# Approximability of Dense Instances of NEAREST CODEWORD Problem 

Cristina Bazgan, Wenceslas Fernandez de La Vega, Marek Karpinski

## - To cite this version:

Cristina Bazgan, Wenceslas Fernandez de La Vega, Marek Karpinski. Approximability of Dense Instances of NEAREST CODEWORD Problem. 8th Scandinavian Workshop on Algorithm Theory, SWAT 2002, Jul 2002, Turku, Finland. pp.298-307, 10.1007/3-540-45471-3_31 . hal-01444194

HAL Id: hal-01444194

## https://hal.science/hal-01444194

Submitted on 23 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Approximability of Dense Instances of Nearest Codeword Problem 

Cristina Bazgan ${ }^{1}$, W. Fernandez de la Vega $^{2}$, Marek Karpinski ${ }^{3}$<br>${ }^{1}$ Université Paris Dauphine, LAMSADE, 75016 Paris, France, bazgan@lamsade.dauphine.fr<br>${ }^{2}$ CNRS, URA 410, Université Paris Sud, LRI, 91405 Orsay, France, lalo@lri.fr<br>${ }^{3}$ Dept. of Computer Science, University of Bonn, 53117 Bonn, Germany<br>marek@cs.uni-bonn.de


#### Abstract

We give a polynomial time approximation scheme (PTAS) for dense instances of the Nearest Codeword problem.


## 1 Introduction

We follow [15] in defining the Nearest Codeword problem as the minimum constraint satisfaction problem for linear equations mod 2 with exactly 3 variables per equation. It is shown in [15] that the restriction imposed on the number of variables per equation (fixing it to be exactly 3 ) does not reduce approximation hardness of the problem. The problem is, for a given set of linear equations $\bmod 2$ to construct an assignment which minimizes the number of unsatisfied equations. We shall use in this paper clearly an equivalent formulation of the problem of minimizing the number of satisfied equations. Adopting the notation of [11] we denote it also as the Min-E3-Lin2 problem. Min-Ek-Lin2 will stand for the $k$-ary version of the Nearest Codeword problem.

The Nearest Codeword problem arises in a number of coding theoretic, and algorithmic contexts, see, e.g., [1], [15], [8], [7]. It is known to be exceedingly hard to approximate; it is known to be NP-hard to approximate to within a factor $n^{\Omega(1) / \log \log n}$. Only recently the first sublinear approximation ratio algorithm has beeing designed for that problem [5]. In this paper we prove that, somewhat surprisingly, the Nearest Codeword problem on dense instances does have a PTAS. We call an instance of Nearest Codeword problem (Min-E2-LIN2) problem dense, if the number of occurrences of each variable in the equations is $\Theta\left(n^{2}\right)$ for $n$ the number of variables. We call an instance of NEAREst Codeword (Min-E2-Lin2) dense in average if the number of equations is $\Theta\left(n^{3}\right)$. Analogously, we define density, and average density, for Min-Ek-Lin2 problems.

It is easy to be seen that the results of [2] and [9] on existence of PTASs for average dense maximum constraint satisfaction problems cannot be applied to their average dense minimum analogs (for a survey paper on approximability of some other dense optimization problems see also [13]). This observation
can be also strenghten for the dense instances of minimum constraint satisfaction by noting that dense instances of Vertex Cover can be expressed as dense instances of minimum constraint satisfaction problem for 2DNF clauses, i.e. conjunctions of 2 literals, and then applying the result of [6], [14] to the effect that there are no PTAS for the dense Vertex Cover. In [10] it was also proven that the dense and average dense instances of $\operatorname{Min} \operatorname{Tsp}(1,2)$ and Longest Path problems do not have polynomial time approximation schemes.

In [2] there were however two dense minimization problems identified as having PTASs, namely dense Bisection, and Min-k Cut. This has lead us to investigate the approximation complexity of dense Nearest Codeword problem. Also recently, PTASs have been designed for dense Min Equivalence and dense Min- $k$ Sat problems (cf. [3], [4]). The main result of this paper is a proof of an existence of a PTAS for the dense Nearest Codeword problem.

