Geometric Barriers for Stable and Online Algorithms for Discrepancy Minimization

David Gamarnik GAMARNIK@MIT.EDU
Sloan School of Management, Massachusetts Institute of Technology

Eren C. Kızıldağ ECK2170@COLUMBIA.EDU
Department of Statistics, Columbia University

Will Perkins MATH@WILLPERKINS.ORG
School of Computer Science, Georgia Institute of Technology

Changji Xu CXU@CMSA.FAS.HARVARD.EDU
Center of Mathematical Sciences and Applications, Harvard University

Editors: Gergely Neu and Lorenzo Rosasco

Abstract

For many computational problems involving randomness, intricate geometric features of the solution space have been used to rigorously rule out powerful classes of algorithms. This is often accomplished through the lens of the multi Overlap Gap Property (m-OGP), a rigorous barrier against algorithms exhibiting input stability. In this paper, we focus on the algorithmic tractability of two models: (i) discrepancy minimization, and (ii) the symmetric binary perceptron (SBP), a random constraint satisfaction problem as well as a toy model of a single-layer neural network.

Our first focus is on the limits of online algorithms. By establishing and leveraging a novel geometrical barrier, we obtain sharp hardness guarantees against online algorithms for both the SBP and discrepancy minimization. Our results match the best known algorithmic guarantees, up to constant factors. Our second focus is on efficiently finding a constant discrepancy solution, given a random matrix $\mathcal{M}\in\mathbb{R}^{M\times n}$. In a smooth setting, where the entries of $\mathcal{M}$ are i.i.d. standard normal, we establish the presence of m-OGP for $n=\Theta(M\log M)$. Consequently, we rule out the class of stable algorithms at this value. These results give the first rigorous evidence towards Altschuler and Niles-Weed (2022, Conjecture 1).

Our methods use the intricate geometry of the solution space to prove tight hardness results for online algorithms. The barrier we establish is a novel variant of the m-OGP. Furthermore, it regards m-tuples of solutions with respect to correlated instances, with growing values of m, $m=\omega(1)$. Importantly, our results rule out online algorithms succeeding even with an exponentially small probability.

Keywords: Discrepancy, Binary perceptron, overlap gap property, statistical-to-computational gap.

1. Introduction

In this paper, we study the *discrepancy minimization* problem and the *perceptron* model. Combinatorial discrepancy theory (Spencer, 1985; Matousek, 1999) is a central topic at the intersection of combinatorics, probability, and algorithms. Given a matrix $\mathcal{M}\in\mathbb{R}^{M\times n}$, the central task in discrepancy theory is computing or bounding the quantity

$$\mathcal{D}(\mathcal{M})\triangleq\min_{\boldsymbol{\sigma}\in\Sigma_n}\left\|\mathcal{M}\boldsymbol{\sigma}\right\|_\infty,$$

known as the *discrepancy* of $\mathcal{M}$.

The *perceptron* is a toy one-layer neural network model storing random patterns as well as a very natural high-dimensional probabilistic model, see Joseph and Hay (1960); Winder (1961); Wendel (1962); Cover (1965) for early works on it. Given random patterns $X_i \in \mathbb{R}^n$, $1 \le i \le M$, *storage* is achieved if one finds a $\boldsymbol{\sigma} \in \mathbb{R}^n$ 'consistent' with all X_i: $\langle \boldsymbol{\sigma}, X_i \rangle \ge 0$ for $1 \le i \le M$. The vector $\boldsymbol{\sigma}$ is interpreted as *synaptic weights*; it can either lie on the sphere in $\mathbb{R}^n$, $\|\boldsymbol{\sigma}\|_2 = \sqrt{n}$, or have binary entries, $\boldsymbol{\sigma} \in \Sigma_n = \{-1,1\}^n$. The former is dubbed as the spherical perceptron, see Gardner (1988); Shcherbina and Tirozzi (2003); Stojnic (2013); Talagrand (2011); Alaoui and Sellke (2020) for relevant work. In this paper, we only focus on the latter, dubbed as the *binary perceptron*. A fundamental object studied in the perceptron literature is the *storage capacity*: the maximum number of (random) patterns that can be stored with a suitable $\boldsymbol{\sigma}$, see Gardner (1987, 1988); Gardner and Derrida (1988). Krauth and Mézard (1989) gave a detailed though non-rigorous characterization of the storage capacity. More recently, perceptron models with an *activation function* $U : \mathbb{R} \to \{0,1\}$ are considered, where a pattern X_i is stored with respect to (w.r.t.) U if $U(\langle \boldsymbol{\sigma}, X_i \rangle) = 1$. Of particular interest to us is the activation $U(x) = \mathbf{1}_{|x| \le \kappa\sqrt{n}}$ which defines the *symmetric binary perceptron* ($\texttt{SBP}$) model proposed by Aubin et al. (2019) (see also Bolthausen et al. (2021); Nakajima and Sun (2023) for results on more general perceptron models and Kızıldağ and Wakhare (2023) for a version of the $\texttt{SBP}$ where random labels independent of the disorder are incorporated to the model). As we see below, the $\texttt{SBP}$ is closely related to discrepancy minimization.

1.1. Discrepancy Minimization

The discrepancy literature pertains to both *worst-case* and *average-case* $\mathcal{M}$. In the worst-case, minimal structure is assumed on $\mathcal{M}$, whereas in the average-case, the entries of $\mathcal{M}$ are random, e.g. i.i.d. Bernoulli, Rademacher, or standard normal. Moreover, both existential as well as algorithmic results are sought in discrepancy theory.

Concerning the worst-case analysis, a landmark result in the area is due to Spencer (1985): $\mathcal{D}(\mathcal{M}) \le 6\sqrt{n}$ if $\mathcal{M} \in \mathbb{R}^{n\times n}$ with $|M_{ij}| \le 1$ for $1 \le i,j \le n$ ('six standard deviations suffice'). The significance of this result is the improvement over the discrepancy guaranteed by the basic probabilistic method: the discrepancy incurred by a random signing is of order $\Theta(\sqrt{n\log n})$ which is substantially larger than $O(\sqrt{n})$. It is worth noting that Spencer's result is worst-case and non-constructive, but recent work by Bansal (2010); Lovett and Meka (2015); Levy et al. (2017); Rothvoss (2017) has given efficient algorithms to find such low discrepancy solutions.

In this paper we focus on average-case discrepancy. Suppose $\mathcal{M} \in \mathbb{R}^{M\times n}$ has i.i.d. $\mathcal{N}(0,1)$ entries and $M = o(n)$. In this case, a line of work initiated in Karmarkar et al. (1986) (for $M = 1$) and subsequently continued in Costello (2009); Turner et al. (2020) (for $M \ge 2$) established that $\mathcal{D}(\mathcal{M}) = \Theta(\sqrt{n}2^{-n/M})$ w.h.p. Algorithmic results in this regime are found in Karmarkar and Karp (1982); Yakir (1996); Turner et al. (2020). In the special case of $M = 1$, Gamarnik and Kızıldağ (2021) gave rigorous evidence that finding a $\boldsymbol{\sigma}$ with $\|\mathcal{M}\boldsymbol{\sigma}\|_\infty = 2^{-\omega(\sqrt{n\log n})}$ may be algorithmically intractable. On the other hand when $M = \Theta(n)$, then it turns out $\mathcal{D}(\mathcal{M}) = \Theta(\sqrt{n})$ and this case is closely related to the $\texttt{SBP}$, see Section 1.3 for more details.

Next suppose the entries of $\mathcal{M}$ are i.i.d. binary, e.g. Rademacher or Bernoulli(p). In this case, while still $\mathcal{D}(\mathcal{M}) = \Theta(\sqrt{n})$ w.h.p. when $M = \Theta(n)$, but it turns out that *constant* discrepancy, in fact $\mathcal{D}(\mathcal{M}) = 1$, is possible when n is much larger than M. The sharpest possible result to this end is due to Altschuler and Niles-Weed (2022) who completely resolved the question of exactly when

$\mathcal{D}(\mathcal{M}) \le 1$: $\mathcal{D}(\mathcal{M}) \le 1$ if $\mathcal{M}$ consists of Bernoulli(p) entries with arbitrary p and $n \ge CM\log M$, where C is any arbitrary constant greater than $(2\log 2)^{-1}$. Their result covers in particular the sparse regime, $p = o(1)$, and is the sharpest possible as $n = \Omega(M\log M)$ is needed for $\mathcal{D}(\mathcal{M})$ to be $O(1)$: if $n = CM\log M$ for $C < (2\log 2)^{-1}$ and $p = 1/2$, then w.h.p. no constant discrepancy solutions exist. Earlier results towards this direction are found in Hoberg and Rothvoss (2019); Franks and Saks (2020); Potukuchi (2018). Equipped with this existential guarantee from Altschuler and Niles-Weed (2022), a natural algorithmic question is whether one can find such a constant discrepancy solution in polynomial time. This task is conjecturally hard (see Altschuler and Niles-Weed, 2022, Conjecture 1). The algorithmic tractability of this problem is a main focuses of the present paper.

1.2. Symmetric Binary Perceptron (SBP)

Fix $\kappa > 0$, $\alpha > 0$, and set $M = \lfloor n\alpha \rfloor \in \mathbb{N}$. Let $X_i \sim \mathcal{N}(0, I_n)$, $1 \le i \le M$, be i.i.d. random vectors, where $\mathcal{N}(0, I_n)$ is the centered multivariate normal distribution in $\mathbb{R}^n$ with identity covariance. Consider the (random) set

$$S_\alpha(\kappa) = \Big\{\boldsymbol{\sigma} \in \Sigma_n : |\langle \boldsymbol{\sigma}, X_i\rangle| \le \kappa\sqrt{n}, 1 \le i \le M\Big\} = \Big\{\boldsymbol{\sigma} \in \Sigma_n : \big\|\mathcal{M}\boldsymbol{\sigma}\big\|_\infty \le \kappa\sqrt{n}\Big\}, \qquad (1)$$

where $\mathcal{M} \in \mathbb{R}^{M\times n}$ with rows $X_1, \dots, X_M$. The word *symmetric* refers to the fact $\boldsymbol{\sigma} \in S_\alpha(\kappa)$ iff $-\boldsymbol{\sigma} \in S_\alpha(\kappa)$. The SBP was put forth by Aubin, Perkins, and Zdeborová (2019) as a symmetric counterpart to the *asymmetric binary perceptron* (ABP), where the constraints are instead of form $\langle \boldsymbol{\sigma}, X\rangle \ge \kappa\sqrt{n}$, $1 \le i \le M$. The ABP turns out very challenging mathematically, see Krauth and Mézard (1989); Kim and Roche (1998); Talagrand (1999); Xu (2021); Ding and Sun (2019); Perkins and Xu (2021); Abbe et al. (2021a); Gamarnik et al. (2022a); Kızıldağ (2022) for relevant work and more details. The SBP, on the other hand, retains pertinent structural properties conjectured for the ABP (Baldassi et al., 2020), while being more amenable to rigorous analysis thanks to the symmetry.

The SBP undergoes a *sharp phase transition*, conjectured in Aubin et al. (2019) and subsequently proven independently by Perkins and Xu (2021) and Abbe et al. (2021b). Let

$$\alpha_c(\kappa) \triangleq -\frac{1}{\log_2 \mathbb{P}[|Z| \le \kappa]}, \quad \text{where} \quad Z \sim \mathcal{N}(0,1). \qquad (2)$$

Then

$$\lim_{n\to\infty} \mathbb{P}\big[S_\alpha(\kappa) \ne \varnothing\big] = \begin{cases} 0, & \text{if } \alpha > \alpha_c(\kappa) \\ 1, & \text{if } \alpha < \alpha_c(\kappa) \end{cases}. \qquad (3)$$

The part $\alpha > \alpha_c(\kappa)$ is due to Aubin et al. (2019) and established via an application of the *first moment method*: $\mathbb{E}\big[|S_\alpha(\kappa)|\big] = o(1)$ if $\alpha > \alpha_c(\kappa)$, so $S_\alpha(\kappa) = \varnothing$ w.h.p. by Markov's inequality. The same paper also studies the case $\alpha < \alpha_c(\kappa)$ and establishes, through the *second moment method*, that $\liminf_{n\to\infty} \mathbb{P}\big[S_\alpha(\kappa) \ne \varnothing\big] > 0$. Boosting this to a high probability guarantee requires more powerful tools, see Perkins and Xu (2021) for a delicate martingale argument and Abbe, Li, and Sly (2021b) for an argument based on a fully connected analog of the small subgraph conditioning method. Furthermore, very recently, the critical window around $\alpha_c(\kappa)$ was shown to be of constant width, see Altschuler (2022); Sah and Sawhney (2023). These facts highlight that the first moment 'prediction' for the precise location of the phase transition is correct, and the transition is very sharp.

Given that $S_\alpha(\kappa)$ is w.h.p. non-empty if $\alpha < \alpha_c(\kappa)$, a natural follow-up question is algorithmic: can a solution $\boldsymbol{\sigma} \in S_\alpha(\kappa)$ be found efficiently? And does the existence of efficient algorithms depend on α? Efficient algorithms at small densities α were given in Kim and Roche (1998) and Abbe et al. (2021a) for the ABP and SBP respectively while on the negative side, Gamarnik et al. (2022a) studied the limits of efficient algorithms (see details below). The works Baldassi et al. (2007, 2015, 2020) put forth possible explanations for the success of efficient algorithms: while almost all solutions are totally frozen (conjectured in Mézard et al. (2005); Huang and Kabashima (2014)), efficient algorithms access rare solutions lying in large clusters. Recent works including Perkins and Xu (2021); Abbe et al. (2021b,a) have studied these structural predictions.

1.3. Connections between Discrepancy Theory and the SBP

In order to explicate the connection between discrepancy minimization and the SBP, we focus on the *proportional regime*, i.e. $M = \Theta(n)$. The discrepancy viewpoint is to take an $\mathcal{M} \in \mathbb{R}^{M\times n}$ with a fixed aspect ratio $\alpha = M/n$, and to seek a $\boldsymbol{\sigma}$ such that $\|\mathcal{M}\boldsymbol{\sigma}\|_\infty$ is as small as possible. The perceptron viewpoint, on the other hand, is the inverse: fix a $\kappa > 0$ first and seek the largest α for which a solution $\boldsymbol{\sigma}$ with $\|\mathcal{M}\boldsymbol{\sigma}\|_\infty \le \kappa\sqrt{n}$ exists. Furthermore, the asymptotic value of the average-case discrepancy in the proportional regime immediately follows from the sharp threshold result for the SBP (3): $\mathcal{D}(\mathcal{M}) = (1+o(1))f(\alpha)\sqrt{n}$ w.h.p., where $f(\alpha)$ is the 'inverse' of $\alpha_c(\kappa)$ (2).

