressed as $\{0,1\}^\ell$ for some ℓ, but rather some efficiently recognizable finite set $\mathcal{Y}_n$. By considering a direct binary encoding of $([-r, r] \cap \gamma\mathbb{Z})^m \supseteq \mathcal{Y}_n$, one can make the codomain $\{0,1\}^\ell$ with a slight loss in parameters.

Corollary 3. *For all $m \in \mathbb{N}$ and all $\gamma, r \in \mathbb{R}_{>0}$, we have the bound*

$$|\mathrm{Ball}_m(r) \cap \gamma\mathbb{Z}^m| \leq \left(\frac{r\sqrt{2\pi e}}{\gamma\sqrt{m}} + \frac{\sqrt{2\pi e}}{2}\right)^m.$$

Proof. By a scaling argument (by $1/\gamma$), we know

$$|\mathrm{Ball}_m(r) \cap \gamma\mathbb{Z}^m| = |\mathrm{Ball}_m(r/\gamma) \cap \mathbb{Z}^m|.$$

Plugging in Lemma 1,

$$|\mathrm{Ball}_m(r) \cap \gamma\mathbb{Z}^m| \leq \mathrm{Vol}\left(\mathrm{Ball}_m\left(\frac{r}{\gamma} + \frac{\sqrt{m}}{2}\right)\right).$$

By using a standard bound that

$$\mathrm{Vol}(\mathrm{Ball}_m(R)) \leq \left(\sqrt{\frac{2\pi e}{m}}\right)^m R^m,$$

we get

$$|\mathrm{Ball}_m(r) \cap \gamma\mathbb{Z}^m| \leq \left(\sqrt{\frac{2\pi e}{m}}\right)^m \cdot \left(\frac{r}{\gamma} + \frac{\sqrt{m}}{2}\right)^m = \left(\frac{r\sqrt{2\pi e}}{\gamma\sqrt{m}} + \frac{\sqrt{2\pi e}}{2}\right)^m,$$

as desired. □

For a matrix $\mathbf{A}$, we let $\|\mathbf{A}\|_2$ denote the standard spectral norm of $\mathbf{A}$, i.e., the largest singular value of $\mathbf{A}$.

Lemma 2 (As in [RV10])**.** *For all $t > 0$, we have*

$$\Pr_{\mathbf{A} \sim \mathcal{N}(0,1)^{m \times n}}\left[\|\mathbf{A}\|_2 \leq \sqrt{m} + \sqrt{n} + t\right] \geq 1 - 2e^{-t^2/2}.$$

In particular, for $n > m$ and setting $t = \sqrt{n}$, we have

$$\Pr_{\mathbf{A} \sim \mathcal{N}(0,1)^{m \times n}}\left[\|\mathbf{A}\|_2 \leq 3\sqrt{n}\right] \geq 1 - 2e^{-n/2}.$$

6.3 Construction

Theorem 11. *Suppose that* CHV *is hard for parameters n, m, B, κ. Then, for $r = 4Bn/\sqrt{m}$ and $\gamma = \kappa/(2\sqrt{m})$, there exists a universal constant C and a robust locality sensitive hash function for $\mathcal{X}_n = ([0, B] \cap \mathbb{Z})^n$ and $\mathcal{Y}_n = \mathrm{Ball}_m(r) \cap \gamma\mathbb{Z}^m$ with parameters*

$$\alpha = 4\sqrt{\frac{n}{m}} \quad \text{and} \quad \beta = \frac{\kappa}{2},$$

as long as

$$(B+1)^n > \left(\frac{CBn}{\kappa\sqrt{m}}\right)^m.$$

In particular, the distortion is

$$\xi = \frac{\alpha}{\beta} = \frac{8}{\kappa}\sqrt{\frac{n}{m}}.$$

We will use the function $\lfloor \cdot \rceil_\gamma : \mathbb{R}^m \to \gamma\mathbb{Z}^m$ to denote coordinate-wise rounding down to the nearest multiple of γ.

Proof of Theorem 11. We construct a robust locality sensitive hash function as described in Figure 5.

Construction of Robust Locality Sensitive Hash Function

- KeyGen(1^n): Sample $\mathbf{A} \sim \mathcal{N}(0, 1/m)^{m \times n}$ and output $\mathsf{k} = \mathbf{A} \in \mathbb{R}^{m \times n}$.[a]
- Hash$_\mathsf{k}$ $(\mathbf{x} \in ([0, B] \cap \mathbb{Z})^n)$: Parse $\mathbf{A} \in \mathbb{R}^{m \times n}$ from k and output

$$\mathbf{y} = \begin{cases} \lfloor \mathbf{Ax} \rceil_\gamma & \text{if } \| \lfloor \mathbf{Ax} \rceil_\gamma \|_2 \le r, \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

By construction, note that $\mathbf{y} \in \mathrm{Ball}_m(r) \cap \gamma\mathbb{Z}^m$.

