Recognizing more random unsatisfiable 3-SAT instances efficiently

Andreas Goerdt André Lanka

Technische Universität Chemnitz, Fakultät für Informatik
Straße der Nationen 62, 09107 Chemnitz, Germany
{goerdt, lanka}@informatik.tu-chemnitz.de

Abstract

We show that random 3-SAT formulas with $\mathrm{poly}(\log n) \cdot n^{3/2} = n^{3/2+o(1)}$ clauses can be efficiently certified as unsatisfiable. This improves a previous bound of $n^{3/2+\varepsilon}$ clauses. There $\varepsilon > 0$ is a constant.

1 Introduction

The efficient certification of the unsatisfiability of random k-SAT instances beyond the satisfiability threshold is a topic of some recent interest, see [GoKr 01], [FrGo 01], [Fe 2002], [FeOf 03] for example. We refer to these papers for a discussion of the nature of efficient certification and its motivation.

From [FrGo 01] it is known that random 3-SAT instances with $n^{3/2+\varepsilon}$ clauses can be efficiently certified as unsatisfiable. If k is even, it is known from [Coja et al] that k-SAT instances with $C \cdot n^{k/2}$ clauses, k even, can be efficiently certified as unsatisfiable. [FeOf 03] gives another proof of this result, but does not treat the case of odd k. At this point the natural

conjecture is that 3-SAT instances with $C \cdot n^{3/2}$ clauses can be efficiently certified as unsatisfiable. The present paper is motivated by this conjecture and obtains the result mentioned above.

Some remark on the approach used seem appropriate. All certification algorithms presented by now rely on the observation that the satisfiability of a propositional formula F implies the existence of a linear size independent set in a graph associated to F. For F random with sufficiently many clauses the graph is sufficiently random and dense such that it has no independent set as required. The absence of a large independent set in random graphs G can be efficiently shown by looking at the Eigenvalues of matrices associated to G or computing the Lovasz number of G, as in [Coja et al].

For even k a k-SAT instance F induces such a graph with $n^{k/2}$ vertices. Moreover, if F is random G is a classical random graph $G_{v,p}$, $v = n^{k/2}$. On such graphs Eigenvalues can be used as long as we have $C \cdot n^{k/2}$ edges, C sufficiently large but constant. This is the reason for the $C \cdot n^{k/2}$ clauses required as each clause induces one edge. The method to calculate the Eigenvalues in this case goes back to [FeKaSz 89] and it is used in [FeOf 03]. The method requires the property that the edges are well distributed over the graph. This property is easily shown by counting arguments for classical random graphs.

In case of a 3-SAT formula F we also get a graph from F which has n^2 vertices. But now the graph has some dependencies between its edges. These dependencies make it difficult to show by a simple counting argument that the edges are well distributed over the graph. Therefore the method of [FeKaSz 89] is not easily applicable.

In case of $n^{3/2+\varepsilon}$ many random 3-clauses the trace method, a method different from [FeKaSz 89] is used in [FrGo 01] to determine the Eigenvalues. It is not obvious if a direct extension of the method from [FrGo 01] yields the bound of $\text{poly}(\log n) \cdot n^{3/2}$ many clauses. In fact in [FrGo 01] the trace method is used with closed walks of bounded length. If $\varepsilon = o(1)$ we need walks of unbounded length we consider. In [FuKo 80] the trace method for walks of unbounded length is considered. Our technique is in some respects inspired by this paper. The dependencies between edges, characteristic for our situation, is an aspect not occurring in [FuKo 80] and thus may be of independent interest.

2 Efficient certification of unsatisfiability

Given a set of n propositional variables $\mathrm{Var}_n = \{v_1, \ldots, v_n\}$ a positive literal over Var_n simply is a variable v_i and a negative literal is a negated variable $\neg v_i$. Moreover, we let $\neg\neg v_i = v_i$. A 3-clause or simply a clause is an ordered 3-tuple $l_1 \vee l_2 \vee l_3$ of (positive or negative) literals. Thus altogether we have $(2n)^3$ clauses. A 3-SAT instance or 3-SAT formula is a set of 3-clauses. We think of a 3-SAT instance as $C_1 \wedge \ldots \wedge C_m$ where each C_i is a 3-clause. Given a truth assignment α of Var_n, that is a mapping assigning $true\,(= 1)$ or $false\,(= 0)$ to each variable, we can assign $true$ or $false$ to a 3-SAT formula as usual. A 3-SAT instance is satisfiable if there exists a truth value assignment α such that this instance evaluates to $true$ under α. Otherwise the instance is unsatisfiable.

The probability space $Form_{n,3,p} = Form_{n,p}$ is the probability space of 3-SAT formulas obtained by picking each of the $(2n)^3$ 3-clauses with probability p independently. There are slightly different ways to define probability spaces of 3-SAT instances. For example with $m \approx p \cdot (2n)^3$ we might consider the uniform distribution of all 3-SAT instances with exactly m different clauses. Note that m is about the expected number of clauses of a random formula from $Form_{n,p}$. One might also define clauses as sets of literals or one might forbid tautological clauses... In line with common usage we assume that it is only a technical matter to transfer our results to any of these possibilities to define random 3-SAT instances, but do not check the details.

Given a 3-SAT instance F over Var_n we assign two multigraphs $G_F = (V, E_F)$ and $G'_F = (V, E'_F)$ to F. $V = \mathrm{Var}_n \times \mathrm{Var}_n$ is the same in both cases, loops and multiedges are allowed. Given (a_1, b_1), $(a_2, b_2) \in V$ we have $(a_1, b_1) - (a_2, b_2) \in E_F$ iff there is a $z \in \mathrm{Var}_n$, such that $(a_1 \vee a_2 \vee z) \in F$ and $(b_1 \vee b_2 \vee \neg z) \in F$ *or* $(a_2 \vee a_1 \vee z) \in F$ and $(b_2 \vee b_1 \vee \neg z) \in F$. We have $(a_1, b_1) - (a_2, b_2) \in E'_F$ iff there is a $z \in \mathrm{Var}_n$, such that $(\neg a_1 \vee \neg a_2 \vee z) \in F$ and $(\neg b_1 \vee \neg b_2 \vee \neg z) \in F$ *or* $(\neg a_2 \vee \neg a_1 \vee z) \in F$ and $(\neg b_2 \vee \neg b_1 \vee \neg z) \in F$.

Note the following points: The b_i come from the clause with the $\neg z$ and are in the second position of the vertices connected.

Given a graph $G = (V, E)$, an independent set of G is a subset $S \subseteq V$ such that we have no edge $\{v, w\} \in E$ with both $v, w \in S$. The independence number of G, $i(G)$, is the maximal size an independent set of G can have. Of course, computing $i(G)$ is $\mathcal{NP}$-hard.

