Low-degree polynomial methods in average-case complexity

organized by
Sam Hopkins - x, Tselil Schramm, and Alex Wein

Workshop Summary

This workshop was devoted to the study of low-degree polynomials as a restricted class of algorithms for high-dimensional statistical problems. This framework has been gaining popularity as a means to rigorously explain and predict statistical-computational tradeoffs. To illustrate the concept of low-degree polynomial (LDP) algorithms, it is instructive to consider the canonical planted clique problem, where the goal is to find a planted k-clique (fully connected subgraph on k vertices), planted in the Erdos-Renyi random graph $G(n, 1/2)$. While it is statistically possible to find a clique of logarithmic (in n) size by brute-force search, the best known polynomial-time algorithms famously require k to be much larger — on the order of $\sqrt{n}$. An even simpler variant of the problem is detection: decide whether a given graph was drawn from the planted clique distribution or simply from $G(n, 1/2)$. A natural approach to this problem is to consider simple statistics such as the total edge count or the number of triangles in the graph. These are LDP algorithms, as they can be represented as low-degree multivariate polynomials (of degree 1 and 3, respectively) in the input variables (in this case, the entries of the adjacency matrix).

The LDP class (with degree logarithmic in n) turns out to be quite powerful, capturing the best known poly-time algorithms for a variety of statistical problems. Results that characterize the limitations of this class of algorithms have by now been widely successful at providing a unifying explanation for suspected "possible-but-hard" regimes in many problems. This workshop aimed to address some of the fundamental questions and current technical challenges present in this area. We summarize below the status of some specific directions that were explored.

Polynomial Threshold Functions

Existing low-degree lower bounds rule out LDP algorithms that meet a specific success criterion based on the first two moments of the polynomial. A direction of ongoing work is to strengthen these lower bounds to rule out other notions of success. One natural goal is to rule out polynomial threshold functions as hypothesis tests, i.e., to prove you cannot distinguish two distributions by thresholding the value of some low-degree polynomial.

The broad open-ended direction that this group worked on was: Let's say you have two distributions P (planted) and N (null) such that $\mathrm{LDLR}_D(P\|N) = o(1)$ for some degree D. What are rigorous implications for concrete algorithms you can derive as a consequence of having small LDLR?

The first basic result we proved in the spirit of the goal is that there is no information-computation gap for distributions on the Boolean hypercube.

(1) Suppose $N = \mathrm{Ber}(p)^n$ for constant p, and P is an S_n-symmetric distribution that has passed through a noisy channel, then $d_{\mathrm{TV}}(P, N) = o(1)$ as long as $D \geq C \log n$ for some large enough constant C depending on the level of noise.
(2) Suppose $N = N(0, Id)^n$, and P has been passed through a noisy channel. Let p be an S_n-symmetric polynomial of degree-$\leq k$ where $k = o(\log n / \log \log n)$, and define the distribution ν as the law of $p(z)$ for $z \sim N$, and π as the law of $p(z)$ for $z \sim P$. Suppose for $D = k^2 \mathrm{polylog} k$, we have $\mathrm{LDLR}_D(P \| N) = o(1/\log n)$, then $d_{\mathrm{TV}}(\nu, \pi) = o(1)$. In fact, this theorem is true for any vector of polynomials of degree-$\leq k$.
(3) Suppose $N = N(0, Id)^{nxn}$, and P has been passed through a noisy channel. Let sigma be any connected subgraph of constant size, and for a matrix M let $\chi_\sigma(M)$ denote the σ subgraph statistic in M. Similarly, define ν and π as the push-forwards of N and P through χ_σ. If for $D \geq \log n \log \log n$, we have $\mathrm{LDLR}_D(P \| N) = o(1)$, then $d_{\mathrm{TV}}(\nu, \pi) = o(1)$.

This group has a manuscript in preparation with these results.

Refuting Colorability

Introduction: For what k can we refute k-colorability efficiently in $G(n, 1/2)$ *via low-degree polynomial tests?* Information-theoretically, a sample from $G(n, 1/2)$ is $\frac{n}{(2 \log n)}(1 + o_n(1))$-colorable. Hence, forgetting about computational efficiency, there is a refutation algorithm for any $k \lesssim \frac{n}{(2 \log n)}(1 + o_n(1))$.

