Two Studies of Constraints in High Dimensions: Entropy Inequalities and the Randomized Symmetric Binary Perceptron

By

Tanay Wakhare

B.S., University of Maryland, College Park (2020)

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2024

Authored by: Tanay Wakhare
Department of Electrical Engineering and Computer Science
January 26, 2024

Certified by: Guy Bresler
Professor of Electrical Engineering and Computer Science
Thesis Supervisor

Accepted by: Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Acknowledgements

I thank my unfailingly patient and supportive advisor Guy Bresler. I also thank my collaborator Eren C. Kızıldağ and my (past and present) office mates Brice Huang, Chenghao Guo, Enric Boix-Adserà, Anzo Teh, Peter Hoffman, Dheeraj Nagaraj, Alessandro Zanardi, Gioele Zardini, Mitchell T. Harris, and Melihcan Erol. Lastly, I thank A. and V. for keeping my spirits up during the course of writing this thesis.

Abstract

We study two constrained problems in high dimensions. We study a high dimensional inequality for the binary entropy. The *perceptron* is a natural model in high-dimensional probability, and a toy shallow neural network which stores random patterns; we also study a randomized variant of the symmetric binary perceptron.

We first consider the $(k+1)$-th derivative of $x^{k-r}H(x^r)$, where $H(x) := -x\log x - (1-x)\log(1-x), 0 \leq x \leq 1$ is the binary entropy and $k \geq r \geq 1$ are integers. Our motivation is the conjectural entropy inequality $\alpha_k H(x^k) \geq x^{k-1}H(x)$, where $0 < \alpha_k < 1$ is given by a functional equation. The $k=2$ case was the key technical tool driving recent breakthroughs on the union-closed sets conjecture, and the $k \to \infty$ case can be considered the "high dimensional limit". We express $\frac{d^{k+1}}{dx^{k+1}}x^{k-r}H(x^r)$ as a rational function, an infinite series, and a sum over generalized Stirling numbers. This allows us to reduce the proof of the entropy inequality for real k to showing that an associated polynomial has only two real roots in the interval $(0,1)$. This reduction allows us to easily verify the inequality for fixed k such as $k = 2, 3, 4$ with a finite calculation, and also allows us to prove the inequality for any fixed fractional exponent such as $k = 3/2$ via a finite calculation. The proof suggests a new framework for proving tight inequalities for the sum of polynomials times the logarithms of polynomials, which converts the inequality into a statement about the real roots of a simpler associated polynomial.

The symmetric binary perceptron (SBP) is a random constraint satisfaction problem (CSP) and a single-layer neural network; it exhibits intriguing features, most notably a sharp phase transition regarding the existence of its satisfying solutions. Secondly, we propose two novel generalizations of the SBP by incorporating random labels. Our proposals admit a natural machine learning interpretation: any satisfying solution to the random CSP is a minimizer of a certain empirical risk. We establish that the expected number of solutions for both models undergoes a sharp phase transition and calculate the location of this transition, which corresponds to the annealed capacity in statistical physics. We then establish, through the Berry-Esseen theorem, a universality result: the location of this transition does not depend on the underlying distribution. We conjecture that both models in fact exhibit an even stronger phase transition akin to the SBP and give rigorous evidence towards this conjecture through the second moment method. Our final focus is on the algorithmic problem of efficiently finding a satisfying solution to our models. We show that both models exhibit the multi Overlap Gap Property (m-OGP), an intricate geometrical property of the solution space which is known to be a rigorous barrier against large classes of algorithms. This gives rigorous evidence of a statistical-to-computational gap for both models. We also show that the m-OGP satisfies a similar universality property.

1. Introduction

In recent years, high dimensional problems have become increasingly important in various fields such as statistics, machine learning, and data science. We study two problems in high dimensional statistics: first, a tight "high dimensional" entropy inequality with applications in statistics, information theory, and combinatorics. Secondly, we study phase transitions in the solution space of the symmetric binary perceptron.

1.1. **Entropy Inequality.**

The binary entropy is a classical information theoretic function which measures the information content of a signal or random variable. The union-closed sets conjecture is a notorious open problem, stating that any set family $\mathcal{F} \subseteq 2^{[n]}$ which is *union-closed* (so that the union of two sets in $\mathcal{F}$ is also in the system) contains a "popular" element of the ground set contained in at least a $1/2$ fraction of the sets of $\mathcal{F}$.

Though still open in general, Gilmer made a recent breakthrough stating that any union-closed set system contains an element in at least an 0.01 fraction of the sets in $\mathcal{F}$. The constant 0.01 was quickly improved to $\frac{3-\sqrt{5}}{2} \approx 0.38197$ [AHS22, Saw22, CL22]. Using more sophisticated coupling arguments suggested by [Saw22], this constant was improved again to ≈ 0.38237, though the method suffers natural limitations [Yu22, Cam22]. The survey [Cam23] summarizes recent progress, but new ideas will be needed to prove the full union-closed sets conjecture.

The $k = 2$ case of the following inequality (1.2) was conjectured by Gilmer, and was one of the key technical tools underlying his breakthrough, while the $k \to \infty$ case can be considered the high dimensional limit. This case was simultaneously proved using computer calculations by [AHS22] and symbolically by [Saw22]. Studying the extension to approximate k-union closed set systems led to [Yus23] conjecturing inequality (1.2) for integer $k \geq 2$ and proving it for $k = 3, 4$. It later emerged that Boppana proved the $k = 2$ case several decades earlier [Bop85]. He recently republished a simplified proof [Bop23], which is the proof we build upon. The main contribution of this work is to reduce the proof of the real $k \geq 1$ case to a conjecture about the roots of an explicit polynomial, which suggests a general framework to prove tight inequalities involving the sum of logarithms of polynomials.

Conjecture 1. Let $k \geq 1$ be real and $0 < \alpha_k < 1$ be the unique solution of

$$\alpha_k = \frac{1}{(1+\alpha_k)^{k-1}} \tag{1.1}$$

in $(0, 1)$. Then

$$\alpha_k H(x^k) \geq x^{k-1} H(x), \quad 0 \leq x \leq 1, \tag{1.2}$$

where $H(x) := -x \log x - (1-x)\log(1-x)$ is the binary entropy. We have equality at $x = 0, \frac{1}{1+\alpha_k}, 1$.

Lemma 25 shows that the functional equation (1.1) has a unique solution satisfying $1/k < \alpha_k < 1$, and Lemma 26 shows that $\alpha_k = \frac{\log k}{k} + O_k\left(\frac{\log\log k}{k}\right)$ asymptotically for large k, which gives the high dimensional scaling limit.

The natural transformation for this problem is $x = \frac{1}{1+y}$, since we now study this inequality over y in $(0, \infty)$ instead of over x in $(0, 1)$, which maps the root at $x = \frac{1}{1+\alpha_k}$ to a root at $y = \alpha_k$. Writing $x_k = \frac{1}{1+\alpha_k}$, this functional equation is equivalent to $x_k + x_k^k = 1$. The equation $x + x^k = 1$ corresponds to the characteristic function of Fibonacci type recurrences like $F_n = F_{n-1} + F_{n-k}$. This explains the appearance of the golden ratio in the $k = 2$ case studied for the union closed sets conjecture, since $x + x^2 = 1$ has roots closely connected to the golden ratio. This also motivates studying x_k, α_k in terms of generalized Fibonacci polynomials, related to [Cig22].

We show that Conjecture 2 implies Conjecture 1. This is a strong statement about polynomial roots, since the following polynomial $p_{k,r}(x)$ has degree $k^2 + kr - r$, but we conjecture it to only have two roots in $(0, 1)$.

The following conjecture also allows us to rigorously prove Conjecture 1 for any rational exponent using a finite calculation, such as for the new case $k = 3/2$.

Conjecture 2. Let $k > r \geq 1$ be integers. Define the **entropy polynomial**

$$h_{k,r}(x) := \sum_{j=0}^{k-1} x^{rj} \sum_{v=0}^{j} \frac{(-1)^{j-v}}{v+1} \binom{rv+k}{k} \binom{k}{j-v} \tag{1.3}$$

and let α_k satisfy the functional equation (1.1). Then the polynomial

$$p_{k,r}(x) := \alpha_{k/r} k(1-x^r)^k h_{k,k}(x) - r(1-x^k)^k h_{k,r}(x) \tag{1.4}$$

has exactly two real roots in $(0, 1)$, counting multiplicity.

Note that Lemma 25 states that $\alpha_{k/r}\frac{k}{r} > 1$, so that the first polynomial is dominant.

Theorem 3. *If Conjecture 2 holds for a particular $k > r$ pair, then inequality* (1.2) *holds for the exponent k/r. If Conjecture 2 holds for all coprime $k > r \geq 1$, then inequality* (1.2) *holds for all real $k \geq 1$.*

For instance, a quick calculation shows that Conjecture 2 holds for $k = 3, r = 2$. A natural approach is to use a special case of Descartes' rules of signs, which states that if a polynomial has two coefficient sign changes, then it has either 0 or 2 positive real roots. Numerically, under the change of variables $x = \frac{1}{1+y}$, the polynomial $(1 + y)^{k^2-kr-r} h_{k,r}\left(\frac{1}{1+y}\right)$ always has two sign changes. The factor of $(1 + y)^{k^2-kr-r}$ ensures that the resulting expression is a polynomial, while only introducing extra roots at $y = -1$. If this has at most two real roots for y in $(0, \infty)$, these correspond to at most two real roots of $h_{k,r}(x) = h_{k,r}\left(\frac{1}{1+y}\right)$ in $(0, 1)$. However, the coefficients in y become unwieldy double or triple sums, from which it is difficult to deduce the sign pattern.

Some example cases are

$$h_{1,1}(x) = 1, \qquad h_{2,2}(x) = 1 + x^2, \qquad h_{3,3}(x) = 1 + 7x^3 + x^6, \qquad h_{4,4}(x) = 1 + 31x^4 + 31x^8 + x^{12}.$$

and

$$\begin{aligned} h_{4,1}(x) &= 1 - \frac{3}{2}x + x^2 - \frac{1}{4}x^3, & h_{4,2}(x) &= 1 + \frac{7}{2}x^2 - \frac{2}{3}x^4 + \frac{1}{6}x^6, \\ h_{4,3}(x) &= 1 + \frac{27}{2}x^3 + 6x^6 - \frac{1}{4}x^9, & h_{4,4}(x) &= 1 + 31x^4 + 31x^8 + x^{12}. \end{aligned}$$

This motivates the study of the binomial sums

$$h_{k,r,j} := \sum_{v=0}^{j} \frac{(-1)^{j-v}}{v+1} \binom{rv+k}{k} \binom{k}{j-v} \tag{1.5}$$

for all values of the parameters k, r, j, which does not appear in the OEIS. For instance, using a variation of the proof of Lemma 16 using finite difference operators, we can show that $h_{k,r,k} = \frac{1}{k+1}\binom{r-1}{k}$ for all $r, k \geq 1$, which is 0 for $r \leq k$. An interesting and related open problem is computing a simple representation for $h_{k,r}(x)$ under the change of variables $x \mapsto 1 - x$ or $x^r \mapsto 1 - x^r$, mirroring the symmetry of the binary entropy $H(x) = H(1 - x)$.

Our key technical tool is several equivalent expansions for the $(k + 1)$-st derivative of $x^{k-r}H(x^r)$, which are all functions of x^r. The first expansion expresses the derivative as a single infinite series, the second factors out a single root at 0 and a root of multiplicity k at 1, which leaves the numerator as a polynomial. The last generalizes and simplifies [Yus23, Lemma 3.5, Lemma 3.8] and rewrites the $(k+1)$-st derivative in terms of a different rational basis, with coefficients given by generalized Stirling numbers.

Theorem 4. *Let $S(k,\ell|\alpha,\beta,\gamma)$ denote the generalized Stirling numbers of Hsu and Shiue, defined in Equation* (2.4)*. Let $k \geq r \geq 1$ be positive integers. For $0 < x < 1$ we have*

$$\left(\frac{d}{dx}\right)^{k+1} x^{k-r}H(x^r) = -r \cdot k! \sum_{\ell=0}^{\infty} \binom{k+r\ell}{k} \frac{1}{\ell+1} x^{r\ell-1} \tag{1.6}$$

$$= -\frac{r \cdot k!}{x(1-x^r)^k} \sum_{j=0}^{k-1} x^{rj} \sum_{v=0}^{j} \frac{(-1)^{j-v}}{v+1} \binom{rv+k}{k} \binom{k}{j-v} \tag{1.7}$$

$$= -\sum_{\ell=0}^{k-1} \ell! S(k,\ell+1|1,r,k-r) r^{\ell+2} \frac{x^{r\ell-1}}{(1-x^r)^{\ell+1}}. \tag{1.8}$$

Corollary 5. *The special case $r=1$ satisfies*

$$\left(\frac{d}{dx}\right)^{k+1} x^{k-1}H(x) = \frac{(k-1)!}{x^2}\left(1 - \frac{1}{(1-x)^k}\right). \tag{1.9}$$

Corollary 6. *The special case $r=k$ has the following additional simplification in terms of s-binomial coefficients defined in Definition* (2.30)*, where $\omega = e^{\frac{2\pi i}{k}}$ is a primitive k-th root of unity:*

$$\left(\frac{d}{dx}\right)^{k+1} H(x^k) = -\frac{k!}{x} \sum_{j=0}^{k-1} \frac{1}{(1-\omega^j x)^k} \tag{1.10}$$

$$= -\frac{k \cdot k!}{x(1-x^k)^k} \sum_{\ell=0}^{k-1} \binom{k}{\ell k}_{k-1} x^{k\ell} \tag{1.11}$$

$$= -k \cdot k! \sum_{\ell=0}^{\infty} \binom{k+k\ell-1}{k-1} x^{k\ell-1}. \tag{1.12}$$

The scaling rv in the inner binomial coefficients is what makes the analysis here difficult. One common classical tool to deal with the rv scaling is the Rothe-Hagen identity [GKP94, Table 202] and its generalizations, for example due to Gould [Gou61]. However, the Rothe-Hagen identity contains binomials of the form $\binom{rv+k}{v}$, where k, r are parameters and v is the summation index. The Lagrange inversion formula also yields series with binomial coefficients $\binom{rv+k}{v}$, such as the expression for $\frac{1}{1+\alpha_k}$ from Lemma 27. Instead, we require binomials of the form $\binom{rv+k}{k}$. We could also compute a Fourier expansion by writing the sum over all v instead of rv, and then inserting $\frac{1}{r}\sum_{j=0}^{r-1} \omega^{jv}$, where ω is a primitive r-th root of unity. However, this only provides a simplification in the case $k=r$, where it is used in the proof of Corollary 6.

This inequality also has an information theoretic interpretation. Letting $X_1, \ldots, X_k \sim \texttt{Bernoulli}(x)$ be Bernoulli distributed bits and $A_j := \wedge_{i=1}^{j} X_i$ denote the binary AND of the first j bits, we have $H(x^k) = H(A_k)$ and $H(A_k|A_{k-1}) = x^{k-1}H(x)$ the conditional entropy of the k-th bit. This gives a strong data processing inequality comparing the entropy of the AND of k bits to the entropy of the AND of k-th bit conditioned on the AND of the previous $k-1$ bits.

The proof suggests a more general framework for proving tight inequalities for logarithms of polynomials of the form

$$f(x) := \sum_i p_i(x) \log(1-q_i(x)) \geq 0,$$

where both $p_i(x)$ and $q_i(x)$ are polynomials in x. Such functions arise as free energies in problems in statistical mechanics or in constraint satisfaction problems, such as in the study of graph k-coloring [COV13, Equation (8)] or boolean k-SAT [MMZ06, Equation (3)]. First, we manually find the roots of $f(x)$. Next, we pass to the $(n+1)$-st derivative, where n is the maximum degree $\max_i(\text{deg}p_i(x) + \text{deg}q_i(x))$. Taking this number of derivatives leads to a rational function. We then use classical methods to identify the number of roots of the $(n+1)$-st derivative, which is a rational function and more tractable than the original logarithmic $f(x)$. We then appeal to Rolle's theorem in the form that a function can have at most one more root than its derivative on a given interval. We use Rolle's theorem to pass from the $(n+1)$-st derivative to the original function, which can have at most $(n+1)$ more roots than the $(n+1)$-st derivative. If we are lucky, then in the first step we identified all roots of $f(x)$. Finally, we show that $f(x)$ takes positive values between each root, which implies that it is positive everywhere. The innovation over previous methods is that if we can control the *multiplicities* of the roots of $f(x)$ and the *multiplicities* of the roots of the $(n+1)$-st derivative, which is now a polynomial, Rolle's theorem allows us to deduce the desired inequality.

1.2. **Randomized Symmetric Binary Perceptron.** We focus on the *perceptron* model, a natural model in high-dimensional probability, and a toy shallow neural network which stores patterns [Win61, Wen62, Cov65]. More concretely, given patterns $X_i \in \mathbb{R}^n$, $1 \le i \le M$, *storage* corresponds to finding a vector $\boldsymbol{\sigma} \in \mathbb{R}^n$ of *synaptic weights* such that $\langle \boldsymbol{\sigma}, X_i\rangle \ge 0$ for $1 \le i \le M$. Our focus is on the *binary* case where $\boldsymbol{\sigma} \in \Sigma_n \triangleq \{-1,1\}^n$, see [Gar88, ST03, Sto13, Tal11, AS20] for the *spherical* case where $\|\boldsymbol{\sigma}\|_2 = \sqrt{n}$. Statistical physics literature provided a very detailed yet non-rigorous characterization of the *storage capacity*, i.e. the maximum number of patterns one can store via a suitable $\boldsymbol{\sigma}$, see [GD88, Gar88, KM89]. More general perceptron models considered recently involve an *activation function* $U : \mathbb{R} \to \{0,1\}$. Here, an $X_i \in \mathbb{R}^n$ is stored with respect to U if $U(\langle \boldsymbol{\sigma}, X\rangle) = 1$. Our particular focus is on *symmetric binary perceptron* [APZ19] defined by $U(x) = \mathbb{1}\{|x| \le \kappa\sqrt{n}\}$ (where $U(x) = 1$ iff $|x| \le \kappa\sqrt{n}$ and $U(x) = 0$ otherwise), see below. For even more general variants, see [BNSX22, NS23].

1.2.1. *Symmetric Binary Perceptron (SBP).* In this section, we mainly follow [GKPX22, GKPX23]. Fix $\kappa > 0$, $\alpha > 0$, and let $M = \lfloor n\alpha \rfloor \in \mathbb{N}$. Generate i.i.d. random vectors $X_i \sim \mathcal{N}(0, I_n)$, $1 \le i \le M$, where $\mathcal{N}(0, I_n)$ is the centered multivariate normal distribution on $\mathbb{R}^n$ with identity covariance. Consider the random set

$$S_\alpha(\kappa) \triangleq \big\{\boldsymbol{\sigma} \in \Sigma_n : |\langle \boldsymbol{\sigma}, X_i\rangle| \le \kappa\sqrt{n}, 1 \le i \le M\big\}. \tag{1.13}$$

Observe that $S_\alpha(\kappa)$ is indeed *symmetric* about the origin: $\boldsymbol{\sigma} \in \Sigma_n$ iff $-\boldsymbol{\sigma} \in \Sigma_n$. Proposed by Aubin, Perkins, and Zdeborová [APZ19], the SBP is a symmetrized analogue of the much studied *asymmetric binary perceptron* (ABP), where the constraints are instead of the form $\langle \boldsymbol{\sigma}, X_i\rangle \ge \kappa\sqrt{n}$, $1 \le i \le M$. The rigorous study of the ABP is an ongoing and difficult mathematical quest, see [KM89, KR98, Tal99, DS19, Xu21, ALS21a] for related work. On the other hand, the SBP exhibits relevant structural properties conjectured for the ABP [BDVLZ20] (see below); at the same time, it is more amenable to rigorous study.

