On the hardness of approximating $\mathcal{NP}$ witnesses

Uriel Feige, Michael Langberg, and Kobbi Nissim

Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, {feige,mikel,kobbi}@wisdom.weizmann.ac.il.

Abstract. The *search* version for $\mathcal{NP}$-complete combinatorial optimization problems asks for finding a solution of optimal value. Such a solution is called a *witness*. We follow a recent paper by Kumar and Sivakumar, and study a relatively new notion of approximate solutions that ignores the value of a solution and instead considers its syntactic representation (under some standard encoding scheme).

The results that we present are of a negative nature. We show that for many of the well known $\mathcal{NP}$-complete problems (such as 3-SAT, CLIQUE, 3-COLORING, SET COVER) it is $\mathcal{NP}$-hard to produce a solution whose Hamming distance from an optimal solution is substantially closer than what one would obtain by just taking a random solution. In fact, we have been able to show similar results for most of Karp's 21 original $\mathcal{NP}$-complete problems. (At the moment, our results are not tight only for UNDIRECTED HAMILTONIAN CYCLE and FEEDBACK EDGE SET.)

1 Introduction

Every language $L \in \mathcal{NP}$ is characterized by a polynomial-time decidable relation $\mathcal{R}_L$ such that $L = \{x | (w,x) \in \mathcal{R}_L \text{ for some } w\}$. Such a characterization intuitively implies that every $x \in L$ has a polynomial *witness* w for being an instance of L. For example in the language 3SAT the witness for a formula ϕ being satisfiable is a satisfying assignment a so that $\phi(a) = 1$.

Let ϕ be a satisfiable 3SAT formula. The problem of constructing a satisfying assignment for ϕ is $\mathcal{NP}$-hard. This motivates an attempt to achieve an efficient *approximation* for a satisfying assignment. There are several possible notions of approximation. For instance, one may attempt to find an assignment that satisfies many of the clauses in ϕ. This is the standard notion of approximation that has been studied extensively. Given a formula ϕ in which each clause is of size exactly three, the trivial algorithm that chooses an assignment at random is expected to satisfy at least 7/8 of the clauses. This is essentially best possible, as it is $\mathcal{NP}$-hard to find an assignment that satisfies more than a $7/8 + \varepsilon$ fraction of the clauses [10]. We consider a different notion of approximation presented in [12] in which one seeks to find an assignment which agrees with some satisfying assignment on many variables. A random assignment is expected to agree with some satisfying assignment on at least half of the variables. We show that it is $\mathcal{NP}$-hard to find an assignment that agrees with some satisfying assignment on significantly more than half of the variables. Our hardness results are based on Karp (many-to-one) reductions and thus also exclude the existence of randomized algorithms that approximate some satisfying assignment with nonnegligible probability (under the assumption that $\mathcal{NP}$ does not have expected polynomial time algorithms).

Recently, a randomized algorithm for finding a satisfying assignment for 3SAT formulas was presented by Schoning [13]. This algorithm has a running time of $O((4/3)^n)$, which is the best currently known. Roughly speaking, the algorithm generates an initial assignment that hopefully agrees with some satisfying assignment on significantly more than half of the variables, and then searches its neighborhood. The initial assignment is chosen at random, with exponentially small success probability. The ability to find such an assignment in polynomial time would improve the running time of this algorithm. Our results essentially show the unlikeliness of improving the algorithm in this way.

A natural question is whether the hardness of approximating witnesses phenomenon holds for $\mathcal{NP}$-complete problems other than 3SAT. The answer depends to some extent on the syntactic representation of witnesses. As shown in [12], if we are given sufficient freedom in how to encode witnesses, then the hardness of approximating witnesses extends to all other $\mathcal{NP}$-complete problems. However, it is much more interesting to examine the witness approximation issue under more *natural* encodings of witnesses.

We examine a large set of $\mathcal{NP}$-complete problems, namely Karp's original list of 21 problems [11]. For each of these problems we consider a "natural" syntactic representation for witnesses. Under this representation, we prove tight hardness results similar to those proved for 3SAT. For example, we encode witnesses for 3-COLORING in ternary rather than binary, representing the colors of the vertices. In this case a random 3-coloring is expected to agree with some valid 3-coloring on only 1/3 of the vertices, rather than 1/2. We show that given a 3-colorable graph $G = (V, E)$ one cannot find a 3-coloring that agrees with some valid 3-coloring of V on significantly more than 1/3 of the vertices. As another example, the encoding that we use for the HAMILTONIAN CYCLE problem is an indicator vector for the edges that participate in a Hamiltonian cycle. We consider graphs that contain roughly $m = 2n$ edges, in which case a random indicator vector is within Hamming distance roughly $m/2$ from an indicator vector of a Hamiltonian cycle. We show that it is NP-hard to find an indicator vector of distance significantly less than $m/2$ from a witness in the directed version of HAMILTONIAN CYCLE.

