

# Algorithms and Barriers in the Symmetric Binary Perceptron Model

David Gamarnik

*Massachusetts Institute of Technology*

gamarnik@mit.edu

Will Perkins

*University of Illinois at Chicago*

math@willperkins.org

Eren C. Kızıldağ

*Massachusetts Institute of Technology*

kizildag@mit.edu

Changji Xu

*Harvard University*

cxu@cmsa.fas.harvard.edu

**Abstract**—The binary (or Ising) perceptron is a toy model of a single-layer neural network and can be viewed as a random constraint satisfaction problem with a high degree of connectivity. The model and its symmetric variant, the symmetric binary perceptron (SBP), have been studied widely in statistical physics, mathematics, and machine learning.

The SBP exhibits a dramatic *statistical-to-computational gap*: the densities at which known efficient algorithms find solutions are far below the threshold for the existence of solutions. Furthermore, the SBP exhibits a striking structural property: at all positive constraint densities almost all of its solutions are ‘totally frozen’ singletons separated by large Hamming distance [1], [2]. This suggests that finding a solution to the SBP may be computationally intractable. At the same time, however, the SBP does admit polynomial-time search algorithms at low enough densities. A conjectural explanation for this conundrum was put forth in [3]: efficient algorithms succeed in the face of freezing by finding exponentially rare clusters of large size. However, it was discovered recently that such rare large clusters exist at all subcritical densities, even at those well above the limits of known efficient algorithms [4]. Thus the driver of the statistical-to-computational gap exhibited by this model remains a mystery.

In this paper, we conduct a different landscape analysis to explain the statistical-to-computational gap exhibited by this problem. We show that at high enough densities the SBP exhibits the multi Overlap Gap Property ( $m$ -OGP), an intricate geometrical property known to be a rigorous barrier for large classes of algorithms. Our analysis shows that the  $m$ -OGP threshold (a) is well below the satisfiability threshold; and (b) matches the best known algorithmic threshold up to logarithmic factors as  $m \rightarrow \infty$ . We then prove that the  $m$ -OGP rules out the class of stable algorithms for the SBP above this threshold. We conjecture that the  $m \rightarrow \infty$  limit of the  $m$ -OGP threshold marks the algorithmic threshold for the problem. Furthermore, we investigate the stability of known efficient algorithms for perceptron models and show that the Kim-Roche algorithm [5], devised for the *asymmetric binary perceptron*, is stable in the sense we consider.

**Index Terms**—Binary perceptron, overlap gap property, statistical-to-computational gap, random CSP, average-case complexity, neural networks.

## I. INTRODUCTION

In this paper, we study the *perceptron model*. Proposed initially in the 1960’s [6]–[9], this is a toy model of one-layer neural network storing random patterns as well as

a very natural model in high-dimensional probability. Let  $X_i \in \mathbb{R}^n$ ,  $1 \leq i \leq M$ , be i.i.d. random patterns to be stored. *Storage* of these patterns is achieved if one finds a vector of synaptic weights  $\sigma \in \mathbb{R}^n$  consistent with all  $X_i$ : that is,  $\langle X_i, \sigma \rangle \geq 0$  for  $1 \leq i \leq M$ . There are two main variants of the perceptron: when the vector  $\sigma$  lies on a sphere in  $\mathbb{R}^n$  (the spherical perceptron) and when  $\sigma \in \mathcal{B}_n \triangleq \{-1, 1\}^n$  (the binary or Ising perceptron). For more on the spherical perceptron see [10]–[14]; in this paper we will focus only on the binary perceptron.

A key quantity associated to the perceptron is the *storage capacity*: the maximum number  $M^*$  of such patterns for which there exists a vector of weights  $\sigma \in \mathcal{B}_n$  that is consistent with all  $X_i$ ,  $1 \leq i \leq M^*$ . Investigations beginning with Gardner [10], [15] and Gardner-Derrida [16] in the statistical physics literature provided a detailed, yet non-rigorous, picture for the storage capacity in the case of patterns distributed as  $n$ -dimensional Gaussian vectors.

More general perceptron models are defined by an activation function  $U : \mathbb{R} \rightarrow \{0, 1\}$  (for an even more general setting see [17]). We say a pattern  $X_i$  is *stored* by  $\sigma$  with respect to  $U$  if  $U(\langle X_i, \sigma \rangle) = 1$ . Much recent work on these models have focused on two classes of activity functions:  $U(x) = \mathbf{1}_{x \geq \kappa\sqrt{n}}$  and  $U(x) = \mathbf{1}_{|x| \leq \kappa\sqrt{n}}$ . The first defines the *asymmetric binary perceptron*, the second the *symmetric binary perceptron*. We now detail some of the previous work on these models.

### A. Perceptron models

1) *Asymmetric Binary Perceptron*: We now define the classic binary perceptron, which we call the *asymmetric binary perceptron* (ABP) throughout. Fix  $\kappa \in \mathbb{R}$ ,  $\alpha > 0$ ; and set  $M = \lfloor n\alpha \rfloor \in \mathbb{N}$ . Let  $X_i \stackrel{d}{=} \mathcal{N}(0, I_n)$ ,  $1 \leq i \leq M$ , be i.i.d. random vectors, where  $\mathcal{N}(0, I_n)$  denotes the  $n$ -dimensional multivariate normal distribution with zero mean and identity covariance. Consider the (random) set

$$S_\alpha^A(\kappa) \triangleq \bigcap_{1 \leq i \leq M} \left\{ \sigma \in \mathcal{B}_n : \langle \sigma, X_i \rangle \geq \kappa\sqrt{n} \right\}. \quad (1)$$

The vectors  $X_i \in \mathbb{R}^n$ ,  $1 \leq i \leq M$ , are collectively referred to as the *disorder*. We will slightly abuse the terminology and use “disorder” to refer to both the vectors  $X_i$ ,  $1 \leq i \leq M$ ; as well as the matrix  $\mathcal{M} \in \mathbb{R}^{M \times n}$  whose rows are  $X_i$ . The set  $S_\alpha^A(\kappa)$  is the *solution space*, a random subset of  $\mathcal{B}_n$ .

The computer science take on the perceptron model is to view it as an instance of a *random constraint satisfaction problem*. Indeed, observe that  $S_\alpha^A(\kappa)$  is an intersection of  $M$  random halfspaces, each defined by the *constraint vector*  $X_i$  (and threshold  $\kappa$ ). Each constraint rules out certain solutions in the space  $\mathcal{B}_n$  of all possible solutions; and the parameter  $\alpha$  plays a role akin to the constraint density in the literature on random  $k$ -SAT, see e.g. [1], [2], [18] for more discussion. For these reasons, we refer to  $\alpha$  as the *constraint density* in the sequel.

Perhaps the most important *structural* question is whether  $S_\alpha^A(\kappa)$  is empty/non-empty (w.h.p., as  $n \rightarrow \infty$ ). Krauth and Mézard conjectured in [19] that the event,  $\{S_\alpha^A(\kappa) \neq \emptyset\}$ , exhibits what is known as a *sharp threshold*: there is an explicit threshold  $\alpha_{\text{KM}}(\kappa)$  such that

$$\lim_{n \rightarrow \infty} \mathbb{P}[S_\alpha^A(\kappa) \neq \emptyset] = \begin{cases} 0, & \text{if } \alpha > \alpha_{\text{KM}}(\kappa) \\ 1, & \text{if } \alpha < \alpha_{\text{KM}}(\kappa) \end{cases}. \quad (2)$$

Using non-rigorous calculations based on the so-called *replica method*, Krauth and Mézard [19] conjecture a precise value of  $\alpha_{\text{KM}}(0)$  around 0.833. It is worth noting that this value deviates significantly from the *first moment threshold*: note that for  $\kappa = 0$ ,  $\mathbb{E}[|S_\alpha(\kappa)|] = \exp_2(n - n\alpha)$ , which is exponentially small (in  $n$ ) only for  $\alpha > 1$ .

The structure of  $S_\alpha^A(\kappa)$  and the aforementioned phase transition still (largely) remain as open problems. Even the very existence of such a sharp phase transition point remains open, though Xu [20], Nakajima-Sun [21] has shown sharpness of the threshold around a possibly  $n$ -dependent value  $\alpha_c^{(n)}(\kappa)$ , as in [22] in the setting of random CSP’s. With that in mind, we can define

$$\alpha_c^*(\kappa) = \inf \left\{ \alpha : \lim_{n \rightarrow \infty} \mathbb{P}[S_\alpha^A(0) = \emptyset] = 1 \right\}.$$

The work by Ding and Sun [23] establishes, using an elegant second-moment argument, that for every  $\alpha \leq \alpha_{\text{KM}}(0)$ ,

$$\liminf_{n \rightarrow \infty} \mathbb{P}[S_\alpha^A(0) \neq \emptyset] > 0.$$

Hence,  $\alpha_c^*(0) \geq \alpha_{\text{KM}}(0)$ . However, a matching upper bound is still missing: the best known bound is due to Kim and Roche [5, Theorem 1.2], which show  $\alpha_c^*(0) \leq 0.9963$ . More precisely, they show for any  $\epsilon < 0.0037$ ,  $\mathbb{P}[S_{1-\epsilon}^A(0) \neq \emptyset] = o(1)$ . For a similar negative result with a stronger convergence guarantee; that is a guarantee of form  $\mathbb{P}[S_{1-\delta}^A(0) \neq \emptyset] \leq \exp(-\delta n)$  for some small  $\delta > 0$  (though potentially worse than 0.0037), see Talagrand [24].

When  $S_\alpha^A(\kappa) \neq \emptyset$  (w.h.p.), a follow-up algorithmic question is whether such a satisfying  $\sigma \in \mathcal{B}_n$  can be found algorithmically (in polynomial time). Regarding such

positive results, the best known guarantee is again due to Kim and Roche. They devise in [5] an (multi-stage majority) algorithm that w.h.p. returns a solution  $\sigma \in S_\alpha^A(0)$  as long as  $\alpha < 0.005$ . (In particular, their algorithm is a constructive proof that  $S_\alpha^A(0) \neq \emptyset$  w.h.p. for  $\alpha < 0.005$ .) Later in Section III-D, we state our main result which shows that their algorithm is stable in an appropriate sense.

2) *Symmetric Binary Perceptron*: Proposed initially by Aubin, Perkins, and Zdeborová in [18]; the symmetric binary perceptron (SBP) model is our main focus in the present paper. Similar to the asymmetric case, fix a  $\kappa > 0$ ,  $\alpha > 0$ ; and set  $M = \lfloor n\alpha \rfloor$ . Let  $X_i \stackrel{d}{=} \mathcal{N}(0, I_n)$ ,  $1 \leq i \leq M$ , be i.i.d. random vectors, and consider

$$\begin{aligned} S_\alpha(\kappa) &\triangleq \bigcap_{1 \leq i \leq M} \left\{ \sigma \in \mathcal{B}_n : |\langle \sigma, X_i \rangle| \leq \kappa\sqrt{n} \right\} \\ &= \left\{ \sigma \in \mathcal{B}_n : \|\mathcal{M}\sigma\|_\infty \leq \kappa\sqrt{n} \right\}, \end{aligned} \quad (3)$$

where  $\mathcal{M} \in \mathbb{R}^{M \times n}$  with rows  $X_1, \dots, X_M$ . This model is called *symmetric* since  $\sigma \in S_\alpha(\kappa)$  iff  $-\sigma \in S_\alpha(\kappa)$ . It turns out that the symmetry makes the SBP more amenable to analysis compared to its asymmetric counterpart, while retaining the relevant conjectural structural properties nearly intact, see [3]. Though not our focus here, it is worth mentioning that this is analogous to the random  $k$ -SAT model. Its symmetric variant, NAE  $k$ -SAT, is mathematically more tractable, yet at the same time exhibits similar structural properties.