[a] Technically, we need to discretize $\mathbf{A}$ to match the syntax of Definition 3. Using $O(\log(n/\kappa))$ bits of precision per entry is sufficient for all of the analysis.

Figure 5: The construction of the robust locality sensitive hash function, as used in Theorem 11. See Definition 3 for the syntax of robust locality sensitive hash functions.

For simplicity, we assume for now that it always holds that $\| \lfloor \mathbf{Ax} \rceil_\gamma \|_2 \le r$, i.e., that $\mathrm{Hash}_\mathsf{k}(\mathbf{x}) = \lfloor \mathbf{Ax} \rceil_\gamma$ for all $\mathbf{x} \in \mathcal{X}_n$. Later in Claim 13, we show that with $1 - \mathsf{negl}(n)$ probability over $\mathbf{A}$, this indeed holds, allowing us to add a $\mathsf{negl}(n)$ term back into the proofs of Items 1 to 3 in Definition 3. We prove that Item 3, Item 2, and Item 1 hold, in that order.

We begin by showing Item 3.

Claim 11 (Computational Non-Contraction)**.** *Item 3 in Definition 3 holds with $\beta = \kappa/2$. That is, assuming the hardness of* CHV*, no p.p.t. algorithm can output $\mathbf{y}, \mathbf{z} \in \mathcal{X}_n$ such that $\|\mathrm{Hash}_\mathsf{k}(\mathbf{y}) - \mathrm{Hash}_\mathsf{k}(\mathbf{z})\|_2 < \kappa/2 \cdot \|\mathbf{y} - \mathbf{z}\|_2$ with non-negligible probability.*

Proof. The reduction from CHV is direct. For an instance of CHV with matrix $\mathbf{A}' \sim \mathcal{N}(0, 1)^{m \times n}$, let $\mathsf{k} = \mathbf{A} = \frac{1}{\sqrt{m}}\mathbf{A}' \sim \mathcal{N}(0, 1/m)^{m \times n}$ be the key for the robust locality sensitive hash function.

For any such violating pair $\mathbf{y}, \mathbf{z}$ (where it must be the case that $\mathbf{y} \neq \mathbf{z}$), the reduction outputs $\mathbf{x} = \mathbf{y} - \mathbf{z} \in ([-B, B] \cap \mathbb{Z})^n$. Let $\mathbf{e}_1 = \mathbf{Ay} - \lfloor \mathbf{Ay} \rceil_\gamma \in [0, \gamma)^m$, $\mathbf{e}_2 = \mathbf{Az} - \lfloor \mathbf{Az} \rceil_\gamma \in [0, \gamma)^m$. We have

$$\begin{aligned} \frac{1}{\sqrt{m}} \|\mathbf{A}'\mathbf{x}\|_2 = \|\mathbf{Ax}\|_2 &= \|\mathbf{Ay} - \mathbf{Az}\|_2 \\ &= \| \lfloor \mathbf{Ay} \rceil_\gamma + \mathbf{e}_1 - (\lfloor \mathbf{Az} \rceil_\gamma + \mathbf{e}_2)\|_2 \\ &\le \| \lfloor \mathbf{Ay} \rceil_\gamma - \lfloor \mathbf{Az} \rceil_\gamma \|_2 + \|\mathbf{e}_1 - \mathbf{e}_2\|_2 \\ &= \|\mathrm{Hash}_\mathsf{k}(\mathbf{y}) - \mathrm{Hash}_\mathsf{k}(\mathbf{z})\|_2 + \|\mathbf{e}_1 - \mathbf{e}_2\|_2 \\ &\le \|\mathrm{Hash}_\mathsf{k}(\mathbf{y}) - \mathrm{Hash}_\mathsf{k}(\mathbf{z})\|_2 + \gamma\sqrt{m} \\ &< \frac{\kappa}{2} \cdot \|\mathbf{y} - \mathbf{z}\|_2 + \frac{\kappa}{2} \\ &\le \frac{\kappa}{2} \cdot \|\mathbf{y} - \mathbf{z}\|_2 + \frac{\kappa}{2} \cdot \|\mathbf{y} - \mathbf{z}\|_2 \\ &= \kappa \cdot \|\mathbf{y} - \mathbf{z}\|_2, \end{aligned}$$

