Algorithmic Obstructions in the Random Number Partitioning Problem

Eren C. Kızıldağ (MIT)

Joint work with David Gamarnik (MIT)

Simons CCSI Reading Group: Overlap Gap Property

September 21, 2021

Overview

1. Introduction
   - Problem Definition
   - Applications
   - *Statistical-to-Computational Gaps*
   - The Overlap Gap Property (OGP)
2. Contributions: Properties of the Landscape of NPP
   - 2-OGP
   - Ensemble-m-OGP with $m = O(1)$
   - Ensemble-m-OGP with $m = \omega(1)$
3. Contributions: Algorithmic Hardness Results
   - Failure of Stable Algorithms
   - Failure of MCMC Methods
4. Conclusion and Future Research
   - Summary of Contributions
   - Future Work

Number Partitioning Problem (NPP): Definition.

- Given n items $X_1, \dots, X_n$; partition them into two "bins" with total weights as close as possible:

$$\min_{A \subset [n]} \left| \sum_{i \in A} X_i - \sum_{i \in A^c} X_i \right|.$$

- Equivalently

$$\min_{\sigma \in \mathcal{B}_n} \left| \langle \sigma, X \rangle \right|, \quad \text{where} \quad \mathcal{B}_n = \{-1, 1\}^n \quad \text{and} \quad \langle \sigma, X \rangle = \sum_{1 \le i \le n} \sigma_i X_i.$$

- **Our focus.** Items X_i are i.i.d. standard normal: $X_i \stackrel{d}{=} \mathcal{N}(0,1)$.

Application of NPP: Design of Randomized Controlled Trials

- **Randomized controlled trials.** Gold standard for clinical trials [KAK19, HSSZ19].
- n persons with covariate info (age, weight, height,...) $X_i \in \mathbb{R}^d$, $1 \le i \le n$.
- Split into two groups (*treatment* and *control*) with similar "features":

$$\min_{\sigma \in \mathcal{B}_n} \left\| X\sigma \right\|_\infty, \quad \text{where} \quad X = (X_1, X_2, \dots, X_n) \in \mathbb{R}^{d \times n}.$$

- **Goal.** Accurate inference for a treatment effect.

More on "Why NPP is interesting to study?"

Vast literature...

- (Many) other applications, including *Multiprocessor scheduling, VLSI design, cryptography...* [CL91]
- Also of theoretical importance, in *theoretical CS* and *statistical mechanics*:
  - **TCS.** One of six basic NP-complete problems by [GJ90].
  - **Statistical Physics.** Locally REM, phase transitions [BCP01, BCMN09a, BCMN09b].
- *Combinatorial discrepancy theory.*

Our Work: *Statistical-to-Computational Gap* of NPP and the OGP

Statistical-to-computational gaps: Gap between existential guarantees and (polynomial-time) algorithmic guarantees.

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

This work:

- Overlap Gap Property (OGP): Intricate geometric property.
- Leverage OGP to *rule out* certain classes of algorithms.

NPP: Available Existential Guarantees

$X_i \in \mathbb{R}^d$, $1 \le i \le n$. Define

$$\mathcal{D}_n \triangleq \min_{\sigma \in \mathcal{B}_n} \|X\sigma\|_\infty \quad \text{where} \quad X = (X_1, \dots, X_n) \in \mathbb{R}^{d \times n}.$$

Worst-case, [Spe85]: For $d = n$ and $\max_i \|X_i\|_\infty \le 1$, $\mathcal{D}_n \le 6\sqrt{n}$. Non-constructive.

Average-case: Assume $X_i \stackrel{d}{=} \mathcal{N}(0, I_d)$, $1 \le i \le n$, i.i.d. For $1 \le d \le o(n)$,

$$\mathcal{D}_n = \Theta\left(\sqrt{n}2^{-n/d}\right), \quad \text{w.h.p.}$$

[KKLO86]: $d = 1$. [Cos09]: $d = O(1)$. [TMR20]: $\omega(1) \le d \le o(n)$.

Average-case, $\mathbb{E}$: [Lue98]: for $d = 1$,

$$\mathbb{E}\left[\mathcal{D}_n\right] = O\left(2^{-cn}\right).$$

NPP: Available (Polynomial-Time) Algorithmic Guarantees

$X_i \overset{d}{=} \mathcal{N}(0, I_d)$, $1 \le i \le n$ i.i.d.

- [KK82]: For $d = 1$; returns $\sigma_{\mathrm{ALG}} \in \mathcal{B}_n$ with
$$|\langle \sigma_{\mathrm{ALG}}, X \rangle| = 2^{-\Theta(\log^2 n)}, \quad \text{w.h.p.}$$
- A simpler heuristic, Largest Differencing Method *(LDM)*. Also good performance [Yak96]:
$$\mathbb{E}[\mathrm{LDM}] = n^{-\Theta(\log n)}.$$
- [TMR20]: For $2 \le d \le O(\sqrt{\log n})$, returns a $\sigma_{\mathrm{ALG}} \in \mathcal{B}_n$ with
$$\left\| X\sigma_{\mathrm{ALG}} \right\|_\infty = \exp\left(-\Omega\left(\frac{\log^2 n}{d}\right)\right), \quad \text{w.h.p.}$$

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

- For $X \stackrel{d}{=} \mathcal{N}(0, I_n)$,

$$\min_{\sigma\in\mathcal{B}_n} |\langle \sigma, X\rangle| = \Theta(\sqrt{n}2^{-n}) \quad \text{vs} \quad |\langle \sigma_{\mathrm{ALG}}, X\rangle| = 2^{-\Theta(\log^2 n)}.$$

- Ignoring $\sqrt{n}$, a striking gap: 2^{-n} vs $2^{-\Theta(\log^2 n)}$.

Source of this gap/hardness?

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, 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:

$$\min_{\theta \in \Theta} \mathcal{L}(\sigma, \xi).$$

(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*.

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

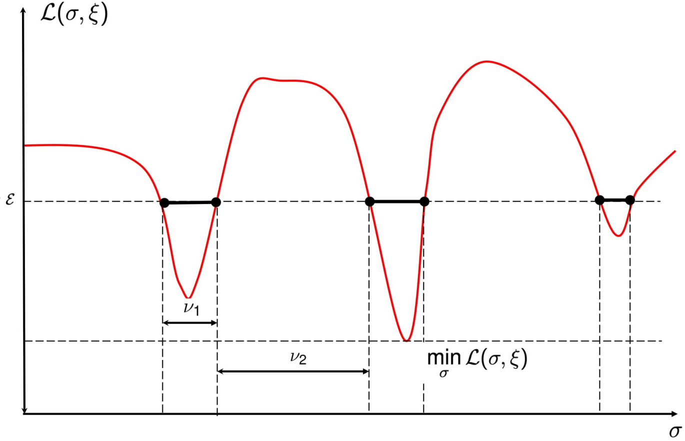
For $\Theta = \mathcal{B}_n = \{-1, 1\}^n$, normalized overlap:

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

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

Overlap Gap Property - A Pictorial Illustration

OGP for $\mathcal{E}$.



$\nu_1 < \nu_2$

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, 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, GJ21, GJS19, GZ19, AWZ20, BH21]...

