2016 IEEE 57th Annual Symposium on Foundations of Computer Science

The number of solutions for random regular NAE-SAT

Allan Sly, Nike Sun, and Yumeng Zhang
Department of Statistics
University of California
Berkeley, CA 94720-3860
Email: {sly,nikesun,ymzhang}@berkeley.edu

***Abstract*—Recent work has made substantial progress in understanding the transitions of random constraint satisfaction problems (CSPs). In particular, for several of these models, the exact satisfiability threshold has been rigorously determined, confirming predictions from the statistical physics literature. Here we revisit one of these models, random regular NAE-SAT: knowing the satisfiability threshold, it is natural to study, in the satisfiable regime, the number of solutions in a typical instance. We prove here that these solutions have a well-defined free energy (limiting exponential growth rate), with explicit value matching the one-step replica symmetry breaking prediction. The proof develops new techniques for analyzing a certain "survey propagation model" associated to this problem. We believe that these methods may be applicable in a wide class of related problems.**

***Keywords*-random constraint satisfaction problem, free energy, replica symmetry, one-step replica symmetry breaking, satisfiability threshold, condensation transition**

I. Introduction

In a general random constraint satisfaction problem (CSP), there are n variables taking values in a finite alphabet $\mathcal{X}$, subject to a random collection of constraints. In previous works on models of this kind, it has emerged that the space of solutions — a random subset of $\mathcal{X}^n$ — can have a complicated structure, posing major obstacles to mathematical analysis.

On this front, major advances were achieved by statistical physicists, who developed powerful analytic heuristics to shed light on the behavior of random CSPs ([1] and references therein). Their insights and methods are fundamental to the current understanding of random CSPs.

One prominent application of the physics heuristic is in giving explicit predictions for the locations of satisfiability thresholds in a large class of random CSPs ([2] and others). Some of these thresholds are established rigorously in recent works [3], [4], [5].

However, the satisfiability threshold is only one aspect of the rich picture that physicists have developed. There are deep conjectures for the behavior of these models inside the satisfiable regime, and it remains an outstanding mathematical challenge to prove them. In this paper we address one part of this challenge, concerning the total number of solutions for a typical instance in the satisfiable regime.

A. Main result

Given a CNF boolean formula, a *not-all-equal*-SAT (hereafter NAE-SAT) solution is an assignment $\underline{\boldsymbol{x}}$ of literals to variables such that both $\underline{\boldsymbol{x}}$ and its negation $\neg\underline{\boldsymbol{x}}$ evaluate to TRUE — equivalently, such that no clause gives the same evaluation to all its variables. A k-NAE-SAT problem is one in which each clause has exactly k literals; it is termed d-*regular* if each variable appears in exactly d clauses. Sampling such a formula in a uniformly random manner gives rise to the *random d-regular k-NAE-SAT model*. We refer to [6] for important early work on the closely related model of random (Erdős–Rényi) NAE-SAT. The appeal of this model is that it has certain symmetries making the analysis particularly tractable, yet it is expected to share most of the interesting qualitative phenomena exhibited by other commonly studied problems, including random k-SAT and random graph colorings.

Following convention, we fix k and then parametrize the model by its clause-to-variable ratio, $\alpha = d/k$. The *partition function* of the model, denoted $Z \equiv Z_n$, is simply the number of valid NAE-SAT assignments for an instance on n variables. It is conjectured that for each $k \geqslant 3$, the model has an exact satisfiability threshold $\alpha_{\text{sat}}(k)$: for $\alpha < \alpha_{\text{sat}}$ it is satisfiable (Z is positive) with high probability, but for $\alpha > \alpha_{\text{sat}}$ it is unsatisfiable (Z is zero) with high probability. (An event is said to hold *with high probability* if its probability tends to one in the limit $n \to \infty$, with k and α fixed.) This has been proved [3] for all k exceeding an absolute constant k_0, together with an explicit formula for α_{sat} which matches the physics prediction. The exact formula (described in [3]) is rather intricate so we omit it here, and note only its approximate value

$$\alpha_{\text{sat}} = \left(2^{k-1} - \frac{1}{2} - \frac{1}{4\ln 2}\right)\ln 2 + \epsilon_k \tag{1}$$

where ϵ_k denotes an error tending to zero as $k \to \infty$.

We say the model has *free energy* $\mathsf{f}(\alpha)$ if $Z^{1/n}$ converges to $\mathsf{f}(\alpha)$ in probability as $n \to \infty$. *A priori*, the limit may not be well-defined. If it exists, however, Markov's inequality and Jensen's inequality imply that it must be upper bounded by the *replica symmetric free energy*

$$\mathsf{f}^{\text{RS}}(\alpha) \equiv (\mathbb{E}Z)^{1/n} = 2(1 - 2/2^k)^\alpha. \tag{2}$$

0272-5428/16 $31.00 © 2016 IEEE
DOI 10.1109/FOCS.2016.82

An intriguing prediction from the physics analysis [7], [8] is that there is a critical value α_{cond} strictly below α_{sat}, such that $\mathsf{f}(\alpha)$ and $\mathsf{f}^{\text{RS}}(\alpha)$ agree up to $\alpha = \alpha_{\text{cond}}$ and diverge thereafter. Since f^{RS} is analytic, f must be non-analytic at α_{cond}. This is the *condensation* or *Kauzmann transition*, to be further described below. For $\alpha \in (\alpha_{\text{cond}}, \alpha_{\text{sat}})$ the conjecture is that $\mathsf{f}(\alpha)$ takes a value $\mathsf{f}^{\text{1RSB}}(\alpha)$ strictly below $\mathsf{f}^{\text{RS}}(\alpha)$. See Figure 1. The function $\mathsf{f}^{\text{1RSB}}(\alpha)$ is explicit though not simple: it is derived via the heuristic of *one-step replica symmetry breaking* (1RSB), and is presented below in Definition I.4.