What about efficient algorithms and lower bounds? It is known that there is an efficient refutation algorithm for $k = o(\sqrt{n})$ – if a graph on n vertices is k-colorable, then it has an independent set of size n/k. One can refute the existence of such an independent set in $G(n, 1/2)$ via a simple semidefinite program (or a spectral method, which can be implemented as a low-degree polynomial). It has been recently shown in [potechin25soslb] that this is optimal within constant-degree SoS programs. That is, degree-$O(1)$ Sum-of-Squares (SoS) programs fail to refute k-colorability when $k \gtrsim \sqrt{n}$. Yet, the best known lower-bound for low-degree polynomials is hardness of k-colorability when $k \gg n^{2/3}$. This leaves the regime of $n^{1/2} \ll k \ll n^{2/3}$ open for low-degree polynomial tests.

The goal of this project is to close the gap for low-degree polynomials. Besides the immediate value of resolving the question, there are several important conceptual aspects:

(1) Could there be a low-degree polynomial that surpasses the SoS lower bound? While this would be extremely surprising, it is perhaps not out of the question. Even after years of attempts, a complete understanding of how SoS and LDP algorithms and lower bounds are related is still mostly lacking.
(2) Or, perhaps, there is a LDP lower bound. Can we use the SoS lower bound of [potechin25soslb] to construct it? The notion of LDP lower bounds itself first originiated from psuedo-calibration-based SoS lower bounds. Again, no formal connection is known here.

Formal Problem Statement: Following [pmlr-v195-kothari23a], we formalize the problem as follows.

[Refutation of property $\mathcal{P}$ via LDP] We say that a low-degree polynomial f of degree at most D refutes property $\mathcal{P}$ in $G(n,1/2)$ if the following holds.

$$f(X) \geq 1 \text{ for any graph } X \text{ satisfying } \mathcal{P} \tag{1}$$

$$_{Y\sim G(n,1/2)}[f(Y)] = 0,\ _{Y\sim G(n,1/2)}[f(Y)^2] = o(1). \tag{2}$$

This naturally gives rise to the following startegy for showing a refutation lower-bound.

[LDP Lower Bound refutation of k-colorability [pmlr-v195-kothari23a]] Construct a distribution Q supported on k-colorable graphs such that for any polynomial f of degree at most D for some $D = \omega(\log n)$,

$$|_{X\sim Q}[f(X)] -_{Y\sim G(n,1/2)}[f(Y)]| = o\Big(\sqrt{_{X\sim Q}[f(X)] +_{Y\sim G(n,1/2)}[f(Y)]}\Big)$$

Prior Lower Bound: The lower-bound of $k = n^{2/3}$ is derived in [pmlr-v195-kothari23a] by considering the following distribution supported on k-colorable graphs. It is a stochastic block model on k^2 communities labeled by pairs $(i,j) \in [k] \times [k]$. Now, each vertex $i \in [n]$ receives a uniformly random label (a_i, b_i). Conditioned on the labels, vertices i and j are adjacent with probability:

$$p_{(a_i,b_i),(a_j,b_j)} = \begin{cases} 0, & a_i = a_j \\ 1, & a_i \neq a_j \text{ and } b_i = b_j \\ 1/2, \text{otherwise.} \end{cases}$$

What we managed to show: One natural way to generalize the above construction is as follows. Suppose that is a stochastic block model over ℓ communities parameterized by a probability vector $p \in [0,1]^\ell$ and a probability matrix $Q \in [0,1]^{\ell\times\ell}$. Each of n vertices receives an iid label according to p. Conditioned on the labels x_i, x_j of vertices i and j, they are adjacent with probability Q_{x_i,x_j}. Suppose further that the k communities are split into k groups $L_1, L_2, \ldots, L_k$ and $Q_{x,y} = 0$ for any $i \in [k], x, y \in L_i$. Then, clearly a sample from the is always k-colorable – color class i corresponding to the vertices with label in L_i.

We call such stochastic block models k-color-splittable.

The construction of [pmlr-v195-kothari23a] is k-color-splittable with $L_i = \{(i,b) \ : \ b \in [k]\}$. Unfortunately, we showed that no k-color-splittable construction improves on $k = n^{2/3}$.

