“1-RSB” clustering and algorithms

Riccardo Zecchina (ICTP, Trieste, Italy)

Marc Mezard (Orsay), Alfredo Braunstein (Trieste ICTP), Thierry Mora (Orsay)

&

Giorgio Parisi (Rome)
Demian Battaglia (Trieste SISSA)
Martin Weigt, Michele Leone, Andrea Pagnani (ISI-Turin)
Plan of the talk:

- Examples of clustering
- Evidence of clustering in random k-sat
- Predictions of the cavity approach
- Survey propagation (SP) algorithm
  - $SP = BP$ over a modified combinatorial problem
- Clusters selection by SP & lossy data compression
Graphical representation (factor graphs) of CSP

Bipartite graph

variable nodes \( \{x_i\} \)

function nodes \( E_a[\{x_i(a)\}] \)

Global cost:
\[
E_{tot} = \sum_{a} E_a[\{x_i(a)\}]
\]
Isolated solutions: XORSAT over (L,K) random uniform graphs

\[ A = (0, 0, 0, ..., 0) \]

Ref. solution: \[ A = (0, 0, 0, ..., 0) \]

Count solutions at distance: \[ dN \]

A random formula needs an even # of edges connected to the “1” sector

\[
\bar{N} = \sum_{s_a=0(\text{even})}^{K} \delta\left(\sum_{a} s_a, LdN\right) \binom{N}{dN} \frac{(LdN)!}{(\prod_{a} s_a)!} \frac{(L(1 - d)N)!}{(\prod_{a} (K - s_a))!} \frac{(K! \frac{L}{K})^{N}}{(dN)!}
\]
\[ \delta(\sum_a s_a, LdN) = \int \frac{d\lambda}{2\pi} e^{-i\lambda(\sum_a s_a - LdN)} \]

\[ \sum_{s_a=0\text{(even)}}^K e^{-i\lambda s_a} = \frac{(1 + e^{i\lambda})^K + (1 - e^{i\lambda})^K}{2s_a!(K - s_a)!} \]

\[ (1 + a)^K = \sum_{s=0}^K \binom{K}{s} a^s, \quad (1 - a)^K = \sum_{s=0}^K \binom{K}{s} (-)^s a^s \]

\[ \bar{N} = e^{N\phi(\mu^*)} \quad (\mu \equiv i\lambda) \]

\[ \phi(\mu) = -\mu Ld + \frac{L}{K} \ln \left( \frac{(1 + e^\mu)^K + (1 - e^\mu)^K}{2} \right) + (L - 1) \left[ d \ln d + (1 - d) \ln(1 - d) \right] \]

\[ \phi'(\mu^*) = 0 \]
weight enumerator function for (3,6) xorsat

\[ \frac{1}{N} \ln \bar{N} \]

Di, Montanari, Urbanke, 04

gap
The case of Poisson degrees: random XORSAT

\[ \bar{N} = \sum_{\ell_1, \ldots, \ell_r, \ldots} \prod_r \left( \binom{N_r}{\ell_r} \right) \delta(dN, \sum_r \ell_r) \sum_{s_a, (\text{even})} \delta(\sum a s_a, \sum_r r \ell_r) \]

\times \frac{(\sum_r r \ell_r)! (\sum_r r(N_r - \ell_r))!}{(\prod_a s_a)! (\prod_a (K - s_a))!} \frac{(K)!^M}{(\sum_r r N_r)!}

\bar{N} = \# \text{ var. of degree } r

\frac{1}{N} \log \bar{N}

(K = 6, \gamma = .9)
Exact solution by leaf removal:

\[ \mathbb{E} \left( D_\ell (T + 1) - D_\ell (T) \mid \{D_i (T)\} \right) = \delta_{\ell,0} - \delta_{\ell,1} + (K - 1) \left( e_{\ell+1} (T) - e_{\ell} (T) \right) \]

\[ e_{\ell} (T) = \frac{1}{3(M - T)} \ell D_\ell (T) \]

\[ \frac{\partial d_\ell (t)}{\partial t} = \delta_{\ell,0} - \delta_{\ell,1} + \frac{(K - 1)}{K} \cdot \frac{(\ell + 1) d_{\ell+1} (t) - \ell d_\ell (t)}{\gamma - t} \]

\[ d_\ell (0) = \frac{(K \gamma)^\ell e^{-K \gamma}}{\ell!} \]