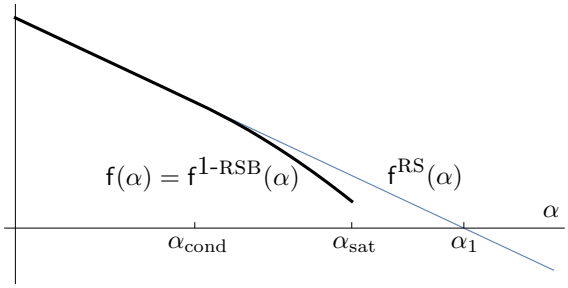

Figure 1: The free energy f is the quantity of interest, while f^{RS} and f^{1RSB} are two candidate explicit formulas for f. For a class of models, the conjecture is that all three functions agree up to α_{cond}, beyond which $\mathsf{f} = \mathsf{f}^{\text{1RSB}}$ diverges from f^{RS}.

Our main result is to prove this prediction for large k:

Theorem 1.
In random regular k-NAE-SAT with $k \geqslant k_0$, for all $\alpha < \alpha_{\text{sat}}(k)$ the free energy $\mathsf{f}(\alpha)$ exists and equals the predicted value $\mathsf{f}^{\text{1RSB}}(\alpha)$.

Remark I.1.
We allow for k_0 to be adjusted as long as it remains an absolute constant (so it need not equal the k_0 from [3]). The result of Theorem 1 is already proved [3] for

$$\alpha \leqslant \alpha_{\text{lbd}} \equiv (2^{k-1} - 2)\ln 2,$$

so we restrict our attention to $\alpha \in (\alpha_{\text{lbd}}, \alpha_{\text{sat}})$, which is a strict superset of the condensation regime $(\alpha_{\text{cond}}, \alpha_{\text{sat}})$.

Of course, for $\alpha > \alpha_{\text{sat}}$, we already know $\mathsf{f}(\alpha) = 0$. The case $\alpha = \alpha_{\text{sat}}$ can arise only if $d_{\text{sat}}(k) \equiv k\alpha_{\text{sat}}(k)$ is integer-valued for some k. We have no reason to believe that this ever occurs; if however it does miraculously occur then the probability for Z to be positive is bounded away from both zero and one [3]. In this situation, our methods would show that $Z^{1/n}$ does not concentrate around a single value but rather on two values, zero and $\lim_{\alpha \uparrow \alpha_{\text{sat}}} \mathsf{f}^{\text{1RSB}}(\alpha)$.

The condensation transition has been actively studied in recent work. The existence of a condensation phenomenon was first established for random NAE-SAT [9], and has since been found in random regular NAE-SAT and independent set [3], [4]. It has been demonstrated to occur even at positive temperature in the problem of hypergraph bicoloring (which is very similar to NAE-SAT) [10]. However, determining the precise location of α_{cond} is challenging, and was first achieved for the random graph coloring model [11] by an impressive and technically challenging analysis. Subsequent work pinpoints α_{cond} for random regular k-SAT (which again is very similar to NAE-SAT) [12]. The main contribution of this paper is to determine for the first time the free energy throughout the condensation regime $(\alpha_{\text{cond}}, \alpha_{\text{sat}})$.

In the remainder of this extended abstract we present some of the physics intuition for this problem, and give an overview of our approach. The proof appears in the full version of this paper, which is available online.
(http://arxiv.org/abs/1604.08546).

B. Statistical physics predictions

According to the statistical physics heuristic, the random regular NAE-SAT model has exactly one level of replica symmetry breaking (1RSB). We refer to [13, Ch. 19] for an expository account. We now summarize some of the key phenomena that are predicted from the 1RSB framework [7], [1], [8]. While part of the following discussion remains conjectural, much of it is rigorously established by the present paper. For this discussion we focus on the leading exponential terms and ignore $\exp\{o(n)\}$ corrections.

Take the NAE-SAT model with k, d fixed, and write $\alpha \equiv d/k$. Abbreviate $0 \equiv$ TRUE, $1 \equiv$ FALSE. For small α, almost all of the solutions lie in a single well-connected subset of $\{0,1\}^n$. This holds until a *clustering transition* α_{clust}, above which the solution space becomes broken up into exponentially many well-separated components, or *clusters*. For k large, α_{clust} is very small relative to α_{sat}. For α above α_{clust}, the number of clusters of size $\exp\{ns\}$ has mean value $\exp\{n\Sigma(s;\alpha)\}$, and further is concentrated about this mean; Σ is sometimes termed the "(entropic) cluster complexity

function." It is common to abbreviate $\Sigma(s) \equiv \Sigma(s;\alpha)$. Summing this prediction over cluster sizes s gives that the total number Z of NAE-SAT solutions has mean

$$\mathbb{E}Z \doteq \sum_s \exp\{n[s+\Sigma(s)]\} \doteq \exp\{n[s_1+\Sigma(s_1)]\},$$

where $\doteq$ indicates equality up to $\exp\{o(n)\}$ factors, and

$$s_1 = \arg\max[s+\Sigma(s)].$$

It is predicted that the function Σ is continuous and strictly concave in s, and that $s+\Sigma(s)$ has a unique maximizer s_1 with $\Sigma'(s_1) = -1$. Note the implicit dependence $s_1 = s_1(\alpha)$, and $\Sigma(s_1) = \Sigma(s_1(\alpha);\alpha)$.

