Storage capacity in symmetric binary perceptrons

Benjamin Aubin, Will Perkins, Lenka Zdeborova

Storage capacity in symmetric binary perceptrons

Benjamin Aubin,[1] Will Perkins,[2] and Lenka Zdeborová[1]
[1)] *Institut de physique théorique, Université Paris Saclay, CNRS, CEA Saclay, F-91191 Gif-sur-Yvette, France*
[2)] *Department of Mathematics, Statistics and Computer Science, University of Illinois, Chicago, USA*

(Dated: 3 January 2019)

We study the problem of determining the capacity of the binary perceptron for two variants of the problem where the corresponding constraint is symmetric. We call these variants the rectangle-binary-perceptron (RPB) and the $u-$function-binary-perceptron (UBP). We show that, unlike for the usual step-function-binary-perceptron, the critical capacity in these symmetric cases is given by the annealed computation in a large region of parameter space (for all rectangular constraints and for narrow enough $u-$function constraints, $K < K^*$). We prove this fact (under two natural assumptions) using the first and second moment methods. We further use the second moment method to conjecture that solutions of the symmetric binary perceptrons are organized in a so-called frozen-1RSB structure, without using the replica method. We then use the replica method to estimate the capacity threshold for the UBP case when the $u-$function is wide $K > K^*$. We conclude that full-step-replica-symmetry breaking would have to be evaluated in order to obtain the exact capacity in this case.

I. INTRODUCTION

In this paper we revisit the problem of computing the capacity of the binary perceptron[1,2] for storing random patterns. This problem lies at the core of early statistical physics studies of neural networks and their learning and generalization properties, for reviews see e.g.[3–6]. While the perceptron problem is motivated by studies of simple artificial neural networks as discussed in detail in the above literature, in this paper we view it as a random constraint satisfaction problem (CSP) where the vector of binary weights $\mathbf{w} \in \{\pm 1\}^N$ (a *solution*) must satisfy M *step* constraints of the type

$$\sum_{i=1}^{N} X_{\mu i} w_i \geq K \,, \tag{1}$$

where $\mu = 1, \dots, M$, $K \in \mathbb{R}$ is the *threshold*, the random variables $X_{\mu i}$ are *iid* Gaussian variables with zero mean and variance $1/N$, and the rows of the matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$ are called patterns. We define an indicator function associated to the perceptron with a step constraint as $\varphi^s(z) = \mathbb{1}_{z \geq K}$.

We say that a given vector $\mathbf{w}$ is a solution of the perceptron instance if all M constraints given by eq. (1) are satisfied. The *storage capacity* is then defined similarly to the satisfiability threshold in random constraint satisfaction problems: we denote the constraint density as $\alpha \equiv M/N$ and define the storage capacity $\alpha_c(K)$ as the infimum of densities α such that in the limit $N \to \infty$, with high probability (over the choice of the matrix $\mathbf{X}$) there are no solutions. It is natural to conjecture that the converse also holds, i.e. the storage capacity $\alpha_c(K)$ equals the supremum of α such that in the limit $N \to \infty$ solutions exist with high probability. In this case we would say the storage capacity is a *sharp threshold*.

Gardner and Derrida in their paper[1] assume the storage capacity $\alpha_c(K)$ is a sharp threshold and they apply the replica calculation to compute it, but reach a result inconsistent with a simple upper bound obtained by the first moment method. Mézard and Krauth[2] found a way to obtain a consistent prediction from the replica calculation and concluded that the storage capacity $\alpha_c^s(K)$ for the step binary perceptron (SBP), i.e. associated to the constraint φ^s, is given by the largest α for which the following quantity, the *entropy* in physics, is positive:

$$\phi^s_{\rm RS}(\alpha, K) = \mathrm{SP}_{q_0, \hat{q}_0} \left\{ \frac{1}{2}\left(q_0 - 1\right)\hat{q}_0 + \int Dt \log\left[2\cosh\left(t\sqrt{\hat{q}_0}\right)\right] + \alpha \int Dt \log\left[\int_{\frac{K - t\sqrt{q_0}}{\sqrt{1-q_0}}}^{\infty} Du\right] \right\} , \tag{2}$$

where $Dt = \int \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$ is a Gaussian measure, and SP stands for "saddle point" meaning that the expression is evaluated where the derivatives on the curl-bracket, with respect to $q_0 \geq 0$ and $\hat{q}_0 \geq 0$, is zero.

Several decades of subsequent research in the statistical physics of disordered systems are consistent with the conjectured Mézard-Krauth formula for the storage capacity of the binary perceptron. Despite the simplicity of the above conjecture and decades of impressive progress in the mathematics of spin glasses and related problems, (see e.g.[7–12] and many others), the storage capacity of the binary perceptron remains an open mathematical problem.

In fact, even the very existence of a sharp threshold, i.e. the fact that in the limit $N \to \infty$ the probability that patterns can be stored drops sharply from one to zero at the capacity, is an open problem. Up to very recently only widely non-matching upper bounds and lower bounds for the storage capacity of the binary perceptron were available[13,14]. As the present work was being finalized Ding and Sun[15] proved in a remarkable paper a lower bound on the capacity that matches the Krauth and Mezard conjecture (note that much like Theorem 4 below, the main theorem in[15] depends on a numerical hypothesis). A matching upper bound remains an open challenge in mathematical physics and probability theory.

In this paper we introduce two simple *symmetric* variants of the binary perceptron problem. Let $z_\mu(\mathbf{w}) = \sum_{i=1}^N X_{\mu i} w_i$. For a threshold $K \in \mathbb{R}^+$, we consider two different types of symmetric constraints:

- The rectangle binary perceptron (RBP) requires $|z_\mu| \leq K, \forall \mu = 1, \dots, M$. Its associated indicator function is $\varphi^r(z) = \mathbb{1}_{|z| \leq K}$.

- The u-function binary perceptron (UBP) requires $|z_\mu| \geq K, \forall \mu = 1, \dots, M$. Its associated indicator function is $\varphi^u(z) = \mathbb{1}_{|z| \geq K}$.

These constraints are symmetric in the sense that if $\mathbf{w}$ is a solution then $-\mathbf{w}$ is a solution as well.

The main result of the present paper, presented in section II, is a proof, subject to a numerical hypothesis, of a formula for the storage capacity, defined in the same way as for the step-function binary perceptron above. In particular, we show that in these symmetric variants the first moment upper bound (corresponding to the annealed capacity in physics) on the storage capacity is tight (except for $K > K^* \simeq 0.817$ for the UBP case). We prove this statement using the second moment method.

Let $Z \sim \mathcal{N}(0,1)$, and for $K \in \mathbb{R}^+$ let $p_{r,K} = \mathbb{P}[|Z| \leq K]$ and $p_{u,K} = \mathbb{P}[|Z| \geq K]$.

- The storage capacity for the rectangle binary perceptron is:

$$\alpha_c^r(K) = \frac{-\log(2)}{\log(p_{r,K})} \qquad \forall K \in \mathbb{R}^+ . \tag{3}$$

- The storage capacity for the $u-$function binary perceptron is:

$$\alpha_c^u(K) = \frac{-\log(2)}{\log(p_{u,K})} \qquad \text{for } \; 0 < K < K^* \simeq 0.817 . \tag{4}$$

The constant $K^* \simeq 0.817$ stems from the properties of the second moment entropy eq. (10). In the physics terms it is defined as the point of intersection between the annealed capacity $\alpha_a^u(K)$ and the local stability of the RS solution $\alpha_{\rm AT}^u(K)$ eq. (17). That is, K^* is the solution of the following equation:

$$\pi p_{u,K}^2 e^{K^2} \log(p_{u,K}) = -2\log(2) K^2 . \tag{5}$$

The two symmetric variants of the perceptron problem considered here share many of the intriguing geometric properties of the original step-function binary perceptron problem. Most significant is the conjectured frozen-1RSB[2] nature of the space of solutions that splits into well separated clusters of vanishing entropy at any $\alpha > 0$. Remarkably, this frozen-1RSB property can be deduced from the form of the second moment entropy as we explain in section III. Our justification of the frozen-1RSB property does not rely on the replica method and is hence of independent interest.

For the UBP and $K > K^*$, the second-moment proof technique fails, and this failure marks tightly the onset of the replica symmetry breaking region. In that region, we evaluate the one-step replica symmetry breaking (1RSB) approximation for the storage capacity, but conclude that full-step replica symmetry breaking (FRSB) would be needed to obtain the exact result. While the FRSB equations can be written along the lines of[16], they are more involved than the ones for the Sherrington-Kirkpatrick model[17–19], and solving them numerically or getting additional insight from them is a challenging task left for future work. We present the replica analysis in section IV. Table I contains the summary of our main results along with the predictions for the step-function perceptron.

Finally let us comment on the simpler and more commonly considered case of spherical perceptron where the binary constraint on the vector $\mathbf{w}$ is replaced by the spherical constraint $\mathbf{w}^\intercal \mathbf{w} = \sum_{i=1}^N w_i^2 = N$. For $K = 0$ the spherical perceptron reduces to the famous problem of intersection of half-spaces with capacity $\alpha_c = 2$ as solved by Wendell[20] and Cover[21]. For $K > 0$ the Gardner-Derrida solution[1] is correct as proven in[22,23]. For $K < 0$ the situation is more challenging and FRSB is needed to compute the storage capacity; for recent progress in physics see[16,24], while mathematical considerations about this case were presented in[25].

Binary perceptron	Constraint	Constraint function	Range of K	Storage capacity
Step-function	$z \geq K$	$\varphi^s(z) = \mathbb{1}_{z \geq K}$	$\forall K \in \mathbb{R}$	RS eq. (2)
Rectangle	$\lvert z\rvert \leq K$	$\varphi^r(z) = \mathbb{1}_{\lvert z\rvert \leq K}$	$\forall K \in \mathbb{R}^+$	Annealed eq. (3)
U-function	$\lvert z\rvert \geq K$	$\varphi^u(z) = \mathbb{1}_{\lvert z\rvert \geq K}$	$0 < K < K^* = 0.817$	Annealed eq. (4)
U-function	$\lvert z\rvert \geq K$	$\varphi^u(z) = \mathbb{1}_{\lvert z\rvert \geq K}$	$\forall K > K^* = 0.817$	FRSB?

TABLE I. This table summarizes results for storage capacity in binary perceptrons with different types of constraints. The result for canonical step-function is from[2]. The results for the rectangle and u-function are obtained in this paper.

II. PROOF OF CORRECTNESS OF THE ANNEALED CAPACITY

To state the main results precisely we introduce some definitions. Let $\mathbf{X}(N, M)$ be the random $M \times N$ pattern matrix. Define the partition functions

$$\mathcal{Z}_r(\mathbf{X}) = \sum_{\mathbf{w}\in\{\pm 1\}^N} \prod_{\mu=1}^{M} \varphi^r(z_\mu(\mathbf{w})) \quad \text{and} \quad \mathcal{Z}_u(\mathbf{X}) = \sum_{\mathbf{w}\in\{\pm 1\}^N} \prod_{\mu=1}^{M} \varphi^u(z_\mu(\mathbf{w}))\,,$$

which count respectively the number of solutions for the rectangle and $u-$function constraints respectively. Let $\mathcal{E}^r(N, M)$ and $\mathcal{E}^u(N, M)$ be the events that $\mathcal{Z}_r(\mathbf{X}) \geq 1$ and $\mathcal{Z}_u(\mathbf{X}) \geq 1$. We formally define the storage capacity.

Defintition 1. *The storage capacity $\alpha_c^r(K)$ is*

$$\alpha_c^r(K) = \inf\{\alpha : \lim_{N\to\infty} \mathbb{P}[\mathcal{E}^r(N, \lfloor \alpha N \rfloor)] = 0\}\,,$$

and likewise for $\alpha_c^u(K)$.

It is believed that there is a sharp threshold for the existence of solutions.

Conjecture 2. *The storage capacity is a sharp threshold:*

$$\alpha_c^r(K) = \sup\{\alpha : \lim_{N\to\infty} \mathbb{P}[\mathcal{E}^r(N, \lfloor \alpha N \rfloor)] = 1\}\,,$$

and likewise for $\alpha_c^u(K)$.

The corresponding conjecture for the random k-SAT model is the celebrated 'satisfiability threshold conjecture' proved for k large by Ding, Sly, and Sun[12].

Next, couple two standard Gaussians Z_1, Z_β by letting Z and Z' be independent standard Gaussians and setting $Z_1 = \sqrt{\beta}Z + \sqrt{1-\beta}Z'$ and $Z_\beta = \sqrt{\beta}Z - \sqrt{1-\beta}Z'$. Let

$$\begin{cases} q_{r,K}(\beta) &= \mathbb{P}[|Z_1| \leq K \wedge |Z_\beta| \leq K] = q_K(\beta)\,, \\ q_{u,K}(\beta) &= \mathbb{P}[|Z_1| \geq K \wedge |Z_\beta| \geq K] = 1 - 2p_{r,K} + q_K(\beta)\,, \end{cases} \tag{6}$$

with $q_K(\beta)$ the probability that two standard Gaussians with correlation $2\beta - 1$ are both at most K in absolute value, that is:

$$q_K(\beta) = \frac{1}{2\pi} \int_{-K}^{K} dy \int_{\frac{-K+(1-2\beta)y}{2\sqrt{\beta(1-\beta)}}}^{\frac{K+(1-2\beta)y}{2\sqrt{\beta(1-\beta)}}} e^{-\frac{x^2+y^2}{2}} dx\,.$$

Note that $q_{t,K}(1) = p_{t,K}$ and $q_{t,K}(1/2) = p_{t,K}^2$ for $t \in \{r, u\}$. We now introduce the functions that dictate the effectiveness of the second moment bound. Let

$$F_{r,K,\alpha}(\beta) = H(\beta) + \alpha \log q_{r,K}(\beta) \tag{7}$$
$$F_{u,K,\alpha}(\beta) = H(\beta) + \alpha \log q_{u,K}(\beta) \tag{8}$$

where $H(\beta) = -\beta \log \beta - (1-\beta)\log(1-\beta)$ is the Shannon entropy function.

We state a numerical hypothesis in terms of the derivatives of these two functions.

Hypothesis 3. *For all choices of $K > 0$ and $\alpha > 0$ so that $F''_{r,K,\alpha}(1/2) < 0$, there is exactly one $\beta \in (1/2, 1)$ so that $F'_{r,K,\alpha}(\beta) = 0$. The same holds for $F_{u,K,\alpha}$.*

Our main theorem is a proof, under Hypothesis 3, that the storage capacity is given by the annealed computation.

Theorem 4. *Under the assumption of Hypothesis 3, the following hold.*

1. *For all $K > 0$, we have $\alpha_c^r(K) = -\log(2)/\log(p_{r,K})$.*
2. *For all $K \in (0, K^*)$, we have $\alpha_c^u(K) = -\log(2)/\log(p_{u,K})$.*

Under our definition of $\alpha_c^r(K)$ and $\alpha_c^u(K)$, we must prove two statements to show that $\alpha_c^r(K) = -\log(2)/\log(p_{r,K})$ (and similarly for $\alpha_c^u(K)$). We use the first moment method to show that for $\alpha > -\log(2)/\log(p_{r,K})$, $\lim_{N\to\infty} \Pr(\mathcal{E}^r(N, M)) = 0$; then we use the second moment method to show that for $\alpha < -\log(2)/\log(p_{r,K})$, $\liminf_{N\to\infty} \Pr(\mathcal{E}^r(N, M)) > 0$ (a result analogous to what Ding and Sun prove for the more challenging step binary perceptron[15]). Conjecture 2 asserts the stronger statement that for $\alpha < -\log(2)/\log(p_{r,K})$, $\lim_{N\to\infty} \Pr(\mathcal{E}^r(N, M)) = 1$.

A. First moment upper bound

Proposition 5.

1. *If $\alpha > \alpha_a^r(K) = \frac{-\log(2)}{\log(p_{r,K})}$, then whp there is no satisfying assignment to the binary perceptron with the rectangle activation function.*
2. *If $\alpha > \alpha_a^u(K) = \frac{-\log(2)}{\log(p_{u,K})}$, then whp there is no satisfying assignment to the binary perceptron with the u-function activation function.*

Proof. We give the proof for the rectangle function as the proof for the u-function is identical. Let $\epsilon = \alpha - \alpha_a^r(K) > 0$. Let $\mathbf{1}$ denote the vector of dimension N with all 1 entries.

$$\begin{aligned}\mathbb{P}[\mathcal{E}^r(N, \alpha N)] &\le \mathbb{E}[\mathcal{Z}_r(\mathbf{X}(N, \alpha N))] = 2^N \mathbb{E}\left[\prod_{\mu=1}^{\alpha N} \mathbb{1}_{|z_\mu(\mathbf{1})| \le K}\right] = 2^N p_{r,K}^{\alpha N} = \exp(N(\log(2) + \alpha \log(p_{r,K})))\\ &= \exp(N\epsilon \log(p_{r,K})) \to 0 \text{ as } N \to \infty\,.\end{aligned}$$

□

B. Second moment lower bound

Proposition 6.

