Modern WalkSAT algorithms

J. Rosti
Summary of workshop presentation
Special research course T-79.7003, Autumn 2007
Phase Transitions in Optimization Problems

ABSTRACT

In this workshop presentation summary we review selected literature related to modern local search algorithms and the random 3-satisfiability problem.

The goal is to review how modern local search algorithms perform in random 3-SAT ensemble and discuss how the structure of the solutions is connected to the typical hardness of the problem.

1 Introduction

The satisfiability problem (SAT) consists of

- A set of n boolean variables $\{x_1, x_2, ..., x_n\}$
- A set of literals. A literal is a variable or a negation of a variable: $x_i, \neg x_i$.
- m distinct clauses $C_1, C_2, ..., C_m$. A clause is set of disjunctions of literals.

A SAT formula is:

$$C_1 \wedge C_2 \wedge ...C_m$$

The goal is to determine if there exists an assignment of truth values to variables, which makes the formula satisfiable.

In the k-satisfiability problem each clause depends on k number of variables. The number of clauses per variable $\alpha = m/n$ is fixed. If $k > 2$ the problem is NP-complete.

In the random ensemble, clauses consists of randomly selected literals. Phase transitions occur as a function of α: common example is *sat-unsat* transition when $\alpha_c = 4.27$ for 3-SAT. *Sat-unsat* means a sharp transition to the phase where there is not satisfiable assignment, when $\alpha > \alpha_c$ and n is large. When number of constraints (clauses) per variable increases the problem becomes harder to solve i.e. typical running time of algorithms increase as a function of α up to α_c.

We review studies on probabilistic local search algorithms. In the local search, one defines a neighbourhood for the problem, guesses a feasible solution and then tries to improve it by selecting a new solution from the neighbourhood. The procedure is repeated until the global minimum of energy is found or the maximum number of guesses is approached. Energy $E(s)$ for 3-SAT is defined conveniently as a number of unsatisfied clauses, thus zero energy corresponds a satisfiable solution.

Local minimums of the function to be minimized make greedy local search procedures perform badly. It is observed that traditional optimization heuristics does not apply well (simulated annealing, tabu search combined with trivial neighbourhood). Thus, a wide range of different heuristics is proposed for SAT. Good results is obtained using probabilistic and focused search algorithms. Focusing in the search heuristics means concentrating only on unsatisfied clauses.

The time to solution of WalkSAT-algorithms has been observed to grow only linearly in the number of variables, for a given clauses-to-variables ratio α_{lin} sufficiently far below the critical satisfiability threshold.

Seitz et al. [1] present numerical results of three focused local search algorithm and their performance. They study execution time of an algorithm They optimize a “random move” -probability and try to extend the linear behaviour of the algorithm when the number of constraints per variable increases. Algorithms under study by Seitz et al are WalkSAT, Focused Metropolis Search (FSM) and Focused Record-to-Record Travel (FRRT) method. Similar approach is taken by Ardelius et al. [2] for Average SAT (ASAT) -algorithm. In this paper we review WalkSAT, FSM and ASAT algorithms.

The well-known focused WalkSAT-algorithm is the following:

```
let s = random truth assignment
while flips < max_flips do:
  if s satisfies: return s
  else:
    let C = random unsat clause
    if variable x' in C can be flipped without breaking any sat clause:
      let x = x'
    else:
       with probability p:
          let x = random variable from C
       else with probability 1-p:
          let x =  x in C that x breaks minimal clauses
    flip x
```

WalkSAT alternates with probability p between greedy and random moves, whereas FSM combines focused greedy moves with the standard Metropolis dynamics. The parameter eta=η is free in the FSM and Seitz et al. [1] use the range $0.2 < \eta < 0.8$. The FSM-algorithm is as follows:

```
let s = random truth assignment
while flips < max_flips do:
  if s satisfies: return s
  let C = random unsat clause
  let x = random in C
  let x' = flip(x), s' = s[x'/x]
```

```
  if E(s') <= E(s) then:
    flip x
  else:
      flip x wiht prob. eta^(E(s')-E(s))
```

ASAT-algorithm is even simpler than FSM. In ASAT Metropolis dynamics -move is replaced with a constant probability flip. This removes direct sensitivity to the heights of the walls around local minima. ASAT in pseudocode is as follows:

```
let s = random truth assignment
while flips < max_flips do:
  if s satisfies: return s
  let C = random unsat clause
  let x = random in C
  let x' = flip(x), s' = s[x'/x]
  if E(s') <= E(s) then:
     flip x,
  else:
    flip x wiht prob. p
```

2 Phase transitions and typical hardness

One motivation to study local search algorithms is to find an effective local search algorithms for random 3-SAT. However, the *sat-unsat* is not the only phase transition when α is varied. Other, geometrical and real, phase transitions exist in the solution space. This raises a question if the structure of the solution space makes solutions very hard to find.