As its asymmetric counterpart, it was conjectured that the SBP also undergoes a *sharp phase transition*. More concretely, it was conjectured that there exists a  $\alpha_c(\kappa)$  such that the event,  $\{S_\alpha(\kappa) \neq \emptyset\}$ , undergoes a sharp phase transition as  $\alpha$  crosses  $\alpha_c(\kappa)$ . Notably,  $\alpha_c(\kappa)$  matches with the first moment prediction:

$$\alpha_c(\kappa) \triangleq -\frac{1}{\log_2 \mathbb{P}[|Z| \leq \kappa]}, \quad \text{where } Z \sim \mathcal{N}(0, 1). \quad (4)$$

It was established in [18] that (a)  $\lim_{n \rightarrow \infty} \mathbb{P}[S_\alpha(\kappa) \neq \emptyset] = 0$  for  $\alpha > \alpha_c(\kappa)$ ; and (b)  $\liminf_{n \rightarrow \infty} \mathbb{P}[S_\alpha(\kappa) \neq \emptyset] > 0$  for  $\alpha < \alpha_c(\kappa)$ . The latter guarantee uses the so-called *second moment method*, though falling short of establishing the high probability guarantee. Subsequent works by Perkins and Xu [1]; and Abbe, Li, and Sly [2] establish that  $\mathbb{P}[S_\alpha(\kappa) \neq \emptyset] = 1 - o(1)$  for all  $\alpha < \alpha_c(\kappa)$ . Namely,  $\alpha_c(\kappa)$  is indeed a sharp threshold for the SBP. Having established the existence and the location of such a sharp phase transition; the next question, once again, is whether such a  $\sigma \in S_\alpha(\kappa)$  can be found *efficiently*; that is, by means of polynomial-time algorithms. This is our main focus in the present paper.

The SBP is closely related to *combinatorial discrepancy theory* [25], [26]. Given a matrix  $\mathcal{M} \in \mathbb{R}^{M \times n}$ , a central problem in discrepancy theory is to compute, approximate, or bound its *discrepancy*  $\mathcal{D}(\mathcal{M})$ , where  $\mathcal{D}(\mathcal{M}) \triangleq \min_{\sigma \in \mathcal{B}_n} \|\mathcal{M}\sigma\|_\infty$ . Several different settings are considered

in the discrepancy literature: *worst-case*  $\mathcal{M}$  and *average-case*  $\mathcal{M}$  (where the entries of  $\mathcal{M}$  either i.i.d. Rademacher or i.i.d. Gaussian); and both existential and algorithmic results are sought. In the *proportional regime*, the discrepancy perspective is to fix the aspect ratio  $\alpha = M/n$  and find a solution  $\sigma$  with small  $\|\mathcal{M}\sigma\|_\infty$ . This is the inverse of the perceptron perspective: fixing  $\kappa > 0$  and finding the largest  $\alpha$  for which a solution  $\sigma$  exists. In particular, the sharp threshold result for the SBP described above settles the question of discrepancy in the random proportional regime: for  $\mathcal{M} \in \mathbb{R}^{M \times n}$  with i.i.d.  $\mathcal{N}(0, 1)$  entries,  $\mathcal{D}(\mathcal{M}) = (1+o(1))f(\alpha)\sqrt{n}$  w.h.p. where  $f(\cdot)$  is the inverse function of  $\alpha_c$ . The first and second moment methods can also be employed to establish the value of discrepancy in the random setting in other regimes, e.g. [27]–[29]. Moreover, as we describe below, discrepancy algorithms (e.g. [30]–[34]) can be employed for the SBP.

## B. Main Results

From an algorithmic point of view, the most striking fact about the SBP is the existence of a large *statistical-to-computational gap*. Explanations for both the algorithmic hardness of the model and for the success of efficient algorithms at low densities have been put forth recently.

*a) A Statistical-to-Computational Gap:* A random constraint satisfaction problem like the SBP is said to exhibit a *statistical-to-computational gap* if the density below which solutions are known to exist w.h.p. is higher than the densities at which known efficient algorithms can find a solution. As we now demonstrate, the SBP exhibits a statistical-to-computational gap for all  $\kappa > 0$ , but this gap is most pronounced in the regime of small  $\kappa$ . In this regime, the best known algorithmic guarantee for finding a solution in the SBP is due to Bansal and Spencer [33] from the literature on combinatorial discrepancy. As we detail in Section III-C and show in Corollary 2, their algorithm works for  $\alpha = O(\kappa^2)$  as  $\kappa \rightarrow 0$ . This stands in stark contrast to the threshold for the existence of solutions. From (4),  $\alpha_c(\kappa)$  behaves like  $\frac{1}{\log_2(1/\kappa)}$ :

$$\alpha_c(\kappa) = -\frac{1}{\log_2\left(\sqrt{\frac{2}{\pi}}\kappa(1+o_\kappa(1))\right)} = \frac{1+o_\kappa(1)}{\log_2(1/\kappa)}.$$

Namely,  $\alpha_c(\kappa)$  is asymptotically much larger than the algorithmic  $\kappa^2$  threshold. The main motivation of the present paper is to inquire into the origins of this gap in the SBP by leveraging insights from statistical physics. In particular, we will establish the presence of a geometric property known as the *Overlap Gap Property* (OGP), and use it to rule out classes of *stable algorithms*, appropriately defined.

*b) Freezing, rare clusters, and algorithms:* The SBP exhibits striking structural properties which are thought to contribute to both the success of polynomial-time algorithms at low densities and the failure of efficient algorithms at higher densities.

On one hand, the model exhibits the “frozen one-step Replica Symmetry Breaking (1-RSB)” scenario at all positive densities  $\alpha < \alpha_c$ . This states that whp over the instance, almost every solution  $\sigma$  is *totally frozen* and isolated: the nearest other solution is at linear Hamming distance to  $\sigma$ . This extreme form of clustering was conjectured to hold for the ABP and SBP in [3], [18], [19], [35], and subsequently established for the SBP in [1], [2]. In light of the earlier works by Mézard, Mora, and Zecchina [36] and Achlioptas and Ricci-Tersenghi [37] positing a link between clustering, freezing, and algorithmic hardness, it is tempting to postulate that finding a solution  $\sigma$  for the SBP is hard for every  $\alpha \in (0, \alpha_c(\kappa))$ , but this is contradicted by the existence of efficient algorithms at low densities such as that of [33], [38]–[41] including the algorithm by Bansal and Spencer discussed above. Combining these facts, we arrive at the conclusion that the SBP exhibits an intriguing phenomenon: polynomial-time algorithms can coexist with the frozen 1-RSB phenomenon. This conundrum challenges the view that clustering and freezing necessarily lead to algorithmic hardness.

In an attempt to explain this apparent conundrum, it was conjectured in [42] that while a  $1 - o(1)$  fraction of all solutions are totally frozen, an exponentially small fraction of solutions appear in clusters of exponential (in  $n$ ) size; and the efficient learning algorithms that manage to find solutions find solutions belonging to such rare clusters, see [1] for further discussion. In this direction, Abbe, Li, and Sly [4] established very recently that whp a connected cluster of solutions of linear diameter does indeed exist at *all* densities  $\alpha < \alpha_c$ . Furthermore, they show that an efficient multi-stage majority algorithm (based on that of [5]) can find such a large cluster at densities  $\alpha = O(\kappa^{10})$  in the  $\kappa \rightarrow 0$  regime<sup>1</sup>.

These results and conjectures prompt several questions regarding the statistical-to-computational gap exhibited by the SBP. If large connected clusters exist at all subcritical densities, what is the reason for the apparent algorithmic hardness? Do the efficient algorithms for densities  $\alpha = O(\kappa^2)$  also find solutions lying in one of these large connected clusters? At what densities are these large clusters algorithmically accessible? In particular, while we now know detailed structural information about the SBP, its statistical-to-computational gap remains a mystery.

*c) Our results on the Overlap Gap Property and failure of stable algorithms:* We investigate the statistical-to-computational gap in the SBP via the *Overlap Gap Property* (OGP), an intricate geometrical property of the solution space that has been used to rigorously rule out large classes of search algorithms for many important random computational problems including random  $k$ -SAT [43]–[45] and independent sets in sparse random graphs [46]–[48], see also the survey paper by Gamarnik [49]. We will describe the OGP in more detail below. At a high level, it

<sup>1</sup>See in particular  $\alpha_0$  appearing in [4, Page 6].

asserts the non-existence of tuples of solutions at prescribed distances in the solution space.

Our first main result establishes the OGP for  $m$ -tuples of solutions (dubbed as  $m$ -OGP) at densities  $\Omega(\kappa^2 \log_2 \frac{1}{\kappa})$ :

**Theorem 1** (Informal, see Theorem 7). *For densities  $\alpha = \Omega(\kappa^2 \log_2 \frac{1}{\kappa})$ , the SBP exhibits the  $m$ -OGP for appropriately chosen parameters.*

We also establish the presence of 2-OGP and the 3-OGP for the SBP in the high  $\kappa$  regime, i.e. when  $\kappa = 1$ , respectively in Theorem 5 and Theorem 6. Moreover, our OGP results enjoy *universality*: they remain valid under milder distributional assumptions on the entries of  $\mathcal{M}$ . We show this in the full version of our paper [50, Theorem 5.2] via the multi-dimensional Berry-Esseen Theorem.

Our next main result shows that the  $m$ -OGP rules out the class of *stable algorithms* formalized in Definition 2. At a high level, an algorithm is stable if a small perturbation of its input results in a small perturbation of the solution  $\sigma$  it outputs. In the literature on other random computational problems, it has been shown that the class of stable algorithms captures powerful classes of algorithms including Approximate Message Passing algorithms [51], low-degree polynomials [45], [52], and low-depth circuits [53].

**Theorem 2** (Informal, see Theorem 8). *The  $m$ -OGP implies the failure of stable algorithms for the SBP.*

Thus, we obtain the following corollary:

**Corollary 1** (Informal, see Theorem 8). *Stable algorithms (with appropriate parameters) fail to find a solution for the SBP for densities  $\alpha = \Omega(\kappa^2 \log_2 \frac{1}{\kappa})$ .*

In particular, this hardness result matches the algorithmic  $\kappa^2$  threshold up to a logarithmic factor. Hence, while the view that freezing implies algorithmic hardness for the SBP breaks down, the rigorous link between the OGP and algorithmic hardness remains intact.

In addition to stable algorithms; we also consider the class of *online algorithms* which includes the Bansal-Spencer algorithm [33]. Informally, an algorithm  $\mathcal{A}$  is online if the  $t^{\text{th}}$  coordinate of the solution it outputs depends only on the first  $t$  columns of  $\mathcal{M}$ .