1. *If $\alpha < \frac{-\log(2)}{\log(p_{r,K})}$, then*

$$\liminf_{N\to\infty} \mathbb{P}[\mathcal{E}^r(N, \alpha N)] > 0.$$

2. *If $K < K^*$ and $\alpha < \frac{-\log(2)}{\log(p_{u,K})}$, then*

$$\liminf_{N\to\infty} \mathbb{P}[\mathcal{E}^u(N, \alpha N)] > 0.$$

To prove Proposition 6 we will apply the second-moment method in a similar fashion to Achlioptas and Moore[26] who determined the satisfiability threshold of random k-SAT to within a factor 2 by considering not-all-equal satisfying assignments (not-all-equal satisfiability (NAE-SAT) constraints are symmetric in the same way the rectangle and u-function constraints are symmetric). Recall the Paley-Zygmund inequality.

Lemma 7. *Let X be a non-negative random variable. Then*

$$\mathbb{P}[X > 0] \ge \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}\,.$$

We will also use the following application of Laplace's method from Achlioptas and Moore[26].

Lemma 8. *Let $g(\beta)$ be a real analytic function on $[0,1]$ and let*

$$G(\beta) = \frac{g(\beta)}{\beta^\beta (1-\beta)^{1-\beta}}\,.$$

If $G(1/2) > G(\beta)$ for all $\beta \neq 1/2$ and $G''(1/2) < 0$, then there exists constants c_1, c_2 so that

$$c_1 G(1/2)^N \leq \sum_{l=0}^{N} \binom{N}{l} g(l/N)^N \leq c_2 G(1/2)^N\,.$$

1. Rectangle binary perceptron

We calculate

$$\mathbb{E}[\mathcal{Z}_r(\mathbf{X})^2] = \sum_{\mathbf{w_1},\mathbf{w_2}\in\{\pm 1\}^N} \mathbb{P}[\mathbf{w_1},\mathbf{w_2} \text{ satisfying}] = 2^N \sum_{\mathbf{w}\in\{\pm 1\}^N} \mathbb{P}[\mathbf{1},\mathbf{w} \text{ satisfying}] = 2^N \sum_{l=0}^{N} \binom{N}{l} q_{r,K}(l/N)^{\alpha N}\,,$$

where we recall $q_{r,K}$ from eq. (6). Define

$$G_{r,K,\alpha}(\beta) \equiv \exp(F_{r,K,\alpha}(\beta)) = \frac{q_{r,K}(\beta)^\alpha}{\beta^\beta(1-\beta)^{1-\beta}}\,, \tag{9}$$

If we can show that $G_{r,K,\alpha}(1/2) > G_{r,K,\alpha}(\beta)$ for all $\beta \neq 1/2$ and $G''_{r,K,\alpha}(1/2) < 0$, then by Lemma 8, we have

$$\begin{aligned}\mathbb{E}[\mathcal{Z}_r(\mathbf{X})^2] &\leq c_2 4^N q_{r,K}(1/2)^{\alpha N} \\ &= c_2 4^N p_{r,K}^{2\alpha N}\,.\end{aligned}$$

Then since $\mathcal{Z}_r(\mathbf{X})$ is integer valued, we have

$$\begin{aligned}\mathbb{P}[\mathcal{Z}_r(\mathbf{X}) \geq 1] &\geq \frac{\mathbb{E}[\mathcal{Z}_r(\mathbf{X})]^2}{\mathbb{E}[\mathcal{Z}_r(\mathbf{X})^2]} = \frac{(2^N p_{r,K}^{\alpha N})^2}{\mathbb{E}[\mathcal{Z}_r(\mathbf{X})^2]} \\ &\geq \frac{(2^N p_{r,K}^{\alpha N})^2}{c_2 4^N p_{r,K}^{2\alpha N}} = 1/c_2 > 0\,.\end{aligned}$$

It remains to show that when $\alpha < \frac{-\log(2)}{\log(p_{r,K})}$, then $G_{r,K,\alpha}(1/2) > G_{r,K,\alpha}(\beta)$ for all $\beta \neq 1/2$ and $G''_{r,K,\alpha}(1/2) < 0$. By eq. (9) and the fact that $G'_{r,K,\alpha}(1/2) = 0$, it is enough to show the same for $F_{r,K,\alpha}$.

Certainly one necessary condition is that $F_{r,K,\alpha}(1/2) > F_{r,K,\alpha}(1)$. This reduces to the condition $2p_{r,K}^{2\alpha} > p_{r,K}^{\alpha}$ or $\alpha < \frac{-\log(2)}{\log(p_{r,K})}$ which is exactly the condition of Proposition 6. Next consider $F''_{r,K,\alpha}(1/2)$.

A calculation shows that

$$F''_{r,K,\alpha}(1/2) = 4\left(-1 + \frac{2}{\pi}\frac{\alpha K^2 e^{-K^2}}{p_{r,K}^2}\right)\,.$$

In particular, $F''_{r,K,\alpha}(1/2) < 0$ if and only if

$$\alpha < \frac{\pi}{2}\frac{p_{r,K}^2}{K^2 e^{-K^2}}\,.$$

But a calculation also shows that

$$-\frac{\log(2)}{\log(p_{r,K})} < \frac{\pi}{2}\frac{p_{r,K}^2}{K^2 e^{-K^2}}$$

for all $K > 0$ and so the condition of Proposition 6 implies that $F''_{r,K,\alpha}(1/2) < 0$.

Moreover, since $F_{r,K,\alpha}(\beta)$ is symmetric around $\beta = 1/2$ and it has a local maximum at $\beta = 1/2$, Hypothesis 3 implies that the global maximum of $F_{r,K,\alpha}(\beta)$ occurs at either $1/2$ or 1, and since $F_{r,K,\alpha}(1/2) > F_{r,K,\alpha}(1)$, we have that $F_{r,K,\alpha}(1/2) > F_{r,K,\alpha}(\beta)$ for all $\beta \neq 1/2$, completing the proof of Proposition 6 for the rectangle binary perceptron.

2. u-function binary perceptron

The proof for the u-function is similar. We can calculate

$$\mathbb{E}[\mathcal{Z}_u(\mathbf{X})^2] = 2^N \sum_{l=0}^{N} \binom{N}{l} q_{u,K}(l/N)^{\alpha N} = \exp\left(N(\log(2) + F_{u,K,\alpha}(\beta))\right) ,$$

where we recall $q_{u,K}$ from eq. (6). Using Lemma 8 and Hypothesis 3 again, it suffices to show that for $0 < K < K^*$ and $\alpha < \frac{-\log(2)}{\log(p_{u,K})}$ we have $F_{u,K,\alpha}(1/2) > F_{u,K,\alpha}(1)$ and $F''_{u,K,\alpha}(1/2) < 0$. The first follows immediately from the fact that $\alpha < \frac{-\log(2)}{\log(p_{u,K})}$. For the second, we have

$$F''_{u,K,\alpha}(1/2) = 4\left(-1 + \frac{2}{\pi}\frac{\alpha K^2 e^{-K^2}}{p_{u,K}^2}\right)$$

and so $F''_{u,K,\alpha}(1/2) < 0$ if and only if

$$\alpha < \frac{\pi}{2}\frac{p_{u,K}^2}{K^2 e^{-K^2}} .$$

Unlike with the rectangle function it is not true that

$$-\frac{\log(2)}{\log(p_{u,K})} < \frac{\pi}{2}\frac{p_{u,K}^2}{K^2 e^{-K^2}} \tag{10}$$

for all K: the left and right sides of the inequality cross at $K = K^*$, which implicitly defines K^*. Thus for $K < K^*$ and $\alpha < -\frac{\log(2)}{\log(p_{u,K})}$ we have $F''_{u,K,\alpha}(1/2) < 0$, which completes the proof of Proposition 6 for the u-function binary perceptron.

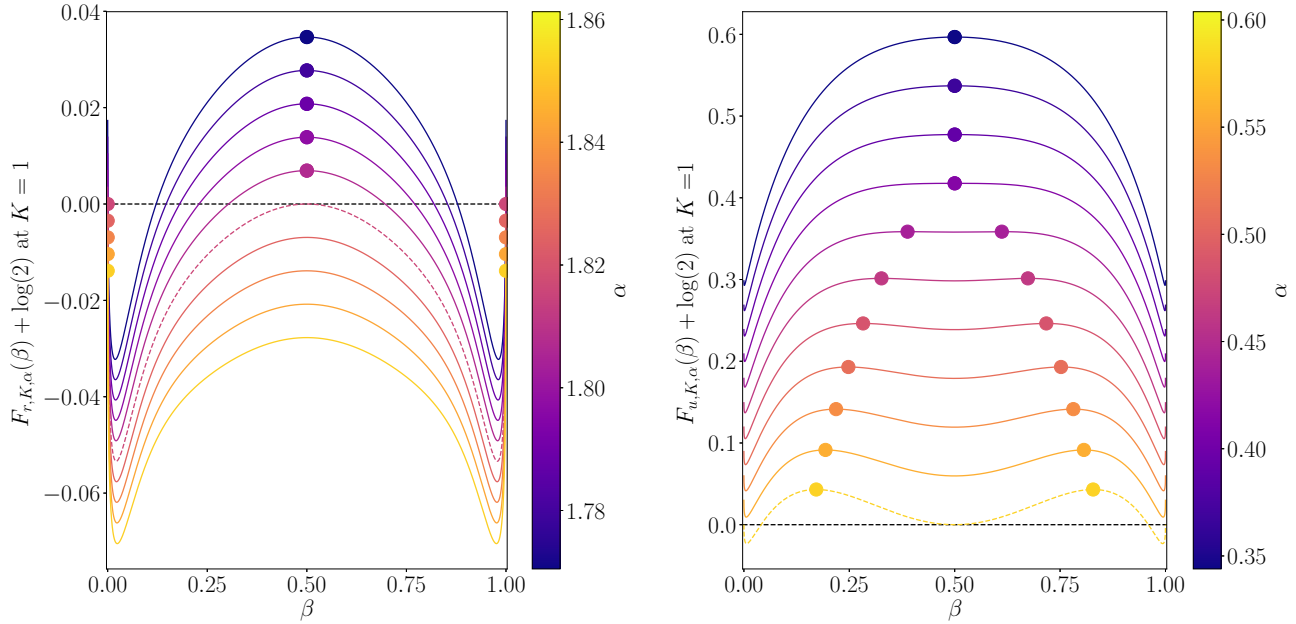

FIG. 1. Second moment entropy densities. **a)**: the rectangle binary perceptron for $\alpha \le \alpha_a^r = 1.816$ (dashed pink), $\beta = \frac{1}{2}$ is the global maximizer. For $\alpha \ge \alpha_a^r$, $\beta = 0$ and $\beta = 1$ are the maximizers. **b)**: the u-function binary perceptron for $\alpha \le \alpha^* = 0.430$, $\beta = \frac{1}{2}$ is the maximizer while for $\alpha^* \le \alpha \le \alpha_a^u = 0.604$ (dashed yellow), the maximizer is non-trivial $\beta \neq 0$.

3. Illustration

As an illustration, we plot the second moment entropy density $\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}[\mathcal{Z}_t^2] = \log(2) + F_{t,K,\alpha}$ for $t \in \{r, u\}$ at $K = 1 > K^*$ in fig. 1. For the rectangle function (**a**), the second moment is tight: the maximum is reached for

$\beta = 1/2$ for all α smaller than the first moment α_a^r (dashed pink). Exactly the same happens for the $u-$function with $K < K^*$. However for $K > K^*$, the second moment method fails (**b**): $\beta = 1/2$ becomes a minimum and the maximum is obtained for non trivial values $\beta \neq 1/2$ for constraint density smaller than the first moment α_a^u (dashed yellow).

III. FROZEN-1RSB STRUCTURE OF SOLUTIONS IN BINARY PERCEPTRONS

One of the most striking properties of the canonical step-function perceptron is the predicted frozen-1RSB[2] nature of the space of solutions. This means that the dominant (measure tending to one) part of the space of solutions splits into well separated clusters each of which has vanishing entropy density at any $\alpha > 0$. This frozen-1RSB scenario and quantitative properties of the solution space were studied in detail recently[27,28]. Following up on conjectures that such a frozen structure of solutions implies computational hardness in diluted constraint satisfaction problems[29], it was argued that finding a satisfying assignment in the binary perceptron should also be algorithmically hard since its solution space is dominated by clusters of vanishing entropy density[28]. Yet this conjecture contradicted empirical results of[30]. This paradox was resolved in[31] where the authors identified that there are subdominant parts (i.e. parts of measure converging to zero as the system size diverges) of the solution space that form extended clusters with large local entropy and all the algorithms that work well always find a solution belonging to one of those large-local-entropy clusters. These sub-dominant clusters are not frozen and somewhat strangely are not captured in the canonical 1RSB calculation[31]. It was argued that existence of these large-local-entropy clusters bears more general consequences on the dynamics of learning algorithms in neural networks, see e.g.[32].

While frozen-1RSB structure has also been identified in constraint satisfaction problems on sparse graphs[33,34], we want to note that its nature in the binary perceptron is of a rather different nature. In sparse systems a simple argument using expansion properties of the underlying graph and properties of the constraints show that each cluster with high probability contains only one solution. In the perceptron model, which has a fully connected bipartite interaction graph, this argument from sparse models does not apply.

In the present paper, we deduce from the second moment calculation of the previous section that the space of solutions in the symmetric binary perceptrons is also of the frozen-1RSB type and this property moreover extends to any finite temperature (with energy being defined as the number of unsatisfied constraints). This is different from the locked constraint satisfaction problems of[29,34] living on diluted hypergraphs, where the solution-clusters have extensive entropy at any non-zero temperature. Another difference is that whereas in the locked constraint satisfaction problems the size of each cluster is one with high probability, in the binary perceptron there are still many solutions in the clusters, it is only their entropy density (i.e. logarithm of their number per variable) that vanishes as $N \to \infty$.

Investigation of the large local entropy clusters and their implications for learning in the symmetric perceptrons is also of great interest, but left for future work. Clearly since mathematically the symmetric perceptrons are simpler than the step-function one, they should also be the proper playground to deepen our understanding of the large local entropy clusters and their relation to learning and generalization.

We present the frozen-1RSB scenario as a conjecture and then below indicate how the second moment calculation gives evidence for this conjecture. Given an instance $\mathbf{X}$ and a solution $\mathbf{w}$, let $\Gamma(\mathbf{w}, d)$ denote the set of solutions $\mathbf{w}'$ with Hamming distance at most d from $\mathbf{w}$.

Conjecture 9. *For every $K > 0$ and every $\alpha \in (0, \alpha_c^r(K))$ there exists $d_{min} > 0$ so that with high probability over the choice of the random instance $\mathbf{X}$ from the RBP, the following property holds: for almost every solution $\mathbf{w}$,*

$$\frac{1}{N} \log |\Gamma(\mathbf{w}, d_{min})| \to 0$$

as $N \to \infty$. The same holds for the UBP for all $K \leq K^*$.

A. The link between the second-moment entropy and size of clusters

In this section we use $t \in \{r, u\}$ and note that the form of the second moment entropy density $\frac{1}{N} \log \mathbb{E}[\mathcal{Z}_t^2]$ has very direct implications on the structure of solutions in the corresponding models. As we defined it above, the second moment entropy is the normalized logarithm of the expected number of pairs of solutions of overlap β.

For problems such as the symmetric binary perceptrons where the quenched and annealed entropies are equal in leading order, there is a striking relation between the planted and the random ensemble of the model[35,36]. The random ensemble is the problem we have considered so far, while the planted ensemble is defined by starting with a configuration of the weights (a solution) and then including only constraints that are satisfied by this *planted* configuration. As long as the quenched and annealed entropies of the random ensemble are equal in leading order

the planted and random ensembles should be contiguous, meaning that high-probability properties that hold in one ensemble also hold in the other. Moreover the planted configuration in the planted ensemble has all the properties of a configuration sampled uniformly at random in the random ensemble. These properties follow on the heuristic level from the cavity method reasoning[36]. They were established fully rigorously in a range of models, see e.g.[35,37,38]. In the present case of symmetric binary perceptrons we have not yet managed to prove contiguity between the random and the planted ensemble, and so we leave a rigorous mathematical result for future work. (In fact the missing ingredient is a version of Friedgut's sharp threshold result[39] suitable for perceptrons; such a result combined with Theorem 4 would also prove Conjecture 2). We hence rely on the above heuristic argument and assume it holds in what follows.

Given a planted solution $\mathbf{w}$ and a configuration $\mathbf{w}_\beta$ that agrees with $\mathbf{w}$ on βN coordinates, the probability that $\mathbf{w}_\beta$ is a solution in the planted model is $(q_{t,K}(\beta)/p_{t,K})^M$, and thus the expected number of solutions at Hamming distance βN from the planted solution in the planted ensemble is

$$\mathbb{E}[\mathcal{Z}_\beta] = \binom{N}{\beta N} (q_{t,K}(\beta)/p_{t,K})^M \,,$$

and its entropy density is

$$\omega_t(\beta) \equiv \lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}[\mathcal{Z}_\beta] = F_{t,K,\alpha}(\beta) - \alpha \log p_{t,K} \text{ for } t \in \{r, u\} \,. \tag{11}$$

Recalling that contiguity implies that the planted solution has the properties of a uniformly chosen solution in the random ensemble then this entropy gives us direct access to properties of the solution space in the random ensemble at equilibrium. Most notably we notice (see derivation in section III B below) that the derivative of $\omega_t(\beta)$ at $\beta = 1$ is $+\infty$ thus implying that $\forall \epsilon > 0$ with high probability there are no solutions at overlap $\beta \in [d_{\min}(\alpha, K), (1-\epsilon)]$. In turn, this means that the dominant (measure converging to one as $N \to \infty$) part of the solution space splits into clusters each of which has vanishing entropy density (i.e. logarithm of the number of solutions in the cluster divided by N goes to zero as $N \to \infty$). The missing ingredient in a full proof of Conjecture 9 is a proof of the contiguity statement.