Overview

Our Contributions: 2−OGP for NPP.

Recall $\min_{\sigma\in\mathcal{B}_n}|\langle\sigma,X\rangle|$, $X\stackrel{d}{=}\mathcal{N}(0,I_n)$, and its gap 2^{-n} vs $2^{-\Theta(\log^2 n)}$.

Theorem (2-OGP)

(Informally) OGP holds below $2^{-\frac{n}{2}}$.
Formally, $\forall\epsilon\in(1/2,1)$, $\exists\rho:=\rho(\epsilon)\in(0,1)$ *such that if* $\sigma,\sigma'\in\mathcal{B}_n$ *achieve*

$$|\langle\sigma,X\rangle|=O(\sqrt{n}2^{-\epsilon n})\quad\textit{and}\quad|\langle\sigma',X\rangle|=O(\sqrt{n}2^{-\epsilon n})$$

then either $\sigma=\sigma'$ *or* $n^{-1}|\langle\sigma,\sigma'\rangle|\le\rho$ *w.h.p. That is,* $n^{-1}|\langle\sigma,\sigma'\rangle|\notin(\rho,\frac{n-2}{n}]$.

- Partitions achieving better than $2^{-\frac{n}{2}}$ are isolated vectors separated by $\Theta(n)$ distance.
- Known as *Frozen 1-RSB*. Similar picture for *Symmetric Ising Perceptron* [PX21, ALS21].
- Yields existence of a *Free Energy Well (FEW)*: failure of **Glauber dynamics** (later).

2-OGP: Proof Sketch via First Moment Method

- Let N count the # of such (σ, σ'): $\mathbb{P}(N \geq 1) \leq \mathbb{E}[N]$.
- Number of σ, σ' with $n^{-1}|\langle \sigma, \sigma' \rangle| \geq \rho$ is $2^{n+nh((1-\rho)/2)}$, where $h(\cdot)$ is binary entropy.
- σ, σ' with $\mathcal{O}(\sigma, \sigma') = \rho$. Let $Y = n^{-\frac{1}{2}}\langle \sigma, X \rangle$ and $Y' = n^{-\frac{1}{2}}\langle \sigma', X \rangle$. Then,

$$\mathbb{P}\Big((Y, Y') \in (-2^{-\epsilon n}, 2^{-\epsilon n})^2\Big) \approx O(2^{-2\epsilon n}).$$

- Hence

$$\mathbb{E}[N] \leq \exp_2\left(n + nh\left(\frac{1-\rho}{2}\right) - 2n\epsilon\right).$$

- As $\epsilon > 1/2$,

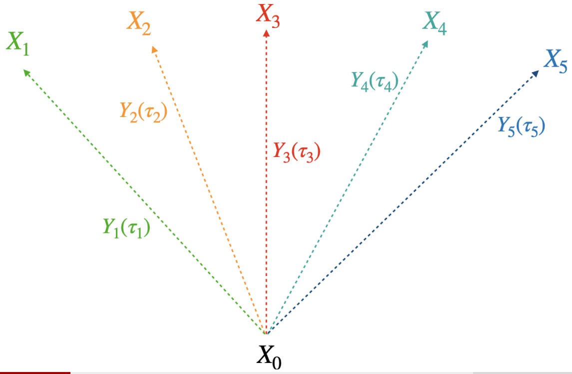
$$1 - 2\epsilon + h((1-\rho)/2) < 0$$

  for a suitable $\rho < 1$.
- Thus, $\mathbb{E}[N] \leq \exp(-\Theta(n))$.

Ensemble-Multi-OGP for NPP.

- $2-$OGP holds below $2^{-\frac{n}{2}}$. Still large gap with $2^{-\Theta(\log^2 n)}$.
- Consider independent instances $X_0,\dots,X_m \stackrel{d}{=} \mathcal{N}(0,I_n)$ i.i.d.; and interpolate

$$Y_i(\tau)=\sqrt{1-\tau^2}X_0+\tau X_i \stackrel{d}{=} \mathcal{N}(0,I_n),\quad \tau\in[0,1],\quad 1\le i\le m.$$



Ensemble-Multi-OGP for NPP

m-tuples $\sigma_i \in \mathcal{B}_n$ ($m-$OGP); each near-optimal w.r.t. $Y_i(\tau_i)$, $\exists \tau_i \in [0,1]$ (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, GJ21, BH21].

Our Contributions: Ensemble $m-$OGP for NPP.

Theorem (Ensemble-multi-OGP)

(Informally) Ensemble $m-$OGP holds below any $2^{-\epsilon n}$, $\epsilon > 0$.
Formally, $\forall \epsilon > 0$, $\forall \mathcal{I} \subset [0,1]$ with $|\mathcal{I}| = 2^{o(n)}$, $\exists m \in \mathbb{N}$, $\exists 1 > \beta > \eta > 0$ s.t. if

$$|\langle \sigma_i, Y_i(\tau_i)\rangle| = O\left(\sqrt{n}2^{-\epsilon n}\right), \quad \tau_i \in \mathcal{I}, \quad 1 \le i \le m$$

then w.h.p. $\exists 1 \le i < j \le m$ such that

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

- No m partitions across interpolated instances of energy $2^{-\epsilon n}$ and overlaps in $(\beta - \eta, \beta)$.
- Proof based on first moment method.

Our Contributions: No $m-$OGP for NPP.

Still striking gap between $2^{-\epsilon n}$ and $2^{-\Theta(\log^2 n)}$.

Theorem (No OGP)

(Informally) No OGP above $2^{-o(n)}$.
Formally, $\forall \omega(1) \le f(n) \le o(n)$, $\forall \beta, \eta \in (0,1)$, and $\forall m \in \mathbb{N}$; w.h.p. $\exists \sigma_i$, $1 \le i \le m$ such that

$$|\langle \sigma_i, X \rangle| = O(\sqrt{n} 2^{-f(n)}) \quad \textit{and} \quad n^{-1} |\langle \sigma_i, \sigma_j \rangle| \in [\beta - \eta, \beta + \eta]$$

- Overlaps of partitions with energy worse than $2^{-o(n)}$ span entire $(0,1)$.
- Proof based on second moment method: let M count such $m-$tuples. Then,

$$\mathbb{P}(M \ge 1) \ge \mathbb{E}[M]^2 / \mathbb{E}[M^2].$$

If $\mathrm{Var}(M) = o(\mathbb{E}[M]^2)$ then $\mathbb{P}(M \ge 1) = 1 - o_n(1)$.

Our Contributions: Ensemble $m-$OGP for $\mathtt{NPP}$ with $m = \omega(1)$.

NEW IDEA: Analyze m growing w.r.t. n.

