COMBINATORICA
Bolyai Society – Springer-Verlag

COMBINATORICA 25 (5) (2005) 519–535

THE HARDNESS OF 3-UNIFORM HYPERGRAPH COLORING

IRIT DINUR*, ODED REGEV†, CLIFFORD SMYTH‡

Received September 29, 2002

We prove that coloring a 3-uniform 2-colorable hypergraph with c colors is NP-hard for any constant c. The best known algorithm [20] colors such a graph using $O(n^{1/5})$ colors. Our result immediately implies that for any constants $k \geq 3$ and $c_2 > c_1 > 1$, coloring a k-uniform c_1-colorable hypergraph with c_2 colors is NP-hard; the case $k=2$, however, remains wide open.

This is the first hardness result for approximately-coloring a 3-uniform hypergraph that is colorable with a constant number of colors. For $k \geq 4$ such a result has been shown by [14], who also discussed the inherent difference between the $k=3$ case and $k \geq 4$.

Our proof presents a new connection between the Long-Code and the Kneser graph, and relies on the high chromatic numbers of the Kneser graph [19,22] and the Schrijver graph [26]. We prove a certain maximization variant of the Kneser conjecture, namely that any coloring of the Kneser graph by fewer colors than its chromatic number, has 'many' non-monochromatic edges.

1. Introduction

Background

A hypergraph $H=(V,E)$ with vertices V and edges $E \subseteq 2^V$ is 3-uniform if every edge in E has exactly 3 vertices. A legal χ-coloring of a hypergraph

Mathematics Subject Classification (2000): 68Q17

* Research supported by NSF grant CCR-9987845.
† Supported by an Alon Fellowship and by NSF grant CCR-9987845.
‡ Work supported in part by NSF grants CCF-9988526 and DMS 9729992, and the State of New Jersery.

H is a function $f : V \to [\chi]$ such that no edge of H is monochromatic. The chromatic number of H is the minimal χ for which such a coloring exists.

During the past decade, significant progress has been made – via the PCP theorem – in understanding the complexity of many combinatorial optimization problems. It is known, for example, that it is hard to approximate the chromatic number of a graph to within a factor of $n^{1-\epsilon}$ [9] (this holds for hypergraphs as well, see [21]). In numerous other cases, the hardness result almost matches the algorithmic result, or perhaps some constant gap separates the two. In contrast, the problem of approximate coloring (where the input is a hypergraph or a graph that is known to be colorable with *very* few colors and the target is to color it with as few colors as possible) retains perhaps the largest gap between the hardness results and the algorithmic results.

The best algorithms for these problems require a polynomial number of colors: for example the best approximate coloring algorithm for 2-colorable 3-uniform hypergraphs requires $O(n^{1/5})$ colors [20], and the best coloring algorithm for 3-colorable graphs, requires $\tilde{O}(n^{3/14})$ colors [5].

On the hardness side, not much is known. For graphs, the best hardness result states that using 4 colors to color a 3-colorable graph is NP-hard [17, 15]. It would already be a significant step to prove that coloring a 3-colorable graph with $O(1)$ colors is NP-hard.

The property of being 2-colorable is well studied in combinatorics and is also referred to as 'property B' (see [14] for further references). Nevertheless, prior to this work no hardness of approximation result was known for 3-uniform hypergraphs, and in fact it wasn't even known if it is NP-hard to color a 3-uniform 2-colorable hypergraph with 3 colors. For 4-uniform hypergraphs and upwards, Guruswami, Håstad, and Sudan [14] were able to show such a separation, i.e., they showed that it is NP-hard to color a 2-colorable 4-uniform hypergraph with any constant number of colors (this result immediately extends to any $k \geq 4$). In their work, an inherent difference between the case $k \geq 4$ and the case $k = 2, 3$ was raised; this was considered evidence that the case $k = 3$ might be harder to understand. This difference has to do with the corresponding maximization problem called 'Set-Splitting', which is the problem of 2-coloring a k-uniform hypergraph while maximizing the number of non-monochromatic hyperedges. This problem exhibits a different behavior for $k \geq 4$ and for $k = 2, 3$. Indeed, for $k \geq 4$, a hardness result by Håstad [16] shows that finding a solution within $1 - 2^{-k+1} + \epsilon$ of the optimal is NP-hard for any $\epsilon > 0$. This result is tight since it is trivially matched by a random assignment. For $k = 2, 3$ the best approximation algorithms use semi-definite programming [11, 27] and have a constant gap from the best gadget-constructed hardness results, see [13].

Guruswami, Håstad, and Sudan [14] also showed that unless NP $\subseteq$ DTIME($n^{O(\log\log n)}$), there is no polynomial-time algorithm for coloring a 2-colorable 4-uniform hypergraph with $c_0 \frac{\log\log n}{\log\log\log n}$ colors for some constant $c_0 > 0$. We can show that unless NP $\subseteq$ DTIME($2^{\mathrm{poly}(\log n)}$), there is no polynomial-time algorithm for coloring a 2-colorable 3-uniform hypergraph with $O(\sqrt[3]{\log\log n})$ colors.

Following our result, Khot [18] was able to prove, via different techniques, that it is NP-hard to color a 3-colorable 3-uniform hypergraph with any constant number of colors. Although this result is already contained in ours, his construction has the extra property that the bad instances do not contain even a small independent set, while in our hypergraph construction, this is not the case.

