Strong lower bounds on the approximability of some NPO PB-complete maximization problems

Viggo Kann
Department of Numerical Analysis and Computing Science
Royal Institute of Technology, S-100 44 Stockholm
viggo@nada.kth.se
Fax: +46-8 790 09 30

January 2, 1995

Abstract

The approximability of several NP maximization problems is investigated and strong lower bounds for the studied problems are proved. For some of the problems the bounds are the best that can be achieved, unless P = NP.

For example we investigate the approximability of Max PB $0-1$ Programming, the problem of finding a binary vector x that satisfies a set of linear relations such that the objective value $\sum c_i x_i$ is maximized, where c_i are binary numbers. We show that, unless P = NP, Max PB $0-1$ Programming is not approximable within the factor $n^{1-\varepsilon}$ for any $\varepsilon > 0$, where n is the number of inequalities, and is not approximable within $m^{1/2-\varepsilon}$ for any $\varepsilon > 0$, where m is the number of variables.

Similar hardness results are shown for other problems on binary linear systems, some problems on the satisfiability of boolean formulas and the longest induced cycle problem.

1 Introduction

Approximation of NP-complete optimization problems is a very interesting and active area of research. Since all NP-complete problems are reducible to each other one could suspect that they should have similar approximation properties, but this is not at all the case.

The range of approximability of NP-complete problems stretches from problems that are approximable within every constant in polynomial time, e.g. the knapsack problem [7], to problems that are not approximable within n^ε for some $\varepsilon > 0$, where n is the size of the input instance, unless P = NP. A problem that is known to be this hard to approximate is the minimum independent dominating set problem (minimum maximal independence number) [6].

Even optimization problems whose objective function is bounded by a polynomial in the size of the input may be hard to approximate. Krentel defined a class of optimization problems called OptP$[\log n]$, that consists of all NP optimization problems that are polynomially bounded [11]. This class, which we will call NPO PB, can be divided into two classes, MAX PB and MIN PB, containing maximization and minimization problems respectively [10]. Berman and Schnitger started to investigate the approximability of MAX PB problems and proved that there are MAX PB-complete problems, i.e. MAX PB problems to which every MAX PB problem can be reduced using an approximation preserving reduction [3]. Several problems are now known to be MAX PB-complete [8].

Later, some minimization problems were shown to be MIN PB-complete [9], and recently Crescenzi, Kann and Trevisan proved that any MIN PB-complete problem is NPO PB-complete and that any MAX PB-complete problem is NPO PB-complete [5]. The classes of MIN PB-complete, MAX PB-complete and NPO PB-complete problems thus coincide.

For every NPO PB-complete problem there is a constant $\alpha > 0$ such that the problem is not approximable within n^{α}, where n is the size of the input instance, unless P = NP. For some problems, for example minimum independent dominating set, this constant can be chosen arbitrarily close to 1, which means that these problems are incredible hard to approximate.

The problems known to be this hard to approximate are mainly minimization problems. Only a few maximization problems are known to have such an extreme nonapproximability bound, and they are all problems on graphs where one looks for a maximum induced connected subgraph [12]. The problem MIN DISTINGUISHED ONES, where one look for a satisfying boolean variable assignment containing as few true variables as possible from some distinguished set of variables, is NPO PB-complete and not approximable within $n^{1-\varepsilon}$, where n is the number of distinguished variables [9]. The corresponding maximization problem is also NPO PB-complete, but no strong lower bound on the approximability is known. One could ask whether minimization problems in some sense can be harder to approximate than maximization problems.

In this paper we will, however, show that this is not true. We will show that several maximization problems, for example MAX DISTINGUISHED ONES, have nonapproximability bounds similar to $n^{1-\varepsilon}$. We will do this by constructing approximation preserving reductions from either MIN INDEPENDENT DOMINATING SET or LONGEST INDUCED PATH and use the fact that these two problems have strong lower bounds on the approximability. We conclude that a convenient way to establish both NPO PB-completess results and strong lower bounds is to reduce from MIN INDEPENDENT DOMINATING SET or LONGEST INDUCED PATH. Note that our results do not need the quite complicated machinery of interactive proofs and the PCP model.

In the appendix all problems mentioned in the text are defined.

1.1 Preliminaries

Definition 1 An NP *optimization problem* A is a fourtuple $(I, sol, m, goal)$ such that

1. I is the set of the *instances* of A and it is recognizable in polynomial time.

2. Given an instance x of I, $sol(x)$ denotes the set of *feasible solutions* of x. These solutions are short, that is, a polynomial p exists such that, for any $y \in sol(x)$, $|y| \leq p(|x|)$. Moreover, it is decidable in polynomial time whether, for any x and for any y such that $|y| \leq p(|x|)$, $y \in sol(x)$.

