A spectral technique for random satisfiable 3CNF formulas

Abraham Flaxman *
Carnegie Mellon University
Mathematical Sciences Department
abie@cmu.edu

Abstract

Let I be a random 3CNF formula generated by choosing a truth assignment ϕ for variables $x_1, \dots, x_n$ uniformly at random and including every clause with i literals set true by ϕ with probability p_i, independently. We show that for any constants $0 \leq \eta_2, \eta_3 \leq 1$ there is a constant d_{min} so that for all $d \geq d_{min}$ a spectral algorithm similar to the graph coloring algorithm of [3] will find a satisfying assignment with high probability for $p_1 = d/n^2$, $p_2 = \eta_2 d/n^2$, and $p_3 = \eta_3 d/n^2$. Appropriately setting the η_i's yields natural distributions on satisfiable 3CNFs, not-all-equal-sat 3CNFs, and exactly-one-sat 3CNFs.

1 Introduction

A 3CNF formula over variables $x_1, \dots, x_n$ consists of clauses $C_1, \dots, C_m$, where each clause is the disjunction of 3 literals, $C_i = \ell_{i_1} \vee \ell_{i_2} \vee \ell_{i_3}$, and each literal is a variable or the negation of a variable. A 3CNF formula is satisfiable if there is an assignment of variables to truth values so that every clause contains at least 1 true literal. Finding a satisfying assignment for a given 3CNF formula is NP-hard, so it is unlikely there is an efficient algorithm (meaning an algorithm with running time bounded by a polynomial function of the input size) which succeeds on all 3CNF formulas [10, 24]. In light of this, it is interesting to investigate efficient algorithms testing

*Supported in part by NSF VIGRE Grant DMS-9819950

for and generating satisfying assignments which work with high probability (meaning with probability tending to 1 as n goes to infinity and abbreviated **whp**) over some reasonable distribution of formulas.

Uniformly random 3CNF formulas have been the focus of extensive research. It is known that there is a sharp threshold in the ratio of clauses to variables; a random 3CNF with clause-to-variable ratio below the threshold is satisfiable **whp** and one with ratio above is not satisfiable **whp** [15]. This threshold is not known exactly, (and not even known to tend to a constant) but it is known to be at least 3.4 (see [20]) and no more than 4.5 (see [21]). Experimental results predict the higher end of this interval [7]. (Much more is known for k-CNF formulas where k is a large constant or slowly growing function [1, 18].

It is conjectured that proving the non-existence of satisfying assignments slightly above the threshold is computationally difficult, which yields nice results in hardness of approximations [11]. One piece of evidence supporting this conjecture is the exponential length of resolution-type proofs refuting such instances [9, 6]. Spectral techniques are effective in efficiently proving the unsatisfiability of formulas an $n^{1/2+\epsilon}$ factor above the threshold [16].

An alternative approach is to investigate exponential-time algorithms which find a satisfying assignment or prove that none exists for all instances. Then the challenge is to make the base of the exponent as small as possible (see, for example, [27, 19]).

1.1 The distribution

In this paper we will consider random 3CNF formulas with a "planted" satisfying assignment. That is, formulas which have a clause-to-variable ratio above the threshold, but which are generated in a way to ensure that they are satisfiable. To be exact, to form instance $I = I_{n,p_1,p_2,p_3}$ we choose a truth assignment ϕ on n variables uniformly at random and include in I each clause with exactly i literals satisfied by ϕ independently with probability p_i. By setting $p_1 = p_2 = p_3$ we obtain the model studied by Motoki and Uehara in [26], which shows a threshold of $p = \Theta(\log n/n^2)$ for 3CNF formulas to have exactly 1 satisfying assignment. An algorithm for $p_1 = p_2 = p_3$ was analyzed by Koutsoupias and Papadimitriou in [22]. They show that a greedy variable assignment rule successfully discovers a satisfying assignment **whp** for instances with $p_1 = p_2 = p_3 = \Omega(\log n/n^2))$. The authors conjecture that some modification of the greedy algorithm will work when $p_1 = p_2 = p_3 = O(1/n^2)$. The present paper shows that their conjecture is correct; in the case where $p_1 = p_2 = p_3$, the spectral phase of the algorithm presented below can be replaced by the greedy assignment rule.

By setting $p_1 = p_2$ and $p_3 = 0$, we obtain a natural distribution on 3CNFs with a planted not-all-equal assignment, a situation where the greedy variable assignment rule generates a random assignment. By setting $p_2 = p_3 = 0$, we obtain 3CNFs with a planted exactly-one-true assignment (which succumb to the greedy algorithm followed by the non-spectral steps below). Also, by correctly adjusting the ratios of p_1, p_2, and p_3, we obtain a variety of (slightly less natural) instance distributions which thwart the greedy algorithm. Carefully selected values of p_1, p_2, and p_3 are considered in [5], where it is conjectured that no algorithm running in polynomial time can solve I_{n,p_1,p_2,p_3} **whp** when $p_i = c_i\alpha/n^2$ and

$$0.077 < c_3 < 0.25 \qquad c_2 = (1-4c_3)/6$$
$$c_1 = (1+2c_3)/6 \qquad \alpha > \frac{4.25}{7}.$$

An implication the present paper is that this conjecture fails when α is a sufficiently large constant.

This paper originally appeared as an extended abstract [14], and subsequent extensions have shown that a similar styled analysis is capable of showing that alternative algorithms succeed in *expected* polynomial time [23] and on *semi-random* instances [13].

In this paper we allow clauses with repeated variables, and formulas with the same clause with the literals in a different order, so there are $8n^3$ possible clauses, $7n^3$ which are consistent with our planted assignment ϕ. The results and proofs hold for other similar models, such as prohibiting clauses with repeated literals or clauses where the same set of literals appear in a different order.

1.2 The algorithm

The main result of this paper is a polynomial time algorithm which returns a satisfying assignment to I_{n,p_1,p_2,p_3} **whp** when $p_1 = d/n^2$, $p_2 = \eta_2 d/n^2$ and $p_3 = \eta_3 d/n^2$, for $0 \le \eta_2, \eta_3 \le 1$, and $d \ge d_{min}$, where d_{min} is a function of η_2, η_3. These restrictions on η_2 and η_3 are more for convenience than necessity, and in Section 2, we will see the matrix equation which dictates the allowable range of η_i's. The unsymmetric feature of this parameterization, that there is no η_1 is not entirely for convenience, however. If we desire any $\eta_1 > \max\{\eta_2, \eta_3\}$, we can renormalize by changing the value of d. Unfortunately, taking $\eta_1 = 0$ is more complicated. The proof of correctness of Step 4 of the algorithm as decribed below relies on there being some positive fraction of clauses with 1 true literal, and it is not clear how to remove this requirement.