**Theorem 3** (Informal, see Theorem 9). *Online algorithms fail to find a solution for the SBP for sufficiently high densities.*

Having established the hardness of stable algorithms for the SBP at the  $m$ -OGP threshold; a natural follow-up question is whether the known efficient algorithms for perceptron models are stable and whether the ABP also exhibits the  $m$ -OGP. To that end, we investigate the stability property of Kim-Roche algorithm [5] for the ABP.

**Theorem 4** (Informal, see Theorem 11). *Kim-Roche algorithm [5] for the ABP is stable in the sense of Definition 2.*

Investigating the stability of the Bansal-Spencer algorithm [33] and whether the ABP also exhibits the OGP are among several open questions we discuss in Section I-D.

### C. Background and Related Work

*a) Statistical-to-Computational Gaps:* As we noted, the SBP model exhibits a *statistical-to-computational gap* (SCG): a gap between what the existential results guarantee (and thus what can be found with unbounded computational power), and what algorithms with bounded computational power (such as polynomial-time algorithms) can promise. Such SCGs are a ubiquitous feature in many algorithmic problems (with *random inputs*) appearing in high-dimensional statistical inference tasks and in the study of random combinatorial structures. A partial list of problems with an SCG includes constraint satisfaction problems [36], [37], [54], optimization problems over random graphs [46], [55], [56] and spin glass models [51]–[53], [57], number partitioning problem [58], principal component analysis [59]–[61], and the “infamous” planted clique problem [62]–[64]; see also the introduction of [58], [65], the recent survey [49]; and the references therein.

Unfortunately, due to the so-called *average-case* nature of these problems, the standard NP-completeness theory often fails to establish hardness for those problems even under the assumption  $P \neq NP$ . (It is worth noting though that a notable exception to this is when the problem exhibits *random self-reducibility*, see e.g. [66] for such a hardness result regarding a spin glass model, conditional on a weaker assumption  $P \neq \#P$ .) Nevertheless, a very fruitful (and still active) line of research proposed certain forms of *rigorous evidences* of algorithmic hardness for such average-case problems. These approaches include the failure of Monte Carlo Markov Chain methods [62], [67], low-degree methods and failure of low-degree polynomials [45], [48], [52], [68], [69], Sum-of-Squares [64], [70]–[72] and Statistical Query [73]–[76] lower bounds, failure of the approximate message passing algorithm (an algorithm that is information-theoretically optimal for certain important problems, see e.g. [77], [78]) [79], [80]; and the reductions from the planted clique problem [59], [81], [82], just to name a few. Yet another very promising such approach is through the *intricate geometry* of the problem, via the so-called *Overlap Gap Property* (OGP).

*b) Overlap Gap Property (OGP):* Implicitly discovered by Mézard, Mora, and Zecchina [36] and Achlioptas and Ricci-Tersenghi [37] (though coined later in [83]), the OGP approach leverages insights from the statistical physics to form a rigorous link between the intricate geometry of the solution space and formal algorithmic hardness. Informally, the OGP is a topological disconnectivity property, and states (in the context of a random combinatorial optimization problem, say over  $\mathcal{B}_n$ ) that (w.h.p. over the randomness) any two near-optimal  $\sigma_1, \sigma_2 \in \mathcal{B}_n$  are either “close” or “far” from each other: there exists  $0 < \nu_1 < \nu_2 < 1$  such that  $n^{-1} \langle \sigma_1, \sigma_2 \rangle \in [0, \nu_1] \cup [\nu_2, 1]$ . That

is, their (normalized) overlaps do not admit *intermediate values*; and no two near-optimal solutions of intermediate distance can be found. It has been shown (see below) that the OGP is a rigorous barrier for large classes of algorithms. See [49] for a survey on OGP.

c) *Algorithmic Implications of OGP*: The line of research relating the OGP to algorithmic hardness was initiated by Gamarnik and Sudan [46], [56]. They consider the problem of finding a large independent set in the sparse random graphs with average degree  $d$ . It is known, see e.g. [84]–[86], that in the double limit (first sending  $n \rightarrow \infty$ , then letting  $d \rightarrow \infty$ ), the largest independent set of this graph is of size  $2^{\frac{\log d}{d}}n$ . On the other hand, the best known polynomial-time algorithm [87] (a very simple greedy protocol) returns an independent set that is half optimal, namely of size  $\frac{\log d}{d}n$ . In order to reconcile this apparent SCG, Gamarnik and Sudan study the space of all large independent sets. They establish that any two independent sets of size greater than  $(1 + 1/\sqrt{2})\frac{\log d}{d}n$  exhibit OGP. By leveraging this, they show, through a contradiction argument, that *local algorithms* (known as the *factors of i.i.d.*) fail to find an independent set of size greater than  $(1 + 1/\sqrt{2})\frac{\log d}{d}n$ . Subsequent research (again via the lens of OGP) extended this hardness result to the class of low-degree polynomials [52]. The extra “oversampling” factor,  $1/\sqrt{2}$ , was removed by inspecting instead the the overlap pattern of many large independent sets (rather than the pairs), therefore establishing hardness all the way down to the algorithmic threshold. This was done by Rahman and Virág [47] for *local algorithms*, and by Wein [48] for *low-degree polynomials*; and is also our focus here (see below). A list of problems where the OGP is leveraged to rule out certain classes of algorithms includes optimization over random graphs and spin glass models [51]–[53], [88], number partitioning problem [58], random constraint satisfaction problems [43], [45].

d) *Multi OGP ( $m$ –OGP)*: As we just mentioned, it was previously observed that by considering more intricate overlap patterns, one can potentially lower the (algorithmic) phase transition points further. This idea was employed for the first time by Rahman and Virág [47] in the context of the aforementioned independent set problem. They managed to “shave off” the extra  $1/\sqrt{2}$  factor present in the earlier result by Gamarnik and Sudan [46], [56], and reached all the way down to the algorithmic threshold,  $\frac{\log d}{d}n$ . In a similar vein, Gamarnik and Sudan [43] studied the overlap structure of  $m$ –tuples  $\sigma^{(i)} \in \mathcal{B}_n$ ,  $1 \leq i \leq m$  of satisfying assignments in the context of the Not-All-Equal (NAE)  $k$ –SAT problem. By showing the presence of OGP for  $m$ –tuples of nearly equidistant points (in  $\mathcal{B}_n$ ), they established nearly tight hardness for *sequential local algorithms*: their results match the computational threshold modulo factors that are polylogarithmic (in  $k$ ). A similar overlap pattern (for  $m$ –tuples consisting of nearly equidistant points) was also considered by Gamarnik and Kızıldağ [58] in the context of random number partitioning

problem (NPP), where they established hardness well below the existential threshold. (It is worth noting that [58] considers  $m$ –tuples where  $m$  itself also grows in  $n$ ,  $m = \omega_n(1)$ .)

More recently,  $m$ –OGP for more intricate forbidden patterns were considered to establish formal hardness in other settings. In particular, by leveraging  $m$ –OGP, Wein [48] showed that low-degree polynomials fail to return a large independent set (in sparse random graphs) of size greater than  $\frac{\log d}{d}n$ , thereby strengthening the earlier result by Gamarnik, Jagannath, and Wein [52]. Wein’s work establishes the *ensemble* variant of OGP (an idea emerged originally in [57]): he considers  $m$ –tuples of independent sets where each set do not necessarily come from the same random graph, but rather from correlated random graphs. The ensemble variant of OGP was also considered in [58] for the NPP. While technically more involved to establish, it appears that the ensemble  $m$ –OGP can be leveraged to rule out virtually any *stable algorithm* (appropriately defined); and will also be our focus here. More recently, by leveraging the ensemble  $m$ –OGP; Bresler and Huang [45] established nearly tight low-degree hardness results for the random  $k$ –SAT problem: they show that low-degree polynomials fail to return a satisfying assignment when the clause density is only a constant factor off by the computational threshold. In yet another work, Huang and Sellke [88] construct a very intricate forbidden structure consisting of an ultrametric tree of solutions, which they refer to as the *branching OGP*. By leveraging this branching OGP, they rule out overlap concentrated algorithms<sup>2</sup> at the algorithmic threshold for the problem of optimizing mixed, even  $p$ –spin model Hamiltonian.

#### D. Open Problems

a) *Location of the Algorithmic Threshold*: We establish in Theorem 7 that the SBP exhibits  $m$ –OGP if  $\alpha = \Omega(\kappa^2 \log_2 \frac{1}{\kappa})$ . On the other hand, we have per Corollary 2 that the Bansal-Spencer algorithm [33] works when  $\alpha = O(\kappa^2)$ . In light of these, we make the following conjecture:

**Conjecture 1.** *As  $\kappa \rightarrow 0$ , the algorithmic threshold for the SBP is at  $\tilde{\Theta}(\kappa^2)$ .*

In particular, we conjecture that up to factors that are polylogarithmic in  $\frac{1}{\kappa}$ , the Bansal-Spencer algorithm is the best possible within the class of efficient algorithms. That is, up to polylogarithmic factors no polynomial-time algorithm succeeds above the  $m$ –OGP threshold. An interesting question is whether the  $\log_2 \frac{1}{\kappa}$  factor is necessary or it can be ‘shaved off’. We believe it might be possible to remove this factor by considering a more intricate overlap pattern, e.g. similar to those considered in [45], [48], [88].

We now make Conjecture 1 more precise. Given an  $m \in \mathbb{N}$  and  $\kappa > 0$ , let  $\alpha_m^*(\kappa)$  be the smallest subcritical

<sup>2</sup>A class that captures  $O(1)$  iterations of gradient descent, approximate message passing; and Langevin Dynamics run for  $O(1)$  time.

density such that the SBP exhibits  $m$ -OGP with appropriate parameters when  $\alpha \geq \alpha_m^*$ . We conjecture that the  $m \rightarrow \infty$  limit of the  $m$ -OGP threshold marks the true algorithmic threshold: for every  $\epsilon > 0$  and  $\kappa$  small enough, there do not exist polynomial-time algorithms for the SBP when  $\alpha \geq (1 + \epsilon) \lim_{m \rightarrow \infty} \alpha_m^*(\kappa)$ . See Conjecture 3 for details. This conjecture is backed up by the evidence that for many random computational problems including random  $k$ -SAT [45], independent sets in sparse random graphs [47], [48], and mixed even  $p$ -spin model [88], the  $m$ -OGP matches or nearly matches the best known algorithmic threshold.

Abbe, Li and Sly ask in [4, Question 1] whether the algorithmic threshold for the SBP coincides with the threshold for the existence of a ‘wide web’: a cluster of solutions with maximum possible diameter  $n$ . One one hand, the existence of a wide web rules out the 2-OGP: pairs of solutions of every possible overlap exist. It would be very interesting to determine whether the threshold for existence of the wide web coincides with the conjectured algorithmic threshold of  $\tilde{\Theta}(\kappa^2)$  above, or even more precisely the limiting  $m$ -OGP threshold  $\lim_{m \rightarrow \infty} \alpha_m^*$  (at least asymptotically as  $\kappa \rightarrow 0$ ).

*b) The Asymmetric Model:* As we noted earlier, the ABP is more challenging from a mathematical perspective, and some of its basic properties are still far from being rigorously understood. In particular, even the very existence of a sharp phase transition and the frozen 1-RSB picture—both rigorously known to hold for the SBP—remain open.

