arXiv:2102.04769v1 [cs.CC] 9 Feb 2021

Constant Approximating k-Clique is W[1]-hard

Bingkai Lin
Nanjing University, China
lin@nju.edu.cn

February 10, 2021

Abstract

For every graph G, let $\omega(G)$ be the largest size of complete subgraph in G. This paper presents a simple algorithm which, on input a graph G, a positive integer k and a small constant $\epsilon > 0$, outputs a graph G' and an integer k' in $2^{\Theta(k^5)} \cdot |G|^{O(1)}$-time such that (1) $k' \leq 2^{\Theta(k^5)}$, (2) if $\omega(G) \geq k$, then $\omega(G') \geq k'$, (3) if $\omega(G) < k$, then $\omega(G') < (1-\epsilon)k'$. This implies that no $f(k) \cdot |G|^{O(1)}$-time algorithm can distinguish between the cases $\omega(G) \geq k$ and $\omega(G) < k/c$ for any constant $c \geq 1$ and computable function f, unless $FPT = W[1]$.

1 Introduction

Given a simple graph G and a positive integer k, the task of k-Clique problem is to decide whether $\omega(G) \geq k$. In parameterized complexity [DF99, FG06], the k-Clique problem with k as its parameter is a canonical $W[1]$-complete problem [DF95]. Unless $W[1] = FPT$, it has no $f(k) \cdot |G|^{O(1)}$-time algorithm (FPT-algorithm) for any computable function $f : \mathbb{N} \to \mathbb{N}$. This problem has been used as a starting point in many reductions and thus plays a fundamental role in the area of parameterized complexity. Yet, it is still not known whether constant approximating k-Clique is also $W[1]$-hard. More precisely, we consider the following question:

Question 1.1. *Is there any algorithm which, on input a graph G and a positive integer k, outputs a new graph G' and a positive integer k' in $f(k) \cdot |G|^{O(1)}$-time for some computable function $f : \mathbb{N} \to \mathbb{N}$ such that,*

- *$k' = g(k)$ for some computable function $g : \mathbb{N} \to \mathbb{N}$,*
- *if $\omega(G) \geq k$, then $\omega(G') \geq k'$,*
- *if $\omega(G) < k$, then $\omega(G') < k'/2$?*

The question above is motivated by the study of FPT-approximation algorithms for k-Clique. For any $c \geq 1$, we say an algorithm is a c-FPT-approximation algorithm for k-Clique if on input a graph G it outputs a clique of size $\omega(G)/c$ in G in $f(\omega(G)) \cdot |G|^{O(1)}$-time for some computable function $f : \mathbb{N} \to \mathbb{N}$. Whether there exists such an algorithm has been repeatedly raised in the literature [Mar08, FGMS, CGG06, DF13]. Previous results for FPT-inapproximability of k-Clique are under strong assumptions which already have a gap [CCK+17, BEKP15]. Proving such results based on standard assumptions is an interesting and important open question. It is wildly believed that the technique needed to resolve this question is closely related to a PCP-theorem for parameterized complexity [CGG06]:

> Non-approximability results in the classical framework were proved for the CLIQUE problem using the PCP-theorem, so it might be necessary to obtain a parameterized version of the PCP-theorem to solve these questions.

This paper gives the first positive answer to Question 1.1. An immediate corollary of our result is the non-existence of FPT-approximation algorithm for the k-Clique problem under the standard parameterized complexity hypothesis $W[1] \neq FPT$.

Theorem 1.2. *Assuming that k-Clique has no FPT-algorithm, there is no FPT-algorithm that can approximate k-Clique to any constant.*

The main contribution of this paper is to show how to create a constant gap for $\omega(G)$ from a $W[1]$-hard problem with no gap.

1.1 Overview of the reduction

Let us illustrate the idea using a toy k-Vector-Sum problem and the Walsh-Hadamard code. Throughout this paper, we will work on finite field $\mathbb{F}$ with characteristic 2. For any $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{F}^d$ and $\vec{a}_1, \ldots, \vec{a}_k \in \mathbb{F}^d$, let $H(\vec{v}_1, \ldots, \vec{v}_k)_{\vec{a}_1,\ldots,\vec{a}_k} = \sum_{i\in[k]} \vec{a}_i \cdot \vec{v}_i$ be the Walsh-Hadamard code of $\vec{v}_1, \ldots, \vec{v}_k$. Here $\vec{a}_i \cdot \vec{v}_i$ denotes the dot product of vectors $\vec{a}_i$ and $\vec{v}_i$. Given k vector $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{F}^d$ and a target vector $\vec{t} \in \mathbb{F}^d$, we want to test whether $\sum_{i\in[k]} \vec{v}_i = \vec{t}$. For that sake, we construct a constraint satisfaction problem (CSP) on variables $\{x_{\vec{a}_1,\ldots,\vec{a}_k} : \vec{a}_1, \ldots, \vec{a}_k \in \mathbb{F}^d\}$. Any assignment to these variables can be seen as a long vector $\vec{x} \in \mathbb{F}^{kd}$, which is supposed to be the Walsh-Hadamard code of $\vec{v}_1, \ldots, \vec{v}_k$. Then we do the following tests.

(T1) To ensure that the vector $\vec{x} \in \mathbb{F}^{kd}$ is a Walsh-Hadamard code of some vector $\vec{v}_1, \ldots, \vec{v}_k \in \mathbb{F}^d$, we check if $\vec{x}_{\vec{a}_1+\vec{b}_1,\ldots,\vec{a}_k+\vec{b}_1} = \vec{x}_{\vec{a}_1,\ldots,\vec{a}_k} + \vec{x}_{\vec{b}_1,\ldots,\vec{b}_k}$ for $\vec{a}_1, \ldots, \vec{a}_k, \vec{b}_1, \ldots, \vec{b}_k \in \mathbb{F}^d$.

(T2) Check if $\vec{x}_{\vec{a}_1,\ldots,\vec{a}_i+\vec{a},\ldots,\vec{a}_k} - \vec{x}_{\vec{a}_1,\ldots,\vec{a}_k} = \vec{v}_i \cdot \vec{a}$ for $i \in [k]$, $\vec{a} \in \mathbb{F}^d$ and $\vec{a}_1, \ldots, \vec{a}_k \in \mathbb{F}^d$.

(T3) Check if $\vec{x}_{\vec{a}_1+\vec{a},\ldots,\vec{a}_k+\vec{a}} - \vec{x}_{\vec{a}_1,\ldots,\vec{a}_k} = \vec{a} \cdot \vec{t}$ for $\vec{a}, \vec{a}_1, \ldots, \vec{a}_k \in \mathbb{F}^d$.

By the linearity testing [BLR93], for any assignment $\vec{x} \in \mathbb{F}^{kd}$, if $\vec{x}$ satisfies δ-fractions of constraints in (T1), then $\vec{x}$ is at most $(1-\delta)$-far from the Walsh-Hadamard code of some vectors $\vec{u}_1, \ldots, \vec{u}_k \in \mathbb{F}^d$. That is, for a δ-fraction of $(\vec{a}_1, \ldots, \vec{a}_k) \in \mathbb{F}^{dk}$, $\vec{x}_{\vec{a}_1,\ldots,\vec{a}_k} = H(\vec{u}_1, \ldots, \vec{u}_k)_{\vec{a}_1,\ldots,\vec{a}_k}$. Next, if $\vec{u}_1 + \cdots + \vec{u}_k \neq \vec{t}$, then for a $(1 - 1/|\mathbb{F}|)$-fraction of $\vec{a} \in \mathbb{F}^d$, $\vec{a} \cdot (\vec{u}_1 + \cdots + \vec{u}_k) \neq \vec{a} \cdot \vec{t}$. If for some $i \in [k]$, $\vec{u}_i \neq \vec{v}_i$, then for a $(1 - 1/|\mathbb{F}|)$-fraction of $\vec{a} \in \mathbb{F}^d$, $\vec{a} \cdot \vec{u}_i \neq \vec{a} \cdot \vec{v}_i$. We conclude that, if $\vec{v}_1 + \cdots + \vec{v}_k = \vec{t}$, then there exists a vector $\vec{x}$ satisfying all the constraints. If $\vec{v}_1 + \cdots + \vec{v}_k \neq \vec{t}$, then for any assignment $\vec{x}$, one of the following must hold:

- a constant fraction of constraints in (T1) are not satisfied.
- for some $i \in [k]$, a constant fraction of constraints in (T2) with respect to i are not satisfied.
- a constant fraction of constraints in (T3) are not satisfied.

Then we construct a graph G from the CSP instance using a modified FGLSS-reduction [FGL+96]. The vertex set of G consists of two parts A and B.

The vertices of A are corresponding to assignments to variables $\{x_{\vec{a}_1,\ldots,\vec{a}_k} : \vec{a}_1, \ldots, \vec{a}_k \in \mathbb{F}^d\}$ of the CSP instance. Two vertices in A are adjacent unless they are corresponding to assignments that are not consistent, i.e., assigning different values to the same variable, or there exists a (T2) or (T3) constraint between them and they do not satisfy that constraint. Note that there are $|\mathbb{F}|^{1+dk}$ vertices in A. They can be partitioned into $|\mathbb{F}|^{dk}$ independent sets of size $|\mathbb{F}|$.

