Theoretical Computer Science 411 (2010) 2513–2519

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

# Inapproximability results for equations over infinite groups

WenBin Chen [a,*], Dengpan Yin [a], Zhengzhang Chen [b]

[a] Department of Computer Science, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, PR China

[b] Department of Computer Science, North Carolina State University, Raleigh, NC 27695, United States

ARTICLE INFO

*Article history:*
Received 20 June 2008
Received in revised form 13 February 2010
Accepted 2 March 2010
Communicated by D.-Z. Du

*Keywords:*
Optimization
Approximation
Groups
Infinite groups
Cyclic groups
Computational complexity
Probabilistically checkable proofs
NP-hardness

ABSTRACT

An *equation* over a group $G$ is an expression of form $w_1 \dots w_k = 1_G$, where each $w_i$ is either a variable, an inverted variable, or a group constant and $1_G$ denotes the identity element; such an equation is *satisfiable* if there is a setting of the variables to values in $G$ such that the equality is realized (Engebretsen et al. (2002) [10]).

In this paper, we study the problem of simultaneously satisfying a family of equations over an infinite group $G$. Let $\mathrm{EQ}_G[k]$ denote the problem of determining the maximum number of simultaneously satisfiable equations in which each equation has occurrences of exactly $k$ different variables. When $G$ is an infinite cyclic group, we show that it is NP-hard to approximate $\mathrm{EQ}^1_G[3]$ to within $48/47 - \epsilon$, where $\mathrm{EQ}^1_G[3]$ denotes the special case of $\mathrm{EQ}_G[3]$ in which a variable may only appear once in each equation; it is NP-hard to approximate $\mathrm{EQ}^1_G[2]$ to within $30/29 - \epsilon$; it is NP-hard to approximate the maximum number of simultaneously satisfiable equations of degree at most $d$ to within $d-\epsilon$ for any $\epsilon$; for any $k \geq 4$, it is NP-hard to approximate $\mathrm{EQ}_G[k]$ within any constant factor. These results extend Håstad's results (Håstad (2001) [17]) and results of (Engebretsen et al. (2002) [10]), who established the inapproximability results for equations over finite Abelian groups and any finite groups respectively.



## 1. Introduction

Since Feige et al. found the links between probabilistic proof systems and inapproximability [12], there has been a lot of work in studying the inapproximability of NP-optimization problems. Since the discovery of the PCP theorem, the inapproximability theory has seen much progress. The inapproximability results for some NP-complete problems have been proven. For instance, Feige proved that approximating set cover to within $\ln n$ is NP-hard [11]. Håstad proved that it is NP-hard to approximate clique problems within $n^{1-\epsilon}$ [16] and it is NP-hard to approximate MAX-3SAT within $8/7 - \epsilon$ [17].

In this paper we study the simultaneous solvability of families of equations over infinite groups, which extends the study of the simultaneous solvability of families of equations over finite groups. Many natural combinatorial optimization problems can be described as questions concerning the simultaneous solvability of families of equations over finite groups. There has been much work on this study. Many strong inapproximability results for problems such as Max Cut, Max Di-Cut, Exact Satisfiability, and Vertex Cover [7–9,15,17,18,20,23] can be obtained from the connection. In [17], Håstad has proved that it is NP-hard to approximate maximum simultaneously satisfiable equations over a finite *Abelian* group $G$. Later, the result is extended to all finite groups [10].

We give some definitions from [10]. An *equation* in variables $x_1, \dots, x_n$ over a group $G$ is an expression of the form $w_1 \dots w_k = 1_G$, where each $w_i$ is either a variable, an inverted variable, or a group constant and $1_G$ denotes the identity

* Corresponding author. Tel.: +86 25 84892801.
*E-mail addresses:* cwbiscas@yahoo.com (W. Chen), yindengpan@gmail.com (D. Yin), zhengzhang.chen@gmail.com (Z. Chen).

0304-3975/$ – see front matter 
doi:10.1016/j.tcs.2010.03.010

element. A *solution* is an assignment of the variables to values in $G$ that realizes the equality. A collection of equations $\mathcal{E}$ over the same variables induces a natural optimization problem, the problem of determining the maximum number of simultaneously satisfiable equations in $\mathcal{E}$. We let $EQ_G$ denote this optimization problem. The special case where a variable may only appear once in each equation is denoted as $EQ_G^1$; when each equation has single occurrences of exactly $k$ variables, the problem is denoted as $EQ_G^1[k]$; when a variable appears at most $d$ times in each equation, this case is denoted as $EQ_G^d$; at the same time, if each equation has single occurrence of exactly $k$ variables, the problem is denoted as $EQ_G^d[k]$. For example, the equation $ae^{-1}abebcb^{-1}cead = 1_G$ could be part of an $EQ_G^3[5]$ instance because five distinct variables occur and no variable occurs more that three times.