Algorithmic Connections The connection between the SBP and discrepancy theory further extends to algorithmic domain: the best known efficient algorithm for the SBP comes from the discrepancy literature. Suppose $\mathcal{M} \in \mathbb{R}^{M\times n}$ has i.i.d. Rademacher entries. Bansal and Spencer (2020) devised an efficient *online algorithm* that finds a $\boldsymbol{\sigma}_{\rm ALG} \in \Sigma_n$ such that $\|\mathcal{M}\boldsymbol{\sigma}_{\rm ALG}\|_\infty = O(\sqrt{M})$ w.h.p. if $n \ge M = \omega(1)$. Informally, an algorithm is online if the $t^{\rm th}$ coordinate of the output $\boldsymbol{\sigma}_{\rm ALG}$ depends only on first t columns of $\mathcal{M}$, see Definition 7 for a formal definition. As an immediate corollary, this yields an efficient algorithm for the SBP that finds a solution $\boldsymbol{\sigma} \in S_\alpha(\kappa)$ w.h.p. if $\alpha = O(\kappa^2)$ (Gamarnik et al., 2022a, Corollary 3.6). In fact, this is the best known algorithmic guarantee both for the SBP and for discrepancy in the random proportional regime, see Gamarnik et al. (2022a, Section 3.3).

In light of these existential and algorithmic results, it appears that the SBP may exhibit a striking *statistical-to-computational gap* (SCG): the density below which solutions exist w.h.p., i.e. $\alpha_c(\kappa)$, is substantially larger than those below which polynomial-time search algorithms work. Further, this SCG is most profound when $\kappa \to 0$. While the Bansal-Spencer algorithm works only when $\alpha = O(\kappa^2)$, solutions do exist w.h.p. below $\alpha_c(\kappa)$ which, per (2), is asymptotically $\frac{1}{\log_2(1/\kappa)}$. Origins of this SCG were investigated in Gamarnik et al. (2022a), where it was shown that the SBP exhibits an intricate geometrical property called the *multi Overlap Gap Property* (m-OGP) when $\alpha = \Omega(\kappa^2 \log_2 \frac{1}{\kappa})$ and consequently *stable algorithms* fail to find a satisfying solution for $\alpha = \Omega(\kappa^2 \log_2 \frac{1}{\kappa})$. It is worth noting, though, that stable algorithms need not include online algorithms, which achieve the computational threshold for the SBP. What the limits of online algorithms are is an open question we undertake in this paper.

In addition to the SBP, the discrepancy minimization problem — in particular the algorithmic problem of efficiently finding a constant discrepancy solution when such solutions exist w.h.p. — also exhibits a similar SCG. To recall, when $\mathcal{M} \in \mathbb{R}^{M\times n}$, then constant discrepancy solutions exist w.h.p. as soon as $n = \Omega(M \log M)$. On the other hand, the best known polynomial-time algorithm succeeds at a dramatically smaller value $M = o(\log n)$ (Bansal, 2022), highlighting another striking $\texttt{SCG}$. This is our second focus in the present paper.

1.4. Main Results

Suppose $\mathcal{M}$ per (1) consists of i.i.d. $\mathcal{N}(0,1)$ entries. Our first main result establishes that online algorithms fail to find a satisfying solution for the $\texttt{SBP}$ at densities $\alpha = \Omega(\kappa^2)$.

Theorem 1 (Informal, see Theorem 9) *For densities $\alpha = \Omega(\kappa^2)$, online algorithms fail to find a solution for the $\texttt{SBP}$ w.p. greater than $e^{-\Theta(n)}$.*

Our next result extends Theorem 1 to the discrepancy minimization problem when $\mathcal{M} \in \mathbb{R}^{M\times n}$ consists of i.i.d. Rademacher or i.i.d. Bernoulli(p) entries.

Theorem 2 (Informal, see Theorems 10-11) *There exists $c > 0$ such that online algorithms fail to return a solution of discrepancy at most $c\sqrt{M}$ w.p. greater than $e^{-\Theta(M)}$.*

If the entries of $\mathcal{M}$ are Rademacher, taking $c = 1/24$ suffices. For Bernoulli case, the implied constant depends on p: it suffices to take $c \triangleq c_p = \sqrt{p-p^2}/24$. Taken together, Theorems 1 and 2 collectively yield that among the class of online algorithms, Bansal-Spencer algorithm (Bansal and Spencer, 2020) is optimal up to constants for both models. Our proof is based on a novel version of m-OGP: we show the non-existence of tuples of solutions agreeing on first $1-\Delta$ fraction of coordinates for a suitable $\Delta \in (0,1)$, for a collection of m correlated instances, see below for details. This barrier is more restricted than m-OGP, which asserts the non-existence of tuples of solutions at a prescribed distance. Additionally, for Theorem 2, one has to consider m-tuples with growing values of m, $m = \omega(1)$; this idea is originally due to Gamarnik and Kızıldağ (2021) for lowering the m-OGP threshold.

To the best of our knowledge, Theorems 1-2 are the first (up-to-constants) tight hardness guarantees via geometrical barriers against classes beyond stable algorithms, see Section 1.5 for details. Furthermore, unlike prior work (Gamarnik et al., 2020; Wein, 2020; Huang and Sellke, 2021; Gamarnik and Kızıldağ, 2021; Gamarnik et al., 2022a), the algorithms ruled out need not succeed w.h.p. or even with a constant probability: an exponentially small success probability suffices. This is made possible by using a clever application of Jensen's inequality, originally due to Gamarnik and Sudan (2017b).

A Technical Remark It is worth mentioning that a lower bound against online algorithms is important also from a technical point of view. Online algorithms need not be stable, so a hardness result for stable algorithms (via, e.g., the m-OGP) does not necessarily imply such a result for online algorithms. For instance, it is not known whether the algorithm by Bansal and Spencer (2020) is stable (in fact, it appears challenging to carry out a stability analysis due to the presence of a certain non-linearity, see Gamarnik et al. (2022a) for details), therefore the hardness result of Gamarnik et al. (2022a) does not apply to this algorithm. On the other hand, Theorems 1-2 yield a tight lower bound against this algorithm through a different geometrical barrier. Furthermore, yet another algorithm for the $\texttt{SBP}$ is due to Abbe et al. (2021a), whose stability is left open in Gamarnik et al. (2022a). It appears that this algorithm is 'almost' online, so the hardness result of Theorem 1 likely covers this algorithm as well, see the discussion following Theorem 9 for details.

Proof Sketch We sketch the proof of Theorem 1, which is based on a new version of m-OGP coupled with a contradiction argument. Suppose that such an online algorithm $\mathcal{A}$ with a success probability of p_s exists. Let $\mathcal{M}_1 \in \mathbb{R}^{M\times n}$ with i.i.d. $\mathcal{N}(0,1)$ entries. Fix an $m \in \mathbb{N}$ and a $\Delta \in (0,1)$, generate random matrices $\mathcal{M}_i \in \mathbb{R}^{M\times n}$, $2 \le i \le m$, by independently resampling the last Δn columns of $\mathcal{M}_1$. Running $\mathcal{A}$ on each $\mathcal{M}_i$, we obtain solutions $\boldsymbol{\sigma}_i \triangleq \mathcal{A}(\mathcal{M}_i) \in \Sigma_n$, $1 \le i \le m$. An application of Jensen's inequality then reveals $\|\mathcal{M}_i\boldsymbol{\sigma}_i\|_\infty \le \kappa\sqrt{n}$, $1 \le i \le m$, w.p. at least p_s^m. Furthermore, since $\mathcal{A}$ is online, it is the case that any $\boldsymbol{\sigma}_i$ and $\boldsymbol{\sigma}_j$ necessarily have identical first $n - \Delta n$ coordinates. Namely, if such an $\mathcal{A}$ exists, then w.p. at least p_s^m, there exists an m-tuple $(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m)$ of satisfying solutions that agree on first $n - \Delta n$ coordinates. We then establish, using the *first moment method*, that for suitably chosen m, Δ; the probability that such an m-tuple exists is in fact strictly less than p_s^m. This is a contradiction. The proof of Theorem 2 is similar, though it requires additional technical steps. In particular, one needs an anti-concentration argument for signed sums of binary variables via Berry-Esseen Theorem.

Our next focus is on the algorithmic problem of efficiently finding a constant discrepancy solution, given a random $\mathcal{M} \in \mathbb{R}^{M\times n}$. To recall, such solutions exist w.h.p. as soon as $n = \Omega(M \log M)$ (Altschuler and Niles-Weed, 2022), while the best known polynomial-time algorithm works only when $M = o(\log n)$ (Bansal, 2022). Further, it was conjectured in Altschuler and Niles-Weed (2022) that this task is algorithmically hard. Towards this conjecture, we focus on a smooth setting where the entries of $\mathcal{M}$ are i.i.d. $\mathcal{N}(0,1)$. Our next main result shows the presence of m-OGP with $m = O(1)$ when $n = \Theta(M \log M)$, giving a rigorous evidence of hardness at the 'boundary' $n = \Theta(M \log M)$.

Theorem 3 (Informal, see Theorem 13) *For $n = \Theta(M \log M)$, the set of constant discrepancy solutions exhibits m-OGP (with constant m) for suitably chosen parameters.*

The regime $\log n \ll M \ll n/\log n$ as well as extensions beyond Gaussian disorder—in particular to the Bernoulli or Rademacher case—are among the open problems we discuss in Section 1.6.

Our final main result leverages the m-OGP to show that *stable algorithms* fail to find a constant discrepancy solution when $n = \Theta(M \log M)$. Informally, an algorithm is stable if a small perturbation of its inputs induces only a small change in its output $\boldsymbol{\sigma}$, see Definition 14 for a formal statement. The class of stable algorithm has been shown to capture powerful classes of algorithms including low-degree polynomials (Gamarnik et al., 2020; Bresler and Huang, 2022), Approximate Message Passing (AMP) (Gamarnik and Jagannath, 2021), and Boolean circuits of low-depth (Gamarnik et al., 2021b).

Theorem 4 (Informal, see Theorem 15) *For $n = \Theta(M \log M)$, stable algorithms fail to find a constant discrepancy solution w.p. greater than a certain constant.*

The proof of Theorem 4 is based on a Ramsey-theoretic argument developed in Gamarnik and Kızıldağ (2021) and also used in Gamarnik et al. (2022a) coupled with the m-OGP result, Theorem 13; it rules out stable algorithms succeeding with a constant probability.

A Brief Summary of Main Results To navigate the reader, we recapitulate the algorithmic problems studied in this paper and summarize the content of each of our main algorithmic hardness results. Our focus is on three algorithmic questions:

- **P1**: Given $\alpha, \kappa > 0$ and an $\mathcal{M} \in \mathbb{R}^{M\times n}$ ($M = \lfloor n\alpha \rfloor$) with i.i.d. entries, find a satisfying solution to the `SBP`, i.e. a $\boldsymbol{\sigma} \in \Sigma_n$ such that $\|\mathcal{M}\boldsymbol{\sigma}\|_\infty \le \kappa\sqrt{n}$. When $\kappa \to 0$, such a $\boldsymbol{\sigma}$ exists when $\alpha = O_\kappa(\frac{1}{\log(1/\kappa)})$ (Perkins and Xu, 2021; Abbe et al., 2021b), whereas the best known polynomial-time algorithm works only when $\alpha = O_\kappa(\kappa^2)$ (Bansal and Spencer, 2020).

- **P2**: Given an $\mathcal{M} \in \mathbb{R}^{M\times n}$ with i.i.d. entries, find a $\boldsymbol{\sigma} \in \Sigma_n$ such that $\|\mathcal{M}\boldsymbol{\sigma}\|_\infty$ is small. When $n \ge M = \omega(1)$, the best known online algorithm[1] returns a solution of objective value $O(\sqrt{M})$ (Bansal and Spencer, 2020).
- **P3**: Given an $\mathcal{M} \in \mathbb{R}^{M\times n}$ with i.i.d. entries, find a $\boldsymbol{\sigma} \in \Sigma_n$ such that $\|\mathcal{M}\boldsymbol{\sigma}\|_\infty = O(1)$, whenever it exists (w.h.p.). Such a $\boldsymbol{\sigma}$ exists (w.h.p.) iff $n = \Omega(M \log M)$ (Altschuler and Niles-Weed, 2022), whereas the best known polynomial-time algorithm works only when $M = o(\log n)$ (Bansal, 2022).

Table 1: Summary of Main Results

Theorem	**Algorithmic Problem**	**Disorder**	**Hardness Against**	**At Value**
Theorem 9	P1	Gaussian	Online Algorithms	$\Omega(\kappa^2)$
Theorem 10	P2	Rademacher	Online Algorithms	$\Omega(\sqrt{M})$
Theorem 11	P2	Bernoulli	Online Algorithms	$\Omega(\sqrt{M})$
Theorem 15	P3	Gaussian	Stable Algorithms	$\Omega(M \log M)$

1.5. Background and Prior Work

Statistical-to-Computational Gaps (SCGs) Both the SBP and discrepancy minimization exhibit an SCG: known efficient algorithms perform strictly worse than the existential guarantee. Such gaps are a universal feature of many *average-case* algorithmic problems arising from random combinatorial structures and high-dimensional statistical inference. A partial list of problems with an SCG include random CSPs (Mézard et al., 2005; Achlioptas and Ricci-Tersenghi, 2006; Achlioptas and Coja-Oghlan, 2008; Gamarnik and Sudan, 2017b; Bresler and Huang, 2022), optimization over random graphs (Gamarnik and Sudan, 2014; Coja-Oghlan and Efthymiou, 2015; Wein, 2020), spin glasses (Gamarnik and Jagannath, 2021; Huang and Sellke, 2021), planted clique (Deshpande and Montanari, 2015; Barak et al., 2019), and tensor decomposition (Wein, 2022), see also the surveys Gamarnik (2021); Gamarnik et al. (2022b). Unfortunately, standard computational complexity theory is often useless due to the average-case nature of such problems[2]. Nevertheless, a very promising line of research proposed various frameworks that provide *rigorous evidence* of hardness. These frameworks include average-case reductions—often from the planted clique (Berthet and Rigollet, 2013; Brennan et al., 2018; Brennan and Bresler, 2019)—as well as unconditional lower bounds against restricted classes of algorithms, including the statistical query algorithms (Diakonikolas et al., 2017; Feldman et al., 2017, 2018), low-degree polynomials (LDP) (Hopkins, 2018; Kunisky et al., 2022; Wein, 2022), sum-of-squares hierarchy (Hopkins et al., 2015, 2017; Raghavendra et al., 2018; Barak et al., 2019), AMP (Zdeborová and Krzakala, 2016; Bandeira et al., 2018), and MCMC (Jerrum, 1992; Dyer et al., 2002). Yet another such approach is based on the intricate geometry of the solution space through the *Overlap Gap Property*.

[1] It is worth noting though that for P2, the class of online algorithms yield the best known polynomial-time algorithmic guarantee only in the proportional regime, $n = \Theta(M)$.

[2] Modulo a few exceptions, see e.g. Ajtai (1996); Boix-Adserà et al. (2021); Gamarnik and Kızıldağ (2021).

Intricate Geometry and the Overlap Gap Property (OGP) Prior work (Mézard et al., 2005; Achlioptas and Ricci-Tersenghi, 2006; Achlioptas and Coja-Oghlan, 2008) discovered a very intriguing connection between intricate geometry and algorithmic hardness in the context of random CSPs: the onset of algorithmic hardness roughly coincides with the emergence of an intricate geometry in the solution space. The OGP framework leverages insights from statistical physics to rigorously link the intricate geometry to formal hardness. In the context of random optimization, the OGP informally states that (w.h.p. over the randomness) any two near-optima are either 'close' or 'far': there exists $0 < \nu_1 < \nu_2 < 1$ such that $n^{-1}\langle \boldsymbol{\sigma}, \boldsymbol{\sigma}' \rangle \in [0, \nu_1] \cup [\nu_2, 1]$ for any pair of near-optima $\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \Sigma_n$. Namely, the region of normalized overlaps is *topologically disconnected*; no pairs of near-optima at *intermediate* distances can be found. The OGP is a rigorous barrier for large classes of algorithms exhibiting input stability—see below. See Gamarnik (2021) for a survey on OGP.