For any $k = o(n^{2/3})$ and any k-color splittable stochastic block model Q, it holds that

$$_{X\sim Q}\mathsf{SC}_4[X] -_{Y\sim G(n,1/2)}\mathsf{SC}_4[Y] = \omega(_{Y\sim G(n,1/2)}\mathsf{SC}_4[Y]^{1/2}),$$

where SC_4 is the signed 4-cycle count.

Proof sketch. If Q is k-splittable, then the expected 4-cycle count is $\Omega(n^4k^{-3})$ via spectral method. $_{Y\sim G(n,1/2)}\mathsf{SC}_4[Y] = 0$ and $_{Y\sim G(n,1/2)}\mathsf{SC}_4[Y] = \Theta(n^4)$. □

Some Ideas and Observations: We observed that the fundamental reason why signed 4-cycle counts have such a large advantage for stochastic block models is that repetitions of labels are possible since labels are independent. In particular, the signed 4-cycle count in a graph where each label appears (at most) ones can be negative (while in an SBM, it is always positive, $n^4 \times ((\sqrt{p}Q\sqrt{p})^4)$).

So, perhaps we can even construct a fixed k-colorable graph H such that Q is a uniformly random permutation of H. To circumvent Proposition , the first question is:

Over all fixed graphs H, if Q is a uniformly random permutation of H, find

$$\min_{X\sim Q}[\mathsf{SC}_4(X)].$$

We resolved this problem showing that the minimum is $-n^3(1-o(1))$ and one minimizer is the graph H with adjacency matrix the Hadamard matrix.

So, perhaps some combination of an SBM and Hadamard can lead to a small number of signed 4-cycles and, even more optimistically, small advantage against $G(n,1/2)$.

Testing Versus Estimation

In classical statistics, a useful approach to proving lower bounds (impossibility results) for estimation tasks is to relate them to lower bounds for hypothesis testing tasks. This method has several manifestations, including Le Cam's two-point inequality, Fano's inequality, and Assouad's inequality, which differ in the number and structure of hypotheses in the testing problem to which one attempts to relate an estimation problem.

Our group's goal was to explore whether similar ideas can be implemented with respect to low-degree polynomial algorithms for estimation and testing. Several previous works have used a reduction "outside of" the low-degree framework to make such arguments: they argue in general that an algorithm for estimation (low-degree polynomial or otherwise) would lead to one for testing of comparable runtime, and then prove low-degree lower bounds against the latter. We wanted to make similar claims but "inside of" the low-degree framework, trying to show that some form of low-degree lower bound against testing directly implies low-degree lower bounds against estimation. That would arguably make such reasoning more convincing, in particular leading to concrete lower bounds for estimation, unlike the "outside" version of the argument.

To be formal, we consider a set of parameters $\Theta \subseteq^M$. To each $X \in \Theta$, we associate a probability measure $_X$ over ^{N}. For a probability measure over Θ, we write for the mixture according to $X \sim$ of the $_x$. We write $(X,Y) \sim$ for the pair sampled by drawing $X \sim$ and then $Y \sim_X$. We are generally interested in how testing problems between distributions of the form are related to estimation problems of X from an unknown $_X$ within the low-degree framework. For the sake of simplicity, we mostly focused on the case where the are given by an additive Gaussian model, outputting $X \sim$ plus Gaussian noise.

The following two ways of measuring how well low-degree polynomials can perform estimation both seem reasonable:

$$\begin{aligned} {}_{\le D}(g,(_X),) &\min_{f\in[Y]^N_{\le D}} {}_{(X,Y)\sim}\|X-f(Y)\|^2 \\ &= \min_{f\in[Y]^N_{\le D}} {}_{X\sim}\,{}_{Y\sim_X}\|X-f(Y)\|^2, \\ {}_{\le D}(g,(_X)) &\min_{f\in[Y]^N_{\le D}} \max_{X\in\Theta} {}_{Y\sim_X}\|X-f(Y)\|^2. \end{aligned}$$

Note that is a "Bayesian quantity" in that it depends on a prior distribution over Θ, which is related to the usual low-degree correlation. On the other hand, is a "minimax quantity" replacing the expectation over $X \in$ with considering the worst-case choice of X. The latter has not been studied as much, but seems better adapted to our question, allowing us to more easily imitate reasoning about minimax error from statistics, so we focused on that.