The ABP also exhibits a statistical-to-computational gap. On one hand, Kim-Roche algorithm [5] finds solutions at low enough densities, specifically when  $\alpha < 0.005$ . On the other hand, the result of Ding-Sun [23] shows that solutions do exist (with probability bounded away from 0) when  $\alpha < \alpha_{\text{KM}}(0) \approx 0.83$ . It would be interesting to show that the ABP exhibits  $m$ -OGP for some  $\alpha < \alpha_{\text{KM}}(0)$ . To understand the statistical-to-computational gap of ABP further, it would be interesting to explore the model in the regime  $\kappa \rightarrow \infty$  and investigate the  $m$ -OGP threshold and threshold for the existence of efficient algorithms. Further, there are other perceptron models one could explore, e.g. the *U-function binary perceptron* introduced in [18].

*c) Stability of Other Algorithms:* We established in Theorem 11 that the Kim-Roche algorithm for ABP is stable. In light of this, we make the following conjecture regarding SBP:

**Conjecture 2.** *There exists a stable algorithm that finds a solution for the SBP w.h.p. when  $\alpha = O(\kappa^2)$ .*

In particular, proving stability of the Bansal-Spencer algorithm would resolve Conjecture 2, but this seems challenging: the presence of a certain non-linear potential function (see [33, Equation 2.5]) renders the stability analysis difficult.

The algorithm of [4] is a variant of the Kim-Roche algorithm that works for the SBP for  $\alpha = O(\kappa^{10})$ . Proving

the stability of this algorithm would be an interesting first step towards resolving Conjecture 2.

*d) Broader Research Agendas on the OGP:* As mentioned above, the OGP is a provable barrier for a broad class of algorithms for many random computational problems. A list of such algorithms includes local/sequential local algorithms, Monte Carlo Markov Chain (MCMC) methods, low-degree polynomials, Langevin dynamics, approximate message passing type algorithms, low-depth circuits, and stable algorithms in general. In many random computational problems (like  $k$ -SAT and independent sets) the OGP coincides with the threshold for the existence of known efficient search algorithms. One might then conjecture (as we do here) that the OGP marks the true algorithmic threshold. It would thus be very surprising and very interesting to find a case where efficient algorithms succeed in the face of the OGP<sup>3</sup>. While random  $k$ -SAT, independent sets in random graphs, and other random CSP’s have been studied for decades without finding such algorithms, algorithms for perceptron models have not been studied as extensively, especially not in the limiting regime  $\kappa \rightarrow 0$  we focus on here, and thus this might be fruitful direction to pursue.

## E. Organization and Notation

*a) Paper Organization:* The rest of the paper is organized as follows. We state our OGP results in Section II. We state our main algorithmic hardness result in Section III-A and formulate a conjecture pertaining the true algorithmic threshold in Section III-C. Finally, we show that the Kim-Roche algorithm is stable in Section III-D. We defer complete proofs and detailed discussions to the full version [50].

*b) Notation:* For any  $n \in \mathbb{N}$ ,  $[n] \triangleq \{1, 2, \dots, n\}$  and  $\mathcal{B}_n \triangleq \{-1, 1\}^n$ . For any  $r > 0$  and  $x \in \mathbb{R}$ ,  $\exp_r(x)$  and  $\log_r(x)$  denote respectively the exponential and logarithm functions base  $r$ . For any  $v, v' \in \mathbb{R}^n$ ,  $\|v\|_\infty = \max_{1 \leq i \leq n} |v_i|$ ,  $\langle v, v' \rangle \triangleq \sum_{1 \leq i \leq n} v_i v'_i$  and  $\mathcal{O}(v, v') \triangleq n^{-1} \langle v, v' \rangle$ . For any  $\sigma, \sigma' \in \mathcal{B}_n$ ,  $d_H(\sigma, \sigma')$  denotes their Hamming distance. We denote the standard normal distribution by  $\mathcal{N}(0, 1)$  and the multivariate normal distribution with mean  $\mu \in \mathbb{R}^n$  and covariance  $\Sigma \in \mathbb{R}^{n \times n}$  by  $\mathcal{N}(\mu, \Sigma)$ . Throughout the paper, we employ the standard asymptotic notation, e.g.  $\Theta(\cdot), O(\cdot), o(\cdot), \Omega(\cdot)$  and  $\tilde{\Theta}(\cdot)$ . If there is no subscript, the asymptotic is with respect to  $n \rightarrow \infty$ . In the case where we consider asymptotics other than  $n \rightarrow \infty$ , we reflect this by a subscript: for instance, if  $f$  is such that  $f(\kappa) \rightarrow \infty$  as  $\kappa \rightarrow 0$ , we write  $f = \omega_\kappa(1)$ .

## II. OGP IN THE SYMMETRIC BINARY PERCEPTRON

Next we establish landscape results, dubbed as *ensemble  $m$ -OGP*, concerning overlap structures of  $m$ -tuples  $(\sigma^{(i)} \in \mathcal{B}_n : 1 \leq i \leq m)$ , that satisfy ‘box constraints’ with respect to correlated instances of Gaussian disorder.

<sup>3</sup>Beyond those cases with a ‘rigid global structure’ such as solving linear equations and independent sets in random bipartite graphs. For instance, algebraic techniques like Gaussian elimination and lattice based methods can find solutions to ‘noiseless’ problems such as solving random linear equations or noiseless clustering [89].

## A. Technical Preliminaries

We next formalize the notion of correlated instances through an appropriate interpolation scheme.

**Definition 1.** Fix a  $\kappa > 0$ , and recall  $\alpha_c(\kappa)$  from (4). Let  $0 < \alpha < \alpha_c(\kappa)$ ,  $m \in \mathbb{N}$ ,  $0 < \eta < \beta < 1$ , and  $\mathcal{I} \subset [0, \pi/2]$ . Set  $M = \lfloor n\alpha \rfloor$  and suppose that  $\mathcal{M}_i \in \mathbb{R}^{M \times n}$ ,  $0 \leq i \leq m$ , is a sequence of i.i.d. random matrices, each having i.i.d.  $\mathcal{N}(0, 1)$  coordinates. Denote by  $\mathcal{S}_\kappa(\beta, \eta, m, \alpha, \mathcal{I})$  the set of all  $m$ -tuples  $(\sigma^{(i)} : 1 \leq i \leq m)$ ,  $\sigma^{(i)} \in \mathcal{B}_n$ , satisfying the following conditions.

(a) **(Pairwise Overlap Condition)** For  $1 \leq i < j \leq m$ ,

$$\beta - \eta \leq \mathcal{O}(\sigma^{(i)}, \sigma^{(j)}) \leq \beta.$$

(b) **(Rectangular Constraints)** There exists  $\tau_i \in \mathcal{I}$ ,  $1 \leq i \leq m$ , such that

$$\|\mathcal{M}_i(\tau_i)\sigma^{(i)}\|_\infty \leq \kappa\sqrt{n}, \quad 1 \leq i \leq m,$$

where  $\mathcal{M}_i(\tau_i) = \cos(\tau_i)\mathcal{M}_0 + \sin(\tau_i)\mathcal{M}_i$ .

The interpretations of the parameters appearing in Definition 1 are as follows. The parameter  $m$  is the size of the tuples we inspect;  $\kappa$  is the constraint threshold; and  $\alpha$  is the constraint density. That is, we consider  $M = \lfloor n\alpha \rfloor$  random constraints. Parameters  $\beta$  and  $\eta$  control the (forbidden) region of pairwise overlaps. Finally, the index set,  $\mathcal{I}$ , is used for generating correlated instances of random constraints via interpolation  $\mathcal{M}_i(\tau_i)$ ,  $\tau_i \in \mathcal{I}$ , defined earlier. This is necessary to study the ensemble OGP, see below.

As a concrete example to Definition 1, consider the toy setting  $m = 2$  and  $\mathcal{I} = \{0\}$ . In this case,  $\mathcal{S}_\kappa(\beta, \eta, 2, \alpha, \{0\})$  is simply the set of all pairs  $(\sigma_1, \sigma_2) \in \mathcal{B}_n \times \mathcal{B}_n$  such that (a)  $\beta - \eta \leq n^{-1} \langle \sigma_1, \sigma_2 \rangle \leq \beta$  and (b)  $\|\mathcal{M}\sigma_i\|_\infty \leq \kappa\sqrt{n}$  for  $i = 1, 2$ ; where  $\mathcal{M} \in \mathbb{R}^{\lfloor \alpha n \rfloor \times n}$  is a random matrix with i.i.d. standard normal entries.

## B. Landscape Results: High $\kappa$ Regime

Our first focus is on the regime where  $\kappa$  is *large*. While we set  $\kappa = 1$  (thus  $\alpha_c(\kappa)$  is approximately 1.8159) for simplicity; our results easily extend to any fixed  $\kappa > 0$ . In this case, we also drop the subscript  $\kappa$  appearing in Definition 1, and simply use the notation  $\mathcal{S}(\beta, \eta, m, \alpha, \mathcal{I})$  to denote  $\mathcal{S}_1(\beta, \eta, m, \alpha, \mathcal{I})$ .

Our first result establishes 2-OGP above  $\alpha \geq 1.71$ .

**Theorem 5.** Let  $1.71 \leq \alpha \leq \alpha_c(1) \approx 1.8159$ . Then, there exists  $0 < \eta_2^* < \beta_2^* < 1$  and a constant  $c_2^* > 0$  such that the following holds. Fix any  $\mathcal{I} \subset [0, \pi/2]$  with  $|\mathcal{I}| \leq \exp_2(c_2^* n)$ . Then,

$$\mathbb{P}[\mathcal{S}(\beta_2^*, \eta_2^*, 2, \alpha, \mathcal{I}) \neq \emptyset] \leq \exp_2(-\Theta(n)).$$

By considering the overlap structure of triples, one can further reduce the threshold (on  $\alpha$ ) to approximately 1.667 above which the overlap gap property takes place.

**Theorem 6.** Let  $1.667 \leq \alpha \leq \alpha_c(1) \approx 1.8159$ . Then, there exists  $0 < \eta_3^* < \beta_3^* < 1$  and a constant  $c_3^* > 0$

such that the following holds. Fix any  $\mathcal{I} \subset [0, \pi/2]$  with  $|\mathcal{I}| \leq \exp_2(c_3^* n)$ . Then,

$$\mathbb{P}[\mathcal{S}(\beta_3^*, \eta_3^*, 3, \alpha, \mathcal{I}) \neq \emptyset] \leq \exp_2(-\Theta(n)).$$

The proof is deferred to the full version [50].

Theorem 6 implies that 3-OGP (with appropriate parameters) takes place for  $\alpha \geq 1.667$ , which is indeed strictly smaller than the corresponding threshold of  $\alpha \geq 1.71$  for 2-OGP per Theorem 5. An inspection of the proof reveals that our choice of  $\eta^*$  satisfies  $\eta^* \ll \beta^*$ . That is, the structure that Theorem 6 rules out corresponds essentially to (nearly) equilateral triangles in Hamming space.

