Symmetric Binary Perceptron Model: Algorithms and Barriers

Eren C. Kızıldağ (MIT)

Joint work with David Gamarnik (MIT), Will Perkins (UIC), and Changji Xu (Harvard)

arXiv:2203.15667

March 30, 2022

Overview

Symmetric Binary Perceptron (SBP): Model.

- Fix $\kappa, \alpha > 0$, and set $M = \lfloor n\alpha \rfloor$. Let $X_i \stackrel{d}{=} \mathcal{N}(0, I_n), 1 \le i \le M$ be i.i.d.
- Consider (random) set

$$S_\alpha(\kappa) \triangleq \bigcap_{1 \le i \le M} \left\{\sigma \in \mathcal{B}_n : \left|\langle \sigma, X_i \rangle\right| \le \kappa\sqrt{n}\right\},$$

where

$$\mathcal{B}_n = \{-1, 1\}^n \qquad \text{and} \qquad \langle \sigma, X_i \rangle = \sum_{1 \le j \le n} \sigma_j X_i(j).$$

- Equivalently, for **disorder** $\mathcal{M} \in \mathbb{R}^{M \times n}$ with rows $X_1, X_2, \dots, X_M \in \mathbb{R}^n$,

$$S_\alpha(\kappa) = \left\{\sigma \in \mathcal{B}_n : \left\|\mathcal{M}\sigma\right\|_\infty \le \kappa\sqrt{n}\right\}.$$

Perceptron Model: Motivation

Toy one-layer **neural network** [Wen62, Cov65].

- **Patterns** $X_i \in \mathbb{R}^n, 1 \le i \le M$ to be **stored**.
- **Storage:** Find a $\sigma \in \mathcal{B}_n$ consistent with all X_i: $\langle \sigma, X_i \rangle \ge 0, 1 \le i \le M$.
- **Storage Capacity:** Maximum number of stored patterns M^*.
  - M^*/n, as $n \to \infty$.
  - Detailed picture by **statistical physicists** [Gar87, Gar88, GD88].

Perceptron Model: Motivation

Connection to *Constraint Satisfaction Problems* (CSPs):

- Each **constraint** $X_i \in \mathbb{R}^n$ rules out certain $\sigma \in \mathcal{B}_n$.
- $\alpha \triangleq M/n$ is **constraint density**.

Symmetric model, SBP [APZ19]: $\sigma \in S_\alpha(\kappa) \iff -\sigma \in S_\alpha(\kappa)$.

- Similar structural properties as **asymmetric** version [BDVLZ20].
- Easier math: analogy with $k-\mathrm{SAT}$ vs. $\mathrm{NAE}-k-\mathrm{SAT}$.

SBP: Also related to **combinatorial discrepancy theory** [CV14, BS20].

Our Work: Two Mysteries in SBP and OGP

Statistical-to-computational gaps:

Gap between existential guarantees and (polynomial-time) algorithmic guarantees.

- SBP has a *statistical-to-computational gap*.
- *Origins of this gap?* **Landscape** of SBP via *statistical physics* lens.

Extreme Clustering:

Typical solutions of SBP are $\Theta(n)$ apart isolated singletons.

- Suggests algorithmic hardness.
- **A Conundrum.** Coincides with polynomial-time algorithms.

This work: **Overlap Gap Property** (OGP). Intricate geometrical property.

Leverage OGP to **rule out** important classes of algorithms.

OGP $\Longrightarrow$ Hardness. **Clustering $\not\Longrightarrow$ Hardness.**

SBP: Main Questions

Recall

$$S_\alpha(\kappa) = \bigcap_{1\le i\le M} \left\{\sigma \in \mathcal{B}_n : \left|\langle \sigma, X_i\rangle\right| \le \kappa\sqrt{n}\right\} = \left\{\sigma \in \mathcal{B}_n : \left\|\mathcal{M}\sigma\right\|_\infty \le \kappa\sqrt{n}\right\}.$$

Proportional Regime: $M, n \to \infty$ while $M/n \to \alpha$. Fix $\kappa > 0$ and vary α.

- **Structural Question.** $S_\alpha(\kappa)$ empty/non-empty (w.h.p.) & its geometry.
- **Algorithmic Question:** Efficient (polynomial-time) algorithms for finding $\sigma \in S_\alpha(\kappa)$.

