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 \( X_1, \ldots, X_n \); partition them into two “bins” with total weights as close as possible:

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

- Equivalently

\[
\min_{\sigma \in B_n} \left| \langle \sigma, X \rangle \right|, \quad \text{where} \quad B_n = \{-1, 1\}^n \quad \text{and} \quad \langle \sigma, X \rangle = \sum_{1 \leq i \leq n} \sigma_i X_i.
\]

- **Our focus.** Items \( X_i \) are i.i.d. standard normal: \( X_i \overset{d}{=} \mathcal{N}(0, 1) \).
Randomized controlled trials. Gold standard for clinical trials [KAK19, HSSZ19].

$n$ persons with covariate info (age, weight, height,...) $X_i \in \mathbb{R}^d$, $1 \leq i \leq n$.

Split into two groups (treatment and control) with similar "features":

$$\min_{\sigma \in B_n} \|X\sigma\|_{\infty}, \quad \text{where} \quad X = (X_1, X_2, \ldots, X_n) \in \mathbb{R}^{d \times n}.$$

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.
\( X_i \in \mathbb{R}^d, 1 \leq i \leq n \). Define

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

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

**Average-case:** Assume \( X_i \overset{d}{=} \mathcal{N}(0, I_d), 1 \leq i \leq n, \text{i.i.d.} \). For \( 1 \leq d \leq o(n) \),

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

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

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

\[
\mathbb{E}[\mathcal{D}_n] = O\left(2^{-cn}\right).
\]
$X_i \overset{d}{=} \mathcal{N}(0, l_d)$, $1 \leq i \leq n$ i.i.d.

- **[KK82]**: For $d = 1$; returns $\sigma_{\text{ALG}} \in \mathcal{B}_n$ with

$$|\langle \sigma_{\text{ALG}}, X \rangle| = 2^{-\Theta(\log^2 n)}, \text{ w.h.p.}$$

- A simpler heuristic, **Largest Differencing Method (LDM)**. Also good performance **[Yak96]**:

$$\mathbb{E}[\text{LDM}] = n^{-\Theta(\log n)}.$$  

- **[TMR20]**: For $2 \leq d \leq O(\sqrt{\log n})$, returns a $\sigma_{\text{ALG}} \in \mathcal{B}_n$ with

$$\|X\sigma_{\text{ALG}}\|_{\infty} = \exp \left( -\Omega \left( \frac{\log^2 n}{d} \right) \right), \text{ w.h.p.}$$
NPP: A *Statistical-to-Computational Gap*

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

- For $X \sim \mathcal{N}(0, I_n)$,
  \[
  \min_{\sigma \in \mathcal{B}_n} |\langle \sigma, X \rangle| = \Theta(\sqrt{n}2^{-n}) \quad \text{vs} \quad |\langle \sigma_{\text{ALG}}, X \rangle| = 2^{-\Theta(\log^2 n)}.
  \]

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

**Source of this gap/hardness?**
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 $P \neq NP$).
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 $\xi$:

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

(Informally) OGP for energy $\mathcal{E}$ if $\exists 0 < \nu_1 < \nu_2$ s.t. $\forall \sigma_1, \sigma_2 \in \Theta$, $\mathcal{L}(\sigma_j, \xi) \leq \mathcal{E} \implies \text{distance}(\sigma_1, \sigma_2) < \nu_1$ or $\text{distance}(\sigma_1, \sigma_2) > \nu_2$.

Any two near optimal $\sigma_1, \sigma_2$ are either too similar or too dissimilar.

**distance($\cdot, \cdot$)**

For $\Theta = B_n = \{-1, 1\}^n$, normalized overlap:

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

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

\[ L(\sigma, \xi) \]

$\nu_1 < \nu_2$

$\min_{\sigma} L(\sigma, \xi)$
Clustering in \( k\)-SAT: Solution space consists of disconnected clusters [MMZ05, ACO08, ACORT11].