$\mathcal{NP}$-complete languages exhibit diverse behavior with respect to the standard notion of approximation. On the other hand, our results indicate that with respect to our notion of witness inapproximability, many $\mathcal{NP}$-complete problems share the property that their witnesses cannot be approximated better than random. The only problems from [11] for which we were unable to fully demonstrate this phenomenon is the undirected version of HAMILTONIAN CYCLE, and the FEEDBACK EDGE SET problem (for the latter we have tight results on graphs with parallel edges). At this point, it is not clear if this is just a matter of constructing more clever gadgets, whether it is an issue of choosing a different encoding scheme for witnesses (e.g., for HAMILTONIAN CYCLE one may encode a solution as an ordered list of vertices rather than as an indicator vector for edges), or whether there is a deeper reason for these two exceptions.

Previous work : As mentioned earlier, Kumar and Sivakumar [12] examine witness approximation properties of any language in $\mathcal{NP}$ when we are free to choose the syntactic representation of witnesses. They show that for any $\mathcal{NP}$ language L there exists a witness relation $\mathcal{R}_L$ such that it is $\mathcal{NP}$-hard given an instance $x \in L$ to produce an *approximate* witness $\hat{w}$ that agrees with some witness w of x on significantly more than half its bits.

In our notion of witness approximation, one has to produce a solution of small Hamming distance from a valid witness, but one need not know in which bits of the solution the difference lies. A different problem that is well studied is whether one can recover with certainty just one bit of a witness (e.g., the value of one variable in a satisfying assignment for a 3SAT formula ϕ). The answer to this is known to be negative (for typical $\mathcal{NP}$-complete problems), due to the self reducibility property (e.g., ϕ can then be reduced to a new formula with one less variable). Making use of the self reducibility property on one particular instance involves a sequence of reductions. Gal *et al.* [5] show a related hardness result via a Karp reduction. They show that a procedure $\mathcal{A}$ that selects any subset of $\sqrt{n}$ variables and outputs their truth values in some satisfying assignment for ϕ may be used for solving SAT, with a single call to $\mathcal{A}$. They show similar results for other problems such as graph isomorphism and shortest lattice vector. Related results of the same nature appear also in [12, 9].

One may also consider the approximation of a specific property of a satisfying assignment for a 3SAT formula. For example, the maximum number of variables assigned a *true* truth value. Zuckerman [15] considered the following maximization version of $\mathcal{NP}$-complete problems. Let $\mathcal{R}_L(w, x)$ be a polynomial-time predicate corresponding to an $\mathcal{NP}$-complete language L so that $L = \{x | \exists w \ (w, x) \in \mathcal{R}_L\}$, where $w = \{0, 1\}^m$. Let $S \subseteq \{1, \ldots, m\}$ and view w as a subset of $\{1, \ldots, m\}$. Given x, S, compute $\max |S \cap w|$ over w satisfying $\mathcal{R}_L(w, x)$. For all the 21 $\mathcal{NP}$-complete problems in Karp's list, Zuckerman shows that this function is $\mathcal{NP}$-hard to approximate within a factor of n^{ε} for some constant $\varepsilon > 0$.

Tools and overview : A major breakthrough that led to many hardness of approximation results is the PCP theorem [4, 2, 1]. Our results follow from weaker tools, and are not based on the PCP theorem. Specifically, our results are based on the generic family of $\mathcal{NP}$ problems for which it is $\mathcal{NP}$-hard to obtain solutions within small Hamming distance of witnesses presented in [12]. This construction, in turn, is based on efficient list decodeable error correcting codes. Such codes are presented in [9].

The standard reductions for proving the $\mathcal{NP}$-completeness of problems preserve the hardness of *exactly* finding witnesses. However, in most cases, these reductions do not preserve the hardness of *approximating* witnesses. In this respect our situation is similar to that encountered for the more standard notion of approximating the value of the objective function. Nevertheless, we typically find that in such reductions, some particular part of the witness (e.g., the value of some of the variables in a satisfying assignment) is hard to approximate within a small Hamming distance. We call this part the *core* of the problem. (The rest of the witness is typically the result of introducing auxiliary variables during the reduction.)

Our reductions identify the *core* of the problem and *amplify* it, so that it becomes almost all of the witness, resulting in the hardness of approximating the whole witness. This amplification is very much problem dependent. For some problems, it is straightforward (e.g., for 3SAT it is a simple duplication of variables), whether for others we construct special gadgets.

In Section 2 we present the relation mentioned in [12] between list decodeable error correcting codes and witness approximation. In Section 3 we prove that given a 3SAT formula it is $\mathcal{NP}$-hard to obtain an assignment which agrees with any of its satisfying assignments on significantly more than 1/2 of the variables. This result combines techniques from [5, 12]. In Sections 4 and 5 we extend this result to the maximum CLIQUE, and minimum CHROMATIC NUMBER problems. In Section 6 we extend our hardness results to the remaining problems in Karp's [11] list of $\mathcal{NP}$-complete problems.