Under the 1RSB framework, physicists propose an explicit (conjectural) formula for Σ. For NAE-SAT and related models, this explicit calculation reveals another critical value $\alpha_{\text{cond}} \in (\alpha_{\text{clust}}, \alpha_{\text{sat}})$, characterized as

$$\alpha_{\text{cond}} = \inf\{\alpha \geqslant \alpha_{\text{clust}} : \Sigma(s_1(\alpha);\alpha) < 0\}.$$

For $\alpha > \alpha_{\text{cond}}$, $\mathbb{E}Z$ is dominated by clusters of size $\exp\{ns_1\}$, whose mean number $\exp\{n\Sigma(s_1)\}$ is exponentially small, meaning they are highly unlikely to appear in a typical realization. Instead, a typical realization is dominated by clusters of size $s_{\max}$ where

$$s_{\max} \equiv s_{\max}(\alpha) \equiv \arg\max\{s+\Sigma(s) : \Sigma(s) \geqslant 0\}.$$

Since $\Sigma(s_{\max}) = 0$, it follows that with high probability

$$Z \doteq \exp\{n[s_{\max}+\Sigma(s_{\max})]\} = \exp\{ns_{\max}\}.$$

According to this picture, we will have (with high probability) that $Z \doteq \mathbb{E}Z$ for $\alpha \leqslant \alpha_{\text{cond}}$, while $Z \ll \mathbb{E}Z$ for $\alpha > \alpha_{\text{cond}}$. See Figure 2.

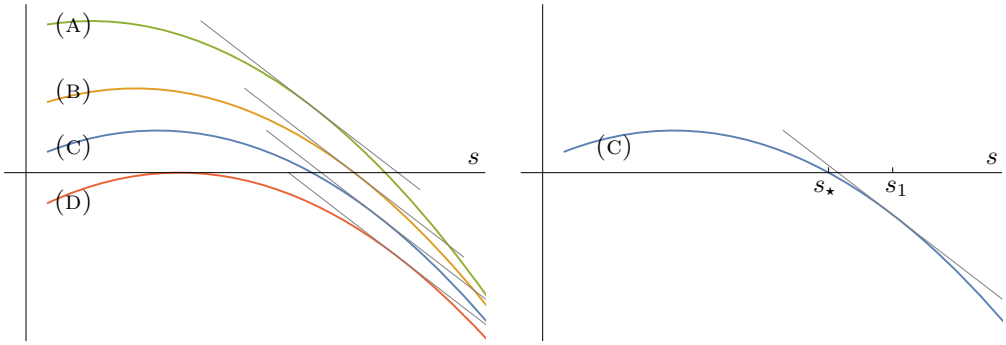

Figure 2: The number of clusters of size roughly $\exp\{ns\}$ concentrates around its mean value $\exp\{n\Sigma(s)\}$. The left panel shows $\Sigma(s) \equiv \Sigma(s;\alpha)$ as a function of s for four different values of α, together with the tangent lines of slope -1. In increasing order of α, the curves indicate (A) $\alpha_{\text{clust}} < \alpha < \alpha_{\text{cond}}$, (B) $\alpha = \alpha_{\text{cond}}$, (C) $\alpha_{\text{cond}} < \alpha < \alpha_{\text{sat}}$, and (D) $\alpha = \alpha_{\text{sat}}$. The right panel shows curve (C) only and indicates the locations of $s_\star$ and s_1.

Thus, for $\alpha > \alpha_{\text{cond}}$, the first moment $\mathbb{E}Z$ fails to capture the typical behavior of Z. This difficulty persists up to and beyond the satisfiability threshold

$$\alpha_{\text{sat}} = \inf\{\alpha \geqslant \alpha_{\text{cond}} : \max_s \Sigma(s;\alpha) < 0\}$$

— indeed, it is well known that there is a non-trivial interval $(\alpha_{\text{sat}}, \alpha_1)$ in which $\mathbb{E}Z \gg 1$ even though $Z = 0$ with high probability.

C. The tilted cluster partition function

Once the function $\Sigma(s;\alpha)$ is determined, it becomes straightforward to derive α_{cond}, α_{sat}, and $\mathsf{f}(\alpha)$. However, prior works have not taken the approach of actually computing Σ. Indeed, α_{sat} was determined [3] by an analysis involving only $\max_s \Sigma(s;\alpha)$, which contains less information than the full curve Σ. In related models, the determination of α_{cond} [11], [12] also avoids Σ, going instead through the so-called "planted model." In order to obtain Σ, consider the λ-tilted partition function

$$\boldsymbol{Z}_\lambda \equiv \sum_{\boldsymbol{\gamma}} |\boldsymbol{\gamma}|^\lambda \tag{3}$$

where the sum is taken over all clusters $\boldsymbol{\gamma}$. According to the physics heuristic as described above, $\mathbb{E}\boldsymbol{Z}_\lambda \doteq \exp\{n\mathfrak{F}(\lambda)\}$ where $\mathfrak{F}$ is the Legendre dual of $-\Sigma$:

$$\mathfrak{F}(\lambda) \equiv (-\Sigma)^\star(\lambda) \equiv \max_s[\lambda s + \Sigma(s)].$$

The physics approach to computing Σ is to first compute $\mathfrak{F}$, and then use the involutive property of the Legendre transform to recover Σ:

$$\Sigma = -\mathfrak{F}^\star.$$

Note that by twice differentiating

$$\mathfrak{F}(\lambda) = n^{-1} \ln \mathbb{E} \boldsymbol{Z}_\lambda$$

with respect to λ, we find that $\mathfrak{F}$ is convex in λ, so the resulting Σ will indeed be concave.

The computation of $\mathfrak{F}(\lambda)$ may seem at first glance quite intractable. Indeed, the reason for NAE-SAT solutions to occur in clusters is that a typical solution has a positive density of variables which are *free*, meaning their value can be changed without violating any clause. Each cluster (connected component of NAE-SAT solutions, where two solutions are connected if they differ by a single bit) may be a complicated subset of $\{0,1\}^n$ — changing the value at one free variable may affect whether its neighbors are free, so a cluster need not be a simple subcube of $\{0,1\}^n$. We then wish to sum over the cluster sizes raised to non-integer powers.

