Approximating coloring and maximum independent sets in 3-uniform hypergraphs *

Michael Krivelevich † *Ram Nathaniel* ‡ *Benny Sudakov* §

Abstract

We discuss approximation algorithms for the coloring problem and the maximum independent set problem in 3-uniform hypergraphs. An algorithm for coloring 3-uniform 2-colorable hypergraphs in $\tilde{O}(n^{1/5})$ colors is presented, improving previously known results. Also, for every fixed $\gamma > 1/2$, we describe an algorithm that, given a 3-uniform hypergraph H on n vertices with an independent set of size γn, finds an independent set of size $\tilde{\Omega}(\min(n, n^{6\gamma-3}))$. For certain values of γ we are able to improve this using the Local Ratio Approach. The results are obtained through Semidefinite Programming relaxations of these optimization problems.

1 Introduction

An *r-uniform hypergraph* H is an ordered pair $H = (V, E)$, where V is a finite non-empty set (the set of *vertices*), and E is a collection of distinct r-subsets of V (the set of *edges*). Thus a 2-uniform hypergraph is just a graph. A subset $I \subseteq V(H)$ is called *independent* if I does not contain any edge of H. The maximal size of an independent set in H is called the *independence number* of H and is denoted by $\alpha(H)$. A *k-coloring* of H is a partition $V(H) = C_1 \cup \ldots \cup C_k$ so that each color class C_i is an independent set. The *chromatic number* of H, denoted by $\chi(H)$, is the minimal k for which H admits a k-coloring. We use standard notation $f(n) = \tilde{O}(g(n))$ if there exists a constant $c > 0$ so that $f(n) = O(g(n) \log^c n)$. Similarly, $f(n) = \tilde{\Omega}(g(n))$ if $g(n) = \tilde{O}(f(n))$.

In this paper we discuss algorithmic problems of approximating the chromatic number and the independence number of k-uniform hypergraphs. The case $k = 2$ corresponds to the very extensively

*An extended abstract of this paper appeared in the Proceedings of the 12^{th} Annual Symposium on Discrete Algorithms (SODA'2001).

†Department of Mathematics, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Email: krivelev@math.tau.ac.il. Supported by a USA-Israeli BSF grant.

‡Department of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel. Email: ramn@math.tau.ac.il.

§Department of Mathematics, Princeton University, Princeton, NJ 08540, USA and Institute for Advanced Study, Princeton, NJ 08540, USA. Email address: bsudakov@math.princeton.edu. Research supported in part by NSF grant and by the State of New Jersey.

studied problems of approximating the chromatic number /the independence number of graphs, two of the most important problems in Combinatorial Optimization. For the coloring problem, the best known result belongs to Halldórsson who obtained in [12] a coloring algorithm with approximation ratio $O(n(\log\log n)^2/\log^3 n)$ for graphs on n vertices. On the negative side, Feige and Kilian showed [9] that, for any fixed $\epsilon > 0$, the chromatic number of graphs on n vertices is not approximable within a factor of $n^{1-\epsilon}$ unless $NP = ZPP$. If a graph on n vertices is 3-colorable, then one can color it using $O(n^{3/14}\log^{O(1)} n)$ colors [5] but it is NP-hard to color it using four colors [17]. As for approximating the independence number of a graph, Boppana and Halldórsson presented an algorithm ([6]) with approximation ratio $O(n/\log^2 n)$ for graphs on n vertices, based on the so called Local Ratio Approach, to be discussed later in this paper. If a graph contains an independent set of size $\gamma n + m$, for a constant $0 < \gamma < 1/2$, then an independent set of size $\tilde{O}(m^{\frac{3\gamma}{\gamma+1}})$ can be found in polynomial time [1]. On the other hand, Håstad proved [14] that it is impossible to approximate $\alpha(G)$ within a factor of $n^{1-\epsilon}$, unless $NP = ZPP$. Even stronger non-approximability result has been obtained recently by Engebretsen and Holmerin [8]. The reader can consult [3] for the account of the state of the art in these two optimization problems.

In contrast, much less is known on the hypergraph versions of these problems. Krivelevich and Sudakov [19] developed a coloring algorithm with approximation ratio $O(n(\log\log n/\log n)^2)$ for r-uniform hypergraphs on n vertices. Algorithms for coloring k-uniform 2-colorable hypergraphs have been proposed in [2, 7, 19]. As for negative results, it is fairly easy to show that for every fixed $k \geq 3$, approximating the chromatic number of a k-uniform hypergraph is at least as hard as the corresponding problem for graphs [15, 19]. Very recently, Guruswami, Håstad and Sudan showed [11] that for any constant c it is NP-hard to color 4-uniform 2-colorable hypergraphs using c colors. Naturally, this result stresses the importance of developing good approximation algorithms for coloring 2-colorable hypergraphs. The only paper on approximating the independence number of uniform hypergraphs we are aware of is [15], whose main result is significantly weaker than those known for graphs. Also, essentially the same reduction as for the chromatic number shows that approximating the independence number of r-uniform hypergraphs is at least as hard as the graph case $r = 2$ [19].