Strikingly, the SBP exhibits a certain *sharp phase transition*, conjectured in [APZ19] and verified independently by Perkins and Xu [PX21] and Abbe, Li, and Sly [ALS21b]. Let
$\alpha_c(\kappa) = -1/\log_2 \mathbb{P}[|\mathcal{N}(0,1)| \le \kappa]$. Then,

$$\lim_{n\to\infty} \mathbb{P}[S_\alpha(\kappa) \ne \varnothing] = \begin{cases} 0, & \text{if } \alpha > \alpha_c(\kappa) \\ 1, & \text{if } \alpha < \alpha_c(\kappa) \end{cases}. \tag{1.14}$$

The part $\alpha > \alpha_c(\kappa)$ is established in [APZ19] through the *first moment method*: when $\alpha > \alpha_c(\kappa)$, $\mathbb{E}|S_\alpha(\kappa)| = o(1)$ and therefore $S_\alpha(\kappa) = \varnothing$ w.h.p. by Markov's inequality, where $|S_\alpha(\kappa)|$ is the cardinality of $S_\alpha(\kappa)$. The same paper also considers $\alpha < \alpha_c(\kappa)$ and shows that $\liminf_{n\to\infty} \mathbb{P}[S_\alpha(\kappa) \ne \varnothing] \ge \delta$ for some $\delta \in (0,1)$. This is based on the *second moment method*; one requires more advanced tools for the high probability guarantee (i.e. for boosting δ to one), see [PX21, ALS21b]. Furthermore, [SS23, Alt22] showed that the aforementioned phase transition is very sharp: the critical window around $\alpha_c(\kappa)$ where the probability increases quickly from $o(1)$ to $1 - o(1)$ is of constant width. So, the first moment method correctly predicts the phase transition point in SBP. This is in stark contrast with the ABP as the conjectured phase transition point [KM89] differs substantially from the first moment prediction, see [DS19].

Recalling that $S_\alpha(\kappa)$ is non-empty when α is below the critical $\alpha_c(\kappa)$ threshold, a natural goal is algorithmically finding a $\boldsymbol{\sigma} \in S_\alpha(\kappa)$. The best known polynomial-time algorithm for the SBP is due to Bansal and Spencer [BS20] from combinatorial discrepancy literature. See [ALS21a] for a different algorithm and [GKPX23, Section 1.3] for details on the connection between the SBP and combinatorial discrepancy. However, both of these algorithms work at densities substantially below $\alpha_c(\kappa)$, highlighting a *statistical-to-computational gap*: for any $\kappa > 0$, there exists an $\alpha_{\rm ALG}(\kappa) \ll \alpha_c(\kappa)$ such that finding a $\boldsymbol{\sigma} \in S_\alpha(\kappa)$ is likely to be computationally intractable when $\alpha_{\rm ALG}(\kappa) < \alpha < \alpha_c(\kappa)$. For small κ, $\alpha_{\rm ALG}(\kappa)$ is of order κ^2, see [GKPX22] for details. Limits of efficient algorithms were recently explored in [GKPX22, GKPX23] and tight lower bounds against stable and online algorithms were obtained. For a more elaborate discussion on the SBP, see [PX21, ALS21a, ALS21b, GKPX22, GKPX23, Kız22].

Notation. Given any $p \in [0,1]$, $\mathrm{Ber}(p)$ denotes the Bernoulli distribution with parameter p. For any $M \in \mathbb{N}$, $[M]$ denotes the set $\{1,\dots,M\}$. For any proposition E, $\mathbb{1}\{E\} \in \{0,1\}$ denotes its indicator. Let $\Sigma_n \triangleq \{-1,1\}^n$ and for any $\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \Sigma_n$, denote their Hamming distance $\sum_{i\le n} \mathbb{1}\{\boldsymbol{\sigma}_i \ne \boldsymbol{\sigma}'_i\}$ by $d_H(\boldsymbol{\sigma},\boldsymbol{\sigma}')$. Given a set S, $|S|$ denotes its cardinality. Given any $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$, denote their inner product $\sum_{1\le i\le n} \boldsymbol{x}_i\boldsymbol{y}_i$ by $\langle \boldsymbol{x},\boldsymbol{y}\rangle$. For any (positive semidefinite) $\boldsymbol{\Sigma}$, $\mathcal{N}(0,\boldsymbol{\Sigma})$ denotes the centered multivariate normal distribution with covariance $\boldsymbol{\Sigma}$; the cases $\boldsymbol{\Sigma} = I_n$ (the identity matrix in $\mathbb{R}^n$) and $\boldsymbol{\Sigma} = \sigma^2$ ($\sigma \in \mathbb{R}^+$) are of particular relevance. For any $r > 0$, $\log_r(\cdot)$ and $\exp_r(\cdot)$ respectively denote the logarithm and the exponential functions base r; we omit the subscript when $r = e$. We often omit floor/ceiling operators for simplicity. We use the standard asymptotic notation, e.g. $\Theta(\cdot), O(\cdot), o(\cdot), \omega(\cdot)$, where the underlying asymptotics are often with respect to $n \to \infty$. We reflect asymptotics other than $n \to \infty$ by a subscript, such as $\Theta_\kappa(\cdot), \Omega_\kappa(\cdot)$.

1.2.2. *Models.* In this section, we propose two novel generalizations of the SBP by incorporating random labels.

Definition 7. Fix $\kappa > 0$, $\alpha > 0$, $p \in [0,1]$, and set $M = n\alpha \in \mathbb{N}$. Let $X_i \sim \mathcal{N}(0, I_n)$, $1 \le i \le M$ be i.i.d. random vectors and $U(x) = \mathbb{1}\{|x| \le \kappa\sqrt{n}\}$ be the activation.

- Let $Y_i \sim \mathrm{Ber}(p)$, $1 \le i \le M$ be i.i.d. Set
$$S_\alpha(\kappa,p) = \big\{\boldsymbol{\sigma} \in \Sigma_n : Y_i = U(\langle \boldsymbol{\sigma}, X_i\rangle), \forall i \in [M]\big\}.$$
- Draw a $\mathcal{I} \subset \{1,2,\dots,M\}$ with $|\mathcal{I}| = Mp$ uniformly at random and let $Y_i = \mathbb{1}\{i \in \mathcal{I}\}$, $1 \le i \le M$. Set
$$\widetilde{S}_\alpha(\kappa,p) = \big\{\boldsymbol{\sigma} \in \Sigma_n : Y_i = U(\langle \boldsymbol{\sigma}, X_i\rangle), \forall i \in [M]\big\}.$$

Several remarks are in order. Note that the SBP is indeed a special case of the models arising in Definition 7, corresponding to the extreme case of $p = 1$. Furthermore, our model also captures the activation $\mathbb{1}\{|x| > \kappa\sqrt{n}\}$ by considering the labels $Y'_i = 1 - Y_i$ instead (equivalently replacing p by $1-p$). This is dubbed the u-function binary perceptron (UBP), see [APZ19] for details. We now highlight some fundamental differences between our models and both the SBP and the UBP. Note that for the SBP (resp. UBP), the solution space gets larger (resp. smaller) as $\kappa \to \infty$ and smaller (resp. larger) as $\kappa \to 0$. Importantly though, for $p \in (0,1)$, the sets $S_\alpha(\kappa,p)$ and $\widetilde{S}_\alpha(\kappa,p)$ shrink both as $\kappa \to 0$ as well as $\kappa \to \infty$.

1.2.3. *Dependence Structure of Labels.* We next compare the two models. On the one hand, they are somewhat similar: if $Y_i \sim \mathrm{Ber}(p)$, $1 \le i \le M$, are i.i.d., then $|\{i : Y_i = 1\}| = Mp + O(\sqrt{M})$ w.h.p. due to concentration of measure. On the other hand, the labels are not independent under the second model. Indeed, while $\mathbb{P}[i \in \mathcal{I}] = p$

for any $i \in [M]$, we have that for any $j \neq i$,

$$\mathbb{P}[j \in \mathcal{I} | i \in \mathcal{I}] = \frac{\binom{M-1}{Mp-1}}{\binom{M}{Mp}} = \frac{Mp-1}{M-1} < p = \mathbb{P}[j \in \mathcal{I}],$$

provided $p < 1$. In the next section, we show that breaking the independence in fact lowers the *critical threshold*. We now provide two interpretations of our models.

1.2.4. *Random CSP Interpretation.* Both the SBP and its generalizations in Definition 7 can be viewed as a random constraint satisfaction problem (CSP): each pair (X_i, Y_i) defines a random constraint $Y_i = \mathbb{1}\{|\langle \boldsymbol{\sigma}, X_i \rangle| \le \kappa\sqrt{n}\}$ and any $\boldsymbol{\sigma} \in S_\alpha(\kappa, p)$ is a satisfying solution to the induced CSP. Random CSPs have been thoroughly studied through various angles, ranging from the existence of solutions to the solution space geometry and the limits of polynomial-time algorithms, see e.g. [PX21] for pointers to relevant literature.

1.2.5. *Machine Learning Interpretation.* Given data consisting of feature/label pairs $(X_i, Y_i) \in \mathbb{R}^n \times \{0,1\}$, $1 \le i \le M$, a canonical task in machine learning is to find a model $f(\cdot, \boldsymbol{\sigma})$, $\boldsymbol{\sigma} \in \theta$ 'accurately explaining' these data, where θ is some domain. This often entails solving the empirical risk minimization (ERM) problem:

$$\min_{\boldsymbol{\sigma} \in \theta} \widehat{\mathcal{L}}(\boldsymbol{\sigma}), \text{ where } \widehat{\mathcal{L}}(\boldsymbol{\sigma}) = \frac{1}{M} \sum_{1 \le i \le M} \ell\big(Y_i; f(X_i, \boldsymbol{\sigma})\big).$$

Here, $\ell : \mathbb{R}^2 \to \mathbb{R}_{\ge 0}$ is a loss function. Note that when $\theta = \Sigma_n$, $\ell(y; x) = \mathbb{1}\{y \neq x\}$ and $f(X_i, \boldsymbol{\sigma}) = U(\langle \boldsymbol{\sigma}, X_i \rangle)$, $S_\alpha(\kappa, p)$ is simply the set of *interpolators*:

$$S_\alpha(\kappa, p) = \big\{\boldsymbol{\sigma} \in \Sigma_n : \widehat{\mathcal{L}}(\boldsymbol{\sigma}) = 0\big\}.$$

The case of random labels as we do here is important both from an optimization viewpoint and as a theoretical toy model in statistics. Closely related to this is the negative spherical perceptron with random labels, where $\|\boldsymbol{\sigma}\|_2 = 1$ and the constraints are of the form $Y_i \langle \boldsymbol{\sigma}, X_i \rangle \ge \kappa$ (note that since $\|\boldsymbol{\sigma}\|_2 = 1$, the right hand side scales as κ instead of $\kappa\sqrt{n}$). See Montanari et al. [MZZ21] for a thorough study of this model, including a rigorous phase transition and the analysis of a certain linear program.

1.2.6. *Annealed and Quenched Free Energies.* We will later apply the *first moment method* to show that the expected size of $S_\alpha(\kappa, p)$ (resp. $\widetilde{S}_\alpha(\kappa, p)$) undergoes a phase transition as α crosses an explicit threshold $\alpha_c(\kappa, p)$ (resp. $\widetilde{\alpha}_c(\kappa, p)$). More precisely, we show that for $S_\alpha(\kappa, p)$,

$$\lim_{n \to \infty} \frac{\log \mathbb{E}\big[|S_\alpha(\kappa, p)|\big]}{n} > 0, \quad \forall \alpha < \alpha_c(\kappa, p) \tag{1.15}$$

$$\lim_{n \to \infty} \frac{\log \mathbb{E}\big[|S_\alpha(\kappa, p)|\big]}{n} < 0, \quad \forall \alpha > \alpha_c(\kappa, p), \tag{1.16}$$

and analogously for $\widetilde{S}_\alpha(\kappa, p)$. This result concerns the quantity $n^{-1} \log \mathbb{E}\big[|S_\alpha(\kappa, p)|\big]$, which is known as the *annealed free energy* in statistical physics literature, see e.g. [MM09, BNSX22]. This should be contrasted with the *quenched free energy*, $n^{-1}\mathbb{E}\big[\log |S_\alpha(\kappa, p)|\big]$ (which is upper bounded by the annealed free energy via Jensen's inequality). An ultimate goal towards which we give some rigorous evidence in Theorem 12 is to show that (a) $S_\alpha(\kappa, p) \neq \varnothing$ with high probability (w.h.p.) if $\alpha < \alpha_c(\kappa)$ and (b) $S_\alpha(\kappa, p) = \varnothing$ (w.h.p.) if $\alpha > \alpha_c(\kappa)$. Note that when $\alpha > \alpha_c(\kappa, p)$, (1.16) yields $S_\alpha(\kappa, p) = \varnothing$ (w.h.p.) via Markov's inequality, see Theorems 8-9 for details. However for $\alpha < \alpha_c(\kappa, p)$, (1.15) does not necessarily imply $S_\alpha(\kappa, p) \neq \varnothing$: it is possible that $\mathbb{E}[|S_\alpha(\kappa, p)|]$ is large, while $|S_\alpha(\kappa, p)|$ is in fact zero w.h.p. To establish $S_\alpha(\kappa, p) \neq \varnothing$ for $\alpha < \alpha_c(\kappa, p)$, it might help studying the quenched free energy instead, e.g. if $n^{-1} \log |S_\alpha(\kappa, p)|$ concentrates around its mean. For the SBP this was done in [PX21], see also [Tal00] for a related result regarding the ABP. For our models, this is left for future work.

For more on the annealed and quenched energies, see [MM09, APZ19]. In light of the preceding discussion, the quantities $\alpha_c(\kappa,p)$ and $\widetilde{\alpha}_c(\kappa,p)$ are dubbed as the *annealed capacity.*

1.3. **Main Results.**

Throughout this section, $q(\kappa)$ denotes $\mathbb{P}[|\mathcal{N}(0,1)|\le\kappa]$, where $\mathcal{N}(0,1)$ is a standard normal.

1.3.1. *Annealed Capacity and a Universality Result.* We begin by studying the annealed capacity. Our first main result addresses the case of i.i.d. labels.

Theorem 8. *Recall $S_\alpha(\kappa,p)$ from Definition 7 and let*

$$\alpha_c(\kappa,p)=-1/\log_2\bigl(pq(\kappa)+(1-p)(1-q(\kappa))\bigr). \tag{1.17}$$

Then

$$\mathbb{E}\bigl[|S_\alpha(\kappa,p)|\bigr]=\begin{cases}\exp(-\Theta(n)), & \text{if } \alpha>\alpha_c(\kappa,p)\\ \exp(\Theta(n)), & \text{if } \alpha<\alpha_c(\kappa,p)\end{cases}.$$

In particular, $\mathbb{P}[S_\alpha(\kappa,p)=\varnothing]\ge 1-e^{-\Theta(n)}$ if $\alpha>\alpha_c(\kappa)$.

Our proof is based on a simple application of the *first moment method*, see Section 3.1.

Our second main result addresses the case where the set $\{i:Y_i=1\}$ is drawn uniformly at random among all subsets of cardinality $\lfloor Mp\rfloor$.

Theorem 9. *Recall $\widetilde{S}_\alpha(\kappa,p)$ from Definition 7 and let*

$$\widetilde{\alpha}_c(\kappa,p)=-1/\bigl(p\log_2 q(\kappa)+(1-p)\log_2(1-q(\kappa)\bigr). \tag{1.18}$$

Then,

$$\mathbb{E}\bigl[|\widetilde{S}_\alpha(\kappa,p)|\bigr]=\begin{cases}\exp(-\Theta(n)), & \text{if } \alpha>\widetilde{\alpha}_c(\kappa,p)\\ \exp(\Theta(n)), & \text{if } \alpha<\widetilde{\alpha}_c(\kappa,p)\end{cases}.$$

In particular, $\mathbb{P}[\widetilde{S}_\alpha(\kappa,p)=\varnothing]\ge 1-e^{-\Theta(n)}$ if $\alpha>\widetilde{\alpha}_c(\kappa,p)$.

Once again, the proof is based on the *first moment method*, see Section 3.2.

1.3.2. *Universality.* We next result establish a *universality* result: under mild assumptions, the quantities $\alpha_c(\kappa,p)$ and $\widetilde{\alpha}_c(\kappa,p)$ do not depend on the distribution of X_i.

Theorem 10. *Theorems 8-9 still hold if $X_i=(X_i(j):j\in[n])\in\mathbb{R}^n$ consists of i.i.d. coordinates with $\mathbb{E}[X_i(1)]=0$, $\mathbb{E}[X_i(1)^2]>0$ and $\mathbb{E}[|X_i(1)|^3]<\infty$.*

Our proof is based on Berry-Esseen Theorem (reproduced below as Theorem 28 for convenience), see Section 3.3.

We note that several related universality results appeared in the literature. In particular, [GKPX22] establishes the universality of a certain intricate geometrical property in the solution space of the SBP and [GKL$^+$22] establishes the universality of the training error for linear classification with random inputs. For a similar universality guarantee regarding solution space geometry, see also Theorem 15 below.

1.3.3. *Comparison of Thresholds in Theorems 8-9.* Inspecting (1.17) and (1.18), observe that Jensen's inequality and the concavity of the map $x\mapsto\log_2 x$ on $(0,\infty)$ collectively yield $\alpha_c(\kappa,p)\ge\widetilde{\alpha}_c(\kappa,p)$. We found it quite remarkable that breaking the independence lowers the critical threshold: the model with independent labels has a higher annealed capacity. We are unaware of any prior work in the random CSP literature that investigates whether and how the critical threshold changes with the dependence structure. We believe that this direction merits further investigation.

1.3.4. *A Sharp Phase Transition Conjecture and a Rigorous Evidence.* Recall that the prior works [PX21, ALS21b] establish a sharp phase transition (1.14) for the `SBP`, and show that the first moment method correctly predicts the location of this transition. Further, Theorems 8-9 collectively yield a phase transition for the first moment itself. In light of these, we conjecture an analogous phase transition for the models we propose.

Conjecture 11. There exists a $\kappa^* > 0$ such that the following holds for every $\kappa < \kappa^*$. The quantity $\mathbb{P}\big[S_\alpha(\kappa,p) \neq \varnothing\big]$ (resp. $\mathbb{P}\big[\widetilde{S}_\alpha(\kappa,p) \neq \varnothing\big]$) undergoes a phase transition at value $\alpha_c(\kappa,p)$ (resp. $\widetilde{\alpha}_c(\kappa,p)$) as $n \to \infty$:

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

$$\lim_{n\to\infty} \mathbb{P}[\widetilde{S}_\alpha(\kappa,p) \neq \varnothing] = \begin{cases} 0, & \text{if } \alpha > \widetilde{\alpha}_c(\kappa,p) \\ 1, & \text{if } \alpha < \widetilde{\alpha}_c(\kappa,p). \end{cases}$$

For the `UBP` (corresponding to $p=0$), [APZ19] shows that the moment method works only for $\kappa < \kappa^* \approx 0.817$. Remarkably, the value 0.817 corresponds to the onset of replica symmetry breaking, see [APZ19] for details. In light of this, we anticipate Conjecture 11 to be valid for small κ, more concretely for $\kappa < \kappa^* \approx 0.817$. The behaviour of our models beyond κ^* is a very interesting open question.

Contingent on a certain assumption, we establish the following result which serves as a rigorous evidence towards Conjecture 11.