Lemma 2.1 *If the 3-SAT formula F over Var_n is satisfiable, then*

$$i(G_F) \geq n^2/4 \quad or \quad i(G'_F) \geq n^2/4\,.$$

Proof: Let α be a satisfying assignment of F and assume that α sets at least $n/2$ variables to true. We show that the set

$$S = \{(a,b)\,|\,\alpha(a) = \alpha(b) = true\}$$

is an independent set of G'_F and as $|S| \geq n^2/4$ the claim holds. Let $(a_1,b_1) - (a_2,b_2) \in E'_F$ then we have a variable z such that $(\neg a_1 \vee \neg a_2 \vee z)$, $(\neg b_1 \vee \neg b_2 \vee \neg z) \in F$ or $(\neg a_2 \vee \neg a_1 \vee z)$, $(\neg b_2 \vee \neg b_1 \vee \neg z) \in F$. As α is a satisfying assignment, in each case both clauses are *true* under α. If $\alpha(z) = 1$ then $\alpha(\neg z) = 0$ and $\alpha(b_1) = 0$ or $\alpha(b_2) = 0$. This means that $(a_1,b_1) \notin S$ or $(a_2,b_2) \notin S$. If $\alpha(z) = 0$ we have the same argument. Finally, if α sets at least $n/2$ variables to false we argue analogously with G_F. □

Let F be a 3-SAT formula over Var_n and let $a, b, z \in \mathrm{Var}_n$. We let

$$B_{a,b,z} = B_{a,b,z,p}(F) = \begin{cases} -1 & \text{if } a \vee b \vee z \in F \\ \frac{p}{1-p} & \text{if } a \vee b \vee z \notin F\,. \end{cases}$$

and

$$C_{a,b,z} = C_{a,b,z,p}(F) = \begin{cases} 1 & \text{if } a \vee b \vee \neg z \in F \\ -\frac{p}{1-p} & \text{if } a \vee b \vee \neg z \notin F\,. \end{cases}$$

For $F \in Form_{n,p}$ the probability $B_{a,b,z} = -1$ is equal to p and the probability of $B_{a,b,z} = \frac{p}{1-p}$ is equal to $1-p$. So of $B_{a,b,z}$ has an expectation of

$$\mathbf{E}[B_{a,b,z}] = -1 \cdot p + \frac{p}{1-p} \cdot (1-p) = -p + p = 0\,.$$

Equally the expectation of $C_{a,b,z}$ is 0.

For $V = \mathrm{Var}_n \times \mathrm{Var}_n$ and $(a_1,b_1),(a_2,b_2) \in V$ we define

$$\mathbf{a}_{a_1,b_1;\,a_2,b_2}(F) = \sum_{z \in \mathrm{Var}_n} B_{a_1,a_2,z} \cdot C_{b_1,b_2,z} \; + \sum_{z \in \mathrm{Var}_n} B_{a_2,a_1,z} \cdot C_{b_2,b_1,z} \qquad (1)$$

Now the $V \times V$-matrix $\mathcal{A} = \mathcal{A}_p(F)$ is given by the entries $\mathbf{a}_{a_1,b_1;\,a_2,b_2}$. Note that $\mathcal{A}$ corresponds to the adjacency matrix of the graph $G_F = (V, E_F)$ in that the non-existence of an edge $(a_1,b_1) - (a_2,b_2)$ is reflected by the fact that the sum for $\mathbf{a}_{a_1,b_1\,;a_2,b_2}$ consists only of terms $p/(1-p)$ or $-(p/(1-p))^2$.

In case of an edge we have at least once the term -1. Concerning the graph $G'_F = (V, E'_F)$ we introduce the analogous matrix $\mathcal{A}' = \mathcal{A}'_p(F)$ based on $B'_{a,b,z}$ and $C'_{a,b,z}$ with their obvious definition.

As $\mathcal{A}$ and $\mathcal{A}'$ are real-valued and symmetric, as can be easily seen from the definition, they have n^2 real-valued Eigenvalues $\lambda_{1,\mathcal{A}}, \lambda_{2,\mathcal{A}}, \ldots, \lambda_{n^2,\mathcal{A}}$ and $\lambda_{1,\mathcal{A}'}, \lambda_{2,\mathcal{A}'} \ldots, \lambda_{n^2,\mathcal{A}'}$, which we consider as ordered by their size

$$\lambda_{1,\mathcal{A}} \geq \ldots \geq \lambda_{n^2,\mathcal{A}} \quad \text{and} \quad \lambda_{1,\mathcal{A}'} \geq \ldots \geq \lambda_{n^2,\mathcal{A}'} .$$

Let

$$\lambda = \lambda_{\mathcal{A}} = \max\{|\lambda_{1,\mathcal{A}}|, \ldots, |\lambda_{n^2,\mathcal{A}}|\} = \max\{|\lambda_{1,\mathcal{A}}|, |\lambda_{n^2,\mathcal{A}}|\}$$

and analogously for $\lambda' = \lambda_{\mathcal{A}'}$. We state our theorems only for $\mathcal{A}$ and λ. They always apply to $\mathcal{A}'$ and λ', too.

For the whole rest of this paper we let

$$p = p(n) = \frac{\ln^6 n}{n^{3/2}} \qquad \text{and} \qquad f = f(n) = \ln^6 n$$

Theorem 2.2 *For $F \in Form_{n,p}$ we have with high probability that*

$$\lambda \leq \ln^5 n \cdot f$$

We prove this theorem in section 2.

In an asymptotic context, as ours is, we say that F is of low discrepancy with respect to

$$S = \{a \vee b \vee z \in F \,|\, a, b, z \in \mathrm{Var}_n\}$$

iff (a) and (b) hold.

(a) $|S| = f \cdot n^{3/2} \cdot (1 + o(1))$.

(b) For each $\varepsilon > 0$ constant, all sufficiently large n and sets $X, Y \subseteq \mathrm{Var}_n$ with $|X| = |Y| = \alpha \cdot n$ and $\varepsilon \leq \alpha \leq 1 - \varepsilon$ we have

$$|E(X)| = \alpha^2 \cdot f \cdot n^{3/2} \cdot (1+o(1)) \quad \text{and} \quad |E(X,Y)| = 2\alpha^2 \cdot f \cdot n^{3/2} \cdot (1+o(1)) ,$$

where

$$E(X) = \{\{a, b\}_{a \vee b \vee z} \,|\, a, b \in X, z \in V\}$$

and

$$E(X,Y) = \{(a, b)_{a \vee b \vee z} \,|\, a \in X, b \in Y, z \in V\} .$$

For sets like $S = \{a \vee b \vee \neg z \in F \,|\, a, b, z \in \mathrm{Var}_n\}$ or $S = \{\neg a \vee \neg b \vee \neg z \in F \,|\, a, b, z \in \mathrm{Var}_n\}$ the analogous notation is used. We say F is of low discrepancy iff F is of low discrepancy with respect to all 8 possible sets S as above.

Theorem 2.3 *If F is of low discrepancy, then $i(G_F) = \Omega(n^2)$ implies*

$$\lambda = \Omega(f^2)$$

.

An analogous statement applies to G'_F and λ'.

We prove this theorem in section 3. Now we can state our algorithm to certify unsatisfiability of 3-SAT formulas.

Algorithm 2.4 *certifies unsatisfiability of a given 3-SAT instance F over Var_n.*

1. *Certify low discrepancy of F using Algorithm 5.2.*
2. *Construct $\mathcal{A} = \mathcal{A}_p(F)$ and $\mathcal{A}' = \mathcal{A}'_p(F)$.*
3. *Determine λ and λ'.*
4. *If $\lambda = o(f^2)$ and $\lambda' = o(f^2)$ then certify F as unsatisfiable. Give an inconclusive answer otherwise.*

If F is satisfiable we have $i(G_F) \geq (n/2)^2$ of $i(G'_F) \geq (n/2)^2$. If in addition F is not of low discrepancy, we get an inconclusive answer in Step 1. If however F is of low discrepancy we get an inconclusive answer in Step 4. because of Theorem 2.3. Thus the algorithm is correct in that it gives no false answers. If $F \in Form_{n,p}$ is a random formula then Algorithm 5.2 certifies low discrepancy with high probability and Theorem 2.2 ensures that the algorithm certifies unsatisfiability. Numerical approximation algorithms allow the approximation of λ and λ' in polynomial time and the algorithm is efficient.