Technique

Our technique is based on the PCP methodology, which we now briefly describe. A PCP system is a collection of variables and tests on them, such that each test is on a constant number of variables (say two). The PCP theorem [2, 1] states that given a PCP system, it is NP-hard to decide whether there exists an assignment that satisfies *all* the tests, or any assignment satisfies only a *small fraction* of the tests. Usually, hardness of approximation results are derived by composing a PCP system (which, in the context of the composition, is called the 'outer-verifier') with a second 'inner-verifier' PCP system. In this composition, each variable in the outer system is encoded via an error-correcting code called the Long-Code [4]. The tests of the outer system are replaced by new 'inner' tests, that are tailored to the specific problem whose hardness is being studied. In our case, these tests will be 3-uniform hyperedges whose coloring must not be monochromatic. The trick is to obtain the correct interplay between the outer and inner PCP systems so as to capture the hardness of the problem. Our construction combines the PCP of [6] as the outer system, and an inner PCP system whose properties are derived from properties of the Kneser and Schrijver graphs. We next describe the PCP of [6], which is known as the layered PCP.

The Layered PCP. PCP systems can be classified according to the types of tests they contain. A PCP system is a *projection* PCP system if all of the tests are projections, i.e., the tests are over two variables, and any assignment to the first variable determines the assignment to the second variable. For technical reasons, non-projection PCP systems do not combine well with Long-Code inner verifiers and hence we focus on projection PCP systems.

The PCP theorem [2,1] combined with Raz's parallel repetition theorem [25] gives a powerful projection PCP system. In this system, there are two types of variables, X and Y. Each test looks at one X-variable and one Y-variable and checks that the assignment given to the Y-variable matches the one determined by the X-variable. Thus, the test graph is bipartite with parts X and Y. In our reduction, we replace each variable by a set of vertices, and each test by a collection of hyperedges. The bipartite-ness of the tests poses a problem because coloring all the X vertices red, and all the Y vertices orange, will be a legal 2-coloring regardless of whether the initial PCP system was satisfiable or not. We overcome the bipartite-ness problem by utilizing a layered PCP that was constructed in [6]. This is a projection PCP system that essentially extends the bipartite nature of the standard PCP described above into being *multipartite*, with many 'types' of variables (rather than only X and Y). Thus, as the number of parts increases, the number of colors required to color an unsatisfiable instance increases as well.

The Long-Code and the Kneser Graph. The Long-Code [4] of a domain R is an important component in numerous hardness of approximation results. One way to view it [7] is as the graph whose $2^{|R|}$ vertices are the subsets of R, and whose edges connect disjoint subsets.[1] In this paper, we consider an induced subgraph, consisting only of vertices that correspond to subsets of a prescribed size (namely, $|R|/2 - O(1)$). This is known as the Kneser graph [19].

We can think of colorings of this graph as ways of encoding values in R. Let us consider the following encoding of an element $a \in R$: color all subsets containing a red, and the rest orange. It is easy to see that *in the Kneser Graph* this coloring has no *red* monochromatic edges, yet has many orange ones, and in particular it is not a 2-coloring. Nevertheless, in our *hypergraph* this coloring corresponds to a legal 2-coloring. We prove the following interesting property of the Kneser graph. Namely, that any coloring of the vertices by a constant number of colors contains one 'special' color that is used in a non-negligible fraction of the vertices and colors two disjoint subsets. This property enables a local list-decoding of the coloring in a manner that ensures global consistency. It is established by proving a maximization variant of the Kneser conjecture showing that using less than the chromatic number colors to color the Kneser graph, leaves 'many' monochromatic edges.

[1] In the original and standard definition of the Long-Code [4], the Long-Code is the collection of all possible Boolean functions over R. In our presentation each such function is viewed as a subset of R.

We use properties of both the Kneser graph and its induced subgraph, the Schrijver graph.

Combining the Two Components. The constructed hypergraph will have a block of vertices per variable of the layered PCP system. Each block is conceptually a copy of the Kneser graph (but we will have hyperedges and not edges).

The heart of the proof lies in translating a legal coloring (with few colors) of the hypergraph into a satisfying assignment for the PCP system. This is done by first translating the coloring within each block, into a short list of elements of R that are supposedly encoded by that coloring. The second and more difficult part is to prove that for distinct blocks, the short lists of decoded values are in fact consistent with each other. This difficulty is resolved using the maximization variant of the Kneser conjecture mentioned above.

2. The Kneser Graph

In this section we define the Kneser graph and describe some of its important properties. For $n \geq 2s+1$, the Kneser graph $KG_{n,s}$ has the set $\binom{[n]}{s} \stackrel{def}{=} \{S \subseteq [n] \mid |S| = s\}$ as its vertex set and two vertices S_1, S_2 are adjacent iff $S_1 \cap S_2 = \phi$. In other words, each vertex corresponds to an s-set and two vertices are adjacent if their corresponding sets are disjoint. In this paper we are mainly interested in the case where s is smaller than $n/2$ by a constant. These graphs have the important property that the chromatic number is high although large independent sets exist. For a discussion of Kneser graphs and other combinatorial problems, see the excellent book by Matoušek [23].

There exists a simple way to color this graph with $n-2s+2$ coloring. In 1955, Kneser [19] conjectured that there is no way to color the graph with less colors, i.e., $\chi(KG_{n,s}) = n-2s+2$. The first to prove this conjecture was Lovász in 1978 [22]. Many other proofs and extensions are known (see, e.g., [8,3,24]) and the latest and simplest one is by Greene [12]. Our goal in this section is to prove the following property of the Kneser graph: in any coloring of the vertices of the Kneser graph with less colors than its chromatic number, there must exist many monochromatic edges. The way we prove this is by considering certain induced subgraphs of the Kneser graph known as Schrijver graphs. The proof follows from the fact that these induced subgraphs have the same chromatic number as the whole Kneser graph.

