Sub-exponential time Sum-of-Squares lower bounds for Principal Components Analysis

Goutham Rajendran

September 29, 2021

Joint work with Aaron Potechin

THE UNIVERSITY OF CHICAGO

Principal Components Analysis

Principal Components Analysis (PCA): A popular data processing and dimension reduction routine, with numerous applications in Machine Learning, Statistics, Engineering, Biology, etc.

Principal Components Analysis

Principal Components Analysis (PCA): A popular data processing and dimension reduction routine, with numerous applications in Machine Learning, Statistics, Engineering, Biology, etc.

Has drawbacks

- Components hard to interpret
- Bad estimators in high dimensions

Principal Components Analysis

Principal Components Analysis (PCA): A popular data processing and dimension reduction routine, with numerous applications in Machine Learning, Statistics, Engineering, Biology, etc.

Has drawbacks

- Components hard to interpret
- Bad estimators in high dimensions

In practice, variants are used

- Sparse PCA: Search for sparse components. Tremendously useful for many applications and has many advantages such as efficiency, interpretability, etc.

Principal Components Analysis

Principal Components Analysis (PCA): A popular data processing and dimension reduction routine, with numerous applications in Machine Learning, Statistics, Engineering, Biology, etc.

Has drawbacks

- Components hard to interpret
- Bad estimators in high dimensions

In practice, variants are used

- Sparse PCA: Search for sparse components. Tremendously useful for many applications and has many advantages such as efficiency, interpretability, etc.
- Tensor PCA

Wishart model of Sparse PCA

A statistical model originally proposed by [JL09]

Wishart model of Sparse PCA

A statistical model originally proposed by [JL09]

Given samples $v_1, \ldots, v_m \in \mathbb{R}^d$ from $\mathcal{N}(0, I_d + \lambda u u^\intercal)$, recover u.
Here,

- $u \in \mathbb{R}^n$ is a unit vector
- u is k-sparse, that is, $\|u\|_0 = k$.
- λ is the signal-to-noise ratio.

Wishart model of Sparse PCA

A statistical model originally proposed by [JL09]

Given samples $v_1, \ldots, v_m \in \mathbb{R}^d$ from $\mathcal{N}(0, I_d + \lambda u u^\intercal)$, recover u.
Here,

- $u \in \mathbb{R}^n$ is a unit vector
- u is k-sparse, that is, $\|u\|_0 = k$.
- λ is the signal-to-noise ratio.

In particular, we will be interested in the regime

$$m \ll \frac{d}{\lambda^2}, \qquad m \ll \frac{k^2}{\lambda^2}$$

Wishart model of Sparse PCA

A statistical model originally proposed by [JL09]

Given samples $v_1, \ldots, v_m \in \mathbb{R}^d$ from $\mathcal{N}(0, I_d + \lambda u u^\intercal)$, recover u.
Here,

- $u \in \mathbb{R}^n$ is a unit vector
- u is k-sparse, that is, $\|u\|_0 = k$.
- λ is the signal-to-noise ratio.

In particular, we will be interested in the regime

$$m \ll \frac{d}{\lambda^2}, \qquad m \ll \frac{k^2}{\lambda^2}$$

In other regimes, algorithms exist for recovery and/or hypothesis testing. But this regime is believed to be hard (various works show conditional hardness results).

The Sum of Squares hierarchy

We will study the performance of the Sum of Squares (SoS) hierarchy of algorithms.

The Sum of Squares hierarchy

We will study the performance of the Sum of Squares (SoS) hierarchy of algorithms.

SoS algorithms have recently revolutionzed robust statistics, settling long-standing open problems such as optimal robust learning of high dimensional Gaussian mixtures.

The Sum of Squares hierarchy

We will study the performance of the Sum of Squares (SoS) hierarchy of algorithms.

SoS algorithms have recently revolutionzed robust statistics, settling long-standing open problems such as optimal robust learning of high dimensional Gaussian mixtures.

SoS algorithms are believed to be the optimal algorithm for many statistical problems

The Sum of Squares hierarchy

We will study the performance of the Sum of Squares (SoS) hierarchy of algorithms.