SBP: Available Structural Results

Sharp Phase Transition. Set $p_\alpha(\kappa) \triangleq \mathbb{P}\left[S_\alpha(\kappa) \neq \varnothing\right]$.

$$\lim_{n\to\infty} p_\alpha(\kappa) = \begin{cases} 1, & \text{if } \alpha < \alpha_c(\kappa) \\ 0, & \text{if } \alpha > \alpha_c(\kappa) \end{cases}, \quad \text{where} \quad \alpha_c(\kappa) = -\frac{1}{\log_2 \mathbb{P}[|\mathcal{N}(0,1)| \le \kappa]}.$$

$\alpha_c(\kappa)$ matches **first moment prediction**: $\mathbb{E}\left[\left|S_\alpha(\kappa)\right|\right] = o(1)$ iff $\alpha > \alpha_c(\kappa)$.

For $\alpha < \alpha_c(\kappa)$:

- [APZ19]: $\liminf_n p_\alpha(\kappa) > 0$ by 2nd **Moment Method**.
- [PX21, ALS21b]: $\lim_n p_\alpha(\kappa) = 1 - o(1)$. More delicate tools.

Asymmetric Model: Available Structural Results

Fix $\alpha > 0$, let $M = \lfloor n\alpha \rfloor$ and $X_i \stackrel{d}{=} \mathcal{N}(0, I_n), 1 \le i \le M$ i.i.d. Set

$$S'_\alpha(\kappa) = \bigcap_{1\le i\le M} \left\{\sigma \in \mathcal{B}_n : \langle \sigma, X_i \rangle \ge \kappa\sqrt{n}\right\} \quad \text{and} \quad p'_\alpha(\kappa) = \mathbb{P}\left[S'_\alpha(\kappa) \ne \varnothing\right].$$

Much less is known (rigorously)!

Conjecture

$\lim_{n\to\infty} p'_\alpha(\kappa)$ *undergoes a sharp phase transition at* $\alpha_{\mathrm{KM}}(\kappa)$.

- $\alpha_{\mathrm{KM}}(0) \approx 0.833$ [KM89]. Significantly different from **moment prediction** ($\alpha = 1$).
- Even the very existence of PT is **open**!
- Lower bound [DS19]: $\liminf_n p'_\alpha(0) > 0$ for $\alpha < \alpha_{\mathrm{KM}}(0)$.
- [KR98]: $p'_\alpha(0) = o(1)$ for $\alpha > 0.9963$; and $p'_\alpha(0) = 1 - o(1)$ for $\alpha < 0.005$ (algorithmic).

SBP: Available Algorithmic Guarantees

Connection to **combinatorial discrepancy theory**:

$$\min_{\sigma \in \mathcal{B}_n} \left\| \mathcal{M}\sigma \right\|_\infty \quad \text{where} \quad \mathcal{M} \in \mathbb{R}^{M \times n}.$$

Many (efficient) algorithms: [Rot17, LRR17, ES18, BS20]...

Theorem (Bansal and Spencer, 2020)

Fix $\kappa > 0$, $\alpha \le K\kappa^2$. Let $\mathcal{M} \in \mathbb{R}^{\lfloor \alpha n \rfloor \times n}$ has i.i.d. Rademacher entries. Then, there exists a polynomial-time algorithm $\mathcal{A}$ such that w.h.p.

$$\left\| \mathcal{M}\mathcal{A}(\mathcal{M}) \right\|_\infty \le \kappa\sqrt{n}.$$

- $K > 0$ absolute constant.
- **Computational Threshold**: $\Theta\left(\kappa^2\right)$. Our benchmark.

SBP: A *Statistical-to-Computational Gap*

Gap between existential guarantees and what polynomial-time algorithms can promise.

Much more **dramatic** for $\kappa \to 0$:

> As $\kappa \to 0$,
> $$\alpha_c(\kappa) = -\frac{1}{\log_2 \mathbb{P}\left[\left|\mathcal{N}(0,1)\right| \le \kappa\right]} \approx -\frac{1}{\log_2 \kappa}.$$

- $S_\alpha(\kappa) \ne \varnothing$ for $\alpha < -\frac{1}{\log_2 \kappa}$. Algorithms exist for $\alpha < K\kappa^2$.
- Ignoring $K > 0$, a striking gap: $-\frac{1}{\log_2 \kappa}$ vs κ^2.

Source of this gap/hardness?

Given $\sigma \in S_\alpha(\kappa)$, if $\sigma^{(i)} \in S_\alpha(\kappa)$ then $1 \le i \le n$ is **free**. Otherwise **frozen**.

Theorem (Perkins-Xu'2021, Abbe-Li-Sly'2021)

Extreme clustering and freezing in SBP: *For any* $0 < \alpha < \alpha_c(\kappa)$*, typical solutions of SBP are isolated (w.h.p.) and the distance to any other solution is* $\Theta(n)$*.*

- Suggests that finding $\sigma \in S_\alpha(\kappa)$ is computationally hard.
- At odds with algorithms [BS20, BZ06, BBBZ07, Bal09, BB15].