2 List decodeable ECC and witness approximation

Let $dist(x, y)$ be the *normalized Hamming distance* between two vectors x and y of equal length, *i.e.* the fraction of characters which differ in x and y. Note that $0 \leq dist(x, y) \leq 1$. An $[n, k, d]_q$ *error correcting code* is a function $\mathcal{C}$ defined as follows.

Definition 1 (Error correcting code - ECC). *Let Σ be an alphabet of size q. An $[n, k, d]_q$ error correcting code is a function $\mathcal{C} : \Sigma^k \to \Sigma^n$ with the property that for every $a, b \in \Sigma^k$ we have that $dist(\mathcal{C}(a), \mathcal{C}(b))$ is greater than or equal to d.*

Given an $[n, k, d]_q$ error correcting code $\mathcal{C}$ and a word $c \in \Sigma^n$, consider the problem of finding all codewords in $\mathcal{C}$ that are *close* to c. This problem is a generalization of the standard decoding problem, and is denoted as *list decoding* ([14, 8, 9]). We say that an $[n, k, d]_q$ error correcting code $\mathcal{C}$ is *list decodeable*, if there is an efficient procedure that given a word $c \in \Sigma^n$ produces a short list containing all words $a \in \Sigma^k$ for which $\mathcal{C}(a)$ is close to c.

Definition 2 (List decodeable ECC). *Let Σ be an alphabet of size q. An $[n, k, d]_q$ error correcting code $\mathcal{C}$ is δ list decodeable if there exists a Turing machine D (the list decoder) which on input $c \in \Sigma^n$ outputs in* $\mathbf{poly}(n)$ *time a list containing* **all** *words $a \in \Sigma^k$ that satisfy $dist(\mathcal{C}(a), c) \leq \delta$.*

In [9] the following theorem regarding list decodeable error correcting codes over finite fields is proven.

Theorem 1 ([9]). *For any finite field $\mathcal{F}_q$ of size q and any constant $\varepsilon > 0$ there exists an $\left[n, k, \left(1-\frac{1}{q}\right)\left(1-\frac{1}{n^{\frac{1}{2}-\varepsilon}}\right)\right]_q$ ECC which is $\left(1-\frac{1}{q}\right)\left(1-\frac{1}{n^{\frac{1}{4}-\varepsilon}}\right)$ list decodeable.*

In the remainder of this section we review the work of [12] which establishes a connection between the existence of list decodeable ECC and the hardness of finding *approximate* witnesses for general $\mathcal{NP}$ relations.

Definition 3 (Approximation within Hamming distance). *Given a polynomial time relation $\mathcal{R}$ and an instance x we say that w'* approximates a witness for x within Hamming distance δ *if there is a witness w (in the sense that $(w, x) \in \mathcal{R}$) for which $dist(w', w) < \delta$.*

Kumar and Sivakumar [12] show that for every $\mathcal{NP}$-complete language L there is a formulation via a relation $\hat{\mathcal{R}}_L$ for which it is $\mathcal{NP}$-hard to approximate a witness within Hamming distance significantly better than $1/2$. Let $\mathcal{R}_L$ be a polynomial relation so that $L = \{x \in \{0,1\}^* \mid \exists w \in \{0,1\}^* \text{ s.t } \mathcal{R}_L(w, x) \text{ holds}\}$. Given x, finding a witness w so that $\mathcal{R}_L(w, x)$ holds is $\mathcal{NP}$-hard. Let $\mathcal{C}$ be the list decodeable error correcting code of Theorem 1 (over GF_2), and let $\hat{\mathcal{R}}_L$ be the relation $\{(\hat{w}, x) \mid \exists w \text{ s.t } \hat{w} = \mathcal{C}(w) \text{ and } \mathcal{R}_L(w, x) \text{ holds}\}$. Note that the language L is equal to $\{x \mid \exists \hat{w} \text{ s.t } \hat{\mathcal{R}}_L(\hat{w}, x) \text{ holds}\}$.

Claim 2 ([12]). *Unless $\mathcal{NP} = \mathcal{P}$ there is no polynomial-time algorithm that given x of size n produces $\hat{w}'$ that approximates a witness for $\hat{\mathcal{R}}_L$ within Hamming distance $1/2 - \varepsilon$ for some $\varepsilon = 1/n^{\Omega(1)}$.*

Proof. Assume the existence of a polynomial time algorithm A that given x produces $\hat{w}'$ as in the claim. Apply the list decoding algorithm for $\mathcal{C}$ on $\hat{w}'$ to compute a list $\mathcal{W}$ (of polynomial length). By definition of $\hat{\mathcal{R}}_L$, there exists $w \in \mathcal{W}$ such that $(w, x) \in \mathcal{R}_L$. As finding such a witness w is $\mathcal{NP}$-hard the assertion is proven. $\square$

3 Witness inapproximability of SAT

In this section we prove the hardness of finding approximate satisfying assignments for Boolean formulas. Namely, we show that given a satisfiable Boolean formula ϕ, it is $\mathcal{NP}$-hard to find an assignment x to the variables of ϕ which is of Hamming distance *significantly* less than $1/2$ to some satisfying assignment of ϕ. We present a proof for 3CSP formulas and conclude in a simple reduction which extends this result to 3SAT formulas as well. This proof combines techniques from [5, 12]. In the proof below and the remaining sections of our work, the parameter ε is set to be $1/n^c$ for some constant $c > 0$, where n is the witness size of the problem considered. During our reductions it might be the case that the witness size changes from n to some polynomial in n. In such cases the constant c above should also change. We do not state this change explicitly, and the notation ε should be treated as $1/n^c$ for some constant $c > 0$ where the constant c may differ from one appearance of ε to another.