Here we propose approximation algorithms for coloring and independent set problems in 3-uniform hypergraphs. It appears that the 3-uniform case stands apart from other uniformity numbers $k \geq 4$, as in this case the powerful machinery of Semidefinite Programming can be applied to produce better approximation algorithms (see [2] for a relevant discussion). In Section 2 we discuss an algorithm for finding a large independent set in hypergraphs on n vertices with an independent set of size at least γn, for a constant $\gamma > 0$. There we propose an algorithm for finding an independent set of size $\tilde{\Omega}(\min(n, n^{6\gamma-3}))$ in such a hypergraph. Then in the next section we show how to improve our results from Section 2 for some values of γ, using the Local Ratio approach. In Section 4 we use algorithm developed in Section 2 as a subroutine to color 3-uniform 2-colorable hypergraphs using $\tilde{O}(n^{1/5})$ colors in polynomial time, thus improving the $\tilde{O}(n^{9/41})$ algorithm from [19] and the previous results from [2],

[7]. The final section is devoted to concluding remarks. Throughout the paper we assume, whenever this is needed, that n is sufficiently large. We also omit all floor and ceiling signs whenever these are not crucial, to simplify the presentation.

2 Finding large independent sets

In this section we discuss an algorithm for approximating the maximum independent set in a 3-uniform hypergraph. We will assume that an input hypergraph H on n vertices contains an independent set of size γn, where $0 < \gamma < 1$ is a constant. The performance of our algorithm depends on γ. The graph version of this problem has been tackled by Boppana and Halldórsson [6] using the subgraph exclusion argument, and then by Alon and Kahale [1] based on the Lovász θ-function. We obtain the following result:

Theorem 1 *Let H be a 3-uniform hypergraph on n vertices, m edges and with an independent set of size at least γn, for some constant $\gamma > 0$. There exists a polynomial time algorithm which finds in H an independent set of size $\tilde{\Omega}(\min(n, n^{3-3\gamma}/m^{2-3\gamma}))$.*

Proof. We will proceed as follows: first a relevant semidefinite program is formulated; solving it will enable us to find a set of unit vectors in R^n with large angles between each pair of them; then a rounding procedure will find a large independent set in the corresponding subset of vertices.

We first formulate a Semidefinite Programming relaxation of the maximum independent set problem as follows:

$$\begin{aligned} \max \quad & \sum_i \frac{1 - v_0^T v_i}{2} \\ \text{s.t.} \quad & \|v_0\| = \|v_i\| = 1 \ , \text{ for } 1 \le i \le n \ , \\ & v_i^T v_j + v_i^T v_k + v_j^T v_k \le v_0^T (v_i + v_j + v_k) \ , \text{ for } \{i,j,k\} \in E(H) \ . \end{aligned}$$

The semidefinite program written above is indeed a relaxation of the independent set problem since, for any independent set I, assigning $v_0 = -1$, $v_i = 1$ if $i \in I$ and $v_i = -1$ otherwise yields that $\sum_i \frac{1-v_0^T v_i}{2} = |I|$. It is also easy to verify that, since I is an independent set, the constraints of the relaxation are satisfied. For every $\epsilon > 0$, the above semidefinite program can be solved within an absolute error of ϵ in time polynomial in n and $\log(1/\epsilon)$ (see e.g. [10]).

Fix $\delta = 1/\log n$ and let $v_0, v_1, \ldots, v_n$ be an optimal solution of the above semidefinite program. For every integer $1 \le t \le \frac{2-2\gamma}{\delta} + 1$, we define

$$S_t = \{1 \le i \le n | -1 + (t-1)\delta \le v_0^T v_i \le -1 + t\delta\} \ .$$

Note that there are at most $\frac{2}{\delta} + 1$ such sets and also the sets S_t split the vectors v_i into subsets with almost equal inner products with v_0.

Lemma 2 *There exists an index t such that $|S_t| \ge \frac{n\delta^2}{2(2+\delta)} = \Omega(\frac{n}{\log^2 n})$.*

Proof. Suppose that for all t, $|S_t| < \frac{n\delta^2}{2(2+\delta)}$. Since there are at most $\frac{2}{\delta}+1$ such sets we obtain that $|\cup_t S_t| < \delta n/2$. Therefore all but at most $\delta n/2$ vectors v_i satisfy that $v_0^T v_i > 1-2\gamma+\delta$. Since for all other vectors $v_0^T v_i \geq -1$, this implies that the value of the semidefinite relaxation is at most

$$\sum_i \frac{1-v_0^T v_i}{2} < (n-\delta n/2)(\gamma-\delta/2)+\delta n/2 = \gamma n - \delta(\gamma-\delta/2)n/2 < \gamma n.$$

This contradicts the fact that the size of the maximum independent set in H is at least γn. □

We have shown that there exists a set S_t of size $|S_t| = \Omega(n/\log^2 n)$. Let S' be such a set, and let H_1 be a subhypergraph of H induced by S'. By definition H_1 is a 3-uniform hypergraph with $n_1 = \tilde{\Omega}(n)$ vertices and $m_1 \leq m$ edges. We show how to find a large independent set in H_1. Note that the scalar products $v_0^T v_i$ are roughly the same for all i from S' and in the worst case their values are at most $(1-2\gamma)+1/\log n$. As $\gamma > 1/2$, the inner products $v_0^T v_i$ are negative. We will show that this observation can be used to find a large independent set inside S'. Let u_i be the normalized projection of the vectors $v_i \in S'$ on the subspace orthogonal to v_0. Then these vectors have the following property.

