On the Complexity of Global Constraint Satisfaction

Cristina Bazgan[*]   Marek Karpinski[†]

Abstract

We study the computational complexity of decision and optimization problems that may be expressed as boolean constraint satisfaction problem with the global cardinality constraints. In this paper we establish a characterization theorem for the decision problems and derive approximation hardness results for the corresponding global optimization problems.

1   Introduction

Constraints of the global nature arise naturally in some optimization problems. For example, MIN BISECTION can be viewed as MIN CUT with the restriction that the two sets of vertices that determine the cut must be of equal size. It is known that MIN CUT is polynomial while MIN BISECTION is NP-hard. MIN BISECTION, MAX BISECTION and other optimization problems can be written as boolean constraint satisfaction problems where a feasible solution is a balanced assignment (where the number of variables set to 1 is the same as the number of variables set to 0). It was an increased interest in global optimization problems recently, cf. [HZ01, FL01, JS04].

In this paper we study the complexity of decision and optimization problems of the balanced versions of boolean constraint satisfaction problems depending on the type of constraints. Schaefer [Sch78] established a dichotomy theorem for the boolean constraint satisfaction problems distinguishing six polynomial time solvable cases. For the decision versions we show that if the set of constraints contains only equations of width 2 or it contains only conjunctions of literals, then the balanced version is polynomial time solvable and otherwise it is NP-complete.

Creigou [Cre95] and Khanna and Sudan [KS96] established a dichotomy theorem for maximization versions of boolean constraint satisfaction problems that classify the problems into polynomially solvable or APX-hard. The balanced versions of these problems where also studied. Sviridenko [Svi01] proved that the balanced version of MAX SAT is $1/(1 - \frac{1}{e})$-approximable. For the balanced version of MAX 2SAT, Blaser and Manthey [BM02] established a 1.514-approximation factor and Hofmeister [Hof03] a 4/3-approximation factor.

Lower bound were also studied for these problems. Holmerin [Hol02] showed that the balanced version of Max E4-0H-Lin2 (see for the definition Section 2) cannot be approximated within 1.0957 in polynomial time, unless $P=NP$. Also Holmerin and Khot [HK03] showed that balanced version of Max E3-0H-Lin2 is hard to approximate within $\frac{4}{3} - \varepsilon$ and in [HK04] they improved their result showing that this problem is hard to approximate within $2 - \varepsilon$, for any $\varepsilon > 0$, if $NP \not\subseteq \cap_{\delta>0} DTIME(2^{n^\delta})$, thus obtaining the best possible inapproximability factor result for this problem. We prove in this paper that all the cases that were considered by Creigou [Cre95] and Khanna and Sudan [KS96] in the dichotomy theorem become APX-hard and also that most of the trivial maximization constraint satisfaction problems have their balanced version APX-hard. In particular, using an inapproximability result for Densest k Subgraph established recently by Khot [Kho04], we prove that the balanced version of Min Monotone-E2Sat has no polynomial time approximation scheme, if $NP \not\subseteq \cap_{\delta>0} BTIME(2^{n^\delta})$, for $BTIME$ denoting randomized polynomial time.

Khanna, Sudan and Trevisan [KST97] established a classification theorem for minimization versions of boolean constraint satisfaction problems. The complexity of approximation of Min Bisection was for long time widely open. Feige and Krautghamer [FK00] established an approximation algorithm for this problem within $O(\log^2 n)$ approximation factor. This result has been recently improved to $O(\log^{1.5} n)$ by the recent result of Arora, Rao and Vazirani [ARV04]. Very recently, Khot [Kho04] established that under the assumption that $NP \not\subseteq \cap_{\delta>0} BTIME(2^{n^\delta})$, Min Bisection has no polynomial time approximation scheme. Under the assumption that refuting SAT formulas is hard to approximate on average, Feige [Fei02] proved also that Min Bisection is hard to approximate below $\frac{4}{3}$.

Holmerin studied the hardness of approximating some generalizations of Min Bisection. In particular he showed [Hol02] that the balanced version of Min E4-1H-Lin2 is not $(2 - \varepsilon)$-approximable for any $\varepsilon > 0$, unless $P=NP$. We prove several inapproximability result for balanced minimization problem. In particular, using the inapproximability result for Densest k Subgraph established by Khot [Kho04], we prove that the balanced version of Min Monotone-E2Sat has no polynomial time approximation scheme, if $NP \not\subseteq \cap_{\delta>0} BTIME(2^{n^\delta})$.

The paper is organized as follows: in Section 2 we introduce some preliminary notation and definitions, Section 3 contains our results on decision problems and Section 4 contains a summary of these results.

In Section 5 and 6 we present our results concerning maximization and minimization optimization problems. In Section 7 we summarize chains of reductions preserving the approximation and in Section 8 we briefly consider dense instances of the problems considered in the chains of reductions. Section 9 contains some open questions.

2   Preliminaries

We refer a general reader to [KST97, KSW97, KSTW01, CKS01] for a background on the boolean constraint satisfaction problems.

A constraint is a boolean function $f : \{0,1\}^k \to \{0,1\}$. A constraint application is a pair $< f, (i_1, \ldots, i_r) >$ where r is the arity of f and the $i_\ell \in [n]$ indicate to which r of the n

boolean variables a given constraint is applied. This constraint application will be denoted in the following by $f(x_{i_1}, \ldots, x_{i_r})$.

Let $\mathcal{F} = \{f_1, \ldots, f_t\}$ be a finite collection of boolean functions. An $\mathcal{F}$-set of constraints on n boolean variables $x_1, \ldots, x_n$ is a collection of constraint applications $\{f_j(x_{j_1}, \ldots, x_{j_{r_j}})\}_{j=1}^m$ for some integer m, where $f_j \in \mathcal{F}$ and r_j is the arity of f_j. We say that an assignment satisfies an $\mathcal{F}$-set of constraints if it satisfies every constraint in the collection.

The satisfiability problem $\mathrm{CSP}(\mathcal{F})$ consists of deciding whether there exists an assignment that satisfies a given $\mathcal{F}$-set of constraints.

$k\mathrm{CSP}(\mathcal{F})$ (respectively, $\mathrm{E}k\mathrm{CSP}(\mathcal{F})$) is the variant of $\mathrm{CSP}(\mathcal{F})$ where each boolean function f_j is a function of at most (respectively, exactly) k variables, for $j \leq t$. The problems MAX (MIN) $\mathrm{CSP}(\mathcal{F})$ consist of finding a boolean assignment that maximizes (minimizes) the number of constraints that are satisfied. MAX (MIN) $k\mathrm{CSP}(\mathcal{F})$ (respectively, MAX (MIN) $\mathrm{E}k\mathrm{CSP}(\mathcal{F})$) are variants of MAX (MIN) $\mathrm{CSP}(\mathcal{F})$ where each constraint depends on at most (respectively, exactly) k literals.

Given a problem A, the BALANCED version of A is the problem A with a new set of feasible solutions being assignments where the number of variables set to $true$ (denoted by 1) is the same as the number of variables set to $false$ (denoted by 0). Such assignments will be called *balanced* assignments.

We consider also a generalization of this problem. Given a problem A, the α-BALANCED version of A, $0 < \alpha < 1$, is the problem A with a new set of feasible solutions being assignments with the number of $true$ variables being an α ratio of the total number of variables. Such assignments will be called in the following α-balanced.

In this paper we study the complexity of decision and optimization problems related to BALANCED $\mathrm{CSP}(\mathcal{F})$ depending on the type of constraints defined by a class $\mathcal{F}$.

We start by defining the types of constraints studied in this paper.

A boolean function f will be called

- 0-valid if $f(0, \ldots, 0) = 1$.

- 1-valid if $f(1, \ldots, 1) = 1$.

- weakly negative (Horn) if f is expressible as a CNF-formula having at most one unnegated variable in each clause.

- weakly positive (anti-Horn) if f is expressible as a CNF-formula having at most one negated variable in each clause.

- affine if f is expressible as a conjunction of linear equations over $GF(2)$ and affine with width 2 if each equation depend on 2 variables.