The vertices in B are corresponding to assignments to three variables satisfying constraints in (T1). Two vertices in B are adjacent unless they are corresponding to inconsistent assignments. Note that the vertex set of B can be partitioned into $|\mathbb{F}|^{2dk}$ disjoint subsets, each forms an independent set of size $|\mathbb{F}|^2$.

We add an edge between any two vertices in A and B if the assignments corresponding to these vertice are consistent. Since the sizes of A and B are not balanced, we assign to each vertex in A weight $|\mathbb{F}|^{kd}$ and each vertex in B weight 1.

If $\vec{v}_1 + \cdots + \vec{v}_k = \vec{t}$, then G contains a clique of weight $2|\mathbb{F}|^{2kd}$. This clique consists of $|\mathbb{F}|^{kd}$ vertices from A and $|\mathbb{F}|^{2kd}$ vertices from B. If $\vec{v}_1 + \cdots + \vec{v}_k \neq \vec{t}$, then for any clique X in G, one of the following must hold:

(i) $|X \cap B| \leq (1-\epsilon)|\mathbb{F}|^{2kd}$,

(ii) there exists $\vec{a} \in \mathbb{F}^d$ and $i \in [k]$ such that the variable set V can be partitioned into two disjoint sets $V = V_0 \cup V_1$ with $V_1 = \{x_{\vec{a}_1,\ldots,\vec{a}_i+\vec{a},\ldots,\vec{a}_k} : x_{\vec{a}_1,\ldots,\vec{a}_k} \in V_0\}$. The assignment corresponding to X satisfies only a $(1-\epsilon)$-fraction of (T2) constraints between V_0 and V_1,

(iii) there exists $\vec{a} \in \mathbb{F}^d$ such that the variable set V can be partitioned into two disjoint sets $V = V_0 \cup V_1$ with $V_1 = \{x_{\vec{a}_1+\vec{a},\ldots,\vec{a}_k+\vec{a}} : x_{\vec{a}_1,\ldots,\vec{a}_k} \in V_0\}$. The assignment corresponding to X satisfies only a $(1-\epsilon)$-fraction of (T3) constraints between V_0 and V_1.

Either (ii) or (iii) implies that $|X \cap A| \leq [1/2+(1-\epsilon)1/2]|\mathbb{F}^{kd}|$. So, in summary, when $\vec{v}_1+\cdots+\vec{v}_k \neq \vec{t}$, we have for every clique of G, either $|X \cap B| \leq (1-\epsilon)|\mathbb{F}|^{2kd}$ or $|X \cap A| \leq (1-\epsilon/2)|\mathbb{F}^{kd}|$. In both cases, every clique in G has at most $(2-\epsilon/2)|\mathbb{F}|^{2kd}$ weight.

There are two problems needed to be solved:

(P1) In the real k-Vector-Sum problem, we are given k sets $V_1, \ldots, V_k$ of vectors instead of k vectors. How to test whether $\vec{x}_{\vec{a}_1,\ldots,\vec{a}_i+\vec{a},\ldots,\vec{a}_k} - \vec{x}_{\vec{a}_1,\ldots,\vec{a}_k} = \vec{v}_i \cdot \vec{a}$ for the same vector $\vec{v}_i \in V_i$? In some bad scenario, it is possible that for each $\vec{a}$, there exists a vector $\vec{v}_{\vec{a}} \in V_i$ such that $\vec{x}_{\vec{a}_1,\ldots,\vec{a}_i+\vec{a},\ldots,\vec{a}_k} - \vec{x}_{\vec{a}_1,\ldots,\vec{a}_k} = \vec{v}_{\vec{a}} \cdot \vec{a}$.

(P2) The k-Vector-Sum problem is $W[1]$-hard when $d = k^{\Omega(1)} \log n$. Applying our reduction directly would take at least $|\mathbb{F}|^{kd} \geq n^{k^{\Omega(1)}}$ time, which we cannot afford.

We handle these problems by sampling ℓ matrices $A_1, \ldots, A_\ell \in \mathbb{F}^{h \times d}$ with $h = k^2$, and replacing each vector $\vec{v} \in V_i$ by an ℓ-tuple $(A_1\vec{v}, \ldots, A_\ell\vec{v})$ of h-dimension vectors. For every $\vec{a}_1, \ldots, \vec{a}_k \in \mathbb{F}^h$, the value of the variable $x_{\vec{a}_1,\ldots,\vec{a}_k}$ becomes a vector in $\mathbb{F}^\ell$ instead of just an element of $\mathbb{F}$. The number of variables become $|\mathbb{F}|^{kh}$ instead of $|\mathbb{F}|^{kd}$. Since $|\mathbb{F}| = O(1)$, the reduction can be done in FPT-time if $\ell \leq O(\log n + h)$. To ensure that the constraints in (T2) still work, we show that when $\ell \geq \Omega(\log n + h)$, with high probability, for all $\vec{a} \in \mathbb{F}^h$ and distinct $\vec{v}, \vec{u} \in V_i$, $(\vec{a}^T A_1 \vec{v}, \ldots, \vec{a}^T A_\ell \vec{v}) \neq (\vec{a}^T A_1 \vec{u}, \ldots, \vec{a}^T A_\ell \vec{u})$ and the bad scenario in (P1) will not occur. For constraints in (T3), we show that with high probability, for all distinct $\vec{t}, \vec{t}' \in \mathbb{F}^d$, $(A_1\vec{t}, \ldots, A_\ell\vec{t}) \neq (A_1\vec{t}', \ldots, A_\ell\vec{t}')$.

1.2 Related work

The k-Clique problem is one of the first known NP-hard problems in [Kar72]. It was showned that approximating k-Clique to a factor of $n^{1-\epsilon}$ is also NP-hard after a long line of research [FGL+96, BGLR93, BS94, Gol98, FK00, Has96, Zuc06]. As pointed out in [CL19], the classical inapproximability results of k-Clique inevitably produce instances with large $\omega(G)$. However, in parameterized complexity, we consider instances with small $\omega(G)$ which does not depend on the size of G. In order to show that k-Clique is still hard to approximate when k is small, a natural idea is to use the PCP-theorem [ALM+98, AS98] to obtain a gap for the SAT problem and then use the method of compressing to reduce the optimum solution size [HKK13]. Unfortunately, since the PCP-theorem causes a polylogarithmic blow-up in the size of SAT instance, this approach cannot rule out FPT-approximation for k-Clique.

To circumvent this, researches used stronger hypothesis to obtain a gap for the SAT problem. Bonnet et al. [BEKP15] used ETH [IPZ01] and the linear PCP conjecture to show constant FPT-inapproximability of k-Clique. Assuming Gap-ETH [Din16, MR16], it was shown that there is no $o(k)$-FPT-approximation for k-Clique [CCK+17].

In [LRSZ20], a weaker conjecture called Parameterized Inapproximability Hypothesis (PIH) was postulated. PIH states that binary CSP parameterized by the number of variables has no constant FPT-approximation. It is easy to see that PIH implies k-Clique has no constant FPT-approximation. Interestingly, it is not known if the hardness of approximation of k-Clique implies PIH [FKLM20].

Assuming a conjecture called DEG-2-SAT, Khot and Shinkar [KS16] used a different approach to rule out FPT-approximation for k-Clique. Although the conjecture turned out to be false [Kay14], their work is still inspiring. The idea of multiplying the input instance with matrices in our reduction is from their paper.

In recent years, several gap-creating techniques have been successfully used to show FPT inapproximabilities [Lin18, LKM19, Lin19, Wło20, KLN21]. We refer the reader to [FKLM20] for a survey of these results.

2 Preliminaries

For every vector $\vec{v} = (v_1, \ldots, v_d) \in \mathbb{F}^d$ and $i \in [d]$, let $\vec{v}[i] = v_i$. For every $\vec{x}, \vec{y} \in \mathbb{F}^n$, we use $\vec{x} \cdot \vec{y} = \sum_{i \in [n]} \vec{x}[i]\vec{y}[i]$ to denote the dot product of $\vec{x}$ and $\vec{y}$. For $X \subseteq \mathbb{F}^d$ and $\vec{v} \in \mathbb{F}^d$, let $X + \vec{v} = \{\vec{x} + \vec{v} : \vec{x} \in X\}$. For $\vec{a} \in \mathbb{F}^n$ and $\vec{b} \in \mathbb{F}^m$, let $\vec{a} \circ \vec{b}$ be the the result of concatenating $\vec{a}$ and $\vec{b}$. For any dn vectors $\vec{v} = (v_1, \ldots, v_{dn}) \in \mathbb{F}^{dn}$ and $\vec{a} = (a_1, \ldots, a_d) \in \mathbb{F}^d$, let

$$F(\vec{a}, \vec{v}) = (\sum_{i \in [d]} a_i v_i, \ldots, \sum_{i \in [d]} a_i v_{i+jd}, \ldots, \sum_{i \in [d]} a_i v_{i+(n-1)d}) \in \mathbb{F}^n.$$

For all $n \in \mathbb{N}$, let $\vec{0}_n$ and $\vec{1}_n$ be the n-dimension all-zero vector and all-one vector respectively.