B. Form of the 2nd moment entropy implying frozen-1RSB

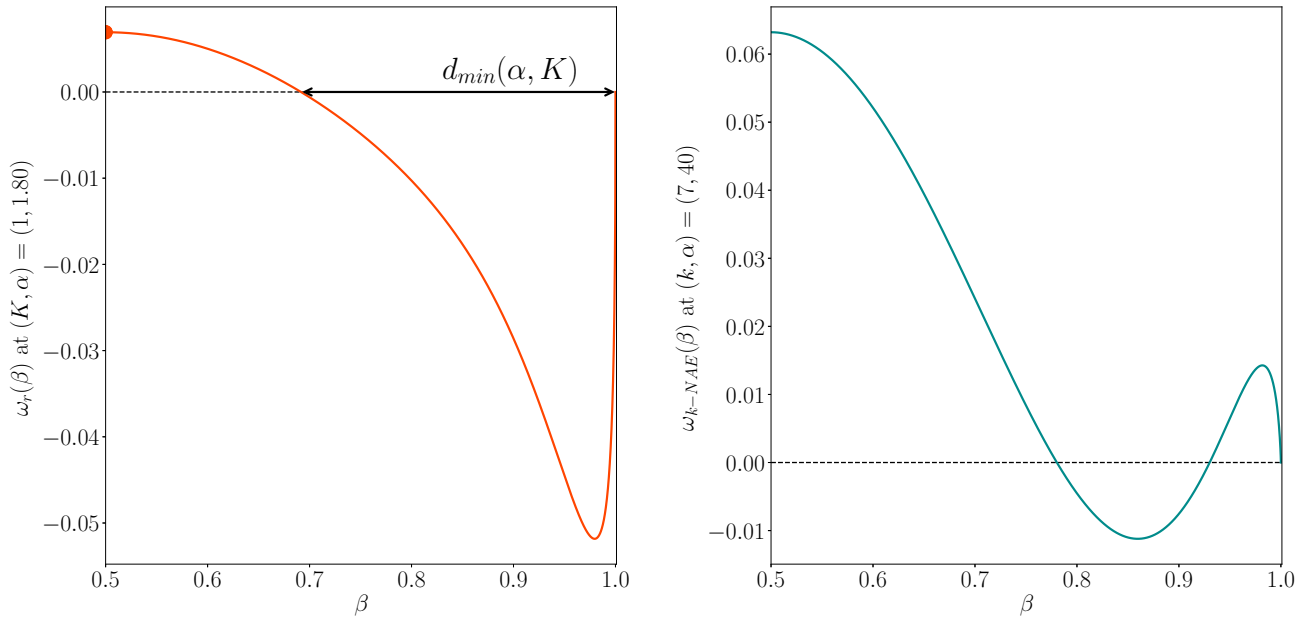

FIG. 2. **a)** Density of the annealed entropy of solutions at overlap β from a random solution in the rectangle binary perceptron at $K = 1$, $\alpha = 1.80 \leq \alpha_c^r(K = 1)$. We see there are no solution in an interval of overlaps $(1 - d_{\min}, 1 - \epsilon)$. This curve is obtained from the second moment entropy and contiguity between the random and planted ensembles. It implies the frozen-1RSB nature of the space of solutions. The same holds for the $u-$function. **b)** To compare we plot the density of the annealed entropy of solutions at overlap β from a random solution in the k-NAE SAT model[26] at $k = 7$, $\alpha = 40$. We see the density is positive in a large region close to $\beta = 1$, showing the absence of frozen-1RSB structure in this problem.

In fig. 2a we plot $\omega_r(\beta)$ for the rectangle binary perceptron, at $K = 1$, $\alpha = 1.80 \leq \alpha_c^r(K = 1)$. Thanks to the contiguity between the planted and random ensembles that holds as long as the second moment entropy density is twice the first moment entropy density, this curve represents also the annealed entropy of solutions at overlap β with a random reference solution. We see notably that there is an interval of distances in which no solutions are present. Analytically we can see from the properties of the functions $F_{t,K,\alpha}(\beta)$ and $\log p_{t,K}$ that $F_{t,K,\alpha}(1) = \alpha \log p_{t,K}$ and the derivative of $F_{t,K,\alpha}(\beta) \to \infty$. This is in contrast with, for instance, the satisfiability problems studied in[26], where the function corresponding to $F_{t,K,\alpha}(\beta)$ would have a negative derivative in $\beta = 1$ (see fig. 2b). There could still be an interval of *forbidden* distance, but the bump in entropy for $\beta \approx 1$ corresponds to the size of the clusters to which typical solutions belong and those would be extensive.

1. Frozen 1RSB in rectangle binary perceptron

In the rectangle binary perceptron, the random and planted ensembles are conjectured to be contiguous for all $K > 0$ and $\alpha \in (0, \alpha_c^r(K))$. Using eq. (8), the first derivative of $\omega_r(\beta)$, eq. (11), is given by (see Appendix VI E)

$$\frac{\partial \omega_r}{\partial \beta} = \frac{\partial F_{r,K,\alpha}}{\partial \beta} = \log\left(\frac{1-\beta}{\beta}\right) + \frac{\alpha}{q_{r,K,T}(\beta)} \frac{1}{\pi\sqrt{\beta(1-\beta)}} \left(e^{-\frac{K^2}{2(1-\beta)}} \left(e^{\frac{(2\beta-1)K^2}{2(1-\beta)\beta}} - 1\right)\right) ,$$

and it diverges for all $K \in \mathbb{R}^+$, $\alpha > 0$ in the limit $\beta \to 1$:

$$\frac{\partial \omega_r}{\partial \beta} \xrightarrow[\beta \to 1]{} +\infty \, . \tag{12}$$

This implies vanishing entropy density of clusters to which typical solutions belong.

2. Frozen 1RSB in the u-function binary perceptron

In the u-function binary perceptron, the random and planted ensembles are conjectured to be contiguous for all $0 < K \leq K^*$ and $\alpha \in (0, \alpha_c^u(K))$. Using eq. (8), the first derivative of $\omega_u(\beta)$ eq. (11), is given by

$$\frac{\partial \omega_u}{\partial \beta} = \frac{\partial F_{u,K,\alpha}}{\partial \beta} = \log\left(\frac{1-\beta}{\beta}\right) + \frac{\alpha}{q_{u,K,T}(\beta)} \frac{1}{\pi\sqrt{\beta(1-\beta)}} \left(e^{-\frac{K^2}{2(1-\beta)}} \left(e^{\frac{(2\beta-1)K^2}{2(1-\beta)\beta}} - 1\right)\right)$$
$$\xrightarrow[\beta \to 1]{} +\infty \, ,$$

thus reaching the same conclusion on presence of frozen-1RSB.

In appendix VI E we extend the second moment calculation to finite temperature (for both the rectangle and $u-$function case). This means that we define energy of a configuration $\mathcal{E}(\mathbf{w})$ as the number of constraints that are violated by this configurations. Then the corresponding partition function is defined $\mathcal{Z}(T) = \sum_{\mathbf{w}} e^{-\mathcal{E}(\mathbf{w})/T}$. There is a one-to-one mapping between the temperature T and energy density $e = \mathcal{E}/N$, consequently the corresponding finite-temperature second moment entropy density counts the number of pairs of solutions at overlap β and energy density e. In appendix VI E we apply the same argument as here connecting the random and planted ensemble, and deduce that the finite-temperature solution space of the models is of also of the frozen-1RSB type for any $T < \infty$.

C. Frozen-1RSB as derived from the replica analysis

We stress that we derived the frozen-1RSB nature of the space of solutions without the use of replicas. For completeness we summarize here how this translates to the properties of the one-step-replica-symmetry breaking solution. This is the way this phenomena was originally discovered and described in[2,27,40]. For readers not familiar with the replica method this section should be read after reading section IV.

In general, three kinds of fixed points of the 1RSB equations are possible:

- The replica symmetric (RS) solution $q_0 = q_1 = q_{\rm RS} < 1$,
- The frozen-1RSB solution (f1RSB) $(q_0, q_1) = (q_{\rm RS}, 1)$,
- The 1RSB solution (q_0, q_1) with $q_1 \neq 1$.

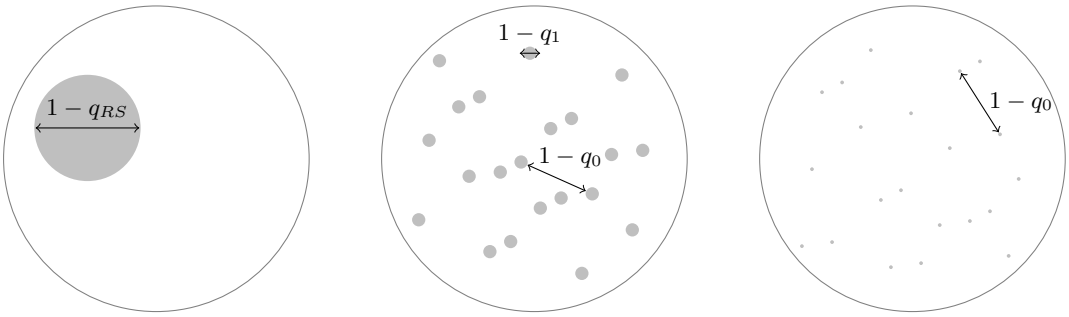

FIG. 3. Illustration of the configuration space for the different phases: **a)**: RS - solutions are concentrated in a single cluster of typical size $1-q_{\rm RS}$. **b)**: 1RSB - solutions form clusters of size $1-q_1$ at a distance $1-q_0$ from each other. **c)**: f1RSB - clusters are point-like ($1-q_1 \simeq 0$) at a distance $1-q_0 = 1-q_{RS}$ from each other.

The frozen-1RSB is characterized by an inner-cluster overlap $q_1 = 1$ and an inter-cluster overlap $q_0 = q_{\rm RS}$, which means that clusters have vanishing entropy density and remain far from each other. Mathematically RS and f1RSB solutions are equivalent in the sense that these solutions have the same free energy eq. (20) $\Phi_{\rm 1RSB}\{q_{\rm RS}, q_{\rm RS}\} = \Phi_{\rm 1RSB}\{q_{\rm RS}, 1\}$, and the complexity of the f1RSB solution equals the RS entropy $\Sigma(\phi = 0) = \phi_{\rm RS}$ eq. (22, 15). However, RS and f1RSB do not share the same configuration space. The RS phase is associated to a single cluster of solution with typical size $1-q_{\rm RS}$, while the f1RSB configuration space is composed of many point-like solutions of size $q_1 \simeq 1$ and at distance $1-q_0 = 1-q_{\rm RS}$ of each other, see fig. 3. From this point of view f1RSB is the correct description of the phase space.

IV. REPLICA CALCULATION OF THE STORAGE CAPACITY

In this section we recall the replica calculation leading to the expression of the storage capacity in the step-function binary perceptron. We show that in the symmetric binary perceptrons the annealed calculation is reproduced by the replica symmetric result. For the $u-$function binary perceptron we show that K^* coincides with the onset of replica symmetry breaking and we evaluate the 1RSB capacity for $K > K^*$.

A. Replica calculation

For the purpose of the calculations, we introduce the constraint function $\mathcal{C}(\mathbf{z})$ that returns 1 if $\mathbf{w}$ satisfies all the constraints $\{\varphi(z_\mu)\}_{\mu=1}^M$ and 0 otherwise

$$\mathcal{C}(\mathbf{z}) = \prod_{\mu=1}^{M} \varphi(z_\mu) \text{ with } z_\mu = \mathbf{X}_\mu \mathbf{w}\,.$$

Recall the partition function $\mathcal{Z}$ is the number of satisfying vectors $\mathbf{w}$, with prior distribution $P_w(\mathbf{w})$, for a given matrix $\mathbf{X}$

$$\mathcal{Z}(\mathbf{X}) = \sum_{\mathbf{w}\in\{\pm 1\}^N} \prod_{\mu=1}^{M} \varphi\left(\mathbf{X}_\mu \mathbf{w}\right) = \int d\mathbf{w} P_w(\mathbf{w}) \int d\mathbf{z}\, \mathcal{C}(\mathbf{z}) \delta(\mathbf{z} - \mathbf{X}\mathbf{w})\,.$$

The replica method allows one to compute explicitly the quenched average $\mathbb{E}_{\mathbf{X}}[\log(\mathcal{Z}(\mathbf{X}))]$[41]. More precisely, using the replica trick, the average of the logarithm can be expressed as the limit $n \to 0$ of the derivative with respect to n of the average of the n-th moment of the partition function. Finally the free entropy reads:

$$\phi(\alpha) \equiv \lim_{N\to+\infty} \frac{1}{N} \mathbb{E}_{\mathbf{X}}[\log(\mathcal{Z}(\mathbf{X}))] = \lim_{N\to+\infty} \lim_{n\to 0} \frac{1}{Nn} \frac{\partial \log\left(\mathbb{E}_{\mathbf{X}}[\mathcal{Z}(\mathbf{X})^n]\right)}{\partial n}\,. \tag{13}$$

Computing the n-th moment of the partition function $\mathcal{Z}$, for $n \in \mathbb{N}$, is equivalent to considering n copies, also called replicas, of the initial system. For a given disorder, these n replicas are non-interacting and $\mathcal{Z}^n$ can be computed easily. However, averaging over the "disorder" with distribution P_X makes the replicas interacting: replicated weight-vectors $\mathbf{w}^a$ and $\mathbf{w}^b$, for $a, b \in [1:n]$, are correlated by the overlap matrix $\mathbf{Q} = (Q_{ab})_{a,b=1}^n = \left(\frac{1}{N}\sum_{i=1}^N w_i^a w_i^b\right)_{a,b=1}^n$.

We show in Appendix VI A that after averaging over the distribution P_X, using an analytical continuation for $n \in \mathbb{R}$ and finally reversing the limits $N \to \infty$ and $n \to 0$, the free entropy ϕ eq. (13) can finally be expressed as a saddle point equation over $n \times n$ symmetric matrices $\mathbf{Q}$ and $\hat{\mathbf{Q}}$

$$\phi(\alpha) = -\mathrm{SP}_{\mathbf{Q},\hat{\mathbf{Q}}} \left\{ \lim_{n\to 0} \frac{\partial S_n(\mathbf{Q},\hat{\mathbf{Q}})}{\partial n} \right\} , \tag{14}$$

where $\hat{\mathbf{Q}}$ is a parameter involved in the change of variable between $\{\mathbf{w}^a, \mathbf{w}^b\}$ and Q_{ab} and with

$$\begin{cases} S_n(\mathbf{Q},\hat{\mathbf{Q}}) = \frac{1}{2}\mathrm{Tr}(\mathbf{Q}\hat{\mathbf{Q}}) - \log(\mathcal{I}_w^n(\hat{\mathbf{Q}})) - \alpha \log\left(\mathcal{I}_z^n(\mathbf{Q})\right) , \\ \mathcal{I}_w^n(\hat{\mathbf{Q}}) = \int_{\mathbb{R}^n} d\tilde{\mathbf{w}} P_{\tilde{w}}(\tilde{\mathbf{w}}) e^{\frac{1}{2}\tilde{\mathbf{w}}^\intercal \hat{\mathbf{Q}} \tilde{\mathbf{w}}} \quad \text{where } \tilde{\mathbf{w}} \in \mathbb{R}^n \text{ and } P_{\tilde{w}}(\tilde{\mathbf{w}}) = \prod_{a=1}^{n} [\delta(\tilde{w}_a - 1) + \delta(\tilde{w}_a + 1)] , \\ \mathcal{I}_z^n(\mathbf{Q}) = \int_{\mathbb{R}^n} d\tilde{\mathbf{z}} P_{\tilde{z}}(\tilde{\mathbf{z}}) \mathcal{C}(\tilde{\mathbf{z}}) \quad \text{where } \tilde{\mathbf{z}} \in \mathbb{R}^n \text{ and } P_{\tilde{z}} \triangleq \mathcal{N}(\mathbf{0}, \mathbf{Q}) . \end{cases}$$

In order to be able to compute the derivative of S_n with respect to n eq. (14), we need an analytical formulation of $\mathbf{Q}$ and $\hat{\mathbf{Q}}$ as a function of n.