The algorithm below is an extension of the 3-coloring algorithm of Alon and Kahale [3]. They describe an algorithm which, in an analogous model of a random 3-colorable graph, finds a proper 3-coloring in polynomial time **whp**. A previous extension of their algorithm by Chen and Frieze [8] adapted the technique to 2-color random 3-uniform bipartite hypergraphs. We follow the same approach:

1. Construct a graph G from the 3CNF.
2. Find the most negative eigenvalue of a matrix related to the adjacency matrix of G.
3. Assign a value to each variable based on the signs of the eigenvector corresponding to the most negative eigenvalue.
4. Iteratively improve the assignment.
5. Perfect the assignment by exhaustive search over a small set containing all the incorrect variables.

We now elaborate on each step:

Step (1): Given 3CNF $I = I_{n,p_1,p_2,p_3}$, where $p_1 = \frac{d}{n^2}$, $p_2 = \eta_2 \frac{d}{n^2}$, and $p_3 = \eta_3 \frac{d}{n^2}$, the graph in step (1) $G = (V, E)$ has $2n$ vertices, corresponding to the literals in I and labeled $\{x_1, \overline{x}_1, \dots x_n, \overline{x}_n\}$. G has an edge between vertices ℓ_i and ℓ_j if I includes a clause with both ℓ_i and ℓ_j (do not add multiple edges).

Step (2): We consider $G' = (V, E')$ formed by deleting all the edges incident to vertices with degree greater than $180d$. Let A be the adjacency matrix of G'. Let λ be the most negative eigenvalue of A and $\mathbf{v}$ be the corresponding eigenvector.

Step (3): There are two assignments to consider, π_+, which is defined by

$$\pi_+(x_i) = \begin{cases} T, & \text{if } \mathbf{v}_i \geq 0; \\ F, & \text{otherwise;} \end{cases}$$

and π_-, which is defined by

$$\pi_-(x) = \neg\pi_+(x).$$

Let π_0 be the better of π_+ and π_- (that is, the assignment which satisfies more clauses). We will argue in the next section that π_0 agrees with ϕ on at least $(1 - C/d)n$ variables for some absolute constant C.

Step (4): For $i = 1, \dots, \log n$ do the following: for each variable x, if x appears in $5\epsilon d$ clauses unsatisfied by π_{i-1}, then set $\pi_i(x) = \neg\pi_{i-1}(x)$, where ϵ is an appropriately chosen constant (taking $\epsilon = 0.1$ works); otherwise set $\pi_i(x) = \pi_{i-1}(x)$.

Step (5): Let $\pi'_0 = \pi_{\log n}$ denote the final assignment generated in step (4). Let $\mathcal{A}_4^{\pi'_0}$ be the set of variables which do not appear in $(3 \pm 4\epsilon)d$ clauses as the only true literal with respect to assignment π'_0, and let $\mathcal{B}$ be the set of variables which do not appear in $(\mu_D \pm \epsilon)d$ clauses, where $\mu_D d = (3+6)d + (6+3)\eta_2 d + 3\eta_3 d + \mathcal{O}(1/n)$ is the expected number of clauses containing variable x. Form partial assignment π'_1 by unassigning all variables in $\mathcal{A}_4^{\pi'_0}$ and $\mathcal{B}$. Now, for $i \geq 1$, if there is a variable x_i which appears in less than $(\mu_D - 2\epsilon)d$ clauses consisting of variables that are all assigned by π'_i, let π'_{i+1} be the partial assignment formed by unassigning x_i in π'_i. Let π' be the partial assignment when this process terminates. Consider the graph Γ with a vertex for each variable that is unassigned in π' and an edge between two variables if they appear in a clause together. If any connected component in Γ is larger than $\log n$ fail. Otherwise, find a satisfying assignment for I by performing an exhaustive search on the variables in each connected component of Γ.

Theorem 1 *For any constants $0 \leq \eta_2, \eta_3 \leq 1$, except $(\eta_2, \eta_3) = (0, 1)$, there exists a constant d_{min} such that for any $d \geq d_{min}$, if $p_1 = d/n^2$, $p_2 = \eta_2 d/n^2$, and $p_3 = \eta_3 d/n^2$ then this polynomial-time algorithm produces a satisfying assignment for random instances drawn from I_{n,p_1,p_2,p_3}* **whp**.

The exception in this theorem is easy to circumvent superficially, since the case where $\eta_2 = 0$ is solvable in worst-case polynomial time by Gaussian elimination. However, the parameterization above obscures the fact that there are instances which appear to be hard in general when $p_2 = \sqrt{d}n$. This is the case sometimes called Gaussian elimination with noise.

1.3 Outline of what follows

We will prove Theorem 1 in the next 2 sections. Section 2 shows that **whp** the eigenvector corresponding to the most negative eigenvalue is close to a satisfying assignment, by showing that π_0 agrees with ϕ on at least $(1 - C/d)n$ clauses, where C is an absolute constant. Section 3 shows that **whp** the iterative reassignment in step (4) correctly assigns a larger fraction of variables, so large that the connected components left after unassignment are size $\mathcal{O}(\log n)$ and the exhaustive search in step (5) can perfect the assignment in polynomial time.

2 Spectral Arguments

The goal of this section is to show the assignment constructed from the eigenvector corresponding to the most negative eigenvalue of A is correct on a $1-C/d$ fraction of variables, where C is a constant independent of d. The intuition behind this result is as follows. Suppose $\phi(x) = T$. Then

- literal x appears in about $3d$ clauses with 2 false literals, $6\eta_2 d$ clauses with 1 false and 1 true literal, and $3\eta_3 d$ clauses with 2 true literals
- literal $\bar{x}$ appears in about $6d$ clauses with 1 false and 1 true literal and $3\eta_2 d$ clauses with 2 true literals.