- bijunctive if f is expressible as a 2CNF-formula

- 2-monotone if f is expressible as a DNF of the type $x_{i_1} \wedge \ldots \wedge x_{i_p}$ or $\bar{x}_{j_1} \wedge \ldots \wedge \bar{x}_{j_q}$ or $(x_{i_1} \wedge \ldots \wedge x_{i_p}) \vee (\bar{x}_{j_1} \wedge \ldots \wedge \bar{x}_{j_q})$.

In the following we give definitions of the problems that we consider in this paper.

kSAT (respectively, EkSAT) is the version of SAT where each clause is of size at most (respectively, exactly) k. kSAT $(B, \bar{B})$ is the variant of kSAT where each literal appears at most B times. MONOTONE-EkSAT is the variant of the EkSAT problem where either all clauses contain only positive literals or all clauses contain only negative literals.

In this paper we use the notation AND instead of DNF for the problem of deciding whether a set of conjunctions of literals has a satisfiable assignment. MONOTONE-EkAND is the variant of the EkAND problem where either all conjunctions contain only positive literals or they contain only negative literals.

The input of the EkLIN2 problem is a set of equations of the type $x_{i_1} \oplus \ldots \oplus x_{i_k} = 0$ or $x_{i_1} \oplus \ldots \oplus x_{i_k} = 1$, and we have to decide if there is a satisfiable assignment. In the homogenous variants of linear equations problem, Ek-bH-LIN2, $b \in \{0, 1\}$, the input consists of a set of equations of the type $x_{i_1} \oplus \ldots \oplus x_{i_k} = b$ on n boolean variables $x_1, \ldots, x_n$ and the problem consists of deciding if there is an assignment satisfying all equations.

MAX kSAT is the problem, given a set of clauses, of constructing an assignment satisfying a maximum number of clauses.

MAX kAND is the problem, given a set of conjunctions of size at most k, of determining an assignment maximizing the number of conjunctions satisfied.

MAX EkLIN2 is the problem, given a set of equations mod 2, of constructing an assignment maximizing the number of equations satisfied.

MIN kSAT is the problem, given a set of clauses, of constructing an assignment satisfying a minimum number of clauses.

MIN kAND is the problem, given a set of conjunctions of size at most k, of determining an assignment minimizing the number of conjunctions satisfied.

MIN EkLIN2 is the problem, given a set of equations mod 2, of constructing an assignment minimizing the number of equations satisfied.

We give now some basic notions on approximation complexity of optimization problems.

Given an NPO optimization problem ([ACG+99]) A and an instance I of A, we use $|I|$ to denote the size of I, $opt(I)$ to denote the optimum value for this instance and $val(I, S)$ to denote the value of a feasible solution S of instance I. The *performance ratio* of S (or *approximation factor*) is $r(I, S) = \max \left\{ \frac{val(I,S)}{opt(I)}, \frac{opt(I)}{val(I,S)} \right\}$. The *error* of S, $\varepsilon(I, S)$, is defined by $\varepsilon(I, S) = r(I, S) - 1$.

For a function f, an algorithm A is an $f(n)$-*approximation*, if for every instance I of the problem A, it returns a solution S such that $r(I, S) \leq f(|I|)$. For a constant c, an algorithm A is a *(randomized) c-approximation*, if for any instance I of the problem, it returns a solution S such that $r(I, S) \leq c$. We say that an optimization problem is *constant factor approximable* if, for some constant c, there exists a polynomial time c-approximation algorithm for it. The class of problems which are constant factor approximable is denoted by APX.

The notion of an E-reduction (*error-preserving* reduction), denoted here by $\leq$, was introduced by Khanna, Motwani, Sudan and Vazirani in [KMSV94]. A problem A is called E-*reducible* to a problem B, if there exist polynomial time computable functions f, g and a constant β such that

- f maps an instance I of A to an instance I' of B such that $opt(I)$ and $opt(I')$ are related by a polynomial factor, i.e. there exists a polynomial $p(n)$ such that $opt(I') \leq p(|I|)opt(I)$,

- g maps solutions S' of I' to solutions S of I such that $\varepsilon(I, S) \leq \beta\varepsilon(I', S')$.

An important property of that reduction is that it can be applied uniformly to all levels of approximability; that is, if A is E-reducible to B and B belongs to $\mathcal{C}$ then A belongs to $\mathcal{C}$ as well, where $\mathcal{C}$ is a class of optimization problems with any kind of approximation guarantee (see also [KMSV94]).

We call two optimization problems A and B, E-equivalent if $A \leq B$ and $B \leq A$.

A problem A is APX-hard, if every problem $B \in APX$ is such that $B \leq A$. A problem is called APX-complete if it belongs to APX, and it is APX-hard.

We formulate our results in terms of E-reductions (note that some authors use, say for minimization problems, different kind of reductions [KSTW01, CP91]). Given two optimization problems A and B we say that B is A-complete if A is E-reducible to B, and vice versa.

3   Complexity of Decision Problems

The decision complexity of boolean constraint satisfaction problems is well established. In particular, Schaefer [Sch78] established the following remarkable dichotomy theorem:

Theorem 1 (Dichotomy Theorem for $\mathrm{CSP}(\mathcal{F})$ **[Sch78])** *Given an $\mathcal{F}$-set of constraints, the problem $\mathrm{CSP}(\mathcal{F})$ is polynomial time computable if $\mathcal{F}$ satisfies one of the conditions below, and $\mathrm{CSP}(\mathcal{F})$ is NP-complete otherwise.*

1. Every function in $\mathcal{F}$ is 0-valid.

2. Every function in $\mathcal{F}$ is 1-valid.

3. Every function in $\mathcal{F}$ is weakly positive.

4. Every function in $\mathcal{F}$ is weakly negative.

5. Every function in $\mathcal{F}$ is affine.

6. Every function in $\mathcal{F}$ is bijunctive.

Motivated by the above result, we aim at formulating analogous result for balanced problems. Firstly we show that for any $\mathcal{F}$-set of constraints, BALANCED $\mathrm{CSP}(\mathcal{F})$ is at least as difficult as $\mathrm{CSP}(\mathcal{F})$.

Lemma 1 *If $\mathrm{CSP}(\mathcal{F})$ is NP-complete, then* BALANCED $\mathrm{CSP}(\mathcal{F})$ *is also NP-complete.*

Proof: We reduce $\mathrm{CSP}(\mathcal{F})$ to BALANCED $\mathrm{CSP}(\mathcal{F})$. Given an instance I of $\mathrm{CSP}(\mathcal{F})$ on n variables $x_1, \ldots, x_n$ and with m constraints, we construct an instance I' of BALANCED $\mathrm{CSP}(\mathcal{F})$ on $2n$ variables $x_1, \ldots, x_n$ and n new variables $y_1, \ldots, y_n$. For each constraint $f_j(x_{j_1}, \ldots, x_{j_{r_j}})$ of I we add to I' the following two constraints: $f_j(x_{j_1}, \ldots, x_{j_{r_j}})$ and $f_j(\bar{y}_{j_1}, \ldots, \bar{y}_{j_{r_j}})$. It is easy to see that if $x_i = v_i, i = 1, \ldots, n$ is an assignment that satisfies I then $x_i = v_i, y_i = \bar{v}_i, i = 1, \ldots, n$ is a balanced assignment that satisfies I', where $\bar{v}$ means the complemented value of v. Conversely, if $x_i = v_i, y_i = w_i, i = 1, \ldots, n$ is a balanced assignment that satisfies I' then $x_i = v_i, i = 1, \ldots, n$ is an assignment that satisfies I. $\qquad\square$

We turn now to a polynomial time case. We formulate our result in slightly more general setting of the α-balanced problems.

Theorem 2 *For any $0 < \alpha < 1$, α-BALANCED E2-LIN2 is solvable in polynomial time.*

Proof: Let us consider first $\alpha = \frac{1}{2}$. Given an instance I of BALANCED E2-LIN2 on n variables and m equations, we construct some equivalence classes on the set of literals by considering the equations one after another as follows. Given an equation $x_i \oplus x_j = 0$ ($x_i \oplus x_j = 1$), we distinguish the following cases.