Let us now define the Schrijver graph $SG_{n,s,\pi}$. Given a permutation π of $[n]$, we say that an s-subset $S \in \binom{[n]}{s}$ is *π-stable* if it does not contain two π-adjacent elements modulo n, i.e., if $\pi(i) \in S$ then $\pi(i+1) \notin S$ and if $\pi(n) \in S$ then $\pi(1) \notin S$. We denote the number of stable s-sets by $\binom{n}{s}_{stab}$ (notice that it is independent of π). The graph $SG_{n,s,\pi}$ contains a vertex for each π-stable s-subset of $[n]$ and two vertices are adjacent if their corresponding sets are disjoint. Clearly, $SG_{n,s,\pi}$ is an induced subgraph of $KG_{n,s}$. Interestingly, the chromatic number of the Schrijver graph is the same, i.e., $\chi(SG_{n,s,\pi}) = n-2s+2$ [26].

We begin with a simple bound on the number of vertices in a Schrijver graph:

Claim 2.1. $\binom{n}{s}_{stab} \le \binom{n}{n-2s}$

Proof. Consider a stable set S according to the permutation $\pi(i)=i$. Define T as the set of all $i \in [n]$ such that $i \notin S$ and $i-1 \notin S$ (and 1 is included in T if $1 \notin S$ and $n \notin S$). Notice that T uniquely defines S. The claim follows from the fact that T is a set of $n-2s$ elements. ∎

The following is the main lemma of this section:

Lemma 2.2. *In any coloring of $KG_{n,s}$ by $n-2s+1$ colors there exists a monochromatic edge whose color is used in at least a $\frac{2}{(n-2s+1)\binom{n}{s}_{stab}} \ge \frac{2}{(n-2s+1)\binom{n}{n-2s}}$ fraction of the vertices.*

Proof. Fix a coloring of $KG_{n,s}$ by $n-2s+1$ colors. For every permutation π, the induced subgraph $SG_{n,s,\pi}$ contains a monochromatic edge since $\chi(SG_{n,s,\pi}) = n-2s+2$. Therefore, there exists a color, say red, such that at least $\frac{1}{n-2s+1}$ of the Schrijver graphs contain a red monochromatic edge. Consider the following distribution on the vertices of $KG_{n,s}$: choose a random permutation π and then a random vertex in $SG_{n,s,\pi}$. The probability of choosing a red vertex is at least $\frac{1}{n-2s+1} \cdot \frac{2}{\binom{n}{s}_{stab}}$ since we first have to choose a Schrijver graph that contains a red monochromatic edge and then one of the two end points of the monochromatic edge. Since each $SG_{n,s,\pi}$ contains the same number of vertices and each vertex of $KG_{n,s}$ is contained in the same number of $SG_{n,s,\pi}$, this distribution is equivalent to the uniform distribution on the vertices of $KG_{n,s}$. Therefore, the fraction of red vertices is at least $\frac{2}{(n-2s+1)\binom{n}{s}_{stab}}$. ∎

3. The Layered PCP

We use the layered PCP construction of Dinur et al. [6]. In an l-layered PCP there are l sets of variables, $X_1,\ldots,X_l$. We refer to X_i as the ith layer. The range of variables in X_i is denoted R_i. For every $1 \le i < j \le l$ there is a set of tests Φ_{ij} where each test $\varphi \in \Phi_{ij}$ depends on exactly one $x \in X_i$ and one $y \in X_j$. For any two variables we denote by $\varphi_{x\to y}$ the test between them if such a test exists. Moreover, the tests in Φ_{ij} are projections from x to y, that is, for every assignment to x there is exactly one assignment to y for which the test accepts.

Theorem 3.1 (Theorem 3.3 in [6]). *For any parameters l,u there exists a reduction from an NP-hard problem of size n to the problem of distinguishing between the following two cases in an l-layered PCP Φ with $n^{O(ul)}$ variables over a range of size $2^{O(ul)}$. Either there exists an assignment that satisfies all the tests or, for every $i<j$, not more than $2^{-\Omega(u)}$ of the tests in Φ_{ij} can be satisfied by an assignment. Moreover, for any $1<m<l$ and for any m layers $i_1<\ldots<i_m$ and sets $S_j \subseteq X_{i_j}$ for $j\in[m]$ such that $S_j \ge \frac{2}{m}|X_{i_j}|$ there exist two sets S_j and $S_{j'}$ such that the number of tests between them is at least $\frac{1}{m^2}$ of the number of tests between the layers X_{i_j} and $X_{i_{j'}}$.*

A sketch of the proof is included in the appendix.

4. The Hypergraph Construction

Theorem 4.1 (Main Theorem). *For any integer $\chi \ge 2$, it is NP-hard to color a 2-colorable 3-uniform hypergraph using χ colors.*

Proof. Let Φ be a PCP instance with layers $X_1,\ldots,X_l$, as described in Section 3, with parameters $l=2\chi^2$, and u to be chosen later. We present a construction of a 3-uniform hypergraph $G=(V,E)$.

Vertices. For each variable x in layer X_i we construct a block of vertices $V[x]$. This block contains a vertex for each subset of R_i of size $\lfloor(|R_i|-\chi)/2\rfloor$, i.e.,

$$V[x] = \binom{R_i}{\lfloor(|R_i|-\chi)/2\rfloor}.$$

Altogether,

$$V = \bigcup_{x\in\cup X_i} V[x]\,.$$

Throughout this section we slightly abuse notation by writing a vertex rather than the subset it represents.