We use these estimates to describe roughly the adjacency matrix A. For each row corresponding to a true literal, we have nonzero columns for about $6\eta_2 d + 6\eta_3 d$ true literals and $6d + 6\eta_2 d$ false literals. Similarly, for each row corresponding to a false literal, we have nonzero columns for about $6d+6\eta_2 d$ true literals and $6d$ false literals. Let $\mathbf{v}_T$ be the vector with $\mathbf{v}_T(\ell) = 1$ if ℓ is a true literal, and 0 if ℓ is a false literal, and let $\mathbf{v}_F = \mathbf{1} - \mathbf{v}_T$. Then we have

$$\begin{aligned} A\mathbf{v}_T &\approx (6\eta_2 d + 6\eta_3 d)\mathbf{v}_T + (6d + 6\eta_2 d)\mathbf{v}_F \\ A\mathbf{v}_F &\approx (6d + 6\eta_2 d)\mathbf{v}_T + 6d\mathbf{v}_F. \end{aligned}$$

So, heuristically, we expect 2 eigenvectors of A to look something like $\beta\mathbf{v}_T + \gamma\mathbf{v}_F$ where β, γ solve the 2-dimensional system

$$\begin{bmatrix} \eta_2 + \eta_3 & 1 + \eta_2 \\ 1 + \eta_2 & 1 \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \alpha \begin{bmatrix} \beta \\ \gamma \end{bmatrix}.$$

When $0 \leq \eta_2, \eta_3 \leq 1$ and $(\eta_2, \eta_3) \neq (0, 1)$, this yields a positive eigenvalue α_+ and a negative eigenvalue α_-. To see this, note that the trace of the matrix equals the sum of the eigenvalues, and then calculate that one of the eigenvalues take the form $(1+\eta_2+\eta_3+\sqrt{(1+\eta_2+\eta_3)^2 - 4(\eta_2+\eta_3-(1+\eta_2)^2)})/2$. The trace of the matrix is $1+\eta_2+\eta_3$ and this eigenvalue is strictly larger provided $\eta_2+\eta_3-(1+\eta_2)^2$ is negative. Simplifying terms shows that is equivalent to having $1+\eta_2+\eta_2^2-\eta_3 > 0$, which is the case for all $0 \leq \eta_2, \eta_3 \leq 1$ besides $(\eta_2, \eta_3) = (0, 1)$.

Let $\begin{bmatrix} \beta_+ \\ \gamma_+ \end{bmatrix}$ be the eigenvector corresponding to α_+ and $\begin{bmatrix} \beta_- \\ \gamma_- \end{bmatrix}$ be the eigenvector corresponding to α_-. Then additional calculation shows that if we normalize so $\beta_-^2 + \gamma_-^2 = 1$, then β_- and γ_- have opposite signs, and $|\beta_-|, |\gamma_-| \geq \sqrt{\frac{1}{17}}$.

Let $\mathbf{v}_+ = \beta_+\mathbf{v}_T + \gamma_+\mathbf{v}_F$ and $\mathbf{v}_- = \beta_-\mathbf{v}_T + \gamma_-\mathbf{v}_F$ denote our heuristic approximation of the eigenvectors.

We will argue that **whp** A has large positive and negative eigenvalues roughly equal to $6d$ times the $\alpha_\pm$ and all the other eigenvalues of A are smaller than $C\sqrt{d}$ in absolute value, so the assignment based on the signs of the most negative eigenvector is close to correct.

To make this precise, let $I = I_{n,p_1,p_2,p_3}$ be a random 3CNF as described above and let $G = (V, E)$ be the graph on the literals of I with edges connecting every pair of literals appearing in a common clause. Let $G' = (V, E')$ be obtained by deleting all edges of G adjacent to a vertex of degree greater than $180d$ and let A be the adjacency matrix of G'. Denote by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{2n}$ the eigenvalues of A and by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{2n}$ a corresponding collection of orthonormal eigenvectors.

Lemma 1 *There is an absolute constant C such that the following hold* **whp***:*

1. $\lambda_1 \geq (6d)\alpha_+ - 2^{-d/C}$
2. $\lambda_{2n} \leq (6d)\alpha_- + 2^{-d/C}$
3. $|\lambda_i| \leq C\sqrt{d}$ *for* $i = 2, \ldots, 2n-1$

Proof The proof is very similar to Lemma 5 of [8] and Proposition 2.1 of [3], which use the techniques of Kahn and Szemerédi from [17]. For a self-contained treatment, see also [12]

Our main tool is Rayleigh's Principle,

$$\lambda_i = \min_{L} \max_{\mathbf{v}\in L, \mathbf{v}\neq \mathbf{0}} \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}} \tag{1}$$

where L ranges over all dimension $2n - i + 1$ subspaces of $\mathbb{R}^{2n}$. (See, for example, [28]).

We partition the matrix A into 4 blocks, $A_{i,j}$, $i, j \in \{T, F\}$, where $A_{T,T}$ corresponds to the literals l with $\phi(l) = T$, and the other $A_{i,j}$ are defined similarly. The edge-set of G has corresponding partition into $E_{i,j}$ for $i, j \in \{T, F\}$.

Proposition 1 *For the edge sets defined above, the following hold* **whp***:* $|E_{T,T}| = (3\eta_3 d + 3\eta_2 d \pm o(1))n$, $|E_{T,F}| = (6\eta_2 d + 6d \pm o(1))n$, $|E_{F,F}| = (3d \pm o(1))n$, *and* $|E \setminus E'| \leq 2^{-2d/C} n$.

These all follow from standard calculations which are omitted here.

We are now ready to prove Lemma 1. To prove part (1) we apply Rayleigh's Principle with $\mathbf{v}_+ = \beta_+ \mathbf{v}_T + \gamma_+ \mathbf{v}_F$.

$$\begin{aligned}
\mathbf{v}_+^T A \mathbf{v}_+ &= \beta_+^2 2|E'_{T,T}| + 2\beta_+\gamma_+|E'_{T,F}| + \gamma_+^2 2|E'_{F,F}| \\
&\geq \beta_+^2 2(|E_{T,T}| - 2^{-2d/C} n) \\
&\quad + 2\beta_+\gamma_+(|E_{T,F}| - 2^{-2d/C} n) \\
&\quad + \gamma_+^2 2(|E_{F,F}| - 2^{-2d/C} n) \\
&\geq \beta_+^2(6\eta_2 d + 6\eta_3 d \pm o(1))n \\
&\quad + \beta_+\gamma_+(12d + 12\eta_2 d \pm o(1))n \\
&\quad + \gamma_+^2(6d \pm o(1))n \\
&\quad - 2(|\beta_+| + |\gamma_+|)^2 2^{-2d/C} n \\
&\geq \begin{bmatrix}\beta_+ & \gamma_+\end{bmatrix} \begin{bmatrix} 6d(\eta_2+\eta_3) & 6d(1+\eta_2) \\ 6d(1+\eta_2) & 6d \end{bmatrix} \begin{bmatrix}\beta_+ \\ \gamma_+\end{bmatrix} n \\
&\quad - 3(|\beta_+| + |\gamma_+|)^2 2^{-2d/C} n \\
&\geq (6d)\alpha_+ n - 2^{-d/C} n
\end{aligned}$$

Since $\mathbf{v}_+^T \mathbf{v}_+ = n$ we conclude that $\lambda_1 \geq (6d)\alpha_+ - 2^{-d/C}$.

To prove part (2) of the lemma, we apply Rayleigh's Principle with $L = \{t\mathbf{v}_+ \colon t \in \mathbb{R}\}$. The calculation is very similar to the one above, and is omitted.