- If literals $x_i, \bar{x}_i, x_j, \bar{x}_j$ do not appear in a class, then we construct a new class and we put together x_i and x_j (x_i and $\bar{x}_j$ respectively).

- If either x_i or $\bar{x}_i$ appears in a class C_k and $x_j, \bar{x}_j$ do not appear in a class, then

  - if $x_i \in C_k$ then we introduce x_j ($\bar{x}_j$ respectively) in C_k.
  - if $\bar{x}_i \in C_k$ then we introduce $\bar{x}_j$ (x_j respectively) in C_k.

- If literals x_i or $\bar{x}_i$ and x_j or $\bar{x}_j$ appear in the same class C_k then I is not satisfiable if $\{x_i, \bar{x}_j\} \subseteq C_k$ or $\{\bar{x}_i, x_j\} \subseteq C_k$ ($\{x_i, x_j\} \subseteq C_k$ or $\{\bar{x}_i, \bar{x}_j\} \subseteq C_k$ respectively).

- If either x_i or $\bar{x}_i$ appears in a class C_k and either x_j or $\bar{x}_j$ appears in a class C_ℓ then

  - if $x_i \in C_k$ and $x_j \in C_\ell$ then we put together the literals of both classes C_k and C_ℓ (we put together the literals of the class C_k with the negated literals of the class C_ℓ).
  - if $x_i \in C_k$ and $\bar{x}_j \in C_\ell$ then we put together the literals of the class C_k with the negated literals of the class C_ℓ (we put together the literals of both classes C_k and C_ℓ).

Suppose that at the end we obtain t equivalence classes $C_1, \ldots, C_t$. Denote by a_{2i-1} and a_{2i} the number of literals that appear positive and respectively negative in C_i. BALANCED E2-LIN2 on I consists of deciding if there exists a partition of these $2t$ integers in two equal size sets P and N such that P and N contain exactly one of a_{2i-1}, a_{2i} for $i = 1, \ldots, t$. This problem in solvable in polynomial time by dynamic programming [GJ76]. If such a partition P, N exists then the following assignment is balanced and satisfies I:

- if $a_{2i-1} \in P$ then we assign to the positive variables of C_i the value 1 and to the negated variables of C_i the value 0.

- if $a_{2i-1} \in N$ then we assign to the positive variables of C_i the value 0 and to the negated variables of C_i the value 1.

If $\alpha \neq \frac{1}{2}$ then as below we construct equivalence classes $C_1, \ldots, C_t$ and compute integers $a_1, \ldots, a_{2t}$. We add two other integers $a_{2t+1} = n|1 - 2\alpha|$, $a_{2t+2} = 0$ and solve the above partition problem on this new instance. $\square$

Before we attack the "Ek"-situation for $k \geq 3$, we establish some auxiliary results.

Proposition 1 BALANCED Ek-0H-LIN2 *and* BALANCED Ek-1H-LIN2 *are polynomial equivalent, for every odd $k \geq 3$.*

Proof: We reduce BALANCED Ek-0H-LIN2 to BALANCED Ek-1H-LIN2. From an instance I of BALANCED Ek-0H-LIN2 on n variables $x_1, \ldots, x_n$ we construct an instance I' on the same set of variables and we associate to each equation $x_{i_1} \oplus \ldots \oplus x_{i_k} = 0$ of I the equation $x_{i_1} \oplus \ldots \oplus x_{i_k} = 1$ in I'. It is easy to see, since k is odd, an assignment satisfies I if and only if the complemented assignment satisfies I'. The reduction in the other direction is similar. $\square$

We define a new problem to be used later.

kONES(E3-bH-LIN2), $b \in \{0, 1\}$

Input: A set of equations of the type $x_{i_1} \oplus x_{i_2} \oplus x_{i_3} = b$ on n boolean variables $x_1, \ldots, x_n$.
Question: Is there an assignment, with exactly k variables set to 1, satisfying all equations ?

MAX ONES($\mathcal{F}$) consists of determining an assignment that satisfies all constraints of $\mathcal{F}$-type and maximizes the number of variables assigned to 1.

Theorem 3 ([KSW97, KSTW01]) *If every function in $\mathcal{F}$ is of the type $x_{i_1} \oplus x_{i_2} \oplus x_{i_3} = 0$ or if every function in $\mathcal{F}$ is of the type $x_{i_1} \oplus x_{i_2} \oplus x_{i_3} = 1$ then the problem* MAX ONES($\mathcal{F}$) *is APX-complete.*

A consequence of the previous theorem is that kONES(E3-0H-LIN2) and kONES(E3-1H-LIN2) are *NP*-complete.

Proposition 2 BALANCED E3-0H-LIN2 *and* BALANCED E3-1H-LIN2 *are both NP-complete.*

Proof: We construct a reduction between kONES(E3-0H-LIN2) and BALANCED E3-0H-LIN2. Given an instance I of kONES(E3-0H-LIN2) on n variables $x_1, \ldots, x_n$ we construct an instance I' on $2n$ variables $x_1, \ldots, x_n, y_1, \ldots, y_{n-k}, z_1, \ldots, z_k$ as follows. For each equation $x_{i_1} \oplus x_{i_2} \oplus x_{i_3} = 0$ from I, we associate in I' the same equation and let us denote in the following this set of equations by A. We add also to I' the following set of equations, called B, $y_i \oplus y_j \oplus z_\ell = 0$ for every $i, j \in \{1, \ldots, n-k\}, \ell \in \{1, \ldots, k\}$, $i \neq j$. It is easy to see that in order that an assignment satisfies B the variables y must have the same value and the variables z must have the same value. Thus, an assignment that satisfies B has $z_i = 0$, $i = 1, \ldots, k$. Suppose that $y_i = 0$, $i = 1, \ldots, n-k$ then since the assignment must be balanced

we have $x_i = 1$, $i = 1, \ldots, n$ but in this case the equations in A are not satisfied. So, $y_i = 1$, $i = 1, \ldots, n - k$ and due to the balanced condition, the restriction of this assignment to variables x satisfies I and contains exactly k variables 1 and $n - k$ variables 0. In the similar way $k\textsc{Ones}(\text{E3-1H-Lin2})$ is reducible to Balanced E3-1H-Lin2. $\square$

Proposition 3 *For every odd $k \geq 3$, $b \in \{0, 1\}$,* Balanced Ek-bH-Lin2 *is NP-complete.*

Proof: For every odd $k \geq 3$, we construct in the following a reduction between Balanced Ek-0H-Lin2 and Balanced E$(k + 2)$-1H-Lin2 and using Propositions 1 and 2 we conclude the *NP*-completeness of Balanced Ek-bH-Lin2, $b \in \{0, 1\}$, for every odd $k \geq 3$. Given an instance I of Balanced Ek-0H-Lin2 on n variables $x_1, \ldots, x_n$ we construct an instance I' on $3n$ variables $x_1, \ldots, x_n, y_1, \ldots, y_n, z_1, \ldots, z_n$ as follows. For each equation $x_{i_1} \oplus \ldots \oplus x_{i_k} = 0$ from I, we associate in I' the following set of equations, called A, $x_{i_1} \oplus \ldots \oplus x_{i_k} \oplus y_j \oplus z_\ell = 1$, for every $j, \ell = 1, \ldots, n$. We add also to I' the following set of equations, called B, $y_{j_1} \oplus \ldots \oplus y_{j_{k+2}} = 1$ for every subset of $k + 2$ variables among $y_1, \ldots, y_n$. It is easy to see that in order for an assignment to satisfy A, the variables y must have the same value and the variables z must have the same value. Since the equations of B are satisfied $y_i = 1, i = 1, \ldots, n$ and since the assignment must be balanced we have $z_i = 0, i = 1, \ldots, n$ and thus I is satisfied by the restriction of this assignment to variables x (that is also balanced). In the similar way Balanced Ek-1H-Lin2 is reducible to Balanced E$(k + 2)$-0H-Lin2. $\square$

Proposition 4 *For every even $k \geq 4$, $b \in \{0, 1\}$,* Balanced Ek-bH-Lin2 *is NP-complete.*

Proof: For every odd $k \geq 3$, we construct a reduction between Balanced Ek-1H-Lin2 and Balanced E$(k + 1)$-1H-Lin2, and a reduction between Balanced Ek-0H-Lin2 and Balanced E$(k + 1)$-0H-Lin2 and using Proposition 3 we conclude the *NP*-completeness of Balanced Ek-H-Lin2 for every even $k \geq 4$.