Theorem 12. *For any $\kappa > 0$, there exists a $p_\kappa^* < 1$ such that the following holds. Fix any $p \in [p_\kappa^*, 1]$ and any $\alpha < \widetilde{\alpha}_c(\kappa,p)$. Then,*

$$\liminf_{n\to\infty} \mathbb{P}\big[\widetilde{S}_\alpha(\kappa,p) \neq \varnothing\big] > 0.$$

Moreover, for any $\kappa \in (0, 0.817)$, there exists a $p_\kappa^{**} > 0$ such that the following holds. Fix any $p \in [0, p_\kappa^{**}]$ and any $\alpha < \widetilde{\alpha}_c(\kappa,p)$. Then,

$$\liminf_{n\to\infty} \mathbb{P}\big[\widetilde{S}_\alpha(\kappa,p) \neq \varnothing\big] > 0.$$

We highlight that our proof is contingent on an assumption regarding (the critical points of) a certain real function, akin to [APZ19, Hypothesis 3]. See Section 3.4 for details. Theorem 12 covers the cases when p is close to 1 (corresponding to `SBP`) and close to 0 (corresponding to `UBP`). We prove Theorem 12 by adapting the *second moment* argument of [APZ19] with a few extra steps.

1.3.5. *Algorithmic Barriers and the Overlap Gap Property.* Theorems 8-9 establish the annealed capacities $\alpha_c(\kappa,p)$, $\widetilde{\alpha}_c(\kappa,p)$, and Theorem 12 gives a rigorous evidence that $\widetilde{S}_\alpha(\kappa,p) \neq \varnothing$ with positive probability when $\alpha < \widetilde{\alpha}_c(\kappa,p)$ and p is close to zero/one. Equipped with these, a natural follow-up question is algorithmic: *for which values of α, κ, p, one can find a $\boldsymbol{\sigma} \in S_\alpha(\kappa,p)$ (or a $\boldsymbol{\sigma} \in \widetilde{S}_\alpha(\kappa,p)$) in polynomial time?* Our focus in this section is on the tractability of this task.

As we mentioned earlier, the `SBP` exhibits a *statistical-to-computational gap* (`SCG`): for any $\kappa > 0$, there exists an $\alpha_{\rm ALG}(\kappa) \ll \alpha_c(\kappa)$, such that while solutions exist for $\alpha < \alpha_c(\kappa)$, the best known polynomial-time search algorithms work only when $\alpha \le \alpha_{\rm ALG}(\kappa)$. The `SBP` is one of many average-case models exhibiting a `SCG`. While the NP-complexity theory is often not helpful for average-case models, an active line of research proposed various frameworks for obtaining 'rigorous evidence' of hardness. We do not review these frameworks here, and instead refer the reader to the excellent surveys [KWB22, Gam21, GMZ22], as well as to the introductions of [Kız22, Hua22, GKPX22, GKPX23]. One such framework is based on the intricate geometry of the solution space utilizing the insights gained from the study of spin glasses [Tal10].

Background on Overlap Gap Property. Introduced by Gamarnik and Sudan [GS14, GS17a], the Overlap Gap Property (OGP) framework rigorously links the intricate geometry of the solution space to algorithmic hardness. At a high level, the OGP asserts the absence of a certain cluster of solutions with pairwise distances prescribed at a certain level. Whenever present, the OGP is a rigorous barrier against broad classes of algorithms exhibiting stability, for many average-case models. The list of algorithms against which the OGP is a barrier includes Langevin dynamics [GJW20, HS21], low-depth Boolean circuits [GJW21], low-degree polynomials [GJW20, Wei20, BH22], approximate message passing algorithms [GJ21], general stable algorithms [GK21, GKPX22], and online algorithms [GKPX23]. Various average-case models exhibiting the OGP include random graphs [GS14, GS17a, RV17, Wei20], random CSPs [GS17b, BH22], spin glass models [GJW20, HS21, HS23, GJK23, Kız23b], random number partitioning problem [GK21, Kız23a], symmetric binary perceptron [GKPX22, GKPX23], discrepancy minimization [GKPX23], and graph alignment problem [DGH23].

1.3.6. *Symmetric m-OGP.* Of particular interest to us is a symmetric version of the multi-OGP (m-OGP) introduced in [GS17b]. This version of the m-OGP asserts that for a suitable $m \in \mathbb{N}$ and $\beta \in (0,1)$, there exists (w.h.p. over the randomness of model) no m-tuple $\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m \in \Sigma_n$ of nearly equidistant solutions with pairwise distances around $n\frac{1-\beta}{2}$. By establishing and leveraging this property, Gamarnik, Kızıldağ, Perkins, and Xu [GKPX22] showed that in the regime $\kappa \to 0$, stable algorithms fail to find a solution for the SBP when $\alpha = \Omega_\kappa(\kappa^2 \log \frac{1}{\kappa})$. Moreover, the same authors devised a novel variant of this barrier and obtained, using this novel barrier, that online algorithms fail to find a solution for the SBP when $\alpha = \Omega_\kappa(\kappa^2)$ [GKPX23]. In light of the fact that the best known polynomial-time algorithm for the SBP is online and it works for $\alpha = O_\kappa(\kappa^2)$ [BS20], the former guarantee is tight modulo the $\log \frac{1}{\kappa}$ factor and the latter guarantee is tight up to absolute constants. For further details, see [GKPX22, GKPX23]. Below, we show that our models also exhibit this version of symmetric m-OGP. (It is worth mentioning though that we do not have any positive algorithmic guarantees for the models we propose. Finding efficient algorithms is among the open problems discussed in Section 1.4.)

OGP in Symmetric Perceptron with Random Labels. In this section, we establish that both $S_\alpha(\kappa, p)$ and $\widetilde{S}_\alpha(\kappa, p)$ introduced in Definition 7 exhibit symmetric m-OGP. We first formalize the set of m-tuples under investigation.

Definition 13. Fix $m \in \mathbb{N}$, $0 < \eta < \beta < 1$, $p \in [0,1]$, $\kappa > 0$, and $\alpha < \alpha_c(\kappa, p)$. Let $\mathcal{F}(\beta, \eta, m, \alpha, \kappa, p)$ be the set of all m-tuples $(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m)$ satisfying the following.

- **(Satisfiability)** For any $1 \le i \le m$, $\boldsymbol{\sigma}_i \in S_\alpha(\kappa, p)$.
- **(Overlap Constraint)** For any $1 \le i < j \le m$, $n^{-1} \langle \boldsymbol{\sigma}_i, \boldsymbol{\sigma}_j \rangle \in [\beta - \eta, \beta]$.

Similarly, for $\alpha < \widetilde{\alpha}_c(\kappa, p)$, let $\widetilde{\mathcal{F}}(\beta, \eta, m, \alpha, \kappa, p)$ be the set of all such m-tuples with $\boldsymbol{\sigma}_i \in \widetilde{S}_\alpha(\kappa, p)$ instead.

Definition 13 regards m-tuples of solutions with a pairwise overlap constraint. In what follows, one can think of $\beta \gg \eta$. That is, any $(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m) \in \mathcal{F}(\beta, \eta, m, \alpha, \kappa, p)$ (or $(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m) \in \widetilde{\mathcal{F}}(\beta, \eta, m, \alpha, \kappa, p)$) is a nearly equidistant m-tuple of satisfying solutions with pairwise Hamming distances around $n\frac{1-\beta}{2}$. The m-OGP mentioned above simply asserts that
$\mathcal{F}(\beta, \eta, m, \alpha, \kappa, p) = \varnothing$ (or $\widetilde{\mathcal{F}}(\beta, \eta, m, \alpha, \kappa, p) = \varnothing$) for a suitable choice of parameters.

Equipped with Definition 13, we now present our next main result.

Theorem 14. *For any $p \in (0,1]$ and $\kappa > 0$, let*

$$\alpha_{\mathrm{OGP}}(\kappa, p) = \frac{10}{p} \kappa^2 \log_2 \frac{1}{\kappa}.$$

There exists a $\kappa_0 > 0$ such that the following holds. For any $\kappa \le \kappa_0$ and $\alpha > \alpha_{\rm OGP}(\kappa,p)$, there exists an $m \in \mathbb{N}$ and $0 < \eta < \beta < 1$ such that

$$\mathbb{P}\big[\mathcal{F}(\beta,\eta,m,\alpha,\kappa,p)=\varnothing\big] \le 2^{-\Theta(n)} \quad \textit{and} \quad \mathbb{P}\big[\widetilde{\mathcal{F}}(\beta,\eta,m,\alpha,\kappa,p)=\varnothing\big] \le 2^{-\Theta(n)}.$$

The proof of Theorem 14 is based on the first moment method, the m-OGP result for the SBP established in [GKPX22, Theorem 2.4], and a careful conditioning argument. See Section 3.7 for details.

Theorem 14 asserts that for every small enough $\kappa > 0$ and any $p \in (0,1]$, both $S_\alpha(\kappa,p)$ and $\widetilde{S}_\alpha(\kappa,p)$ exhibit symmetric m-OGP when $\alpha = \Omega_{p,\kappa}(\frac{1}{p}\kappa^2\log\frac{1}{\kappa})$. Observe that for any fixed $p \in (0,1]$, $\alpha_{\rm OGP}(\kappa,p) = o_\kappa\big(\alpha_c(\kappa,p)\big)$ and $\alpha_{\rm OGP}(\kappa,p) = o_\kappa\big(\widetilde{\alpha}_c(\kappa,p)\big)$ as $\kappa \to 0$. That is, the onset of the m-OGP is asymptotically below the critical threshold per (1.17), (1.18). In light of prior discussion, this gives a strong evidence of algorithmic hardness for the regimes $\frac{10}{p}\kappa^2\log\frac{1}{\kappa} < \alpha < \alpha_c(\kappa,p)$ and $\frac{10}{p}\kappa^2\log\frac{1}{\kappa} < \alpha < \widetilde{\alpha}_c(\kappa,p)$. That is, while the solutions exist with positive probability in these regimes, polynomial-time search algorithms likely fail. In particular, one can rigorously show that sufficiently stable algorithms[1] fail to find a $\boldsymbol{\sigma} \in S_\alpha(\kappa,p)$ or a $\boldsymbol{\sigma} \in \widetilde{S}_\alpha(\kappa,p)$ for $\alpha = \Omega_{p,\kappa}(\frac{1}{p}\kappa^2\log\frac{1}{\kappa})$. This can be done by adapting the techniques of [GKPX22, Theorem 3.2] verbatim. For this reason and for keeping our exposition clean, we do not pursue this improvement herein.

1.3.7. *Universality for the OGP.*

Our last result is a universality property for the m-OGP.

Theorem 15. *Let $\mathcal{D}$ be a distribution on $\mathbb{R}$ such that*

$$\mathbb{E}_{T\sim\mathcal{D}}[T]=0, \quad \mathbb{E}_{T\sim\mathcal{D}}[T^2]>0, \quad \textit{and} \quad \mathbb{E}_{T\sim\mathcal{D}}[|T|^3]<\infty.$$

Suppose that $X_i \in \mathbb{R}^n$, $1 \le i \le M$ are i.i.d. with entries drawn from $\mathcal{D}$. Then, Theorem 14 still remains valid.

Theorem 15 shows that our proposed models still exhibit the m-OGP under mild distributional assumptions. The proof of Theorem 15 is quite similar to that of [GKPX22, Theorem 5.2]. In particular, it is based on the multi-dimensional version of the Berry-Esseen theorem. See Section 3.10 for an outline of the proof.

1.4. **Conclusion and Open Problems.**

In this paper, we proposed two novel generalizations of the SBP that involve random labels. Our models form a natural link between the SBP and machine learning: any satisfying solution is a minimizer of a certain empirical risk. We then calculated the critical capacity for both models, showed a certain universality property for the critical capacity, and established, through the second moment method, that solutions exist with positive probability below the critical capacity. We lastly showed that our models exhibit an intricate geometrical property known as the Overlap Gap Property (OGP), and that the onset of the OGP is well below the critical capacity. The OGP is a rigorous barrier for large classes of search algorithms, and that it also enjoys a universality property.

We now provide an extensive list of open problems.

1.4.1. *Sharp Phase Transition.*

In light of earlier discussion, we conjecture that both models exhibit a sharp phase transition (Conjecture 11). It is plausible that Conjecture 11 can be resolved by employing an argument similar to [PX21, ALS21b]; we leave this as an open problem.

1.4.2. *Interplay between the Critical Threshold and Dependence Structure.*

Recall that $\alpha_c(\kappa,p) \ge \widetilde{\alpha}_c(\kappa,p)$ for any $\kappa > 0$ and $p \in [0,1]$. The interplay between the critical threshold and the dependence structure in the context of other random CSPs or neural network models (such as the Hopfield model) is an interesting question for future work.

[1] Informally, an algorithm is stable if a small perturbation of its input changes its output only by a small amount. For a formal definition, see [GKPX22, Definition 3.1]

1.4.3. *Other Perceptron Models.* It would be very interesting to extend our results to the spherical case ($\|\boldsymbol{\sigma}\|_2 = 1$). We believe that the arguments of [MZZ21] may transfer. Similarly, it would be interesting to consider different activations $U(x)$ and more general perceptron models [BNSX22, NS23].

1.4.4. *Algorithms.* While [BS20] and [ALS21a] devise efficient algorithms for finding solutions of the SBP and the UBP at sufficiently low densities, it is not clear whether they apply to our models. Let $\mathcal{I} = \{i : Y_i = 1\}$, $\mathcal{M} \in \mathbb{R}^{|\mathcal{I}| \times n}$ with rows $X_i \in \mathbb{R}^n$, $i \in \mathcal{I}$, and $\overline{\mathcal{M}} \in \mathbb{R}^{(M-|\mathcal{I}|)\times n}$ with rows $X_i \in \mathbb{R}^n$, $i \in [M] \setminus \mathcal{I}$. Note that when $0 < p < 1$ holds strictly, both $\mathcal{I}$ and $\mathcal{I}^c$ are w.h.p. non-empty. Observe that finding a $\boldsymbol{\sigma} \in S_\alpha(\kappa, p)$ (or a $\boldsymbol{\sigma} \in \widetilde{S}_\alpha(\kappa, p)$) amounts to finding a $\boldsymbol{\sigma}$ such that both $\|\mathcal{M}\boldsymbol{\sigma}\|_\infty \le \kappa\sqrt{n}$ and $\min_i \left|\left(\overline{\mathcal{M}}\boldsymbol{\sigma}\right)_i\right| > \kappa\sqrt{n}$ hold. To that end, one can potentially run (a) the discrepancy minimization algorithm to find a $\boldsymbol{\sigma}_1 \in \Sigma_n$ with $\|\mathcal{M}\boldsymbol{\sigma}\|_\infty \le \kappa\sqrt{n}$ and (b) the algorithm of Abbe, Li, and Sly [ALS21a] to find a $\boldsymbol{\sigma}_2 \in \Sigma_n$ with $\min_i \left|\left(\overline{\mathcal{M}}\boldsymbol{\sigma}\right)_i\right| > \kappa\sqrt{n}$. It is, however, unclear if these algorithms return the same solution (i.e. $\boldsymbol{\sigma}_1 = \boldsymbol{\sigma}_2$) even at very low densities. Assuming that solutions do exist for densities below the critical threshold, it is a very interesting open question to find efficient algorithms finding these solutions at certain densities.

1.4.5. *Solution Space Geometry.* A large body of literature on random CSPs is devoted to the study of their solution space geometry [PX21]. Intricate geometrical properties of their solution spaces are linked to the failure of algorithms, see [ART06, ACO08, PX21, GKPX22, GKPX23] for a discussion. [GKPX22] studied the solution space geometry of the SBP and established the presence of the multi Overlap Gap Property (m-OGP) in order to obtain nearly tight lower bounds against the class of stable algorithms. More recently, the same authors established in [GKPX23] a different intricate geometrical property and leveraged it to obtain tight hardness guarantees against online algorithms. The class of online algorithms captures, in particular, the best known algorithm for the SBP [BS20]. In Theorem 14, we established that for small enough κ, both models we propose exhibit the m-OGP. It would be very interesting to extend this result to moderate values of κ, as well as to the case $\kappa \to \infty$. We anticipate that the fact that $S_\alpha(\kappa, p)$ (and $\widetilde{S}_\alpha(\kappa, p)$) shrinks as $\kappa \to \infty$ may simplify the analysis in the latter case.

2. Entropy Inequality

In Section 2.1 we introduce the generalized Stirling numbers of Hsu and Shiue, and demonstrate a connection to generalized Bernoulli and Eulerian numbers. In Section 2.2 we evaluate $h_{k,r,j}$ in certain regimes of k, r, j. In Section 2.3 we prove the entropy expansions of Theorem 4. In Section 2.4 we simplify the special cases of Corollaries 5 and 6. In Section 2.5 we prove Theorem 3, an equivalence between our main entropy inequality and counting real roots of $h_{k,r}(x)$ in $(0,1)$. This allows us to verify the inequality for small k, and prove new cases such as $k = 3/2$ via a finite calculation. Finally, in Section 2.6 we study the scaling constant α_k.

2.1. **Definitions.** The ubiquitous **Stirling numbers of the second kind** are defined as the solutions to the recurrence [GKP94, Equation (6.3)] and have the closed form [GKP94, Equation (6.19)]:

$$S(n+1,\ell) = S(n,\ell-1) + \ell S(n,\ell), \tag{2.1}$$

$$\ell! S(n,\ell) = \sum_{v=0}^{\ell} (-1)^{\ell-v} \binom{\ell}{v} v^n. \tag{2.2}$$

Define the **scaled Pochhammer** symbol $(z|\alpha)_n := z(z-\alpha)\cdots(z-(n-1)\alpha), n \geq 1$. Hsu and Shiue [HS88] introduced generalized Stirling numbers using these scaled Pochhammer symbols. Let $\alpha, \beta, \gamma \in \mathbb{R}$. Define **generalized Stirling numbers** $S(n,\ell|\alpha,\beta,\gamma)$ via the change of basis relation

$$(z|\alpha)_n = \sum_{\ell=0}^{n} S(n,\ell|\alpha,\beta,\gamma)(z-\gamma|\beta)_n \tag{2.3}$$

and initial conditions $S(0,0|\alpha,\beta,\gamma) = 1, S(n,0|\alpha,\beta,\gamma) = (\gamma|\alpha)_n$.

Many properties of these, discovered by subsequent researchers, are surveyed in the book [MS16]. We will require the following recurrence and closed form [MS16, Theorem (4.51), Theorem (4.52)]:

$$S(n+1,\ell|\alpha,\beta,\gamma) = S(n,\ell-1|\alpha,\beta,\gamma) + (\ell\beta - n\alpha + \gamma)S(n,\ell|\alpha,\beta,\gamma), \tag{2.4}$$

$$S(n,\ell|\alpha,\beta,\gamma) = \frac{(-1)^\ell}{\beta^\ell \ell!} \sum_{j=0}^{\ell} (-1)^j \binom{\ell}{j} (\beta j + \gamma|\alpha)_n. \tag{2.5}$$

In the remainder of this chapter, we often take $\alpha = 1$ in the definition of the generalized Stirling numbers, which gives

$$\ell! S(n,\ell|1,\beta,\gamma)\beta^\ell = \sum_{v=0}^{\ell} (-1)^{\ell-v} \binom{\ell}{v} n! \binom{\beta v + \gamma}{n}. \tag{2.6}$$

We also require the Dobiński type formula [HS88, Equation (27)] which factors e^x out of the series:

$$\sum_{\ell=0}^{\infty} \frac{x^\ell}{\ell!} \binom{r\ell + s}{n} = \frac{1}{n!} \sum_{\ell=0}^{n} S(n,\ell|1,r,s) r^\ell e^x x^\ell. \tag{2.7}$$

Note that by comparing recurrences and initial conditions, we can show that

$$C(k,t,j) = \frac{k^j}{(k-1)!} S(t,j|1,k,0),$$

where $C(k,t,j)$ are the rational coefficients introduced by Yuster [Yus23] in his work on the k-union closed sets conjecture, which considers the $r = 1, k$ cases of our result. Furthermore, the $C(k,t,j)$ coefficients have been studied before and are exactly the generalized factorial coefficients of [Cha02, Definition 8.2], since they satisfy the same recurrence and initial conditions.