Definition 2.1 (Distance)**.** *For every $d \in \mathbb{N}$ and $\vec{x}, \vec{y} \in \mathbb{F}^d$, let*

$$dist(\vec{x}, \vec{y}) = \frac{|\{i \in [d] : \vec{x}[i] \neq \vec{y}[i]\}|}{d}.$$

For ease of notation, let $dist(\vec{x}) = dist(\vec{x}, \vec{0}_d)$.

Let G and H be two groups and $+$ be the group operator. For any $\delta \in [0, 1]$ and $f, g : G \to H$, we say f is δ-far from g if $\Pr_x[f(x) \neq g(x)] = \delta$. A function $f : G \to H$ is a homomorphism if $f(x) + f(y) = f(x + y)$ for all $x, y \in G$. The follow theorem is from [BLR93, Gol16].

Theorem 2.2 (Linearity test)**.** *If a function $f : G \to H$ satisfies $\Pr_{x,y}[f(x) + f(y) = f(x + y)] \geq (1 - \delta/2)$ for some small δ, then there exists a homomorphism function $g : G \to H$ such that f is at most δ-far from g, i.e., $\Pr_x[f(x) = g(x)] \geq (1 - \delta)$.*

Definition 2.3 (Constraint Satisfaction Problem (CSP))**.** *Given an alphabet Γ, an instance of constraint satisfaction problem contains a set of variables $V = \{v_1, \ldots, v_k\}$ and constraints $\{C_1, C_2, \ldots, C_m\}$. For every $i \in [m]$, $C_i = (\vec{s}_i, R_i)$, where $\vec{s}_i = (v_{j_1}, \ldots, v_{j_{\ell_i}})$ is an ℓ_i-tuple of variables for some $\ell_i \in [k]$ and $R_i \subseteq \Gamma^{\ell_i}$. The goal is to find an assignment $\sigma : V \to \Gamma$ such that*

- *for all $i \in [m]$, $\sigma(\vec{s}_i) \in R_i$.*

In parameterized complexity, the hypothesis of $W[1] \neq FPT$ states that no algorithm can, on input a graph G and a positive integer k, decide whether $\omega(G) \geq k$ in $f(k) \cdot |G|^{O(1)}$ time for any computable function $f : \mathbb{N} \to \mathbb{N}$. A parameterized problem L is $W[1]$-hard if there is a reduction from k-Clique to this problem such that for every instance (G, k) of k-Clique, the reduction outputs an instance (x, k') of L in $f(k) \cdot |G|^{O(1)}$-time for some computable function $f : \mathbb{N} \to \mathbb{N}$ and:

- (G, k) is a yes-instance of k-Clique if and only if (x, k') is a yes-instance of L,
- $k' \leq g(k)$ for some computable function $g : \mathbb{N} \to \mathbb{N}$.

Obviously, if a parameterized problem L is $W[1]$-hard, then no $f(k) \cdot |x|^{O(1)}$-time algorithm can decide whether (x, k) is a yes-instance of L unless $W[1] = FPT$. We say approximating k-Clique to a factor of c is $W[1]$-hard if the existence of $f(k) \cdot |G|^{O(1)}$-time algorithm that can distinguish $\omega(G) \geq k$ and $\omega(G) < k/c$ would imply $W[1] = FPT$.

Definition 2.4 (k-Vector-Sum)**.** *Given k sets $V_1, \ldots, V_k$ of vectors and a target vector $\vec{t}$ in $\mathbb{F}^m$, the goal of k-vector-sum problem is to decide whether there exist $\vec{v}_1 \in V_1, \ldots, \vec{v}_k \in V_k$ such that*

$$\sum_{i \in [k]} \vec{v}_i = \vec{t}.$$

The $W[1]$-hardness of k-Vector-Sum was proved in [ALW13]. For the convenience of the reader, we include a proof in the Appendix.

Theorem 2.5. *k-Vector-Sum with $\mathbb{F} = \mathbb{F}_2$ and $m = \Theta(k^2 \log n)$ is $W[1]$-hard parameterized by k.*

3 Gap-reduction from k-Vector-Sum to k-Clique

Given an instance $(V_1, V_2, \ldots, V_k, \vec{t})$ of k-Vector-Sum over $\mathbb{F}^m$. Let $V = \bigcup_{i \in [k]} V_i$, $n = |V|$ and $h = \Theta(k^2)$. By Theorem 2.5, we can assume that $m = h \log n$. Let $\mathbb{F}$ be a finite field with $|\mathbb{F}| = 4$. Vectors in the hardness instances from Theorem 2.5 can still be treated as vectors in $\mathbb{F}^m$. Since V only contains vectors in $\{0,1\}^m$, we have

$$\vec{v} \neq c\vec{u} \text{ for any distinct } \vec{v}, \vec{u} \in V \text{ and nonzero } c \in \mathbb{F}. \tag{1}$$

For any $\ell \in \mathbb{N}$, select ℓ matrices $A_1, A_2, \ldots, A_\ell \in \mathbb{F}^{h \times m}$ randomly and independently. For every $\vec{v} \in \mathbb{F}^m$, let

$$g(\vec{v}) = (A_1\vec{v}, \cdots, A_\ell\vec{v}) \in \mathbb{F}^{h\ell}.$$

For every vector $\vec{\alpha} \in \mathbb{F}^h$ and $\vec{v} \in \mathbb{F}^m$, let

$$f(\vec{\alpha}, \vec{v}) = (\vec{\alpha}^T A_1 \vec{v}, \cdots, \vec{\alpha}^T A_\ell \vec{v}) \in \mathbb{F}^\ell.$$

Recall that for any $d \cdot n$ vectors $\vec{v} = (v_1, \ldots, v_{dn}) \in \mathbb{F}^{dn}$ and $\vec{a} = (a_1, \ldots, a_d) \in \mathbb{F}^d$,

$$F(\vec{a}, \vec{v}) = (\sum_{i \in [d]} a_i v_i, \ldots, \sum_{i \in [d]} a_i v_{i+jd}, \ldots, \sum_{i \in [d]} a_i v_{i+(n-1)d}) \in \mathbb{F}^n.$$

It follows that $f(\vec{\alpha}, \vec{v}) = F(\vec{\alpha}, g(\vec{v}))$. Note that for every $\vec{v} \in \mathbb{F}^{dn}$, $F(\cdot, \vec{v}) : \mathbb{F}^d \to \mathbb{F}^n$ is a homomorphism from $\mathbb{F}^d$ to $\mathbb{F}^n$. Every homomorphism f from $\mathbb{F}^d$ to $\mathbb{F}^n$ is also a bitwise linear function, so it can be written as $f(\cdot) = F(\cdot, \vec{v})$ for some $\vec{v} \in \mathbb{F}^{dn}$.

Lemma 3.1. *If $1/10 > (1/|\mathbb{F}|)^{\ell h} \cdot 2^m$, then with probability at least $9/10$, for all nonzero vector $\vec{v} \in \mathbb{F}^m$, $g(\vec{v})) \neq \vec{0}_{\ell h}$.*

Proof. For any nonzero vector $\vec{v} \in \mathbb{F}^m$, $\Pr[A_i\vec{v} = \vec{0}_h] = (1/|\mathbb{F}|)^h$.

$$\Pr[g(\vec{v}) = \vec{0}_{\ell h}] = \prod_{i \in [\ell]} \Pr[A_i \vec{v} = \vec{0}_h] = (1/|\mathbb{F}|)^{\ell h}.$$

With probability at least

$$1 - (1/|\mathbb{F}|)^{\ell h} \cdot 2^m \geq 9/10,$$

$g(\vec{v}) \neq \vec{0}_{\ell h}$ for all nonzero $\vec{v} \in \mathbb{F}^m$. □

Lemma 3.2. *If $A \in \mathbb{F}^{h \times m}$ is a random matrix, then for any nonzero vectors $\vec{b}, \vec{c} \in \mathbb{F}^h$ and distinct $\vec{v}, \vec{u} \in \mathbb{F}^m$ with $\vec{v} \neq a\vec{u}$ for any $a \in \mathbb{F} \setminus \{0\}$,*

$$\Pr[\vec{b}^T A \vec{v} = \vec{c}^T A \vec{u}] = 1/|\mathbb{F}|. \tag{2}$$

Proof. Let B be an $h \times m$ matrix with $B_{ij} = \vec{b}[i] \cdot \vec{v}[j]$. Let C be an $h \times m$ matrix with $C_{ij} = \vec{c}[i] \cdot \vec{u}[j]$. We can treat A, B as a vector of length hm and use $A \cdot B$ denote their dot product. It follows that $\vec{b}^T A \vec{v} = B \cdot A$ and $\vec{c}^T A \vec{u} = C \cdot A$. Since $\vec{v} \neq a\vec{u}$ for any nonzero $a \in \mathbb{F}$, we have $B - C$ is not a zero matrix. Therefore,

$$\Pr[\vec{b}^T A \vec{v} = \vec{c}^T A \vec{u}] = \Pr[B \cdot A = C \cdot A] = \Pr[(B - C) \cdot A = 0] = 1/|\mathbb{F}|.$$

