The Overlap Gap Property in Inference: A Short Survey.

Ilias Zadik (MIT)

Simons workshop

September 13, 2021

Introduction

Computational gaps/trade-offs appear frequently in random environments.

Introduction

Computational gaps/trade-offs appear frequently in random environments.

Two (rough) main categories for search problems.

Introduction

Computational gaps/trade-offs appear frequently in random environments.

Two (rough) main categories for search problems.

(1) **Optimization**
$\mathsf{OPT} = \min_{\beta \in \Sigma} \mathsf{H}(\beta)$.

Introduction

Computational gaps/trade-offs appear frequently in random environments.

Two (rough) main categories for search problems.

(1) **Optimization**
$\mathrm{OPT} = \min_{\beta \in \Sigma} \mathrm{H}(\beta)$.
e.g. spin glasses, k-SAT, max independent set in random graphs.

Introduction

Computational gaps/trade-offs appear frequently in random environments.

Two (rough) main categories for search problems.

(1) **Optimization**
$\mathrm{OPT} = \min_{\beta \in \Sigma} \mathrm{H}(\beta)$.
e.g. spin glasses, k-SAT, max independent set in random graphs.
Gap: efficient methods achieve $\geq (1+\alpha) \times \mathrm{OPT}, \alpha > 0$.

Introduction

Computational gaps/trade-offs appear frequently in random environments.

Two (rough) main categories for search problems.

(1) **Optimization**
$\mathrm{OPT} = \min_{\beta \in \Sigma} \mathrm{H}(\beta)$.
e.g. spin glasses, k-SAT, max independent set in random graphs.
Gap: efficient methods achieve $\geq (1+\alpha) \times \mathrm{OPT}, \alpha > 0$.

(2) **Inference/Estimation**
Observe $\mathcal{D} \sim \mathrm{P}(\cdot|\beta^*)$, infer β^* (signal-to-noise ratio (SNR) λ).

Introduction

Computational gaps/trade-offs appear frequently in random environments.

Two (rough) main categories for search problems.

(1) **Optimization**
$\mathrm{OPT} = \min_{\beta \in \Sigma} \mathrm{H}(\beta)$.
e.g. spin glasses, k-SAT, max independent set in random graphs.
Gap: efficient methods achieve $\geq (1+\alpha) \times \mathrm{OPT}, \alpha > 0$.

(2) **Inference/Estimation**
Observe $\mathcal{D} \sim \mathrm{P}(\cdot|\beta^*)$, infer β^* (signal-to-noise ratio (SNR) λ).
e.g. planted clique, sparse regression, PCA.

Introduction

Computational gaps/trade-offs appear frequently in random environments.

Two (rough) main categories for search problems.

(1) **Optimization**
$\mathrm{OPT} = \min_{\beta \in \Sigma} \mathrm{H}(\beta)$.
e.g. spin glasses, k-SAT, max independent set in random graphs.
Gap: efficient methods achieve $\geq (1+\alpha) \times \mathrm{OPT}, \alpha > 0$.

(2) **Inference/Estimation**
Observe $\mathcal{D} \sim \mathrm{P}(\cdot|\beta^*)$, infer β^* (signal-to-noise ratio (SNR) λ).
e.g. planted clique, sparse regression, PCA.
Gap: Info-theory SNR: λ_1, efficient algorithms need SNR $\lambda_2 > \lambda_1$.

Introduction

Computational gaps/trade-offs appear frequently in random environments.

Two (rough) main categories for search problems.

(1) **Optimization**
$\mathrm{OPT} = \min_{\beta \in \Sigma} \mathrm{H}(\beta)$.
e.g. spin glasses, k-SAT, max independent set in random graphs.
Gap: efficient methods achieve $\geq (1+\alpha) \times \mathrm{OPT}, \alpha > 0$.

(2) **Inference/Estimation**
Observe $\mathcal{D} \sim \mathrm{P}(\cdot|\beta^*)$, infer β^* (signal-to-noise ratio (SNR) λ).
e.g. planted clique, sparse regression, PCA.
Gap: Info-theory SNR: λ_1, efficient algorithms need SNR $\lambda_2 > \lambda_1$.

Can **geometrical phase transitions** explain these gaps?

OGP for random optimization

$\mathsf{OPT} = \min_{\beta \in \Sigma} \mathsf{H}(\beta)$, $\Sigma \subseteq \mathsf{S}^{p-1}$.
Gap: efficient methods achieve $\geq (1+\alpha) \times \mathsf{OPT}, \alpha > 0$.

OGP for random optimization

$\mathrm{OPT} = \min_{\beta \in \Sigma} H(\beta)$, $\Sigma \subseteq S^{p-1}$.
Gap: efficient methods achieve $\geq (1+\alpha) \times \mathrm{OPT}, \alpha > 0$.