The approximation schemes developed in this paper for the dense Nearest CODEWORD problem use some novel density sampler techniques for graphs, and $k$-uniform hypergraphs, and extend available up to now approximation techniques for attacking dense instances of minimum constraint satisfaction problems.

The Nearest Codeword problem in its bounded arity $(=3)$ form was proven to be approximation hard for its unbounded arity version in [15] (Lemma 37). This results in $n^{\Omega(1) / \log \log n}$ approximation lower bound for the NEAREST Codeword problem by [8], [7], where $n$ is the number of variables. It is also easy to show that Nearest Codeword is hard to approximate to within a factor $n^{\Omega(1) / \log \log n}$ on average dense instances.

The paper is organized as follows. In Section 2 we give the necessary definitions and prove the NP-hardness of dense instances of Min-E3-Lin2 in exact setting, and in Section 3 we give a polynomial time approximation scheme for the dense instances of Min-E3-Lin2.

## 2 Preliminaries

We begin with basic definitions.
Approximability. A minimization problem has a polynomial time approximation scheme (a PTAS, in short) if there exists a polynomial time approximation algorithm that gives for each instance $x$ of the problem a solution $y$ of value $m(x, y)$ such that $m(x, y) \leq(1+\varepsilon) \operatorname{opt}(x)$ for every constant $\varepsilon>0$ where opt $(x)$ is the value of an optimum solution.

Nearest Codeword Problem (Min-E3-Lin2)
Input: A set of $m$ equations in boolean variables $x_{1}, \ldots, x_{n}$ where each equation has the form $x_{i_{1}} \oplus x_{i_{2}} \oplus x_{i_{3}}=0$ or $x_{i_{1}} \oplus x_{i_{2}} \oplus x_{i_{3}}=1$.
Output: An assignment to the variables that minimizes the number of equations satisfied.

Density. A set of instances of Min-E3-Lin2 is $\delta$-dense if for each variable $x$, the total number of occurrences of $x$ is at least $\delta n^{2}$ in each instance. A class of instances of Min-E3-Lin2 is dense, if there is a constant $\delta$ such that the class is $\delta$-dense.

Let us show now that dense Min-E3-Lin2 is NP-hard in exact setting. The reduction is from Min-E3-Lin2, which is approximation hard for a ratio $n^{\Omega(1) / \log \log n}[8],[7]$, where $n$ is the number of variables. Given an instance $I$ of Min-E3-Lin2 on a set of $n$ variables $X=\left\{x_{1}, \ldots, x_{n}\right\}$ with $m$ equations $x_{t_{1}} \oplus x_{t_{2}} \oplus x_{t_{3}}=b$, where $b \in\{0,1\}$, we construct an instance $I^{\prime}$ of dense Min-E3-Lin2 as follows: we extend the set of variables $X$ by two disjoint sets $Y=\left\{y_{1}, \ldots, y_{n}\right\}$ and $Z=\left\{z_{1}, \ldots, z_{n}\right\} . I^{\prime}$ contains aside from the equations of $I$, the equations of the form $x_{i} \oplus y_{j} \oplus z_{h}=0$ and $x_{i} \oplus y_{j} \oplus z_{h}=1$ for all $1 \leq i, j, h \leq n$. Note that the system $I^{\prime}$ is dense. We note also that exactly $n^{3}$ of the added equations are satisfied independently of the values of the variables in $X, Y$ and $Z$. Thus $o p t\left(I^{\prime}\right)=o p t(I)+n^{3}$, proving the claimed reduction.

## 3 Dense Min-E3-Lin2 has a PTAS

Let the system $\mathcal{S}=\left\{E_{1}, \ldots, E_{m}\right\}$ be a $\delta$-dense instance of Min-E3-Lin2, on a set $X$ of $n$ variables $\left\{x_{1}, \ldots, x_{n}\right\}$.