3. Given an instance x and a feasible solution y of x, $m(x,y)$ denotes the positive integer *measure* of y (often also called the value of y). The function m is computable in polynomial time and is also called the *objective* function.

4. $goal \in \{\max, \min\}$.

The *class* NPO is the set of all NP optimization problems.

The goal of an NPO problem with respect to an instance x is to find an *optimum solution*, that is, a feasible solution y such that

$$m(x,y) = goal\{m(x,y') \ : \ y' \in sol(x)\}.$$

In the following *opt* will denote the function mapping an instance x to the measure of an optimum solution.

An NPO problem is said to be *polynomially bounded* if a polynomial q exists such that, for any instance x and for any solution y of x, $m(x,y) \leq q(|x|)$. The *class* NPO PB is the set of all polynomially bounded NPO problems. NPO PB = MAX PB$\cup$MIN PB where MAX PB is the set of all maximization problems in NPO PB and MIN PB is the set of all minimization problems in NPO PB.

Definition 2 Let A be an NPO problem. Given an instance x and a feasible solution y of x, we define the *performance ratio of y with respect to x* as

$$R(x,y) = \max\left\{\frac{m(x,y)}{m^*(x)}, \frac{m^*(x)}{m(x,y)}\right\}.$$

The performance ratio is always a number greater than 1 and is as close to 1 as the feasible solution is close to the optimum one.

Definition 3 Let A be an NPO problem and let T be an algorithm that, for any instance x of A, returns a feasible solution $T(x)$. Given an arbitrary function $r : N \rightarrow (1,\infty)$, we say that T is an *$r(n)$-approximate algorithm for A* if, for any instance x, the performance ratio of the feasible solution $T(x)$ with respect to x verifies the following inequality:

$$R(x, T(x)) \leq r(|x|).$$

Definition 4 Let A and B be two NPO problems. A is said to be PTAS-*reducible* to B, in symbols $A \leq_{\mathrm{PTAS}} B$, if three functions f, g, and c exist such that:

1. For any $x \in I_A$ and for any $r \in (1,\infty)$, $f(x,r) \in I_B$ is computable in time polynomial with respect to $|x|$.

2. For any $x \in I_A$, for any $y \in sol_B(f(x,r))$, and for any $r \in (1,\infty)$, $g(x,y,r) \in sol_A(x)$ is computable in time polynomial with respect to both $|x|$ and $|y|$.

3. $c : (1,\infty) \to (1,\infty)$ is invertible.

4. For any $x \in I_A$, for any $y \in sol_B(f(x,r))$, and for any $r \in (1,\infty)$,

$$R_B(f(x,r),y) \leq c(r) \text{ implies } R_A(x,g(x,y,r)) \leq r.$$

Proposition 1 *If $A \leq_{\text{PTAS}} B$ and $B \in$ APX (respectively, $B \in$ PTAS), then $A \in$ APX (respectively, $A \in$ PTAS).*

Definition 5 A problem $A \in$ NPO is NPO-*complete* if, for any $B \in$ NPO, $B \leq_{\text{PTAS}} A$. Similarly, a problem $A \in$ NPO PB is NPO PB-*complete* if, for any $B \in$ NPO PB, $B \leq_{\text{PTAS}} A$.

In the same way Max PB-complete and Min PB-complete problems can be defined. Recently was shown that a problem that is Max PB-complete or Min PB-complete is at the same time complete for the whole of NPO PB.

Proposition 2 ([5]) *Any* Min PB-*complete problem is* NPO PB-*complete, and any* Max PB-*complete problem is* NPO PB-*complete.*

The approximability for problems that are not approximable within a constant is usually described as a function of the size of the problem instance, or more precisely, as a function of some size parameter, like the number of nodes or edges in an input graph. The Ptas-reduction does not preserve size parameters, so this reduction cannot be used when investigating the approximability (or nonapproximability) of problems that are very hard to approximate, like NPO PB-complete problems. For such problems it is not relevant whether the reduction increases the performance ratio by a constant factor. We would like a reduction that just guarantees that the performance ratio is preserved within *some* constant factor, but has full control over the increase of the size if the problem instance. We will use the S-reduction, defined in [9].