Between easy and hard regime **"an abrupt change in the geometry of the space of (near-optimal) solutions"** [Achlioptas, Coga-Oghlan '08].

OGP for random optimization

$\mathrm{OPT} = \min_{\beta \in \Sigma} \mathrm{H}(\beta)$, $\Sigma \subseteq \mathrm{S}^{\mathrm{p}-1}$.
Gap: efficient methods achieve $\geq (1+\alpha) \times \mathrm{OPT}, \alpha > 0$.

Between easy and hard regime **"an abrupt change in the geometry of the space of (near-optimal) solutions"** [Achlioptas, Coga-Oghlan '08].

Shattering, Condensation, Frozen Variables,
Replica Symmetry Breaking, *Overlap Gap Property (OGP)*

OGP for random optimization

$\mathsf{OPT} = \min_{\beta \in \Sigma} \mathsf{H}(\beta)$, $\Sigma \subseteq \mathsf{S}^{\mathsf{p}-1}$.
Gap: efficient methods achieve $\geq (1+\alpha) \times \mathsf{OPT}, \alpha > 0$.

Between easy and hard regime **"an abrupt change in the geometry of the space of (near-optimal) solutions"** [Achlioptas, Coga-Oghlan '08].

Shattering, Condensation, Frozen Variables,
Replica Symmetry Breaking, *Overlap Gap Property (OGP)*

OGP [Gamarnik, Sudan '14]

Set $\mathcal{T}_\epsilon = \{\langle \beta, \beta' \rangle : \mathsf{H}(\beta), \mathsf{H}(\beta') \leq (1+\epsilon) \min_{\beta \in \Sigma} \mathsf{H}(\beta)\} \subseteq \mathbb{R}$.

Algorithmically easy *if and only if* $\mathcal{T}_\epsilon$ is an "interval".

OGP for random optimization

$\mathrm{OPT} = \min_{\beta \in \Sigma} \mathrm{H}(\beta)$, $\Sigma \subseteq \mathrm{S}^{\mathrm{p}-1}$.
Gap: efficient methods achieve $\geq (1+\alpha) \times \mathrm{OPT}, \alpha > 0$.

Between easy and hard regime **"an abrupt change in the geometry of the space of (near-optimal) solutions"** [Achlioptas, Coga-Oghlan '08].

Shattering, Condensation, Frozen Variables,
Replica Symmetry Breaking, *Overlap Gap Property (OGP)*

OGP [Gamarnik, Sudan '14]

Set $\mathcal{T}_\epsilon = \{\langle \beta, \beta' \rangle : \mathrm{H}(\beta), \mathrm{H}(\beta') \leq (1+\epsilon) \min_{\beta \in \Sigma} \mathrm{H}(\beta)\} \subseteq \mathbb{R}$.

Algorithmically easy *if and only if* $\mathcal{T}_\epsilon$ is an "interval".

- (Gamarnik, Jagannath, Wein '20) MIS, (Bresler, Huang '21) k-SAT.
- OGP implies failure of stable (low-degree) algorithms.

This talk: a short survey

Overlap Gap Property (OGP) for **inference**. [Gamarnik, **Z** '17]

(1) *Definition* and *overview*

(2) Two case-studies:

- *sparse regression*
- *planted clique* (more involved)

The OGP for Inference

Data: $\mathcal{D} \sim \mathbb{P}_{\beta^*}$, $\beta^* \in \Theta \subseteq \mathsf{S}^{\mathsf{p}-1}$.

The OGP for Inference

Data: $\mathcal{D} \sim \mathbb{P}_{\beta^*}$, $\beta^* \in \Theta \subseteq \mathsf{S}^{p-1}$.
For some "informative loss function" $\mathcal{L}$, $\min_{\beta \in \Theta} \mathcal{L}\left(\mathcal{D}, \beta | \beta^*\right)$.

The OGP for Inference

Data: $\mathcal{D} \sim \mathbb{P}_{\beta^*}$, $\beta^* \in \Theta \subseteq \mathsf{S}^{p-1}$.
For some "informative loss function" $\mathcal{L}$, $\min_{\beta \in \Theta} \mathcal{L}(\mathcal{D}, \beta | \beta^*)$.

Belief: A canonical loss function's (e.g. likelihood's) landscape captures the inference hardness.

The OGP for Inference

Data: $\mathcal{D} \sim \mathbb{P}_{\beta^*}$, $\beta^* \in \Theta \subseteq \mathsf{S}^{p-1}$.
For some "informative loss function" $\mathcal{L}$, $\min_{\beta \in \Theta} \mathcal{L}(\mathcal{D}, \beta | \beta^*)$.

Belief: A canonical loss function's (e.g. likelihood's) landscape captures the inference hardness.