A number of familiar optimization problems are related to $EQ_G$. For instance, when $G = \mathbf{Z_2}$, the familiar Max Cut problem corresponds to the instances of $EQ_{\mathbf{Z_2}}^1[2]$, i.e. exactly two variables occur in each clause. For when $G = S_3$, the non-Abelian symmetric group on three letters, M. Goldmann and A. Russell reduce the problem of maximizing the number of bichromatic edges in a 3-coloring of a given graph to $EQ_G$ [14]. Håstad [17] and Zwick [23] describe some other examples. The general problem has also been studied by Barrington et al. [5]. They have studied the computational complexity of determining whether a system of equations over a fixed finite monoid has a solution. Finally, a number of well-studied combinatorial enumeration problems can be produced from the $EQ_G$ problem: see, e.g., [6,13,21,22]).

For $G$ an Abelian group, Håstad [17] proved that it is NP-hard to approximate $EQ_G^1[3]$ to within $|G| - \epsilon$ for any $\epsilon > 0$. Engebretsen et al. extended the result to all finite groups. They proved that if $P \neq NP$ and $G$ is any finite group, no polynomial time approximation algorithm can approximate $EQ_G^1[3]$ to within any $|G| - \epsilon$ for any $\epsilon > 0$ [10].

For $EQ_G^1$ over a finite group, the trivial randomized approximation algorithm which independently assigns each variable to a uniformly selected value in $G$ satisfies an expected fraction $|G|^{-1}$ of the equations. According to the method of conditional expectation [13], this algorithm can be efficiently derandomized.

For when $G$ is the $\mathbf{R}$ field, Amaldi et al. [1] and Arora et al. [3] all study the complexity of the $EQ_{\mathbf{R}}$ problem. They proved that it is NP-hard to approximate $EQ_{\mathbf{R}}$ within any constant factor and within a factor $n^{\epsilon}$ for some $\epsilon > 0$, where $n$ is the number of equations.

To our knowledge, nobody has studied the complexity of simultaneously satisfying a family of equations over an infinite group $G$. Since the structure of an infinite group is not obvious, we consider the simple situation where $G$ is an infinite cyclic group in this paper. We obtain some complexity results which are shown as follows.

**Our result**

In this paper, we show that, when $G$ is an infinite cyclic group, it is NP-hard to approximate $EQ_G^1[3]$ to within $48/47 - \epsilon$; it is NP-hard to approximate $EQ_G^1[2]$ to within $30/29 - \epsilon$; it is NP-hard to approximate $EQ_G[k]$ within any constant factor; it is NP-hard to approximate $EQ_G^d[k]$ to within $d - \epsilon$ for any $\epsilon$ ($k \geq 4$).

**Technique**

Since an infinite cyclic group $G$ is isomorphic to the additive integer group $\mathbf{Z}$, we study the problem $EQ_{\mathbf{Z}}$. In order to obtain the inapproximability result of $EQ_G^1[3]$, we give a gap-preserving reduction from Max-E3-Sat to $EQ_G^1[3]$. Similarly, we give a gap-preserving reduction from Max-E2-Sat to $EQ_G^1[2]$.

We give a polynomial time reduction from $EQ_{\mathbf{Z}_d}^1[3]$ to $EQ_{\mathbf{Z}}^d[4]$, where $\mathbf{Z}_d$ is the cyclic group with $d$ elements. Since it is NP-hard to approximate $EQ_{\mathbf{Z}_d}^1[3]$ within $d - \epsilon$ [17], it is NP-hard to approximate $EQ_{\mathbf{Z}}^d[4]$ within $d - \epsilon$. Similarly, we can give a reduction from $EQ_{\mathbf{Z}_d}^1[k]$ to $EQ_{\mathbf{Z}}^d[k+1]$ ($k > 4$).

**Structure of the paper**

In Section 2, we introduce some definitions. Section 3 describes the reduction from the Max-E3-Sat problem to $EQ_{\mathbf{Z}}^1[3]$, establishing the hardness of approximating $EQ_{\mathbf{Z}}^1[3]$. Section 4 describes the reduction from the Max-E2-Sat problem to $EQ_{\mathbf{Z}}^1[2]$, establishing the hardness of approximating $EQ_{\mathbf{Z}}^1[2]$. In Section 5 we prove that approximating $EQ_{\mathbf{Z}}^d[4]$ is NP-hard within $d - \epsilon$ by reducing $EQ_{\mathbf{Z}_d}^1[3]$ to it. Finally, in Section 6 we present some conclusions and some open problems.