B. RS entropy

The simplest ansatz is to assume that the overlap matrix $\mathbf{Q}$ is Replica Symmetric (RS), which means that all replicas play the same role: the correlation between two arbitrary, but different, replicas is denoted q_0, and therefore the RS ansatz reads:

$$\forall (a,b) \in [1:n] \times [1:n], \ \frac{1}{N}(\mathbf{w}^a \cdot \mathbf{w}^b) = \begin{cases} q_0 \text{ if } a \neq b , \\ Q = 1 \text{ if } a = b . \end{cases}$$

It enforces the matrix $\hat{\mathbf{Q}}$ to present the same symmetry, respectively with parameters $\hat{q}_0$ and $\hat{Q} = 1$. Using this ansatz and the $n \to 0$ limit, the Replica Symmetric (RS) entropy can be expressed as a set of saddle point equations over scalar parameters q_0 and $\hat{q}_0$, evaluated at the saddle point (Appendix VI B):

$$\phi_{\mathrm{RS}}(\alpha) = \mathrm{SP}_{q_0,\hat{q}_0} \left\{ -\frac{1}{2} + \frac{1}{2}(q_0\hat{q}_0 - 1) + \mathcal{I}_{\mathrm{RS}}^w(\hat{q}_0) + \alpha \mathcal{I}_{\mathrm{RS}}^z(q_0) \right\} , \tag{15}$$

$$\text{with } \begin{cases} \mathcal{I}_{\mathrm{RS}}^w(\hat{q}_0) \equiv \int Dt \log\left(g_0^w(t,\hat{q}_0)\right) , \\ \mathcal{I}_{\mathrm{RS}}^z(q_0) \equiv \int Dt \log\left(f_0^z(t,q_0)\right) , \end{cases} \text{ and for } i \in \mathbb{N} \begin{cases} g_i^w(t,\hat{q}_0) \equiv \int dw\, w^i P_w(w) \exp\left(\frac{(1-\hat{q}_0)}{2} w^2 + t\sqrt{\hat{q}_0} w\right) , \\ f_i^z(t,q_0) \equiv \int Dz\, z^i \varphi(\sqrt{q_0}t + \sqrt{1-q_0}z) . \end{cases} \tag{16}$$

Note that above and in what follows $Dt = \int \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$. In the binary perceptron case, the function P_w is defined as $P_w(w) = [\delta(w-1) + \delta(w+1)]$ (note that this is not a probability distribution because of the normalization), and recall $\varphi(z)$ is the indicator function, checking that a constraint on the argument is satisfied (e.g in the step case, $\varphi^s(z) = 1$ if $z > K$).

While in the step binary perceptron (SBP) the fixed point solution $(q_0, \hat{q}_0)$ is non-trivial, the symmetry of the activation function in the RBP and UBP cases enforces the configuration space to be symmetric and the fixed point $(q_0, \hat{q}_0) = (0,0)$ to exist. If this symmetric fixed point is stable and has the lowest free energy, the RS free entropy matches the annealed entropy $\phi_a^t(\alpha) = \log(2) + \alpha \log(p_{t,K}) = \frac{1}{N} \log \mathbb{E}_{\mathbf{X}}[\mathcal{Z}_t(\mathbf{X})]$ from section II A with $t \in \{r, u\}$.

1. Rectangle

Solving numerically the corresponding saddle point equations leads to the single symmetric fixed point $(q_0, \hat{q}_0) = (0,0)$. Hence the RS entropy saturates the first moment bound:

$$\phi_{\mathrm{RS}}^r(\alpha) = \log(2) + \alpha \log\left(p_{r,K}\right) = \phi_a^r(\alpha) ,$$

and the RS capacity equals the annealed capacity eq. (II A):

$$\alpha_{\mathrm{RS}}^r(K) = \alpha_a^r(K) = \frac{-\log(2)}{\log\left(p_{r,K}\right)} .$$

2. U-function

- For $K \leq K^*$, only the symmetric fixed point $(q_0, \hat{q}_0) = (0, 0)$ exists, which leads again to the annealed free entropy:

$$\phi^u_{\rm RS}(\alpha) = \log(2) + \alpha \log\left(p_{u,K}\right) = \phi^u_a(\alpha)\,,$$

  and annealed capacity eq. (II A):

$$\alpha^u_{\rm RS}(K) = \alpha^u_a(K) = \frac{-\log(2)}{\log\left(p_{u,K}\right)}\,.$$

- For $K > K^*$, the RS entropy does not match the annealed entropy because the fixed point $(q_0, \hat{q}_0) \neq (0, 0)$ corresponds to a lower free energy than the symmetric fixed point $(0, 0)$. The symmetric fixed point becomes unstable for $K > K^*$, where K^* is remarkably given by the same value as in the independent section II B 2. Hence it naturally verifies eq. (5) even though its definition derives from the stability of the RS solution, that we study in the next section.

C. Stability

The local stability of the RS solution can be studied using de Almeida and Thouless (AT) method[42], based on the positivity of the Hessian of $S_n(\mathbf{Q}, \hat{\mathbf{Q}})$. The replica symmetric AT-line $\alpha_{\rm AT}$ is given by the solution of the following implicit equation (Appendix VI D):

$$\frac{1}{\alpha} = \frac{1}{(1-q_0(\alpha))^2}\int Dt \frac{\left(f_0^z(f_0^z - f_2^z) + (f_1^z)^2\right)^2}{(f_0^z)^4}(t, q_0(\alpha)) \int Dt \frac{\left(g_0^w g_2^w - (g_1^w)^2\right)^2}{(g_0^w)^4}(t, \hat{q}_0(\alpha))\,.$$

As illustrated above, for the rectangle and $u-$function, the symmetry of the weights P_w and the constraint φ imposes the existence of the symmetric fixed point $(q_0, \hat{q}_0) = (0, 0)$. This simplifies the previous condition and becomes equivalent to the linear stability condition of the symmetric fixed point $(q_0, \hat{q}_0) = (0, 0)$ (see Appendix VI D):

$$\frac{1}{\alpha_{\rm AT}} = \left(\frac{\tilde{f}_2^z - \tilde{f}_0^z}{\tilde{f}_0^z}\right)^2 \left(\frac{\tilde{g}_2^w}{\tilde{g}_0^w}\right)^2\,, \text{ where for } i \in \mathbb{N}\text{: } \begin{cases} \tilde{g}_i^w = \int dw w^i P_w(w) e^{\frac{w^2}{2}}\,, \\ \tilde{f}_i^z = \int Dz z^i \varphi(z)\,. \end{cases}$$

We plotted the annealed capacity, the replica symmetric capacity and the AT-line for the step, rectangle and u-function binary perceptrons as functions of K in fig. 4, 5, 6.

1. Step binary perceptron

We note that for the step binary perceptron the RS solution is always stable towards 1RSB, even for negative threshold $K < 0$. This is interesting in the view of recent work on the spherical perceptron with negative threshold where the replica symmetry breaks for all $K < 0$, and full-step RSB is needed to evaluate the storage capacity[16].

2. Rectangle

As the RS capacity $\alpha^r_{\rm RS}$ is always below the AT line $\alpha^r_{\rm AT}$, the RS solution is always locally stable.

3. u-function

There is a crossing between the values of the RS capacity $\alpha^u_{\rm RS}$ and the AT-line $\alpha^u_{\rm AT}$, which defines implicitly the value $K^* \simeq 0.817$, and matches the equality in eq. (10):

$$\frac{-\log(2)}{\log\left(p_{u,K^*}\right)} = \frac{\pi}{2}\frac{\left(p_{u,K^*}\right)^2}{e^{-(K^*)^2}(K^*)^2}\,. \tag{17}$$

For $K \leq K^*$, the RS solution is locally stable, while for $K > K^*$ the RS solution becomes unstable, and a symmetry breaking solution appears.

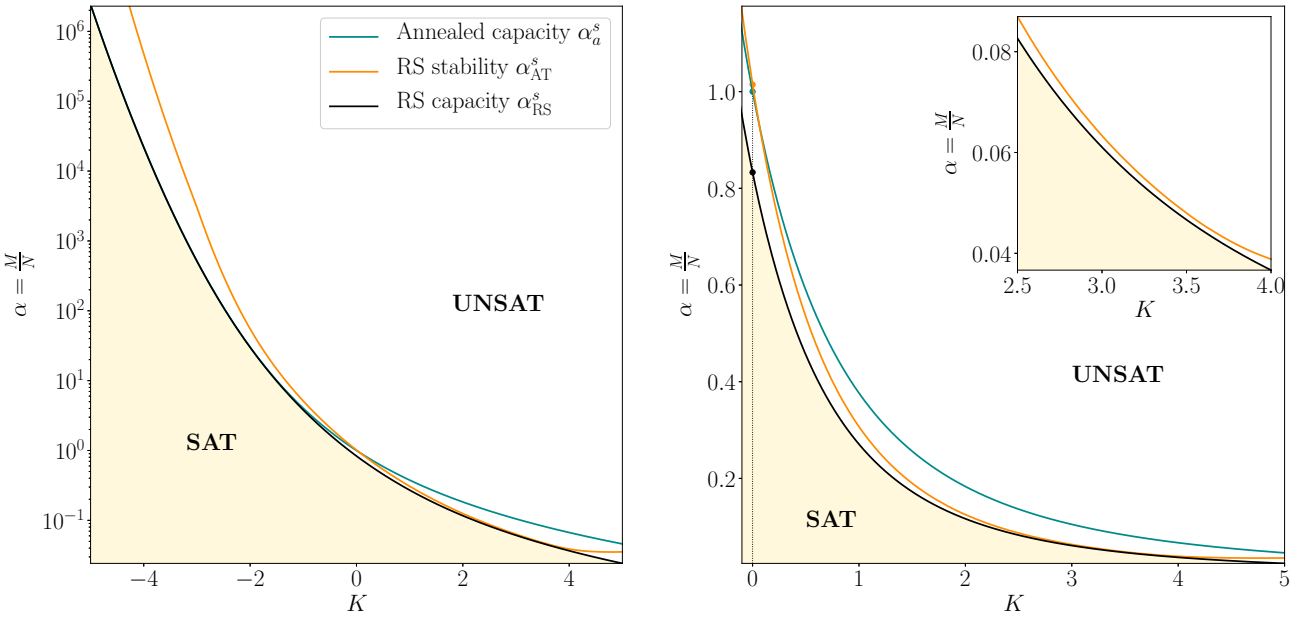

FIG. 4. Step binary perceptron (SBP): the RS capacity $\alpha_{\rm RS}^s$ (black) does not match the annealed capacity α_a^s (blue) and is always below the AT-line $\alpha_{\rm AT}^s$ (orange). The AT-line is closest to the annealed capacity for $K_{\rm min} \simeq 3.62$ where the difference $\alpha_{\rm AT}^s - \alpha_a^s \simeq 0.0012$. For $K = 0$, we retrieve well known results[2]: $\alpha_{\rm RS}^r \simeq 0.833$, $\alpha_{\rm AT}^r \simeq 1.015$ and $\alpha_a^r = 1$. The left and right hand sides, and the inset, represent the same data on different scales. The satisfiable (SAT) phase is represented by the beige shaded area and is located below the RS capacity, while the unsatisfiable (UNSAT) starts at the capacity (black line) and extends for a larger number of constraints.

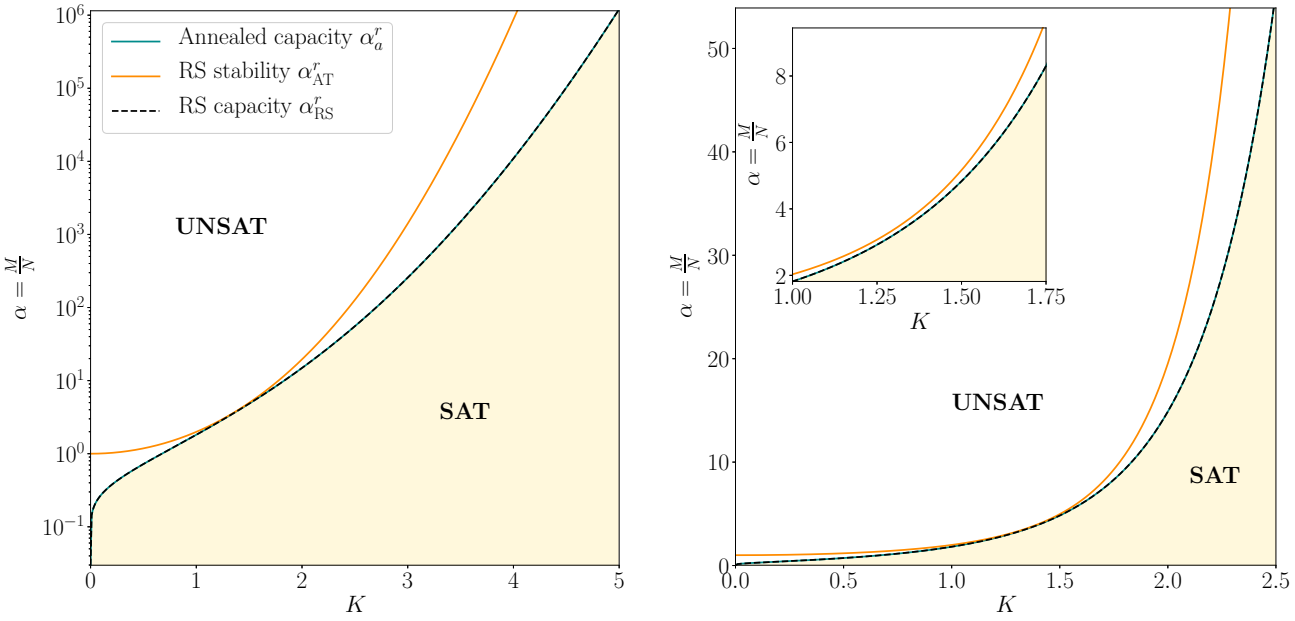

FIG. 5. Rectangle binary perceptron (RBP): the RS capacity $\alpha_{\rm RS}^r$ (black) matches the annealed bound α_a^r (blue), and the RS solution is locally stable for all K: $\alpha_{\rm RS}^r < \alpha_{\rm AT}^r$. The AT-line (orange) is closest to the annealed capacity for $K_{\rm min} \simeq 1.24$ where the difference $\alpha_{\rm AT}^s - \alpha_a^s \simeq 0.15$. The left and right hand sides, and the inset, represent the same data on different scales. The satisfiable (SAT) phase is represented by the beige shaded area and is located below the RS capacity, while the unsatisfiable (UNSAT) starts at the capacity (black line) and extends for a larger number of constraints.

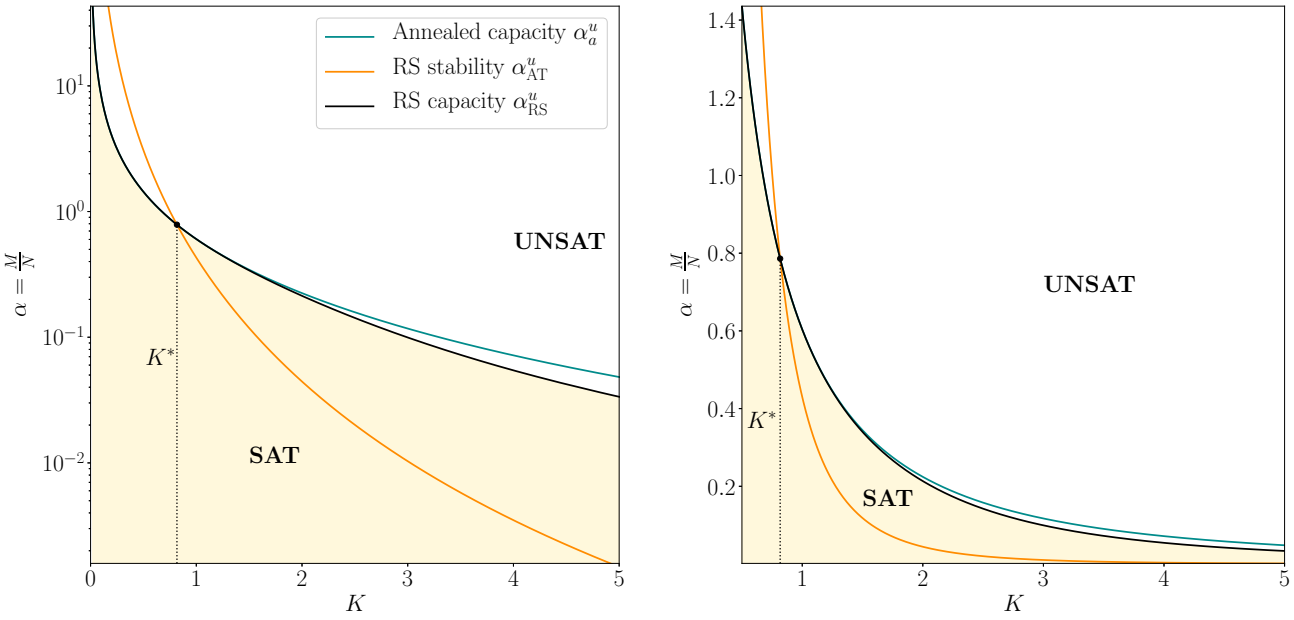

FIG. 6. $U-$function binary perceptron (UBP): the RS capacity black) matches the annealed bound (blue) for $K < K^*$. At $K = K^*$, the RS capacity crosses the AT-line (orange). For $K > K^*$, the RS solution is unstable and the RS capacity deviates from the annealed capacity. The left and right hand sides, and the inset, represent the same data on different scales. The satisfiable (SAT) phase is represented by the beige shaded area and is located below the RS capacity, while the unsatisfiable (UNSAT) starts at the capacity (black line) and extends for a larger number of constraints.

D. 1RSB calculation

In the previous section we concluded that the replica symmetric solution is unstable in the $u-$function binary perceptron for $K > K^*$, we analyze therefore the first-step of replica symmetry breaking (1RSB) ansatz in this section. This ansatz and calculations is due to seminal works of G. Parisi and is classic in the field of disordered systems and well presented in the literature[17–19,43], we thus mainly give the key formulas and defer the details into the Appendix VI C.