Theorem (Ensemble-multi-OGP, $m = \omega(1)$)

(Informally) Ensemble $m-$OGP holds below $2^{-\omega(\sqrt{n \log n})}$ for super-constant m.
Formally, $\forall \omega(\sqrt{n \log n}) \le E_n \le o(n)$, $\forall \mathcal{I} \subset [0,1]$ *with* $|\mathcal{I}| = n^{O(1)}$, $\exists m_n \in \mathbb{N}$, $\exists 1 > \beta_n > \eta_n > 0$ *s.t. if*

$$|\langle \sigma_i, Y_i(\tau_i)\rangle| \le \sqrt{n} 2^{-E_n}, \quad \tau_i \in \mathcal{I}, \quad 1 \le i \le m_n$$

then w.h.p. $\exists 1 \le i < j \le m_n$ *such that*

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

- **First** $m-$OGP result with $m = \omega_n(1)$.
- The rate $\omega(\sqrt{n \log n})$ appears **unimprovable**.

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 [GJ21].
- 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], planted clique problem [GZ19], sparse PCA [AWZ20].

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

Stable Algorithms: Formal Definition

- Algorithm $\mathcal{A}$, $\mathcal{A}(X) = \sigma \in \mathcal{B}_n$.
- Potentially **randomized**.
- **Informal**: $\mathcal{A}$ is stable if small change in X yields small change in $\mathcal{A}(X)$.

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

Definition

(a) **Success**:

$$\mathbb{P}\left(n^{-\frac{1}{2}}|\langle X, \mathcal{A}(X)\rangle| \le E\right) \ge 1 - p_f.$$

(b) **Stability**: $\exists \rho \in (0,1]$, $X, Y \stackrel{d}{=} \mathcal{N}(0, I_n)$ with $\mathrm{Cov}(X, Y) = \rho I_n$;

$$\mathbb{P}\left(d_H\left(\mathcal{A}(X), \mathcal{A}(Y)\right) \le f + L\|X - Y\|_2^2\right) \ge 1 - p_{\mathrm{st}}.$$

Stable Algorithms: Which Algorithms are Stable?

Stable algorithms include

- Approximate message passing type algorithms [GJ21].
- Low-degree polynomial based algorithms [GJW20].

Conjecture

Largest differencing (LDM) algorithm is stable.

Verified by simulations.

OGP implies Failure of Stable Algorithms

Theorem (Stable Algorithms Fail for $\mathtt{NPP}$)

Stable algorithms can't achieve value better than

$$\exp\left(-\omega\left(\frac{n}{\log^{1/5} n}\right)\right):$$

$\forall \epsilon \in (0, 1/5)$, $\forall \omega(n \log^{-1/5+\epsilon} n) \le E_n \le o(n)$, *there is no stable $\mathcal{A}$ that w.h.p. returns a σ with energy 2^{-E_n} (with appropriate $f, \rho', p_f, p_{\mathrm{st}}$).*

- For **extreme** case, $E_n = \Theta(n)$: rule out $p_f, p_{\mathrm{st}} = O(1)$.
- **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.
- Rate $2^{-\omega(n \log^{-1/5} n)}$: Via Ramsey Theory.

An MCMC Dynamics for NPP

- Let $X \stackrel{d}{=} \mathcal{N}(0, I_n)$; and define **Hamiltonian** $H(\sigma) \triangleq n^{-\frac{1}{2}}|\langle \sigma, X\rangle|$.
- Define **Gibbs distribution** at inverse temperature $\beta > 0$ on $\mathcal{B}_n$:

$$\pi_\beta(\sigma) = \frac{1}{Z_\beta}\exp(-\beta H(\sigma)) \quad \text{where} \quad Z_\beta = \sum_{\tau \in \mathcal{B}_n} \exp(-\beta H(\tau)).$$

- **Fact:** As $\beta \to \infty$, π_β **concentrates** on

$$\left\{\sigma : H(\sigma) = \min_{\tau \in \mathcal{B}_n} H(\tau)\right\}.$$

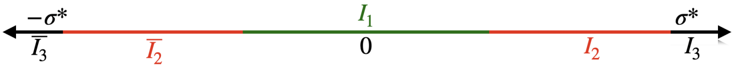
- Construct $\mathbb{G} = (V, E)$ with $V = \mathcal{B}_n$ and $(\sigma, \sigma') \in E \iff d_H(\sigma, \sigma') = 1$.
- Consider any nearest neighbor MC $(X_t)_{t \geq 0}$ on $\mathbb{G}$ reversible w.r.t. π_β.

OGP implies FEW

- Let $(\pm)\sigma^* = \min_{\sigma\in\mathcal{B}_n} |\langle \sigma, X\rangle|$.
- For $\epsilon \in (1/2, 1)$, let $\rho := \rho(\epsilon)$ be 2-OGP parameter.
- Define

$$I_1 = \left\{\sigma : -\rho \le \frac{1}{n}\langle \sigma,\sigma^*\rangle \le \rho\right\}, \quad I_2 = \left\{\sigma : \rho \le \frac{1}{n}\langle \sigma,\sigma^*\rangle \le \frac{n-2}{n}\right\}, \quad \text{and} \quad I_3 = \{\sigma^*\}.$$

- Finally, let $\overline{I_2} := -I_2$ and $\overline{I_3} := -I_3$.



OGP implies FEW

Theorem (Free Energy Well in NPP)

For $\beta = \Omega(n2^{n\epsilon})$, w.h.p. (w.r.t. $X \stackrel{d}{=} \mathcal{N}(0, I_n)$),

$$\min\left\{\pi_\beta\left(I_1\right), \pi_\beta\left(I_3\right)\right\} \ge e^{\Omega(n)}\pi_\beta\left(I_2\right).$$

- I_2 is a FEW with exponentially small **Gibbs** mass separating I_3 and $I_1 \cup \overline{I_2} \cup \overline{I_3}$.
- Consequence of $2-$OGP.
- *Exit time* from well is *exponential*: **Slow mixing**.

FEW: Proof Sketch

- Recall $H(\sigma^*) = H(-\sigma^*) = \Theta(2^{-n})$. Absorbing constants into $\beta > 0$,

$$\pi_\beta(I_3) = \pi_\beta(\overline{I_3}) = \exp\left(-\beta 2^{-n}\right)/Z_\beta.$$

- Due to 2−**OGP**, $\min_{\sigma \in I_2} H(\sigma) = \Omega(2^{-\epsilon n})$. Moreover, $|I_2| \sim 2^{nh((1-\rho)/2)}$. Hence,

$$\pi_\beta(I_2) = \sum_{\sigma \in I_2} \pi_\beta(\sigma) \le \frac{|I_2| \exp(-\beta 2^{-\epsilon n})}{Z_\beta} \sim \frac{1}{Z_\beta} \exp_2\left(nh\left(\frac{1-\rho}{2}\right) - \beta 2^{-\epsilon n}\right).$$

- Fix $\epsilon' \in (\epsilon, 1)$. By [KKLO86, Thm 3.1], w.p. $1 - O(1/n)$, $\exists \sigma'$ with $H(\sigma') = \Theta(2^{-\epsilon' n})$. Via $\bigcup$−bound, $\sigma' \in I_1$ w.h.p. Hence,

$$\pi_\beta(I_1) \ge \pi_\beta(\sigma') = \exp(-\beta 2^{-\epsilon' n})/Z_\beta.$$

- Combining, we get for $\beta = \Omega(n 2^{\epsilon n})$, $\pi_\beta(I_1) \wedge \pi_\beta(I_3) \ge e^{\Theta(n)} \pi_\beta(I_2)$.