Algorithmic Implications of OGP The first work establishing and leveraging OGP to rule out algorithms is due to Gamarnik and Sudan (2014, 2017a). Their focus is on the problem finding a large independent set in sparse random graphs on n vertices with average degree d, which exhibits an SCG: the largest such independent set is of size $2\frac{\log d}{d}n$ (Frieze and Łuczak, 1992; Bayati et al., 2010), whereas the best known efficient algorithm finds an independent set of size $\frac{\log d}{d}n$. They establish that any pair of independent sets of size larger than $(1 + 1/\sqrt{2})\frac{\log d}{d}n$ exhibit the OGP. By leveraging the OGP, they then show that *local algorithms* fail to find an independent set of size larger than $(1 + 1/\sqrt{2})\frac{\log d}{d}n$. Subsequent research established and leveraged the OGP to rule out other classes of algorithms (e.g., AMP (Gamarnik and Jagannath, 2021), low-degree polynomials (Gamarnik et al., 2020; Wein, 2020; Bresler and Huang, 2022), Langevin dynamics (Gamarnik et al., 2020; Huang and Sellke, 2021), low-depth circuits (Gamarnik et al., 2021b)) for various other models (e.g., random graphs (Rahman and Virag, 2017; Gamarnik et al., 2020; Wein, 2020), spin glass models (Chen et al., 2019; Gamarnik and Jagannath, 2021; Huang and Sellke, 2021), random CSPs (Gamarnik and Sudan, 2017b; Bresler and Huang, 2022; Gamarnik et al., 2022a)). A very important feature found across the algorithms ruled out by the OGP and other versions of intricate geometry is input stability, similar to Definition 14 (apart from the failure of Monte Carlo Markov Chain methods in *planted* models, e.g. Jerrum (1992); Arous et al. (2020); Gamarnik et al. (2021a)). Our work marks the first instance of intricate geometry yielding tight algorithmic hardness against classes beyond stable algorithms.

Multi OGP (m-OGP) The prior work by Gamarnik and Sudan (2014) discussed above establish the failure of local algorithms at value $(1 + 1/\sqrt{2})\frac{\log d}{d}n$. By considering a certain overlap pattern involving many large independent sets, Rahman and Virag (2017) subsequently removed the additional $1/\sqrt{2}$ term; they showed that the onset of OGP precisely coincides with the algorithmic $\frac{\log d}{d}n$ value. That is, one can potentially lower the onset of the OGP and rule out algorithms for a broader range of parameters through more intricate overlap patterns. In a similar vein, Gamarnik and Sudan (2017b) studied the Not-All-Equal k-SAT problem and showed the presence of the OGP for the m-tuples of nearly equidistant satisfying assignments. Consequently, they obtained nearly tight hardness guarantees against sequential local algorithms. A similar m-OGP was also employed in Gamarnik and Kızıldağ (2021); Gamarnik et al. (2022a), and is also our focus here.

Recently, m-OGP for more intricate patterns were proposed. These forbidden patterns regard m-tuples of solutions where for any $2 \le i \le m$, the i^{th} solution has 'intermediate' overlap with the first $i - 1$ solutions. By doing so, tight hardness guarantees against low-degree polynomials were obtained for finding independent sets in sparse random graphs by Wein (2020) and for the random k-SAT by Bresler and Huang (2022). Similarly, Huang and Sellke (2021) construct a very intricate forbidden structure consisting of an ultrametric tree of solutions dubbed as the *Branching OGP*. By leveraging the branching OGP, they obtain tight hardness guarantees against Lipschitz algorithms for the p-spin model. Moreover, these papers establish the *Ensemble m-OGP* which regards m-tuples that are near-optimal w.r.t. correlated instances. The Ensemble OGP emerged in Chen et al. (2019); it has been instrumental in ruling out stable algorithms since. The investigation of m-tuples of solutions w.r.t. correlated instances is at the core of our paper. Furthermore, we inspect m-tuples with super-constant m, $m = \omega(1)$, to rule out online algorithms in Theorems 10-11. This idea originated in Gamarnik and Kızıldağ (2021) for further lowering the m-OGP threshold.

Online Setting We highlight that online algorithms are not only meant to be simply one algorithmic approach to solving a problem. Instead, the online setting is a very important computational model meant to address real-world decision making under uncertainty. This setting has been studied extensively in the literature, in particular in machine learning (Rakhlin et al., 2010, 2011a,b; Rakhlin and Sridharan, 2013) and convex optimization (Hazan et al., 2016). Moreover, the online setting is of great importance also in the discrepancy literature, see e.g. Bansal et al. (2020, 2021); Bansal and Spencer (2020) and the references therein. In particular, as we already mentioned, for both the SBP and the random discrepancy in the proportional regime, the best known polynomial-time algorithmic guarantee is online (Bansal and Spencer, 2020). Furthermore, many of the best known polynomial-time algorithms for optimization in random structures, including the largest independent set problem in a random graph and random constraint satisfaction problems (such as random k-SAT), are greedy algorithms, which are implemented in an online fashion and thus are special cases of online algorithms. Our results yield the first essentially tight unconditional lower bounds in the online setting; they also yield a toolkit for establishing lower bounds against the class of online algorithms for other average-case models exhibiting similar statistical-to-computational trade-offs and transfer to different settings such as *learning*.

1.6. Open Problems

Geometrical Barriers for other Classes of Algorithms Prior work on OGP showed that intricate geometry is a signature of algorithmic hardness, and gave lower bounds against stable algorithms. Theorems 9, 10 and 11 extend this beyond stable algorithms; they leverage intricate geometry to rule out online algorithms. It would be very interesting to rule out other classes of algorithms via similar geometrical barriers.

Discrepancy Minimization beyond Gaussian Disorder Theorem 13 shows that, for $\mathcal{M} \in \mathbb{R}^{M\times n}$ with i.i.d. $\mathcal{N}(0,1)$ entries and $n = \Theta(M \log M)$, the set of constant discrepancy solutions exhibits the m-OGP. A very interesting question is whether m-OGP still holds when the entries of $\mathcal{M}$ are binary. Prior work established OGP both for models with discrete disorder (e.g., random k-SAT (Gamarnik and Sudan, 2017b; Bresler and Huang, 2022), random graphs (Gamarnik and Sudan, 2014, 2017a; Gamarnik et al., 2020; Wein, 2020)) as well as for models with continuous disorder (e.g., spin glass models (Gamarnik and Jagannath, 2021; Huang and Sellke, 2021), number partitioning (Gamarnik and Kızıldağ, 2021), the SBP (Gamarnik et al., 2022a)). These results suggest that OGP exhibits universality: the distributional details of the disorder are immaterial[3]. In light of these, we make the following conjecture:

Conjecture 5 *For $\mathcal{M} \in \mathbb{R}^{M\times n}$ with i.i.d. Rademacher or Bernoulli(p) entries and $n = \Theta(M \log M)$, set of constant discrepancy solutions exhibits m-OGP with suitable parameters.*

Resolving Conjecture 5 may require understanding a probability term of the form $\mathbb{P}[M\boldsymbol{v} = \boldsymbol{x}]$ for a random $\boldsymbol{v}$ with i.i.d. binary entries and a deterministic $M \in \{-1,1\}^{m\times n}$ whose rows $\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m$ satisfy $n^{-1}\langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j\rangle = \beta$ for some fixed β and every $1 \le i < j \le m$. One direction is to employ local limit arguments; we leave this as an open problem for future investigation.

Discrepancy Minimization beyond $n = \Theta(M \log M)$ Recall that constant discrepancy solutions exist as soon as $n = \Theta(M\log M)$, i.e. when $M = O(n/\log n)$, while the best known polynomial-time algorithm works only when $M = o(\log n)$. In light of these, Theorem 13 provides rigorous evidence of hardness, yet only at the 'boundary'. The regime $\log n \ll M \ll n/\log n$ is an interesting direction left for future work. A potential avenue would be to consider a more intricate overlap pattern, such as those in Wein (2020); Huang and Sellke (2021) or the *branching OGP* (Huang and Sellke, 2021). To this end, we discover an intriguing phase transition (proof omitted):

Theorem 6 *Let $\mathcal{M} \in \mathbb{R}^{M\times n}$ with i.i.d. $\mathcal{N}(0,1)$ entries. Fix a $K > 0$ and let $S(m,\delta,K)$ be the set of $(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m)$ such that $d_H(\boldsymbol{\sigma}_i,\boldsymbol{\sigma}_j) = \delta$ and $\max_{i\le m}\|\mathcal{M}\boldsymbol{\sigma}_i\|_\infty \le K$.*

(a) If $M = \omega(\log n)$ then $\mathbb{E}\big[\big|S\big(n\log^{-O(1)} n, \log^{O(1)} n, K\big)\big|\big] = o(1)$.

(b) If $M = o(\log n)$, then $\mathbb{E}\big[\big|S\big(n\log^{-O(1)} n, \log^{O(1)} n, K\big)\big|\big] = \omega(1)$.

Namely, the value $M = \log n$ is the threshold at which the (expected) number of m-tuples of constant discrepancy solutions at distance $\log^{O(1)} n$ with $m = \widetilde{\Theta}(n)$ undergoes a phase transition. Whether this phase transition at value $\log n$ is coincidental or a signature of algorithmic hardness is an open problem left for future work.

Paper Organization and Notation The rest of the paper is organized as follows. In Section 2, we formalize the class of online algorithms and state our hardness results. Section 3 is devoted to the algorithmic problem of finding constant discrepancy solutions. See Section 3.1 for preliminaries and the definition of set of m-tuples regarding OGP, Section 3.2 for the main m-OGP result and Section 3.3 for the hardness result against stable algorithms. Finally, see Appendix A for complete proofs. Our notation is fairly standard, see the beginning of Appendix A for details.

2. Tight Hardness Guarantees for Online Algorithms

In this section, we explore the limits of *online algorithms* in the context of our two-models: the `SBP` and average-case discrepancy. We begin by formalizing the class of online algorithms in the context of these two models.

Definition 7

- **(`SBP`)** *Fix a $\kappa > 0$ and an $\alpha < \alpha_c(\kappa)$. Let $\mathcal{M} \in \mathbb{R}^{\alpha n\times n}$ with i.i.d. $\mathcal{N}(0,1)$ entries.*

[3] In fact, such a universality result has already been established for the `SBP`, see Gamarnik et al. (2022a, Theorem 5.2).

- **(Discrepancy)** *Fix $n \ge M$, let $\mathcal{M} \in \mathbb{R}^{M\times n}$ has i.i.d. Rademacher or Bernoulli(p) entries.*

Fix a $p_s > 0$ and a $K > 0$. An algorithm $\mathcal{A}$ is (p_s, α)-online for the `SBP` *or (p_s, K) online for discrepancy if it satisfies the following. Let $\mathcal{A}(\mathcal{M}) \triangleq \boldsymbol{\sigma}_{\rm ALG} = (\boldsymbol{\sigma}_{\rm ALG}(i) : 1 \le i \le n) \in \Sigma_n$.*

- **(Success)** *We have $\mathbb{P}\bigl[\bigl\|\mathcal{M}\boldsymbol{\sigma}_{\rm ALG}\bigr\|_\infty \le \kappa\sqrt{n}\bigr] \ge p_s$ (for the* `SBP`*) and $\mathbb{P}\bigl[\bigl\|\mathcal{M}\boldsymbol{\sigma}_{\rm ALG}\bigr\|_\infty \le K\bigr] \ge p_s$ (for discrepancy).*
- **(Online)** *Let $\mathcal{C}_1, \dots, \mathcal{C}_n$ be the columns of $\mathcal{M}$. There exists deterministic functions $f_1, \dots, f_n$ such that for $1 \le t \le n$, $\boldsymbol{\sigma}_{\rm ALG}(t) = f_t\bigl(\mathcal{C}_1, \dots, \mathcal{C}_t\bigr) \in \{-1, 1\}$.*

The parameter p_s is the success guarantee of the algorithm, where the probability is taken w.r.t. the randomness in $\mathcal{M}$. The online nature of the algorithm admits the following interpretation. Columns $\mathcal{C}_i$ arrive at a time. At the end of round $t-1$, the signs $\boldsymbol{\sigma}(i) \in \{-1,1\}$, $1 \le i \le t-1$ are assigned, and a new column $\mathcal{C}_t$ arrives. The sign $\boldsymbol{\sigma}(t)$ then depends only on the previous decisions $\boldsymbol{\sigma}(i)$, $1 \le i \le t-1$ and $\mathcal{C}_t$. That is, $\boldsymbol{\sigma}(t)$ depends only on $\mathcal{C}_i$, $1 \le i \le t$. This abstraction captures, in particular, the Bansal-Spencer algorithm:

Theorem 8 *(Bansal and Spencer, 2020, Theorem 3.4) Let $n \ge M$ and $\mathcal{M} \in \{-1,1\}^{M\times n}$ has i.i.d. Rademacher entries. Then, there exists absolute constants $C > 0$ and $\gamma < 1$, and an online algorithm $\mathcal{A}$ admitting $\mathcal{M}$ as its input and returning a $\boldsymbol{\sigma} \triangleq \mathcal{A}(\mathcal{M})$ such that*

$$\mathbb{P}\bigl[\bigl\|\mathcal{M}\boldsymbol{\sigma}\bigr\|_\infty \le C\sqrt{M}\bigr] \ge 1 - e^{-\Theta(M^\gamma)}.$$

Theorem 8 immediately yields an efficient algorithm for the `SBP` when $\alpha \le \kappa^2/C^2$, see Gamarnik et al. (2022a, Corollary 4.6). As mentioned in the introduction, the Bansal-Spencer algorithm is the best known polynomial-time algorithm both for the `SBP` and for the discrepancy minimization in random proportional regime. In the sense of Definition 7, it is a $(1 - e^{-\Theta(n^\gamma)}, \kappa^2/C^2)$-online algorithm for the `SBP` and a $(1 - e^{-\Theta(M^\gamma)}, C\sqrt{M})$-online algorithm for the discrepancy.

Online Algorithms for the `SBP` Our first main result focuses on the `SBP` in the regime $\kappa \to 0$ and establishes the following hardness for the class of online algorithms.

Theorem 9 *Fix any small enough $\kappa > 0$ and any $\alpha \ge 4\kappa^2$. Then there exists an $n_0 \in \mathbb{N}$ and an absolute constant $c > 0$ such that the following holds. For any $n \ge n_0$, there exists no (e^{-cn}, α)-online algorithm for the* `SBP` *in the sense of Definition 7.*

We prove Theorem 9 in Appendix A.1. Several remarks are in order.