SoS algorithms have recently revolutionzed robust statistics, settling long-standing open problems such as optimal robust learning of high dimensional Gaussian mixtures.

SoS algorithms are believed to be the optimal algorithm for many statistical problems

Therefore, we ask (also posed as an open problem in several prior works)

Can Sum-of-Squares algorithms beat known algorithms for Sparse PCA?

The Sum of Squares hierarchy

We will study the performance of the Sum of Squares (SoS) hierarchy of algorithms.

SoS algorithms have recently revolutionzed robust statistics, settling long-standing open problems such as optimal robust learning of high dimensional Gaussian mixtures.

SoS algorithms are believed to be the optimal algorithm for many statistical problems

Therefore, we ask (also posed as an open problem in several prior works)

Can Sum-of-Squares algorithms beat known algorithms for Sparse PCA?

In this work, we answer this negatively.

The Sum of Squares hierarchy

More concretely, the SoS hierarchy is a family of semidefinite relaxations parameterized by the degree or level d.

The Sum of Squares hierarchy

More concretely, the SoS hierarchy is a family of semidefinite relaxations parameterized by the degree or level d.

The programs get stronger as d increases, and degree-d SoS can be solved in $n^{O(d)}$ time, where n is the input size.

The Sum of Squares hierarchy

More concretely, the SoS hierarchy is a family of semidefinite relaxations parameterized by the degree or level d.

The programs get stronger as d increases, and degree-d SoS can be solved in $n^{O(d)}$ time, where n is the input size.

Constant degree d corresponds to polynomial runtime.

The Sum of Squares hierarchy

More concretely, the SoS hierarchy is a family of semidefinite relaxations parameterized by the degree or level d.

The programs get stronger as d increases, and degree-d SoS can be solved in $n^{O(d)}$ time, where n is the input size.

Constant degree d corresponds to polynomial runtime.

In this work, our lower bounds focus on $d \approx n^{\epsilon}$, corresponding to subexponential runtime.

Why SoS lower bounds

SoS captures the guarantees of many known algorithms and also obtains new guarantees, in optimization and algorithms (e.g. MaxCut[GW95], Sparsest cut [ARV04], Tensor PCA [HSS15]).

Why SoS lower bounds

SoS captures the guarantees of many known algorithms and also obtains new guarantees, in optimization and algorithms (e.g. MaxCut[GW95], Sparsest cut [ARV04], Tensor PCA [HSS15]).

A striking result of Raghavendra [Rag08] shows that SoS is optimal for all constraint satisfaction problems, assuming the Unique Games conjecture.

Why SoS lower bounds

SoS captures the guarantees of many known algorithms and also obtains new guarantees, in optimization and algorithms (e.g. MaxCut[GW95], Sparsest cut [ARV04], Tensor PCA [HSS15]).

A striking result of Raghavendra [Rag08] shows that SoS is optimal for all constraint satisfaction problems, assuming the Unique Games conjecture.

Therefore, showing unconditional SoS lower bounds serves as strong evidence of hardness of computation.

Why SoS lower bounds

SoS captures the guarantees of many known algorithms and also obtains new guarantees, in optimization and algorithms (e.g. MaxCut[GW95], Sparsest cut [ARV04], Tensor PCA [HSS15]).

A striking result of Raghavendra [Rag08] shows that SoS is optimal for all constraint satisfaction problems, assuming the Unique Games conjecture.

Therefore, showing unconditional SoS lower bounds serves as strong evidence of hardness of computation.

This is especially for average-case problems where NP-hardness remains out of reach of our current techniques.

Why SoS lower bounds

SoS captures the guarantees of many known algorithms and also obtains new guarantees, in optimization and algorithms (e.g. MaxCut[GW95], Sparsest cut [ARV04], Tensor PCA [HSS15]).

A striking result of Raghavendra [Rag08] shows that SoS is optimal for all constraint satisfaction problems, assuming the Unique Games conjecture.

Therefore, showing unconditional SoS lower bounds serves as strong evidence of hardness of computation.