Consider the search version of the 3CSP (3SAT) problem in which we are given a formula ϕ where each clause is some Boolean function (disjunction) over three variables and we wish to obtain a satisfying assignment for ϕ.

Claim 3. *Approximating satisfying assignments for 3SAT (and 3CSP) formulas within Hamming distance $1/2 - \varepsilon$ is $\mathcal{NP}$-hard.*

Proof. We start by proving the hardness of approximating a satisfying assignment for 3CSP formulas within Hamming distance less than 1/2. Let $\mathcal{R} = \{(w, \phi_0) \mid \exists a \text{ s.t. } w = \mathcal{C}(a) \text{ and } \phi_0(a) \text{ is true}\}$ where ϕ_0 is a SAT formula. Denote the size of a witness w to ϕ_0 in $\mathcal{R}$ by n. We follow the proof paradigm described in the introduction. We use the relation $\mathcal{R}$ above as the base of the reduction, present a *core* instance ϕ_1 with the property that some variables of ϕ_1 are hard to approximate, and then *amplify* these hard variables (by duplicating them) to obtain our final formula ϕ_2.

Using Claim 2, we have that unless $\mathcal{NP} = \mathcal{P}$ there is no polynomial-time algorithm that given ϕ_0 produces w' that approximates a witness for ϕ_0 in $\mathcal{R}$ within Hamming distance $1/2 - \varepsilon$.

Core : Using any standard reduction (e.g., Cook's theorem) transform $\mathcal{R}$ into a 3SAT formula ϕ_1. Denote the variables of ϕ_1 as $\{w_1, \ldots, w_n; z_1, \ldots, z_l\}$, where the value of $w_1, \ldots, w_n$ in a satisfying assignment to ϕ_1 represent a witness w to ϕ_0 in $\mathcal{R}$, the variables $z_1, \ldots, z_l$ are auxiliary variables added by the reduction, and $l = \mathbf{poly}(n)$. We have that an assignment a to ϕ_0 is a satisfying assignment iff there exists an assignment to z such that $\phi_1(\mathcal{C}(a), z) = 1$. *I.e.* ϕ_1 checks that $\mathcal{C}(a)$ is a codeword and that $\phi_0(a) = 1$.

Amplification : Let $m = \mathbf{poly}(l)$. Since $l = \mathbf{poly}(n)$ it holds that $m = \mathbf{poly}(n)$ as well. Construct a 3CSP formula ϕ_2 by duplicating the variables $w_1, \ldots, w_n$ so that the number of auxiliary variables $z_1, \ldots, z_l$ is small with respect to the total number of variables in ϕ_2.

Define $\phi_2(w_1, \ldots, w_n; w_1^1, \ldots, w_n^m; z_1, \ldots, z_l)$ to be

$$\phi_2 = \phi_1(w_1, \ldots, w_n; z_1, \ldots, z_l) \wedge \bigwedge_{i=1}^{n} (w_i = w_i^1) \wedge \bigwedge_{i=1}^{n} \bigwedge_{j=1}^{m-1} (w_i^j = w_i^{j+1})$$

Assume that one can efficiently find an assignment x that approximates a satisfying assignment for ϕ_2 within Hamming distance $1/2 - \varepsilon$. For $j = 1 \ldots m$ let x^j be the restriction of x over the variables $w_1^j \ldots w_n^j$. As x is within Hamming distance $1/2 - \varepsilon$ of a satisfying assignment to ϕ_2, we conclude by definition of ϕ_2 that there exists some j such that x^j is within Hamming distance $1/2 - \varepsilon$ from $\mathcal{C}(a)$ where a is an assignment satisfying ϕ_0 (we have neglected the auxiliary variables $z_1, \ldots, z_l$ as l is significantly smaller than m). This contradicts the witness inapproximability of the relation $\mathcal{R}$ above, and concludes our assertion regarding 3CSP formulas.

By translating clauses of the form $(w_i^j = w_i^{j+1})$ in ϕ_2 to $(\bar{w}_i^j \vee w_i^{j+1}) \wedge (w_i^j \vee \bar{w}_i^{j+1})$ we obtain a 3SAT formula with the desired inapproximability properties as well. $\square$

4 Witness inapproximability of CLIQUE

In this section we consider the search version of CLIQUE in which we are given a graph $G = (V, E)$ and we wish to obtain an indicator vector $x \in \{0,1\}^{|V|}$ so that $\{v \in V | x_v = 1\}$ induces a maximum clique in G.