Proving (3) takes more work. Fortunately we can use a reduction very similar to Lemma 5(iii) in [8]. Recall that $\mathbf{v}_T$ is the vector with $\mathbf{v}_T(\ell) = 1$ if literal ℓ is true and 0 otherwise. Also, recall that $\mathbf{v}_F = \mathbf{1} - \mathbf{v}_T$. We begin by showing

Proposition 2 *For any* $\mathbf{v}$ *with* $\mathbf{v}^T\mathbf{v} = 1$, $\mathbf{v}^T\mathbf{v}_T = 0$, *and* $\mathbf{v}^T\mathbf{v}_F = 0$, *we have* $|\mathbf{v}^T A\mathbf{v}| \le C\sqrt{d}$.

Proof Observe that the entries of $A_{F,F}$ are 1 independently with probability $1-(1-p_1)^{3n} \sim 3d_1/n$. Unfortunately, all the other $A_{i,j}$ have dependencies because the edges are added a triangle at a time. To work around this, for each clause, we randomly color the edges of the triangle corresponding to the clause, one edge red, one green, and one blue. We add each to the appropriately colored graph G^r, G^g, or G^b (but add an edge to at most 1 graph). Note that the edges of a particular G^c appear independently, as each edge is contributed by a different clause. Let A^c be the adjacency matrix of G^c for $c \in \{r, g, b\}$. Note that $A = A^r + A^g + A^b$. Let $A^c_{i,j}$ for $i, j \in \{T, F\}$ be the submatricies of A^c corresponding to the submatricies of A defined above. Then we have

$$\begin{aligned}|\mathbf{v}^T A\mathbf{v}| &\le \sum_{c\in\{r,g,b\}} |\mathbf{v}^T A^c\mathbf{v}| \\ &\le \sum_{c\in\{r,g,b\}} \sum_{i,j\in\{T,F\}} |\mathbf{v}_i^T A^c_{i,j}\mathbf{v}_j|.\end{aligned}$$

The edges of $A^c_{i,j}$ occur with different probabilities for different combinations of i, j, but, provided we have made $d_{min}(\eta_2, \eta_3)$ large enough, all submatricies with non-zero edge probabilities have edge probabilities exceeding D/n and we can use an argument identical to Lemma 2.4 of [3] (except with $5d$ changed to $180d$) to show that any unit vectors $\mathbf{v}$ with $\mathbf{v}^T\mathbf{v}_T = 0$ and $\mathbf{v}^T\mathbf{v}_F = 0$ has $|\mathbf{v}_i^T A^c_{i,j}\mathbf{v}_j| \le (C/12)\sqrt{d}$, which implies $|\mathbf{v}^T A\mathbf{v}| \le C\sqrt{d}$. □

We also need

Proposition 3 *The following hold* **whp**

$$\|(A-(6d)\alpha_+ I)\mathbf{v}_+\|^2 \le d\|\mathbf{v}_+\|^2,$$
$$\|(A-(6d)\alpha_- I)\mathbf{v}_-\|^2 \le d\|\mathbf{v}_-\|^2.$$

Proof We work with $\mathbf{v}_+$, as the bound on $\mathbf{v}_-$ is calculated analogously.

Setting $\mathbf{y} = (A-(6d)\alpha_+ I)\,\mathbf{v}_+$, we see that

$$\begin{aligned}\mathbf{y}^T\mathbf{y} &= \mathbf{v}_+^T(A-(6d)\alpha_+ I)^T(A-(6d)\alpha_+ I)\mathbf{v}_+\\ &= \mathbf{v}_+^T A^2\mathbf{v}_+ - 2(6d)\alpha_+\mathbf{v}_+^T A\mathbf{v}_+ + ((6d)\alpha_+)^2\mathbf{v}_+^T\mathbf{v}_+.\end{aligned}$$

We know that, **whp**, $\mathbf{v}_+^T A\mathbf{v}_+ \ge (6d)\alpha_+ n - 2^{-d/C}n$ from the proof of Lemma 1 (1). By a similar calculation, we have $\mathbf{v}_+^T A\mathbf{v}_+ \le (6d)\alpha_+ n + 2^{-d/C}n$. To complete the proposition, we calculate that, **whp**, $\mathbf{v}_+^T A^2\mathbf{v}_+ = ((6d)\alpha_+)^2 n \pm 2^{-d/C}n$. (To see this, write $\mathbf{v}_+$ as a linear combination of the eigenvectors of A, and note that the coefficient of the eigenvector corresponding to eigenvalue $(6d)\alpha_+$ must have most of the weight in the sum in order for $\mathbf{v}_+^T A\mathbf{v}_+$ to be close to $(6d)\alpha_+$). Summing up, we see that **whp**

$$\begin{aligned}\mathbf{y}^T\mathbf{y} &= ((6d)\alpha_+)^2 n \pm 2^{-d/C}n - 2(6d)\alpha_+((6d)\alpha_+ n \pm 2^{-d/C}n) + ((6d)\alpha_+)^2 n\\ &= \pm 3\cdot 2^{-d/C}n.\end{aligned}$$

□

We can now complete the lemma. To show $\lambda_2 \le (C+2)\sqrt{d}$, we apply Rayleigh's Principle with $L = \{\mathbf{v}\colon\ \mathbf{v}^T\mathbf{v}_+ = 0\}$. Then we write $\mathbf{v} \in L$ as $t\mathbf{v}_- + \mathbf{w}$, where $\mathbf{w}^T\mathbf{v}_+ = 0$ and $\mathbf{w}^T\mathbf{v}_- = 0$. So

$$\begin{aligned}\mathbf{v}^T A\mathbf{v} &= t^2\mathbf{v}_-^T A\mathbf{v}_- + 2t\mathbf{w}^T A\mathbf{v}_- + \mathbf{w}^T A\mathbf{w}\\ &= t^2\mathbf{v}_-^T A\mathbf{v}_- + 2t\mathbf{w}^T(A-(6d)\alpha_- I)\mathbf{v}_- + \mathbf{w}^T A\mathbf{w}\\ &\le t^2((6d)\alpha_- + 2^{-d/C})\|\mathbf{v}_-\|^2 + 2t\sqrt{d}\|\mathbf{w}\|\|\mathbf{v}_-\|\\ &\qquad + C\sqrt{d}\|\mathbf{w}\|^2\\ &\le (C+2)\sqrt{d}\|\mathbf{v}\|^2,\end{aligned}$$

where the final inequality follows from $\alpha_- < 0$, $\|\mathbf{w}\| \le \|\mathbf{v}\|$, and $t\|\mathbf{v}_-\| \le \|\mathbf{v}\|$.

