A Curious Case of Symmetric Binary Perceptron Model: Algorithms and Barriers

David Gamarnik [1] Eren C. Kızıldağ [2] Will Perkins [3] Changji Xu [4]

[1]MIT ORC [2]MIT EECS & LIDS [3]University of Illinois at Chicago [4]Harvard CMSA

Symmetric Binary Perceptron (SBP)

Setup: Fix $\kappa,\alpha>0$, set $M=\lfloor n\alpha\rfloor\in\mathbb{N}$. Generate i.i.d. $X_i\stackrel{d}{=}\mathcal{N}(0,I_n)$, $1\le i\le M$. Define

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

where $\mathcal{B}_n=\{-1,1\}^n$ and $\mathcal{M}\in\mathbb{R}^{M\times n}$ is the matrix of **disorder** with rows $X_1,\dots,X_M\in\mathbb{R}^n$.

Algorithmic Goal: Find a $\sigma\in S_\alpha(\kappa)$ in **polynomial-time** whenever $S_\alpha(\kappa)\ne\varnothing$ (whp).

Motivation

Neural Networks: Toy one-layer neural network (Wendel'62, Cover'65).

- Patterns $X_i\in\mathbb{R}^n$ to be **stored**.
- **Storage:** Find $\sigma\in\mathcal{B}_n$ "consistent" with X_i's: $\langle\sigma,X_i\rangle\ge 0$.

Constraint Satisfaction Problems: X_i rules out certain $\sigma\in\mathcal{B}_n$. **Constraint Density:** $\alpha=M/n$.

Discrepancy Theory: Given $\mathcal{M}\in\mathbb{R}^{M\times n}$, explore its discrepancy $\min_{\sigma\in\mathcal{B}_n}\|\mathcal{M}\sigma\|_\infty$.

Existential and Algorithmic Guarantees

Sharp Phase Transition. Let $\alpha_c(\kappa)=-1/\log_2\mathbb{P}\left[|\mathcal{N}(0,1)|\le\kappa\right]$. Perkins-Xu'21, Abbe-Li-Sly'21:

$$S_\alpha(\kappa)\ne\varnothing\text{ (whp) if }\alpha<\alpha_c(\kappa).\quad S_\alpha(\kappa)=\varnothing\text{ (whp) if }\alpha>\alpha_c(\kappa).$$

Algorithmic (Polynomial-Time). Bansal-Spencer'20: for $\alpha=O(\kappa^2)$, outputs a $\sigma_{\rm ALG}\in S_\alpha(\kappa)$ (whp).

A Statistical-to-Computational Gap

Gap between existential guarantee and the best polynomial-time algorithmic guarantee.

Most pronounced for $\kappa\to 0$:

- $S_\alpha(\kappa)\ne\varnothing$ (whp) iff $\alpha<-1/\log_2\kappa$. Algorithms exist for $\alpha=O(\kappa^2)$.
- A striking gap: $-1/\log_2\kappa$ vs κ^2.

Source of this gap/hardness?

Extreme Clustering and Freezing

Also known as Frozen 1-RSB in physics. For **any** $0<\alpha<\alpha_c(\kappa)$:

- **Typical** solutions of SBP are isolated (whp). Distance to nearest solution is $\Theta(n)$.
- Suggests algorithmic hardness (Achlioptas & Coja-Oghlan'08).

A Conundrum: Extreme clustering/freezing coexist with polynomial-time algorithms.

Study of Statistical-to-Computational Gap

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...*

Average-Case Problems: No analogue of worst-case theory (such as $P\ne NP$).

Rigorous Evidences of Hardness: *low-degree methods, reductions from the planted clique, failure of MCMC, failure of BP/AMP, SoS/SQ lower bounds,...*

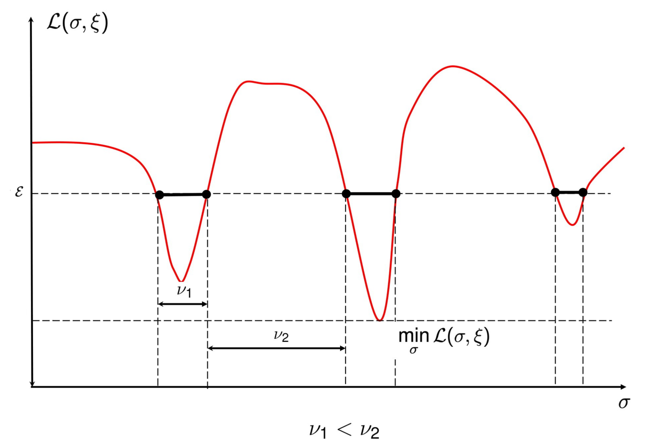
Overlap Gap Property (OGP)

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

- Generic optimization problem with random instance ξ: $\min_{\sigma\in\Theta}\mathcal{L}(\sigma,\xi)$.
- (Informally) OGP for **energy** $\mathcal{E}$ if $\exists 0<\nu_1<\nu_2$ s.t. w.h.p. over ξ, $\forall\sigma_1,\sigma_2\in\Theta$,

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

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



First algorithmic implication: Finding maximum independent set in $\mathbb{G}_d(n)$. (Gamarnik-Sudan'13).