Existence of polynomial-time algorithms coincides with clustering.

Existence of polynomial-time algorithms coincides with clustering.

One Explanation.

Rare clusters ($o(1)$ fraction) with **positive entropy density** ($e^{\Theta(n)}$ size) [ALS21a].

However:

- Large clusters exists at all $0 < \alpha < \alpha_c(\kappa)$.
- **No** algorithmic **improvement** over κ^2.

Statistical-to-Computational Gaps

Common feature in many algorithmic problems in high-dimensional statistics & random combinatorial structures:

Random k-SAT, optimization over random graphs, p-spin model, number partitioning, planted clique, matrix PCA, linear regression, spiked tensor, largest submatrix problem...

No analogue of worst-case theory (such as $P \neq NP$).

Statistical-to-Computational Gaps

Various forms of *rigorous evidences*:

- **Low-degree methods:** [Hop18, KWB19, Wei20]...
- **Reductions from the planted clique:** [BR13, BBH18, BB19]...
- Many more: **Failure of MCMC, Failure of BP/AMP, Methods from Statistical Physics, SoS Lower Bounds,...** [Jer92, HSS15, LKZ15, ZK16, HKP+17, DKS17, BHK+19]...

Another approach (spin glass theory): **Overlap Gap Property**.

The Overlap Gap Property (OGP)

Generic optimization problem with random ξ: $\min_{\theta\in\Theta}\mathcal{L}(\sigma,\xi)$. For SBP,

$$\mathcal{L}(\sigma,\mathcal{M}) \triangleq \sum_{1\le i\le M} \mathbb{1}\left\{\left|\langle \sigma, X_i\rangle\right| > \kappa\sqrt{n}\right\}. \quad (\text{\# of violated constraints.})$$

(Informally) OGP for **energy** $\mathcal{E}$ if $\exists 0<\nu_1<\nu_2$ s.t. $\forall \sigma_1,\sigma_2\in\Theta$,

$$\mathcal{L}(\sigma_j,\xi)\le \mathcal{E} \implies \mathrm{distance}(\sigma_1,\sigma_2)<\nu_1 \quad \text{or} \quad \mathrm{distance}(\sigma_1,\sigma_2)>\nu_2.$$

Any two **near optimal** σ_1,σ_2 are either *too similar* or *too dissimilar*.

$\mathrm{distance}(\cdot,\cdot)$

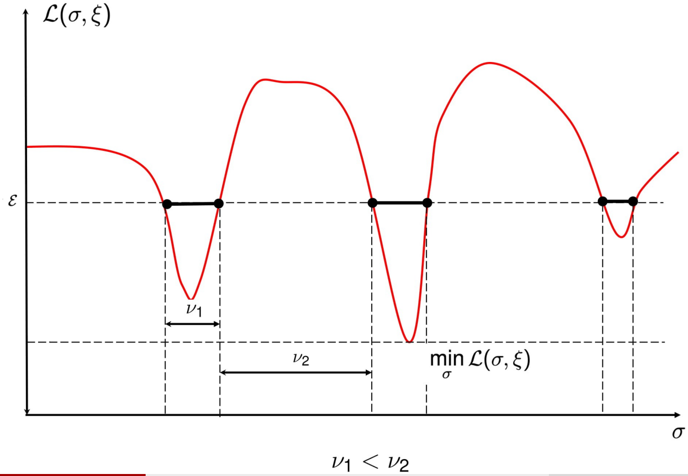
For $\Theta=\mathcal{B}_n=\{-1,1\}^n$, normalized overlap:

$$\mathcal{O}(\sigma,\sigma')=n^{-1}\langle\sigma,\sigma'\rangle\in[0,1].$$

$$\textbf{Large } \mathcal{O} \iff \textbf{Small } d_H \iff \textbf{Similar } \sigma\approx\sigma'.$$

Overlap Gap Property - A Pictorial Illustration

OGP for $\mathcal{E}$.



OGP: Prior Work

- **Clustering in $k-$SAT**: Solution space consists of disconnected clusters [MMZ05, ACO08, ACORT11].
- **First algorithmic implication**: Max independent set in random $d-$regular graph $\mathbb{G}_d(n)$. [GS17a].
- OGP: Any large $\mathcal{I}_1, \mathcal{I}_2$ either have **significant** intersection, or **no** intersection at all.
- Local algorithms fail to return a large $\mathcal{I}$.