OGP for inference [Gamarnik, **Z** '17]

$\mathcal{T}_r = \{\langle \beta, \beta^* \rangle : \mathcal{L}(\mathcal{D}, \beta|\beta^*) \leq \min_\beta \mathcal{L}(\mathcal{D}, \beta|\beta^*) + r\}$ interval.
Easy *if and only if* $\mathcal{T}_r$ is an "interval" *for all* r.

The OGP for Inference

Data: $\mathcal{D} \sim \mathbb{P}_{\beta^*}$, $\beta^* \in \Theta \subseteq \mathrm{S}^{\mathrm{p}-1}$.
For some "informative loss function" $\mathcal{L}$, $\min_{\beta \in \Theta} \mathcal{L}(\mathcal{D}, \beta|\beta^*)$.

Belief: A canonical loss function's (e.g. likelihood's) landscape captures the inference hardness.

OGP for inference [Gamarnik, **Z** '17]

$\mathcal{T}_{\mathrm{r}} = \{\langle \beta, \beta^* \rangle : \mathcal{L}(\mathcal{D}, \beta|\beta^*) \leq \min_{\beta} \mathcal{L}(\mathcal{D}, \beta|\beta^*) + \mathrm{r}\}$ interval.
Easy *if and only if* $\mathcal{T}_{\mathrm{r}}$ is an "interval" *for all* r.

- *Sparse Linear Regression* [Gamarnik, **Z** '17a, '17b]
- *Planted Clique* [Gamarnik, **Z** '19].
- *Sparse PCA* [Gamarnik, Jagannath, Sen'19], [Ben Arous, Wein,**Z**'20]
- *Tensor PCA* [Ben Arous, Gheissari, Jagannath '18]
- *Group Testing* [Iliopoulos, **Z** '21]

The OGP for Inference — part 2

OGP for inference [Gamarnik, **Z** '17]

$\mathcal{T}_{\mathrm{r}} = \{\langle \beta, \beta^* \rangle : \mathcal{L}(\mathcal{D}, \beta | \beta^*) \leq \min_\beta \mathcal{L}(\mathcal{D}, \beta | \beta^*) + \mathrm{r}\}$ interval.
Easy *if and only if* $\mathcal{T}_{\mathrm{r}}$ is an "interval" *for all* r.

The OGP for Inference — part 2

OGP for inference [Gamarnik, Z '17]

$\mathcal{T}_r = \{\langle \beta, \beta^* \rangle : \mathcal{L}(\mathcal{D}, \beta|\beta^*) \leq \min_\beta \mathcal{L}(\mathcal{D}, \beta|\beta^*) + r\}$ interval.
Easy *if and only if* $\mathcal{T}_r$ is an "interval" *for all* r.

$\Gamma(\zeta) = \min_{\beta \in \Theta, \langle \beta, \beta^* \rangle = \zeta} \mathcal{L}(\mathcal{D}, \beta|\beta^*), \zeta \in [0, 1].$

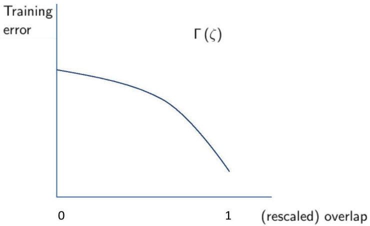
The OGP for Inference — part 2

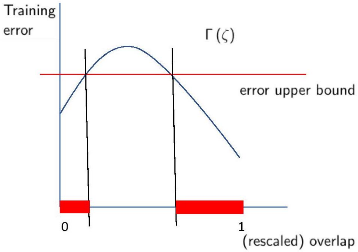
OGP for inference [Gamarnik, **Z** '17]

$\mathcal{T}_r = \{\langle \beta, \beta^* \rangle : \mathcal{L}(\mathcal{D}, \beta | \beta^*) \leq \min_\beta \mathcal{L}(\mathcal{D}, \beta|\beta^*) + r\}$ interval.
Easy *if and only if* $\mathcal{T}_r$ is an "interval" *for all* r.

$\Gamma(\zeta) = \min_{\beta \in \Theta, \langle \beta, \beta^* \rangle = \zeta} \mathcal{L}(\mathcal{D}, \beta|\beta^*), \zeta \in [0, 1]$.

Lemma: OGP if and only if Γ is non-monotonic.





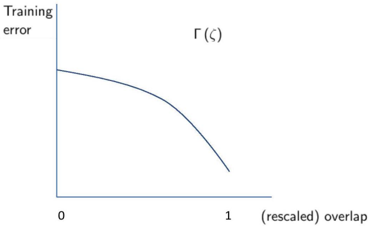
The OGP for Inference — part 2

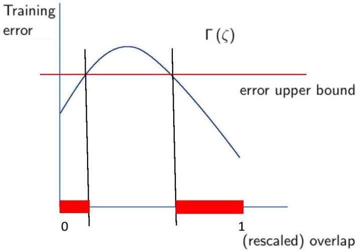
OGP for inference [Gamarnik, **Z** '17]

$\mathcal{T}_r = \{\langle \beta, \beta^* \rangle : \mathcal{L}(\mathcal{D}, \beta | \beta^*) \leq \min_\beta \mathcal{L}(\mathcal{D}, \beta | \beta^*) + r\}$ interval.
Easy *if and only if* $\mathcal{T}_r$ is an "interval" *for all* r.

$$\Gamma(\zeta) = \min_{\beta \in \Theta, \langle \beta, \beta^* \rangle = \zeta} \mathcal{L}(\mathcal{D}, \beta | \beta^*), \zeta \in [0, 1].$$

Lemma: OGP if and only if Γ is non-monotonic.