Hyperedges. Before explicitly specifying the hyperedges, let us explain the idea in constructing them. This is closely related to the (simple) proof of completeness (Lemma 4.2). Consider the following *natural block-coloring* of a block $V[x]$ according to an assignment a for x. Simply color all the vertices that contain a red, and all the rest orange. For two variables x, y sharing a test $\varphi_{x\to y}$, we add all possible hyperedges on the vertices of $V[x] \cup V[y]$ such that all natural block-colorings corresponding to an assignment a for x and b for y satisfying the test, remain legal colorings. In fact, we take only a proper subset of these hyperedges as they already suffice for our soundness argument.

More precisely, we construct hyperedges between blocks $V[x]$ and $V[y]$ only if there exists a test $\varphi_{x\to y} \in \Phi$. We put a hyperedge between any $v_1, v_2 \in V[x]$ and $u \in V[y]$ whenever $v_1 \cap v_2 = \phi$ and $\varphi_{x\to y}(R_i \setminus (v_1 \cup v_2)) \subseteq u$. In summation,

$$E = \bigcup_{\varphi_{x\to y}\in\Phi} \Big\{ \{v_1, v_2, u\} \;\Big|\; v_1, v_2 \in V[x], \quad u \in V[y], \\ v_1 \cap v_2 = \phi \;\text{ and }\; \varphi_{x\to y}(R_i \setminus (v_1 \cup v_2)) \subseteq u \Big\}.$$

The best way to get more intuition on this definition is to consider the case where every block is colored by a natural block-coloring, and to notice that this is a legal coloring of the hypergraph iff the underlying assignment satisfies all of the tests. See the proof of completeness, shortly below.

Note that our hyperedges are always between two layers X_i and X_j of the PCP. Moreover, they are 'directed' in the sense that if $i < j$ then two vertices are chosen from a block in X_i and one vertex is chosen from a block in X_j.

Remark. Interestingly, it is easy to see that this construction has rather large independent sets (consisting of almost half the number of the vertices) no matter what the underlying PCP is. For example, we can choose from each block all the vertices that contain a certain assignment, say the first one. Moreover, it is possible to construct two independent sets (i.e., two colors) that cover almost all of the hypergraph, leaving out only a sub-constant part. We briefly sketch this example, which originally appeared in [10]. Consider any $x \in X_i$, $i \in [l]$ and let T_i be any subset of R_i of size $\frac{|R_i|}{2}$. The first independent set contains all the vertices in $V[x]$ whose intersection with T_i is of size more than $\frac{|T_i|}{2}$. Similarly, the second independent set contains all the vertices in $V[x]$ whose intersection with $R_i \setminus T_i$ is more than $\frac{|T_i|}{2}$. These are indeed independent sets because two vertices in the first independent set intersect on T_i and similarly for the second independent set. Also, most of the vertices in $V[x]$ are in one of the independent sets. Informally, this is true because the size of the intersection of a vertex in $V[x]$ and T_i has a standard deviation of $\Theta(\sqrt{|R_i|})$.

Lemma 4.2 **(Completeness).** *If Φ is satisfiable then G is 2-colorable.*

Proof. Let A be a satisfying assignment for Φ, i.e., A maps each variable $x\in X_i$ to an assignment in R_i such that all the tests are satisfied. In the block $V[x]$ we color all the vertices that contain the assignment $A(x)$ red and all the rest orange (i.e., the natural block-coloring corresponding to $A(x)$). There are no red monochromatic edges because two red vertices inside the same block have a non-empty intersection. Now we show that there are no orange monochromatic edges. Let $\{v_1,v_2,u\}$ be an arbitrary hyperedge and let x,y,i be such that $v_1,v_2\in V[x]$, $x\in X_i$ and $u\in V[y]$. Assume that both v_1 and v_2 are orange. Since there exists a hyperedge between them, v_1 and v_2 are disjoint. Therefore, $A(x)\in R_i\setminus(v_1\cup v_2)$ which implies that $\varphi_{x\to y}(A(x))\in\varphi_{x\to y}(R_i\setminus(v_1\cup v_2))\subseteq u$ where the last containment is again because $\{v_1,v_2,u\}$ is a hyperedge. But $\varphi_{x\to y}(A(x))=A(y)$ since A is a satisfying assignment and therefore $A(y)\in u$ and u is red. ∎

Lemma 4.3 **(Soundness).** *If G is χ-colorable then Φ is satisfiable.*

Proof. Fix a coloring of the graph G. The first step in our proof is to 'list-decode' this coloring, i.e., to find a list of candidate-assignments for each variable. For any variable $x\in X_i$, $i\in[l]$, consider the vertices inside the block $V[x]$. We can map them to the vertices of the Kneser graph $KG_{|R_i|,\lfloor(|R_i|-\chi)/2\rfloor}$. According to Lemma 2.2 we can find a color c_x with the following two properties:

1. The subset $U[x]\subseteq V[x]$ of vertices colored c_x contains at least an $\Omega(\frac{1}{\chi}|R_i|^{-(\chi+1)})$ fraction of the vertices in $V[x]$.
2. There are two vertices $v_{x,1},v_{x,2}\in U[x]$ such that $v_{x,1}\cap v_{x,2}=\phi$.

Let $B(x)$ be the set $R_i\setminus(v_{x,1}\cup v_{x,2})$. This is the list of 'decoded' values. Notice that it contains at most $\chi+1$ assignments. We say that the variable x is *colored with* c_x, and denote as above by $U[x]\subseteq V[x]$ the set of vertices $V[x]$ colored with c_x.

The next step in the proof is to establish consistency. Define the color of a layer $i\in[l]$ to be the color in which the largest number of its variables are colored. Notice that at least $\frac{1}{\chi}$ of the variables of the layer must be colored with the color of the layer. Finally, we can find $\frac{l}{\chi}=2\chi$ layers that are colored with the same color, say red. Using the properties of the PCP with the set of red layers and the red variables within each layer, we conclude that there exist two layers X_i and X_j such that $\frac{1}{4\chi^2}$ of the tests between them are tests between red variables. Let us denote by X the red variables in X_i and by Y the red variables in X_j.