Definition 6 ([9]) Let A and B be two NPO problems. A is said to be *S-reducible* to B with size amplification a if there exist three functions f, g, a, and a positive constant c such that:

1. For any $x \in I_A$, $f(x,r) \in I_B$ is computable in time polynomial with respect to $|x|$.

2. For any $x \in I_A$, for any $y \in sol_B(f(x))$, $g(x,y) \in sol_A(x)$ is computable in time polynomial with respect to both $|x|$ and $|y|$.

3. $a : R^+ \to R^+$ is a monotonously increasing positive and computable function.

4. For any $x \in I_A$, for any $y \in sol_B(f(x))$,

$$R_A(x, g(x,y)) \leq R_B(f(x), y).$$

5. For any $x \in I_A$, $|f(x)| \leq a(|x|)$.

Proposition 3 ([9]) *Given two* NPO *problems F and G, if there is an S-reduction from F to G with size amplification $a(n)$ and G is approximable within some monotonously increasing positive function $u(n)$ of the size of the input instance, then F is approximable within $c \cdot u(a(n))$, which is a monotonously increasing positive function. Conversely, if F is not approximable within $c \cdot u(a(n))$, then G is not approximable within $u(n)$.*

For constant and polylogarithmic approximable problems the S-reduction preserves approximability within a constant for any polynomial size amplification, since $c \log^k(n^p) = p^k c \log^k n = O(\log^k n)$. For n^c approximable problems it only does this for size amplification $O(n)$, since $c \cdot (O(n))^c = O(n^c)$.

2 Lower bounds

In this section we will prove lover bounds on the approximability of the following NPO PB-complete problems:

- Max PB $0-1$ Programming [3],
- Max Number of Satisfiable Formulas [13],
- Max Distinguished Ones [13],
- Max Ones [13],
- Max C Bin Sat$^{\mathcal{R}_1;\mathcal{R}_2}$ (maximum constrained binary satisfiable linear subsystem) [1],
- Max Bin Irrelevant Sat$^{\mathcal{R}}$ (maximum irrelevant binary variables in linear system) [2],
- Longest Induced Cycle.

In the references above the problems are defined and are shown to be Max PB-complete. The problems are therefore NPO PB-complete by proposition 2. Formal definitions of the problems can be found in the appendix.

We first show that Max PB $0-1$ Programming is hard to approximate. This result was obtained as a side-effect in the proof of proposition 2 in [5]. We will then modify this proof to prove hardness results for the other problems.

Theorem 4 Max PB $0-1$ Programming *is not approximable within $n^{1-\varepsilon}$ for any $\varepsilon > 0$, where n is the number of inequalities, and is not approximable within $m^{1/2-\varepsilon}$ for any $\varepsilon > 0$, where m is the number of variables.*

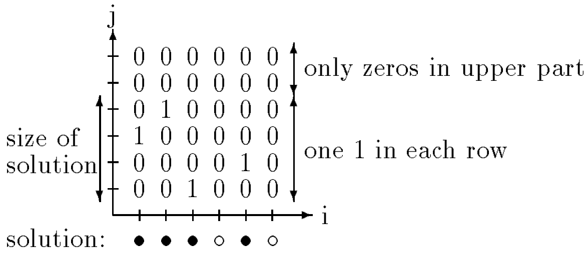

Figure 1: *The idea of the reduction from* Min Independent Dominating Set *to* Max PB $0-1$ Programming. *The variable $x_i^j = 1$ if and only if v_i is the jth node in the solution. There is at most one 1 in each column and in each row.*

Proof Halldórsson has proved that, unless P = NP, Min Independent Dominating Set is not approximable within $n^{1-\varepsilon}$ for any $\varepsilon > 0$, where n is the sum of the number of nodes and edges in the graph [6]. We will use this fact to show that Max PB $0-1$ Programming is hard to approximate.

We will construct a reduction from Min Independent Dominating Set to Max PB $0-1$ Programming using the following idea. The objective function, i.e. the number of nodes in the independent dominating set is encoded by introducing an order of the nodes in the solution. The order is encoded by a squared number of $0-1$ variables in the programming problem, see figure 1. A solution of size 1 shall correspond to the $0-1$ programming objective value n, and a solution of size p shall correspond to an objective value of $\lfloor n/p \rfloor$.

