Low-Degree Hardness of Maximum Independent Set

Alex Wein Courant Institute, New York University

Joint work with:



David Gamarnik MIT



Aukosh Jagannath _{Waterloo}

Examples:

Examples:

Max clique in a random graph



Examples:

- Max clique in a random graph
- Max-k-SAT on a random formula

Examples:

- Max clique in a random graph
- Max-k-SAT on a random formula
- Maximizing a random degree-p polynomial over the sphere

Examples:

- Max clique in a random graph
- Max-k-SAT on a random formula
- Maximizing a random degree-p polynomial over the sphere

Note: no planted solution

Examples:

- Max clique in a random graph
- Max-k-SAT on a random formula
- Maximizing a random degree-p polynomial over the sphere

Note: no planted solution

Q: What is the typical value of the optimum (OPT)?

Examples:

- Max clique in a random graph
- Max-k-SAT on a random formula
- Maximizing a random degree-p polynomial over the sphere

Note: no planted solution

- Q: What is the typical value of the optimum (OPT)?
- Q: What objective value can be reached algorithmically (ALG)?

Examples:

- Max clique in a random graph
- Max-k-SAT on a random formula
- Maximizing a random degree-p polynomial over the sphere

Note: no planted solution

- Q: What is the typical value of the optimum (OPT)?
- Q: What objective value can be reached algorithmically (ALG)?

 $Q{:}$ In cases where it seems hard to reach a particular objective value, can we understand why?

Examples:

- Max clique in a random graph
- Max-k-SAT on a random formula
- Maximizing a random degree-p polynomial over the sphere

Note: no planted solution

- Q: What is the typical value of the optimum (OPT)?
- Q: What objective value can be reached algorithmically (ALG)?

Q: In cases where it seems hard to reach a particular objective value, can we understand why? In a $\underline{unified}$ way?

Example (max independent set): given sparse graph G(n, d/n),

 $\max_{S \subseteq [n]} |S| \quad s.t. \quad S \text{ independent}$

Example (max independent set): given sparse graph G(n, d/n),

$$\max_{S \subseteq [n]} |S| \quad s.t. \quad S \text{ independent}$$

Example (max independent set): given sparse graph G(n, d/n),

 $\max_{S \subseteq [n]} |S| \quad s.t. \quad S \text{ independent}$

$$OPT = 2 \frac{\log d}{d} n$$
[Frieze '90]

(d large constant)

Example (max independent set): given sparse graph G(n, d/n),

$$\max_{S \subseteq [n]} |S| \quad s.t. \quad S \text{ independent}$$

$$OPT = 2 \frac{\log d}{d} n \qquad ALG = \frac{\log d}{d} n \qquad (d \text{ large constant})$$
[Frieze '90]

Example (max independent set): given sparse graph G(n, d/n),

$$\max_{S \subseteq [n]} |S| \quad s.t. \quad S \text{ independent}$$

$$OPT = 2 \frac{\log d}{d} n \qquad ALG = \frac{\log d}{d} n \qquad (d \text{ large constant})$$
[Frieze '90]

[Karp '76]: Is there a better algorithm?

Example (max independent set): given sparse graph G(n, d/n),

$$\max_{S \subseteq [n]} |S| \quad s.t. \quad S \text{ independent}$$

$$OPT = 2 \frac{\log d}{d} n \qquad ALG = \frac{\log d}{d} n \qquad (d \text{ large constant})$$
[Frieze '90]

[Karp '76]: Is there a better algorithm?

Structural evidence suggests no! [Achlioptas, Coja-Oghlan '08; Coja-Oghlan, Efthymiou '10]

Example (max independent set): given sparse graph G(n, d/n),

$$\max_{S \subseteq [n]} |S| \quad s.t. \quad S \text{ independent}$$

$$OPT = 2 \frac{\log d}{d} n \qquad ALG = \frac{\log d}{d} n \qquad (d \text{ large constant})$$
[Frieze '90]

[Karp '76]: Is there a better algorithm?