## 2. Preliminaries

We briefly introduce some notation (see [4]).

**Definition 1.** An *optimization problem* $\Pi$ is a set $\mathcal{I} \subseteq \{0, 1\}^*$, a set $\mathcal{S} \subseteq \{0, 1\}^*$ of feasible solutions on input $I \in \mathcal{I}$, and a polynomial time computable measure $m : \mathcal{I} \times \mathcal{S} \to \mathbf{R}_+$, that assigns to each tuple of instance $I$ and solution $S$ a positive real number $m(I, S)$, called the *value* of the solution $S$. The optimization problem is to find, for a given input $I \in \mathcal{I}$, a solution $S \in \mathcal{S}$ such that $m(I, S)$ is optimum over all possible $S \in \mathcal{S}$.

If the optimum is $min_{S \in \mathcal{S}}\{m(I, S)\}$ (resp. $max_{S \in \mathcal{S}}\{m(I, S)\}$), we refer to $\Pi$ as a minimization (resp. maximization) problem.

**Definition 2.** For an input $I$ of a maximization problem $\Pi$ whose optimum solution has value $opt(I)$, an algorithm $A$ is said to *approximate* $opt(I)$ *within a factor* $f(I)$ iff

$$1 \leq \frac{opt(I)}{A(I)} \leq f(I),$$

where $f(I) \geq 1$ and $A(I) > 0$.

For studying the hardness of approximation problems we introduce the following reduction due to Arora [2].

**Definition 3.** Let $\Pi$ and $\Pi'$ be two maximization optimization problems and $\rho, \rho' \geq 1$. A *gap-preserving reduction from* $\Pi$ *to* $\Pi'$ *with parameters* $((c, \rho), (c', \rho'))$ is a polynomial transformation $\tau$ mapping every instance $I$ of $\Pi$ to an instance $I' = \tau(I)$ of $\Pi'$ such that for the optima $opt_\Pi(I)$ and $opt_{\Pi'}(I')$ of $I$ and $I'$, respectively, the following hold:

$$opt_\Pi(I) \geq c \Longrightarrow opt_{\Pi'}(I') \geq c'$$
$$opt_\Pi(I) \leq c/\rho \Longrightarrow opt_{\Pi'}(I') \leq c'/\rho',$$

where $c$, $\rho$ and $c'$, $\rho'$ depend on the instance sizes $I$ and $I'$, respectively.

## 3. The hardness of approximating $EQ^1_{\mathbf{Z}}[3]$

In this section, we show that it is NP-hard to approximate $EQ^1_{\mathbf{Z}}[3]$ within $48/47 - \epsilon$. The proof is by gap-preserving Given a finite set $X$ of variables and a set $C = \{C_1, \ldots, C_m\}$ of disjunctive clauses with exactly three literals in each clause, find a truth assignment for $X$ that satisfies as many clauses of $C$ as possible. In the famous paper [17], Håstad proved the following lemmas.

**Lemma 2** (*Theorem 6.1 of [17]*). *For any $\epsilon$ it is NP-hard to approximate Max-E3-Sat within a factor $8/7 - \epsilon$.*

**Lemma 3** (*Theorem 6.5 of [17]*). *For any $\epsilon$ it is NP-hard to distinguish satisfiable E3-CNF formulas from $(7/8 + \epsilon)$-satisfiable E3-CNF formulas.*

In the following, we show that it is NP-hard to approximate $EQ^1_{\mathbf{Z}}[3]$ within $48/47 - \epsilon$.

**Theorem 4.** *It is NP-hard to approximate $EQ^1_{\mathbf{Z}}[3]$ within $48/47 - \epsilon$ for any $\epsilon$.*

**Proof.** We give a gap-preserving reduction from Max-E3-Sat to $EQ^1_{\mathbf{Z}}[3]$. Let $(X, C)$ with $C = \{C_1, \ldots, C_m\}$ be an arbitrary instance of Max-E3-Sat. For each clause $C_i$, $1 \leq i \leq m$, containing three variables $x_{i1}$, $x_{i2}$ and $x_{i3}$, we construct the following equations:

$$a_{i1}x_{i1} + a_{i2}x_{i2} + a_{i3}x_{i3} = 3 \quad (5)$$
$$a_{i1}x_{i1} + a_{i2}x_{i2} + a_{i3}x_{i3} = 1 \quad (6)$$
$$a_{i1}x_{i1} + a_{i2}x_{i2} + a_{i3}x_{i3} = -1 \quad (7)$$
$$x_{i1} + y_i + z_i = 1 \quad (8)$$
$$x_{i1} + y_i + z_i = -1 \quad (9)$$
$$x_{i2} + y_i + z_i = 1 \quad (10)$$
$$x_{i2} + y_i + z_i = -1 \quad (11)$$
$$x_{i3} + y_i + z_i = 1 \quad (12)$$
$$x_{i3} + y_i + z_i = -1 \quad (13)$$
$$y_i + z_i + w_i = 0 \quad (14)$$
$$y_i + z_i - w_i = 0 \quad (15)$$

where $a_{ij} = 1$ if $x_{ij}$ occurs positively in $C_i$ and $a_{ij} = -1$ if $x_{ij}$ occurs negatively ($j = 1, 2, 3$). Thus we have a system with $11m$ equations.

Given a truth assignment which satisfies $s$ clauses of Max-E3-Sat, we immediately obtain a solution $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w})$ that satisfies $5m + s$ equations of the above $EQ^1_{\mathbf{Z}}[3]$ instance. This is simply achieved by setting the variables $x_j$ to 1 if the corresponding boolean variable is TRUE in the assignment and otherwise setting $x_j$ to $-1$ and setting $y_i = z_i = w_i = 0$.

Consider any solution $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w})$ of the above $EQ^1_{\mathbf{Z}}[3]$ instance. For each $i$, $1 \leq i \leq m$, at most six equations can be simultaneously satisfied: at most one of (5)–(7) and at most five of (8)–(15). If any component of $\mathbf{x}$ is neither 1 nor $-1$, we can set it to 1; if $w_i$ is not 0 and $y_i+z_i$ is not 0, we can set them to 0; for old values of $y_i, z_i$, since $y_i+z_i$ is not 0, at most one of (14) and (15) is satisfied. But for new values $y_i = z_i = w_i = 0$, (14) and (15) are both satisfied. Thus, for (14) and (15), at least one of the equations satisfied is added. But for the new values of $x_i$, it is possible that the equation of (5)–(7) satisfied becomes unsatisfied. However, the number is at most 1. Hence, new values do not decrease the number of equations satisfied. Thus, we can assume without loss of generality that any component of $\mathbf{x}$ is either 1 or $-1$ and $y_i = z_i = w_i = 0$. Suppose that the solution $(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w})$ satisfies $5m + s$ equations. We set the corresponding boolean variable to TRUE if $x_j = 1$ and otherwise to FALSE. It is easy to see that the alignment satisfies $s$ clauses of the Max-E3-Sat. Consequently, we have a correspondence between solutions of the Max-E3-Sat instance satisfying $s$ clauses and solutions of the $EQ^1_{\mathbf{Z}}[3]$ instance fulfilling $5m + s$ equations. Thus we get

$$opt_{Max\text{-}E3\text{-}Sat}(I) \geq c \Longrightarrow opt_{EQ^1_{\mathbf{Z}}[3]}(I') \geq 5m + c$$
$$opt_{Max\text{-}E3\text{-}Sat}(I) \leq c/\rho \Longrightarrow opt_{EQ^1_{\mathbf{Z}}[3]}(I') \leq 5m + c/\rho,$$

where $I$ is the instance of Max-E3-Sat and $I'$ is the instance of $EQ^1_{\mathbf{Z}}[3]$.

Since it is NP-hard to distinguish satisfiable E3-CNF formulas from $(7/8 + \epsilon)$-satisfiable E3-CNF formulas by Lemma 3, it is NP-hard to distinguish the $EQ^1_{\mathbf{Z}}[3]$ instance fulfilling $6m$ equations and the $EQ^1_{\mathbf{Z}}[3]$ instance fulfilling $5m + (7/8 + \epsilon)m$ equations. Thus the inapproximability factor is $6m/(5m + (7/8 + \epsilon)m) < 48/47 - \epsilon$. □

## 4. The hardness of approximating $EQ^1_{\mathbf{Z}}[2]$

In this section, we show that it is NP-hard to approximate $EQ^1_{\mathbf{Z}}[2]$ within $30/29 - \epsilon$. The proof method is similar to the method in Theorem 4. The proof is by gap-preserving reduction from the known Max-E2-Sat problem that is defined as follows. Given a finite set $X$ of variables and a set $C = \{C_1, \ldots, C_m\}$ of disjunctive clauses with exactly two literals in each clause, find a truth assignment for $X$ that satisfies as many clauses of $C$ as possible. By the Lemma 5.13 and theorem 6.16 of [17], we have the following conclusion:

**Lemma 4.** *For any $\epsilon$ it is NP-hard to distinguish $(11 - \epsilon)$-satisfiable E2-CNF formulas from $(11 - \frac{1}{2} + \epsilon)$-satisfiable E2-CNF formulas.*

In the following, we show that it is NP-hard to approximate $EQ^1_{\mathbf{Z}}[2]$ within $30/29 - \epsilon$. The following reduction is similar to the reduction from Max-2-Sat to MAX $FLS^=$ given in [1].