First algorithmic implication: Max independent set in random \( d\)-regular graph \( G_d(n) \). [GS17a].

OGP: Any large \( I_1, I_2 \) either have significant intersection, or no intersection at all.

Local algorithms fail to return a large \( I \).
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

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
Recall $\min_{\sigma \in B_n} |\langle \sigma, X \rangle|$, $X \overset{d}{=} \mathcal{N}(0, I_n)$, 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 $\sigma, \sigma' \in B_n$ achieve

$$|\langle \sigma, X \rangle| = O(\sqrt{n2^{-\epsilon n}}) \quad \text{and} \quad |\langle \sigma', X \rangle| = O(\sqrt{n2^{-\epsilon n}})$$

then either $\sigma = \sigma'$ or $n^{-1}|\langle \sigma, \sigma' \rangle| \leq \rho$ 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).
Let $N$ count the $\#$ of such $(\sigma, \sigma')$: $\Pr(N \geq 1) \leq \mathbb{E}[N]$.  

Number of $\sigma, \sigma'$ with $n^{-1}|\langle \sigma, \sigma' \rangle| \geq \rho$ is $2^{n + nh((1 - \rho)/2)}$, where $h(\cdot)$ is binary entropy.  

$\sigma, \sigma'$ with $O(\sigma, \sigma') = \rho$. Let $Y = n^{-\frac{1}{2}}\langle \sigma, X \rangle$ and $Y' = n^{-\frac{1}{2}}\langle \sigma', X \rangle$. Then, 

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

Hence 

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

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 $X_0, \ldots, X_m \overset{d}{=} \mathcal{N}(0, I_n)$ i.i.d.; and interpolate

$$Y_i(\tau) = \sqrt{1 - \tau^2} X_0 + \tau X_i \overset{d}{=} \mathcal{N}(0, I_n), \quad \tau \in [0, 1], \quad 1 \leq i \leq m.$$
$m$-tuples $\sigma_i \in B_n$ ($m$–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 $G_d(n)$.
  
  - **Computational threshold** ($\log d/d)n$, 2-OGP rules out $|I| \geq (1 + 1/\sqrt{2})(\log d/d)n$.
  
  - [RV17]: Study instead $m$-tuples $I_i$, $1 \leq i \leq m$: 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].

---

**Ensemble-Multi-OGP for NPP**
Our Contributions: Ensemble $m$–OGP for NPP.

Theorem (Ensemble-multi-OGP)

*(Informally)* Ensemble $m$–OGP holds below any $2^{-\epsilon n}$, $\epsilon > 0$.

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

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

then w.h.p. $\exists 1 \leq i < j \leq m$ 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, $\forall \omega(1) \leq f(n) \leq o(n)$, $\forall \beta, \eta \in (0, 1)$, and $\forall m \in \mathbb{N}$; w.h.p. $\exists \sigma_i$, $1 \leq i \leq m$ such that

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

- 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,

\[
P(M \geq 1) \geq \frac{\mathbb{E}[M]^2}{\mathbb{E}[M^2]}.
\]

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

E. C. Kızıldağ (MIT)

OGP in the NPP
Our Contributions: Ensemble $m$–OGP for 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, $\forall \omega(\sqrt{n\log n}) \leq E_n \leq o(n)$, $\forall \mathcal{I} \subset [0, 1]$ with $|\mathcal{I}| = n^{O(1)}$, $\exists m_n \in \mathbb{N}$, $\exists 1 > \beta_n > \eta_n > 0$ s.t. if

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

then w.h.p. $\exists 1 \leq i < j \leq m_n$ 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

Introduction
- Problem Definition
- Applications
- Statistical-to-Computational Gaps
- The Overlap Gap Property (OGP)

Contributions: Properties of the Landscape of NPP
- 2-OGP
- Ensemble-\(m\)-OGP with \(m = O(1)\)
- Ensemble-\(m\)-OGP with \(m = \omega(1)\)

Contributions: Algorithmic Hardness Results
- Failure of Stable Algorithms
- Failure of MCMC Methods