3 Proof of Theorem 2.2

As $\lambda = \lambda_{\mathcal{A}_p(F)}$, λ is a random variable defined on $Form_{n,p}$, Theorem 2.2 will follow from

Theorem 3.1

$$\mathbf{E}[\lambda^k] \le \left(\ln^4 n \cdot f\right)^k$$

provided k is the smallest even integer greater than $\ln n$.

Proof of Theorem 2.2: Using the Markov inequality we get

$$\begin{aligned}\Pr[\lambda \ge \ln^5 n \cdot f] = \Pr[\lambda^k \ge \ln^{5k} n \cdot f^k] &\le \frac{\mathbf{E}[\lambda^k]}{\ln^{5k} n \cdot f^k} \\ &\le \left(\frac{\ln^4 n \cdot f}{\ln^5 n \cdot f}\right)^k \le \left(\frac{1}{\ln n}\right)^k \le (\ln n)^{-\ln n} \\ &= o(1)\end{aligned}$$

and the assertion follows. □

In the rest of this section we prove Theorem 3.1. The proof uses the trace method inspired by the techniques of Füredi et al. [FuKo 80]. In our case we have dependencies between different entries in $\mathcal{A} = \mathcal{A}_p(F)$ which causes additional complications. We simplify $\mathrm{Var}_n = \{1, \ldots, n\} = [n]$ and $\mathcal{A}$ becomes an $([n] \times [n]) \times ([n] \times [n])$-matrix. Given F and an integer $k \ge 1$, the k'th power of $\mathcal{A}$ is denoted by $\mathcal{A}^k = (\mathbf{a}^k_{b,c;\, b',c'})_{1 \le b,c,b',c' \le n}$ where

$$\begin{aligned}\mathbf{a}^k_{b,c;\, b',c'} = \sum_{b_1=1}^{n} \sum_{c_1=1}^{n} \cdot \ldots \cdot \sum_{b_{k-1}=1}^{n} \sum_{c_{k-1}=1}^{n} &\mathbf{a}_{b,c;\, b_1,c_1} \cdot \mathbf{a}_{b_1,c_1;\, b_2,c_2} \cdot \ldots \\ &\cdot \mathbf{a}_{b_{k-2},c_{k-2};\, b_{k-1},c_{k-1}} \cdot \mathbf{a}_{b_{k-1},c_{k-1};\, b',c'}. \qquad (2)\end{aligned}$$

The trace of the matrix $\mathcal{A}^k$ is the sum of the entries on the main diagonal of $\mathcal{A}^k$ and we have

$$\mathrm{Trace}[\mathcal{A}^k] = \sum_{b=1}^{n} \sum_{c=1}^{n} \mathbf{a}^k_{b,c;\, b,c} = \sum_{i=1}^{n^2} \lambda^k_{i,\mathcal{A}}\,.$$

From now on we assume that k is even and we have

$$\lambda^k \leq \sum_{i=1}^{n^2} \lambda_{i,\mathcal{A}}^k = \text{Trace}[\mathcal{A}^k].$$

We bound $\text{Trace}[\mathcal{A}^k]$.

$$\begin{aligned}
&\text{Trace}[\mathcal{A}^k] \\
&= \sum_{b_1=1}^{n}\sum_{c_1=1}^{n} \cdot \ldots \cdot \sum_{b_k=1}^{n}\sum_{c_k=1}^{n} \mathbf{a}_{b_1,c_1;\, b_2,c_2} \cdot \mathbf{a}_{b_2,c_2;\, b_3,c_3} \cdot \ldots \cdot \mathbf{a}_{b_k,c_k;\, b_1,c_1} \\
&= \sum_{b_1,c_1,\ldots,b_k,c_k} \left(\sum_{z_1=1}^{n} \left(B_{b_1,b_2,z_1} \cdot C_{c_1,c_2,z_1} + B_{b_2,b_1,z_1} \cdot C_{c_2,c_1,z_1}\right) \right) \cdot \ldots \\
&\qquad\qquad \cdot \left(\sum_{z_k=1}^{n} \left(B_{b_k,b_1,z_k} \cdot C_{c_k,c_1,z_k} + B_{b_1,b_k,z_k} \cdot C_{c_1,c_k,z_k}\right) \right) \\
&= \sum_{b_1,\ldots,b_k} \sum_{c_1,\ldots,c_k} \sum_{z_1,\ldots,z_k} \left(B_{b_1,b_2,z_1} \cdot C_{c_1,c_2,z_1} + B_{b_2,b_1,z_1} \cdot C_{c_2,c_1,z_1}\right) \cdot \ldots \\
&\qquad\qquad \cdot \left(B_{b_k,b_1,z_k} \cdot C_{c_k,c_1,z_k} + B_{b_1,b_k,z_k} \cdot C_{c_1,c_k,z_k}\right)
\end{aligned}$$

We abbreviate $B = (b_1, \ldots, b_k)$, $C = (c_1, \ldots, c_k)$, $Z = (z_1, \ldots, z_k)$ and let

$$\begin{aligned}
\mathcal{P}(B,C,Z) = &\left(B_{b_1,b_2,z_1} \cdot C_{c_1,c_2,z_1} + B_{b_2,b_1,z_1} \cdot C_{c_2,c_1,z_1}\right) \cdot \ldots \\
&\cdot \left(B_{b_k,b_1,z_k} \cdot C_{c_k,c_1,z_k} + B_{b_1,b_k,z_k} \cdot C_{c_1,c_k,z_k}\right).
\end{aligned}$$

Given $B = (b_1, \ldots, b_k)$ we let $|B| = |\{b_1, \ldots, b_k\}|$ be the number of different members of B, and analogously for C and Z. Clearly $1 \leq |B|, |C|, |Z| \leq k$ and we can rewrite

$$\text{Trace}[\mathcal{A}^k] = \sum_{b=1}^{k}\sum_{c=1}^{k}\sum_{z=1}^{k} \sum_{\substack{B \\ |B|=b}} \sum_{\substack{C \\ |C|=c}} \sum_{\substack{Z \\ |Z|=z}} \mathcal{P}(B,C,Z) \tag{3}$$

The preceding considerations apply to any F and concerning the expectation we get

$$\mathbf{E}[\lambda^k] \leq \mathbf{E}[\text{Trace}[A^k]] = \sum_{b=1}^{k}\sum_{c=1}^{k}\sum_{z=1}^{k} \sum_{B,|B|=b} \sum_{C,|C|=c} \sum_{Z,|Z|=z} \mathbf{E}[\mathcal{P}(B,C,Z)]. \tag{4}$$

Computing the expectation $\mathbf{E}[\mathcal{P}(B,C,Z)]$ we can restrict $|B|$, $|C|$ and $|Z|$.

Lemma 3.2 *If* $|B| \geq k/2+2$, $|C| \geq k/2+2$ *or* $|Z| \geq k/2+1$ *then*

$$\mathbf{E}[\mathcal{P}(B,C,Z)] = 0.$$

Proof: First we consider $Z = (z_1, \ldots, z_k)$. If $|Z| \geq k/2+1$ there is a z_m, $1 \leq m \leq k$, which occurs only once among the z_i of Z. In $\mathcal{P}(B,C,Z)$ the factor corresponding to z_m is $B_{b_m,b_{m+1},z_m} \cdot C_{c_m,c_{m+1},z_m} + B_{b_{m+1},b_m,z_m} \cdot C_{c_{m+1},c_m,z_m}$ if $m < k$, and 1 instead of $m+1$ if $m = k$. The expectation of this factor is 0 as $\mathbf{E}[B_\beta] = \mathbf{E}[C_\gamma] = 0$ and B_β and C_γ are stochastically independent for any possible sequence of indices for β and γ. Now, as z_m occurs only once the factor corresponding to z_m is independent of the rest of $\mathcal{P}(B,C,Z)$ and $\mathbf{E}[\mathcal{P}(B,C,Z)] = 0$.