In the Fig. 1 a phase structure proposed by Krzakala et al. [3] is shown. The phase structure is for the q-coloring problem, but it is expected to be same for k-SAT as well.

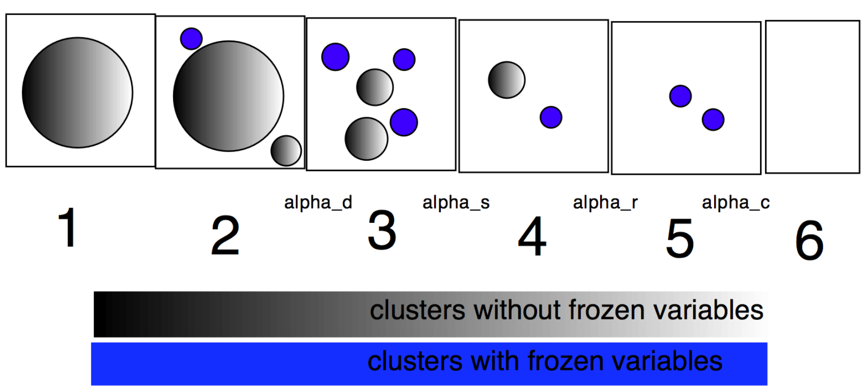

Figure 1: Structual phases of solutions. 1. Unique, 2. Irrelevant, 3. Clustered, 4. Condensed, 5. Rigid, 6. unsat.

Variable is *frozen* if it is forced for a truth value in the satisfying truth assignment. You cannot flip it and find the solution. A cluster is a set of solutions, which are connected via path through satisfiable solutions. A path is defined as a set of assignments which are obtained by flipping a single truth value in the assignment.

In the Fig. 1 numbers are explained to correspond following phases:

1. All solutions belong to a unique cluster.
2. The solution space contains irrelevant clusters, which are very small compared to dominating one.
3. In the clustered phase is when $\alpha_d < \alpha < \alpha_s$: It consists of large amount of dominating clusters. α_d is geometric phase transition i.e. the number of solutions is not non-analytical.
4. Condensed phase is $\alpha_s < \alpha < \alpha_c$: solutions are found from finite number of the largest clusters. Free energy is non-analytical in the clustered-condensed limit i.e. at α_s. The condensation happens due to increasing constraints.
5. Rigid phase is $\alpha_r < \alpha < \alpha_c$. In the rigid phase a finite fraction of variables is frozen inside dominant clusters. It is not known if $\alpha_r < \alpha_s$ or $\alpha_s < \alpha_r$ for k-SAT.
6. Unsat phase is $\alpha > \alpha_c$.

The fundamental question is: if the picture is correct, is the typical hardness related to the phase diagram? It might be. For typical local search algorithms average running time is linear up to α_{lin} and for even completely random flipping of truth values leads to non-zero α. For the RANDOMWALKSAT-algorithm, where one pick unsatisfied clause and flip random variable, the $\alpha_{lin} \approx 2.7$ (3-SAT). Different heuristics can push α_{lin} far beyond RANDOMWALKSAT.

3 Review of results

The Fig. 2 illustrates experiments by Seitz. The median running time (errors as quartiles) is plotted against α. The linear region of the running time extends towards large α, when the parameter p is increased towards optimal value $p = 0.57$. In the Fig. 3 are normalized solution times as a function of parameter p, which allows one find the optimum. Authors find α_{lin}=4.2 with p=0.57. [1]

Fig. 4 presents cumulative solution time distributions for WalkSAT with p=0.55. Distrubutions are narrow at α=4.15 (left figure). Heavy tail starts to develop when one increases α (right figure). Note that the parameter p=0.55 is non-optimal.