Conclusion and Future Research
- Summary of Contributions
- Future Work
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 $\Rightarrow$ FEW $\Rightarrow$ 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**: 
\[ \mathbb{P}\left(n^{-\frac{1}{2}}|\langle X, \mathcal{A}(X) \rangle| \leq E\right) \geq 1 - p_f. \]

(b) **Stability**: $\exists \rho \in (0, 1]$, $X, Y \overset{d}{=} \mathcal{N}(0, I_n)$ with $\text{Cov}(X, Y) = \rho I_n$;
\[ \mathbb{P}\left(d_H(\mathcal{A}(X), \mathcal{A}(Y)) \leq f + L\|X - Y\|_2^2\right) \geq 1 - p_{st}. \]
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 NPP)

Stable algorithms can’t achieve value better than

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

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

- For extreme case, $E_n = \Theta(n)$: rule out $p_f, p_{\text{st}} = O(1)$.
- **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.
Let $X \overset{d}{=} \mathcal{N}(0, I_n)$; and define Hamiltonian $H(\sigma) \triangleq n^{-\frac{1}{2}}|\langle \sigma, X \rangle|$.  

Define Gibbs distribution at inverse temperature $\beta > 0$ on $B_n$:

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

Fact: As $\beta \to \infty$, $\pi_\beta$ concentrates on

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

Construct $G = (V, E)$ with $V = B_n$ and $(\sigma, \sigma') \in E \iff d_H(\sigma, \sigma') = 1$.

Consider any nearest neighbor MC $(X_t)_{t \geq 0}$ on $G$ reversible w.r.t. $\pi_\beta$. 

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
  $$I_1 = \left\{ \sigma : -\rho \leq \frac{1}{n} \langle \sigma, \sigma^* \rangle \leq \rho \right\}, \quad I_2 = \left\{ \sigma : \rho \leq \frac{1}{n} \langle \sigma, \sigma^* \rangle \leq \frac{n-2}{n} \right\}, \quad \text{and} \quad I_3 = \{\sigma^*\}.$$
- 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. $X \overset{d}{=} \mathcal{N}(0, I_n)$),

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

- $I_2$ 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.

![Diagram showing the free energy well with $I_1$, $I_2$, and $I_3$]
Recall $H(\sigma^*) = H(-\sigma^*) = \Theta(2^{-n})$. Absorbing constants into $\beta > 0$,

$$\pi_\beta(I_3) = \pi_\beta(\overline{I}_3) = \exp(-\beta 2^{-n}) / Z_\beta.$$  

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,

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

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,

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

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

**E. C. Kızıldağ (MIT)**
From OGP to MCMC

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

OGP $\implies$ FEW.

∴ 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 $X_0 \overset{d}{=} \pi_\beta(\cdot \mid I_3 \cup \partial S)$. Define escape time

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

Theorem (Slow Mixing)

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

- $\pi_\beta(I_1 \cup I_3) \geq (1 + o_n(1))/2.$
- $\tau_\beta = e^{\Theta(n)}.$
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
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.
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 $\beta$).
  - For different initialization, e.g. uniform case.
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 VI


References VII


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 $X_i \sim \text{Unif}[0, 1]$. Each operation reduce size by $1/n$. Recurse $\sim \log n$ rounds: $n^{-\log n}$. 

E. C. Kızıldağ (MIT)
Algorithm $A : \mathbb{R}^n \times \Omega \rightarrow B_n$. $(\Omega, P_\omega)$ coin flips of $A$.

- $X \overset{d}{=} \mathcal{N}(0, I_n)$. Success guarantee w.r.t. $\mathcal{N}(0, I_n) \otimes P_\omega$:
  \[
  \mathbb{P}(X, \omega) \sim \mathcal{N}(0, I_n) \otimes P_\omega \left( n^{-\frac{1}{2}} | \langle X, A(X, \omega) \rangle | \leq E \right) \geq 1 - p_f.
  \]