The 1RSB ansatz assumes that the space of configurations splits into states. Consequently replicas are not symmetric anymore and instead n replicas are organized in $\frac{n}{m}$ groups containing m replicas each:

$$\forall (a,b) \in [1:n]\times[1:n],\ \frac{1}{N}(\mathbf{w}^a\cdot\mathbf{w}^b) = \begin{cases} q_1 \text{ if } a,b \text{ belong to the same state,} \\ q_0 \text{ if } a,b \text{ do not belong to the same state,} \\ Q = 1 \text{ if } a = b\,. \end{cases} \tag{18}$$

Following[44], the partition function $\mathcal{Z}_m$ associated to m replicas falling in the same state is expressed as a sum over all possible states Ψ weighted by their corresponding free entropy ϕ:

$$\mathcal{Z}_m = \sum_{\{\Psi\}} \exp(Nm\phi(\Psi)) = \sum_{\{\phi\}} \mathcal{N}_\phi \exp(Nm\phi) = \sum_{\{\phi\}} \exp(N\Sigma(\phi))\exp(Nm\phi) \sim \int d\phi \exp(N(m\phi + \Sigma(\phi))\,,$$

where we introduced the number of states at a given free entropy ϕ: $\mathcal{N}_\phi \equiv \exp(N\Sigma(\phi))$ and the complexity $\Sigma(\phi)$, also called the configurational entropy.

Using the saddle point method in the $N\to\infty$ limit, the 1RSB replicated free entropy $\Phi_{\rm 1RSB}$ is written as a function of the Parisi parameter m, the free entropy ϕ and the complexity $\Sigma(\phi)$:

$$\Phi_{\rm 1RSB}(m,\alpha) \equiv \lim_{N\to\infty}\frac{1}{N}\mathbb{E}_{\mathbf{X}}\left[\log(\mathcal{Z}_m(\mathbf{X}))\right] = m\phi + \Sigma(\phi)\,. \tag{19}$$

Injecting the 1RSB ansatz eq. (18) in the replica derivation eq. (14), the 1RSB replicated free entropy $\Phi_{\rm 1RSB}$ is written as a saddle point equation over $\mathbf{q} = (q_0, q_1)$ and $\hat{\mathbf{q}} = (\hat{q}_0, \hat{q}_1)$ (see Appendix VI C):

$$\Phi_{\rm 1RSB}(m,\alpha) = \underset{\mathbf{q},\hat{\mathbf{q}}}{\rm SP}\left\{\frac{m}{2}(q_1\hat{q}_1 - 1) + \frac{m^2}{2}(q_0\hat{q}_0 - q_1\hat{q}_1) + m\mathcal{I}^w_{\rm 1RSB}(\hat{\mathbf{q}}) + \alpha m\mathcal{I}^z_{\rm 1RSB}(\mathbf{q})\right\} \tag{20}$$

$$\text{with } \begin{cases} \mathcal{I}^w_{\text{1RSB}}(\hat{\mathbf{q}}) = \frac{1}{m}\int Dt_0 \log\left(\int Dt_1 g_0^w(\mathbf{t},\hat{\mathbf{q}})^m\right), \\ \mathcal{I}^z_{\text{1RSB}}(\mathbf{q}) = \frac{1}{m}\int Dt_0 \log\left(\int Dt_1 f_0^z(\mathbf{t},\mathbf{q})^m\right), \end{cases}$$

$$\text{denoting } \mathbf{t}=(t_0,t_1), \text{ and for } i\in\mathbb{N}: \begin{cases} g_i^w(\mathbf{t},\hat{\mathbf{q}}) = \int dw\, w^i P_w(w) \exp\left(\frac{(1-\hat{q}_1)}{2}w^2 + \left(\sqrt{\hat{q}_0}t_0 + \sqrt{\hat{q}_1-\hat{q}_0}t_1\right)w\right), \\ f_i^z(\mathbf{t},\mathbf{q}) = \int Dz\, z^i \varphi(\sqrt{q_0}t_0 + \sqrt{q_1-q_0}t_1 + \sqrt{1-q_1}z). \end{cases} \tag{21}$$

Taking the derivative of Φ_{1RSB} with respect to m, the free entropy ϕ and complexity Σ can be written as:

$$\begin{cases} \phi(\alpha) = \frac{\partial\Phi_{\text{1RSB}}(m,\alpha)}{\partial m} = \underset{\mathbf{q},\hat{\mathbf{q}}}{\text{SP}}\left\{\frac{1}{2}(q_1\hat{q}_1 - 1) + m\left(q_0\hat{q}_0 - q_1\hat{q}_1\right) + \mathcal{J}^w_{\text{1RSB}}(\hat{\mathbf{q}}) + \alpha\mathcal{J}^z_{\text{1RSB}}(\mathbf{q})\right\}, \\ \Sigma(\phi) = \Phi_{\text{1RSB}} - m\phi = \underset{\mathbf{q},\hat{\mathbf{q}}}{\text{SP}}\left\{\frac{m^2}{2}(q_1\hat{q}_1 - q_0\hat{q}_0) + m(\mathcal{I}^w_{\text{1RSB}} - \mathcal{J}^w_{\text{1RSB}})(\hat{\mathbf{q}}) + m\alpha(\mathcal{I}^z_{\text{1RSB}} - \mathcal{J}^z_{\text{1RSB}})(\mathbf{q})\right\}, \end{cases} \tag{22}$$

$$\text{with } \begin{cases} \mathcal{J}^w_{\text{1RSB}}(\hat{\mathbf{q}}) = \frac{\partial(m\mathcal{I}^w_{\text{1RSB}})}{\partial m} = \int Dt_0 \frac{\int Dt_1 \log(g_0^w(\mathbf{t},\hat{\mathbf{q}}))g_0^w(\mathbf{t},\hat{\mathbf{q}})^m}{\int Dt_1 g_0^w(\mathbf{t},\hat{\mathbf{q}})^m}, \\ \mathcal{J}^z_{\text{1RSB}}(\mathbf{q}) = \frac{\partial(m\mathcal{I}^z_{\text{1RSB}})}{\partial m} = \int Dt_0 \frac{\int Dt_1 \log(f_0^z(\mathbf{t},\mathbf{q}))f_0^z(\mathbf{t},\mathbf{q})^m}{\int Dt_1 f_0^z(\mathbf{t},\mathbf{q})^m}. \end{cases}$$

E. 1RSB results for UBP

From now on we only consider the $u-$function binary perceptron, whose RS solution is unstable for $K > K^*$.

The Parisi parameter m is fixed to its equilibrium value by maximizing the total entropy in the SAT phase, $\phi_{\text{tot}} = \phi + \Sigma(\phi)$, under the constraint that the free entropy and complexity are both positive $\phi \geq 0$ and $\Sigma(\phi) \geq 0$:

$$m_{\text{eq}} = \max_{m|\phi\geq 0,\Sigma\geq 0} \phi + \Sigma(\phi).$$

Using the expressions eq. (22) and varying the Parisi parameter $m \in [0;1]$, we obtain the curve of the complexity $\Sigma(\phi)$ as shown in fig. 7. At $m = 1$, the complexity is negative. Decreasing m, the complexity increases and becomes positive at the value m_{eq}. Besides for small values of m, an unphysical (convex) branch appears, as commonly observed in other systems solved by the replica method.

We note that at α increases both the equilibrium complexity and free entropy decrease. In constraint satisfaction problems such as K-satisfiability or random graph coloring the mechanism in which the satisfiability threshold appears is that the maximum of the complexity becomes negative. In the present UBP problem it is actually both the free entropy and the complexity that vanish together, as illustrated in fig. 7.

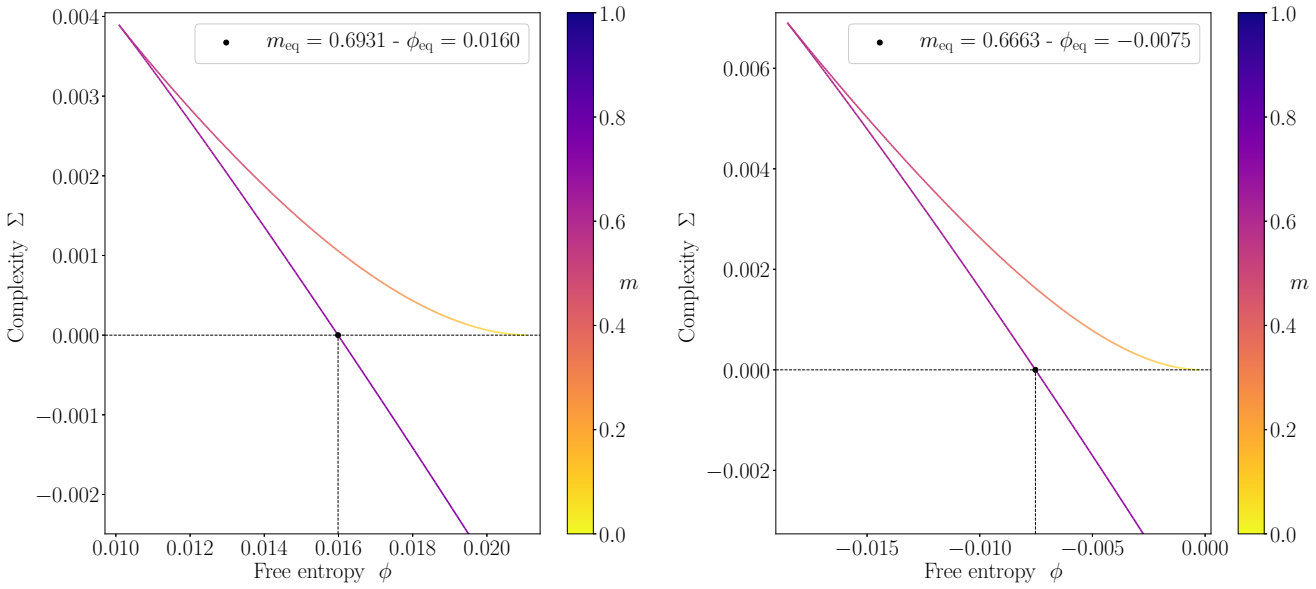

FIG. 7. Complexity $\Sigma(\phi)$ as a function of the free entropy ϕ for the $u-$function binary perceptron at $K = 1.5 > K^*$. Complexity reaches $\Sigma = 0$ (black dot) at $m_{\rm eq}$. For $K = 1.5$ and $\alpha = 0.33$ **a)** the free-entropy corresponding to $m_{\rm eq}$ is positive $\phi_{\rm eq} > 0$, whereas for $\alpha = 0.34$ **b)** the free entropy at $m_{\rm eq}$ is negative $\phi_{\rm eq} < 0$ and therefore there is no part of the curve where both complexity and free entropy are positive: thus this value of α is beyond the 1RSB storage capacity, and the capacity is in the interval $[0.33; 0.34]$.

Computing the equilibrium value $m_{\rm eq}(\alpha)$, we have access to the corresponding equilibrium overlaps q_0^* and q_1^*, that we may compare with the RS solution $q_{\rm RS}$. All these are depicted in fig. 8. We also compute the 1RSB entropy $\phi^u_{\rm 1RSB} \leq \phi^u_{\rm RS}$ which vanishes at the 1RSB capacity $\alpha^u_{\rm 1RSB}$ as depicted in fig. 9a. The 1RSB solution provides a small correction to the RS result for storage capacity, as illustrated in fig. 9b, where we plotted the difference between the annealed upper bound and the capacity for the RS and 1RSB solutions: $\alpha^u_a - \alpha^u_{\rm RS}$ and $\alpha^u_a - \alpha^u_{\rm 1RSB}$.

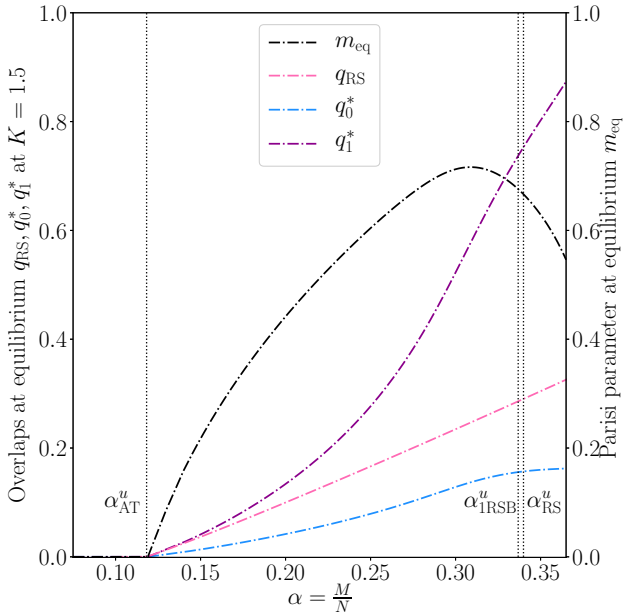

FIG. 8. Equilibrium values of the overlap $q_0^* \neq q_{\rm RS}$, q_1^* and the Parisi parameter $m_{\rm eq}$ for the UBP at $K = 1.5$. For $K < K^*$, the RS solution is stable and the only fixed point is $q_0 = q_1 = q_{RS} = 0$.

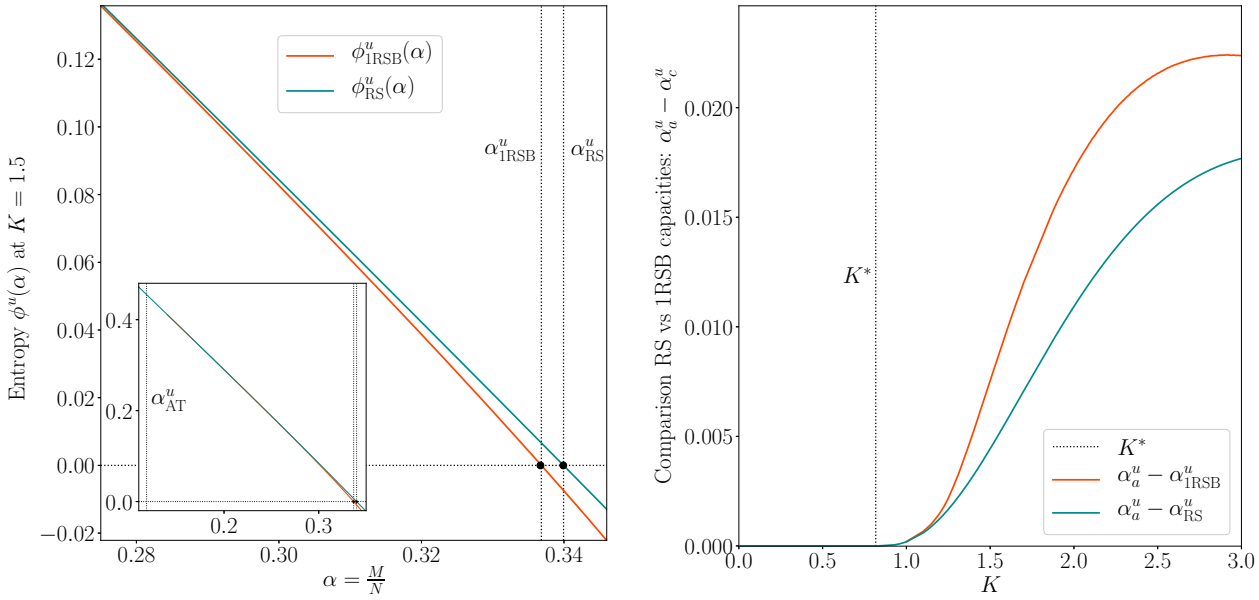

FIG. 9. **a)**: Comparison of the RS (blue) and 1RSB (orange) entropy for the UBP at $K = 1.5$. For $\alpha < \alpha_{\rm AT} \simeq 0.118$, RS and 1RSB entropies are equalled. For $\alpha > \alpha_{\rm AT}$, 1RSB entropy deviates slightly of the RS entropy before vanishing respectively at $\alpha^u_{\rm 1RSB} \simeq 0.337$ and $\alpha^u_{\rm RS} \simeq 0.334$. The inset represents the same data on a different scale. **b)**: Difference between the annealed upper bound and the 1RSB capacity $\alpha^u_a - \alpha^u_{\rm 1RSB}$ (orange) and the RS capacity $\alpha^u_a - \alpha^u_{\rm RS}$ (blue). Below K^* the RS solution is stable: RS and 1RSB entropies match exactly. Above K^*, the RS solution is unstable: the 1RSB entropy deviates slightly from the RS solution.

F. 1RSB Stability

In the previous section we evaluated the 1RSB storage capacity of the $u-$function binary perceptron for $K > K^*$. In this section we will argue that this cannot be an exact solution to the problem.

