# Polynomial Data Structure Lower Bounds in the Group Model

Alexander Golovnev* Gleb Posobin† Oded Regev‡ Omri Weinstein§

**Abstract**

Proving super-logarithmic data structure lower bounds in the static *group model* has been a fundamental challenge in computational geometry since the early 80's. We prove a polynomial ($n^{\Omega(1)}$) lower bound for an explicit range counting problem of $n^3$ convex polygons in $\mathbb{R}^2$ (each with $n^{\tilde{O}(1)}$ facets/semialgebraic-complexity), against linear storage arithmetic data structures in the group model. Our construction and analysis are based on a combination of techniques in Diophantine approximation, pseudorandomness, and compressed sensing—in particular, on the existence and partial derandomization of optimal *binary* compressed sensing matrices in the polynomial sparsity regime ($k = n^{1-\delta}$). As a byproduct, this establishes a (logarithmic) separation between compressed sensing matrices and the stronger RIP property.

*Harvard University. Email: alexgolovnev@gmail.com. Research supported by a Rabin Postdoctoral Fellowship.

†Columbia University. Email: posobin@gmail.com. Research supported by NSF CAREER award CCF-1844887.

‡Courant Institute of Mathematical Sciences, New York University. Research supported by the Simons Collaboration on Algorithms and Geometry, a Simons Investigator Award, and by the National Science Foundation (NSF) under Grant No. CCF-1814524.

§Columbia University. Email: omri@cs.columbia.edu. Research supported by NSF CAREER award CCF-1844887.

# 1 Introduction

Understanding the tradeoff between the query time and storage space of data structures has been a long and active research endeavor for several decades. In the static model, which is the focus of our paper, the goal is to represent a database of $n$ elements using the least amount of space $s \geq n$ (measured in memory words), so that queries $q \in \mathcal{R}$ on the input data can be answered quickly, in time $t$. The most compelling model for proving such tradeoffs is the *cell-probe* model [Yao82], in which "query time" is measured only by the *number $t$ of memory accesses* required to retrieve the answer to $q \in \mathcal{R}$ (as in query complexity), whereas all computations on "probed" memory cells are completely free of charge. Time-space lower bounds in this model are therefore a purely information-theoretic question, making *unconditional* lower bounds a viable possibility. Unfortunately, while a simple counting argument [Mil93] shows that almost all static data structure problems with $|\mathcal{R}|$ queries require either near-trivial time $t \geq n^{0.99}$ ($t = n$ is easy by reading all the input database) or near-trivial space $s \geq |\mathcal{R}|^{0.99}$ ($s = |\mathcal{R}|$ is easy by storing all answers), the highest *explicit* cell-probe lower bound known to date is merely

$$t \geq \Omega\left(\frac{\log(|\mathcal{R}|/n)}{\log(s/n)}\right). \tag{1}$$

In the interesting and realistic regime where $|\mathcal{R}| = \text{poly}(n)$ and the data structure is allowed to use *linear storage space* ($s = O(n)$), (1) yields a $t \geq \Omega(\log n)$ lower bound on the query time of several important data structure problems, such as evaluating hash functions, near-neighbor search, and 2D range counting to mention a few (see [Sie04, Păt08, Lar12] and references therein). Proving an unconditional $\omega(\log n)$ lower bound for *any* explicit static problem (with polynomially many queries) against linear-space data structures, remains an outstanding open question in the field.

This gloomy state of affairs remains true even in restricted *arithmetic* models of data structures, which have been studied extensively over the past several decades, primarily in computational geometry and spatial databases [Mat94, Aga17]. The most well-studied model in this line of work is the *group model* and its more restrictive *semigroup* variant [Fre81, Cha90], where the canonical problem is that of geometric *range counting*. In the group model, the input database is a set of $n$ points $x_1, \ldots, x_n \in \mathbb{R}^d$ (where $d$ is typically a small constant), each associated with a weight $w(x_i) \in G$ from an arbitrary commutative group (e.g., $G = \mathbb{Z}_m$), and the goal is to preprocess these points into small memory (by storing an arbitrary collection of $s$ group elements), so that the weighted sum $\sum_{i:x_i \in r} w(x_i)$ of points in a given query "range" $r \in \mathcal{R}$ can be reported efficiently, where summation is over the underlying group. At query time, the data structure can only manipulate weights through the black-box *addition* and *subtraction* of weights stored in memory, and must work for any choice of the underlying group. The query time is the number of algebraic operations performed (additions/subtractions of memory elements). In particular, multiplication by scalars other than $\pm 1$ is not allowed at query time. Any other computation, e.g., planning algebraic operations based on coordinates, is free of charge. The weaker *semigroup model* is defined in the same way, with the crucial difference that only *addition* is allowed, but not subtraction.

Some of the most natural and well-studied examples of range searching problems are orthogonal range counting (where ranges $\mathcal{R}$ are axis-parallel boxes), or counting with respect to more complex geometric objects such as balls, halfspaces or simplices. In its most general form, a range counting problem is defined by a family $\mathcal{R}$ of subsets of $\mathbb{R}^d$, where a common measure of the "complexity" of the problem is the *semialgebraic complexity* of $\mathcal{R}$ [AM94], i.e., the number of polynomial inequalities defining a range $r \in \mathcal{R}$ (Definition 2.3).

Essentially all known range searching data structures can be implemented in the group model[1] (see [Mat94, Aga17] and references therein). Two classic examples of such data structures for *orthogonal* range counting in $d$ dimensions are *range trees* [Lue78], which solve the problem using $s = O(n \log^d n)$ words of space and query time $t = O(\log^{d-1} n)$, and *Kd-trees* [Ben75], which have linear space $s = O(n)$ at the price of polynomial $t \approx n^{1-1/d}$ query time. The latter upper bound can be achieved for *any* range counting problem where ranges have constant semialgebraic complexity [AMS12].

[1] Up to $\text{poly}(\log \log n)$ factors, which can be shaved-off using standard word-level parallelism on the RAM.

For general polytopes or simplex range counting problems [Aga17], such polynomial query times are believed to be inevitable unless near-trivial storage is used ($s \approx |\mathcal{R}|$), but this was only proved in rather weak arithmetic models (or pointer machines [CR95]). In the *semigroup* model, where only additions are allowed, the aforementioned data structures are known to be essentially optimal [Cha90, Afs19]—in particular, this is the only arithmetic model where *polynomial* lower bounds are known on the query time of static range searching problems. By contrast, in the group model, where both additions and subtractions are allowed, the highest static lower bound (against linear storage) remains $\tilde{\Omega}(\log n)$, for 2D orthogonal range counting [Păt07]. This challenge and disparity between the models was summarized by Pătraşcu [Păt07] as follows:

*"Philosophically speaking, the difference in the type of reasoning behind* **semigroup** *lower bounds and* **group** *lower bounds is parallel to the difference between understanding geometry and understanding computation. Since we have been vastly more successful at the former, it should not come as a surprise that progress outside the semigroup model has been extremely slow."*

Indeed, lower bounds in the semigroup model ultimately boil down to arguing that not all query ranges can be "covered" with a small number of subsets of input objects [Aga17, Kol04]. Unfortunately, no such property holds for the group model, which makes proving lower bounds in the group model much harder.

Our main result is a polynomial lower bound for an explicit range counting problem in the static group model for data structures with linear storage:

**Theorem 1.1.** *There is an explicit set $\mathcal{P}_n$ of $m = n^3$ convex polygons in the plane $\mathbb{R}^2$, and a prime $p$, such that any group-model data structure of size $s$ and query time $t$ for range counting with respect to $\mathcal{P}_n$ (over $\mathbb{Z}_p$), must have*

$$t^s \geq n^{n/7}.$$

*In particular, any linear storage ($s = O(n)$) data structure for $\mathcal{P}_n$(over $\mathbb{Z}_p$) must have $n^{\Omega(1)}$ query time. Moreover, the semialgebraic complexity of $\mathcal{P}_n$, i.e., number of facets of each convex polygon, is $n^{O((\log\log n)^2)}$.*

We remark that Theorem 1.1 holds even when preprocessing is allowed to depend on the specific group, i.e., for every group and input configuration, the data structure is allowed to store an arbitrary set of $s$ group elements. In contrast, the standard model in range counting literature (as well as all known upper bounds) is *oblivious*: preprocessing uses the group as a black-box. We also note that $n$-facet polygons arise naturally in computational geometry, for example in the *planar point location* problem [CP09] (i.e., 2D nearest-neighbor search), where the input itself is a collection of $O(n)$-facet *disjoint* polygons (Voronoi diagrams).

Lastly, we note that in the stronger *linear model* of data structures, where *multiplication* (by scalars) is allowed as well as addition and subtraction (hence the group is actually a ring), a recent line of results shows that proving super-logarithmic lower bounds for *any* explicit (range counting) problem would have dramatic implications to arithmetic circuit lower bounds and matrix rigidity [DGW19, RR20]. In that sense, the group model is the strongest arithmetic model in which polynomial lower bounds fall short of major circuit lower bounds, and Theorem 1.1 meets precisely this frontier.

The conceptual message of this paper is that a range counting problem in $\mathbb{R}^d$ is hard for the group model *if (a subset of) its ranges $\mathcal{R}$ forms a good pseudorandom generator against affine hyperplanes in $\mathbb{R}^{|\mathcal{R}|}$*. This will be explained in the following section. While this may not be the *only* source of hardness of range counting in the group model, our work provides a new technique for analyzing such problems and the first substantial step toward proving (exponentially higher) lower bounds on more natural range counting problems, i.e., with lower semialgebraic complexity: If such pseudorandom-generators (PRGs) can be realized as (indicators of) "simple" geometric ranges in reasonably low dimension, then any linear-storage arithmetic data structure should have high query time for the underlying range counting problem. In Section 4 we formalize this by defining the notion of *Geometric PRGs*, which exploit the "geometry" of the $d$-dimensional input seed in a simple way. We prove the existence of such PRGs with nontrivial parameters, as a step toward polynomial lower bounds for *halfspace* range-counting in very high yet nontrivial dimension. We believe the concept of Geometric PRGs may have further applications in computational geometry and is of independent interest.

**Techniques.** Our proof introduces a new way to analyze arithmetic data structures (excluding multiplications), by combining ideas from Diophantine approximation, compressed sensing, and pseudorandomness (specifically, derandomization of anti-concentration theorems). Our results and analysis of "Geometric PRGs" (Theorem 4.7) rely on Fourier-analysis and hypercontractivity to establish pseudorandom properties of halfspace range counting. In the next two sections, we elaborate on these components and the role they play in our proof.

## 1.1 Binary compressed sensing and sparse recovery

A key ingredient in the proof of Theorem 1.1 is the construction of *binary* matrices $M \in \{\pm 1\}^{m \times n}$ with $m = n^c$ and $c > 1$ where every $k$ rows are linearly independent over $\mathbb{R}$ (matrices with such property are commonly known as *compressed sensing*[2] matrices [CT06, DSV12]). If $M$ is allowed to have $\mathrm{poly}(n)$-bit entries, the optimal value of $k = n$ is easily achievable, for example by taking an $n^c \times n$ Vandermonde matrix. However, with binary (or any constant-size) entries, which is crucially the case for our application, the answer is not clear [Ind08]. One way to construct such explicit binary matrices *over the finite field* $\mathbb{F}_2$ (which is harder than over $\mathbb{R}$) is to take the parity-check matrix of rate-optimal binary error correcting codes (e.g., expander codes [SS96, DSV12], see Appendix A), but unfortunately when the number of rows is polynomial ($m > n^{1+\delta}$) this approach only yields $k = O(n/\log n)$-wise independence, and this is well known to be tight [CGH$^+$85] (in fact the latter shows $k = O(n/\log_q n)$ for any constant size field $\mathbb{F}_q$). Other binary constructions in the compressed-sensing and LDPC literature (e.g., [Ind08, XH07, GLW08]) achieve even worse parameters. As explained in Section 1.3, for the proof of Theorem 1.1 it is crucial to obtain the optimal value of $k = \Omega(n)$ (even settling for a much weaker $\mathrm{poly}\log(n)$ query lower bound would still require $k = \omega(n/\log n)$).

This raises a more basic question: *Do optimal ($k = \Omega(n)$) binary compressed-sensing matrices over $\mathbb{R}$ even exist*? Our first step is showing that a *random binary* $\mathrm{poly}(n) \times n$ matrix is indeed $\Omega(n)$-wise independent over $\mathbb{R}$ (and in fact, with some more effort, also over any large enough finite field):

**Theorem 1.2** (Optimal Binary Compressed Sensing Matrices)**.** *For every $c \geq 1$, and every large enough $n$, a uniformly random binary matrix $M \in \{\pm 1\}^{n^c \times n}$ is $\Omega(n/c)$-wise independent over $\mathbb{R}$ w.h.p. Moreover, such binary matrix $\widetilde{M}$ can be constructed using only $O(cn \log n (\log\log n)^2)$ random bits (instead of $n^{c+1}$).*

Note that a naïve union bound will not work here: The probability that a fixed set of $k$ rows is linearly independent is, at best, $2^{-n}$ (the probability 2 rows are equal), which is not small enough to union bound over all $\binom{n^c}{k} = n^{\Theta(n)}$ choices of $k$-subsets for $k = \Omega(n)$. Hence, the proof requires a different analysis that somehow exploits the fact that we are working over fields of large characteristic, while overcoming the challenge that there is an unbounded number of linear combinations over $\mathbb{R}$.

A key observation is that our analysis for random matrices turns out to carry over even to certain *pseudorandom* binary matrices. This is one of the main insights of this paper, which leads to the partial derandomization ($O(cn \log n (\log\log n)^2)$ random bits) in Theorem 1.2. While a fully explicit construction of optimal compressed sensing matrices with constant-size entries remains an intriguing question in the context of sparse recovery (see the next paragraph), the construction in Theorem 1.2 turns out to be enough for obtaining a *fully explicit* data structure lower bound for range counting problems with reasonably low ($n^{\tilde{O}(1)}$) semialgebraic complexity. We elaborate on this step in Section 1.3 below.