Theorem 9 establishes that in the regime $\kappa \to 0$, online algorithms fail to find a satisfying solution for the `SBP` for densities $\alpha = \Omega(\kappa^2)$. This substantially improves upon an earlier result in Gamarnik et al. (2022a, Theorem 7.4), which showed the failure of online algorithms only when α is sufficiently close to the satisfiability threshold $\alpha_c(\kappa)$. Further, in light of Theorem 8, Theorem 9 is the sharpest possible: Bansal-Spencer algorithm is optimal (up to constants) among online algorithms; no online algorithm, in the sense of Definition 7, can improve upon it.

The algorithms Theorem 9 rules out need not succeed w.h.p. or even with a constant probability: an exponentially small success guarantee suffices. This is a particular strength of Theorem 9; we are unaware of any similar hardness guarantees for algorithms that succeed w.p. $o(1)$. This is based on a clever application of Jensen's inequality that is originally due to Gamarnik and Sudan (2017b).

We next remark on a polynomial-time algorithm devised by Abbe et al. (2021a) for the SBP. It is not clear whether this algorithm is stable in the sense of Definition 14 below (in fact, the stability of this algorithm is among the open questions raised in Gamarnik et al. (2022a)), therefore the hardness result in Gamarnik et al. (2022a) does not apply to this algorithm. However, an inspection of this algorithm reveals that it is in fact 'almost' online. The coordinates are iteratively determined in chunks that are not so large in size; the assignment is based on previous columns and aims at compensating for large discrepancies induced by the previous rounds. Consequently, it appears that Theorem 9 might also be made to apply to this algorithm modulo minor tweaks. (We thank the anonymous reviewer for this nice argument.)

Online Algorithms for the Discrepancy Minimization Our second main result extends Theorem 9 to discrepancy minimization for the case when the entries of $\mathcal{M}$ are binary.

Theorem 10 *Let $c < 1/2$ be arbitrary, $n \ge M = \omega(1)$, and $\mathcal{M} \in \{-1,1\}^{M\times n}$ with i.i.d. Rademacher entries. Then there exists an $M_0 \in \mathbb{N}$ such that the following holds. For every $M \ge M_0$, there exists no $\left(e^{-cM}, \sqrt{M}/24\right)$-online algorithm for discrepancy in the sense of Definition 7.*

Furthermore, Theorem 10 remains valid even when the entries of $\mathcal{M}$ are Bernoulli(p).

Theorem 11 *Let $c < 1/2$ be arbitrary, $n \ge M = \omega(1)$, and $\mathcal{M} \in \{0,1\}^{M\times n}$ with i.i.d. Bernoulli(p) entries. Then there exists an $M_0 \in \mathbb{N}$ such that the following holds. For every $(p-p^2)M \ge M_0$, there exists no $\left(e^{-cM}, \sqrt{M(p-p^2)}/24\right)$-online algorithm for discrepancy in the sense of Definition 7.*

We prove Theorem 10 in Appendix A.2 and give the extension to Theorem 11 in Appendix A.3.

Theorems 10-11 collectively establish that up to constant factors the Bansal-Spencer algorithm is optimal within the class of online algorithms for the discrepancy minimization problem. Once again, the algorithms ruled out can succeed even with an exponentially small probability.

At a technical level, Theorems 10-11 are established by showing the non-existence of certain m-tuples of solutions described earlier with growing values of m, $m = \omega_M(1)$. The idea of considering the 'landscape' of m-tuples with $m = \omega(1)$ was introduced in the context of random number partitioning problem (Gamarnik and Kızıldağ, 2021). By doing so, the authors subsequently lowered the m-OGP threshold and ruled out stable algorithms for a broader range of parameters than what one can get for constant m. Ours is the first work leveraging such a barrier with growing values of m beyond stable algorithms; it further illustrates the potential gain of considering super-constant tuples for random computational problems. Another key ingredient of our proof is an anti-concentration inequality for signed sum of Bernoulli/Rademacher variables, via the Berry-Esseen Theorem.

3. Algorithmic Barriers in Finding Constant Discrepancy Solutions

In this section, we focus on the algorithmic problem of finding a constant discrepancy solution. More concretely, given a random $\mathcal{M} \in \mathbb{R}^{M\times n}$ we ask the following question: for what values of M and n, can a solution $\boldsymbol{\sigma} \in \Sigma_n$ of constant discrepancy, $\|\mathcal{M}\boldsymbol{\sigma}\|_\infty = O(1)$, be found efficiently?

To begin with, a simple first-moment calculation shows that $n = \Omega(M \log M)$ is necessary for such solutions to exist. This condition turns out to be sufficient, as well; Altschuler and Niles-Weed (2022) showed that if $\mathcal{M}$ has i.i.d. Bernoulli(p) entries then $\mathcal{D}(\mathcal{M}) \le 1$ w.h.p. if $n \ge CM\log M$ where C is any arbitrary constant greater than $(2\log 2)^{-1}$. On the other hand, the best known polynomial-time algorithm finding such a solution works only when $M = o(\log n)$ (Bansal, 2022). This highlights a striking statistical-to-computational gap (SCG).

In this section, we study the nature of this SCG in a smooth setting where the entries of $\mathcal{M}$ are i.i.d. $\mathcal{N}(0,1)$, near the existential boundary $n = \Theta(M\log M)$. We first focus on the 'landscape' of the set of constant discrepancy solutions, and show the presence of Ensemble m-OGP, an intricate geometrical property. We then leverage m-OGP to rule out the class of stable algorithms.

3.1. Technical Preliminaries

We formalize the set of tuples of constant discrepancy solutions under investigation.

Definition 12 *Fix a $K > 0$, an $m \in \mathbb{N}$, $0 < \eta < \beta < 1$, and $\mathcal{I} \subset [0,\pi/2]$. Let $\mathcal{M}_i \in \mathbb{R}^{M\times n}$, $0 \le i \le m$, be i.i.d. random matrices, each having i.i.d. $\mathcal{N}(0,1)$ entries. Denote by $\mathcal{S}(K,m,\beta,\eta,\mathcal{I})$ the set of all m-tuples $\boldsymbol{\sigma}_i \in \Sigma_n$, $1 \le i \le m$, satisfying the following:*

- **(Pairwise Overlap Condition)** *For any $1 \le i < j \le m$, $\beta - \eta \le n^{-1}\langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j\rangle \le \beta$.*
- **(Constant Discrepancy Condition)** *There exists $\tau_1,\dots,\tau_m \in \mathcal{I}$ such that $\max_{1\le i\le m}\big\|\mathcal{M}_i(\tau_i)\boldsymbol{\sigma}_i\big\|_\infty \le K$, where $\mathcal{M}_i(\tau_i) = \cos(\tau_i)\mathcal{M}_0 + \sin(\tau_i)\mathcal{M}_i \in \mathbb{R}^{M\times n}$.*

Definition 12 concerns tuples of solutions of discrepancy at most K. The parameter m is the size of tuples under consideration, and β and η collectively control the (forbidden) region of overlaps. Finally, the set $\mathcal{I}$ is employed for generating correlated instances; this is necessary for establishing the Ensemble m-OGP to rule out stable algorithms, see below for details.

3.2. Ensemble m-OGP in Discrepancy Minimization

Our next main result shows that the set of constant discrepancy solutions exhibits the m-OGP.

Theorem 13 *Fix arbitrary constants $C_1 > c_2 > 0$ and a $K > 0$, suppose that $C_1 M\log_2 M \ge n \ge c_2 M\log_2 M$. Then, there exists an $m \in \mathbb{N}$, a $c > 0$ and $0 < \eta < \beta < 1$ such that the following holds. Fix any $\mathcal{I} \subset [0,\pi/2]$ with $|\mathcal{I}| \le 2^{cn}$. Then, $\mathbb{P}\big[\mathcal{S}(K,m,\beta,\eta,\mathcal{I}) \ne \varnothing\big] \le 2^{-\Theta(n)}$.*

We prove Theorem 13 in Appendix A.4. Several remarks are in order. Theorem 13 shows that for any $K > 0$ and throughout the entire regime $n = \Theta(M\log M)$, the set of solutions with discrepancy at most K exhibits the m-OGP, for suitable m, β and η. In light of prior work discussed earlier, this gives some rigorous evidence for algorithmic hardness at the boundary $n = \Theta(M\log M)$, and constitutes a first step towards Altschuler and Niles-Weed (2022, Conjecture 1).

Our proof is based on the first moment method: we show that the expected number of such m-tuples is exponentially small for suitably chosen m, β, η and apply Markov's inequality. Further, our proof reveals that $\beta \gg \eta$: Theorem 13 rules out m-tuples of constant discrepancy solutions that are nearly equidistant. Moreover, the solutions need not be of constant discrepancy w.r.t. the same instance: $\mathcal{M}_i(\tau_i)$ appearing in Definition 12 are potentially correlated. This is known as the Ensemble m-OGP and instrumental in ruling out stable algorithms in Theorem 15.

3.3. m-OGP Implies Failure of Stable Algorithms

In this section, we show that the Ensemble m-OGP established in Theorem 13 implies the failure of *stable algorithms* in finding a constant discrepancy solution. We begin by elaborating on the algorithmic setting and formalizing the class of stable algorithms we investigate.

Algorithmic Setting An algorithm $\mathcal{A}$ is a mapping between $\mathbb{R}^{M\times n}$ and Σ_n, where randomization is allowed: we assume there exists a probability space $(\Omega, \mathbb{P}_\omega)$ such that $\mathcal{A} : \mathbb{R}^{M\times n} \times \Omega \to \Sigma_n$. For any $\omega \in \Omega$ and $\mathcal{M} \in \mathbb{R}^{M\times n}$, we want $\|\mathcal{M}\boldsymbol{\sigma}_{\rm ALG}\|_\infty = O(1)$, where $\boldsymbol{\sigma}_{\rm ALG} = \mathcal{A}(\mathcal{M}, \omega) \in \Sigma_n$. The class of stable algorithms is formalized as follows.

Definition 14 *Fix a $K > 0$. An algorithm $\mathcal{A} : \mathbb{R}^{M\times n} \times \Omega \to \Sigma_n$ is called $(K, \rho, p_f, p_{\rm st}, f, L)$-stable (for discrepancy minimization) if it satisfies the following for all sufficiently large M.*

- **(Success)** *For $\mathcal{M}$ with i.i.d. $\mathcal{N}(0, 1)$ entries, $\mathbb{P}_{(\mathcal{M},\omega)}\left[\left\|\mathcal{M}\mathcal{A}(\mathcal{M}, \omega)\right\|_\infty \le K\right] \ge 1 - p_f$.*
- **(Stability)** *Let $\mathcal{M}, \overline{\mathcal{M}} \in \mathbb{R}^{M\times n}$ be random matrices, each with i.i.d. $\mathcal{N}(0, 1)$ entries, such that $\mathbb{E}\left[\mathcal{M}_{ij}\overline{\mathcal{M}}_{ij}\right] = \rho$ for $1 \le i \le M$ and $1 \le j \le n$. Then,*

$$\mathbb{P}_{(\mathcal{M},\overline{M},\omega)}\left[d_H\big(\mathcal{A}(\mathcal{M}, \omega), \mathcal{A}(\overline{\mathcal{M}}, \omega)\big) \le f + L\|\mathcal{M} - \overline{\mathcal{M}}\|_F\right] \ge 1 - p_{\rm st}.$$

Definition 14 is the same as Gamarnik et al. (2022a, Definition 3.1). W.p. at least $1 - p_f$, $\mathcal{A}$ finds a solution of discrepancy below K. $\mathcal{A}$ can tolerate an input correlation value of ρ; and the parameters f and L quantify the sensitivity of the output of $\mathcal{A}$ to changes in its input. The stability guarantee is probabilistic—w.r.t. both $\mathcal{M}, \overline{\mathcal{M}}$ and to the randomness ω of $\mathcal{A}$—holding w.p. at least $1 - p_{\rm st}$. Finally, the term f makes our negative result only stronger: $\mathcal{A}$ is allowed to make f flips even when $\mathcal{M}$ and $\overline{\mathcal{M}}$ are 'too close'. Our final main result is as follows.

Theorem 15 *Fix a $K > 0$, $C_1 > c_2 > 0$ and a $\mathcal{L} > 0$. Suppose $C_1 M \log_2 M \ge n \ge c_2 M \log_2 M$. Let $m \in \mathbb{N}$ and $0 < \eta < \beta < 1$ be the m-OGP parameters prescribed by Theorem 13. Set*

$$C = \frac{\eta^2}{1600}, \quad Q = \frac{4800\mathcal{L}\pi}{\eta^2}, \quad \text{and} \quad T = \exp_2\left(2^{4mQ\log_2 Q}\right). \tag{4}$$

Then, there exists an $n_0 \in \mathbb{N}$ such that the following holds. For every $n \ge n_0$, there exists no randomized algorithm $\mathcal{A} : \mathbb{R}^{M\times n} \times \Omega \to \Sigma_n$ which, in the sense of Definition 14, is

$$\left(K, \cos\left(\frac{\pi}{2Q}\right), \frac{1}{9(Q+1)T}, \frac{1}{9Q(T+1)}, Cn, \mathcal{L}\sqrt{\frac{n}{M}}\right) - \textit{stable}.$$

The proof of Theorem 15 is almost identical to that of Gamarnik et al. (2022a, Theorem 3.2), and omitted for brevity. Several remarks are in order.

Firstly, there is no restriction on the running time of $\mathcal{A}$: Theorem 15 rules out any $\mathcal{A}$ that is stable with suitable parameters in the sense of Definition 14. Secondly, observe that $m, \beta, \eta, \mathcal{L}$ are all $O(1)$ (as $n \to \infty$). Hence, C, Q, T per (4) are all $O(1)$, as well. This is an important feature of our result: Theorem 15 rules out algorithms with a constant success/stability guarantee. Lastly, since $C = O(1)$, $\mathcal{A}$ is still allowed to make $\Theta(n)$ bit flips even when $\mathcal{M}$ and $\overline{\mathcal{M}}$ are nearly identical.

Acknowledgments

We thank the anonymous reviewers for their helpful feedback. The first author is supported in part by NSF grant DMS-2015517. The second author is supported by a Columbia University, Distinguished Postdoctoral Fellowship in Statistics. The third author is supported in part by NSF grant DMS-1847451.

References

Emmanuel Abbe, Shuangping Li, and Allan Sly. Binary perceptron: efficient algorithms can find solutions in a rare well-connected cluster. *arXiv preprint arXiv:2111.03084*, 2021a.

Emmanuel Abbe, Shuangping Li, and Allan Sly. Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron. *arXiv preprint arXiv:2102.13069*, 2021b.

Dimitris Achlioptas and Amin Coja-Oghlan. Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 793–802. IEEE, 2008.

Dimitris Achlioptas and Federico 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*, pages 130–139, 2006.

Miklós Ajtai. Generating hard instances of lattice problems. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, pages 99–108, 1996.

Ahmed El Alaoui and Mark Sellke. Algorithmic pure states for the negative spherical perceptron. *arXiv preprint arXiv:2010.15811*, 2020.

Dylan J Altschuler. Fluctuations of the symmetric perceptron. *arXiv preprint arXiv:2205.02319*, 2022.

Dylan J Altschuler and Jonathan Niles-Weed. The discrepancy of random rectangular matrices. *Random Structures & Algorithms*, 60(4):551–593, 2022.

Gérard Ben Arous, Alexander S Wein, and Ilias Zadik. Free energy wells and overlap gap property in sparse pca. In *Conference on Learning Theory*, pages 479–482. PMLR, 2020.