where the last inequality comes from the fact that $\mathbf{y} \neq \mathbf{z}$ and $\mathbf{y}, \mathbf{z} \in \mathbb{Z}^n$. Multiplying both sides by $\sqrt{m}$ gives

$$\|\mathbf{A}'\mathbf{x}\|_2 < \kappa\sqrt{m} \cdot \|\mathbf{y} - \mathbf{z}\|_2 = \kappa\sqrt{m} \cdot \|\mathbf{x}\|_2,$$

solving CHV with instance $\mathbf{A}'$. □

Next, we show Item 2.

Claim 12 (Statistical Non-Expansion)**.** *Item 2 in Definition 3 holds with*

$$\alpha = 4\sqrt{\frac{n}{m}}.$$

More explicitly, with probability at least $1 - 2e^{-n/2}$*, for all* $\mathbf{y}, \mathbf{z} \in \mathcal{X}_n$*,*

$$\|\mathrm{Hash}_{\mathsf{k}}(\mathbf{y}) - \mathrm{Hash}_{\mathsf{k}}(\mathbf{z})\|_2 < 4\sqrt{\frac{n}{m}} \cdot \|\mathbf{y} - \mathbf{z}\|_2.$$

Proof. Up to rounding concerns, this is equivalent to upper bounding $\|\mathbf{A}\mathbf{x}\|_2/\|\mathbf{x}\|_2$ over all $\mathbf{x} \in [-B, B]^n \cap \mathbb{Z}^n \setminus \{\mathbf{0}\}$. We can bound this directly by the spectral norm of $\mathbf{A}$:

$$\max_{\mathbf{x}\in([-B,B]\cap\mathbb{Z})^n\setminus\{\mathbf{0}\}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \sup_{\mathbf{x}\in\mathbb{R}^n\setminus\{\mathbf{0}\}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \|\mathbf{A}\|_2.$$

Since $\sqrt{m} \cdot \mathbf{A} \sim \mathcal{N}(0,1)^{m\times n}$, by Lemma 2, it follows that with probability at least $1 - 2e^{-n/2}$,

$$\max_{\mathbf{x}\in([-B,B]\cap\mathbb{Z})^n\setminus\{\mathbf{0}\}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \|\mathbf{A}\|_2 \leq 3\sqrt{\frac{n}{m}}.$$

Finally, we incorporate the rounding. For $\mathbf{y} = \mathbf{z}$, the desired inequality is trivially true, so we assume $\mathbf{y} \neq \mathbf{z}$. Let $\mathbf{e}_1 = \mathbf{A}\mathbf{y} - \lfloor \mathbf{A}\mathbf{y} \rfloor_\gamma \in [0, \gamma)^m$ and $\mathbf{e}_2 = \mathbf{A}\mathbf{z} - \lfloor \mathbf{A}\mathbf{z} \rfloor_\gamma \in [0, \gamma)^m$. For $\mathbf{x} = \mathbf{y} - \mathbf{z} \neq \mathbf{0}$, we have

$$\begin{aligned}
\|\mathrm{Hash}_{\mathsf{k}}(\mathbf{y}) - \mathrm{Hash}_{\mathsf{k}}(\mathbf{z})\|_2 &= \|\lfloor \mathbf{A}\mathbf{y} \rfloor_\gamma - \lfloor \mathbf{A}\mathbf{z} \rfloor_\gamma\|_2 \\
&= \|\mathbf{A}\mathbf{y} - \mathbf{e}_1 - (\mathbf{A}\mathbf{z} - \mathbf{e}_2)\|_2 \\
&\leq \|\mathbf{A}\mathbf{y} - \mathbf{A}\mathbf{z}\|_2 + \|\mathbf{e}_1 - \mathbf{e}_2\|_2 \\
&= \|\mathbf{A}\mathbf{x}\|_2 + \|\mathbf{e}_1 - \mathbf{e}_2\|_2 \\
&\leq 3\sqrt{\frac{n}{m}} \cdot \|\mathbf{x}\|_2 + \gamma\sqrt{m} \\
&\leq 3\sqrt{\frac{n}{m}} \cdot \|\mathbf{x}\|_2 + \gamma\sqrt{m} \cdot \|\mathbf{x}\|_2 \\
&= \left(3\sqrt{\frac{n}{m}} + \frac{\kappa}{2}\right) \cdot \|\mathbf{x}\|_2,
\end{aligned}$$

