Spectral Refutations of Semirandom k-LIN over Larger Fields

Nicholas Kocurek
University of Washington, Seattle, WA, USA

Peter Manohar
The Institute for Advanced Study, Princeton, NJ, USA

Abstract

We study the problem of strongly refuting semirandom k-$\mathsf{LIN}(\mathbb{F})$ instances: systems of k-sparse inhomogeneous linear equations over a finite field $\mathbb{F}$. For the case of $\mathbb{F} = \mathbb{F}_2$, this is the well-studied problem of refuting semirandom instances of k-XOR, where the works of [18, 20] establish a tight trade-off between runtime and clause density for refutation: for any choice of a parameter ℓ, they give an $n^{O(\ell)}$-time algorithm to certify that there is no assignment that can satisfy more than $\frac{1}{2} + \varepsilon$-fraction of constraints in a semirandom k-XOR instance, provided that the instance has $O(n) \cdot \left(\frac{n}{\ell}\right)^{k/2-1} \log n/\varepsilon^4$ constraints, and the work of [28] provides good evidence that this tight up to a polylog(n) factor via lower bounds for the Sum-of-Squares hierarchy. However, for larger fields, the only known results for this problem are established via black-box reductions to the case of $\mathbb{F}_2$, resulting in a $|\mathbb{F}|^{3k}$ gap between the current best upper and lower bounds.

In this paper, we give an algorithm for refuting semirandom k-$\mathsf{LIN}(\mathbb{F})$ instances with the "correct" dependence on the field size $|\mathbb{F}|$. For any choice of a parameter ℓ, our algorithm runs in $(|\mathbb{F}|n)^{O(\ell)}$-time and strongly refutes semirandom k-$\mathsf{LIN}(\mathbb{F})$ instances with at least $O(n) \cdot \left(\frac{|\mathbb{F}^*|n}{\ell}\right)^{k/2-1} \log(n|\mathbb{F}^*|)/\varepsilon^4$ constraints. We give good evidence that this dependence on the field size $|\mathbb{F}|$ is optimal by proving a lower bound for the Sum-of-Squares hierarchy that matches this threshold up to a polylog$(n|\mathbb{F}^*|)$ factor. Our results also extend beyond finite fields to the more general case of $\mathbb{Z}_m$ and arbitrary finite Abelian groups. Our key technical innovation is a generalization of the "$\mathbb{F}_2$ Kikuchi matrices" of [36, 18] to larger fields, and finite Abelian groups more generally.

2012 ACM Subject Classification Theory of computation → Approximation algorithms analysis

Keywords and phrases Spectral Algorithms, CSP Refutation, Kikuchi Matrices

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2025.17

Category APPROX

Related Version *Full Version*: `https://arxiv.org/abs/2508.18185`

Funding This material is based upon work supported by the National Science Foundation under Grant No. DMS-1926686.

1 Introduction

A k-$\mathsf{LIN}(\mathbb{F})$ instance over a finite field $\mathbb{F}$ is a collection of k-sparse $\mathbb{F}$-linear inhomogeneous equations in n variables $x_1, \ldots, x_n$. That is, each equation has the form $\sum_{i \in I} \alpha_i x_i = b_I$, where $|I| = k$, $\alpha_i \in \mathbb{F} \setminus \{0\}$, and $b_I \in \mathbb{F}$. For worst-case instances, there has been a long line of work on developing algorithms for and proving hardness of determining the value, i.e., the maximum fraction of satisfiable constraints, of a k-$\mathsf{LIN}(\mathbb{F})$ instance, with a special focus on the case of $\mathbb{F} = \mathbb{F}_2$, where the k-$\mathsf{LIN}(\mathbb{F}_2)$ problem is commonly referred to as "k-XOR". For k-$\mathsf{LIN}(\mathbb{F})$ instances with $O(n)$ constraints, Håstad's PCP [19] shows that for $k \geq 3$, it is NP-hard to decide if the instance has value $\leq \frac{1}{|\mathbb{F}|} + \varepsilon$ (a random assignment is near-optimal) or $\geq 1 - \varepsilon$ (nearly satisfiable). On the algorithmic side, the work of [6] gives a PTAS for

© Nicholas Kocurek and Peter Manohar;
licensed under Creative Commons License CC-BY 4.0
Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2025).
Editors: Alina Ene and Eshan Chattopadhyay; Article No. 17; pp. 17:1–17:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

k-XOR instances with $n^{O(k)}$ constraints ("maximally dense" instances). And, assuming the exponential time hypothesis [21], the work of [15] gives an essentially tight "runtime vs. number of constraints" trade-off for worst-case instances. For the case of $k = 2$, the 2-$\mathsf{LIN}(\mathbb{F})$ problem is closely related to the Unique Games Conjecture [25], which conjectures that deciding if a 2-$\mathsf{LIN}(\mathbb{F})$ instance has value $\leq \frac{1}{|\mathbb{F}|} + \varepsilon$ or $\geq 1 - \varepsilon$ is NP-hard when $|\mathbb{F}|$ is sufficiently large as a function of ε.

Despite its worst-case hardness, there have been many works on designing algorithms for *random* k-XOR instances. As random k-XOR instances with $\gg n$ constraints have value $\leq \frac{1}{2} + \varepsilon$ with high probability, the natural problem to consider is the task of (strong) *refutation*, where the algorithmic goal is to output a certificate that the value of the random instance is at most $\frac{1}{2} + \varepsilon$. This problem has been the focus of several works [16, 10, 2, 34], with the ultimate goal being to understand the trade-off between "runtime" and "number of constraints". That is, given n and a "time budget" of $n^{O(\ell)}$ for a parameter ℓ (which may be super-constant, e.g., n^{δ} for constant $\delta > 0$), how many constraints, as a function of n and ℓ, are required to refute a random k-XOR in $n^{O(\ell)}$ time? This question was essentially answered by [34], which gives for any ℓ, an $n^{O(\ell)}$-time algorithm that, with high probability over a random instance, certifies that a random k-XOR instance has maximum value at most $\frac{1}{2} + \varepsilon$ if it has at least $n \cdot (n/\ell)^{k/2-1}\mathsf{poly}(\log(n), \varepsilon^{-1})$ constraints. This trade-off between runtime and number of constraints is conjectured to be essentially optimal, with evidence coming in the form of lower bounds in various restricted computational models [17, 13, 35, 9, 32, 29, 7, 28]. More recently, there has been a flurry of work [1, 18, 20] on designing algorithms for k-XOR in the harder *semirandom* model, where the "left-hand sides" of the equations are worst case, and only the "right-hand sides" b_I are chosen at random. These works show that one can refute semirandom instances at the same runtime vs. number of constraints trade-off as shown in [34]. That is, semirandom instances are just as easy to refute as fully random ones.

Thus, for the Boolean case of $\mathbb{F} = \mathbb{F}_2$, we have a near-complete understanding: for any choice of the parameter ℓ, if the number of constraints in the semirandom k-XOR instance is at least $n \cdot (n/\ell)^{k/2-1}\mathsf{poly}(\log(n), \varepsilon^{-1})$, then the algorithm of [1, 18, 20] can refute the instance in $n^{O(\ell)}$ time, and if the number of constraints is smaller than $n \cdot (n/\ell)^{k/2-1}\mathsf{poly}(\log(n), \varepsilon^{-1})$, the lower bound of, e.g., [28] provides good evidence that there is no algorithm to refute in $n^{O(\ell)}$ time, even for random instances.

What can we say about semirandom k-LIN over finite fields $\mathbb{F} \neq \mathbb{F}_2$? There is a simple reduction to the Boolean case (see Appendix B in [2]) that loses an $|\mathbb{F}|^{3k}$ factor in the number of constraints. That is, the reduction gives an algorithm to refute such instances with at least $|\mathbb{F}|^{3k} \cdot n \cdot (n/\ell)^{k/2-1}\mathsf{poly}(\log(n), \varepsilon^{-1})$ constraints. On the side of lower bounds, the work [28] establishes the same lower bound as in the case of $\mathbb{F}_2$, i.e., with no field-dependent factor. Thus, for e.g., $\mathbb{F}$ with $|\mathbb{F}| = n^{\delta}$, there a large gap between the upper and lower bounds.