OGP in Other Problems & OGP as a Provable Algorithmic Barrier

Many other problems with OGP:

random k-SAT, NAE-k-SAT, p-spin model, number partitioning, sparse PCA, largest submatrix problem, max-CUT, planted clique...

OGP as a *provable barrier* to algorithms:

WALKSAT, local algorithms, stable algorithms, low-degree polynomials, AMP, MCMC...

[COHH17, GS17b, GJW20, Wei20, Gam21, GJS19, GK21, BH21]...

Overview

High κ Regime

Recall the setup. $M = \lfloor n\alpha \rfloor$ and $X_i \stackrel{d}{=} \mathcal{N}(0, I_n), 1 \le i \le M$ i.i.d.

$$S_\alpha(\kappa) = \bigcap_{1\le i\le M} \left\{\sigma \in \mathcal{B}_n : \left|\langle \sigma, X_i\rangle\right| \le \kappa\sqrt{n}\right\} = \left\{\sigma \in \mathcal{B}_n : \left\|\mathcal{M}\sigma\right\|_\infty \le \kappa\sqrt{n}\right\}.$$

High κ Regime. $\kappa = 1$ as running example.

$$\alpha_c(\kappa) = -\frac{1}{\log_2 \mathbb{P}\left[|\mathcal{N}(0,1)| \le 1\right]} \approx 1.8158.$$

Namely, $S_\alpha(1) \ne \varnothing$ (w.h.p.) for $\alpha < 1.8158$.

Question.

Is it possible to efficiently find a $\sigma \in S_\alpha(1)$ when $\alpha < 1.8158$?

Our Contributions: 2−OGP for SBP when $\kappa = 1$.

Let $\mathcal{M} \in \mathbb{R}^{M \times n}$ with rows $X_1, \ldots, X_M \in \mathbb{R}^n$, $M = \lfloor \alpha n \rfloor$.

Theorem (Gamarnik, K., Perkins, and Xu, 2022+)

(Informally) 2-OGP (for $\kappa = 1$) holds if $\alpha \geq 1.71$.
Formally, $\forall \alpha \geq 1.71$, $\exists 1 > \beta > \eta > 0$ such that w.h.p. for any distinct $\sigma_1, \sigma_2 \in \mathcal{B}_n$ with $\|\mathcal{M}\sigma_1\|_\infty \leq \sqrt{n}$ and $\|\mathcal{M}\sigma_2\|_\infty \leq \sqrt{n}$; $n^{-1}\langle \sigma_1, \sigma_2 \rangle \notin [\beta - \eta, \beta]$.

- Suggests that finding a $\sigma \in S_\alpha(1)$ is computationally **intractable** if $\alpha \geq 1.71$.
- Proof based on first moment method and Gaussian comparison inequality [Sid68].

Proof Sketch. Let N count $\#$ such (σ_1, σ_2). Show $\mathbb{E}[N] = o(1)$ for appropriate $1 > \beta > \eta > 0$ and apply Markov's inequality $\mathbb{P}[N \geq 1] \leq \mathbb{E}[N]$.

Ensemble-Multi-OGP for SBP.

- Consider i.i.d. matrices $\mathcal{M}_i \in \mathbb{R}^{M\times n}, 0 \le i \le m$.
- Each $\mathcal{M}_i$ have i.i.d. $\mathcal{N}(0,1)$ entries.
- Consider interpolation

$$\mathcal{M}_i(\tau) = \cos(\tau)\mathcal{M}_0 + \sin(\tau)\mathcal{M}_i \in \mathbb{R}^{M\times n}, \quad \tau \in \left[0, \frac{\pi}{2}\right], \quad 1 \le i \le m.$$

- For each $0 \le \tau \le \frac{\pi}{2}$ and $1 \le i \le m$, $\mathcal{M}_i(\tau)$ has i.i.d. $\mathcal{N}(0,1)$ entries.

Ensemble-Multi-OGP for SBP

m-tuples $\sigma_i \in \mathcal{B}_n$ ($m-$OGP); each satisfy constraints $\mathcal{M}_i(\tau_i)$, $\exists \tau_i \in [0, \pi/2]$ (ensemble).

- m-**OGP**: Reduces thresholds further: Max independent set in $\mathbb{G}_d(n)$.
  - **Computational threshold** $(\log d/d)n$, 2-OGP rules out $|\mathcal{I}| \ge (1 + 1/\sqrt{2})(\log d/d)n$.
  - [RV17]: Study instead m-tuples $\mathcal{I}_i$, $1 \le i \le m$: hit $(\log d/d)n$.
  - Similar story for NAE-k-SAT [GS17b].
- **Ensemble OGP:** Can rule out any sufficiently stable algorithm [GJW20, Wei20, Gam21, GK21, BH21, HS21].