- Need two $X, Y \overset{d}{=} \mathcal{N}(0, I_n)$ to talk about stability. To specify $P_{X, Y}$, need Cov:
  $\text{Cov}(X, Y) = \rho I$. Then, with respect to $(X, Y, \omega) \sim P_{X, Y} \otimes P_\omega$,
  \[
  \mathbb{P}(X, Y, \omega) : X \sim \rho Y, \omega \sim P_\omega \left( d_H(A(X, \omega), A(Y, \omega)) \leq f + L \| X - Y \|_2^2 \right) \geq 1 - p_{st}.
  \]
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:

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

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

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

Any \( c' \) greater than this value (and less than 1) works.

- For \( \epsilon = 1/5 \) (\( E_n = \Theta(n) \)), \( c' \to 0 \):

\[
p_f, p_{st} = O(1) \quad \text{suffice.}
\]
Stable Algorithms Fail for NPP: Proof Sketch

- Fix $E_n$. $m$–OGP holds with $(m, \beta, \eta)$: $[\beta - \eta, \beta]$ is the forbidden region.

- Discretization $Q$, required for $\eta$. $T$ “replicas”.

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

Proof by contradiction: Suppose randomized $A$ exists, reduce to deterministic $A$.

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

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

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

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

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

E. C. Kızıldağ (MIT)
Stable Algorithms Fail for NPP: Proof Sketch

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

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

(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. $O^{(i_S, j_S)}(\cdot)$ eventually below $\beta - \eta$.

(7) Stability of $O(\cdot) \implies$

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

Intuitively, $O$ can’t change abruptly.
Stable Algorithms Fail for NPP: Proof Sketch, Graph Construction

(8) Construct $G = (V, E)$: $V = \{1, 2, \ldots, T\}$.
- $(i, j) \in E$ iff $\exists k \in \{1, \ldots, Q\}$: $O^{(ij)}(\tau_k) \in (\beta - \eta, \beta)$.
- Color $(i, j) \in E$ with first $t \in \{1, \ldots, Q\}$ s.t., $O^{(ij)}(\tau_t) \in (\beta - \eta, \beta)$ for first time.

(9) Independence number of $G$ is bounded: $\alpha(G) \leq m - 1$.

(10) Apply Ramsey Theory twice:
- Extract a large clique $C_M$ of $G$. 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 $P$’s via $U$-bound: $P(\exists$ 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 $\delta'$ 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 $(m_n, \beta_n, \eta_n)$, 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 $m_n \eta_n = \Theta(g(n)) = \omega(1)$, 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 $\Sigma^{-1}$ to exist, $\eta \lesssim \nu_n/m$.
- For $[\beta - \eta, \beta]$ to be non-vacuous, $m\eta = \Omega(1)$ (as $n \times \text{Overlap} \in \mathbb{Z}$). Hence,
  
  \[ m\eta = \Omega(1) \implies n\nu_n/m = \Omega \implies n\nu_n = \Omega(m). \]

- Computing exponent of $\mathbb{E}[:]$:
  - $\mathbb{P}$ 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
    
    \[ 2^n \left( \frac{n}{n^{1-\beta}} \right)^{m-1} \sim \exp_2 (n - m\nu_n \log \nu_n). \]

- Combining, the exponent is
  
  \[ n - m\nu_n \log \nu_n - mE_n. \]
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, $E_n = \Omega(n\nu_n \log(1/\nu_n))$.

Using $\log 1/\nu_n = \omega(1)$, we need

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

- Slightly more delicate analysis yields extra $\sqrt{\log n}$ factor.
Derrida’s REM Model

- 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 $Z_\sigma$, $\sigma \in \mathcal{B}_n$, are i.i.d. standard normal.
- Perhaps the simplest model of “random disorder”.
- Back to 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.
Let \( X_i, 1 \leq i \leq n \), be i.i.d. uniform over \( \{0, 1, \ldots, A\} \) 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 \( \kappa_c \) to be around 0.96.