Given an instance of Min Independent Dominating Set, i.e. a graph with nodes $V = \{v_1, \ldots, v_m\}$ and edges E, construct m^2 variables x_i^j, $1 \le i, j \le m$, n variables y_p, $1 \le p \le n$, and the following inequalities:

$$\forall i \in [1..m] \qquad \sum_{j=1}^{m} x_i^j \le 1 \qquad \text{(at most one 1 in each column)} \tag{1}$$

$$\forall j \in [1..m] \qquad \sum_{i=1}^{m} x_i^j \le 1 \qquad \text{(at most one 1 in each row)} \tag{2}$$

$$\forall j \in [1..m-1] \qquad \sum_{i=1}^{m} x_i^j - \sum_{i=1}^{m} x_i^{j+1} \ge 0 \qquad \text{(only zeros in upper part)} \tag{3}$$

$$\forall (v_i, v_j) \in E \qquad \sum_{k=1}^{m} x_i^k + \sum_{k=1}^{m} x_j^k \le 1 \qquad \text{(independence)} \tag{4}$$

$$\forall i \in [1..m] \qquad \sum_{k=1}^{m} x_i^k + \sum_{\substack{j:(v_i,v_j)\in E \\ k\in[1..m]}} x_j^k \ge 1 \qquad \text{(domination)} \tag{5}$$

$$\forall i \in [1..n] \qquad y_i \le \sum_{k=1}^{m} x_k^{\lfloor n/i \rfloor} \qquad \text{(objective variables)} \tag{6}$$

$$y_i \ge \sum_{k=1}^{m} x_k^{\lfloor n/i \rfloor} \tag{7}$$

The objective function is defined as $\sum_{p=1}^{n} y_p$.

One can now verify that an independent dominating set of size s will exactly correspond to a solution of the $0-1$ programming problem with objective value $\lfloor n/s \rfloor$ and vice versa.

Suppose that the minimum independent dominating set has size M, then the performance ratio s/M for the independent dominating set problem will correspond to the performance ratio

$$\frac{\lfloor \frac{n}{M} \rfloor}{\lfloor \frac{n}{s} \rfloor} = \frac{s}{M}\left(1 \pm \frac{m}{n}\right)$$

for the $0-1$ programming problem, where $\frac{m}{n}$ is the relative error due to the floor operation. By choosing n large enough the relative error can be made arbitrarily small, but for proving the theorem it is enough to choose $n = m$ to obtain a reduction that preserves the approximability within a factor of 2.

The reduction is obviously an S-reduction with amplification $O(m + |E|)$ with respect to the number of inequalities and amplification $O(m^2)$ with respect to the number of variables. The theorem now follows from proposition 3 together with the fact that Min Independent Dominating Set is not approximable within $n^{1-\varepsilon}$ for any $\varepsilon > 0$, where n is the sum of the number of nodes and edges in the graph [6]. □

Theorem 5 Max Number of Satisfiable Formulas *is not approximable within* $n^{1-\varepsilon}$ *for any* $\varepsilon > 0$*, where* n *is the number of formulas, unless* P = NP.

Proof We modify the construction of the proof of Theorem 4 in the following way.

Given a graph with nodes $V = \{v_1, \ldots, v_m\}$ and edges E, construct m^2 variables x_i^j, $1 \le i, j \le m$, m variables y_p, $1 \le p \le m$, and the m boolean formulas

$$\Phi_p = y_p \wedge \varphi,\ 1 \le p \le m,$$

where φ is the conjunction of the following formulas:

$$\forall i \in [1..m], 1 \le j < k \le m \qquad \overline{x_i^j} \vee \overline{x_i^k} \tag{8}$$

$$\forall i \in [1..m], 1 \le j < k \le m \qquad \overline{x_j^i} \vee \overline{x_k^i} \tag{9}$$

$$\forall j \in [1..m-1] \quad \left(\bigvee_{k=1}^m x_k^{j+1}\right) \Rightarrow \left(\bigvee_{k=1}^m x_k^j\right) \tag{10}$$

$$\forall (v_i, v_j) \in E, \forall k, l \in [1..m] \qquad \overline{x_i^k} \vee \overline{x_j^l} \tag{11}$$

$$\forall i \in [1..m] \qquad \bigvee_{k=1}^m x_i^k \vee \bigvee_{\substack{j:(v_i,v_j)\in E \\ k\in[1..m]}} x_j^k \tag{12}$$

$$\forall p \in [1..m] \qquad y_p \equiv \bigvee_{k=1}^m x_k^{\lfloor m/i \rfloor} \tag{13}$$