- **1st MM:** lower bound on Γ. **2nd MM:** upper bound on Γ.

This talk: a short survey

Overlap Gap Property (OGP) for **inference**. [Gamarnik, **Z** '17]

(1) *Definition* and *overview*
(2) Two case-studies:
   - *sparse regression*
   - *planted clique* (more involved)

The Sparse Regression Model

Setup

Let $\beta^* \in \{0,1\}^p$ be a **binary** k-**sparse** vector (regime $k = o(p)$.)
For

- $X \in \mathbb{R}^{n \times p}$ consisting of i.i.d $\mathcal{N}(0,1)$ entries
- $W \in \mathbb{R}^n$ consisting of i.i.d. $\mathcal{N}(0,\sigma^2)$ entries

we get n **noisy linear samples** of β^*, $Y \in \mathbb{R}^n$, given by,

$$Y := X\beta^* + W.$$

The Sparse Regression Model

Setup

Let $\beta^* \in \{0,1\}^p$ be a **binary** k-**sparse** vector (regime $k = o(p)$.)
For

- $X \in \mathbb{R}^{n \times p}$ consisting of i.i.d $\mathcal{N}(0,1)$ entries
- $W \in \mathbb{R}^n$ consisting of i.i.d. $\mathcal{N}(0,\sigma^2)$ entries

we get n **noisy linear samples** of β^*, $Y \in \mathbb{R}^n$, given by,

$$Y := X\beta^* + W.$$

Goal: Statistical and Computational Limit

Minimum $n = n_p$: given (Y, X) recover β^* w.h.p. as $p \to +\infty$.

Computational-Statistical Gap

Under $k/\sigma^2 = \omega(1)$:

$$n^* = \frac{2k \log \frac{p}{k}}{\log(k/\sigma^2+1)}, \; n_{\text{alg}} = 2k \log \frac{p}{k}.$$

Computational-Statistical Gap

Under $k/\sigma^2 = \omega(1)$:

$n^* = \frac{2k \log \frac{p}{k}}{\log(k/\sigma^2+1)}$, $n_{\text{alg}} = 2k \log \frac{p}{k}$. *Recall:* **OGP:** loss funct+ monotonicity.

Computational-Statistical Gap

Under $k/\sigma^2 = \omega(1)$:

$n^* = \frac{2k\log\frac{p}{k}}{\log(k/\sigma^2+1)}$, $n_{\text{alg}} = 2k\log\frac{p}{k}$. *Recall:* **OGP:** loss funct+ monotonicity.

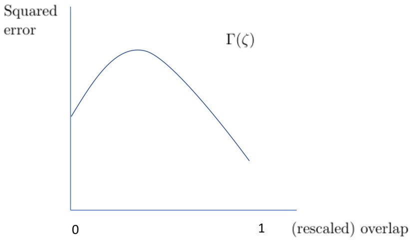
OGP via likelihood loss function

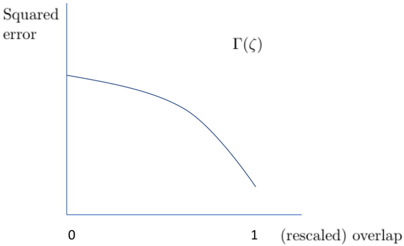
$$\min_{\beta\in\{0,1\}^p, \|\beta\|_0=k} n^{-\frac{1}{2}}\|Y - X\beta\|_2$$

The OGP for Regression — Phase Transition

For $\zeta \in [0,1]$,

$$\Gamma(\zeta) = \min_{\beta \in \{0,1\}^p, \|\beta\|_0 = k, \langle \beta, \beta^* \rangle = \zeta k} n^{-\frac{1}{2}} \|Y - X\beta\|_2$$