However, in the regime of interest $\alpha \geqslant \alpha_{\text{lbd}}$ (see Remark I.1), the analysis of NAE-SAT solution clusters is greatly simplified by the fact that in a typical satisfying assignment the vast majority of variables are *frozen* rather than free. The result of this, roughly speaking, is that a cluster can be encoded by a configuration $\underline{x} \in \{0,1,\mathtt{f}\}^n$ (representing its circumscribed subcube, so $x_v = \mathtt{f}$ indicates a free variable) with no essential loss of information. We call $\underline{x}$ the *frozen configuration* representing the cluster. It turns out that the frozen configurations can be regarded as the solutions of a certain CSP lifted from the original NAE-SAT problem — so the physics heuristics can be applied again to the new CSP. Variations on this idea appear in several places in the physics literature; in the specific context of random CSPs we refer to [14], [15], [16].

Analyzing the *number* of frozen configurations — corresponding to (3) with $\lambda = 0$ — leads to the sharp satisfiability threshold for this model [3]. To analyze (3) for general λ requires a deeper investigation of the arrangement of free and frozen variables in the frozen configurations $\underline{x}$. In fact, the majority of free variables are simply isolated vertices. A smaller fraction occur in linked pairs, and a yet smaller fraction occur in components of size three or more. Each free component $\boldsymbol{T}$ is surrounded by frozen variables, and we let $z(\boldsymbol{T})$ count the number of NAE-SAT assignments on $\boldsymbol{T}$ which are consistent with the frozen boundary. Then the total size of the cluster represented by $\underline{x}$ is simply the product of $z(\boldsymbol{T})$ over all the free components $\boldsymbol{T}$ of $\underline{x}$.

The random NAE-SAT graph has few short cycles, so almost all of the free components are *trees*, and so their weights $z(\boldsymbol{T})$ can be evaluated recursively by the method of *belief propagation* (BP). To implement this, we must replace variable spins by "messages," which are indexed by the directed edges of the graph and so are more natural for tree recursions. The message $\mathtt{m}_{v\to a}$ from variable v to clause a represents the state of v "in absence of a." It is also necessary to introduce a richer alphabet of symbols for these messages, replacing $\{0,1,\mathtt{f}\}$ with probability measures on $\{0,1\}$ (where any non-degenerate measure will project to $\mathtt{f}$). Thus the message $\mathtt{m}_{v\to a}$ represents the distribution at v (within the cluster) in absence of clause a. The messages are related to one another via local consistency equations, which are precisely the BP equations. The configuration $\underline{\mathtt{m}}$ encodes the same cluster as $\underline{x}$, with the key advantage that *the cluster size can be readily deduced from* $\underline{\mathtt{m}}$*, as a certain product of local functions.* For the cluster size raised to power λ, simply raise each local function to power λ. Thus the configurations $\underline{\mathtt{m}}$ with λ-tilted weights form a *spin system* (Markov random field), whose partition function is the quantity of interest (3). The new spin system is sometimes termed the "auxiliary model" [13, Ch. 19].

D. One-step replica symmetry breaking

Above, we asserted informally that each BP solution $\underline{\mathtt{m}}$ encodes a cluster of NAE-SAT solutions. An important caveat is that this is only rigorous if the free variables in $\underline{\mathtt{m}}$ occur in trees, separated by frozen regions where we must have messages $\mathtt{m}_{v\to a}$ that are degenerate (supported on either on $\mathtt{0}$ or on $\mathtt{1}$). Otherwise, one always has the trivial "replica symmetric" BP solution where every $\mathtt{m}_{v\to a}$ is $\mathrm{unif}(\{0,1\})$, and this is not a "meaningful" solution for large α. One way to understand this is via the physics calculation of $\mathsf{f}^{\text{RS}}(\alpha)$, which we now describe by way of motivating the more complicated expression for $\mathsf{f}^{\text{1RSB}}(\alpha)$.

Given a random regular NAE-SAT instance $\mathscr{G}$ on n variables, choose k uniformly random variables $v_1,\dots,v_k$, and assume for simplicity that no two of these share a clause. Then (1) remove the k variables along with their kd incident clauses, producing an instance $\mathscr{G}''$, and (2) add $d(k-1)$ new clauses to $\mathscr{G}''$, producing $\mathscr{G}'$. Then $\mathscr{G}'$ is distributed as a random regular NAE-SAT instance on $n-k$ variables. If the free energy exists, then

$$\mathsf{f}(\alpha)^n \doteq Z \doteq \left(\frac{Z(\mathscr{G})}{Z(\mathscr{G}')}\right)^{n/k}. \tag{4}$$

Suppose u is a variable in $\mathscr{G}'$ of degree $d-1$, meaning it was a neighbor of a clause a which was deleted from $\mathscr{G}$. The interpretation of $\underline{\mathtt{m}}$ is that in $\mathscr{G}''$, the spin at u has law $\mathtt{m}_{u\to a}$, and the different $u's$ are independent. If every $\mathtt{m}_{u\to a}$ is $\mathrm{unif}(\{0,1\})$, then

$$\begin{aligned}\left(\frac{Z(\mathscr{G})}{Z(\mathscr{G}'')}\right)^{1/k} &= 2(1-2/2^k)^d,\\ \left(\frac{Z(\mathscr{G}')}{Z(\mathscr{G}'')}\right)^{1/k} &= (1-2/2^k)^{\alpha(k-1)},\end{aligned} \tag{5}$$

Taking the ratio of these and substituting into (4) gives the prediction $\mathsf{f}(\alpha) \doteq \mathsf{f}^{\text{RS}}(\alpha)$, which we know to be false for large α. Thus the replica symmetric $\underline{\mathtt{m}}$ gives the incorrect prediction. The reason for this failure is that in reality the u's are *not* independent in $\mathscr{G}''$, but rather are significantly

correlated even though they are typically far apart in $\mathscr{G}''$. This phenomenon of long-range dependence may be taken as a definition of replica symmetry breaking, and it is expected to occur precisely for $\alpha > \alpha_{\text{cond}}$.