□

Lemma 3.3. *If $|V|^2 \cdot |\mathbb{F}|^h \cdot (1/|\mathbb{F}|)^\ell < 1/10$, then with probability at least $9/10$, $f(\vec{\alpha},\vec{v}) \neq f(\vec{\alpha},\vec{u})$ for any distinct $\vec{v},\vec{u} \in V$ and nonzero $\vec{\alpha} \in \mathbb{F}^h$.*

Proof. By (1), $\vec{v} \neq c\vec{u}$ for any nonzero $c \in \mathbb{F}$. Apply Lemma 3.2 with $\vec{b} = \vec{c} = \vec{\alpha}$, we get

$$\Pr[f(\vec{\alpha},\vec{v}) = f(\vec{\alpha},\vec{u})] = \prod_{i\in[\ell]} \Pr[\vec{\alpha}^T A_i \vec{v} = \vec{\alpha}^T A_i \vec{u}] = (1/|\mathbb{F}|)^\ell.$$

There are at most $|V|^2$ pairs of $(\vec{v},\vec{u})$ and at most $|\mathbb{F}|^d$ choices of $\vec{\alpha}$. Since $|V|^2 \cdot |\mathbb{F}|^d \cdot (1/|\mathbb{F}|)^\ell < 1/10$, with probability at least $9/10$, $f(\vec{\alpha},\vec{v}) \neq f(\vec{\alpha},\vec{u})$ for all nonzero $\vec{\alpha} \in \mathbb{F}^h$ and distinct $\vec{v},\vec{u} \in V$. $\square$

Lemma 3.4. *If $|V|^3 \cdot |\mathbb{F}|^{2h} \cdot (1/|\mathbb{F}|)^\ell < 1/10$, then with probability at least $9/10$, $f(\vec{\alpha},\vec{v}) + f(\vec{\alpha}',\vec{u}) \neq f(\vec{\alpha}+\vec{\alpha}',\vec{w})$ for any distinct $\vec{v},\vec{u},\vec{w} \in V$ and nonzero $\vec{\alpha},\vec{\alpha}'$.*

Proof. Observe that $\Pr[f(\vec{\alpha},\vec{v}) + f(\vec{\alpha}',\vec{u}) = f(\vec{\alpha}+\vec{\alpha'},\vec{w})] = \Pr[f(\vec{\alpha},\vec{v}-\vec{w}) = f(\vec{\alpha'},\vec{w}-\vec{u})]$ and $\vec{w}-\vec{u} = \vec{w}+\vec{u} \neq \vec{v}+\vec{w} = \vec{v}-\vec{w}$. Since $\vec{w}-\vec{u}$ and $\vec{v}-\vec{w}$ are vectors in $\{0,1\}^m \subseteq \mathbb{F}^m$, $\vec{w}-\vec{u} \neq (\vec{v}-\vec{w})$ implies $\vec{w}-\vec{u} \neq a(\vec{v}-\vec{w})$ for any $a \in \mathbb{F} \setminus \{0\}$. By Lemma 3.2, $\Pr[f(\vec{\alpha},\vec{v}-\vec{w}) = f(\vec{\alpha'},\vec{w}-\vec{u})] \leq (1/|\mathbb{F}|)^\ell$. There are at most $|V|^3$ pairs of $(\vec{v},\vec{u},\vec{w})$ and at most $|\mathbb{F}|^{2d}$ choices of $\vec{\alpha},\vec{\alpha}'$. Since $|V|^3 \cdot |\mathbb{F}|^{2d} \cdot (1/|\mathbb{F}|)^\ell < 1/10$, with probability at least $9/10$, $f(\vec{\alpha},\vec{v}) + f(\vec{\alpha}',\vec{u}) \neq f(\vec{\alpha}+\vec{\alpha'},\vec{w})$ for any distinct $\vec{v},\vec{u},\vec{w} \in V$ and nonzero $\vec{\alpha},\vec{\alpha}'$. $\square$

Construction of the CSP. Let $\ell = 2\log n + 2h$. Then for large n,

$$(1/|\mathbb{F}|)^{\ell h} \cdot 2^{h\log n} = 4^{-2h\log n - 2h} \cdot 2^{h\log n} \leq 2^{-3h\log n} < 1/10,$$

and

$$|V|^3 \cdot |\mathbb{F}|^{2h} \cdot (1/|\mathbb{F}|)^\ell = n^3 \cdot 4^{2h} \cdot 4^{-2logn-2h} \leq 1/n \leq 1/10.$$

By Lemma 3.3, Lemma 3.4 and Lemma 3.1, with probability at least $7/10$, $g(\vec{v}) \neq \vec{0}_{\ell h}$ for all $\vec{v} \in \mathbb{F}^{h\log n}$, $f(\vec{\alpha},\vec{v}) \neq f(\vec{\alpha},\vec{u})$ and $f(\vec{\alpha},\vec{v}) + f(\vec{\alpha}',\vec{u}) \neq f(\vec{\alpha}+\vec{\alpha'},\vec{w})$ for all distinct $\vec{v},\vec{u},\vec{w} \in V$ and nonzero $\vec{\alpha},\vec{\alpha}'$.

Construct a CSP instance I with $|\mathbb{F}|^{kh}$ variables $\{x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k} : \vec{\alpha}_1 \dots \vec{\alpha}_k \in \mathbb{F}^h\}$. The alphabet of this CSP is $\Gamma = \mathbb{F}^\ell$. If the instance is a yes-instance, then each $x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k}$ is expected to take the value $f(\vec{\alpha}_1,\vec{v}_1) + \dots + f(\vec{\alpha}_k,\vec{v}_k)$ for some a solution $\vec{v}_1,\dots,\vec{v}_k$ to the k-Vector-Sum problem. We now describe three types of constraints.

(C1) For all $\vec{\alpha}_1,\dots,\vec{\alpha}_k$ and $\vec{\beta}_1,\dots,\vec{\beta}_k$, check if $x_{\vec{\alpha}_1+\vec{\beta}_1,\dots,\vec{\alpha}_k+\vec{\beta}_k} = x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k} + x_{\vec{\beta}_1,\dots,\vec{\beta}_k}$. In other words, the relation for variable tuple $(x_{\vec{\alpha}_1+\vec{\beta}_1,\dots,\vec{\alpha}_k+\vec{\beta}_k}, x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k}, x_{\vec{\beta}_1,\dots,\vec{\beta}_k})$ is

$$R_{x_{\vec{\alpha}_1+\vec{\beta}_1,\dots,\vec{\alpha}_k+\vec{\beta}_k}, x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k}, x_{\vec{\beta}_1,\dots,\vec{\beta}_k}} = \{(\vec{a},\vec{b},\vec{c}) \in \Gamma^3 : \vec{a} = \vec{b} + \vec{c}\}.$$

(C2) For every $i \in [k]$ and $\vec{\alpha}_1,\dots,\vec{\alpha}_k,\vec{\alpha} \in \mathbb{F}^h$, check if $x_{\vec{\alpha}_1,\dots,\vec{\alpha}_i+\vec{\alpha},\dots,\vec{\alpha}_k} - x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k} = f(\vec{\alpha},\vec{v})$ for some $\vec{v} \in V_i$. In other words, the constraint between $x_{\vec{\alpha}_1,\dots,\vec{\alpha}_i+\vec{\alpha},\dots,\vec{\alpha}_k}$ and $x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k}$ is

$$R_{x_{\vec{\alpha}_1,\dots,\vec{\alpha}_i+\vec{\alpha},\dots,\vec{\alpha}_k}, x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k}} = \{(\vec{a},\vec{b}) \in \Gamma^2 : \vec{a} - \vec{b} = f(\vec{\alpha},\vec{v}) \text{ for some } \vec{v} \in V_i\}.$$

(C3) For all $\vec{\alpha}_1,\dots,\vec{\alpha}_k,\vec{\alpha} \in \mathbb{F}^h$, check if $x_{\vec{\alpha}_1+\vec{\alpha},\dots,\vec{\alpha}_k+\vec{\alpha}} - x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k} = f(\vec{\alpha},\vec{t})$. That is,

$$R_{x_{\vec{\alpha}_1+\vec{\alpha},\dots,\vec{\alpha}_k+\vec{\alpha}}, x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k}} = \{(\vec{a},\vec{b}) \in \Gamma^2 : \vec{a} - \vec{b} = f(\vec{\alpha},\vec{t})\}.$$

Constraints of the form $x_{\vec{\alpha}_1,\dots,\vec{\alpha}_i+\vec{\alpha},\dots,\vec{\alpha}_k} - x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k} = f(\vec{\alpha},\vec{v})$ in (C2) are also called (C2)-i-type or (C2)-i-$\vec{\alpha}$-type constraints. Similarly, constraints of the form $x_{\vec{\alpha}_1+\vec{\alpha},\dots,\vec{\alpha}_k+\vec{\alpha}} - x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k} = f(\vec{\alpha},\vec{t})$ are called (C3)-$\vec{\alpha}$-type constraints.