We can now define an assignment to the variables in X and Y such that many of the tests between them are satisfied. For a variable $x \in X$ we choose a random assignment from the set $B(x)$. For a variable $y \in Y$ we choose the assignment that agrees with the maximal number of $B(x)$'s,

$$A(y) = \text{maxvar}_{a \in R_Y} |\{x \in X \mid \varphi_{x\to y} \in \Phi \text{ and } a \in \varphi_{x\to y}(B(x))\}| \,.$$

The heart of the proof lies in the following claim.

Claim 4.4. *Fix $y \in Y$. Consider the collection*[2] *of subsets of R_j*

$$\{\varphi_{x\to y}(B(x)) \mid x \in X \text{ shares a test with } y\}\ . \tag{1}$$

Then there is some $a \in R_j$ contained in at least $\alpha = \Omega(\chi^{-2} \cdot 3^{-\chi} \cdot (ul)^{-1})$ of these sets.

Let us first see that this claim completes the proof. Indeed, it follows that the expected fraction of X, Y tests satisfied by A is at least $\alpha/(\chi+1)$ because with probability $1/|B(x)| \geq (\chi+1)^{-1}$ the element of $B(x)$ that is assigned to x is consistent with the assignment for y. Recalling that tests between X and Y represent $\Omega(\chi^{-2})$ of the tests between layers X_i and X_j, we get that $\Omega(\chi^{-3}\alpha) = \Omega(\chi^{-5}3^{-\chi}(ul)^{-1})$ of the tests between layers X_i and X_j are satisfied. Choosing u to be $c \cdot \chi$ with a big enough constant c, this is more than $2^{-\Omega(u)}$ and hence Φ is satisfiable.

Proof of Claim 4.4. We first prove in Claim 4.5 that the family defined in (1) cannot contain too many pairwise disjoint sets. We then rely on a simple combinatorial claim (Claim 4.6) to show that there must be one element present in many of the sets.

Why aren't there many pairwise disjoint sets in the collection (1)? If there were too many such sets, this would prevent $U[y]$ from being large. Indeed, consider a variable x such that the test $\varphi_{x\to y}$ exists. Since the vertices $v_{x,1}, v_{x,2}$ and the vertices in $U[y]$ are colored red, there are no hyperedges between them. Therefore, by definition of the hyperedges, all the vertices $u \in U[y]$ must not contain $\varphi_{x\to y}(B(x))$. This poses a restriction on the size of $U[y]$. We make this formal in the following claim. The reader might want to skip its proof at first read.

[2] One should think of this collection as a multiset, i.e., we allow the same subset to appear more than once.

Claim 4.5. *Let $A_1,\dots,A_q \subseteq [n]$ be pairwise disjoint, $|A_i| \le s$. Let $\mathcal{F} = \{F \in \binom{[n]}{k} \mid \forall i\, F \not\supseteq A_i\}$. Then if $\frac{n}{3} \le k \le \frac{2n}{3}$ we have*

$$q \le 3^s \left[\frac{1}{2}\log n + c_0 - \log\left(|\mathcal{F}| / \binom{n}{k}\right)\right]$$

where c_0 is an absolute constant.

Proof. Consider the probability distribution μ on $2^{[n]}$ where each element $i \in [n]$ is chosen to be in the set with probability $\frac{k}{n}$ and out of it with probability $1 - \frac{k}{n}$. Thus, $\mu(F) = (\frac{k}{n})^{|F|}(1-\frac{k}{n})^{n-|F|}$ and for $\mathcal{F} \subseteq 2^{[n]}$ we define $\mu(\mathcal{F}) = \sum_{F\in\mathcal{F}} \mu(F)$.

The probability that a fixed A_i is not completely contained in a μ-random subset is at most $1-(k/n)^s$. Since the A_i's are pairwise disjoint,

$$\mu(\mathcal{F}) \le \left(1 - \left(\frac{k}{n}\right)^s\right)^q \le \exp\left(-q\left(\frac{k}{n}\right)^s\right).$$

Since all k-sets appear with equal probability, we have

$$|\mathcal{F}| \Big/ \binom{n}{k} = \mu\left(\mathcal{F}\right) \Big/ \mu\left(\binom{[n]}{k}\right).$$

From Stirling's formula we get $n! = \Theta(n^n e^{-n}\sqrt{n})$ and hence

$$\mu\left(\binom{[n]}{k}\right) = \binom{n}{k}\left(\frac{k}{n}\right)^k\left(1-\frac{k}{n}\right)^{n-k} = \Theta\left(\sqrt{\frac{n}{k(n-k)}}\right)$$

which for $\frac{n}{3} \le k \le \frac{2n}{3}$ equals $\Theta(1/\sqrt{n})$. Combining the above equations means that

$$|\mathcal{F}| / \binom{n}{k} = O\left(\sqrt{n}\exp\left(-q\left(1-\frac{k}{n}\right)^s\right)\right).$$

Taking logarithms we get

$$q\left(1-\frac{k}{n}\right)^s \le \frac{1}{2}\log n + c_0 - \log\left(|\mathcal{F}| \Big/ \binom{n}{k}\right)$$

for some constant $c_0 > 0$. Since $\frac{n}{3} \le k \le \frac{2n}{3}$ this implies the conclusion of the claim. ∎

We use the above claim with $A_1,\dots,A_q$ a maximum collection of pairwise disjoint sets in Equation (1), $n = |R_j|, k = \lfloor(|R_j| - \chi)/2\rfloor$, and $s = \chi + 1$.