The idea of 1RSB is that, in passing from the original NAE-SAT model to the (seemingly far more complicated) "auxiliary model" of weighted BP solutions, we in fact return to replica symmetry, provided $\Sigma(s_\lambda)$ is positive for

$$s_\lambda \equiv \arg\max_s\{\lambda s + \Sigma(s)\}. \tag{6}$$

That is, for such λ, the auxiliary model is predicted to have correlation decay, in contrast with the long-range correlations of the original model. The implication is that in this context, the above heuristic ((4) and (5)) *is* expected to yield the correct answer. The replica symmetric BP solution for the auxiliary model will be a certain measure $\dot{q}_\lambda$ over messages $\mathtt{m}$. Taking $\dot{q}_{v\to a} \equiv \dot{q}_\lambda$ is the precise analogue, in the auxiliary model, of taking $\mathtt{m}_{v\to a} \equiv \text{unif}(\{0,1\})$ on every $v \to a$ in the original model. Under the assumption that the auxiliary model has strong correlation decay, (4) and (5) give an expression for $\mathfrak{F}(\lambda)$ in terms of $\dot{q}_\lambda$.

E. The 1RSB free energy prediction

Having described the heuristic reasoning, we now proceed to formally state the 1RSB free energy prediction.

We first describe $\dot{q}_\lambda$ as a certain discrete probability measure over $\mathtt{m}$. Since $\mathtt{m}$ is a probability measure over $\{0,1\}$, we can encode it by a single real number $x \equiv \mathtt{m}(1) \in [0,1]$. A measure q on $\mathtt{m}$ can thus be encoded by an element $\mu \in \mathscr{P}$ where $\mathscr{P}$ is the set of discrete probability measures on $[0,1]$.

Now, for any measurable $B \subseteq [0,1]$, define

$$\begin{aligned}
\hat{\mathscr{R}}_\lambda\mu(B) &\equiv \hat{\mathscr{Z}}(\mu)^{-1}\int\left(2-\prod_{i=1}^{k-1}x_i-\prod_{i=1}^{k-1}(1-x_i)\right)^\lambda \mathbf{1}\left\{\frac{1-\prod_{i=1}^{k-1}x_i}{2-\prod_{i=1}^{k-1}x_i-\prod_{i=1}^{k-1}(1-x_i)}\in B\right\}\prod_{i=1}^{k-1}\mu(dx_i),\\
\dot{\mathscr{R}}_\lambda\mu(B) &\equiv \dot{\mathscr{Z}}(\mu)^{-1}\int\left(\prod_{i=1}^{d-1}y_i+\prod_{i=1}^{d-1}(1-y_i)\right)^\lambda \mathbf{1}\left\{\frac{\prod_{i=1}^{d-1}y_i}{\prod_{i=1}^{d-1}y_i+\prod_{i=1}^{d-1}(1-y_i)}\in B\right\}\prod_{i=1}^{d-1}\mu(dy_i),
\end{aligned} \tag{7}$$

where $\hat{\mathscr{Z}}(\mu)$ and $\dot{\mathscr{Z}}(\mu)$ are the normalizing constants such that $\hat{\mathscr{R}}_\lambda\mu$ and $\dot{\mathscr{R}}_\lambda\mu$ are also probability measures on $[0,1]$. (For $\lambda = 0$ we make the convention that $0^0 = 0$.) Denote

$$\mathscr{R}_\lambda \equiv \dot{\mathscr{R}}_\lambda \circ \hat{\mathscr{R}}_\lambda : \mathscr{P} \to \mathscr{P}.$$

The map $\mathscr{R}_\lambda$ represents the BP recursion for the auxiliary model. We now present a fixed point of this recursion in the regime

$$(2^{k-1}-2)\ln 2 \equiv \alpha_{\text{lbd}} \leqslant \alpha \leqslant \alpha_{\text{ubd}} \equiv 2^{k-1}\ln 2,$$

which we recall is a superset of $(\alpha_{\text{cond}}, \alpha_{\text{sat}})$.

Definition I.2.
For any $\lambda \in [0,1]$, let $\dot{\mu}_{\lambda,l} \in \mathscr{P}$ be the sequence of probability measures defined by

$$\dot{\mu}_{\lambda,0} \equiv \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1, \quad \text{and} \quad \dot{\mu}_{\lambda,l+1} = \mathscr{R}_\lambda\dot{\mu}_{\lambda,l}$$

for all $l \geqslant 0$.

To specify the topology of convergence, let

$$S_l \equiv (\operatorname{supp}\dot{\mu}_{\lambda,l})\backslash(\operatorname{supp}(\dot{\mu}_{\lambda,0}+\ldots+\dot{\mu}_{\lambda,l-1})),$$

so S_l is a finite subset of $[0,1]$. Regard $\dot{\mu}_{\lambda,l}$ as an infinite sequence indexed by the elements of S_1 in increasing order, followed by the elements of S_2 in increasing order, and so on. We then have the following convergence result:

Proposition I.3.
For $k \geqslant k_0$ and $\alpha_{\text{lbd}} \leqslant \alpha \leqslant \alpha_{\text{ubd}}$, in the limit $l \to \infty$, $\dot{\mu}_{\lambda,l}$ converges in the ℓ^1 sequence space to a limit $\dot{\mu}_\lambda \in \mathscr{P}$. The limit is a fixed point of the recursion, $\dot{\mu}_\lambda = \mathscr{R}_\lambda\dot{\mu}_\lambda$. It satisfies the symmetry condition $\dot{\mu}_\lambda(dx) = \dot{\mu}_\lambda(d(1-x))$. It is mostly supported on $\{0,1\}$ with $\dot{\mu}_\lambda((0,1)) \leqslant 7/2^k$.