Our Contributions: Ensemble $3-$OGP for SBP when $\kappa = 1$.

Theorem (Gamarnik, K., Perkins, and Xu, 2022+)

(Informally) Ensemble $3-$OGP (for $\kappa = 1$) holds if $\alpha \ge 1.667$
Formally, $\forall \alpha \ge 1.667$, $\forall \mathcal{I} \subset [0, \pi/2]$ with $|\mathcal{I}| = 2^{O(n)}$, $\exists 1 > \beta > \eta > 0$ s.t. w.h.p. if

$$\left\|\mathcal{M}_i(\tau_i)\sigma_i\right\|_\infty \le \sqrt{n}, \quad \tau_i \in \mathcal{I}, \quad 1 \le i \le 3$$

then $\exists 1 \le i < j \le 3$ such that

$$n^{-1}\langle \sigma_i, \sigma_j \rangle \notin (\beta - \eta, \beta).$$

- $\beta \gg \eta$: rules out **equilateral triangles** in Hamming space.
- Proof based again on first moment method and Gaussian comparison inequality [Sid68].

Low κ regime: Ensemble $m-$OGP beyond $m = 3$

More **intricate** structure $\implies$ Hardness for broader α.

$\kappa = 1$

- $\alpha_c(1) \approx 1.8158$. $2-$OGP for $\alpha \geq 1.71$ & $3-$OGP for $\alpha \geq 1.667$.
- 3-OGP requires **exact counting** (up to lower order terms).
- Counting term **intractable** for $m \geq 4$.

How about small κ regime?

$\kappa \to 0$

- For $\sigma_i \in \mathcal{B}_n$ and correlated $X_i \stackrel{d}{=} \mathcal{N}(0, I_n)$, $1 \leq i \leq m$; $\mathbb{P}\big[|\langle \sigma_i, X_i \rangle| \leq \kappa, 1 \leq i \leq m\big]$ controlled by volume of $[-\kappa, \kappa]^m$.
- **Main Idea**: If κ small, $(2\kappa)^m$ shrinks further with m large.

Our Contributions: Ensemble-m-OGP for SBP when $\kappa \to 0$.

Theorem (Gamarnik, K., Perkins, and Xu, 2022+)

(Informally) Ensemble $m-$OGP (as $\kappa \to 0$, for appropriate $m \in \mathbb{N}$) holds if $\alpha = \Omega\left(\kappa^2 \log_2 \frac{1}{\kappa}\right)$. Formally, $\forall \kappa > 0$ small enough, $\forall \alpha \ge 10\kappa^2 \log_2 \frac{1}{\kappa}$, $\forall \mathcal{I} \subset [0, \pi/2]$ with $|\mathcal{I}| = 2^{O(n)}$, $\exists m \in \mathbb{N}$, $\exists 1 > \beta > \eta > 0$ s.t. w.h.p. if

$$\left\|\mathcal{M}_i(\tau_i)\sigma_i\right\|_\infty \le \kappa\sqrt{n}, \quad \tau_i \in \mathcal{I}, \quad 1 \le i \le m$$

then $\exists 1 \le i < j \le m$ such that

$$n^{-1}\langle \sigma_i, \sigma_j \rangle \notin (\beta - \eta, \beta).$$

- Ensemble $m-$OGP **above** $\kappa^2 \log_2 \frac{1}{\kappa}$.
- Nearly **tight**: almost matches **algorithmic threshold** κ^2 (modulo $\log_2 \frac{1}{\kappa}$ factor)
- $\beta \gg \eta$: $m-$tuple of **equidistant** points. **Best rate** with this approach.

Overview

Problems with OGP and Algorithms Hardness Results

- Random walk type algorithms for random k-SAT [COHH17].
- Low-degree polynomials for random k-SAT [BH21].
- Sequential local algorithms for NAE-k-SAT [GS17b].
- Low-degree polynomials and Langevin dynamics [GJW20, Wei20].
- AMP for optimizing p-spin model Hamiltonian [GJ21a].
- Overlap concentrated algorithms [1] for mixed, even $p-$spin model Hamiltonian [HS21]
- Low-depth circuits for even $p-$spin model Hamiltonian [GJW21].
- OGP $\implies$ FEW $\implies$ Failure of MCMC: Principle submatrix problem [GJS19].

[1] Includes $O(1)$ iteration of GD, AMP; and Langevin Dynamics run for $O(1)$ time.