ODE: probability of finding a vertex of degree \( \ell \) after \( t \) N links have been removed
\[ \lambda (t) \equiv 3 \left[ \gamma (\gamma - t)^2 \right]^{\frac{1}{3}} \quad (K = 3) \]

\[ d_\ell (t) = e^{-\lambda(t)} \frac{\lambda(t)^\ell}{\ell!} \quad \ell \geq 2 \]

\[ d_1 (t) = \lambda (t) \left[ e^{-\lambda(t)} - 1 + \left( \frac{\lambda (t)}{3\gamma} \right)^{\frac{1}{2}} \right] \]

\[ d_0 (t) = 1 - \sum_{\ell=1}^{\infty} d_\ell (t) \]

\[ t_{final} (\gamma) \equiv t^* : d_1 (t^*) = 0 \]

\[ \gamma_{final} = \frac{1}{\gamma - t^*} \left[ 1 - (1 + \lambda (t^*)) e^{-\lambda(t^*)} \right] \]
\[
S(\gamma) = \log(2)(1 - \gamma)
\]
\[
S_{\text{int}}(\gamma) = S(\gamma) - \Sigma(\gamma)
\]
\[
\Sigma(\gamma) = \begin{cases} 
0 & \text{if } \gamma < \gamma_d \\
\log(2) \left[ 1 - \frac{\lambda}{3} - e^{-\lambda}(1 + \frac{2}{3}\lambda) \right] & \text{if } \gamma_d > \gamma < \gamma_c
\end{cases}
\]

Same results with the 1-rsb cavity method (also for max-xorsat)
Clustered structure = core & non-core variables

An assignment to the core variables is the **seed** of the cluster:

- Hamming distance between two solutions on the core is $\Omega(N)$
- Two solutions with the same seed can be joined by a path of solutions with jumps of Hamming distance $O(1)$
- By linearity clusters are identical

$\# \text{clusters} = e^{N\Sigma} = \# \bullet = 2^{N_c-M_c}$

$\bullet = \text{core solution}$

non-core variables give the intra-cluster entropy

$S_{nc} = S - S_c = S - \Sigma$

Ricci-Tersenghi, Weight, Zecchina (01)
Franz, Leone, Ricci-Tersenghi, Zecchina (02)
Dubois, Mandler (02)
Mezard, Ricci-Tersenghi, Zecchina (03)
Cocco, Dubois, Mandler, Monasson (03)
Clusters in XORSAT-like are identical: fluctuating variables are always the same in all clusters (identified by the graph structure).

In general the variables which fluctuate depend on the cluster (e.g. in random k-sat, q-coloring).

Cavity (and replica) method still provide results that are conjectured to be exact in the more general case.
Random K-SAT

- Let $C_K(N)$ be the set of all $2^K \binom{N}{K}$ possible K-clauses on $\sigma_1, \sigma_2, \ldots, \sigma_N$

- Select uniformly, independently and with replacements $M = \alpha N$ clauses from $C_K(N)$ to generate a K-cnf formula $F_N(K, \alpha)$

$$\left( x_1 \lor x_27 \lor \bar{x}_3 \right) \land \left( \bar{x}_{11} \lor x_3 \lor x_2 \right) \land \ldots \land \left( x_9 \lor \bar{x}_8 \lor \bar{x}_{30} \right)$$

**Question:** does $F_N(K, \alpha)$ have a truth assignment?

**Conjecture:** for each $K$ there exist a limit value for $\alpha = \alpha_c(K)$ separating the sat and unsat phases
**Evidence of clustering : x-Satisfiability**