From OGP to MCMC

FEW $\implies$ Failure of MCMC: tensor PCA [AGJ20].

OGP $\implies$ FEW.

$\therefore$ OGP $\implies$ Failure of MCMC:

sparse PCA [AWZ20], principal submatrix recovery [GJS19], planted clique [GZ19].

OGP implies FEW, which implies Failure of MCMC

Let $\partial S := \{\sigma : d_H(\sigma,\sigma^*) = 1\}$. Initialize $X_0 \stackrel{d}{=} \pi_\beta(\cdot \mid I_3 \cup \partial S)$. Define **escape time**

$$\tau_\beta := \inf\Big\{t \ge 1 : X_t \notin I_3 \cup \partial S \mid X_0 \sim \pi_\beta(\cdot \mid I_3 \cup \partial S)\Big\}.$$

Theorem (Slow Mixing)

$\forall \epsilon \in (1/2, 1)$ and $\beta = \Omega(n2^{n\epsilon})$, the following holds w.h.p. as $n \to \infty$, w.r.t. $X \stackrel{d}{=} \mathcal{N}(0, I_n)$:

- $\pi_\beta\left(I_1 \cup \overline{I_3}\right) \ge (1 + o_n(1))/2$.
- $\tau_\beta = e^{\Theta(n)}$.

Overview

Main Contributions

Statistical-to-Computational Gap of NPP: 2^{-n} vs $2^{-\Theta(\log^2 n)}$.

- Landscape of NPP:
  - Presence of $2-$OGP and (Ensemble) $m-$OGP (with $m = O(1)$ and $m = \omega(1)$).
  - Absence of $m-$OGP.
  - Presence of a FEW.
- Algorithmic hardness.
  - Stable algorithms fail to solve NPP with objective value below $2^{-\omega(n \log^{-1/5} n)}$.
  - Glauber dynamics mixes **slowly** for sufficiently small temperature.
- Expected number of local optima: $e^{\Theta(n)}$. First moment **evidence** for failure of **Greedy**.

Future Work

Some major challenges.

- Formally verifying stability of **LDM**.
- Proving algorithmic hardness all the way to $2^{-\omega(\sqrt{n\log n})}$.
  - Rate $2^{-\omega(n\log^{-1/5} n)}$ unimprovable by Ramsey.
- Still a significant gap $2^{-\omega(\sqrt{n\log n})}$ vs $2^{-\Theta(\log^2 n)}$.
  - Either prove hardness for $2^{-\omega(\log^2 n)}$: OGP not applicable.
  - Or devise a better (polynomial-time) algorithm achieving $2^{-\omega(\log^2 n)}$.
- *Slow mixing*
  - For **higher** temperatures (smaller β).
  - For **different** initialization, e.g. *uniform* case.

Future Work

Bigger Challenges:

- OGP rules out **stable** algorithms.
- *Can OGP rule out* **all** *polynomial-time algorithms?*
- 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.
- Gerard Ben Arous, Reza Gheissari, and Aukosh Jagannath, *Algorithmic thresholds for tensor pca*, Annals of Probability **48** (2020), no. 4, 2052–2087.
- Noga Alon, Michael Krivelevich, and Benny Sudakov, *Finding a large hidden clique in a random graph*, Random Structures & Algorithms **13** (1998), no. 3-4, 457–466.
- Emmanuel Abbe, Shuangping Li, and Allan Sly, *Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron*, arXiv preprint arXiv:2102.13069 (2021).

References II

Dimitris Achlioptas and Cristopher Moore, *Random k-sat: Two moments suffice to cross a sharp threshold*, SIAM Journal on Computing **36** (2006), no. 3, 740–762.

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

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

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

Christian Borgs, Jennifer Chayes, Stephan Mertens, and Chandra Nair, *Proof of the local rem conjecture for number partitioning. i: Constant energy scales*, Random Structures & Algorithms **34** (2009), no. 2, 217–240.

References III

———, *Proof of the local rem conjecture for number partitioning. ii. growing energy scales*, Random Structures & Algorithms **34** (2009), no. 2, 241–284.

Christian Borgs, Jennifer Chayes, and Boris Pittel, *Phase transition and finite-size scaling for the integer partitioning problem*, Random Structures & Algorithms **19** (2001), no. 3-4, 247–288.

Mohsen Bayati, David Gamarnik, and Prasad Tetali, *Combinatorial approach to the interpolation method and scaling limits in sparse random graphs*, Proceedings of the forty-second ACM symposium on Theory of computing, 2010, pp. 105–114.

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).
- Edward Grady Coffman and George S Lueker, *Probabilistic analysis of packing and partitioning algorithms*, Wiley-Interscience, 1991.
- Amin Coja-Oghlan, Amir Haqshenas, and Samuel Hetterich, *Walksat stalls well below satisfiability*, SIAM Journal on Discrete Mathematics **31** (2017), no. 2, 1160–1173.
- Amin Coja-Oglan and Konstantinos Panagiotou, *Catching the k-naesat threshold*, Proceedings of the forty-fourth annual ACM symposium on Theory of computing, 2012, pp. 899–908.

References V

Kevin P Costello, *Balancing gaussian vectors*, Israel Journal of Mathematics **172** (2009), no. 1, 145–156.

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.

Alan M Frieze and T Luczak, *On the independence and chromatic numbers of random regular graphs*, Journal of Combinatorial Theory, Series B **54** (1992), no. 1, 123–132.

Alan M Frieze, *On the independence number of random graphs*, Discrete Mathematics **81** (1990), no. 2, 171–175.

Michael R. Garey and David S. Johnson, *Computers and intractability; a guide to the theory of np-completeness*, W. H. Freeman & Co., USA, 1990.

References VI

- David Gamarnik and Aukosh Jagannath, *The overlap gap property and approximate message passing algorithms for p-spin models*, 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).
- David Gamarnik and Madhu Sudan, *Limits of local algorithms over sparse random graphs*, Ann. Probab. **45** (2017), no. 4, 2353–2376.

References VII

- ———, *Performance of sequential local algorithms for the random nae-k-sat problem*, SIAM Journal on Computing **46** (2017), no. 2, 590–619.
- Ian P Gent and Toby Walsh, *Phase transitions and annealed theories: Number partitioning as a case study'*, ECAI, PITMAN, 1996, pp. 170–174.
- David Gamarnik and Ilias Zadik, *The landscape of the planted clique problem: Dense subgraphs and the overlap gap property*, arXiv preprint arXiv:1904.07174 (2019).
- 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 VIII

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.

Christopher Harshaw, Fredrik Sävje, Daniel Spielman, and Peng Zhang, *Balancing covariates in randomized experiments using the gram-schmidt walk*, arXiv preprint arXiv:1911.03071 (2019).

Mark Jerrum, *Large cliques elude the metropolis process*, Random Structures & Algorithms **3** (1992), no. 4, 347–359.

Abba M Krieger, David Azriel, and Adam Kapelner, *Nearly random designs with greatly improved balance*, Biometrika **106** (2019), no. 3, 695–701.

Richard M Karp, *The probabilistic analysis of some combinatorial search algorithms.*

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

References IX

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

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.