The limit $\dot{\mu}_\lambda$ of Proposition I.3 encodes the desired replica symmetric solution $\dot{q}_\lambda$ for the auxiliary model. We can then express $\mathfrak{F}(\lambda)$ in terms of $\dot{\mu}_\lambda$ as follows. Writing $\hat{\mu}_\lambda \equiv \mathscr{R}_\lambda\dot{\mu}_\lambda$, let $\dot{w}_\lambda, \hat{w}_\lambda, \bar{w}_\lambda \in \mathscr{P}$ be defined by

$$\begin{aligned}
\dot{w}_\lambda(B) &= (\dot{\mathfrak{Z}}_\lambda)^{-1}\int\left(\prod_{i=1}^{d}y_i+\prod_{i=1}^{d}(1-y_i)\right)^\lambda \mathbf{1}\left\{\prod_{i=1}^{d}y_i+\prod_{i=1}^{d}(1-y_i)\in B\right\}\prod_{i=1}^{d}\hat{\mu}_\lambda(dy_i),\\
\hat{w}_\lambda(B) &= (\hat{\mathfrak{Z}}_\lambda)^{-1}\int\left(1-\prod_{i=1}^{k}x_i-\prod_{i=1}^{k}(1-x_i)\right)^\lambda \mathbf{1}\left\{1-\prod_{i=1}^{k}x_i-\prod_{i=1}^{k}(1-x_i)\in B\right\}\prod_{i=1}^{k}\dot{\mu}_\lambda(dx_i),\\
\bar{w}_\lambda(B) &= (\bar{\mathfrak{Z}}_\lambda)^{-1}\iint\Big(xy+(1-x)(1-y)\Big)^\lambda \mathbf{1}\Big\{xy+(1-x)(1-y)\in B\Big\}\dot{\mu}_\lambda(dx)\hat{\mu}_\lambda(dy),
\end{aligned} \tag{8}$$

with $\dot{\mathfrak{Z}}_\lambda, \hat{\mathfrak{Z}}_\lambda, \bar{\mathfrak{Z}}_\lambda$ the normalizing constants. The analogue of (5) for this model is

$$\left(\frac{\boldsymbol{Z}_\lambda(\mathscr{G})}{\boldsymbol{Z}_\lambda(\mathscr{G}'')}\right)^{1/k} = \dot{\mathfrak{Z}}_\lambda(\hat{\mathfrak{Z}}_\lambda/\bar{\mathfrak{Z}}_\lambda)^d,$$
$$\left(\frac{\boldsymbol{Z}_\lambda(\mathscr{G}')}{\boldsymbol{Z}_\lambda(\mathscr{G}'')}\right)^{1/k} = (\hat{\mathfrak{Z}}_\lambda)^{\alpha(k-1)},$$

and substituting this into (4) gives the 1RSB prediction:

$$\boldsymbol{Z}_\lambda \doteq \exp\{\mathfrak{F}(\lambda)\}$$

with high probability, where

$$\mathfrak{F}(\lambda) \equiv \mathfrak{F}(\lambda;\alpha) \equiv \ln \dot{\mathfrak{Z}}_\lambda + \alpha \ln \hat{\mathfrak{Z}}_\lambda - k\alpha \ln \bar{\mathfrak{Z}}_\lambda. \tag{9}$$

Further, the maximizer of (6) is predicted to be given by

$$s_\lambda \equiv s_\lambda(\alpha) \equiv \int \ln(x)\dot{w}_\lambda(dx) + \alpha \int \ln(x)\hat{w}_\lambda(dx) - k\alpha \int \ln(x)\bar{w}_\lambda(dx). \tag{10}$$

If $s = s_\lambda$ for $\lambda \in [0,1]$ we define

$$\Sigma(s) \equiv \Sigma(s;\alpha) \equiv \mathfrak{F}(\lambda;\alpha) - \lambda s_\lambda(\alpha).$$

This yields the predicted thresholds

$$\alpha_{\text{cond}} \equiv \sup\{\alpha : \Sigma(s_1;\alpha) > 0\},$$
$$\alpha_{\text{sat}} \equiv \sup\{\alpha : \Sigma(s_0;\alpha) > 0\},$$

and we can now formally state the predicted free energy of the original NAE-SAT model:

Definition I.4.
For $\alpha \in k^{-1}\mathbb{Z}$, 1RSB free energy prediction $\mathsf{f}^{\text{1RSB}}(\alpha)$ is defined as

$$\mathsf{f}^{\text{1RSB}}(\alpha) = \begin{cases} \mathsf{f}^{\text{RS}}(\alpha) = 2(1-2/2^k)^\alpha & \alpha \leqslant \alpha_{\text{cond}}, \\ \exp[\sup\{s : \Sigma(s) \geqslant 0\}] & \alpha_{\text{cond}} \leqslant \alpha < \alpha_{\text{sat}}, \\ 0 & \alpha > \alpha_{\text{sat}}. \end{cases} \tag{11}$$

(In regular k-NAE-SAT we must have integer $d = k\alpha$, so we need not consider $\alpha \notin k^{-1}\mathbb{Z}$.)

Proposition I.5.
Consider $\alpha \in A \equiv [\alpha_{\text{lbd}}, \alpha_{\text{ubd}}] \cap (k^{-1}\mathbb{Z})$. For $k \geqslant k_0$ and $\alpha \in A$, the function $\Sigma(s) \equiv \Sigma(s;\alpha)$ is well-defined, continuous, and strictly decreasing in s, so that $\mathsf{f}^{\text{RS}}(\alpha)$ is well-defined.

Proposition I.6.
For $k \geqslant k_0$ and $\lambda \in [0,1]$,

$$\Sigma(s_\lambda;\alpha) \equiv \mathfrak{F}(\lambda) - \lambda s_\lambda$$

is strictly decreasing as a function of $\alpha \in A$. There is a unique $\alpha_\lambda \in A$ such that $\Sigma(s_\lambda;\alpha)$ is non-negative for all $\alpha \leqslant \alpha_\lambda$, and is negative for all $\alpha > \alpha_\lambda$. In particular

$$\alpha_{\text{cond}} = \alpha_1 = (2^{k-1}-1)\ln 2 + \text{err},$$
$$\alpha_{\text{sat}} = \alpha_0 = \left(2^{k-1} - \frac{1}{2} - \frac{1}{4\ln 2}\right)\ln 2 + \text{err}.$$

We remark that the asymptotic expansion of α_{sat} matches the previously mentioned result (1) from [3]. The asymptotic expansion of α_{cond} matches an earlier result of [17], which was obtained for a slightly different but closely related model.