To show $\lambda_{2n-1} \ge -C\sqrt{d}$, we let L be any 2 dimensional subspace of $\mathbb{R}^{2n}$. If L is spanned by $\{\mathbf{v}_-, \mathbf{v}_+\}$, we take the $\mathbf{v}$ in Rayleigh's Principle to be

$\mathbf{v} = \mathbf{v}_+$, and we have $\max_{\mathbf{v}\in L, \mathbf{v}\neq \mathbf{0}} \mathbf{v}^T A\mathbf{v}/(\mathbf{v}^T\mathbf{v}) > 0$. Otherwise, there is some nonzero vector $\mathbf{v} \in L$ that is orthogonal to $\mathbf{v}_+$ and $\mathbf{v}_-$, and so Proposition 2 shows that $\max_{\mathbf{v}\in L, \mathbf{v}\neq \mathbf{0}} \mathbf{v}^T A\mathbf{v}/(\mathbf{v}^T\mathbf{v}) \geq -C\sqrt{d}$. □

We conclude the section by proving

Lemma 2 *Let λ_{2n} be the most negative eigenvalue of A and let $\mathbf{v}_{2n}$ be the corresponding eigenvector. Let $\mathbf{v}_- = \beta_- \mathbf{v}_T + \gamma_- \mathbf{v}_F$ as above. Then the sign of $\mathbf{v}_{2n}$ or $-\mathbf{v}_{2n}$ disagrees with the sign of $\mathbf{v}_-$ on at most $(C/d)n$ coordinates.*

Proof Expand $\mathbf{v}_-$ as a linear combination of orthonormal eigenvectors of A, so that we have $\mathbf{v}_- = \sum_{i=1}^{2n} c_i \mathbf{v}_i$. Then

$$((6d)\alpha_- I - A)\mathbf{v}_- = \sum_{i=1}^{2n} ((6d)\alpha_- - \lambda_i) c_i \mathbf{v}_i$$

and

$$\begin{aligned}
\|((6d)\alpha_- I - A)\mathbf{v}_-\|^2 &= \sum_{i=1}^{2n} c_i^2 ((6d)\alpha_- - \lambda_i)^2 \\
&\geq c_1^2((6d)\alpha_- - \lambda_1)^2 + \sum_{i=2}^{2n-1} c_i^2((6d)\alpha_- - C\sqrt{d})^2 \\
&\geq ((3d)\alpha_-)^2 \sum_{i=1}^{2n-1} c_i^2,
\end{aligned}$$

since $\alpha_- < 0$, $\lambda_1 > 0$, and $\lambda_i < C\sqrt{d}$ for $i = 2, \ldots, 2n-1$.

We know from Proposition 3 above that

$$d\|\mathbf{v}_-\|^2 \geq \|((6d)\alpha_- I - A)\mathbf{v}_-\|^2,$$

so

$$\sum_{i=1}^{2n-1} c_i^2 \leq \frac{d}{((3d)\alpha_-)^2}\|\mathbf{v}_-\|^2 \leq \frac{1}{(9d)\alpha_-^2}\|\mathbf{v}_-\|^2 = \frac{1}{(9d)\alpha_-^2} n.$$

Let $\tilde{\mathbf{v}} = \sum_{i=1}^{2n-1} c_i \mathbf{v}_i$, and we have $c_{2n}\mathbf{v}_{2n} = \mathbf{v}_- - \tilde{\mathbf{v}}$. Each entry of $\mathbf{v}_-$ is at least $\sqrt{\frac{1}{17}}$ in absolute value, so $c_{2n}\mathbf{v}_{2n}(\ell)$ may have sign opposite of $\mathbf{v}_-(\ell)$ for at most $\frac{17}{(9d)\alpha_-^2} n$ coordinates. □

Corollary 1 *After step (3) at least $(1 - C/d)n$ variables are set correctly.*

3 Non-spectral Arguments

This section completes the main theorem by analyzing steps (4) and (5) of the algorithm. We choose d_{min} large enough so the truth assignment π produced in step (3) is correct on all but δn variables, where δ is a sufficiently small constant (like 0.001).

To simplify the following discussion we make a few definitions; in the following ψ is a partial truth assignment. Recall that ϕ is the satisfying assignment we used to generate the instance I.

- We say variable x *supports clause* C *with respect to assignment* ψ if x is the only true literal in C with respect to ψ or $\overline{x}$ is the only true literal in C with respect to ψ.
- Let $\mathcal{A}_k$ be the set of variables x such that there are $(3 \pm k\epsilon)d$ clauses which x supports with respect to ϕ. Here ϵ is a sufficiently small constant (eg. $\epsilon = 0.1$).
- Let $\mathcal{B}$ be the set of variables x such that x appears in $(\mu_D \pm \epsilon)d$ clauses, where $\mu_D d$ is the expected number of clauses containing x, which is $(3+6)d + (6+3)\eta_2 d + 3\eta_3 d + \mathcal{O}(1/n)$.

We will be concerned with $\mathcal{A}_1$ and $\mathcal{A}_4$, so we may think of $\mathcal{A}_1$ as the variables that support "about the right number of clauses" and $\mathcal{A}_4$ as the variables that support "almost about the right number of clauses" (with respect to ϕ).

We now list some useful properties as in [8] which hold for I **qs**[1]:

Useful Property 1 $|\mathcal{A}_1 \cap \mathcal{B}| \geq n(1 - e^{-d/C})$.

Useful Property 2 There is no subset of variables U such that $|U| \leq 2\delta n$ and at least $\frac{1}{9}\epsilon d|U|$ clauses contain two variables from U.

These follows from standards calculations which are omitted here.

Now we show step (4) improves the assignment found in step (3).

Lemma 3 *After step (4) at least* $(1 - 2^{-d/C})n$ *variables are set correctly* **whp**.

[1]We say a sequence of events $\mathcal{E}_n$ holds *quite surely* (**qs**) if the probability $\Pr[\mathcal{E}_n] = o(n^{-C})$ for any constant C.

Proof Define the set of variables H as follows:

1. Let $H_1 = \mathcal{A}_1 \cap \mathcal{B}$. Let B be the remaining variables.
2. While there is a variable $a_i \in H_i$ which is in less than $(\mu_D - 2\epsilon)d$ clauses with only variables in H_i, define H_{i+1} to be $H_i \setminus \{a_i\}$.
3. Let a_m be the last variable removed in step (2) and let $H = H_m$.

Proposition 4 $|H| \geq (1 - 2^{-d/C})n$ **qs**.