Lemma 3.5. *If the k-Vector-Sum instance has a solution, then so does I. If the k-Vector-Sum instance has no solution, then there exists a constant $\epsilon > 0$ such that for every assignment to the variables of I one of the followings must hold:*

- *$\epsilon/2$-fraction of constraints of (C1) are not satisfied,*
- *there exists $i \in [k]$ such that ϵ^2-fraction of (C2)-i-type constraints are not satisfied.*
- *ϵ-fraction of constraints of (C3) are not satisfied.*

Proof. If the k-Vector-Sum instance has a solution $\vec{v}_1 \in V_1, \dots, \vec{v}_k \in V_k$, then let $x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k} = f(\vec{\alpha}_1, \vec{v}_1) + \cdots + f(\vec{\alpha}_k, \vec{v}_k)$. It is easy to check that all the constraints are satisfied.

Now suppose that the k-Vector-Sum instance has no solution. Fix any assignment $\vec{x} \in \mathbb{F}^{\ell}$. If $\epsilon/2$-fraction of constraints in (C1) are not satisfied, then we are done. Otherwise $(1-\epsilon/2)$-fraction of (C1) constraints are satisfied. By the linearity test [BLR93] and (C1), there exist $\vec{c}_1, \dots, \vec{c}_k \in \mathbb{F}^{h\ell}$ such that for $(1-\epsilon)|\mathbb{F}^{kh}|$ many choices of $(\vec{\alpha}_1, \dots, \vec{\alpha}_k) \in \mathbb{F}^{kh}$, $\vec{x}_{\vec{\alpha}_1,\dots,\vec{\alpha}_k} = F(\vec{\alpha}_1, \vec{c}_1) + F(\vec{\alpha}_2, \vec{c}_2) + \cdots + F(\vec{\alpha}_k, \vec{c}_k)$. Let

$$A = \{(\vec{\alpha}_1, \dots, \vec{\alpha}_k) \in \mathbb{F}^{kh} : \vec{x}_{\vec{\alpha}_1,\dots,\vec{\alpha}_k} = F(\vec{\alpha}_1, \vec{c}_1) + F(\vec{\alpha}_2, \vec{c}_2) + \cdots + F(\vec{\alpha}_k, \vec{c}_k)\}.$$

We have that $|A| \geq (1-\epsilon)|\mathbb{F}|^{kh}$.

Obviously, there are two cases:

- Either for every $i \in [k]$, there exists $\vec{v}_i \in V_i$ such that $\vec{c}_i = g(\vec{v}_i)$,
- or there exists $i \in [k]$ such that $\vec{c}_i \neq g(\vec{v})$ for all $\vec{v} \in V_i$.

In the later case, we will show that at least an ϵ^2-fraction of (C2)-i-type constraints are not satisfied. Call a vector $\vec{\alpha} \in \mathbb{F}^h$ good if at least $(1-\epsilon)$-fraction of (C2)-i-$\vec{\alpha}$-type constraints are satisfied. For every nonzero vector $\vec{\alpha} \in \mathbb{F}^h$, $\mathbb{F}^{kh}$ can be partitioned into two disjoint sets $X^-_{\vec{\alpha}}$ and X^+_{α} such that $X^+_{\vec{\alpha}} = \{(\vec{a}_1, \dots, \vec{a}_i + \vec{\alpha}, \dots, \vec{a}_k) : (\vec{a}_1, \dots, \vec{a}_k) \in X^-_{\vec{\alpha}}\}$. Now suppose $\vec{\alpha}$ is a good vector. There are $|\mathbb{F}^{kh}|/2$ constraints of (C2)-i-$\vec{\alpha}$-type. We construct a bipartite graph on $X^-_{\vec{\alpha}}$ and $X^+_{\vec{\alpha}}$. Two vertices $(\vec{a}_1, \dots, \vec{a}_k) \in X^-_{\vec{\alpha}}$ and $(\vec{a}_1, \dots, \vec{a}_i + \vec{\alpha}, \dots, \vec{a}_k) \in X^+_{\vec{\alpha}}$ are adjacent if $(\vec{x}_{\vec{a}_1,\dots,\vec{a}_k}, \vec{x}_{\vec{a}_1,\dots,\vec{a}_i+\vec{\alpha},\dots,\vec{a}_k})$ satisfies the constraint between them. Since $\vec{\alpha}$ is a good vector, there are $(1-\epsilon)|\mathbb{F}^{kh}|/2$ edges between $X^-_{\vec{\alpha}}$ and $X^+_{\vec{\alpha}}$. These edges form a matching M. Observe that $\min\{|A \cap X_1|, |A \cap X_2|\} \geq (1/2-\epsilon)|\mathbb{F}^{kh}|$. Since the size of matching M is $(1-\epsilon)|\mathbb{F}^{kh}|/2$, when $6\epsilon < 1$, M contains an edge whose endpoints are both in A. In other words, there exists $(\vec{a}_1, \dots, \vec{a}_k) \in A$ such that $(\vec{a}_1, \dots, \vec{a}_i + \vec{\alpha}, \dots, \vec{a}_k) \in A$ and $(\vec{x}_{\vec{a}_1,\dots,\vec{a}_i+\vec{\alpha},\dots,\vec{a}_k}, \vec{x}_{\vec{a}_1,\dots,\vec{a}_k})$ satisfies the (C2) constraint. By (C2) and Lemma 3.3, we deduce that $F(\vec{\alpha}, \vec{c}_i) = f(\vec{\alpha}, \vec{v})$ for some unique $\vec{v} \in V_i$. To summarize, for every good vector $\vec{\alpha}$, there exists a vector $\vec{v}_{\vec{\alpha}} \in V_i$ such that $F(\vec{\alpha}, \vec{c}_i) = f(\vec{\alpha}, \vec{v}_{\vec{\alpha}})$.

Next, we show that there are at most $(1-\epsilon)|\mathbb{F}^h|$ good vectors, and hence at most $(1-\epsilon) + \epsilon(1-\epsilon) = (1-\epsilon^2)$-fraction of constraints of (C2)-i-type are satisfied. Otherwise, pick an arbitrary good vector $\vec{\alpha}$. All the vectors in $\mathbb{F}^h - \{0, \vec{\alpha}\}$ can be partitioned into two sets X_1 and X_2 such that $X_1 = X_2 + \vec{\alpha}$. There exists $Y_1 \subseteq X_1$ such that Y_1 is a set of good vectors and $|Y_1| \geq (1/2-\epsilon)|\mathbb{F}^h|$. Since $\vec{c}_i \neq g(\vec{v}_{\vec{\alpha}})$, there is a set X of size at most $|\mathbb{F}|^{h-1}$ such that for all $\vec{x} \in \mathbb{F}^h \setminus X$, $F(\vec{x}, \vec{c}_i) \neq F(\vec{x}, g(\vec{v}_{\vec{\alpha}}))$. Since $(1-\epsilon)|\mathbb{F}^h|$ vectors are good, there exists a set $Z \subseteq Y_2 = Y_1 + \vec{\alpha}$ such that $|Z| \leq \epsilon|\mathbb{F}|^h$ and all the bad vectors of Y_2 are in Z. When $2\epsilon + 1/|\mathbb{F}| < 1/2$, we have $|Y_1| - |X| - |Z| > 0$. Thus there exists a vector $\vec{\alpha}' \in Y_1 - X - (Z + \vec{\alpha})$. According to the definitions, $\vec{\alpha}'$ and $\vec{\alpha} + \vec{\alpha}'$ are good and $\vec{v}_{\vec{\alpha}} \neq \vec{v}_{\vec{\alpha}'}$. Since $\vec{\alpha} + \vec{\alpha}'$ is good, there exists $\vec{u} \in V_i$ such that $F(\vec{\alpha}' + \vec{\alpha}, \vec{c}_i) = f(\vec{\alpha}' + \vec{\alpha}, \vec{u})$. Note that $\vec{u} \neq \vec{v}_{\vec{\alpha}}$, otherwise by $F(\vec{\alpha}, \vec{c}_i) = f(\vec{\alpha}, \vec{v}_{\vec{\alpha}})$, we can deduce that $F(\vec{\alpha}', \vec{c}_i) = f(\vec{\alpha}', \vec{u})$, which implies $\vec{u} = \vec{v}_{\vec{\alpha}'} \neq \vec{v}_{\vec{\alpha}}$, that is impossible. So we have $f(\vec{\alpha}' + \vec{\alpha}, \vec{u}) = F(\vec{\alpha}, \vec{c}_i) + F(\vec{\alpha}', \vec{c}_i) = f(\vec{\alpha}, \vec{v}_{\vec{\alpha}}) + f(\vec{\alpha}', \vec{v}_{\vec{\alpha}'})$, where $\vec{\alpha}, \vec{\alpha}'$ are nonzero vectors and $\vec{v}_{\vec{\alpha}}, \vec{v}_{\vec{\alpha}'}, \vec{u}$ are distinct, contradicting Lemma 3.4.