Benjamin Aubin, Will Perkins, and Lenka Zdeborová. Storage capacity in symmetric binary perceptrons. *Journal of Physics A: Mathematical and Theoretical*, 52(29):294003, 2019.

Carlo Baldassi, Alfredo Braunstein, Nicolas Brunel, and Riccardo Zecchina. Efficient supervised learning in networks with binary synapses. *Proceedings of the National Academy of Sciences*, 104(26):11079–11084, 2007.

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.

Carlo Baldassi, Riccardo Della Vecchia, Carlo Lucibello, and Riccardo Zecchina. Clustering of solutions in the symmetric binary perceptron. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(7):073303, 2020.

Afonso S Bandeira, Amelia Perry, and Alexander S Wein. Notes on computational-to-statistical gaps: predictions using statistical physics. *Portugaliae Mathematica*, 75(2):159–186, 2018.

Nikhil Bansal. Constructive algorithms for discrepancy minimization. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 3–10. IEEE, 2010.

Nikhil Bansal. personal communication, 2022.

Nikhil Bansal and Joel H. Spencer. On-line balancing of random inputs. *Random Structures and Algorithms*, 57(4):879–891, December 2020. ISSN 1042-9832. doi: 10.1002/rsa.20955.

Nikhil Bansal, Haotian Jiang, Sahil Singla, and Makrand Sinha. Online vector balancing and geometric discrepancy. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1139–1152, 2020.

Nikhil Bansal, Haotian Jiang, Raghu Meka, Sahil Singla, and Makrand Sinha. Online discrepancy minimization for stochastic arrivals. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2842–2861. SIAM, 2021.

Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.

Mohsen Bayati, David Gamarnik, and Prasad 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*, pages 105–114, 2010.

Quentin Berthet and Philippe Rigollet. Computational lower bounds for sparse PCA. *arXiv preprint arXiv:1304.0828*, 2013.

Enric Boix-Adserà, Matthew Brennan, and Guy Bresler. The average-case complexity of counting cliques in Erdös–Rényi hypergraphs. *SIAM Journal on Computing*, (0):FOCS19–39, 2021.

Erwin Bolthausen, Shuta Nakajima, Nike Sun, and Changji Xu. Gardner formula for Ising perceptron models at small densities. *arXiv preprint arXiv:2111.02855*, 2021.

Matthew Brennan and Guy Bresler. Optimal average-case reductions to sparse pca: From weak assumptions to strong hardness. *arXiv preprint arXiv:1902.07380*, 2019.

Matthew Brennan, Guy Bresler, and Wasim Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. *arXiv preprint arXiv:1806.07508*, 2018.

Guy Bresler and Brice Huang. The algorithmic phase transition of random k-sat for low degree polynomials. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 298–309. IEEE, 2022.

Wei-Kuo Chen, David Gamarnik, Dmitry Panchenko, and Mustazee Rahman. Suboptimality of local algorithms for a class of max-cut problems. *The Annals of Probability*, 47(3):1587–1618, 2019.

Amin Coja-Oghlan and Charilaos Efthymiou. On independent sets in random graphs. *Random Structures & Algorithms*, 47(3):436–486, 2015.

Kevin P Costello. Balancing gaussian vectors. *Israel Journal of Mathematics*, 172(1):145–156, 2009.

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.

Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *Conference on Learning Theory*, pages 523–562, 2015.

Ilias Diakonikolas, Daniel M Kane, and Alistair 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)*, pages 73–84. IEEE, 2017.

Jian Ding and Nike Sun. Capacity lower bound for the Ising perceptron. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 816–827, 2019.

Martin Dyer, Alan Frieze, and Mark Jerrum. On counting independent sets in sparse graphs. *SIAM Journal on Computing*, 31(5):1527–1541, 2002.

Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM (JACM)*, 64(2): 1–37, 2017.

Vitaly Feldman, Will Perkins, and Santosh Vempala. On the complexity of random satisfiability problems with planted solutions. *SIAM Journal on Computing*, 47(4):1294–1338, 2018.

Cole Franks and Michael Saks. On the discrepancy of random matrices with many columns. *Random Structures & Algorithms*, 57(1):64–96, 2020.

Alan M Frieze and T Łuczak. On the independence and chromatic numbers of random regular graphs. *Journal of Combinatorial Theory, Series B*, 54(1):123–132, 1992.

David Gamarnik. The overlap gap property: A topological barrier to optimizing over random structures. *Proceedings of the National Academy of Sciences*, 118(41), 2021.

David Gamarnik and Aukosh Jagannath. The overlap gap property and approximate message passing algorithms for p-spin models. *The Annals of Probability*, 49(1):180–205, 2021.

David Gamarnik and Eren C Kızıldağ. Algorithmic obstructions in the random number partitioning problem. *arXiv preprint arXiv:2103.01369*, 2021.

David Gamarnik and Eren C. Kızıldağ. Computing the partition function of the Sherrington–Kirkpatrick model is hard on average. *The Annals of Applied Probability*, 31(3):1474 – 1504, 2021. doi: 10.1214/20-AAP1625. URL https://doi.org/10.1214/20-AAP1625.

David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. In *Proceedings of the 5th conference on Innovations in theoretical computer science*, pages 369–376, 2014.

David Gamarnik and Madhu Sudan. Limits of local algorithms over sparse random graphs. *Ann. Probab.*, 45(4):2353–2376, 07 2017a. doi: 10.1214/16-AOP1114. URL https://doi.org/10.1214/16-AOP1114.

David Gamarnik and Madhu Sudan. Performance of sequential local algorithms for the random NAE-K-SAT problem. *SIAM Journal on Computing*, 46(2):590–619, 2017b.

David Gamarnik, Aukosh Jagannath, and Alexander S Wein. Low-degree hardness of random optimization problems. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 131–140. IEEE, 2020.

David Gamarnik, Aukosh Jagannath, and Subhabrata Sen. The overlap gap property in principal submatrix recovery. *Probability Theory and Related Fields*, 181:757–814, 2021a.

David Gamarnik, Aukosh Jagannath, and Alexander S Wein. Circuit lower bounds for the p-spin optimization problem. *arXiv preprint arXiv:2109.01342*, 2021b.

David Gamarnik, Eren C Kızıldağ, Will Perkins, and Changji Xu. Algorithms and barriers in the symmetric binary perceptron model. *arXiv preprint arXiv:2203.15667*, 2022a.

David Gamarnik, Cristopher Moore, and Lenka Zdeborová. Disordered systems insights on computational hardness. *Journal of Statistical Mechanics: Theory and Experiment*, 2022(11):114015, 2022b.

Elizabeth Gardner. Maximum storage capacity in neural networks. *EPL (Europhysics Letters)*, 4 (4):481, 1987.

Elizabeth Gardner. The space of interactions in neural network models. *Journal of physics A: Mathematical and general*, 21(1):257, 1988.

Elizabeth Gardner and Bernard Derrida. Optimal storage properties of neural network models. *Journal of Physics A: Mathematical and general*, 21(1):271, 1988.

Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016.

Rebecca Hoberg and Thomas Rothvoss. A fourier-analytic approach for the discrepancy of random set systems. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2547–2556. SIAM, 2019.

AJ Hoffman and HW Wielandt. The variation of the spectrum of a normal matrix. *Duke Mathematical Journal*, 20(1):37–39, 1953.

Samuel B Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In *Conference on Learning Theory*, pages 956–1006, 2015.

Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 720–731. IEEE, 2017.

Samuel Brink Klevit Hopkins. Statistical inference and the sum of squares method. 2018.

Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge University Press, 2012.

Brice Huang and Mark Sellke. Tight lipschitz hardness for optimizing mean field spin glasses. *arXiv preprint arXiv:2110.07847*, 2021.

Haiping Huang and Yoshiyuki Kabashima. Origin of the computational hardness for learning with binary synapses. *Physical Review E*, 90(5):052813, 2014.

Mark Jerrum. Large cliques elude the metropolis process. *Random Structures & Algorithms*, 3(4): 347–359, 1992.

Roger David Joseph and Louise Hay. The number of orthants in n-space intersected by an s-dimensional subspace. Technical report, CORNELL AERONAUTICAL LAB INC BUFFALO NY, 1960.

Narendra Karmarkar and Richard M Karp. *The differencing method of set partitioning*. Computer Science Division (EECS), University of California Berkeley, 1982.

Narendra Karmarkar, Richard M Karp, George S Lueker, and Andrew M Odlyzko. Probabilistic analysis of optimum partitioning. *Journal of Applied probability*, 23(3):626–645, 1986.

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.

Eren C Kızıldağ and Tanay Wakhare. Symmetric perceptron with random labels. *In submission*, 2023.

Eren C Kızıldağ. *Algorithms and Algorithmic Barriers in High-Dimensional Statistics and Random Combinatorial Structures*. PhD thesis, Massachusetts Institute of Technology, 2022.

Werner Krauth and Marc Mézard. Storage capacity of memory networks with binary couplings. *Journal de Physique*, 50(20):3057–3066, 1989.

Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *Mathematical Analysis, its Applications and Computation: ISAAC 2019, Aveiro, Portugal, July 29–August 2*, pages 1–50. Springer, 2022.

Avi Levy, Harishchandra Ramadas, and Thomas Rothvoss. Deterministic discrepancy minimization via the multiplicative weight update method. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 380–391. Springer, 2017.

Shachar Lovett and Raghu Meka. Constructive discrepancy minimization by walking on the edges. *SIAM Journal on Computing*, 44(5):1573–1582, 2015.

Jiri Matousek. *Geometric discrepancy: An illustrated guide*, volume 18. Springer Science & Business Media, 1999.

Marc Mézard, Thierry Mora, and Riccardo Zecchina. Clustering of solutions in the random satisfiability problem. *Physical Review Letters*, 94(19):197205, 2005.

Shuta Nakajima and Nike Sun. Sharp threshold sequence and universality for ising perceptron models. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 638–674. SIAM, 2023.

Will Perkins and Changji Xu. Frozen 1-RSB structure of the symmetric Ising perceptron. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1579–1588, 2021.

Aditya Potukuchi. Discrepancy in random hypergraph models. *arXiv preprint arXiv:1811.01491*, 2018.

Prasad Raghavendra, Tselil Schramm, and David Steurer. High-dimensional estimation via sum-of-squares proofs. *arXiv preprint arXiv:1807.11419*, 6, 2018.

Mustazee Rahman and Balint Virag. Local algorithms for independent sets are half-optimal. *The Annals of Probability*, 45(3):1543–1577, 2017.

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. In *Conference on Learning Theory*, pages 993–1019. PMLR, 2013.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: Random averages, combinatorial parameters, and learnability. *Advances in Neural Information Processing Systems*, 23, 2010.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: Beyond regret. In *Proceedings of the 24th Annual Conference on Learning Theory*, pages 559–594. JMLR Workshop and Conference Proceedings, 2011a.

Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning: Stochastic, constrained, and smoothed adversaries. *Advances in neural information processing systems*, 24, 2011b.

Thomas Rothvoss. Constructive discrepancy minimization for convex sets. *SIAM Journal on Computing*, 46(1):224–234, 2017.

Ashwin Sah and Mehtaab Sawhney. Distribution of the threshold for the symmetric perceptron. *arXiv preprint arXiv:2301.10701*, 2023.

Mariya Shcherbina and Brunello Tirozzi. Rigorous solution of the Gardner problem. *Communications in mathematical physics*, 234(3):383–422, 2003.

Zbynek Sidák. On multivariate normal probabilities of rectangles: their dependence on correlations. *The Annals of Mathematical Statistics*, 39(5):1425–1434, 1968.

Joel Spencer. Six standard deviations suffice. *Transactions of the American mathematical society*, 289(2):679–706, 1985.

Mihailo Stojnic. Another look at the Gardner problem. *arXiv preprint arXiv:1306.3979*, 2013.

Michel Talagrand. Intersecting random half cubes. *Random Structures & Algorithms*, 15(3-4): 436–449, 1999.

Michel Talagrand. *Mean Field Models for Spin Glasses: Advanced replica-symmetry and low temperature*. Springer, 2011.

Paxton Turner, Raghu Meka, and Philippe Rigollet. Balancing Gaussian vectors in high dimension. In *Conference on Learning Theory*, pages 3455–3486. PMLR, 2020.

Alexander S Wein. Optimal low-degree hardness of maximum independent set. *arXiv preprint arXiv:2010.06563*, 2020.

Alexander S Wein. Average-case complexity of tensor decomposition for low-degree polynomials. *arXiv preprint arXiv:2211.05274*, 2022.

James G Wendel. A problem in geometric probability. *Mathematica Scandinavica*, 11(1):109–111, 1962.

Robert O Winder. Single stage threshold logic. In *2nd Annual Symposium on Switching Circuit Theory and Logical Design (SWCT 1961)*, pages 321–332. IEEE, 1961.

Changji Xu. Sharp threshold for the ising perceptron model. *The Annals of Probability*, 49(5): 2399–2415, 2021.

Benjamin Yakir. The differencing algorithm ldm for partitioning: a proof of a conjecture of karmarkar and karp. *Mathematics of Operations Research*, 21(1):85–99, 1996.

Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.

Appendix A. Proofs

Additional Notation We commence this section with an additional list of notation. For any set A, $|A|$ denotes its cardinality. Given any event E, denote its indicator by $\mathbb{1}\{E\}$. For any $v = (v(i) : 1 \le i \le n) \in \mathbb{R}^n$ and $p > 0$, $\|v\|_p = \left(\sum_{1\le i\le n} |v(i)|^p\right)^{1/p}$ and $\|v\|_\infty = \max_{1\le i\le n} |v(i)|$. For $v, v' \in \mathbb{R}^n$, $\langle v, v'\rangle \triangleq \sum_{1\le i\le n} v(i)v'(i)$. For any $\boldsymbol{\sigma}, \boldsymbol{\sigma'} \in \Sigma_n \triangleq \{-1,1\}^n$, $d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma'})$ denotes their Hamming distance: $d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma'}) \triangleq \sum_{1\le i\le n} \mathbb{1}\{\boldsymbol{\sigma}(i) \ne \boldsymbol{\sigma'}(i)\}$. For any $r > 0$, $\log_r(\cdot)$ and $\exp_r(\cdot)$ denote, respectively, the logarithm and exponential functions base r; when $r = e$, we omit the subscript. For $p \in [0,1]$, $h_b(p) \triangleq -p\log_2 p - (1-p)\log_2(1-p)$. Denote by I_k the $k \times k$ identity matrix, and by $\mathbf{1}$ the vector of all ones whose dimension will be clear from the context. Given $\boldsymbol{\mu} \in \mathbb{R}^k$ and $\Sigma \in \mathbb{R}^{k\times k}$, denote by $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ the multivariate normal distribution in $\mathbb{R}^k$ with mean $\boldsymbol{\mu}$ and covariance Σ. Given a matrix $\mathcal{M}$, $\|\mathcal{M}\|_F$, $\|\mathcal{M}\|_2$, and $|\mathcal{M}|$ denote, respectively, the Frobenius norm, the spectral norm, and the determinant of $\mathcal{M}$.