We could investigate the stability of 1RSB towards further levels of replica symmetry breaking along the same lines we did for the RS solution. However, in the present case we do not need to do that to see that the obtained solution cannot be correct. The explanations lies in the breaking of the up-down symmetry in the problem. This symmetry must either be broken explicitly as in the ferromagnet, where the system would acquire an overall magnetization, but we have not observed any trace of this in the present problem. Or this up-down symmetry must be conserved in the final correct solution. The conservation of the up-down symmetry is manifested in the value $q_0 = 0$ in the replica symmetric phase. The fact that in the 1RSB solution evaluated above we do not observe $q_0 = 0$, but instead $q_0 > 0$ is a sign of the fact that we are evaluating a wrong solution. The only possible way to obtain an exact solution we foresee is to evaluate the full-step replica symmetry breaking with a continuity of overlaps $q(x)$, the smallest one of them should be 0 in order to restore the up-down symmetry. We let the evaluation of the full-RSB for future work.

Finally let us note that the 1RSB solution obtained in the previous section can be interpreted as frozen-2RSB. In 2RSB we would have 3 kinds of overlaps, q_0, q_1 and q_2. In frozen 2RSB we would have $q_2 = 1$, $q_1 = q_1^{\rm 1RSB}$, $q_0 = q_0^{\rm 1RSB}$.

V. CONCLUSION

The step-function binary perceptron has thus far eluded a rigorous establishment of the conjectured storage capacity, eq. (2). This prediction is expected to be exact because of the frozen-1RSB nature of the problem[2,27]. At the same time the work of[31] sheds light on the fact that the structure of the space of solutions is not fully described by the frozen-1RSB picture, and that rare dense and unfrozen regions exist and in fact are amenable to dynamical procedures searching for solutions. It remains to be understood how is it possible that the 1RSB calculation does not capture these dense unfrozen regions of solutions[31]. They do not dominate the equilibrium, but the RSB calculation is expected to describe rare events via their large deviations, which in this case it does not.

In this paper we focus on two cases of the binary perceptron with symmetric constraints, the rectangle binary perceptron and the $u-$function binary perceptron. We prove (up to a numerical assumption) using the second

moment method that the storage capacity agrees in those cases with the annealed upper bound, except for the $u-$function binary perceptron for $K > K^*$ eq. (5). We analyze the 1RSB solution in that case and indeed obtain a lower prediction for the storage capacity. However, we do not expect the 1RSB to provide the exact solution because it does not respect the up-down symmetry of the problem. Though the precise nature of the satisfiable phase for the $u-$function binary perceptron for $K > K^*$ remains illusive, we can conjecture it is full-RSB[17–19]. Establishing this rigorously would provide much deeper understanding and remains a challenging subject for future work.

ACKNOWLEDGEMENT

We thank Florent Krzakala, Joe Neeman, and Pierfrancesco Urbani for useful discussions. We acknowledge funding from the ERC under the European Unions Horizon 2020 Research and Innovation Programme Grant Agreement 714608-SMiLe. WP was supported in part by EPSRC grant EP/P009913/1.

[1]E. Gardner & B. Derrida. Optimal storage properties of neural network models. *J. Phys. A: Math. and Gen*, 1988.
[2]W. Krauth & M. Mézard. Storage capacity of memory networks with binary couplings. *J. Phys. France*, 1989.
[3]Timothy LH Watkin, Albrecht Rau, and Michael Biehl. The statistical mechanics of learning a rule. *Reviews of Modern Physics*, 65(2):499, 1993.
[4]HS Seung, Haim Sompolinsky, and N Tishby. Statistical mechanics of learning from examples. *Physical Review A*, 45(8):6056, 1992.
[5]A. Engel & C. Van den Broeck. *Statistical mechanics of learning*. Cambridge university press, 2001.
[6]H. Nishimori. *Statistical Physics of Spin Glasses and Information Processing: An Introduction*. Oxford University Press, Oxford, UK, 2001.
[7]Michel Talagrand. The Parisi formula. *Annals of mathematics*, pages 221–263, 2006.
[8]Michel Talagrand. *Spin glasses: a challenge for mathematicians: cavity and mean field models*, volume 46. Springer Science & Business Media, 2003.
[9]M. Mézard & A. Montanari. *Information, Physics, and Computation*. Oxford Graduate Texts, 2009.
[10]Dimitris Achlioptas, Amin Coja-Oghlan, and Federico Ricci-Tersenghi. On the solution-space geometry of random constraint satisfaction problems. *Random Structures & Algorithms*, 38(3):251–268, 2011.
[11]Dmitry Panchenko. The Parisi formula for mixed p-spin models. *The Annals of Probability*, 42(3):946–958, 2014.
[12]Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 59–68. ACM, 2015.
[13]Jeong Han Kim and James R Roche. Covering cubes by random half cubes, with applications to binary neural networks. *Journal of Computer and System Sciences*, 56(2):223–252, 1998.
[14]Mihailo Stojnic. Discrete perceptrons. *arXiv preprint arXiv:1306.4375*, 2013.
[15]Jian Ding and Nike Sun. Capacity lower bound for the Ising perceptron. *arXiv preprint arXiv:1809.07742*, 2018.
[16]S. Franz, G. Parisi, M. Sevelev, P. Urbani, and F. Zamponi. Universality of the SAT-UNSAT (jamming) threshold in non-convex continuous constraint satisfaction problems. *SciPost Phys*, 2017.
[17]Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979.
[18]Giorgio Parisi. A sequence of approximated solutions to the sk model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115, 1980.
[19]Giorgio Parisi. The order parameter for spin glasses: a function on the interval 0-1. *Journal of Physics A: Mathematical and General*, 13(3):1101, 1980.
[20]James G Wendel. A problem in geometric probability. *Math. Scand*, 11:109–111, 1962.
[21]Thomas M Cover. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE transactions on electronic computers*, (3):326–334, 1965.
[22]Mariya Shcherbina and Brunello Tirozzi. Rigorous solution of the Gardner problem. *Communications in mathematical physics*, 234(3):383–422, 2003.
[23]Mihailo Stojnic. Another look at the Gardner problem. *arXiv preprint arXiv:1306.3979*, 2013.
[24]Silvio Franz and Giorgio Parisi. The simplest model of jamming. *Journal of Physics A: Mathematical and Theoretical*, 49(14):145001, 2016.
[25]Mihailo Stojnic. Negative spherical perceptron. *arXiv preprint arXiv:1306.3980*, 2013.
[26]Dimitris Achlioptas and Cristopher Moore. The asymptotic order of the random k-SAT threshold. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 779–788. IEEE, 2002.
[27]K.Y.M Wong & Y. Kabashima H. Huang. Entropy landscape of solutions in the binary perceptron problem. *Journal of Physics A: Mathematical and Theoretical*, 2013.
[28]Haiping Huang and Yoshiyuki Kabashima. Origin of the computational hardness for learning with binary synapses. *Physical Review E*, 90(5):052813, 2014.
[29]Lenka Zdeborová and Marc Mézard. Constraint satisfaction problems with isolated solutions are hard. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(12):P12004, 2008.
[30]Alfredo Braunstein and Riccardo Zecchina. Learning by message passing in networks of discrete synapses. *Physical review letters*, 96(3):030201, 2006.
[31]Carlo Baldassi, Alessandro Ingrosso, Carlo Lucibello, Luca Saglietti, and Riccardo Zecchina. Subdominant dense clusters allow for simple learning and high computational performance in neural networks with discrete synapses. *Physical review letters*, 115(12):128101, 2015.
[32]Carlo Baldassi, Christian Borgs, Jennifer T Chayes, Alessandro Ingrosso, Carlo Lucibello, Luca Saglietti, and Riccardo Zecchina. Unreasonable effectiveness of learning neural networks: From accessible states and robust ensembles to basic algorithmic schemes. *Proceedings of the National Academy of Sciences*, 113(48):E7655–E7662, 2016.
[33]Lenka Zdeborová and Marc Mézard. Locked constraint satisfaction problems. *Physical review letters*, 101(7):078702, 2008.
[34]Lenka Zdeborová and Florent Krzakala. Quiet planting in the locked constraint satisfaction problems. *SIAM Journal on Discrete Mathematics*, 25(2):750–770, 2011.

[35]Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *Foundations of Computer Science, 2008. FOCS'08. IEEE 49th Annual IEEE Symposium on*, pages 793–802. IEEE, 2008.
[36]Florent Krzakala and Lenka Zdeborová. Hiding quiet solutions in random constraint satisfaction problems. *Physical review letters*, 102(23):238701, 2009.
[37]Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461, 2015.
[38]Amin Coja-Oghlan, Florent Krzakala, Will Perkins, and Lenka Zdeborová. Information-theoretic thresholds from the cavity method. *Advances in Mathematics*, 333:694–795, 2018.
[39]Ehud Friedgut. Sharp thresholds of graph properties, and the k-SAT problem. *Journal of the American mathematical Society*, 12(4):1017–1054, 1999.
[40]OC Martin, M Mézard, and O Rivoire. Frozen glass phase in the multi-index matching problem. *Physical review letters*, 93(21):217205, 2004.
[41]C. Schülke. *Statistical physics of linear and bilinear inference problems*. PhD thesis, Université Paris Diderot - La Sapienza, 2016.
[42]J.R.L de Almeida and D.J Thouless. Stability of the Sherrington-Kirkpatrick solution of a spin glass model. *J. Phys. A: Math. Gen*, 1978.
[43]& M.Virasoro M.Mézard, G.Parisi. *Spin glasses and beyond*. World Science, Singapore, 1987.
[44]R. Monasson. Structural glass transition and the entropy of the metastable states. *Physical Review Letter*, (75 2847), 1995.

VI. APPENDICES

A. General replica calculation

We present here the replica computation for general prior distribution P_w and constraint function φ. In order to compute the quenched average of the free entropy, we consider the partition function of $n \in \mathbb{N}$ identical copies of the initial system. Using the replica trick, and an analytical continuation, the averaged free entropy ϕ of the initial system reads:

$$\phi(\alpha) \equiv \lim_{N\to+\infty} \frac{1}{N}\mathbb{E}_{\mathbf{X}}[\log(\mathcal{Z}(\mathbf{X}))] = \lim_{N\to+\infty}\lim_{n\to 0}\frac{1}{N}\frac{\partial \log\left(\mathbb{E}_{\mathbf{X}}[\mathcal{Z}(\mathbf{X})^n]\right)}{\partial n}\,, \tag{23}$$

where the replicated partition function can be written as

$$\mathbb{E}_{\mathbf{X}}[\mathcal{Z}(\mathbf{X})^n] = \int d\mathbf{X} P_X(\mathbf{X})\mathcal{Z}(\mathbf{X})^n = \int d\mathbf{X}P_X(\mathbf{X})\prod_{a=1}^{n}\int d\mathbf{w}^a P_w(\mathbf{w}^a)\int d\mathbf{z}^a \mathcal{C}(\mathbf{z}^a)\delta(\mathbf{z}^a - \mathbf{X}\mathbf{w}^a)\,, \tag{24}$$

with the global constraint function $\mathcal{C}(\mathbf{z}) = \prod_{\mu=1}^{M}\varphi(z_\mu)$.

We suppose that inputs are *iid* distributed from $P_X \triangleq \mathcal{N}\left(0, \frac{1}{N}\right)$. More precisely, for $i,j \in [1:N]$, $\mu,\nu \in [1:M]$, $\mathbb{E}_{\mathbf{X}}[X_{i\mu}X_{j\nu}] = \frac{1}{N}\delta_{\mu\nu}\delta_{ij}$. Hence $z_\mu^a = \sum_{i=1}^N X_{i\mu}w_i^a$ is the sum of *iid* random variables. The central limit theorem insures that $z_\mu^a \sim \mathcal{N}\left(\mathbb{E}_{\mathbf{X}}[z_\mu^a], \mathbb{E}_{\mathbf{X}}[z_\mu^a z_\mu^b]\right)$, with two first moments:

$$\begin{cases}\mathbb{E}_{\mathbf{X}}[z_\mu^a] = \sum_{i=1}^N \mathbb{E}_{\mathbf{X}}[X_{i\mu}]w_i^a = 0\\ \mathbb{E}_{\mathbf{X}}[z_\mu^a z_\mu^b] = \sum_{ij}\mathbb{E}_{\mathbf{X}}[X_{i\mu}X_{j\mu}]w_i^a w_j^b = \frac{1}{N}\sum_{ij}\delta_{ij}w_i^a w_j^b = \frac{1}{N}\sum_{i=1}^N w_i^a w_i^b\,.\end{cases} \tag{25}$$

In the following we introduce the symmetric overlap matrix $\mathbf{Q} \equiv (\frac{1}{N}\sum_{i=1}^N w_i^a w_i^b)_{a,b=1..n}$. Define $\tilde{\mathbf{z}}_\mu \equiv (z_\mu^a)_{a=1..n}$ and $\tilde{\mathbf{w}}_i \equiv (w_i^a)_{a=1..n}$. $\tilde{\mathbf{z}}_\mu$ follows a multivariate gaussian distribution $\tilde{\mathbf{z}}_\mu \sim P_{\tilde{z}} \triangleq \mathcal{N}(\mathbf{0},\mathbf{Q})$ and $P_{\tilde{w}}(\tilde{\mathbf{w}}) = \prod_{a=1}^n[\delta(\tilde{w}_a - 1) + \delta(\tilde{w}_a + 1)]$. Introducing the change of variable and the Fourier representation of the δ-Dirac function that involves a new parameter $\hat{\mathbf{Q}}$:

$$1 = \int d\mathbf{Q}\prod_{a\le b}\delta\left(NQ_{ab} - \sum_{i=1}^N w_i^a w_i^b\right) = \int d\mathbf{Q}\int d\hat{\mathbf{Q}}\exp\left(-\frac{N}{2}Tr(\mathbf{Q}\hat{\mathbf{Q}})\right)\exp\left(\frac{1}{2}\sum_{i=1}^N \tilde{\mathbf{w}}_i^\intercal\hat{\mathbf{Q}}\tilde{\mathbf{w}}_i\right),$$

the replicated partition function becomes an integral over the matrix parameters $\mathbf{Q}$ and $\hat{\mathbf{Q}}$, that can be evaluated using Laplace method in the $N\to\infty$ limit,

$$\mathbb{E}_{\mathbf{X}}[\mathcal{Z}(\mathbf{X})^n] = \int d\mathbf{Q}d\hat{\mathbf{Q}}e^{-N\left(\frac{1}{2}\mathrm{Tr}(\mathbf{Q}\hat{\mathbf{Q}}) - \log\left(\int d\tilde{\mathbf{w}}P_{\tilde{w}}(\tilde{\mathbf{w}})e^{\frac{1}{2}\tilde{\mathbf{w}}^\intercal\hat{\mathbf{Q}}\tilde{\mathbf{w}}}\right) - \alpha\log\left(\int d\tilde{\mathbf{z}}P_{\tilde{z}}(\tilde{\mathbf{z}})\mathcal{C}(\tilde{\mathbf{z}})\right)\right)} \tag{26}$$

$$= \int d\mathbf{Q}d\hat{\mathbf{Q}}e^{-NS_n(\mathbf{Q},\hat{\mathbf{Q}})} \underset{N\to\infty}{\simeq} e^{-N\cdot\mathrm{SP}_{\mathbf{Q},\hat{\mathbf{Q}}}\left\{S_n(\mathbf{Q},\hat{\mathbf{Q}})\right\}}, \tag{27}$$

where SP states for saddle point and we defined

$$\begin{cases} S_n(\mathbf{Q},\hat{\mathbf{Q}}) = \frac{1}{2}\mathrm{Tr}(\mathbf{Q}\hat{\mathbf{Q}}) - \log(\mathcal{I}_n^w(\hat{\mathbf{Q}})) - \alpha\log\left(\mathcal{I}_n^z(\mathbf{Q})\right) \\ \mathcal{I}_n^w(\hat{\mathbf{Q}}) = \int_{\mathbb{R}^n} d\tilde{\mathbf{w}} P_{\tilde{w}}(\tilde{\mathbf{w}}) e^{\frac{1}{2}\tilde{\mathbf{w}}^\intercal\hat{\mathbf{Q}}\tilde{\mathbf{w}}} \\ \mathcal{I}_n^z(\mathbf{Q}) = \int_{\mathbb{R}^n} d\tilde{\mathbf{z}} P_{\tilde{z}}(\tilde{\mathbf{z}})\mathcal{C}(\tilde{\mathbf{z}}). \end{cases} \tag{28}$$

Finally, using eq. (23) and switching the two limits $n \to 0$ and $N \to \infty$, the quenched free entropy ϕ simplifies as a saddle point equation

$$\phi(\alpha) = -\mathrm{SP}_{\mathbf{Q},\hat{\mathbf{Q}}}\left\{\lim_{n\to 0}\frac{\partial S_n(\mathbf{Q},\hat{\mathbf{Q}})}{\partial n}\right\}, \tag{29}$$

over general symmetric matrices $\mathbf{Q}$ and $\hat{\mathbf{Q}}$. In the following we will assume simple ansatz for these matrices that allows to get analytic expressions in n in order to take the derivative.