**Implications to Sparse Recovery and separation from RIP.** An interesting consequence of Theorem 1.2 is that it provides a (tight) separation between binary *compressed sensing* matrices in which any $k$ rows are linearly independent, and the stronger *restricted isometry property* (RIP), which instead requires any $k$ rows to be "nearly orthogonal". Indeed, in the polynomial sparsity regime $k = n^{1-\delta}$, it is well known that *any* RIP matrix must have $\Theta(k \log k)$ rows (and this is tight for random binary matrices, e.g. [GLW08, BLL$^+$19]). By contrast, Theorem 1.2 asserts that binary compressed sensing matrices only require

[2]Indeed, it is not hard to see that if the rows of $M \in \mathbb{R}^{m \times n}$ are $k$-wise linearly independent, then $M^\top$ is a compressed sensing matrix for $(k/2)$-sparse $m$-dimensional vectors, with $n$ "measurements", hence the problem we study is equivalent.

$O(k)$ linear measurements. To the best of our knowledge, previous binary constructions, even non-explicit ones, were only guaranteed to work with $O(k \log k)$ measurements (see [Ind08] and references therein).

## 1.2 Related Work in the Group Model

A sequence of papers initiated by Chazelle in the early 90's proved essentially tight lower bounds in the *semigroup* model, for orthogonal and halfspace range counting in any constant dimension [Cha90, BCP93]. Similar bounds were shown in the *pointer machine* model [CR95]. In the *offline* group model, where no preprocessing is allowed, Chazelle [Cha94] proved an $\Omega(n \log n)$ lower bound for halfspaces. No super-logarithmic data structure lower bounds were known in the static (i.e., online) setting.

In the *dynamic* group model [Fre81], where the data structure needs to support insertions and deletions of points in $\mathbb{R}^d$, *polynomial* lower bounds ($\Omega(n^{1-1/d})$) have been proved only recently in a breakthrough result of Larsen [Lar14] using combinatorial discrepancy arguments, but as the author notes himself, this technique does not apply to the static case.[3] We remark that amortized dynamic lower bounds for "decomposable" data-structure problems[4], such as range counting, also imply (up to a logarithmic loss) lower bounds for static data structures *with efficient preprocessing time* $p(n) \approx s$ [OvL81]. Alas, in the group model this is a very severe restriction, since storing an arbitrary group element generally requires $O(|G|)$ group operations, which in the dynamic setting is prohibitive and trivialized the aforementioned reduction. For a broader overview of arithmetic lower bounds we refer the reader to [Aga17, Section 3.4] and [Lar14].

## 1.3 Technical Overview

We now provide a streamlined overview of our main result, Theorem 1.1. Recall that a data structure in the group model must solve the given problem with respect to *any* underlying group. We will carefully design a range counting problem such that no (group model) data structure with space $s$ and query time $t$ can solve our problem even in the additive group of integers modulo $p$, $\mathbb{Z}_p$, for an explicit prime $p = \Theta(t)^s$.

**Step one: Diophantine approximations.** Let us fix the $n$ input points $\boldsymbol{x} \in (\mathbb{R}^d)^n$ to a range counting problem, and the weights of these input points $\boldsymbol{w} \in \mathbb{Z}_p^n$. For this input, the data structure computes and stores $S \in \mathbb{Z}_p^s$, a set of $s$ elements of the group. Later, for every query range $r \in \mathcal{R}$, the data structure will output a sum of at most $t$ stored elements (or their negations). The first step in our lower bound uses a basic fact in Diophantine approximation (known as *Dirichlet's simultaneous approximation theorem*, which is an easy application of the pigeonhole principle), which implies that for every set $|S| = s$ of integers from $\{0, \ldots, p-1\}$, there is some $1 \leq q < p := \Theta(t)^s$ such that after multiplication by $q$, *all* those $s$ numbers become smaller than $p/(4t)$ modulo $p$. This of course means that even $t$-*sums* of these numbers (and their negations) remain small modulo $p$ (this is the only but crucial place where we exploit the restriction that only $\pm 1$ coefficients are allowed in the group model at query time). Therefore, every arithmetic data structure with only $s$ memory cells, on *every* set of input weights from $\mathbb{Z}_p^n$, must output a set which is small modulo $p$ after multiplication by some $q = q(S) < p$.

The obvious next step is to construct an explicit range counting problem $\mathcal{R}$ which does *not* have this property, i.e., for some configuration of the $n$ inputs points in $\mathbb{R}^d$, there exists an assignment of weights $(w_1, \ldots, w_n) \in \mathbb{Z}_p^n$, s.t. the output set *does not* become simultaneously too small after multiplication by *any* $1 \leq q < p$. We formalize this with the notion of *diversity*: A set of $m$ elements $\boldsymbol{y} \in \mathbb{Z}_p^m$ (corresponding to $m$ range counting queries) is *diverse*, if for every $1 \leq q < p$ there is some element (query answer) $i \in [m]$ such that $y_i \cdot q \in \left[\frac{p}{4}, \frac{3p}{4}\right] \mod p$. Note that, since we aim for a polynomial ($t > n^{\varepsilon}$) lower bound on query time, and trivially $s \geq n$, Dirichlet's theorem dictates that $p = \Omega(t)^s = 2^{\Omega(n \log n)}$. This fact will soon be

[3] The discrepancy argument relies on the fact that dynamic arithmetic data structures induce a factorization of the query matrix into a product of two sparse matrices, which is not the case in the static setting.

[4] A data structure problem $\mathcal{P}$ is *decomposable* if the answer to a query on the union of two input databases $X, Y$ can be computed as a black-box from the marginal answers $\mathcal{P}(A, B) = f(\mathcal{P}(A), \mathcal{P}(B))$.

important. We also remark that constructing an arbitrary diverse vector explicitly is actually quite easy[5]—the challenge is that this vector must correspond to an output of a *range counting* problem, and indeed this is our next goal.

**Step two: Diversity from $k$-wise independence.** For a range counting problem specified by a family of $m$ ranges $\mathcal{R}$, and a set of input points $\boldsymbol{x} = (x_1, \ldots, x_n) \in (\mathbb{R}^d)^n$, consider the *binary incidence matrix* $\mathcal{I}_{\mathcal{R}}(\boldsymbol{x}) \in \{0,1\}^{m \times n}$ whose $(i,j)$th entry indicates whether point $x_j$ lies in the $i$th range $r_i \in \mathcal{R}$. Note that for the weights $\boldsymbol{w} \in \mathbb{Z}_p^n$, the desired outputs to the $m$ range queries are $\mathcal{I}_{\mathcal{R}}(\boldsymbol{x}) \cdot \boldsymbol{w} \in \mathbb{Z}_p^m$. Now our goal is to find a set of points $\boldsymbol{x}$, weights $\boldsymbol{w}$, and ranges $\mathcal{R}$ for which $\mathcal{I}_{\mathcal{R}}(\boldsymbol{x}) \cdot \boldsymbol{w}$ is a diverse vector. Note that only the set of ranges $\mathcal{R}$ needs to be explicit. The set of weights $\boldsymbol{w}$ and the set of input points $\boldsymbol{x}$ do not need to be explicit, because the data structure must solve the problem for any choice of input weights $\boldsymbol{w} \in \mathbb{Z}_p^n$ and points $\boldsymbol{x} \in (\mathbb{R}^d)^n$. We observe that *if for some input* $\boldsymbol{x}$, the rows of $\mathcal{I}_{\mathcal{R}}(\boldsymbol{x})$ are *$k$-wise independent* over the field $\mathbb{F}_p$ for a $k$ satisfying $(m/k)^{k/2} > p$, then there exist weights $\widetilde{\boldsymbol{w}} \in \mathbb{Z}_p^n$ such that the answer vector $\mathcal{I}_{\mathcal{R}}(\boldsymbol{x}) \cdot \widetilde{\boldsymbol{w}} \in \mathbb{Z}_p^m$ is diverse. This is an easy consequence of Chernoff bounds for $k$-wise independent random variables [BR94]. Hence, our problem boils down to designing a $k$-wise independent *binary* matrix over a large field $\mathbb{F}_p$, where $k$ must satisfy $(m/k)^{k/2} > p = 2^{\Theta(n \log n)}$. With polynomially many queries ($m = |\mathcal{R}| = n^{O(1)}$), this condition requires that $k = \Omega(n)$ in order to get any useful lower bound[6], which is as high as $k$ can possibly be.

**Step three: Binary compressed sensing matrices.** As discussed in Section 1.1, such optimal ("compressed sensing") matrices with constant-size entries were not even know to *exist* in the polynomial sparsity regime $m = \mathrm{poly}(k)$, and indeed the next step of our proof is showing that a *random* $n^c \times n$ binary matrix $M$ is $\Omega(n)$-wise independent with high probability (the first premise of Theorem 1.2). The intuition for why random matrices work *over* $\mathbb{R}$ is as follows: For any fixed nonzero linear combination of some fixed $k \approx n$ rows of $M$, the probability that each coordinate $i \in [n]$ of this linear combination is equal to 0, is $\approx 1/\sqrt{n}$, by the Littlewood-Offord Lemma [Erd45]. (Note that this exploits the fact that we are working over a field of large characteristic.) Since columns are independent, the overall probability that this linear combination is the all-0 vector (and hence linearly dependent) is $\sim (\frac{1}{\sqrt{n}})^n = 2^{-\Theta(n \log n)}$, which is enough to union-bound over all $\binom{n^c}{k} = 2^{\Theta(n \log n)}$ subsets of $k$ rows. Alas, there are two substantial flaws in this argument: Most importantly, it does not rule out *sparse* linear combinations of rows—indeed, the probability that *two* rows are identical is $2^{-n}$, which already dooms the entire claim. Fortunately, there are only $O(n^{2c})$ such pairs, so intuitively a more careful union bound should work. Indeed, using a certain *chaining argument*, we can group the linear combinations by sparsity and handle them separately using some case analysis. The second flaw is that we "forgot" to union-bound over all possible linear combinations in $\mathbb{R}$ as well, which appears daunting. We circumvent this hurdle using an approach (dating back to Komlós [Kom67]) which exploits the observation that linear dependency of rows can be "charged" to the existence of a (minimal) square sub-matrix which is not full rank, making the union bound possible.

**Step four: Derandomization.** While this argument only shows that a random incidence matrix works (which implies a lower bound for a "trivial" range-counting problem[7]), the key observation of our proof is that the above analysis works even for certain *pseudo-random* matrices. Indeed, essentially all the above proof uses about $M$ is that (i) columns are statistically independent, and (ii) that weighted sums of coordinates in each column are *anti-concentrated* for any choice of real-valued weights, i.e., each column *$\varepsilon$-fools affine hyperplanes in $\mathbb{R}^m$* (see Definition 3.9), for $\varepsilon = n^{-0.1}$ (say). Fortunately, explicit pseudo-random generators (PRGs) against affine halfspaces (and therefore hyperplanes, see Proposition 3.11) are known, with almost

[5] Let $p = 2^{0.5n \log n} - 1, m = 0.5n \log n$, and $\boldsymbol{y} \in \mathbb{Z}_p^m$ be a vector with $\boldsymbol{y}_i = 2^i$ for $0 \le i < m$. It is easy to see that $y$ is diverse (indeed, while multiplying $q$ by all powers of 2, we must hit the interval $[p/4, 3p/4]$). Alas, $\boldsymbol{y}$ is not an output of any *group model* problem, as generating it requires summing input weights with *multiplicities*.

[6] Even settling for exponentially weaker $t > \mathrm{poly}\log(n)$ lower bounds using this approach would still require $p > 2^{O(n \log \log n)}$, i.e., $k > \Omega(n \log \log n / \log n)$, which is already beyond known compressed sensing constructions with constant-size alphabets.

[7] I.e., when ranges have semi-algebraic complexity $2^m$, which is the support-size of the columns of a random $m \times n$ matrix.

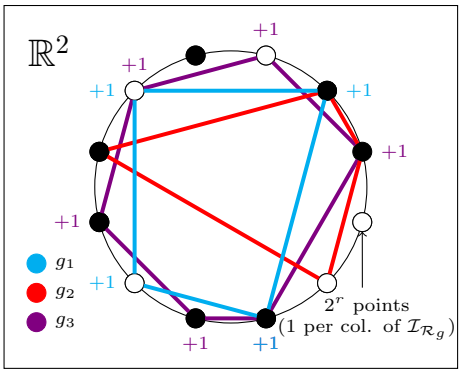


Figure 1: The family of $m = n^3$ convex polygons $\mathcal{R} = \{g_i\}_{i=1}^m$ is generated by mapping each of the $2^r = n^{O((\log\log n)^2)}$ columns of the incidence matrix $\widetilde{M}$ to a point $p(j)$ on a circle in $\mathbb{R}^2$. For each $i \in [m]$, $g_i$ is the convex hull of all the points (columns) $p(j)$ on which $g_i$ outputs 1: $\{p_j \in [2^r] : g_i(s_j) = +1\}$, where $g_i$ is the $i$'th output bit of the [GKM18] PRG and $s_j \in \{0,1\}^r$ is the input seed corresponding to the $j$th column of $\widetilde{M}$. The figure illustrates this for $m = 3$ polygons. Black points on the circle denote the (unknown) $n$ input points—the submatrix $\mathcal{I}_{\mathcal{R}}(\boldsymbol{x}) \subset \widetilde{M}$—for which the range counting problem is guaranteed to be hard.

optimal seed-length $r = \tilde{O}(\log(m/\varepsilon)) = \tilde{O}(\log n)$ random bits [GKM18]. As such, we can derandomize $M$ by choosing each of the $n$ columns independently using this PRG $g(s_i)$ with independent seeds, one per column. In other words, we can generate this $\widetilde{M}$ by sampling $n$ columns from a *fully explicit* incidence matrix supported on only $2^r = n^{\tilde{O}(1)}$ columns. The "ranges" corresponding to this matrix are defined by its rows, i.e., $\mathcal{R} = \{g_i\}_{i=1}^m$, where $g_i$ is the $i$th output bit of the PRG. Our theorem now guarantees that $\widetilde{M}$ contains an induced sub-matrix $\mathcal{I}_{\mathcal{R}} \subset \widetilde{M}$ on $n$ columns, which is $\Omega(n)$-wise independent. Recall that these $n$ columns need *not* be explicit, so long as they are sampled from a small enough superset (which is indeed the case for $\widetilde{M}$).

**Step five: Embedding into $\mathbb{R}^2$.** The final step of the proof is a simple geometric embedding of the ranges $\mathcal{R}$ into the plane. This is done by laying out $2^r = n^{\tilde{O}(1)}$ points on a circle in $\mathbb{R}^2$, each corresponding to a column of $\widetilde{M}$, and observing that for any $g_i \in \mathcal{R}$, the subset of columns corresponding to inputs $x$ for which $g_i(x) = 1$, can be trivially encoded as a convex polygon with at most $2^r$ facets, hence the semi-algebraic complexity of $\mathcal{R}$ is $n^{\tilde{O}(1)}$, which completes the proof (see Figure 1).