F. Proof approach

Since $\mathsf{f} = \mathsf{f}(\alpha)$ is *a priori* not well-defined, the statement $\mathsf{f} \leqslant \mathsf{g}$ means formally that for all $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(Z^{1/n} \geqslant \mathsf{g} + \epsilon) = 0.$$

With this notation in mind, we will prove separately the upper bound $\mathsf{f}(\alpha) \leqslant \mathsf{f}^{\text{1RSB}}(\alpha)$ and the matching lower bound $\mathsf{f}(\alpha) \geqslant \mathsf{f}^{\text{1RSB}}(\alpha)$. This implies the main result Theorem 1: the free energy $\mathsf{f}(\alpha)$ is indeed well-defined, and equals $\mathsf{f}^{\text{1RSB}}(\alpha)$.

The upper bound is proved by an interpolation argument. This builds on similar bounds for spin glasses on Erdős–Rényi graphs [18], [19], together with ideas from [20] for interpolation in random regular models. Write $Z_n(\beta)$ for the partition function of NAE-SAT at inverse temperature $\beta > 0$. The interpolation method yields an upper bound on $\mathbb{E}\ln Z_n(\beta)$ which is expressed as the infimum of a certain function $\mathcal{P}(\mu;\beta)$, with μ ranging over probability measures on $[0,1]$. We then choose μ according to Proposition I.3, and take $\beta \to \infty$ to obtain the desired bound $\mathsf{f}(\alpha) \leqslant \mathsf{f}^{\text{1RSB}}(\alpha)$.

Most of the paper is devoted to establishing the matching lower bound. The proof is inspired by the physics picture described above, and at a high level proceeds as follows. Take any λ for which the (predicted) value of $\Sigma(s_\lambda)$ is non-negative, and let $\boldsymbol{Y}_\lambda$ be the number of clusters of size $\doteq \exp\{ns_\lambda\}$. The informal statement of what we show is that

$$\boldsymbol{Y}_\lambda \doteq \exp\{n[\lambda s_\lambda + \Sigma(s_\lambda)]\}. \tag{12}$$

Adjusting λ as indicated by (11) then proves the desired bound $\mathsf{f}(\alpha) \geqslant \mathsf{f}^{\text{1RSB}}(\alpha)$.

Proving a formalized version of (12) occupies a significant part of the present paper. We introduce a slightly modified version of the messages $\mathtt{m}$ which record the topologies of the free trees $\boldsymbol{T}$. We then restrict to free trees with fewer than T variables, which limits the distance that information can propagate between free variables. We prove a version of (12) for every fixed T, and show that this yields the sharp lower bound in the limit $T \to \infty$. The proof of (12) for fixed T is via the moment method for the auxiliary model, which boils down to a complicated optimization problem over many dimensions. It is known (see e.g. [3, Lem. 3.6]) that stationary points of the optimization problem correspond to "generalized" BP fixed points — these are measures $Q_{v\to a}(\mathtt{m}_{v\to a}, \mathtt{m}_{a\to v})$, rather than the simpler "one-sided" measures $q_{v\to a}(\mathtt{m}_{v\to a})$ considered in the 1RSB heuristic.

The one-sided property is a crucial simplification, but is challenging to prove in general. One contribution of this work that we wish to highlight is a novel resampling argument which yields a reduction to one-sided messages,

and allows us to solve the moment optimization problem. (We are helped here by the truncation on the sizes of free trees.) Furthermore, the approach allows us to bring in methods from large deviations theory. With these we can show that the objective function has negative-definite Hessian at the optimizer, which is necessary for the second moment method. This resampling approach is quite general and should apply in a broad range of models.

G. Open problems

Beyond the free energy, it remains a challenge to establish the full picture predicted by statistical physicists for $\alpha \leqslant \alpha_{\text{sat}}$. We refer the reader to several recent works targeted at a broad class of models in the regime $\alpha \leqslant \alpha_{\text{cond}}$ [21], [22], [23]. In the condensation regime $(\alpha_{\text{cond}}, \alpha_{\text{sat}})$, an initial step would be to show that most solutions lie within a bounded number of clusters. A much more refined prediction is that the mass distribution among the largest clusters forms a Poisson–Dirichlet process. Another question is to show that on a typical problem instance over n variables, if $\underline{\boldsymbol{x}}^1, \underline{\boldsymbol{x}}^2$ are sampled independently and uniformly at random from the solutions of that instance, then the normalized overlap

$$R_{1,2} \equiv n^{-1}\{v : \boldsymbol{x}_v^1 = \boldsymbol{x}_v^2\}$$

concentrates on two values (corresponding roughly to the two cases that $\underline{\boldsymbol{x}}^1, \underline{\boldsymbol{x}}^2$ come from the same cluster, or from different clusters). This criterion is sometimes taken as the precise definition of 1RSB, and so would be quite interesting to prove for models in the condensation regime.

Beyond the immediate context of random CSPs, understanding the condensation transition may deepen our understanding of the stochastic block model, a model for random networks with underlying community structure. Here again ideas from statistical physics have played an important role [24]. A great deal is now known rigorously for the case of two blocks [25], [26], where there is no condensation regime. For models with more than two blocks, however, it is predicted that the condensation can occur, and may define a regime where detection is information-theoretically possible but computationally intractable.