2.1.1. *Classical analogy.* Comparing the two closed form expressions for Stirling numbers (2.2) and (2.6) shows that, up to normalization by β, we replace the term v^n with

$$n!\binom{\beta v+\gamma}{n} = (\beta v+\gamma)(\beta v+\gamma-1)\cdots(\beta v+\gamma-n+1).$$

If we further specialize $\gamma = 0$, we can take the limit

$$\lim_{\beta\to\infty} \frac{S(n,\ell|1,\beta,0)\beta^{\ell}}{\beta^n} = \lim_{\beta\to\infty} \frac{1}{\ell!\beta^n}\sum_{v=0}^{\ell}(-1)^{\ell-v}\binom{\ell}{v}n!\binom{\beta v}{n} \tag{2.8}$$

$$= \frac{1}{\ell!}\sum_{v=0}^{\ell}(-1)^{\ell-v}\binom{\ell}{v}v^n \tag{2.9}$$

$$= S(n,\ell). \tag{2.10}$$

The **Eulerian numbers** $A_{n,k}$ have the closed form [GKP94, Equation (6.38)]

$$A_{n,\ell} = \sum_{v=0}^{\ell}(-1)^{\ell-v}\binom{n+1}{\ell-v}(v+1)^n. \tag{2.11}$$

We introduce the companion sequence of **generalized Eulerian numbers**

$$A_{n,\ell}^{(r,s)} = n!\sum_{v=0}^{\ell}(-1)^{\ell-v}\binom{n+1}{\ell-v}\binom{(v+1)r+s}{n}. \tag{2.12}$$

By a similar argument to the Stirling case we see that

$$\lim_{r\to\infty}\frac{A_{n,\ell}^{(r,0)}}{r^n} = A_{n,\ell}.$$

Note that if $\gamma \neq 0$ or $s \neq 0$ we do not reduce to standard Eulerian and Stirling numbers in the large β limit.

There are many classical relations linking Stirling numbers, sum of powers, Eulerian numbers, and Bernoulli numbers. One of the key identities linking moments, Stirling numbers, and Eulerian numbers [GKP94, Equation (7.46)] is

$$\left(z\frac{d}{dz}\right)^n\frac{1}{1-z} = \sum_{\ell=1}^{\infty}\ell^n z^{\ell} = \sum_{j=0}^{n} j!S(n,j)\frac{z^j}{(1-z)^{j+1}} = \frac{z}{(1-z)^{n+1}}\sum_{j=0}^{n}A_{n,j}z^j. \tag{2.13}$$

Combining Lemma 2.24 with some further calculations gives a generalization of this transformation involving the parameters r, s:

$$n!\sum_{\ell=1}^{\infty}\binom{r\ell+s}{n}z^{\ell} = \sum_{j=0}^{n} j!S(n,j|1,r,s)r^j\frac{z^j}{(1-z)^{j+1}} = \frac{z}{(1-z)^{n+1}}\sum_{j=0}^{n}A_{n,j}^{(r,s)}z^j. \tag{2.14}$$

The existence of this transformation is what leads to the definition of $A_{n,j}^{(r,s)}$. These definitions of generalized Stirling and Eulerian numbers suggest an analog of the classical calculus where we systematically replace powers v^t with the scaled Pochhammer $t!\binom{\beta v}{t} = (\beta v)(\beta v-1)\cdots(\beta v-t+1)$. This introduces the free parameter β, and we can recover all of the classical sequences by normalizing and taking the $\beta\to\infty$ limit.

Beginning with the closed form for Bernoulli numbers [DLMF, Equation (24.6.9)]

$$B_n = \sum_{\ell=0}^{n}\sum_{v=0}^{\ell}\frac{(-1)^v}{\ell+1}\binom{\ell}{v}v^n, \tag{2.15}$$

we then define **generalized Bernoulli numbers** by the closed form

$$B_n^{(r,s)} := n! \sum_{\ell=0}^{n} \sum_{v=0}^{\ell} \frac{(-1)^v}{\ell+1} \binom{\ell}{v} \binom{rv+s}{n}. \tag{2.16}$$

We leave it as an open question to explore the links between these generalizations of Bernoulli, Eulerian, and Stirling numbers.

2.2. **Finite differences.** Recall that we defined the key binomial sum (1.5)

$$h_{k,r,j} := \sum_{v=0}^{j} \frac{(-1)^{j-v}}{v+1} \binom{rv+k}{k} \binom{k}{j-v}.$$

We will need to understand $h_{k,r,j}$ in more detail, in particular its vanishing for $j \geq k \geq r \geq 1$. Using finite difference operators, we can provide an expression for $h_{k,r,j}$ which sums over $k-j$ terms instead of j terms. This gives explicit expressions for the leading coefficients of $h_{k,r}(x)$, since for example we have $h_{k,r,k-1} = (-1)^r \frac{1}{\binom{k}{r}}$ and $(-1)^k h_{k,r,k-2} = (-1)^{r+1} \frac{k}{\binom{k}{r}} + \binom{k-2r}{k}$. Note the appearance of binomial coefficients with a negative upper index.

Lemma 16. *Consider integer $k \geq r \geq 1$. If $j \geq k$, then $h_{k,r,j} = 0$. If $1 \leq j < k$, then*

$$h_{k,r,j} = (-1)^{j+r+1} \frac{\binom{k}{j+1}}{\binom{k}{r}} + \sum_{v=2}^{k-j} \frac{(-1)^{j+v}}{v-1} \binom{k}{j+v} \binom{k-rv}{k}. \tag{2.17}$$

Proof. Define the polynomial $q(x) = \frac{1}{x+1}\binom{rx+k}{k}$. For $r \leq k$ the numerator contains the factor $rx+r$ which cancels with $x+1$ in the denominator, so that $q(x)$ is a polynomial of degree $\leq k-1$. We apply the finite difference operator Δ defined by $\Delta q(x) := q(x+1) - q(x)$, which lowers the degree of a polynomial $q(x)$ by 1. Therefore, we have the key identity

$$\Delta^k q(x) = \sum_{v=0}^{k} (-1)^{k-v} \binom{k}{v} q(x+v) = \sum_{v=0}^{k} (-1)^{k-v} \binom{k}{v} \frac{1}{x+v+1} \binom{rx+rv+k}{k} = 0, \tag{2.18}$$

which is 0 because we have lowered the degree of a degree $k-1$ polynomial k times.

Consider the case $j \geq k \geq r \geq 1$. We rewrite the definition of $h_{k,r,j}$ as

$$\begin{aligned} h_{k,r,j} &= \sum_{v=0}^{j} \frac{(-1)^{j-v}}{v+1} \binom{rv+k}{k} \binom{k}{j-v} \\ &= \sum_{v=j-k}^{j} \frac{(-1)^{j-v}}{v+1} \binom{rv+k}{k} \binom{k}{j-v} \\ &= \sum_{v=0}^{k} \frac{(-1)^{k-v}}{j-k+v+1} \binom{rv+k+r(j-k)}{k} \binom{k}{k-v} \\ &= \sum_{v=0}^{k} \frac{(-1)^{k-v}}{j-k+v+1} \binom{rv+k+r(j-k)}{k} \binom{k}{v}. \end{aligned}$$

We first truncated the sum from $v = j-k \geq 0$ to $v = k$ since $\binom{k}{j-v}$ vanishes outside this range. We then shifted the v sum by $j-k$ and used the symmetry $\binom{k}{k-v} = \binom{k}{v}$. Comparing against Equation (2.18), we see that this is exactly $\Delta^k q(x)|_{x=j-k} = 0$, since setting $x = j-k \geq 0$ does not lead to any singular terms.

Now consider the case $k \geq r \geq 1$ and $k > j \geq 1$. Consider the limit $x \to -j$ in Equation (2.18), so that the only singular term is at $v = j-1$, where we have

$$\begin{aligned}
\lim_{x\to -j}(-1)^{k-j-1}\binom{k}{j-1}\frac{1}{x+j}\binom{rx+rj+k-r}{k} &= (-1)^{k-j-1}\binom{k}{j-1}\lim_{x\to 0}\frac{1}{x}\binom{rx+k-r}{k}\\
&= (-1)^{k-j-1}\binom{k}{j-1}\frac{r}{k!}\lim_{x\to 0}\frac{(rx+k-r)(rx+k-r-1)\cdots(rx-r+1)}{rx}\\
&= (-1)^{k-j-1}\binom{k}{j-1}\frac{r}{k!}\cdot(k-r)!(r-1)!(-1)^{r-1}\\
&= (-1)^{k-j-r}\frac{\binom{k}{j-1}}{\binom{k}{r}}.
\end{aligned}$$

The rx terms in the numerator and denominator cancelled, so that substituting $x=0$ was well-defined. Hence the finite difference result (2.18) reduces to

$$\begin{aligned}
0 = \Delta^k q(x)\big|_{x=-j} &= (-1)^{k-j-r}\frac{\binom{k}{j-1}}{\binom{k}{r}} + \sum_{v=0}^{j-2}\frac{(-1)^{k-v}}{v-j+1}\binom{k}{v}\binom{rv-rj+k}{k} + \sum_{v=j}^{k}\frac{(-1)^{k-v}}{v-j+1}\binom{k}{v}\binom{rv-rj+k}{k}\\
&= (-1)^{k-j-r}\frac{\binom{k}{j-1}}{\binom{k}{r}} + h_{k,r,k-j} + \sum_{v=0}^{j-2}\frac{(-1)^{k-v}}{v-j+1}\binom{k}{v}\binom{rv-rj+k}{k}. \qquad (2.19)
\end{aligned}$$

Here, we rewrote $h_{k,r,k-j}$ as

$$\begin{aligned}
h_{k,r,k-j} :&= \sum_{v=0}^{k-j}\frac{(-1)^{k-j-v}}{v+1}\binom{rv+k}{k}\binom{k}{k-j-v}\\
&= \sum_{v=j}^{k}\frac{(-1)^{k-v}}{v-j+1}\binom{rv-rj+k}{k}\binom{k}{k-v}\\
&= \sum_{v=j}^{k}\frac{(-1)^{k-v}}{v-j+1}\binom{rv-rj+k}{k}\binom{k}{v},
\end{aligned}$$

where we shifted the v summation by j, reversed the order of summation, and used the symmetry $\binom{k}{k-v} = \binom{k}{v}$.

We also have

$$\begin{aligned}
\sum_{v=0}^{j-2}\frac{(-1)^{k-v}}{v-j+1}\binom{k}{v}\binom{rv-rj+k}{k} &= \sum_{v=0}^{j-2}\frac{(-1)^{k+j-v}}{-(v+1)}\binom{k}{j-v-2}\binom{-rv-2r+k}{k}\\
&= \sum_{v=2}^{j}\frac{(-1)^{k+j-v+1}}{v-1}\binom{k}{j-v}\binom{k-rv}{k},
\end{aligned}$$

where we reversed the order of summation and then shifted the summation over v by 2. Finally, reversing $j \mapsto k-j$ in Equation (2.19) and noting $\binom{k}{k-j-v} = \binom{k}{j+v}$ completes the proof. □

2.3. **Entropy derivative closed forms.** This subsection proves the closed forms for iterated entropy derivatives from Theorem 4. We begin with a fundamental infinite series expansion for the binary entropy. The key is that we consider $x^k \log x$ as an analytic function around 0, which we do not series expand, while we series expand $\log(1-x^k)$.

Lemma 17. *For integer $k \geq 1$ and real $0 \leq x \leq 1$ we have the expansion*

$$H(x^k) = -x^k \log x^k - (1-x^k)\log(1-x^k) = -kx^k\log x + x^k - \sum_{\ell=1}^{\infty}\frac{x^{k(\ell+1)}}{\ell(\ell+1)}. \qquad (2.20)$$

Proof. Recall $-\log(1-x) = \sum_{\ell=1}^{\infty} \frac{x^\ell}{\ell}$. Consider the series

$$\begin{aligned}\sum_{\ell=1}^{\infty} \frac{x^{\ell+1}}{\ell(\ell+1)} &= \sum_{\ell=1}^{\infty} x^{\ell+1}\left(\frac{1}{\ell} - \frac{1}{\ell+1}\right)\\ &= x\sum_{\ell=1}^{\infty}\frac{x^\ell}{\ell} - \sum_{\ell=2}^{\infty}\frac{x^\ell}{\ell}\\ &= -x\log(1-x) + (x + \log(1-x))\\ &= x + (1-x)\log(1-x).\end{aligned}$$

Mapping $x \mapsto x^k$ and substituting into the definition of $H(x^k) = -x^k \log x^k - (1-x^k)\log(1-x^k)$ finishes the proof.

Note that at $x = 0$ this approaches $H(0) = 0$ and at $x = 1$ we can telescope

$$\sum_{\ell=1}^{\infty} \frac{1}{\ell(\ell+1)} = \sum_{\ell=1}^{\infty}\left(\frac{1}{\ell} - \frac{1}{\ell+1}\right) = 1,$$

so that the series converges for $0 \le x \le 1$. □

We now differentiate termwise to obtain an expression for the $(k+1)$-st derivative.

Lemma 18. *For integer $k \ge r \ge 1$ and real $0 < x < 1$ we have*

$$\left(\frac{d}{dx}\right)^{k+1} x^{k-r}H(x^r) = -r\cdot k!\sum_{\ell=0}^{\infty}\binom{k+r\ell}{k}\frac{1}{\ell+1}x^{r\ell-1}. \tag{2.21}$$

Proof. We begin with Lemma 17 in the form

$$x^{k-r}H(x^r) = -rx^k\log x + x^k - \sum_{\ell=1}^{\infty}\frac{x^{k+r\ell}}{\ell(\ell+1)}. \tag{2.22}$$

Note that

$$\left(\frac{d}{dx}\right)^{k+1} x^k \log x = \frac{k!}{x}$$

and

$$\left(\frac{d}{dx}\right)^{k+1} x^{k+r\ell} = \frac{(k+r\ell)!}{(r\ell-1)!}x^{r\ell-1},$$

so that after differentiating $(k+1)$ times termwise we have

$$\left(\frac{d}{dx}\right)^{k+1} x^{k-r}H(x^r) = -\frac{r\cdot k!}{x} - \sum_{\ell=1}^{\infty}\frac{(k+r\ell)!}{(r\ell-1)!(\ell)(\ell+1)}x^{r\ell-1}. \tag{2.23}$$

Rewrite the factorials as

$$\frac{(k+r\ell)!}{(r\ell-1)!(\ell)(\ell+1)} = r\cdot k!\frac{(k+r\ell)!}{(r\ell-1)!k!(r\ell)(\ell+1)} = r\cdot k!\binom{k+r\ell}{k}\frac{1}{\ell+1},$$

and then recognize $-\frac{r\cdot k!}{x}$ as the $\ell = 0$ term of the sum. The key observation is that the factorial ratio cancels nontrivially. Finally, the sum in Equation (2.23) becomes

$$\left(\frac{d}{dx}\right)^{k+1} x^{k-r}H(x^r) = -r\cdot k!\sum_{\ell=0}^{\infty}\binom{k+r\ell}{k}\frac{1}{\ell+1}x^{r\ell-1}$$

and we are done. □

To show that the $(k+1)$-st derivative is a rational function in x, we consider the product with $(1-x^r)^k$ and show that this is a polynomial, which is not obvious.

Lemma 19. *For integer $k \geq 1$ and real $0 < x < 1$ we have*

$$\left(\frac{d}{dx}\right)^{k+1} x^{k-r} H(x^r) = -\frac{r \cdot k!}{x(1-x^r)^k} \sum_{j=0}^{k-1} x^{rj} \sum_{v=0}^{j} \frac{(-1)^{j-v}}{v+1} \binom{rv+k}{k} \binom{k}{j-v}.$$

Proof. For $0 < x < 1$, where the entropy series converges, reindex the product

$$\begin{aligned}(1-x^r)^k \sum_{\ell=0}^{\infty} \binom{k+r\ell}{k} \frac{1}{\ell+1} x^{r\ell} &= \sum_{m=0}^{k} (-1)^m \binom{k}{m} x^{rm} \sum_{\ell=0}^{\infty} \binom{k+r\ell}{k} \frac{1}{\ell+1} x^{r\ell} \\ &= \sum_{j=0}^{\infty} x^{rj} \sum_{v=0}^{j} \frac{(-1)^{k-v}}{v+1} \binom{k+rv}{k} \binom{k}{j-v} \\ &= \sum_{j=0}^{\infty} x^{rj} h_{k,r,j},\end{aligned}$$

where we recall the definition of $h_{k,r,j}$ in Equation (1.5). Now Lemma 16 says that for $k \geq r \geq 1$ and $j \geq k$, we have $h_{k,r,j} = 0$. Therefore this sum is actually a polynomial, and

$$\sum_{\ell=0}^{\infty} \binom{k+r\ell}{k} \frac{1}{\ell+1} x^{r\ell} = \frac{1}{(1-x^r)^k} \sum_{j=0}^{k-1} x^{rj} h_{k,r,j}.$$

Comparing with Lemma 18 completes the proof. □

Lemma 20. *Let $r, n \geq 1$ be integers and $r \leq s \leq n+r-1$ an integer. For complex w with $\Re(w) < 1$ we have*

$$\sum_{\ell=0}^{\infty} w^{r\ell-1} \binom{r\ell+s}{n} = \frac{1}{n!} \sum_{\ell=0}^{n} \ell! S(n,\ell|1,r,s) r^{\ell} \frac{w^{r\ell-1}}{(1-w^r)^{\ell+1}} \tag{2.24}$$

and

$$\sum_{\ell=0}^{\infty} \frac{w^{r\ell-1}}{\ell+1} \binom{r\ell+s}{n} = \frac{1}{n!} \sum_{\ell=0}^{n} \ell! S(n,\ell+1|1,r,s-r) r^{\ell+1} \frac{w^{r\ell-1}}{(1-w^r)^{\ell+1}}, \tag{2.25}$$

.

Proof. We begin with the Dobiński-type formula of Equation (2.7):

$$\sum_{\ell=0}^{\infty} \frac{x^{\ell}}{\ell!} \binom{r\ell+s}{n} = \frac{1}{n!} \sum_{\ell=0}^{n} S(n,\ell|1,r,s) r^{\ell} e^x x^{\ell}. \tag{2.26}$$

We will take Laplace transforms of both sides. Note that the Laplace transform with $\Re(w) > 1$ acts on monomials as

$$\int_0^{\infty} e^{-wx} x^{\ell} dx = \frac{\ell!}{w^{\ell+1}},$$

so that

$$\int_0^{\infty} e^{-wx} e^x x^{\ell} dx = \frac{\ell!}{(w-1)^{\ell+1}}.$$

Laplace transforming both sides gives

$$\sum_{\ell=0}^{\infty} \frac{1}{w^{\ell+1}} \binom{r\ell+s}{n} = \frac{1}{n!} \sum_{\ell=0}^{n} S(n,\ell|1,r,s) r^{\ell} \int_0^{\infty} e^{(1-w)x} x^{\ell} dx \tag{2.27}$$

$$= \frac{1}{n!} \sum_{\ell=0}^{n} S(n,\ell|1,r,s) r^{\ell} \frac{\ell!}{(w-1)^{\ell+1}} \tag{2.28}$$

with $\Re(w) > 1$. Now mapping $w \mapsto 1/w$ gives

$$\sum_{\ell=0}^{\infty} w^{\ell+1}\binom{r\ell+s}{n} = \frac{1}{n!}\sum_{\ell=0}^{n} \ell! S(n,\ell|1,r,s) r^{\ell} \frac{w^{\ell+1}}{(1-w)^{\ell+1}}$$

with $\Re(w) < 1$. Dividing by w, mapping $w \mapsto w^r$, and dividing by w again gives the first result.