**Geometric PRGs.** The proof of Theorem 1.1 uses PRGs in a black-box manner – Since PRGs are agnostic to the representation of the input seed (i.e., treat it as a random bit-string with no underlying geometry), the semi-algebraic complexity of our range counting problem in Figure 1 (=number of facets of polygons) can only be as low as $2^r$ where $r$ is the minimal seed length of the PRG, which for affine halfspaces in $\mathbb{R}^m$ must be $r \geq \Omega(\log m)$ [MZ13]. A natural question toward lower bounds for more natural range counting problems (i.e., with lower semi-algebraic complexity), is whether our lower bound approach can exploit the *dimensionality* of the input space, to decrease the complexity (number of polynomial equations) defining the ranges $\mathcal{R}$ of the underlying range counting problem. Motivated by this question, we define the notion of *Geometric PRGs*: Here, the input "seed" is represented as a (random) point in a $d$-dimensional grid $[B]^d \subset \mathbb{R}^d$ as opposed to the standard bit-string representation, and each of the output bits $g_i$ of the PRG is computed as a *low-degree $d$-variate polynomial threshold function (PTF)*. Such functions generalize (the incidence function of) geometric ranges such as $d$-dimensional *halfspaces*, which are among the most natural and well-studied range counting problems.

In contrast to standard PRGs, the mere existence of Geometric PRGs (in sublinear dimension $d \ll n$) is a nontrivial question. We establish the existence of geometric PRGs in nontrivial ($d = n^{\varepsilon}$) dimension and

polynomial ($\sim n^{\varepsilon}$) degree against affine halfspaces, and make a significant next step en-route to polynomial lower bounds in the group model for halfspace range counting (*degree-1* PTFs): A well known result of Diakonikolas et al. [DGJ$^+$10] asserts that *any* $k$-wise (statistically) independent distribution over $\{\pm 1\}^m$ $\tilde{\Omega}(1/\sqrt{k})$-fools halfspaces in $\mathbb{R}^m$. Our main technical result is showing that, for a typical collection $\mathcal{H}$ of $m = \text{poly}(d)$ hyperplanes in $\mathbb{R}^d$, the *signed inner-products* of $\mathcal{H}$ with a random hyperplane $\mathbf{r} \in_{\mathrm{R}} \{\pm 1\}^{\mathrm{d}}$ (the "seed"), forms a distribution on $\{\pm 1\}^m$ which is *almost* $k$-wise independent for any $k \leq d^{1/6}$, i.e. $\|\{\mathsf{sign}\langle h_i, \mathbf{r}\rangle\}_{\mathrm{i}\in[\mathrm{k}]} - \mathcal{U}_{\mathrm{d}}\|_1 = 1/\text{poly(d)}$ for any subset $\mathcal{H}' = \{h_1, \ldots, h_k\} \subset \mathcal{H}$. This statement is essentially tight. We prove it by adapting a theorem due to Klartag and Regev [KR11] about random projections in *Gaussian* spaces to the setting of caps in the hypercube, using Fourier-analytic techniques. The main difficulty is that in Boolean space, the hypercontractivity argument of [KR11] is not enough to bound the desired measure of intersection, hence we use a convolution with the *majority* function and a "mirroring" trick to bound this measure (see Theorem 4.7). We leave open the question of finding an *explicit* construction of low-degree geometric PRGs, which could lead to group-model lower bounds for more natural range searching problems using our approach.

# 2 Preliminaries

We identify the finite field $\mathbb{F}_p$ with the set $\{0, \ldots, p-1\}$. For a non-negative integer $m$, we denote by $[m]$ the set $\{1, \ldots, m\}$. For a real number $x$, we denote by $\|x\|$ the distance from $x$ to the nearest integer. All logarithms are to the base 2. We write $|b| < c \bmod p$ to mean that $b$ is $c$-close to an integer divisible by $p$, i.e., that there exists an integer $k$ such that $|b - kp| < c$, or in other words $\|b/p\| < c/p$. All logarithms in the paper are base 2 unless otherwise stated, and we use the shorthand $\log x = \log_2 x$.

**Definition 2.1** ($k$-wise independent matrices)**.** *A matrix $M \in \mathbb{F}^{m\times n}$ is* $k$-wise independent *if every set of $k$ rows of $M$ is linearly independent (over $\mathbb{F}$).*

## 2.1 Range counting in the group model

**Definition 2.2** (Range counting in group model)**.** *Let $\mathcal{R} = \{\mathcal{R}_i \subset \mathbb{R}^d, i \in [m]\}$ be a range counting problem in $\mathbb{R}^d$. A data structure $\mathcal{D} = (P : (\mathbb{R}^d)^n \times G^n \to G^s, Q : G^s \to G^m)$ is said to be an $(s,t)$ data structure in the group model for $\mathcal{R}$ over an abelian group $G$, if:*

1. *For any configuration of $n$ points $x_1, \ldots, x_n$ in $\mathbb{R}^d$, any tuple of $n$ weights $\boldsymbol{w} = (w_1, \ldots, w_n) \in G^n$, and any index $i \in [m]$, it holds that $Q_i \circ P((x_k)_{k=1}^n, \boldsymbol{w}) = \sum_{j:x_j\in\mathcal{R}_i} w_j$. That is, the answer to the $i$-th query is the sum of the weights of the points that fall into the $i$-th range.*

2. *Each query function $Q_i$ computes a group sum of* at most *$t$ memory elements or their negations, possibly with repetitions: $Q_i(g_1, \ldots, g_s) = \Sigma_{j\leq t}\xi_j g_{i_j}$, with $\xi_j \in \{\pm 1\}$. Note that the function $Q_i$ can be adaptive (i.e., a depth-$t$ decision-tree over the $s$ memory elements).*

While this arithmetic model seems rather restrictive, all known data structures for range searching problems fall into this setting[8]. We remark that here we allow the preprocessing function $P$ to depend on the specific group, i.e., for every group and every input $\boldsymbol{x} \in \mathbb{R}^d$, the data structure is allowed to compute and store an arbitrary set of $s$ group elements. By contrast, in the standard "oblivious" definition of the group model [Aga17, Cha90], as well as in all known upper bounds, the preprocessing function treats the group as a black-box: the $s$ group elements are computed using a fixed set of group operations.

We use the common notion of semialgebraic sets to quantify the complexity of a range counting problem:

**Definition 2.3** (Semialgebraic Sets [AMS12, Aga17])**.** *A* semialgebraic set *is a subset of $\mathbb{R}^d$ obtained from a finite number of sets of the form $\{\boldsymbol{x} \in \mathbb{R}^d \mid p(\boldsymbol{x}) \geq 0\}$, where $p$ is a $d$-variate polynomial with integer*

[8] See Footnote 1.

*coefficients, using arbitrary intersections, unions, and complementations. $\Gamma_{d,\Delta,s}$ denotes the collection of all the semialgebraic sets in $\mathbb{R}^d$ obtained from at most $s$ $d$-variate polynomials of total degree at most $\Delta$ each.*

*For brevity, when we say that a range counting problem $\mathcal{R} \subseteq \mathbb{R}^d$ has* semialgebraic complexity $K$ *we mean that the number of linear inequalities defining each range $r \in \mathcal{R}$ is at most $K$. For example, a $K$-facet polytope in $\mathbb{R}^d$ has semialgebraic complexity $K$ under this convention.*

The ranges we construct in Theorem 1.1 are (convex) polygons in two dimensions with at most $n^{(\log\log n)^2}$ sides, so they will lie in $\Gamma_{2,1,n^{(\log\log n)^2}}$.

### 2.2 Anti-concentration inequalities

We will use the following versions of the Littlewood-Offord Lemma.

**Lemma 2.4** ([Erd45])**.** *Suppose that $x_1, \ldots, x_k \in \mathbb{R} \setminus \{0\}, y \in \mathbb{R}$. Then*

$$\left|\left\{\varepsilon \in \{-1,1\}^k \middle| \sum_{i=1}^k \varepsilon_i x_i = y\right\}\right| \leq \binom{k}{\lfloor k/2 \rfloor} \leq \frac{2^k}{\sqrt{k}}\,.$$

**Lemma 2.5** ([NW18, Theorem 6.3], [NP18, Theorem A.21])**.** *Let $\mathbb{F}_p$ be a prime field of order $p > 2$. Suppose that $x_1, \ldots, x_k \in \mathbb{F}_p \setminus \{0\}$ and $y \in \mathbb{F}_p$. Then*

$$\left|\left\{\varepsilon \in \{-1,1\}^k \middle| \sum_{i=1}^k \varepsilon_i x_i = y\right\}\right| \leq 2^k \left(\frac{1}{p} + \sqrt{\frac{8}{k}}\right)\,.$$

## 3 Proof of Theorem 1.1

### 3.1 Efficient Data Structures are Not Diverse

Let us fix the input points $(x_1, \ldots, x_n) \in (\mathbb{R}^d)^n$, and vary their weights. Let us look at the possible answers that the composition $Q \circ P((x_1, \ldots, x_n), \cdot)$ can produce. We will show that $Q \circ P$ always has structure (the resulting vectors are not "diverse"), while some range counting problems do not have this structure (are "diverse"), thus, cannot be computed by efficient data structures in the group model.

**Definition 3.1** (Diversity)**.** *We call a set $S \subset \mathbb{Z}_p$* diverse *if for any $q \in \mathbb{Z}_p \setminus \{0\}$, $q \cdot S \not\subseteq [-p/4, p/4] \bmod p$. Similarly, we say that a vector $\boldsymbol{w} \in \mathbb{Z}_p^m$ is* diverse *if the set of its coordinates is diverse.*

**Lemma 3.2** (Outputs of an efficient DS are not diverse)**.** *For any $(4t)^s < p$, any $(s,t)$ data structure $\mathcal{D} = (P : ([n]^d)^n \times \mathbb{Z}_p^n \to \mathbb{Z}_p^s, Q : \mathbb{Z}_p^s \to \mathbb{Z}_p^m)$ in the group model, any points $x_1, \ldots, x_n \in [n]^d$, and any weights $w_1, \ldots, w_n \in G$, the set of answers to queries*

$$\{Q_j \circ P((x_i)_{i=1}^n, \boldsymbol{w})\}_{j=1}^m$$

*is not diverse.*

To compute each output coordinate, $Q$ only adds and subtracts $t$ of the memory cells, counting repetitions, so it is enough to show that there is a $1 \leq q \leq (4t)^s$ such that each of the memory cells is at most $p/(4t)$ modulo $p$ after multiplication by $q$. The following lemma shows that such a $q$ exists for any set of $s$ integers.

**Lemma 3.3** (Dirichlet's Simultaneous Approximation Theorem)**.** *For any set of $s$ integer numbers $x_1, \ldots, x_s$, a natural number $Q \geq 1$, and a natural number $p > 1$, there exists an integer $1 \leq q \leq Q^s$ such that*

$$\forall i \in [s],\ |x_i \cdot q| < p/Q \bmod p\,.$$

*Or, in greater generality: given $s$ real numbers $\boldsymbol{r} = (r_1, \ldots, r_s)$ and an integer $Q$, there is an integer $1 \leq q \leq Q^s$ such that for every $i$, $\|q \cdot r_i\| < 1/Q$, i.e. all the given real numbers can be approximated by rational fractions $k_i/q$ with some denominator $1 \leq q \leq Q^s$ up to an error strictly less than $1/(qQ)$.*

*Proof.* The second statement implies the first one when $r_i := x_i/p$.

Split the interval $[0,1)$ into $Q$ sub-intervals $[k/Q,(k+1)/Q)$. Multiply the vector $\boldsymbol{r}$ by numbers $0,\ldots,Q^s$, take the fractional part of each coordinate, and map it to the index of the sub-interval it falls into. There are only $Q^s$ distinct vectors of this form, but we got $Q^s+1$ vectors corresponding to $0,\boldsymbol{r},2\cdot\boldsymbol{r},\ldots,Q^s\cdot\boldsymbol{r}$, so two of them will be the same. Let those two vectors be $q_1\cdot\boldsymbol{r}$ and $q_2\cdot\boldsymbol{r}$ for $q_1<q_2$. Then, $1\le(q_2-q_1)\le Q^s$ and $\|(q_2-q_1)\cdot r_i\|<1/Q$ for all $i$, as desired. □

## 3.2 $k$-wise Independence Implies Diversity

We have shown above that any set of outputs from an $(s,t)$ data structure for range counting in the group model is not diverse if $(4s)^t<p$. Now we show that it is possible to construct a matrix $M\in\mathbb{Z}_p^{m\times n}$, such that if it is an incidence matrix of $m$ shapes and $n$ points, then there is a way to assign weights $\boldsymbol{w}$ to those points so that the answers $M\boldsymbol{w}$ to the $m$ queries form a diverse set.

The idea is that if $\boldsymbol{w}$ is uniformly random, then for a fixed $1\le q<p$ the probability of $q\cdot(M\boldsymbol{w})_i$ being at most $p/4$ modulo $p$ is $1/2$, and if the coordinates of $M\boldsymbol{w}$ were independent, the probability of all of $q\cdot(M\boldsymbol{w})_i$ being at most $p/4$ would be $2^{-m}$. Of course, if $n<m$, some coordinates of $M\boldsymbol{w}$ will be linearly, and thus statistically, dependent, but we can make do with a weaker requirement: that the coordinates are $k$-wise independent. Then the probability that all $q\cdot(M\boldsymbol{w})_i$ are at most $p/4$ modulo $p$ can be bounded by $1/p$ by a version of a Chernoff bound for $k$-wise independent variables. After applying union bound over all possible $q\in[1,p-1]$, this gives the existence of $\boldsymbol{w}$ that produces a diverse $M\boldsymbol{w}$. And the $k$-wise independence of the coordinates of $M\boldsymbol{w}$ follows from the $k$-wise independence of the rows of the matrix $M$ if it is viewed as a matrix in $\mathbb{F}_p^{m\times n}$, so then we are left with the problem of finding a binary matrix $M$ with $k$-wise linearly independent rows over $\mathbb{F}_p$ and a set of shapes and points for which $M$ is an incidence matrix.

We will use the following form of the Chernoff bound for $k$-wise independent variables.