Now assume that $\vec{c}_i = g(\vec{v}_i)$ for every $i \in [k]$. Note that $\vec{v}_1 + \cdots + \vec{v}_k \neq \vec{t}$. By Lemma 3.1, $g(\sum_{i\in[k]} \vec{v}_i) \neq g(\vec{t})$. There exists a set $B \subseteq \mathbb{F}^h$ such that $|B| \geq (1-1/|\mathbb{F}|) \cdot |\mathbb{F}^h|$ and for all $\vec{\alpha} \in B$,

$$\sum_{i\in[k]} F(\vec{\alpha}, \vec{c}_i) = \sum_{i\in[k]} F(\vec{\alpha}, g(\vec{v}_i)) = F(\vec{\alpha}, \sum_{i\in[k]} g(\vec{v}_i)) \neq F(\vec{\alpha}, g(\vec{t})) = f(\vec{\alpha}, \vec{t}).$$

Notice that $|A| \geq (1-\epsilon)|\mathbb{F}|^{kh}$. We have

$$|\{(\vec{\alpha}_1, \dots, \vec{\alpha}_k, \vec{\alpha}) : (\vec{\alpha}_1, \dots, \vec{\alpha}_k), (\vec{\alpha}_1 + \vec{\alpha}, \dots, \vec{\alpha}_k + \vec{\alpha}) \in A, \vec{\alpha} \in B\}| \geq (1-1/|\mathbb{F}|) \cdot (1-2\epsilon) \cdot |\mathbb{F}|^{k(h+1)}.$$

This implies that at least $(1-1/|\mathbb{F}|)\cdot(1-\epsilon)|\mathbb{F}|^{k(h+1)} > \epsilon|\mathbb{F}|^{k(h+1)}$ constraints in (C3) are not satisfied. $\square$

Construction of the Gap-clique instance. For every $\vec{\alpha}_1,\dots,\vec{\alpha}_k \in \mathbb{F}^h$ and $\vec{\beta}_1,\dots,\vec{\beta}_k \in \mathbb{F}^h$, introduce a vertex set $V_{\vec{\alpha}_1,\dots,\vec{\alpha}_k,\vec{\beta}_1,\dots,\vec{\beta}_k} = \{(\vec{x},\vec{y},\vec{z}) : \vec{x},\vec{y},\vec{z} \in \mathbb{F}^\ell, \vec{x} = \vec{y}+\vec{z}\}$. Each vertex in this set is corresponding to an assignment to three variables $x_{\vec{\alpha}_1+\vec{\beta}_1,\dots,\vec{\alpha}_k+\vec{\beta}_k}$, $x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k}$ and $x_{\vec{\beta}_1,\dots,\vec{\beta}_k}$ which satisfies the constraint of (C1). For every variables $x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k}$ and $i \in [|\mathbb{F}|^{kh}]$, introduce a vertex set $V_{\vec{\alpha}_1,\dots,\vec{\alpha}_k,i} = \mathbb{F}^\ell$. Each vertex in $V_{\vec{\alpha}_1,\dots,\vec{\alpha}_k,i}$ is an assignment to the variable $x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k}$.

Construct a graph G' on vertice $(\bigcup_{\vec{\alpha}_1,\dots,\vec{\alpha}_k,\vec{\beta}_1,\dots,\vec{\beta}_k} V_{\vec{\alpha}_1,\dots,\vec{\alpha}_k,\vec{\beta}_1,\dots,\vec{\beta}_k}) \cup (\bigcup_{\vec{\alpha}_1,\dots,\vec{\alpha}_k,i} V_{\vec{\alpha}_1,\dots,\vec{\alpha}_k,i})$. Two vertices in G' are adjacent unless they are corresponding to inconsistent assignments or they do not satisfy the constraint between the variables they assigned to.

Lemma 3.6. *If the k-Vector-Sum instance has a solution, then the graph G' contains a clique of size $2|\mathbb{F}|^{2kh}$.*

Proof. For every vertex set, select the vertex corresponding to the assignment. According to the definition, these vertices form a clique of size $2|\mathbb{F}|^{2kh}$. $\square$

Lemma 3.7. *If the k-Vector-Sum instance has no solution, then the graph G' contains no clique of size $(1-\epsilon')2|\mathbb{F}|^{2kh}$ for some small constant $\epsilon' > 0$.*

Proof. Let ϵ be the constant in Lemma 3.5. Pick a small ϵ' such that $4\epsilon' < \min\{\epsilon/2, \epsilon^2\}$. Let X be the clique in the graph of size larger than $(1-\epsilon')2|\mathbb{F}|^{2kh}$. We have

$$|X \cap \bigcup_{\vec{\alpha}_1,\dots,\vec{\alpha}_k,\vec{\beta}_1,\dots,\vec{\beta}_k \in \mathbb{F}^h} V_{\vec{\alpha}_1,\dots,\vec{\alpha}_k,\vec{\beta}_1,\dots,\vec{\beta}_k}| \geq (1-2\epsilon')|\mathbb{F}|^{2kh}. \tag{3}$$

In addition, since $|X| > (1-\epsilon')2|\mathbb{F}|^{2kh}$, there exists an index $i \in [|\mathbb{F}^{kh}|]$ such that X contains more than $(1-2\epsilon')|\mathbb{F}^{kh}|$ vertices with respect to index i. We will prove that this is impossible using the following Claim 1.

Claim 1. For every $i \in [|\mathbb{F}^{kh}|]$, $|X \cap \bigcup_{\vec{\alpha}_1,\dots,\vec{\alpha}_k \in \mathbb{F}^h} V_{\vec{\alpha}_1,\dots,\vec{\alpha}_k,i}| \leq (1-2\epsilon')|\mathbb{F}^{kh}|$.

Proof of Claim 1. Fix an index $i \in [|\mathbb{F}|^{kh}]$. Define an assignment σ_X as follows. For every $\vec{\alpha}_1,\dots,\vec{\alpha}_k \in \mathbb{F}^h$, $\sigma_X(x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k}) = v$ if $\{v\} = X \cap V_{\vec{\alpha}_1,\dots,\vec{\alpha}_k,i}$, otherwise $\sigma_X(x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k}) = \vec{0}_\ell$. By (3) and the definition of edge set, δ_X satisfies $(1-2\epsilon')$-fraction of constraints in (C1). By Lemma 3.5, either σ_X satisfies at most $(1-4\epsilon')$-fraction of constraints in (C3) or there exists an $j \in [k]$ such that σ_X satisfies at most $(1-4\epsilon')$-fraction of constraints of (C2)-j type.

- Assume that σ_X satisfies at most $(1-4\epsilon')$-fraction of constraints in (C3). We say a vector $\vec{\alpha} \in \mathbb{F}^h$ is δ-good if more than δ-fraction of (C3)-$\vec{\alpha}$-type constraints are satisfied by σ_X. There exists a vector $\vec{\alpha} \in \mathbb{F}^h$ that is not $(1-4\epsilon')$-good, otherwise δ_X satisfies more than $(1-4\epsilon')$-fraction of constraints in (C3), contradicting our assumption. Now consider a partition $\mathbb{F}^{kh} = V^-_{\vec{\alpha}} \cup V^+_{\vec{\alpha}}$ with $V^+_{\vec{\alpha}} = V^-_{\vec{\alpha}} + (\vec{\alpha},\dots,\vec{\alpha})$. Since $\vec{\alpha}$ is not $(1-4\epsilon')$-good, there are at most $(1-4\epsilon')|\mathbb{F}^{kh}|/2$ tuples $(\vec{\alpha}_1,\dots,\vec{\alpha}_k)$ in $V^-_{\vec{\alpha}}$ such that δ_X satisfies the (C3) constraint between $x_{\vec{\alpha}_1,\dots,\vec{\alpha}_k}$ and $x_{\vec{\alpha}_1+\vec{\alpha},\dots,\vec{\alpha}_k+\vec{\alpha}}$. Let

$$X^-_{\vec{\alpha}} = X \cap \bigcup_{(\vec{\alpha}_1,\dots,\vec{\alpha}_k)\in V^-_{\vec{\alpha}}} V_{\vec{\alpha}_1,\dots,\vec{\alpha}_k,i} \text{ and } X^+_{\vec{\alpha}} = X \cap \bigcup_{(\vec{\alpha}_1,\dots,\vec{\alpha}_k)\in V^+_{\vec{\alpha}}} V_{\vec{\alpha}_1,\dots,\vec{\alpha}_k,i}$$

  It follows that $\min\{|X^-_{\vec{\alpha}}|, |X^+_{\vec{\alpha}}|\} \leq (1-4\epsilon')|\mathbb{F}|^h/2$. Thus $|X| \leq (1/2 + (1-4\epsilon')/2)|\mathbb{F}^{kh}| = (1-2\epsilon')|\mathbb{F}|^{kh}$.