Claim 4. *Approximating the indicator vector of the maximum CLIQUE within Hamming distance $1/2 - \varepsilon$ is $\mathcal{NP}$-hard.*

Proof. Let ϕ be a 3SAT formula for which it is $\mathcal{NP}$ hard to approximate a satisfying assignment beyond Hamming distance $1/2 - \varepsilon$ (Claim 3). Given ϕ we present a core instance H, and then amplify certain parts of H.

Core : Let H be the graph obtained by the reduction from 3SAT to CLIQUE described in [4]. For every clause in ϕ create a set of vertices corresponding to assignments to the clause variables that set the clause to *true*. *I.e.* clauses of the form $(l_1 \vee l_2 \vee l_3)$ yield seven vertices, each corresponding to an assignment that satisfies the clause. Connect by an edge every two consistent vertices, *i.e.* vertices that correspond to non-contradicting assignments to their respective variables. It is not hard to verify that any satisfying assignment for ϕ corresponds to a maximum clique in H, and vice versa.

Amplification : Let $y_1 \dots y_n$ be the variables of ϕ and $m = \mathbf{poly}(n)$. For each variable y_i of ϕ add 2 new sets of vertices to H. A set of m vertices $K_i = \{v_i^1 \dots v_i^m\}$ corresponding to a *true* truth value for the variable y_i, and a set of m vertices $\bar{K}_i = \{\bar{v}_i^1 \dots \bar{v}_i^m\}$ corresponding to a *false* truth value for y_i. Let V be the vertex set obtained by this addition. Let $G = (V, E)$ where the edge set E consists of an edge between every two consistent vertices of V (note that E includes the original edge set of H).

Let x be an indicator vector of some maximum clique in G. A maximum clique in G corresponds to a satisfying assignment of ϕ. Specifically, given a satisfying assignment a to ϕ, the set of vertices in G consistent with a form a maximum clique in G, and each maximum clique in G is of this nature. Note that for each i this set of vertices includes one of the sets K_i or $\bar{K}_i$.

Let x' be an indicator vector which approximates the vector x within Hamming distance $1/2 - \varepsilon$. Using an averaging argument similar to the one used in Claim 3, we conclude that one of the following hold. There exists some index j such that the values of the vector x restricted to the vertices $v_1^j \dots v_n^j$ agrees with the value of x' restricted to these vertices on at least a $1/2 + \varepsilon$ fraction of their values, or such an agreement exists on the vertices $\bar{v}_1^j \dots \bar{v}_n^j$. As the values of x restricted to $v_1^j \dots v_n^j$ or $\bar{v}_1^j \dots \bar{v}_n^j$ correspond to a satisfying assignment a of ϕ (or its bit wise inverse) we conclude our claim. □

5 Witness inapproximability of CHROMATIC NUMBER

Given a graph $G = (V, E)$, the CHROMATIC NUMBER problem is the problem of finding a coloring vector $\sigma \in \{1, \dots, k\}^{|V|}$ such that for all $(u, v) \in E$ it holds that $\sigma_u \neq \sigma_v$ and k is of minimum cardinality.

Claim 5. *Approximating the coloring vector of a* 3 *colorable graph within Hamming distance* $2/3-\varepsilon$ *is* $\mathcal{NP}-$*hard.*

Proof. Let $\mathcal{C}$ be the error correcting code over $\Sigma=\{0,1,2\}$ implied by Theorem 1. Recall that $\mathcal{C}$ is $\delta=2/3-\varepsilon$ list decodeable. Let ϕ_0 be a 3SAT formula, $\mathcal{R}=\{(w,\phi_0)\mid\exists a \text{ s.t. } \mathcal{C}(a)=w \text{ and } \phi_0(a)=\text{true}\}$, and n be the size of a witness w to ϕ_0 in $\mathcal{R}$. As in Claim 2, given ϕ_0 it can be seen that one cannot approximate a witness for $\mathcal{R}$ within Hamming distance $2/3-\varepsilon$ (unless $\mathcal{NP}=\mathcal{P}$). Let $\tau:\{0,1,2\}\rightarrow\{0,1\}^2$ be the mapping which associates with any character in $\{0,1,2\}$ its binary representation and denote the component-wise image of $\mathcal{C}(a)$ under this mapping as $\tau(\mathcal{C}(a))$. Let $\hat{\mathcal{R}}=\{(y,\phi_0)\mid\exists w \text{ s.t. } \tau(w)=y \text{ and } \mathcal{R}(w,\phi_0)\}$ be a *binary representation* of $\mathcal{R}$. Using any standard reduction (e.g., Cook's theorem) transform $\hat{\mathcal{R}}$ in to a 3SAT formula ϕ_1. Denote the variables of ϕ_1 as $\{y_1^0,y_1^1,\ldots,y_n^0,y_n^1;z_1,\ldots,z_l\}$ where each pair y_i^0,y_i^1 in a satisfying assignment to ϕ_1 represents the image of a single character of $\mathcal{C}(a)$ under τ for some satisfying assignment a of ϕ_0. The variables z_j are auxiliary variables added by the reduction, and $l=\mathbf{poly}(n)$. The formula ϕ_1 is reduced to a graph $G=(V,E)$ as follows.