We employ standard Bachmann-Landau asymptotic notation throughout, e.g. $\Theta(\cdot)$, $O(\cdot)$, $o(\cdot)$, $\Omega(\cdot)$, where the underlying asymptotics will often be clear from the context. In certain cases where a confusion is possible, we reflect the underlying asymptotics as a subscript, e.g. $\Theta_\kappa(\cdot)$. All floor/ceiling operators are omitted for the sake of simplicity.

A.1. Proof of Theorem 9

Let $\kappa, \alpha > 0$, $M = n\alpha$, $m \in \mathbb{N}$, and $\Delta \in (0, \frac{1}{2})$. Suppose $\mathcal{M}_1 \in \mathbb{R}^{M\times n}$ has i.i.d. $\mathcal{N}(0,1)$ entries and let $\mathcal{M}_2, \dots, \mathcal{M}_m \in \mathbb{R}^{M\times n}$ be random matrices obtained from $\mathcal{M}_1$ by independently resampling the last Δn columns of $\mathcal{M}_1$. Denote by $\Xi(m, \Delta)$ the set of all m-tuples satisfying the following:

- $\max_{1\le i\le m} \left\|\mathcal{M}_i \boldsymbol{\sigma}_i\right\|_\infty \le \kappa\sqrt{n}$.
- For $1 \le i < j \le n$ and $1 \le k \le n - \Delta n$, $\boldsymbol{\sigma}_i(k) = \boldsymbol{\sigma}_j(k)$.

We establish the following proposition.

Proposition 16 *Fix any $\kappa > 0$ small enough and let $\alpha \ge 4\kappa^2$. Then, there exists an $m \in \mathbb{N}$ and a $\Delta \in (0, 1/2)$ such that*

$$\mathbb{P}\big[\Xi(m, \Delta) = \varnothing\big] \ge 1 - e^{-\Theta(n)}.$$

We first assume Proposition 16 and show how to deduce Theorem 9. Fix a $c > 0$ and suppose, for the sake of contradiction, that an (e^{-cn}, α)-online $\mathcal{A}$ exists. For $\mathcal{M}_1, \dots, \mathcal{M}_m \in \mathbb{R}^{M\times n}$ described above, set

$$\boldsymbol{\sigma}_i \triangleq \mathcal{A}(\mathcal{M}_i) \in \Sigma_n, \quad 1 \le i \le m. \tag{5}$$

Note that for any $1 \le i < j \le m$, the first $n - \Delta n$ columns of $\mathcal{M}_i$ and $\mathcal{M}_j$ are identical. Consequently,

$$\boldsymbol{\sigma}_i(k) = \boldsymbol{\sigma}_j(k) \quad \text{for} \quad 1 \le i < j \le m \quad \text{and} \quad 1 \le k \le n - \Delta n.$$

Next, we establish the following probability guarantee.

Lemma 17

$$\mathbb{P}\left[\max_{1\le i\le m} \left\|\mathcal{M}_i \boldsymbol{\sigma}_i\right\|_\infty \le \kappa\sqrt{n}\right] \ge p_s^m.$$

Proof Our argument is based on a clever application of Jensen's inequality, due to Gamarnik and Sudan (2017b, Lemma 5.3). Denote by ζ the first $(1-\Delta)n$ columns of $\mathcal{M}_1$. That is, ζ is the 'common randomness' shared by $\mathcal{M}_1, \dots, \mathcal{M}_m$. Set

$$I_i = \mathbb{1}\big\{\|\mathcal{M}_i \boldsymbol{\sigma}_i\|_\infty \le \kappa\sqrt{n}\big\}.$$

Then,

$$\mathbb{P}\left[\max_{1\le i\le m} \left\|\mathcal{M}_i \sigma_i\right\|_\infty \le \kappa\sqrt{n}\right] = \mathbb{E}\big[I_1 \cdots I_m\big].$$

We then complete the proof by noticing

$$\mathbb{E}\big[I_1 \cdots I_m\big] = \mathbb{E}_\zeta\Big[\mathbb{E}\big[I_1 \cdots I_m | \zeta\big]\Big] = \mathbb{E}_\zeta\left[\mathbb{E}[I_1|\zeta]^m\right] \ge \big(\mathbb{E}_\zeta\left[\mathbb{E}[I_1|\zeta]\right]\big)^m = \mathbb{E}[I_1]^m = p_s^m,$$

where we used fact that $I_1, \dots, I_m$ are independent conditional on ζ and Jensen's inequality. ■

Note that clearly $(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m) \in \Xi(m, \Delta)$, in particular $\Xi(m, \Delta)$ is non-empty w.p. at least e^{-cmn}. Using Proposition 16, we obtain

$$e^{-\Theta(n)} \ge \mathbb{P}\big[\Xi(m, \Delta) \ne \varnothing\big] \ge e^{-cmn}.$$

If m is constant and $c > 0$ is sufficiently small, this is a contradiction for all large enough n. Therefore, it suffices to establish Proposition 16.

Proof [of Proposition 16] Our proof is based on the *first moment method*. Let

$$\mathcal{S} = \big\{(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m) : \boldsymbol{\sigma}_i(k) = \boldsymbol{\sigma}_j(k), 1 \le i < j \le m, 1 \le k \le n - \Delta n\big\}.$$

Observe that

$$\big|\Xi(m,\Delta)\big| = \sum_{(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m)\in\mathcal{S}} \mathbb{1}\left\{\max_{1\le i\le m}\|\mathcal{M}_i\boldsymbol{\sigma}_i\|_\infty \le \kappa\sqrt{n}\right\}. \tag{6}$$

In what follows, we show that for a suitable $m \in \mathbb{N}$ and $\Delta \in (0,1/2)$,

$$\mathbb{E}\big[|\Xi(m,\Delta)|\big] \le e^{-\Theta(n)}.$$

Counting Estimate We bound $|\mathcal{S}|$. There are 2^n choices for $\boldsymbol{\sigma}_1 \in \Sigma_n$. Having chosen a $\boldsymbol{\sigma}_1$, there are $2^{\Delta n}$ choices for any $\boldsymbol{\sigma}_i$, $2 \le i \le m$. So,

$$|\mathcal{S}| \le 2^n\big(2^{\Delta n}\big)^{m-1} \le \exp_2\big(n + nm\Delta\big) \tag{7}$$

Probability Estimate Fix any $(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m) \in \mathcal{S}$. Denote by $R_1,\dots,R_m \in \mathbb{R}^n$ the first rows of $\mathcal{M}_1,\dots,\mathcal{M}_m$, respectively; and set

$$Z_i = \frac{1}{\sqrt{n}}\langle R_i,\boldsymbol{\sigma}_i\rangle \overset{d}{=} \mathcal{N}(0,1), \quad 1 \le i \le m.$$

Observe that if $k \ne k'$ or $n - \Delta n + 1 \le k = k' \le n$, $\mathbb{E}\big[R_i(k)R_j(k')\big] = 0$. Using this fact, we immediately conclude that $\mathbb{E}[Z_iZ_j] = 1 - \Delta$. In particular, $(Z_1,\dots,Z_m) \in \mathbb{R}^m$ is a centered multivariate normal random vector with covariance Σ, where

$$\Sigma = \Delta I_m + (1-\Delta)\mathbf{1}\mathbf{1}^T \in \mathbb{R}^{m\times m},$$

where $\mathbf{1} \in \mathbb{R}^m$ is the vector of all ones. In particular, the spectrum of Σ consists of the eigenvalue $\Delta + (1-\Delta)m$ with multiplicity one and the eigenvalue Δ with multiplicity $m-1$. We then obtain

$$\begin{aligned}\mathbb{P}\left[\max_{1\le i\le m}\|\mathcal{M}_i\boldsymbol{\sigma}_i\|_\infty \le \kappa\sqrt{n}\right] &\le \mathbb{P}\left[\max_{1\le i\le m}\big|\langle R_i,\boldsymbol{\sigma}_i\rangle\big| \le \kappa\sqrt{n}\right]^{\alpha n} \\ &= \left((2\pi)^{-\frac{m}{2}}|\Sigma|^{-\frac12}\int_{\boldsymbol{z}\in[-\kappa,\kappa]^m}\exp\left(-\frac{\boldsymbol{z}^T\Sigma^{-1}\boldsymbol{z}}{2}\right)\right)^{\alpha n} \\ &\le \Big((2\pi)^{-\frac{m}{2}}\big(\Delta + (1-\Delta)m\big)^{-\frac12}\Delta^{-\frac{m-1}{2}}(2\kappa)^m\Big)^{\alpha n}. \end{aligned} \tag{8}$$

Estimating $\mathbb{E}[|\Xi(\Delta,m)|]$ We now combine (7) and (8) to arrive at

$$\mathbb{E}\big[\big|\Xi(\Delta,m)\big|\big] \le \exp_2\Big(n\Psi(\Delta,m,\alpha)\Big), \tag{9}$$

where

$$\Psi(\Delta,m,\alpha) = 1 + m\Delta - \frac{\alpha m}{2}\log_2(2\pi) + \alpha m\log_2(2\kappa) - \frac{\alpha(m-1)}{2}\log_2\Delta - \frac{\alpha}{2}\log_2\big(\Delta + (1-\Delta)m\big).$$

Using the fact $\log_2\frac{1}{\Delta} > 0$ if $\Delta < \frac12$, we further arrive at the bound

$$\Psi(\Delta,m,\alpha) \le m\left(\frac{1}{m} - \frac{\alpha}{2m}\log_2\big(\Delta + (1-\Delta)m\big) + \Upsilon(\Delta,\alpha)\right), \tag{10}$$

for

$$\Upsilon(\Delta,\alpha) = \Delta - \frac{\alpha}{2}\log_2(2\pi) + \alpha\log_2(2\kappa) - \frac{\alpha}{2}\log_2\Delta.$$

Analyzing $\Upsilon(\Delta,\alpha)$ We set $\Delta=(2\kappa)^2$, so that

$$\alpha\log_2(2\kappa)-\frac{\alpha}{2}\log_2\Delta=0.$$

Next, fix any $\alpha\ge 4\kappa^2$. Then,

$$\Upsilon(\Delta,\alpha)=\Delta-\frac{\alpha}{2}\log_2(2\pi)\le 4\kappa^2-2\kappa^2\log_2(2\pi)=-\kappa^2\big(2\log_2(2\pi)-4\big)=-\Theta_\kappa(\kappa^2). \quad (11)$$

Combining everything For fixed small $\kappa>0$, $\alpha\ge 4\kappa^2$, and $\Delta=(2\kappa)^2$; we have $\Upsilon(\Delta,\alpha)=-\Theta_\kappa(\kappa^2)<0$. Furthermore,

$$\frac{1}{m}-\frac{\alpha}{2m}\log_2\big(\Delta+(1-\Delta)m\big)=o_m(1)$$

as $m\to\infty$. Note that $\Upsilon(\Delta,\alpha)$ depends only on α,κ. So, for $m\in\mathbb{N}$ sufficiently large, (11) yields

$$\frac{1}{m}-\frac{\alpha}{2m}\log_2\big(\Delta+(1-\Delta)m\big)+\Upsilon(\Delta,\alpha)<0.$$

Hence, combining (9) and (10), we get

$$\mathbb{E}\big[\big|\Xi(m,\Delta)\big|\big]\le e^{-\Theta(n)}.$$

From here, we conclude by Markov's inequality as

$$\mathbb{P}\big[\big|\Xi(\Delta,m)\big|\ge 1\big]\le\mathbb{E}\big[\big|\Xi(\Delta,m)\big|\big]=\exp\big(-\Theta(n)\big).$$

■

A.2. Proof of Theorem 10

The proof of Theorem 10 is similar to that of Theorem 9. We first establish the following proposition.

Proposition 18 *Let $n\ge M=\omega(1)$, $\mathcal{M}_1\in\{-1,1\}^{M\times n}$ with i.i.d. Rademacher entries and $m=\lceil\frac{2n}{M}\rceil$. Generate $\mathcal{M}_2,\dots,\mathcal{M}_m\in\{-1,1\}^{M\times n}$ by independently resampling the last M columns of $\mathcal{M}_1$. Denote by $\Xi_d(m,M)$ the set of all m-tuples $\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m\in\Sigma_n$ satisfying the following:*

- $\max_{1\le i\le m}\|\mathcal{M}_i\boldsymbol{\sigma}_i\|\le C_u\sqrt{M}$, *where* $C_u\triangleq\frac{1}{24}$.
- *For* $1\le i<j\le m$ *and* $1\le k\le n-M$, $\boldsymbol{\sigma}_i(k)=\boldsymbol{\sigma}_j(k)$.

Then,

$$\mathbb{P}\big[\Xi_d(m,M)=\varnothing\big]\ge 1-e^{-n}.$$

Before proving Proposition 18, we highlight that if $n = \omega(M)$ then $m = \omega_M(1)$ and the fraction $\Delta = M/n$ of the resampled columns is vanishing. We first show how Proposition 18 yields Theorem 10. Suppose, for the sake of contradiction, that an $\mathcal{A} : \{-1,1\}^{M\times n} \to \Sigma_n$ which is $(e^{-cM}, \sqrt{M}/24)$-optimal (with $c < 1/2$ arbitrary) exists. For $\mathcal{M}_i$, $1 \le i \le m$, as in the proposition, define

$$\boldsymbol{\sigma}_i \triangleq \mathcal{A}(\mathcal{M}_i) \in \Sigma_n, \quad 1 \le i \le n$$

and observe that

$$\boldsymbol{\sigma}_i(k) = \boldsymbol{\sigma}_j(k), \quad \text{for all} \quad 1 \le i < j \le m \quad \text{and} \quad 1 \le k \le n - M$$

as $\mathcal{A}$ is online. We then establish

Lemma 19

$$\mathbb{P}\left[\max_{1\le i\le m} \|\mathcal{M}_i\boldsymbol{\sigma}_i\| \le \frac{\sqrt{M}}{24}\right] \ge \left(e^{-cM}\right)^m \ge e^{-2cn}.$$

Proof of Lemma 19 is identical to Lemma 17. So, under the assumption that such an $\mathcal{A}$ exists, we obtain $\Xi_d(m,\Delta) \ne \varnothing$ w.p. at least e^{-2cn}. Finally using Proposition 18,

$$e^{-n} \ge \mathbb{P}\big[\Xi_d(m,M) \ne \varnothing\big] \ge e^{-2cn}$$

which is a contradiction since $c < 1/2$. Hence, it suffices to establish Proposition 18.