We will run two distinct algorithms, algorithm A and algorithm B, and select the solution with the minimum value. Algorithm A gives a good approximate solution for the instances whose minimum value is at least $\alpha n^{3}$. Algorithm B gives a good approximate solution for the instances whose minimum value is less than $\alpha n^{3}$, where $\alpha$ is a constant depending both on $\delta$ and the required accuracy $\epsilon$.

### 3.1 Algorithm A

Algorithm A depends on formulating the problem as a Smooth Integer Program [2] as follows.

A smooth degree-3 polynomial (with smoothness $e$ ) has the form

$$
\sum a_{i j h} x_{i} x_{j} x_{h}+\sum b_{i j} x_{i} x_{j}+\sum c_{i} x_{i}+d
$$

where each $\left|a_{i j h}\right| \leq e,\left|b_{i j}\right| \leq e n,\left|c_{i}\right| \leq e n^{2},|d| \leq e n^{3}$ (cf. [2]).
For each equation $x_{i} \oplus y_{i} \oplus z_{i}=b_{i}$ in $\mathcal{S}$, we construct the smooth polynomial

$$
P_{i} \equiv\left(1-x_{i}\right)\left(1-y_{i}\right)\left(1-z_{i}\right)+x_{i} y_{i}\left(1-z_{i}\right)+y_{i} z_{i}\left(1-x_{i}\right)+z_{i} x_{i}\left(1-y_{i}\right)
$$

if $b_{i}=0$, and

$$
P_{i} \equiv x_{i}\left(1-y_{i}\right)\left(1-z_{i}\right)+y_{i}\left(1-x_{i}\right)\left(1-z_{i}\right)+z_{i}\left(1-x_{i}\right)\left(1-y_{i}\right)+x_{i} y_{i} z_{i}
$$

if $b_{i}=1$. We have then the Smooth Integer Program IP:

$$
\left\{\begin{array}{l}
\min \sum_{j=1}^{m} P_{i} \\
\text { s. t. } x_{i}, y_{i}, z_{i} \in\{0,1\} \forall i, 1 \leq i \leq n .
\end{array}\right.
$$

A result of [2] can be used now to approximate in polynomial time the minimum value of IP with additive error $\epsilon n^{3}$ for every $\epsilon>0$. This provides an approximation ratio $1+\epsilon$ whenever the optimum value is $\Omega\left(n^{3}\right)$.

### 3.2 Algorithm B

The algorithm B is guaranteed to give, as we will show, approximation ratio $1+\epsilon$ for each fixed $\epsilon$, whenever the optimum is at most $\alpha n^{3}$ for a fixed $\alpha$, depending on $\epsilon$ and on the density.

## Algorithm B

Input: A dense system $\mathcal{S}$ of linear equations in GF[2] over a set $X$ of $n$ variables with exactly 3 variables per equation.

1. Pick two disjoint random samples $S_{1}, S_{2} \subseteq X$ of size $m=\Theta\left(\frac{\log n}{\epsilon^{2}}\right)$;
2. For each possible assignment $a \in\{0,1\}^{\left|S_{1} \cup S_{2}\right|}$ for the variables $y$ in $S_{1} \cup S_{2}$ ( $y^{a}$ will stand for the boolean value of $y$ for the assignement $a$ ) do the following:
2.1 For each variable $x \notin S_{1} \cup S_{2}$ do the following:

Let $H_{x, 0}^{a}$ and $H_{x, 1}^{a}$ be the bipartite graphs with common vertex set $V\left(H_{x, 0}^{a}\right)=$ $V\left(H_{x, 1}^{a}\right)=S_{1} \cup S_{2}$ and edge sets

$$
E\left(H_{x, 0}^{a}\right)=\left\{\{y, z\}: \chi_{S_{1}}(y) \oplus \chi_{S_{1}}(z)=1 \wedge x \oplus y \oplus z=b \in \mathcal{S} \wedge y^{a} \oplus z^{a}=b\right\}
$$

and

$$
E\left(H_{x, 1}^{a}\right)=\left\{\{y, z\}: \chi_{S_{1}}(y) \oplus \chi_{S_{1}}(z)=1 \wedge x \oplus y \oplus z=b \in \mathcal{S} \wedge 1 \oplus y^{a} \oplus z^{a}=b\right\}
$$