**Theorem 3.4** ([BR94])**.** *Let $k\ge4$ be an even integer. Suppose that $X_1,\ldots,X_n$ are $k$-wise independent random variables taking values in $[0,1]$. Let $X=X_1+\cdots+X_n$, and $\mu=E[X]$, and let $A>0$. Then*

$$\Pr[|X-\mu|>A]<2\left(\frac{nk}{A^2}\right)^{k/2}.$$

**Theorem 3.5** ($k$-wise independence implies diverse weights)**.** *Let $M\in\mathbb{F}_p^{m\times n}$ be $k$-wise independent and $p\le0.5\left(\frac{m}{16k}\right)^{k/2}$, then there exists $\boldsymbol{w}\in\mathbb{F}_p^n$ such that $M\boldsymbol{w}$ is diverse.*

*Proof.* For a fixed $1\le q<p$, let $P_q$ be the probability that $q$ is a "witness" of non-diversity of a random input $\boldsymbol{w}$. That is,

$$P_q\coloneqq\Pr_{\boldsymbol{w}}\left[\forall i\in[m],\ |q\cdot(M\boldsymbol{w})_i|<p/4 \bmod p\right].$$

We will show that $P_q\le1/p$ which, by the union bound, will imply the theorem statement. Let $W_{q,i}$ be the indicator variable of the event $|q\cdot(M\boldsymbol{w})_i|\ge p/4 \bmod p$. Then $\mathbb{E}_{\boldsymbol{w}}[W_{q,i}]>1/2$ since there is at least one non-zero coordinate in $M_i$, and $P_q=\Pr[\sum_{i=1}^m W_{q,i}=0]$. Note that from $k$-wise independence of $M$, we have $k$-wise independence of $W_{q,i}$: indeed, $k$-wise independence of $M$ implies that any $k$ rows $M_{i_1},\ldots,M_{i_k}$ of $M$ are linearly independent, so the tuple $(M_{i_1}\boldsymbol{w},\ldots,M_{i_k}\boldsymbol{w})$ is distributed uniformly in $\mathbb{F}_p^k$. From Theorem 3.4 with $A=m/4$, we have

$$P_q=\Pr\left[\sum_{i=1}^m W_{q,i}=0\right]<2(16k/m)^{k/2}\le1/p.$$

□

## 3.3 Existence of $k$-wise Independent Binary Matrices

In this section, we prove that almost all matrices from $\{\pm1\}^{n^c\times n}$ are $\Omega(n)$-wise independent (for every $c\ge1$).

**Theorem 3.6.** *For every $\varepsilon \in (0, 1/2]$, every large enough $n \le m \le n^{\frac{1}{2\varepsilon}}$, every prime field $\mathbb{F}_p$ of order $p \ge \sqrt{n}$, a uniformly random matrix $M \in \{-1,1\}^{m\times n}$ is $k := \varepsilon n$-wise independent over $\mathbb{F}_p$ with probability $1 - o(1)$.*

We remark that $k$-wise independence of $M$ over $\mathbb{F}_p$ implies $k$-wise independence over $\mathbb{R}$. Indeed, $k$-wise independence of $M$ implies that any $k$ rows of $M$ contain a non-singular $k \times k$ submatrix over $\mathbb{F}_p$, this submatrix is also non-singular over $\mathbb{R}$, which in turn implies $k$-wise independence over $\mathbb{R}$.

In the proof of Theorem 3.6 we will use the following claim.

**Claim 3.7.** *Any $k$-dimensional subspace $L \subset \mathbb{F}_p^n$ contains at most $2^k$ vertices of the hypercube $\{\pm 1\}^n$.*

*Proof.* Fix any $k$ spanning vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ of $L$ and find $k$ coordinates $m_1, \dots, m_k \in [n]$ such that the matrix $C \in \mathbb{F}_p^{k\times k}$, where $C_{i,j} = (\boldsymbol{x}_{i,m_j})$ is invertible. Then there are exactly $2^k$ vectors $\boldsymbol{v}$ in $L$ whose coordinates $\boldsymbol{v}_{m_j}$ are all $\pm 1$. □

Now we are ready to prove the main result of this section.

*Proof of Theorem 3.6.* For a matrix $M \in \{-1,1\}^{m\times n}$, denote by $M_i \in \{-1,1\}^n$ its $i$th row, and denote by $M_{A\times B}$ a submatrix of $M$ formed by taking cells with indices in the cartesian product $A \times B$. Let $P$ be the probability that a random matrix $M \in \{-1,1\}^{m\times n}$ is *not* $k$-wise independent. That is equivalent to saying that there is a subset $S \subset [m]$ of indices of size $|S| \le k$ such that $M_{S\times[n]}$ is not full-rank.

Let us take a minimal such $S$: $M_{S\times[n]}$ is not full-rank, while all its submatrices obtained by removing one of the rows *are* full-rank. In particular, this implies that $M_{S\times[n]}$ has rank $|S| - 1$, but it is a stronger condition than that. To describe this event succinctly, denote by $\mathsf{sp}(H)$ the row spark of a matrix: the smallest number of linearly dependent rows of a matrix $H$, defined to be $\infty$ when all the rows are linearly-independent.[9] Then the condition above is equivalent to $\mathsf{sp}(M_{S\times[n]}) = |S|$. We will bound the probability of that happening for some fixed subset $S$ of rows, and then use union bound to bound $P$.

Let $P_S$ for $S \subset [m]$ be the probability of the event $\mathsf{sp}(M_{S\times[n]}) = |S|$. For this event to happen, $M_{S\times[n]}$ has to at least have a full-rank $(|S|-1) \times (|S|-1)$ submatrix $M_{S'\times T}$ in its first $|S|-1$ rows, where $S'$ is obtained from $S$ by removing the largest element, and $T$ is some $(|S|-1)$-element subset of $[n]$.

Let $P_{S,T}$, with $S \subset [m], T \subset [n]$ and $|T| = |S|-1$, be the probability of the intersection of $\mathsf{sp}(M_{S\times[n]}) = |S|$ and the submatrix $M_{S'\times T}$ being full-rank. By union bound:

$$P_S \le \sum_T P_{S,T} = \binom{n}{|S|-1} P_{S,[|S|-1]} \ ,$$

$$P \le \sum_{|S|\le k} P_S = \sum_{i=2}^{k} \binom{m}{i}\binom{n}{i-1} P_{[i],[i-1]} \ .$$

Now it remains to bound $P_{[i],[i-1]}$. Since the corresponding event depends only on the first $i$ rows of the matrix, let us look at those rows separately.

Consider the matrix $A = M_{[i]\times[n]} \in \{-1,1\}^{i\times n}$, and call its upper-left $(i-1)\times(i-1)$ submatrix $Q$. $P_{[i],[i-1]}$ is the intersection of three events:

- $\mathcal{E}_Q$ — $Q$ is a full-rank $\pm 1$-matrix;
- $\mathcal{E}_{\mathsf{sp}\ge i}$ — any $i-1$ rows of $A$ are linearly independent;
- $\mathcal{E}_{\mathsf{rk}=i-1}$ — the rank of $A$ is $i-1$.

[9] Note that spark is often defined as the smallest number of linearly dependent *columns*, not rows.

To upper bound $P_{[i],[i-1]} = \Pr[\mathcal{E}_Q \wedge \mathcal{E}_{\mathsf{sp}\geq i} \wedge \mathcal{E}_{\mathsf{rk}=i-1}]$, we replace it with conditional probability: $\Pr[\mathcal{E}_Q \wedge \mathcal{E}_{\mathsf{sp}\geq i} \wedge \mathcal{E}_{\mathsf{rk}=i-1}] \leq \Pr[\mathcal{E}_{\mathsf{sp}\geq i} \wedge \mathcal{E}_{\mathsf{rk}=i-1} | \mathcal{E}_Q]$.

Assuming $\mathcal{E}_Q$ holds, i.e. assuming that the submatrix $Q$ is full-rank, there is exactly one way to combine the rows of the invertible matrix $Q$ to get the first $i-1$ coordinates of the $i$th row of $A$. That is, there is exactly one set of coefficients $\alpha_1, \ldots, \alpha_{i-1}, \alpha_i$ with $\alpha_i = 1$ such that $\sum_{k=1}^{i} \alpha_k A_{\{k\}\times[i-1]} = A_{\{i\}\times[i-1]} + \sum_{k=1}^{i-1} \alpha_k Q_k$ is zero. Introduce the new event $\mathcal{E}_\alpha$: intersection of $\mathcal{E}_Q$ and the event that *all* the coefficients $\alpha_1, \ldots, \alpha_i$ are non-zero. Conditioned on $\mathcal{E}_Q$, the intersection $\mathcal{E}_{\mathsf{sp}\geq i} \wedge \mathcal{E}_{\mathsf{rk}=i-1}$ implies $\mathcal{E}_\alpha$: since spark of $A$ is exactly its number of rows, there is a linear combination with non-zero coefficients of the $A$'s first $i-1$ rows that gives $A_i$, and since $Q$ is invertible, this combination gives the only possible list of $\alpha_1, \ldots, \alpha_i$. So conditioning on $\mathcal{E}_\alpha$ can only increase the probability of $\mathcal{E}_{\mathsf{sp}\geq i} \wedge \mathcal{E}_{\mathsf{rk}=i-1}$.

Moreover, $\mathcal{E}_\alpha$ and $\mathcal{E}_Q$ are determined only by the first $i-1$ columns of the matrix $A$. So

$$\begin{aligned}
\Pr\left[\mathcal{E}_{\mathsf{sp}\geq i} \wedge \mathcal{E}_{\mathsf{rk}=i-1} | \mathcal{E}_Q\right] &\leq \Pr\left[\mathcal{E}_{\mathsf{sp}\geq i} \wedge \mathcal{E}_{\mathsf{rk}=i-1} | \mathcal{E}_Q \wedge \mathcal{E}_\alpha\right] \\
&\leq \Pr\left[\mathcal{E}_{\mathsf{rk}=i-1} | \mathcal{E}_\alpha\right] \\
&\leq \Pr\left[\sum_{k=1}^{i} \alpha_k A_{\{k\}\times([n]\setminus[i-1])} = 0 \;\middle|\; \mathcal{E}_\alpha\right],
\end{aligned}$$

where $\alpha_1, \ldots, \alpha_i$ with $\alpha_i = 1$ are determined by $\mathcal{E}_\alpha$. Note that $\mathcal{E}_\alpha$ depends only on the first $i-1$ columns of the matrix, while the event under the probability sign is determined by all the other columns and coefficients $\alpha$ obtained from the first $i-1$ columns. Also note that we have not relied on any properties of the distribution of $M$ so far.

Now we use the fact that the columns are i.i.d.:

$$P_{[i]\times[i-1]} \leq \Pr\left[\sum_{k=1}^{i} \alpha_k A_{\{k\}\times([n]\setminus[i-1])} = 0 \;\middle|\; \mathcal{E}_\alpha\right] = \Pr\left[\sum_{k=1}^{i} \alpha_k \xi_k = 0 \;\middle|\; \mathcal{E}_\alpha\right]^{n-i+1}, \tag{2}$$

where $\xi_k$ are the entries in a column of $A$, and all $\alpha_k$ are non-zero by $\mathcal{E}_\alpha$. We highlight again the fact that the vector $\xi$ is independent from the condition $\mathcal{E}_\alpha$ and the vector of coefficients $\alpha_1, \ldots, \alpha_k$ as long as the columns are independent. The upper bound on this exact probability for the uniform $\pm 1$ vector $\boldsymbol{\xi}$ is then given by a version of the Littlewood-Offord lemma for finite fields:

For fixed non-zero $\alpha_1, \ldots, \alpha_i \in \mathbb{F}_p \setminus \{0\}$, let $R_i$ be the probability that a uniformly random column $A_{[i]\times\{t\}} \in \{-1, 1\}^{i\times 1}$ satisfies $\sum_{j=1}^{i} \alpha_j A_{j,t} = 0$. Then we have

$$P_{[i]} \leq \binom{n}{i-1} \cdot R_i^{n-(i-1)}.$$

Now we show that $P \leq \sum_{i=2}^{k} \binom{m}{i} P_{[i]} = o(1)$ by showing that each term in this sum is $o(1/k) = o(1/(\varepsilon n))$.

- For $2 \leq i \leq n/(4\log m)$, for every choice of the values of some $(i-1)$ linearly independent rows, by Claim 3.7, there exist at most $2^{i-1}$ values of the $i$-th row which will create a linear dependence. Therefore, $P_{[i]} \leq \frac{2^{i-1}}{2^n}$, and

$$\binom{m}{i} P_{[i]} \leq \frac{m^i}{i!} \cdot \frac{2^{i-1}}{2^n} \leq 2^{i\log m - n} \leq 2^{-3n/4}.$$

- For $n/(4\log m) < i \leq k = \varepsilon n$, from the Littlewood-Offord Lemma over finite fields (Lemma 2.5), we have that

$$R_i \leq \left(\frac{1}{p} + \sqrt{\frac{8}{i}}\right) \leq \left(\frac{1}{\sqrt{n}} + \sqrt{\frac{32\log m}{n}}\right) \leq \frac{O(\log m)}{\sqrt{n}}.$$

Then

$$P_{[i]} \leq \binom{n}{i-1} R_i^{n-(i-1)} \leq 2^n \cdot R_i^{(1-\varepsilon)n} \leq 2^{O(n\log\log m) - 0.5(1-\varepsilon)n\log n}.$$

From $i \le \varepsilon n$ and $m \le n^{\frac{1}{2\varepsilon}}$, we have that

$$\binom{m}{i} \le \left(\frac{m}{\varepsilon n}\right)^{\varepsilon n} = 2^{O(n)+\varepsilon n \log\left(\frac{m}{n}\right)} = 2^{O(n)+(0.5-\varepsilon)n\log n}.$$

Finally,

$$\binom{m}{i} P_{[i]} \le 2^{O(n\log\log m)+(0.5-\varepsilon)n\log n - 0.5(1-\varepsilon)n\log n} = 2^{O(n\log\log n)-0.5\varepsilon n\log n} = o(1/k).$$

□

Now we observe that the result of Theorem 3.6 holds for $\{0,1\}$-matrices as well. While we cannot just take a $\{\pm 1\}$ matrix from Theorem 3.6, increment each entry by one, and divide by two, we can perform a similar operation while preserving $k$-wise independence.

