STAT 206A: Gibbs Measures	Elchanan Mossel
Lecture 8	
Lecture date: Sep 21	*Scribe: Arnab Sen*

Today we will sketch the analysis of unit clause propagation algorithm for 3-satisfiability problem in more detail to get a lower bound of the satisfiability threshold. Before going to the proof, let us rephrase the algorithm so that it becomes more amenable to analysis.

1 The Unit Clause Propagation Algorithm (UC)

Notation: For any set V of Boolean variables we $L(V)$ will denote the set of $2|V|$ literals on the variables in V. Given a literal l, $\mathrm{var}(l)$ denotes its underlying variable and $\bar{l}$ is the complementary literal of l.

Input: A set of N Boolean variables $V = \{x_1, x_2, \cdots, x_N\}$ and a SAT formula ψ on V.

1. (forced step) If there is any 1-clauses choose a 1-clause l uniformly at random from them. Set $V \leftarrow V \setminus \{\mathrm{var}(l)\}$.
2. (free step) Else choose l uniformly at random from $L(V)$. Set $V \leftarrow V \setminus \{\mathrm{var}(l)\}$.
3. Remove all clauses containing l. Remove all occurrences of $\bar{l}$ from all the clauses.
4. Repeat unless V is empty.

Output: If there is no 0-clauses, then ψ is SAT. Otherwise cannot determine.

Let $V(t)$ be the (random) set of unassigned variables at time t and $\mathcal{C}_i(t)$ be the (random) set of i-clauses remaining at time t, $0 \le t \le N,\ i = 0, 1, 2, 3$. Denote the size of $\mathcal{C}_i(t)$ by $c_i(t)$. Clearly, $|L(V(t))| = 2|V(t)| = 2(N - t)$. Note that we make the algorithm run for exactly N steps to simplify our analysis.

We will show that for the $\mathrm{SAT}_N(3, \alpha)$ ensemble, if $\alpha < \frac{8}{3}$, the probability that the *Unit Clause Propagation (UC)* algorithm generates no 0-clauses until some time $t_e \equiv t_e(N)$, remains bounded away from zero and also the remaining $N - t_e$ unassigned variables can be dealt with quite easily so that finally we reach a SAT assignment with nontrivial probability as $N \to \infty$.

Claim 1: For any $t \in \{0, 1, \cdots, N\}$, each clause in $\mathcal{C}_i(t)$ is uniformly distributed among the i-clauses over the literals in $L(V(t))$, $i = 1, 2, 3$.

By induction on t. To see the induction step note that the literal at time t is chosen uniformly at random from $L(V(t-1))$. For free step it is obvious and for forced step observe that $\mathcal{C}_1(t-1)$ is a set of randomly chosen 1-clauses over $V(t-1)$. $\square$

Claim 2: Fix a small $\varepsilon > 0$ and take $t_e = (1-\varepsilon)N$. If w.h.p.(with high probability) $\sum_{t=0}^{t_e-1} c_1(t) < MN$ and $c_1(t) < \varepsilon^2 N$ for some constant M, then $\mathbf{P}(c_0(t_e) = 0) > \Omega(1)$.

If $c_1(t-1) = 0$ then no 0-clause can be generated in step t since every clauses loses at most one literal in each step. If $c_1(t-1) > 0$ then in step t UC picks some literal $l \in \mathcal{C}_1(t-1)$ at random and satisfy it. Clearly, if $\bar{l} \notin \mathcal{C}_1(t-1)$ then no 0-clause will be generated at step t. Hence, by uniform randomness, we get that the conditional on $c_1(t-1) = a > 0$, the probability that no 0-clause is generated in step t is given by $(1 - \frac{1}{2(n-t)})^{a-1}$. So, at any time $t \le t_e$, probability of not getting an 0-clause is at least $(1 - \frac{1}{2\varepsilon n})^{c_1(t-1)}$. Also note that given $c_1(t-1)$, the event that no 0-clause is generated in step t is independent of the event that no 0-clauses in generated at time $1, 2, \cdots, t-1$. Combining the above facts, we get $\mathbf{P}(c_0(t_e) = 0) \ge \prod (1 - \frac{c_1(t)}{2\varepsilon N})^{Mn} \ge \exp(-\Omega(1) \sum c_1(t)/(\epsilon N)) \ge \Omega(1)$. $\square$

Let $\mathbf{H}(t)$ be $2 \times (t+1)$ matrix whose j-th column is $(c_2(j-1), c_3(j-1))^T$. Thus $\mathbf{H}(t)$ describes the entire history of the number of 2-clauses and 3-clauses up to time t. Let $\Delta c_i(t) = c_i(t+1) - c_i(t)$.