Understanding the optimal dependence on the field size for refuting semirandom k-LIN instances has many potential applications. One immediate application is obtaining better attacks on the (sparse) learning parity with noise (LPN) assumption commonly used in cryptography. The k-sparse LPN assumption[1] is the distinguishing variant of the refutation problem for k-$\mathsf{LIN}(\mathbb{F})$ – the "dense" LPN assumption removes the sparsity requirement on the equations – and is considered a foundational assumption in cryptography. While many cryptographic applications only use this assumption over the field $\mathbb{F}_2$ [5], many applications such as constructing indistinguishability obfuscation [22, 23, 33] and others [12, 11] require fields of much larger (even superpolynomial) size.

[1] The phrase "sparse LPN" sometimes refers to the case when the secret is sparse. Here, we mean that the *equations* are sparse.

As another application, one could hope to prove stronger lower bounds for information-theoretic private information retrieval (PIR) schemes. Recent work of [4] has led to a flurry of improvements in lower bounds for *binary* locally decodable [4, 8, 24] and locally correctable codes [26, 37, 3, 27] by establishing a connection between these lower bounds and refuting "semirandom-like" instances of k-LIN over $\mathbb{F}_2$. While these results can be extended to larger, constant-sized alphabets (see, e.g., Appendix A in [4]), information-theoretic PIR schemes are essentially equivalent to locally decodable codes over alphabets of $\mathsf{poly}(n)$ size. Thus, the large loss in $|\mathbb{F}|$ above prevents the approach of [4] from being able to prove stronger PIR lower bounds.

1.1 Our results

In this paper, we investigate the dependence on the field size in the number of constraints required to refute semirandom k-LIN($\mathbb{F}$) instances. As our main results, we give both an algorithm and a matching Sum-of-Squares lower bound with the "correct" polynomial dependence on the field size $|\mathbb{F}|$.

Before stating our main results, we formally define semirandom k-LIN instances.

▶ **Definition 1** ((Semirandom) k-LIN over $\mathbb{F}$)**.** *An instance of k-LIN($\mathbb{F}$) is $\mathcal{I} = (\mathcal{H}, \{b_v\}_{v\in\mathcal{H}})$, where $\mathcal{H}$ is a set of k-sparse vectors*[2] *in $\mathbb{F}^n$ and $b_v \in \mathbb{F}$ for all $v \in \mathcal{H}$. We view $\mathcal{I}$ as representing the system of linear equations with variables $x_1, \ldots, x_n$ specified by $\sum_{i=1}^n v_i x_i = b_v$ for each $v \in \mathcal{H}$. The value of the instance, which we denote by val($\mathcal{I}$), is the maximum over $x \in \mathbb{F}^n$ of the fraction of constraints satisfied by x. That is, $\mathrm{val}(\mathcal{I}) = \max_{x\in\mathbb{F}^n} \frac{1}{|\mathcal{H}|} \sum_{v\in\mathcal{H}} 1(\sum_{i=1}^n v_i x_i = b_v)$.*

An instance of k-LIN is random *if $\mathcal{H}$ is a random subset of k-sparse vectors and each b_v is drawn independently and uniformly from $\mathbb{F}$. An instance of k-LIN is* semirandom *if each b_v is drawn independently and uniformly from $\mathbb{F}$ (but $\mathcal{H}$ may be arbitrary).*

The first main result of this paper gives a refutation algorithm for semirandom k-LIN($\mathbb{F}$) for any field $\mathbb{F}$.

▶ **Theorem 2** (Tight refutation of semirandom k-LIN($\mathbb{F}$))**.** *Fix $\ell \geq k/2$. There is an algorithm that takes as input a k-LIN($\mathbb{F}$) instance $\mathcal{I} = (\mathcal{H}, \{b_v\}_{v\in\mathcal{H}})$ in n variables and outputs a number algval($\mathcal{I}$) $\in [0,1]$ in time $(|\mathbb{F}|n)^{O(\ell)}$ with the following two guarantees:*

(1) *algval($\mathcal{I}$) $\geq$ val($\mathcal{I}$) for every instance $\mathcal{I}$;*

(2) *If $|\mathcal{H}| \geq \Omega(n) \cdot \left(\frac{n|\mathbb{F}^*|}{\ell}\right)^{k/2-1} \cdot \log(|\mathbb{F}^*|n) \cdot \varepsilon^{-4}$ and $\mathcal{I}$ is drawn from the semirandom distribution described in Definition 1, then with probability $\geq 1 - \frac{1}{\mathsf{poly}(n)}$ over the draw of the semirandom instance, i.e., the randomness of $\{b_v\}_{v\in\mathcal{H}}$, it holds that algval($\mathcal{I}$) $\leq \frac{1}{|\mathbb{F}|} + \varepsilon$.*

As a byproduct of the analysis of Theorem 2, we also establish an extremal combinatorics statement on the existence of short linear dependencies in any sufficiently dense collection of k-sparse vectors $\mathcal{H}$ over a finite field $\mathbb{F}$.

▶ **Theorem 3** (Short linear dependencies in k-sparse vectors over $\mathbb{F}$)**.** *Let $\mathcal{H}$ be a set of k-sparse vectors in $\mathbb{F}^n$ with $|\mathcal{H}| \geq \Omega(n) \cdot \left(\frac{n|\mathbb{F}^*|}{\ell}\right)^{k/2-1} \cdot \log(|\mathbb{F}^*|n)$. Then, there exists a set $\mathcal{V} \subseteq \mathcal{H}$ with $|\mathcal{V}| \leq \ell \log(|\mathbb{F}^*|n)$ and non-zero coefficients $\{\alpha_v\}_{v\in\mathcal{V}}$ in $\mathbb{F}^*$ such that:*

$$\sum_{v\in\mathcal{V}} \alpha_v \cdot v = 0\,.$$

That is, $\mathcal{V}$ is a linearly dependent subset of $\mathcal{H}$.

[2] A vector $v \in \mathbb{F}^n$ is k-sparse if $|\{i : v_i \neq 0\}| = k$.

Theorem 3 is a generalization of the hypergraph Moore bound, or Feige's conjecture on the existence of short even covers in hypergraphs [14] (first proven in [18] for the case of $\mathbb{F}_2$) to arbitrary finite fields. The hypergraph Moore bound establishes a rate vs. distance trade-off for binary LDPC codes (see [30]). One can similarly view Theorem 3 as establishing such a trade-off for LDPC codes over general finite fields.

The key technical innovation in our proofs of Theorems 2 and 3 is the introduction of a new Kikuchi matrix for any finite field $\mathbb{F}$ (Definition 10). Our Kikuchi matrices are a generalization of the "$\mathbb{F}_2$ Kikuchi matrices" of [36, 18] to other fields, and finite Abelian groups more generally. As we point out after Definition 10, the natural generalization of the $\mathbb{F}_2$ Kikuchi matrix is not sufficient, and we have to add an additional condition to the matrix to make the analysis work (see Remark 11).

In our second main result, we prove a Sum-of-Squares lower bound for refuting k-$\mathsf{LIN}(\mathbb{F})$ instances that nearly matches the threshold in Theorem 2.

▶ **Theorem 4** (Sum-of-Squares lower bounds for refuting random k-LIN, informal). *Fix $k \geq 3$ and $\frac{n}{\max(|\mathbb{F}^*|,k)} \geq \ell \geq k$. Let $\mathcal{I}$ be a random k-LIN($\mathbb{F}$) instance $|\mathcal{H}| \leq O(n) \cdot \left(\frac{n|\mathbb{F}^*|}{\ell}\right)^{k/2-1} \cdot \varepsilon^{-2}$. Then, with large probability over the draw of $\mathcal{I}$, it holds that*

(1) *$val(\mathcal{I}) \leq \frac{1}{|\mathbb{F}|} + \varepsilon$;*

(2) *The degree-$\tilde{O}(\ell)$ Sum-of-Squares relaxation for k-LIN($\mathbb{F}$) fails to refute $\mathcal{I}$.*

We note that Theorem 4 requires $k \geq 3$ as Theorem 2 gives a polynomial-time algorithm when $|\mathcal{H}| = O(n)$ and $k = 2$.