It is clear that the boolean formulas (8–13) restrict the variables in exactly the same way as the inequalities (1–7) and that the number of satisfied formulas $y_p \wedge \varphi, 1 \le p \le m$ is the same as the sum $\sum_{p=1}^m y_p$ in Max PB $0-1$ Programming. Therefore Max Number of Satisfiable Formulas is not approximable within $n^{1-\varepsilon}$ for any $\varepsilon > 0$, where n is the number of formulas, unless P = NP. □

Theorem 6 MAX DISTINGUISHED ONES *is not approximable within* $n^{1-\varepsilon}$ *for any* $\varepsilon > 0$*, where* n *is the number of distinguished variables, and* MAX ONES *is not approximable within* $n^{1/3-\varepsilon}$ *for any* $\varepsilon > 0$*, where* n *is the number of variables, unless* P = NP*.*

PROOF First we prove the result for MAX DISTINGUISHED ONES. This time we modify the construction of the proof of Theorem 5. We first construct a formula with clauses of unbounded size and then rewrite the clauses as 3-SAT clauses.

Let $y_p, 1 \le p \le m$ be the distinguished variables. Apart from the $m^2 + m$ variables m_i^j and y_p we will need m variables t_p, $1 \le p \le m$, and m variables z_p, $1 \le p \le m$. We define these variables by

$$\forall p \in [1..m] \quad t_p \equiv \bigvee_{k=1}^{m} x_k^p \tag{14}$$

$$\forall p \in [1..m] \quad z_p \equiv \bigvee_{k=1}^{m} x_p^k \tag{15}$$

We then can reformulate the formulas (10), (12), and (13) as, respectively,

$$\forall j \in [1..m-1] \qquad \overline{t_{j+1}} \vee t_j \tag{16}$$

$$\forall i \in [1..m] \quad z_i \vee \bigvee_{j:(v_i,v_j)\in E} z_j \tag{17}$$

$$\forall i \in [1..m] \qquad y_i \equiv t_{\lfloor m/i \rfloor} \tag{18}$$

We can now rewrite the equivalences in (14), (15), and (18) as clauses using the rewriting rule $u \equiv v \longrightarrow (u \vee \overline{v}) \wedge (\overline{u} \vee v)$. All clauses now consist of 2 literals except (14), (15), and (17). We use the standard method of rewriting these as clauses of 3 literals. In this process we introduce $O(m^2 + |E|)$ new variables.

The conjunction of all these clauses is equivalent to the formula φ in the proof of theorem 5, and the number of true distinguished variables in a variable assignment will exactly correspond to the number of satisfied formulas in the other proof.

Since the number of distinguished variables is m, MAX DISTINGUISHED ONES is not approximable within $m^{1-\varepsilon}$ for any $\varepsilon > 0$, unless P = NP.

In order to formulate the problem as a MAX ONES problem instance we use an idea in [13]. We need to make a lot of copies of the distinguished variables to make each such variable more valuable than all the nondistinguished variables together. If there are N nondistinguished variables we introduce, for each distinguished variable y_p, $2N$ new variables $y_p^1, \ldots, y_p^{2N}$ and $2N$ pairs of clauses $(y_p^i \vee \overline{y}_p^{i+1}) \wedge (\overline{y}_p^i \vee y_p^{i+1})$ to assure that $y_p = y_p^1 = \cdots = y_p^{2N}$. A solution of this MAX ONES problem with objective value M can be transformed to a solution of the original MAX DISTINGUISHED ONES problem with objective value $\lfloor M/(2N+1) \rfloor$.

The constructed instance will have $2Nm + N = O(m^3)$ variables so the reduction from MIN INDEPENDENT DOMINATING SET to MAX ONES is an S-reduction with amplification $O(m^3)$. Thus MAX ONES is not approximable within $n^{1/3-\varepsilon}$ for any $\varepsilon > 0$, where n is the number of variables, unless P = NP.
□

Theorem 7 MAX C BIN SAT$^{\mathcal{R}_1;\mathcal{R}_2}$ *is not approximable within* $n^{1-\varepsilon}$ *where* n *is the number of optional relations, and not within* $n^{1/2-\varepsilon}$ *where* n *is the number of variables, for any* $\varepsilon > 0$, *and for any* $\mathcal{R}_1, \mathcal{R}_2 \in \{=, \geq, >, \neq\}$, *unless* P = NP.

PROOF We first show the result for the problem MAX C BIN SAT$^{\geq;\geq}$ and then extend the result to the other variants.

We construct an S-reduction from the MAX PB $0-1$ PROGRAMMING instances constructed in the proof of theorem 4 as follows.