Let $m_{0}^{a}=\left|E\left(H_{x, 0}^{a}\right)\right|, m_{1}^{a}=\left|E\left(H_{x, 1}^{a}\right)\right|$.
If $m_{0}^{a} \geq \frac{2}{3}\left(m_{0}^{a}+m_{1}^{a}\right)$, then set $x$ to 1 .
If $m_{1}^{a} \geq \frac{2}{3}\left(m_{0}^{a}+m_{1}^{a}\right)$, then set $x$ to 0 .
Otherwise, set $x$ to be undefined.
2.2 In this stage, we assign values to the variables which are undefined after the completion of stage 2.1. Let $D^{a}$ be the set of variables assigned in stage 2.1, and let $U^{a}=S_{1} \cup S_{2} \cup D^{a}$. $V^{a}=X \backslash U^{a}$ denotes the set of undefined variables. For each undefined variable $y$, let $S_{y}$ denote the set of equations which contain $y$ and two variables in $U^{a}$. Let $k_{0}^{a}$ (resp. $k_{1}^{a}$ ) denote the number of equations in $S_{y}$ satisfied by $a$ and by setting $y$ to 0 (resp. to 1 ).

If $k_{0}^{a} \leq k_{1}^{a}$, then set $y$ to 0 . Else, set $y$ to 1 .
Let $X^{a}$ denote the overall assignment produced by the end of this stage.
3. Among all the assignments $X^{a}$ pick one which satisfies the minimum number of equations of $\mathcal{S}$.

Output that assignment.

## 4 Proof of the correctness of algorithm B when the value of the instance is "small"

We will use the following graph density sampling lemma. Recall that the density $d$ of a graph $G=(V, E)$ is defined by

$$
d=\frac{|E|}{\binom{|V|}{2}} .
$$

Lemma 1. Let $d$ and $\epsilon$ be fixed and let the graph $G=(V, E)$ have $|V|=n$ vertices and density $d$. Let $m=\Theta\left(1 / d \epsilon^{-2} \log n\right)$. Let $X=\left\{x_{1}, \ldots, x_{m}\right\}$ and $Y=\left\{y_{1}, \ldots, y_{m}\right\}$ be two random disjoint subsets of $V(G)$ with $|X|=|Y|=m$ and let $e(X, Y)$ be the number of edges of $G$ between $X$ to $Y$. Then, for each sufficiently large $n$, we have

$$
\operatorname{Pr}\left[\left|e(X, Y)-m^{2} d\right| \leq \epsilon m^{2} d\right]=1-o(1 / n)
$$

Proof. We will use the following inequality due to Hoeffding [12]. Let $X_{1}, \ldots, X_{m}$ be independent and identically distributed. Let $\mu=E\left(X_{1}\right)$ and assume that $X_{1}$ satisfies $0 \leq X_{1} \leq \Delta$. Let $S_{m}=\sum_{i=1}^{m} X_{i}$. Then:

$$
\begin{equation*}
\operatorname{Pr}\left(\left|S_{m}-\mu m\right| \geq \epsilon \Delta m\right) \leq 2 \exp \left(-2 \epsilon^{2} m\right) \tag{1}
\end{equation*}
$$

Clearly

$$
E(e(X, Y))=m^{2} d
$$

For each $z \in V \backslash X$, write

$$
T_{z}=|\Gamma(z) \cap X|
$$

Let $T=\sum_{z \in V \backslash X} T_{z}$. Then, $T=T^{\prime}+\Delta$ where $\Delta \leq m(m-1) / 2$, and $T^{\prime}$ is the sum of $m$ randomly chosen valencies from the set of valencies of $G$. Thus using (1),

$$
\operatorname{Pr}\left[\left|T^{\prime}-m n d\right| \leq \epsilon m n+m(m-1)^{2} / 2\right] \geq 1-2 \exp \left(-O\left(\epsilon^{2} m\right)\right)
$$