**Corollary 3.8.** *For every $\varepsilon \in (0, 1/2]$, every large enough $n \le m \le n^{\frac{1}{2\varepsilon}}$, every prime field $\mathbb{F}_p$ of order $p \ge \sqrt{n}$, a uniformly random matrix $M \in \{0,1\}^{(m-1)\times n}$ is $(k-1) \coloneqq (\varepsilon n - 1)$-wise independent over $\mathbb{F}_p$ and $\mathbb{R}$ with probability $1 - o(1)$.*

*Proof.* Choose a uniformly random matrix $M \in \{\pm 1\}^{m\times n}$, and multiply some of the columns by $-1$ so that the first row contains only 1s. The rest of the matrix remains uniformly random, and multiplying columns by scalars does not introduce new linear dependencies among the rows. For $i \in [m-1]$, let $B_i = (M_{i+1} + M_1)/2$. This gives a uniformly random $\{0,1\}^{(m-1)\times n}$ matrix $B$. Take any $k-1$ rows $B_S = (B_{i_1}, \ldots, B_{i_{k-1}})$, and consider the rows $M_{\{1\}\cup(1+S)} = (M_1, M_{1+i_1}, \ldots, M_{1+i_{k-1}})$. Linear independence of $M_{\{1\}\cup(1+S)}$ implies linear independence of $B_S$, because $B_S$ is obtained from $M_{\{1\}\cup(1+S)}$ by row operations and removing the first row, so if $M$ is $k$-wise independent, then $B$ is $(k-1)$-wise independent. □

## 3.4 Derandomization

In this section we reduce the randomness used in the proof above from $mn$ bits to $O\left(n \log m (\log\log m)^2\right)$ bits: in effect instead of sampling the columns from the set of all the possible columns in $\{\pm 1\}^m$, we show that we can limit ourselves to a much smaller subset and still get a very high probability of sampling a $k$-wise independent matrix. This will dramatically (exponentially) reduce the semi-algebraic complexity of the resulting range-counting problem in our final construction.

As discussed in the introduction, the main observation here is that the analysis of Theorem 3.6 for random binary matrices, also carries over to the case where each column is chosen independently according to a distribution on $\{\pm 1\}^m$ which fools any *affine hyperplane* in $\mathbb{R}^m$. We now formalize this statement.

**Definition 3.9.** *We say that a deterministic function $F : \{0,1\}^r \to \{\pm 1\}^m$ $\varepsilon$-fools a family of sets $\mathcal{S} \subset 2^{\{\pm 1\}^m}$, if for any $s \in \mathcal{S}$:*

$$\left| \Pr_{\boldsymbol{x} \leftarrow \mathrm{Unif}\{\pm 1\}^m} [\boldsymbol{x} \in s] - \Pr_{\boldsymbol{x} \leftarrow F} [\boldsymbol{x} \in s] \right| \le \varepsilon,$$

*where the second $\boldsymbol{x}$ is the output of $F$ when it is passed a uniformly random input string from $\{0,1\}^r$ as an input. We also call such an $F$ an $\varepsilon$-*PRG against $\mathcal{S}$.

Our derandomization will use the following pseudo-random generator (PRG) of Gopalan, Kane and Meka [GKM18], which $\varepsilon$-fools affine halfspaces with almost optimal seed-length.

**Lemma 3.10** ([GKM18, Corollary 1.2])**.** *For every $\varepsilon > 0$ there exists a PRG $F_{\mathsf{GKM}} : \{0,1\}^r \to \{\pm 1\}^m$ that uses $r = O\left(\log(m/\varepsilon) \cdot (\log\log(m/\varepsilon))^2\right)$ uniformly random input bits and $\varepsilon$-fools affine halfspaces, i.e., the family of sets*

$$\left\{\boldsymbol{x} \in \{\pm 1\}^m \mid \langle \boldsymbol{x}, \alpha \rangle \le \theta \right\}_{\boldsymbol{\alpha} \in \mathbb{R}^m, \theta \in \mathbb{R}}.$$

**Proposition 3.11** (Fooling halfspaces ⇒ Fooling hyperplanes)**.** *If $F$ is an $\varepsilon$-PRG against affine halfspaces in $\mathbb{R}^m$, then it is a $2\varepsilon$-PRG against hyperplanes in $\mathbb{R}^m$.*

*Proof.* Let $\boldsymbol{\alpha} \in \mathbb{R}^m$ be a fixed vector. Let $F_1$ and $F_2$ be the CDFs of the distributions of $\langle \boldsymbol{x}, \boldsymbol{\alpha} \rangle$ for a uniform $\boldsymbol{x} \in \{-1, 1\}^m$, and for $\boldsymbol{x}$ generated from $F$, respectively. From Definition 3.9, we have that the Kolmogorov-Smirnov distance between their distributions is at most $\varepsilon$: $\forall t,\ |F_1(t) - F_2(t)| \leq \varepsilon$. This implies that the probability of $\langle \boldsymbol{x}, \boldsymbol{\alpha} \rangle$ attaining a certain value can differ by at most $2\varepsilon$ between the two distributions:

$$\begin{aligned}\left| \Pr_{\boldsymbol{x} \leftarrow \mathrm{Unif}}[\langle \boldsymbol{x}, \boldsymbol{\alpha} \rangle = \theta] - \Pr_{\boldsymbol{x} \leftarrow F}[\langle \boldsymbol{x}, \boldsymbol{\alpha} \rangle = \theta] \right| &= \left| (F_1(\theta) - F_1(\theta^-)) - (F_2(\theta) - F_2(\theta^-)) \right| \\ &\leq |F_1(\theta) - F_2(\theta)| + \left| F_1(\theta^-) - F_2(\theta^-) \right| \\ &\leq 2\varepsilon\,.\end{aligned}$$

□

Using this proposition, along with the GKM PRG for affine hyperplanes, we shall prove the following:

**Theorem 3.12.** *For every $\varepsilon \in (0, 1/2]$, every large enough $n \leq m \leq n^{\frac{1}{2\varepsilon}}$, there is a polynomial-time algorithm that uses $O\left(n \log m (\log\log m)^2\right)$ random bits and outputs a matrix $W \in \{\pm 1\}^{m \times n}$, s.t. $W$ is $k := \varepsilon n$-wise independent over $\mathbb{R}$ with probability $1 - o(1)$. As a consequence, $W$ is also $k$-wise independent over $\mathbb{F}_p$ for any prime $p > (\varepsilon n)!$.*

*Proof.* Let $F_{\mathsf{GKM}} \colon \{0,1\}^r \to \{\pm 1\}^m$ be the PRG from Lemma 3.10 with seed length $r = O\left(\log m (\log\log m)^2\right)$ which $(1/\sqrt{m})$-fools hyperplanes. The algorithm will generate each column of $W$ as the output of $F_{\mathsf{GKM}}$ with a fresh seed of length $r$. Clearly, the algorithm uses $nr = O\left(n \log m (\log\log m)^2\right)$ random bits. In order to prove that $W$ is $k$-wise independent with probability $1 - o(1)$, we inspect the proof of Theorem 3.6.

Recall that until the equation (2) the proof was distribution-independent. Then, in equation (2) we used the independence of the columns (and the columns of $W$ remain independent for our algorithm), bounded the probability $R_i$ of a column satisfying a linear constraint $\vec{\alpha}$ of Hamming weight $\|\vec{\alpha}\|_0 = i$, and bounded each term in the sum $\sum_{i=2}^{k} \binom{m}{i} P_{[i]}$.

- For $2 \leq i \leq n/(4 \log m)$, recall that for a uniformly random matrix $M$, the Littlewood-Offord Lemma over the reals (Lemma 2.4) implies that the probability that a random column satisfies a fixed linear equation with $i$ non-zero coefficients is bounded by $R_i^{\mathsf{uni}} = \frac{1}{2^i}\binom{i}{\lfloor i/2 \rfloor} \leq 1/2$. Since every column of $W$ $(1/\sqrt{m})$-fools hyperplanes, the probability that a random column of $W$ satisifes a fixed linear equation is bounded by
$$R_i \leq R_i^{\mathsf{uni}} + \frac{1}{\sqrt{m}} \leq \frac{2}{3}\,.$$
This gives us the following bound on $\binom{m}{i} P_{[i]}$:
$$\binom{m}{i} P_{[i]} \leq \binom{m}{i}\binom{n}{i-1} R_i^{n-i-1} \leq m^{2i} \left(\frac{2}{3}\right)^{n-o(n)} \leq 2^{n/2} \cdot \left(\frac{2}{3}\right)^{n-o(n)} \leq 2^{-\Omega(n)}.$$
- For $n/(4 \log m) < i \leq k = \varepsilon n$, by Lemma 2.4, $R_i^{\mathsf{uni}} \leq 1/\sqrt{i}$. Therefore, for a PRG that $(1/\sqrt{m})$-fools hyperplanes,
$$R_i \leq \frac{1}{\sqrt{i}} + \frac{1}{\sqrt{m}} \leq \frac{O(\log m)}{\sqrt{n}}$$
as in the proof of Theorem 3.6, giving the same bound $o(1/k)$ for each term as before.

Now we show that $k$-wise independence of $W$ over $\mathbb{R}$ implies its $k$-wise independence over $\mathbb{F}_p$ for every $p > (\varepsilon n)!$. Indeed, from $k$-wise independence of $W$ over the reals, we have that every set of $k$ rows contains a non-singular $k \times k$ submatrix. Since the determinant of an $\varepsilon n \times \varepsilon n$ square $\pm 1$ matrix is at most $(\varepsilon n)!$, the deteminant stays non-zero after taking it modulo $p$. Therefore, every set of $k$ rows of $W$ contains a non-singular $k \times k$ submatrix over $F_p$, which implies $k$-wise independence over $\mathbb{F}_p$. □

We observe that the matrix $W \in \{\pm 1\}^{m \times n}$ from the previous theorem can be also transformed into a $(k-1)$-wise independent matrix from $W' \in \{0,1\}^{m \times n}$.

**Corollary 3.13.** *For every $\varepsilon \in (0, 1/2]$, every large enough $n \leq m \leq n^{\frac{1}{2\varepsilon}}$, there is a polynomial-time algorithm that uses $O\left(n \log m (\log\log m)^2\right)$ random bits and outputs a matrix $W' \in \{0,1\}^{m \times n}$, s.t. $W'$ is $(k-1) \coloneqq (\varepsilon n - 1)$-wise independent over $\mathbb{R}$ (and over $\mathbb{F}_p$ for any prime $p > (\varepsilon n)!$) with probability $1 - o(1)$.*

*Proof.* First, we generate a matrix $W$ as in Theorem 3.12 with the same parameters $n, m, \varepsilon$. Then we multiply each column by a fresh independent and uniform number from $\pm 1$. Finally, we add 1 to all the entries in the matrix, and divide the result by 2. Let us denote the resulting matrix by $W' \in \{0,1\}^{m \times n}$.

Consider the following distribution of matrices. Take the matrix $W$ and prepend a uniformly random $\pm 1$ row to it. In effect, we are replacing the $F_{\mathsf{GKM}}$ PRG with a new PRG $F'_{\mathsf{GKM}} : \{0,1\}^{r+1} \to \{\pm 1\}^{m+1}$ that generates the first output bit according to the first input bit, and generates the rest using $F_{\mathsf{GKM}}$ applied to all the other $r$ input bits. It is easy to see that $F'_{\mathsf{GKM}}$ $\varepsilon$-fools hyperplanes if $F_{\mathsf{GKM}}$ $\varepsilon$-fools hyperplanes, so Theorem 3.12 also holds for $F'_{\mathsf{GKM}}$ (with $n$ additional bits of randomness in total). We multiply some columns of this matrix by $-1$ so that the first row contains only 1s. After adding the first row to all other rows of the matrix, and dividing them by 2, the resulting matrix from $\{0,1\}^{m \times n}$ is $(k-1)$-wise independent (by the reasoning from Corollary 3.8). Note that the distribution of these matrices is identical to the distribution of matrices $W'$ described above. □

## 3.5 Putting it all together

Now we are ready to prove Theorem 1.1.

**Theorem 1.1.** *There is an explicit set $\mathcal{P}_n$ of $m = n^3$ convex polygons in the plane $\mathbb{R}^2$, and a prime $p$, such that any group-model data structure of size $s$ and query time $t$ for range counting with respect to $\mathcal{P}_n$ (over $\mathbb{Z}_p$), must have*

$$t^s \geq n^{n/7}.$$

*In particular, any linear storage ($s = O(n)$) data structure for $\mathcal{P}_n$(over $\mathbb{Z}_p$) must have $n^{\Omega(1)}$ query time. Moreover, the semialgebraic complexity of $\mathcal{P}_n$, i.e., number of facets of each convex polygon, is $n^{O((\log\log n)^2)}$.*

*Proof.* Let $m = n^3$, and $G\colon \{0,1\}^r \to \{0,1\}^m$ for $r = O\left(\log m (\log\log m)^2\right)$ be the PRG used in Corollary 3.13 for generating columns.[10] Let $R = 2^r$, and $a_1, \ldots, a_R$ be an arbitrary set of $R$ distinct points on a circle in $\mathbb{R}^2$. We will associate these points with binary string of length $r$—all possible values of $G$'s seed.

The set $\mathcal{P}_n$ of $m$ convex polygons in the plane is defined as follows: the $i$th polygon has $a_j$ as its vertex if and only if the $i$th bit of the output of $G(a_j)$ is 1.

We constructed an explicit set $\mathcal{P}_n$ of $m$ convex polygons, and since each polygon has at most $R$ facets, its semialgebraic complexity is bounded from above by $R = n^{O((\log\log n)^2)}$. It remains to show that any $(s,t)$ group-model data structure must satisfy $t^s \geq n^{n/7}$.

Let us fix $\varepsilon = 1/6$, and let $p$ be a prime in the range $[2^{\varepsilon n \log n}, 2^{\varepsilon n \log n + 1}]$. For a set of input points $\boldsymbol{x} = (a_{i_1}, \ldots, a_{i_n})$, let $\mathcal{I}_{\mathcal{P}}(\boldsymbol{x}) \in \{0,1\}^{m \times n}$ be a matrix whose $(i,j)$'th entry indicates whether point $x_j$ lies in the $i$th polygon $r_i \in \mathcal{P}_n$. Since $p > (\varepsilon n)!$ and $m = n^3 \leq n^{1/(2\varepsilon)}$, by Corollary 3.8, there exists a set of input points $\boldsymbol{x}$ such that $\mathcal{I}_{\mathcal{P}}(\boldsymbol{x})$ is $\varepsilon n$-wise independent.

