# Random subcubes as a toy model for constraint satisfaction problems 

Lenka Zdeborová
with Thierry Mora
 (LPTMS, Orsay, France)

## Constraint satisfaction problems

"The boolean satisfiability problems (SAT) is of central importance in various areas of computer science, including theoretical computer science, algorithmics, artificial intelligence, hardware design, electronic design automation, and verification." (from Wikipedia - Boolean satisfiability problem)

$$
\left(x_{1} \vee \neg x_{2} \vee x_{3}\right) \wedge\left(x_{2} \vee \neg x_{4} \vee \neg x_{5}\right) \wedge\left(\neg x_{1} \vee x_{4} \vee x_{5}\right)
$$

"The problem of coloring a graph has found a number of applications. Some of them are scheduling, register allocation in compilers, frequency assignment in mobile radios, and pattern matching." (from Wikipedia - Graph coloring)


NP complete in the worst case - but many instances are easy! Where the really hard problems are?

## Phase diagram of random K-SAT and q-COL



## The results of statistical physics (1RSB)



- Dynamical transition
- Condensation transition
- SAT/UNSAT threshold
F. Krząkała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Z.:

Proc. Natl. Acad. Sci. U.S.A., 104, 10318 (2007).

## So where are we in understanding clusters?

- Definition: Using concepts of reconstruction on trees, Gibbs measure extremality (Krzakala, Montanari, et al, PNAS, 2007)
- Good news! Definition exists.
- Good property: "Equivalent to the cavity clusters"
- Useful to prove things about the un-clustered phase. But not yet very useful to understand the clustered phase.
- Existence: Geometrically separated clusters exist (Mezard,Mora,Zecchina, 2005, Achlioptas, Ricci-Tersenghi,2006), far too strong definition used here.
- Properties: Via the cavity method,
- Satisfactory for physicist, but still complicated.
- Not that satisfactory for mathematicians.


## So where are we in understanding clusters?

- Definition: Using concepts of reconstruction on trees, Gibbs measure extremality (Krzakala, Montanari, et al, PNAS, 2007)
- Good news! Definition exists.
- Good property: "Equivalent to the cavity clusters"
- Useful to prove things about the un-clustered phase. But not yet very useful to understand the clustered phase.
- Existence: Geometrically separated clusters exist (Mezard,Mora,Zecchina, 2005, Achlioptas, Ricci-Tersenghi,2006), far too strong definition used here.
- Properties: Via the cavity method,
- Satisfactory for physicist, but still complicated.
- Not that satisfactory for mathematicians.


## Conclusion: We are in need of a "REM-like" model here!

## Random Subcubes Model

Definition: Random Cluster $\mathbf{A}$ is a subcube of $\{0,1\}^{N}$
where

$$
A=\left\{\underline{\sigma} \mid \forall i \in\{1 \ldots N\}, \sigma_{i} \in \pi_{A}(i)\right\}
$$

$$
\begin{array}{rll}
\pi_{A}(i)=\{0\} & \text { prob } & p / 2 \\
\pi_{A}(i)=\{1\} & \text { prob } & p / 2 \\
\pi_{A}(i)=\{0,1\} & \text { prob } & 1-p
\end{array}
$$

Choose independently $\mathcal{N}=2^{(1-\alpha) N}$ random cluster.

Aim: Study the structure of the set of "solutions" parameters $p$ and $\alpha$

$$
\mathcal{S}=\cup_{a=1}^{\mathcal{N}} A_{a}
$$

## Distribution of sizes of clusters

Internal entropy s of a cluster: fraction of unfrozen variables in the cluster.
Typical entropy: $s_{\text {typ }}=1-p$

Definition of complexity function: \#\{clusters of entropy $s\}=2^{N \Sigma(s)}$

$$
\begin{aligned}
2^{N \Sigma(s)} & =2^{(1-\alpha) N}\binom{N}{s N}(1-p)^{s N} p^{(1-s) N} \\
\Sigma(s) & =(1-\alpha)-D(s \| 1-p) \\
D(s \| 1-p) & =-s \log _{2} \frac{s}{1-p}-(1-s) \log _{2} \frac{1-s}{p}
\end{aligned}
$$