Stable Algorithms: Formal Definition

- Algorithm $\mathcal{A}:\mathbb{R}^{M\times n}\to\mathcal{B}_n$ ($M=\lfloor \alpha n\rfloor$ and $\alpha<\alpha_c(\kappa)$). $\mathcal{A}(\mathcal{M})=\sigma\in\mathcal{B}_n$.
- Potentially **randomized**.
- **Informal**: $\mathcal{A}$ is stable if small change in $\mathcal{M}$ yields small change in $\mathcal{A}(\mathcal{M})$.

Semi-formally, $\mathcal{A}$ satisfies

Definition

(a) **Success**:

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

(b) **Stability**: $\exists \rho\in(0,1]$, $\mathcal{M},\overline{\mathcal{M}}\in\mathbb{R}^{M\times n}$ having $\mathcal{N}(0,1)$ entries with $\mathrm{Cov}\left(\mathcal{M}_{ij},\overline{\mathcal{M}}_{ij}\right)=\rho$;

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

Stable Algorithms: Which Algorithms are Stable?

Stable algorithms include

- Approximate message passing type algorithms [GJ21b, Gam21].
- Low-degree polynomial based algorithms [GJW20].

Kim-Roche Algorithm for (Asymmetric) Perceptron [KR98].

- $S'_\alpha(0) \ne \varnothing$ (w.h.p.) when $\alpha < 0.005$.
- Proof is **algorithmic**: multi-stage majority.
- Recently extended to SBP by Abbe, Li, and Sly [ALS21a].

Is multi-stage majority stable?

Our Contributions: Kim-Roche Algorithm is Stable.

- Let $\alpha < 0.005$, $M = \lfloor n\alpha \rfloor$. $\mathcal{M}, \mathcal{M}' \in \mathbb{R}^{M\times n}$ i.i.d. with i.i.d $\mathcal{N}(0,1)$ entries; and

$$\mathcal{M}(\tau) = \cos(\tau)\mathcal{M} + \sin(\tau)\mathcal{M}' \in \mathbb{R}^{M\times n}.$$

Theorem (Gamarnik, K., Perkins, and Xu, 2022+)

(Informal) Kim-Roche algorithm, $\mathcal{A}_{\rm KR}$ [KR98] is stable.
Set $\tau = n^{-0.02}$. Then,

$$\mathbb{P}\Big[d_H\big(\mathcal{A}_{\rm KR}(\mathcal{M}), \mathcal{A}_{\rm KR}(\mathcal{M}(\tau))\big) = o(n)\Big] \ge 1 - O\big(n^{-\frac{1}{41}}\big).$$

Namely, $\mathcal{A}_{\rm KR}$ is stable with

$$p_f = o\big(n^{-1}\big),\ p_{\rm st} = O\big(n^{-1/41}\big),\ \rho = \cos\big(n^{-0.02}\big),$$

and **any** $f = \Theta(n), L > 0$.

Our Contributions: OGP implies Failure of Stable Algorithms

Theorem (Gamarnik, K., Perkins, and Xu, 2022+)

For $\alpha = \Omega\left(\kappa^2 \log_2 \frac{1}{\kappa}\right)$*, there is* **no stable** $\mathcal{A}$ *for SBP (with appropriate* $f, \rho, p_f, p_{\rm st}$*).*

- Rule out $p_f, p_{\rm st} = O(1)$. **No** need for **high-probability** guarantee.
- **Proof Idea.** By contradiction. Suppose $\exists \mathcal{A}$.
  - m-OGP: a structure occurs with *vanishing probability*.
  - Run $\mathcal{A}$ on correlated instances. Show that w.p. > 0, *forbidden* structure occurs.
- Proof based on Ramsey Theory [GK21].

Our Contributions: Failure of Online Algorithms

Disorder $\mathcal{M} \in \mathbb{R}^{M \times n}$, columns $\mathcal{C}_i \in \mathbb{R}^M$, $1 \le i \le n$. $\mathcal{A}(\mathcal{M}) = (\sigma_i : 1 \le i \le n) \in \mathcal{B}_n$.

Definition

$\mathcal{A}$ is **online** if $\exists f_t$ such that $\sigma_t = f_t(\mathcal{C}_i : 1 \le i \le t)$ for $1 \le t \le n$.

Online Algorithms: Bansal-Spencer [BS20].

Theorem (Gamarnik, K., Perkins, and Xu, 2022+)

$\exists \epsilon > 0$ *such that for* $\alpha \ge \alpha_c(\kappa) - \epsilon$, *there is* **no online** $\mathcal{A}$ *for SBP.*

Proof Idea. By contradiction. A **forbidden structure** different than OGP.