Next we consider $B = (b_1, \ldots, b_k)$, where we denote $b = |B|$. Let $C = (c_1, \ldots, c_k)$ and $Z = (z_1, \ldots, z_k)$ be fixed but otherwise arbitrary.

$$\begin{aligned}\mathcal{P}(B,C,Z) = &(B_{b_1,b_2,z_1} \cdot C_{c_1,c_2,z_1} + B_{b_2,b_1,z_1} \cdot C_{c_2,c_1,z_1}) \cdot \ldots \\ &\cdot (B_{b_k,b_1,z_k} \cdot C_{c_k,c_1,z_k} + B_{b_1,b_k,z_k} \cdot C_{c_1,c_k,z_k})\end{aligned}$$

can be naturally be represented as a sum of 2^k terms. Let

$$X = B_{\beta_1} \cdot C_{\gamma_1} \cdot B_{\beta_2} \cdot C_{\gamma_2} \cdot \ldots \cdot B_{\beta_k} \cdot C_{\gamma_k}$$

be such a term. That means $\beta_i = (b_i, b_{i+1}, z_i)$ and $\gamma_i = (c_i, c_{i+1}, z_i)$ or $\beta_i = (b_{i+1}, b_i, z_i)$ and $\gamma_i = (c_{i+1}, c_i, z_i)$ for $1 \leq i < k$ and analogously for β_k and γ_k with 1 instead of $k+1$.

We claim that X must have at least $b-1$ different B-factors. To see this we go from left to right over the B-factors. The first B-factor needs at most 2 of the b different b_i's which do not occur before. Each of the remaining B-factors can have at most one b_i which has not already occurred. As the B-factors with different indices are distinct, we have at least $b-1$ different B-factors.

Now, if $b \geq k/2+2$ we have that $b-1 \geq k/2+1$ and at least one B-factor occurs only once in X, By independence $\mathbf{E}[X] = 0$. As the same argument applies to all 2^k terms of $\mathcal{P}(B,C,Z)$, the claim of the lemma holds. For $C = (c_1, \ldots, c_k)$ we can argue in the same manner. □

It is easy to construct an example sich that the claim does not hold for $|B| = k/2+1$. For example $k = 4$ and consider $B_{b_1,b_2,-}$, $B_{b_2,b_3,-}$, $B_{b_2,b_3,-}$, $B_{b_1,b_2,-}$ which can occur together.

Lemma 3.3 *Let B, C, Z be k-tuples of variables and let $b = |B|$, $c = |C|$ and $z = |Z|$. Then we have that*

$$\mathbf{E}[\mathcal{P}(B, C, Z)] \leq 2^k \cdot (2p)^{m_1+m_2}$$

where $m_1 = \max(b-1, z)$ and $m_2 = \max(c-1, z)$.

Proof: Let $B = (b_1, \ldots, b_k)$, $C = (c_1, \ldots, c_k)$ and $Z = (z_1, \ldots, z_k)$. The random variable

$$\begin{aligned}\mathcal{P}(B, C, Z) = &(B_{b_1,b_2,z_1} \cdot C_{c_1,c_2,z_1} + B_{b_2,b_1,z_1} \cdot C_{c_2,c_1,z_1}) \cdot \ldots \\ &\cdot (B_{b_k,b_1,z_k} \cdot C_{c_k,c_1,z_k} + B_{b_1,b_k,z_k} \cdot C_{c_1,c_k,z_k})\end{aligned}$$

can be naturally written as a sum with 2^k terms. Let again

$$X = B_{\beta_1} \cdot C_{\gamma_1} \cdot B_{\beta_2} \cdot C_{\gamma_2} \cdot \ldots \cdot B_{\beta_k} \cdot C_{\gamma_k}$$

be such a term. To calculate the expectation $\mathbf{E}[X]$ observe that the factors are independent unless they are equal. For $\beta = \beta_i$ for $1 \leq i \leq k$ and $r \geq 1$ we have

$$\begin{aligned}\mathbf{E}[B_\beta^r] &= p \cdot (-1)^r + (1-p) \cdot \left(\frac{p}{1-p}\right)^r \\ &\leq p + \frac{p^r}{(1-p)^{r-1}} \\ &\leq 2 \cdot p\end{aligned}$$

as $p \leq 1/2$ and we assume that n is sufficiently large. The number of different B-factors is at least $m_1 = \max(b-1, z)$ (refer to the proof of Lemma 3.2). Therefore

$$\mathbf{E}[B_{\beta_1} \cdot \ldots \cdot B_{\beta_k}] \leq (2p)^{m_1}$$

as $2p \leq 1$. In the same way we get with $m_2 = \max(c-1, z)$

$$\mathbf{E}[C_{\gamma_1} \cdot \ldots \cdot C_{\gamma_k}] \leq (2p)^{m_2}.$$

From linearity of expectation we finally get

$$\mathbf{E}[\mathcal{P}(B, C, Z)] \leq 2^k \cdot (2p)^{m_1+m_2}.$$

□

The subsequent estimate of $\mathbf{E}[\mathrm{Trace}[\mathcal{A}^k]]$ together with (4) yields the proof of Theorem 3.1.

Theorem 3.4

$$\mathbf{E}[\mathrm{Trace}[\mathcal{A}^k]] \leq \left(\ln^4 n \cdot f\right)^k$$

where k is the smallest even integer greater than $\ln n$.

Proof: With Lemma 3.2 and (3) we have that

$$\mathbf{E}[\mathrm{Trace}[\mathcal{A}^k]] = \sum_{b=1}^{k/2+1} \sum_{c=1}^{k/2+1} \sum_{z=1}^{k/2} \sum_{\substack{B \\ |B|=b}} \sum_{\substack{C \\ |C|=c}} \sum_{\substack{Z \\ |Z|=z}} \mathbf{E}[\mathcal{P}(B,C,Z)].$$

For given b, c and z and $m_1 = \max(b-1, z)$ and $m_2 = \max(c-1, z)$ we have with Lemma 3.3

$$\mathbf{E}[\mathcal{P}(B,C,Z)] \leq 2^k \cdot (2p)^{m_1+m_2}.$$

As this is independent of the actual sequences B, C and Z we count the number of such k-tuples. The number of different k-tuples B with $|B| = b$ is at most

$$\binom{k}{b} \cdot n^b \cdot b^{k-b}.$$

This is because each possible B can be obtained from the following choosing process:

1. Pick the b positions among the k available altogether in which one of the b variables occurs for the first time when going from left to right over the k positions available:

$$\binom{k}{b} \text{ possibilities.}$$

2. Pick the b different variables and place them into the slots picked in 1.:

$$\binom{n}{b} \cdot b! \leq n^b \text{ possibilities.}$$

3. Fill the remaining slots with the b variables picked in 2.:

$$\leq b^{k-b} \text{ possibilities.}$$

As $1 \leq b \leq k$ we can bound the number of different k-tuples B by

$$\begin{aligned}\binom{k}{b} \cdot n^b \cdot b^{k-b} &\leq 2^k \cdot n^b \cdot b^{k-b} \\ &\leq n^b \cdot 2^k \cdot k^k\end{aligned}$$