The OGP for Regression — Phase Transition

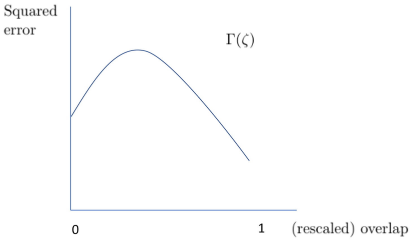
For $\zeta \in [0, 1]$,

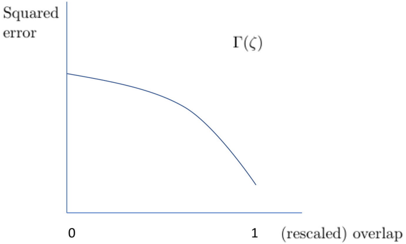
$$\Gamma(\zeta) = \min_{\beta \in \{0,1\}^p, \|\beta\|_0 = k, \langle \beta, \beta^* \rangle = \zeta k} n^{-\frac{1}{2}} \|Y - X\beta\|_2$$

Theorem (Gamarnik, Z '17)

Suppose $k \leq \exp(\sqrt{\log p})$. *There exists* $C > 1 > c > 0$ *such that,*

- *If* $n^* < n < c n_{alg}$ *then w.h.p.* Γ *is not monotonic (OGP).*
- *If* $n > C n_{alg}$ *then w.h.p.* Γ *is monotonic (no-OGP).*





The OGP for Regression — Phase Transition

For $\zeta \in [0,1]$,

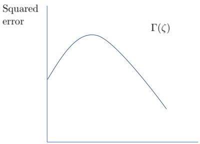
$$\Gamma(\zeta) = \min_{\beta \in \{0,1\}^p, \|\beta\|_0 = k, \langle \beta, \beta^* \rangle = \zeta k} n^{-\frac{1}{2}} \|Y - X\beta\|_2$$

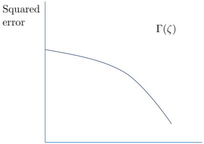
Theorem (Gamarnik, Z '17)

Suppose $k \leq \exp(\sqrt{\log p})$. *There exists* $C > 1 > c > 0$ *such that,*

- *If* $n^* < n < cn_{alg}$ *then w.h.p.* Γ *is not monotonic (OGP).*
- *If* $n > Cn_{alg}$ *then w.h.p.* Γ *is monotonic (no-OGP).*

OGP coincides with the failure of **convex relaxation** and **compressed sensing** methods!





The Planted Clique Model

The Planted Clique Model [Jerrum '92], [Kučera '95]

Generating Assumptions for $G(n, 1/2, k)$**:**

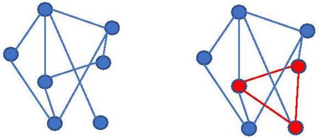
The Planted Clique Model

The Planted Clique Model [Jerrum '92], [Kučera '95]

Generating Assumptions for $G(n, 1/2, k)$**:**

- Stage 1: $\mathcal{G}_0$ is an Erdos-Renyi $G(n, 1/2)$.

$n = 7, k = 3$, $\mathcal{G}_0$ (left) and $\mathcal{G}$ (right) :



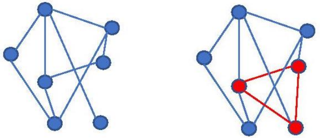
The Planted Clique Model

The Planted Clique Model [Jerrum '92], [Kučera '95]

Generating Assumptions for $G(n, 1/2, k)$**:**

- Stage 1: $\mathcal{G}_0$ is an Erdos-Renyi $G(n, 1/2)$.
- Stage 2: k out of the n vertices of $\mathcal{G}_0$ are chosen u.a.r. to form a k-vertex clique, $\mathcal{PC}$. Call $\mathcal{G}$ the final graph.

$n = 7, k = 3$, $\mathcal{G}_0$ (left) and $\mathcal{G}$ (right) :



The Planted Clique Model

The Planted Clique Model [Jerrum '92], [Kučera '95]

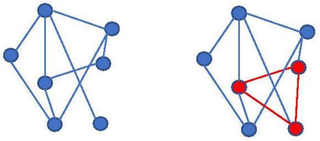
Generating Assumptions for $G(n, 1/2, k)$**:**

- Stage 1: $\mathcal{G}_0$ is an Erdos-Renyi $G(n, 1/2)$.
- Stage 2: k out of the n vertices of $\mathcal{G}_0$ are chosen u.a.r. to form a k-vertex clique, $\mathcal{PC}$. Call $\mathcal{G}$ the final graph.

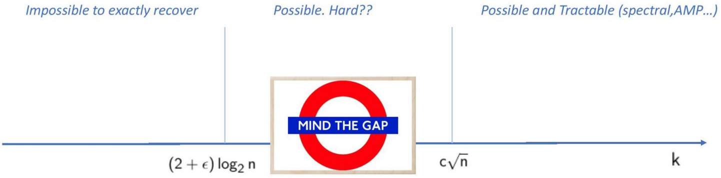
Goal: Recover $\mathcal{PC}$ from observing $\mathcal{G} \sim G(n, 1/2, k)$.

Question: For how small $k = k_n$ can we recover?