Theorem 6 is established using the *first moment method*. More specifically, let the random variable  $N$  count the number of such triples. We show that  $\mathbb{E}[N] = \exp(-\Theta(n))$  under an appropriate choice of parameters and conclude by Markov's inequality since  $\mathbb{P}[N \geq 1] \leq \mathbb{E}[N] = o(1)$ . It is worth noting though that unlike [58], our counting bound is exact (up to lower-order terms). This appears necessary to improve upon Theorem 5, see [50] for more details.

As noted earlier, we do not pursue the  $m$ -OGP improvement for  $m \geq 4$  in the high  $\kappa$  regime since the first moment method actually fails as  $m$  gets larger. This, of course, is only a failure of the first moment method; it does not necessarily imply that the  $m$ -OGP itself yields a worse threshold. In fact, given the prior work as well as the fact that  $m$ -OGP deals with a more nested structure, it indeed makes sense that  $m$ -OGP (for  $m \geq 4$ ) should hold for a much broader range of  $\alpha$ .<sup>4</sup> For this reason, it is plausible to conjecture that considering  $m$ -OGP beyond  $m \in \{2, 3\}$  lowers the threshold on  $\alpha$ . We leave the formal verification of this for future investigation.

Finally, we remark that Baldassi et al. established in [3] similar OGP results for the high  $\kappa$  case, see [50] for details.

## C. Landscape Results: The Regime $\kappa \rightarrow 0$

We now turn to our results in the regime  $\kappa \rightarrow 0$ . Observe that for any fixed  $\kappa > 0$ , the volume of the “rectangular box”  $[-\kappa, \kappa]^m$  (which eventually controls the probabilistic term) appearing in Definition 1 is  $(2\kappa)^m$ . When  $\kappa \rightarrow 0$ , this term actually shrinks further by increasing  $m$ . Thus, one can hope to pursue the  $m$ -OGP improvement. Our main result to that end is as follows.

**Theorem 7.** Let

$$\alpha_{\text{OGP}}(\kappa) \triangleq 10\kappa^2 \log \frac{1}{\kappa}. \quad (5)$$

Then, for every sufficiently small  $\kappa > 0$  and  $\alpha \geq \alpha_{\text{OGP}}(\kappa)$ , there exist  $0 < \eta < \beta < 1$ ,  $c > 0$ , and an  $m \in \mathbb{N}$  such that the following holds. Fix any  $\mathcal{I} \subset [0, \pi/2]$  with  $|\mathcal{I}| \leq \exp_2(cn)$ . Then,

$$\mathbb{P}[\mathcal{S}_\kappa(\beta, \eta, m, \alpha, \mathcal{I}) \neq \emptyset] \leq \exp_2(-\Theta(n)).$$

<sup>4</sup>Such a strict monotonicity in  $m$  has also been conjectured by Ben Arous and Jagannath in the context of spherical spin glass models [90].

The proof is deferred to the full version [50].

Recall from our earlier discussion (also see Section III-C and Corollary 2 therein) that the algorithm by Bansal and Spencer [33] works for  $\alpha = O(\kappa^2)$ . On the other hand, no (efficient) algorithm is known for  $\alpha = \Omega(\kappa^2)$ . Namely, the current known algorithmic threshold for the symmetric binary perceptron model is  $\Theta(\kappa^2)$ . In light of these facts, Theorem 7 shows that the OGP threshold  $\alpha_{\text{OGP}}(\kappa)$  is *nearly matching*: the onset of OGP coincides up to logarithmic (in  $\kappa$ ) factors with the threshold above which no polynomial-time algorithms are known to work.

We now comment on the “extra”  $\log_2 \frac{1}{\kappa}$  factor appearing in (5). As we detail in [50, Section 4], the exponent of the first moment of the cardinality term,  $|S_\kappa(\beta, \eta, m, \alpha, \mathcal{I})|$ , appears to be strictly positive (for every  $\beta, \eta, m$ ) if  $\alpha = O(\kappa^2 \log_2(1/\kappa))$ . That is, Theorem 7 is in a sense the best possible using our techniques. However, it is plausible that by considering a more delicate forbidden structure (akin to those studied in [45], [48], [88]), one may be able to remove this logarithmic factor. This suggests two conjectures: (a) in the regime  $\kappa \rightarrow 0$ , the algorithm by Bansal and Spencer [33] is best possible (up to constant factors); and that (b) the OGP marks the onset of algorithmic hardness.

### III. ALGORITHMIC BARRIERS FOR THE PERCEPTRON MODEL

#### A. $m$ —Overlap Gap Property Implies Failure of Stable Algorithms

We commence this section by recalling our setup. Fix a  $\kappa > 0$ , and an  $\alpha < \alpha_c(\kappa)$  so that w.h.p. as  $n \rightarrow \infty$ , there exists a  $\sigma \in S_\alpha(\kappa)$ , where  $S_\alpha(\kappa)$  is the (random) set in (3). Having ensured that  $S_\alpha(\kappa)$  is (w.h.p.) non-empty; our focus in this section is on the problem of finding such a  $\sigma$  by using *stable algorithms*, formalized below.

*a) Algorithmic Setting:* We interpret an algorithm  $\mathcal{A}$  as a mapping from  $\mathbb{R}^{M \times n}$  to  $\mathcal{B}_n$ . We allow  $\mathcal{A}$  to be potentially randomized: we assume there exists an underlying probability space  $(\Omega, \mathbb{P}_\omega)$  such that  $\mathcal{A} : \mathbb{R}^{M \times n} \times \Omega \rightarrow \mathcal{B}_n$ . That is, for any  $\omega \in \Omega$  and disorder matrix  $\mathcal{M} \in \mathbb{R}^{M \times n}$ ;  $\mathcal{A}(\cdot, \omega)$  returns a  $\sigma_{\text{ALG}} \triangleq \mathcal{A}(\mathcal{M}, \omega) \in \mathcal{B}_n$ ; and we want  $\sigma_{\text{ALG}}$  to satisfy  $\|\mathcal{M}\sigma_{\text{ALG}}\|_\infty \leq \kappa\sqrt{n}$ .

We now formalize the class of stable algorithms that we investigate in the present paper.

**Definition 2.** Fix a  $\kappa > 0$ , an  $\alpha < \alpha_c(\kappa)$ ; and set  $M = \lfloor n\alpha \rfloor$ . An algorithm  $\mathcal{A} : \mathbb{R}^{M \times n} \times \Omega \rightarrow \mathcal{B}_n$  is called  $(\rho, p_f, p_{\text{st}}, f, L)$ —stable for the SBP model, if it satisfies the following for all sufficiently large  $n$ .

- **(Success)** Let  $\mathcal{M} \in \mathbb{R}^{M \times n}$  be a disorder matrix with i.i.d  $\mathcal{N}(0, 1)$  entries. Then,

$$\mathbb{P}_{(\mathcal{M}, \omega)} \left[ \|\mathcal{M}\mathcal{A}(\mathcal{M}, \omega)\|_\infty \leq \kappa\sqrt{n} \right] \geq 1 - p_f.$$

- **(Stability)** Let  $\mathcal{M}, \overline{\mathcal{M}} \in \mathbb{R}^{M \times n}$  have i.i.d.  $\mathcal{N}(0, 1)$  coordinates such that  $\mathbb{E}[\mathcal{M}_{ij}\overline{\mathcal{M}}_{ij}] = \rho$  for  $1 \leq i \leq M$  and  $1 \leq j \leq n$ . Then,

$$\begin{aligned} \mathbb{P}_{(\mathcal{M}, \overline{\mathcal{M}}, \omega)} \left[ d_H(\mathcal{A}(\mathcal{M}, \omega), \mathcal{A}(\overline{\mathcal{M}}, \omega)) \leq f + L\|\mathcal{M} - \overline{\mathcal{M}}\|_F \right] \geq 1 - p_{\text{st}}. \end{aligned}$$

Definition 2 is similar to the notion of stability considered in [58, Definition 3.1]. Moreover, Definition 2 applies also to deterministic algorithms where the probabilities are taken w.r.t.  $\mathcal{M}$  and  $\overline{\mathcal{M}}$  only. In the remainder of the paper, we often abuse the notation by dropping  $\omega$  and simply referring to  $\mathcal{A} : \mathbb{R}^{M \times n} \rightarrow \mathcal{B}_n$  as a randomized algorithm.

We next highlight the operational parameters appearing in Definition 2.  $\kappa$  is the “width” of the “rectangles” defined by the constraints.  $\alpha$  is the constraint density (also known as the aspect ratio). That is,  $M = \lfloor n\alpha \rfloor$  is the number of constraints. The parameter  $p_f$  controls the success guarantee. The parameters  $\rho, p_{\text{st}}, f$  and  $L$  collectively control the stability guarantee. The parameter  $\rho$  essentially controls the amount of correlation. Stability parameters  $p_{\text{st}}, f$  and  $L$  describe the amount of sensitivity of the algorithm’s output to the correlation values. Our stability guarantee is probabilistic, where the probability is taken with respect to the joint randomness in  $\mathcal{M}, \overline{\mathcal{M}}$  as well as to the coin flips  $\omega$  of  $\mathcal{A}$ . The “extra room” of  $f$  bits makes our negative result only stronger: even when  $\mathcal{M}$  and  $\overline{\mathcal{M}}$  are very close, the algorithm is still allowed to make roughly  $f$  flips.

We now state our next main result.

**Theorem 8.** Fix any sufficiently small  $\kappa > 0$ ,  $\alpha \geq \alpha_{\text{OGP}}(\kappa) = 10\kappa^2 \log_2 \frac{1}{\kappa}$ , and  $L > 0$ . Let  $m \in \mathbb{N}$  and  $0 < \eta < \beta < 1$  be the  $m$ —OGP parameters prescribed by Theorem 7. Set

$$C = \frac{\eta^2}{1600}, \quad Q \triangleq \frac{4800L\pi}{\eta^2} \sqrt{\alpha}, \quad T = \exp_2 \left( 2^{4mQ \log_2 Q} \right). \quad (6)$$

Then, there exists an  $n_0 \in \mathbb{N}$  such that the following holds. For every  $n \geq n_0$ , there exists no randomized algorithm  $\mathcal{A} : \mathbb{R}^{M \times n} \rightarrow \mathcal{B}_n$  that is

$$\left( \cos \left( \frac{\pi}{2Q} \right), \frac{1}{9(Q+1)T}, \frac{1}{9Q(T+1)}, Cn, L \right) - \text{stable}$$

for the SBP, in the sense of Definition 2.

We defer the proof to full version [50]. Several brief remarks are now in order, see [50, Page 17] for more details.

Firstly, observe that there is no restriction on the running time of  $\mathcal{A}$ : as long as it is stable in the sense of Definition 2 with appropriate parameters, Theorem 8 applies. Secondly, as  $\alpha, L, m$  and  $\eta$  are all  $O(1)$ ,  $p_f$  and  $p_{\text{st}}$  are of constant order: the algorithms that we rule out have a *constant probability* of success/stability and need not have a high probability guarantee. Thirdly,  $\mathcal{A}$  is still allowed to make  $\Theta(n)$  flips even when  $\mathcal{M}$  and  $\overline{\mathcal{M}}$  are “nearly identical”. Lastly, while we establish Theorem 8 for the case  $L = O(1)$  for simplicity; our arguments extend to  $L = O(\frac{\log n}{\log \log n})$ .

### B. Failure of Online Algorithms for SBP

Our next focus is on the class of *online algorithms*, formalized below.