B. RS entropy

Let's compute the functional $S_n(\mathbf{Q},\hat{\mathbf{Q}})$ appearing in the free entropy eq. (29) in the simplest ansatz: the Replica Symmetric ansatz. This later assumes that all replica remain equivalent with a common overlap $q_0 = \frac{1}{N}\sum_{i=1}^N w_i^a w_i^b$ for $a \neq b$ and a norm $Q = \frac{1}{N}\sum_{i=1}^N w_i^a w_i^a$, leading to the following expressions of the matrices $\mathbf{Q}$ and $\hat{\mathbf{Q}} \in \mathbb{R}^{n\times n}$:

$$\mathbf{Q} = \begin{pmatrix} Q & q_0 & \dots & q_0 \\ q_0 & Q & \dots & \dots \\ \dots & \dots & \dots & q_0 \\ q_0 & \dots & q_0 & Q \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{Q}} = \begin{pmatrix} \hat{Q} & \hat{q}_0 & \dots & \hat{q}_0 \\ \hat{q}_0 & \hat{Q} & \dots & \dots \\ \dots & \dots & \dots & \hat{q}_0 \\ \hat{q}_0 & \dots & \hat{q}_0 & \hat{Q} \end{pmatrix}. \tag{30}$$

Let's compute separately the terms involved in the functional $S_n(\mathbf{Q},\hat{\mathbf{Q}})$ eq. (28): the first is a trace term, the second a term of prior $\mathcal{I}_n^w$ and finally the third a term depending on the constraint $\mathcal{I}_n^z$.

a. Trace term The trace term can be easily computed and takes the following form:

$$\frac{1}{2}\mathrm{Tr}(\mathbf{Q}\hat{\mathbf{Q}})\Big|_{\mathrm{RS}} = \frac{1}{2}\left(nQ\hat{Q} + n(n-1)q_0\hat{q}_0\right). \tag{31}$$

b. Prior integral Evaluated at the RS fixed point, and using a gaussian identity also known as a Hubbard-Stratonovich transformation, the prior integral can be further simplified

$$\mathcal{I}_n^w(\hat{\mathbf{Q}})\Big|_{\mathrm{RS}} = \int d\tilde{\mathbf{w}} P_{\tilde{w}}(\tilde{\mathbf{w}}) e^{\frac{1}{2}\tilde{\mathbf{w}}^\intercal\hat{\mathbf{Q}}\tilde{\mathbf{w}}} = \int d\tilde{\mathbf{w}} P_{\tilde{w}}(\tilde{\mathbf{w}}) \exp\left(\frac{(\hat{Q}-\hat{q}_0)}{2}\sum_{a=1}^n(\tilde{w}^a)^2\right)\exp\left(\hat{q}_0\left(\sum_{a=1}^n \tilde{w}^a\right)^2\right) \tag{32}$$

$$= \int Dt\left[\int dw P_w(w)\exp\left(\frac{(\hat{Q}-\hat{q}_0)}{2}w^2 + t\sqrt{\hat{q}_0}w\right)\right]^n. \tag{33}$$

c. Constraint integral Recall the vector $\tilde{\mathbf{z}} \sim P_{\tilde{z}} \triangleq \mathcal{N}(\mathbf{0},\mathbf{Q})$ follows a gaussian distribution with zero mean and covariance matrix $\mathbf{Q}$. In the RS ansatz, the covariance can be rewritten as a linear combination of the identity $\mathbf{I}$ and $\mathbf{J}$ the matrix with all ones entries of size $n \times n$: $\mathbf{Q}|_{\mathrm{RS}} = (Q-q_0)\mathbf{I} + q_0\mathbf{J}$, that allows to split the variable $z^a = \sqrt{q_0}t + \sqrt{Q-q_0}u^a$ with $t \sim \mathcal{N}(0,1)$ and $\forall a,\ u_a \sim \mathcal{N}(0,1)$. Finally, the constraint integral reads:

$$\mathcal{I}_n^z(\mathbf{Q})|_{\mathrm{RS}} = \int d\tilde{\mathbf{z}} P_{\tilde{z}}(\tilde{\mathbf{z}})\mathcal{C}(\tilde{\mathbf{z}}) = \int Dt \int \prod_{a=1}^n Du^a \varphi\left(\sqrt{q_0}t + \sqrt{Q-q_0}u^a\right) \tag{34}$$

$$= \int Dt\left[\int Du\varphi\left(\sqrt{q_0}t + \sqrt{Q-q_0}u\right)\right]^n. \tag{35}$$

d. Summary and RS free entropy $\phi_{\rm RS}$ Finally putting pieces together, the functional S_n taken at the RS fixed point has an explicit formula and dependency in n:

$$S_n(\mathbf{Q},\hat{\mathbf{Q}})\Big|_{\rm RS} = \frac{1}{2}\mathrm{Tr}(\mathbf{Q}\hat{\mathbf{Q}}) - \log(\mathcal{I}_w^n(\hat{\mathbf{Q}})) - \alpha\log\left(\mathcal{I}_z^n(\mathbf{Q})\right)\Big|_{\rm RS} \tag{36}$$

$$\underset{n\to 0}{\simeq} \frac{1}{2}\left(nQ\hat{Q} + n(n-1)q_0\hat{q}_0\right) - n\int Dt\log\left(\int dwP_w(w)\exp\left(\frac{(\hat{Q}-\hat{q}_0)}{2}w^2 + t\sqrt{\hat{q}_0}w\right)\right) \tag{37}$$

$$- n\alpha\int Dt\log\left(\int Du\varphi\left(y,\sqrt{q_0}t+\sqrt{Q-q_0}u\right)\right). \tag{38}$$

Finally taking the derivative with respect to n and the $n\to 0$ limit, the RS free entropy has a simple expression

$$\phi_{\rm RS}(\alpha) = \mathrm{SP}_{q_0,\hat{q}_0}\left\{-\frac{1}{2}Q\hat{Q} + \frac{1}{2}q_0\hat{q}_0 + \mathcal{I}^w_{\rm RS}(\hat{q}_0) + \alpha\mathcal{I}^z_{\rm RS}(q_0)\right\}, \tag{39}$$

with $Q=\hat{Q}=1$ and the following notations,

$$\begin{cases} \mathcal{I}^w_{\rm RS}(\hat{q}_0) \equiv \int Dt\log\left(\int dwP_w(w)\exp\left(\frac{(\hat{Q}-\hat{q}_0)}{2}w^2 + t\sqrt{\hat{q}_0}w\right)\right) \\ \mathcal{I}^z_{\rm RS}(q_0) \equiv \int Dt\log\left(\int Dz\varphi\left(\sqrt{q_0}t+\sqrt{Q-q_0}z\right)\right) \end{cases}. \tag{40}$$

C. 1RSB entropy

The free entropy eq. (23) can also be evaluated at the simplest non trivial fixed point: the one step Replica Symmetry Breaking ansatz (1RSB). Instead assuming that replicas are equivalent, it assumes that the symmetry between replica is broken and that replicas are clustered in different states, with inner overlap q_1 and outer overlap q_0. Translating this in a matrix formulation, the matrices can be expressed as

$$\mathbf{Q} = q_0\mathbf{J}_n + (q_1-q_0)\,\mathbf{I}_{\frac{n}{m}}\otimes\mathbf{J}_m + (Q-q_1)\,\mathbf{I}_n \quad\text{and}\quad \hat{\mathbf{Q}} = \hat{q}_0\mathbf{J}_n + (\hat{q}_1-\hat{q}_0)\,\mathbf{I}_{\frac{n}{m}}\otimes\mathbf{J}_m + \left(\hat{Q}-\hat{q}_1\right)\mathbf{I}_n\,. \tag{41}$$

a. Trace term Again, the trace term can be easily computed

$$\frac{1}{2}\mathrm{Tr}(\mathbf{Q}\hat{\mathbf{Q}})\Big|_{\rm 1RSB} = \frac{1}{2}\left(nQ\hat{Q} + n(m-1)q_1\hat{q}_1 + n(n-m)q_0\hat{q}_0\right). \tag{42}$$

b. Prior integral Separating replicas with different overlaps, the prior integral can be written as

$$\mathcal{I}_n^w(\hat{\mathbf{Q}})\Big|_{\rm 1RSB} = \int d\tilde{\mathbf{w}}P_{\tilde{w}}(\tilde{\mathbf{w}})e^{\frac{(\hat{Q}-\hat{q}_1)}{2}\sum_{a=1}^n(\tilde{w}^a)^2 + \frac{(\hat{q}_1-\hat{q}_0)}{2}\sum_{k=1}^{\frac{n}{m}}\sum_{a,b=(k-1)m+1}^{km}\tilde{w}^a\tilde{w}^b + \frac{\hat{q}_0}{2}\left(\sum_{a=1}^n\tilde{w}^a\right)^2} \tag{43}$$

$$= \int Dt_0\left[\int Dt_1\left[\int dwP_w(w)\exp\left(\frac{(\hat{Q}-\hat{q}_1)}{2}w^2 + \left(\sqrt{\hat{q}_0}t_0+\sqrt{\hat{q}_1-\hat{q}_0}t_1\right)w\right)\right]^m\right]^{\frac{n}{m}} \tag{44}$$

c. Constraint integral Again the vector $\tilde{\mathbf{z}}\sim P_{\tilde{z}}\triangleq\mathcal{N}(\mathbf{0},\mathbf{Q})$ follows a gaussian vector with zero mean and covariance $\mathbf{Q}|_{\rm 1RSB} = q_0\mathbf{J}_n + (q_1-q_0)\,\mathbf{I}_{\frac{n}{m}}\otimes\mathbf{J}_m + (Q-q_1)\,\mathbf{I}_n$. The gaussian vector of covariance $\mathbf{Q}|_{\rm 1RSB}$ can be decomposed in a sum of normal gaussian vectors $t_0\sim\mathcal{N}(0,1)$, $\forall k\in[1:\frac{n}{m}], t_k\sim\mathcal{N}(0,1)$ and $\forall a\in[(k-1)m+1:km]$, $u_a\sim\mathcal{N}(0,1)$: $z^a = \sqrt{q_0}t_0 + \sqrt{q_1-q_0}t_k + \sqrt{Q-q_1}u_a$. Finally the constraint integral reads

$$\mathcal{I}_n^z(\mathbf{Q})|_{\rm 1RSB} = \int Dt_0\int\prod_{k=1}^{\frac{n}{m}}Dt_k\int\prod_{a=(k-1)m+1}^{km}Du_a\varphi(\sqrt{q_0}t_0+\sqrt{q_1-q_0}t_k+\sqrt{Q-q_1}u_a) \tag{45}$$

$$= \int Dt_0\left[\int Dt_1\left[\int Du\varphi(\sqrt{q_0}t_0+\sqrt{q_1-q_0}t_1+\sqrt{Q-q_1}u)\right]^m\right]^{\frac{n}{m}}. \tag{46}$$

d. Summary and 1RSB free entropy $\phi_{\rm 1RSB}$ Gathering the previous computations eq. (42, 44, 46), the functional S_n evaluated at the 1RSB fixed point reads:

$$S_n(\mathbf{Q},\hat{\mathbf{Q}})\Big|_{\rm 1RSB} = \frac{1}{2}\mathrm{Tr}(\mathbf{Q}\hat{\mathbf{Q}}) - \log(\mathcal{I}_w^n(\hat{\mathbf{Q}})) - \alpha\log\left(\mathcal{I}_z^n(\mathbf{Q})\right)\Big|_{\rm 1RSB} \tag{47}$$

$$\underset{n\to 0}{\simeq} \frac{1}{2}\left(nQ\hat{Q} + n(m-1)q_1\hat{q}_1 + n(n-m)q_0\hat{q}_0\right) \tag{48}$$

$$-\frac{n}{m}\int Dt_0 \log\left(\int Dt_1\left[\int d\tilde{w}P_w(\tilde{w})\exp\left(\frac{(\hat{Q}-\hat{q}_1)}{2}\tilde{w}^2 + \left(\sqrt{\hat{q}_0}t_0 + \sqrt{\hat{q}_1-\hat{q}_0}t_1\right)\tilde{w}\right)\right]^m\right) \tag{49}$$

$$-\alpha\frac{n}{m}\int dy\int Dt_0\log\left(\int Dt_1\left[\int Du\varphi(y,\sqrt{q_0}t_0+\sqrt{q_1-q_0}t_1+\sqrt{Q-q_1}u)\right]^m\right). \tag{50}$$

Let's introduce the replicated free entropy following[44]. We consider m reals replicas of the same system and we imagine we put a small field, that allows the m replicas to fall in the same state. The replicated free entropy is the free entropy corresponding to these m uncorrelated copies in the limit of zero coupling. To compute it, we consider $n' = \frac{n}{m}$ replicas. Denoting $\mathbf{q} = (q_0, q_1)$ and $\hat{\mathbf{q}} = (\hat{q}_0, \hat{q}_1)$, the replicated free entropy reads as m times the free entropy of n replicas with 1RSB structure:

$$\Phi^{\rm 1RSB}(\alpha) := \left(\lim_{N\to\infty}\frac{1}{N}\mathbb{E}_{\mathbf{X}}\left[\log(\mathcal{Z}_m(\mathbf{X})\right]\right) \simeq \lim_{N\to\infty}\frac{1}{N}\lim_{n'\to 0}\frac{\partial\log\left(\mathbb{E}_{\mathbf{X}}[\mathcal{Z}^{mn'}(\mathbf{X})]\right)}{\partial n'} \tag{51}$$

$$= m\left(\lim_{N\to\infty}\lim_{n\to 0}\frac{1}{N}\frac{\partial\log\left(\mathbb{E}[\mathcal{Z}^n(\mathbf{X})]_{\mathbf{X}}\right)}{\partial n}\right) = m\left(-\mathrm{SP}_{\mathbf{Q},\hat{\mathbf{Q}}}\left\{\lim_{n\to 0}\frac{\partial S_n(\mathbf{Q},\hat{\mathbf{Q}})}{\partial n}\right\}\right) \tag{52}$$

$$= \underset{\mathbf{q},\hat{\mathbf{q}}}{\mathrm{SP}}\left\{\frac{m}{2}\left(q_1\hat{q}_1 - Q\hat{Q}\right) + \frac{m^2}{2}\left(q_0\hat{q}_0 - q_1\hat{q}_1\right) + m\mathcal{I}^w_{\rm 1RSB}(\hat{\mathbf{q}}) + \alpha m\mathcal{I}^z_{\rm 1RSB}(\mathbf{q})\right\}. \tag{53}$$

with $\mathbf{t} = (t_0, t_1)$, g_0^w and f_0^z defined in eq. (21) and

$$\mathcal{I}^w_{\rm 1RSB}(\hat{\mathbf{q}}) = \frac{1}{m}\int Dt_0\log\left(\int Dt_1 g_0^w\left(\mathbf{t},\hat{\mathbf{q}}\right)^m\right) \quad \text{and} \quad \mathcal{I}^z_{\rm 1RSB}(\mathbf{q}) = \frac{1}{m}\int Dt_0\log\left(\int Dt_1 f_0^z\left(\mathbf{t},\mathbf{q}\right)^m\right). \tag{54}$$

D. RS Stability

1. De Almeida Thouless RS Stability

The stability of a given saddle point ansatz is related to the positivity the hessian of the functional S_n. This stability analysis has first been done by de Almeida Thouless and following[1,5,42], replicons eigenvalues of the RS ansatz λ_3^A and λ_3^B can be expressed as functions of $\{g_i^w, f_i^z\}_{i=0}^2$ defined in eq. (16):

$$\lambda_3^A(q_0) = \frac{1}{(Q-q_0)^2}\int Dt\frac{\left(f_0^z(f_0^z-f_2^z)+(f_1^z)^2\right)^2}{(f_0^z)^4}(t,q_0), \quad \text{and} \quad \lambda_3^B(\hat{q}_0) = \int Dt\frac{\left(g_0^w g_2^w - (g_1^w)^2\right)^2}{(g_0^w)^4}(t,\hat{q}_0). \tag{55}$$

The instability AT-line is defined when the determinant of the hessian vanishes that translates as an implicit equation over α, where $q_0, \hat{q}_0$ are solution of the saddle point equations eq. (15) at $\alpha = \alpha_{AT}$:

$$\frac{1}{\alpha_{AT}} = \lambda_3^A\left(q_0(\alpha_{AT}),\beta\right)\lambda_3^B\left(\hat{q}_0(\alpha_{AT})\right). \tag{56}$$

However for $\alpha < \alpha_{AT}$, $(q_0,\hat{q}_0) = (0,0)$ is the only solution. Using $\{\tilde{f}_i^z, \tilde{g}_i^w\}_{i=0}^2$ defined eq. (58), this expression simplifies because of the symmetry of the prior distribution P_w and the constraints φ in the rectangle and $u-$function cases. In fact the symmetry imposes $\tilde{f}_1^z = 0$ and $\tilde{g}_1^w = 0$ and the condition reads:

$$\frac{1}{\alpha_{AT}} = \left(\frac{\tilde{f}_2^z - \tilde{f}_0^z}{\tilde{f}_0^z}\right)^2\left(\frac{\tilde{g}_2^w}{\tilde{g}_0^w}\right)^2. \tag{57}$$

2. Existence and stability of the RS fixed point $(q_0, \hat{q}_0) = (0, 0)$

We provide an alternative approach to get the instability condition of the RS solution for symmetric prior and constraint. In this symmetric case, the stability can be derived from the existence and stability of the symmetric fixed point $(q_0, \hat{q}_0) = (0, 0)$. Let's define

$$\begin{cases} F(q_0) \equiv \alpha \int Dt \frac{(f_1^z)^2 - 2t\sqrt{q_0} f_0^z f_1^z + q_0 t^2 (f_0^z)^2}{(1-q_0)^2 (f_0^z)^2}(t, q_0)\,, \\ G(\hat{q}_0) \equiv \int Dt \frac{g_2^w - t\hat{q}_0^{-1/2} g_1^w}{g_0^w}(t, \hat{q}_0)\,, \end{cases} \quad \text{with} \quad \begin{cases} \tilde{f}_i^z(y) \equiv \int Dz z^i \varphi(z)\,, \\ \tilde{g}_i^w \equiv \int dw w^i P_w(w) e^{\frac{w^2}{2}}\,. \end{cases} \tag{58}$$