**def.** A formula \( F_N(K, \alpha) \) is **x-sat** if a pair of solutions \( \sigma, \tau \) at a distance \( d_{\sigma \tau} \) specified by \( x \) exists

\[
d_{\sigma \tau} = \frac{N}{2} - \frac{1}{2} \sum_{i=1}^{N} \sigma_i \tau_i = \frac{1 - x}{2} N
\]

\( x \) is the fraction of common values ("overlap")

(Mezard, Mora, Zecchina 05)
x-Satisfiability Threshold conjecture

For all $K \geq 3$, there exists $\alpha_1(K), \alpha_2(K)$, with $\alpha_1(K) < \alpha_2(K)$, such that:
for all $\alpha_1(K) < \alpha < \alpha_2(K)$, there exist $x_1(K, \alpha) < x_2(K, \alpha) < x_3(K, \alpha)$ such that:

- for all $x \in [x_1(K, \alpha), x_2(K, \alpha)] \cup [x_3(K, \alpha), 1]$, a random formula $F_N(K, \alpha)$ is $x$-satisfiable w.h.p.
- for all $x \in [0, x_1(K, \alpha)] \cup [x_2(K, \alpha), x_3(K, \alpha)]$, a random formula $F_N(K, \alpha)$ is $x$-unsatisfiable w.h.p.

1. For all $K \geq 2$ and for all $x$, $0 < x < 1$, there exists an $\alpha_c(K, x)$ such that:
   - if $\alpha < \alpha_c(x)$, $F_N(K, \alpha)$ is $x$-satisfiable w.h.p.
   - if $\alpha > \alpha_c(x)$, $F_N(K, \alpha)$ is $x$-unsatisfiable w.h.p.

2. for $K \geq 3$, as a function of $x$, the function $\alpha_c(K, x)$ is non monotonous and has two local maxima: a local maximum at a value $x_M(K) < 1$, and a global maximum at $x = 1$. 
Bounds by first and second moments

By first and second moment methods we obtain two functions, \( \alpha_{UB}(K, x) \) and \( \alpha_{LB}(K, x) \), such that:

- For \( \alpha > \alpha_{UB}(K, x) \), a random K-CNF is \( x \)-unsatisfiable w.h.p.
- For \( \alpha < \alpha_{LB}(K, x) \), a random K-CNF is \( x \)-satisfiable with a finite probability

Upper bound by first moment:

\[
\Pr(Z \geq 1) \leq \mathbb{E}(Z)
\]

\[
Z = \sum_{\bar{\sigma}, \bar{\tau}} \delta \left( \frac{d_{\sigma \tau}}{N} = (1 - x) \right) \delta(\bar{\sigma}, \bar{\tau} \in S(F))
\]

\( S(F) \) is the set of solutions of \( F_N(K, \alpha) \)
\[
\mathbb{E}(Z(F)) = 2^N \binom{N}{N} \mathbb{E} [\delta(\vec{\sigma}, \vec{\tau} \in S(F))] = 2^N \binom{N}{N} \mathbb{E} [(\vec{\sigma}, \vec{\tau} \in S(c))]^M
\]

\[
\lim_{N \to \infty} \ln \mathbb{E}(Z(F))/N \to \Phi_1(x, \alpha)
\]

\[
\Phi_1(x, \alpha) = \ln 2 + H_2(x) + \alpha \ln \left(1 - 2^{1-K} + 2^{-K} x^K \right)
\]

\[
\alpha_{UB}(K, x) = -\frac{\ln 2 + H_2(x)}{\ln(1 - 2^{1-K} + 2^{-K} x^K)},
\]

\textit{x-unsat} if \( \alpha > \alpha_{UB}(K, x) \)
Lower bound by second moment (refs. Achlioptas, Moore, Peres, Naor,...):

\[ P(Z > 0) \geq \frac{\mathbb{E}(Z)^2}{\mathbb{E}(Z^2)} \quad (Z \geq 0) \]

\[ Z = \sum_{\vec{\sigma}, \vec{\tau}} \delta \left( \frac{d_{\sigma\tau}}{N} = (1 - x) \right) W(\vec{\sigma}, \vec{\tau}, F) \]

\[ Z > 0 \leftrightarrow \text{existence of pairs of solutions} \]

\[ \alpha_{LB}(K, x) = \inf_{\vec{a} \in V} \frac{\ln 2 + 2H_2(x) - H_8(\vec{a})}{\ln f_2(\vec{a}) - 2 \ln f_1(x)} \]