Clearly,

$$
\begin{aligned}
e(X, Y) & =\sum_{z \in Y} T_{z} \\
& =\sum_{1 \leq i \leq m} \delta_{i}
\end{aligned}
$$

say. Assume now, with negligible error, that the vertices of $Y$ are produced by independent trials. Then, the $\delta_{i}$ are independent random variables with the same distribution as $\delta_{1}$, defined by

$$
\operatorname{Pr}\left[\delta_{1}=k\right]=\frac{1}{n-m}\left|\left\{z \in V(G) \mid T_{z}=k\right\}\right|, \quad 0 \leq k \leq m
$$

Conditionally on $\theta$ where $\theta \in \operatorname{mnd}(1 \pm \epsilon)$ and $E\left(\delta_{1}\right)=\theta$, and using again (1),

$$
\operatorname{Pr}\left[\left|e(X, Y)-\frac{m \theta}{n}\right| \leq \epsilon m^{2}\right] \geq 1-2 \exp \left(-2 \epsilon^{2} m\right)
$$

or

$$
\operatorname{Pr}\left[\left|e(X, Y)-\frac{m \theta}{n}\right| \leq \epsilon m^{2} d\right] \geq 1-2 \exp \left(-2 \epsilon^{2} d^{2} m\right)
$$

The conditioning event has probability at least $1-2 \exp \left(-2 \epsilon^{2} m^{2} d\right)$. We have thus, without any conditioning,

$$
\begin{aligned}
\operatorname{Pr}\left[\left|e(X, Y)-\frac{m \theta}{n}\right| \leq \epsilon m^{2} d\right] & \geq 1-2 \exp \left(-2 \epsilon^{2} d^{2} m\right)-2 \exp \left(-2 \epsilon^{2} m^{2} d\right) \\
& \geq 1-3 \exp \left(-2 \epsilon^{2} d^{2} m\right)
\end{aligned}
$$

This completes the proof.
We now return to our proof of correctness. We assume, as we can, that $a$ is the restriction to $S_{1} \cup S_{2}$ of an optimal assignment $a^{*}$. For each $y \in X$, we let $y^{a^{*}}$ denote the value of $y$ in $a^{*}$. Let $x \in X \backslash\left(S_{1} \cup S_{2}\right)$.

Let $G_{x, 0}$ and $G_{x, 1}$ be the graphs with common vertex set $V\left(G_{x, 0}\right)=V\left(G_{x, 1}\right)=$ $X$ and edge sets

$$
E\left(G_{x, 0}\right)=\left\{\{y, z\}: x \oplus y \oplus z=b \in \mathcal{S} \wedge y^{a^{*}} \oplus z^{a^{*}}=b\right\}
$$

and

$$
E\left(G_{x, 1}\right)=\left\{\{y, z\}: x \oplus y \oplus z=b \in \mathcal{S} \wedge 1 \oplus y^{a^{*}} \oplus z^{a^{*}}=b\right\}
$$

Let $n_{0}^{a^{*}}=\left|E\left(G_{x, 0}\right)\right|, n_{1}^{a^{*}}=\left|E\left(G_{x, 1}\right)\right|, n^{a^{*}}=n_{0}^{a^{*}}+n_{1}^{a^{*}}$. Also, let $m^{a}=$ $m_{0}^{a}+m_{1}^{a}$.

Lemma 2. (i) Assume that $x$ is such that we have

$$
n_{0}^{a^{*}} \geq \frac{3\left(n_{0}^{a^{*}}+n_{1}^{a^{*}}\right)}{4}
$$

Then, with probability $1-o(1 / n), x$ is assigned (correctly) to 1 in step 2.1 of algorithm B.
(ii) Assume that $x$ is such that we have

$$
n_{1}^{a^{*}} \geq \frac{3\left(n_{0}^{a^{*}}+n_{1}^{a^{*}}\right)}{4}
$$

Then, with probability $1-o(1 / n), x$ is assigned (correctly) to 0 in step 2.1 of algorithm B.
(iii) With probability $1-o(1)$, each variable $y \in D^{a}$ is assigned to its correct value $y^{a^{*}}$ by the algorithm $B$.