This is especially for average-case problems where NP-hardness remains out of reach of our current techniques.

Often, the very task of constructing such lower bounds sheds light on the aspects of the problem that makes it hard.

SoS lower bounds in prior works

Early works showed degree-2 lower bounds (e.g. Max-Cut [FS02], k-clique [FK00]).

SoS lower bounds in prior works

Early works showed degree-2 lower bounds (e.g. Max-Cut [FS02], k-clique [FK00]).

Grigoriev and Schoenebeck showed linear degree lower bounds for k-XOR, k-SAT and knapsack [Gri01a, Gri01b], which were reduced to other problems [Tul09, BCG$^+$12].

SoS lower bounds in prior works

Early works showed degree-2 lower bounds (e.g. Max-Cut [FS02], k-clique [FK00]).

Grigoriev and Schoenebeck showed linear degree lower bounds for k-XOR, k-SAT and knapsack [Gri01a, Gri01b], which were reduced to other problems [Tul09, BCG^{+}12].

Other works include SoS lower bounds for maximum clique on random graphs [BHK^{+}16], Max k-CSPs on random instances [KMOW17], etc.

SoS lower bounds for Sparse PCA

Theorem: SoS lower bounds for Sparse PCA

For the Wishart model of Sparse PCA, sub-exponential time SoS algorithms fail to recover the principal component when the number of samples $m \ll \min(\frac{d}{\lambda^2}, \frac{k^2}{\lambda^2})$

SoS lower bounds for Sparse PCA

Theorem: SoS lower bounds for Sparse PCA

For the Wishart model of Sparse PCA, sub-exponential time SoS algorithms fail to recover the principal component when the number of samples $m \ll \min(\frac{d}{\lambda^2}, \frac{k^2}{\lambda^2})$

Prior work on algorithms: In other regimes, algorithms are known, e.g. diagonal thresholding, covariance thresholding, SoS algorithms, etc.

SoS lower bounds for Sparse PCA

Theorem: SoS lower bounds for Sparse PCA

For the Wishart model of Sparse PCA, sub-exponential time SoS algorithms fail to recover the principal component when the number of samples $m \ll \min(\frac{d}{\lambda^2}, \frac{k^2}{\lambda^2})$

Prior work on algorithms: In other regimes, algorithms are known, e.g. diagonal thresholding, covariance thresholding, SoS algorithms, etc.

Prior work on inapproximability:

- Conditional: Reductions from other conjectures, Landscape behavior.
- Unconditional: Statistical query lower bounds have also been studied, degree 2 and degree 4 (weak) SoS lower bounds
- SoS lower bounds for the related Wigner model (our techniques also recover these)