\text{x-sat if } \alpha < \alpha_{LB}(K, x) \]
... at least for $K>7$ there is evidence of clustering in random K-SAT

(Mezard, Mora, Zecchina 05)
Random K-SAT as a diluted spin glass

\[ E = \sum_{\ell=1}^{\alpha N} E_\ell \] (# violated clauses)

\[ E_\ell = \frac{1}{2K} \prod_{i \in \ell} (1 - J_{i}^\ell \sigma_i) \]

Phase transition: SAT/UNSAT ⇔ \((E_0 = 0)/(E_0 > 0)\)

\[ \alpha = \frac{M}{N} \]
Some rigorous facts about spin glasses (SK & p-spin) & random k-sat:

Guerra, Guerra & Toninelli, Talagrand

1. Free energy:
   \[ f \equiv \lim_{N \to \infty} \frac{1}{N} \log Z_J \quad \rightarrow \quad \text{exists} \]

2. Concentration of measure:
   \[
   \text{Prob}[|\frac{1}{N} \log Z_J - \frac{1}{N} \log Z_J| > \epsilon] \leq e^{-g(\epsilon)N}, \quad g(\epsilon) > 0
   \]

3. \( f_a \) = free energies from cavity or replica methods
   \[ f - f_a = R \geq 0 \]

(for K-SAT proved for K even (S. Franz, M. Leone, 2003); K odd ...).
T=0 Cavity method

Cavity spins are log(N)-distant and weakly correlated

\[ E(\sigma_0) = A - \sum_{a=1}^{k} w_a(g_a, h_a) - \sigma_0 \sum_{a=1}^{k} u_a(g_a, h_a) \]

marginal probability giving the fraction of SAT assignment

\[ h_0 = \sum_{a=1}^{k} u_a, \quad P_{cavity}(\sigma_0) \propto e^{-\beta h_0 \sigma_0} \]
Message-passing and cavity analysis

- BP = cavity RS = recursive equations for marginals over sat assignments (absence correlations hypothesis). Also known as Sum/Product algorithms for trees.

- Warning Propagation (WP): discretized BP

\[ P_{i\rightarrow a}(J^a_i h_{i\rightarrow a} > 0) = \mathcal{F} \left[ \{ P_{j\rightarrow b}(J^b_j h_{j\rightarrow b} > 0) \}_{b \in V(i) \setminus a} \right] \]

- WP with many clusters does not converge (clusters imply correlations).

- Survey Propagation (SP) = cavity 1-rsb= histogram of WP over clusters

*next talk and Mezard on Friday*
Absence of correlations $\iff$ one cluster (RS)

exact on trees (or random graphs of small enough average degree)

$$P(h_0) = \int \prod_k dg_k dh_k P(g_k)P(h_k) \delta(h_0 - f(\{g_k, h_k\}))$$

... easy to solve (integers fields) ...

$$P(h) = \sum_r p_r \delta(h - r)$$

$$p_0 = e^{-3\alpha(1-c_0)}I_0(3\alpha(1-c_0)), \quad c_0 = 1 - \left(\frac{1-p_0}{2}\right)^2$$

$$p_r = p_{-r} = f_r(p_0)$$

... but wrong in the interesting region: **How to deal with correlations?**
1-RSB scenario & cavity assumptions