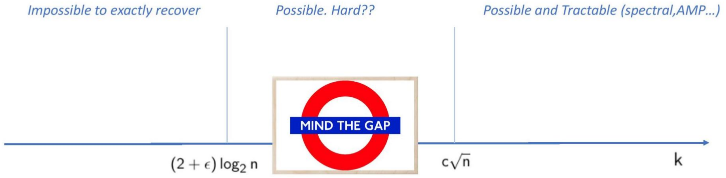
$n = 7, k = 3$, $\mathcal{G}_0$ (left) and $\mathcal{G}$ (right) :



The Planted Clique Model - Brief Literature Review

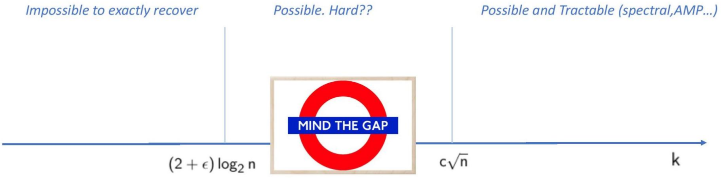


The Planted Clique Model - Brief Literature Review



- Likelihood “Dirac” (unique k-clique): **no landscape!**

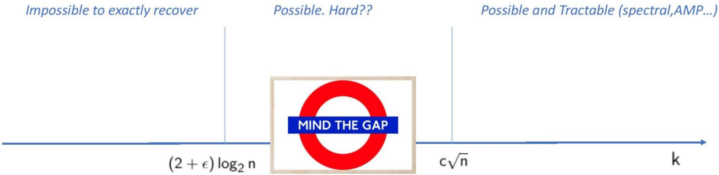
The Planted Clique Model - Brief Literature Review



- Likelihood "Dirac" (unique k-clique): **no landscape!**
- Proxy (densest subgraph problem):

$$\min_{C \subseteq V(G), |C|=k} \mathcal{L}(G, C|\mathcal{PC}) = \binom{k}{2} - |E_G[C]|.$$

The Planted Clique Model - Brief Literature Review



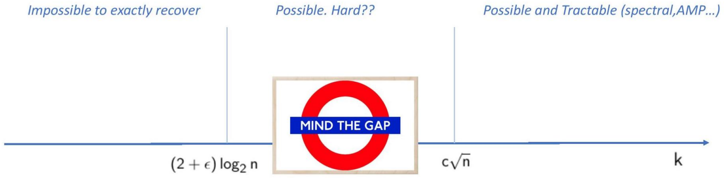
- Likelihood "Dirac" (unique k-clique): **no landscape!**
- Proxy (densest subgraph problem):

$$\min_{C\subseteq V(G),|C|=k} \mathcal{L}(G,C|\mathcal{PC}) = \binom{k}{2} - |E_G[C]|.$$

- Overparametrized densest subgraph problem $\bar{k} \geq k$:

$$\max_{C\subseteq V(G),|C|=\bar{k}} |E_G[C]|.$$

The Planted Clique Model - Brief Literature Review



- Likelihood "Dirac" (unique k-clique): **no landscape!**
- Proxy (densest subgraph problem):

$$\min_{C\subseteq V(G),|C|=k} \mathcal{L}(G, C|\mathcal{PC}) = \binom{k}{2} - |E_G[C]|.$$

- Overparametrized densest subgraph problem $\bar{k} \geq k$:

$$\max_{C\subseteq V(G),|C|=\bar{k}} |E_G[C]|.$$

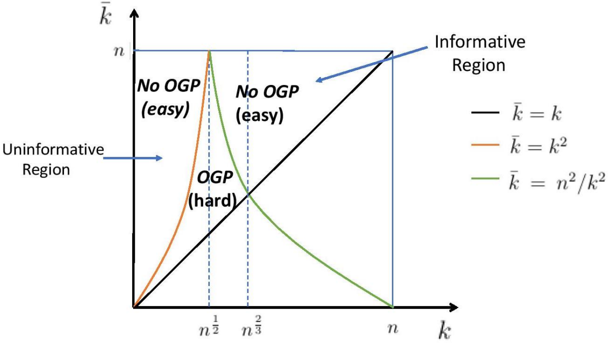
For $z \in [\bar{k}k/n, k]$, let $\Gamma_{\bar{k}}(z) = \max_{C\subseteq V(G),|C|=\bar{k},|C\cap\mathcal{PC}|=z} |E[C]|$.

The Planted Clique Model - First moment OGP diagram

Fix $z \in [\bar{k}k/n, k]$ and let $\Gamma_{\bar{k}}(z) = \max_{C \subseteq V(G), |C|=\bar{k}, |C \cap \mathcal{PC}|=z} |E[C]|$.