Proof Useful Property 1 shows that $|H_1| \geq (1 - e^{-d/C})n$ **qs**. Suppose that $m \geq m_0 = e^{-d/C}n$. Let $U = \{a_1, \ldots, a_{m_0}\} \cup B$. Each a_i appears at least $(\mu_D - \epsilon)d$ clauses but at most $(\mu_D - 2\epsilon)d$ clauses with only variables in H, so for each a_i there must be at least ϵd clauses containing a_i and some other variable of U. But each clause can account for at most 3 of the $\epsilon m_0 d$ pairs, so the total number of clauses containing two variable from U is at least $\frac{1}{3}\epsilon|U|/2$, contradicting Useful Property 2.

Therefore, $|H| \geq (1 - 2e^{-d/C})n \geq (1 - 2^{-d/C})n$ **qs**. □

Proposition 5 *Let B_i be the incorrectly assigned variables in H at the i-th iteration of step (4). Then $|B_i| \leq |B_{i-1}|/2$* **qs**.

Proof We will assume not and use Useful Property 2 to derive a contradiction. We know $|B_0| \leq \delta n$ **qs** because step (3) works. If $x \in B_i$, there are 2 cases to consider.

If $x \in B_{i-1}$ then $\pi_{i-1}(x) = \pi_i(x)$ so x appears in at most $5\epsilon d$ clauses unsatisfied by π_{i-1}. But $x \in H$, so there are at least $(3 - \epsilon)d$ clauses which x supports with respect to ϕ, so at least $(3 - 6\epsilon)d$ of these clauses contain some other variable y with $\pi_{i-1}(y) \neq \phi(y)$. Also because $x \in H$, x appears in at most $(\mu_D + \epsilon)d$ clauses, at least $(\mu_D - 2\epsilon)d$ of which include only variables also in H. So x appears in at most $3\epsilon d$ clauses with variables not in H. This means of the $(3-6\epsilon)d$ clauses containing another variable which is assigned incorrectly by π_{i-1}, at least $(3 - 9\epsilon)d$ of them contain a variable in B_{i-1}.

If $x \notin B_{i-1}$ then, since it is in B_i, it must be in $5\epsilon d$ clauses which are unsatisfied with respect to π_{i-1}. Since all these clauses are satisfied with respect to ϕ, they must each contain some variable y which has $\pi_{i-1}(y) \neq \phi(y)$. But x is in at most $3\epsilon d$ clauses with variables outside of H, so x is in at least $2\epsilon d$ clauses with some variable in B_{i-1}.

In either case, every variable in B_i appears in at least $2\epsilon d$ clauses with some variable of B_{i-1}. Setting $U = B_i \cup B_{i-1}$, we have at least $\frac{1}{3}2\epsilon d|B_i|$ clauses containing two variables from U. If $|B_i| \geq |B_{i-1}|/2$ then the bound on number of clauses above exceeds $\frac{4}{9}\epsilon d|U|$, which contradicts Useful Property 2. □

This shows that **whp** all literals in H are assigned correctly in $\log n$ iterations, which completes the proof of Lemma 3. □

Lemma 4 *After unassignment in step (5) all variables in H remain assigned and no variable which remains assigned is assigned incorrectly* **whp***.*

Proof *All variables in H remains assigned:* all $x \in H$ are assigned correctly at the end of step (4), and there are at most $3\epsilon d$ clauses containing x and variables outside of H, so there are at least $(3-4\epsilon)d$ clauses (the ones in H) which x supports and no more than $(3+4\epsilon)d$ clauses which x supports. In addition, we know $H \subseteq \mathcal{A}_4$, so all $x \in H$ remain assigned in π'_1. To see that no x is unassigned in later π'_i, note that for $x \in H$, x is in at least $(\mu_D - 2\epsilon)d$ clauses consisting only of other variables in H.

Any variable still assigned after unassignment is assigned correctly: Let U be the set of variables that are assigned incorrectly after unassignment. Suppose $x \in U$. Then x appears in at most $(\mu_D + \epsilon)d$ clauses, of which at most $3\epsilon d$ contain an unassigned variable. Also, x supports at least $(3-4\epsilon)d$ clauses, so x supports at least $(3-7\epsilon)d$ clauses containing no unassigned variables. In the correct assignment, x is opposite its current value and all the clauses are satisfied, so each of these $(3-7\epsilon)d$ assigned clauses has some other assigned variable set incorrectly. Thus x appears in $(3-7\epsilon)d$ clauses with some other variable from U. Since each clause can account for at most 3 such pairs, we have at least $\frac{1}{3}(3-7\epsilon)d|U|$ clauses containing two variables of U. $|U| \leq 2^{-d/C}n$ so this contradicts Useful Property 2. □

For the final piece of the argument, consider the graph Γ with a vertex for each variable and an edge between two unassigned variables if they appear in a common clause. We will show Γ has connected components of size at most $\log n$ **whp**. This is proved similarly to Proposition 4 of [8]. The argument is based on a calculation of the expected number of $(\log n)$-sized trees covered by the clauses of I that are disjoint from H.

Lemma 5 *No connected component of* Γ *has size larger than* $\log n$ **whp**.

Proof Let T' be a fixed tree on $\log n$ vertices, and let T be a fixed collection of clauses such that each edge of T' appears in some clause of T. We call T minimal if deleting any clause results in a set which does not cover T'. Let $V(T)$ denote the set of variables appearing in some clause of T and $V(T')$ denote the set of variables appearing in T'. We wish to show that $\Pr[T \subseteq I \text{ and } V(T') \cap H = \emptyset]$ is small. Let J be the subset of variables of $V(T')$ which appear in at most 6 clauses of T.

Proposition 6 $|J| \geq |V(T')|/2$

Proof Suppose $|J| < |V(T')|/2$. Then at least $|V(T')|/2$ variables appear in more than 6 clauses of T. So $|T| \geq \frac{1}{3} \cdot 6 \cdot |V(T')|/2 = |V(T')|$. But since T is minimal, each clause of T covers at least 1 unique edge of T', so $|T| \leq |V(T')| - 1$. Contradiction. □

We define the set of variables H' by the following iterative procedure (which is similar to the procedure we used to generate H, but depends on $V(T) \setminus J$):

1. Set H'_1 to be the set of variables x such that x supports at least $(3-\epsilon)d$ clauses and at most $(3+\epsilon)d - 6$ clauses with respect to ϕ, x appears in at least $(\mu_D - \epsilon)d$ clauses and at most $(\mu_D + \epsilon)d - 6$ clauses, and x is not in $V(T) \setminus J$
2. While there exists x_i appearing in less than $(\mu_D - 2\epsilon)d$ clauses with only variables from H'_i, set $H'_{i+1} = H'_i \setminus \{x_i\}$.
3. Set H' to H'_m, the final result of the previous step.

Proposition 7 *Let F be a set of clauses and let $H(F \cup T)$ be the value of H if $I = F \cup T$ and let $H'(F)$ be the value of H' if $I = F$. Then* $H'(F) \subseteq H(F \cup T)$.