Structural evidence suggests no! [Achlioptas, Coja-Oghlan '08; Coja-Oghlan, Efthymiou '10]

Local algorithms achieve value ALG <u>and no better</u> [Gamarnik, Sudan '13; Rahman, Virág '14]

Example (spherical *p*-spin model): for $Y \in (\mathbb{R}^n)^{\otimes p}$ i.i.d. $\mathcal{N}(0,1)$,

$$\max_{\|v\|=1} \frac{1}{\sqrt{n}} \langle Y, v^{\otimes p} \rangle$$

(maximize random degree-*p* polynomial over the sphere)

Example (spherical *p*-spin model): for $Y \in (\mathbb{R}^n)^{\otimes p}$ i.i.d. $\mathcal{N}(0, 1)$,

$$\max_{\|v\|=1} \frac{1}{\sqrt{n}} \langle Y, v^{\otimes p} \rangle$$

(maximize random degree-*p* polynomial over the sphere)

 $OPT_p = \Theta(1)$ [Auffinger, Ben Arous, Černý '13]

Example (spherical *p*-spin model): for $Y \in (\mathbb{R}^n)^{\otimes p}$ i.i.d. $\mathcal{N}(0, 1)$,

$$\max_{\|\mathbf{v}\|=1} \frac{1}{\sqrt{n}} \langle Y, \mathbf{v}^{\otimes p} \rangle$$

(maximize random degree-*p* polynomial over the sphere)

 $OPT_p = \Theta(1)$ [Auffinger, Ben Arous, Černý '13]

 $ALG_p = \Theta(1)$ [Subag '18]

Example (spherical *p*-spin model): for $Y \in (\mathbb{R}^n)^{\otimes p}$ i.i.d. $\mathcal{N}(0, 1)$,

$$\max_{\|\mathbf{v}\|=1} \frac{1}{\sqrt{n}} \langle Y, \mathbf{v}^{\otimes p} \rangle$$

(maximize random degree-*p* polynomial over the sphere)

 $OPT_p = \Theta(1)$ [Auffinger, Ben Arous, Černý '13]

 $ALG_p = \Theta(1)$ [Subag '18]

 $ALG_p < OPT_p$ (for $p \ge 3$)

Example (spherical *p*-spin model): for $Y \in (\mathbb{R}^n)^{\otimes p}$ i.i.d. $\mathcal{N}(0, 1)$,

$$\max_{\|\mathbf{v}\|=1}\frac{1}{\sqrt{n}}\langle Y, \mathbf{v}^{\otimes p}\rangle$$

(maximize random degree-*p* polynomial over the sphere)

 $OPT_p = \Theta(1)$ [Auffinger, Ben Arous, Černý '13]

 $ALG_p = \Theta(1)$ [Subag '18]

 $ALG_p < OPT_p$ (for $p \ge 3$)

Approximate message passing (AMP) algorithms achieve valueALG_p and no better[El Alaoui, Montanari, Sellke '20]

How to give the best "evidence" that there are no better algorithms?

How to give the best "evidence" that there are no better algorithms?

Prior work rules out certain classes of algorithms (local, AMP), but do we expect these to be optimal?

How to give the best "evidence" that there are no better algorithms?

Prior work rules out certain classes of algorithms (local, AMP), but do we expect these to be optimal?

AMP is not optimal for tensor PCA [Montanari, Richard '14]

How to give the best "evidence" that there are no better algorithms?

Prior work rules out certain classes of algorithms (local, AMP), but do we expect these to be optimal?

AMP is not optimal for tensor PCA [Montanari, Richard '14]

Would like a unified framework for lower bounds

How to give the best "evidence" that there are no better algorithms?

Prior work rules out certain classes of algorithms (local, AMP), but do we expect these to be optimal?

AMP is not optimal for tensor PCA [Montanari, Richard '14]

Would like a unified framework for lower bounds

Local algorithms only make sense on sparse graphs

How to give the best "evidence" that there are no better algorithms?

Prior work rules out certain classes of algorithms (local, AMP), but do we expect these to be optimal?

AMP is not optimal for tensor PCA [Montanari, Richard '14]

Would like a unified framework for lower bounds

Local algorithms only make sense on sparse graphs

<u>Solution</u>: lower bounds against a larger class of algorithms (low-degree polynomials) that contains both local and AMP algorithms

Study a restricted class of algorithms: low-degree polynomials

Study a restricted class of algorithms: low-degree polynomials

• Multivariate polynomial $f : \mathbb{R}^M \to \mathbb{R}^N$

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^M \to \mathbb{R}^N$
 - ▶ Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^M \to \mathbb{R}^N$
 - ▶ Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
 - ▶ Output: e.g. $v \in \mathbb{R}^n$

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^M \to \mathbb{R}^N$
 - ▶ Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
 - ▶ Output: e.g. $v \in \mathbb{R}^n$