where the last inequality comes from the fact that $\mathbf{x} \neq 0$. Since $m < n$ and $\kappa < 1$, we can continue the above inequality to see that

$$\begin{aligned}\|\mathrm{Hash}_{\mathsf{k}}(\mathbf{y}) - \mathrm{Hash}_{\mathsf{k}}(\mathbf{z})\|_2 &\leq \left(3\sqrt{\frac{n}{m}} + \frac{\kappa}{2}\right) \cdot \|\mathbf{x}\|_2 \\ &\leq \left(3\sqrt{\frac{n}{m}} + \frac{1}{2}\right) \cdot \|\mathbf{x}\|_2 \\ &< 4\sqrt{\frac{n}{m}} \cdot \|\mathbf{x}\|_2,\end{aligned}$$

as desired. □

As a consequence of Claim 12, we have the following:

Claim 13 (Output Norm Bound)**.** *With probability at least* $1 - (2B+1)^{-n}$*, for all* $\mathbf{y} \in \mathcal{X}_n$*,*

$$\|\lfloor \mathbf{A}\mathbf{y} \rceil_\gamma\|_2 < \frac{4Bn}{\sqrt{m}}.$$

Proof. This follows from Claim 12 by setting $\mathbf{z} = \mathbf{0}$ and using the bound $\|\mathbf{y}\|_2 \leq B\sqrt{n}$. □

Lastly, we show Item 1.

Claim 14 (Compression)**.** *Item 1 holds in Definition 3. More specifically, the function* $\mathrm{Hash}_{\mathsf{k}}$ *is compressing (by a factor of at least* 2*) if*

$$(B+1)^n > 2\left(\frac{Bn\sqrt{128\pi e}}{\kappa\sqrt{m}} + \frac{\sqrt{2\pi e}}{2}\right)^m.$$

Proof. Recall that $\mathcal{X}_n = ([0,B] \cap \mathbb{Z})^n$ and $\mathcal{Y}_n = \mathrm{Ball}_m(r) \cap \gamma\mathbb{Z}^m$. The domain has cardinality

$$|\mathcal{X}_n| = |([0,B] \cap \mathbb{Z})^n| = (B+1)^n.$$

By Corollary 3, we know the codomain $\mathcal{Y}_n$ has cardinality

$$|\mathcal{Y}_n| = |\mathrm{Ball}_m(r) \cap \gamma\mathbb{Z}^m| \leq \left(\frac{r\sqrt{2\pi e}}{\gamma\sqrt{m}} + \frac{\sqrt{2\pi e}}{2}\right)^m.$$

Plugging in our value of r from Claim 13, this becomes

$$|\mathcal{Y}_n| \leq \left(\frac{Bn\sqrt{32\pi e}}{\gamma m} + \frac{\sqrt{2\pi e}}{2}\right)^m.$$

Therefore, since $\gamma = \kappa/(2\sqrt{m})$, for the function to be compressing by a factor of at least 2, it suffices that

$$(B+1)^n > 2\left(\frac{Bn\sqrt{128\pi e}}{\kappa\sqrt{m}} + \frac{\sqrt{2\pi e}}{2}\right)^m.$$

□

□

7 Statistical Threshold

In this section, we describe the statistical threshold for CHV by classifying when the expected number of solutions $\mathbf{x}$ is at least 1. For more precise bounds that take into account more than the first moment (for related variants of this problem), we direct the reader to the works by Aubin, Perkins and Zdeborová [APZ19], Perkins and Xu [PX21], and Abbe, Li and Sly [ALS21], which confirm that the first moment bound accurately gives a sharp statistical threshold for binary symmetric perceptron models.

We begin by stating a lower tail bound on the χ^2_m distribution.

Lemma 3. *There exist universal constants $C_1, C_2 > 0$ such that for all $\kappa \le 1/2$ and $m \in \mathbb{N}$,*

$$(C_1 \cdot \kappa)^m \le \Pr_{Z \sim \chi^2_m}\left[Z \le \kappa^2 m\right] \le (C_2 \cdot \kappa)^m.$$

Proof. We first show the lower bound. Recall that we can characterize $Z \sim \chi^2_m$ by

$$Z = \sum_{i \in [m]} X_i^2$$

for i.i.d. $X_i \sim \mathcal{N}(0,1)$. If for all $i \in [m]$ it holds that $|X_i| \le \kappa$, then

$$Z = \sum_{i \in [m]} X_i^2 \le \kappa^2 m.$$

Therefore,

$$\Pr_{Z \sim \chi^2_m}\left[Z \le \kappa^2 m\right] \ge \Pr\left[\forall i \in [m],\ |X_i| \le \kappa\right] = \Pr_{X \sim \mathcal{N}(0,1)}\left[|X| \le \kappa\right]^m \ge (C_1 \cdot \kappa)^m,$$

for some universal constant C_1, where the right-hand-most inequality holds since the measure of $\mathcal{N}(0,1)$ is at least $e^{-1/8}/\sqrt{2\pi}$ on $[-1/2, 1/2]$ and $\kappa \le 1/2$.