(i) Cavity variables uncorrelated within clusters

(ii) Cluster proliferation $\mathcal{N}(e) \sim \exp(N\Sigma(e))$

Study the survey of local fields in the states (clusters) of given energy density

$$P^e_i(h) = C \sum_{s \in \text{states}} \delta(h - h^s_i)\delta(e - E^s/N)$$

Functional T=0 cavity equations for the order parameter:

$$\{P_i(h)\} \rightarrow \mathcal{P}[P(h)]$$

Order parameter: probability measure in a space of functions
3-SAT Survey Propagation equations

\[
\begin{aligned}
P_{j \to a}(h) &= c_{j \to a} \int \prod_{b \in V(j) \setminus a} du_b Q_{b \to j}(u_b) \delta(h - \sum_b u_b) \chi[\{u\}] \\
P_{k \to a}(g) &= \ldots
\end{aligned}
\]

\[
Q_{a \to i}(u) = \int dh dg P_{j \to a}(h) P_{k \to a}(g) \delta(u - \hat{u}(h, g)) \quad \text{“surveys”}
\]

link contribution to cavity fields

\[
\hat{u}(h, g) = J^a_i \theta(J^a_j h) \theta(J^a_k g) \in \{0, J^a_i\}
\]

re-weighting of states

\[
\chi[\{u\}] = \exp(y(|\sum u| - \sum |u|))
\]
Statistical analysis of SP equations: average over the random factor graphs

\[ Q_{a \rightarrow i}(u) = \begin{cases} 
\delta(u) \text{ prob. } t & \text{(dangling nodes)} \\
(1 - \eta_{a \rightarrow i}) \delta(u) + \eta_{a \rightarrow i} \delta(u \pm 1) & \text{prob. } \frac{1-t}{2}
\end{cases} \]

Functional integral equations for link-to-link fluctuations of surveys

\[ \rho(\eta) = \mathcal{F}[\rho(\eta)] \implies \overline{\Phi(y)} = e - \Sigma(\alpha, e)/y \]

population dynamics \hspace{1cm} average free-energy

\[ e \rightarrow \text{energy density} \]

\[ \Sigma \rightarrow \frac{1}{N} \log ( \# \text{ clusters}) \]

“complexity”
Critical points:

\[
\begin{array}{|c|c|c|}
\hline
K & \alpha_d & \alpha_c \\
\hline
3 & 3.92 & 4.2667 \\
4 & 8.29 & 9.931 \\
5 & 16.1 & 21.117 \\
6 & 30.4 & 43.37 \\
7 & 57.2 & 87.79 \\
8 & 107.2 & - \\
9 & 201 & - \\
10 & 379 & - \\
\hline
\end{array}
\]

Large \( K \)

\[
\alpha_c(K) = 2^K \ln 2 - \frac{1 + \ln 2}{2} + O(2^{-K})
\]

\[
\alpha_d = \frac{C(K)}{K} 2^K
\]


3-SAT phase diagram (1-rsb)
SAT

RS

1-\text{rsb}

unstable

1-\text{rsb}

stable

0

\alpha_d

\alpha_c
finite “y” cavity analysis: (E>0) MAX-K-SAT and metastable clusters
Survey Propagation Algorithm: SP over a given instance

What is the role of var. $\sigma_{i=317}$ in a system of size $10^7$?

Fixed point for: $\{Q_{a\rightarrow i}(u)\}$ or $\{P_{i\rightarrow a}(h)\}$

$Q_{a\rightarrow i}(u) = (1 - \eta^+_a - \eta^-_a)\delta(u) + \eta^+_a \delta(u - 1) + \eta^-_a \delta(u + 1)$

$\eta^+_a = \mathcal{F}[\{\eta^+_b\}]$

$j \in V(a) - i$, $b \in V(j) - a$

converge rapidly
Bias over clusters of sat assignments

\[ P_i(H) = \int \prod_{a \in V(i)} d\mathbf{u}_a \mathcal{Q}_{i \rightarrow a}(\mathbf{u}_a) \delta(H - \sum_a J_i^a \mathbf{u}_a) \chi[\{\mathbf{u}\}] \]

\begin{align*}
W_i^{(+)} &= \int_{H>0} P_i(H) \\
W_i^{(-)} &= \int_{H<0} P_i(H) \\
W_i^{(0)} &= P_i(0)
\end{align*}

“Probability to find variable i frozen + or -, or unfrozen in a cluster”

Other quantities of interest: complexity, correlations, ...
SP decimation algorithm (SAT phase, \(y = \infty\))