"Low" degree means O(log n) where n is dimension

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^M \to \mathbb{R}^N$
 - ▶ Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
 - Output: e.g. $v \in \mathbb{R}^n$
- "Low" degree means O(log n) where n is dimension

Examples of low-degree algorithms:

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^M \to \mathbb{R}^N$
 - ▶ Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
 - ▶ Output: e.g. $v \in \mathbb{R}^n$
- "Low" degree means O(log n) where n is dimension

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^M \to \mathbb{R}^N$
 - ▶ Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
 - ▶ Output: e.g. $v \in \mathbb{R}^n$
- ▶ "Low" degree means *O*(log *n*) where *n* is dimension

Examples of low-degree algorithms: input $Y \in \mathbb{R}^{n \times n}$

Power iteration: Y^k1

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^M \to \mathbb{R}^N$
 - ▶ Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
 - ▶ Output: e.g. $v \in \mathbb{R}^n$
- "Low" degree means O(log n) where n is dimension

- Power iteration: Y^k1
- Approximate message passing

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^M \to \mathbb{R}^N$
 - ▶ Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
 - ▶ Output: e.g. $v \in \mathbb{R}^n$
- "Low" degree means O(log n) where n is dimension

- Power iteration: Y^k1
- Approximate message passing
- Local algorithms on sparse graphs

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^M \to \mathbb{R}^N$
 - ▶ Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
 - Output: e.g. $v \in \mathbb{R}^n$

"Low" degree means O(log n) where n is dimension

- Power iteration: Y^k1
- Approximate message passing
- Local algorithms on sparse graphs
- Or any of the above applied to $\tilde{Y} = g(Y)$

Low-degree algorithms are already well-studied for problems with a planted signal

[Barak, Hopkins, Kelner, Kothari, Moitra, Potechin '16] [Hopkins, Steurer '17] [Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer '17] [Hopkins '18] (PhD thesis)

Low-degree algorithms are already well-studied for problems with a planted signal

[Barak, Hopkins, Kelner, Kothari, Moitra, Potechin '16] [Hopkins, Steurer '17]

[Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer '17]

[Hopkins '18] (PhD thesis)

[Kunisky, W., Bandeira '19] (survey)

Low-degree algorithms are already well-studied for problems with a planted signal

[Barak, Hopkins, Kelner, Kothari, Moitra, Potechin '16] [Hopkins, Steurer '17] [Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer '17] [Hopkins '18] (PhD thesis) [Kunisky, W., Bandeira '19] (survey)

For a wide range of planted problems, $O(\log n)$ -degree polynomials are as powerful as the best known poly-time algorithms

Low-degree algorithms are already well-studied for problems with a planted signal

[Barak, Hopkins, Kelner, Kothari, Moitra, Potechin '16] [Hopkins, Steurer '17] [Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer '17] [Hopkins '18] (PhD thesis) [Kunisky, W., Bandeira '19] (survey)

For a wide range of planted problems, $O(\log n)$ -degree polynomials are as powerful as the best known poly-time algorithms

Planted clique, sparse PCA, community detection, tensor PCA, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, ... [BHKKMP16,HS17,HKPRSS17,Hop18,BK**W**19,K**W**B19,DK**W**B19,S**W**20,...]

Low-degree algorithms are already well-studied for problems with a planted signal

[Barak, Hopkins, Kelner, Kothari, Moitra, Potechin '16] [Hopkins, Steurer '17] [Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer '17] [Hopkins '18] (PhD thesis) [Kunisky, W., Bandeira '19] (survey)

For a wide range of planted problems, $O(\log n)$ -degree polynomials are as powerful as the best known poly-time algorithms

Planted clique, sparse PCA, community detection, tensor PCA, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, ... [BHKKMP16,HS17,HKPRSS17,Hop18,BK**W**19,K**W**B19,DK**W**B19,S**W**20,...]

This work: extend low-degree framework to non-planted setting

Results: Max Independent Set

Example (max independent set): given sparse graph G(n, d/n),

$$\max_{S \subseteq [n]} |S| \quad s.t. \quad S \text{ independent}$$

OPT = 2 $\frac{\log d}{d}n \qquad$ ALG = $\frac{\log d}{d}n$

Results: Max Independent Set

Example (max independent set): given sparse graph G(n, d/n),

$$\max_{S \subseteq [n]} |S| \quad s.t. \quad S \text{ independent}$$

$$OPT = 2 \frac{\log d}{d}n \qquad ALG = \frac{\log d}{d}n$$

<u>Result</u>: no low-degree polynomial can achieve $(1 + \epsilon) \frac{\log d}{d} n$