With analogous considerations we can bound the number of k-tuples C with $|C| = c$ (resp. Z with $|Z| = z$) by $n^c \cdot 2^k \cdot k^k$ (resp. $n^z \cdot 2^k \cdot k^k$). Thus we get that

$$\mathbf{E}[\mathrm{Trace}[\mathcal{A}^k]] \leq \sum_{b=1}^{k/2+1} \sum_{c=1}^{k/2+1} \sum_{z=1}^{k/2} 2^{3k} \cdot k^{3k} \cdot n^{b+c+z} \cdot 2^k \cdot (2p)^{m_1+m_2}. \tag{5}$$

Further below we bound $n^{b+c+z} \cdot (2p)^{m_1+m_2} \leq (2f)^k \cdot n^2$. This yields the claim as we get from (5)

$$\begin{aligned}\mathbf{E}[\mathrm{Trace}[\mathcal{A}^k]] \leq& \sum_{b=1}^{k/2+1} \sum_{c=1}^{k/2+1} \sum_{z=1}^{k/2} 2^{4k} \cdot k^{3k} \cdot (2f)^k \cdot n^2 \\ \leq& (k/2+1)(k/2+1)(k/2) \cdot 2^{5k} \cdot k^{3k} \cdot f^k \cdot n^2\end{aligned}$$

and under assumption k is the smallest even integer greater than $\ln n$, we get with n sufficiently large

$$\mathbf{E}[\mathrm{Trace}[\mathcal{A}^k]] \leq \left(\ln^4 n \cdot f\right)^k .$$

To show the bound $n^{b+c+z} \cdot (2p)^{m_1+m_2} \leq (2f)^k \cdot n^2$ we calculate three cases

1. $z \geq b-1 \geq c-1$
2. $b-1 > z \geq c-1$
3. $b-1 \geq c-1 > z$.

The remaining three cases, where $c > b$ are analogous.

1. For $z \geq b-1 \geq c-1$ we get

$$\begin{aligned}(2p)^{m_1+m_2} \cdot n^{b+c+z} &= (2p)^{2z} \cdot n^{b-1+c-1+z} \cdot n^2 \\ &\leq (4p^2)^z \cdot n^{3z} \cdot n^2 \\ &\leq (4p^2 \cdot n^3)^{k/2} \cdot n^2 \\ &= (2pn^{3/2})^k \cdot n^2 \\ &= (2f)^k \cdot n^2\end{aligned}$$

2. For $b-1 > z \geq c-1$ we get

$$\begin{aligned}(2p)^{m_1+m_2} \cdot n^{b+c+z} &= (2p)^{b-1+z} \cdot n^{b-1+c-1/2+z} \cdot n^{3/2} \\ &\leq (2p)^{b-1+z} \cdot n^{3/2\cdot(b-1+z)} \cdot n^2 \\ &\leq (2pn^{3/2})^k \cdot n^2 \\ &= (2f)^k \cdot n^2\end{aligned}$$

3. For $b-1 \geq c-1 > z$ we get

$$\begin{aligned}(2p)^{m_1+m_2} \cdot n^{b+c+z} &= (2p)^{b-1+c-1} \cdot n^{b-1+c-1+z+1} \cdot n \\ &\leq (2p)^{b-1+c-1} \cdot n^{3/2\cdot(b-1+c-1)} \cdot n \\ &\leq (2pn^{3/2})^k \cdot n^2 \\ &\leq (2f)^k \cdot n^2\end{aligned}$$

□

4 Proof of Theorem 2.3

Throughout this section we let F be a 3-SAT instance over Var_n which has low discrepancy. Again we identify Var_n with $[n] = \{1, \ldots, n\}$. We consider the graph $G_F = (V, E_F)$ with $V = [n] \times [n]$. $\mathcal{A} = \mathcal{A}_p(F)$ is the $V \times V$-matrix associated to F as in (1) on page 4.

The vectors we consider are column vectors whose coordinates are indexed by V. The value of a vector v at coordinate (a, b) is denoted by $v_{a,b}$. For $W \subseteq V$ we let $\chi = \chi_W$ be the characteristic vector of W. That is

$$\chi_{(a,b)} = \begin{cases} 1 & \text{if } (a,b) \in W \\ 0 & \text{otherwise .} \end{cases}$$

The subsequent relations allow us to relate $\lambda = \lambda_{\mathcal{A}}$ and $i(G_F)$. Let $\Psi = \mathcal{A} \cdot \chi_W$, then for $(a, b) \in V$ we have

$$\Psi_{(a,b)} = \sum_{(a',b') \in W} \mathbf{a}_{a,b;\, a',b'}.$$

When χ_W^{tr} is the transpose of χ_W we have that

$$\begin{aligned}
&\chi_W^{tr} \cdot \Psi \\
&= \sum_{(a,b)\in W} \Psi_{(a,b)} \\
&= \sum_{(a,b)\in W} \sum_{(a',b')\in W} \mathbf{a}_{a,b;\,a',b'} \\
&= \sum_{(a,b)\in W} \sum_{(a',b')\in W} \left(\sum_{z\in \mathrm{Var}_n} B_{a,a',z} \cdot C_{b,b',z} + \sum_{z\in \mathrm{Var}_n} B_{a',a,z} \cdot C_{b',b,z} \right) \\
&= \sum_{(a,b)\in W} \sum_{(a',b')\in W} \sum_{z\in \mathrm{Var}_n} B_{a,a',z} \cdot C_{b,b',z} + \sum_{(a,b)\in W} \sum_{(a',b')\in W} \sum_{z\in \mathrm{Var}_n} B_{a',a,z} \cdot C_{b',b,z} \\
&= 2 \cdot \sum_{(a,b)\in W} \sum_{(a',b')\in W} \sum_{z\in \mathrm{Var}_n} B_{a,a',z} \cdot C_{b,b',z}
\end{aligned}$$

The last equation holds since (a, b) and (a', b') run through the same set W.

The next lemma is included for expository reasons only, in order to point out the ideas of the proof of Theorem 2.3.

Lemma 4.1 *Let $c > 1$ be a constant and let*

$$I = \{(a, b) \,|\, 1 \leq a, b \leq n/c\}$$

be an independent set of G_F, then

$$\lambda = \Omega(f^2).$$

Proof: Let $\chi = \chi_I$ be the characteristic vector of I. From the min-max characterization of Eigenvalues [Pr 94] of symmetric matrices we have that

$$\lambda \geq \lambda_{1,\mathcal{A}} = \max_{v\neq 0} \frac{v^{tr} \cdot \mathcal{A} \cdot v}{v^{tr} \cdot v} \geq \frac{\chi^{tr} \cdot \mathcal{A} \cdot \chi}{\chi^{tr} \cdot \chi} \,.$$

The denominator of the preceding fraction is $D = (n/c)^2 = |I|$. The enumerator is

$$N = 2 \cdot \sum_{(a,b)\in I} \sum_{(a',b')\in I} \sum_{z\in \mathrm{Var}_n} B_{a,a',z} \cdot C_{b,b',z}$$

as shown above. We show that $N = \Omega(f^2 \cdot n^2)$. From the simple structure of I we get

$$\sum_{(a,b)\in I} \sum_{(a',b')\in I} \sum_{z\in \mathrm{Var}_n} B_{a,a',z} \cdot C_{b,b',z} = \sum_{1\le a,a',b,b'\le n/c} \sum_{1\le z\le n} B_{a,a',z} \cdot C_{b,b',z} \qquad (6)$$

In principle we have four possibilities for $B_{a,a',z}$ and $C_{b,b',z}$ which yield three different values $B_{a,a',z} \cdot C_{b,b',z}$:

1. If $B_{a,a',z} = -1$, $C_{b,b',z} = 1$ then $B_{a,a',z} \cdot C_{b,b',z} = -1$
2. If $B_{a,a',z} = -1$, $C_{b,b',z} = -p/(1-p)$ then $B_{a,a',z} \cdot C_{b,b',z} = p/(1-p)$
3. If $B_{a,a',z} = p/(1-p)$, $C_{b,b',z} = 1$ then $B_{a,a',z} \cdot C_{b,b',z} = p/(1-p)$
4. If $B_{a,a',z} = p/(1-p)$, $C_{b,b',z} = -p/(1-p)$ then $B_{a,a',z} \cdot C_{b,b',z} = -(p/(1-p))^2$

We count how often each of the possibilities occur in the sum (6).