**Theorem 5.** *It is NP-hard to approximate $EQ^1_{\mathbf{Z}}[2]$ within $30/29 - \epsilon$ for any $\epsilon$.*

**Proof.** We give a gap-preserving reduction from Max-E2-Sat to $EQ^1_{\mathbf{Z}}[2]$. Let $(X, C)$ with $C = \{C_1, \ldots, C_m\}$ be an arbitrary instance of Max-E2-Sat. For each clause $C_i$, $1 \leq i \leq m$, containing two variables $x_{i1}$ and $x_{i2}$, we construct the following equations:

$$a_{i1}x_{i1} + a_{i2}x_{i2} = 2 \quad (16)$$
$$a_{i1}x_{i1} + a_{i2}x_{i2} = 0 \quad (17)$$
$$x_{i1} + y_i = 1 \quad (18)$$
$$x_{i1} + y_i = -1 \quad (19)$$
$$x_{i2} + y_i = 1 \quad (20)$$
$$x_{i2} + y_i = -1 \quad (21)$$
$$y_i + z_i = 0 \quad (22)$$
$$y_i - z_i = 0 \quad (23)$$

where $a_{ij} = 1$ if $x_{ij}$ occurs positively in $C_i$ and $a_{ij} = -1$ if $x_{ij}$ occurs negatively $(j = 1, 2)$. Thus we have a system with $8m$ equations.

Given a truth assignment which satisfies $s$ clauses of Max-E2-Sat, we immediately obtain a solution $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ that satisfies $4m + s$ equations of the above $EQ^1_{\mathbf{Z}}[2]$ instance. This is simply achieved by setting the variables $x_j$ to 1 if the corresponding boolean variable is TRUE in the assignment and otherwise setting $x_j$ to $-1$ and setting $y_i = z_i = 0$.

Consider any solution $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ of the above $EQ^1_{\mathbf{Z}}[2]$ instance. For each $i$, $1 \leq i \leq m$, at most five equations can be simultaneously satisfied: at most one of (16)–(17) and at most four of (18)–(23). If any component of $\mathbf{x}$ is neither 1 nor $-1$, we can set it to 1; if $z_i$ is not 0 and $y_i$ is not 0, we can set them to 0; that doesn't decrease the number of satisfied equations. Consequently, we have a correspondence between solutions of the Max-E2-Sat instance satisfying $s$ clauses and solutions of the $EQ^1_{\mathbf{Z}}[2]$ instance fulfilling $4m + s$ equations.

Since it is NP-hard to distinguish $(11 - \epsilon)$-satisfiable E2-CNF formulas from $(11 - \frac{1}{2} + \epsilon)$-satisfiable E2-CNF formulas by Lemma 4, it is NP-hard to distinguish the $EQ^1_{\mathbf{Z}}[2]$ instance fulfilling $4m + (11 - \epsilon)m$ equations from the $EQ^1_{\mathbf{Z}}[2]$ instance fulfilling $4m + (11 - \frac{1}{2} + \epsilon)m$ equations. Thus the inapproximability factor is $\frac{4m+(11-\epsilon)m}{4m+(11-\frac{1}{2}+\epsilon)m} < 30/29 - \epsilon$. □

## 5. The hardness of approximating $EQ_G^d[4]$

Since an infinite cyclic group $G$ is isomorphic to the additive integer group $\mathbf{Z}$, we study the problem $EQ_{\mathbf{Z}}$. In this section, we shall prove that approximating $EQ_{\mathbf{Z}}^d[4]$ is NP-hard within $d-\epsilon$ by reducing $EQ_{\mathbf{Z}_d}^1[3]$ to it. Using the PCP theorem and Raz's parallel repetition theorem [19], Håstad proved the following lemma [17].

**Lemma 1.** *It is NP-hard to approximate $EQ_{\mathbf{Z}_d}^1[3]$ within $d-\epsilon$ for any $\epsilon$.*

In the following, we give a gap-preserving reduction from $EQ_{\mathbf{Z}_d}^1[3]$ to $EQ_{\mathbf{Z}}^d[4]$.