Given an instance I of Balanced Ek-1H-Lin2 on n variables $x_1, \ldots, x_n$ we construct an instance I' of Balanced E$(k + 1)$-1H-Lin2 on $3n$ variables $x_1, \ldots, x_n, y_1, \ldots, y_n, z_1, \ldots, z_n$ as follows. For each equation $x_{i_1} \oplus \ldots \oplus x_{i_k} = 1$ from I, we associate in I' the following set of equations, called A, $x_{i_1} \oplus \ldots \oplus x_{i_k} \oplus z_\ell = 1$, for every $\ell = 1, \ldots, n$. We add also to I' the following set of equations, called B, $z_{j_1} \oplus \ldots \oplus z_{j_k} \oplus y_\ell = 1$ for every subset of size k among variables $z_1, \ldots, z_n$ and $\ell = 1, \ldots, n$. It is easy to see that in order for an assignment to satisfy A, the variables z must have the same value. Since the equations of B are satisfied, the variables y must have the same value. Since the assignment must be balanced we have either $y_i = 1, i = 1, \ldots, n$ and $z_i = 0, i = 1, \ldots, n$ or $y_i = 0, i = 1, \ldots, n$ and $z_i = 1, i = 1, \ldots, n$. If the first case appears then the restriction of this assignment to variables x satisfies I, otherwise the complement assignment satisfies I.

Given an instance I of Balanced Ek-0H-Lin2 on n variables $x_1, \ldots, x_n$ we construct an instance I' of Balanced E$(k + 1)$-0H-Lin2 on $3n$ variables $x_1, \ldots, x_n, y_1, \ldots, y_n, z_1, \ldots, z_n$ as follows. For each equation $x_{i_1} \oplus \ldots \oplus x_{i_k} = 0$ from I, we associate to I' the following set of equations, called A, $x_{i_1} \oplus \ldots \oplus x_{i_k} \oplus z_\ell = 0$, for every $\ell = 1, \ldots, n$. We add also to I' the following set of equations, called B, $z_{j_1} \oplus \ldots \oplus z_{j_{k-1}} \oplus y_i \oplus y_\ell = 1$ for every subset of size $k - 1$ among variables $z_1, \ldots, z_n$, and every subset of size 2 among variables $y_1, \ldots, y_n$. It is easy to see that in order that an assignment satisfies A, all the variables z must have the same value.

Since the equations of B are satisfied, the variables y must have the same value. Since the assignment must be balanced we have either $y_i = 1, i = 1, \ldots, n$ and $z_i = 0, i = 1, \ldots, n$ or $y_i = 0, i = 1, \ldots, n$ and $z_i = 1, i = 1, \ldots, n$. If the first case, the restriction of this assignment to variables x satisfies I, otherwise the complementary assignment satisfies I. $\qquad \square$

Theorem 4 *For any $k \geq 3$, $b \in \{0, 1\}$,* BALANCED Ek-bH-LIN2 *is NP-complete.*

Proof: This is a consequence of Propositions 3 and 4. $\qquad \square$

MONOTONE-2SAT is a trivial problem. In contrast to this, we show that α-BALANCED MONOTONE-E2SAT is, in fact, *NP*-hard.

Theorem 5 α-BALANCED MONOTONE-E2SAT *is NP-complete, for any $\alpha > 0$.*

Proof: We reduce α-CLIQUE (cf. [GJ76]) to α-BALANCED MONOTONE-E2SAT. An instance of α-CLIQUE has an input a graph on n vertices and we have to decide if it contains a clique of size at least αn. The reduction is as follows: given a graph $G = (V, E)$ on n vertices, we construct an instance I on n boolean variables $x_1, \ldots, x_n$, one for each vertex of G. For any $i, j \in V$ such that $(i, j) \notin E$, we add the clause $\bar{x}_i \vee \bar{x}_j$. It is clear that if C is a clique in G of size αn, then the assignment $x_i = 1$ if $i \in C$ and $x_i = 0$ if $i \notin C$ satisfies each clause of I since for each $(i, j) \notin E$, x_i or x_j is false. Conversely, if an α-balanced assignment satisfies I, then the set $C = \{i : x_i = 1\}$ is a clique of size αn. Since α-CLIQUE is *NP*-hard [GJ76], α-BALANCED MONOTONE-E2SAT is *NP*-hard as well. $\qquad \square$

Theorem 6 BALANCED MONOTONE-EkSAT *is NP-complete, for any $k \geq 3$.*

Proof: We reduce EkSAT to BALANCED MONOTONE-EkSAT. From an instance I on n variables $x_1, \ldots, x_n$, instance of EkSAT, we construct an instance I' on $2n$ variables $x_1, \ldots, x_n$, $y_1, \ldots, y_n$, instance of BALANCED MONOTONE-EkSAT as follows: we preserve in I' the clauses of I in which we replace every negative literal $\bar{x}_i$ by y_i, for $i = 1, \ldots, n$. We add to I' for $i = 1, \ldots, n$, the clauses $x_i \vee y_i \vee z_1 \vee \ldots \vee z_{k-2}$ where variables $z_1, \ldots, z_{k-2}$ are any subset of $k - 2$ variables among variables x_j, y_ℓ, $j, \ell = 1, \ldots, n$, $j, \ell \neq i$. It is easy to see that if I is satisfiable then I' is satisfiable by a balanced assignment where x_j and y_j have different values, for $j = 1, \ldots, n$. If I' is satisfiable by a balanced assignment then x_j and y_j have different values, for $j = 1, \ldots, n$, since, otherwise if there exits a pair of variables such that $x_i = y_i = 0$ then there is a clause $x_i \vee y_i \vee z_1 \vee \ldots \vee z_{k-2}$ where $z_j = 0$, for $j = 1, \ldots, k - 2$, that is not satisfied. $\qquad \square$

Since BALANCED AND is trivial we can formulate the following

Theorem 7 (Characterization Theorem for BALANCED CSP($\mathcal{F}$)**)** *Given an $\mathcal{F}$-set of constraints, the problem α-BALANCED CSP($\mathcal{F}$) is polynomial time solvable (if every function in $\mathcal{F}$ is affine with width 2 or if every function in $\mathcal{F}$ is a conjunction of literals), otherwise it is NP-complete.*

4 Summary of Balanced Decision Problems

We summarize the above results concerning balanced decision problem for a given $\mathcal{F}$-set in the following table:

Polynomial Time	NP-complete
Every function in $\mathcal{F}$ is affine with width 2	Every function in $\mathcal{F}$ is 0-valid
Every function in $\mathcal{F}$ is a conjunction of literals	Every function in $\mathcal{F}$ is 1-valid
	Every function in $\mathcal{F}$ is weakly positive
	Every function in $\mathcal{F}$ is weakly negative
	Every function in $\mathcal{F}$ is bijunctive

5 Approximation of Global Maximum Constraint Satisfaction

We state first the following known classification theorem of MAX CSP($\mathcal{F}$) (cf. [Cre95, KS96]).

Theorem 8 (Characterization Theorem for MAX CSP($\mathcal{F}$) [Cre95, KS96]) MAX CSP($\mathcal{F}$) *is either polynomial time computable or is APX-complete. Moreover, it is in P if and only if $\mathcal{F}$ is either 0-valid or 1-valid or 2-monotone.*

Some upper bounds have been established for these balanced versions of MAX CSP($\mathcal{F}$).

Theorem 9 ([Svi01]) α-BALANCED MAX SAT *is $1/(1 - \frac{1}{e})$-approximable.*

Theorem 10 ([BM02]) α-BALANCED MAX 2SAT *is 1.514-approximable.*

Theorem 11 ([Hof03]) BALANCED MAX 2SAT *is 4/3-approximable.*

We state first the following direct lemma:

Lemma 2 MAX CSP($\mathcal{F}$) *is E-reducible to* BALANCED MAX CSP($\mathcal{F}$).