For the second result with the $\frac{1}{\ell+1}$ factor, we again begin with

$$\sum_{\ell=0}^{\infty} \frac{x^{\ell}}{\ell!}\binom{r\ell+s}{n} = \frac{1}{n!}\sum_{\ell=0}^{n} S(n,\ell|1,r,s) r^{\ell} e^{x} x^{\ell},$$

separate out the $\ell = 0$ terms on both sides, shift ℓ by 1, and divide through by x:

$$\binom{s}{n} + \sum_{\ell=1}^{\infty} \frac{x^{\ell}}{\ell!}\binom{r\ell+s}{n} = \frac{1}{n!} S(n,0|1,r,s) e^{x} + \frac{1}{n!}\sum_{\ell=1}^{n} S(n,\ell|1,r,s) r^{\ell} e^{x} x^{\ell}$$

and

$$\frac{1}{x}\binom{s}{n} + \sum_{\ell=0}^{\infty} \frac{x^{\ell}}{(\ell+1)!}\binom{r\ell+r+s}{n} = \frac{1}{n!} S(n,0|1,r,s)\frac{e^{x}}{x} + \frac{1}{n!}\sum_{\ell=0}^{n-1} S(n,\ell+1|1,r,s) r^{\ell+1} e^{x} x^{\ell}.$$

Note that the Laplace transform of $1/x$ does not exist, so we need both of the initial terms to drop. When $0 \le s < n$ is an integer, the binomial coefficient evaluates to 0 and $S(n,0|1,r,s) = s(s-1)\cdots(s-n+1) = 0$. Now Laplace transform both sides with $\Re(w) > 1$:

$$\sum_{\ell=0}^{\infty} \frac{1}{(\ell+1)w^{\ell+1}}\binom{r\ell+r+s}{n} = \frac{1}{n!}\sum_{\ell=0}^{n-1} S(n,\ell+1|1,r,s) r^{\ell+1} \frac{\ell!}{(w-1)^{\ell+1}}.$$

Map $w \mapsto 1/w$, so $\Re(w) < 1$, divide by w, and map $s \mapsto s - r$ so that $r \le s < n + r$:

$$\sum_{\ell=0}^{\infty} \frac{w^{\ell}}{\ell+1}\binom{r\ell+s}{n} = \frac{1}{n!}\sum_{\ell=0}^{n-1} S(n,\ell+1|1,r,s-r) r^{\ell+1} \ell! \frac{w^{\ell}}{(1-w)^{\ell+1}}. \tag{2.29}$$

Map $w \mapsto w^r$ and divide by w:

$$\sum_{\ell=0}^{\infty} \frac{w^{r\ell-1}}{\ell+1}\binom{r\ell+s}{n} = \frac{1}{n!}\sum_{\ell=0}^{n-1} \ell! S(n,\ell+1|1,r,s-r) r^{\ell+1} \frac{w^{r\ell-1}}{(1-w^{r})^{\ell+1}}$$

to finish. □

An alternate proof of Lemma 19 proceeds by starting with Equation (2.25), clearing denominators by $(1-w^r)^{n+1}$, using the binomial theorem on $(1-w^r)^{n-\ell}$, and inserting the closed form expression for generalized Stirling numbers from Equation (2.6). Then we can switch the order of summation in the triple sum and evaluate the innermost sum using a classical binomial identity to get back down to a double sum.

Combining all of these lemmas proves Theorem 4, giving closed forms for the $(k+1)$-st derivative of $x^{k-r}H(x^r)$.

2.4. **Special cases.**

We will simplify the cases $r = 1, k$ which Yuster originally studied in [Yus23]. This will prove Corollaries 56. The following sequence, which has been studied many times, makes an appearance.

Definition 21. Define the ***s*-binomial coefficients** through the generating function

$$\sum_{\ell=0}^{ks} \binom{k}{\ell}_{s} x^{\ell} := (1 + x + x^2 + \cdots + x^s)^k = \left(\frac{1-x^{s+1}}{1-x}\right)^{k}. \tag{2.30}$$

A 1731 result of de Moivre [dM31] gives the closed form

$$\binom{k}{\ell}_{s-1} = \sum_{v=0}^{\lfloor \ell/s \rfloor} (-1)^v \binom{k}{v}\binom{\ell - vs + k - 1}{k-1}, \tag{2.31}$$

where the restriction $v \le \lfloor \ell/s \rfloor$ comes from setting $\ell - s + k - 1 \ge k - 1$ so that the second binomial coefficient is positive.

We repeat the statement of Corollary 6. Equation (2.34) proves an observation of Yuster that the coefficients are given by OEIS sequence A108267, which is $\binom{k}{\ell k}_{k-1}$.

Corollary 22. *Consider real $0 < x < 1$ and $\omega = e^{\frac{2\pi i}{k}}$ a primitive k-th root of unity. In terms of s-binomial coefficients defined in Definition* (2.30)*,*

$$\left(\frac{d}{dx}\right)^{k+1} H(x^k) = -k \cdot k! \sum_{\ell=0}^{\infty} \binom{k + k\ell - 1}{k-1} x^{k\ell - 1} \tag{2.32}$$

$$= -\frac{k!}{x} \sum_{j=0}^{k-1} \frac{1}{(1-\omega^j x)^k} \tag{2.33}$$

$$= -\frac{k \cdot k!}{x(1-x^k)^k} \sum_{\ell=0}^{k-1} \binom{k}{\ell k}_{k-1} x^{k\ell}. \tag{2.34}$$

Proof. Specializing Theorem 4 to $r = k$ and noting the binomial coefficient identity

$$\binom{k+k\ell}{k}\frac{1}{\ell+1} = \frac{k+k\ell}{k}\binom{k+k\ell-1}{k-1}\frac{1}{\ell+1} = \binom{k+k\ell-1}{k-1}$$

proves Equation (2.32). Now let $\omega = e^{2\pi i/k}$ be a primitive k-th root of unity and write

$$\begin{aligned}\left(\frac{d}{dx}\right)^{k+1} H(x^k) &= -\frac{k \cdot k!}{x} \sum_{\ell=0}^{\infty} \binom{k+k\ell-1}{k\ell} x^{k\ell} \\ &= -\frac{k \cdot k!}{x} \sum_{\ell=0}^{\infty} \binom{k+\ell-1}{\ell} x^{\ell} \mathbb{1}\left[\ell \equiv 0 \pmod{k}\right] \\ &= -\frac{k!}{x} \sum_{\ell=0}^{\infty} x^{\ell} \binom{k+\ell-1}{\ell} \sum_{j=0}^{k-1} \omega^{j\ell}.\end{aligned}$$

We divided by k since the inner sum along roots of unity is zero unless $\ell \equiv 0 \pmod{k}$, in which case it is k. Now, we use the generalized binomial theorem to deduce Equation (2.33)

$$-\frac{k!}{x}\sum_{\ell=0}^{\infty} x^{\ell}\binom{k+\ell-1}{\ell}\sum_{j=0}^{k-1}\omega^{j\ell} = -\frac{k!}{x}\sum_{j=0}^{k-1}\sum_{\ell=0}^{\infty}\binom{k+\ell-1}{\ell}(\omega^j x)^{\ell} = -\frac{k!}{x}\sum_{j=0}^{k-1}\frac{1}{(1-\omega^j x)^k}.$$

Now note that if $F(z) = \sum_{n=0}^{\infty} a_n z^n$, we have $\sum_{n=0}^{\infty} a_{kn} z^{kn} = \frac{1}{k}\sum_{j=0}^{k-1} F(w^j z)$, where ω is a primitive k-th root of unity. Then by setting $F(x) = \left(\frac{1-x^k}{1-x}\right)^k$ to be the generating function of $\binom{k}{\ell}_{k-1}$, we have

$$\sum_{\ell=0}^{k}\binom{k}{k\ell}_{k-1} x^{k\ell} = \frac{1}{k}\sum_{j=0}^{k-1}\left(\frac{1-(\omega^j x)^k}{1-\omega^j x}\right)^k = \frac{(1-x^k)^k}{k}\sum_{j=0}^{k-1}\frac{1}{(1-\omega^j x)^k}, \tag{2.35}$$

which proves Equation (2.34). Note that this essentially computed the Fourier expansion of the $(k-1)$-binomial generating function. □

Corollary 23. *We have*

$$\left(\frac{d}{dx}\right)^{k+1} x^{k-1}H(x) = \frac{(k-1)!}{x^2}\left(1-\frac{1}{(1-x)^k}\right). \tag{2.36}$$

Proof. Consider Equation (1.6) with $r=1$, so that

$$\left(\frac{d}{dx}\right)^{k+1} x^{k-1}H(x) = -k!\sum_{\ell=0}^{\infty}\binom{k+\ell}{k}\frac{1}{\ell+1}x^{\ell-1}.$$

Now use the generalized binomial theorem to show

$$\sum_{\ell=0}^{\infty}\binom{k+\ell}{k}\frac{1}{\ell+1}x^{\ell} = \frac{1}{k}\sum_{\ell=0}^{\infty}\binom{k+\ell}{\ell+1}x^{\ell} = \frac{1}{k}\sum_{\ell=1}^{\infty}\binom{k+\ell-1}{\ell}x^{\ell-1} = \frac{1}{kx}\left(\frac{1}{(1-x)^k}-1\right),$$

and we are done. □

2.5. **Real rootedness reduction.** We finally show that our main real rootedness conjecture inequality (1.2) for real exponents.

Theorem 24. *The real rootedness Conjecture 2 implies the entropy inequality of Conjecture 1 for all real $k \geq 1$.*

Proof. Our proof follows the framework of [Yus23], but with the extra parameter r. The flexibility given by the extra r parameter is crucial to proving the reduction for real exponents, as opposed to integer exponents. Consider the function

$$f_{k,r}(x) := \alpha H(x^k) - x^{k-r}H(x^r),$$

where $\alpha := \alpha_{k/r}$ satisfies the function equation (1.1) with parameter k/r, which is equivalent to

$$\alpha^r = \frac{1}{(1+\alpha)^{k-r}}. \tag{2.37}$$

We omit the subscript in $\alpha_{k/r}$ for clarity. Our goal is to compute all roots of $f_{k,r}(x)$ in $[0,1]$.

We have a trivial root at $x=1$ since $H(1)=0$.

We have a double root at $\frac{1}{(1+\alpha)^{1/r}}$ since we can calculate that $f_{k,r}\left(\frac{1}{(1+\alpha)^{1/r}}\right) = f'_{k,r}\left(\frac{1}{(1+\alpha)^{1/r}}\right) = 0$. Using the symmetry $H(x)=H(1-x)$ and the functional equation for α, we have

$$\begin{aligned} f_{k,r}\left(\frac{1}{(1+\alpha)^{1/r}}\right) &= \alpha H\left(\frac{1}{(1+\alpha)^{k/r}}\right) - \frac{1}{(1+\alpha)^{k/r-1}}H\left(\frac{1}{1+\alpha}\right) \\ &= \alpha H\left(\frac{\alpha}{1+\alpha}\right) - \alpha H\left(\frac{1}{1+\alpha}\right) \\ &= 0. \end{aligned}$$

We now compute the derivative

$$\frac{1}{x^{k-r-1}}\frac{d}{dx}f_{k,r}(x) = \alpha k x^r \log\left(\frac{1-x^k}{x^k}\right) - kx^r\log\left(\frac{1-x^r}{x^r}\right) + (k-r)\log(1-x^r). \tag{2.38}$$

Using the functional equation for α several times, at $x = \frac{1}{(1+\alpha)^{1/r}}$ we have

$$1-x^r = \frac{\alpha}{1+\alpha}, \quad \frac{1-x^r}{x^r} = \alpha, \quad \frac{1-x^k}{x^k} = (1+\alpha)^{k/r}-1 = \frac{1+\alpha}{\alpha}-1 = \frac{1}{\alpha}.$$

Now note that

$$(k-r)\log\left(\frac{\alpha}{1+\alpha}\right) = k\log\left(\frac{\alpha}{1+\alpha}\right) - \log\left(\frac{\alpha}{1+\alpha}\right)^r = k\log\frac{\alpha}{1+\alpha} - \log\frac{1}{(1+\alpha)^k} = k\log\alpha.$$

Substituting this into the derivative (2.38) gives

$$\begin{aligned}(1+\alpha)^{\frac{k-r-1}{r}} f'_{k,r}\left(\frac{1}{(1+\alpha)^{1/r}}\right) &= k\frac{\alpha}{1+\alpha}\log\frac{1}{\alpha} - \frac{k}{1+\alpha}\log\alpha + (k-r)\log\left(\frac{\alpha}{1+\alpha}\right)\\ &= -k\frac{\alpha}{1+\alpha}\log\alpha - \frac{k}{1+\alpha}\log\alpha + k\log\alpha\\ &= 0.\end{aligned}$$

We also have a root of multiplicity k at $x=0$. Equation (2.22) states that

$$x^{k-r}H(x^r) = -rx^k\log x + x^k - \sum_{\ell=1}^{\infty}\frac{x^{k+r\ell}}{\ell(\ell+1)},$$

so that for $0 \le t \le k-1$ we have

$$\left(\frac{d}{dx}\right)^t x^{k-r}H(x^r)\bigg|_{x=0} = -r\left(\frac{d}{dx}\right)^t x^k\log x\bigg|_{x=0}.$$

Using the iterated product rule and separating the term at $\ell = 0$ gives

$$\begin{aligned}\left(\frac{d}{dx}\right)^t x^k\log x &= -r\sum_{\ell=0}^{t}\binom{t}{\ell}\left(\frac{d}{dx}\right)^{t-\ell}x^k\cdot\left(\frac{d}{dx}\right)^{\ell}\log x\\ &= -r\frac{k!}{(k-t)!}x^{k-t}\log x - r\sum_{\ell=1}^{t}\binom{t}{\ell}\frac{k!}{(k-t+\ell)!}x^{k-t+\ell}\frac{(-1)^{\ell-1}}{x^{\ell}}\\ &= -r\frac{k!}{(k-t)!}x^{k-t}\log x - r\sum_{\ell=1}^{t}(-1)^{\ell-1}\binom{t}{\ell}\frac{k!}{(k-t+\ell)!}x^{k-t}.\end{aligned}$$

Irrespective of the value of r, for $0 \le t \le k-1$ we have $\lim_{x\to 0} x^{k-t}\log x = 0$, which in turn means that

$$\left(\frac{d}{dx}\right)^t f_{k,r}(x)\bigg|_{x=0} = 0.$$

Now, we appeal to Theorem 4, which states that

$$\begin{aligned}\left(\frac{d}{dx}\right)^{k+1} f_{k,r}(x) &= -\alpha\frac{k\cdot k!}{x(1-x^k)^k}h_{k,k}(x) + \frac{r\cdot k!}{x(1-x^r)^k}h_{k,r}(x)\\ &= -\frac{k!}{x(1-x^r)^k(1-x^k)^k}\left(\alpha k(1-x^r)^k h_{k,k}(x) - r(1-x^k)^k h_{k,r}(x)\right),\end{aligned}$$

where

$$h_{k,r}(x) = \sum_{j=0}^{k-1}x^{rj}\sum_{v=0}^{j}\frac{(-1)^{j-v}}{v+1}\binom{rv+k}{k}\binom{k}{j-v}$$

as before. Now assume the conjecture that the numerator has two real roots in $0 < x < 1$. By Rolle's theorem applied $k+1$ times to the $(k+1)$-st derivative, it follows that $f_{k,r}(x)$ contains at most $k+3$ roots in $[0,1]$, counting multiplicity. We have a trivial root at $x=1$, a double root at $x = \frac{1}{(1+\alpha)^{1/r}}$, and a root of multiplicity k at $x=0$. Therefore, we have found all $k+3$ roots of $f_{k,r}(x)$ in $[0,1]$.

Because $f_{k,r}(x)$ has a double root at $\frac{1}{(1+\alpha)^{1/r}}$, and the other roots are at the endpoints of the interval $[0,1]$, it must be either non-positive or non-negative on $[0,1]$. Yuster [Yus23, Lemma 3.3] showed that there is a small ε such that $f_{k,1}(x) > 0$ for $0 < x < \varepsilon$ and integers $k \ge 2$. The exact same proof shows that there is an ε_r so that $f_{k,r}(x^{1/r}) > 0$ for $0 < x < \varepsilon_r$ and $k/r > 1$. Since $f_{k,r}$ takes a positive value, it must be non-negative on $[0,1]$.

Given that $f_{k,r}(x) = \alpha_{k/r}H(x^k) - x^{k-r}H(x^r) \ge 0, 0 \le x \le 1$, we now map $x \mapsto x^{1/r}$, which sends $[0,1]$ to $[0,1]$. Therefore $\alpha_{k/r}H\left(x^{k/r}\right) - x^{k/r-1}H(x) \ge 0$. However we picked $k > r \ge 1$ as arbitrary coprime integers, so that k/r runs through all rationals greater than 1, and the inequality $\alpha_q H(x^q) - x^{q-1}H(x) \geq 0$ holds for all rational $q > 1$. Since each term $\alpha_q, H(x^q), x^{q-1}$ is continuous in $q > 1$, the inequality must also hold for all real $q > 1$. The inequality is also trivial at $q = 1$, which finishes the proof. □

The previous proof shows that if we can verify that $p_{k,r}(x) := \alpha_{k/r} k(1-x^r)^k h_{k,k}(x) - r(1-x^k)^k h_{k,r}(x)$ has two roots in $(0,1)$ for a fixed pair of integers k, r, then we have verified inequality (1.2) for the rational exponent k/r. For instance, at $k = 3, r = 2, \alpha_{3/2} \approx 0.754878$, this polynomial is

$$\begin{aligned} p_{3,2}(x) &= 3\alpha_{3/2}\left(-x^{12} + 3x^{10} - 7x^9 - 3x^8 + 21x^7 - 21x^5 + 3x^4 + 7x^3 - 3x^2 + 1\right) \\ &\quad - \left(\frac{2x^{13}}{3} - 4x^{11} - 2x^{10} - 2x^9 + 12x^8 + 2x^7 + 6x^6 - 12x^5 - \frac{2x^4}{3} - 6x^3 + 4x^2 + 2\right). \end{aligned}$$

This has two real roots in $(0,1)$ at $\approx 0.204863, 0.74186$, which proves the main entropy inequality (1.2) for the fractional exponent $3/2$.

Also note that we can factor $(1-x)^k$ out of $p_{k,r}(x)$, while still leaving a polynomial. Equivalently, we can to show that

$$\alpha_{k/r} k \left(\frac{1-x^r}{1-x}\right)^k h_{k,k}(x) - r\left(\frac{1-x^k}{1-x}\right)^k h_{k,r}(x)$$

has two real roots in $(0,1)$, counting multiplicity. The term $\left(\frac{1-x^r}{1-x}\right)^k$ is the generating function for $(r-1)$-binomial coefficients given in Definition (2.30). The $r = 1$ case of this factored polynomial is exactly the polynomial $p_k(x)$ of Yuster [Yus23, Corollary 3.7] which arose in his study of inequality (1.2) for integer k. The $k = 2, r = 1$ case is additionally the polynomial $p(x)$ of Boppana [Bop23].