The threshold in Theorem 4 matches the threshold in Theorem 2 up to the "lower order" $\mathsf{poly}(\log n, \varepsilon^{-1})$ factor. However, it has a one limitation: the degree ℓ of the Sum-of-Squares relaxation can only be at most $n/|\mathbb{F}^*|$, rather than the whole of range of $O(1)$ to $\Omega(n)$. That is, Theorem 4 can only give a subexponential-time lower bound of $2^{n^{1-\delta}}$ for the Sum-of-Squares algorithm when $|\mathbb{F}| = n^\delta$. It turns out that this is nearly necessary, as there is a very simple refutation algorithm, implementable within degree $\tilde{O}(\ell)$ Sum-of-Squares, that out-performs the algorithm in Theorem 2 when $|\mathbb{F}|$ is very large.

▶ **Theorem 5** (Simple refutation algorithm). *Fix $n/2 \geq \ell \geq k$. There is an algorithm that takes as input a k-LIN($\mathbb{F}$) instance $\mathcal{I} = (\mathcal{H}, \{b_v\}_{v \in \mathcal{H}})$ in n variables and outputs a number $algval(\mathcal{I}) \in [0,1]$ in time $(|\mathbb{F}|n)^{O(\ell)}$ with the following two guarantees:*

(1) *$algval(\mathcal{I}) \geq val(\mathcal{I})$ for every instance $\mathcal{I}$;*

(2) *If $|\mathcal{H}| \geq \Omega(n) \cdot \left(\frac{n}{\ell}\right)^{k-1} \cdot \log n \cdot \varepsilon^{-2}$ and $\mathcal{I}$ is drawn from the fully random distribution described in Definition 1, then with probability $\geq 1 - \frac{1}{\mathsf{poly}(n)}$ over the draw of the random instance, it holds that $algval(\mathcal{I}) \leq \frac{1}{|\mathbb{F}|} + \varepsilon$;*

(3) *If $|\mathcal{H}| \geq \Omega(n) \cdot \left(\frac{n}{\ell}\right)^{k-1} \cdot \log n \cdot \varepsilon^{-3}$ and $\mathcal{I}$ is drawn from the semirandom distribution described in Definition 1, then with probability $\geq 1 - \frac{1}{\mathsf{poly}(n)}$ over the draw of the semirandom instance, i.e., the randomness of $\{b_v\}_{v \in \mathcal{H}}$, it holds that $algval(\mathcal{I}) \leq \frac{1}{|\mathbb{F}|} + \varepsilon$.*

Theorem 5 is nearly identical to Theorem 2: the main difference is that, ignoring the lower order $\mathsf{poly}(\log n, \varepsilon^{-1})$ term, the "constraint threshold" is now $n \cdot (n/\ell)^{k-1}$ instead of $n \cdot (n|\mathbb{F}^*|/\ell)^{k/2-1}$, and this is smaller when $\ell \geq n/|\mathbb{F}^*|^{1-2/k}$. As this algorithm is "captured" by the Sum-of-Squares hierarchy, the Sum-of-Squares lower bound in Theorem 4 is false when $\ell \geq n/|\mathbb{F}^*|^{1-2/k}$. Thus, the range of ℓ where the lower bound in Theorem 4 holds is nearly tight: we show it holds for $\ell \leq n/|\mathbb{F}^*|$, and it is false for $\ell \geq n/|\mathbb{F}^*|^{1-2/k}$.

We also extend our refutation result for k-$\mathsf{LIN}(\mathbb{F})$ to k-$\mathsf{LIN}(\mathbb{Z}_m)$ for composite m, and more generally to any finite Abelian group G. Below, we define the natural extension of Definition 1 to Abelian groups, and then state our generalization of Theorem 2 to this setting. Recall that by the fundamental theorem of finite Abelian groups, any finite Abelian group G is isomorphic to $\bigotimes_{i=1}^{r} \mathbb{Z}_{m_i}$ for some $m_1, ..., m_r \in \mathbb{N}$.

▶ **Definition 6** ((Semirandom) k-LIN over an Abelian group G)**.** *Let $G = \bigotimes_{i=1}^{r} \mathbb{Z}_{m_i}$ for some $m_1, ..., m_r \in \mathbb{N}$. An instance of k-LIN(G) is $\mathcal{I} = (\mathcal{H}, \{b_v\}_{v \in \mathcal{H}})$, where $\mathcal{H}$ is a set of k-sparse vectors in G^n and $b_v \in G$ for all $v \in \mathcal{H}$. We view $\mathcal{I}$ as representing the system of linear equations with variables $x_1, \ldots, x_n$ specified by $\langle v, x \rangle = b_v$ for each $v \in \mathcal{H}$. Note that the inner product notation here represents $\langle v, x \rangle = \sum_{i=1}^{n} v_i \cdot x_i$ where the multiplication is direct product multiplication over each $\mathbb{Z}_{m_i}$. The value of the instance, which we denote by val$(\mathcal{I})$, is the maximum over $x \in \mathbb{F}^n$ of the fraction of constraints satisfied by x. That is, val$(\mathcal{I}) = \max_{x \in G^n} \frac{1}{|\mathcal{H}|} \sum_{v \in \mathcal{H}} 1(\langle v, x \rangle = b_v)$.*

An instance of k-LIN is random *if $\mathcal{H}$ is a random subset of k-sparse vectors and each b_v is drawn independently and uniformly from G.*

An instance of k-LIN is semirandom *if each b_v is drawn independently and uniformly from G (but $\mathcal{H}$ may be arbitrary).*

Definition 6 allows each equation to have coefficients on the variables, which was natural in the case of finite fields (Definition 1), but may appear strange in the case of a general finite Abelian group, where it is perhaps more natural to only have coefficients that are 1. However, because we are working with semirandom instances, the "left-hand sides" of the equations are arbitrary, and so semirandom instances "capture" the special case where the coefficients are all 1. We choose this perhaps nonstandard definition because it is more general; it seamlessly captures both the case of a finite field $\mathbb{F}$ or a ring $\mathbb{Z}_m$, where coefficients are natural, and also a finite Abelian group.

Our final result generalizes Theorem 2 to the case of k-LIN(G).

▶ **Theorem 7** (Tight refutation of semirandom k-LIN(G))**.** *Fix $\ell \geq k/2$. There is an algorithm that takes as input a k-LIN(G) instance $\mathcal{I} = (\mathcal{H}, \{b_v\}_{v \in \mathcal{H}}$ in n variables and outputs a number algval$(\mathcal{I}) \in [0, 1]$ in time $(|G|n)^{O(\ell)}$ with the following two guarantees:*

(1) *algval$(\mathcal{I}) \geq$ val$(\mathcal{I})$ for every instance $\mathcal{I}$;*

(2) *If $|\mathcal{H}| \geq \Omega(n) \cdot \left(\frac{n|G|}{\ell}\right)^{k/2-1} \cdot \log(|G|n) \cdot \varepsilon^{-5}$ and $\mathcal{I}$ is drawn from the semirandom distribution described in Definition 6, then with probability $\geq 1 - \frac{1}{\mathsf{poly}(n)}$ over the draw of the semirandom instance, i.e., the randomness of $\{b_v\}_{v \in \mathcal{H}}$, it holds that algval$(\mathcal{I}) \leq \frac{1}{|G|} + \varepsilon$.*

The proof of Theorem 7 encounters additional technical difficulties compared to Theorem 2, arising from zero divisors in $\mathbb{Z}_m$ for composite m. Roughly, this allows a semirandom instance to embed equations within a subgroup of $\mathbb{Z}_m$, by, e.g., choosing only equations where the coefficients are divisible by some integer $d \geq 2$, and handling this issue requires an additional step and an extra factor ε^{-1}.

2 Preliminaries

2.1 Basic notation

We let $[n]$ denote the set $\{1, \ldots, n\}$. For two subsets $S, T \subseteq [n]$, we let $S \oplus T$ denote the symmetric difference of S and T, i.e., $S \oplus T := \{i : (i \in S \wedge i \notin T) \vee (i \notin S \wedge i \in T)\}$. For a natural number $t \in \mathbb{N}$, we let $\binom{[n]}{t}$ be the collection of subsets of $[n]$ of size exactly t.

For a rectangular matrix $A \in \mathbb{C}^{m \times n}$, we let $\|A\|_2 := \max_{x \in \mathbb{C}^m, y \in \mathbb{C}^n : \|x\|_2 = \|y\|_2 = 1} x^\dagger A y$ denote the spectral norm of A.