Proof: We consider the reduction given in Lemma 1. Given an instance I of MAX CSP($\mathcal{F}$) on n variables $x_1, \ldots, x_n$ and with m constraints we construct an instance I' of BALANCED MAX CSP($\mathcal{F}$) on $2n$ variables $x_1, \ldots, x_n$ and n new variables $y_1, \ldots, y_n$. For each constraint $f_j(x_{j_1}, \ldots, x_{j_{r_j}})$ of I we add in I' the following two constraints: $f_j(x_{j_1}, \ldots, x_{j_{r_j}})$ and $f_j(\bar{y}_{j_1}, \ldots, \bar{y}_{j_{r_j}})$. It is easy to see that $opt(I') \geq 2opt(I)$. Given a balanced assignment of I' we can transform it into a balanced assignment v' that satisfies the same number of constraints in I as in the constraints of I' that are on variables y. The restriction of v' to variables x, called v, satisfies in I, $val(I, v) = \frac{val(I', v')}{2}$ constraints. Thus, $opt(I') = 2opt(I)$ and $\varepsilon(I, v) = \varepsilon(I', v')$. $\qquad\square$

The following lemma shows that the three polynomial cases for MAX CSP($\mathcal{F}$) became difficult for the balanced version.

Lemma 3 Balanced Max Monotone-E2Sat *is APX-hard.*

Proof: We E-reduce Max E2Sat$(B, \bar{B})$ (see for definition Section 2) to Balanced Max Monotone-E2Sat. From an instance I on n variables $x_1, \ldots, x_n$, instance of Max E2Sat$(B, \bar{B})$ we construct an instance I' on $2n$ variables $x_1, \ldots, x_n, y_1, \ldots, y_n$, instance of Balanced Max Monotone-E2Sat as follows : we preserve in I' the clauses of I where the negations of variables $\bar{x}_i$ are replaced by y_i. We add in I', B times each of the following clauses $x_i \vee y_i$, $i = 1, \ldots, n$. It is easy to see that from an optimal assignment of I satisfying $opt(I)$ clauses, we can define a balanced assignment for I' where x_i and y_i have opposed values that satisfy $opt(I) + n \times B$ clauses. Given a balanced assignment of I', if there exists a pair x_i, y_i with the same value 1 then there is another one x_j, y_j with the same value 0, and so changing the value of y_i and y_j will not decrease the number of clauses satisfied in I'. Thus each balanced assignment of I' can be transformed in a balanced assignment of I' that satisfies at least as many clauses and where in each pair of variables x_i and y_i have opposed values. If a balanced assignment of I' where x_i and y_i have opposed values satisfies $val + B \times n$ clauses then this assignment satisfies val clauses in I. $\square$

We can also generalize the previous lemma.

Theorem 12 Balanced Max Monotone-EkSat *is APX-hard, for $k \geq 3$.*

Proof: We can E-reduce Max EkSat$(B, \bar{B})$ to Balanced Max Monotone-EkSat. From an instance I of Max EkSat$(B, \bar{B})$ on n variables $x_1, \ldots, x_n$, we construct an instance I' of Balanced Max Monotone-EkSat on $2n$ variables $x_1, \ldots, x_n, y_1, \ldots, y_n$, as follows : we preserve in I' the clauses of I in which we replace each negative literal $\bar{x}_i$ by y_i, for $i = 1, \ldots, n$. We add to I', for $i = 1, \ldots, n$, the clauses $x_i \vee y_i \vee z_1 \vee \ldots \vee z_{k-2}$ where variables $z_1, \ldots, z_{k-2}$ form any subset of $k - 2$ variables among variables x_j, y_ℓ, $j, \ell = 1, \ldots, n$; $j, \ell \neq i$. This last subset is called A. It is easy to see that from an optimum assignment that satisfies $opt(I)$ clauses in I, we can construct a balanced assignment in I' where x_j and y_j have different values for $j = 1, \ldots, n$ that satisfies $opt(I') = opt(I) + |A|$. From a balanced assignment of I' where there exits a pair of variables such that $x_i = y_i = 0$ and another pair such that $x_j = y_j = 1$, we can construct a balanced assignment that satisfies at least as many clauses and where x_i, y_i and x_j, y_j have different values. This is possible since there are $\Theta(n^{k-2})$ clauses in A of type $x_i \vee y_i \vee z_1 \vee \ldots \vee z_{k-2}$ where $z_j = 0$, for $j = 1, \ldots, k-2$, that are not satisfied when $x_i = y_i = 0$ and become satisfied when we flip the value of one of these 2 variables and the number of clauses that could become non satisfied after the flip is at most $4B$. $\square$

A particular case of the following problem is equivalent to a Balanced Max CSP$(\mathcal{F})$ problem for some particular $\mathcal{F}$ as it will be proved later.

We introduce now a new problem.

Densest k Subgraph

Input: A graph $G = (V, E)$ on n vertices where n is even.
Output: A subset $S \subseteq V$ of size k that maximize the number of edges with both extremities in S.

The hardness of the approximation of Densest k Subgraph remained open for long time. Recently, Khot was able to establish a such result using a special PCP technique.

Theorem 13 ([Kho04]) DENSEST k SUBGRAPH *has no polynomial time approximation scheme if $NP \not\subseteq \cap_{\delta>0} BTIME(2^{n^\delta})$.*

More precisely, Khot proved the previous result for DENSEST k SUBGRAPH when $k = cn$ and $c < \frac{1}{2}$.

Proposition 5 DENSEST k SUBGRAPH *is E-reducible to* DENSEST $\frac{n}{2}$ SUBGRAPH.

Proof: Given a graph G on n vertices, an instance of DENSEST k SUBGRAPH, where $k = cn$ and $c < \frac{1}{2}$, we construct a graph G', instance of DENSEST $\frac{n}{2}$ SUBGRAPH, from G by adding a complete graph on $n - 2k$ vertices. We have $opt(G') = opt(G) +$ *the number of edges of the complete graph*. Given a solution of G' we can transform it in another one, with a better value, that contains the complete graph and a subset of vertices from G. It is easy to verify that this is an E-reduction. $\qquad\square$

Proposition 6 BALANCED MAX MONOTONE-E2AND *is E-equivalent to* DENSEST $\frac{n}{2}$ SUBGRAPH.

Proof: Given an instance I of BALANCED MAX MONOTONE-E2AND on n variables $x_1, \ldots, x_n$ and m clauses, we construct an instance G of DENSEST $\frac{n}{2}$ SUBGRAPH on n vertices and m edges as follows : we associate a vertex i for each variable x_i and an edge (i, j) for each conjunction $x_i \wedge x_j$. Given an optimal balanced solution for I, consider S the set of variables set to 1 in this assignment. We have $opt(G) \geq opt(I)$. Given a set S of size $\frac{n}{2}$ in G of value $val(I', S)$, the following assignment (called v) $x_i = 1$ if $i \in S$ and $x_i = 0$ otherwise, satisfies $val(I', S)$ conjunctions in I. Thus $opt(G) = opt(I)$ and $\varepsilon(I, v) = \varepsilon(I', S)$. The reduction in the other direction is similar. $\qquad\square$

Proposition 7 BALANCED MAX MONOTONE-EkAND *is E-reducible to* BALANCED MAX MONOTONE-E$(k+1)$AND*, for $k \geq 2$.*

Proof: Given an instance I of BALANCED MAX MONOTONE-EkAND on n variables $x_1, \ldots, x_n$ and m conjunctions, we construct an instance I' of BALANCED MAX MONOTONE-E$(k+1)$AND on the same set of variables at which we add two variables y and z as follows. For each clause $x_{i_1} \wedge \ldots \wedge x_{i_k}$ from I we add in I' the clause $x_{i_1} \wedge \ldots \wedge x_{i_k} \wedge y$. It is easy to see that from an optimal balanced assignment that satisfies $opt(I)$ conjunctions in I we can construct a balanced assignment in I' (considering the same assignment for variables x and considering $z = 0$ and $y = 1$) that satisfies also $opt(I)$ conjunctions. Thus $opt(I') \leq opt(I)$. Given a balanced assignment of I', if y and z have different values then we can assume that $z = 0$ and $y = 1$. Otherwise if y and z have the same value 1 then we can swap the value of z and swap also the value of a variable x that is set to 0 obtaining a balanced assignment with a better value. The restriction of this assignment to variables x is a balanced assignment satisfying the same number of conjunctions. $\qquad\square$

Theorem 14 BALANCED MAX MONOTONE-EkAND*, $k \geq 2$, has no polynomial time approximation scheme if $NP \not\subseteq \cap_{\delta>0} BTIME(2^{n^\delta})$.*

Proof: The result is a consequence of Propositions 5, 6 and 7 and Theorem 13 (Khot's result [Kho04]). □

We consider in the following the balanced version of affine constraints.