Core : Using a standard reduction from 3SAT to 3-coloring (for details see [3]), we define a graph H which corresponds to the formula ϕ_1. This reduction guarantees that the graph H is 3-colorable iff the formula ϕ_1 is satisfiable. Furthermore, in such a case an assignment satisfying ϕ_1 can be obtained from a 3-coloring of H by setting the truth value of the variables in ϕ_1 to be equal to the colors of their corresponding vertices in H. In this reduction, these vertices will be colored by one of two colors. We denote these colors as 1 and 0 and the remaining color as 2.

Translation : Let y_i^0,y_i^1 be the two vertices in H corresponding to the two variables y_i^0,y_i^1 in ϕ_1. Let σ be a valid 3-coloring of H which corresponds to a satisfying assignment to the formula ϕ_1 and let $\theta_i=\tau^{-1}(y_i^0,y_i^1)$ be the value of the binary representation of the colors assigned to y_i^0,y_i^1 in σ. Note that $\theta_i\in\{0,1,2\}$. We would like to add a new vertex w_i^1 to H which, given such a coloring, must be colored in the color θ_i. Such a vertex will *translate* the Boolean colors assigned to y_i^0,y_i^1 into a corresponding color in $\{0,1,2\}$. This translation will be later used in our proof. We construct the gadget *(a)* in Figure 1 for each pair y_i^0,y_i^1. This gadget receives as input the bits y_i^0 and y_i^1 and fixes the value of w_i^1 to θ_i. The components of gadget *(a)* are the gadgets *(i)*,*(ii)* and *(iii)* presented in Figure 1. In these gadgets we assume that some vertices are colored by a specific color. The ability to do so is yet another property of the reduction used to construct the core graph H.

Amplification : For each i, we now add $m=\mathbf{poly}(n)$ copies of each vertex w_i^1 using gadget *(b)* of Figure 1. Denote the resulting graph by G. Note that in any valid 3-coloring of G the colors of the vertices w_i^j for $j=1,\ldots,m$ are identical.

Let ϕ_0 be a satisfiable formula, G be the graph above, and σ be a coloring vector of a valid 3-coloring of G. By our definition of ϕ_1 and G, if we set

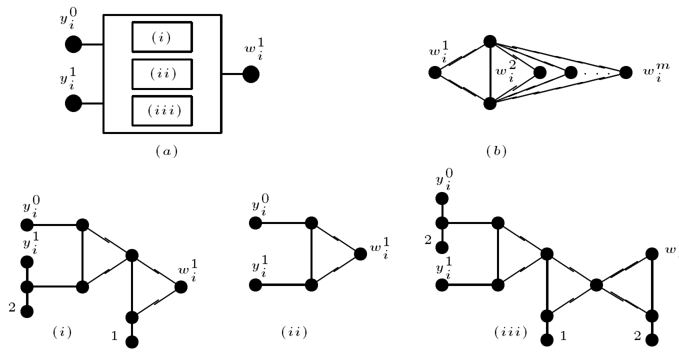

Fig. 1. Translation and amplification of the core graph H. Gadget *(a)* uses gadgets (i), (ii) and (iii) in order to set the value of w_i^1 to be equal to the binary representation of y_i^0 and y_i^1. Gadget (i) sets w_i^1 to be 1 whenever $y_i^0 = 0$ and $y_i^1 = 1$ otherwise it does not restrict w_i^1. Similarly, gadget (ii) sets w_i^1 to be 0 whenever $y_i^0 = 0$ and $y_i^1 = 0$ and gadget (iii) sets w_i^1 to be 2 whenever $y_i^0 = 1$ and $y_i^1 = 0$. Gadget *(b)* adds m new copies of w_i^1 and makes sure that they are colored by the same color.

$w \in \{0, 1, 2\}^n$ to be the value of σ restricted to the entries corresponding to the vertices $w_1^1, \ldots, w_n^1$, we obtain a witness for ϕ_0 in $\mathcal{R}$. By following the line of proof given in Claim 4, we conclude that approximating the coloring vector σ within Hamming distance $2/3 - \varepsilon$ yields an approximation of a witness w to ϕ_0 in the relation $\mathcal{R}$ within Hamming distance $2/3 - \varepsilon$. Detailed proof is omitted. □

6 Remaining problems

In the following we consider the search version of the remaining problems from Karp's list of 21 $\mathcal{NP}$-complete decision problems.