Notice that in this case $U[y] \subseteq \mathcal{F}$ and recall that $U[y]$ contains at least an $\Omega(\frac{1}{\chi}|R_j|^{-(\chi+1)})$ fraction of the vertices in $V[y]$. Hence,

$$-\log\left(|\mathcal{F}|/\binom{|R_j|}{\lfloor(|R_j|-\chi)/2\rfloor}\right) \leq -\log\left(\frac{|U[y]|}{|V[y]|}\right)$$
$$\leq (\chi+1)\log|R_j| + \log\chi + O(1) = O(\chi\log|R_j|).$$

Therefore, using the claim,

$$q \leq 3^{\chi+1}\Big(\frac{1}{2}\log|R_j| + c_0 + O(\chi\log|R_j|)\Big) = O(\chi\cdot 3^{\chi}\cdot\log|R_j|) = O(\chi\cdot 3^{\chi}\cdot ul)\,.$$

Having bounded q, we complete the proof with the following simple claim:

Claim 4.6. *Let $A_1,\ldots,A_n$ be a collection of n sets of size at most s such that there are at most q pairwise disjoint sets in the collection. Then, there must be an element contained in at least n/qs sets in this collection.*

Proof. Let $A_{i_1},\ldots,A_{i_{q'}}$ be any maximal sub-collection of pairwise disjoint sets. By assumption, $q' \leq q$. Each of the remaining $n-q'$ subsets must intersect $\cup_{m=1}^{q'} A_{i_m}$. Since $|\cup A_{i_m}| \leq qs$, there must be an element contained in $(n-q')/(qs)+1 \geq n/qs$ subsets. ∎

Claim 4.6 with $A_1,\ldots,A_n$ the sets in (1) and $s=\chi+1$, implies that there exists an assignment for y that is contained in at least a fraction $\alpha = q^{-1}(\chi+1)^{-1} = \Omega(\chi^{-1}\cdot 3^{-\chi}\cdot(ul)^{-1}\cdot(\chi+1)^{-1})$ of the sets in (1). ∎

This completes the proof of Lemma 4.3. ∎

It is worthwhile to mention that we rely on the quantitative aspects of the parallel repetition theorem [25]. In particular, it is essential for our proof that after u repetitions the soundness error becomes smaller than $1/u^{O(1)}$ (of course Raz proved that it decreases exponentially in u which is all the better). This is in contrast to the usual scenario where it only matters that the soundness error goes to zero as u goes to infinity. ∎

Theorem 4.7. *Assuming* NP $\not\subseteq$ DTIME$(2^{\mathrm{poly}(\log n)})$, *there is no polynomial time algorithm that colors a 2-colorable 3-uniform hypergraph using $O(\sqrt[3]{\log\log N})$ colors where N is the number of vertices in the hypergraph.*

Proof. We note that in the previous proof, we can take $\chi = c\sqrt[3]{\log\log n}$ where $c>0$ is any constant and n is the size of the NP-hard problem instance from which the reduction begins. The parameters we chose are $l=O(\chi^2)$ and $u=O(\chi)$. Therefore, the size of the hypergraph we construct is $N = n^{O(ul)}2^{2^{O(ul)}} = 2^{\mathrm{poly}(\log n)}$. The proof is completed by noting that $\log\log N = \log(\mathrm{poly}(\log n)) = O(\log\log n)$. ∎

5. Discussion

The construction we presented relies on the properties of the Kneser graph in a strong way. This allowed us to prove the hardness of coloring in a direct way and not via the size of the maximal independent set, as was done in essentially all previous results. We believe that the Kneser graph might be useful in understanding the hardness of approximate coloring for graphs, a problem that is notoriously difficult.

6. Acknowledgements

We would like to thank Benny Sudakov for introducing us to the Kneser graph, and to both Benny Sudakov and Noga Alon, for insightful discussions. We also thank the anonymous referees for many helpful comments.

References

[1] S. Arora, C. Lund, R. Motwani, M. Sudan and M. Szegedy: Proof verification and the hardness of approximation problems, *J. ACM* **45(3)** (1998), 501–555.

[2] S. Arora and S. Safra: Probabilistic checking of proofs: a new characterization of NP, *J. ACM* **45(1)** (1998), 70–122.

[3] I. Bárány: A short proof of Kneser's conjecture, *J. Combin. Theory Ser. A* **25(3)** (1978), 325–326.

[4] M. Bellare, O. Goldreich and M. Sudan: Free bits, PCPs, and nonapproximability – towards tight results; *SIAM J. Comput.* **27(3)** (1998), 804–915.

[5] A. Blum and D. Karger: An $\tilde{O}(n^{3/14})$-coloring algorithm for 3-colorable graphs, *Inform. Process. Lett.* **61(1)** (1997), 49–53.

[6] I. Dinur, V. Guruswami, S. Khot and O. Regev: A new multilayered PCP and the hardness of hypergraph vertex cover, in *Proc. 35th ACM Symp. on Theory of Computing (STOC)*, pages 595–601, 2003.

[7] I. Dinur and S. Safra: The importance of being biased, in *Proc. 34th ACM Symp. on Theory of Computing (STOC)*, pages 33–42, 2002.

[8] V. L. Dol'nikov: Transversals of families of sets, in *Studies in the theory of functions of several real variables (Russian)*, pages 30–36, 109. Yaroslav. Gos. Univ., Yaroslavl', 1981.

[9] U. Feige and J. Kilian: Zero knowledge and the chromatic number, *J. Comput. System Sci.* **57(2)** (1998), 187–199.

