Phase transitions in the random satisfiability problem

Marc Mézard

20th January 2004

Collaborators:

• G. Parisi (Univ. Roma 1), R. Zecchina (ICTP Trieste)

• A. Braunstein (Trieste), S. Mertens (Magdeburg), F. Ricci-Tersenghi (Roma), M. Weigt (Göttingen).
Message passing (between disciplines)

• Many 'simple' variables

• Interactions $= \{Constraints\}$, involving a small fraction of the variables

• Find the values of variables compatible with all constraints (or which minimize the number of violated constraints)

Ubiquitous situation (statistical physics, combinatorial optimization, error correcting codes, statistical inference,...) $\rightarrow$ Message passing procedures:

• Organize the constraints and variables in a graph

• Exchange messages, of probabilistic nature, along the graph
SATISFIABILITY: an example

“You are chief of protocol for the embassy ball. The crown prince instructs you either to invite Peru or to exclude Qatar. The queen asks you to invite either Qatar or Romania or both. The king, in a spiteful mood, wants to snub either Romania or Peru or both. Is there a guest list that will satisfy the whims of the entire royal family?” (from B. Hayes, American Scientist 1997).

Configurations = assignments of Boolean variables: \[ P, Q, R \in \{0, 1\} \]

Constraints = clauses: \[ P \lor \overline{Q}, \ Q \lor R, \ \overline{R} \lor \overline{P} \]

Is there a choice of \( P, Q, R \) such that all constraints are satisfied (SAT)?
SATISFIABILITY: an important problem

$2^N$ Configurations $=$ assignments of $N$ Boolean variables: $x_i \in \{0, 1\}$

$M$ Constraints $=$ clauses like $x_1 \lor x_2 \lor \bar{x}_3, \ x_11 \lor x_2, \ldots$

Decision problem: is there a choice of the Boolean variables such that all constraints are satisfied (SAT)?

Generic (conjunctive normal form $(x_1 \lor x_2 \lor \bar{x}_3) \land (\bar{x}_11 \lor x_2) \land \ldots$).

Theorem (Cook, 1971): The SATISFIABILITY problem is NP-complete.

“NP”: a solution can be checked in polynomial time ($\sim N^r$); vast class (Travelling salesman, Hamiltonian path, Protein folding, Spin glass,...)

“Complete”: any other NP problem can be mapped to SATISFIABILITY in polynomial time $\rightarrow$ the hardest NP problems
“Worst-case” versus “Typical-case” complexity

Computational complexity = worst case analysis.

Experimental complexity = typical case analysis: $\rightarrow$ class of samples (probability measure on instances).

Ex: Complexity of the random 3SAT problem. Three variables per clause, chosen randomly in $\{x_1, \ldots, x_N\}$, negated randomly with probability $1/2$:

$$(x_1 \lor x_{27} \lor \overline{x_3}) \land (\overline{x_{11}} \lor x_3 \lor x_2) \land \ldots \land (x_9 \lor \overline{x_8} \lor \overline{x_{30}})$$

Control parameter: $\alpha = \frac{M}{N} = \text{Constraints/Variables}$.

Numerically: Threshold phenomenon at $\alpha_c \sim 4.26$.

Numerics Mitchell Selman Levesque Kirkpatrick Crawford Auton...;

Threshold Friedgut; Bounds Kaporis Kirousis Lalas Dubois Boufkhad...
Threshold phenomenon $\rightarrow$ Phase transition

generically SAT for $\alpha < \alpha_c$

generically UNSAT for $\alpha > \alpha_c$
Threshold phenomenon → Phase transition

Easy, and generically SAT, for $\alpha < \alpha_c$

Hard, in the region $\alpha \sim \alpha_c$

Easy, and generically UNSAT, for $\alpha > \alpha_c$
Statistical physics of the random 3-SAT problem


1- Analytic result: Three phases

2- New algorithm: Survey propagation

Increasing $\alpha$:

Easy-SAT, Hard-SAT, UNSAT
Main steps

• Graphical representation: 3SAT as a random factor graph

• Elementary message passing procedure: exchange of warnings between constraints and variables.