- Now assume that σ_X satisfies at most $(1-4\epsilon')$-fraction of constraints of (C2)-j type for some $j \in [k]$. Similarly, for every $\vec{\alpha} \in \mathbb{F}^h$, we say $\vec{\alpha}$ is δ-good if more than δ-fraction of (C2)-j-$\vec{\alpha}$-type constraints are satisfied by σ_X. There exists $\vec{\alpha} \in \mathbb{F}^h$ that is not $(1-4\epsilon')$-good, otherwise δ_X satisfies more than $(1-4\epsilon')$-fraction of constraints of type (C2)-j, contradicting our assumption. Now consider a partition $\mathbb{F}^{kh} = V^-_{\vec{\alpha}} \cup V^+_{\vec{\alpha}}$ with

$$V^+_{\vec{\alpha}} = \{(\vec{\alpha}_1, \ldots, \vec{\alpha}_i + \vec{\alpha}, \ldots, \vec{\alpha}_k) : (\vec{\alpha}_1, \ldots, \vec{\alpha}_k) \in V^-_{\vec{\alpha}}\}.$$

Let

$$X^-_{\vec{\alpha}} = X \cap \bigcup_{(\vec{\alpha}_1,\ldots,\vec{\alpha}_k)\in V^-_{\vec{\alpha}}} V_{\vec{\alpha}_1,\ldots,\vec{\alpha}_k,i} \text{ and } X^+_{\vec{\alpha}} = X \cap \bigcup_{(\vec{\alpha}_1,\ldots,\vec{\alpha}_k)\in V^+_{\vec{\alpha}}} V_{\vec{\alpha}_1,\ldots,\vec{\alpha}_k,i}.$$

Since $\vec{\alpha}$ is not $(1-4\epsilon')$-good, we have $\min\{|X^-_{\vec{\alpha}}|, |X^+_{\vec{\alpha}}|\} \le (1-4\epsilon')|\mathbb{F}|^h/2$. Thus $|X| \le (1/2 + (1-4\epsilon')/2)|\mathbb{F}^{kh}| = (1-2\epsilon')|\mathbb{F}|^{kh}$.

□

3.1 Putting all together

For any instance (G, k) of k-Clique, we use Theorem 2.5 to reduce it to an instance $(k', V_1, \ldots, V_{k'}, \vec{t})$ of k'-Vector-Sum with $k' = \Theta(k^2)$. Then we use the reduction describe to obtain a graph G' with $h = k'^2 = \Theta(k^4)$ and small $\epsilon > 0$ in $2^{k^{O(1)}} \cdot |G|^{O(1)}$-time. By Lemma 3.6 and Lemma 3.7, we have

- if $\omega(G) \ge k$, then $\omega(G') \ge 2^{4kh+1}$,
- if $\omega(G) < k$, then with probability at least $7/10$, $\omega(G') < (1-\epsilon)2^{4kh+1}$.

Using the graph product method, we can amplify the gap to any constant.

3.2 Derandomization

To derandomize the reduction, we need to construct $O(\log n + h)$ matrices $A_1, \ldots, A_\ell \in \mathbb{F}^{h\times m}$ such that the following conditions are satisfied. For all nonzero $\vec{v} \in \mathbb{F}^m = \mathbb{F}^{h\log n}$,

$$(A_1\vec{v}, \ldots, A_\ell\vec{v}) \ne \vec{0}_{\ell h} \tag{4}$$

For all distinct $\vec{v}, \vec{u} \in V$ and nonzero $\vec{\alpha} \in \mathbb{F}^h$,

$$(\vec{\alpha}^T A_1\vec{v}, \ldots, \vec{\alpha}^T A_\ell\vec{v}) \ne (\vec{\alpha}^T A_1\vec{u}, \ldots, \vec{\alpha}^T A_\ell\vec{u}) \tag{5}$$

For all distinct $\vec{v}, \vec{u}, \vec{w} \in V$ and nonzero $\vec{\alpha}, \vec{\alpha'} \in \mathbb{F}^h$

$$(\vec{\alpha}^T A_1(\vec{v}+\vec{w}), \ldots, \vec{\alpha}^T A_\ell(\vec{v}+\vec{w})) \ne (\vec{\alpha'}^T A_1(\vec{u}+\vec{w}), \ldots, \vec{\alpha'}^T A_\ell(\vec{u}+\vec{w})) \tag{6}$$

Let $A_i \in \mathbb{F}^{h\times m}$ be the matrix such that $A_i\vec{v}$ is the projection of $\vec{v}$ onto the its subvector with coordinates between $1+(i-1)h$ and ih. Then $A_1, \ldots, A_{\log n}$ satisfy (4). It remains to construct another $O(\log n + h)$ matrices $A'_1, \ldots, A'_\ell \in \mathbb{F}^{h\times m}$ satisfying (5) and (6). Then their union $A_1, \ldots, A_{\log n}, A'_1, \ldots, A'_\ell$ would satisfy all the conditions. Note that we can think of a matrix in $\mathbb{F}^{h\times m}$ as an hm-dimension vector. The task can be formulated as given $N = |\mathbb{F}|^{2h} \cdot n^{O(1)}$ vectors $C_1, \ldots, C_N \in \mathbb{F}^{hm}$, find $O(\log n + h)$ vectors $A'_1, \ldots, A'_\ell \in \mathbb{F}^{hm}$ such that for every $i \in [N]$, there exists A'_j such that $A'_j \cdot C_i \ne 0$.

We show that, in $N^{O(1)}$-time, a vector $A \in \mathbb{F}^{hm}$ can be found such that there are at most $N/|\mathbb{F}|$ indices $i \in [N]$ satisfying $A \cdot C_i = 0$. Then we apply this algorithm $\log N/\log|\mathbb{F}|$ times to obtain the vectors $A'_1, \ldots, A'_\ell$. The vector A can be found using the method of conditional probabilities [Juk11, AS04]. Let A be a vector with $A[i]$ selected randomly and independently from $\mathbb{F}$. Define a random variable $X = |\{i \in [N] : A \cdot C_i = 0\}|$. We have $E[X] = N/|\mathbb{F}|$. For $a_1, \ldots, a_i \in \mathbb{F}$, let $X|a_1, \ldots, a_i = |\{i \in [N] : A\cdot C_i = 0, A[1] = a_1, \ldots, A[i] = a_i\}|$. We have

$$E[X|a_1, \ldots, a_i] = \sum_{x\in\mathbb{F}} E[X|a_1, \ldots, a_i, x]/|\mathbb{F}| \ge \min\{E[X|a_1, \ldots, a_i, x] : x \in \mathbb{F}\}.$$

For each $i \in [hm]$, $E[X|a_1, \ldots, a_i]$ can be computed in $N^{O(1)}$-time. For each $i \in [hm]$, we pick the value a_i to minimize $E[X|a_1, \ldots, a_i]$. We have $E[X|a_1, \ldots, a_{hm}] \le N/|\mathbb{F}|$ and the vector A with $A[i] = a_i (i \in [hm])$ is our target vector.

4 Conclusion

This paper constructs a PCP verifier which always accepts yes-instances and with probability $\Theta(1/k)$ rejects no-instances for a $W[1]$-hard problem and shows how to create a constant gap for $\omega(G)$ using this PCP. I hope that the technique of this paper will help obtain a parameterized version of PCP theorem, e.g. the Parameterized Inapproximability Hypothesis (PIH) [LRSZ20].

References

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM (JACM)*, 45(3):501–555, 1998.

[ALW13] Amir Abboud, Kevin Lewi, and Ryan Williams. On the parameterized complexity of k-sum. *CoRR, abs/1311.3054*, 2013.

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM (JACM)*, 45(1):70–122, 1998.

[AS04] Noga Alon and Joel H Spencer. *The probabilistic method*. John Wiley & Sons, 2004.

[BEKP15] Edouard Bonnet, Bruno Escoffier, Eun Jung Kim, and Vangelis Th Paschos. On subexponential and fpt-time inapproximability. *Algorithmica*, 71(3):541–565, 2015.

[BGLR93] Mihir Bellare, Shafi Goldwasser, Carsten Lund, and Alexander Russell. Efficient probabilistically checkable proofs and applications to approximations. In *Proceedings of the twenty-fifth annual ACM symposium on Theory of computing*, pages 294–304, 1993.

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of computer and system sciences*, 47(3):549–595, 1993.

[BS94] Mihir Bellare and Madhu Sudan. Improved non-approximability results. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 184–193, 1994.

[CCK+17] Parinya Chalermsook, Marek Cygan, Guy Kortsarz, Bundit Laekhanukit, Pasin Manurangsi, Danupon Nanongkai, and Luca Trevisan. From gap-ETH to FPT-inapproximability: Clique, dominating set, and more. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*, pages 743–754. IEEE, 2017.