[10] P. Frankl and Z. Füredi: Extremal problems concerning Kneser graphs, *J. Combin. Theory Ser. B* **40(3)** (1986), 270–284.

[11] M. X. Goemans and D. P. Williamson: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming, *J. Assoc. Comput. Mach.* **42(6)** (1995), 1115–1145.

[12] J. E. Greene: A new short proof of Kneser's conjecture, *Amer. Math. Monthly* **109(10)** (2002), 918–920.

[13] V. Guruswami: Inapproximability results for set splitting and satisfiability problems with no mixed clauses, in *Approximation algorithms for combinatorial optimization (Saarbrücken, 2000)*, **volume 1913** of *Lecture Notes in Comput. Sci.*, pages 155–166, Springer, Berlin, 2000.
[14] V. Guruswami, J. Håstad and M. Sudan: Hardness of approximate hypergraph coloring, *SIAM J. Comput.* **31(6)** (2002), 1663–1686.
[15] V. Guruswami and S. Khanna: On the hardness of 4-coloring a 3-colorable graph, in *15th Annual IEEE Conference on Computational Complexity (Florence, 2000)*, pages 188–197, IEEE Computer Soc., Los Alamitos, CA, 2000.
[16] J. Håstad: Some optimal inapproximability results, *J. ACM* **48(4)** (2001), 798–859.
[17] S. Khanna, N. Linial and S. Safra: On the hardness of approximating the chromatic number, *Combinatorica* **20(3)** (2000), 393–415.
[18] S. Khot: Hardness of coloring 3-colorable 3-uniform hypergraphs, in *Proc. 43rd Annual IEEE Symp. on Foundations of Computer Science (FOCS)*, pages 23–32, 2002.
[19] M. Kneser: Aufgabe 360, *Jahresbericht der Deutschen Mathematiker-Vereinigung* **58(2)** (1955), Abteilung, S. 27.
[20] M. Krivelevich, R. Nathaniel and B. Sudakov: Approximating coloring and maximum independent sets in 3-uniform hypergraphs, *J. Algorithms* **41(1)** (2001), 99–113.
[21] M. Krivelevich and B. Sudakov: Approximate coloring of uniform hypergraphs, *J. Algorithms* **49(1)** (2003), 2–12.
[22] L. Lovász: Kneser's conjecture, chromatic number, and homotopy; *J. Combin. Theory Ser. A* **25(3)** (1978), 319–324.
[23] J. Matoušek: *Using the Borsuk–Ulam theorem*, Universitext, Springer-Verlag, Berlin, 2003. Lectures on topological methods in combinatorics and geometry, written in cooperation with Anders Björner and Günter M. Ziegler.
[24] J. Matoušek: A combinatorial proof of Kneser's conjecture, *Combinatorica* **24(1)** (2004), 163–170.
[25] R. Raz: A parallel repetition theorem, *SIAM J. Comput.* **27(3)** (1998), 763–803.
[26] A. Schrijver: Vertex-critical subgraphs of Kneser graphs, *Nieuw Arch. Wisk. (3)* **26(3)** (1978), 454–461.
[27] U. Zwick: Outward rotations: a tool for rounding solutions of semidefinite programming relaxations, with applications to MAX CUT and other problems; in *Annual ACM Symposium on Theory of Computing (Atlanta, GA, 1999)*, pages 679–687, ACM, New York, 1999.

Appendix

We include a proof of Theorem 3.1. This theorem was proven in Dinur et al. [6], and we repeat the proof for self-containment only.

Theorem 3.1 (Theorem 3.3 in [6]). *For any parameters l, u there exists a reduction from an NP-hard problem of size n to the problem of distinguishing between the following two cases in an l-layered PCP Φ with $n^{O(ul)}$ variables over a range of size $2^{O(ul)}$. Either there exists an assignment that satisfies all the tests or, for every $i<j$, not more than $2^{-\Omega(u)}$ of the tests in Φ_{ij} can*

be satisfied by an assignment. Moreover, for any $1<m<l$ and for any m layers $i_1<\ldots<i_m$ and sets $S_j \subseteq X_{i_j}$ for $j\in[m]$ such that $S_j \geq \frac{2}{m}|X_{i_j}|$ there exist two sets S_j and $S_{j'}$ such that the number of tests between them is at least $\frac{1}{m^2}$ of the number of tests between the layers X_{i_j} and $X_{i_{j'}}$.

Proof Sketch. We begin with the parallel repetition lemma [25] applied with parameter u to the 3SAT5 problem of [1]. This provides us with a reduction from an NP-hard problem of size n to a PCP Ψ over two sets of variables Y (with range R_Y of size 7^u) and Z (with range R_Z of size 2^u) where the number of variables is $n^{O(u)}$. The tests $\psi_{y,z}$ are only between a variable $y\in Y$ and a variable $z\in Z$ and are projections from R_Y to R_Z, i.e., for any given assignment to $y\in Y$ there exists exactly one consistent assignment to $z\in Z$. The problem is to decide whether there exists an assignment that satisfies all the tests or no assignment satisfies more than $2^{-\Omega(u)}$ fraction of the tests. One property of the resulting PCP that we use is its uniformity: the distribution created by uniformly taking a variable $y\in Y$ and then uniformly choosing one of the variables in $z\in Z$ with which it has a test is a uniform distribution on Z. This property follows from the fact that 3SAT5 instances are SAT instances in which every clause contains exactly 3 variables and every variables appears in exactly 5 clauses.