Proof. We first prove (iii). Suppose that $y$ is assigned to 1 in stage 2.1. The case where $y$ is assigned to 0 is similar. We have to prove that $n_{0}^{a^{*}} \geq n_{1}^{a^{*}}$ with probability $1-o(1 / n)$ since if in an optimum solution $x_{i}=1$ then $n_{0}^{a^{*}} \geq n_{1}^{a^{*}}$.

Thus, Lemma 1 applied to the graph $G_{x, 0}$ with $d=\frac{2 n_{0}^{a^{*}}}{n(n-1)}$ and the samples $S_{1}$ and $S_{2}$ gives

$$
\operatorname{Pr}\left(m_{0}^{a} \leq \frac{8 \cdot 2 n_{0}^{a^{*}} m^{2}}{7 n(n-1)}\right)=1-o(1 / n)
$$

and so,

$$
\operatorname{Pr}\left(n_{0}^{a^{*}} \geq \frac{7 m_{0}^{a} n(n-1)}{2 \cdot 8 m^{2}}\right)=1-o(1 / n)
$$

Also, Lemma 1 applied to the union of the graphs $G_{x, 0}$ and $G_{x, 1}$ with $d=\frac{2 n^{a^{*}}}{n(n-1)}$ and the samples $S_{1}$ and $S_{2}$ gives

$$
\operatorname{Pr}\left(m^{a} \geq \frac{8 \cdot 2 n^{a^{*}} m^{2}}{9 n(n-1)}\right)=1-o(1 / n)
$$

and so,

$$
\operatorname{Pr}\left(n^{a^{*}} \leq \frac{9 m^{a} n(n-1)}{2 \cdot 8 m^{2}}\right)=1-o(1 / n)
$$

Since $y$ takes value 1 in stage 2.1 and $m_{0}^{a} \geq 2 / 3 m^{a}$,

$$
\operatorname{Pr}\left(\frac{n_{0}^{a^{*}}}{n^{a^{*}}} \geq \frac{7 \cdot 2}{9 \cdot 3}\right)=1-o(1 / n)
$$

and so ,

$$
\operatorname{Pr}\left(\frac{n_{0}^{a^{*}}}{n^{a^{*}}} \geq \frac{1}{2}\right)=1-o(1 / n)
$$

Assertion (iii) follows.
Now we prove (i). The proof of (ii) is completely similar to that of (i). Lemma 1 applied to the graph $G_{x, 0}$ with $d=\frac{2 n_{0}^{a^{*}}}{n(n-1)}$ and the samples $S_{1}$ and $S_{2}$ gives

$$
\operatorname{Pr}\left(m_{0}^{a} \geq(1-\epsilon) \frac{2 m^{2}}{n(n-1)} n_{0}^{a *}\right)=1-o(1 / n)
$$

Let $m^{a}=m_{0}^{a}+m_{1}^{a}$. We apply now Lemma 1 to the union of the graphs $G_{x, 0}$ and $G_{x, 1}$. This gives

$$
\operatorname{Pr}\left(m^{a} \leq(1+\epsilon) \frac{2 m^{2}}{n(n-1)} n^{a^{*}}\right)=1-o(1 / n)
$$

Substraction gives
$\operatorname{Pr}\left(m_{0}^{a}-\frac{2 m^{a}}{3} \geq \frac{2 m^{2}}{n(n-1)}\left((1-\epsilon) n_{o}^{a^{*}}-(1+\epsilon) \frac{2\left(n_{0}^{a^{*}}+n_{1}^{a^{*}}\right)}{3}\right)\right)=1-o(1 / n)$.