Computational barrier diagram for Sparse PCA

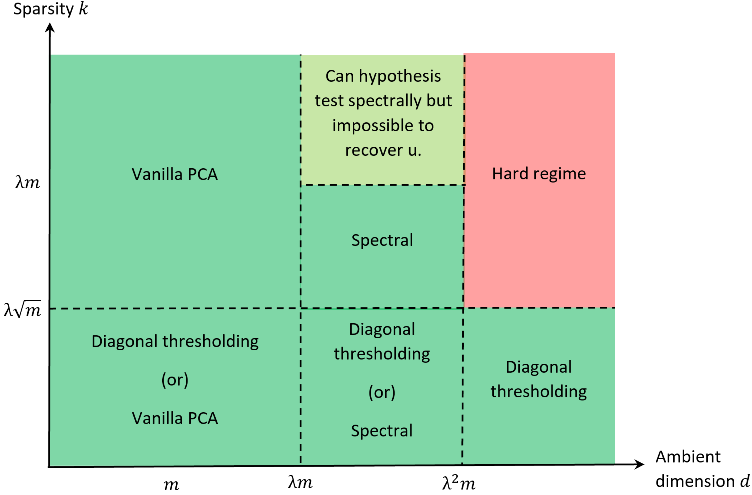

Figure: The computational barrier diagram when $\lambda \geq 1$

Computational barrier diagram for Sparse PCA

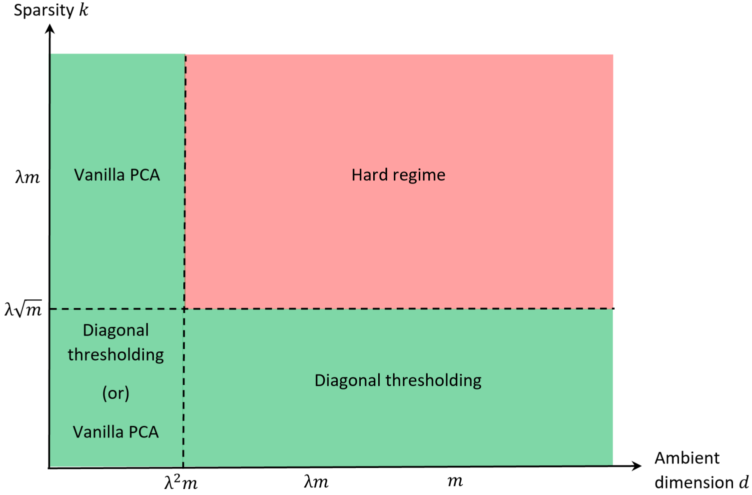

Figure: The computational barrier diagram when $\lambda < 1$

A proof sketch

We proceed as in prior works, via the technique of *pseudo-calibration* and nonlinear random matrix (*graph matrix*) analysis

A proof sketch

We proceed as in prior works, via the technique of *pseudo-calibration* and nonlinear random matrix (*graph matrix*) analysis

Our main innovation lies in the latter part, we introduce and prove a general meta-theorem to show SoS lower bounds

A proof sketch

We proceed as in prior works, via the technique of *pseudo-calibration* and nonlinear random matrix (*graph matrix*) analysis

Our main innovation lies in the latter part, we introduce and prove a general meta-theorem to show SoS lower bounds

We instantiate on our desired applications of Sparse PCA and Tensor PCA

A proof sketch

We proceed as in prior works, via the technique of *pseudo-calibration* and nonlinear random matrix (*graph matrix*) analysis

Our main innovation lies in the latter part, we introduce and prove a general meta-theorem to show SoS lower bounds

We instantiate on our desired applications of Sparse PCA and Tensor PCA

Our work is also a potential step towards the *low-degree likelihood ratio conjecture*

Summary

This work: Almost-tight SoS lower bounds for the fundamental Sparse PCA problem

Summary

This work: Almost-tight SoS lower bounds for the fundamental Sparse PCA problem

Our techniques also work for Tensor PCA, another PCA variant on higher order tensors

Potential future work:

- Analyzing SoS for other statistical problems such as Mixture Modeling and Non-Gaussian Component Analysis
- Low-degree likelihood ratio hypothesis

Summary

This work: Almost-tight SoS lower bounds for the fundamental Sparse PCA problem

Our techniques also work for Tensor PCA, another PCA variant on higher order tensors

Potential future work:

- Analyzing SoS for other statistical problems such as Mixture Modeling and Non-Gaussian Component Analysis
- Low-degree likelihood ratio hypothesis

Thank You

Sanjeev Arora, Satish Rao, and Umesh Vazirani.
Expander flows and a $\sqrt{\log n}$-approximation to sparsest cut.
In *Proceedings of the 36th ACM Symposium on Theory of Computing*, 2004.

Aditya Bhaskara, Moses Charikar, Venkatesan Guruswami, Aravindan Vijayaraghavan, and Yuan Zhou.
Polynomial integrality gaps for strong sdp relaxations of densest k-subgraph.
In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 388–405. SIAM, 2012.

Vijay Bhattiprolu, Mrinalkanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani.
Weak decoupling, polynomial folds and approximate optimization over the sphere.
In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1008–1019. IEEE, 2017.

Vijay Bhattiprolu, Venkatesan Guruswami, and Euiwoong Lee.

Sum-of-squares certificates for maxima of random tensors on the sphere.
arXiv preprint arXiv:1605.00903, 2016.