George S Lueker, *Exponentially small bounds on the expected optimum of the partition and subset sum problems*, Random Structures & Algorithms **12** (1998), no. 1, 51–62.

References X

- Stephan Mertens, *Phase transition in the number partitioning problem*, Physical Review Letters **81** (1998), no. 20, 4281.
- Marc Mézard, Thierry Mora, and Riccardo Zecchina, *Clustering of solutions in the random satisfiability problem*, Physical Review Letters **94** (2005), no. 19, 197205.
- Will Perkins and Changji Xu, *Frozen 1-rsb structure of the symmetric ising perceptron*, arXiv preprint arXiv:2102.05163 (2021).
- Mustazee Rahman and Bálint Virág, *Local algorithms for independent sets are half-optimal*, Ann. Probab. **45** (2017), no. 3, 1543–1577.
- Joel Spencer, *Six standard deviations suffice*, Transactions of the American mathematical society **289** (1985), no. 2, 679–706.
- Paxton Turner, Raghu Meka, and Philippe Rigollet, *Balancing gaussian vectors in high dimension*, Conference on Learning Theory, PMLR, 2020, pp. 3455–3486.

References XI

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

Benjamin Yakir, *The differencing algorithm ldm for partitioning: a proof of a conjecture of karmarkar and karp*, Mathematics of Operations Research **21** (1996), no. 1, 85–99.

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

Details on LDM and PDM

LDM.

- Sort X_i: $X'_1 < X'_2 < \cdots < X'_n$.
- Apply differencing on X'_n and X'_{n-1}. Consider the list $L' = \{X'_1, \ldots, X'_{n-2}, |X'_n - X'_{n-1}|\}$.
- Recurse.

PDM.

- Sort X_i: $X'_1 < X'_2 < \cdots < X'_n$.
- Applying differencing on pairs (X'_n, X'_{n-1}), (X'_{n-2}, X'_{n-3}), and so on.
- Obtain a list of $\lfloor n/2 \rfloor$ items. Recurse.

A Heuristic Reasoning. Consider PDM when $X_i \stackrel{d}{=} \mathrm{Unif}[0,1]$. Each operation reduce size by $1/n$. Recurse $\sim \log n$ rounds: $n^{-\log n}$.

Details on Stable Algorithms

Algorithm $\mathcal{A} : \mathbb{R}^n \times \Omega \to \mathcal{B}_n$. $(\Omega, \mathbb{P}_\omega)$ *coin flips* of $\mathcal{A}$.

- $X \overset{d}{=} \mathcal{N}(0, I_n)$. Success guarantee w.r.t. $\mathcal{N}(0, I_n) \otimes \mathbb{P}_\omega$:

$$\mathbb{P}_{(X,\omega)\sim\mathcal{N}(0,I_n)\otimes\mathbb{P}_\omega}\left(n^{-\frac{1}{2}}\big|\langle X, \mathcal{A}(X,\omega)\rangle\big| \le E\right) \ge 1 - p_f.$$

- Need two $X, Y \overset{d}{=} \mathcal{N}(0, I_n)$ to talk about *stability*. To specify $\mathbb{P}_{X,Y}$, need Cov: $\mathrm{Cov}(X, Y) = \rho I$. Then, with respect to $(X, Y, \omega) \sim \mathbb{P}_{X,Y} \otimes \mathbb{P}_\omega$,

$$\mathbb{P}_{(X,Y,\omega):X\sim_\rho Y,\omega\sim\mathbb{P}_\omega}\left(d_H\big(\mathcal{A}(X,\omega), \mathcal{A}(Y,\omega)\big) \le f + L\|X - Y\|_2^2\right) \ge 1 - p_{\mathrm{st}}.$$

Details on Algorithmic Hardness Result for Stable Algorithms

- f turns out to be $c_1 n \log^{-O(1)} n$ for some $c_1 > 0$.
- p_f, p_{st} sub-exponential:

$$p_f, p_{\mathrm{st}} \simeq \exp_2\left(-2^{o(\log^{c'} n)}\right), \quad c' \in (0,1).$$

- For $E_n = \omega\left(n \log^{-1/5+\epsilon} n\right)$, $0 < \epsilon < 1/5$, explicit trade-off between c' and ϵ:

$$c' \simeq \left(\frac{1}{5} - \epsilon\right)\left(5 + \frac{\epsilon}{2}\right) = 1 - \frac{49\epsilon}{10} + \Theta(\epsilon^2).$$

  Any c' greater than this value (and less than 1) works.
- For $\epsilon = 1/5$ ($E_n = \Theta(n)$), $c' \to 0$:

$$p_f, p_{\mathrm{st}} = O(1) \qquad \text{suffice.}$$

Stable Algorithms Fail for NPP: Proof Sketch

- Fix E_n. $m-$OGP holds with (m,β,η): $[\beta-\eta,\beta]$ is the **forbidden** region.
- **Discretization** Q, required for η. T **"replicas"**.

$$Q \sim (n/E_n)^{4+\frac{\epsilon}{4}} \sim \log^{O(1)} n \quad \text{and} \quad T \sim \exp_2\left(2^{4mQ\log_2 Q}\right) \sim 2^{o(n)}.$$

Proof by contradiction: Suppose **randomized** $\mathcal{A}$ exists, reduce to **deterministic** $\mathcal{A}$.

Idea: Show a structure (contradicting with $m-$OGP) appears w.p.>0.

(1) Let $X_i \stackrel{d}{=} \mathcal{N}(0,I_n)$, $0\le i\le T$ i.i.d. Interpolate:

$$Y_i(\tau) \triangleq \sqrt{1-\tau^2}X_0+\tau X_i, \quad \tau\in[0,1], \quad 1\le i\le T.$$

(2) Let $\sigma_i(\tau) \triangleq \mathcal{A}(Y_i(\tau)) \in \mathcal{B}_n$. Define $\mathcal{O}^{(ij)}(\tau) \triangleq n^{-1}\langle \sigma_i(\tau),\sigma_j(\tau)\rangle \in [-1,1]$.

(3) **Discretize** $[0,1]$**:** $0=\tau_0<\tau_1<\cdots<\tau_Q=1$.

Stable Algorithms Fail for NPP: Proof Sketch

(4) Stability of $\mathcal{A}$ + Concentration $\implies$ Stability of $\mathcal{O}^{(ij)}(\tau)$:

$$\left|\mathcal{O}^{(ij)}(\tau_k) - \mathcal{O}^{(ij)}(\tau_{k+1})\right| \quad \text{is small,} \quad \text{for all} \quad 1 \le i < j \le T, 0 \le k \le Q-1.$$

(5) $\sigma_i(\tau)$ identical at $\tau = 0$: Overlaps all one. $Y_i(\tau)$ independent at $\tau = 1$.

(6) $\forall S \subset T$ with $|S| = m$, $\exists i_S, j_S \in S$ s.t. $\mathcal{O}^{(i_S,j_S)}(\cdot)$ eventually below $\beta - \eta$.