For a vector $v \in \mathbb{F}^n$, we let $\mathrm{supp}(v) := \{i : v_i \neq 0\}$ and $\mathrm{wt}(v) := |\mathrm{supp}(v)|$. For a field $\mathbb{F}$ with $\mathrm{char}(\mathbb{F}) = p$, we let $\mathrm{Tr}(\cdot)$ denote the trace map of $\mathbb{F}$ over $\mathbb{F}_p$.

For a matrix $A \in \mathbb{C}^{n\times n}$, we let $\mathrm{tr}(A)$ be the trace of A, i.e., $\sum_{i=1}^{n} A_{i,i}$. This should not be confused with the trace map for field elements, which we denote by $\mathrm{Tr}(\cdot)$. For two vectors $x, y \in \mathbb{C}^n$ we define the following inner product:

$$\langle x, y\rangle = x^\dagger y = \sum_{i=1}^{n} \overline{x_i} \cdot y_i \,.$$

2.2 Fourier analysis

Let G be an Abelian group isomorphic to $\mathbb{Z}_{m_1} \times ... \times \mathbb{Z}_{m_r}$ via the isomorphism ψ. For $m \in \mathbb{N}$, we let $\omega_m := e^{\frac{2\pi i}{m}}$. For $\alpha, x \in G$, we define

$$\chi_\alpha(x) = \prod_{i=1}^{r} \omega_{m_i}^{\psi(\alpha)_i \psi(x)_i} \,.$$

These functions form a Fourier basis for G, as shown in [31]. This extends to a Fourier basis for G^n as follows. For $v, x \in G^n$, we define

$$\chi_v(x) = \prod_{i=1}^{n} \chi_{v_i}(x_i) \,.$$

For a function $f\colon G^n \to \mathbb{C}$, we have that for each $x \in G^n$,

$$f(x) = \sum_{v\in G^n} \hat{f}(v) \cdot \chi_v(x) \,,$$

where $\hat{f}(v) = \mathbb{E}_{x\in G^n}\left[f(x) \cdot \overline{\chi_v(x)}\right]$.

For the special case of functions $f\colon \mathbb{F}^n \to \mathbb{C}$ with $\mathrm{char}(\mathbb{F}) = p$, we note that the standard Fourier basis is simply

$$\chi_v(x) = \omega_p^{\mathrm{Tr}(\langle v,x\rangle)} \,.$$

2.3 Binomial coefficient inequalities

In this section, we state and prove the following fact about binomial coefficients that we will use.

▶ **Proposition 8.** *Let n, ℓ, q be positive integers with $\ell \le n$. Let q be constant and ℓ, n be asymptotically large with $\ell \le n/2$. Then,*

$$\frac{\binom{n}{\ell-q}}{\binom{n}{\ell}} = \Theta\left(\left(\frac{\ell}{n}\right)^q\right) \,,$$

$$\frac{\binom{n-q}{\ell}}{\binom{n}{\ell}} = \Theta(1) \,.$$

Proof. We have that

$$\frac{\binom{n}{\ell-q}}{\binom{n}{\ell}} = \frac{\binom{\ell}{q}}{\binom{n-\ell+q}{q}} \,.$$

Using that $\left(\frac{a}{b}\right)^b \le \binom{a}{b} \le \left(\frac{ea}{b}\right)^b$ finishes the proof of the first equation.

We also have that

$$\frac{\binom{n-q}{\ell}}{\binom{n}{\ell}} = \frac{(n-q)!(n-\ell)!}{n!(n-\ell-q)!} = \prod_{i=0}^{q-1} \frac{n-\ell-i}{n-i} = \prod_{i=0}^{q-1}\left(1 - \frac{\ell}{n-i}\right) \,,$$

and this is $\Theta(1)$ since $\ell \le n/2$ and q is constant. ◀

3 Spectral Algorithms for Refuting Semirandom k-LIN($\mathbb{F}$) for Even k

As our proof overview, we will give a complete proof of Theorem 2 in the case when k is even. As in [18, 20], the proof is substantially simpler in the case of even k. We will assume familiarity with the notation and conventions defined in Section 2.

Our refutation algorithm for semirandom k-LIN roughly follows the framework established in [18, 20]. The main technical tool we use is a generalization of the Kikuchi matrix of [36] for $\mathbb{F}_2$ to arbitrary finite fields $\mathbb{F}$. Analyzing the spectral norm of this matrix requires a more complicated trace moment calculation as compared to the case of $\mathbb{F}_2$, and requires a careful choice of the Kikuchi matrix (see Remark 11).

We let $\mathrm{char}(\mathbb{F}) = p$ and $\omega_p = e^{2\pi i/p}$ denote a primitive p-th root of unity in $\mathbb{C}$.

3.1 Step 1: Expressing a k-LIN($\mathbb{F}$) instance as a polynomial in $\mathbb{C}$

As the first step in the proof, we make the following observation, which shows that we can express the fraction of constraints satisfied by an assignment $x \in \mathbb{F}^n$ as a polynomial in n variables in $\mathbb{C}$.

▶ **Observation 9.** *For a k-LIN($\mathbb{F}$) instance $\mathcal{I} = (\mathcal{H}, \{b_v\}_{v\in\mathcal{H}})$, let $val(\mathcal{I}, x)$ denote the fraction of constraints satisfied by an assignment $x \in \mathbb{F}^n$. Then, we can express $val(\mathcal{I}, x)$ as a polynomial in $\mathbb{C}$. That is,*

$$val(\mathcal{I}, x) = \frac{1}{|\mathbb{F}|} + \frac{1}{|\mathcal{H}||\mathbb{F}|} \sum_{v\in\mathcal{H}} \sum_{\beta\in\mathbb{F}^*} \omega_p^{\mathrm{Tr}(\beta b_v)} \cdot \overline{\chi_{\beta v}(x)} := \frac{1}{|\mathbb{F}|} + \Phi(x)\,.$$

Proof. Recall that a constraint in $\mathcal{I}$ takes the form $\langle v, x\rangle = b_v$ for $v \in \mathcal{H}$, where $x \in \mathbb{F}^n$ are the variables. The indicator variable for this event is simply:

$$1(\langle v, x\rangle = b_v) = \mathbb{E}_{\beta\sim\mathbb{F}}\left[\omega_p^{\mathrm{Tr}(\beta b_v - \beta\langle v,x\rangle)}\right] = \frac{1}{|\mathbb{F}|}\sum_{\beta\in\mathbb{F}} \omega_p^{\mathrm{Tr}(\beta b_v)} \cdot \overline{\chi_{\beta v}(x)}\,.$$

where $p = \mathrm{char}(\mathbb{F})$. Indeed, if $\langle v, x\rangle = b_v$, then $\mathrm{Tr}(\beta b_v - \beta\langle v, x\rangle) = 0$ for all $\beta \in \mathbb{F}$. If $b_v - \langle v, x\rangle \neq 0$, i.e., it is some $\alpha \in \mathbb{F}^*$, then $\mathbb{E}_{\beta\sim\mathbb{F}}\left[\omega_p^{\mathrm{Tr}(\beta\alpha)}\right] = \mathbb{E}_{\beta\sim\mathbb{F}}\left[\omega_p^{\mathrm{Tr}(\beta)}\right] = 0$. Hence, it follows that

$$\begin{aligned}\mathrm{val}(\mathcal{I}, x) &= \frac{1}{|\mathcal{H}|}\sum_{v\in\mathcal{H}} 1(\langle v, x\rangle = b_v) = \frac{1}{|\mathcal{H}|}\sum_{v\in\mathcal{H}}\frac{1}{|\mathbb{F}|}\sum_{\beta\in\mathbb{F}} \omega_p^{\mathrm{Tr}(\beta b_v)} \cdot \overline{\chi_{\beta v}(x)} \\ &= \frac{1}{|\mathbb{F}|} + \frac{1}{|\mathcal{H}||\mathbb{F}|}\sum_{v\in\mathcal{H}}\sum_{\beta\in\mathbb{F}^*} \omega_p^{\mathrm{Tr}(\beta b_v)} \cdot \overline{\chi_{\beta v}(x)}\,,\end{aligned}$$

which finishes the proof. ◀

3.2 Step 2: Expressing $\Phi(x)$ as a quadratic form on a Kikuchi matrix

In light of Observation 9, it thus remains to find a certificate that bounds $\max_{x\in\mathbb{F}^n} \Phi(x)$. We do this by generalizing the analysis of [18] and constructing a Kikuchi matrix whose spectral norm provides a certificate bounding the maximum value of Φ.