We construct Φ as follows. The variables X_i of layer $i\in[l]$ are the elements of the set Z^iY^{l-i}, i.e., all l-tuples where the first i elements are Z variables and the last $l-i$ elements are Y variables. The variables in layer i have assignments from the set $R_i=R_Z^iR_Y^{l-i}$ corresponding to an assignment to each variable of Ψ in the l-tuple. It is easy to see that $|R_i|\leq 2^{O(ul)}$ for any $i\in[l]$ and that the total number of variables is $n^{O(ul)}$. For any $1\leq i<j\leq l$ we define the tests in Φ_{ij} as follows. A test exists between a variable $x_i\in X_i$ and a variable $x_j\in X_j$ if they contain the same Ψ variables in the first i and the last $l-j$ elements of their l-tuples. Moreover, for any $i<k\leq j$ there should be a test in Ψ between $x_{i,k}$ and $x_{j,k}$. More formally,

$$\begin{aligned}\Phi_{ij}=\{\varphi_{x_i,x_j} \mid x_i\in X_i,\ \ x_j\in X_j,\ \ & \forall k\in[l]\setminus\{i+1,\ldots,j\}\ x_{i,k}=x_{j,k},\\ & \forall k\in\{i+1,\ldots,j\}\ \psi_{x_{i,k},x_{j,k}}\in\Psi\}.\end{aligned}$$

As promised, the tests φ_{x_i,x_j} are projections. Given as assignment a to x_i, we define the consistent assignment b to x_j as $b_k=\psi_{x_{i,k},x_{j,k}}(a_k)$ for $k\in\{i+1,\ldots,j\}$ and $b_k=a_k$ otherwise.

The completeness of Φ follows easily from the completeness of Ψ. That is, assume we are given an assignment $A:Y\cup Z\to R_Y\cup R_Z$ that satisfies all the tests. Then, the assignment $B:\bigcup X_i\to\bigcup R_i$ defined by $B(x_1,\ldots,x_l)=(A(x_1),\ldots,A(x_l))$ is a satisfying assignment. For the soundness part, assume that there exist two layers $i<j$ and an assignment B that satisfies more than $2^{-\Omega(u)}$ of the tests in Φ_{ij}. We partition X_i into classes such that two variables in X_i are in the same class iff they are identical except possibly on coordinate j. The variables in X_j are also partitioned according to coordinate j. Since more than $2^{-\Omega(u)}$ of the tests in Φ_{ij} are satisfied, it must be the case that there exist a class $x_{i,1},\ldots,x_{i,j-1},x_{i,j+1},\ldots,x_{i,l}$ in the partition of X_i and a class $x_{j,1},\ldots,x_{j,j-1},x_{j,j+1},\ldots,x_{j,l}$ in the partition of X_j between which there exist tests and the fraction of satisfied tests is more than $2^{-\Omega(u)}$. We define an assignment to Ψ as $A(y)=(B(x_{i,1},\ldots,x_{i,j-1},y,x_{i,j+1},\ldots,x_{i,l}))_j$ for $y\in Y$ and as $A(z)=(B(x_{j,1},\ldots,x_{j,j-1},z,x_{j,j+1},\ldots,x_{j,l}))_j$ for $z\in Z$. Notice that there is a one-to-one and onto correspondence between the tests in Ψ and the tests between the two chosen classes in Φ. Moreover, if the test is Φ is satisfied, then the test in Ψ is also satisfied. Therefore, A is an assignment to Ψ that satisfies more than $2^{-\Omega(u)}$ of the tests.

To prove the second part of the theorem let $1<m<l$, and set $\delta=2/m$. Take any m layers $i_1<\ldots<i_m$ and sets $S_j\subseteq X_{i_j}$ for $j\in[m]$ such that $S_j\geq\delta|X_{i_j}|$. Consider a random walk beginning from a uniformly chosen variable $x_1\in X_1$ and proceeding to a variable $x_2\in X_2$ chosen uniformly among the variables with which x_1 has a test. The random walk continues in a similar way to a variable $x_3\in X_3$ chosen uniformly among the variables with which x_2 has a test and so on up to a variable in X_l. Denote by E_j the indicator variable of the event that the random walk hits a variable in S_j when in layer X_{i_j}. From the uniformity of Ψ it follows that for every j, $P(E_j)\geq\delta=\frac{2}{m}$. Moreover, by using the inclusion-exclusion principle, we get:

$$1\geq P(\bigvee E_j)\geq\sum_j P(E_j)-\sum_{j<k}P(E_j\wedge E_k)\geq 2-\binom{m}{2}\max_{j<k}P(E_j\wedge E_k)$$

which implies

$$\max_{j<k}P(E_j\wedge E_k)\geq 1/\binom{m}{2}>\frac{1}{m^2}.$$

Fix j and k such that $P(E_j\wedge E_k)\geq\frac{1}{m^2}$ and consider a shorter random walk beginning from a random variable in X_{i_j} and proceeding to the next layer and so on until hitting layer i_k. Since E_j is uniform on X_{i_j} we still have that $P(E_j\wedge E_k)\geq\frac{1}{m^2}$ where the probability is taken over the random walks between X_{i_j} and X_{i_k}. Also, notice that there is a one-to-one and onto mapping from the set of all random walks between X_{i_j} and X_{i_k} to the set Φ_{i_j,i_k}. Therefore, $\frac{1}{m^2}$ of the tests between X_{i_j} and X_{i_k} are between S_j and S_k. ∎

Irit Dinur

School of Computer Science and Engineering
The Hebrew University of Jerusalem
Jerusalem 91904
Israel
dinuri@cs.huji.ac.il

Oded Regev

Department of Computer Science
Tel-Aviv University
Tel-Aviv 69978
Israel

Clifford Smyth

Department of Mathematics
Carnegie Mellon University
Pittsburgh, PA 15213-3890
USA
csmyth@andrew.cmu.edu