Given a graph $G = (V, E)$ the INDEPENDENT SET problem is the problem of finding an indicator vector x so that $\{v \in V | x_v = 1\}$ is a maximum independent set in G. The VERTEX COVER problem is the problem of finding an indicator vector x so that $\{v \in V | x_v = 1\}$ is a minimum vertex cover in G.

Claim 6. *Approximating the indicator vector of the INDEPENDENT SET and VERTEX COVER problems within Hamming distance $1/2 - \varepsilon$ is $\mathcal{NP}$-hard.*

Proof. Inapproximability of INDEPENDENT SET and VERTEX COVER follow from the inapproximability of CLIQUE by standard reductions. For INDEPENDENT SET, given a graph G, construct the complement graph $\bar{G}$. The indicator vector corresponding to the maximum clique in $\bar{G}$ is identical to the indicator vector representing the maximum independent set in G. For VERTEX COVER, given a graph G, the indicator vector corresponding to the maximum independent set in G is the bitwise complement of the indicator vector corresponding to the minimum vertex cover of G. □

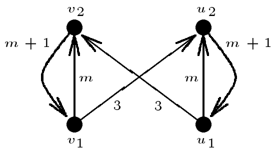

Fig. 2. The reduced form of an edge (u, v) from the original graph G. The numbers appearing by each edge represent its multiplicity. The parameter m is some polynomial in the size of G.

Given a directed graph $G = (V, E)$ the FEEDBACK EDGE SET problem is the problem of finding an indicator vector $x \in \{0,1\}^{|E|}$ such that $\{e \in E \mid x_e = 1\}$ is a minimum subset of edges which intersects every directed cycle in H. We present a hardness result for the FEEDBACK EDGE SET problem in which we allow the given graph H to have parallel edges. We do not know whether a similar hardness result holds if parallel edges are not allowed.

Claim 7. *Approximating the indicator vector for the FEEDBACK EDGE SET within Hamming distance* $1/2 - \varepsilon$ *is* $\mathcal{NP}$*-hard.*

Proof. We enhance the standard reduction from VERTEX COVER to FEEDBACK EDGE SET presented in [11] by adding an amplification gadget. Let $G = (V, E)$ be a given instance of the VERTEX COVER problem in which $|V| = n$, and H be the reduced instance for the FEEDBACK EDGE SET problem. Each vertex v in G is transformed into two vertices v_1 and v_2 in H. These vertices are connected by a set of directed edges, $m = \mathbf{poly}(n)$ edges from v_1 to v_2 and $m + 1$ edges from v_2 to v_1. In addition each undirected edge (u, v) in G is transformed into 6 edges in H, 3 edges of the form (v_1, u_2) and three edges of the form (u_1, v_2). A fraction of the graph H corresponding to one original edge (u, v) of G is presented in Figure 2.

It can be seen that a minimum feedback edge set in G consists of the edges (v_1, v_2) or the edges (v_2, v_1) only (that is *crossing* edges of the form (u_1, v_2) will not appear in a minimum solution). Furthermore, for each pair of vertices v_1, v_2 in H the $m+1$ edges (v_2, v_1) appear in a minimum feedback edge set in H iff the vertex v appears in a minimum vertex cover of G. Using this correspondence, our claim is proven using techniques similar to those presented in Claim 4 and Claim 5. Detailed proof is omitted. □

Let $G = (V, E)$ be an undirected (directed) graph. The UNDIRECTED (DIRECTED) HAMILTONIAN CYCLE problem is the problem of finding an indicator vector x of edges in E comprising a directed (undirected) Hamiltonian cycle.

Claim 8. *Approximating the indicator vector of the DIRECTED HAMILTONIAN CYCLE and HAMILTONIAN CYCLE within Hamming distance* $1/2 - \varepsilon$ *and* $2/5 - \varepsilon$ *respectively is* $\mathcal{NP}$*-hard.*

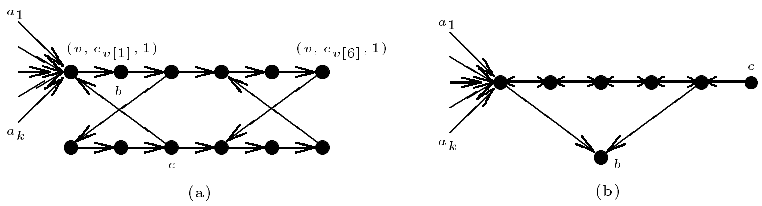

Fig. 3. *(a)* The original 'cover testing gadget', *(b)* The gadget replacing the vertex $(v, e_{v[1]}, 1)$ from *(a)* with $m = 4$.