[CGG06] Yijia Chen, Martin Grohe, and Magdalena Grüber. On parameterized approximability. In *International Workshop on Parameterized and Exact Computation*, pages 109–120. Springer, 2006.

[CL19] Yijia Chen and Bingkai Lin. The constant inapproximability of the parameterized dominating set problem. *SIAM Journal on Computing*, 48(2):513–533, 2019.

[DF95] Rodney G Downey and Michael R Fellows. Fixed-parameter tractability and completeness II: On completeness for W[1]. *Theoretical Computer Science*, 141(1-2):109–131, 1995.

[DF99] Rodney G Downey and Michael R Fellows. *Parameterized Complexity*. Springer-Verlag, 1999.

[DF13] Rodney G Downey and Michael R Fellows. *Fundamentals of parameterized complexity*, volume 4. Springer, 2013.

[Din16] Irit Dinur. Mildly exponential reduction from gap 3sat to polynomial-gap label-cover. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 23, 2016.

[FG06] Jörg Flum and Martin Grohe. *Parameterized Complexity Theory.* Springer, 2006.

[FGL+96] Uriel Feige, Shafi Goldwasser, Laszlo Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM (JACM)*, 43(2):268–292, 1996.

[FGMS] Michael R Fellows, Jiong Guo, Dániel Marx, and Saket Saurabh. Data reductions and problem kernels. In *Dahstuhl Seminar*, number 12241. Citeseer.

[FK00] Uriel Feige and Joe Kilian. Two-prover protocols—low error at affordable rates. *SIAM Journal on Computing*, 30(1):324–346, 2000.

[FKLM20] Andreas Emil Feldmann, C.S. Karthik, Euiwoong Lee, and Pasin Manurangsi. A survey on approximation in parameterized complexity: Hardness and algorithms. *Algorithms*, 13(6):146, 2020.

[Gol98] Shafi Goldwasser. Introduction to special section on probabilistic proof systems. *SIAM Journal on Computing*, 27(3):737, 1998.

[Gol16] Oded Goldreich. Lecture notes on linearity (group homomorphism) testing. 2016.

[Has96] Johan Hastad. Clique is hard to approximate within $n^{1-\epsilon}$. In *Proceedings of 37th Conference on Foundations of Computer Science*, pages 627–636. IEEE, 1996.

[HKK13] M. T. Hajiaghayi, R. Khandekar, and G. Kortsarz. Fixed parameter inapproximability for Clique and SetCover in time super-exponential in OPT. *CoRR*, abs/1310.2711, 2013.

[IPZ01] Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. Which problems have strongly exponential complexity? *Journal of Computer and System Sciences*, 63(4):512–530, 2001.

[Juk11] Stasys Jukna. *Extremal combinatorics: with applications in computer science*. Springer Science & Business Media, 2011.

[Kar72] R. M. Karp. Reducibility among combinatorial problems. In *Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York.*, pages 85–103, 1972.

[Kay14] Neeraj Kayal. Solvability of systems of polynomial equations over finite fields. *A talk given by Neeraj Kayal at the Simons Institute for the Theory of Computing, Berkeley, CA [Accessed: 2017/20/7]*, page 1, 2014.

[KLN21] C.S. Karthik and Inbal Livni-Navon. On hardness of approximation of parameterized set cover and label cover: Threshold graphs from error correcting codes. *SOSA*, 2021.

[KS16] Subhash Khot and Igor Shinkar. On hardness of approximating the parameterized clique problem. In *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pages 37–45, 2016.

[Lin18] Bingkai Lin. The parameterized complexity of the k-biclique problem. *Journal of the ACM (JACM)*, 65(5):1–23, 2018.

[Lin19] Bingkai Lin. A simple gap-producing reduction for the parameterized set cover problem. In *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[LKM19] Bundit Laekhanukit, C.S. Karthik, and Pasin Manurangsi. On the parameterized complexity of approximating dominating set. *Journal of the ACM (JACM)*, 66(5):33, 2019.

[LRSZ20] Daniel Lokshtanov, MS Ramanujan, Saket Saurab, and Meirav Zehavi. Parameterized complexity and approximability of directed odd cycle transversal. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2181–2200. SIAM, 2020.

[Mar08] Dániel Marx. Parameterized complexity and approximation algorithms. *The Computer Journal*, 51(1):60–78, 2008.

[MR16] Pasin Manurangsi and Prasad Raghavendra. A birthday repetition theorem and complexity of approximating dense csps. *arXiv preprint arXiv:1607.02986*, 2016.

[Wło20] Michał Włodarczyk. Parameterized inapproximability for steiner orientation by gap amplification. In *47th International Colloquium on Automata, Languages, and Programming (ICALP 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

[Zuc06] David Zuckerman. Linear degree extractors and the inapproximability of max clique and chromatic number. In *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pages 681–690, 2006.

Appendix

Theorem 4.1 (Theorem 2.5 restated)**.** *k-Vector-Sum with $\mathbb{F} = \mathbb{F}_2$ and $m = \Theta(k^2 \log n)$ is $W[1]$-hard parameterized by k.*

Proof. We construct a reduction from k-Multi-Color-Clique to $(k + \binom{k}{2})$-Vector-Sum. Let (G, k) be an instance of k-Multi-Color-Clique with $V(G) = V_1 \cup V_2 \cup \cdots \cup V_k$. Set $n = |V(G)| + 1$. For every $v \in V(G)$, let $\sigma(v) \in \{0,1\}^{\log n}$ be the binary encoding of v. Since $n > V(G)$, we can assume that $\sigma(v) \neq \vec{0}_{\log n}$ for all $v \in V(G)$. Let $f : \binom{[k]}{2} \to [\binom{k}{2}]$ be a bijection. For every $i \in [k]$ and $v \in V(G)$, let

$$\vec{\eta}_{v,i} = (\underbrace{\overbrace{0, \cdots, 0}^{(i-1)\log n}, \sigma(v), 0, \cdots, 0}_{k \log n}) \text{ and } \vec{\gamma}_{v,i} = (\underbrace{\overbrace{\sigma(v), \ldots, \sigma(v)}^{(i-1)\log n}, \vec{0}_{\log n}, \sigma(v), \ldots, \sigma(v)}_{k \log n})$$

For any distinct $i, j \in [k]$, let

$$\vec{\theta}_{i,j} = (\underbrace{\overbrace{0, \cdots, 0}^{f(\{i,j\})-1}, 1, 0, \cdots, 0}_{k(k-1)/2}) \text{ and } \vec{\delta}_i = (\underbrace{\overbrace{0, \cdots, 0}^{i-1}, 1, 0, \cdots, 0}_{k}).$$

For every edge $e = \{v, u\}$ with $v \in V_i$ and $u \in V_j$, let

$$\vec{w}_e = \vec{0}_k \circ \vec{\theta}_{i,j} \circ (\underbrace{\overbrace{\overbrace{0, \ldots, 0}^{(i-1)k\log n}, \vec{\eta}_{v,j}, 0, \ldots, 0}^{(j-1)k\log n}, \vec{\eta}_{u,i}, 0, \ldots, 0}_{k^2 \log n}).$$

For every $v \in V_i$, let

$$\vec{w}_v = \vec{\delta}_i \circ \vec{0}_{k(k-1)/2} \circ (\underbrace{\overbrace{0, \ldots, 0}^{(i-1)k\log n}, \vec{\gamma}_{v,i}, 0, \ldots, 0}_{k^2 \log n})$$

The instance of vector sum is defined as follows.

- The target vector is $\vec{t} = \vec{1}_{k+k(k-1)/2} \circ \vec{0}_{k^2 \log n}$.
- There are $k(k-1)/2 + k$ sets of vectors.
  - For every $\{i, j\} \in \binom{[k]}{2}$, let
  $$W_{ij} = \{\vec{w}_e : e = \{v, u\} \text{ is an edge in } G \text{ with } v \in V_i \text{ and } u \in V_j\}.$$
  - For every $i \in [k]$, let
  $$W_i = \{\vec{w}_v : v \in V_i\}.$$

If (G, k) is a yes-instance, then there exist $v_1 \in V_1, \ldots, v_k \in V_k$ such that $\{v_1, \ldots, v_k\}$ induces a k-clique in G. It is easy to check that the sum of $\vec{w}_{v_i v_j}$'s and $\vec{w}_{v_i}$'s is equal to $\vec{t}$.

On the other hand, if there exist $\vec{w}_{ij} \in W_{ij}$ and $\vec{w}_i \in W_i$ such that

$$\sum \vec{w}_{ij} + \sum \vec{w}_i = \vec{t}.$$

Each w_i is corresponding to a vertex $v_i \in V_i$. Each w_{ij} is corresponding to an edge e_{ij} between V_i and V_j. It is easy to see that v_i is an endpoint of e_{ij} for all $j \in [k] \setminus \{i\}$. Therefore they form a clique of size k. □