[Mer98] argued \( \kappa_c = 1 + o_n(1) \).

Rigorously confirmed by [BCP01].
Common feature in many algorithmic problems in high-dimensional statistics & random combinatorial structures.

**Largest clique/independent set problem.**

- $\mathcal{G}(n, 1/2)$.
- 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 $d > 0$,
  \[
  \frac{1}{n} |I_n| \to \alpha_d \quad \text{for some sequence } \alpha_d, \text{ where } \alpha_d = 2(1 + o_d(1)) \frac{\log d}{d} \quad \text{as } d \to \infty.
  \]
- 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 $\epsilon$), then by interpolation one can create “forbidden structures”.
- Yields a contradiction with OGP.
Consider $\mathbb{G}(n, \frac{d}{n})$ or $\mathbb{G}_d(n)$. Recall $\alpha(\mathbb{G}) \chi(\mathbb{G}) \geq n$.

- [Fri90, FL92, BGT10]: $\alpha(\mathbb{G}) \simeq 2(1 + o_d(1))^{\frac{\log d}{d}} n$.

- $\chi^* \triangleq \chi(\mathbb{G}) \simeq \frac{(1 + o_d(1))d}{2 \log d}$. Simple algorithm for $q \geq 2\chi^*$.

- Space of $\{1, 2, \ldots, q\}^n$:
  - Connected large ball if $q \geq 2\chi^*$.
  - Exponentially many isolated clusters large ball if $q \leq (2 - \epsilon)\chi^*$.

[ACO08].

- Factor 2 Gap: Analogous to gap in large clique for dense random graphs.
\[ X \in \mathbb{R}^{n \times p}, \ \beta^* \in \mathbb{R}^{p \times 1}, \ W \in \mathbb{R}^n \text{ i.i.d. } \mathcal{N}(0, \sigma^2). \] Observe \( Y = X\beta^* + W. \)

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

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

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

For

\[ n < cn_{\text{ALG}}, \text{ where } c > 0 \text{ is sufficiently small} \]

**OGP** takes place.
$\Theta(\sqrt{n}2^{-n})$: A Heuristic Calculation

- Let $X = (X_i : 1 \leq i \leq n) \overset{d}{=} \mathcal{N}(0, I_n)$. 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
  $$|\langle \sigma, X \rangle| = O(\sqrt{n}2^{-n}).$$
OGP: NAE-k-SAT Problem

- **n Boolean variables** $x_i$, $1 \leq i \leq n$.
- Each **clause** $C_i = x_{i_1} \lor \overline{x}_{i_2} \lor \cdots \lor x_{i_k}$ with $k$ literals.
- $C_i$, $1 \leq i \leq M$ with $M = dn$, $d$ **Density**.
- **k-SAT**: satisfy all $C_i$. **NAE-k-SAT**: Satisfy a $C_i$ and unsatisfy a $C_j$
- **Information-Theoretic Threshold**: let $d_s := 2^{k-1} \ln 2 + O_K(1)$. Then,
  $$\mathbb{P}[\exists (x_1, \ldots, x_n) \text{ satisfying } C_i, \forall i] = 1 \quad \text{for} \quad d < d_s \quad \text{and is} \quad = 0 \quad \text{for} \quad d > d_s.$$ 
  [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

- $G(n, \frac{1}{2})$. 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 $G$. 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 $k = \Omega(\sqrt{n \log_2 n})$. Based on observation: when $k = \Omega(\sqrt{n \log_2 n})$, $k$–largest degree vertices are w.h.p. vertices of $\mathcal{PC}$.
- No polynomial-time algorithm when $k = o(\sqrt{n})$. 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 $\mathcal{P}_{CD}(n, k, p)$: 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 Erdős-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 $A_n : G_n \rightarrow \{0, 1\}$ and $k_n$ be a sequence of positive integers with $\limsup_{n \rightarrow \infty} \log_n k_n < \frac{1}{2}$. Then if $G$ is an instance of $\mathcal{P}_{CD}(n, k, p)$, it holds that

$$\liminf_{n \rightarrow \infty} \left( \mathbb{P}_{H_0}(A_n(G) = 1) + \mathbb{P}_{H_1}(A_n(G) = 0) \right) \geq 1.$$  

