References

R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, L. Troyansky,
Determining computational complexity from characteristic 'phase transitions'

O.C. Martin, R. Monasson, R. Zecchina,
Statistical mechanics methods and phase transitions in optimization problems

D. Achlioptas, A. Chtcherba, G. Istrate, C. Moore
The Phase Transition in 1-in-k SAT and NAE 3-SAT

Y. Gao, J. Culberson
The Phase Transition in NK Landscapes is Easy

Phases and transitions

- Liquid-gas transition
  - at T_c coexistence of liquid and gas phases
  - jump in order parameter $\Delta\rho$
  - first-order phase transition
- Ferromagnetic-paramagnetic transition
  - at T_c no phase coexistence
  - order parameter (magnetization) changes continuously
  - second-order phase transition

Random K-SAT

- N boolean variables x_i, $i = 1, ..., N$
- Clauses C are formed of K randomly chosen variables, each of them negated with probability $1/2$
- $M = \alpha N$ independently drawn random clauses C
- 2-SAT exhibits phase transition at $\alpha_c(K = 2) = 1$ (exact value)
- $P_N(\alpha, K)$ – prob. to find satisfying set of x_i
  - $P_N \to 1$ for $\alpha < \alpha_c(K)$
  - $P_N \to 0$ for $\alpha > \alpha_c(K)$ when $N \to \infty$
- For low values of α the problem is underconstrained and for high values overconstrained.

- A sat-unsat threshold exists for any value of K.
- Its exact location is a difficult open problem.
- For characteristic heuristic search 'easy-hard-easy' pattern in computational costs:
  - low values of α – easy to find satisfying assignment
  - α close to α_c – difficult to find satisfying assignment or show unsatisfiability
  - $\alpha > \alpha_c$ exponential in N with a coefficient of N decreasing as a power law in α ($K > 2$)
- Tools for analytical studies from statistical physics.

OPTIMIZATION	STATISTICAL PHYSICS
instance	sample
cost function	energy
optimal configuration	ground state
minimal cost	ground state energy

Partition function

- The system is in configuration C with probability $p(C)$

$$p(C) = \frac{1}{Z} \exp\left(-\frac{1}{T} E(C)\right)$$

where E is the energy and Z is the partition function

$$Z = \sum_{C} \exp\left(-\frac{1}{T} E(C)\right)$$

Replica method

- Consider a generic model with N spins σ_i and an energy function $E(C, J)$ depending on a set of random couplings J
- Assume that $F(J)$ is self-averaging and we want to calculate its quenched average value $\overline{F(J)}$.
- Computation of $\overline{\ln Z(J)}$ is a very hard task
  $\rightarrow$ use replica method
- Start with the following expansion, valid for any J and small n:

$$Z(J)^n = 1 + n \ln Z(J) + O(n^2)$$

- Now the problem is to find the following limit

$$\overline{F(J)} = -T \lim_{n \to 0} \left(\frac{\overline{Z(J)^n} - 1}{n} \right)$$

- If we restrict to integer n

$$\overline{Z(J)^n} = \overline{\left[\sum_C \exp\left(-E(C,J)/T\right) \right]^n} = \sum_{C^1,\ldots,C^n} \overline{\exp\left(\frac{1}{T} \sum_{a=1}^{n} E(C^a, J) \right)}$$

- We have n copies, or replicas, of the initial problem.
- Random couplings disappear after averiging.
- We end up in computing the partition function of a system of N vectorial spins $\vec{\sigma}_i = (\sigma_i^1, ..., \sigma_i^n)$.

- To leading order in N, n^{th} moment of Z can be written as

$$\overline{Z(J)^n} = \exp(-Nf(n)/T)$$

  where $f(n)$ correspnds to the extremum of a functional optimization problem, solution of which can be saught in so-called replica symmetric (RS) subspace.

- At the final step one performs the analytic continuation $n \to 0$.