Lemma 3 *If* $(i,j,k) \in E(H_1)$, *then*

$$u_i^T u_j + u_i^T u_k + u_j^T u_k \leq \frac{3(1-2\gamma)}{2(1-\gamma)} + O(1/\log n).$$

Proof. By the definition of S' there exists a real $a \leq 1-2\gamma$ such that $a \leq v_0^T v_i \leq a+1/\log n$ for all $i \in S'$. In addition, as $\{i,j,k\} \subset S'$ is an edge of H, we have that $v_i^T v_j + v_i^T v_k + v_j^T v_k \leq v_0^T(v_i+v_j+v_k) \leq 3a + O(1/\log n)$. As $v_i^T vj + v_i^T v_k + v_j^T v_k = (\|v_i+v_j+v_k\|^2 - \|v_i\|^2 - \|v_j\|^2 - \|v_k\|^2)/2 \geq -3/2$, we may assume that $a \geq -1/2 + O(1/\log n)$, since otherwise S' contains no edges and we are done. Let a vector $u_i' = v_i - (v_0^T v_i)v_0$ be the projection of v_i on the subspace orthogonal to v_0. The vectors u_i are obtained by normalizing the vectors u_i'. Since the scalar products $v_0^T v_i$ are roughly equal a, this implies that $||u_i'||^2 \geq 1 - a^2 - O(1/\log n)$. Therefore

$$\begin{aligned}
\frac{u_i^T u_j + u_i^T u_k + u_j^T u_k)}{1-a^2} - O\left(\frac{1}{\log n}\right) &\leq u_i'^T u_j' + u_i'^T u_k' + u_j'^T u_k' \\
&= v_i^T v_j - (v_i^T v_0)(v_j^T v_0) + v_i^T v_k - (v_i^T v_0)(v_k^T v_0) + v_j^T v_k - (v_j^T v_0)(v_k^T v_0) \\
&\leq v_i^T v_j + v_i^T v_k + v_j^T v_k - 3a^2 + O\left(\frac{1}{\log n}\right) \\
&\leq v_0^T(v_i+v_j+v_k) - 3a^2 + O\left(\frac{1}{\log n}\right) \\
&\leq 3a - 3a^2 + O\left(\frac{1}{\log n}\right).
\end{aligned}$$

This yields

$$u_i^T u_j + u_i^T u_k + u_j^T u_k \leq \frac{3a-3a^2}{1-a^2} + O\left(\frac{1}{\log n}\right) = \frac{3a}{1+a} + O\left(\frac{1}{\log n}\right) \leq \frac{3(1-2\gamma)}{2(1-\gamma)} + O\left(\frac{1}{\log n}\right),$$

where the last inequality follows from the facts that $-1 < a \le 1 - 2\gamma$ and $f(a) = 3a/(1+a)$ is an increasing function of a in this range. □

Notice that as u_i, u_j, u_k are all unit vectors, we have $0 \le \|u_i+u_j+u_k\|^2 = 3+2(u_i^T u_j+u_i^T u_k+u_j^T u_k)$, implying that $u_i^T u_j + u_i^T u_k + u_j^T u_k \ge -3/2$. Then it follows from the above lemma that for $\gamma > 2/3$ the set S' contains no edges, i.e. is independent. Thus in the remaining part of this section we assume that $\gamma \le 2/3$. We can also assume that $m_1 > 3n_1$, since otherwise a simple greedy algorithm, which picks each time a vertex of minimal degree and deletes all vertices closing an edge with two already chosen ones, finds a linear size independent set in H_1.

Next we show how to find a large independent set in S', using the vectors u_i. This rounding algorithm is very similar to an algorithm of Karger, Motwani and Sudan [16] for coloring k-colorable graphs. Choose a random vector r according to the standard n-dimensional normal distribution, i.e. each component of r is an independent random variable with the standard normal distribution. Let $I = \{i \in S' \mid u_i^T r \ge c\}$, for some c which we specify later. Let $n' = |I|$ be the size of I and let $m' = |\{(i,j,k) \in E \mid i,j,k \in I\}|$ be the number of edges contained in I. An independent set I' of size $n' - m'$ is then easily obtained by removing one vertex from each edge contained in I. To finish the proof we will show that there exists a c such that the expected size of I' is $\tilde{\Omega}(\min(n_1, n_1^{3-3\gamma}/m_1^{2-3\gamma})) = \tilde{\Omega}(\min(n, n^{3-3\gamma}/m^{2-3\gamma}))$.