1. This possibility does not occur as I is an independent set of G_F.
2. Let $M = \{(a,a',z) \mid B_{a,a',z} = -1\}$. For each $B_{a,a',z}$ we have exactly $(n/c)^2$ possible factors $C_{b,b',z}$. From low discrepancy of F we know that $|M| = 1/c^2 \cdot f \cdot n^{3/2} \cdot (1 + o(1))$. Therefore we get an asymptotic contribution of
$$\frac{p}{1-p} \cdot \frac{1}{c^2} \cdot f \cdot n^{3/2} \cdot \frac{n^2}{c^2} \ge \frac{f^2 n^2}{c^4}$$
to the sum, as $p = f/n^{3/2}$.
3. In the same way as in (b) we get using low discrepancy of the clauses $b \vee b' \vee \neg z \in F$, $b, b', z \in \mathrm{Var}_n$, a contribution of
$$\ge \frac{f^2 \cdot n^2}{c^4}$$
4. We have at most n^5/c^4 possibilities to choose a, a', b, b', z. Each possibility gives an addend of $-(p/(1-p))^2$ and we get at least
$$-\frac{n^5}{c^4} \cdot \frac{p^2}{(1-p)^2} = -\frac{n^2 f^2}{c^4 \cdot (1-o(1))^2} = -\frac{n^2 f^2 \cdot (1+o(1))}{c^4},$$
as $p = o(1)$.

Thus we get asymptotically

$$\begin{aligned} N &= 2 \cdot \sum_{(a,b)\in I} \sum_{(a',b')\in I} \sum_{z\in \mathrm{Var}_n} B_{a,a',z} \cdot C_{b,b',z} \\ &\geq 2 \cdot \left(2 \cdot \frac{f^2 \cdot n^2}{c^4} - \frac{f^2 \cdot n^2}{c^4} \right) \\ &= \Omega(f^2 \cdot n^2)\,. \end{aligned}$$

The claim of the lemma holds as

$$\lambda \geq \frac{N}{D} = \frac{\Omega(f^2 \cdot n^2)}{(n/c)^2} = \Omega(f^2).$$

□

Theorem 2.3 is the analogue of Lemma 4.1 for independent sets I whose structure may be arbitrary. The key in the proof of Lemma 4.1 is equation (6), that

$$\sum_{(a,b)\in I} \sum_{(a',b')\in I} \sum_{z\in \mathrm{Var}_n} \text{ can be replaced by } \sum_{1\leq a,a',b,b'\leq n/c} \sum_{1\leq z\leq n}.$$

To apply the principle from the proof of Lemma 4.1 to arbitrary sets $I \subseteq V$ we introduce some structure.

Definition 4.2

(a) For $a \in Var_n$ an s-section of I with respect to a in the first position is a subset

$$S = \{(a, b_1), (a, b_2), \ldots, (a, b_s)\} \subseteq I$$

where the b_i are all distinct. An s-section S' with respect to b in the second position is defined analogously as

$$S' = \{(a_1, b), (a_2, b), \ldots, (a_s, b)\} \subseteq I$$

(b) For $1 \leq i \leq s$ let S_i be an s-section of I with respect to a_i in the first position and the a_i are pairwise distinct. An $s \times s$-tile of I with respect to the first position is a subset

$$T = S_1 \cup \ldots \cup S_s \subseteq I.$$

With respect to the second position the notion of an $s \times s$-tile is defined analogously.

The structure we impose on I is a decomposition into disjoint tiles plus a small and furthermore irrelevant rest which does not contain a tile anymore.

Lemma 4.3 *Let $c > 1$ be a constant and $\varepsilon > 0$ be a constant which we assume sufficiently small and let $s = \lfloor \varepsilon \cdot n \rfloor$. Let $I \subseteq V$ with $|I| = (n/c)^2$. Then I contains at least $9/(10c^2\varepsilon^2)$ many $s \times s$-tiles with respect to the first position which are pairwise disjoint. The same statement applies to the second position.*

Proof: We consider the following process.

1. Pick s variables $a_1, \ldots, a_s$ such that for each $a = a_i$ we can pick at least s vertices $(a, -) \in I$. If no such vertices can be found, the process stops.

2. Consider the vertices picked in 1. as an $s \times s$-tile and delete them from I.

3. Continue the process at step 1.

We claim that any set $W \subseteq V$ with $|W| \geq 2\varepsilon \cdot n^2 \geq 2s \cdot n$ contains an $s \times s$-tile. This follows because if W contains no $s \times s$-tile we have at most $s - 1$ variables a such that we have s or more vertices $(a, -) \in W$. As we can have at most n vertices $(a, -) \in W$ and for all the remaining $n - (s-1)$ variables a we have less than s vertices $(a, -) \in W$ we get

$$\begin{aligned} |W| &\leq (s-1) \cdot n + (n - (s-1)) \cdot (s-1) \\ &= sn - n + n2 - n - s^2 + s + s - 1 \\ &= 2ns - 2n - s^2 + 2s - 1 \\ &\leq 2ns \end{aligned}$$

as $\varepsilon \leq 1$ and $2s \leq 2\varepsilon n \leq 2n$.

The number of $s \times s$-tiles found by the process is at least

$$\begin{aligned}
&\frac{n^2/c^2 - 2ns}{s^2} \\
&= \frac{n^2}{c^2 s^2} - \frac{2n}{s} \\
&\geq \frac{n^2}{c^2 \cdot (\varepsilon n)^2} - \frac{2n}{\varepsilon n - 1} \\
&\geq \frac{1}{c^2 \cdot \varepsilon^2} - \frac{2n}{\varepsilon n - 1} \\
&\geq \frac{1}{c^2 \cdot \varepsilon^2} - \frac{2}{\varepsilon - 1/n} \\
&\geq \frac{9}{10c^2 \cdot \varepsilon^2}
\end{aligned}$$

which implies the claim if only we pick ε sufficiently small ($\varepsilon < 1/(20c^2)$ is enough) and n is large. □

Now we come to the proof of Theorem 2.3: Let F have low discrepancy and let I with $|I| = (n/c)^2$ be an independent set of G_F. As in the proof of Lemma 4.1 we get that

$$\lambda = \lambda_{\mathcal{A}_p(F)} \geq \frac{N}{D}$$

with

$$N = 2 \cdot \sum_{(a,b) \in I} \sum_{(a',b') \in I} \sum_{z \in \mathrm{Var}_n} B_{a,a',z} \cdot C_{b,b',z}$$

and $D = (n/c)^2$.