K-SAT energy and partition function

- Map boolean variables x_i to spins S_i
  $S_i = +1$ if x_i =true
  $S_i = -1$ if x_i =false
- Map random clauses into $M \times N$ matrix C_{li}
  $C_{li} = -1$ if C_l includes $\bar{x_i}$
  $C_{li} = +1$ if C_l includes x_i
  $C_{li} = 0$ otherwise
- $\sum_{i=1}^{N} C_{li} S_i =$ wrong literals in clause l
- Cost-function $E[\mathbf{C}, \mathbf{S}]$ defined as number of literals that not satisfied

$$E[\mathbf{C}, \mathbf{S}] = \sum_{l=1}^{M} \delta \left(\sum_{i=1}^{N} C_{li} S_i + K \right)$$

- The minimum (or ground state) $E[\mathbf{C}]$ is a random variable that concentrates around $E_{GS} \equiv \overline{E[\mathbf{C}]}$ as $N \to \infty$
- $E_{GS} = 0$ in sat region and $E_{GS} > 0$ in unsat region.
- The knowledge of E_{GS} as a function of α determines $\alpha_c(K)$.

$$E_{GS} = -T\overline{\log Z[\mathbf{C}]} + O(T^2)$$

when $T \to 0$ and where

$$Z[\mathbf{C}] = \sum_{\mathbf{S}} \exp(-E[\mathbf{C}, \mathbf{S}]/T)$$

- E_{GS} can be found with the replica method.

- Define average of spin S_i over all ground state configurations

$$m_i = \frac{1}{N_{GS}} \sum_{g=1}^{N_{GS}} S_i^g$$

- Clearly $-1 \leq m_i \leq +1$.
- $m_i = -1$ means that the corresponding x_i is always false in all ground states
- $m_i = +1$ means that the corresponding x_i is always true in all ground states

- The distribution $P(m)$ of all m_i describes the microscopic structure of the ground state.
- $P(m = \pm 1)$ represents a 'backbone' of completely constrained variables.
- $P(m \approx 0)$ describes weakly constrained variables.
- If the RS solution is global optimum of the problem arising in finding $f(n)$, the solution of this problem is $P(m)$
- Replica symmetric solution leads to an order parameter which is precisely $P(m)$.

Sat phase

- The typical number of solutions can be obtained from the ground state entropy density $s_{GS}(\alpha)$.
- $s_{GS}(\alpha = 0) = \ln 2$ ($F = E - ST$, $S(E) = \ln N(E)$)
- From Taylor expansion of s_{GS} around $\alpha = 0$:
  $s_{GS}(\alpha_c = 1) = 0.38$ for 2-SAT and
  $s_{GS}(\alpha = 4.2) = 0.1$ for 3-SAT
- Solutions are exponentially numerous.
- The RS calculations are belived to be exact at low α ratios.
- According to analytical (and numerical) calculations for 3-SAT the RS theory breaks down at $\alpha_{RSB} < \alpha_c$.
- At α_{RSB} the solutions start to be organized into distinct clusters.

- The space of N spins configuration – the N-dimensional hypercube.
- Optimal solutions are a subset of 2^N vertices of the hypercube.
- RS assumes that any pair of vertices are separated by the same Hamming distance d – the fraction of distinct spins. Solutions are in a single cluster of diameter dN.
- This holds for $\alpha < \alpha_{RSB}$ where solutions are characterized by a single $P(m)$.
- At $\alpha_{RSB} \simeq 4.0$ the space of solutions breaks into a large number (polynomial in N) of diffrent clusters.

$\rightarrow$ space of solutions has a highly organized structure

Unsat phase

- It is expected that $O(N)$ variables become totally constrained.
- $P(m)$ develops Dirac peaks at $m = \pm 1$.
- Taking into account this effect RS gives $\alpha_c(2) = 1$ correctly.
- $\alpha_c(3) \simeq 4.6$ is slightly larger than the value from numerical simulations $\alpha_c(3) \simeq 4.25$.
- The RS theory provides an upper bound for the thresholds for any $K > 2$.