Note that for the set of input points $\boldsymbol{x}$ and their weights $\boldsymbol{w} \in \mathbb{Z}_p^n$, the outputs to the $m$ queries must be $\mathcal{I}_{\mathcal{P}}(\boldsymbol{x})\boldsymbol{w}$. Since $\mathcal{I}_{\mathcal{P}}(\boldsymbol{x})$ is $\varepsilon n$-wise independent and $p < 0.5\left(\frac{m}{16\varepsilon n}\right)^{\varepsilon n/2}$, by Theorem 3.5, there exists a set of weights $\boldsymbol{w}$, such that the vector $\mathcal{I}_{\mathcal{P}}(\boldsymbol{x})\boldsymbol{w}$ is diverse.

Finally, by Lemma 3.2, for any group-model data structure $\mathcal{D}$ over $\mathbb{Z}_p$ that uses $s$ cells of space and has query time $t$, the set of its $m$ possible answers can be diverse only if $(4t)^s \geq p$. Therefore,

$$t^s \geq p^{6/7} \geq 2^{6\varepsilon n \log n/7} = n^{n/7}\,.$$

□

[10] The seed length of $G$ is exactly the seed length of the $\mathsf{GKM}$ PRG plus one bit used for changing the sign of the column.

We remark that a construction of a PRG against halfspaces with optimal seed length $O_\varepsilon(\log m)$ would improve the semialgebraic complexity of the polygons in the theorem statement: instead of polygons with $n^{O((\log\log n)^2)}$ facets, one could construct a set of $n^3$ polygons with $n^{O(1)}$ facets.

**Corollary 3.14.** *If there exists an explicit PRG $G\colon \{0,1\}^r \to \{0,1\}^m$ which $\varepsilon$-fools halfspaces and uses optimal seed $r = O(\log(m/\varepsilon))$, then there exists an explicit set $\mathcal{P}_n$ of $m = n^3$ convex polygons in the plane $\mathbb{R}^2$ with* poly$(n)$ facets/semialgebraic complexity*, such that any group-model data structure of size $s$ and query time $t$ for weighted range counting with respect to $\mathcal{P}_n$ (over some $\mathbb{Z}_p$), must have $t^s \geq n^{\Omega(n)}$.*

# 4 Geometric PRGs

The construction of the incidence matrix $\mathcal{I}_\mathcal{R}$ in the proof of Theorem 1.1 uses affine PRGs in a *black-box* manner: the $i$th row of $\mathcal{I}_\mathcal{R}$ encodes the truth table of the *$i$th output bit* of the (GKM) PRG $g_i : \{\pm1\}^r \mapsto \{\pm1\}$, hence the complexity of the $i$th "range" $g_i \in \mathcal{R}$ (number of polytope facets) is trivially bounded by $2^r = n^{O(\log^2\log n)}$. This is essentially inevitable if the PRG is applied in a black-box fashion (since $r > \log m$ is the best one could hope for [MZ13]). However, we might hope to obtain lower bounds against more "natural" range counting problems (with lower semi-algebraic complexity) using our approach, by constructing PRGs that take advantage of the *underlying representation* of the input seed. Indeed, the input to a range counting problem is a point in a $d$-dimensional space, hence it is natural to ask if the PRG can exploit higher dimensionality in a "simple" way so as to reduce the semi-algebraic complexity of the underlying hard problem. This is the motivation of our next definition.

**Definition 4.1** (Geometric PRGs)**.** *A function $g : [B]^d \to \{\pm1\}^m$ is said to be a $(B, d, s, \varepsilon)$-*Geometric PRG *(GPRG) against a family of sets $\mathcal{S} \subset 2^{\{\pm1\}^m}$, if $g$ $\varepsilon$-fools $\mathcal{S}$, and furthermore every output bit of $g$ can be encoded as a low-degree polynomial threshold function (PTF), i.e.,*

$$\forall\, i \in [m]\,, \quad g_i(r_1, \ldots, r_d) = \mathsf{sign}\left(P_i(r_1, \ldots, r_d)\right),$$

*where each $P_i$ is a $d$-variate polynomial of (total) degree $s$.*

Note that this definition generalizes the (incidence function of) halfspaces ($s = 1$), and their natural extension to ranges specified by low-degree PTFs. We now show that an explicit construction of Geometric PRGs would yield a data structure lower bound against a corresponding class of ranges.

To this end, note that in the proof of Theorem 3.12, we showed that a PRG that $1/\sqrt{m}$ -fools halfspaces yields an $\varepsilon n$-wise independent matrix (which, by Theorem 1.1, implies a data structure lower bound). By inspection of the proof of Theorem 3.12, we see that for any constant $\gamma > 0$, even a PRG that $1/m^\gamma$-fools halfspaces still gives us an $\Omega_\gamma(n)$-wise independent matrix, yielding a data structure lower bound of $t^s \geq n^{\Omega(n)}$ by Theorem 1.1.

Using the preceding remark, a $(B, d, s, 1/m^\gamma)$-Geometric PRG $g : [B]^d \to \{\pm1\}^m$ gives an explicit set of $m$ ranges, each of which defined by a single $d$-variate polynomial of degree $s$, such that range counting with respect to these ranges is hard for linear storage data structures:

**Lemma 4.2.** *Suppose $g : [B]^d \to \{\pm1\}^m$ is a $(B, d, s, 1/m^\gamma)$-Geometric PRG (GPRG) against affine halfspaces in $\mathbb{R}^m$ for a constant $\gamma > 0$. Then there exists a range counting problem $\mathcal{P}_\mathcal{R}$ with $|\mathcal{R}| = m$ ranges, each of semi-algebraic complexity $\Gamma_{d,s,1}$, such that any $S$-space data structure for $\mathcal{P}_\mathcal{R}$ has query time $t \geq n^{\Omega(n/S)}$ in the static group model.*

## 4.1 Existence of high-degree Geometric PRGs

In contrast to the standard definition of PRGs (Definition 3.9), whose mere *existence* with small seed length follows a rather simple counting argument [MZ13], the existence of GPRGs (with sublinear seedlength) against affine halfspaces, is a nontrivial question. The remainder of this section is devoted to this question.

Our first result shows that GPRGs against halfspaces exist in nontrivial dimension $d = n^\varepsilon$ (for any $\varepsilon > 0$), albeit with high degree $s \approx n^\varepsilon$ (which corresponds to ranges being high-degree PTFs). In fact, we will show an *explicit* construction of such GPRGs, by "adapting" BCH codes from $\mathbb{F}_2$ to $\mathbb{R}$.

An important ingredient of the proof is the following theorem due to Diakonikolas et al. [DGJ+10], asserting that *any* $k$-wise (statistically) independent distribution over $\{\pm 1\}^m$ $\tilde{\Omega}(1/\sqrt{k})$-fools halfspaces:

**Theorem 4.3** (Bounded independence fools halfspaces [DGJ+10])**.** *Let $\mathcal{D}$ be a $k$-wise independent distribution on $\{\pm 1\}^m$. Then $\mathcal{D}$ $\delta$-fools halfspaces as long as*

$$k \geq \frac{C}{\delta^2} \log^2 \left(\frac{1}{\delta}\right)$$

*for some universal constant $C$.*

By setting $\delta$ in Theorem 4.3 to $m^{-0.05}$, an immediate corollary is that any $k = (Cm^{0.1} \log^2 m)$-wise independent distribution will fool halfspaces with small enough error for Lemma 4.2 to yields data structure lower bounds. Thus, it remains to construct a GPRG whose output is $Cm^{0.1} \log^2 m$-wise statistically independent.

A $k$-wise independent distribution with uniform marginals is easy to construct from a parity check matrix $M$ of a linear code of minimal distance $> k$ over $\mathbb{F}_2^m$: since any $k$ of the $M$'s columns are linearly independent over $\mathbb{F}_2$, taking the sum of a uniformly random subset of the rows of $M$ gives a vector of length $m$ whose entries are statistically $k$-wise independent. In fact, this is the optimal way to obtain such a $k$-wise independent distribution, as Alon, Babai, Itai showed:

**Theorem 4.4** ([ABI86])**.** *Suppose $(\xi_i)_{i=1}^m$ is a collection of $k$-wise independent random variables $\xi_i : \Omega \to \mathbb{R}$ each of which takes at least two distinct values with non-zero probability. Then the size of the sample space is at least the size of the Hamming ball of radius $\lfloor k/2 \rfloor$: $|\Omega| \geq \sum_{j=0}^{\lfloor k/2 \rfloor} \binom{m}{j}$.*

*Moreover, when $m = 2^t - 1$ and $k = 2s+1$, this bound is essentially achieved for $\xi_i = \langle H^i, x \rangle_{\mathbb{F}_2}$ where $H^i$ is the $i$-th column of the parity check matrix $H \in \mathbb{F}_2^{(1+st)\times m}$ of a binary BCH code and $x$ is chosen uniformly at random from $\mathbb{F}_2^{1+st}$.*

To make this construction "geometric", treat $H$ as a binary matrix over the reals and $x$ as a binary vector over the reals. Then the entries of the vector $H^T x$ are integers that are $k$-wise independent modulo 2, and are between 0 and $k \log m/2$. Let $p(x) := \prod_{i=1}^{k \log m/2} (i - 1/2 - x)$, then $\mathsf{sign}(p(x)) = (-1)^x$ for $x \in \{0, \ldots, k \log m/2\}$, giving our desired theorem:

**Theorem 4.5** (Explicit high-degree GPRGs)**.** *Let $\mathcal{R}_i(x) = p(\langle H^i, x \rangle)$, where $H$ is the parity check matrix defined above treated as a binary matrix over $\mathbb{R}$, and $x \in \{0,1\}^r$ for $r = O(m^{0.1} \log^3 m)$. Then $\mathcal{R} : \{0,1\}^r \to \{\pm 1\}^m$ is a $(2, O(m^{0.1} \log^3 m), O(m^{0.1} \log^3 m), 1/\mathrm{poly}(n))$-GPRG against affine halfspaces in $\mathbb{R}^m$.*

This theorem asserts a polynomial lower bound for range counting of $n^{O_\varepsilon(1)}$ explicit PTFs in $\mathbb{R}^{n^\varepsilon}$, each of degree $\tilde{O}(n^{0.1})$, against linear storage group-model data structures. Is it possible to construct GPRGs using *constant degree* PTFs? This is the content of the next subsection.

## 4.2 Can Halfspaces Fool Halfspaces ?

The special case of degree-1 PTFs is particularly interesting in our context, as it corresponds to the very natural problem of *halfspace* range counting, which is among the most well-studied range counting problems in the literature [Cha90, Aga17]. While Theorem 4.5 produces an *explicit* high-degree GPRG, for degree $s = 1$ it is not even clear whether GPRGs *exist* in nontrivial dimension, i.e., whether "halfspaces fool halfspaces" in sublinear dimension :

**Problem 4.6.** *Is there any set $\mathcal{H}$ of $m = \mathrm{poly}(d)$ halfspaces in $\mathbb{R}^d$ for $d \ll n$, such that the function given by $g_i(r) := (\mathsf{sign}\langle h_i, r \rangle)_{i \in [m]}$ for $r \in_R \{\pm 1\}^d$, $(1/\mathrm{poly}(n))$-fools affine halfspaces in $\mathbb{R}^m$?*

We note that the (explicit) BCH-code construction in Theorem 4.5 shows these parameters are achievable *over* $\mathbb{F}_2$, i.e., had we taken the *parity* of the inner-products $\langle h_i, r\rangle$ instead of their *signs*. Intuitively, threshold functions are much less sensitive than the parity operator (in particular, thresholds are monotone function and hence have low influence in sharp contrast to parities). Hence, achieving $n^{\Omega(1)}$-wise independence (by Theorem 4.3) with signs of inner products seems to require a substantially different construction and analysis.

The main technical result of this section is showing that it is possible to construct an *almost* $k$-wise independent distribution, i.e., a $(\delta/2^k, k)$-independent distribution[11] over $\{\pm 1\}^m$, with $\delta = 1/\mathrm{poly}(d)$, from $d$-dimensional halfspaces over the Boolean hypercube, which implies that the same distribution is $\delta$-far from $k = d^{\Omega(1)}$-wise independence, i.e. all the $k$-marginals are $\delta$-close to uniform in statistical distance:

**Theorem 4.7** (Random halfspaces are almost $k$-wise independent)**.** *For a fixed number $p$, pick $m = d^p$ uniformly random vectors $v_1, \ldots, v_m$ in $\{\pm 1\}^{2d+1}$ for large enough $d$, and let $\varepsilon < 1/6$. Then for a random vector $\boldsymbol{x} \in_R \{\pm 1\}^{2d+1}$, the distribution of the binary vector $f(\boldsymbol{x}) := \{\mathsf{sign}(\langle v_i, \boldsymbol{x}\rangle)\}_{i=1}^m$ is $(d^{1/6}k/2^k, k)$-independent with high probability (over the choice of the $v_i$'s), where $k = d^\varepsilon$.*

We need to show that all the $k$-marginals of $f$ are close to uniform $d$-bit vectors in $L_\infty$ norm. We will bound the probability that some fixed $k$-marginal is far from a uniform and union bound over the $\binom{m}{k}$ possible choices of $k$ coordinates. This will be implied by a bound on the measure of the intersection of $k$ random halfspaces with the hypercube — we want to say that that measure is very close to $2^{-k}$ with high probability.

We will need the following basic notions from Fourier analysis over the hypercube to prove the above theorem. For a Boolean function $f : \{\pm 1\}^n \to \mathbb{R}$ its Fourier coefficients are defined by $\widehat{f}(\omega) = \mathbb{E}_x[\chi_\omega(x) f(x)]$, where $\omega \subseteq [n]$ and $\chi_\omega(x) = \prod_{i \in \omega} x_i$. The characters $\chi_\omega$ are orthonormal with respect to $\langle f, g\rangle = \mathbb{E}_x[f(x)g(x)]$ and $f$ can be uniquely rewritten in the Fourier basis:

$$f(x) = \sum_{\omega \subseteq [n]} \chi_\omega(x)\widehat{f}(\omega).$$

The following Bonami-Gross-Beckner hypercontractive inequality for the Boolean hypercube is well-known:

**Theorem 4.8** ([Bon70, Gro75, Bec75])**.** *Let $f : \{\pm 1\}^n \to \mathbb{R}$. Define $T_\rho(f)(x) = \sum_{\omega \subseteq [n]} \rho^{|\omega|}\chi_\omega(x)\widehat{f}(x)$ — the noise operator with parameter $\rho$, and let $\|f\|_p = \mathbb{E}_x[|f(x)|^p]^{1/p}$. Then for any $1 \le p \le 2$,*

$$\|T_{\sqrt{p-1}} f\|_2 \le \|f\|_p$$

We will need the following corollary of Theorem 4.8, which also appears in [GKK+08, dW08, KR11]:

**Lemma 4.9.** *Let $f' : \{\pm 1\}^n \to \{-1, 0, 1\}$ with support $A := \mathrm{supp}(f') = \{x : f'(x) \neq 0\}$, and let $f = f'/\mu(A)$ be the normalized $f'$. Then the bound on the Fourier mass on level $k$ holds:*

$$\sum_{|\omega|=k} \widehat{f}(\omega)^2 \le \left(e \max\left(1, \frac{2}{k} \ln \frac{1}{\mu(A)}\right)\right)^k,$$

*where $\mu$ denotes the uniform measure on the hypercube.*

*Proof.* Applying Theorem 4.8 we get:

$$\begin{aligned}
\sum_{|\omega|=k} \widehat{f}(\omega)^2 &\le \frac{1}{(p-1)^k} \sum_{\omega \subseteq [n]} (p-1)^{|\omega|} \widehat{f}(\omega)^2 \\
&= \frac{1}{(p-1)^k} \|T_{\sqrt{p-1}} f\|_2^2 \\
&\le \frac{1}{(p-1)^k} \|f\|_p^2 = \frac{1}{(p-1)^k} \mu(A)^{-2(1-1/p)}.
\end{aligned}$$

[11] This is a distribution where all marginals on $k$ coordinates are $\delta/2^k$-close in $L_\infty$ norm to uniform, see [AGHP92] for details and its wide applications.