1. Iterate SP equations

2. Compute biases and the average bias

\[
W_i = \{W_i^{(-)}, W_i^{(0)}, W_i^{(+)}\}
\]

\[
<\text{bias}> = \frac{1}{N_t} \sum_{i \in V_t} \max\{W_i^{(+)}, W_i^{(-)}\}
\]

3. If \(<\text{bias}> = 0\ (\ < \epsilon)\) then
   
   3.1. Output sub-problem
   
   3.2. Call an external algorithm to solve
   
   3.3. Exit

4. Sort and Fix the most biased variables

5. Go to (1)

Efficient \((N \log N)\) for hard problems (SAT, Coloring, ...
### Performance on random 3-sat

<table>
<thead>
<tr>
<th>N</th>
<th>2.5 \cdot 10^4</th>
<th>5.0 \cdot 10^4</th>
<th>1.0 \cdot 10^5</th>
</tr>
</thead>
<tbody>
<tr>
<td>\alpha</td>
<td>\beta</td>
<td>\gamma</td>
<td>\delta</td>
</tr>
<tr>
<td>4%</td>
<td>86%</td>
<td>66%</td>
<td>28%</td>
</tr>
<tr>
<td>2%</td>
<td>100%</td>
<td>86%</td>
<td>50%</td>
</tr>
<tr>
<td>1%</td>
<td>94%</td>
<td>78%</td>
<td>32%</td>
</tr>
<tr>
<td>0.5%</td>
<td>98%</td>
<td>88%</td>
<td>50%</td>
</tr>
<tr>
<td>0.25%</td>
<td>100%</td>
<td>90%</td>
<td>60%</td>
</tr>
<tr>
<td>0.125%</td>
<td>94%</td>
<td>60%</td>
<td></td>
</tr>
<tr>
<td>\langle t \rangle</td>
<td>1369</td>
<td>2428</td>
<td>4635</td>
</tr>
</tbody>
</table>

SP-y: good performance also on random max-3-sat (Phys. Rev. E 70, 036107 (2004))
SP = BP on a modified combinatorial problem

New state "*": variable not forced (unfrozen) in a cluster

\[ \sigma_i \in \{-1, 1, \} \implies s_i \in \{-1, *, 1, \} \]

A configuration in \(\{-1, *, 1\}^N\) is a solution if:

(i) all the clauses are satisfied: each clause should have at least one true literal or one *

(ii) all the so-called frozen variables \(\{+1,-1\}\) should be constrained by at least one clause (a variable is constrained by a clause when it is the only one that agrees with the clause)

(iii) no * variable should be constrained by any clause.
K-SAT ($\{-1,+1\}^N$) \hspace{1cm} Modified problem ($\{-1,*,+1\}^N$)

\[ F = \prod_{a \in A} C_a \] \hspace{1cm} \[ \Rightarrow \] \hspace{1cm} \[ G = \prod_i V_i. \]

\[ C_a = 1 - E_a, \]

\[ V_i = \delta_{s_i, *} \prod_{a \in \hat{i}} C_a^{-1} \left( -1 \right)^{s_i} + \sum_{\sigma = \pm 1} \delta_{s_i, \sigma} \prod_{a \in \hat{i}} C_a^{\sigma} \left( 1 - \prod_{a \in \hat{i}} C_a^{-\sigma} \right) \]

\[ C_a^{i,x}(s) = C_a(s^{(i,x)}) \]

$s^{(i,x)}$ such that $s^{(i,x)}_i = x$ and $s^{(i,x)}_j = s_j$ for $j \neq i$. 
\[ \hat{E} = \sum_{a=1}^{M} \hat{E}_a + \sum_{i=1}^{N} A_i \]

Geometrical constraints enforcing the “don’t care” state:

\[ A_i = \delta_{s_i,*} \left( 1 - \delta_{E_i^{-1},E_i^1} \right) + \sum_{\sigma=\pm 1} \delta_{s_i,\sigma} \theta \left( E_i^{\sigma} - E_i^{-\sigma} \right) \]

\[ E_i^{\sigma} = \sum_{a \in i} \left( 1 - C_{a,\sigma}^i \right) \]
duality
Simple proven statements