**Theorem 1.** *It is NP-hard to approximate $EQ_{\mathbf{Z}}^d[4]$ within $d-\epsilon$ for any $\epsilon$.*

**Proof.** Suppose the following system of equations is an instance $I$ of $EQ_{\mathbf{Z}_d}^1[3]$:

$$\begin{aligned} &x_{11}+x_{12}+x_{13}=b_1 \bmod (d)\\ &\vdots\\ &x_{m1}+x_{m2}+x_{m3}=b_m \bmod (d) \end{aligned} \tag{1}$$

We construct an instance $I'$ of $EQ_{\mathbf{Z}}^d[4]$ as follows:

$$\begin{aligned} &x_{11}+x_{12}+x_{13}+d\cdot y_1=b_1\\ &\vdots\\ &x_{m1}+x_{m2}+x_{m3}+d\cdot y_m=b_m \end{aligned} \tag{2}$$

We claim that the reduction is gap-preserving. If a single equation $x_{i1}+x_{i2}+x_{i3}=b_i \bmod (d)$ has a solution, then there is a $y$ such that $x_{i1}+x_{i2}+x_{i3}+d\cdot y=b_i$. So the single equation $x_{i1}+x_{i2}+x_{i3}+d\cdot y_i=b_i$ also has a solution. Thus the number of simultaneously satisfiable equations of (1) is equal to the number of simultaneously satisfiable equations of (2). So $opt(I)=opt(I')$. Hence the reduction is gap-preserving. By Lemma 1, it is NP-hard to approximate $EQ_{\mathbf{Z}_d}^1[3]$ within $d-\epsilon$ for any $\epsilon$. So it is NP-hard to approximate $EQ_{\mathbf{Z}}^d[4]$ within $d-\epsilon$ for any $\epsilon$. □

Similarly, for any $k>4$, we can give a gap-preserving reduction from $EQ_{\mathbf{Z}_d}^1[k]$ to $EQ_{\mathbf{Z}}^d$[k+1]. Since it is NP-hard to approximate $EQ_{\mathbf{Z}_d}^1[k]$ within $d-\epsilon$ for any $k\geq 4$, we come to the following conclusion:

**Theorem 2.** *For any $k>4$, it is NP-hard to approximate $EQ_{\mathbf{Z}}^d[k]$ within $d-\epsilon$.*

By the Theorem 2, we can obtain the following conclusion.

**Theorem 3.** *For any $k\geq 4$, it is NP-hard to approximate $EQ_{\mathbf{Z}}[k]$ within any constant factor.*

**Proof.** For any constant $d$, the set of instances of $EQ_{\mathbf{Z}}^d[k]$ is a subset of the set of instances of $EQ_{\mathbf{Z}}[k]$, so it is NP-hard to approximate $EQ_{\mathbf{Z}}[k]$ within $d-\epsilon$ for any $k\geq 4$ by Theorems 1 and 2. Since $d$ is any constant, it is NP-hard to approximate $EQ_{\mathbf{Z}}[k]$ within any constant factor. □

On the basis of the result of Theorem 3, we have some corollaries.

**Corollary 1.** *If $G\approx \mathbf{Z}\times\mathbf{Z}$, where $G$ is isomorphic to $\mathbf{Z}\times\mathbf{Z}$, then it is NP-hard to approximate $EQ_G[k]$ within any constant factor for any $k\geq 4$.*

**Proof.** We give a gap-preserving reduction from $EQ_{\mathbf{Z}}[k]$ to $EQ_G[k]$. Suppose the following system of equations is an instance $I$ of $EQ_{\mathbf{Z}}[k]$:

$$\begin{aligned} &a_{11}x_{11}+\cdots+a_{1k}x_{1k}=b_1\\ &\vdots\\ &a_{m1}x_{m1}+\cdots+a_{mk}x_{mk}=b_m \end{aligned} \tag{3}$$

We construct an instance $I'$ of $EQ_G[k]$ as follows:

$$\begin{aligned} &a_{11}<x_{11},y_{11}>+\cdots+a_{1k}<x_{1k},y_{1k}>=<b_1,b_1>\\ &\vdots\\ &a_{m1}<x_{m1},y_{m1}>+\cdots+a_{mk}<x_{mk},y_{mk}>=<b_m,b_m> \end{aligned} \tag{4}$$

Now we show that the reduction is gap-preserving.