In fact the saddle point equations at the RS fixed point eq. (15) can be written using the functions F, G, and can be reduced to a single fixed point equation over q_0:

$$\begin{cases} q_0 = G(\hat{q}_0)\,, \\ \hat{q}_0 = F(q_0)\,, \end{cases} \Rightarrow \Big\{ q_0 = G \circ F(q_0) \equiv H(q_0)\,. \tag{59}$$

As stressed above, the RS stability is equivalent to the existence and stability of the fixed point $q_0 = 0$. According to that, let's compute the stability of the above fixed point equation eq. (59). Computing F, F', G, G' in the limit $(q_0, \hat{q}_0) \to (0, 0)$, expanding $\{f_i^z, g_i^w\}_i$ as functions of $\{\tilde{f}_i^z, \tilde{g}_i^w\}_i$ and finally using the symmetry that implies $\tilde{f}_1^z = 0$ and $\tilde{g}_1^w = 0$:

$$\begin{cases} F(q_0) \underset{q_0 \to 0}{=} \alpha \left[\left(\frac{\tilde{f}_1^z}{\tilde{f}_0^z} \right)^2 + q_0 \left(\frac{(\tilde{f}_2^z - \tilde{f}_0^z)^2}{(\tilde{f}_0^z)^2} + 3\frac{(\tilde{f}_1^z)^4}{(\tilde{f}_0^z)^4} - 4\frac{(\tilde{f}_1^z)^2 (\tilde{f}_2^z - \tilde{f}_0^z)}{(\tilde{f}_0^z)^3} \right) + \mathcal{O}(q_0^2) \right] \sim \alpha q_0 \left(\frac{\tilde{f}_2^z - \tilde{f}_0^z}{\tilde{f}_0^z} \right)^2 \underset{q_0 \to 0}{\longrightarrow} 0\,, \\ \frac{\partial F}{\partial q_0}(q_0) \underset{q_0 \to 0}{=} \alpha \left[\left(\frac{\tilde{f}_2^z - \tilde{f}_0^z}{\tilde{f}_0^z} \right)^2 + \left(\frac{\tilde{f}_1^z}{\tilde{f}_0^z} \right)^2 \left(3\frac{(\tilde{f}_1^z)^2}{(\tilde{f}_0^z)^2} - 4\frac{\tilde{f}_2^z - \tilde{f}_0^z}{\tilde{f}_0^z} \right) + \mathcal{O}(q_0) \right] \underset{q_0 \to 0}{\longrightarrow} \alpha \left(\frac{\tilde{f}_2^z - \tilde{f}_0^z}{\tilde{f}_0^z} \right)^2\,, \\ G(\hat{q}_0) \underset{\hat{q}_0 \to 0}{=} \left(\frac{\tilde{g}_1^w}{\tilde{g}_0^w} \right)^2 + \hat{q}_0 \left(\left(\frac{\tilde{g}_2^w}{\tilde{g}_0^w} \right)^2 + \frac{\tilde{g}_1^w}{\tilde{g}_0^w} \left(3 \left(\frac{\tilde{g}_1^w}{\tilde{g}_0^w} \right)^3 - 4\frac{\tilde{g}_1^w \tilde{g}_2^w}{(\tilde{g}_0^w)^2} \right) \right) + \mathcal{O}(\hat{q}_0^{3/2}) \underset{\hat{q}_0 \to 0}{\longrightarrow} 0\,, \\ \frac{\partial G}{\partial \hat{q}_0} \underset{\hat{q}_0 \to 0}{=} \left(\frac{\tilde{g}_2^w}{\tilde{g}_0^w} \right)^2 + \frac{\tilde{g}_1^w}{\tilde{g}_0^w} \left(3 \left(\frac{\tilde{g}_1^w}{\tilde{g}_0^w} \right)^3 - 4\frac{\tilde{g}_1^w \tilde{g}_2^w}{(\tilde{g}_0^w)^2} \right) + \mathcal{O}(\sqrt{\hat{q}_0}) \underset{\hat{q}_0 \to 0}{\longrightarrow} \left(\frac{\tilde{g}_2^w}{\tilde{g}_0^w} \right)^2\,. \end{cases} \tag{60}$$

Finally, the existence and stability conditions of the fixed point $(q_0, \hat{q}_0) = (0, 0)$ translate as an explicit condition over α that defines α_{AT}

$$\begin{cases} H(q_0) = G \circ F(q_0) \underset{q_0 \to 0}{\to} 0 \\ \left. \frac{\partial H}{\partial q_0} \right|_{q_0 = 0} = \left. \frac{\partial G}{\partial \hat{q}_0} \right|_{\hat{q}_0 = 0} \left. \frac{\partial F}{\partial q_0} \right|_{q_0 = 0} \leq 1\,, \end{cases} \Rightarrow \quad \alpha \leq \left[\left(\frac{\tilde{f}_2^z - \tilde{f}_0^z}{\tilde{f}_0^z} \right)^2 \left(\frac{\tilde{g}_2^w}{\tilde{g}_0^w} \right)^2 \right]^{-1} \equiv \alpha_{AT}\,. \tag{61}$$

E. Moments at finite temperature

In this section we generalize the definition of the partition function for any temperature T. The energy of a configuration $\mathbf{w}$ is defined as the number of unsatisfied constraints and the corresponding partition function is defined by $\mathcal{Z}(\mathbf{X}, T) = \sum_{\mathbf{w} \in \{\pm 1\}^N} e^{-\mathcal{E}(\mathbf{w})/T}$. In particular for the rectangle and $u-$function constraints, the partition functions at temperature T read

$$\mathcal{Z}_r(\mathbf{X}, T) = \sum_{\mathbf{w} \in \{\pm 1\}^N} \prod_{\mu=1}^{M} e^{-\frac{1}{T}\left(1 - \mathbb{1}\left|z_\mu(\mathbf{w})\right|_{\leq K}\right)} \quad \text{and} \quad \mathcal{Z}_u(\mathbf{X}, T) = \sum_{\mathbf{w} \in \{\pm 1\}^N} \prod_{\mu=1}^{M} e^{-\frac{1}{T}\left(1 - \mathbb{1}\left|z_\mu(\mathbf{w})\right|_{\geq K}\right)}\,. \tag{62}$$

We define the probabilities that constraints are satisfied at temperature T:

$$\begin{cases} p_{r,K,T} \equiv \int Dz e^{-\frac{1}{T}\left(1 - \mathbb{1}|z|_{\leq K}\right)} = e^{-\frac{1}{T}} + (1 - e^{-\frac{1}{T}}) p_{r,K}\,, \\ p_{u,K,T} \equiv \int Dz e^{-\frac{1}{T}\left(1 - \mathbb{1}|z|_{\geq K}\right)} = e^{-\frac{1}{T}} + (1 - e^{-\frac{1}{T}}) p_{u,K}\,, \\ p_{s,K,T} \equiv \int Dz e^{-\frac{1}{T}\left(1 - \mathbb{1} z \geq K\right)} = e^{-\frac{1}{T}} + (1 - e^{-\frac{1}{T}}) p_{s,K}\,. \end{cases} \tag{63}$$

1. First moment at finite temperature

Let $\mathcal{E}^r(N,M,T)$ the event that $\mathcal{Z}_r(\mathbf{X},T)\geq 1$. Let's compute the first moment in the rectangle case,

$$\mathbb{P}[\mathcal{E}^r(N,\alpha N,T)]\leq \mathbb{E}[\mathcal{Z}_r(\mathbf{X}(N,\alpha N),T)] = 2^N\mathbb{E}\left[\prod_{\mu=1}^{\alpha N} e^{-\frac{1}{T}\left(1-\mathbb{1}_{|z_\mu(\mathbf{1})|\leq K}\right)}\right] \tag{64}$$

$$= 2^N p_{r,K,T}^{\alpha N} = \exp(N(\log(2)+\alpha\log(p_{r,K,T})))\,. \tag{65}$$

and this derivation holds similarly for the step and $u-$function.

2. Second moment at finite temperature

Again we show the computation for the rectangle and it can be done similarly for the $u-$function.

a. Expression of $F_{r,K,\alpha,T}$

$$\mathbb{E}[\mathcal{Z}_r(\mathbf{X}(N,\alpha N),T)^2] = \sum_{\mathbf{w_1},\mathbf{w_2}\in\{\pm1\}^N}\mathbb{E}\left[\prod_{\mu=1}^{\alpha N} e^{-\frac{1}{T}\left(1-\mathbb{1}_{|z_\mu(\mathbf{w_1})|\leq K}\right)}e^{-\frac{1}{T}\left(1-\mathbb{1}_{|z_\mu(\mathbf{w_2})|\leq K}\right)}\right] \tag{66}$$

$$= 2^N\sum_{\mathbf{w}\in\{\pm1\}^N}\prod_{\mu=1}^{\alpha N}\mathbb{E}\left[e^{-\frac{1}{T}\left\{\left(1-\mathbb{1}_{|z_\mu(\mathbf{1})|\leq K}\right)+\left(1-\mathbb{1}_{|z_\mu(\mathbf{w})|\leq K}\right)\right\}}\right] \tag{67}$$

$$= 2^N\sum_{l=0}^{N}\binom{N}{l}q_{r,K,T}(l/N)^{\alpha N}\equiv\exp(N(\log(2)+F_{r,K,\alpha,T}))\,, \tag{68}$$

where we defined $q_{r,K,T}$ the probability that two standard Gaussians with correlation β are both at most K in absolute value at temperature T. Defining $\rho(\beta)=1-2\beta$ and

$$\mathcal{I}_{\alpha_1,\beta_1}^{\alpha_2,\beta_2}(\rho)\equiv\int_{\alpha_1}^{\beta_1}\int_{\alpha_2}^{\beta_2}dxdy\frac{e^{-\frac{1}{2}(x^2+y^2+2\rho xy)}}{2\pi\sqrt{1-\rho^2}} = \frac{1}{2\pi}\int_{\alpha_2}^{\beta_2}\int_{\frac{\alpha_1+\rho y}{\sqrt{1-\rho^2}}}^{\frac{\beta_1+\rho y}{\sqrt{1-\rho^2}}}dydxe^{-\frac{y^2+x^2}{2}}\,, \tag{69}$$

the function $F_{r,K,\alpha,T}$ at finite temperature can be written

$$F_{r,K,\alpha,T} = H(\beta)+\alpha\log q_{r,K,T}(\beta)\,,$$

where

$$q_{r,K,T}(\beta)\equiv\int_{\mathbb{R}^2}dxdy\frac{e^{-\frac{1}{2}(x^2+y^2+2\rho(\beta)xy)}}{2\pi\sqrt{1-\rho(\beta)^2}}e^{-\frac{1}{T}\left(\left(1-\mathbb{1}_{|z_\mu(\mathbf{1})|\leq K}\right)+\left(1-\mathbb{1}_{|z_\mu(\mathbf{w})|\leq K}\right)\right)} \tag{70}$$

$$= \mathcal{I}_{-K,K}^{-K,K}+e^{-\frac{1}{T}}\left(\mathcal{I}_{-\infty,-K}^{-K,K}+\mathcal{I}_{K,+\infty}^{-K,K}+\mathcal{I}_{-K,K}^{-\infty,-K}+\mathcal{I}_{-K,K}^{K,+\infty}\right)+e^{-\frac{2}{T}}\left(\mathcal{I}_{-\infty,-K}^{-\infty,-K}+\mathcal{I}_{-\infty,-K}^{K,+\infty}+\mathcal{I}_{K,+\infty}^{-\infty,-K}+\mathcal{I}_{K,+\infty}^{K,+\infty}\right)\,. \tag{71}$$

b. Expression of $\partial_\beta F_{r,K,\alpha,T}$

To compute the derivative of $q_{r,K,T}$, we first introduce

$$\mathcal{G}_\gamma^{\alpha_2,\beta_2}(\rho)\equiv\frac{1}{2\pi}\int_{\alpha_2}^{\beta_2}dye^{-\frac{y^2}{2}}e^{-\frac{1}{2}\frac{(\gamma+\rho y)}{1-\rho^2}}(y+\gamma\rho)\,.$$

The derivative of each integral involved in eq. (71) can be easily computed as

$$\partial_\beta\mathcal{I}_{\alpha_1,\beta_1}^{\alpha_2,\beta_2}(\rho(\beta)) = -\frac{1}{4(\beta(1-\beta))^{3/2}}\left(\mathcal{G}_{\beta_1}^{\alpha_2,\beta_2}-\mathcal{G}_{\alpha_1}^{\alpha_2,\beta_2}\right)(\rho(\beta))\,. \tag{72}$$

Hence taking the derivative of each term of the form $\mathcal{I}_{\alpha_1,\beta_1}^{\alpha_2,\beta_2}$ and simplifying it, the probability $q_{r,K,T}$ reads:

$$q_{r,K,T}(\beta) = -\frac{1}{4(\beta(1-\beta))^{3/2}}\left(\mathcal{G}_K^{-K,K}-\mathcal{G}_{-K}^{-K,K}\right)(\rho)(1-e^{-1/T})^2 = \frac{(1-e^{-1/T})^2}{\pi\sqrt{\beta(1-\beta)}}\left(e^{-\frac{K^2}{2(1-\beta)}}\left(e^{\frac{(2\beta-1)K^2}{2(1-\beta)\beta}}-1\right)\right)\,.$$

In the end, the derivative of the second moment can be evaluated for $\beta = 0$ and $\beta = 1$ at all temperature T :

$$\frac{\partial F_{r,K,\alpha,T}}{\partial \beta}(\beta) = \log\left(\frac{1-\beta}{\beta}\right) + \frac{\alpha}{q_{r,K,T}}\frac{\partial q_{r,K,T}(\beta)}{\partial \beta} \tag{73}$$

$$= \log\left(\frac{1-\beta}{\beta}\right) + \frac{\alpha}{q_{r,K,T}(\beta)}\frac{(1-e^{-1/T})^2}{\pi\sqrt{\beta(1-\beta)}}\left(e^{-\frac{K^2}{2(1-\beta)}}\left(e^{\frac{(2\beta-1)K^2}{2(1-\beta)\beta}}-1\right)\right) \xrightarrow[\beta\to 1/2\pm 1/2]{} \pm\infty\,. \tag{74}$$

In particular at $T = 0$,

$$\frac{\partial F_{r,K,\alpha}}{\partial \beta}(\beta) = \log\left(\frac{1-\beta}{\beta}\right) + \frac{\alpha}{q_{r,K,T}(\beta)}\frac{1}{\pi\sqrt{\beta(1-\beta)}}\left(e^{-\frac{K^2}{2(1-\beta)}}\left(e^{\frac{(2\beta-1)K^2}{2(1-\beta)\beta}}-1\right)\right)\,. \tag{75}$$

c. Expression of $\partial_\beta F_{u,K,\alpha,T}$

Adapting the previous steps and using

$$q_{u,K,T}(\beta) \equiv \int_{\mathbb{R}^2} dxdy \frac{e^{-\frac{1}{2}(x^2+y^2+2\rho(\beta)xy)}}{2\pi\sqrt{1-\rho(\beta)^2}} e^{-\frac{1}{T}\left(\left(1-\mathbb{1}_{|z_\mu(\mathbf{1})|\le K}\right)+\left(1-\mathbb{1}_{|z_\mu(\mathbf{w})|\le K}\right)\right)}$$
$$= \left(\mathcal{I}^{-\infty,-K}_{-\infty,-K} + \mathcal{I}^{K,+\infty}_{-\infty,-K} + \mathcal{I}^{-\infty,-K}_{K,+\infty} + \mathcal{I}^{K,+\infty}_{K,+\infty}\right) + e^{-\frac{1}{T}}\left(\mathcal{I}^{-K,K}_{-\infty,-K} + \mathcal{I}^{-K,K}_{K,+\infty} + \mathcal{I}^{-\infty,-K}_{-K,K} + \mathcal{I}^{K,+\infty}_{-K,K}\right) + e^{-\frac{2}{T}}\left(\mathcal{I}^{-K,K}_{-K,K}\right)$$
$$= q_{r,K,-T}e^{-\frac{2}{T}}\,,$$

and eq. (74) the derivative for the $u-$function is straightforward to compute and is given by

$$\frac{\partial F_{u,K,\alpha,T}}{\partial \beta}(\beta) = \log\left(\frac{1-\beta}{\beta}\right) + \frac{\alpha}{q_{u,K,T}(\beta)}\frac{\partial q_{u,K,T}}{\partial \beta}(\beta)$$
$$= \log\left(\frac{1-\beta}{\beta}\right) + \frac{\alpha}{q_{u,K,T}(\beta)}\frac{(e^{-1/T}-1)^2}{\pi\sqrt{\beta(1-\beta)}}\left(e^{-\frac{K^2}{2(1-\beta)}}\left(e^{\frac{(2\beta-1)K^2}{2(1-\beta)\beta}}-1\right)\right)$$
$$= \xrightarrow[\beta\to 1/2\pm 1/2]{} \pm\infty\,.$$