(7) Stability of $\mathcal{O}(\cdot)$ $\implies$

$$\exists 1 \le k \le Q : \quad \mathcal{O}^{(i_S,j_S)}(\tau_k) \in (\beta - \eta, \beta).$$

Intuitively, $\mathcal{O}$ **can't change abruptly**.

Stable Algorithms Fail for NPP: Proof Sketch, Graph Construction

(8) Construct $\mathbb{G} = (V, E)$: $V = \{1, 2, \ldots, T\}$.

- $(i, j) \in E$ iff $\exists k \in \{1, \ldots, Q\}$: $\mathcal{O}^{(ij)}(\tau_k) \in (\beta - \eta, \beta)$.
- Color $(i, j) \in E$ with first $t \in \{1, \ldots, Q\}$ s.t., $\mathcal{O}^{(ij)}(\tau_t) \in (\beta - \eta, \beta)$ for first time.

(9) **Independence number** of $\mathbb{G}$ is bounded: $\alpha(\mathbb{G}) \le m - 1$.

(10) Apply Ramsey Theory twice:

- Extract a large clique C_M of $\mathbb{G}$. Edges colored one of Q colors.
- Extract a monochromatic $m-$clique C_m from C_M.

(11) C_m contradicts with $m-$OGP.

(12) Track $\mathbb{P}$'s via $\cup$-bound: $\mathbb{P}(\exists \text{ monochromatic } C_m) > 0$.

A Concrete Execution of $m-$OGP result.

- Suppose we want to **rule out** exponent $E_n = n^{1-\delta}$, $\delta \in (0, 1/2)$.
- Set $g(n) = n^{\delta'}$ for some δ' with $\delta' + 2\delta < 1$. In fact, any $g(n)$ satisfying below works:

$$g(n) \in \omega(1) \quad \text{and} \quad g(n) \in o\left(\frac{E_n^2}{n \log n}\right).$$

- Then, $m-$OGP holds with (m_n, β_n, η_n), where

$$m_n = \frac{2n}{E_n} = 2n^{\delta}, \quad \beta_n = 1 - 2\frac{g(n)}{E_n} = 1 - 2n^{\delta'+\delta-1}, \quad \text{and} \quad \eta_n = \frac{g(n)}{2n} = \frac{1}{2}n^{-1+\delta'}.$$

- Note that $n\eta_n = \Theta(g(n)) = \omega(1)$, hence $(\beta_n - \eta_n, \beta_n)$ is non-vacuous.

The Rate $\omega(\sqrt{n \log n})$ is Tight: First Moment Method Fails Beyond

- We need $\beta = 1 - o_n(1)$: set $\beta = 1 - 2\nu_n$. For Σ^{-1} to exist, $\eta \lesssim \nu_n/m$.
- For $[\beta - \eta, \beta]$ to be **non-vacuous**, $n\eta = \Omega(1)$ (as $n \times \mathrm{Overlap} \in \mathbb{Z}$). Hence,

$$n\eta = \Omega(1) \implies n\nu_n/m = \Omega \implies n\nu_n = \Omega(m).$$

- Computing exponent of $\mathbb{E}[\cdot]$:
  - $\mathbb{P}$ term contributes $-mE_n$ via 2^{-E_n}.
  - $\log_2 \binom{n}{k} = (1 + o_n(1))k \log_2 \frac{n}{k}$ for $k = o(n)$. Hence, $\#$ term contributes

$$2^n \binom{n}{n\frac{1-\beta}{2}}^{m-1} \sim \exp_2\left(n - mn\nu_n \log \nu_n\right).$$

- Combining, the exponent is

$$n - mn\nu_n \log \nu_n - mE_n.$$

The Rate $\omega(\sqrt{n\log n})$ is Tight: First Moment Method Fails Beyond

- 1^{st} moment works only if $-\xi(n)=\omega_n(1)$, where $\xi(n)=n-mn\nu_n\log\nu_n-mE_n$.
  - $mE_n=\Omega(n)$. As $n\nu_n=\Omega(m)$, we get $n\nu_n=\Omega(n/E_n)$.
  - $mE_n=\Omega(mn\nu_n\log(1/\nu_n))$. That is, $E_n=\Omega(n\nu_n\log\frac{1}{\nu_n})$.
- Using $\log 1/\nu_n=\omega(1)$, we need

$$E_n=\omega\left(n\nu_n\right)=\omega\left(n/E_n\right)\implies E_n=\omega(\sqrt{n}).$$

- Slightly more delicate analysis yields extra $\sqrt{\log n}$ factor.

Derrida's REM Model

- $\texttt{NPP}$ is the first system for which **local REM conjecture** is established.
- **Derrida's REM Model.** A simple stochastic process: assign, to each $\sigma \in \mathcal{B}_n$, a random variable $X_\sigma = -\sqrt{n} Z_\sigma$ where Z_σ, $\sigma \in \mathcal{B}_n$, are i.i.d. standard normal.
- Perhaps the simplest model of *"random disorder"*.
- Back to $\texttt{NPP}$: for $\sigma \in \mathcal{B}_n$, denote $E(\sigma) \triangleq n^{-1/2}|\langle \sigma, X\rangle|$. Note that $E(\sigma) = E(-\sigma)$.
- For each pair $(\sigma, -\sigma)$; keep exactly one. Let $N \triangleq 2^{n-1}$, $E^{(1)} < \cdots < E^{(N)}$ be energies sorted; and $\sigma^{(i)}$ be the "spin configuration" with $E(\sigma^{(i)}) = E^{(i)}$.

Theorem

(Informal) If i and i' are nearby, then (a) $E^{(i)}$ and $E^{(i')}$ are uncorrelated; and (b) $\sigma^{(i)}$ and $\sigma^{(i')}$ are nearly orthogonal.

Namely, the system "locally" behaves like REM.

A Phase Transition in NPP: Integer-valued X_i

- Let X_i, $1 \le i \le n$, be i.i.d. uniform over $\{0, 1, \dots, A\}$ where $A = \lfloor 2^{n\kappa} \rfloor$.
- [GW96] argued the existence of a phase transition:
  - For $\kappa < \kappa_c$, there exists (exponentially many) *perfect partitions*: with discrepancy 0 or 1 depending on parity of $\sum_i X_i$.
  - For $\kappa > \kappa_c$, w.h.p. no such partitions exist.

  They predicted κ_c to be around 0.96.
- [Mer98] argued $\kappa_c = 1 + o_n(1)$.
- Rigorously confirmed by [BCP01].

Statistical-to-computational gaps: Largest Clique in Dense ER Graph

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

Largest clique/independent set problem.

- $\mathbb{G}(n, 1/2)$.
- Largest clique $\sim 2\log_2 n$, trivial *greedy* returns $\sim \log_2 n$.
- **Open problem** [Kar76]: Find a better polynomial-time algorithm.
- Open since...

Independent Sets in Random Sparse Graphs

- Both random $d-$regular graph and $\mathbb{G}(n, d/n)$ behave essentially the same.
- As $n \to \infty$, for $d > 0$,

$$\frac{1}{n}|\mathcal{I}_n| \to \alpha_d \quad \text{for some sequence } \alpha_d, \text{ where} \quad \alpha_d = 2(1 + o_d(1))\frac{\log d}{d} \quad \text{as} \quad d \to \infty.$$

- If there is a $\mathcal{A}$ returning, w.h.p., an independent set of size $(1 + c)(\log d/d)n$ (c can be $1/\sqrt{2}$ or ϵ), then by interpolation one can create "forbidden structures".
- Yields a contradiction with OGP.