For $p = 1 + \frac{k}{2\ln(1/\mu(A))}$ when $k \le 2\ln(1/\mu(A))$ this is at most $\left(\frac{2e}{k}\ln(1/\mu(A))\right)^k$, and when $k \ge 2\ln(1/\mu(A))$ the sum of all squared Fourier coefficients is $\|f\|_2^2 = 1/\mu(A) \le e^k$ as desired. □

We shall use the following estimate on the Fourier coefficients of the Majority function[12] on $2d+1$ inputs (the proof of this claim is provided in Appendix B).

**Claim 4.10.** *The Fourier decomposition of the majority function* $\mathrm{Maj}(x) := [\sum_i x_i \ge 0]$ *on* $\{\pm 1\}^{2d+1}$ *is:*

$$\mathrm{Maj}(x) = \frac{1}{2} + \mu_1 \sum_{|\omega|=1} \chi_\omega(x) + \mu_2 \sum_{|\omega|=2} \chi_\omega(x) + \ldots + \mu_{2d+1} \sum_{|\omega|=2d+1} \chi_\omega(x),$$

*where each* $\mu_{2k} = 0$ *for* $k > 0$*, and the odd coefficients are:*

$$\mu_{2k+1} = \binom{2d}{d}\binom{d}{k}\binom{2d}{2k}^{-1}\frac{(-1)^k}{2^{d+1}}.$$

The following lemma, whose proof follows closely the argument of an analogous statement by Klartag and Regev [KR11] but for "equators" in *Gaussian spaces* (instead of halfspaces in the hypercube), almost finishes the proof using an inductive application:

**Lemma 4.11.** *Let* $A, B$ *be two subsets of* $\{\pm 1\}^{2d+1}$ *of measure at least* $\exp(-d^{1/3})$*. Then*

$$\Pr_{y\sim B, x\sim \mathrm{Cap}(y)}[x \in A] \in (1 \pm d^{-1/6})\mu(A),$$

*where* $\mathrm{Cap}(y) := \{x \in \{\pm 1\}^{2d+1} : \langle x, y\rangle > 0\}$ *is the set of points at Hamming distance at most* $d$ *from* $y$*.*

*Proof.* Let $f := \mathbf{1}_A/\mu(A)$, $g := \mathbf{1}_B/\mu(B)$ be the normalized indicator functions of the two sets. Define the operator $R^{\mathrm{Cap}}$ on Boolean functions by $R^{\mathrm{Cap}}(h)(y) := \mathbb{E}_{x\sim\mathrm{Cap}(y)}[h(x)]$. Then

$$\Pr_{\substack{y\sim B\\ x\sim\mathrm{Cap}(y)}}[x \in A] = \mu(A)\mathbb{E}_y[g(y)R^{\mathrm{Cap}}(f)(y)] = \mu(A)\sum_\omega \widehat{g}(\omega)\widehat{R^{\mathrm{Cap}}(f)}(\omega),$$

so we need to bound this sum. Notice that $R^{\mathrm{Cap}}$ is just a convolution with the majority function defined above: $R^{\mathrm{Cap}}(f)(y) = 2\,\mathbb{E}_x[f(x \odot y)\mathrm{Maj}(x)]$, so in the Fourier basis $R^{\mathrm{Cap}}$ is diagonal with the Fourier coefficients of 2Maj as eigenvalues, giving:

$$\mu(A)^{-1}\Pr_{\substack{y\sim B\\ x\sim\mathrm{Cap}(y)}}[x \in A] = 1 + 2\mu_1\sum_{|\omega|=1}\widehat{f}(\omega)\widehat{g}(\omega) + 2\mu_3\sum_{|\omega|=3}\widehat{f}(\omega)\widehat{g}(\omega) + \ldots$$

Now we need to bound the sums on each level. By Cauchy-Schwarz to bound $\sum \widehat{f}(\omega)\widehat{g}(\omega)$ it is enough to bound $\sum \widehat{f}(\omega)^2$ and $\sum \widehat{g}(\omega)^2$. Lemma 4.9 implies

$$\sum_{|\omega|=2k+1}\widehat{f}(\omega)^2 \le \left(e\max\left(1, \frac{2}{2k+1}\ln\frac{1}{\mu(A)}\right)\right)^{2k+1}$$

Multiplying $f$ by the parity on all the coordinates $\chi_{[2d+1]}$ by the same Lemma 4.9 we also get the symmetric upper bound

$$\sum_{|\omega|=2k+1}\widehat{f}(\omega)^2 \le \left(e\max\left(1, \frac{2}{2d-2k}\ln\frac{1}{\mu(A)}\right)\right)^{2d-2k}.$$

[12]Note that we define Maj to take values 0 and 1, even though it takes input vectors with ±1 coordinates

We split the sum into four parts: with $|\omega| \le \delta d$, $\delta d \le |\omega| \le (2-\delta)d$, $(2-\delta)d \le |\omega| \le 2d$, and the one term with $|\omega| = 2d+1$, for small enough universal constant $\delta$ to be chosen.

We now upper bound the ratio of the term for $|\omega| = 2k+3$ to the term for $|\omega| = 2k+1$, showing that the terms are geometrically decreasing. First, the ratio of eigenvalues is

$$\mu_{2k+3}/\mu_{2k+1} = \binom{d}{k+1}\binom{2d}{2k}\binom{d}{k}^{-1}\binom{2d}{2k+2}^{-1} = \frac{(d-k)(2k+2)(2k+1)}{(k+1)(2d-2k)(2d-2k-1)} = \frac{2k+1}{2d-2k-1}$$

The ratio of Fourier masses at two adjacent non-empty levels is:

$$\left(e\max\left(1, \frac{2\ln(1/\mu(A))}{2k+3}\right)\right)^{2k+3} \Bigg/ \left(e\max\left(1, \frac{2\ln(1/\mu(A))}{2k+1}\right)\right)^{2k+1} \le e^2 \max\left(1, \frac{2\ln(1/\mu(A))}{2k+1}\right)^2$$

Combining the above bound with Cauchy-Schwarz, the ratio of the product is bounded by

$$e^2 \frac{2k+1}{2d-2k-1} \max\left(1, \frac{2\ln(1/\mu(A))}{2k+1}\right) \max\left(1, \frac{2\ln(1/\mu(B))}{2k+1}\right),$$

which is at most $1/2$ for a small enough constant $\delta$ since $\ln(1/\mu(A))\ln(1/\mu(B)) \le d^{2/3}$.

Since the sum for the first-level Fourier coefficients is:

$$\mu_1 \sum_{i\in[n]} \widehat{f}(\{i\})^2 \le \frac{2e}{\sqrt{2d}} \ln\frac{1}{\mu(A)} \le O(d^{-1/6}),$$

the contribution of the first part to the total sum can be bounded by $O(d^{-1/6})$.

For the middle part $\delta d \le 2k+1 = |\omega| \le (2-\delta)d$, notice that the absolute values of $\mu_{2k+1}$ are decreasing until $2k = d$ and are (almost) symmetric around $k = d/2$, so we can bound the contribution of the middle terms to the sum by

$$|\mu_{\delta d}| \sum_{\delta d \le |\omega| \le (2-\delta)d} \widehat{f}(\omega)\widehat{g}(\omega) \le \frac{|\lambda^{2d}_{\delta d}|}{\sqrt{\mu(A)\mu(B)}}.$$

Approximating $|\mu_{\delta d}| \le \binom{d}{\delta d}\binom{2d}{2\delta d}^{-1}/\sqrt{2d} = \exp(-dH(\delta) + o(d))$ where $H$ is the entropy function in nats $H(x) = -x\ln x - (1-x)\ln(1-x)$, we get an exponentially small bound on the middle terms.

The bound on the third part is the same as for the first part, but the first non-zero weight from the end is even smaller in this case because there is no weight on the level $2d$.

Now we need to bound the largest coefficient $|\lambda^{2d}_{2d}|\widehat{f}([2d+1])^2 \le \widehat{f}([2d+1])^2/\sqrt{2d}$, but $\widehat{f}(\omega) = \mathbb{E}_x[\chi_\omega(x)f(x)]$ is at most $\mathbb{E}_x[\mathbf{1}_A/\mu A] = 1$ for any set $\omega$, so we are done. □

With this lemma in hand we can prove Theorem 4.7.

*Proof of Theorem 4.7.* Take $k = d^\varepsilon$ uniformly random vectors $y_1, \dots, y_k \in \{\pm 1\}^{2d+1}$. Denote by $A_t$ the intersection of the first $t$ halfspaces: $A_t = \mathrm{Cap}(y_1) \cap \mathrm{Cap}(y_2) \cap \cdots \cap \mathrm{Cap}(y_t)$. We want to show that the measure of $A_t$ is within $[\exp(-2td^{-1/6}), \exp(td^{-1/6})]2^{-t}$ for all $t$ with high probability. Notice that $A_t = \mathrm{Cap}(y_t) \cap A_{t-1}$, so if

$$\mu(A_{t-1}) \in \frac{1}{2^{t-1}}\left[\exp(-2(t-1)d^{-1/6}), \exp((t-1)d^{-1/6})\right] < \exp(d^{-1/3}), \tag{3}$$

then by Lemma 4.11 the probability of $\mu(A_i) \notin (1 \pm d^{-1/6})\mu(A_{i-1})/2$ is at most $\exp(-d^{1/3})$. Indeed, we let $A := A_{t-1}$ and let $B$ be the set of all points $y$ whose corresponding halfspaces intersect $A$ by too much (same works for the halfspaces whose intersection is too small):

$$B := \left\{y : \mu(A_{t-1} \cap \mathrm{Cap}(y)) > (1+d^{-1/6})\mu(A_{t-1})/2\right\}.$$

The measure of such $B$ by Lemma 4.11 can be at most $\exp(-d^{1/3})$, so conditioned on the set $A_{t-1}$ having the appropriate measure as in (3), $A_t$ will have measure

$$\mu(A_t) \in \frac{1 \pm d^{-1/6}}{2^t}\left[\exp(-2(t-1)d^{-1/6}),\ \exp((t-1)d^{-1/6})\right] \subset \frac{1}{2^t}\left[\exp(-2td^{-1/6}),\ \exp(td^{-1/6})\right]$$

with probability at least $1-2\exp(-d^{1/3})$. Here we have used $e^{-2t} \leq 1-t$ for small enough $t$ and $1+t \leq \exp(t)$ for all $t$.

By union bound the probability that $\mu(A_k)$ will lie in the small interval around $2^{-k}$ is at least $1 - 2k\exp(-d^{1/3})$. Moreover, we can also bound the probability that *all* the $2^k$ possible intersections of halfspaces (taking either $\mathrm{Cap}(y_t)$ or $\overline{\mathrm{Cap}(y_t)}$ for each $t$) have measure close to $2^{-k}$. Indeed, we can imagine a complete binary tree with $k+1$ layers, and at a node $v$ at level $t-1$ we are choosing whether to take $H = \mathrm{Cap}(y)$ or $H = \overline{\mathrm{Cap}(y)}$, and travel down along the corresponding edge and write the new set $A_v \cap H$. Conditioned on $A_v$ being close to correct size as in (3), the same reasoning applies as above, so by union bound the probability that on at least one edge in the tree the difference in measures will be too large or too small is at most $2 \cdot 2^k \exp(d^{-1/3})$. This means that the $L_\infty$ distance between uniform distribution on $\{0,1\}^k$ and the vector of indicators $\big([x \in \mathrm{Cap}(y_1)], [x \in \mathrm{Cap}(y_2)], \ldots, [x \in \mathrm{Cap}(y_k)]\big)$ when $x \in \{\pm 1\}^{2d+1}$ is chosen uniformly at random will be at most

$$\frac{1}{2^k}\max\left(1 - \exp(-2kd^{1/6}), \exp(kd^{1/6}) - 1\right) \leq \frac{2kd^{1/6}}{2^k}$$

with probability at least $1 - 2 \cdot 2^k \exp(-d^{1/3})$.

Applying the union bound over all possible choices of $k$ out of $m$ coordinates of the vector $f(x) = (\mathsf{sign}(\langle y_i, x\rangle))_{i=1}^m$ from the theorem statement, we get that the probability that all the $k$-marginals of this vector are $\frac{2kd^{1/6}}{2^k}$-close to the uniform when evaluated on a uniformly random point $x$ is at least

$$1 - 2 \cdot 2^k \exp(-d^{1/3})\binom{m}{k} \geq 1 - \exp\left(-d^{1/3} + k\ln 2 + k\ln(em) - k\ln k\right),$$

which is exponentially close to 1 since $k = d^\varepsilon < d^{1/6}$ and $m$ is polynomial in $d$. □

If Theorem 4.3 were "robust" to distributions *statistically-close* to being $k$-wise independent, then Theorem 4.7 would have established the existence of a degree-1 (halfspace) GPRG against halfspaces in $n^\varepsilon$ dimensions. Unfortunately, the proof of Theorem 4.3 only applies to distributions which are *exponentially* close (in $k$) to $k$-wise independence (whereas the distance $\delta = 1/\mathrm{poly}(k)$ obtained in Theorem 4.7 is essentially tight by considering the correlation between just 2 coordinates). Nonetheless, Theorem 4.3 is a necessary milestone toward the fooling question, and we believe the proof technique of Theorem 4.7 may be useful to directly settle the existence of GPRGs from halfspaces.