MAX E2-1H-LIN2, that is MAX CUT, is known to be APX-hard [PY91] and BALANCED MAX E2-1H-LIN2 that is MAX BISECTION is known to be APX-hard [PY91, Has97]. Each instance of MAX E2-0H-LIN2 is satisfied by the trivial assignment 0. We show a relation between the complexity of BALANCED MAX MONOTONE-E2AND and BALANCED MAX E2-0H-LIN2 (or BALANCED MAX UNCUT).

Proposition 8 BALANCED MAX MONOTONE-E2AND *is E-reducible to* BALANCED MAX E2-0H-LIN2.

Proof: Given an instance I of BALANCED MAX MONOTONE-E2AND on n variables $x_1, \ldots, x_n$ and m conjunctions, we construct an instance I' of BALANCED MAX E2-0H-LIN2 on $n + 2$ variables $x_1, \ldots, x_n$ and two new variables y and z and $3m$ equations as follows: for each conjunction $x_1 \wedge x_2$ we add 3 equations $x_1 \oplus x_2 = 0, x_1 \oplus y = 0, x_2 \oplus y = 0$. We have $opt(I') \geq 2opt(I) + m$ since the assignment satisfying $opt(I)$ conjunctions in I and $z = 0$ and $y = 1$ satisfies $2opt(I) + m$ equations in I'. Given a balanced assignment v for I' satisfying val' equations, we can consider $y = 1$. If $z = 0$ then, the restriction of v on x variables is balanced and satisfies $\frac{val' - m}{2}$ conjunctions. If $z = 1$ then the restriction of v on x variables satisfies $\frac{val' - m}{2}$ conjunctions in I but is not balanced. Observe that the balanced assignment obtained by changing the value of an x variable from 0 to 1 satisfies at least $\frac{val' - m}{2}$ conjunctions. □

Thus we establish the first inapproximability result for BALANCED MAX UNCUT.

Theorem 15 BALANCED MAX UNCUT *has no polynomial time approximation scheme if NP* $\not\subseteq \cap_{\delta > 0} BTIME(2^{n^\delta})$.

Proof: The result is a consequence of Propositions 6, 8 and Theorem 13 (Khot's result [Kho04]). □

When k is odd, MAX Ek-bH-LIN2 is trivial since the assignment b for all variables satisfies all equations. When k is even, MAX Ek-0H-LIN2 is also trivial since the assignment 0 for all variables satisfies all equations. For $k \geq 4$ even, MAX Ek-1H-LIN2 is not know to be hard to approximate.

Theorem 16 BALANCED MAX Ek-bH-LIN2 *is APX-hard, for $k \geq 3$, $b \in \{0,1\}$*.

Proof: We construct an E-reduction between BALANCED MAX E2-1H-LIN2 and BALANCED MAX E3-1H-LIN2. Given an instance I of BALANCED MAX E2-1H-LIN2 on n variables $x_1, \ldots, x_n$ and m equations, we construct an instance I' on $3n$ variables $x_1, \ldots, x_n$, $y_1, \ldots, y_n$, $z_1, \ldots, z_n$ as follows. For each equation $x_{i_1} \oplus x_{i_2} = 1$ from I, we associate in I' the equations $x_{i_1} \oplus x_{i_2} \oplus z_\ell = 1$, for $\ell = 1, \ldots, n$ and let us call in the following this set of equations A. We add also to I' the following equations $z_i \oplus z_j \oplus y_\ell = 1$ for $i \neq j$, $i, j, \ell \in \{1, \ldots, n\}$. This last set of equations is called B. It is easy to see that $opt(I') \geq n \times opt(I) + |B|$ since

the same assignment for variables x, $z_i = 0$, $i = 1, \ldots, n$ and $y_i = 1$, $i = 1, \ldots, n$ satisfies in A, $n \times opt(I)$ equations. Since there are $\Theta(n^2)$ equations of the type $z_i \oplus z_j \oplus y_\ell = 1$ and $z_i \oplus z_j \oplus y_t = 1$ for some fixed i and j and $\ell, t \in \{1, \ldots, n\}$ then $y_\ell = y_t$ and so all variables y have the same values and this value is 1. We can prove similarly that all variables z have the same values. Since the assignment is balanced, then variables z have values 0 and thus the variables x form a balanced solution.

Propositions 1, 3 and 4 can be adapted in order to prove the same results for maximization versions. $\qquad\square$

BALANCED MAX Ek-bH-LIN2 was studied for particular cases of k and $b = 0$. More precisely, Holmerin [Hol02] proved that BALANCED MAX E4-0H-LIN2 cannot be approximated within 1.0957 in polynomial time, unless $P = NP$. Also Holmerin and Khot showed in [HK03] that BALANCED MAX E3-0H-LIN2 is hard to approximate within $\frac{4}{3} - \varepsilon$ and in [HK04] they improved their result showing that BALANCED MAX E3-0H-LIN2 is hard to approximate within $2 - \varepsilon$ if $NP \not\subseteq \cap_{\delta > 0} DTIME(2^{n^\delta})$, thus obtaining the best possible inapproximability bound result for this problem (under this assumption).

Theorem 17 (Characterization Theorem for BALANCED MAX CSP($\mathcal{F}$)) BALANCED MAX CSP($\mathcal{F}$) *is APX-hard.*

6 Approximation of Global Minimum Constraint Satisfaction

A classification theorem for MIN CSP($\mathcal{F}$) was formulated in [KST97].

We can show directly, like for the decision and maximization constraint satisfaction problems, that the balanced version of a minimization problem is at least as hard as an underlying problem.

Lemma 4 MIN CSP($\mathcal{F}$) *is E-reducible to* BALANCED MIN CSP($\mathcal{F}$).

Proof: The reduction given in the proof of Theorem 2 is also an E-reduction between MIN CSP($\mathcal{F}$) and BALANCED MIN CSP($\mathcal{F}$). $\qquad\square$

MIN MONOTONE-EkSAT for $k \geq 2$ are trivial problems since the assignment 0 for all variables satisfies no clause. For the balanced situation we formulate

Proposition 9 BALANCED MAX MONOTONE-E2AND *is E-reducible to* BALANCED MIN MONOTONE-E2SAT.

Proof: Given an instance I of BALANCED MAX MONOTONE-E2AND on n variables $x_1, \ldots, x_n$ and m conjunctions, we construct an instance I' on the same set of variables as follows. For each conjunction $\bar{x}_{i_1} \wedge \bar{x}_{i_2}$ from I we add in I' the clause $x_{i_1} \vee x_{i_2}$. We can check easily that this is an E-reduction. $\qquad\square$

14

Theorem 18 BALANCED MIN MONOTONE-E2SAT *has no polynomial time approximation scheme if NP* $\not\subseteq \cap_{\delta>0}$ *BTIME($2^{n^{\delta}}$)*.

Proof: The result is a consequence of Propositions 6, 9 and Theorem 13 (Khot's result [Kho04]). □

For sparse instances, we show that BALANCED MIN MONOTONE-E2SAT and DENSEST $\frac{n}{2}$ SUBGRAPH are E-equivalent.

Proposition 10 SPARSE BALANCED MIN MONOTONE-E2SAT *is E-reducible to* SPARSE DENSEST $\frac{n}{2}$ SUBGRAPH.