OGP in Coloring of Sparse Random Graphs

Consider $\mathbb{G}(n, \frac{d}{n})$ or $\mathbb{G}_d(n)$. Recall $\alpha(\mathbb{G})\mathcal{X}(\mathbb{G}) \ge n$.

- [Fri90, FL92, BGT10]: $\alpha(\mathbb{G}) \simeq 2(1 + o_d(1))\frac{\log d}{d} n$.
- $\mathcal{X}^* \triangleq \mathcal{X}(\mathbb{G}) \simeq \frac{(1+o_d(1))d}{2\log d}$. Simple algorithm for $q \ge 2\mathcal{X}^*$.
- Space of $\{1, 2, \ldots, q\}^n$:
  - Connected large ball if $q \ge 2\mathcal{X}^*$.
  - Exponentially many isolated clusters large ball if $q \le (2 - \epsilon)\mathcal{X}^*$.

  [ACO08].
- Factor 2 Gap: Analogous to gap in large clique for dense random graphs.

OGP in Sparse Regression

$X \in \mathbb{R}^{n\times p}$, $\beta^* \in \mathbb{R}^{p\times 1}$, $W \in \mathbb{R}^n$ i.i.d. $\mathcal{N}(0,\sigma^2)$. Observe $Y = X\beta^* + W$.

Goal: Recover β^* from (Y,X). $\|\beta\|_0 \le k$.

- Convex optimization solves for $n > n_{\rm ALG} := \Omega(k\log p)$.
- Brute force works iff $n > n_{\rm INF} := \Omega(k\log p/\log(1+k/\sigma^2))$.

Again a **statistical-to-computational gap!**

For

$$n < cn_{\rm ALG}, \quad \text{where} \quad c>0 \quad \text{is sufficiently small}$$

OGP takes place.

$\Theta(\sqrt{n}2^{-n})$: A Heuristic Calculation

- Let $X = (X_i : 1 \le i \le n) \stackrel{d}{=} \mathcal{N}(0, I_n)$. Consider $a \in \{0,1\}^n$ and $S(a) = \langle a, X \rangle$.
- Due to **concentration of measure**, for **many** a, $S(a) = \Theta(\sqrt{n})$.
- Roughly 2^n such a. By **Pigeonhole**, there are (distinct) $a, a' \in \{0,1\}^n$ such that

$$|S(a) - S(a')| = O(\sqrt{n}2^{-n}).$$

- Set $\sigma := a - a' \in \{-1,0,1\}^n$. Then

$$\left|\langle \sigma, X \rangle\right| = O(\sqrt{n}2^{-n}).$$

OGP: NAE-k-SAT Problem

- n **Boolean variables** x_i, $1 \le i \le n$.
- Each **clause** $\mathcal{C}_i = x_{i_1} \vee \bar{x}_{i_2} \vee \cdots \vee x_{i_k}$ with k literals.
- $\mathcal{C}_i$, $1 \le i \le M$ with $M = dn$, d **density**.
- k-SAT: satisfy **all** $\mathcal{C}_i$. NAE-k-SAT: Satisfy a $\mathcal{C}_i$ **and** unsatisfy a $\mathcal{C}_j$
- **Information-Theoretic Threshold**: let $d_s := 2^{k-1}\ln 2 + O_K(1)$. Then,

$$\mathbb{P}[\exists (x_1,\dots,x_n) \text{ satisfying } \mathcal{C}_i, \forall i] = 1 \quad \text{for} \quad d < d_s \quad \text{and is} \quad = 0 \quad \text{for} \quad d > d_s.$$

  [AM06, COP12].
- **Computational Threshold**: Unit clause returns an $(x_i : i \in [n])$ if $d < d_s/k$.
- For $d > (d_s/k)\ln^2 k$, sequential local alg fail [GS17b]; and WALKSAT fails [COHH17].

Statistical-to-computational gaps: Planted Clique

Same story with **planted clique problem**...

- $\mathbb{G}(n, 1/2)$, plant a clique $\mathcal{PC}$ of size k.
- **Problem.** Observe graph, recover $\mathcal{PC}$.
- Impossible for $k < 2\log_2 n$. Possible in polynomial-time if $k = \Omega(\sqrt{n})$ [AKS98]
- **Hard regime.** No polynomial-time algorithm known for $2\log_2 n < k = o(\sqrt{n})$.

The (Infamous) Planted Clique Problem

- $\mathbb{G}(n, \frac{1}{2})$. Largest clique $\sim 2\log_2 n$.
- Select k vertices (u.a.r.). Deterministically "plant" all $\binom{k}{2}$ edges between them ($\mathcal{PC}$).
- Inference Problem. Recover $\mathcal{PC}$ from $\mathbb{G}$. *Various regimes on k:*
  - Information-theoretically impossible if $k < 2\log_2 n$.
  - *Brute-force* succeeds when $k \geq (2+\epsilon)\log_2 n$.
- What about polynomial-time algorithms?
  - Kučera [1995] A very simple algorithm for $k = \Omega\left(\sqrt{n\log_2 n}\right)$. Based on observation: when $k = \Omega\left(\sqrt{n\log_2 n}\right)$, $k-$largest degree vertices are w.h.p. vertices of $\mathcal{PC}$.
  - Alon, Krivelevich, and Sudakov [1998] A spectral algorithm for $k = \Omega(\sqrt{n})$.
- No polynomial-time algorithm when $k = o\left(\sqrt{n}\right)$. Again a gap.

Kucera's argument

Let $k \geq C\sqrt{n \log n}$ for some $C > 1$. We claim w.h.p. the k-nodes having the largest number of neighbours are those from the planted clique.
Let $I_i^{(j)}$, $1 \leq i \leq n$ and $1 \leq j \leq n$ be i.i.d. Bernoulli with $I_i^{(j)}$, $1 \leq i \leq n$, being the "status" of the neighbours of node j. It suffices to show

$$\mathbb{P}\left(\sum_i I_i^{(j)} \geq \frac{n}{2} + C\sqrt{n \log n}, 1 \leq j \leq n\right) = o_n(1).$$

Applying Bernoulli concentration,

$$\mathbb{P}\left(\sum_i \left|I_i^{(j)} - \frac{1}{2}\right| \geq C\sqrt{n \log n}\right) \leq \exp\left(-\frac{C^2 n \log n}{n}\right) = n^{-C^2}.$$

Taking a union bound over $1 \leq j \leq n$, it follows this probability is n^{-C^2+1}, which is $o_n(1)$ provided $C > 1$.