Overview

Main Contributions

Two Conundrums in SBP

- A *Statistical-to-Computational Gap*: $-\frac{1}{\log_2 \kappa}$ vs κ^2.
- Clustering coincides with efficient algorithms.

Landscape of SBP

- High κ $(\kappa = O(1))$: Presence of 2−OGP and (Ensemble) 3−OGP, strictly below $\alpha_c(\kappa)$.
- Low κ $(\kappa \to 0)$: Presence of (Ensemble) m−OGP for $\alpha = \Omega(\kappa^2 \log_2 \frac{1}{\kappa})$.

Algorithmic Results

- Kim-Roche algorithm is stable.
- Stable algorithms fail to find a solution if $\alpha = \Omega(\kappa^2 \log_2 \frac{1}{\kappa})$.
- Online algorithms fail to find a solution for sufficiently large densities.

Universality: Gaussianity of **disorder** $\mathcal{M}$ immaterial. Extends via **Berry-Esseen**

Future Work

Some major challenges.

- (Ensemble) Multi-OGP for $\kappa = O(1)$.
- Closing the gap, κ^2 vs $\kappa^2 \log_2 \frac{1}{\kappa}$.
  - Rate $\kappa^2 \log_2 \frac{1}{\kappa}$ **best possible** with our approach.
  - More delicate structure [BH21, HS21]?
- Establishing stability of Bansal-Spencer algorithm [BS20].
- True algorithmic threshold.

A Conjecture: True Algorithmic Threshold

$\alpha_m^*(\kappa)$: smallest $\alpha < \alpha_c(\kappa)$ such that $m-$OGP holds (with proper $0 < \eta < \beta < 1$). Set

$$\alpha_{\mathrm{ALG}}(\kappa) = \lim_{m \to \infty} \alpha_m^*(\kappa).$$

Conjecture (Existence Threshold for Efficient Algorithms)

Let $\alpha > \alpha_{\mathrm{ALG}}(\kappa)$, $\mathcal{M} \in \mathbb{R}^{M \times n}$ ($M = \lfloor n\alpha \rfloor$) with i.i.d.$\mathcal{N}(0,1)$ entries. Then, **no** *(randomized) efficient algorithm $\mathcal{A}$ such that*

$$\mathbb{P}\left[\left\|\mathcal{M}\mathcal{A}(\mathcal{M})\right\|_\infty \le \kappa\sqrt{n}\right] \ge 1 - o(1).$$

Future Work: Bigger Challenges

What is the largest class of algorithms ruled out by OGP?

Naturally includes **stable** algorithms and **MCMC**.

Is there a problem ***with*** *OGP* ***yet*** *admitting a polynomial-time algorithm?*

Thank you!

References I

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.

Dimitris Achlioptas, Amin Coja-Oghlan, and Federico Ricci-Tersenghi, *On the solution-space geometry of random constraint satisfaction problems*, Random Structures & Algorithms **38** (2011), no. 3, 251–268.

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).

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

References II

- 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.
- Carlo Baldassi, *Generalization learning in a perceptron with binary synapses*, Journal of Statistical Physics **136** (2009), no. 5, 902–916.
- Carlo Baldassi and Alfredo Braunstein, *A max-sum algorithm for training discrete neural networks*, Journal of Statistical Mechanics: Theory and Experiment **2015** (2015), no. 8, P08008.
- Matthew Brennan and Guy Bresler, *Optimal average-case reductions to sparse pca: From weak assumptions to strong hardness*, arXiv preprint arXiv:1902.07380 (2019).

References III

- Carlo Baldassi, Alfredo Braunstein, Nicolas Brunel, and Riccardo Zecchina, *Efficient supervised learning in networks with binary synapses*, Proceedings of the National Academy of Sciences **104** (2007), no. 26, 11079–11084.
- Matthew Brennan, Guy Bresler, and Wasim Huleihel, *Reducibility and computational lower bounds for problems with planted sparse structure*, arXiv preprint arXiv:1806.07508 (2018).
- 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.
- Guy Bresler and Brice Huang, *The algorithmic phase transition of random k-sat for low degree polynomials*, arXiv preprint arXiv:2106.02129 (2021).

References IV

Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin, *A nearly tight sum-of-squares lower bound for the planted clique problem*, SIAM Journal on Computing **48** (2019), no. 2, 687–735.

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

Nikhil Bansal and Joel H. Spencer, *On-line balancing of random inputs*, Random Structures and Algorithms **57** (2020), no. 4, 879–891 (English (US)), Funding Information: This research was supported by the NWO Vici grant, 639.023.812. ERC Consolidator grant, 617951 [N.B.]. Funding information Publisher Copyright: © 2020 Wiley Periodicals LLC.