(I) Equivalent (1-RSB) / (extended RS) iterations:

\[ P_{a \rightarrow i}^{dbp} (\{ P_{k \rightarrow b}^\tau \}) \equiv P_{a \rightarrow i}^\tau (\{ P_{j \rightarrow a}^{sp} \}) \]

(II) extended RS marginals \sim\ SP marginals

\[ P_{i}^{BP} (s_i = -1, *, 1) = P_{i}^{SP} (H_i < 0, H_i = 0, H_i > 0) \]

(III) Bethe entropy \sim 1-RSB complexity (SP):

\[ S = \sum \]
\[ S = \sum_{\{s_i = -1, *, 1\}} P(\{s_i\}) \log P(\{s_i\}) \ , \ P = \prod_a P_{a} \prod_i P_{i}^{1-n_i} \]

(IV) SP over the dual graph: 2-RSB message-passing algorithms

(C. Baldassi, RZ, 05)

Related work: E. Maneva, E. Mossel, M. Wainwright, SODA 05 (next talk!)
Statistical analysis of SP equilibrium equations

A. Braunstein, C. Moore, T. Mora, R.Z. (05)

First moment entropy

\[ \sum' \]

1-RSB complexity

just one root

unphysical branch

2.40502
SP equations at finite N:

- clustering operation based on the "*" state

![hyper-cubic covers of clusters]

- complete enumeration results on finite (N~150) random systems show that "solutions in the extended space which are SP-stable are extensible to solutions of the original problem"

\[ (-1, 1, *, 1, 1, *, *, 1) \Rightarrow (-1, 1, -1, 1, 1, -1, -1, 1) \]

**Problem:** starting from a solution we can “whiten” \((\pm 1 \rightarrow *)\) some variables. Quite often (for small K) the all “*” state is reached!

**Question:** can the Local Weak Convergence method say something about this RS problem?
Exploring clusters: SP with external driving surveys

Correct but inefficient:

\[ E = E_{SAT} + \frac{\lambda}{\beta} \sum_{i} \sigma_i \xi_i \]
\[ Z = \sum_{\sigma|\sigma=F} e^{-\lambda} \sum_{i} \sigma_i \xi_i \]

Efficient approximation

\[ Q_i^{\text{ext}}(u) = (1 - \eta) \delta(u) + \eta \delta(u - \xi_i) \]

\[ dH(\sigma^{SP}, \xi) < dH_{\text{typical}} \]
Applicati

Source Coding (lossy data compression)

Shannon’s bound: $R < 1/(1-H[D])$

Source: $x = \{x_i = 0, 1\}, \ i = 1, ..., N$

$$P(x) = \prod_{i} P_i(x_i), \quad P_i(x_i) = (1 - b_i)\delta(x_i) + b\delta(x_i - 1) \ (b_i = 1/2)$$

Encoder: $y = \{y_j = 0, 1\}, \ j = 1, ...M$

Decoder: $x' = \{x'_i = 0, 1\}, \ i = 1, ..., N$

$$R = \frac{1}{1 - H[D]}$$

$$D(x, x') = \frac{1}{N} \sum_i (1 - \delta_{x_i, x'_i})$$

$R = N/M > 1, \ H[a] = -a \log a - (1-a) \log(1-a)$
Packing of pure states & lossy compression

**Encoding**: SP with external local fields to find a cluster close to string X & store $1/R$ bits ("spins")

**Decoding**: SP with $N/R$ forced bits

\[
Q_{i}^{\text{ext}}(u) = (1 - \eta^{\text{ext}})\delta(u) + \eta^{\text{ext}}\delta(u - x_{i})
\]

(\(\eta^{\text{ext}} = 1 \Rightarrow \sigma_{i} = x_{i}\))

\[
P_{i}(H) = c_{i} \int \prod_{b \in V(i)} du_{b} Q_{b \rightarrow i}(u_{b}) du Q_{i}^{\text{ext}}(u)\delta(H - \sum_{b} u_{b} - u)\chi[\{u_{b}\}, u]
\]