▶ **Definition 10.** *(Even-arity Kikuchi matrix over $\mathbb{F}$). Let $k/2 \le \ell \le n/2$ be a parameter,*[3] *and let $N = |\mathbb{F}^*|^\ell \binom{n}{\ell}$. For each k-sparse vector $v \in \mathbb{F}^n$ and $\beta \in \mathbb{F}^*$, we define a matrix $A_{v,\beta} \in \mathbb{C}^{N\times N}$ as follows. First, we identify N with the set of ℓ-sparse vectors in $\mathbb{F}^n$. Then, for ℓ-sparse vectors $U, V \in \mathbb{F}^n$, we let*

$$A_{v,\beta}(U,V) = \begin{cases} 1 & U \xrightarrow{v,\beta} V \\ 0 & \textit{otherwise} \end{cases}$$

where we say $U \xrightarrow{v,\beta} V$ if $U - V = \beta v$ and $\mathrm{supp}(U) \oplus \mathrm{supp}(V) = \mathrm{supp}(v)$.

Let $\Phi(x) = \frac{1}{|\mathbb{F}||\mathcal{H}|}\sum_{v\in\mathcal{H}}\sum_{\beta\in\mathbb{F}^*} c_{v,\beta}\cdot\chi_{\beta v}$ be a polynomial defined by a set $\mathcal{H}$ of k-sparse vectors from $\mathbb{F}^n$ and complex coefficients $\{c_{v,\beta}\}_{\substack{v\in\mathcal{H}\\ \beta\in\mathbb{F}^*}}$. We define the level-$\ell$ Kikuchi matrix for this polynomial to be $A = \sum_{v\in\mathcal{H}}\sum_{\beta\in\mathbb{F}^*} c_{v,\beta}\cdot A_{v,\beta}$. We refer to the graph (with complex weights) defined by the underlying adjacency matrix as the Kikuchi graph.

▶ Remark 11. Our Kikuchi matrix in Definition 10 has an additional condition that $\mathrm{supp}(U)\oplus\mathrm{supp}(V) = \mathrm{supp}(v)$. A perhaps more natural generalization of the $\mathbb{F}_2$ Kikuchi matrix of [36, 18] would be the matrix where this condition is removed, i.e., we only require that $U - V = \beta v$. As we shall see, when we do the trace moment calculation at the end of Section 3.3, it is crucial that for any U and v, there is at most one $\beta \in \mathbb{F}^*$ such that $U \xrightarrow{v,\beta} V$ for some V. That is, the number of edges adjacent to U "coming from" a constraint v is at most 1. If this uniqueness of β did not hold, we would lose an additional factor of $\mathbb{F}^*$ in the number of constraints that we require in Theorem 2. This is a substantial increase, as e.g., this would increase the number of constraints when $k = 2$ to $\sim n|\mathbb{F}^*|$ when the correct dependence is $\sim n$. The condition that $\mathrm{supp}(U) \oplus \mathrm{supp}(V) = \mathrm{supp}(v)$ implies uniqueness of β above, and without this condition we could have a U with $\mathrm{supp}(v) \subseteq U$, in which case $U - \alpha v = V$ for an ℓ-sparse vector V in $\mathbb{F}^n$ for all $\alpha \in \mathbb{F}$.

▶ Remark 12. We note that in Definition 10, we have $A_{v,\beta} = A_{\beta v,1}$. The reason we use the above definition with two parameters v and β is that it will be more convenient when counting walks in the matrix A, as it makes explicit the choice of v and β. Note that in $\mathcal{H}$, there could exist v and v' with $\beta v = v'$ for some $\beta \in \mathbb{F}^*$, and we need to count these terms separately.

▶ **Observation 13.** *The Kikuchi matrix A is always Hermitian.*

Proof. To see this note that $U - V = \beta v \iff V - U = -\beta v$, $\overline{\chi_\beta} = \chi_{-\beta}$, and $\oplus$ is commutative. ◀

The following observation shows that we can express $\Phi(x)$ as a quadratic form on the matrix A defined in Definition 10. Thus, $\|A\|_2$ bounds $\max_{x\in\mathbb{F}^n}\Phi(x)$.

▶ **Observation 14.** *For $x \in \mathbb{F}^n$ define $y \in \mathbb{C}^N$ as follows. For each ℓ-sparse $U \in \mathbb{F}^n$, we set $y_U = \chi_U(x)$. Then*

$$\Phi(x) = \frac{1}{|\mathcal{H}||\mathbb{F}|\Delta} y^\dagger A y\,,$$

where $\Delta := \binom{k}{k/2}\binom{n-k}{\ell-k/2}|\mathbb{F}^*|^{\ell-k/2}$.

[3] Note that it suffices to prove Theorem 2 for ℓ in this range.

Proof.

$$\begin{aligned}
y^\dagger A y &= \sum_{\substack{U,V\in\mathbb{F}^n\\ \mathrm{wt}(U)=\mathrm{wt}(V)=\ell}} A(U,V)\cdot\overline{\chi_U(x)}\cdot\chi_V(x)\\
&= \sum_{\substack{U,V\in\mathbb{F}^n\\ \mathrm{wt}(U)=\mathrm{wt}(U)=\ell}} \sum_{v\in\mathcal{H},\beta\in\mathbb{F}^*} 1\left(U\xrightarrow{v,\beta}V\right)\cdot c_{v,\beta}\cdot\overline{\chi_U(x)}\cdot\chi_V(x)\\
&= \sum_{\substack{U,V\in\mathbb{F}^n\\ \mathrm{wt}(U)=\mathrm{wt}(U)=\ell}} \sum_{v\in\mathcal{H},\beta\in\mathbb{F}^*} 1\left(U\xrightarrow{v,\beta}V\right)\cdot c_{v,\beta}\cdot\overline{\chi_{U-V}(x)}\\
&= \sum_{\substack{U,V\in\mathbb{F}^n\\ \mathrm{wt}(U)=\mathrm{wt}(U)=\ell}} \sum_{v\in\mathcal{H},\beta\in\mathbb{F}^*} 1\left(U\xrightarrow{v,\beta}V\right)\cdot c_{v,\beta}\cdot\overline{\chi_{\beta v}(x)}\,.
\end{aligned}$$

For each $v\in\mathcal{H}$ and $\beta\in\mathbb{F}^*$, the term $c_{v,\beta}\cdot\overline{\chi_{\beta v}(x)}$ appears once for each pair of vertices (U,V) with $U\xrightarrow{v,\beta}V$. Let us now argue that the number of such pairs (U,V) is exactly $\Delta=\binom{k}{k/2}\binom{n-k}{\ell-k/2}|\mathbb{F}^*|^{\ell-k/2}$. We will count the number of pairs (U,V) by first specifying $\mathrm{supp}(U)$ and $\mathrm{supp}(V)$, and then by specifying U_i for each $i\in\mathrm{supp}(U)$ (and same for V). We first require that $\mathrm{supp}(U)\oplus\mathrm{supp}(V)=\mathrm{supp}(v)$, which in turn means that $\mathrm{supp}(U)$ has intersection exactly $k/2$ with $\mathrm{supp}(v)$ and likewise for $\mathrm{supp}(V)$. Thus, we can pay $\binom{k}{k/2}$ to count the number of ways to split $\mathrm{supp}(v)$ into two equal parts. Second, we need to specify $\mathrm{supp}(U)\setminus\mathrm{supp}(v)$, which is equal to $\mathrm{supp}(V)\setminus\mathrm{supp}(v)$, which is $\binom{n-k}{\ell-k/2}$ choices. Finally, we need to specify U_i for each $i\in\mathrm{supp}(U)$ and V_i for each $i\in\mathrm{supp}(V)$. For each $i\in\mathrm{supp}(U)\cap\mathrm{supp}(v)$, we set $U_i=(\beta v)_i$, and for each $i\in\mathrm{supp}(U)\setminus\mathrm{supp}(v)$, we can set U_i to be any element in $\mathbb{F}^*$. Note that specifying U then determines V, so we have $|\mathbb{F}^*|^{\ell-k/2}$ choices. This finishes the proof. ◀

Next, we compute the average degree (or number of non-zero entries) in a row/column in A.