We now show the upper bound. Using the moment-generating function of the χ^2_m distribution and the proof of Cramér's theorem, for all $\kappa < 1$, we have

$$\begin{aligned}\Pr_{Z \sim \chi^2_m}\left[Z \le \kappa^2 m\right] &\le \exp\left(m \cdot \frac{\ln(\kappa^2) - \kappa^2 + 1}{2}\right) \\ &= \exp\left(m \cdot \left(\ln(\kappa) - \frac{\kappa^2}{2} + \frac{1}{2}\right)\right) \\ &= \left(\kappa \cdot \exp\left(-\frac{\kappa^2}{2} + \frac{1}{2}\right)\right)^m \\ &\le (\kappa \cdot C_2)^m\end{aligned}$$

for some universal constant C_2, since the function $\exp(-\kappa^2/2 + 1/2)$ lies in the interval $[e^{3/8}, e^{1/2}]$ for all $\kappa \in [0, 1/2]$. □

We proceed to compute the statistical bound. For $\mathbf{A} \sim \mathcal{N}(0,1)^{m \times n}$ and fixed $\mathbf{x} \in [-B, B]^n \cap \mathbb{Z}^n$, let $I_{\mathbf{x}}$ be the indicator random variable given by

$$I_{\mathbf{x}} := \mathbb{1}\left[\|\mathbf{A}\mathbf{x}\|_2 < \kappa\sqrt{m} \cdot \|\mathbf{x}\|_2\right] \in \{0, 1\}.$$

By linearity of expectation, we know that the expected number of solutions is given by

$$\begin{aligned}\mathbb{E}\left[\left|\left\{\mathbf{x} \in [-B,B]^n \cap \mathbb{Z}^n : \|\mathbf{A}\mathbf{x}\|_2 < \kappa\sqrt{m} \cdot \|\mathbf{x}\|_2\right\}\right|\right] &= \sum_{\mathbf{x}\in[-B,B]^n\cap\mathbb{Z}^n} \mathbb{E}\left[I_{\mathbf{x}}\right] \\ &= \sum_{\mathbf{x}\in[-B,B]^n\cap\mathbb{Z}^n} \Pr\left[\|\mathbf{A}\mathbf{x}\|_2 < \kappa\sqrt{m} \cdot \|\mathbf{x}\|_2\right].\end{aligned}$$

For $\mathbf{A} \sim \mathcal{N}(0,1)^{m\times n}$ and fixed $\mathbf{x} \in [-B,B]^n \cap \mathbb{Z}^n \setminus \{0^n\}$, by standard properties of the Gaussian distribution, the distribution of $\mathbf{A}\mathbf{x}$ is $\mathcal{N}(0, \|\mathbf{x}\|_2^2)^m$. Therefore, $\mathbf{A}\mathbf{x}/\|\mathbf{x}\|_2 \sim \mathcal{N}(0,1)^m$, so

$$\frac{\|\mathbf{A}\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} \sim \chi_m^2.$$

It follows that

$$\Pr\left[\|\mathbf{A}\mathbf{x}\|_2 < \kappa\sqrt{m} \cdot \|\mathbf{x}\|_2\right] = \Pr\left[\frac{\|\mathbf{A}\mathbf{x}\|_2^2}{\|\mathbf{x}\|_2^2} < \kappa^2 m\right] = \Pr_{Z\sim\chi_m^2}\left[Z < \kappa^2 m\right].$$

For $\kappa \leq 1/2$, by Lemma 3, we have

$$\Pr_{Z\sim\chi_m^2}\left[Z < \kappa^2 m\right] = \Theta\left(\kappa\right)^m.$$

Since $\mathbf{x} = 0^n$ is never a solution, it follows that

$$\begin{aligned}\mathbb{E}\left[\left|\left\{\mathbf{x} \in [-B,B]^n \cap \mathbb{Z}^n : \|\mathbf{A}\mathbf{x}\|_2 < \kappa\sqrt{m} \cdot \|\mathbf{x}\|_2\right\}\right|\right] &= ((2B+1)^n - 1) \cdot \Pr_{Z\sim\chi_m^2}\left[Z < \kappa^2 m\right] \\ &= ((2B+1)^n - 1) \cdot \Theta(\kappa)^m.\end{aligned}$$

Setting this to 1 gives the statistical threshold

$$\kappa = \Theta\left((2B+1)^{-n/m}\right).$$

Acknowledgements. We thank Avi Wigderson for advice, and we are grateful to Aayush Jain, Thuy-Duong Vuong, and Justin Y. Chen for useful discussions.