Proof. We use the standard reduction presented in [6] from the VERTEX COVER problem on a given graph G of vertex size n, to the DIRECTED HAMILTONIAN CYCLE problem, and amplify some of the 'cover testing gadgets' used in this reduction. Using the notation from [6], a single cover testing gadget is presented in Figure 3*(a)*. In the reduction of [6], a vertex cover in the original graph G is obtained from a Hamiltonian cycle in the reduced graph by setting the vertex v in G to be in the vertex cover if the Hamiltonian cycle in the reduced graph enters the vertex $(v, e_{v[1]}, 1)$ from a vertex a_i (for some i). On the other hand, if the Hamiltonian cycle in the reduced graph enters $(v, e_{v[1]}, 1)$ from the vertex c then in the original graph G the vertex v is not included in the vertex cover. In order to amplify this difference, we replace each vertex $(v, e_{v[1]}, 1)$ in the original cover testing gadget by the additional gadget of size $m = \mathbf{poly}(n)$ presented in Figure 3*(b)*.

It can be seen that if the Hamiltonian cycle enters the vertex $(v, e_{v[1]}, 1)$ from a_i on its way to the vertex b in Figure 3*(a)* then it must use the edges pointing right in Figure 3*(b)*. Otherwise, the Hamiltonian path must enter $(v, e_{v[1]}, 1)$ from the vertex c in Figure 3*(a)*. In this case the edges pointing left in Figure 3*(b)* will be used in order to reach the vertex b from c.

Using the above it can be seen that approximating the indicator vector of the DIRECTED HAMILTONIAN CYCLE within Hamming distance $1/2 - \varepsilon$ is $\mathcal{NP}$-hard. By replacing the gadget in Figure 3*(b)* by a different undirected gadget, one may achieve the hardness result of $2/5 - \varepsilon$ for the undirected HAMILTONIAN CYCLE problem. Possibly, a hardness result of $1/2 - \varepsilon$ can also be proven for the undirected HAMILTONIAN CYCLE problem, though we have not been able to do so. $\square$

The remaining problems in Karp's list of 21 $\mathcal{NP}$-complete decision problems are : 3-DIMENSIONAL MATCHING (3DM), STEINER TREE, CLIQUE COVER, EXACT COVER, SET PACKING, SET COVER, HITTING SET, PARTITION, KNAPSACK, INTEGER PROGRAMMING, JOB SEQUENCING, MAX CUT, and FEEDBACK NODE SET.

For all the above problems one can define *natural* witnesses, and prove *tight* witness inapproximability results by slight adjustments to the standard reductions appearing in [11, 6, 7]. Full proofs will appear in a future version of our paper.

Acknowledgments

The first author is the Incumbent of the Joseph and Celia Reskin Career Development Chair. This research was supported in part by a Minerva grant, project 8354.

References

1. S. Arora, C. Lund, R. Motwani, M. Sudan, M. Szegedy. Proofs Verification and Hardness of Approximation Problems. Journal of the ACM, Vol. 45(3), pp. 501-555, 1998.
2. S. Arora, S. Safra. Probabilistic Checking of Proofs: A new Characterization of NP. Journal of the ACM, Vol. 45(1), pp. 70-122, 1998.
3. T. H. Cormen, C. E. Leiserson, R. L. Rivest. Introduction to Algorithms. MIT press, 1990.
4. U. Feige, S. Goldwasser, L. Lovasz, S. Safra, M. Szegedy. Interactive Proofs and the Hardness of Approximating Cliques. Journal of ACM, Vol. 43(2), pp. 268-292, 1996.
5. A. Gal and S. Halevi and R. Lipton and E. Petrank. Computing from partial solutions. Proc. 14th IEEE Conference on Computational Complexity, pp. 34-45, 1999.
6. M.R. Garey and D.S. Johnson. Computers and Intractability: A guide to the theory of NP-Completeness. W.H. Freeman, 1979.
7. M.R. Garey, D.S. Johnson, L. Stockmeyer. Some simplified NP-complete graph problems. Theoretical Computer Science 1, pp.237-267, 1976.
8. O. Goldreich, D. Ron and M. Sudan. Chinese Remaindering with Errors. Proc. 31st Symp. Theory of Computing, pp. 225-234, 1999.
9. V. Guruswami and M. Sudan. List Decoding Algorithms for certain Concatenated Codes. Proc. 32nd Symp. Theory of Computing, pp. 181-190, 2000.
10. J. Håstad. Some optimal inapproximability results. Proc. 28th Symp. Theory of Computing, pp. 1-10, 1997.
11. R. M. Karp. Reducibility among Combinatorial Problems. Proc. Symp. Complexity of Computer Computations. Ed. R.E. Miller, J.W. Thatcher, pp. 85-103, 1972.
12. R. Kumar and D. Sivakumar. Proofs, Codes, and polynomial-Time Reducibilities. Proc. 14th IEEE Conference on Computational Complexity, pp. 46-53, 1999.
13. U. Schoning. A probabilistic Algorithm for k-SAT and Constraint Satisfaction Problems. Proc. 40th IEEE Symp. on Foundations of Computer Science, pp. 410-414, 1999.
14. M. Sudan. Decoding of Reed Solomon codes beyond the error-correction bound. Journal of Complexity 13(1), pp. 180-193, 1997.
15. D. Zuckerman. On Unapproximable Versions of NP-Complete Problems. SIAM J. Comput. 25(6), pp. 1293-1304, 1996