Results: Max Independent Set

Example (max independent set): given sparse graph G(n, d/n),

$$\max_{S \subseteq [n]} |S| \quad s.t. \quad S \text{ independent}$$

$$OPT = 2 \frac{\log d}{d}n \qquad ALG = \frac{\log d}{d}n$$

<u>Result</u>: no low-degree polynomial can achieve $(1 + \epsilon) \frac{\log d}{d}n$

Theorem [Gamarnik, Jagannath, W. '20; W. '20] No polynomial $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}^n$ of degree $\operatorname{polylog}(n)$ achieves both of the following with probability $1 - \exp(-n^{\Omega(1)})$:

▶
$$f_i(Y) \in [0, 1/3] \cup [2/3, 1]$$
 for most *i*

• $\{i: f_i(Y) \in [2/3, 1]\}$ is a near-indep set of size $(1 + \epsilon) \frac{\log d}{d} n$

How to prove failure of low-degree polynomials?

How to prove failure of low-degree polynomials?

For problems with a planted signal:

- Detection: linear algebra [BHKKMP'16; HS'17; HKPRSS'17]
- Recovery: Jensen + linear algebra [Schramm, W. '20]

How to prove failure of low-degree polynomials?

For problems with a planted signal:

- Detection: linear algebra [BHKKMP'16; HS'17; HKPRSS'17]
- Recovery: Jensen + linear algebra [Schramm, W. '20]

For random optimization problems, need different approach:

How to prove failure of low-degree polynomials?

For problems with a planted signal:

- Detection: linear algebra [BHKKMP'16; HS'17; HKPRSS'17]
- Recovery: Jensen + linear algebra [Schramm, W. '20]

For random optimization problems, need different approach:

Stability of low-degree polynomials

How to prove failure of low-degree polynomials?

For problems with a planted signal:

- Detection: linear algebra [BHKKMP'16; HS'17; HKPRSS'17]
- Recovery: Jensen + linear algebra [Schramm, W. '20]

For random optimization problems, need different approach:

Stability of low-degree polynomials

Overlap gap property (OGP)
 [Gamarnik, Sudan '13]
 [Rahman, Virág '14]
 [Chen, Gamarnik, Panchenko, Rahman '17]
 [Gamarnik, Jagannath '19]

 $Y \sim \text{i.i.d. Bernoulli}(p)$

 $Y \sim \text{i.i.d. Bernoulli}(p)$

Interpolation path: $Y^{(0)}$ $Y^{(1)}$ $Y^{(2)}$ \cdots $Y^{(m-1)}$ $Y^{(m)}$

 $Y \sim \text{i.i.d. Bernoulli}(p)$

Interpolation path: $Y^{(0)}$ $Y^{(1)}$ $Y^{(2)}$ \cdots $Y^{(m-1)}$ $Y^{(m)}$

Fix $f: \{0,1\}^m \to \mathbb{R}^n$ degree D

 $Y \sim \text{i.i.d. Bernoulli}(p)$

Interpolation path: $Y^{(0)}$ $Y^{(1)}$ $Y^{(2)}$ \cdots $Y^{(m-1)}$ $Y^{(m)}$

Fix $f: \{0,1\}^m \to \mathbb{R}^n$ degree D

Definition: Step *i* is "*c*-bad" if

$$||f(Y^{(i)}) - f(Y^{(i-1)})||^2 > c \mathop{\mathbb{E}}_{Y} ||f(Y)||^2$$

 $Y \sim \text{i.i.d. Bernoulli}(p)$

Interpolation path: $Y^{(0)}$ $Y^{(1)}$ $Y^{(2)}$ \cdots $Y^{(m-1)}$ $Y^{(m)}$

Fix $f: \{0,1\}^m \to \mathbb{R}^n$ degree D

Definition: Step *i* is "*c*-bad" if

$$\|f(Y^{(i)}) - f(Y^{(i-1)})\|^2 > c \mathop{\mathbb{E}}_{Y} \|f(Y)\|^2$$

Theorem [Gamarnik, Jagannath, W. '20]

$$\Pr_{Y^{(0)},...,Y^{(m)}}[\nexists c\text{-bad } i] \ge p^{4D/c}$$

With non-trivial probability (over path), f's output is "smooth"

Overlap Gap Property [Gamarnik, Sudan '13]