As the mandatory relations we use the MAX PB $0-1$ PROGRAMMING inequalities. For $1 \leq p \leq m$ we construct the optional inequality

$$y_p \geq 1. \tag{19}$$

Thus we have a system with $m^2 + m$ binary variables, m optional relations and $O(m + |E|)$ mandatory ones.

A solution of this system that satisfies s optional relations will also be a solution of the MAX PB $0-1$ PROGRAMMING problem with the objective value s and vice versa. We have got an S-reduction from MIN INDEPENDENT DOMINATING SET to MAX C BIN SAT$^{\geq;\geq}$ with amplification $O(m + |E|)$ with respect to the number of inequalities and amplification $O(m^2)$ with respect to the number of variables. Once again the theorem follows from the fact that MIN INDEPENDENT DOMINATING SET is not approximable within $n^{1-\varepsilon}$ for any $\varepsilon > 0$, where n is the sum of the number of nodes and edges in the graph [6].

For the other constrained problems, MAX C BIN SAT$^{\mathcal{R}_1;\mathcal{R}_2}$, we use the same reduction as above but we express each inequality in another way.

For example, (1) can be expressed in the following ways.

$$\forall i \in [1..m] \quad \sum_{j=1}^{m} x_i^j \leq 1 \tag{20}$$

$$\forall i \in [1..m] \quad \sum_{j=1}^{m} x_i^j < 2 \tag{21}$$

$$\forall i \in [1..m] \quad \sum_{j=1}^{m} x_i^j = z_i \tag{22}$$

$$\forall i \in [1..m], 1 \leq j < k \leq m \quad x_i^j + x_i^k \neq 2 \tag{23}$$

The variables z_i in (22) are new slack variables not occurring in any other relation. In the same way we can express all mandatory relations as strict inequalities, equalities and nonequalities. The total number of slack variables needed in the case of mandatory equalities is $O(m + |E|)$.

We express the optional relations in the following way.

type $\geq$	type $>$	type $=$	type $\neq$
$y_p \geq 1$	$y_p > 0$	$y_p = 1$	$y_p \neq 0$

In each case we get $O(m^2)$ variables and m optional relations in the constructed instance. Hence the theorem follows.

We can also note that in each case except for mandatory nonequalities we get a total number of $O(m + |E|)$ relations. For these cases the problem is therefore not approximable within $n^{1-\varepsilon}$ for any $\varepsilon > 0$, where n is the total number of relations, unless P = NP. $\square$

Theorem 8 MAX BIN IRRELEVANT SAT$^{\mathcal{R}}$ *is not approximable within* $n^{1/3-\varepsilon}$ *for any* $\varepsilon > 0$*, and for any* $\mathcal{R} \in \{=, \geq, >, \neq\}$*, where* n *is the sum of the number of variables and the number of relations, unless* P = NP.

PROOF We extend the reduction from MIN INDEPENDENT DOMINATING SET to MAX C BIN SAT$^{\mathcal{R}_1;\mathcal{R}_2}$ in the proof of theorem 7 to a reduction from MIN INDEPENDENT DOMINATING SET to MAX BIN IRRELEVANT SAT$^{\mathcal{R}}$.

We start from the mandatory relations of the constructed instance with $\mathcal{R}_1 = \mathcal{R}_2 = \mathcal{R}$. Observe that the instance consists of $O(m^2)$ variables and $O(m + |E|)$ mandatory relations, except for the nonequality case where we have $O(m^3)$ mandatory relations. Since we count irrelevant variables, i.e. zero variables, we have to change every occurrence of y_p to $1 - y_p$, for $1 \leq p \leq m$.

We then use the duplication technique from the proof of theorem 6 to make each y_p variable more valuable than all the other $N = O(m^2)$ variables together. It is easy to express $y_p = y_p^1 = \cdots = y_p^{2N}$ by adding, for $1 \leq i \leq 2N$, relations of type $\mathcal{R}$ as in the following table.

type $\geq$	type $>$	type $=$	type $\neq$
$y_p \geq y_p^i$	$y_p > y_p^i - 1$	$y_p = y_p^i$	$y_p \neq 1 - y_p^i$
$y_p \leq y_p^i$	$y_p < y_p^i + 1$		

The constructed instance will have $2Nm + N = O(m^3)$ variables and $O(mN) = O(m^3)$ relations, so the reduction from MIN INDEPENDENT DOMINATING SET to MAX BIN IRRELEVANT SAT$^{\mathcal{R}}$ is an S-reduction with amplification $O(m^3)$ both in the number of variables and relations. MAX BIN IRRELEVANT SAT$^{\mathcal{R}}$ is therefore not approximable within $n^{1/3-\varepsilon}$ for any $\varepsilon > 0$, where n is the sum of the number of variables and the number of relations, unless P = NP. $\square$

Theorem 9 LONGEST INDUCED CYCLE *is not approximable within* $n^{1-\varepsilon}$ *for any* $\varepsilon > 0$*, where* n *is the number of nodes, unless* P = NP.