Using the inequality $n_{0}^{a^{*}}+n_{1}^{a^{*}} \leq \frac{4 n_{0}^{a^{*}}}{3}$, we obtain

$$
\operatorname{Pr}\left(m_{0}^{a}-\frac{2 m^{a}}{3} \geq \frac{2 m^{2}}{n(n-1)} \frac{1-20 \epsilon}{9} n_{o}^{a^{*}}\right)=1-o(1 / n)
$$

and fixing $\epsilon=1 / 20$,

$$
\operatorname{Pr}\left(m_{0}^{a}-\frac{2 m^{a}}{3} \geq 0\right)=1-o(1 / n)
$$

concluding the proof.
Lemma 3. With probability $1-o(1)$, the number of variables undefined after the completion of stage 2.1 satisfies

$$
\left|V^{a}\right| \leq \frac{4 \mathrm{opt}}{\delta n^{2}}
$$

Proof. Assume that $x$ is undefined. We have thus simultaneously $n_{0}^{a^{*}}<\frac{3}{4}\left(n_{0}^{a^{*}}+\right.$ $\left.n_{1}^{a^{*}}\right)$ and $n_{1}^{a^{*}}<\frac{3}{4}\left(n_{0}^{a^{*}}+n_{1}^{a^{*}}\right)$ and so $n_{1}^{a^{*}}>\frac{1}{4}\left(n_{0}^{a^{*}}+n_{1}^{a^{*}}\right)$ and $n_{0}^{a^{*}}>\frac{1}{4}\left(n_{0}^{a^{*}}+n_{1}^{a^{*}}\right)$. Since $x$ appears in at least $\delta n^{2}$ equations, $n_{0}^{a^{*}}+n_{1}^{a^{*}} \geq \delta n^{2}$. Thus,

$$
o p t \geq \min \left\{n_{0}^{a^{*}}, n_{1}^{a^{*}}\right\} \cdot\left|V^{a}\right| \geq \frac{\delta n^{2}}{4}\left|V^{a}\right|
$$

The assertion of the lemma follows.
We can now complete the correctness proof. Let val denote the value of the solution given by our algorithm and let opt be the value of an optimum solution.

Theorem 1. Let $\epsilon$ be fixed. If opt $\leq \alpha n^{3}$ where $\alpha$ is sufficiently small, then we have that val $\leq(1+\epsilon)$ opt.

Proof. Let us write

$$
\mathrm{val}=\operatorname{val}_{1}+\operatorname{val}_{2}+\operatorname{val}_{3}+\operatorname{val}_{4}
$$

where:

- val $_{1}$ is the number of satisfied equations with all variables in $U^{a}$
- $\mathrm{val}_{2}$ is the number of satisfied equations with all variables in $V^{a}$
- $\mathrm{val}_{3}$ is the number of satisfied equations with two variables in $U^{a}$ and one in $V^{a}$
- $\operatorname{val}_{4}$ is the number of satisfied equations with one variable in $U^{a}$ and two in $V^{a}$.

With an obvious intended meaning, we write also

$$
\mathrm{opt}=\mathrm{opt}_{1}+\mathrm{opt}_{2}+\mathrm{opt}_{3}+\mathrm{opt}_{4}
$$

We have clearly val ${ }_{1}=$ opt $_{1}$ and $\operatorname{val}_{3} \leq$ opt $_{3}$. Thus,

$$
\begin{aligned}
\mathrm{val} & \leq \mathrm{opt}+\mathrm{val}_{2}-\mathrm{opt}_{2}+\mathrm{val}_{4}-\mathrm{opt}_{4} \\
& \leq \mathrm{opt}+\mathrm{val}_{2}+\mathrm{val}_{4} \\
& \leq \mathrm{opt}+\frac{\left|V^{a}\right|^{3}}{6}+n \frac{\left|V^{a}\right|^{2}}{2}
\end{aligned}
$$

and, using Lemma 3,

$$
\begin{aligned}
\mathrm{val} & \leq \mathrm{opt}+\frac{4^{3} \mathrm{opt}^{3}}{6 \delta^{3} n^{6}}+n \frac{4^{2} \mathrm{opt}^{2}}{2 \delta^{2} n^{4}} \\
& \leq \mathrm{opt}\left(1+\frac{32 \mathrm{opt}^{2}}{3 \delta^{3} n^{6}}+\frac{8 \mathrm{opt}}{\delta^{2} n^{3}}\right)
\end{aligned}
$$

Since opt $\leq \alpha n^{3}$ then,

$$
\begin{aligned}
\mathrm{val} & \leq \operatorname{opt}\left(1+\frac{32 \alpha^{2}}{3 \delta^{3}}+\frac{8 \alpha}{\delta^{2}}\right) \\
& \leq \operatorname{opt}(1+\epsilon)
\end{aligned}
$$

for $\alpha \leq \frac{\delta^{2} \epsilon}{9}$ and sufficiently small $\epsilon$.