Proof [of Proposition 18] The proof of Proposition 18 is similar to Proposition 16; it is based in particular on the first moment method. Let

$$\bar{\mathcal{S}} = \big\{(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m) : \boldsymbol{\sigma}_i(k) = \boldsymbol{\sigma}_j(k), 1 \le i < j \le m, 1 \le k \le n - M\big\}.$$

Note that

$$\big|\Xi_d(m,M)\big| = \sum_{(\sigma_1,\dots,\sigma_m)\in\bar{\mathcal{S}}} \mathbb{1}\left\{\max_{1\le i\le m}\|\mathcal{M}_i\boldsymbol{\sigma}_i\|_\infty \le C_u\sqrt{M}\right\}, \quad \text{where} \quad C_u = \frac{1}{24}. \tag{12}$$

Counting term We bound $|\bar{\mathcal{S}}|$. There are 2^n choices for $\boldsymbol{\sigma}_1$ and having fixed it, there are 2^M choices for any $\boldsymbol{\sigma}_i$, $2 \le i \le m$. So,

$$|\bar{\mathcal{S}}| \le 2^n(2^M)^{m-1} \le \exp_2\big(n + mM\big). \tag{13}$$

Probability term. Fix an arbitrary $(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m) \in \bar{\mathcal{S}}$. Let $R_i \in \{\pm 1\}^n$, $1 \le i \le m$, denote respectively the first rows of $\mathcal{M}_i$, $1 \le i \le m$. For each fixed i, the rows of $\mathcal{M}_i$ are independent. So,

$$\mathbb{P}\left[\max_{1\le i\le m}\|\mathcal{M}_i\boldsymbol{\sigma}_i\|_\infty \le C_u\sqrt{M}\right] = \mathbb{P}\left[\max_{1\le i\le m}|\langle R_i,\boldsymbol{\sigma}_i\rangle| \le C_u\sqrt{M}\right]^M. \tag{14}$$

Next, let

$$R_i = \big(R_{ik} : 1 \le k \le n\big), \quad 1 \le i \le m.$$

Fix any $1 \le i < j \le m$. Observe that the random vectors

$$\big(R_{ik} : 1 \le k \le n - M\big) \quad \text{and} \quad \big(R_{jk} : 1 \le k \le n - M\big)$$

are identical. For this reason, we drop the first index and use $\big(R_k : 1 \le k \le n-M\big)$ instead. Next, fix any $\boldsymbol{v} = (v_1, \dots, v_{n-M}) \in \{-1,1\}^{n-M}$ and define

$$\Delta_i(\boldsymbol{v}) \triangleq \sum_{1\le k\le n-M} v_k \boldsymbol{\sigma}_i(k) \qquad \text{and} \qquad \Sigma_i \triangleq \sum_{n-M+1\le k\le n} R_{ik}\boldsymbol{\sigma}_i(k). \tag{15}$$

Our goal is to control the right hand side in (14). To that end, our strategy is to condition on $R_1, \dots, R_{n-M}$ and apply Berry-Esseen inequality for Σ_i. We establish the following auxiliary result.

Lemma 20 *Let $Z_1, \dots, Z_M$ be i.i.d. Rademacher random variables, $\epsilon_i \in \{-1,1\}$, $1 \le i \le M$, be deterministic signs, and $I \subset \mathbb{R}$ be an interval of length $|I| = \omega_M(1)$. Then*

$$\mathbb{P}\big[Z_1\epsilon_1 + \cdots + Z_M\epsilon_M \in I\big] \le \frac{3|I|}{\sqrt{M}}$$

for every large enough M.

Lemma 20 essentially rederives a classical Littlewood-Offord result; we provide a proof below for completeness.

Proof [of Lemma 20] Let $\frac{1}{\sqrt{M}}I$ denotes the set $\{c/\sqrt{M} : c \in I\}$. By the Central Limit Theorem,

$$\frac{1}{\sqrt{M}} \sum_{1\le i\le M} Z_i\epsilon_i \Rightarrow \mathcal{N}(0,1)$$

in distribution, where the speed of convergence is controlled by the Berry-Esseen inequality:

$$\left|\mathbb{P}\left[\sum_{1\le i\le M} Z_i\epsilon_i \in I\right] - \mathbb{P}\left[\mathcal{N}(0,1) \in \frac{1}{\sqrt{M}}I\right]\right| \le \frac{\mathcal{C}_{\rm be}}{\sqrt{M}}. \tag{16}$$

Here, $\mathcal{C}_{\rm be} > 0$ is an absolute constant. Furthermore, we have

$$\mathbb{P}\left[\mathcal{N}(0,1) \in \frac{1}{\sqrt{M}}I\right] = \frac{1}{\sqrt{2\pi}}\int_{u\in\frac{1}{\sqrt{M}}I} \exp(-u^2/2)\,du \le \frac{|I|}{\sqrt{2\pi M}}. \tag{17}$$

Combining (16) and (17) via triangle inequality, we obtain that for all large enough M,

$$\mathbb{P}\left[\sum_{1\le i\le M} Z_i\epsilon_i \in I\right] \le \frac{1}{\sqrt{M}}\left(\mathcal{C}_{\rm be} + \frac{|I|}{\sqrt{2\pi}}\right) \le \frac{3|I|}{\sqrt{M}}, \tag{18}$$

where we recalled $|I| = \omega_M(1)$ and $\mathcal{C}_{\rm be} = O_M(1)$. This establishes Lemma 20. ■

Next, fix any $\boldsymbol{v} \in \{-1,1\}^{n-M}$ and set

$$I_i(\boldsymbol{v}) = \left[-C_u\sqrt{M} - \Delta_i(\boldsymbol{v}), C_u\sqrt{M} - \Delta_i(\boldsymbol{v})\right],$$

where we recall $\Delta_i(\boldsymbol{v})$ from (15). In particular,

$$\big|I_i(\boldsymbol{v})\big| = 2C_u\sqrt{M}, \quad \text{for all} \quad 1 \le i \le m \quad \text{and} \quad v \in \{-1,1\}^{n-M}. \tag{19}$$

Next fix a $1 \le i \le m$ and recall Σ_i per (15). Applying Lemma 20, we conclude that

$$\max_{1\le i\le m}\max_{\boldsymbol{v}\in\{-1,1\}^{n-M}} \mathbb{P}\Big[\Sigma_i \in I_i(\boldsymbol{v})\Big] \le 6C_u. \tag{20}$$

We are ready to bound the probability term (14) by conditioning on $R_1,\dots,R_{n-M}$.

$$\begin{aligned}
&\mathbb{P}\left[\max_{1\le i\le m}|\langle R_i,\boldsymbol{\sigma}_i\rangle| \le C_u\sqrt{M}\right] \\
&= \sum_{\boldsymbol{v}\in\{-1,1\}^{n-M}} \mathbb{P}\Big[\Sigma_i\in I_i(\boldsymbol{v}), 1\le i\le m \Big| (R_1,\dots,R_{n-M})=\boldsymbol{v}\Big] \underbrace{\mathbb{P}\left[(R_1,\dots,R_{n-M})=\boldsymbol{v}\right]}_{=2^{-(n-M)}} && (21)\\
&= 2^{-(n-M)} \sum_{\boldsymbol{v}\in\{-1,1\}^{n-M}} \mathbb{P}\Big[\Sigma_i\in I_i(\boldsymbol{v}), 1\le i\le m\Big] && (22)\\
&= 2^{-(n-M)} \sum_{v\in\{-1,1\}^{n-M}} \prod_{1\le i\le m} \mathbb{P}\big[\Sigma_i\in I_i(\boldsymbol{v})\big] && (23)\\
&\le (6C_u)^m. && (24)
\end{aligned}$$

We now justify the lines above. Equation (21) follows by conditioning on the 'common randomness' $R_1,\dots,R_{n-M}$ and recalling that they are uniform over $\{-1,1\}^{n-M}$. Equation (22) uses the fact for any fixed $1\le i\le m$, Σ_i is independent of $R_1,\dots,R_{n-M}$, and (23) uses the fact $\Sigma_1,\dots,\Sigma_m$ is also a collection of independent random variables. Finally, (24) uses (20).

Combining (14) with (24), we thus conclude

$$\max_{(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m)\in\bar{\mathcal{S}}} \mathbb{P}\left[\max_{1\le i\le m}\|\mathcal{M}_i\boldsymbol{\sigma}_i\|_\infty \le C_u\sqrt{M}\right] \le (6C_u)^{mM}. \tag{25}$$

Bounding $\mathbb{E}\big[|\Xi_d(m,M)|\big]$ We are ready to estimate $\mathbb{E}\big[|\Xi_d(m,M)|\big]$. Using (12), (13) and (25),

$$\mathbb{E}\big[|\Xi_d(m,M)|\big] \le \exp_2\left(n + mM - mM\log_2\frac{1}{6C_u}\right).$$

Inserting the values $C_u = 1/24$ and $m \ge 2n/M$, we obtain

$$n + mM - mM\log_2\frac{1}{6C_u} \le -n,$$

so that $\mathbb{E}\big[|\Xi_d(m,M)|\big] \le e^{-n}$. Finally, we conclude by applying Markov's inequality:

$$\mathbb{P}\big[\Xi_d(m,M)\ne\varnothing\big] = \mathbb{P}\big[|\Xi_d(m,M)|\ge 1\big] \le \mathbb{E}\big[|\Xi_d(m,M)|\big] \le e^{-n}.$$

■

A.3. Proof Sketch for Theorem 11

The proof of Theorem 11 is quite similar to Theorem 10; we only highlight the necessary changes. Let $\mathcal{M}_1$ consists of i.i.d. Bernoulli(p) entries. Suppose that there exists an $\mathcal{A}:\mathbb{R}^{M\times n}\to\Sigma_n$ that is $(e^{-cM}, C_u'\sqrt{M})$-online in the sense of Definition 7, where $c < 1/2$ is arbitrary and

$$C_u' = \frac{\sqrt{p-p^2}}{24}.$$

We set $m = 2n/M$ and show how to adapt Proposition 18 to this case. Once this is done, the rest follows verbatim from Theorem 10. First, all instances of C_u in the proof of Theorem 10 are replaced with $C_u' = C_u\sqrt{p-p^2}$. Next, the counting estimate per (13) remains intact. Lemma 20, on the other hand, is replaced with the following.

Lemma 21 *Let $Z_1,\dots,Z_M$ be i.i.d. Bernoulli(p) random variables, $\epsilon_i \in \{-1,1\}$, $1 \le i \le M$, be deterministic signs, and $I \subset \mathbb{R}$ be an interval of length $|I| = \omega_{(p-p^2)M}(1)$. Then*

$$\mathbb{P}\big[Z_1\epsilon_1 + \cdots + Z_M\epsilon_M \in I\big] \le \frac{3|I|}{\sqrt{M(p-p^2)}},$$

for every large enough M.

Similar to Lemma 20, Lemma 21 also follows from classical Littlewood-Offord results; we provide a proof for completeness.

Proof Observe that $\mathbb{E}[Z_i\epsilon_i] = p\epsilon_i$ and

$$\mathrm{Var}(Z_i\epsilon_i) = \mathbb{E}[Z_i^2\epsilon_i^2] - p^2\epsilon_i^2 = p - p^2,$$

as $\epsilon_i \in \{-1,1\}$. Thus by the CLT,

$$\frac{1}{\sqrt{(p-p^2)M}}\left(\sum_{1\le i\le M} Z_i\epsilon_i - p\left\langle \mathbf{1},\boldsymbol{\epsilon}\right\rangle\right) \Rightarrow \mathcal{N}(0,1)$$

in distribution, where $\mathbf{1} \in \mathbb{R}^M$ is the vector of all ones and $\boldsymbol{\epsilon} = (\epsilon_i : 1 \le i \le M) \in \{-1,1\}^M$. Further, by the Berry-Esseen inequality, we have that

$$\left|\mathbb{P}\left[\sum_{1\le i\le M} Z_i\epsilon_i \in I\right] - \mathbb{P}\left[\mathcal{N}(0,1) \in \frac{I - p\left\langle \mathbf{1},\boldsymbol{\epsilon}\right\rangle}{\sqrt{(p-p^2)M}}\right]\right| \le \frac{\mathcal{C}_{\rm be}'}{\sqrt{(p-p^2)M}}$$

for some absolute constant $\mathcal{C}_{\rm be}' > 0$. Here,

$$\frac{I - p\left\langle \mathbf{1},\boldsymbol{\epsilon}\right\rangle}{\sqrt{(p-p^2)M}} = \left\{\frac{c - p\left\langle \mathbf{1},\boldsymbol{\epsilon}\right\rangle}{\sqrt{(p-p^2)M}} : c \in I\right\},$$

so that

$$\left|\frac{I - p\left\langle \mathbf{1},\boldsymbol{\epsilon}\right\rangle}{\sqrt{(p-p^2)M}}\right| = \frac{|I|}{\sqrt{(p-p^2)M}},$$

using the translation invariance of Lebesgue measure. From here, proceeding in the exact same way as in the proof of Lemma 20, we establish Lemma 21. ■

Equipped with Lemma 21 and using the exact same notation, (20) modifies to (26) where

$$\max_{1\le i\le m}\max_{\boldsymbol{v}\in\{-1,1\}^{n-M}} \mathbb{P}\Big[\Sigma_i \in I_i(\boldsymbol{v})\Big] \le \frac{6C_u'}{\sqrt{p-p^2}} = \frac{1}{4}. \tag{26}$$

We now proceed analogously to lines (21)-(24). Note that for any arbitrary $\boldsymbol{v} \in \{0,1\}^{n-M}$,

$$\mathbb{P}\Big[\Sigma_i \in I_i(\boldsymbol{v}), 1 \le i \le m \Big| (R_1, \dots, R_{n-M}) = \boldsymbol{v}\Big] \le 2^{-2m}, \tag{27}$$

using (26). Hence,

$$\begin{aligned}
&\mathbb{P}\left[\max_{1 \le i \le m} |\langle R_i, \boldsymbol{\sigma}_i \rangle| \le C_u' \sqrt{M}\right] \\
&\le \sum_{\boldsymbol{v} \in \{0,1\}^{n-M}} \mathbb{P}\Big[\Sigma_i \in I_i(\boldsymbol{v}), 1 \le i \le m \Big| (R_1, \dots, R_{n-M}) = \boldsymbol{v}\Big] \mathbb{P}\Big[(R_1, \dots, R_{n-M}) = \boldsymbol{v}\Big] \\
&\le 2^{-2m} \sum_{\boldsymbol{v} \in \{0,1\}^{n-M}} \mathbb{P}\Big[(R_1, \dots, R_{n-M}) = \boldsymbol{v}\Big] \\
&= 2^{-2m},
\end{aligned}$$

where we used (27) in the penultimate line. This is precisely the same bound as (24), so the rest of the proof remains intact. This completes the proof of Theorem 11.

A.4. Proof of Theorem 13

Fix a $K > 0$, $C_1 > c_2 > 0$, and suppose

$$C_1 M \log_2 M \ge n \ge c_2 M \log_2 M. \tag{28}$$

We establish our result via the first-moment method. Notice that by Markov's inequality,

$$\mathbb{P}\big[\mathcal{S}(K, m, \beta, \eta, \mathcal{I}) \ne \varnothing\big] = \mathbb{P}\big[\big|\mathcal{S}(K, m, \beta, \eta, \mathcal{I})\big| \ge 1\big] \le \mathbb{E}\big[\big|\mathcal{S}(K, m, \beta, \eta, \mathcal{I})\big|\big].$$

So, it suffices to prove that

$$\mathbb{E}\big[\big|\mathcal{S}(K, m, \beta, \eta, \mathcal{I})\big|\big] \le 2^{-\Theta(n)}.$$

We now estimate $\mathbb{E}\big[\big|\mathcal{S}(K, m, \beta, \eta, \mathcal{I})\big|\big]$.

Counting term Fix $m \in \mathbb{N}$, $0 < \eta < \beta < 1$ and denote by $M(m, \beta, \eta)$ the number of m-tuples $(\boldsymbol{\sigma}_i \in \Sigma_n : 1 \le i \le m)$ such that $\beta - \eta \le n^{-1} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle \le \beta$ for $1 \le i < j \le m$. We establish

Lemma 22 *For $m = O(1)$ as $n \to \infty$,*

$$M(m, \beta, \eta) \le \exp_2\left(n + n(m-1) h_b\left(\frac{1 - \beta + \eta}{2}\right) + O(\log_2 n)\right).$$

Lemma 22 is verbatim from Gamarnik et al. (2022a, Lemma 6.7), we include the proof for completeness.