Proof: Given an instance I of BALANCED MIN MONOTONE-E2SAT on n variables $x_1, \ldots, x_n$ and m clauses, we construct an instance G of DENSEST $\frac{n}{2}$ SUBGRAPH on n vertices and m edges as follows : we associate a vertex i for each variable x_i and an edge (i,j) for each clause $x_i \vee x_j$. Given an optimum balanced solution for I, consider S the set of variables set to 0 in this assignment. Thus $opt(G) \geq m - opt(I)$. Given a set S of size $\frac{n}{2}$ in G of value $val(I',S)$, the following assignment (called v) $x_i = 0$ if $i \in S$ and $x_i = 1$ otherwise, satisfies $m - val(I',S)$ clauses in I. Thus $opt(G) = m - opt(I) \leq c'opt(I)$ for a constant c' since $m = cn$ and $opt(I) \geq \frac{n}{4}$. Also $\varepsilon(I,v) = \frac{val(I,v)}{opt(I)} - 1 \leq \varepsilon(I',S)\frac{val(I',S)}{m-opt(I')} \leq c''\varepsilon(I',S)$ since $val(I',S) \leq m, opt(I) \leq \frac{n}{4}$ and I is a sparse instance. □

We first show that a hardness approximation result for BALANCED MIN MONOTONE-E2SAT implies a hardness approximation result for MIN BISECTION.

Proposition 11 BALANCED MIN MONOTONE-E2SAT *is E-reducible to* MIN BISECTION.

Proof: Given an instance I of BALANCED MIN MONOTONE-E2SAT on n variables $x_1, \ldots, x_n$ and m clauses, we construct an instance I' of MIN BISECTION on $n + 2$ variables $x_1, \ldots, x_n$ and two new variables y and z and $3m$ equations as follows : for each clause $x_1 \vee x_2$ we add 3 equations $x_1 \oplus x_2 = 1, x_1 \oplus z = 1, x_2 \oplus z = 1$. We have $opt(I') \leq 2opt(I)$ since the assignment satisfying $opt(I)$ clauses in I and $z = 0$ and $y = 1$ satisfies $2opt(I)$ equations in I'. Given a balanced assignment v for I' satisfying val' equations, we can consider $z = 0$. If $y = 1$ then, the restriction of v on x variables is balanced and satisfies $\frac{val'}{2}$ clauses. If $y = 0$ then the restriction of v on x variables satisfies $\frac{val'}{2}$ clauses in I but is not balanced. Observe that the balanced assignment obtained by changing the value of an x variable from 1 to 0 satisfies at most $\frac{val'}{2}$ clauses. □

We establish now an E-reduction between BALANCED MAX UNCUT and MIN BISECTION.

Proposition 12 BALANCED MAX E2-0H-LIN2 *is E-reducible to* BALANCED MIN E2-1H-LIN2.

Proof: Given an instance I of BALANCED MAX E2-0H-LIN2 on n variables $x_1, \ldots, x_n$ and m equations, we construct an instance I' on the same set of variables as follows. For each equation $x_{i_1} \oplus x_{i_2} = 0$ from I we add in I' the equation $x_{i_1} \oplus x_{i_2} = 1$. We can check easily that this is an E-reduction. □

Proposition 13 BALANCED MIN MONOTONE-EkSAT *is E-reducible to* BALANCED MIN MONOTONE-E$(k+1)$SAT, *for* $k \geq 2$.

Proof: Given an instance I of BALANCED MIN MONOTONE-EkSAT on n variables $x_1, \ldots, x_n$ and m clauses, we construct an instance I' of BALANCED MIN MONOTONE-E$(k+1)$SAT on the same set of variables at which we add two variables y and z as follows. For each clause $x_{i_1} \vee \ldots \vee x_{i_k}$ from I we add in I' the clause $x_{i_1} \vee \ldots \vee x_{i_k} \vee z$. It is easy to see that from an optimal balanced assignment that satisfies $opt(I)$ clauses in I we can construct a balanced assignment in I' (considering the same assignment for variables x and considering $z = 0$ and $y = 1$) that satisfies also $opt(I)$ clauses. Thus $opt(I') \leq opt(I)$. Given a balanced assignment of I', if y and z have different values then we can suppose that $z = 0$ and $y = 1$. Otherwise if y and z have the same value 0 then we can swap the value of y and swap also the value of a variable x that is set to 1 obtaining a balanced assignment with a better value. The restriction of this assignment to variables x is a balanced assignment satisfying the same number of clauses. $\square$

Theorem 19 BALANCED MIN MONOTONE-EkSAT, $k \geq 2$, *has no polynomial time approximation scheme if* $NP \not\subseteq \cap_{\delta>0} BTIME(2^{n^\delta})$.

Proof: This is a consequence of Theorem 18 and Proposition 13. $\square$

MIN E2-0H-LIN2 is MIN UNCUT that is known to be APX-hard by [GVY93] and thus BALANCED MIN E2-0H-LIN2 is BALANCED MIN UNCUT is also APX-hard. MIN E2-1H-LIN2 that is MIN CUT is polynomial solvable. BALANCED MIN E2-1H-LIN2 is MIN BISECTION for which the hardness of approximation was proved very recently [Kho04]. For $k \geq 3$, MIN Ek-1H-LIN2 is trivial since the assignment 0 for all variables satisfies no equation. When k is odd, MIN Ek-0H-LIN2 is also trivial since the assignment 1 for all variables satisfies no equation and when $k \geq 4$ is even, it is not known if MIN Ek-0H-LIN2 is hard to approximate.

Theorem 20 ([HK03]) BALANCED MIN E3-bH-LIN2, $b \in \{0,1\}$, *is NP-hard to approximate within any constant factor.*

The proof of the above result uses a PCP technique. We can prove now without using directly a PCP method a somewhat weaker result:

Theorem 21 BALANCED MIN Ek-bH-LIN2, $b \in \{0,1\}$, *is APX-hard for every* $k \geq 3$.

Proof: We construct an E-reduction from BALANCED MAX E3-0H-LIN2 to BALANCED MIN E3-1H-LIN2 and in a similar way an E-reduction from BALANCED MAX E3-1H-LIN2 to BALANCED MIN E3-0H-LIN2. Given an instance I of BALANCED MAX E3-0H-LIN2 on n variables $x_1, \ldots, x_n$ and m equations, we construct an instance I' of BALANCED MIN E3-1H-LIN2 on the same set of variables as follows. For each equation $x_{i_1} \oplus x_{i_2} \oplus x_{i_3} = 0$ from I we add in I' the equation $x_{i_1} \oplus x_{i_2} \oplus x_{i_3} = 1$. We can check easily that this is an E-reduction. $\square$

7   Summary of the Chains of Reductions

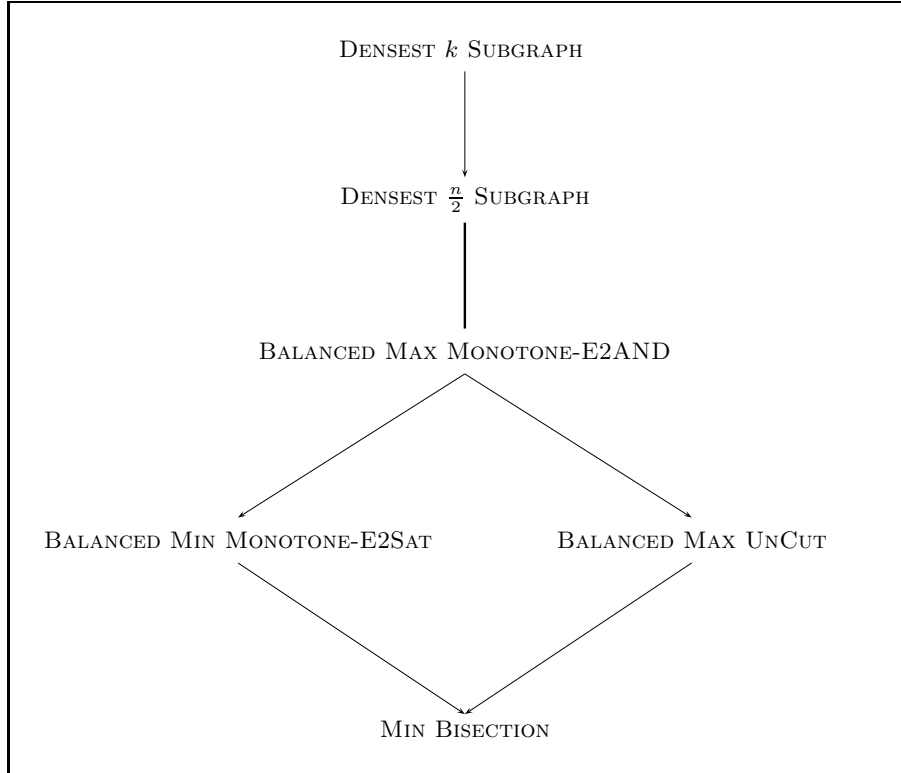We summarise our inapproximability results by the following chains of reductions (Figure 1).