▶ **Observation 15.** *For $U\in\mathbb{F}^n$ with $\mathrm{wt}(U)=\ell$, we define the graph degree as normal:*

$$\deg(U):=\left|\{\beta v\mid\beta\in\mathbb{F}^*,v\in\mathcal{H}\ \textit{s.t.}\ \exists V\in\mathbb{F}^n,\mathrm{wt}(V)=\ell,U\xrightarrow{v,\beta}V\}\right|.$$

Then $\mathbb{E}[\deg(U)]\geq\frac{|\mathbb{F}^*|}{2}\left(\frac{\ell}{|\mathbb{F}^*|n}\right)^{k/2}\cdot|\mathcal{H}|$.

Proof. Each $v\in\mathcal{H}$ contributes $|\mathbb{F}^*|\Delta$ to the total degree, so the average degree is $\mathbb{E}[\deg(U)]=\frac{|\mathcal{H}||\mathbb{F}^*|\Delta}{N}$. We then have:

$$\mathbb{E}[\deg(U)]=\frac{|\mathbb{F}^*|\Delta}{N}\cdot|\mathcal{H}|=\frac{|\mathbb{F}^*|^{\ell-k/2+1}\binom{k}{k/2}\binom{n-k}{\ell-k/2}}{|\mathbb{F}^*|^\ell\binom{n}{\ell}}\cdot|\mathcal{H}|\geq\frac{|\mathbb{F}^*|}{2}\left(\frac{\ell}{|\mathbb{F}^*|n}\right)^{k/2}\cdot|\mathcal{H}|\,,$$

where the last inequality follows from Proposition 8. ◀

3.3 Step 3: Bounding the spectral norm of A via the trace moment method

The following spectral norm bound now implies Theorem 2.

▶ **Lemma 16.** *Let A be the level-ℓ Kikuchi matrix over $\mathbb{F}^n$ defined in Definition 10 for the k-LIN instance $\mathcal{I} = (\mathcal{H}, \{b_v\}_{v\in\mathcal{H}})$. Let $\Gamma \in \mathbb{C}^{N\times N}$ be the diagonal matrix $\Gamma = D + d\mathbb{I}$ where $D_{U,U} := \deg(U)$ and $d = \mathbb{E}[\deg(U)]$. Suppose that the b_v's are drawn independently and uniformly from $\mathbb{F}$, i.e., the instance $\mathcal{I}$ is* semirandom *(Definition 1). Then, with probability $\geq 1 - \frac{1}{\mathsf{poly}(n)}$, it holds that*

$$\|\Gamma^{-1/2}A\Gamma^{-1/2}\|_2 \leq O\left(\sqrt{\frac{\ell\log(|\mathbb{F}^*|n)}{d}}\right).$$

We postpone the proof of Lemma 16 to the end of this section, and now finish the proof of Theorem 2.

Proof of Theorem 2 from Lemma 16. Let $\mathcal{I} = (\mathcal{H}, \{b_v\}_{v\in\mathcal{H}})$ be the input to the algorithm. Given ℓ, the algorithm constructs the matrix A and computes $\mathrm{algval}(\mathcal{I}) = \frac{1}{|\mathbb{F}|} + \frac{2|\mathbb{F}^*|}{|\mathbb{F}|}\|\tilde{A}\|_2$, where $\tilde{A} = \Gamma^{-1/2}A\Gamma^{-1/2}$. It remains to argue that this quantity has the desired properties.

Let $\Phi(x)$ be the polynomial defined in Observation 9. For each $x \in \mathbb{F}^n$, letting $y \in \mathbb{C}^n$ be the vector defined in Observation 14, we have

$$\begin{aligned}\Phi(x) &= \frac{1}{|\mathbb{F}||\mathcal{H}|\Delta}\cdot y^\dagger Ay = \frac{1}{|\mathbb{F}||\mathcal{H}|\Delta}\cdot(\Gamma^{1/2}y)^\dagger\tilde{A}(\Gamma^{1/2}y) \leq \frac{1}{|\mathbb{F}||\mathcal{H}|\Delta}\cdot\|\tilde{A}\|_2\|\Gamma^{1/2}y\|_2^2\\ &= \frac{1}{|\mathbb{F}||\mathcal{H}|\Delta}\cdot\|\tilde{A}\|_2\cdot\mathrm{tr}(\Gamma) = \frac{2|\mathbb{F}^*|}{|\mathbb{F}|}\|\tilde{A}\|_2,\end{aligned}$$

where we use that $\|\Gamma^{1/2}y\|_2^2 = y^\dagger\Gamma y = \sum_U \Gamma_U|y_U|^2 = \sum_U \Gamma_U = \mathrm{tr}(\Gamma)$ since $|y_U| = 1$ for all U, and that $\mathrm{tr}(\Gamma) = 2|\mathcal{H}||\mathbb{F}^*|\Delta$. Hence,

$$\mathrm{val}(\mathcal{I}) = \frac{1}{|\mathbb{F}|} + \max_{x\in\mathbb{F}^n}\Phi(x) \leq \frac{1}{|\mathbb{F}|} + \frac{2|\mathbb{F}^*|}{|\mathbb{F}|}\|\tilde{A}\|_2,$$

which proves Item 1 in Theorem 2.

To prove Item 2, we observe that by Lemma 16, if $\mathcal{I}$ is semirandom, then with high probability over the draw of the b_v's, it holds that

$$\|\tilde{A}\|_2 \leq O\left(\sqrt{\frac{\ell\log(|\mathbb{F}^*|n)}{d}}\right).$$

From Observation 15, we have $d \geq \frac{|\mathbb{F}^*|}{2}\left(\frac{\ell}{|\mathbb{F}^*|n}\right)^{k/2}\cdot|\mathcal{H}|$. Hence, if we have that $|\mathcal{H}| \geq Cn\log(|\mathbb{F}^*|n)\left(\frac{|\mathbb{F}^*|n}{\ell}\right)^{k/2-1}\varepsilon^{-2}$ for a sufficiently large constant C, then $\|\tilde{A}\|_2 \leq \varepsilon$ with probability $1 - 1/\mathsf{poly}(n)$. This proves Item 2. ◀

It remains to prove Lemma 16, which we do using the trace moment method.

Proof of Lemma 16. By Observation 13, we have that $\|\tilde{A}\|_2 \leq \mathrm{tr}((\Gamma^{-1}A)^{2t})^{1/2t}$ for any positive integer t. Because the b_v's are drawn independently from $\mathbb{F}$, the matrix $\tilde{A}$ is a random matrix. By Markov's inequality,

$$\Pr\left[\mathrm{tr}((\Gamma^{-1}A)^{2t}) \geq N\cdot\mathbb{E}[\mathrm{tr}((\Gamma^{-1}A)^{2t})]\right] \leq \frac{1}{N}.$$

We note this event is the same as $\mathrm{tr}((\Gamma^{-1}A)^{2t})^{1/2t} \geq N^{1/2t}\cdot\mathbb{E}[\mathrm{tr}((\Gamma^{-1}A)^{2t})]^{1/2t}$, and for $2t \geq \log N$ we have $N^{1/2t} \leq O(1)$. This immediately gives us that with probability $\geq 1 - \frac{1}{N}$, $\|\tilde{A}\|_2 \leq O\left(\mathbb{E}[\mathrm{tr}((\Gamma^{-1}A)^{2t})]^{1/2t}\right)$. We then have that

$$\begin{aligned}
\mathbb{E}\left[\operatorname{tr}\left(\left(\Gamma^{-1}A\right)^{2t}\right)\right] &= \mathbb{E}\left[\operatorname{tr}\left(\left(\Gamma^{-1}\sum_{v\in\mathcal{H},\beta\in\mathbb{F}^*} c_{v,\beta}\cdot A_{v,\beta}\right)^{2t}\right)\right] \\
&= \mathbb{E}\left[\operatorname{tr}\left(\sum_{(v_1,\beta_1),\ldots,(v_{2t},\beta_{2t})\in\mathcal{H}\times\mathbb{F}^*}\prod_{i=1}^{2t}\Gamma^{-1}\cdot c_{v_i,\beta_i}\cdot A_{v_i,\beta_i}\right)\right] \\
&= \sum_{(v_1,\beta_1),\ldots,(v_{2t},\beta_{2t})\in\mathcal{H}\times\mathbb{F}^*}\mathbb{E}\left[\operatorname{tr}\left(\prod_{i=1}^{2t}\Gamma^{-1}\cdot c_{v_i,\beta_i}\cdot A_{v_i,\beta_i}\right)\right] \\
&= \sum_{(v_1,\beta_1),\ldots,(v_{2t},\beta_{2t})\in\mathcal{H}\times\mathbb{F}^*}\mathbb{E}\left[\prod_{i=1}^{2t}c_{v_i,\beta_i}\right]\cdot\operatorname{tr}\left(\prod_{i=1}^{2t}\Gamma^{-1}A_{v_i,\beta_i}\right).
\end{aligned}$$