Namely, one cannot “beat” the random guessing.
Problem 1 (Think of Planted Clique):

\[ H_0 : X \sim P_X^0 \quad \text{and} \quad H_1 : X \sim P_X^1. \]

Problem 2 (Think of spiked Wigner):

\[ H_0 : Y \sim P_Y^0 \quad \text{and} \quad H_1 : Y \sim P_Y^1. \]

**Goal:** Find a kernel \( W_{Y|X} \) such that

\[
d_{TV}(W_{Y|X} P_X, P_Y) \to 0,
\]

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:**
  \[ H_0 : Y \sim \mathcal{Q} \quad \text{and} \quad H_1 : Y \sim \mathcal{P}. \]

  **Planted clique:** Graph \( Y \). \( \mathcal{Q} = \mathcal{G}(n, 1/2) \) and \( \mathcal{P} = \mathcal{G}(n, k, 1/2) \).

- **Goal:** Distinguish \( H_0 \) and \( H_1 \) with error probability \( o(1) \).

- **Likelihood ratio:**
  \[ L(Y) := \frac{d\mathcal{P}}{d\mathcal{Q}}(Y). \]

- Do with degree \( \leq D \) polynomials.

\[ \text{Adv}_{\leq D} := \max_{f : \deg(f) \leq D} \frac{\mathbb{E}_P[f(Y)]}{\sqrt{\mathbb{E}_Q[f(Y)^2]}}. \]
Recall

\[ \mathbb{E}_P[f(Y)] = \mathbb{E}_Q[L(Y)f(Y)]. \]

Inner product

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

Then

\[ \text{Adv}_{\leq D} = \max_{f: \deg(f) \leq D} \langle L(Y), \hat{f}(Y) \rangle, \quad \text{where} \quad \hat{f}(Y) = f(Y)/\|f(Y)\|. \]

Turns out

\[ \text{Adv}_{\leq D} := \|L^{\leq D}\|. \]

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 $n^{\tilde{\Theta}(D)}$.
- $D = \log n$ proxy for polynomial-time algorithms:
  
  If $\|L^{\leq D}\| = O(1)$ for some $D = \omega(\log n)$ then no poly-time algorithm.

- Intuition from spectral methods: If $Y$ has largest eigenvalue $\lambda_1$, then
  
  $\operatorname{tr}(Y^k) \approx \lambda_1^k$ for $k \approx O(\log n)$.

- Captures many known thresholds: $\mathcal{PC}$, sparse PCA, Kesten-Stigum threshold in SBM...
Case Study: \( \mathcal{PC} \).
- If \( k = \Omega(\sqrt{n}) \) then \( \|L^{\leq D}\| = \omega(1) \) for some \( D \approx O(\log n) \).
- If \( k = O(n^{\alpha - \epsilon}) \) then \( \|L^{\leq D}\| = O(1) \) for all \( D \approx O(\log n) \).

Even More Refined Thresholds:
- If smallest \( \leq D \) with \( \|L^{\leq D}\| = \omega(1) \) is \( n^\delta \) \( (\delta \in (0, 1)) \) 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 $\Omega$, define **total variation**

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

$(X_t)_{t \geq 1}$ MC with states $\Omega$, kernel $P$ and stationary distribution $\pi$. Let

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

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

$$t_{\text{mix}}(\epsilon) := \inf \{ t \geq 1 : d(t) \leq \epsilon \}.$$
Markov Chain Mixing: Interpretation of Our Result

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

\[
|\mu P^t(A) - \pi(A)| \leq \epsilon
\]

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,

\[
t_{\text{mix}}(A) \geq \tau_\beta \quad \forall \epsilon < \frac{1}{2}.
\]