B. Barak, S. B. Hopkins, J. Kelner, P. Kothari, A. Moitra, and A. Potechin.
A nearly tight sum-of-squares lower bound for the planted clique problem.
In *Proceedings of the 57th IEEE Symposium on Foundations of Computer Science*, pages 428–437, 2016.

Quentin Berthet, Philippe Rigollet, et al.
Optimal detection of sparse principal components in high dimension.
The Annals of Statistics, 41(4):1780–1815, 2013.

Amir Dembo, Andrea Montanari, and Subhabrata Sen.
Extremal cuts of sparse random graphs.
The Annals of Probability, 45(2):1190–1217, 2017.

Uriel Feige and Robert Krauthgamer.
Finding and certifying a large hidden clique in a semirandom graph.
Random Structures & Algorithms, 16(2):195–208, 2000.

Uriel Feige and Gideon Schechtman.
On the optimality of the random hyperplane rounding technique for max cut.
Random Structures & Algorithms, 20(3):403–440, 2002.

Dima Grigoriev.
Complexity of positivstellensatz proofs for the knapsack.
computational complexity, 10(2):139–154, 2001.

Dima Grigoriev.
Linear lower bound on degrees of positivstellensatz calculus proofs for the parity.
Theor. Comput. Sci., 259(1-2):613–622, 2001.

M.X. Goemans and D.P. Williamson.
Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming.
Journal of the ACM, 42(6):1115–1145, 1995.
Preliminary version in *Proc. of STOC'94*.

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.
In *Proceedings of the 58th IEEE Symposium on Foundations of Computer Science*, pages 720–731. IEEE, 2017.

Samuel Brink Klevit Hopkins.
Statistical inference and the sum of squares method.
2018.

Samuel B Hopkins, Jonathan Shi, and David Steurer.
Tensor principal component analysis via sum-of-squares proofs.
In *Conference on Learning Theory*, pages 956–1006, 2015.

Iain M Johnstone and Arthur Yu Lu.
Sparse principal components analysis.
arXiv preprint arXiv:0901.4392, 2009.

Pravesh Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer.
Sum of squares lower bounds for refuting any CSP.

In *Proceedings of the 49th ACM Symposium on Theory of Computing*, 2017.

Robert Krauthgamer, Boaz Nadler, Dan Vilenchik, et al.
Do semidefinite relaxations solve sparse pca up to the information limit?
The Annals of Statistics, 43(3):1300–1322, 2015.

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.

Jean B Lasserre.
Global optimization with polynomials and the problem of moments.
SIAM Journal on optimization, 11(3):796–817, 2001.

Sidhanth Mohanty, Prasad Raghavendra, and Jeff Xu.
Lifting sum-of-squares lower bounds: degree-2 to degree-4.
In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 840–853, 2020.

Tengyu Ma and Avi Wigderson.
Sum-of-squares lower bounds for sparse pca.
In *Advances in Neural Information Processing Systems*, pages 1612–1620, 2015.

Pablo A Parrilo.
Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization.
PhD thesis, California Institute of Technology, 2000.

Aaron Potechin and Goutham Rajendran.
Machinery for proving sum-of-squares lower bounds on certification problems.
arXiv preprint arXiv:2011.04253, 2020.

Prasad Raghavendra.
Optimal algorithms and inapproximability results for every CSP?
In *Proceedings of the 40th ACM Symposium on Theory of Computing*, pages 245–254, 2008.

Emile Richard and Andrea Montanari.

A statistical model for tensor pca.
In *Advances in Neural Information Processing Systems*, pages 2897–2905, 2014.

Ryota Tomioka and Taiji Suzuki.
Spectral norm of random tensors.
arXiv preprint arXiv:1407.1870, 2014.

Madhur Tulsiani.
CSP gaps and reductions in the Lasserre hierarchy.
In *Proceedings of the 41st ACM Symposium on Theory of Computing*, 2009.

Lenka Zdeborová and Stefan Boettcher.
A conjecture on the maximum cut and bisection width in random regular graphs.
Journal of Statistical Mechanics: Theory and Experiment, 2010(02):P02020, 2010.