Let $N(x) = \int_x^\infty \phi(y)dy$, where $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, denote the tail of the standard normal distribution. It is well known that $(\frac{1}{x} - \frac{1}{x^3})\phi(x) \le N(x) \le \frac{1}{x}\phi(x)$, for every $x > 0$. It is also known that if v is an arbitrary unit vector in $\mathbb{R}^n$, and r is a random vector chosen according to the standard n-dimensional normal distribution, then the inner product $v^t r$ is distributed according to the standard one dimensional normal distribution. Therefore

$$\begin{aligned} E[n'] &= n_1 \Pr[u_1^T r \ge c] \;=\; n_1 N(c)\ , \\ E[m'] &= \sum_{(i,j,k)\in E(H_1)} \Pr[u_i^T r \ge c \text{ and } u_j^T r \ge c \text{ and } u_k^T r \ge c] \\ &\le \sum_{(i,j,k)\in E(H_1)} \Pr[(u_i+u_j+u_k)^T r \ge 3c]\ . \end{aligned}$$

Fix an edge $(i,j,k) \in E(H_1)$. Our aim is to estimate from above the probability $\Pr[(u_i+u_j+u_k)^T r \ge 3c]$. If $\|u_i+u_j+u_k\| = 0$, then for any positive c one has $\Pr[(u_i+u_j+u_k)^T r \ge 3c] = 0$. Therefore we may assume that $\|u_i+u_j+u_k\| > 0$. In this case according to the said above

$$\Pr[(u_i+u_j+u_k)^T r \ge c] = N\left(\frac{3c}{\|u_i+u_j+u_k\|}\right)\ .$$

By Lemma 3 we know that

$$\|u_i+u_j+u_k\|^2 \le 3 + 3\cdot\frac{1-2\gamma}{1-\gamma} + O(1/\log n) = 3\cdot\frac{2-3\gamma}{1-\gamma} + O(1/\log n).$$

Consider first the case $\gamma = 2/3$. In this case we obtain $\|u_i + u_j + u_k\|^2 = O(1/\log n)$, implying:

$$\Pr[(u_i + u_j + u_k)^T r \geq 3c] \leq N(3c \cdot \Omega(\log^{1/2} n)) = e^{-\Omega(c^2 \log n)} .$$

Choosing c to be a sufficiently large constant we derive:

$$\begin{aligned} E[n'] &= \Theta(n_1) = \Omega\left(\frac{n}{\log^2 n}\right) ; \\ E[m'] &= m_1 \cdot e^{-\Omega(c^2 \log n)} = o(n_1) . \end{aligned}$$

and thus $E[I'] = \Omega(n/\log^2 n)$, proving the desired bound.

If $\gamma < 2/3$, then for every edge $(i,j,k) \in E(H_1)$ it follows that

$$\Pr[i,j,k \in I] \leq N\left(3c\left(\sqrt{(1-\gamma)/(3(2-3\gamma))} + O(1/\log n)\right)\right) .$$

Therefore

$$E[|I'|] \geq E[n'] - E[m'] \geq n_1 N(c) - m_1 N\left(3c\left(\left(\frac{1-\gamma}{3(2-3\gamma)}\right)^{\frac{1}{2}} + O(1/\log n)\right)\right) .$$

Fix c to be $\sqrt{(4-6\gamma)\ln(m_1/n_1) + \Theta(1)}$. Then by choosing an appropriate constant $\Theta(1)$ we can make $E[m'] \leq n_1 N(c)/2 = E[n']/2$, and thus

$$\begin{aligned} E[|I'|] &\geq \frac{n_1}{2} N(c) \geq \frac{n_1}{2}\left(\frac{1}{c} - \frac{1}{c^3}\right)\frac{1}{\sqrt{2\pi}} e^{-c^2/2} \\ &= \tilde{\Omega}\left(\frac{n_1^{3-3\gamma}}{m_1^{2-3\gamma}}\right) = \tilde{\Omega}\left(\frac{n^{3-3\gamma}}{m^{2-3\gamma}}\right) , \end{aligned}$$

as promised. □

Note that even though our algorithm is randomized, it can be easily derandomized using the technique of Mahajan and Ramesh [20].

We will finish this section with the following immediate consequence of the last theorem.

Corollary 4 *If $\gamma > 1/2$ and H is a 3-uniform hypergraph on n vertices with an independent set of size at least γn, then there is a polynomial time algorithm which finds in H an independent set of size $\tilde{\Omega}(\min(n, n^{6\gamma-3}))$.*

Proof. Note that in Theorem 1 the value of m in the worst case is at most $O(n^3)$. □

3 Using the Local Ratio Approach

The so called *Local Ratio Approach* was pioneered by Bar-Yehuda and Even [4], who used it to develop an approximate graph vertex cover algorithm. Boppana and Halldórsson [6] used the Local Ratio Approach in their algorithm for approximating the independence number of a graph. In [18], this approach has been applied to a hypergraph approximation problem. Here we use the Local Ratio Lemma to improve the result of Theorem 1 for certain values of the parameter γ. Our approach is similar in spirit to that of [6].

Notation. For a hypergraph $H=(V,E)$ the *local ratio* of H is defined by:

$$lr(H)=\frac{\alpha(H)}{|V(H)|}\ .$$

Lemma 5 (**Local Ratio Lemma.**) *Let $H_0=(V_0,E_0)$ be a fixed hypergraph with $lr(H_0)=\gamma$. Let $H=(V,E)$ be a hypergraph on n vertices with $\alpha(H)\geq\gamma n+m$. Start with $H'=H$ and as long as H' contains a copy of H_0, delete from H' all vertices of this copy. Then the obtained hypergraph H' has the following properties:*

1. *H' does not contain a copy of H_0;*
2. *$|V(H')|\geq\alpha(H')\geq\gamma|V(H')|+m\geq m$.*

Proof. Property 1 is obvious from the construction of H'.