References

[ABD$^+$21] Noga Alon, Omri Ben-Eliezer, Yuval Dagan, Shay Moran, Moni Naor, and Eylon Yogev. Adversarial laws of large numbers and optimal regret in online classification. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 447–455. ACM, 2021. 1, 6

[AC09] Nir Ailon and Bernard Chazelle. The fast Johnson–Lindenstrauss transform and approximate nearest neighbors. *SIAM J. Comput.*, 39(1):302–322, 2009. 1

[ACSS24] Idan Attias, Edith Cohen, Moshe Shechner, and Uri Stemmer. A framework for adversarial streaming via differential privacy and difference estimators. *Algorithmica*, 86(11):3339–3394, 2024. 1, 6

[ALS21] Emmanuel Abbe, Shuangping Li, and Allan Sly. Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron, 2021. 6, 28

[ALS22] Emmanuel Abbe, Shuangping Li, and Allan Sly. Binary perceptron: efficient algorithms can find solutions in a rare well-connected cluster. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 860–873. ACM, 2022. 6

[APZ19] Benjamin Aubin, Will Perkins, and Lenka Zdeborová. Storage capacity in symmetric binary perceptrons. *Journal of Physics A: Mathematical and Theoretical*, 52(29):294003, June 2019. 6, 28

[BDVLZ20] Carlo Baldassi, Riccardo Della Vecchia, Carlo Lucibello, and Riccardo Zecchina. Clustering of solutions in the symmetric binary perceptron. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(7):073303, 2020. 6

[BEAKZ24] Damien Barbier, Ahmed El Alaoui, Florent Krzakala, and Lenka Zdeborová. On the atypical solutions of the symmetric binary perceptron. *Journal of Physics A: Mathematical and Theoretical*, 57(19):195202, April 2024. 6

[BEKMR23] Omri Ben-Eliezer, Esty Kelman, Uri Meir, and Sofya Raskhodnikova. Property testing with online adversaries, 2023. 1, 6

[BKM$^+$22] Amos Beimel, Haim Kaplan, Yishay Mansour, Kobbi Nissim, Thatchaphol Saranurak, and Uri Stemmer. Dynamic algorithms against an adaptive adversary: generic constructions and lower bounds. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 1671–1684. ACM, 2022. 1, 6

[BLP$^+$13] Zvika Brakerski, Adeline Langlois, Chris Peikert, Oded Regev, and Damien Stehlé. Classical hardness of learning with errors. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 575–584. ACM, 2013. 21

[BLV19] Elette Boyle, Rio LaVigne, and Vinod Vaikuntanathan. Adversarially robust property-preserving hash functions. In Avrim Blum, editor, *10th Innovations in Theoretical Computer Science Conference, ITCS 2019, January 10-12, 2019, San Diego, California, USA*, volume 124 of *LIPIcs*, pages 16:1–16:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019. 5, 22

[BR23] Andrej Bogdanov and Alon Rosen. Nondeterministic Interactive Refutations for Nearest Boolean Vector. In Kousha Etessami, Uriel Feige, and Gabriele Puppis, editors, *50th International Colloquium on Automata, Languages, and Programming (ICALP 2023)*, volume 261 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 28:1–28:14, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 6

[BRST21] Joan Bruna, Oded Regev, Min Jae Song, and Yi Tang. Continuous LWE. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM*

SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021, pages 694–707. ACM, 2021. 8, 20, 21

[BS20] Nikhil Bansal and Joel H. Spencer. On-line balancing of random inputs. *Random Struct. Algorithms*, 57(4):879–891, 2020. 2, 3, 4, 10, 11

[BY20] Omri Ben-Eliezer and Eylon Yogev. The adversarial robustness of sampling. In Dan Suciu, Yufei Tao, and Zhewei Wei, editors, *Proceedings of the 39th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2020, Portland, OR, USA, June 14-19, 2020*, pages 49–62. ACM, 2020. 1, 6