Abrupt vs. smooth phase transition

- An qualitative difference between 2-SAT and 3-SAT is the way $P(m)$ changes at the threshold.
- This discrepancy can be seen in the fraction $\gamma(K, \alpha)$ of boolean variables that become fully constrained, at and above the threshold.
- $\gamma(K, \alpha)$ is directly computable in the RS theory.
- $\gamma(K, \alpha)$ vanishes in the sat region.
- Two kinds of scenarii when entering the unsat region:
  1. For 2-SAT $\gamma(2, \alpha)$ smoothly increases above the threshold.
  2. For 3-SAT $\gamma(3, \alpha)$ has a discontinuous jump to a finite value γ_c slightly above the threshold.

$(2+p)$-SAT model

- A mixed model, which continuously interpolates between 2-SAT and 3-SAT.
- A fraction p (resp. $1-p$) clauses of length three (resp. two).
- $p = 0$ corresponds to 2-SAT and $p = 1$ to 3-SAT.
- The sat-unsat transition becomes abrupt when $p > p_0 \simeq 0.4$.
- When $p < p_0$ the transition is smooth.
- For $p < p_0$ $(2+p)$-SAT shares the physical features of 2-SAT problem and for $p > p_0$ of 3-SAT.

Scaling of computational costs

- Numerical experiments: easy-hard-easy pattern for the typical search cost with a peak of complexity close to α_c.
- The peak scales polynomially with N for 2-SAT and exponentially with N for 3-SAT.
- This suggests the connection between the nature of the phase transition and the computational costs.

1-in-K SAT

- Exact location of threshold for $K \geq 3$ at $\alpha_{1,K} = 1/\binom{K}{2}$.
- However, there is no jump in phase transition.

NAE 3-SAT

- NAE caluse (a, b, c) is equvalent to $(a \vee b \vee c) \wedge (\bar{a} \vee \bar{b} \vee \bar{c})$
- According to numerical experiments it seems that:
  $2 \times \alpha_c(3)^{NAE} = \alpha_c(3)^{SAT}$

NK landscapes

- An NK landscape $f(x) = \sum_{i=1}^{n} f_i(x_i, \Pi(x_i))$, where $n > 0$ and $x = (x_1, ..., x_n) \in \{0,1\}^n$
  f_i local fitness function
  neighbourhood of x_i: $\Pi(x_i) \subset \{x_1, ..., x_n\} \backslash \{x_i\}$
- Main parameters: n and $k = |\Pi(x_i)|$
- Consider the random neighbourhood where the k variables are chosen randomly from the set $\{x_1, ..., x_n\} \backslash \{x_i\}$.
- 'Is the optimum of $f(x)$ equal to n?' is NP-complete for $k \geq 2$.

NK landscapes with random neighbourhoods

- In the uniform probability model $\overline{N}(n,k,p)$ $f_i(y)=0$ with probability $p$and $f_i(y)=1$ with probability $1-p$, where $y \in Dom(f_i)=\{0,1\}^{k+1}$.
- The fixed ratio model $N(n,k,z)$: the parameter z takes on values from $[0,2^{k+1}]$.
  - z is integer: z tuples of possible assignments $Y=(y_1,...,y_z)$ from $Dom(f_i)=\{0,1\}^{k+1}$ and defining $f_i(y)=0$ if $y \in Y$, $f_i(y)=1$ otherwise.
  - non-integer $z=(1-\alpha)[z]+\alpha[z+1]$ choose randomly $[(1-\alpha)n]$ local fitness functions and determine their values according to $N(n,k,[z])$. The rest of f_i according to $N(n,k,[z]+1)$.

The uniform probability model:

- For any $p(n)$ such that $\lim_n p(n)n^{1/(2^{k+1})}$ exists, k fixed, there is a polynomial algrithm that finds solution with probability asymptotic to 1 as $n \to \infty$.
- If p does not decrease very quickly with n, then asymptotically there will be at least one $f_i = 0$, making the whole decision problem insoluble.

The fixed ratio model:

- Upper bounds for insolubility $z > 3$ and $z > 2.837$.
- Polynomial algorthims also for the insoluble phase.
- Experiments: the poblem is also easy around and below the threshold.

$\to$ is this 'smooth' transition?