Claim 3: For all $0 \le t \le N-3$, conditional on $\mathbf{H}(t)$,

$$\Delta c_3(t) = -B_1(t) \quad \text{and} \quad \Delta c_2(t) = B_2(t) - B_3(t),$$

where $B_1(t) \stackrel{d}{=} \mathrm{Bin}\big(c_3(t), \frac{3}{N-t}\big)$, $B_2(t) \stackrel{d}{=} \mathrm{Bin}\big(c_3(t), \frac{3}{2(N-t)}\big)$, $B_3(t) \stackrel{d}{=} \mathrm{Bin}\big(c_2(t), \frac{2}{N-t}\big)$.

Suppose literal l is chosen at time $t+1$ among the literals in $L(V(t))$. By claim 1, every clause $a \in \mathcal{C}_i(t),\ i = 2, 3$ contains var(l), independently of all other clauses and with the same probability. Since there are $N-t$ unassigned variables and clause a contains i literals, this probability is $i/(N-t)$. This gives the negative terms in $\Delta c_3(t)$ $\Delta c_2(t)$ as each clause containing var(l) is going to be removed from the set in which it belongs at time t. Note that, also by claim 1, the probability that any clause $a \in \mathcal{C}_3(t)$ contains $\bar{l}$ is $3/2(N-t)$ independent of other clauses, which gives rise to the positive term. $\square$

Evolution of the number of 2 and 3 clauses has certain nice properties [1] which make it possible to approximate $\{\frac{c_i(xN)}{N} : 0 \le x \le 1-\varepsilon\}, i = 2, 3$ (for any fixed $\varepsilon > 0$) by solutions of certain differential equations as $N \to \infty$.

[1] (i) the conditional change in each step is highly concentrated around its expectation, (ii) Knowing the parameters of the process $(t, c_2(t), c_3(t))$ within o(N) suffices to determine $(\Delta c_2(t), \Delta c_3(t))$ within o(1). In other words $\Delta c_i(t)$ is a 'smooth' function of $t, c_2(t)$ and $c_3(t)$.

$$\mathbf{E}(\Delta c_3(t)\big|\mathbf{H}(t)) = -\frac{3c_3(t)}{N-t}, \quad c_3(0) = \alpha N \quad \frac{\mathrm{d}\eta_3}{\mathrm{d}x} = -\frac{3\eta_3(x)}{1-x}, \eta_3(0) = \alpha$$

$$\mathbf{E}(\Delta c_2(t)\big|\mathbf{H}(t)) = \frac{3c_3(t)}{2(N-t)} - \frac{2c_2(t)}{(N-t)}, \quad c_2(0) = 0 \quad \frac{\mathrm{d}\eta_2}{\mathrm{d}x} = \frac{3\eta_3(x)}{2(1-x)} - \frac{2\eta_2(x)}{1-x}, \eta_2(0) = 0$$

Solving the two differential equations above, we get

$$\eta_3(x) = \alpha(1-x)^3, \quad \eta_2(x) = \frac{3}{2}\alpha x(1-x)^2 \tag{1}$$

The following claim is a consequence of a theorem of Wormald [**?**].

Claim 4 Fix $\varepsilon > 0$ and let $t_e = (1-\varepsilon)N$. Then w.h.p.

$$\max_{t \le t_e}\big|c_i(t) - \eta_i(t/N).N\big| = \mathrm{o}(N), \ i = 2, 3.$$

Claim 5 Fixed $t \le N-1$. Then conditional on $\mathbf{H}(t)$, $c_1(t+1) \le \max(c_1(t)-1, 0) + B_4(t)$ where $B_4(t) \overset{d}{=} \mathrm{Bin}(c_2(t), \frac{1}{N-t})$.

If $c_1(t) > 0$ then the forced step ensures that at least one of the clauses in $\mathcal{C}_1(t)$ gets removed from the set of 1-clauses. Obviously, $B_4(t)$ denotes the number of 1-clauses that generates from $\mathcal{C}_2(t)$ at time t. If we have chosen literal l at time $t+1$ then, by claim 1, the probability that any clause in $\mathcal{C}_2(t)$ contains $\bar{l}$ is $2/2(N-t)$, independent of others. $\square$

Let us fix $\varepsilon = \frac{1}{10}$ yielding $t_e = (9/10)N$ and take $\alpha = \frac{8}{3}(1-\delta)$ where $\delta > 0$ is any constant. Since $\frac{\eta_2(x)}{1-x} \le \frac{3\alpha}{8}$, using claim 4, we get w.h.p. $c_2(t) < (1-\frac{\delta}{2})(N-t)$ for all $t \le t_e$ which implies that $\mathbf{E}\big(B_4(t)\big|c_2(t)\big) \le 1 - \frac{\delta}{2}$ for all $t \le t_e$.