For the proof of Property 2, note that the first and the last inequalities hold trivially, so we need to prove only the middle inequality. Denote $|V(H_0)|=n_0$ and suppose we have deleted $s\geq 0$ copies of H_0 with vertex sets $U_1,\ldots,U_s$. Denote $U^*=\bigcup_{i=1}^{s}U_i$. As all U_i are pairwise disjoint, we get $|U^*|=n_0s$ and thus $|V(H')|=n-n_0s$. Let $I\subseteq V(H)$ be an independent set in H of size $|I|=\gamma n+m$. Then for every $1\leq i\leq s$ one has $|I\cap U_i|\leq\alpha(H_0)=\gamma n_0$. Hence $|I\cap U^*|\leq\gamma n_0 s$. This implies that

$$\alpha(H')\geq|I\setminus U^*|=|I|-|I\cap U^*|\geq(\gamma n+m)-\gamma n_0s=\gamma(n-n_0s)+m=\gamma|V(H')|+m,$$

thus establishing Property 2. □

Let $\{H_k\}_{k=1}^{\infty}$ be the family of 3-uniform hypergraphs defined by the following recursive construction: $V(H_1)=\{1,2,3\}$, $E(H_1)=\{(1,2,3)\}$; for all $k\geq 2$, $V(H_k)=\{1,2,\ldots,2k+1\}$, $E(H_k)=E(H_{k-1})\cup\{(i,2k,2k+1)|1\leq i\leq 2k-1\}$. Clearly the number of vertices in H_k is $2k+1$ and it is easy to prove by induction that $\alpha(H_k)=k+1$. Indeed, first note that the set $\{1,3,\ldots,2k-1,2k+1\}$ is independent in H_k. Next, if I is independent in H_k, then either $2k,2k+1\in I$ and then $I\cap\{1,\ldots,2k-1\}=\emptyset$, or $|I\cap\{2k,2k+1\}|\leq 1$ and then by induction $|I|\leq 1+\alpha(H_{k-1})\leq k+1$. Therefore $lr(H_k)=\frac{k+1}{2k+1}$. Note that the hypergraph H_k contains all hypergraphs H_i, $1\leq i\leq k-1$, as subhypergraphs.

For a fixed integer k, let t_k be the constant whose value is defined by the recursion:

$$s_{k,1}=1,\qquad s_{k,i+1}=\frac{(k+2)s_{k,i}}{k-1+(2k+1)s_{k,i}},\quad 1\leq i\leq k-1,\qquad t_k=s_{k,k}.$$

We prove the following theorem.

Theorem 6 *Let H be a 3-uniform hypergraph with n vertices and an independent set of size at least $(\frac{k+1}{2k+1}+\epsilon(n))n$, where $\epsilon(n)=\Omega(1/\log^{\Theta(1)} n)$. Then there exists a polynomial time algorithm which finds in H an independent set of size $\tilde{\Omega}(n^{t_k})$.*

Proof. Since $lr(H_k)=(k+1)/(2k+1)$, by the Local Ratio Lemma we can exclude all copies of H_k from the hypergraph H and obtain a hypergraph H' with $n'=\tilde{\Omega}(n)$ vertices and an independent set of size at least $\frac{k+1}{2k+1}n'+\epsilon(n)n$.

Now we show by induction that, for all $1\le i\le k$, if H' is a hypergraph on n vertices, not containing a copy of the hypergraph H_i, then there is a polynomial time algorithm which either finds in H' an independent set of size $\tilde{\Omega}(n^{s_{k,i}})$, or outputs a subset $U\subseteq V(H')$ so that $lr(H'[U])<(k+1)/(2k+1)$.

Basis of induction. If $i=1$, then by the definition of H_1 we have that H' contains no edges. Therefore the vertex set of H' is an independent set of size $n=n^{s_{k,1}}$, as desired.

Induction step. Suppose first that the number of edges in H' satisfies:

$$m\le n^{\frac{3k+(4k+2)s_{k,i}}{k-1+(2k+1)s_{k,i}}}.$$

If $lr(H')\ge(k+1)/(2k+1)$, then by applying our algorithm from Section 2 we can find an independent set of size

$$\tilde{\Omega}\left(\frac{n^{\frac{3k}{2k+1}}}{m^{\frac{k-1}{2k+1}}}\right)=\tilde{\Omega}\left(n^{\frac{(k+2)s_{k,i}}{k-1+(2k+1)s_{k,i}}}\right)=\tilde{\Omega}(n^{s_{k,i+1}}).$$

If the algorithm fails to output an independent set if size $\tilde{\Omega}(n^{s_{k,i+1}})$, then H' has local ratio $lr(H')<(k+1)/(2k+1)$, and we are done again.