## The total entropy (dominating clusters)

The entropy of dominating clusters $\mathrm{s}^{*}$, and the total entropy:

$$
\begin{aligned}
& s^{*}=\arg \max _{s}[(\Sigma(s)+s) \mid \Sigma(s) \geq 0] \\
& s^{\mathrm{tot}}=\max _{s}[(\Sigma(s)+s) \mid \Sigma(s) \geq 0]=\Sigma\left(s^{*}\right)+s^{*}
\end{aligned}
$$

## 4 different phases

- Liquid phase: $s^{\text {tot }}=1$
- Clustered phase: $s^{\text {tot }}<1 \quad \Sigma\left(s^{*}\right)>0$
- Condensed phase: $s^{\text {tot }}<1 \quad \Sigma\left(s^{*}\right)=0$
- Unsatisfiable phase:

$$
\Sigma(s)<0 \quad \forall s
$$



$$
\alpha_{d}=\log _{2}(2-p) \quad \alpha_{c}=p /(2-p)+\log _{2}(2-p)
$$

## About the transitions

- The dynamical transition = ergodicity breaking (for a random walk in the solution space).
- Total entropy has a discontinuity in the second derivative at the condensation transition (glass analog: jump in specific heat at the Kauzmann transition).
- The slope of the function $\partial \Sigma(s) / \partial s=-m$ is the Parisi parameter in the cavity (replica) method
- Condensed phase: the relative sizes of clusters follow the Poisson-Dirichlet process with parameter m .


## Main differences between subcubes and K-SAT

- In subcubes: no underlying geometry, no graph of constraints.
- SAT: Unfrozen variables are not independent.
- SAT: Clusters do not have to contain frozen variables, might contain only very much correlated variables.


## Subcubes are a limit case of K-SAT and q-COL

REM

$$
=\text { is } p \text {-spin for } p \longrightarrow \infty
$$

Random Cluster Model
= large q limit of q-coloring of random graphs
= large K limit of random K-satisfiability

## Subcubes model = large K limit of K-satisfiability

Reparametrisation in the subcubes model:

$$
\begin{array}{ll}
p=1-\epsilon & \epsilon \ll 1 \\
\alpha=1+\epsilon \frac{1+\gamma}{\ln 2} & \gamma=\Theta(1)
\end{array}
$$

K-SAT in the large K limit constant near to the satisfiability threshold:

$$
\begin{aligned}
& \epsilon=\frac{1}{2^{K+1}} \quad \frac{M}{N}=2^{K} \ln 2-\frac{\ln 2}{2}+\frac{\gamma}{2} \\
& \Sigma(s) \ln 2=s\left(1-\ln \frac{s}{\epsilon}\right)-\epsilon(2+\gamma)+o(\epsilon)
\end{aligned}
$$

Consequence: Clusters in large K limit near to the satisfiability threshold fill the subcube!

## Energy landscape

Create $e^{N \Sigma\left(E_{0} / N\right)}$ random subcubes with energy $E_{0}$, where $\Sigma\left(e_{0}\right)$ is a nice increasing function.

Define: energy of a configuration $\vec{\sigma}$

$$
E(\vec{\sigma}):=\min _{V}\left[E_{0}(V)+d(\vec{\sigma}, V)\right]
$$

Crucial properties: Lipschitz continuity of the landscape. Gradient descent meaningful. Energetic barriers are extensive.

Glassy dynamics in subcube energy landscape



## Conclusions: Random Subcubes Model

- A toy to understand clustering.
- The same phase transitions as in K-SAT/q-COL.
- Which properties of SAT/COL are probabilistic and which are due to the structure of constraints.
- TO DO: Finite size corrections.
- In the large K/q limit clusters in K-SAT/q-COL are subcube-like. TO DO: Prove that.
- Playground to study dynamics on a glassy energy landscape. TO DO: Different definition of the landscape to reproduce aging, rejuvenation, memory ...

Mora T., Zdeborová L., arXiv:0710.3804v1 [cs.CC]
Random subcubes as a toy model for constraint satisfaction problems