[CJN18] Michael B. Cohen, T. S. Jayram, and Jelani Nelson. Simple analyses of the sparse Johnson-Lindenstrauss transform. In Raimund Seidel, editor, *1st Symposium on Simplicity in Algorithms, SOSA 2018, January 7-10, 2018, New Orleans, LA, USA*, volume 61 of *OASIcs*, pages 15:1–15:9. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. 1

[CLN$^+$22] Edith Cohen, Xin Lyu, Jelani Nelson, Tamás Sarlós, Moshe Shechner, and Uri Stemmer. On the robustness of countsketch to adaptive inputs. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato, editors, *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pages 4112–4140. PMLR, 2022. 1, 6

[CNS$^+$24] Edith Cohen, Jelani Nelson, Tamás Sarlós, Mihir Singhal, and Uri Stemmer. One attack to rule them all: Tight quadratic bounds for adaptive queries on cardinality sketches. *CoRR*, abs/2411.06370, 2024. 1, 6

[CSS25] Edith Cohen, Mihir Singhal, and Uri Stemmer. Breaking the quadratic barrier: Robust cardinality sketches for adaptive queries. *CoRR*, abs/2502.05723, 2025. 1, 6

[EIO02] Lars Engebretsen, Piotr Indyk, and Ryan O'Donnell. Derandomized dimensionality reduction with applications. In David Eppstein, editor, *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms, January 6-8, 2002, San Francisco, CA, USA*, pages 705–712. ACM/SIAM, 2002. 1

[Gam21] David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41), October 2021. 4

[GJJ$^+$20] Mrinalkanti Ghosh, Fernando Granha Jeronimo, Chris Jones, Aaron Potechin, and Goutham Rajendran. Sum-of-squares lower bounds for sherrington-kirkpatrick via planted affine planes. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 954–965. IEEE, 2020. 6

[GKPX22] David Gamarnik, Eren C. Kizildag, Will Perkins, and Changji Xu. Algorithms and barriers in the symmetric binary perceptron model. In *63rd IEEE Annual Symposium* *on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 576–587. IEEE, 2022. 6, 8, 16

[GKPX23] David Gamarnik, Eren C. Kizildag, Will Perkins, and Changji Xu. Geometric barriers for stable and online algorithms for discrepancy minimization. In Gergely Neu and Lorenzo Rosasco, editors, *The Thirty Sixth Annual Conference on Learning Theory, COLT 2023, 12-15 July 2023, Bangalore, India*, volume 195 of *Proceedings of Machine Learning Research*, pages 3231–3263. PMLR, 2023. 6

[GLW+24] Elena Gribelyuk, Honghao Lin, David P. Woodruff, Huacheng Yu, and Samson Zhou. A strong separation for adversarially robust l0 estimation for linear sketches. In *65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024*, pages 2318–2343. IEEE, 2024. 1, 6

[GLW+25] Elena Gribelyuk, Honghao Lin, David P Woodruff, Huacheng Yu, and Samson Zhou. Lifting linear sketches: Optimal bounds and adversarial robustness. *arXiv preprint arXiv:2503.19629*, 2025. To appear at STOC 2025. 1, 6

[GVV22] Aparna Gupte, Neekon Vafa, and Vinod Vaikuntanathan. Continuous LWE is as hard as LWE & applications to learning gaussian mixtures. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 1162–1173. IEEE, 2022. 8, 20, 21

[HKM+20] Avinatan Hassidim, Haim Kaplan, Yishay Mansour, Yossi Matias, and Uri Stemmer. Adversarially robust streaming algorithms via differential privacy. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. 1, 6

[Hug23] Hugging Face. Semantic search with faiss, 2023. 5

[HW13] Moritz Hardt and David P. Woodruff. How robust are linear sketches to adaptive inputs? In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 121–130. ACM, 2013. 1

[IM98] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In Jeffrey Scott Vitter, editor, *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*, pages 604–613. ACM, 1998. 1

[JL84] William B Johnson and Joram Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemporary Mathematics*, 26:189–206, 1984. 1

[JPS+22] Vishesh Jain, Natesh S Pillai, Ashwin Sah, Mehtaab Sawhney, and Aaron Smith. Fast and memory-optimal dimension reduction using kac's walk. *The Annals of Applied Probability*, 32(5):4038–4064, 2022. 1

[KF15] Paul Kirchner and Pierre-Alain Fouque. An improved BKW algorithm for LWE with applications to cryptography and lattices. In Rosario Gennaro and Matthew Robshaw, editors, *Advances in Cryptology - CRYPTO 2015 - 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part I*, volume 9215 of *Lecture Notes in Computer Science*, pages 43–62. Springer, 2015. 23