Proof First, we argue that $H'_1(F) \subseteq H_1(F \cup T)$. For $x \notin H_1(F \cup T)$ consider the following cases:

1. If x appears in more than $(\mu_D + \epsilon)d$ clauses of $F \cup T$ or supports more than $(3+\epsilon)d$ clauses of $F \cup T$ with respect to ϕ then it is not included in $H_1(F \cup T)$; we argue x is also not in $H'_1(F)$ by examining 2 cases:
   (a) $x \in V(T) \setminus J$. Then x is not included in $H'_1(F)$.
   (b) $x \notin V(T) \setminus J$. Then x appears in most 6 clauses of T, so it appears in more than $(\mu_D + \epsilon)d - 6$ clauses of F or x supports more than $(3 + \epsilon)d - 6$ clauses of F with respect to ϕ and hence is not included in $H'_1(F)$.
2. If x appears in less than $(\mu_D - \epsilon)d$ clauses of $F \cup T$ or supports less than $(3-\epsilon)d$ clauses of $F \cup T$ with respect to ϕ then, since it appears in no more clauses of F and supports no more clauses of F with respect to ϕ, it is not included in $H_1(F \cup T)$ or $H'_1(F)$.

We proceed by showing that if $H'_i(F) \subseteq H_i(F \cup T)$ then $H'_{i+1}(F) \subseteq H_{i+1}(F \cup T)$: if x_i appears in less than $(\mu_D - 2\epsilon)d$ clauses of $F \cup T$ with only variables of $H_i(F \cup T)$ then it also appears in less than $(\mu_D - 2\epsilon)d$ clauses of F with only variables of $H'_i(F)$.

Thus, we conclude that $H'(F) \subseteq H(F \cup T)$. □

Proposition 8 $\Pr[T \subseteq I \textit{ and } V(T') \cap H = \emptyset] \leq \Pr[T \subseteq I] \Pr[J \cap H' = \emptyset]$

Proof It is sufficient to show that

$$\Pr[J \cap H = \emptyset \mid T \subseteq I] \leq \Pr[J \cap H' = \emptyset].$$

We do this now:

$$\begin{aligned} \Pr[J \cap H' = \emptyset] &= \sum_{F:\ J \cap H'(F) = \emptyset} \Pr[I = F] \\ &\geq \sum_{F:\ J \cap H(F \cup T) = \emptyset} \Pr[I = F], \end{aligned}$$

where the inequality follows from $H'(F) \subseteq H(F \cup T)$. Now, we break each set of clauses F into $F' = F \setminus T$ and $F'' = F \cap T$. We rewrite the value above as

$$
\begin{aligned}
&\sum_{F:\ J\cap H(F\cup T)=\emptyset} \Pr[I=F] \\
&= \sum_{\substack{F':\ F'\cap T=\emptyset,\\ J\cap H(F'\cup T)=\emptyset}} \ \sum_{F'':\ F''\subseteq T} \Pr\left[I\setminus T=F' \wedge I\cap T=F''\right] \\
&= \left(\sum_{\substack{F':\ F'\cap T=\emptyset,\\ J\cap H(F'\cup T)=\emptyset}} \Pr[I\setminus T=F']\right)\left(\sum_{F'':\ F''\subseteq T} \Pr[I\cap T=F'']\right) \\
&= \sum_{\substack{F':\ F'\cap T=\emptyset,\\ J\cap H(F'\cup T)=\emptyset}} \Pr[I\setminus T=F'] \\
&= \sum_{\substack{F':\ F'\cap T=\emptyset,\\ J\cap H(F'\cup T)=\emptyset}} \Pr[I\setminus T=F' \mid T\subseteq I] \\
&= \Pr[J\cap H=\emptyset \mid T\subseteq I].
\end{aligned}
$$

□

Proposition 9 $\Pr[J\cap H'=\emptyset]\le 2n^{-d/2C}$

Proof Although H' is formed by a complicated iterative procedure, this procedure does not depend in any way on J, and so the probability that J does not intersect H' is the same as the probability that any set of size $j=|J|$ does not intersect H'. Conditioned on $|H'|$, this is given by

$$
\Pr\left[J\cap H'=\emptyset \mid |H'|=h\right]=\binom{n-j}{h}\Big/\binom{n}{h}\le(n-h)^j.
$$

It follows from the same arguments as in Proposition 4 that $|H'|>(1-2^{-d/C})n$ **qs**.

Therefore the unconditional probability that $J\cap H'=\emptyset$ is at most $2^{-jd/C}+n^{-d/2C}$. Since $j=|J|\ge|V(T')|/2=(\log n)/2$, the desired bound on the probability holds. □

Let $k=\log n$, and let $N_{T',s}$ denote the number of ways to pair $2s$ edges of T' to form s clauses which each cover 2 edges. Since there are 6 ways to permute the order of the variables in each clause, 8 ways to set the negations, and

$k-1-s$ clauses total, it follows that there are at most $N_{T',s}(48n)^{k-1-2s}$ ways to cover tree T' with a minimial set of clauses such that s clauses cover 2 edges and $k-1-2s$ clauses cover 1 edge. Let T be such a set of clauses. Then we have $\Pr[T \subseteq I] = (d/n^2)^{k-1-s}$. Above we showed that $\Pr[J \cap H'] \le e^{-k(d/2C)}$. Thus, the probability that the random instance I contains a set of clauses which cover a k-tree that is disjoint from H is at most

$$\sum_{k\text{-trees } T'} \sum_{s=0}^{k/2} N_{T',s}(48n)^{k-1-2s} \left(d/n^2\right)^{k-1-s} e^{-k(d/2C)}$$
$$\le \sum_{k\text{-trees } T'} \left(\sum_{s=0}^{k/2} N_{T',s} \right) (48d)^k n^{1-k} e^{-k(d/2C)}$$