If $opt(I) \geq c$, then there is a solution $\mathbf{x}$ satisfying at least $c$ equations of (3). So $< \mathbf{x}, \mathbf{x} >$ also satisfy at least $c$ equations of (4). Thus $opt(I') \geq c$.

Suppose $opt(I) \leq c/\rho$. When a single equation $a_{i1} < x_{i1}, y_{i1} > + \cdots + a_{1k} < x_{ik}, y_{ik} > = < b_i, b_i >$ is satisfiable, the single equation $a_{i1}x_{i1} + \cdots + a_{ik}x_{ik} = b_i$ is also satisfiable. So $opt(I') \leq opt(I) \leq c/\rho$. So the reduction is gap-preserving. Thus by Theorem 3, it is NP-hard to approximate $EQ_G[k]$ within any constant factor for any $k \geq 4$. □

By a method similar to the above, we come to the following conclusions.

**Corollary 2.** *If $G \approx \mathbf{Z} \times \mathbf{Z}_d$, where $G$ is isomorphic to $\mathbf{Z} \times \mathbf{Z}_d$, then it is NP-hard to approximate $EQ_G[k]$ within any constant factor for any $k \geq 4$.*

**Proof.** We give a gap-preserving reduction from $EQ_{\mathbf{Z}}[k]$ to $EQ_G[k]$. Suppose the following systems of equation is an instance $I$ of $EQ_{\mathbf{Z}}[k]$:

$$\begin{aligned} &a_{11}x_{11} + \cdots + a_{1k}x_{1k} = b_1 \\ &\vdots \\ &a_{m1}x_{m1} + \cdots + a_{mk}x_{mk} = b_m \end{aligned} \tag{3'}$$

We construct an instance $I'$ of $EQ_G[k]$ as follows:

$$\begin{aligned} &a_{11} < x_{11}, y_{11} > + \cdots + a_{1k} < x_{1k}, y_{1k} > = < b_1, 0 > \\ &\vdots \\ &a_{m1} < x_{m1}, y_{m1} > + \cdots + a_{mk} < x_{mk}, y_{mk} > = < b_m, 0 > \end{aligned} \tag{4'}$$

Now we show that the reduction is gap-preserving.

If $opt(I) \geq c$, then there is a solution $\mathbf{x}$ satisfying at least $c$ equations of (3′). So $< \mathbf{x}, \mathbf{0} >$ also satisfies at least $c$ equations of (4′). Thus $opt(I') \geq c$.

Suppose $opt(I) \leq c/\rho$. When a single equation $a_{i1} < x_{i1}, y_{i1} > + \cdots + a_{1k} < x_{ik}, y_{ik} > = < b_i, 0 >$ is satisfiable, the single equation $a_{i1}x_{i1} + \cdots + a_{ik}x_{ik} = b_i$ is also satisfiable. So $opt(I') \leq opt(I) \leq c/\rho$. So the reduction is gap-preserving. Thus by Theorem 3, it is NP-hard to approximate $EQ_G[k]$ within any constant factor for any $k \geq 4$. □

Similarly, we can obtain the following conclusion.

**Corollary 3.** *If $G \approx \mathbf{Z} \times G_1$, where $G_1$ is a finite group, then it is NP-hard to approximate $EQ_G[k]$ within any constant factor for any $k \geq 4$.*

## 6. Conclusion

In this paper, we show that when $G$ is an infinite cyclic group, it is NP-hard to approximate $EQ_G^1[3]$ to within $48/47 - \epsilon$; it is NP-hard to approximate $EQ_G^1[2]$ to within $30/29 - \epsilon$; it is NP-hard to approximate $EQ_G[k]$ within any constant factor; it is NP-hard to approximate $EQ_G^d[k]$ to within $d - \epsilon$ for any $\epsilon$ ($k \geq 4$).

We guess that it is NP-hard to approximate $EQ_G^1[k]$ within any constant factor if $G$ is a infinite group. Another open problem is obtaining polynomial factor hardness of approximating $EQ_G^1[k]$ (i.e. $n^\epsilon$ for some $\epsilon > 0$, where $n$ is the number of equations). New techniques seem to be required in order to attack the open problem.

## Acknowledgements

I would like to thank the anonymous referees for their careful readings of the manuscripts and many useful suggestions. The research was partially supported by young teacher research funds.