**Definition 3.** Fix a  $\kappa > 0$ , an  $\alpha < \alpha_c(\kappa)$ ; and set  $M = \lfloor n\alpha \rfloor \in \mathbb{N}$ . Let  $\mathcal{M} \in \mathbb{R}^{M \times n}$  with columns  $\mathcal{C}_1, \mathcal{C}_2, \dots, \mathcal{C}_n \in \mathbb{R}^M$ , and  $\mathcal{A} : \mathbb{R}^{M \times n} \rightarrow \mathcal{B}_n$  be an algorithm where

$$\mathcal{A}(\mathcal{M}) = \sigma = (\sigma_1, \sigma_2, \dots, \sigma_n) \in \mathcal{B}_n.$$

We call  $\mathcal{A}$   $p_f$ -online if the following hold.

- **(Success)** For  $\mathcal{M}$  consisting of i.i.d.  $\mathcal{N}(0, 1)$  entries,

$$\mathbb{P}[\|\mathcal{M}\mathcal{A}(\mathcal{M})\|_\infty \leq \kappa\sqrt{n}] \geq 1 - p_f.$$

- **(Online)** There exists deterministic functions  $f_t$ ,  $1 \leq t \leq n$  such that

$$\sigma_t = f_t(\mathcal{C}_i : 1 \leq i \leq t) \in \{-1, 1\} \quad \text{for } 1 \leq t \leq n.$$

Several remarks are now in order. The parameter  $p_f$  is the failure probability of  $\mathcal{A}$ :  $\mathcal{A}(\mathcal{M}) \in \mathcal{S}_\alpha(\kappa)$  w.p. at least  $1 - p_f$ . The second condition states that for all  $1 \leq t \leq n$ ,  $\sigma_t$  is a function of  $\mathcal{C}_1, \dots, \mathcal{C}_t$  only. More precisely, the signs  $\sigma_i \in \{-1, 1\}$ ,  $1 \leq i \leq t - 1$ , have been assigned at the end of round  $t - 1$ . A new column  $\mathcal{C}_t \in \mathbb{R}^M$  arrives in the beginning of round  $t$ , and  $\mathcal{A}$  assigns a  $\sigma_t \in \{-1, 1\}$  depending only on the previous decisions. This highlights the online nature of  $\mathcal{A}$ .

Definition 3 is an abstraction that captures, in particular, the algorithm by Bansal and Spencer [33]. Our next result establishes that online algorithms fail to return a  $\sigma \in \mathcal{S}_\alpha(\kappa)$  for densities  $\alpha$  close to the critical threshold  $\alpha_c(\kappa)$ . As in Section II-B, we stick to the case  $\kappa = 1$  for simplicity, even though our argument easily extends to arbitrary  $\kappa > 0$ .

**Theorem 9.** Let  $1.77 \leq \alpha \leq \alpha_c(1) \approx 1.8159$ . Then, there exists a constant  $c_f > 0$  such that the following holds. For any  $p_f < 1 - \exp(-c_f n)$ , there exists no  $\mathcal{A}$  for SBP which is  $p_f$ -online in the sense of Definition 3.

The proof is based on a contradiction argument (slightly different than 2-OGP) and is deferred to full version [50].

Note that the lower bound  $\alpha \geq 1.77$  appearing in Theorem 9 is strictly larger than the corresponding 2-OGP threshold, i.e.  $\alpha \geq 1.71$ , for the same setting ( $\kappa = 1$ ) per Theorem 5. Moreover, we rule out online algorithms that succeed even with an exponentially small probability. This is based on a clever application of Jensen's inequality, originally due to Gamarnik and Sudan [43, Lemma 5.3].

### C. Algorithmic Threshold in SBP

a) *Algorithmic Lower Bound in SBP*: Heretofore, we used  $\Theta(\kappa^2)$  as our baseline for the current computational threshold for the SBP. Namely, against this threshold; we (a) formulated the aforementioned *statistical-to-computational gap* and (b) compared our hardness result, Theorem 8, for the stable algorithms established through the  $m$ -OGP. In this section, we justify this choice for the algorithmic threshold, from the lower bound perspective.

As we mentioned in the introduction, the SBP is closely related to the well-known problem of minimizing the discrepancy of a matrix (or set system). The discrepancy minimization problem received much attention in the field of combinatorics and theoretical computer science; several efficient algorithms have been devised for it, see e.g. [33], [91]–[93]. In what follows, we use the algorithm of [33] as our baseline for postulating a computational threshold on  $\alpha$  as one varies  $\kappa$ ; though several of the algorithms cited above essentially yield the same  $\Theta(\kappa^2)$  guarantee modulo different absolute constants. Before we proceed with the result of [33]; it is worth noting that there is yet another complementary line of research focusing on the so-called online guarantees, see e.g. [94]–[97]. However, all of these algorithms suffer from extra polylogarithmic factors: their implied guarantees on  $\alpha$  are poorer. That is they provably work only for  $\alpha$  asymptotically much smaller than  $\kappa^2$ .

The work by Bansal and Spencer (see in particular [33, Section 3.3]) establishes the following.

**Theorem 10.** [33, Theorem 3.4] Let  $T \in \mathbb{N}$  be an arbitrary time horizon, and  $v_i \sim \text{Unif}(\mathcal{B}_M)$ ,  $1 \leq i \leq T$ , be i.i.d. random vectors. Then there exists a value  $K > 0$  and an algorithm that returns signs  $s_1, \dots, s_T \in \{-1, 1\}$  in  $\text{Poly}(M, T)$  time such that

$$\mathbb{P}[\|\sum_{i \leq T} s_i v_i\|_\infty \leq K\sqrt{M}] \geq 1 - \exp(-cM).$$

Here,  $c, K > 0$  are absolute constants independent of  $M, T$ .

**Corollary 2.** There exists an absolute constant  $K > 0$  such that the following holds. Fix any  $\kappa > 0$ ,  $\alpha < (\kappa/K)^2$ ; and consider the matrix  $\mathcal{M} \in \mathbb{R}^{\alpha n \times n}$  with i.i.d. entries  $\mathcal{M}_{ij}$ ,  $i \in [\alpha n]$  and  $j \in [n]$ , where  $\mathbb{P}[\mathcal{M}_{ij} = +1] = \frac{1}{2} = \mathbb{P}[\mathcal{M}_{ij} = -1]$ . Then, there exists an algorithm  $\mathcal{A}$ , running in  $\text{poly}(n)$  time, such that w.h.p.  $\|\mathcal{M} \cdot \mathcal{A}(\mathcal{M})\|_\infty \leq \kappa\sqrt{n}$ .

Corollary 2 is a direct consequence of Theorem 10. Indeed, consider  $\mathcal{M} \in \{\pm 1\}^{\alpha T \times T}$  with  $\alpha = n/T$ , whose columns are  $v_i$ ,  $1 \leq i \leq T$ . Then one can find, in polynomial (in  $n, T$ ) time, a  $\sigma \in \mathcal{B}_T$  such that  $\|\mathcal{M}\sigma\|_\infty \leq K\sqrt{n} = K\sqrt{\alpha T}$ . Since  $\alpha < (\kappa/K)^2$ , the claim follows.

Admittedly, their result is established for the case of i.i.d. Rademacher disorder. Nevertheless, due to the aforementioned universality guarantees encountered in perceptron-like models, it is expected that the exact same guarantee (perhaps with a modified constant  $K$ ) remains true for the case of i.i.d. standard normal disorder.

b) *A Conjecture on the Algorithmic Threshold*: Recall from above that for many random computational problems, the  $m$ -OGP threshold coincides (or nearly coincides) with conjectured algorithmic threshold. Examples include finding the largest independent set in random sparse graphs [47], [48], NAE- $k$ -SAT [43], random  $k$ -SAT [45], mixed even  $p$ -spin model [88], and so on. In light of the preceding discussion, this is also the case for the SBP model: the limit of known algorithms is at  $\Theta(\kappa^2)$ , whereas,

as we establish in Theorem 7, the ensemble  $m$ -OGP holds for densities  $\Omega(\kappa^2 \log_2 \frac{1}{\kappa})$  in the regime  $\kappa \rightarrow 0$ .

On the other hand, unlike models such as the independent set problem,  $k$ -SAT, or the planted clique; prior to this work no conjectures were proposed regarding the threshold for algorithmic hardness in SBP model in the  $\kappa \rightarrow 0$  regime. Here, we do put forward such a conjecture. To that end, let

$$\alpha_m^*(\kappa) \triangleq \inf \left\{ \alpha \in [0, \alpha_c(\kappa)] : \exists 1 > \beta > \eta > 0, \right. \\ \left. \liminf_{n \rightarrow \infty} \mathbb{P} \left[ \mathcal{S}_\kappa(\beta, \eta, m, \alpha, \{0\}) = \emptyset \right] = 1 \right\}.$$

That is,  $\alpha_m^*(\kappa)$  is the threshold for the  $m$ -OGP (with appropriate  $\beta, \eta$ ). Let

$$\alpha_\infty^*(\kappa) \triangleq \lim_{m \rightarrow \infty} \alpha_m^*(\kappa),$$

where the limit is well-defined since  $(\alpha_m^*)_{m \geq 1}$  is a non-increasing sequence of non-negative real numbers. Then we conjecture  $\alpha_\infty^*(\kappa)$  marks the true algorithmic threshold for this problem.

**Conjecture 3.** *For any  $\epsilon > 0$ , there exists a  $\kappa^*(\epsilon) > 0$  such that the following hold for every  $\kappa \leq \kappa^*(\epsilon)$ :*

- *There exists no polynomial-time search algorithms for the SBP if  $\alpha > (1 + \epsilon)\alpha_\infty^*(\kappa)$ .*
- *There exists a polynomial-time search algorithm for the SBP if  $\alpha < (1 - \epsilon)\alpha_\infty^*(\kappa)$ .*

Recall that per Theorem 7,  $\alpha_\infty^*(\kappa) = O(\kappa^2 \log_2 \frac{1}{\kappa})$  as  $\kappa \rightarrow 0$ . Notice that the  $\alpha_m^*(\kappa)$  (hence the  $\alpha_\infty^*(\kappa)$ ) are defined for the non-ensemble variant of  $m$ -OGP,  $\mathcal{I} = \{0\}$ . That is,  $\sigma^{(i)}$ ,  $1 \leq i \leq m$ , satisfy constraints dictated by the rows of the **same** disorder matrix  $\mathcal{M} \in \mathbb{R}^{M \times n}$  with i.i.d.  $\mathcal{N}(0, 1)$  (or Rademacher) entries, where  $M = \lfloor \kappa n \rfloor$ . This is merely for simplicity: the ensemble  $m$ -OGP and the non-ensemble  $m$ -OGP often take place at the **exact same** threshold. The former, on the other hand, is just technically more involved; and is necessary to rule out certain classes of algorithms via an interpolation/contradiction argument as we do in this paper. The structural property implied by the non-ensemble OGP already suffices to predict the desired algorithmic threshold.

#### D. Stability of the Kim-Roche Algorithm

Having established that the  $m$ -OGP is a provable barrier for the class of stable algorithms, it is then natural to inquire whether the class of stable algorithms captures the implementations of known algorithms for perceptron models. Here we investigate this question for a certain algorithm devised for the *asymmetric* model, which we recall from (1).