Let us now make the following observation. Let $(v_1,\beta_1),\ldots,(v_{2t},\beta_{2t})\in\mathcal{H}\times\mathbb{F}^*$ be a term in the above sum. Fix $v\in\mathcal{H}$, and let $R(v)$ denote the set of $i\in[2t]$ such that $v_i=v$. We observe that if for some $v\in\mathcal{H}$, $\sum_{i\in R(v)}\beta_i\neq 0$, then $\mathbb{E}\left[\prod_{i=1}^{2t}c_{v_i,\beta_i}\right]=0$. Indeed, this is because b_v is independent for each $v\in\mathcal{H}$, and so $\mathbb{E}\left[\prod_{i=1}^{2t}c_{v_i,\beta_i}\right]=\prod_{v\in\mathcal{H}}\mathbb{E}\left[\prod_{i\in R(v)}c_{v,\beta_i}\right]$, and

$$\mathbb{E}\left[\prod_{i\in R(v)}c_{v,\beta_i}\right]=\mathbb{E}\left[\prod_{i\in R(v)}\omega_p^{\operatorname{Tr}(\beta_i b_v)}\right]=\mathbb{E}\left[\omega_p^{\operatorname{Tr}((\sum_{i\in R(v)}\beta_i)b_v)}\right].$$

Then, since b_v is uniform from $\mathbb{F}$, it follows that $\mathbb{E}\left[\omega_p^{\operatorname{Tr}((\sum_{i\in R(v)}\beta_i)b_v)}\right]=0$ if $\sum_{i\in R(v)}\beta_i\neq 0$, and $\mathbb{E}\left[\omega_p^{\operatorname{Tr}((\sum_{i\in R(v)}\beta_i)b_v)}\right]=1$ if $\sum_{i\in R(v)}\beta_i=0$. This motivates the following definition.

▶ **Definition 17** (Trivially closed sequence). *Let $(v_1,\beta_1),\ldots,(v_{2t},\beta_{2t})\in\mathcal{H}\times\mathbb{F}^*$. We say that $(v_1,\beta_1),\ldots,(v_{2t},\beta_{2t})\in\mathcal{H}\times\mathbb{F}^*$ is trivially closed with respect to v if it holds that $\sum_{i\in R(v)}\beta_i=0$. We say that the sequence is trivially closed if it is trivially closed with respect to all $v\in\mathcal{H}$.*

With the above definition in hand, we have shown that

$$\mathbb{E}\left[\operatorname{tr}\left(\left(\Gamma^{-1}A\right)^{2t}\right)\right]=\sum_{\substack{(v_1,\beta_1),\ldots,(v_{2t},\beta_{2t})\\ \text{trivially closed}}}\operatorname{tr}\left(\prod_{i=1}^{2t}\Gamma^{-1}A_{v_i,\beta_i}\right).$$

The following lemma yields the desired bound on $\mathbb{E}[\operatorname{tr}((\Gamma^{-1}A)^{2t})]$.

▶ **Lemma 18.** $\sum_{\substack{(v_1,\beta_1),\ldots,(v_{2t},\beta_{2t})\\ \textit{trivially closed}}}\operatorname{tr}\left(\prod_{i=1}^{2t}\Gamma^{-1}A_{v_i,\beta_i}\right)\leq N\cdot 2^{2t}\cdot\left(\frac{2t}{d}\right)^t$.

With Lemma 18, we thus have the desired bound $\mathbb{E}[\operatorname{tr}((\Gamma^{-1}A)^{2t})]$. Taking t to be $c\log_2 N$ for a sufficiently large constant c and applying Markov's inequality finishes the proof. ◀

Proof of Lemma 18. We bound the sum as follows. First, we observe that for a trivially closed sequence $(v_1,\beta_1),\ldots,(v_{2t},\beta_{2t})$, we have

$$\operatorname{tr}\left(\prod_{i=1}^{2t}\Gamma^{-1}A_{v_i,\beta_i}\right)=\sum_{U_0,U_1,\ldots,U_{2t-1}}\prod_{i=0}^{2t-1}\Gamma^{-1}_{U_i}\cdot 1\left(U_i\xrightarrow{v_{i+1},\beta_{i+1}}U_{i+1}\right),$$

where we define $U_{2t} = U_0$. Thus, the sum that we wish to bound in Lemma 18 simply counts the total weight of "trivially closed walks" $U_0, v_1, \beta_1, U_1, \ldots, U_{2t-1}, v_{2t}, \beta_{2t}, U_{2t}$ (where $U_{2t} = U_0$) in the Kikuchi graph A, where the weight of a walk is simply $\prod_{i=0}^{2t-1} \Gamma_{U_i}^{-1}$.

Let us now bound this total weight by encoding a walk $U_0, v_1, \beta_1, U_1, \ldots, U_{2t-1}, v_{2t}, \beta_{2t}, U_{2t}$ as follows.

- First, we write down the start vertex U_0.
- For $i = 1, \ldots, 2t$, we let z_i be 1 if $v_i = v_j$ for some $j < i$. In this case, we say that the edge is "old". Otherwise $z_i = 0$, and we say that the edge is "new".
- For $i = 1, \ldots, 2t$, if z_i is 1 then we encode U_i by writing down the smallest $j \in [2t]$ such that $v_i = v_j$. We note that we *do not* need to specify the element β_i, as for any vertex U, there is at most one V and one $\beta \in \mathbb{F}^*$ such that $1(U \xrightarrow{v_i, \beta} V)$. As pointed out in Remark 11, this crucially saves us a factor of $|\mathbb{F}^*|$ in the total number of constraints that we require.
- For $i = 1, \ldots, 2t$, if z_i is 0 then we encode U_i by writing down an integer in $1, \ldots, \deg(U_{i-1})$ that specifies the edge we take to move to U_i from U_{i-1} (we associate $[\deg(U_{i-1})]$ to the edges adjacent to U_{i-1} with an arbitrary fixed map).

With the above encoding, we can now bound the total weight of all trivially closed walks as follows. First, let us consider the total weight of walks for some fixed choice of $z_1, \ldots, z_{2t}$. We have N choices for the start vertex U_0. For each $i = 1, \ldots, 2t$ where $z_i = 0$, we have $\deg(U_{i-1})$ choices for U_i, and we multiply by a weight of $\Gamma_{U_{i-1}}^{-1} \leq \frac{1}{\deg(U_{i-1})}$. For each $i = 1, \ldots, 2t$ where $z_i = 1$, we have at most $2t$ choices for the index $j < i$, and we multiply by a weight of $\Gamma_{U_{i-1}}^{-1} \leq \frac{1}{d}$. Hence, the total weight for a specific $z_1, \ldots, z_{2t}$ is at most $N \left(\frac{2t}{d}\right)^r$, where r is the number of z_i such that $z_i = 1$.

Finally, we observe that any trivially closed walk must have $r \geq t$. Hence, after summing over all $z_1, \ldots, z_{2t}$, we have the final bound of $N 2^{2t} \left(\frac{2t}{d}\right)^t$, which finishes the proof. ◀

References

1. Jackson Abascal, Venkatesan Guruswami, and Pravesh K. Kothari. Strongly refuting all semi-random Boolean CSPs. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 – 13, 2021*, pages 454–472. SIAM, 2021. doi:10.1137/1.9781611976465.28.
2. Sarah R. Allen, Ryan O'Donnell, and David Witmer. How to Refute a Random CSP. In *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 October, 2015*, pages 689–708. IEEE Computer Society, 2015. doi:10.1109/FOCS.2015.48.
3. Omar Alrabiah and Venkatesan Guruswami. Near-tight bounds for 3-query locally correctable binary linear codes via rainbow cycles. In *65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024*. IEEE, 2024. doi:10.1109/FOCS61266.2024.00112.
4. Omar Alrabiah, Venkatesan Guruswami, Pravesh K. Kothari, and Peter Manohar. A near-cubic lower bound for 3-query locally decodable codes from semirandom CSP refutation. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 1438–1448. ACM, 2023. doi:10.1145/3564246.3585143.
5. Benny Applebaum, Boaz Barak, and Avi Wigderson. Public-key cryptography from different assumptions. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*, pages 171–180. ACM, 2010. doi:10.1145/1806689.1806715.