The Fig. 5 shows the time course of solution of one instance by ASAT at α=4.22 and N=10^6. It shows different decay processes and different time scales and finally approaches zero energy via long lasting energy plateaus. For ASAT the linear running time extends up to α_{lin}=4.21. [2]

3.1 Whitening

A solution is called completely “”white“” if it does not contain frozen variables. Focused search algorithms find only completely white solutions in linear time [1]. However, it is conjenctured that almost all solutions are completely white up to α_c [4]. α_c is expected to be very close to α_r in 3-SAT.

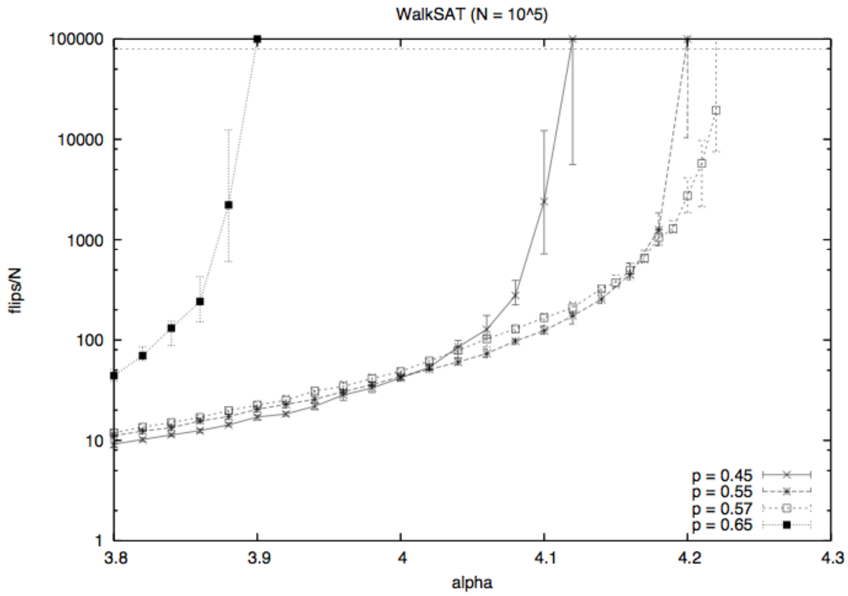

Figure 2: Median time to solution as a function of an α. [1]

4 Conclusions

Very simple WalkSAT -like algorithms extend much beyond random walk $\alpha_{lin} = 2.7$. Local search strategies seem to perform linearly at least up to $\alpha_d = 4.2$, but message passing algorithms, like survey propagation, still outperform WalkSAT when α is close to α_c. Message passing algorithms have been developed by using tools from statistical physics. Survey propagation involves, in a sense, a sophisticated probabilistic analysis of the properties of the solution of the problem instance under consideration.

The upper limit for dynamical paramater α_{lin} for focused local search algorithms may be related to rigidity and α_r or condensation transition. A precise picture is missing and the role of frozen variables and constraints is unclear.

REFERENCES

[1] Sakari Seitz, Mikko Alava and Pekka Orponen, *Focused local search for random 3-satisfiability*, J. Stat. Mech. P06006 (2006)

[2] John Ardelius and Erik Aurell, *Behavior of heuristics on large and hard satisfiability problems* Phys. Rev. Vol. E 74, 037702 (2006)

[3] Florent Krzakala and Lenka Zdeborova, *The Colorings of a Large Random Graph: From the Phase Diagram to Computational Hardness*, PNAS, Vol. 104, 25, 10318 (2007)

[4] E. Maneva, E. Mossel, and M. J. Wainwright, *A new look at survey propagation and its generalizations*, Technical report cs.CC/0409012, arXiv.org (2004)