## References

[1] E. Amaldi, V. Kann, The complexity and approximability of finding feasible subsystems of linear relations, Theoretical Computer Science 147 (1995) 181–210.
[2] S. Arora, Probabilistic checking of proofs and the hardness of approximation problems. Ph.D. Thesis, UC Berkeley, 1994.
[3] S. Arora, L. Babai, J. Stern, E.Z. Sweedyk, The hardness of approximate optima in lattices, codes, and systems of linear equations, Journal of Computer and System Sciences 54 (1997) 317–331.
[4] G. Ausiello, P. Crescenzi, M. Protasi, Approximation solutions of NP-optimization problems, Theoretical Computer Science 150 (1995) 1–55.
[5] D.A.M. Barrington, P. McKenzie, C. Moore, P. Tesson, D. Thérien, Equation satisfiability and program satisfiability for finite monoids, in: Proceedings of the 25th Annual Symposium on Mathematical Foundations of Computer Science, Bratislava, Slovak Republic, 2000, pp. 172–181.
[6] F. Bédard, A. Goupil, The poset of conjugacy classes and decomposition of products in the symmetric group, Canadian Mathematical Bulletin 35 (2) (1992) 152–160.

[7] I. Dinur, S. Safra, The importance of being biased, in: Proceedings of the 34th Annual ACM Symposium on Theory of Computing, Montréal, Quebéc, Canada, 2002, pp. 33–42.
[8] L. Engebretsen, The non-approxiability of non-Boolean predicates, in: M. Goemans, K. Jansen, J.D.P. Rolim, L. Trevisan (Eds.), Proceedings of 5th International Workshop on Randomization and Approximation Techniques in Computer Science, in: Lecture Notes in Computer Science, vol. 2129, Springer-Verlag, Berkeley, CA, USA, 2001, pp. 241–248.
[9] L. Engebretsen, J. Holmerin, Clique is hard to approximate within $n^{1-o(n)}$, in: U. Montanari, J.D.P. Rolim, E. Welzl (Eds.), Proceedings of 27th International Colloquium on Automata, Languages and Programming, in: Lecture Notes in Computer Science, vol. 1853, Springer-Verlag, Geneva, 2000, pp. 2–12.
[10] L. Engebretsen, J. Holmerin, A. Russell, Inapproximability results for equations over finite groups, ICALP 2002, pp. 73–84.
[11] U. Feige, A threshold of ln $n$ for approximating set cover, Journal of ACM 45 (4) (1998) 634–652.
[12] U. Feige, S. Goldwasser, L Lovász, S. Safra, M. Szegedy, Interactive proofs and the hardness of approximating cliques, Journal of ACM 43 (2) (1996) 268–292.
[13] H. Finkelstein, K.I. Mandelberg, On solutions of ''equations in symmetric groups'', Journal of Combinatorial Theory, Series A 25 (2) (1978) 142–152.
[14] M. Goldmann, A. Russell, The complexity of solving equations over finite groups, in: Proceedings of the Fourteenth Annual IEEE Conference on Computational Complexity, Atlanta, Georgia, 1999, pp. 80–86.
[15] V. Guruswami, J. Håstad, M. Sudan, Hardness of approximate hypergraph coloring, in: 41st Annual Symposium on Foundations of Computer Science, IEEE, Redondo Beach, California, 2000, pp. 149–158.
[16] J. Håstad, Clique is hard to approximate within $n^{1-\epsilon}$, in: Proc. 37th Symposium on Foundation of Computer Science, 1996, pp. 627–636.
[17] J. Håstad, Some optimal inapproximability results, Journal of ACM 48 (4) (2001) 798–859.
[18] S. Khot, Improved inapproximability results for maxclique, chromatic number and approximate graph coloring, in: 42nd Annual Symposium on Foundations of Computer Science, IEEE, Las Vegas, Nevada, 2001, pp. 600–609.
[19] R. Raz, A parallel repetition theorem, SIAM Journal on Computing 27 (3) (1998) 763–803.
[20] A. Samorodnitsky, L. Trevisan, A PCP characterization of NP with optimal amortized query complexity, in: Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, Portland, Oregon, 2000, pp. 191–199.
[21] R.P. Stanley, Enumerative Combinatorics, volume 2, in: Cambridge Studies in Advances Mathematics, vol. 62, Cambridge University Press, Cambridge, 1999.
[22] S.P. Strunkov, K teorii uravnenij na konečnyh gruppah, Izvestiâ Rossijskoj Akademii Nauk, Seriâ Matematičeskaâ 59 (6) (1995) 171–180. (English translation: On the theory of equations in finite groups. Izvestiya: Mathematics, 59 (6): 1273–1282, 1995).
[23] U. Zwick, Approximations algorithms for constraint satisfaction programs involving at most three variables per constraint, in: Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, San Francisco, CA, 1998, pp. 201–210.