Let

$$J = \{((a,b),(a',b'),z) \,|\, (a,b),(a',b') \in I, z \in \mathrm{Var}_n\}$$

be the set of indices of our sum. We partition J into four sets:

$$\begin{aligned}
J_1 &= \{((a,b),(a',b'),z) \in J \,|\, a \vee a' \vee z, b \vee b' \vee z \in F\} \\
J_2 &= \{((a,b),(a',b'),z) \in J \,|\, a \vee a' \vee z \in F, b \vee b' \vee z \notin F\} \\
J_3 &= \{((a,b),(a',b'),z) \in J \,|\, a \vee a' \vee z \notin F, b \vee b' \vee z \in F\} \\
J_4 &= \{((a,b),(a',b'),z) \in J \,|\, a \vee a' \vee z, b \vee b' \vee z \notin F\}
\end{aligned}$$

We consider the four statements to be proved below.

$$\begin{aligned}
&|J_1| = 0\\
&|J_2| \geq (9/(10c^2))^2 \cdot f \cdot n^{3/2} \cdot n^2\\
&|J_3| \geq (9/(10c^2))^2 \cdot f \cdot n^{3/2} \cdot n^2\\
&|J_4| \leq (n^2/c^2)^2 \cdot n
\end{aligned}$$

From these statements the theorem follows:

$$\begin{aligned}
N &= \sum_{J_1} B_{a,a',z} \cdot C_{b,b',z} + \sum_{J_2} B_{a,a',z} \cdot C_{b,b',z}\\
&\qquad + \sum_{J_3} B_{a,a',z} \cdot C_{b,b',z} + \sum_{J_4} B_{a,a',z} \cdot C_{b,b',z}\\
&\geq 0 + \left(\frac{9}{10c^2}\right)^2 \cdot f \cdot n^{3/2} \cdot n^2 \cdot \frac{p}{1-p}\\
&\qquad + \left(\frac{9}{10c^2}\right)^2 \cdot f \cdot n^{3/2} \cdot n^2 \cdot \frac{p}{1-p} - \left(\frac{n^2}{c^2}\right)^2 \cdot n \cdot \left(\frac{p}{1-p}\right)^2\\
&\geq 2 \cdot \left(\frac{9}{10c^2}\right)^2 \cdot f^2 \cdot n^2 - \frac{n^2 \cdot f^2 \cdot (1+o(1))}{c^4}\\
&\geq \frac{1}{2c^4} \cdot f^2 \cdot n^2\\
&= \Omega(f^2 n^2)
\end{aligned}$$

The statement $|J_1| = 0$ is true because I is an independent set.

Now we come to J_2. Here we need our tiles. First let $s = \lfloor \varepsilon \cdot n \rfloor$, where ε is a constant, and let $S, T \subseteq I$ be two disjoint $s \times s$-tiles with respect to the first position. Let

$$K_S = \{((a,b),(a',b'),z) \in J_2 \,|\, (a,b),(a',b') \in S, z \in \mathrm{Var}_n\}$$

and

$$\begin{aligned}
K_{S,T} = \{((a,b),(a',b'),z) \in J_2 \,|\, &(a,b) \in S, (a',b') \in T \text{ or}\\
&(a',b') \in S, (a,b) \in T, z \in \mathrm{Var}_n\}.
\end{aligned}$$

Note that $K_{S,T} = K_{T,S}$. We prove below that asymptotically

$$|K_S| = \varepsilon^2 \cdot f \cdot n^{3/2} \cdot s^2 \tag{7}$$

$$|K_{S,T}| = 2\varepsilon^2 \cdot f \cdot n^{3/2} \cdot s^2 \,. \tag{8}$$

With this two statements we argue as follows. For $s = \lfloor \varepsilon \cdot n \rfloor$ let ε sufficiently small, such that Lemma 4.3 can be applied. Then we get $l \geq 9/(10c^2\varepsilon^2)$ many disjoint $s \times s$-tiles $T_1, \ldots, T_l$ inside of I. Then we have that all sets K_{T_i} and $K_{T_iT_j}$ with $i < j$ are disjoint. Moreover,

$$J_2 \supseteq K_{T_1} \,\dot\cup\, \ldots \,\dot\cup\, K_{T_l} \,\dot\cup\, \dot{\bigcup}_{1 \leq i < j \leq l} K_{T_i,T_j}.$$

From (7) and (8) we get that asymptotically

$$\begin{aligned}
|J_2| &\geq l \cdot \varepsilon^2 \cdot f \cdot n^{3/2} \cdot s^2 + \binom{l}{2} \cdot 2 \cdot \varepsilon^2 \cdot f \cdot n^{3/2} \cdot s^2 \\
&\geq l^2 \cdot \varepsilon^2 \cdot f \cdot n^{3/2} \cdot s^2 \\
&\geq \left(\frac{9}{10c^2\varepsilon^2}\right)^2 \cdot \varepsilon^2 \cdot f \cdot n^{3/2} \cdot (\varepsilon n - 1)^2 \\
&\geq \left(\frac{9}{10c^2}\right)^2 \cdot f \cdot n^{3/2} \cdot n^2
\end{aligned}$$

We need to calculate $|K_S|$. Let

$$S = \begin{matrix}
(a_1, b_{1,1}) & (a_1, b_{1,2}) & \ldots & (a_1, b_{1,s}) \\
(a_2, b_{2,1}) & (a_2, b_{2,2}) & \ldots & (a_2, b_{2,s}) \\
\vdots & \vdots & & \vdots \\
(a_s, b_{s,1}) & (a_s, b_{s,2}) & \ldots & (a_s, b_{s,s})
\end{matrix} \subseteq J_2 \,.$$

The a_i are all distinct as are the $b_{i,j}$. Let $1 \leq i, j \leq s$ and assume $a_i \vee a_i' \vee z \in F$ for a $z \in \mathrm{Var}_n$. Then for all $1 \leq j, j' \leq s$

$$b_{i,j} \vee b_{i',j'} \vee z \notin F$$

as I is independent. These are s^2 clauses. Therefore for all $1 \leq j, j' \leq s$

$$((a_i, b_{i,j}), (a_{i'}, b_{i',j'}), z) \in K_S \,.$$

These are s^2 different indices j, j'. From low discrepancy of F the number of clauses $a_i \vee a_{i'} \vee - \in F$ where $1 \leq i, i' \leq s$ is asymptotically $\varepsilon^2 \cdot f \cdot n^{3/2}$.

As the contributions to K_S induced by different clauses $a_i \vee a_{i'} \vee - \in F$ are disjoint sets of s^2 indices asymptotically

$$|K_S| = \varepsilon^2 \cdot f \cdot n^{3/2} \cdot s^2 \,.$$

Now we come to $K_{S,T}$. Here the typical situation can be sketched by

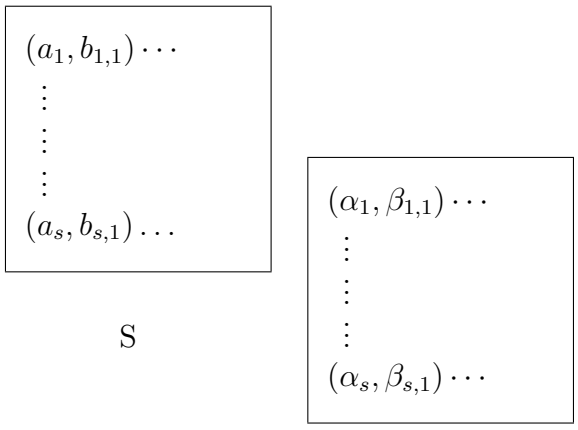

Figure 1

The technical problem here is the overlap in the variable in the first position. That is, it may be that $a_j = \alpha_\nu$. Let

$$X = \{a_1, \ldots, a_s\} \quad \text{and} \quad Y = \{\alpha_1, \ldots, \alpha_s\} \,.$$

Let $a_i \in X \setminus Y$ and $a_i \vee \alpha_\nu \vee z \in F$. In this case we have that for all $1 \leq j, \mu \leq s$ $b_{i,j} \vee \beta_{\nu,\mu} \vee z \notin F$ as I is independent. Thus get a contribution of s^2 elements to $K_{S,T}$. The same argument applies for $\alpha_\nu \in Y \setminus X$.