Kim and Roche devised in [5] an algorithm which takes an  $\mathcal{M} \in \mathbb{R}^{k \times n}$  with i.i.d. entries as its input and returns a  $\sigma \in \mathcal{B}_n$  such that  $\mathcal{M}\sigma \geq 0$  entry-wise as long as  $k < 0.005n$ . (We use  $k$  in place of  $M$  so as to be consistent with their notation.) In particular,  $S_\alpha^A(0) \neq \emptyset$  w.h.p. if  $\alpha <$

0.005. We denote their algorithm by  $\mathcal{A}_{\text{KR}} : \mathbb{R}^{k \times n} \rightarrow \mathcal{B}_n$ . It is worth noting that while their results are established for the case where  $\mathcal{M}$  consists of i.i.d. Rademacher entries, they easily extend to the case of Gaussian  $\mathcal{N}(0, 1)$  entries, which will be our focus here.  $\mathcal{A}_{\text{KR}}$  takes  $O(\log \log n)$  steps, each requiring  $\text{poly}(n)$  time. Namely,  $\mathcal{A}_{\text{KR}}$  is an *efficient* algorithm that provably works in the so-called *linear regime*,  $k = \Theta(n)$ . Admittedly,  $\mathcal{A}_{\text{KR}}$  is tailored for the ABP. Nevertheless, there are only a few known algorithms with rigorous guarantees for perceptron models; thus it is indeed natural to explore the stability of  $\mathcal{A}_{\text{KR}}$ .

We defer an informal description of  $\mathcal{A}_{\text{KR}}$  and its operational parameters to the full version [50, Section 3.4]. Our next main result establishes that  $\mathcal{A}_{\text{KR}}$  is indeed stable in the sense of Definition 2.

**Theorem 11.** *Let  $\mathcal{M}, \mathcal{M}' \in \mathbb{R}^{k \times n}$  be i.i.d. random matrices, each with i.i.d.  $\mathcal{N}(0, 1)$  entries. For  $\tau = n^{-0.02}$ , let*

$$\overline{\mathcal{M}}(\tau) \triangleq \cos(\tau)\mathcal{M} + \sin(\tau)\mathcal{M}'. \quad (7)$$

*Let  $\sigma = \mathcal{A}_{\text{KR}}(\mathcal{M}) \in \mathcal{B}_n$ ,  $\overline{\sigma} \triangleq \mathcal{A}_{\text{KR}}(\mathcal{M}(\tau)) \in \mathcal{B}_n$ . Then,*

$$\mathbb{P} \left[ d_H(\sigma, \overline{\sigma}) = o(n) \right] \geq 1 - O(n^{-\frac{1}{41}}).$$

The proof is deferred to the full version [50].

As a result, the Kim-Roche algorithm is

$$(\cos(n^{-0.02}), o(1/n), O(n^{-1/41}), Cn, L) - \text{stable}$$

in the sense of Definition 2 for any  $C > 0$  and  $L > 0$  (see [50, Page 21] for further details). Note though that this parameter scaling is not comparable with Theorem 7 since Theorem 7 pertains to the symmetric model, see below for more details. Recalling  $\mathcal{O}(\sigma, \overline{\sigma}) = n^{-1} \langle \sigma, \overline{\sigma} \rangle = 1 - 2d_H(\sigma, \overline{\sigma})/n$ , it follows that in the setting of Theorem 11,  $\mathcal{O}(\sigma, \overline{\sigma}) = 1 - o(1)$ . That is,  $\sigma$  and  $\overline{\sigma}$  agree on all but a vanishing fraction of coordinates. Informally, this suggests that  $\mathcal{A}_{\text{KR}}$  cannot overcome the overlap barrier of  $\eta$  appearing in Theorems 5, 6, and 7 as  $\eta = O(1)$ . However, we established the OGP results for the symmetric case as opposed to the asymmetric model for which  $\mathcal{A}_{\text{KR}}$  is devised. Thus Theorem 11 is not exactly compatible with the hardness result, Theorem 8. A more compelling picture would be to show that the OGP takes place also for the asymmetric model, with an  $\eta$  that is of order  $O(1)$ ; and then couple such result with Theorem 11. We leave this as a very interesting direction for future work.

Lastly, it would also be very interesting to prove that the algorithm by Abbe, Li and Sly [4] devised for the SBP is also stable in the relevant sense. While an inspection of [4] reveals that several of the key steps are indeed similar to that of [5], there are a few differences which prevent immediate verification of stability, see [50, Page 23] for a detailed discussion. We leave the formal investigation of this as a very interesting direction for future research.

## REFERENCES

- [1] W. Perkins and C. Xu, “Frozen 1-RSB structure of the symmetric Ising perceptron,” in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, 2021, pp. 1579–1588.
- [2] E. Abbe, S. Li, and A. Sly, “Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron,” *arXiv preprint arXiv:2102.13069*, 2021.
- [3] C. Baldassi, R. Della Vecchia, C. Lucibello, and R. Zecchina, “Clustering of solutions in the symmetric binary perceptron,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2020, no. 7, p. 073303, 2020.
- [4] E. Abbe, S. Li, and A. Sly, “Binary perceptron: efficient algorithms can find solutions in a rare well-connected cluster,” *arXiv preprint arXiv:2111.03084*, 2021.
- [5] J. H. Kim and J. R. Roche, “Covering cubes by random half cubes, with applications to binary neural networks,” *Journal of Computer and System Sciences*, vol. 56, no. 2, pp. 223–252, 1998.
- [6] R. D. Joseph and L. Hay, “The number of orthants in n-space intersected by an s-dimensional subspace,” CORNELL AERONAUTICAL LAB INC BUFFALO NY, Tech. Rep., 1960.
- [7] R. O. Winder, “Single stage threshold logic,” in *2nd Annual Symposium on Switching Circuit Theory and Logical Design (SWCT 1961)*. IEEE, 1961, pp. 321–332.
- [8] J. G. Wendel, “A problem in geometric probability,” *Mathematica Scandinavica*, vol. 11, no. 1, pp. 109–111, 1962.
- [9] T. M. Cover, “Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition,” *IEEE transactions on electronic computers*, no. 3, pp. 326–334, 1965.
- [10] E. Gardner, “The space of interactions in neural network models,” *Journal of physics A: Mathematical and general*, vol. 21, no. 1, p. 257, 1988.
- [11] M. Shcherbina and B. Tirozzi, “Rigorous solution of the Gardner problem,” *Communications in mathematical physics*, vol. 234, no. 3, pp. 383–422, 2003.
- [12] M. Stojnic, “Another look at the Gardner problem,” *arXiv preprint arXiv:1306.3979*, 2013.
- [13] M. Talagrand, *Mean Field Models for Spin Glasses: Advanced replica-symmetry and low temperature*. Springer, 2011.
- [14] A. E. Alaoui and M. Sellke, “Algorithmic pure states for the negative spherical perceptron,” *arXiv preprint arXiv:2010.15811*, 2020.
- [15] E. Gardner, “Maximum storage capacity in neural networks,” *EPL (Europhysics Letters)*, vol. 4, no. 4, p. 481, 1987.
- [16] E. Gardner and B. Derrida, “Optimal storage properties of neural network models,” *Journal of Physics A: Mathematical and general*, vol. 21, no. 1, p. 271, 1988.
- [17] E. Bolthausen, S. Nakajima, N. Sun, and C. Xu, “Gardner formula for Ising perceptron models at small densities,” *arXiv preprint arXiv:2111.02855*, 2021.
- [18] B. Aubin, W. Perkins, and L. Zdeborová, “Storage capacity in symmetric binary perceptrons,” *Journal of Physics A: Mathematical and Theoretical*, vol. 52, no. 29, p. 294003, 2019.
- [19] W. Krauth and M. Mézard, “Storage capacity of memory networks with binary couplings,” *Journal de Physique*, vol. 50, no. 20, pp. 3057–3066, 1989.
- [20] C. Xu, “Sharp threshold for the Ising perceptron model,” *arXiv preprint arXiv:1905.05978*, 2019.
- [21] S. Nakajima and N. Sun, “Sharp threshold sequence and universality for ising perceptron models,” *arXiv preprint arXiv:2204.03469*, 2022.
- [22] E. Friedgut, “Sharp thresholds of graph properties, and the k-SAT problem,” *Journal of the American mathematical Society*, vol. 12, no. 4, pp. 1017–1054, 1999.
- [23] J. Ding and N. Sun, “Capacity lower bound for the Ising perceptron,” in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, 2019, pp. 816–827.
- [24] M. Talagrand, “Intersecting random half cubes,” *Random Structures & Algorithms*, vol. 15, no. 3-4, pp. 436–449, 1999.
- [25] J. Spencer, “Six standard deviations suffice,” *Transactions of the American mathematical society*, vol. 289, no. 2, pp. 679–706, 1985.
- [26] J. Matousek, *Geometric discrepancy: An illustrated guide*. Springer Science & Business Media, 1999, vol. 18.
- [27] A. Potukuchi, “Discrepancy in random hypergraph models,” *arXiv preprint arXiv:1811.01491*, 2018.
- [28] P. Turner, R. Meka, and P. Rigollet, “Balancing Gaussian vectors in high dimension,” in *Conference on Learning Theory*. PMLR, 2020, pp. 3455–3486.
- [29] D. J. Altschuler and J. Niles-Weed, “The discrepancy of random rectangular matrices,” *Random Structures & Algorithms*, 2021.
- [30] N. Bansal, “Constructive algorithms for discrepancy minimization,” in *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. IEEE, 2010, pp. 3–10.
- [31] S. Lovett and R. Meka, “Constructive discrepancy minimization by walking on the edges,” *SIAM Journal on Computing*, vol. 44, no. 5, pp. 1573–1582, 2015.
- [32] K. Chandrasekaran and S. S. Vempala, “Integer feasibility of random polytopes: random integer programs,” in *Proceedings of the 5th conference on Innovations in theoretical computer science*, 2014, pp. 449–458.
- [33] N. Bansal and J. Spencer, “On-line balancing of random inputs,” *Random Structures and Algorithms*, vol. 57, no. 4, pp. 879–891, Dec. 2020.
- [34] A. Potukuchi, “A Spectral Bound on Hypergraph Discrepancy,” in *47th International Colloquium on Automata, Languages, and Programming (ICALP 2020)*, ser. Leibniz International Proceedings in Informatics (LIPIcs), A. Czumaj, A. Dawar, and E. Merelli, Eds., vol. 168. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2020, pp. 93:1–93:14.
- [35] H. Huang, K. M. Wong, and Y. Kabashima, “Entropy landscape of solutions in the binary perceptron problem,” *Journal of Physics A: Mathematical and Theoretical*, vol. 46, no. 37, p. 375002, 2013.
- [36] M. Mézard, T. Mora, and R. Zecchina, “Clustering of solutions in the random satisfiability problem,” *Physical Review Letters*, vol. 94, no. 19, p. 197205, 2005.
- [37] D. Achlioptas and F. Ricci-Tersenghi, “On the solution-space geometry of random constraint satisfaction problems,” in *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, 2006, pp. 130–139.
- [38] A. Braunstein and R. Zecchina, “Learning by message passing in networks of discrete synapses,” *Physical review letters*, vol. 96, no. 3, p. 030201, 2006.
- [39] C. Baldassi, A. Braunstein, N. Brunel, and R. Zecchina, “Efficient supervised learning in networks with binary synapses,” *Proceedings of the National Academy of Sciences*, vol. 104, no. 26, pp. 11 079–11 084, 2007.
- [40] C. Baldassi, “Generalization learning in a perceptron with binary synapses,” *Journal of Statistical Physics*, vol. 136, no. 5, pp. 902–916, 2009.
- [41] C. Baldassi and A. Braunstein, “A max-sum algorithm for training discrete neural networks,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2015, no. 8, p. P08008, 2015.
- [42] C. Baldassi, A. Ingrosso, C. Lucibello, L. Saglietti, and R. Zecchina, “Subdominant dense clusters allow for simple learning and high computational performance in neural networks with discrete synapses,” *Physical review letters*, vol. 115, no. 12, p. 128101, 2015.
- [43] D. Gamarnik and M. Sudan, “Performance of sequential local algorithms for the random NAE-K-SAT problem,” *SIAM Journal on Computing*, vol. 46, no. 2, pp. 590–619, 2017.
- [44] A. Coja-Oghlan, A. Haghshenas, and S. Hetterich, “Walksat stalls well below satisfiability,” *SIAM Journal on Discrete Mathematics*, vol. 31, no. 2, pp. 1160–1173, 2017.
- [45] G. Bresler and B. Huang, “The algorithmic phase transition of random  $k$ -sat for low degree polynomials,” *arXiv preprint arXiv:2106.02129*, 2021.
- [46] D. Gamarnik and M. Sudan, “Limits of local algorithms over sparse random graphs,” in *Proceedings of the 5th conference on Innovations in theoretical computer science*, 2014, pp. 369–376.
- [47] M. Rahman and B. Virág, “Local algorithms for independent sets are half-optimal,” *The Annals of Probability*, vol. 45, no. 3, pp. 1543–1577, 2017.
- [48] A. S. Wein, “Optimal low-degree hardness of maximum independent set,” *arXiv preprint arXiv:2010.06563*, 2020.
- [49] D. Gamarnik, “The overlap gap property: A topological barrier to optimizing over random structures,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 41, 2021.
- [50] D. Gamarnik, E. C. Kızıldağ, W. Perkins, and C. Xu, “Algorithms and barriers in the symmetric binary perceptron model,” *arXiv preprint arXiv:2203.15667*, 2022.