• In the presence of many clusters: cavity method $\rightarrow$ messages $=$ surveys of elementary messages in all clusters.
Graphical representation: “factor graph”

One clause $a$:

Boolean: $\bar{x}_1 \lor x_2 \lor \bar{x}_3$

\[
(\bar{x}_1 \lor \bar{x}_2 \lor \bar{x}_4) \land (x_1 \lor \bar{x}_2) \land (x_2 \lor x_4 \lor x_5) \\
\land (x_1 \lor x_2 \lor \bar{x}_5) \land (x_1 \lor \bar{x}_3 \lor x_5)
\]
Simple message passing: warning propagation

Message $u_{a \rightarrow 1} \in \{0, 1\}$
sent from clause $a$
to variable $1$
Warning $u_{a \to i} = 1$:

"According to the messages I received, you should take the value which satisfies me!".
Simple message passing: warning propagation

No warning $u_{a\rightarrow i} = 0$:

“No problem, take any value!”

Warning propagation converges and gives the correct answer on a tree: SAT
iff no contradictory message
Basic underlying idea: uncorrelated inputs

Joint 'cavity' probability $P^{(a)}(x_2, x_3)$

$a$ absent implies $x_2$ far from $x_3$:

$$P^{(a)}(x_2, x_3) \sim P^{(a)}(x_2)P^{(a)}(x_3)$$

'Belief propagation':

$$P_{a \to 1}(x_1) = \sum_{x_2, x_3} C_a(x_1, x_2, x_3)P^{(a)}(x_2)P^{(a)}(x_3)$$

$$P^{(b)}(x_1) \propto \prod_{a \in V(1) \setminus b} P_{a \to 1}(x_1)$$

Limit of hard constraints $\rightarrow$ 'Warning propagation'.
Proliferation of states

Warning propagation works if one can neglect the correlations between the input fields (tree).

Random 3SAT “locally tree-like”: generically, $x_2$ and $x_3$ are very far away (distance $O(\log(N))$) → Uncorrelated if only one cluster / pure state.

→ OK in Easy-SAT phase.

→ Wrong in Hard-SAT phase
From warning propagation to survey propagation

Hard SAT phase: Message = Survey of the elementary messages in the various clusters of SAT configurations: \( \eta_{a \rightarrow i} \) = probability of a warning being sent from constraint \( a \) to variable \( i \), when a cluster is picked up at random.

Propagate the surveys along the graph. Converges!

→ Results on the phase diagram (generic sample), but also information on a single sample.
Survey propagation

\( \eta_{a \rightarrow 1} \): known exactly from joint probability of incoming warnings.

**SP approximation**: this joint probability factorizes
Three phases

Number of clusters at 'energy' $E$ is exponential: $N(E) = \exp(N\Sigma(E))$

NB: Stability to further sub-clustering: stable in a finite region around $\alpha_c$. Conjecture: $\alpha_c = 4.267\ldots$ is the exact threshold.
Single sample analysis: a new algorithm

Order parameter = Survey of local polarizations, in all states → Algorithm for the Hard-SAT phase. Survey Inspired Decimation: fix the variable which is most biased, rerun the survey propagation, iterate...

Solves typical random 3sat up to $N = 10^7$ at $\alpha = 4.23$, complexity $O(N \log N)$.

Local surveys of magnetic fields → a lot of information. Probably possible to invent other algorithms based on the surveys.
Summary

• Analytic result on the generic samples of random 3sat: Phase diagram.

• Slowdown of algorithms near to $\alpha_c = 4.267$ due to the existence of a Hard SAT phase at $\alpha \in [\sim 3.9, 4.267]$, with exponentially many states.

• The whole construction can be checked versus rigorous computations on the “random XORSAT” (or “3 spin glass”) problem. (MM, Ricci-Tersenghi, Zecchina; Cocco, Dubois, Mandler, Monasson).

• Single sample analysis: Survey propagation converges, yields non trivial information on the sample (diversity of sites) $\rightarrow$ Survey Inspired Decimation: a very efficient algorithm for solving random 3sat problems. Also applicable to many “constrained satisfaction problems”, e.g. graph colouring (Mulet, Pagnani, Weigt, Zecchina).
References