The remaining case is that for $a_i \in Y$ and $\alpha_\nu \in X$, that is $a_i, \alpha_\nu \in X \cap Y$, $a_i \vee \alpha_\nu \vee z \in F$. In this case we get that for all $1 \leq j, \mu \leq s$ (reading $a_i \in X, \alpha_\nu \in Y$)

$$b_{i,j} \vee \beta_{\nu,\mu} \vee z \notin F$$

and

$$\beta_{\nu,\mu} \vee b_{i,j} \vee z \notin F$$

(reading $a_i \in Y, \alpha_\nu \in X$). We get a contribution of $2s^2$ elements to $K_{S,T}$.

What is the overall contribution to $K_{S,T}$ of all clauses $a_i \vee a_\nu \vee z \in F$? We claim that it is $|E(X,Y)| \cdot s^2$. Here $E(X,Y)$ refers to the projection of the all-positive clauses of F. From this claim to be proved below we get (8) as asymptotically

$$|E(X,Y)| = 2 \cdot \varepsilon^2 \cdot f \cdot n^{3/2}$$

by low discrepancy.

The claim above holds because in $E(X,Y)$ each edge of the projection $\{a_i, \alpha_\nu\}_{a_i \vee \alpha_\nu \vee z}$ where $a_i, \alpha_\nu \in X \cap Y$ occurs twice as $(a_j, \alpha_\nu)_{a_i \vee \alpha_\nu \vee z}$ and $(\alpha_\nu, a_j)_{a_i \vee \alpha_\nu \vee z}$ whereas for $a_i \in X \setminus Y$ we only have $(a_i, \alpha_\nu)_{a_i \vee \alpha_\nu \vee z}$ and the same for $a_\nu \in Y \setminus X$,

J_3 is treated as J_2 using tiles with respect to the second coordinate. The claim for J_4 is obvious.

5 Discrepancy considerations

Given a set of 3-clauses S over Var_n the projection (onto coordinates 1 and 2)of S is the labelled multigraph $G = (V, E)$ wit $V = \mathrm{Var}_n$ and

$$\{a, b\}_{a \vee b \vee z} \in E \quad \text{iff} \quad a \vee b \vee z \in S\,.$$

That is edges are labelled by clauses and we get one edge for each clause. Loops are also possible. Let $d = d(n)$. In our asymptotic context we say that a projection is almost d-regular iff the degree of each vertex of the projection is $d(n) \cdot (1 + o(1))$. The projection $G = (V, E)$ is of low discrepancy if for any constant $\varepsilon > 0$ we have that

$$|E(X,Y)| = 2 \cdot e \cdot \alpha \cdot \beta \cdot (1 + o(1))$$

where $X, Y \subseteq V$ with $|X| = \alpha n$, $|Y| = \beta n$ and $\varepsilon \leq \alpha, \beta \leq 1 - \varepsilon$.

Low discrepancy can be certified by Eigenvalue methods. Given a projection, the adjacency and the Laplacian matrix $A = A(G)$and $L = L(G)$ are well defined. Edges are counted with their multiplicity in $A(G)$. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ be the Eigenvalues of $I - L(G)$. Then $\lambda_1 = 1$, we let $\lambda = \max(|\lambda_2|, |\lambda_n|)$. We refer to [Ch 97]. The natural extension to multigraphs of Theorem 5.1 on page 71 of [Ch 97] gives

Fact 5.1 *Let $G = (V, E)$ be a projection which is almost d-regular. For $X, Y \subseteq V$, $|X| = \alpha \cdot n$, $Y = \beta \cdot n$ we have*

$$|E(X,Y) - 2 \cdot e \cdot \alpha \cdot \beta| \leq \lambda \cdot e \cdot \sqrt{\alpha \cdot \beta}$$

where $e = |E|$ is the number of labelled edges of G.

Thus $\lambda = o(1)$ implies low discrepancy of the projection. As in [Coja et al] $\lambda = o(1)$ can be shown by showing that $\mathrm{Trace}[A^k] = d^k + o(d^k)$.

Algorithm 5.2 *certifies low discrepancy. Input is a set of clauses S.*

1. *Construct the projection $G = (V, E)$ of S.*
2. *Check almost regularity and determine a suitable d.*
3. *Compute* $\mathrm{Trace}[A^6]$.
4. *Check if* $\mathrm{Trace}[A^6] = d^6 + o(d^6)$. *If positive, certify low discrepancy otherwise give an inconclusive answer.*

The correctness of the algorithm follows from Fact 5.1. The algorithm is complete if S is the set of all positive clauses of a random $F \in Form_{n,p}$, where $p = f/n^{3/2}$. In this case the projection is almost a random graph with n vertices and edge probability $p \cdot n = f/n^{1/2}$. We refer to [Coja et al] for more details. Of course the algorithm is also complete if S is the set of all clauses of F, whose first and second literal is positive and the third is negative and also for the other six possibilities for S.

References

[Ch 97] Fan R. K. Chung, *Spectral Graph Theory*, CBMS Regional Conference Series in Mathematics, Vol. 92. American Mathematical Society, Providence, R.I., 1997

[Coja et al] A. Coja-Oghlan, A. Goerdt, A. Lanka, F. Schädlich *Certifying Unsatisfiability of Random 2k-SAT Formulas using Approximation Techniques*, In Proceedings FCT 2003

[Fe 2002] Uriel Feige. *Relations between average case complexity and approximation complexity*, In Proceedings STOC 2002

[FeKaSz 89] J. Friedman, J. Kahn, E. Szemeredi *On the second eigenvalue in random regular graphs*, In Proceedings of the Twenty First Annual ACM Smyposium on Theory of Computing, 1989, 587-598

[FeOf 03] U. Feige, E. Ofek, *Spectral Techniques Applied to Sparse Random Graphs*, 2003, available at
`http://wisdomarchive.wisdom.weizmann.ac.il:81/archive/00000307/`

[Fr 99] E. Friedgut *Necessary and sufficient conditions for sharp tresholds of graph properties and the k-SAT problem*, Journal of the American Mathematical Society 12, 1999, 1017-1054

[FrGo 01] J. Friedman, A. Goerdt, *Recognizing more Unsatisfiable Random 3-Sat Instances efficiently*, In Proceedings ICALP 2001, LNCS 2076, 310–321

[FuKo 80] Z. Füredi, J. Komlós, *The eigenvalues of random symmetric matrices*, Combinatorica 1, 1981, 233–241

[GoJu 02] Andreas Goerdt, Tomasz Jurdzinski. *Some Results on Random Unsatisfiable k-Sat Instances and Approximation Algorithms Applied to Random Structures.* In Proceedings MFCS 2002, LNCS 2420, 280–291

[GoKr 01] A. Goerdt, M. Krivelevich, *Efficient recognition of random unsatisfiable k-SAT instances by spectral methods*, STACS 2001, Lecture Notes in Computer Science 2010, 294–304

[Pr 94] V. V. Prasolov, *Problems and Theorems in Linear Algebra*, 1994, ISBN 0-8218-0236-4, 63-64, 93