Overlap gap property (OGP): with high probability over $Y \sim G(n, d/n)$, there does not exist $S, T \subseteq [n]$ such that

► *S*, *T* independent sets

$$|S|, |T| \ge (1 + \frac{1}{\sqrt{2}})\Phi \qquad \Phi := \frac{\log d}{d}n$$
$$|S \cap T| \approx \Phi$$

Proof: first moment method [Gamarnik, Sudan '13]

Ensemble OGP [CGPR'17, GJ'19]

Ensemble OGP: with high probability over

 $Y^{(0)}$ $Y^{(1)}$ $Y^{(2)}$ \cdots $Y^{(m-1)}$ $Y^{(m)}$

there does not exist $S, T \subseteq [n]$ such that

- S independent set in some $Y^{(i)}$
- ► T independent set in some Y^(j)
- $\blacktriangleright |S|, |T| \ge (1 + \frac{1}{\sqrt{2}}) \Phi \qquad \Phi := \frac{\log d}{d} n$

$$\blacktriangleright |S \cap T| \approx \Phi$$

Proof [Gamarnik, Jagannath, W. '20]

Proof that low-degree polynomials cannot reach $(1 + \frac{1}{\sqrt{2}})\Phi$: Suppose f(Y) outputs independent sets of size $(1 + \frac{1}{\sqrt{2}})\Phi$

 $Y^{(0)}$ $Y^{(1)}$ $Y^{(2)}$ \ldots $Y^{(m-1)}$ $Y^{(m)}$

 $\begin{array}{l} \underline{\text{Separation}}: \ f(Y^{(0)}) \ \text{and} \ f(Y^{(m)}) \ \text{are "far apart"} \\ \underline{\text{Stability}}: \ \text{with probability} \gtrsim n^{-D}, \ \text{there are no big "jumps"} \\ f(Y^{(i)}) \rightarrow f(Y^{(i+1)}) \end{array} \end{array}$

Contradicts OGP

• Improvement to $(1 + \epsilon) \frac{\log d}{d} n$ via ensemble multi-OGP [W. '20]

Inspired by [Rahman, Virág '14]

• Improvement to $(1 + \epsilon) \frac{\log d}{d} n$ via ensemble multi-OGP [W. '20]

Inspired by [Rahman, Virág '14]

Similar results for *p*-spin model

▶ Improvement to $(1 + \epsilon) \frac{\log d}{d} n$ via ensemble multi-OGP [W. '20]

Inspired by [Rahman, Virág '14]

 Similar results for *p*-spin model
 Proof of OGP for *p*-spin (for *p* ≥ 4 even) [Chen, Sen '15; Auffinger, Chen '17]

• Improvement to $(1 + \epsilon) \frac{\log d}{d} n$ via ensemble multi-OGP [W. '20]

Inspired by [Rahman, Virág '14]

Similar results for *p*-spin model

▶ Proof of OGP for *p*-spin (for *p* ≥ 4 even) [Chen, Sen '15; Auffinger, Chen '17]

Also rule out Langevin dynamics

• Improvement to $(1 + \epsilon) \frac{\log d}{d} n$ via ensemble multi-OGP [W. '20]

Inspired by [Rahman, Virág '14]

Similar results for *p*-spin model
 Proof of OGP for *p*-spin (for *p* > 4 even)

[Chen, Sen '15; Auffinger, Chen '17]

Also rule out Langevin dynamics

▶ Follow-up work on random k-SAT [Bresler, Huang '21]

• Improvement to $(1 + \epsilon) \frac{\log d}{d} n$ via ensemble multi-OGP [W. '20]

Inspired by [Rahman, Virág '14]

Similar results for *p*-spin model
 Proof of OGP for *p*-spin (for *p* ≥ 4 even)

[Chen, Sen '15; Auffinger, Chen '17]

Also rule out Langevin dynamics

▶ Follow-up work on random *k*-SAT [Bresler, Huang '21]

Connections between heuristics

▶ OGP \rightarrow Low-Degree

• Improvement to $(1 + \epsilon) \frac{\log d}{d} n$ via ensemble multi-OGP [W. '20]

Inspired by [Rahman, Virág '14]

 Similar results for *p*-spin model
 Proof of OGP for *p*-spin (for *p* ≥ 4 even) [Chen, Sen '15; Auffinger, Chen '17]

Also rule out Langevin dynamics

▶ Follow-up work on random k-SAT [Bresler, Huang '21]

Connections between heuristics

► OGP \rightarrow Low-Degree

Thanks!