Problems with OGP: Many, *random k-SAT, p−spin model, number partitioning...*

OGP as a Provable Barrier to Algorithms: *WALKSAT, local algorithms, stable algorithms, low-degree polynomials, approximate message passing (AMP), MCMC, low-depth circuits, QAOA...*

Landscape Results: Presence of OGP

Consider i.i.d. $\mathcal{M}_i\in\mathbb{R}^{M\times n}$, $0\le i\le m$, each with i.i.d. $\mathcal{N}(0,1)$ entries. Interpolate:

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

Fix $\kappa>0$. SBP exhibits **Ensemble $m-$OGP** with $(m,\beta,\eta,\mathcal{I})$, if for any $\sigma_1,\dots,\sigma_m\in\mathcal{B}_n$ with

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

there exists $1\le i<j\le m$ such that $n^{-1}\langle\sigma_i,\sigma_j\rangle\notin(\beta-\eta,\beta)$.

$m-$tuples: Hardness for broader range of parameters (i.e. lower threshold for α).

Ensemble: Correlated instances. Rule out any sufficiently stable algorithm.

Small κ regime, $\kappa\to 0$: Statistical-to-Computational Gap is most pronounced.

Theorem. $\forall\kappa>0$ small and $\mathcal{I}\subset[0,\pi/2]$ with $|\mathcal{I}|\le\exp(O(n))$, there exists $m\in\mathbb{N}$ and $1>\beta>\eta>0$ such that the SBP exhibits (whp) the Ensemble $m-$OGP with $(m,\beta,\eta,\mathcal{I})$ for $\alpha=\Omega(\kappa^2\log\frac{1}{\kappa})$.

- Nearly **tight**: Matches algorithmic κ^2 threshold up to $\log\frac{1}{\kappa}$ factor.
- $\beta\gg\eta$: no equidistant $m-$tuples each satisfying constraint $\mathcal{M}_i(\tau_i)$, $1\le i\le m$.

Large κ regime: Set $\kappa=1$, $\alpha_c(\kappa)\approx 1.8158$. Thus $S_\alpha(\kappa)\ne\varnothing$ (whp) iff $\alpha<1.8158$.

Theorem. Let $\kappa=1$. $\exists 0<\beta_2,\beta_3,\eta_2,\eta_3<1$ (where $\beta_i>\eta_i$) such that the following holds whp:

- SBP exhibits Ensemble $2-$OGP with $(2,\beta_2,\eta_2,\mathcal{I})$ for $\alpha\ge 1.71$.
- SBP exhibits Ensemble $3-$OGP with $(3,\beta_3,\eta_3,\mathcal{I})$ for $\alpha\ge 1.67$.

Algorithmic Hardness Results

Algorithm $\mathcal{A}:\mathbb{R}^{M\times n}\to\mathcal{B}_n$, potentially randomized.

Stable Algorithms. Informally, $\mathcal{A}$ is stable if small change in X yields small change in $\mathcal{A}(X)$.

Success:

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

Stability: $\exists\rho\in(0,1]$ such that for i.i.d. $\mathcal{M},\overline{\mathcal{M}}\in\mathbb{R}^{M\times n}$ with ${\rm Cov}(\mathcal{M}_{ij},\overline{\mathcal{M}}_{ij})=\rho$

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

AMP and low-degree polynomials are stable (Gamarnik-Jagannath-Wein'20).

Question: *"Are known efficient algorithms for perceptron models stable?"*

Theorem. Kim-Roche algorithm (Kim-Roche'98) for the asymmetric perceptron is stable.

$m-$OGP $\implies$ Failure of Stable Algorithms.

Theorem. Stable algorithms fail to find a solution for the SBP for $\alpha=\Omega(\kappa^2\log\frac{1}{\kappa})$.

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.
- Uses Ramsey Theory (Gamarnik-Kızıldağ'21).

Failure of Online Algorithms for High Densities:

Columns of $\mathcal{M}$: $\mathcal{C}_1,\dots,\mathcal{C}_n\in\mathbb{R}^M$. $\mathcal{A}$ is **online** if $\exists f_t$ s.t. $\sigma_t=f_t(\mathcal{C}_i:1\le i\le t)$ for $1\le t\le n$.

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

Future Directions

Algorithmic Threshold: Let $\alpha_m(\kappa)$ be the smallest density such that for some $0<\eta<\beta<1$, SBP exhibits (whp) Ensemble $m-$OGP with $(m,\beta,\eta,\{0\})$ for $\alpha\ge\alpha_m(\kappa)$. Define

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

Conjecture. $\alpha_\infty^*(\kappa)$ marks the true algorithmic threshold of SBP.

- Bansal-Spencer algorithm is likely optimal (up to logaritmic factors).
- $\log\frac{1}{\kappa}$ factor? More delicate structure (Wein'20, Bresler-Huang'21, Huang-Sellke'21).

Stability of Other Algorithms: *"Is Bansal-Spencer algorithm stable? Other discrepancy algorithms?"*

Asymmetric Perceptron: Many open problems.

- Existence/Location of sharp phase transition point. *Krauth-Mézard (89) prediction.*
- Rigorously verifying Frozen 1-RSB picture.
- OGP and failure of stable algorithms.

More Enthusiastic Questions on OGP.

- Largest class of algorithms ruled out by OGP: Includes stable algorithms, MCMC, etc.
- Counterexample to OGP: Is there a model where efficient algorithms coexist with OGP?