PROOF We reduce from the problem LONGEST INDUCED PATH that is known to be NPO PB-complete [3] and not approximable within $n^{1-\varepsilon}$ for any $\varepsilon > 0$, where n is the number of nodes, unless P = NP [12].

Given an input graph to LONGEST INDUCED PATH, $G = (V, E)$. We extend this graph by adding one new node w_i for each node $v_i \in V$ and add edges (w_i, v_i) for $1 \leq i \leq |V|$. We also add one special node w_0 that is connected to all new nodes, i.e. we have edges (w_0, w_i) for $1 \leq i \leq |V|$.

Every induced path in G (say, starting in node v_s and ending in node v_e) can now be extended to an induced cycle in the new graph by adding the nodes w_s, w_e and w_0. The cycle will have length 4+(length of induced path).

On the other hand, every induced cycle in the new graph that contains the node w_0 will give us an induced path in G containing 4 edges less than the cycle. An induced cycle that does not contain the special vertex cannot contain any w node, so by removing any node in the cycle we will get an induced path in G containing 2 edges less than the cycle.

Thus this is an S-reduction with node amplification $2|V|+1 = O(|V|)$, and thus LONGEST INDUCED CYCLE is not approximable within $n^{1-\varepsilon}$ for any $\varepsilon > 0$, where n is the number of nodes, unless P = NP. $\square$

3 Summary

The following table summarizes the nonapproximability results in the paper. For each result we give the name of the problem, the nonapproximability exponent b and the size parameter n, saying that the problem is not approximable within $n^{b-\varepsilon}$ for any $\varepsilon > 0$, unless P = NP.

problem	b	size parameter n
MAX PB $0-1$ PROGRAMMING	1	inequalities
MAX PB $0-1$ PROGRAMMING	1/2	variables
MAX NUMBER OF SATISFIABLE FORMULAS	1	formulas
MAX DISTINGUISHED ONES	1	distinguished variables
MAX DISTINGUISHED ONES	1/2	variables
MAX ONES	1/3	variables
MAX C BIN SAT$^{\mathcal{R}_1;\mathcal{R}_2}$	1	optional relations
MAX C BIN SAT$^{\mathcal{R}_1;\mathcal{R}_2}$	1/2	variables
MAX BIN IRRELEVANT SAT$^{\mathcal{R}}$	1/3	variables+relations
LONGEST INDUCED CYCLE	1	nodes

References

[1] E. Amaldi and V. Kann. The complexity and approximability of finding maximum feasible subsystems of linear relations. Technical Report ORWP-11-93, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne, 1993. Theoretical Comput. Sci., to appear.

[2] E. Amaldi and V. Kann. On the approximability of removing the smallest number of relations from linear systems to achieve feasability. Technical Report ORWP-6-94, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne and Technical Report TRITA-NA-9402, ISSN 0348-2952, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, 1994. Presented at 15th International Symposium on Mathematical Programming, Ann Arbor, Michigan, August.

[3] P. Berman and G. Schnitger. On the complexity of approximating the independent set problem. *Inform. and Comput.*, 96:77–94, 1992.

[4] P. Crescenzi and V. Kann. A compendium of NP optimization problems. Continuously updated manuscript. Available by anonymous ftp from `nada.kth.se/Theory/Viggo-Kann/compendium.ps.Z`, 1994.

[5] P. Crescenzi, V. Kann, and L. Trevisan. Natural complete and intermediate problems in approximation classes. Manuscript, 1994.

[6] M. M. Halldórsson. Approximating the minimum maximal independence number. *Inform. Process. Lett.*, 46:169–172, 1993.

[7] O. H. Ibarra and C. E. Kim. Fast approximation for the knapsack and sum of subset problems. *Journal of the ACM*, 22(4):463–468, 1975.

[8] V. Kann. *On the Approximability of NP-complete Optimization Problems.* PhD thesis, Department of Numerical Analysis and Computing Science, Royal Institute of Technology, Stockholm, 1992.