2.6. **Functional equation.** We now collect some useful properties of α_k, including basic bounds and first order asymptotics. Recall that α_k satisfies the functional equation (1.1)

$$\alpha_k = \frac{1}{(1+\alpha_k)^{k-1}}.$$

Note that the following result is tight since $\lim_{k\to 1^+} \alpha_k = 1$.

Lemma 25. *For real $k > 1$, α_k monotonically decreases in k and satisfies*

$$\frac{1}{k} < \alpha_k < 1. \tag{2.39}$$

Proof. Consider the functional equation $x_k + x_k^k = 1$, written in terms of $x_k = \frac{1}{1+\alpha_k}$. This is monotonic in $0 < x_k < 1$ so has a unique solution in $(0,1)$, which corresponds to a unique value of α_k in $(0,1)$ satisfying (1.1). If k increases, the power $0 < x_k^k < 1$ decreases, so x_k must monotonically increase. Then $\alpha_k = \frac{1}{x_k} - 1$ monotonically decreases. Noting that $\lim_{k\to 1^+} \alpha_k = 1$ gives the upper bound.

Assume $\alpha_k \leq 1/k$, then $x_k = \frac{1}{1+\alpha_k} \geq \frac{k}{k+1}$. Then we apply Bernoulli's (strict) inequality to $x_k + x_k^k \geq \frac{k}{k+1} + \left(1 - \frac{1}{k+1}\right)^k > \frac{k}{k+1} + \frac{1}{k+1} = 1$, which contradicts the functional equation $x_k + x_k^k = 1$ and gives the lower bound. □

We can compute the large k asymptotics of α_k. Note that $b_k \approx \log\log k$ to first order, but there are multiplicative corrections of order $\frac{1}{\log k}, \frac{1}{\log^2 k}, \ldots$. The point of making b_k the solution to an exact equation is that the remaining error term in Lemma 26 is much smaller.

Lemma 26. *Let b_k be the unique solution to*

$$b_k - \log\left(1 - \frac{b_k}{\log k}\right) = \log\log k. \tag{2.40}$$

In the large k limit, we have

$$\alpha_k = \frac{\log k - b_k}{k} + O\left(\frac{\log^2 k}{k^2}\right) \tag{2.41}$$

$$= \frac{\log k}{k} + O\left(\frac{\log\log k}{k}\right). \tag{2.42}$$

Proof. We will do our calculations in $x_k = \frac{1}{1+\alpha_k}$, which is the unique solution of $x_k + x_k^k = 1$.

We will guess for now that $x_k = 1 - \frac{\log k - \delta}{k}$ for $\delta \in [0, 2\log\log k]$. We will see below that there is a solution x_k of this form, which must be the unique solution. We calculate

$$\log x_k = \log\left(1 - \frac{\log k - \delta}{k}\right) = -\frac{\log k - \delta}{k} + O\left(\frac{\log^2 k}{k^2}\right),$$

$$\log x_k^k = \delta - \log k + O\left(\frac{\log^2 k}{k}\right).$$

Moreover

$$\log(1 - x_k) = \log(\log k - \delta) - \log k = \log\log k + \log\left(1 - \frac{\delta}{\log k}\right) - \log k.$$

The equation $x_k + x_k^k = 1$ implies $\log x_k^k = \log(1 - x_k)$, so

$$\delta - \log k + O\left(\frac{\log^2 k}{k}\right) = \log\log k + \log\left(1 - \frac{\delta}{\log k}\right) - \log k,$$

which rearranges to

$$\delta - \log\left(1 - \frac{\delta}{\log k}\right) = \log\log k + O\left(\frac{\log^2 k}{k}\right).$$

This equation has a solution $\delta \in [0, 2\log\log k]$ by the intermediate value theorem, and by inspection $\delta = b_k + O(\log^2 k/k)$. Therefore

$$x_k = 1 - \frac{\log k - b_k + O(\log^2 k/k)}{k},$$

which implies the estimate on $\alpha_k = \frac{1-x_k}{x_k}$. □

Finally, we can give a series expansion for x_k using Lagrange inversion. Note that the lower index of the binomial coefficient is j, as opposed to the kj which appears in the definition of $h_{k,r}(x)$.

Lemma 27. *We have the following series expansion for α_k:*

$$x_k^N = \frac{1}{(1+\alpha_k)^N} = \sum_{j=0}^{\infty} (-1)^j \frac{N}{(k-1)j + N}\binom{kj + N - 1}{j}. \tag{2.43}$$

Proof. Rewrite the functional equation $x_k + x_k^k = 1$ as $x_k = \frac{1}{1+x_k^{k-1}}$. Consider $x_k(z)$ given as the solution of

$$x_k(z) = \frac{z}{1 + x_k(z)^{k-1}}.$$

We now perform Lagrange inversion along the variable z in $x_k(z)$ before setting $z = 1$, following [Ges16, Equation (2.2.1)]. We have

$$[z^n]x_k(z)^N = \frac{N}{n}\left[t^{n-N}\right]\frac{1}{(1+t^{k-1})^n} = \frac{N}{n}\left[t^{n-N}\right]\sum_{j=0}^{\infty}\binom{n-1+j}{j}(-1)^j t^{(k-1)j}.$$

The inner coefficient is only nonzero when $n - N = (k-1)j$, or when $n = (k-1)j + N$ for some j. Therefore

$$\begin{aligned} x_k(z)^N &= \sum_{n=0}^{\infty} z^n \frac{N}{n} \left[t^{n-N}\right] \frac{1}{(1+t^{k-1})^n} \\ &= \sum_{j=0}^{\infty} z^{(k-1)j+N} \frac{N}{(k-1)j+N} \binom{kj+N-1}{j} (-1)^j. \end{aligned}$$

Now setting $z = 1$ recovers x_k^N. □

3. Randomized Symmetric Binary Perceptron

We provide proofs of all results mentioned in the introduction, as well as plots of numerical experiments.

3.1. **Proof of Theorem 8.**

Proof of Theorem 8. Our proof is based on the *first moment method*: note that by Markov's inequality,

$$\mathbb{P}\big[\big|S_\alpha(\kappa,p)\big|\ge 1\big]\le \mathbb{E}[\big|S_\alpha(\kappa,p)\big|],$$

so that $S_\alpha(\kappa,p)=\varnothing$ w.h.p. if $\mathbb{E}\big[\big|S_\alpha(\kappa,p)\big|\big]=o(1)$. So, the remainder of proof estimates $\mathbb{E}\big[\big|S_\alpha(\kappa,p)\big|\big]$. Fix any $\boldsymbol{\sigma}\in\Sigma_n$ and let $Z_i(\boldsymbol{\sigma})=\mathbb{1}\big\{Y_i=U(\langle\boldsymbol{\sigma},X_i\rangle)\big\}$. Then,

$$|S_\alpha(\kappa,p)|=\sum_{\boldsymbol{\sigma}\in\Sigma_n}Z(\boldsymbol{\sigma}),\quad\text{where}\quad Z(\boldsymbol{\sigma})=\prod_{1\le i\le M}Z_i(\boldsymbol{\sigma}).$$

Now fix any $\boldsymbol{\sigma}\in\Sigma_n$ and observe that $Z_1(\boldsymbol{\sigma}),\dots,Z_M(\boldsymbol{\sigma})$ are i.i.d. Bernoulli. Moreover, $\langle\boldsymbol{\sigma},X_i\rangle\sim\mathcal{N}(0,n)$. So,

$$\begin{aligned}\mathbb{P}[Z_i(\boldsymbol{\sigma})=1]&=\mathbb{P}[Z_i(\boldsymbol{\sigma})=1|Y_i=1]\mathbb{P}[Y_i=1]+\mathbb{P}[Z_i(\boldsymbol{\sigma})=1|Y_i=0]\mathbb{P}[Y_i=0]\\&=p\mathbb{P}[|\langle\boldsymbol{\sigma},X_i\rangle|\le\kappa\sqrt{n}]+(1-p)\mathbb{P}[|\langle\boldsymbol{\sigma},X_i\rangle|>\kappa\sqrt{n}]\\&=pq(\kappa)+(1-p)(1-q(\kappa)).\end{aligned}$$

Thus, $\mathbb{E}\big[\big|S_\alpha(\kappa,p)\big|\big]=\exp_2\big(nf(\alpha,p,\kappa)\big)$ where

$$f(\alpha,p,\kappa)=1+\alpha\log_2\big(pq(\kappa)+(1-p)(1-q(\kappa))\big).$$

As $f(\alpha,p,\kappa)>0$ iff $\alpha<\alpha_c(\kappa)$ the proof is complete. □

3.2. **Proof of Theorem 9.**

Proof of Theorem 9. The proof is quite similar to that of Theorem 8; we only point out necessary modifications. Define $\widetilde{Z}(\boldsymbol{\sigma})=\prod_{1\le i\le M}\widetilde{Z}_i(\boldsymbol{\sigma})$, where $\widetilde{Z}_i(\boldsymbol{\sigma})=\mathbb{1}\{Y_i=U(\langle\boldsymbol{\sigma},X_i\rangle)\}$ for $1\le i\le M$. Let $\mathcal{I}_t$, $1\le t\le\binom{M}{Mp}$ be the subsets of $[M]$ of size Mp. Notice that $\mathbb{P}[\widetilde{Z}(\boldsymbol{\sigma})=1|\mathcal{I}=\mathcal{I}_t]=\prod_{i\in\mathcal{I}_t}\mathbb{P}[|\langle\boldsymbol{\sigma},X_i\rangle|\le\kappa\sqrt{n}]\cdot\prod_{i\in[M]\setminus\mathcal{I}_t}\mathbb{P}[|\langle\boldsymbol{\sigma},X_i\rangle|>\kappa\sqrt{n}]=q(\kappa)^{Mp}(1-q(\kappa))^{M(1-p)}$, using the fact $\langle\boldsymbol{\sigma},X_i\rangle\sim\mathcal{N}(0,n)$ and the independence of $X_1,\dots,X_M$. Hence,

$$\mathbb{P}[\widetilde{Z}(\boldsymbol{\sigma})=1]=\sum_{t=1}^{\binom{M}{Mp}}\binom{M}{Mp}^{-1}\mathbb{P}[\widetilde{Z}(\boldsymbol{\sigma})=1|\mathcal{I}=\mathcal{I}_t]=q(\kappa)^{Mp}(1-q(\kappa))^{M(1-p)}.$$

As $M=\alpha n$, we immediately obtain $\mathbb{E}\big[\big|\widetilde{S}_\alpha(\kappa,p)\big|\big]=\exp_2\big(n\tilde{f}(\alpha,p,\kappa)\big)$, where

$$\tilde{f}(\alpha,p,\kappa)=1+\alpha\big(p\log_2 q(\kappa)+(1-p)\log_2(1-q(\kappa))\big).$$

This yields Theorem 9. □

3.3. **Proof of Theorem 10.**

Proof of Theorem 10. We show the extension for Theorem 8; that of Theorem 9 is analogous. Our argument is based on the Berry-Esseen inequality [Ber41, Ess42], reproduced below for convenience.

Theorem 28. *There exists an absolute constant $C>0$ such that the following holds. Let $T_1,\dots,T_n$ be i.i.d. random variables with $\mathbb{E}[T_1]=0$, $\mathbb{E}[T_1^2]=\sigma^2>0$ and $\mathbb{E}[|T_1|^3]=\rho<\infty$. Then, for $Z\sim\mathcal{N}(0,1)$,*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left[\frac{T_1+\cdots+T_n}{\sigma\sqrt{n}}\le x\right]-\mathbb{P}[Z\le x]\right|\le\frac{C\rho}{\sigma^3\sqrt{n}}.$$

Equipped with Theorem 28, fix any $i \in [M]$ and let $X_i = (X_i(j) : j \in [n])$ with $\mathbb{E}[X_i(j)] = 0$, $\mathbb{E}[X_i(j)^2] = \sigma^2$ (where $\sigma > 0$) and $\mathbb{E}[|X_i(j)|^3] = \rho < \infty$. Note that

$$\begin{aligned}\mathbb{P}[Y_i = U(\langle \boldsymbol{\sigma}, X_i \rangle)] &= \mathbb{P}[Y_i = U(\langle \boldsymbol{\sigma}, X_i \rangle) | Y_i = 1]\mathbb{P}[Y_i = 1] + \mathbb{P}[Y_i = U(\langle \boldsymbol{\sigma}, X_i \rangle) | Y_i = 0]\mathbb{P}[Y_i = 0] \\ &= \mathbb{P}\big[|\langle \boldsymbol{\sigma}, X_i \rangle| \le \kappa\sqrt{n}\big]p + \mathbb{P}\big[|\langle \boldsymbol{\sigma}, X_i \rangle| > \kappa\sqrt{n}\big](1-p). \qquad (3.1)\end{aligned}$$

Let $q(\kappa) = \mathbb{P}[-\kappa \le Z \le \kappa]$ where $Z \sim \mathcal{N}(0,1)$. Applying Theorem 28 to $X_i(j)$, $j \in [n]$, together with the triangle inequality, we obtain

$$\Big|\mathbb{P}\big[|\langle \boldsymbol{\sigma}, X_i \rangle| \le \kappa\sqrt{n}\big] - q(\kappa)\Big| \le \frac{2C\rho}{\sigma^3\sqrt{n}} \triangleq \frac{\mathcal{C}}{\sqrt{n}}, \qquad (3.2)$$

where $\mathcal{C} = \frac{2C\rho}{\sigma^3} = O(1)$. Combining (3.1) and (3.2), we obtain

$$\mathbb{P}[Y_i = U(\langle \boldsymbol{\sigma}, X_i \rangle)] \le q(\kappa)p + (1 - q(\kappa))(1-p) + \frac{\mathcal{C}}{\sqrt{n}}. \qquad (3.3)$$

Recall now the Taylor expansion for logarithm: as $x \to 0$,

$$\log_2(1+x) = -\frac{x}{\log 2} + O(x^2). \qquad (3.4)$$

We now combine (3.3) with (3.4) to obtain

$$\begin{aligned}&\mathbb{P}[Y_i = U(\langle \boldsymbol{\sigma}, X_i \rangle)]^{\alpha n} \\ &\le \big(q(\kappa)p + (1-q(\kappa))(1-p)\big)^{\alpha n} \times \left(1 + \frac{\mathcal{C}}{\sqrt{n}\big(q(\kappa)p + (1-q(\kappa))(1-p)\big)}\right)^{\alpha n} \\ &= \exp_2\left(\alpha n \log_2\big(q(\kappa)p + (1-q(\kappa))(1-p)\big) + \alpha n \log_2\left(1 + \frac{\mathcal{C}}{\sqrt{n}\big(q(\kappa)p + (1-q(\kappa))(1-p)\big)}\right)\right) \\ &= \exp_2\Big(\alpha n \log_2\big(q(\kappa)p + (1-q(\kappa))(1-p)\big) + \Theta(\sqrt{n})\Big).\end{aligned}$$

With this, we obtain immediately that

$$\mathbb{E}\big|S_\alpha(\kappa, p)\big| = \exp_2\big(nf(\alpha, p, \kappa) + \Theta(\sqrt{n})\big).$$

The extension for Theorem 9 is similar. □

3.4. **Proof of Theorem 12.** We prove Theorem 12 contingent on an assumption regarding a certain real-valued function. It is worth noting that various related results in the field mentioned earlier were also established contingent on an analogous assumption, see e.g. [APZ19, Hypothesis 3], [PX21, Assumption 1], and [DS19, Condition 1.2].

Assumption 3.1. Following the notation in [APZ19], let

$$\begin{aligned}F_{r,\kappa,\alpha}(\beta) &\triangleq h(\beta) + \alpha \log_2 \mathbb{P}[|Z_1| \le \kappa, |Z_\beta| \le \kappa] \\ F_{u,\kappa,\alpha}(\beta) &\triangleq h(\beta) + \alpha \log_2 \mathbb{P}[|Z_1| > \kappa, |Z_\beta| > \kappa],\end{aligned}$$

where $Z_1, Z_\beta \sim \mathcal{N}(0,1)$ with correlation $2\beta - 1$ and $h(\beta)$ is the binary entropy function:

$$h(\beta) = -\beta \log_2 \beta - (1-\beta)\log_2(1-\beta).$$

Fix any $p \in [0,1]$ and set

$$F_{\kappa,\alpha,p}(\beta) = pF_{r,\kappa,\alpha}(\beta) + (1-p)F_{u,\kappa,\alpha}(\beta).$$

For any $\kappa > 0$ and $\alpha > 0$ with $F''_{\kappa,\alpha,p}(1/2) < 0$, there is at most one $\beta \in (1/2, 1)$ such that $F'_{\kappa,\alpha,p}(\beta) = 0$.

Several remarks are in order. Assumption 3.1 is analogous to [APZ19, Hypothesis 3], adopted both for the SBP (corresponding to $p = 1$ in our model) and for the UBP (corresponding to $p = 0$ in our model) therein. Furthermore, for the SBP, [APZ19, Hypothesis 3] has been verified by Abbe, Li, and Sly [ALS21b]. It is likely that their techniques adapt also to the UBP for a range of κ values, e.g. when $\kappa < \kappa^* \approx 0.817^2$. In light of these facts, as well as numerical studies reported in Section 3.11, Assumption 3.1 is indeed plausible. A rigorous verification is left for future work.

Equipped with these, we now start formally proving Theorem 12.

Proof of Theorem 12. Our proof is very similar to that of [APZ19, Proposition 6], and we use the identical notation whenever appropriate. Furthermore, we only prove the first part as the second part is identical.

The proof is based on the *second moment method.*

Lemma 29. *Let Z be an integer-valued random variable with $\mathbb{P}[Z \ge 0] = 1$. Then*

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

Lemma 29 is known as the Paley-Zygmund inequality, we provide a proof for completeness.

Proof of Lemma 29. Let $I = \mathbb{1}\{Z > 0\}$, thus $\mathbb{P}[Z > 0] = \mathbb{E}[I] = \mathbb{E}[I^2]$. We then conclude by applying Cauchy-Schwarz inequality:

$$\mathbb{P}[Z > 0]\mathbb{E}[Z^2] = \mathbb{E}[I^2]\mathbb{E}[Z^2] \ge \mathbb{E}[Z\mathbb{1}\{Z > 0\}]^2 = \mathbb{E}[Z]^2.$$

□

We next provide an auxiliary lemma, originally due to Achlioptas and Moore [AM02, Lemma 2]. The version below is reproduced from [APZ19, Lemma 8].

Lemma 30. *Let $g(\beta)$ be a real analytic function on $[0,1]$ and let*

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

Suppose that (a) $G(1/2) > G(\beta)$ for every $\beta \ne 1/2$ and (b) $G''(1/2) < 0$. Then, there exists constants $c_2 > c_1 > 0$ such that

$$c_2 G(1/2)^n \ge \sum_{0\le \ell \le n} \binom{n}{\ell} g(\ell/n)^n \ge c_1 G(1/2)^n.$$

In the remainder of the proof, we let $q(\kappa) \triangleq \mathbb{P}[|Z| \le \kappa]$ where $Z \sim \mathcal{N}(0,1)$.

Equipped with Lemmas 29 and 30, we let

$$Z = \big|\widetilde{S}_\alpha(\kappa,p)\big| = \sum_{\boldsymbol{\sigma}\in\Sigma_n} \mathbb{1}\big\{\boldsymbol{\sigma} \in \widetilde{S}_\alpha(\kappa,p)\big\}.$$

Theorem II.3 from the main text yields

$$\mathbb{E}[Z] = 2^n q(\kappa)^{p\alpha n}(1-q(\kappa))^{(1-p)\alpha n}. \tag{3.5}$$

3.5. **Second Moment Calculation.** Next, fix any $\beta \in [0,1]$, let $Z \sim \mathcal{N}(0,1)$ and $Z_\beta \sim \mathcal{N}(0,1)$ with $\mathbb{E}[Z_\beta Z] = 2\beta - 1$. Define

$$q_{r,\kappa}(\beta) = \mathbb{P}[|Z| \le \kappa, |Z_\beta| \le \kappa] \tag{3.6}$$

$$q_{u,\kappa}(\beta) = \mathbb{P}[|Z| > \kappa, |Z_\beta| > \kappa]. \tag{3.7}$$

[2] Above κ^*, the model exhibits *replica symmetry breaking* behaviour, see [APZ19] for details.