To obtain useful upper bounds on the sum $\left(\sum_s N_{T',s}\right)$, we fix a degree sequence $(d_1, \ldots, d_k)$ for T', and consider the following procedure for pairing edges so that triangles can cover the edge pairs. For each vertex, we specify a permutation of the edges incident to that vertex. Then we iterate through the vertices, and for each vertex, we iterate through the edges and pair up each unpaired edge with the edge given by the permutation associated with the current vertex (and leave the edge unpaired if the permutation sends the edge to itself). Any pairing of edges which can be covered by clauses can be generated this way by choosing the permutations to transpose each pair of edges to be covered by a single clause and to leave fixed all the other edges. Since there are $d_i!$ different permutations for vertex i, we have

$$\sum_{s=0}^{k/2} N_{T',s} \le \prod_{i=1}^{k} (d_i!)\,.$$

Prüfer codes give a bijection between the set $[k]^{k-2}$ and labeled trees on k vertices. They have the additional nice property that the degree of vertex i in the tree corresponding to code $c \in [k]^{k-2}$ is exactly 1 less than the number of times i appears in c. It follows that the number of k-trees with degree sequence $(d_1, \ldots, d_k)$ equals $\binom{k-2}{d_1-1,\ldots,d_k-1}$ (see, for example, [25, Section 4.1, p. 33]). There are $\binom{n}{k}$ ways to choose the k vertices of the tree. So the probability above is at most

$$\sum_{d_1+\ldots+d_k=2(k-1)} \binom{k-2}{d_1-1,\ldots,d_k-1}\binom{n}{k}\left(\prod_{i=1}^{k}(d_i!)\right)(48d)^k n^{1-k} e^{-k(d/2C)}$$

$$\leq \sum_{d_1+\ldots+d_k=2(k-1)} k^2 e^k \left(\prod_{i=1}^{k} d_i\right)(48d)^k n e^{-k(d/2C)}.$$

For $(d_1,\ldots,d_k)$ with $d_1+\ldots+d_k = 2(k-1)$, the product $\prod_{i=1}^{k} d_i$ is maximized when $d_1 = \ldots = d_k$ and so $\prod_{i=1}^{k} d_i < 2^k$. The number of ways to choose positive integers $(d_1,\ldots,d_k)$ so that $d_1+\ldots+d_k = 2(k-1)$ is less than $\binom{2k-1}{k-1}$, which is less than 2^{2k}. So, provided we have chosen the constant d sufficiently large, we find that the probability that I contains a set of clauses which covers a $(\log n)$-tree disjoint from H is at most

$$2^{2\log n}(\log n)^2 e^{\log n} 2^{\log n}(48d)^{\log n} n e^{-\log n(d/2C)} = o(1).$$

$\square$

4 Acknowledgments

Thanks to Alan Frieze for recommending average case analysis of 3SAT as an interesting area of study and encouraging me to pursue the spectral approach. Thanks also to Luis von Ahn for helpful comments and discussion.

References

[1] D. Achlioptas and Y. Peres, The threshold for random k-SAT is $2^k \log 2 - O(k)$, *Proc. of the 35th ACM STOC* (2003) 223-231.

[2] M. Alekhnovich, More on average case vs approximation complexity, *Proc. of the 44rd IEEE FOCS* (2003) 298-307.

[3] N. Alon and N. Kahale, A Spectral Technique for Coloring Random 3-Colorable Graphs, *DIMACS TR-94-35* (1994).

[4] N. Alon and J. Spencer, *The Probabilistic Method*, John Wiley & Sons (2000).

[5] W. Barthel, A. K. Hartmann, M. Leone, F. Ricci-Tersenghi, M. Weigt, and R. Zecchina, Hiding solutions in random satisfiability problems: A statistical mechanics approach, *Phys. Rev. Lett. 88*, 188701 (2002)

[6] E. Ben-Sasson and A. Wigderson, Short proofs are narrow – Resolution made simple, *J. Assoc. Comp. Mach.* (1999) 517-526.

[7] J. Crawford and L. Auton. Experimental results on the crossover point in random 3-SAT, *Artificial Intelligence* 81 (1996) 31-57.

[8] H. Chen and A. Frieze, Coloring Bipartite Hypergraphs, *Proc. 5th IPCO* (1996) 345-358.

[9] V. Chvátal and E. Szemerédi, Many hard examples for resolution, *J. Assoc. Comp. Mach.* (1988) 759-768.

[10] S. Cook, The Complexity of theorem-proving procedures, *Proc. 3rd FOCS* (1971) 151-158.

[11] U. Feige, Relations between average case complexity and approximation complexity, *Proc. 34th ACM STOC* (2002).

[12] U. Feige and E. Ofek, Spectral techniques applied to sparse random graphs, *Random Structures and Algorithms*, 27 (2) 251-275.

[13] U. Feige and D. Vilenchik, A local search algorithm for 3SAT, *Technical Report MCS04-07 of the Weizmann Institute*, 2004.

[14] A. D. Flaxman, A spectral technique for random satisfiable 3CNF formulas, *Proc. 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* (2003) 357-363.

[15] E. Friedgut, Necessary and sufficient conditions for sharp thresholds of graph properties, and the k-SAT problem. *J. Amer. Math. Soc. 12* (1999) 1017-1054.

[16] J. Friedman and A. Goerdt, Recognizing More Unsatisfiable Random 3-SAT Instances Efficiently, *Proc. 28th ICALP* (2001) 310-321

[17] J. Friedman, J. Kahn, and E. Szemerédi, On the second eigenvalue in random regular graphs, *Proc. 21st ACM STOC* (1989) 587-598.

[18] A. M. Freize and N. C. Wormald, Random k-SAT: a tight threshold for moderately growing k. In *Proc. 5th International Symposium on Theory and Applications of Satisfiability Testing* (2002) 1-6.

[19] T. Hofmeister, W. Schöning, R. Schuler, and O. Watanabe, A Probabilistic 3-SAT Algorithm Further Improved, *Proc. of the 19th Annual Symposium on Theoretical Aspects of Computer Science (STACS '02)* LCNS 2285 (2002) 192-202.

[20] A. Kaporis, L. Kirousis, and E. Lalas, The probabilistic analysis of a greedy satisfiability algorithm, *Proc. 5th Satisfiability Testing Workshop* (2002) 362-276.

[21] A. Kaporis, L. Kirousis, Y. Stamatiou, M. Vamvakari, M. Zito, Coupon Collectors, q-Binomial Coefficients and the Unsatisfiability Threshold. *Proc. 7th ICTCS* (2001) 328-338

[22] E. Koutsoupias and C. Papadimitriou, On the greedy algorithm for satisfiability, *Info. Proc. Let.*, 43 (1992) 53-55.

[23] M. Krivelevich and D. Vilenchik, Solving random satisfiable 3CNF formulas in expected polynomial time, to appear in *Proc. 17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* (2003).

[24] L. Levin, Universal search problems, *Prob. Info. Trans.* 9 (1973) 265-266.

[25] L. Lovász, *Combinatorial problems and exercises*, 2nd Ed., Elsevier Science Publishers, Amsterdam, 1993.

[26] M. Motoki and R. Uehara, Unique Solution Instance Generation for the 3- Satisfiability (3SAT) Problem, *IEICE Technical Report, COMP98-54* (1998) 25-32.

[27] U. Schoning, A Probabilistic Algorithm for k-SAT and Constraint Satisfaction Problems, *Proc. 40th Symp. on Foundations of Computer Science* (1999) 410-414

[28] G. Strang, *Linear algebra and its applications*, Hardcourt Brace Jovanovich Publishing (1988).