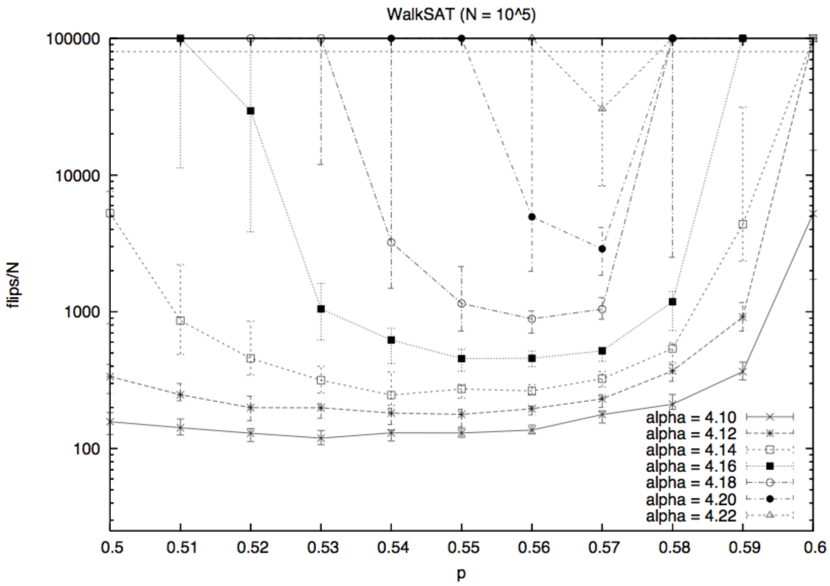

Figure 3: Time to solution as a function of p. [1]

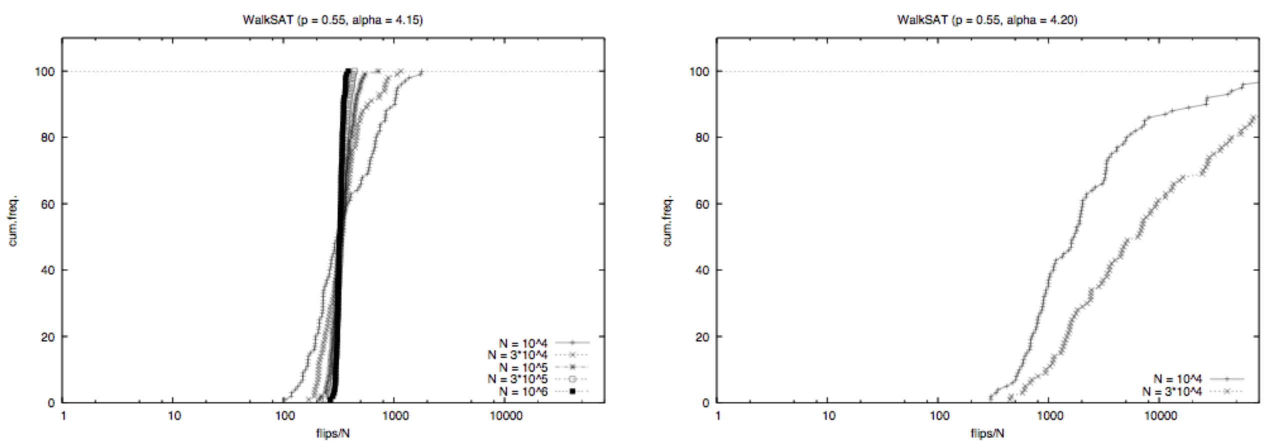

Figure 4: Cumulative solution time distribution for WalkSAT. [1]

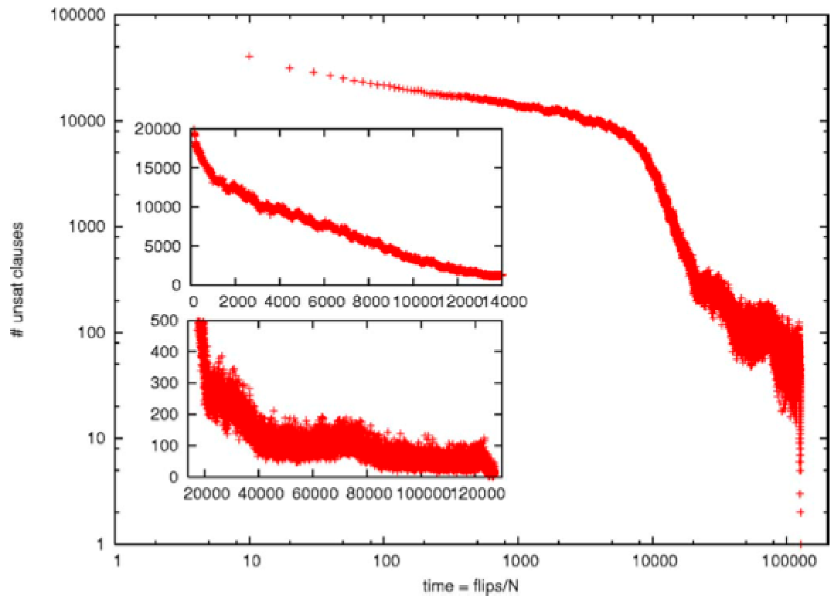

Figure 5: ASAT approaching solution slowly above α_{lin}, $\alpha = 4.22$, insets are different time scales. [2]