Acknowledgements

We thank Amir Dembo, Jian Ding, Andrea Montanari, and Lenka Zdeborová for many helpful conversations. We are also grateful for the hospitality of the Simons Institute at Berkeley where part of this work was completed. A.S. acknowledges support from NSF grants (DMS-1208338 and DMS-1352013), and the Sloan Fellowship. N.S. acknowledges support from the NSF MSPRF grant (DMS-1401123).

References

[1] F. Krząkała, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, "Gibbs states and the set of solutions of random constraint satisfaction problems," *Proc. Natl. Acad. Sci. USA*, vol. 104, no. 25, pp. 10 318–10 323 (electronic), 2007. [Online]. Available: http://dx.doi.org/10.1073/pnas.0703685104

[2] S. Mertens, M. Mézard, and R. Zecchina, "Threshold values of random k-SAT from the cavity method," *Random Structures Algorithms*, vol. 28, no. 3, pp. 340–373, 2006.

[3] J. Ding, A. Sly, and N. Sun, "Satisfiability threshold for random regular NAE-SAT," in *Proc. 46th STOC*. New York, NY, USA: ACM, 2014.

[4] ——, "Maximum independent sets on random regular graphs," arXiv:1310.4787, 2013.

[5] ——, "Proof of the satisfiability conjecture for large k," in *Proc. 47th STOC*. New York: ACM, 2015, pp. 59–68.

[6] D. Achlioptas and C. Moore, "Random k-SAT: two moments suffice to cross a sharp threshold," *SIAM J. Comput.*, vol. 36, no. 3, pp. 740–762 (electronic), 2006. [Online]. Available: http://dx.doi.org/10.1137/S0097539703434231

[7] L. Zdeborova and F. Krząkała, "Phase transitions in the coloring of random graphs," *Phys. Rev. E*, vol. 76, no. 3, p. 031131, 2007.

[8] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian, "Clusters of solutions and replica symmetry breaking in random k-satisfiability," *J. Stat. Mech.*, vol. 2008, no. 04, p. P04004, 2008. [Online]. Available: http://stacks.iop.org/1742-5468/2008/i=04/a=P04004

[9] A. Coja-Oghlan and K. Panagiotou, "Catching the k-NAE-SAT threshold," in *Proc. 44th STOC*. New York: ACM, 2012, pp. 899–907. [Online]. Available: http://dx.doi.org/10.1145/2213977.2214058

[10] V. Bapst, A. Coja-Oghlan, and F. Raßmann, "A positive temperature phase transition in random hypergraph 2-coloring," arXiv:1410.2190, 2014.

[11] V. Bapst, A. Coja-Oghlan, S. Hetterich, F. Raßmann, and D. Vilenchik, "The condensation phase transition in random graph coloring," *Comm. Math. Phys.*, vol. 341, no. 2, pp. 543–606, 2016. [Online]. Available: http://dx.doi.org/10.1007/s00220-015-2464-z

[12] V. Bapst and A. Coja-Oghlan, "The condensation phase transition in the regular k-SAT model," arXiv:1507.03512, 2015.

[13] M. Mézard and A. Montanari, *Information, physics, and computation*, ser. Oxford Graduate Texts. Oxford University Press, Oxford, 2009. [Online]. Available: http://dx.doi.org/10.1093/acprof:oso/9780198570837.001.0001

[14] G. Parisi, "On local equilibrium equations for clustering states," arXiv:cs/0212047, 2002.

[15] A. Braunstein, M. Mézard, and R. Zecchina, "Survey propagation: an algorithm for satisfiability," *Random Structures Algorithms*, vol. 27, no. 2, pp. 201–226, 2005. [Online]. Available: http://dx.doi.org/10.1002/rsa.20057

[16] E. Maneva, E. Mossel, and M. J. Wainwright, "A new look at survey propagation and its generalizations," *J. ACM*, vol. 54, no. 4, pp. Art. 17, 41, 2007. [Online]. Available: http://dx.doi.org/10.1145/1255443.1255445

[17] A. Coja-Oghlan and L. Zdeborová, "The condensation transition in random hypergraph 2-coloring," in *Proc. 23rd SODA*. ACM, New York, 2012, pp. 241–250.

[18] S. Franz and M. Leone, "Replica bounds for optimization problems and diluted spin systems," *J. Statist. Phys.*, vol. 111, no. 3-4, pp. 535–564, 2003. [Online]. Available: http://dx.doi.org/10.1023/A:1022885828956

[19] D. Panchenko and M. Talagrand, "Bounds for diluted mean-fields spin glass models," *Probab. Theory Related Fields*, vol. 130, no. 3, pp. 319–336, 2004. [Online]. Available: http://dx.doi.org/10.1007/s00440-004-0342-2

[20] M. Bayati, D. Gamarnik, and P. Tetali, "Combinatorial approach to the interpolation method and scaling limits in sparse random graphs," *Ann. Probab.*, vol. 41, no. 6, pp. 4080–4115, 2013. [Online]. Available: http://dx.doi.org/10.1214/12-AOP816

[21] V. Bapst and A. Coja-Oghlan, "Harnessing the Bethe free energy," arXiv:1504.03975, 2015.

[22] A. Coja-Oghlan, W. Perkins, and K. Skubch, "Limits of discrete distributions and Gibbs measures on random graphs," arXiv:1512.06798, 2015.

[23] A. Coja-Oghlan and W. Perkins, "Belief propagation on replica symmetric random factor graph models," arXiv:1603.08191, 2016.

[24] A. Decelle, F. Krząkała, C. Moore, and L. Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Phys. Rev. E*, vol. 84, no. 6, p. 066106, 2011.

[25] L. Massoulié, "Community detection thresholds and the weak Ramanujan property," in *Proc. 46th STOC*. New York: ACM, 2014, pp. 694–703.

[26] E. Mossel, J. Neeman, and A. Sly, "Reconstruction and estimation in the planted partition model," *Probab. Theory Related Fields*, vol. 162, no. 3-4, pp. 431–461, 2015. [Online]. Available: http://dx.doi.org/10.1007/s00440-014-0576-6