These are precisely the same quantities appearing in [APZ19, Equation 6]. Note that

$$Z^2 = \sum_{\boldsymbol{\sigma},\boldsymbol{\sigma'}\in\Sigma_n} \mathbb{1}\left\{\boldsymbol{\sigma}\in\widetilde{S}_\alpha(\kappa,p),\boldsymbol{\sigma'}\in\widetilde{S}_\alpha(\kappa,p)\right\}.$$

Taking expectations of both sides, we obtain

$$\mathbb{E}[Z^2] = 2^n\sum_{0\le\ell\le n}\binom{n}{\ell}q_{r,\kappa}(\beta)^{p\alpha n}q_{u,\kappa}(\beta)^{(1-p)\alpha n}.$$

Soon, we will apply Lemma 30 to $G_{\kappa,\alpha,p}(\beta)$ where

$$G_{\kappa,\alpha,p}(\beta) = \frac{q_{r,\kappa}(\beta)^{p\alpha}q_{u,\kappa}(\beta)^{(1-p)\alpha}}{\beta^\beta(1-\beta)^{1-\beta}}. \tag{3.8}$$

Suppose first that $G_{\kappa,\alpha,p}(\cdot)$ satisfies the conditions of Lemma 30. Then, we immediately obtain

$$\mathbb{E}[Z^2]\le c_2\cdot 2^n\cdot G(1/2)^n = c_2\cdot 4^n\cdot q(\kappa)^{2p\alpha n}\cdot(1-q(\kappa))^{2(1-p)\alpha n},$$

for some $c_2>0$. Observe that

$$q_{r,\kappa}(1/2) = q(\kappa)^2 \quad\text{and}\quad q_{u,\kappa}(1/2) = \big(1-q(\kappa)\big)^2.$$

Now recalling (3.5) and applying Lemma 29, we establish the desired result:

$$\liminf_{n\to\infty}\mathbb{P}\big[\widetilde{S}_\alpha(\kappa,p)\ne\varnothing\big] = \liminf_{n\to\infty}\mathbb{P}[Z>0]\ge\frac{\mathbb{E}[Z]^2}{\mathbb{E}[Z^2]}\ge\frac{1}{c_2}>0.$$

3.6. **Verifying Conditions of Lemma 30.** Hence, it suffices to verify that $G_{\kappa,\alpha}(\beta)$ defined in (3.8) satisfies the conditions of Lemma 30. We proceed analogously to [APZ19]. To that end, we let

$$G_{r,\kappa,\alpha}(\beta) = \frac{q_{r,\kappa}(\beta)^\alpha}{\beta^\beta(1-\beta)^{1-\beta}}$$

and

$$G_{u,\kappa,\alpha}(\beta) = \frac{q_{u,\kappa}(\beta)^\alpha}{\beta^\beta(1-\beta)^{1-\beta}},$$

and obtain

$$G_{\kappa,\alpha,p}(\beta) = G_{r,\kappa,\alpha}(\beta)^p G_{u,\kappa,\alpha}(\beta)^{1-p}. \tag{3.9}$$

We then set $G_{\kappa,\alpha,p}(\beta) = \exp\big(F_{\kappa,\alpha,p}(\beta)\big)$ as in the proof of [APZ19, Proposition 6] and observe, using (3.9), that

$$F_{\kappa,\alpha,p}(\beta) = pF_{r,\kappa,\alpha}(\beta)+(1-p)F_{u,\kappa,\alpha}(\beta), \tag{3.10}$$

where $F_{r,\kappa,\alpha}(\beta)$ is precisely the term arising in [APZ19, Equation 9] and $F_{u,\kappa,\alpha}(\beta)$ is the term defined in [APZ19, Section 2.2.2]. Note that a necessary condition is $F_{\kappa,\alpha,p}(1/2) > F_{\kappa,\alpha,p}(1)$ for all p, which boils down to the condition

$$\alpha < -\frac{1}{p\log_2 q(\kappa)+(1-p)\log_2\big(1-q(\kappa)\big)} = \widetilde{\alpha}_c(\kappa,p). \tag{3.11}$$

Next, we have $F^{''}_{\kappa,\alpha,p}(1/2) = pF^{''}_{r,\kappa,\alpha}(1/2)+(1-p)F^{''}_{u,\kappa,\alpha}(1/2)$. Using the expressions for $F^{''}_{r,\kappa,\alpha}(1/2)$ and $F^{''}_{u,\kappa,\alpha}(1/2)$ derived in [APZ19], we get

$$\begin{aligned}F^{''}_{\kappa,\alpha,p}(1/2) &= 4p\left(-1+\frac{2}{\pi}\frac{\alpha\kappa^2e^{-\kappa^2}}{q(\kappa)^2}\right)+4(1-p)\left(-1+\frac{2}{\pi}\frac{\alpha\kappa^2e^{-\kappa^2}}{\big(1-q(\kappa)\big)^2}\right)\\ &= -4+\alpha\cdot\frac{8}{\pi}\kappa^2e^{-\kappa^2}\left(\frac{p}{q(\kappa)^2}+\frac{1-p}{\big(1-q(\kappa)\big)^2}\right).\end{aligned}$$

So, it suffices to verify that

$$\alpha < \frac{\pi}{2\kappa^2 e^{-\kappa^2}}\left(\frac{p}{q(\kappa)^2}+\frac{1-p}{\big(1-q(\kappa)\big)^2}\right)^{-1} \tag{3.12}$$

to ensure $F''_{\kappa,\alpha,p}(1/2)<0$. We now establish our claim. Fix any $\kappa>0$. Note that the argument of [APZ19] shows

$$-\frac{1}{\log_2 q(\kappa)} < \frac{\pi}{2\kappa^2 e^{-\kappa^2}} q(\kappa)^2. \tag{3.13}$$

Define

$$\zeta(p,\kappa) = -\frac{1}{p\log_2 q(\kappa)+(1-p)\log_2\big(1-q(\kappa)\big)} - \frac{\pi}{2\kappa^2 e^{-\kappa^2}}\left(\frac{p}{q(\kappa)^2}+\frac{1-p}{\big(1-q(\kappa)\big)^2}\right)^{-1}. \tag{3.14}$$

Note that for any fixed $\kappa>0$, $p\mapsto\zeta(p,\kappa)$ is continuous. Furthermore, (3.13) yields $\zeta(1,\kappa)<0$. So, for any fixed κ, there is a p_κ^* for which $\zeta(p,\kappa)<0$ for every $p\in[p_\kappa^*,1]$. Now if $\zeta(p,\kappa)<0$, then we have

$$\zeta(p,\kappa)=\widetilde{\alpha}_c(\kappa,p)-\frac{\pi}{2\kappa^2 e^{-\kappa^2}}\left(\frac{p}{q(\kappa)^2}+\frac{1-p}{\big(1-q(\kappa)\big)^2}\right)^{-1}<0,$$

so that for any $\alpha<\widetilde{\alpha}_c(\kappa,p)$, (3.12) holds. We now verify that $F_{\kappa,\alpha,p}(\beta)$ is maximized at $\beta=1/2$, under Assumption 3.1. As $F_{\kappa,\alpha,p}$ is symmetric around $\beta=\frac{1}{2}$, it suffices to consider $\beta\in[1/2,1]$. Since $F'_{\kappa,\alpha,p}(1/2)=0$ and $F''_{\kappa,\alpha,p}(1/2)<0$, and $F_{\kappa,\alpha,p}$ has at most one critical point in $(1/2,1)$, it must attain its maxima either at $\beta=1/2$ or at $\beta=1$. Since $F_{\kappa,\alpha,p}(1/2)>F_{\kappa,\alpha,p}(1)$, as verified in (3.11), the conditions of Lemma 30 are satisfied.

The second part of the Theorem 12 is established similarly. In this case, an inequality analogous to (3.13) holds only when $\kappa<\kappa^*=0.817$, marking the onset of replica symmetric breaking, see [APZ19] for details. □

3.7. **Proof of Theorem 14.** Fix $\kappa>0$ and $\alpha>0$. Let $\Xi_\kappa(\beta,\eta,m,\alpha)$ be the set of all m-tuples $(\boldsymbol{\sigma}_1,\dots,\boldsymbol{\sigma}_m)$ such that:

- For any $1\le i\le\lfloor n\alpha\rfloor$ and $1\le j\le m$, $|\langle X_i,\boldsymbol{\sigma}_j\rangle|\le\kappa\sqrt{n}$, where $X_i\sim\mathcal{N}(0,I_n)$ are i.i.d.
- For $1\le i<i'\le m$, $n^{-1}\langle\boldsymbol{\sigma}_i,\boldsymbol{\sigma}_{i'}\rangle\in[\beta-\eta,\beta]$.

In other words, $\Xi_\kappa(\beta,\eta,m,\alpha)=\mathcal{F}(\beta,\eta,m,\alpha,\kappa,1)$. We next record the following result from [GKPX22, Theorem 2.4].

Theorem 31. [GKPX22, Theorem 2.4] *Let $\alpha^*_{\rm OGP}(\kappa)=10\kappa^2\log_2\frac{1}{\kappa}$. For any small enough κ and $\alpha\ge\alpha^*_{\rm OGP}(\kappa)$, there exists $m\in\mathbb{N}$ and $0<\eta<\beta<1$ such that*

$$\mathbb{P}\big[\Xi_\kappa(\beta,\eta,m,\alpha)\ne\varnothing\big]\le 2^{-\Theta(n)}.$$

3.8. **Part I: $\mathcal{F}(\beta,\eta,m,\alpha,\kappa,p)$.** Fix $\kappa>0$ small, $p\in[0,1]$, and $\alpha>\alpha_{\rm OGP}(\kappa,p)=(10\kappa^2/p)(\log_2(1/\kappa))$. Note first that $p\alpha>\alpha^*_{\rm OGP}(\kappa)=10\kappa^2\log_2\frac{1}{\kappa}$, thus there exists an $\epsilon>0$ such that

$$p\alpha(1-\epsilon)\ge\alpha^*_{\rm OGP}(\kappa).$$

Next, for i.i.d. $Y_i\sim{\rm Ber}(p)$, $1\le i\le M$, a standard concentration bound [Ver10, Ver18] yields

$$\mathbb{P}\left[\left|\frac{1}{M}\sum_{1\le i\le M}Y_i-p\right|\le p\epsilon\right]\ge 1-2^{-\Theta(M)}. \tag{3.15}$$

Set

$$T=\sum_{k\in\mathbb{N}:|k/M-p|\le p\epsilon}\binom{M}{k},$$

and let $\mathcal{I}_1, \dots, \mathcal{I}_T$ be an enumeration of all $A \subset [M]$ with $|A| \in [Mp(1-\epsilon), Mp(1+\epsilon)]$. Define the random variable $\mathcal{I} = \{i : Y_i = 1\}$. Using (3.15), we have

$$\mathbb{P}\big[\mathcal{I} \ne \mathcal{I}_i, 1 \le i \le T\big] \le 2^{-\Theta(M)}. \tag{3.16}$$

Next, for any $m \in \mathbb{N}$, $0 < \eta < \beta < 1$, we have

$$\mathbb{P}\big[\mathcal{F}(\beta, \eta, m, \alpha, \kappa, p) \ne \varnothing \mid \mathcal{I} = \mathcal{I}_t\big] \le \mathbb{P}\big[\Xi_\kappa(\beta, \eta, m, p\alpha(1-\epsilon)) \ne \varnothing\big]. \tag{3.17}$$

Here, (3.17) follows by combining the following facts:

- The labels Y_i are independent of X_i.
- $|\mathcal{I}_t| \ge p\alpha(1-\epsilon)n \ge \alpha^*_{\text{OGP}}(\kappa)n$.
- Conditional on $\mathcal{I} = \mathcal{I}_t$, any $(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m) \in \mathcal{F}(\beta, \eta, m, \alpha, \kappa, p)$ satisfies
$$|\langle X_i, \boldsymbol{\sigma}_j \rangle| \le \kappa\sqrt{n}, \quad \forall i \in \mathcal{I}_t, \ \forall 1 \le j \le m$$
where $X_i \sim \mathcal{N}(0, I_n), i \in \mathcal{I}_t$ are i.i.d.

As $p\alpha(1-\epsilon) \ge \alpha^*_{\text{OGP}}(\kappa)$, Theorem 31 immediately yields the existence of an $m^* \in \mathbb{N}$ and $0 < \eta^* < \beta^* < 1$ for which

$$\mathbb{P}\big[\Xi_\kappa(\beta^*, \eta^*, m^*, p\alpha(1-\epsilon)) \ne \varnothing\big] \le 2^{-\Theta(n)}.$$

Since the right hand side of (3.17) is independent of the choice of $\mathcal{I}_t$, we have

$$\max_{1 \le t \le T} \mathbb{P}\big[\mathcal{F}(\beta^*, \eta^*, m^*, \alpha, \kappa, p) \ne \varnothing \mid \mathcal{I} = \mathcal{I}_t\big] \le \mathbb{P}\big[\Xi_\kappa(\beta^*, \eta^*, m^*, p\alpha(1-\epsilon)) \ne \varnothing\big] \le 2^{-\Theta(n)}. \tag{3.18}$$

Now,

$$\begin{aligned}
&\mathbb{P}\big[\mathcal{F}(\beta^*, \eta^*, m^*, \alpha, \kappa, p) \ne \varnothing\big] \\
&= \sum_{1 \le i \le T} \mathbb{P}\big[\mathcal{F}(\beta^*, \eta^*, m^*, \alpha, \kappa, p) \ne \varnothing \mid \mathcal{I} = \mathcal{I}_i\big]\mathbb{P}[\mathcal{I} = \mathcal{I}_i] + \mathbb{P}\big[\mathcal{I} \ne \mathcal{I}_i, 1 \le i \le T\big] && (3.19) \\
&\le 2^{-\Theta(n)} \sum_{1 \le i \le M} \mathbb{P}[\mathcal{I} = \mathcal{I}_i] + 2^{-\Theta(M)} && (3.20) \\
&= 2^{-\Theta(n)} \cdot \mathbb{P}\big[\mathcal{I} \in \{\mathcal{I}_1, \dots, \mathcal{I}_T\}\big] + 2^{-\Theta(M)} && (3.21) \\
&= 2^{-\Theta(n)}, && (3.22)
\end{aligned}$$

where (3.19) follows from the fact that the events $\{\mathcal{I} = \mathcal{I}_i\}, 1 \le i \le T$ and $\{\mathcal{I} \ne \mathcal{I}_i, 1 \le i \le T\}$ collectively partition the probability space; (3.20) follows by combining (3.18) and (3.16); (3.21) uses the fact that the events $\{\mathcal{I} = \mathcal{I}_i\}$ are pairwise disjoint; and (3.22) uses the fact that $M = p\alpha n = \Theta(n)$ for $p, \alpha = O(1)$. This establishes Theorem 14 for $\mathcal{F}(\beta, \eta, m, \alpha, \kappa, p)$.

3.9. **Part II: $\widetilde{\mathcal{F}}(\beta, \eta, m, \alpha, \kappa, p)$.** Fix $\kappa > 0$ small, $p \in [0,1]$ and $\alpha \ge \alpha_{\text{OGP}}(\kappa, p)$. Similar to above, let $\mathcal{I}_t, 1 \le t \le \binom{M}{Mp}$ be the subsets of $[M]$ of size Mp and $\mathcal{I} = \{i : Y_i = 1\}$. Observe that for any $m \in \mathbb{N}$, $0 < \eta < \beta < 1$,

$$\mathbb{P}\big[\widetilde{\mathcal{F}}(\beta, \eta, m, \alpha, \kappa, p) \ne \varnothing \mid \mathcal{I} = \mathcal{I}_t\big] \le \mathbb{P}\big[\Xi_\kappa(\beta, \eta, m, p\alpha) \ne \varnothing\big], \tag{3.23}$$

using the facts (a) that the labels Y_i are independent of X_i and (b) that on $\mathcal{I} = \mathcal{I}_t$, $|\mathcal{I}_t| = p\alpha n$ and any $(\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_m) \in \widetilde{\mathcal{F}}(\beta, \eta, m, \alpha, \kappa, p)$ satisfies

$$|\langle X_i, \boldsymbol{\sigma}_j \rangle| \le \kappa\sqrt{n}, \quad \forall i \in \mathcal{I}_t, \forall 1 \le j \le m.$$

For $\alpha \geq \alpha_{\text{OGP}}(\kappa) = \frac{10}{p}\kappa^2 \log_2 \frac{1}{\kappa}$, Theorem 31 immediately yields the existence of an $m' \in \mathbb{N}$ and $0 < \eta' < \beta' < 1$ for which

$$\mathbb{P}\big[\Xi_\kappa(\beta', \eta', m', p\alpha) \neq \varnothing\big] \leq 2^{-\Theta(n)}.$$

Notice that the right hand side of (3.23) is independent of t. With this, we establish Theorem 14 for the $\widetilde{F}(\beta, \eta, m, \alpha, \kappa, p)$ through the law of total expectation and obtain

$$\mathbb{P}\big[\widetilde{\mathcal{F}}(\beta', \eta', m', \alpha, \kappa, p) \neq \varnothing\big] \leq 2^{-\Theta(n)}.$$

3.10. **Proof Sketch for Theorem 15.** Let $X_i \sim \mathcal{D}^{\otimes n}$ denotes the random vector $X_i \in \mathbb{R}^n$ with i.i.d. coordinates drawn from $\mathcal{D}$. For $\Xi_\kappa(\beta, \eta, m, \alpha)$ defined in Section 3.7, [GKPX22] establishes the following:

Theorem 32. [GKPX22, Theorem 5.2]

$$\mathbb{E}_{X_i \sim \mathcal{D}^{\otimes n}, 1 \leq i \leq M \ i.i.d.}\big[\big|\Xi_\kappa(\beta, \eta, m, \alpha)\big|\big] \leq \mathbb{E}_{X_i \sim \mathcal{N}(0, I_n), 1 \leq i \leq M \ i.i.d.}\big[\big|\Xi_\kappa(\beta, \eta, m, \alpha)\big|\big] \exp\big(\Theta(\sqrt{n}\big).$$

Theorem 32 is established using a multi-dimensional version of the Berry-Esseen theorem, see [GKPX22, Section 5] for details. Next, [GKPX22, Theorem 2.4] shows that

$$\mathbb{E}_{X_i \sim \mathcal{N}(0, I_n), 1 \leq i \leq M \text{ i.i.d.}}\big[\big|\Xi_\kappa(\beta, \eta, m, \alpha)\big|\big] = \exp\big(-\Theta(n)\big),$$

for $\alpha \geq \alpha^*_{\text{OGP}}(\kappa)$. This, together with Theorem 32, immediately yield that Theorem 31 still remains valid if $X_i \sim \mathcal{D}^{\otimes n}$, $1 \leq i \leq M$ are i.i.d., and $\mathcal{D}$ satisfies the assumptions in Theorem 15. With this, a reasoning identical to that in the proof of Theorem 14 yields Theorem 15.

3.11. **Numerical Experiments.** In this section, we report numerical experiments that support Assumption 3.1.

3.11.1. *The Function* $\zeta(\kappa, p)$. See Figure 3.1 for a plot of $\zeta(\kappa, p)$ appearing in (3.14).