On the other hand, if the number of edges satisfies

$$m\ge n^{\frac{3k+(4k+2)s_{k,i}}{k-1+(2k+1)s_{k,i}}},$$

then we clearly can find a pair of vertices u and v in H' such that $d(u,v)\ge m/n^2$. Note that by definition, the subhypergraph induced by $N(u,v)$ does not contains a copy of H_i. Therefore by applying the induction hypothesis to this hypergraph we will either find an independent set of size

$$\tilde{\Omega}\left(|N(u,v)|^{s_{k,i}}\right)\ge\tilde{\Omega}\left(\left(\frac{m}{n^2}\right)^{s_{k,i}}\right)\ge\tilde{\Omega}(n^{s_{k,i+1}}),$$

or a subset U so that $lr(H'[U])<(k+1)/(2k+1)$. This completes the proof of the induction step.

Now we apply repeatedly the above algorithm with $i=k$. If the algorithm finds a subset U for which $lr(H'[U])<(k+1)/(2k+1)$, then we delete U and proceed. Notice that by the Local Ratio Lemma deleting from H' such a subset U leaves the local ratio of the obtained hypergraph above $(k+1)/(2k+1)$. Recall that by our assumption $lr(H')\ge(\frac{k+1}{2k+1}+\epsilon(n))n$. Therefore during the repeated calls of the algorithm the number of vertices in the hypergraph is always as large as $\epsilon(n)n=\tilde{\Omega}(n)$.

Hence at some point the algorithm will return an independent set of size at least $\tilde{\Omega}((\epsilon(n)n)^{t_k}) = \tilde{\Omega}(n^{t_k})$, as desired. $\square$

It is easy to show by induction that t_k is always greater than $3/(2k+1)$ for all $k \geq 2$. This implies that the result of Theorem 6 improves Corollary 4. To conclude, we present a table which compares quantitatively these two results for the first few values of k.

$\gamma = \frac{\alpha(H)}{\|V(H)\|} = \frac{k+1}{2k+1}$	$\frac{2}{3}$	$\frac{3}{5}$	$\frac{4}{7}$	$\frac{5}{9}$	$\frac{6}{11}$
Result of Corollary 4	1	$\frac{3}{5} = 0.6$	$\frac{3}{7} = 0.428\ldots$	$\frac{1}{3} = 0.333\ldots$	$\frac{3}{11} = 0.272\ldots$
Result of Theorem 6	1	$\frac{2}{3} = 0.666\ldots$	$\frac{25}{53} = 0.471\ldots$	$\frac{4}{11} = 0.363\ldots$	$\frac{2401}{8121} = 0.295\ldots$

Table 1: The exponents of the algorithms of Corollary 4 and of Theorem 6.

4 Coloring 2-colorable hypergraphs

In this section we present an algorithm for coloring 3-uniform 2-colorable hypergraphs, which improves the previous results from [19], [2], [7]. We will prove the following statement.

Theorem 7 *Let H be a 3-uniform 2-colorable hypergraph on n vertices. Then there is a polynomial time algorithm which colors H using $\tilde{O}(n^{1/5})$ colors.*

Proof. Let us start by defining terminology and notation to be used later. Given a 3-uniform hypergraph $H = (V, E)$, for a pair of vertices $v, u \in V$ we denote $N(u,v) = \{w \in V : (u,v,w) \in E\}$ and also $d(u,v) = |N(u,v)|$. H is *linear* if every pair of edges of H has at most one vertex in common, that is, $d(u,v) \leq 1$ for every $u, v \in V$. Also, the *neighborhood* of $v \in V$ is defined as $N(v) = \{u \in V \setminus \{v\} : \exists w \in V, (u,v,w) \in E(H)\}$.

The algorithm is obtained by combining Corollary 4 and the ideas from [19]. To simplify the presentation we avoid specific constants and polylogarithmic factors, hiding them in the standard $\tilde{O}, \tilde{\Omega}$-notation. The exact form of these expressions can be easily figured out. As usual, to produce an $\tilde{O}(n^{\frac{1}{5}})$ coloring of H, it is enough to be able to find an independent set of size $\tilde{\Omega}(n^{4/5})$. Here is our plan: we first reduce all codegrees $d(u,v)$ to $n^{4/5}$; then in the obtained hypergraph we find a linear subhypergraph with all degrees at least $\Omega(n^{4/5})$; finally we use this subhypergraph to find a subset $V_0 \subseteq V(H)$ of size

$|V_0| = \Omega(n^{4/5})$ so that V_0 contains an independent set two thirds of its size; finally we use Corollary 4 to find an independent set of size $\tilde{\Omega}(n^{4/5})$ inside V_0.

In our algorithm we will use the semidefinite programming subroutine of [7] and [2] to find large independent sets. The analysis in these papers yields the existence of the following procedure:

Procedure $Semidef(H)$
Input: A 3-uniform 2-colorable hypergraph $H = (V, E)$ on n vertices with $m \geq n$ edges.
Output: An independent set I of size $|I| = \tilde{\Omega}(\frac{n^{9/8}}{m^{1/8}})$.

(Both papers [7] and [2] show that a 3-uniform 2-colorable hypergraph $H = (V, E)$ with a maximal degree Δ can be colored in $\tilde{O}(\Delta^{1/8})$ colors (see, e.g. display (4) of [7] or Corollary 1 of [2]). Given a 3-uniform 2-colorable hypergraph $H = (V, E)$ with n vertices and m edges, we first, as long as H contains a vertex of degree at least m/n, delete such a vertex and all incident edges. This procedure will stop with a 3-uniform hypergraph $H' \subseteq H$ with $\Theta(n)$ vertices and with all degrees at most m/n. Applying to this hypergraph the above mentioned result of [7], [2], we get the above stated procedure $Semidef(H)$.)