[9] V. Kann. Polynomially bounded minimization problems which are hard to approximate. In *Proc. of 20th International Colloquium on Automata, Languages and Programming*, pages 52–63. Springer-Verlag, 1993. Lecture Notes in Comput. Sci. 700. Nordic J. Computing, to appear.

[10] P. G. Kolaitis and M. N. Thakur. Logical definability of NP optimization problems. Technical Report UCSC-CRL-93-10, Board of Studies in Computer and Information Sciences, University of California at Santa Cruz, 1993. To be published in Information and Computation.

[11] M. W. Krentel. The complexity of optimization problems. *J. Comput. System Sci.*, 36:490–509, 1988.

[12] C. Lund and M. Yannakakis. The approximation of maximum subgraph problems. In *Proc. of 20th International Colloquium on Automata, Languages and Programming*, pages 40–51. Springer-Verlag, 1993. Lecture Notes in Comput. Sci. 700.

[13] A. Panconesi and D. Ranjan. Quantifiers and approximation. *Theoretical Comput. Sci.*, 107:145–163, 1993.

Here follows a list of definitions of problems mentioned in the text. A much larger list of NPO problems can be found in [4].

A A List of NPO Problems

A.1 Min Independent Dominating Set

INSTANCE: Graph $G = (V, E)$.
SOLUTION: An independent dominating set for G, i.e., a subset $V' \subseteq V$ such that for all $u \in V - V'$ there is a $v \in V'$ for which $(u, v) \in E$, and such that no two vertices in V' are joined by an edge in E.
MEASURE: Cardinality of the independent dominating set, i.e., $|V'|$.

A.2 Max PB $0-1$ Programming

Instance: Integer $m \times n$-matrix $A \in Z^{m \cdot n}$, integer m-vector $b \in Z^m$, nonnegative binary n-vector $c \in \{0,1\}^n$.
Solution: A binary n-vector $x \in \{0,1\}^n$ such that $Ax \geq b$.
Measure: The scalar product of c and x, i.e., $\sum_{i=1}^{n} c_i x_i$.

A.3 Max Number of Satisfiable Formulas

Instance: Set U of variables, collection C of 3CNF formulas.
Solution: A subset $C' \subseteq C$ of the formulas such that there is a truth assignment for U that satisfies every formula in C'.
Measure: Number of satisfied formulas, i.e., $|C'|$.

A.4 Max Distinguished Ones

Instance: Disjoint sets X, Z of variables, collection C of disjunctive clauses of at most 3 literals, where a literal is a variable or a negated variable in $X \cup Z$.
Solution: Truth assignment for X and Z that satisfies every clause in C.
Measure: The number of Z variables that are set to true in the assignment.

A.5 Max Ones

Instance: Set X of variables, collection C of disjunctive clauses of at most 3 literals, where a literal is a variable or a negated variable.
Solution: Truth assignment that satisfies every clause in C.
Measure: The number of variables that are set to true in the assignment.

A.6 Max C Bin Sat$^{\mathcal{R}_1;\mathcal{R}_2}$

Instance: $\mathcal{R}_1, \mathcal{R}_2 \in \{=, \geq, >, \neq\}$ defining the types of relations. Systems $A_1 x_1 \mathcal{R}_1 b_1$ and $A_2 x_2 \mathcal{R}_2 b_2$ of linear relations, where A_1 and A_2 are integer matrices, and b_1 and b_2 are integer vectors. The first system is called the *mandatory* relations and the second system the *optional* relations.
Solution: Two vectors x_1 and x_2 of binary numbers such that all mandatory relations, $A_1 x_1 \mathcal{R}_1 b_1$, are satisfied.
Measure: The number of optional relations, $A_2 x_2 \mathcal{R}_2 b_2$, that are satisfied by x_2.

A.7 Max Bin Irrelevant Sat$^{\mathcal{R}}$

Instance: $\mathcal{R} \in \{=, \geq, >, \neq\}$ defining the type of relations. System $Ax\mathcal{R}b$ of linear relations, where A is an integer matrix, and b is an integer vector.
Solution: A vector x of binary numbers such that all relations, $Ax\mathcal{R}b$, are satisfied.
Measure: The number of zero elements in x, i.e., $|\{i : x_i = 0\}|$.

A.8 Longest Induced Cycle

Instance: Graph $G = \langle V, E \rangle$.
Solution: A subset $V' \subseteq V$ such that the subgraph induced by V' is a cycle.
Measure: Length of the cycle, i.e., $|V'|$.