[KN12] Daniel M. Kane and Jelani Nelson. Sparser Johnson-Lindenstrauss transforms. In Yuval Rabani, editor, *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 1195–1206. SIAM, 2012. 1

[LN16] Kasper Green Larsen and Jelani Nelson. The Johnson-Lindenstrauss lemma is optimal for linear dimensionality reduction. In Ioannis Chatzigiannakis, Michael Mitzenmacher, Yuval Rabani, and Davide Sangiorgi, editors, *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, volume 55 of *LIPIcs*, pages 82:1–82:11. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016. 1

[LN17] Kasper Green Larsen and Jelani Nelson. Optimality of the Johnson-Lindenstrauss lemma. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 633–638. IEEE Computer Society, 2017. 1

[MCCD13] Tomás Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In Yoshua Bengio and Yann LeCun, editors, *1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings*, 2013. 5

[MNS08] Ilya Mironov, Moni Naor, and Gil Segev. Sketching in adversarial environments. In Cynthia Dwork, editor, *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 651–660. ACM, 2008. 1, 6

[MRX20] Sidhanth Mohanty, Prasad Raghavendra, and Jeff Xu. Lifting sum-of-squares lower bounds: Degree-2 to degree-4. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, page 840–853, New York, NY, USA, 2020. Association for Computing Machinery. 6

[MTMR23] Niklas Muennighoff, Nouamane Tazi, Loïc Magne, and Nils Reimers. Mteb: Massive text embedding benchmark, 2023. 5

[NN19] Shyam Narayanan and Jelani Nelson. Optimal terminal dimensionality reduction in euclidean space. In Moses Charikar and Edith Cohen, editors, *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 1064–1069. ACM, 2019. 1

[NST23] Kobbi Nissim, Uri Stemmer, and Eliad Tsfadia. Adaptive data analysis in a balanced adversarial model. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. 1

[NY19] Moni Naor and Eylon Yogev. Bloom filters in adversarial environments, 2019. 6

[OM90] A.M. Odlyzko and J.E. Mazo. Lattice points in high-dimensional spheres. *Monatshefte für Mathematik*, 110(1):47–62, 1990. 17

[PTVW22] Aaron Potechin, Paxton Turner, Prayaag Venkat, and Alexander S. Wein. Near-optimal fitting of ellipsoids to random points, 2022. 6

[PX21] Will Perkins and Changji Xu. Frozen 1-rsb structure of the symmetric ising perceptron. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1579–1588. ACM, 2021. 6, 28

[Reg09] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *J. ACM*, 56(6):34:1–34:40, 2009. 21

[RG19] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. *CoRR*, abs/1908.10084, 2019. 5

[RKH$^+$21] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021. 5

[RV10] Mark Rudelson and Roman Vershynin. Non-asymptotic theory of random matrices: extreme singular values. In *Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures*, pages 1576–1602. World Scientific, 2010. 24

[Sim98] Daniel R. Simon. Finding collisions on a one-way street: Can secure hash functions be based on general assumptions? In Kaisa Nyberg, editor, *Advances in Cryptology - EUROCRYPT '98, International Conference on the Theory and Application of Cryptographic Techniques, Espoo, Finland, May 31 - June 4, 1998, Proceeding*, volume 1403 of *Lecture Notes in Computer Science*, pages 334–345. Springer, 1998. 6

[SRS20] Congzheng Song, Alexander M. Rush, and Vitaly Shmatikov. Adversarial semantic collisions. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 4198–4210. Association for Computational Linguistics, 2020. 5

[TJS23] Shengbang Tong, Erik Jones, and Jacob Steinhardt. Mass-producing failures of multimodal systems with language models. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine, editors, *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information*

Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023, 2023. 5

[Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018. 15

[VV25] Neekon Vafa and Vinod Vaikuntanathan. Symmetric perceptrons, number partitioning and lattices. *IACR Cryptol. ePrint Arch.*, page 130, 2025. To appear at STOC 2025. 4, 8, 20

[ZJBS24] Tingwei Zhang, Rishi D. Jha, Eugene Bagdasaryan, and Vitaly Shmatikov. Adversarial illusions in multi-modal embeddings. In Davide Balzarotti and Wenyuan Xu, editors, *33rd USENIX Security Symposium, USENIX Security 2024, Philadelphia, PA, USA, August 14-16, 2024*. USENIX Association, 2024. 5