# Acknowledgement

We are grateful to Chin Ho-Lee, Rocco Servedio, Eshan Chattopadhyay, Swastik Kopparty and Jarek Błasiok for helpful pointers and insightful discussions on this work.

# References

[ABI86] Noga Alon, László Babai, and Alon Itai. A fast and simple randomized parallel algorithm for the maximal independent set problem. *J. Algorithms*, 7(4):567–583, 1986.

[ABN+92] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M. Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Trans. Inf. Theory*, 38(2):509–516, 1992.

[Afs19] Peyman Afshani. A new lower bound for semigroup orthogonal range searching. In *SoCG 2019*, pages 3:1–3:14. LIPIcs, 2019.

[Aga17] Pankaj K. Agarwal. Simplex range searching and its variants: A review. In *A Journey Through Discrete Mathematics*, pages 1–30. Springer, 2017.

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple construction of almost $k$-wise independent random variables. *Random Struct. Algorithms*, 3(3):289–304, 1992.

[AM94] Pankaj K. Agarwal and Jirí Matoušek. On range searching with semialgebraic sets. *Discrete Comput. Geom.*, 11:393–418, 1994.

[AMS12] Pankaj K. Agarwal, Jirí Matoušek, and Micha Sharir. On range searching with semialgebraic sets II. In *FOCS 2012*, pages 420–429. IEEE, 2012.

[BCP93] Hervé Brönnimann, Bernard Chazelle, and János Pach. How hard is half-space range searching. *Discrete Comput. Geom.*, 10:143–155, 1993.

[Bec75] William Beckner. Inequalities in Fourier analysis on $R^n$. *Proc. Natl. Acad. Sci. USA*, 72:638–641, 1975.

[Ben75] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18:509–517, September 1975.

[BLL+19] Jarosław Błasiok, Patrick Lopatto, Kyle Luh, Jake Marcinek, and Shravas Rao. An improved lower bound for sparse reconstruction from subsampled Hadamard matrices. In *FOCS 2019*, pages 1564–1567. IEEE, 2019.

[Bon70] Aline Bonami. Étude des coefficients de Fourier des fonctions de $L^p(G)$. *Annales de l'institut Fourier*, 20(2):335–402, 1970.

[BR94] Mihir Bellare and John Rompel. Randomness-efficient oblivious sampling. In *FOCS 1994*, pages 276–287. IEEE, 1994.

[CGH+85] B. Chor, O. Goldreich, J. Hasted, J. Freidmann, S. Rudich, and R. Smolensky. The bit extraction problem or $t$-resilient functions. In *SFCS 1985*, pages 396–407. IEEE, 1985.

[Cha90] Bernard Chazelle. Lower bounds for orthogonal range searching: Part II. The arithmetic model. *J. ACM*, 37(3):439–463, 1990.

[Cha94] Bernard Chazelle. A spectral approach to lower bounds. In *FOCS 1994*, pages 674–682. IEEE, 1994.

[CP09] Timothy M. Chan and Mihai Patrascu. Transdichotomous results in computational geometry, I: point location in sublogarithmic time. *SIAM J. Comput.*, 39(2):703–729, 2009.

[CR95] Bernard Chazelle and Burton Rosenberg. Simplex range reporting on a pointer machine. *Comput. Geom.*, 5:237–247, 1995.

[CT06] Emmanuel J. Candès and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory*, 52(12):5406–5425, 2006.

[DGJ+10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. Bounded independence fools halfspaces. *SIAM J. Comput.*, 39(8):3441–3462, 2010.

[DGW19] Zeev Dvir, Alexander Golovnev, and Omri Weinstein. Static data structure lower bounds imply rigidity. In *STOC 2019*, pages 967–978. ACM, 2019.

[DSV12] Alexandros G. Dimakis, Roxana Smarandache, and Pascal O. Vontobel. LDPC codes for compressed sensing. *IEEE Trans. Inf. Theory*, 58(5):3093–3114, 5 2012.

[dW08] Ronald de Wolf. A brief introduction to Fourier analysis on the Boolean cube. *Theory Comput., Graduate Surveys*, 1:1–20, 2008.

[Erd45] Paul Erdös. On a lemma of Littlewood and Offord. *Bull. Am. Math. Soc.*, 51(12):898–902, 1945.

[Fre81] Michael L. Fredman. A lower bound on the complexity of orthogonal range queries. *J. ACM*, 28(4):696–705, 1981.

[GKK+08] Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf. Exponential separation for one-way quantum communication complexity, with applications to cryptography. *SIAM J. Comput.*, 38(5):1695–1708, 2008.

[GKM18] Parikshit Gopalan, Daniel M. Kane, and Raghu Meka. Pseudorandomness via the discrete Fourier transform. *SIAM J. Comput.*, 47(6):2451–2487, 2018.

[GLW08] Venkatesan Guruswami, James R. Lee, and Avi Wigderson. Euclidean sections of $\ell_1^N$ with sublinear randomness and error-correction over the reals. In *RANDOM 2008*, pages 444–454. Springer, 2008.

[Gro75] Leonard Gross. Logarithmic Sobolev inequalities. *Am. J. Math.*, 97(4):1061–1083, 1975.

[Ind08] Piotr Indyk. Explicit constructions for compressed sensing of sparse signals. In *SODA 2008*, pages 30–33, 2008.

[Jus72] Jørn Justesen. Class of constructive asymptotically good algebraic codes. *IEEE Trans. Inf. Theory*, 18(5):652–656, 1972.

[Kol04] Vladlen Koltun. Sharp bounds for vertical decompositions of linear arrangements in four dimensions. *Discrete Comput. Geom.*, 31:2004, 2004.

[Kom67] János Komlós. On the determinant of $(0, 1)$ matrices. *Studia Sci. Math. Hungar.*, 2:7–21, 1967.

[KR11] Bo'az Klartag and Oded Regev. Quantum one-way communication can be exponentially stronger than classical communication. In *STOC 2011*, pages 31–40. ACM, 2011.

[Lar12] Kasper Green Larsen. Higher cell probe lower bounds for evaluating polynomials. In *FOCS 2012*, pages 293–301, 2012.

[Lar14] Kasper Green Larsen. On range searching in the group model and combinatorial discrepancy. *SIAM J. Comput.*, 43(2):673–686, 2014.

[Lue78] George S. Lueker. A data structure for orthogonal range queries. *SFCS 1978*, pages 28–34, 1978.

[Mat94] Jirí Matoušek. Geometric range searching. *ACM Comput. Surv.*, 26:421–461, 1994.

[Mil93] Peter Bro Miltersen. The bit probe complexity measure revisited. In *STACS 1993*, pages 662–671, 1993.

[MS77] Florence Jessie MacWilliams and Neil James Alexander Sloane. *The theory of error-correcting codes*. Elsevier, 1977.

[MZ13] Raghu Meka and David Zuckerman. Pseudorandom generators for polynomial threshold functions. *SIAM J. Comput.*, 42(3):1275–1301, 2013.

[NP18] Hoi H. Nguyen and Elliot Paquette. Surjectivity of near-square random matrices. *Comb. Probab. Comput.*, pages 1–26, 2018.

[NW18] Hoi H. Nguyen and Melanie Matchett Wood. Random integral matrices: universality of surjectivity and the cokernel. *arXiv:1806.00596*, 2018.

[OvL81] Mark H. Overmars and Jan van Leeuwen. Dynamization of decomposable searching problems yielding good worst-case bounds. In *Theor. Comput. Sci.*, pages 224–233. Springer, 1981.

[Păt07] Mihai Pătraşcu. Lower bounds for 2-dimensional range counting. In *STOC 2007*, pages 40–46, 2007.

[Păt08] Mihai Pătraşcu. Unifying the landscape of cell-probe lower bounds. In *FOCS 2008*, pages 434–443, 2008.

[RR20] Sivaramakrishnan Natarajan Ramamoorthy and Cyrus Rashtchian. Equivalence of systematic linear data structures and matrix rigidity. In *ITCS 2020*, pages 35:1–35:20. LIPIcs, 2020.

[Sie04] Alan Siegel. On universal classes of extremely random constant-time hash functions. *SIAM J. Comput.*, 33(3):505–543, 2004.

[SS96] Michael Sipser and Daniel A. Spielman. Expander codes. *IEEE Trans. Inf. Theory*, 42(6):1710–1722, 1996.

[XH07] Weiyu Xu and Babak Hassibi. Efficient compressive sensing with deterministic guarantees using expander graphs. In *ITW 2007*, pages 414–419, Sep. 2007.

[Yao82] Andrew Chi-Chih Yao. Theory and applications of trapdoor functions (extended abstract). In *FOCS 1982*, pages 80–91, 1982.

# A Explicit $(n/\log m)$-wise independent matrices

In order to construct an explicit binary matrix $M \in \{0,1\}^{m\times n}$ which is $k = \frac{n}{\log m}$-wise independent over the reals, it suffices to take the parity-check matrix of an optimal linear error correcting code over $\mathbb{F}_2$ (see, e.g., [Jus72, MS77, ABN$^+$92, SS96, DSV12]). While such constructions are typically quite sophisticated, they also provide a stronger property: the resulting matrix is $k$-wise independent even over $\mathbb{F}_2$. We give an explicit folklore construction of a binary matrix which is $k$-wise independent over the reals (and over $\mathbb{F}_2$). To construct such a matrix, we take a Vandermonde matrix over a finite field and binarize it.

Let $\ell = \lceil \log m \rceil + 1$ and $k = n/\ell$. We assume that $n$ is divisible by $\ell$. Choose a prime $2m > p \geq m$, and take the rectangular Vandermonde matrix $M \in \mathbb{F}_p^{m\times k}$: $M_{i,j} = i^j \mod p$. This matrix is $k$-wise independent over $\mathbb{F}_p$, as any $k$ rows of $M$ form a full-rank Vandermonde matrix . Now apply procedure $\mathrm{Bin} : \mathbb{F}_p \to \{0,1\}^\ell$ to each entry of the matrix, and concatenate the resulting vectors in each row to get $M' \in \{0,1\}^{m\times n}$. The procedure Bin takes an element $x$ of the field, finds 0/1 coefficients $d_0, \ldots, d_{\ell-1}$ such that $\sum_{i=0}^{\ell-1} d_i 2^i = x \mod p$, and returns $(d_0, \ldots, d_{\ell-1})$.

**Proposition A.1** (Folklore)**.** *The matrix $M' \in \{0,1\}^{m\times n}$ constructed above is $k$-wise independent over $\mathbb{F}_p$ and, consequently, over $\mathbb{R}$.*

*Proof.* There is a matrix $H \in \mathbb{F}_p^{n\times k}$ such that $M = M'H$ since there is a linear inverse of the map Bin. If some $k$ rows are linearly dependent in $M'$, they are also linearly dependent in $M$. But any $k$ rows of $M$ are independent, since they form a $k \times n$ Vandermonde matrix with distinct rows (and $k \leq n$).

Linear independence over $\mathbb{F}_p$ implies linear independence over $\mathbb{R}$: if a subset of rows is linearly independent, there is an invertible square submatrix $X$ in those rows, and $\det X \neq 0 \mod p$ implies $\det X \neq 0$ over the reals. □

# B The Fourier coefficients of Maj

**Claim 4.10.** *The Fourier decomposition of the majority function* $\mathrm{Maj}(x) := [\sum_i x_i \geq 0]$ *on* $\{\pm 1\}^{2d+1}$ *is:*

$$\mathrm{Maj}(x) = \frac{1}{2} + \mu_1 \sum_{|\omega|=1} \chi_\omega(x) + \mu_2 \sum_{|\omega|=2} \chi_\omega(x) + \ldots + \mu_{2d+1} \sum_{|\omega|=2d+1} \chi_\omega(x),$$

*where each* $\mu_{2k} = 0$ *for* $k > 0$*, and the odd coefficients are:*

$$\mu_{2k+1} = \binom{2d}{d}\binom{d}{k}\binom{2d}{2k}^{-1} \frac{(-1)^k}{2^{d+1}}\,.$$

*Proof.* For $\omega \subset [2d+1]$, the Fourier coefficient in front of $\chi_\omega$ in the Fourier decomposition is

$$\widehat{\mathrm{Maj}}(\omega) = \mathop{\mathbb{E}}_x[\chi_\omega(x)\mathrm{Maj}(x)] = \frac{1}{2^{2d+1}} \sum_{w(x)\leq d} \chi_\omega(x).$$

Suppose $i \in \omega$. Then $\chi_\omega(x \odot e_i) = -\chi_\omega(x)$, where $\odot$ denotes coordinate-wise multiplication. So if both $x$ and $x \odot e_i$ have Hamming weight at most $d$, the corresponding terms in the sum will cancel each other out. The terms that remain will correspond to the points $x$ that have Hamming weight $w(x) = d$ with $x_i = 1$ giving

$$\widehat{\mathrm{Maj}}(\omega) = \frac{1}{2^{2d+1}} \sum_{\substack{w(x)=d \\ x_i=0}} \chi_\omega(x). \tag{4}$$

It is clear that this sum is the same for the Fourier coefficients of the same weight. Summing over all the characters $\gamma$ of weight $|\omega|$ which contain coordinate $i$ we get:

$$\begin{aligned}\binom{2d}{|\omega|-1}\widehat{\mathrm{Maj}}(\omega) &= \sum_{\substack{|\gamma|=|\omega| \\ i\in\gamma}} \frac{1}{2^{2d+1}} \sum_{\substack{w(x)=d \\ x_i=0}} \chi_\gamma(x) = \frac{1}{2^{2d+1}} \sum_{\substack{w(x)=d \\ x_i=0}} \sum_{\substack{|\gamma|=|\omega| \\ i\in\gamma}} \chi_\gamma(x) \\ &= \frac{1}{2^{2d+1}}\binom{2d}{d} \sum_{h=0}^{d} (-1)^h \binom{d}{h}\binom{d}{|\omega|-h-1},\end{aligned}$$

where $h$ in the last sum corresponds to the number of $-1$ coordinates in $x$ that fall into $\gamma$. The last sum can be computed by noticing that it is exactly the coefficient in front of $x^{|\omega|-1}$ in the polynomial $(1-x)^d(1+x)^d = (1-x^2)^d$, in particular we get that when $|\omega|$ is even, the coefficient is 0.

The constant Fourier coefficient is the average value of Maj over the hypercube, so it is $1/2$. □