## 5 Extensions to dense Min-Ek-Lin2

We are able to extend our result to arbitrary $k$-ary versions of the problem, i.e. to dense Min-E $k$-Lin2 for arbitrary $k$. This requires a bit more subtle construction, and the design of a density sampler for $(k-1)$-uniform hypergraphs. This extension appear in the final version of the paper [4].

## Acknowledgments

The authors thank Madhu Sudan and Luca Trevisan for stimulating remarks and discussions.

## References

1. S. Arora, L. Babai, J. Stern and Z. Sweedyk, The Hardness of Approximate Optima in Lattice, Codes, and Systems of Linear Equations, Proc. of 34th IEEE FOCS, 1993, 724-733.
2. S. Arora, D. Karger and M. Karpinski, Polynomial Time Approximation Schemes for Dense Instances of NP-hard Problems, Proc. of 27th ACM STOC, 1995, 284293; the full paper appeared in Journal of Computer and System Sciences 58 (1999), 193-210.
3. C. Bazgan and W. Fernandez de la Vega, A Polynomial Time Approximation Scheme for Dense Min 2Sat, Proc. Fundamentals of Computation Theory, LNCS 1684, Springer, 1999, 91-99.
4. C. Bazgan, W. Fernandez de la Vega and M. Karpinski, Polynomial Time Approximation Schemes for Dense Instances of Minimum Constraint Satisfaction, ECCC Technical Report TR01-034, 2001.
5. P. Berman and M. Karpinski, Approximating Minimum Unsatisfiability of Linear Equations, Proc. 13th ACM-SIAM SODA, 2002, 514-516.
6. A.E.F. Clementi and L. Trevisan, Improved Non-Approximability Results for Vertex Cover with Density Constraints, Proc. of 2nd Conference on Computing and Combinatorics, 1996, Springer, 1996, 333-342.
7. I. Dinur, G. Kindler, R. Raz and S. Safra, An Improved Lower Bound for Approximating CVP, 2000, submitted.
8. I. Dinur, G. Kindler and S. Safra, Approximating CVP to Within Almost Polynomial Factors is NP-Hard, Proc. of 39th IEEE FOCS, 1998, 99-109.
9. W. Fernandez de la Vega, Max-Cut Has a Randomized Approximation Scheme in Dense Graphs, Random Structures and Algorithms, 8(3) (1996), 187-198.
10. W. Fernandez de la Vega and M. Karpinski, On Approximation Hardness of Dense TSP and Other Path Problem, Information Processing Letters 70, 1999, 53-55.
11. J. Hastad, Some Optimal Inapproximability Results, Proc. of 29th ACM STOC, 1997, 1-10.
12. W. Hoeffding, Probability Inequalities for Sums of Bounded Random Variables, Journal of the American Statistical Association, 58(301), 1964, 13-30.
13. M. Karpinski, Polynomial Time Approximation Schemes for Some Dense Instances of NP-Hard Optimization Problems, Randomization and Approximation Techniques in Computer Science, LNCS 1269, Springer, 1997, 1-14.
14. M. Karpinski and A. Zelikovsky, Approximating Dense Cases of Covering Problems, ECCC Technical Report TR 97-004, 1997, appeared also in DIMACS Series in Discrete Mathematics and Theoretical Computer Science, vol. 40, 1998, 169-178.
15. S. Khanna, M. Sudan and L. Trevisan, Constraint Satisfaction: The Approximability of Minimization Problems, Proc. of 12th IEEE Computational Complexity, 1997, 282-296.