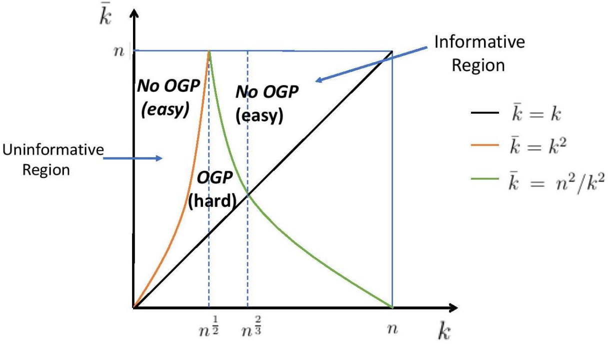
The Planted Clique Model - First moment OGP diagram

Fix $z \in [\bar{k}k/n, k]$ and let $\Gamma_{\bar{k}}(z) = \max_{C \subseteq V(G), |C| = \bar{k}, |C \cap \mathcal{PC}| = z} |E[C]|$.



The Planted Clique Model - First moment OGP diagram

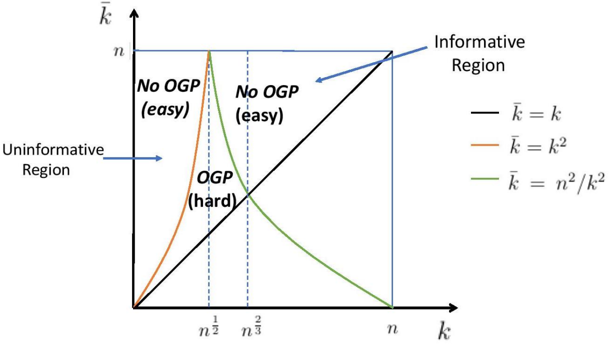
Fix $z \in [\bar{k}k/n, k]$ and let $\Gamma_{\bar{k}}(z) = \max_{C \subseteq V(G), |C| = \bar{k}, |C \cap \mathcal{PC}| = z} |E[C]|$.



- $k = n^{2/3}$-OGP transition for "standard" landscape! (Maria's talk)

The Planted Clique Model - First moment OGP diagram

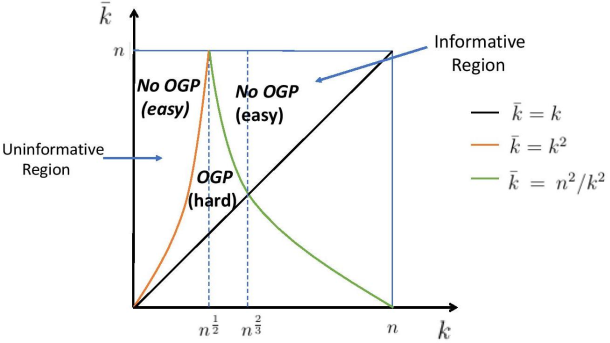
Fix $z \in [\bar{k}k/n, k]$ and let $\Gamma_{\bar{k}}(z) = \max_{C \subseteq V(G), |C| = \bar{k}, |C \cap \mathcal{PC}| = z} |E[C]|$.



- $k = n^{2/3}$-OGP transition for "standard" landscape! (Maria's talk)
- New OGP transition at $\sqrt{n}$: need to lift!

The Planted Clique Model - First moment OGP diagram

Fix $z \in [\bar{k}k/n, k]$ and let $\Gamma_{\bar{k}}(z) = \max_{C \subseteq V(G), |C|=\bar{k}, |C \cap \mathcal{PC}|=z} |E[C]|$.



- $k = n^{2/3}$-OGP transition for "standard" landscape! (Maria's talk)
- New OGP transition at $\sqrt{n}$: need to lift! (agrees with algos)

This talk: a short survey

Overlap Gap Property (OGP) for **inference**. [Gamarnik, **Z** '17]

(1) *Definition* and *overview*

(2) Two case-studies:

- *sparse regression*
- *planted clique* (more involved)

Other/Future Work

(1) OGP for inference implies *"local" MCMC lower bounds.*
(Ben Arous, Wein, **Z** '20), (Gamarnik, Jagannath, Sen '19).

Other/Future Work

(1) OGP for inference implies *"local" MCMC lower bounds.*
(Ben Arous, Wein, **Z** '20), (Gamarnik, Jagannath, Sen '19).
(BAW**Z**'20): In sparse PCA subexponential-time predictions via OGP **identical** with low-degree method.

Other/Future Work

(1) OGP for inference implies *"local" MCMC lower bounds.*
(Ben Arous, Wein, **Z** '20), (Gamarnik, Jagannath, Sen '19).
(BAW**Z**'20): In sparse PCA subexponential-time predictions via OGP **identical** with low-degree method.

(a) OGP for inference gives valid hardness predictions!

Other/Future Work

(1) OGP for inference implies *"local" MCMC lower bounds.*
(Ben Arous, Wein, **Z** '20), (Gamarnik, Jagannath, Sen '19).
(BAW**Z**'20): In sparse PCA subexponential-time predictions via OGP **identical** with low-degree method.