So it is intuitively clear that since the expected number of 1-clauses generated at time t is bounded (w.h.p.) by $1-\frac{\delta}{2}$ where as the number of 1-clauses that gets removed from $\mathcal{C}_1(t)$ is at least 1 as long as $c_1(t) > 0$, no accumulation of 1-clauses should take place during the run of UC. In fact, under this condition, $c_1(t)$ behaves very much like the queue size in a stable server system. We can formalize this in our next claim which can be proved rigorously using the *Lazy-server lemma* (see Achlioptas [**?**]).

Claim 6 W.h.p. $(i)\sum_{t=1}^{t_e} c_1(t) \le MN$ and $(ii)\max_{0\le t\le t_e} c_1(t) \le \log^K N$ for some constants M and K.

The Claim 4 and 6 together assert that with positive probability there are no 0-clauses or 1 clauses at time t_e. Furthermore, from claim 4, we can easily check w.h.p. $c_2(t_e) + c_3(t_e) < (3/4)(N-t_e)$. Now, to conclude the proof, we argue as follows. Given ψ in $\mathrm{SAT}_N(3, \alpha)$ with α fixed as above we run UC for exactly t_e steps. With positive probability the remaining formula will have no 0 or 1 clauses. By uniformity, the remaining formula is the union of one random 2-SAT formula and an independent random 3-SAT formula (with $|V(t_e)| = N - t_e = \Omega(N)$ variables and fewer (w.h.p.) than $(3/4)(N - t_e)$ clauses). It is easy to see that such a formula is satisfiable with high probability since a 2-SAT fromula with the same number of clauses is satisfiable with high probability.

2 Method of Second Moment

In this section we will introduce yet another method for finding the lower bound of satisfiability threshold for a random k-SAT problem. Given a SAT formula ψ we consider a function $U(\psi)$ such that

$$U(\psi) \begin{cases} = 0 & \text{if } \psi \text{ is UNSAT} \\ > 0 & \text{otherwise} \end{cases} \tag{2}$$

The next lemma will be very useful.

Lemma 1

$$\mathbf{P}(\psi \text{ is SAT}) = \mathbf{P}(U(\psi) > 0) \geq \frac{[\mathbf{E}U(\psi)]^2}{\mathbf{E}[U(\psi)]^2}.$$

Proof $\mathbf{E}^2(U(\psi)) = [\mathbf{E}\{U(\psi)\mathrm{I}(U(\psi) > 0)\}]^2 \leq \mathbf{E}[U(\psi)]^2\mathbf{P}(U(\psi) > 0)$ by Cauchy-Schwartz. $\square$

The success of this method depends heavily on the choice of suitable function U such that ratio in the right side is bounded away from zero as $N \to \infty$. In most of the cases the choice of U turns out to be rather delicate. For example, the simple choice $U(\psi) = Z(\psi) =$ number of satisfying assignments does not give any significant lower bound. More precisely, the ratio $[\mathbf{E}U(\psi)]^2/\mathbf{E}[U(\psi)]^2$ becomes exponentially small in N for any $\alpha > 0$ (**exercise**).

Define

$$U(\psi) = \sum_{\mathbf{x}} \prod_{a=1}^{M} W(\mathbf{x}, a) \tag{3}$$

Here the sum is taken over all the 2^N assignments, and $W(\mathbf{x}, a)$ is the weight associated with the clause a with the following property.

$$W(\mathbf{x}, a) \begin{cases} = 0 & \text{if the assignment } \mathbf{x} \text{ does not satisfy clause } a \\ > 0 & \text{otherwise} \end{cases}$$

We take

$$W(\mathbf{x}, a) = \begin{cases} 0 & \text{if } \mathbf{x} \text{ does not satisfy clause } a \\ \varphi(r(\mathbf{x}, a)) & \text{if } \mathbf{x} \text{ satisfies clause } a \end{cases} \tag{4}$$

where $r(\mathbf{x}, a)$ = the number of variables in $\mathbf{x}$ satisfying clause a.

References

[1] Achlioptas, D. , "Lower Bounds for Random 3-SAT via Differential Equations ," *Theoretical Computer Science*, 2001, pp. 159-185.

[2] Chao, M. and Franco, J., "Probabilistic analysis of two heuristic for the 3-satisfiablity problem, " *SIAM J. Comput.,15*, 1986, pp. 1106-1118.

[3] Wormald, N.C. , "Differential equations for random processes and random graphs, " *Ann. Appl. Probab. 5 (4)*, 1995, pp. 1217-1235