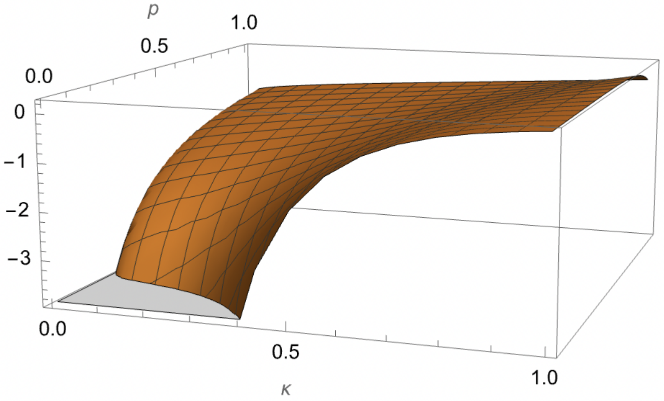

Figure 3.1. Plot of $\zeta(\kappa, p)$, truncated as $\zeta(0, 0) \to -\infty$.

Furthermore, Figure 3.2 shows the region of (κ, p) pairs for which $\zeta(\kappa, p) < 0$. Recall that for any given $\kappa > 0$, we establish Theorem 12 for a range of p values, i.e. $p \in [p^*_\kappa, 1]$ for a suitable p^*_κ. For any fixed $\kappa > 0$, the corresponding p^*_κ can be read off directly from Figure 3.2.

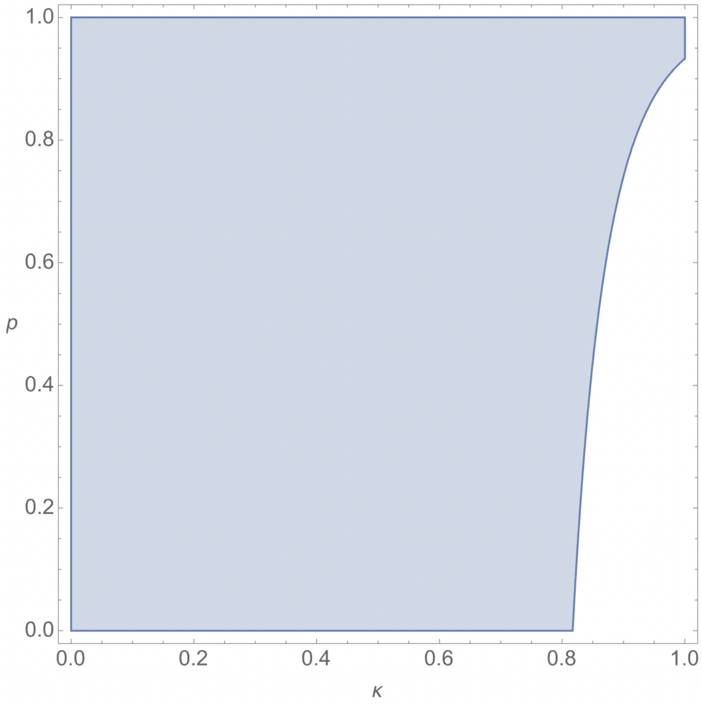

FIGURE 3.2. Region of (κ, p) pairs with $\zeta(\kappa, p) < 0$.

3.11.2. *The Function $F_{\kappa,\alpha,p}(\beta)$.* We now plot $F_{\kappa,\alpha,p}(\beta)$ in Assumption 3.1, where the axes correspond to p and β. We plotted $F_{\kappa,\alpha,p}$ across p for a broad range of (κ, α) pairs, see Figure 3.3 for $(\kappa, \alpha) = (0.6, 1)$. This demonstrates typical behavior: Assumption 3.1 is satisfied for all values of p. (3.10)

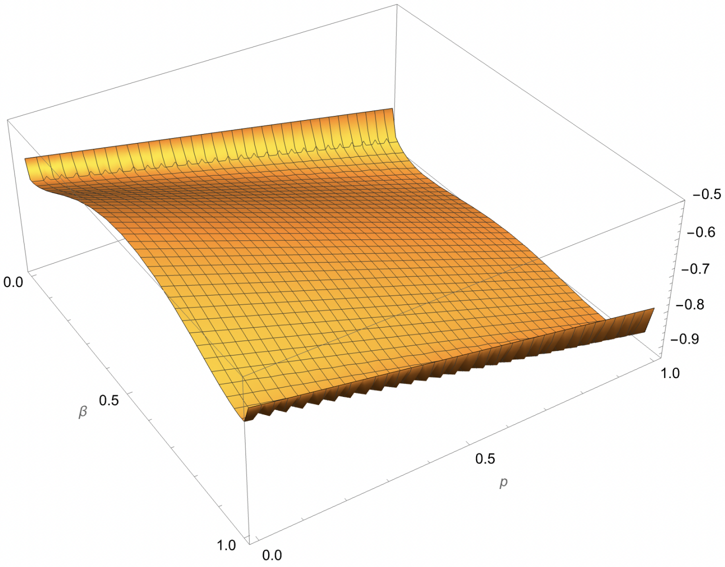

FIGURE 3.3. $F_{\kappa,\alpha,p}(\beta)$ for $\kappa = 0.6, \alpha = 1$.

See Figure 3.4 for a plot of $F_{\kappa,\alpha,p}(\beta)$ for $\kappa = 1.8, \alpha = 0.5$, where the axes correspond to p and β. This demonstrates a phase transition, where Assumption 3.1 is only satisfied for $p \in [p^*_\kappa, 1]$ for a suitable p^*_κ. At $p = 0$, corresponding to the `UBP`, $F_{\kappa,\alpha,p}(1/2)$ is not a local maximum. However, at $p = 1$, corresponding to the `SBP`, $F_{\kappa,\alpha,p}(1/2)$ is a local maximum.

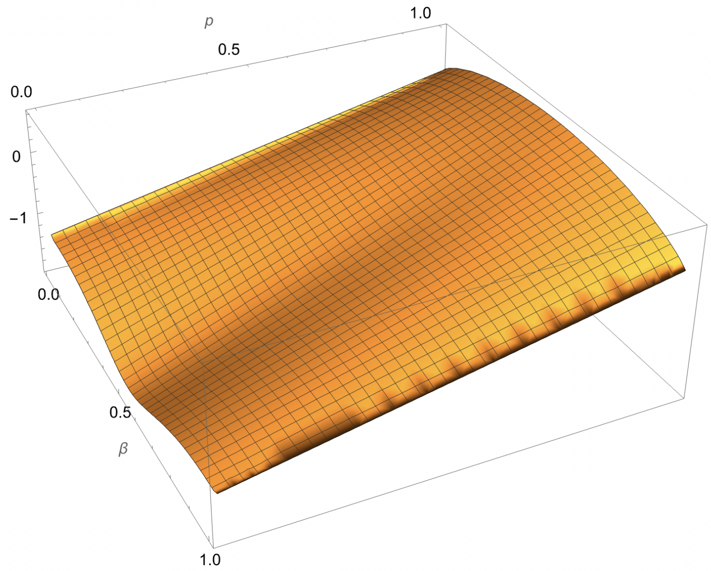

Figure 3.4. $F_{\kappa,\alpha,p}(\beta)$ for $\kappa = 1.8, \alpha = 0.5$.

References

[ACO08] Dimitris Achlioptas and Amin Coja-Oghlan, *Algorithmic barriers from phase transitions*, 2008 49th Annual IEEE Symposium on Foundations of Computer Science, IEEE, 2008, pp. 793–802.

[AHS22] Ryan Alweiss, Brice Huang, and Mark Sellke, *Improved lower bound for frankl's union-closed sets conjecture*, arXiv preprint arXiv:2211.11731 (2022).

[ALS21a] 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 (2021).

[ALS21b] ______, *Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron*, arXiv preprint arXiv:2102.13069 (2021).

[Alt22] Dylan J Altschuler, *Fluctuations of the symmetric perceptron*, arXiv preprint arXiv:2205.02319 (2022).

[AM02] Dimitris Achlioptas and Cristopher Moore, *The asymptotic order of the random k-sat threshold*, The 43rd Annual IEEE Symposium on Foundations of Computer Science, 2002. Proceedings., IEEE, 2002, pp. 779–788.

[APZ19] Benjamin Aubin, Will Perkins, and Lenka Zdeborová, *Storage capacity in symmetric binary perceptrons*, Journal of Physics A: Mathematical and Theoretical **52** (2019), no. 29, 294003.

[ART06] Dimitris Achlioptas and Federico Ricci-Tersenghi, *On the solution-space geometry of random constraint satisfaction problems*, Proceedings of the thirty-eighth annual ACM symposium on Theory of computing, 2006, pp. 130–139.

[AS20] Ahmed El Alaoui and Mark Sellke, *Algorithmic pure states for the negative spherical perceptron*, arXiv preprint arXiv:2010.15811 (2020).

[BDVLZ20] 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** (2020), no. 7, 073303.

[Ber41] Andrew C Berry, *The accuracy of the gaussian approximation to the sum of independent variates*, Transactions of the american mathematical society **49** (1941), no. 1, 122–136.

[BH22] Guy Bresler and Brice Huang, *The algorithmic phase transition of random k-sat for low degree polynomials*, 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2022, pp. 298–309.

[BNSX22] Erwin Bolthausen, Shuta Nakajima, Nike Sun, and Changji Xu, *Gardner formula for ising perceptron models at small densities*, Conference on Learning Theory, PMLR, 2022, pp. 1787–1911.

[Bop85] Ravi B. Boppana, *Amplification of probabilistic boolean formulas*, 26th Annual Symposium on Foundations of Computer Science, 1985, pp. 449–458.

[Bop23] ———, *A useful inequality for the binary entropy function*, arXiv preprint arXiv:2301.09664 (2023).

[BS20] Nikhil Bansal and Joel H. Spencer, *On-line balancing of random inputs*, Random Structures and Algorithms **57** (2020), no. 4, 879–891 (English (US)).

[Cam22] Stijn Cambie, *Better bounds for the union-closed sets conjecture using the entropy approach*, arXiv preprint arXiv:2212.12500 (2022).

[Cam23] ———, *Progress on the union-closed conjecture and offsprings in winter 2022-2023*, arXiv preprint arXiv:2306.12351 (2023).

[Cha02] Charalambos A. Charalambides, *Enumerative combinatorics*, Chapman & Hall/CRC, 2002.

[Cig22] Johann Cigler, *Recurrences for certain sequences of binomial sums in terms of (generalized) fibonacci and lucas polynomials*, arXiv preprint arXiv:2212.02118 (2022).

[CL22] Zachary Chase and Shachar Lovett, *Approximate union closed conjecture*, arXiv preprint arXiv:2211.11689 (2022).

[Cov65] Thomas M Cover, *Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition*, IEEE transactions on electronic computers (1965), no. 3, 326–334.

[COV13] Amin Coja-Oghlan and Dan Vilenchik, *Chasing the k-colorability threshold*, 2013 IEEE 54th Annual Symposium on Foundations of Computer Science, 2013, pp. 380–389.

[DGH23] Hang Du, Shuyang Gong, and Rundong Huang, *The algorithmic phase transition of random graph alignment problem*, arXiv preprint arXiv:2307.06590 (2023).

[DLMF] *NIST Digital Library of Mathematical Functions*, F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds.

[dM31] A de Moivre, *Miscellanca analytica de scrichus et quadraturis*, Tomson and J. Watts, London (1731).

[DS19] Jian Ding and Nike Sun, *Capacity lower bound for the Ising perceptron*, Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, 2019, pp. 816–827.

[Ess42] Carl-Gustav Esseen, *On the liapunov limit error in the theory of probability*, Ark. Mat. Astr. Fys. **28** (1942), 1–19.

[Gam21] David Gamarnik, *The overlap gap property: A topological barrier to optimizing over random structures*, Proceedings of the National Academy of Sciences **118** (2021), no. 41.

[Gar88] Elizabeth Gardner, *The space of interactions in neural network models*, Journal of physics A: Mathematical and general **21** (1988), no. 1, 257.

[GD88] Elizabeth Gardner and Bernard Derrida, *Optimal storage properties of neural network models*, Journal of Physics A: Mathematical and general **21** (1988), no. 1, 271.

[Ges16] Ira M. Gessel, *Lagrange inversion*, Journal of Combinatorial Theory, Series A **144** (2016), 212–249.

[GJ21] David Gamarnik and Aukosh Jagannath, *The overlap gap property and approximate message passing algorithms for p-spin models*, The Annals of Probability **49** (2021), no. 1, 180–205.

[GJK23] David Gamarnik, Aukosh Jagannath, and Eren C Kızıldağ, *Shattering in the ising pure p-spin model*, arXiv preprint arXiv:2307.07461 (2023).

[GJW20] David Gamarnik, Aukosh Jagannath, and Alexander S Wein, *Low-degree hardness of random optimization problems*, 2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2020, pp. 131–140.

[GJW21] ———, *Circuit lower bounds for the p-spin optimization problem*, arXiv preprint arXiv:2109.01342 (2021).

[GK21] David Gamarnik and Eren C Kızıldağ, *Algorithmic obstructions in the random number partitioning problem*, arXiv preprint arXiv:2103.01369 (2021).

[GKL$^+$22] Federica Gerace, Florent Krzakala, Bruno Loureiro, Ludovic Stephan, and Lenka Zdeborová, *Gaussian universality of linear classifiers with random labels in high-dimension*, arXiv preprint arXiv:2205.13303 (2022).

[GKP94] Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete mathematics*, second ed., Addison-Wesley Publishing Company, Reading, MA, 1994.

[GKPX22] David Gamarnik, Eren C Kızıldağ, Will Perkins, and Changji Xu, *Algorithms and barriers in the symmetric binary perceptron model*, 2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2022, pp. 576–587.

[GKPX23] David Gamarnik, Eren C Kizildağ, Will Perkins, and Changji Xu, *Geometric barriers for stable and online algorithms for discrepancy minimization*, The Thirty Sixth Annual Conference on Learning Theory, PMLR, 2023, pp. 3231–3263.

[GMZ22] David Gamarnik, Cristopher Moore, and Lenka Zdeborová, *Disordered systems insights on computational hardness*, Journal of Statistical Mechanics: Theory and Experiment **2022** (2022), no. 11, 114015.

[Gou61] Henry W. Gould, *A series transformation for finding convolution identities*, Duke Math. J. **28** (1961), 193–202.

[GS14] David Gamarnik and Madhu Sudan, *Limits of local algorithms over sparse random graphs*, Proceedings of the 5th conference on Innovations in theoretical computer science, 2014, pp. 369–376.

[GS17a] ———, *Limits of local algorithms over sparse random graphs*, Ann. Probab. **45** (2017), no. 4, 2353–2376.

[GS17b] ———, *Performance of sequential local algorithms for the random NAE-K-SAT problem*, SIAM Journal on Computing **46** (2017), no. 2, 590–619.

[HS88] Leetsch C. Hsu and Peter Jau-Shyong Shiue, *A unified approach to generalized Stirling numbers*, Advances in Applied Mathematics (1988).

[HS21] Brice Huang and Mark Sellke, *Tight lipschitz hardness for optimizing mean field spin glasses*, arXiv preprint arXiv:2110.07847 (2021).

[HS23] ———, *Algorithmic threshold for multi-species spherical spin glasses*, arXiv preprint arXiv:2303.12172 (2023).

[Hua22] Brice Huang, *Computational hardness in random optimization problems from the overlap gap property*, Ph.D. thesis, Massachusetts Institute of Technology, 2022.

[Kız22] Eren C. Kızıldağ, *Algorithms and algorithmic barriers in high-dimensional statistics and random combinatorial structures*, Ph.D. thesis, Massachusetts Institute of Technology, 2022.

[Kız23a] Eren C Kızıldağ, *Planted random number partitioning problem*, arXiv preprint arXiv:2309.15115 (2023).

[Kız23b] ———, *Sharp phase transition for multi overlap gap property in ising p-spin glass and random k-sat models*, arXiv preprint arXiv:2309.09913 (2023).

[KM89] Werner Krauth and Marc Mézard, *Storage capacity of memory networks with binary couplings*, Journal de Physique **50** (1989), no. 20, 3057–3066.

[KR98] 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** (1998), no. 2, 223–252.

[KWB22] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira, *Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio*, Mathematical Analysis, its Applications and Computation: ISAAC 2019, Aveiro, Portugal, July 29–August 2, Springer, 2022, pp. 1–50.

[MM09] Marc Mezard and Andrea Montanari, *Information, physics, and computation*, Oxford University Press, 2009.

[MMZ06] Stephan Mertens, Marc Mézard, and Riccardo Zecchina, *Threshold values of random k-sat from the cavity method*, Random Structures & Algorithms **28** (2006), no. 3, 340–373.

[MS16] Toufik Mansour and Matthias Schork, *Commutation relations, normal ordering, and Stirling numbers*, CRC Press, 2016.

[MZZ21] Andrea Montanari, Yiqiao Zhong, and Kangjie Zhou, *Tractability from overparametrization: The example of the negative perceptron*, arXiv preprint arXiv:2110.15824 (2021).

[NS23] Shuta Nakajima and Nike Sun, *Sharp threshold sequence and universality for ising perceptron models*, Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), SIAM, 2023, pp. 638–674.

[PX21] Will Perkins and Changji Xu, *Frozen 1-RSB structure of the symmetric Ising perceptron*, Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, 2021, pp. 1579–1588.

[RV17] Mustazee Rahman and Balint Virag, *Local algorithms for independent sets are half-optimal*, The Annals of Probability **45** (2017), no. 3, 1543–1577.

[Saw22] Will Sawin, *An improved lower bound for the union-closed set conjecture*, arXiv preprint arXiv:2211.11504 (2022).

[SS23] Ashwin Sah and Mehtaab Sawhney, *Distribution of the threshold for the symmetric perceptron*, arXiv preprint arXiv:2301.10701 (2023).

[ST03] Mariya Shcherbina and Brunello Tirozzi, *Rigorous solution of the Gardner problem*, Communications in mathematical physics **234** (2003), no. 3, 383–422.

[Sto13] Mihailo Stojnic, *Another look at the Gardner problem*, arXiv preprint arXiv:1306.3979 (2013).

[Tal99] Michel Talagrand, *Intersecting random half cubes*, Random Structures & Algorithms **15** (1999), no. 3-4, 436–449.

[Tal00] ———, *Intersecting random half-spaces: toward the gardner-derrida formula*, The Annals of Probability **28** (2000), no. 2, 725–758.

[Tal10] ———, *Mean field models for spin glasses: Volume i: Basic examples*, vol. 54, Springer Science & Business Media, 2010.

[Tal11] ———, *Mean field models for spin glasses: Advanced replica-symmetry and low temperature*, Springer, 2011.

[Ver10] Roman Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, arXiv preprint arXiv:1011.3027 (2010).

[Ver18] ———, *High-dimensional probability: An introduction with applications in data science*, vol. 47, Cambridge university press, 2018.

[Wei20] Alexander S Wein, *Optimal low-degree hardness of maximum independent set*, arXiv preprint arXiv:2010.06563 (2020).

[Wen62] James G Wendel, *A problem in geometric probability*, Mathematica Scandinavica **11** (1962), no. 1, 109–111.

[Win61] Robert O Winder, *Single stage threshold logic*, 2nd Annual Symposium on Switching Circuit Theory and Logical Design (SWCT 1961), IEEE, 1961, pp. 321–332.

[Xu21] Changji Xu, *Sharp threshold for the ising perceptron model*, The Annals of Probability **49** (2021), no. 5, 2399–2415.

[Yu22] Lei Yu, *Dimension-free bounds for the union-closed sets conjecture*, arXiv preprint arXiv:2212.00658 (2022).

[Yus23] Raphael Yuster, *Almost k-union closed set systems*, arXiv preprint arXiv:2302.12276 (2023).