(a) OGP for inference gives valid hardness predictions!
Optimal loss function?: *Overparametrization in PC.*

Other/Future Work

(1) OGP for inference implies *"local" MCMC lower bounds.*
(Ben Arous, Wein, **Z** '20), (Gamarnik, Jagannath, Sen '19).
(BAW**Z**'20): In sparse PCA subexponential-time predictions via OGP **identical** with low-degree method.

(a) OGP for inference gives valid hardness predictions!
Optimal loss function?: *Overparametrization in PC.*

(b) Does **OGP** $\Rightarrow$ failure of stable algorithms (beyond MCMC)?
Similar to *OGP in optimization* (GJW '20).

Other/Future Work

(1) OGP for inference implies *"local" MCMC lower bounds.*
(Ben Arous, Wein, **Z** '20), (Gamarnik, Jagannath, Sen '19).
(BAW**Z**'20): In sparse PCA subexponential-time predictions via OGP **identical** with low-degree method.

(a) OGP for inference gives valid hardness predictions!
Optimal loss function?: *Overparametrization in PC.*

(b) Does **OGP** $\Rightarrow$ failure of stable algorithms (beyond MCMC)?
Similar to *OGP in optimization* (GJW '20).

(c) Does **absence of OGP** $\Rightarrow$ success of "local" methods?
Similar to *no-OGP (FRSB) in spin glasses.*
(Subag '18), (Montanari, El Alaoui, Sellke '19 -'21).

Other/Future Work

(1) OGP for inference implies *"local" MCMC lower bounds.*
(Ben Arous, Wein, **Z** '20), (Gamarnik, Jagannath, Sen '19).
(BAW**Z**'20): In sparse PCA subexponential-time predictions via OGP **identical** with low-degree method.

(a) OGP for inference gives valid hardness predictions!
Optimal loss function?: *Overparametrization in PC.*

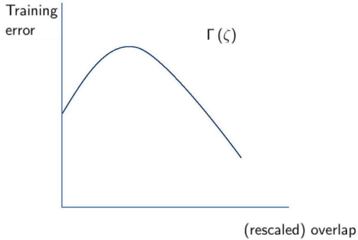
(b) Does **OGP** $\Rightarrow$ failure of stable algorithms (beyond MCMC)?
Similar to *OGP in optimization* (GJW '20).

(c) Does **absence of OGP** $\Rightarrow$ success of "local" methods?
Similar to *no-OGP (FRSB) in spin glasses.*
(Subag '18), (Montanari, El Alaoui, Sellke '19 -'21).

Thank you!!

MCMC lower bounds

$\min_{\beta \in \Theta} \mathcal{L}(\mathcal{D}, \beta | \beta^*)$. $\Gamma(\zeta) = \min_{\beta \in \Theta, \langle \beta, \beta^* \rangle = \zeta} \mathcal{L}(\mathcal{D}, \beta | \beta^*)$, $\zeta \in [0, 1]$.

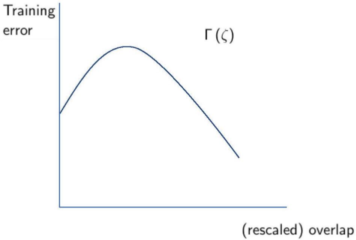


MCMC lower bounds

$\min_{\beta \in \Theta} \mathcal{L}(\mathcal{D}, \beta | \beta^*)$. $\Gamma(\zeta) = \min_{\beta \in \Theta, \langle \beta, \beta^* \rangle = \zeta} \mathcal{L}(\mathcal{D}, \beta | \beta^*), \zeta \in [0,1]$.

[Ben Arous, Wein, **Z** '20], [Gamarnik, Jagannath, Sen '19]

Under OGP **of height** D and for "small" $T > 0$
any "local" MC with stationary $\mu(\beta) \propto e^{-T^{-1} \mathcal{L}(\mathcal{D}, \beta | \beta^*)}, \beta \in \Theta$
(worst-case initialization) needs time exp(D).



MCMC lower bounds

$\min_{\beta\in\Theta} \mathcal{L}(\mathcal{D}, \beta|\beta^*)$. $\Gamma(\zeta) = \min_{\beta\in\Theta, \langle\beta,\beta^*\rangle=\zeta} \mathcal{L}(\mathcal{D}, \beta|\beta^*)$, $\zeta \in [0, 1]$.

[Ben Arous, Wein, **Z** '20], [Gamarnik, Jagannath, Sen '19]

Under OGP **of height** D and for "small" $T > 0$
any "local" MC with stationary $\mu(\beta) \propto e^{-T^{-1}\mathcal{L}(\mathcal{D},\beta|\beta^*)}$, $\beta \in \Theta$
(worst-case initialization) needs time exp(D).

Sparse PCA: OGP height is low-degree sub-exponential time prediction.