Figure 1: Chains of E-reductions (bold line denotes an E-equivalence)

8   Dense Instances

In the previous chains of E-reductions the *everywhere density* is preserved except for the first reduction. Thus, since MIN BISECTION has a polynomial time approximation scheme on everywhere instances [AKK95], all the above problems have a polynomial time approximation scheme on everywhere dense instances. Observe that the maximization problems DENSEST k SUBGRAPH with $k = \Theta(n)$, BALANCED MAX MONOTONE-E2AND and BALANCED MAX UNCUT have a polynomial time approximation scheme even for average dense instances, while BALANCED MIN MONOTONE-E2SAT and MIN BISECTION have no polynomial time approximation scheme for average dense instances [AKK95, BFK03].

9   Further Research

The results of our paper shed some light on the global classes of optimization problems connected to the *constraint satisfaction problems*. There are still several problems which are open. Can some of our inapproximability bounds be dramatically improved ?

Another important question is to find an alternative (to Khot's technique [Kho04]) method of proving hardness results on some level of our chains in Section 7. Perhaps some "pure" and explicit reduction methods can be developed for that. Another intriguing question is to construct approximation algorithms for global optimization problems studied in our paper with more *satisfactory* approximation factors than the best known up to now.

References

[AKK95] S. Arora, D. Karger and M. Karpinski, *Polynomial time approximation schemes for dense instances of NP-hard problems*, Proceedings of 27th Annual ACM Symposium on the Theory of Computing, 1995, 284–293; also published in Journal of Computer and System Sciences 58, 1999, 193–210.

[ARV04] S. Arora, S. Rao and U. Vazirani, *Expander flows and a $\sqrt{\log n}$-approximation to sparsest cut*, Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC 2004), 222 – 231.

[ACG+99] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protasi, *Complexity and approximation. Combinatorial optimization problems and their approximability properties*, Springer, 1999.

[BFK03] C. Bazgan, W. Fernandez de la Vega and M. Karpinski, *Polynomial time approximation schemes for dense instances of the minimum constraint satisfaction*, Random Structures and Algorithms, 23(1), 2003, 73–91.

[BM02] M. Blaser and B. Manthey, *Improved Approximation Algorithms for Max 2Sat with Cardinality Constraint*, The 13th Annual International Symposium on Algorithms and Computation (ISAAC 2002), LNCS 2518, 2002, 187–198.

[Cre95] N. Creigou, *A dichotomy theorem for maximum generalized satisfiability problems*, Journal of Computer and System Sciences, 51(3), 1995, 511–522.

[CKS01] N. Creignou, S. Khanna and M. Sudan, *Complexity Classifications of Boolean Constraint Satisfaction Problems*, SIAM Monographs on Discrete Mathematics and Applications, 2001.

[CP91] P. Crescenzi and A. Panconesi, *Completeness in approximation classes*, Information and Computation, 93, 1991, 241–262.

[DMS99] I. Dumer, D. Micciancio and M. Sudan, *Hardness of Approximating the Minimum Distance of a Linear Code*, Proceedings of the 40th Annual Symposium on Foundations of Computer Science, 1999, 475–484.

[Fei02] U. Feige, *Relations between average case complexity and approximation complexity*, Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC 2002), 534–543.

[FK00] U. Feige and R. Krauthgamer, *A polylogarithmic approximation of the Minimum Bisection*, Proceedings of the 41st Annual Symposium on Foundations of Computer Science, 2000, 105–115.

[FL01]   U. Feige and M. Langberg, *Approximation Algorithms for Maximization Problems arising in Graph Partitioning*, Journal of Algorithms 41 (2001), 174–211.

[GJ76]   M. R. Garey and D. S. Johnson, *Computers and intractability. A guide to the theory of NP-completeness*, Freeman, C.A. San Francisco (1979).

[GVY93]   N. Garg, V. Vazirani and M. Yannakakis, *Approximate max-flow min-(multi)cut theorems and their applications*, Proceedings of the 25th Annual ACM Symposium on Theory of Computing (STOC 1993), 698–707.

[HZ01]   E. Halperin, U. Zwick, *A unified framework for obtaining improved approximation algorithms for maximum graph bisection problems*, Proceedings of the 8th Conference on Integer Programming and Combinatorial Optimization (IPCO 2001), 210–225.

[Has97]   J. Håstad, *Some optimal inapproximability results*, Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC 1997), 1-10.

[Hof03]   T. Hofmeister, *An Approximation Algorithm for Max 2Sat with Cardinality Constraint*, Proceedings of the 11th Annual European Symposium on Algorithms, 301–312, 2003.

[Hol02]   J. Holmerin, *PCP with Global Constraints – Balanced Homogeneous Linear Equations*, Manuscript, 2002.

[HK03]   J. Holmerin and S. Khot, *A strong inapproximability result for a generalization of Minimum Bisection*, Proceedings of the 18th IEEE Conference on Computational Complexity, 371–378, 2003.

[HK04]   J. Holmerin and S. Khot, *A new PCP Outer Verifier with Applications to Homogeneous Linear Equations and Max-Bisection*, Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC 2004), 11–17.

[JS04]   G. Jger, A. Srivastav, *Improved Approximation Algorithms for Maximum Graph Partitioning Problems*, Proceedings of the 24th International Conference Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2004), 348–359.

[KMSV94] S. Khanna, R. Motwani, M. Sudan and U. Vazirani, *On syntactic versus computational views of approximability*, Proceedings of the 35th Annual IEEE Annual Symposium on Foundations of Computer Science, 819–830, 1994, Also published in SIAM Journal on Computing, 28(1), 1999, 164–191.

[KS96]   S. Khanna and M. Sudan, *The optimization complexity of constraint satisfaction problems*, Technical note STAN-CS-TN-96-29, Stanford University, CA, 1996.

[KST97]   S. Khanna, M. Sudan and L. Trevisan, *Constraint Satisfaction: the Approximability of Minimization Problems*, Proceedings of 12th IEEE Computational Complexity, 1997, 282–296.

[KSTW01] S. Khanna, M. Sudan, L. Trevisan and D. P. Williamson, *The Approximability of Constraint Satisfaction Problems*, SIAM Journal of Computing, 30(6), 1863-1920 (2001).

[KSW97] S. Khanna, M. Sudan and D. Williamson, *A Complete Classification of the Approximability of Maximization Problems Derived from Boolean Constraint Satisfaction*, Proceedings of 29th ACM Symposium on Theory of Computing (STOC 1997), 11–20.

[Kho04] S. Khot, *Ruling Out PTAS for Graph Min-Bisection, Densest Subgraph and Bipartite Clique*, Proceedings of the 45th Annual IEEE Annual Symposium on Foundations of Computer Science (FOCS 2004), 136–145.

[PY91] C. H. Papadimitriou and M. Yannakakis, *Optimization, approximation, and complexity classes*, Journal of Computing and System Science, 43 (1991), 425-440.

[Sch78] T. Schaefer, *The complexity of satisfiability problems*, In Conference Record of the 10th Annual ACM Symposium on Theory and Computing (STOC 1978), 216–226.

[Svi01] M. I. Sviridenko, *Best possible approximation algorithm for MAX SAT with cardinality constraint*, Algorithmica, 30(3), 2001, 398–405.