Our algorithm consists of three main steps.

Step 1. Consider a pair of vertices u, v such that $d(u, v) \geq n^{4/5}$. If there is no such pair proceed to Step 2. If $N(u, v)$ contains no edges of H, we are done, since we have found an independent set of size $n^{4/5}$. Otherwise in any legal 2-coloring of H the colors of u and v are different. Delete all edges of H containing both u and v and replace them by an edge (u, v). The obtained hypergraph is still 2-colorable. Repeat this procedure while there is such a pair of vertices as above. Note that even though this process creates edges of size two, this can be easily overcome. We simply ignore all edges of size two that appear in the course of the algorithm execution. After having colored the hypergraph, for every color class we are left with a bipartite graph that can easily be colored using two colors. We therefore need at most double the number of colors to reach a legal coloring of H. Denote the 3-uniform hypergraph obtained in the end of the first step by H_1. It has a property that for any pair of vertices $d(u, v) < n^{4/5}$.

Step 2. Consider the hypergraph H_1. If $|E(H_1)| = O(n^{13/5})$, then by applying $Semidef$ we can find an independent set of size $\tilde{\Omega}(n^{4/5})$. Otherwise, H_1 is a hypergraph with $\Omega(n^{13/5})$ edges in which there are at most $n^{4/5}$ edges containing any given pair of vertices. Therefore H_1 contains a linear subhypergraph H_1' with the same vertex set and $\Omega(n^{9/5})$ edges. This subhypergraph can be found by a simple greedy procedure: start with $E(H_1') = \emptyset$. As long as there exists an edge $e \in E(H_1) \setminus E(H_1')$ such that after adding e to $E(H_1')$, H_1' will still be linear, add e to $E(H_1')$. If this procedure stops with $|E(H_1')| = t$ edges, then $3tn^{4/5} \geq |E(H_1)| - t$ since for every edge in $E(H_1')$ there are at most $3n^{4/5}$ other edges intersecting it in two vertices. This yields that $t = \Omega(n^{9/5})$. We now repeatedly delete from H_1' vertices with degree at most $t/2n$, for as long as such vertices exist. The resulting subhypergraph is denoted by H_2. The total number of deleted edges does not exceed $(t/2n)n = t/2$ and therefore $E(H_2) \neq \emptyset$. Also, every vertex in H_2 has degree at least $t/2n = \Omega(n^{4/5})$ and clearly H_2 is a linear hypergraph as well.

Step 3. Since H is a 2-colorable hypergraph, we can fix some 2-coloration $V(H) = C_1 \cup C_2$ of H. For a vertex $v \in V(H_2)$, let $C(v) \in \{C_1, C_2\}$ denote the color class of v in this coloring. Consider the following sum

$$\sum_{v \in V(H_2)} (|N_{H_2}(v) \setminus C(v)| - 2|N_{H_2}(v) \cap C(v)|) = \sum_{v \in V(H_2)} |N_{H_2} \setminus C(v)| - \sum_{v \in V(H_2)} 2|N_{H_2}(v) \cap C(v)| \ .$$

Note that every edge $e \in E(H_2)$ has exactly two vertices of one color and exactly one vertex of the opposite color. Also, since H_2 is linear, for every $v \in V(H_2)$ every vertex $u \in N(v)$ belongs to exactly one edge incident with v. Therefore every edge e contributes 4 to the first sum of the right hand side of the above equality and the same amount to the second sum. This observation shows that the sum above is equal to zero, implying in turn that at least one of the summands in the left hand side is non-negative. We infer that there exists a vertex $v_0 \in V(H_2)$ such that $N_{H_2}(v_0) = \Omega(n^{4/5})$ and $|N_{H_2}(v_0) \setminus C(v_0)| \geq (2/3)|N_{H_2}(v_0)|$. This implies that at least 2/3 of the vertices of $N_{H_2}(v_0)$ belong to the same color class and thus form an independent set in H. Let $H_3 = H[N_{H_2}(v_0)]$ be the induced subhypergraph of H, $|V(H_3)| = \Omega(n^{4/5})$. Then H_3 is 3-uniform, 2-colorable and satisfies $\alpha(H_3) \geq (2/3)|V(H_3)|$. Suppose we have this vertex v_0 at hand (we can check all the vertices in polynomial time). Then we can use Corollary 4 from Section 2, with $\gamma = 2/3$, to find an independent set of H of size $\tilde{\Omega}(n^{4/5})$ in $V(H_3)$. This completes the proof of the theorem. □

Remark. It is easy to see that our algorithm can be also used to color 2-colorable hypergraph H of *dimension* 3 (i.e. all the edges are of size 2 or 3) using only $\tilde{O}(n^{1/5})$ colors. Here n is the number of vertices of H. Indeed, we simply ignore all edges of size two in the course of the algorithm execution. After having colored the hypergraph, for every color class we are left with a bipartite graph that can easily be colored using two colors. We therefore need at most double the number of colors to reach a legal coloring of the hypergraph.