Alfredo Braunstein and Riccardo Zecchina, *Learning by message passing in networks of discrete synapses*, Physical review letters **96** (2006), no. 3, 030201.

References V

- Amin Coja-Oghlan, Amir Haqshenas, and Samuel Hetterich, *Walksat stalls well below satisfiability*, SIAM Journal on Discrete Mathematics **31** (2017), no. 2, 1160–1173.
- 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.
- Karthekeyan Chandrasekaran and Santosh S Vempala, *Integer feasibility of random polytopes: random integer programs*, Proceedings of the 5th conference on Innovations in theoretical computer science, 2014, pp. 449–458.
- Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart, *Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures*, 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2017, pp. 73–84.

References VI

- 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.
- Ronen Eldan and Mohit Singh, *Efficient algorithms for discrepancy minimization in convex sets*, Random Structures & Algorithms **53** (2018), no. 2, 289–307.
- 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.
- Elizabeth Gardner, *Maximum storage capacity in neural networks*, EPL (Europhysics Letters) **4** (1987), no. 4, 481.
- ______, *The space of interactions in neural network models*, Journal of physics A: Mathematical and general **21** (1988), no. 1, 257.
- Elizabeth Gardner and Bernard Derrida, *Optimal storage properties of neural network models*, Journal of Physics A: Mathematical and general **21** (1988), no. 1, 271.

References VII

- 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.
- ______, *The overlap gap property and approximate message passing algorithms for p-spin models*, The Annals of Probability **49** (2021), no. 1, 180–205.
- David Gamarnik, Aukosh Jagannath, and Subhabrata Sen, *The overlap gap property in principal submatrix recovery*, arXiv preprint arXiv:1908.09959 (2019).
- 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), 2020.
- ______, *Circuit lower bounds for the p-spin optimization problem*, arXiv preprint arXiv:2109.01342 (2021).

References VIII

- David Gamarnik and Eren C Kızıldağ, *Algorithmic obstructions in the random number partitioning problem*, arXiv preprint arXiv:2103.01369 (2021).
- David Gamarnik and Madhu Sudan, *Limits of local algorithms over sparse random graphs*, Ann. Probab. **45** (2017), no. 4, 2353–2376.
- ______, *Performance of sequential local algorithms for the random nae-k-sat problem*, SIAM Journal on Computing **46** (2017), no. 2, 590–619.
- Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer, *The power of sum-of-squares for detecting hidden structures*, 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2017, pp. 720–731.
- Samuel Brink Klevit Hopkins, *Statistical inference and the sum of squares method*.

References IX

- Brice Huang and Mark Sellke, *Tight lipschitz hardness for optimizing mean field spin glasses*, arXiv preprint arXiv:2110.07847 (2021).
- Samuel B Hopkins, Jonathan Shi, and David Steurer, *Tensor principal component analysis via sum-of-square proofs*, Conference on Learning Theory, 2015, pp. 956–1006.
- Mark Jerrum, *Large cliques elude the metropolis process*, Random Structures & Algorithms **3** (1992), no. 4, 347–359.
- Werner Krauth and Marc Mézard, *Storage capacity of memory networks with binary couplings*, Journal de Physique **50** (1989), no. 20, 3057–3066.
- 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.

References X

Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira, *Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio*, arXiv preprint arXiv:1907.11636 (2019).

Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová, *Phase transitions in sparse pca*, 2015 IEEE International Symposium on Information Theory (ISIT), IEEE, 2015, pp. 1635–1639.

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

Marc Mézard, Thierry Mora, and Riccardo Zecchina, *Clustering of solutions in the random satisfiability problem*, Physical Review Letters **94** (2005), no. 19, 197205.

References XI

- 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.
- Thomas Rothvoss, *Constructive discrepancy minimization for convex sets*, SIAM Journal on Computing **46** (2017), no. 1, 224–234.
- Mustazee Rahman and Balint Virag, *Local algorithms for independent sets are half-optimal*, The Annals of Probability **45** (2017), no. 3, 1543–1577.
- Zbynek Sidák, *On multivariate normal probabilities of rectangles: their dependence on correlations*, The Annals of Mathematical Statistics **39** (1968), no. 5, 1425–1434.
- Alexander S Wein, *Optimal low-degree hardness of maximum independent set*, arXiv preprint arXiv:2010.06563 (2020).

References XII

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

Lenka Zdeborová and Florent Krzakala, *Statistical physics of inference: Thresholds and algorithms*, Advances in Physics **65** (2016), no. 5, 453–552.