Proof Observe that $\langle \boldsymbol{\sigma}, \boldsymbol{\sigma}' \rangle = n - 2d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ for any $\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \Sigma_n$. There are 2^n choices for $\boldsymbol{\sigma}_1$. Having fixed a $\boldsymbol{\sigma}_1$, there are

$$\sum_{\substack{\rho : \frac{1-\beta}{2} \le \rho \le \frac{1-\beta+\eta}{2} \\ \rho n \in \mathbb{N}}} \binom{n}{n\rho} \le \binom{n}{n\frac{1-\beta+\eta}{2}} n^{O(1)},$$

choices for any $\boldsymbol{\sigma}_i$, $2 \le i \le m$, under the constraint $\beta - \eta \le n^{-1} \langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_i \rangle \le \beta$. Next, for any $\rho \in (0,1)$, $\binom{n}{n\rho} = \exp_2\big(nh(\rho) + O\big(\log_2 n\big)\big)$ by Stirling's approximation. Combining these and recalling $m = O_n(1)$, we obtain Lemma 22. ■

Probability estimate Fix any $(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m)$ with

$$\frac{1}{n} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle = \beta - \eta_{ij}, \quad 1 \le i < j \le m.$$

Clearly $0 \le \eta_{ij} \le \eta$. Further, let $\boldsymbol{\eta} = (\eta_{ij} : 1 \le i < j \le m) \in \mathbb{R}^{m(m-1)/2}$. Then $\|\boldsymbol{\eta}\|_\infty \le \eta$. Our eventual choice of parameters β, η and m will ensure

$$\eta = \frac{1-\beta}{2m}. \tag{29}$$

We now control the probability term.

Lemma 23 *Let $\Sigma(\boldsymbol{\eta}) \in \mathbb{R}^{m \times m}$ with unit diagonal entries such that for $1 \le i < j \le m$,*

$$\big(\Sigma(\boldsymbol{\eta})\big)_{ij} = \big(\Sigma(\boldsymbol{\eta})\big)_{ji} = \beta - \eta_{ij}.$$

Then, the following holds.

(a) $\Sigma(\boldsymbol{\eta})$ is positive definite (PD) if η satisfies (29).

(b) Suppose that η satisfies (29). Then,

$$\mathbb{P}\left[\exists \tau_1, \dots, \tau_m \in \mathcal{I} : \max_{1 \le i \le m} \big\|\mathcal{M}_i(\tau_i)\boldsymbol{\sigma}_i\big\|_\infty \le K\right] \le |\mathcal{I}|^m (2\pi)^{-\frac{mM}{2}} \left(\frac{1-\beta}{2}\right)^{-\frac{Mm}{2}} \left(\frac{2K}{\sqrt{n}}\right)^{Mm}.$$

Proof

Part (a) Let $E \in \mathbb{R}^{m \times m}$ such that $E_{ii} = 0$ and $E_{ij} = E_{ji} = -\eta_{ij}$ for $1 \le i < j \le m$. Then,

$$\Sigma(\boldsymbol{\eta}) = (1-\beta)I + \beta \mathbf{1}\mathbf{1}^T + E.$$

Note that the smallest eigenvalue of $(1-\beta)I + \beta\mathbf{1}\mathbf{1}^T$ is $1-\beta$ and $\|E\|_2 \le \|E\|_F < \eta m$. So, $\Sigma(\boldsymbol{\eta})$ is invertible if $\eta < (1-\beta)/m$. Recalling the fact it is a covariance matrix, so in particular positive semidefinite, we establish part (a).

Part (b) As a first step, we take a union bound over $\mathcal{I}$ to obtain

$$\mathbb{P}\left[\exists \tau_1, \dots, \tau_m \in \mathcal{I} : \max_{1 \le i \le m} \big\|\mathcal{M}_i(\tau_i)\boldsymbol{\sigma}_i\big\|_\infty \le K\right] \le |\mathcal{I}|^m \max_{\tau_i \in \mathcal{I}, 1 \le i \le m} \mathbb{P}\left[\max_{1 \le i \le m} \big\|\mathcal{M}_i(\tau_i)\boldsymbol{\sigma}_i\big\|_\infty \le K\right]. \tag{30}$$

Next, denote by $R_i \sim \mathcal{N}(0, I_n)$ the first row of $\mathcal{M}_i(\tau_i) \in \mathbb{R}^{M \times n}$, $1 \le i \le m$. Observe that using the fact each $\mathcal{M}_i(\tau_i)$ has independent rows,

$$\mathbb{P}\left[\max_{1 \le i \le m} \big\|\mathcal{M}_i(\tau_i)\boldsymbol{\sigma}_i\big\|_\infty \le K\right] \le \mathbb{P}\left[\max_{1 \le i \le m} n^{-\frac{1}{2}} |\langle R_i, \boldsymbol{\sigma}_i \rangle| \le \frac{K}{\sqrt{n}}\right]^M. \tag{31}$$

Next, we consider the multivariate normal random vector $\big(n^{-1/2} \langle R_i, \boldsymbol{\sigma}_i \rangle : 1 \le i \le m\big)$ consisting of standard normal coordinates. Let $\overline{\Sigma}$ denotes its covariance matrix, which depends on the choice of $\tau_1, \dots, \tau_m$. Observe that for $1 \le i < j \le m$,

$$\overline{\Sigma}_{ij} = \frac{1}{n} \mathbb{E}\big[\langle R_i, \boldsymbol{\sigma}_i \rangle \langle R_j, \boldsymbol{\sigma}_j \rangle\big] = \frac{1}{n} (\boldsymbol{\sigma}_i)^T \underbrace{\mathbb{E}[R_i R_j^T]}_{=\cos(\tau_i)\cos(\tau_j) I_m} \boldsymbol{\sigma}_j = \cos(\tau_i)\cos(\tau_j)(\beta - \eta_{ij}).$$

We now remove the dependence on τ_i by relying on a Gaussian comparison inequality, due to Sidák (1968, Corollary 1). The version below is reproduced from Gamarnik et al. (2022a, Theorem 6.5).

Theorem 24 *Let $(X_1,\dots,X_k)\in\mathbb{R}^k$ be a centered multivariate normal random vector. Suppose that its covariance matrix $\Sigma\in\mathbb{R}^{k\times k}$ has unit diagonal entries has the following form: there exists $0\le\lambda_i\le 1$, $1\le i\le k$, such that for any $1\le i\ne j\le k$, $\Sigma_{ij}=\lambda_i\lambda_j\rho_{ij}$ where $(\rho_{ij}:1\le i\ne j\le k)$ is a fixed arbitrary covariance matrix. Fix values $c_1,\dots,c_k>0$, and set*

$$P(\lambda_1,\dots,\lambda_k)=\mathbb{P}\big[|X_1|<c_1,|X_2|<c_2,\dots,|X_k|<c_k\big].$$

Then, $P(\lambda_1,\dots,\lambda_k)$ is a non-decreasing function of each λ_i, $i=1,2,\dots,k$, $0\le\lambda_i\le 1$. That is,

$$P(\lambda_1,\lambda_2,\dots,\lambda_k)\le P(1,1,\dots,1).$$

We now let $(Z_1,\dots,Z_m)$ to be a centered multivariate normal random vector with covariance $\Sigma(\boldsymbol{\eta})$. Observe that

$$\begin{aligned}\max_{\tau_1,\dots,\tau_m\in\mathcal{I}}\mathbb{P}\left[\max_{1\le i\le m} n^{-\frac12}\,|\langle R_i,\boldsymbol{\sigma}_i\rangle|\le\frac{K}{\sqrt{n}}\right]&\le\mathbb{P}\left[\max_{1\le i\le m}|Z_i|\le\frac{K}{\sqrt{n}}\right] &(32)\\ &=(2\pi)^{-\frac m2}|\Sigma(\boldsymbol{\eta})|^{-\frac12}\int_{\boldsymbol{z}\in\left[-\frac{K}{\sqrt{n}},\frac{K}{\sqrt{n}}\right]^m}\exp\left(-\frac{\boldsymbol{z}^T\Sigma(\boldsymbol{\eta})^{-1}\boldsymbol{z}}{2}\right)\,d\boldsymbol{z}\\ &\le(2\pi)^{-\frac m2}|\Sigma(\boldsymbol{\eta})|^{-\frac12}\left(\frac{2K}{\sqrt{n}}\right)^m, &(33)\end{aligned}$$

where (32) follows from Theorem 24 and (33) follows from the trivial fact $\exp\left(-\frac{\boldsymbol{z}^T\Sigma(\boldsymbol{\eta})^{-1}\boldsymbol{z}}{2}\right)\le 1$.

We lastly bound $|\Sigma(\boldsymbol{\eta})|$. For this, we rely on the following tool from matrix analysis.

Theorem 25 (Hoffman-Wielandt Inequality) *Let $A\in\mathbb{R}^{m\times m}$ and $A+E\in\mathbb{R}^{m\times m}$ be two symmetric matrices with eigenvalues*

$$\lambda_1(A)\ge\cdots\ge\lambda_m(A)\quad\textit{and}\quad\lambda_1(A+E)\ge\cdots\ge\lambda_m(A+E).$$

Then,

$$\sum_{1\le i\le m}\big(\lambda_i(A+E)-\lambda_i(A)\big)^2\le\|E\|_F.$$

See Horn and Johnson (2012, Corollary 6.3.8) for a reference, and Hoffman and Wielandt (1953) for the original paper. We apply Theorem 25 to $\Sigma(\boldsymbol{\eta})$. Let $A=(1-\beta)I+\beta\mathbf{1}\mathbf{1}^T$ with eigenvalues $\lambda_1=1-\beta+\beta m>\lambda_2=\cdots=\lambda_m=1-\beta$ and E be as above. Suppose that the eigenvalues of $A+E$ are $\mu_1\ge\cdots\ge\mu_m$. Fix any $2\le i\le m$. Theorem 25 yields

$$\big|\mu_i-(1-\beta)\big|\le\|E\|_F\le\eta m=\frac{1-\beta}{2},$$

yielding

$$\mu_i\ge\frac{1-\beta}{2},\quad 2\le i\le m.$$

Furthermore, this bounds extends to μ_1, too, as

$$\mu_1\ge 1-\beta+\beta m-\frac{1-\beta}{2}>\frac{1-\beta}{2}.$$

Since $\boldsymbol{\eta} \in \mathbb{R}^{m(m-1)/2}$ is arbitrary with $\|\boldsymbol{\eta}\|_\infty \le \eta \le \frac{1-\beta}{2m}$, we obtain

$$\inf_{\substack{\boldsymbol{\eta}\in\mathbb{R}^{m(m-1)/2} \\ \|\boldsymbol{\eta}\|_\infty \le \frac{1-\beta}{2m}}} \big|\Sigma(\boldsymbol{\eta})\big| = \prod_{1\le i\le m} \mu_i \ge \left(\frac{1-\beta}{2}\right)^m . \tag{34}$$

Finally, combining (30), (31), (33), and (34) we establish the proof of part (b). ■

Estimating the expectation Let $\mathcal{F}(m,\beta,\eta)$ be the set of all m-tuples $(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m)$ such that $\beta-\eta \le n^{-1}\langle \boldsymbol{\sigma}_i,\boldsymbol{\sigma}_j\rangle \le \beta$, $1\le i<j\le m$. Then

$$\big|\mathcal{S}(K,m,\beta,\eta,\mathcal{I})\big| = \sum_{(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m)\in\mathcal{F}(m,\beta,\eta)} \mathbb{1}\left\{\exists \tau_1,\dots,\tau_m \in \mathcal{I} : \max_{1\le i\le m} \big\|\mathcal{M}_i(\tau_i)\boldsymbol{\sigma}_i\big\|_\infty \le K\right\}.$$

Using linearity of expectation, Lemma 22, Lemma 23, and the fact $\log_2|\mathcal{I}| \le cn$, we obtain

$$\mathbb{E}\big[\big|\mathcal{S}(K,m,\beta,\eta,\mathcal{I})\big|\big] \le \exp_2\Big(\Psi(m,\beta,\eta,c) + O(\log_2 n)\Big), \tag{35}$$

where

$$\Psi(m,\beta,\eta,c) = n + mnh_b\left(\frac{1-\beta+\eta}{2}\right) + cmn + \frac{mM}{2}\log_2\frac{4K^2}{\pi(1-\beta)} - \frac{Mm}{2}\log_2 n. \tag{36}$$

We set η and c as

$$\eta = \frac{1-\beta}{2m} \quad \text{and} \quad c = \frac{1}{m}, \tag{37}$$

where we recalled η from (29); parameters β and m are to be tuned soon. We now recall the scaling on n from (28). In particular,

$$\log_2 n \ge \log_2 c_2 + \log_2 M + \log_2\log_2 M \ge \log_2 c_2 + \log_2 M.$$

With this, we arrive at

$$\begin{aligned}\Psi\left(m,\beta,\frac{1-\beta}{2m},\frac{1}{m}\right) &\le 2C_1 M\log_2 M + mC_1 M\log_2 M \cdot h_b\left(\frac{1-\beta}{2}+\frac{1-\beta}{4m}\right) \\ &\quad + \frac{mM}{2}\log_2\frac{4K^2}{\pi(1-\beta)c_2} - \frac{Mm}{2}\log_2 M.\end{aligned} \tag{38}$$

Note that if $\beta\in(1/2,1)$ and $m\in\mathbb{N}$, we clearly have

$$h_b\left(\frac{1-\beta}{2}+\frac{1-\beta}{4m}\right) \le h_b(1-\beta).$$

We choose $\beta^* > 1/2$ such that

$$h_b(1-\beta^*) = \min\left\{\frac{1}{4C_1},\frac{1}{2}\right\}.$$

So,

$$mC_1M\log_2 M\cdot h_b\left(\frac{1-\beta^*}{2}+\frac{1-\beta^*}{4m}\right)\le\frac{Mm}{4}\log_2 M. \tag{39}$$

Combining (38) and (39), we further upper bound

$$\Psi\left(m,\beta^*,\frac{1-\beta^*}{2m},\frac{1}{m}\right)\le 2C_1M\log_2 M-\frac{Mm}{4}\log_2 M+\Theta(mM). \tag{40}$$

Finally, taking $m=m^*=\max\{2,16C_1\}$, we get

$$\Psi\left(m^*,\beta^*,\frac{1-\beta^*}{2m^*},\frac{1}{m^*}\right)=-\Theta(M\log_2 M). \tag{41}$$

Combining (35) with the fact $O(\log_2 n)=O(\log_2 M)=o(M\log_2 M)$ as $M=\omega(1)$, we conclude that

$$\mathbb{E}\big[\big|\mathcal{S}(K,m^*,\beta^*,\eta^*,\mathcal{I})\big|\big]\le\exp_2\left(\Psi\left(m^*,\beta^*,\frac{1-\beta^*}{2m^*},\frac{1}{m^*}\right)+o(M\log_2 M)\right)=2^{-\Theta(M\log M)}=2^{-\Theta(n)}.$$

This completes the proof of Theorem 13.