6 Sanjeev Arora, David R. Karger, and Marek Karpinski. Polynomial time approximation schemes for dense instances of *NP*-hard problems. In *Proceedings of the Twenty-Seventh Annual ACM Symposium on Theory of Computing, 29 May-1 June 1995, Las Vegas, Nevada, USA*, pages 284–293. ACM, 1995. doi:10.1145/225058.225140.

7 Boaz Barak, Siu On Chan, and Pravesh K. Kothari. Sum of Squares Lower Bounds from Pairwise Independence. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 97–106. ACM, 2015. doi:10.1145/2746539.2746625.

8 Arpon Basu, Jun-Ting Hsieh, Pravesh Kothari, and Andrew Lin. Improved lower bounds for all odd-query locally decodable codes. *Electron. Colloquium Comput. Complex.*, pages TR24–189, 2024. URL: https://eccc.weizmann.ac.il/report/2024/189.

9 Siavosh Benabbas, Konstantinos Georgiou, Avner Magen, and Madhur Tulsiani. SDP gaps from pairwise independence. *Theory of Computing*, 8(1):269–289, 2012. doi:10.4086/TOC.2012.V008A012.

10 Amin Coja-Oghlan, Andreas Goerdt, and André Lanka. Strong refutation heuristics for random k-SAT. *Combinatorics, Probability & Computing*, 16(1):5, 2007.

11 Henry Corrigan-Gibbs, Alexandra Henzinger, Yael Kalai, and Vinod Vaikuntanathan. Somewhat homomorphic encryption from linear homomorphism and sparse LPN. *IACR Cryptol. ePrint Arch.*, page 1760, 2024. URL: https://eprint.iacr.org/2024/1760.

12 Quang Dao, Yuval Ishai, Aayush Jain, and Huijia Lin. Multi-party homomorphic secret sharing and sublinear MPC from sparse LPN. In *Advances in Cryptology – CRYPTO 2023 – 43rd Annual International Cryptology Conference, CRYPTO 2023, Santa Barbara, CA, USA, August 20-24, 2023, Proceedings, Part II*, volume 14082 of *Lecture Notes in Computer Science*, pages 315–348. Springer, 2023. doi:10.1007/978-3-031-38545-2_11.

13 Uriel Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 534–543, 2002. doi:10.1145/509907.509985.

14 Uriel Feige. Small linear dependencies for binary vectors of low weight. In *Building Bridges: Between Mathematics and Computer Science*, pages 283–307. Springer, 2008.

15 Dimitris Fotakis, Michael Lampis, and Vangelis Th. Paschos. Sub-exponential Approximation Schemes for CSPs: From Dense to Almost Sparse. In *33rd Symposium on Theoretical Aspects of Computer Science, STACS 2016, February 17-20, 2016, Orléans, France*, volume 47 of *LIPIcs*, pages 37:1–37:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2016. doi:10.4230/LIPICS.STACS.2016.37.

16 Andreas Goerdt and André Lanka. Recognizing more random unsatisfiable 3-sat instances efficiently. *Electron. Notes Discret. Math.*, 16:21–46, 2003. doi:10.1016/S1571-0653(04)00461-5.

17 Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1):613–622, 2001. doi:10.1016/S0304-3975(00)00157-2.

18 Venkatesan Guruswami, Pravesh K. Kothari, and Peter Manohar. Algorithms and certificates for Boolean CSP refutation: smoothed is no harder than random. In *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 – 24, 2022*, pages 678–689. ACM, 2022. doi:10.1145/3519935.3519955.

19 Johan Håstad. Some optimal inapproximability results. *Journal of the ACM (JACM)*, 48(4):798–859, 2001. doi:10.1145/502090.502098.

20 Jun-Ting Hsieh, Pravesh K. Kothari, and Sidhanth Mohanty. A simple and sharper proof of the hypergraph Moore bound. In *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pages 2324–2344. SIAM, 2023. doi:10.1137/1.9781611977554.CH89.

21 Russell Impagliazzo and Ramamohan Paturi. On the Complexity of k-SAT. *J. Comput. Syst. Sci.*, 62(2):367–375, 2001. doi:10.1006/JCSS.2000.1727.

22 Aayush Jain, Huijia Lin, and Amit Sahai. Indistinguishability obfuscation from well-founded assumptions. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 60–73. ACM, 2021. doi:10.1145/3406325.3451093.

23 Aayush Jain, Huijia Lin, and Amit Sahai. Indistinguishability obfuscation from LPN over $\mathbb{F}_p$, DLIN, and PRGs in NC^0. In *Advances in Cryptology – EUROCRYPT 2022 – 41st Annual International Conference on the Theory and Applications of Cryptographic Techniques, Trondheim, Norway, May 30 – June 3, 2022, Proceedings, Part I*, volume 13275 of *Lecture Notes in Computer Science*, pages 670–699. Springer, 2022. doi:10.1007/978-3-031-06944-4_23.

24 Oliver Janzer and Peter Manohar. A $k^{\frac{q}{q-2}}$ lower bound for odd query locally decodable codes from bipartite kikuchi graphs. *Electron. Colloquium Comput. Complex.*, pages TR24–187, 2024. URL: https://eccc.weizmann.ac.il/report/2024/187.

25 Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 767–775. ACM, 2002. doi:10.1145/509907.510017.

26 Pravesh K. Kothari and Peter Manohar. An exponential lower bound for linear 3-query locally correctable codes. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, Vancouver, BC, Canada, June 24-28, 2024*, pages 776–787. ACM, 2024. doi:10.1145/3618260.3649640.

27 Pravesh K. Kothari and Peter Manohar. Exponential lower bounds for smooth 3-lccs and sharp bounds for designs. In *65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024*. IEEE, 2024. doi:10.1109/FOCS61266.2024.00110.

28 Pravesh K. Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any CSP. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 132–145. ACM, 2017. doi:10.1145/3055399.3055485.

29 Ryuhei Mori and David Witmer. Lower Bounds for CSP Refutation by SDP Hierarchies. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2016, September 7-9, 2016, Paris, France*, volume 60 of *LIPIcs*, pages 41:1–41:30, 2016. doi:10.4230/LIPICS.APPROX-RANDOM.2016.41.

30 Assaf Naor and Jacques Verstraëte. Parity check matrices and product representations of squares. *Combinatorica*, 28(2):163–185, 2008. doi:10.1007/S00493-008-2195-2.

31 Ryan O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

32 Ryan O'Donnell and David Witmer. Goldreich's PRG: evidence for near-optimal polynomial stretch. In *2014 IEEE 29th Conference on Computational Complexity (CCC)*, pages 1–12. IEEE, 2014. doi:10.1109/CCC.2014.9.

33 Seyoon Ragavan, Neekon Vafa, and Vinod Vaikuntanathan. Indistinguishability obfuscation from bilinear maps and LPN variants. In *Theory of Cryptography – 22nd International Conference, TCC 2024, Milan, Italy, December 2-6, 2024, Proceedings, Part IV*, volume 15367 of *Lecture Notes in Computer Science*, pages 3–36. Springer, 2024. doi:10.1007/978-3-031-78023-3_1.

34 Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random CSPs below the spectral threshold. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 121–131. ACM, 2017. doi:10.1145/3055399.3055417.

35 Grant Schoenebeck. Linear level lasserre lower bounds for certain k-csps. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 593–602. IEEE Computer Society, 2008. doi:10.1109/FOCS.2008.74.

36 Alexander S. Wein, Ahmed El Alaoui, and Cristopher Moore. The Kikuchi Hierarchy and Tensor PCA. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019, Baltimore, Maryland, USA, November 9-12, 2019*, pages 1446–1468. IEEE Computer Society, 2019. doi:10.1109/FOCS.2019.000-2.

37 Tal Yankovitz. A stronger bound for linear 3-lcc. In *65th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2024, Chicago, IL, USA, October 27-30, 2024*. IEEE, 2024. doi:10.1109/FOCS61266.2024.00109.