[51] D. Gamarnik and A. Jagannath, “The overlap gap property and approximate message passing algorithms for  $p$ -spin models,” *The Annals of Probability*, vol. 49, no. 1, pp. 180–205, 2021.

[52] D. Gamarnik, A. Jagannath, and A. S. Wein, “Low-degree hardness of random optimization problems,” in *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2020, pp. 131–140.

[53] ———, “Circuit lower bounds for the  $p$ -spin optimization problem,” *arXiv preprint arXiv:2109.01342*, 2021.

[54] D. Achlioptas and A. Coja-Oghlan, “Algorithmic barriers from phase transitions,” in *2008 49th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, 2008, pp. 793–802.

[55] A. Coja-Oghlan and C. Efthymiou, “On independent sets in random graphs,” *Random Structures & Algorithms*, vol. 47, no. 3, pp. 436–486, 2015.

[56] D. Gamarnik and M. Sudan, “Limits of local algorithms over sparse random graphs,” *Ann. Probab.*, vol. 45, no. 4, pp. 2353–2376, 07 2017. [Online]. Available: <https://doi.org/10.1214/16-AOP1114>

[57] W.-K. Chen, D. Gamarnik, D. Panchenko, and M. Rahman, “Suboptimality of local algorithms for a class of max-cut problems,” *The Annals of Probability*, vol. 47, no. 3, pp. 1587–1618, 2019.

[58] D. Gamarnik and E. C. Kizildag, “Algorithmic obstructions in the random number partitioning problem,” *arXiv preprint arXiv:2103.01369*, 2021.

[59] Q. Berthet and P. Rigollet, “Computational lower bounds for sparse PCA,” *arXiv preprint arXiv:1304.0828*, 2013.

[60] T. Lesieur, F. Krzakala, and L. Zdeborová, “Mmse of probabilistic low-rank matrix estimation: Universality with respect to the output channel,” in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2015, pp. 680–687.

[61] ———, “Phase transitions in sparse pca,” in *2015 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2015, pp. 1635–1639.

[62] M. Jerrum, “Large cliques elude the metropolis process,” *Random Structures & Algorithms*, vol. 3, no. 4, pp. 347–359, 1992.

[63] Y. Deshpande and A. Montanari, “Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems,” in *Conference on Learning Theory*, 2015, pp. 523–562.

[64] B. Barak, S. Hopkins, J. Kelner, P. K. Kothari, A. Moitra, and A. Potechin, “A nearly tight sum-of-squares lower bound for the planted clique problem,” *SIAM Journal on Computing*, vol. 48, no. 2, pp. 687–735, 2019.

[65] E. C. Kizildag, “Algorithms and algorithmic barriers in high-dimensional statistics and random combinatorial structures,” Ph.D. dissertation, Massachusetts Institute of Technology, 2022.

[66] D. Gamarnik and E. C. Kizildag, “Computing the partition function of the Sherrington–Kirkpatrick model is hard on average,” *The Annals of Applied Probability*, vol. 31, no. 3, pp. 1474 – 1504, 2021. [Online]. Available: <https://doi.org/10.1214/20-AAP1625>

[67] M. Dyer, A. Frieze, and M. Jerrum, “On counting independent sets in sparse graphs,” *SIAM Journal on Computing*, vol. 31, no. 5, pp. 1527–1541, 2002.

[68] S. B. K. Hopkins, “Statistical inference and the sum of squares method,” 2018.

[69] D. Kunisky, A. S. Wein, and A. S. Bandeira, “Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio,” *arXiv preprint arXiv:1907.11636*, 2019.

[70] S. B. Hopkins, J. Shi, and D. Steurer, “Tensor principal component analysis via sum-of-square proofs,” in *Conference on Learning Theory*, 2015, pp. 956–1006.

[71] S. B. Hopkins, P. K. Kothari, A. Potechin, P. Raghavendra, T. Schramm, and D. Steurer, “The power of sum-of-squares for detecting hidden structures,” in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2017, pp. 720–731.

[72] P. Raghavendra, T. Schramm, and D. Steurer, “High-dimensional estimation via sum-of-squares proofs,” *arXiv preprint arXiv:1807.11419*, vol. 6, 2018.

[73] M. Kearns, “Efficient noise-tolerant learning from statistical queries,” *Journal of the ACM (JACM)*, vol. 45, no. 6, pp. 983–1006, 1998.

[74] I. Diakonikolas, D. M. Kane, and A. Stewart, “Statistical query lower bounds for robust estimation of high-dimensional Gaussians and Gaussian mixtures,” in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2017, pp. 73–84.

[75] V. Feldman, E. Grigorescu, L. Reyzin, S. S. Vempala, and Y. Xiao, “Statistical algorithms and a lower bound for detecting planted cliques,” *Journal of the ACM (JACM)*, vol. 64, no. 2, pp. 1–37, 2017.

[76] V. Feldman, W. Perkins, and S. Vempala, “On the complexity of random satisfiability problems with planted solutions,” *SIAM Journal on Computing*, vol. 47, no. 4, pp. 1294–1338, 2018.

[77] Y. Deshpande and A. Montanari, “Information-theoretically optimal sparse PCA,” in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 2197–2201.

[78] Y. Deshpande, E. Abbe, and A. Montanari, “Asymptotic mutual information for the balanced binary stochastic block model,” *Information and Inference: A Journal of the IMA*, vol. 6, no. 2, pp. 125–170, 2017.

[79] L. Zdeborová and F. Krzakala, “Statistical physics of inference: Thresholds and algorithms,” *Advances in Physics*, vol. 65, no. 5, pp. 453–552, 2016.

[80] A. S. Bandeira, A. Perry, and A. S. Wein, “Notes on computational-to-statistical gaps: predictions using statistical physics,” *arXiv preprint arXiv:1803.11132*, 2018.

[81] M. Brennan, G. Bresler, and W. Huleihel, “Reducibility and computational lower bounds for problems with planted sparse structure,” *arXiv preprint arXiv:1806.07508*, 2018.

[82] M. Brennan and G. Bresler, “Optimal average-case reductions to sparse pca: From weak assumptions to strong hardness,” *arXiv preprint arXiv:1902.07380*, 2019.

[83] D. Gamarnik and Q. Li, “Finding a large submatrix of a Gaussian random matrix,” *The Annals of Statistics*, vol. 46, no. 6A, pp. 2511–2561, 2018.

[84] A. M. Frieze, “On the independence number of random graphs,” *Discrete Mathematics*, vol. 81, no. 2, pp. 171–175, 1990.

[85] A. M. Frieze and T. Łuczak, “On the independence and chromatic numbers of random regular graphs,” *Journal of Combinatorial Theory, Series B*, vol. 54, no. 1, pp. 123–132, 1992.

[86] M. Bayati, D. Gamarnik, and P. Tetali, “Combinatorial approach to the interpolation method and scaling limits in sparse random graphs,” in *Proceedings of the forty-second ACM symposium on Theory of computing*, 2010, pp. 105–114.

[87] R. M. Karp, “The probabilistic analysis of some combinatorial search algorithms.” 1976.

[88] B. Huang and M. Sellke, “Tight lipschitz hardness for optimizing mean field spin glasses,” *arXiv preprint arXiv:2110.07847*, 2021.

[89] I. Zadik, M. J. Song, A. S. Wein, and J. Bruna, “Lattice-based methods surpass sum-of-squares in clustering,” *arXiv preprint arXiv:2112.03898*, 2021.

[90] G. B. Arous and A. Jagannath, “Shattering versus metastability in spin glasses,” *arXiv preprint arXiv:2104.08299*, 2021.

[91] T. Rothvoss, “Constructive discrepancy minimization for convex sets,” *SIAM Journal on Computing*, vol. 46, no. 1, pp. 224–234, 2017.

[92] A. Levy, H. Ramadas, and T. Rothvoss, “Deterministic discrepancy minimization via the multiplicative weight update method,” in *International Conference on Integer Programming and Combinatorial Optimization*. Springer, 2017, pp. 380–391.

[93] R. Eldan and M. Singh, “Efficient algorithms for discrepancy minimization in convex sets,” *Random Structures & Algorithms*, vol. 53, no. 2, pp. 289–307, 2018.

[94] N. Bansal, H. Jiang, S. Singla, and M. Sinha, “Online vector balancing and geometric discrepancy,” in *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, 2020, pp. 1139–1152.

[95] N. Bansal, H. Jiang, R. Meka, S. Singla, and M. Sinha, “Online discrepancy minimization for stochastic arrivals,” in *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 2021, pp. 2842–2861.

[96] R. Alweiss, Y. P. Liu, and M. Sawhney, “Discrepancy minimization via a self-balancing walk,” in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, 2021, pp. 14–20.

[97] Y. P. Liu, A. Sah, and M. Sawhney, “A Gaussian fixed point random walk,” *arXiv preprint arXiv:2104.07009*, 2021.