Planted Clique Conjecture

An instance of $\mathcal{PC}_D(n,k,p)$: Suppose $p \in (0,1)$,

$$H_0 \sim \mathbb{G}(n,p) \quad \text{and} \quad H_1 \sim \mathbb{G}(n,k,p).$$

Here, H_0 is the hypothesis that a graph is **Erdös-Rényi**; whereas H_1 is the hypothesis that the graph contains a **planted clique** of size k. Informally, one cannot recover the planted clique if $k \ll \sqrt{n}$. Formally,

Conjecture (Conjecture 2.1 in [BBH18])

Let $\{A_n\}$ be a sequence of (randomized) polynomial time algorithms $A_n : \mathcal{G}_n \to \{0,1\}$ and k_n be a sequence of positive integers with $\limsup_{n\to\infty} \log_n k_n < \frac{1}{2}$. Then if G is an instance of $PC_D(n,k,p)$, it holds that

$$\liminf_{n\to\infty}\Big(\mathbb{P}_{H_0}(A_n(G)=1)+\mathbb{P}_{H_1}(A_n(G)=0)\Big) \ge 1.$$

Namely, one cannot "beat" the random guessing.

Statistics Preserving Reductions: Reduction from Planted Clique

Problem 1 (Think of Planted Clique):

$$H_0 : X \sim P_X^0 \qquad \text{and} \qquad H_1 : X \sim P_X^1.$$

Problem 2 (Think of spiked Wigner):

$$H_0 : Y \sim P_Y^0 \qquad \text{and} \qquad H_1 : Y \sim P_Y^1.$$

Goal: Find a *kernel* $W_{Y|X}$ such that

$$d_{\mathrm{TV}}\Big(W_{Y|X}P_X, P_Y\Big) \to 0,$$

as $n \to \infty$, under both H_0 and H_1.

A complication: By DPI, one loses "information": recall many such problems have a signal parameter.

Low-Degree Methods

- **Hypothesis testing:**

$$H_0 : Y \sim \mathbb{Q} \qquad \text{and} \qquad H_1 : Y \sim \mathbb{P}.$$

  Planted clique: Graph Y. $\mathbb{Q} = \mathbb{G}(n, 1/2)$ and $\mathbb{P} = \mathbb{G}(n, k, 1/2)$.
- **Goal:** Distinguish H_0 and H_1 with error probability $o(1)$.
- **Likelihood ratio:**

$$L(Y) := \frac{d\mathbb{P}}{d\mathbb{Q}}(Y).$$

- Do with degree $\leq D$ polynomials.

$$\mathrm{Adv}_{\leq D} := \max_{f:\deg(f)\leq D} \frac{\mathbb{E}_{\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f(Y)^2]}}.$$

Low-Degree Methods

- Recall

$$\mathbb{E}_{\mathbb{P}}[f(Y)] = \mathbb{E}_{\mathbb{Q}}[L(Y)f(Y)].$$

- Inner product

$$\langle f, g \rangle := \mathbb{E}_{\mathbb{Q}}[f(Y)g(Y)].$$

- Then

$$\mathrm{Adv}_{\le D} = \max_{f:\deg(f)\le D} \langle L(Y), \widehat{f}(Y) \rangle, \quad \text{where} \quad \widehat{f}(Y) = f(Y)/\|f(Y)\|.$$

- Turns out

$$\mathrm{Adv}_{\le D} := \left\| L^{\le D} \right\|.$$

- Easily computable if Q is **product measure**: if $Q \sim \mathcal{N}(0, I_n)$ then take Hermite coefficients.

Low-Degree Methods

- Informally, if $\|L^{\leq D}\| = \omega(1)$ then "easy": degree $\leq D$ can distinguish.
- If $\|L^{\leq D}\| = O(1)$ then "hard": degree $\leq D$ fails to distinguish.
- If $\|L^{\leq D}\| = O(1)$ for D, then no algorithm with running time $n^{\widetilde{\Theta}(D)}$.
- $D = \log n$ proxy for polynomial-time algorithms:

$$\text{If} \quad \|L^{\leq D}\| = O(1) \quad \text{for some} \quad D = \omega(\log n) \quad \text{then no poly-time algorithm.}$$

- Intuition from spectral methods: If Y has largest eigenvalue λ_1, then

$$\mathrm{tr}(Y^k) \approx \lambda_1^k \qquad \text{for} \qquad k \approx O(\log n).$$

- Captures many known thresholds: $\mathcal{PC}$, sparse PCA, Kesten-Stigum threshold in SBM...

Low-Degree Methods

Case Study: $\mathcal{PC}$.

- If $k = \Omega(\sqrt{n})$ then $\|L^{\leq D}\| = \omega(1)$ for some $D \simeq O(\log n)$.
- If $k = O(n^{\frac{1}{2}-\epsilon})$ then $\|L^{\leq D}\| = O(1)$ for all $D \simeq O(\log n)$.

Even More Refined Thresholds:

- If smallest $\leq D$ with $\|L^{\leq D}\| = \omega(1)$ is n^δ $(\delta \in (0,1))$ then need $\exp(n^{\delta+o(1)})$ time.

Some advantages:

- Precise trade-off: D versus runtime.
- Easy to compute. Rigorous evidence for failure of **spectral methods**.
- Many alg. (power iteration, AMP,...) realized as low-degree polynomials.

Some drawbacks:

- Applicable **almost solely** to hypothesis testing.
- Need to know **orthogonal polynomials** in null $\mathbb{Q}$: e.g. when null is $\mathbb{G}_d(n)$.

Markov Chain Mixing: Main Definitions

For Q, R on Ω, define **total variation**

$$\|Q-R\|_{\mathrm{TV}} := \sup_{A\subset\Omega}\Big|Q(A)-R(A)\Big|.$$

$(X_t)_{t\ge 1}$ MC with states Ω, kernel P and stationary distribution π. Let

$$d(t) := \sup_{x\in\Omega}\Big\|P^t(x,\cdot)-\pi(\cdot)\Big\|_{\mathrm{TV}} = \sup_{\mu\in\mathcal{P}}\Big\|\mu P^t-\pi\Big\|_{\mathrm{TV}}.$$

$d(t)$ called **distance to stationarity**. Finally,

$$t_{\mathrm{mix}}(\epsilon) := \inf\Big\{t\ge 1 : d(t)\le\epsilon\Big\}.$$

Markov Chain Mixing: Interpretation of Our Result

$\mathcal{P}$: space of **probability measures** on Ω. Namely, $t_{\rm mix}(\epsilon)$ is **first** t s.t.

$$\left|\mu P^t(A) - \pi(A)\right| \le \epsilon$$

for **all initialization** $\mu \in \mathcal{P}$ and **all states** $A \subset \Omega$.

Our result: for $X_0 \sim \pi_\beta(\cdot|I_3 \cup \partial S)$, and $t < \tau_\beta$, $X_t \in I_3 \cup \partial S$.

- Let $\mu = \pi_\beta(\cdot|I_3 \cup \partial S)$ and $A = I_1 \cup \overline{I_3}$.
- $I_1 \cup \overline{I_3}$ and $I_3 \cup \partial S$ disjoint $\implies$ at $t = \tau_\beta - 1$, $\mu P^t(A) = 0$.
- $\pi(A) = \frac{1}{2}(1 + o_n(1))$ (part (a) of Thm). Hence,

$$t_{\rm mix}(A) \ge \tau_\beta \qquad \forall \epsilon < \frac{1}{2}.$$

.