5 Concluding remarks

We have presented approximation algorithms for coloring and finding a large independent set in 3-uniform hypergraphs. Our coloring algorithm colors a 3-uniform 2-colorable hypergraph on n vertices is $\tilde{O}(n^{1/5})$ colors. For the independent set problem, we can find an independent set of size $\tilde{\Omega}(\min(n, n^{6\gamma-3}))$ in a 3-uniform hypergraph on n vertices with an independent set of size γn, where $\gamma > 0$ is a constant. Both these algorithms start by formulating and solving a Semidefinite Programming relaxation of the corresponding problem and then round an optimal vector solution in order to find a good approximation of the relevant parameter. For certain values of the parameter γ we are able to improve the latter algorithm, using the Local Ratio Approach.

A challenging question which remains open is to determine if there exists a polynomial time algorithm which finds an independent set of size at least n^{ϵ}, for some $\epsilon > 0$ in a 3-uniform hypergraph on n vertices and with independent set of size γn, where $\gamma \leq 1/2$.

Also, it would be very interesting to develop good approximation algorithms for coloring and independent set problems for k-uniform hypergraphs with $k \geq 4$.

References

[1] N. Alon and N. Kahale, Approximating the independence number via the θ-function, *Math. Programming* 80 (1998), 253–264.

[2] N. Alon, P. Kelsen, S. Mahajan and H. Ramesh, Coloring 2-colorable hypergraphs with a sublinear number of colors, *Nordic J. Comput.* 3 (1996), 425–439.

[3] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetii-Spaccamela and M. Protasi, **Complexity and approximation**, Springer-Verlag, Berlin, 1999. See also: http://www.nada.kth.se/∼viggo/problemlist/compendium.html.

[4] R. Bar-Yehuda and S. Even, A local ratio theorem for approximating the weighted vertex cover problem, in: *Analysis and design of algorithms for combinatorial problems* (G. Ausiello and M. Lucertini, eds.), Annals of Discrete Math. Vol. 25, Elsevier, Amsterdam, 1985, 27–46.

[5] A. Blum and D. Karger, an $\tilde{O}(n^{3/14})$-coloring algorithm for 3-colorable graphs, *Inform. Process. Lett.* 61 (1997), 49–53.

[6] R. Boppana and M. M. Halldórsson, Approximating maximum independent sets by excluding subgraphs, *BIT* 32 (1992), 180–196.

[7] H. Chen and A. Frieze, Coloring bipartite hypergraphs, *Proc. 5^{th} Intern. IPCO Conf.*, Lecture Notes in Comp. Sci., Vol. 1084 (1996), 345–358.

[8] L. Engebretsen and J. Holmerin, *Clique is hard to approximate within $n^{1-o(1)}$*, Proc. 27^{th} Intern. Colloquium on Automata, Languages and Programming (ICALP 2000), Lecture Notes in Comp. Sci., Vol. 1853, 2–12.

[9] U. Feige and J. Kilian, Zero knowledge and the chromatic number, *Proc. 11^{th} Annual IEEE Conf. on Computational Complexity*, 1996. Also: J. Comput. System Sci. 57 (1998), 187–199.

[10] M. Grötschel, L. Lovász and A. Schrijver, The ellipsoid method and its consequences in combinatorial optimization, *Combinatorica* 1 (1981), 169–197.

[11] V. Guruswami, J. Håstad and M. Sudan, Hardness of approximate hypergraph coloring, *Proc. 41^{st} IEEE FOCS*, IEEE (2000), 149–158.

[12] M. M. Halldórsson, A still better performance guarantee for approximate graph coloring, *Inform. Process. Lett.* 45 (1993), 19–23.

[13] E. Halperin, Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs, *Proc. 11^{th} Symposium on Discrete Algorithms (SODA'2000)*, 329–337.

[14] J. Håstad, Clique is hard to approximate within $n^{1-\epsilon}$, *Proc. 37^{th} IEEE FOCS*, IEEE (1996), 627–636. Also: Acta Math. 182 (1999), 105–142.

[15] T. Hofmeister and H. Lefmann, Approximating maximum independence sets in uniform hypergraphs, *Proc. 23^{rd} Symp. on Math. Found. Comp. Sci.*, Lecture Notes in Comp. Sci., Vol. 1450, 562–570.

[16] D. Karger, R. Motwani and M. Sudan, Approximate Graph coloring by semidefinite programming, *Journal of the ACM*, 45 (1998), 246–265.

[17] S. Khanna, N. Linial and M. Safra, On the hardness of approximating the chromatic number, *Proc. 2^{nd} Israeli Symposium on Theor. Comp. Sci.*, IEEE (1992), 250–260.

[18] M. Krivelevich, Approximate set covering in uniform hypergraphs, *J. Algorithms* 25 (1997), 118–143.

[19] M. Krivelevich and B. Sudakov, Approximate coloring of uniform hypergraphs, *Proc. 6^{th} European Symposium on Algorithms (ESA'98)*, Lecture Notes in Comp. Sci., Vol. 1461 (1998), 477–489.

[20] S. Mahajan and H. Ramesh, Derandomizing approximation algorithms based on semidefinite programming, *SIAM J. Computing* 28 (1999), 1641–1663.