We successfully related to both the low-degree advantage between a "two-point" pair of distributions $_{X_1}$,$_{X_2}$ and a "point-mixture" pair $_{X_1}$,. However, neither of these results gives

strong lower bounds against estimation in natural models like spiked matrix models. The issue seems to be that two-point and point-mixture testing problems are not hard enough. For instance, testing between two instances of spiked matrix model where one model's prior is just on a single spike direction (which is known to the testing algorithm) is too easy to give useful information about the difficulty of estimating an unkonwn spike in general, since the testing algorithm can probe in the direction of that spike. In the two-point setting this issue is particularly severe; we did find that the point-mixture argument can give some soft but non-trivial information about the scaling of the best possible correlation achievable by low-degree polynomial estimators in some spiked matrix models. However, both types of argument are far from, say, characterizing the precise best possible correlation or the weak recovery threshold sharply with correct constants.

It seems that comparing two mixtures $_1$,$_2$ would be more appropriate, but we encountered technical issues in trying to carry out such an argument. We also explored some other alternative approaches, including looking for a useful direct definition of multiple hypothesis testing within the low-degree framework (which perhaps could replace using mixture distributions in the above setup), and in arguing more directly that a good low-degree estimator can be used to construct a low-degree test by, say, plugging it into a low-degree proxy for the likelihood function (that is, the likelihood of X given the observation Y, not to be confused with the low-degree likelihood ratio of two distributions).

Lower Bounds for Null Optimization

The low-degree polynomial framework was developed in the context of *planted* recovery poblems, where the goal is to recover a planted signal obscured by random noise. Many such problems can be reformulated as optimization problems with a random input. In the *null* version, the goal is instead to optimize an objective over pure noise, in the absence of a planted signal.

The presence of information-computation gaps persists in null optimization problems, and following the influential work of Gamarnik, Jagganath, and Wein in 2020 [GJW-ld], we can substantiate many of these information-computation gaps with low-degree lower bounds.

However, it is not as clear whether low-degree lower bounds are as convincing for null problems as they are for planted problems. Partly this is because the solutions to null optimization problems are not noise-stable, so there is not a principled reason to take low-degree polynomials as a canonical algorithm in the first place. But even further, we do not know how to simulate as many null optimization algorithms with low-degree polynomials. Local algorithms in sparse graphs can be simulated by low-degree polynomials, as can $O(1)$ iterations of AMP, spectral algorithms, or gradient descent. But (n) steps of iterative algorithms are not known to be captured by low-degree polynomials, and neither are greedy algorithms for maximum clique or independent set. In fact, it seems possible, or even likely, that low-degree polynomials cannot simulate some of these algorithms in the easy regime.

During the workshop, we focused on the specific example of finding cliques in $G(n, 1/2)$. The state-of-the-art polynomial time algorithm is a greedy algorithm, which iteratively adds vertices to the clique until no more can be added. A simple analysis proves that this algorithm succeeds in finding a clique of size $(1 - o(1)) \log n$ with high probability.

Prove that no polynomial of degree $o(\log^2 n)$ can find a clique of size $0.9 \log n$ in G(n,1/2).

During the workshop, we were able to prove a significantly weaker version of the statement above, but which stops short of conclusively resolving the question. Using noise stability of low-degree functions, we could show that when restricting to functions $f : \{0,1\}^{\binom{n}{2}} \to^n$ which are (1) **symmetric**, in the sense that for any permutation $\pi \in S_n$, $f(\pi(G)) = \pi(f(G))$, and (2) **of fixed degree**,

$$\left[\max_{S \text{ a } k \text{ clique in } G} \langle 1_S, f(G)\rangle\right] \leq 0.9\sqrt{k}[\|f\|_2].$$

For $k = 0.9 \log n$.

This statement shows that, on the one hand, no such f can be well-correlated with a clique in G. However, the statement falls short of answering the problem conclusively for two reasons: firstly, and most simply, it only rules out degree $O(1)$ polynomials. But perhaps more problematically, one would like to allow f to break symmetry, either by being asymmetric to start or by incorporating a *rounding* step. The reason is that there are so many cliques of size k in a typical $G(n, 1/2)$ that symmetry breaking would seem to be an essential part of such an algorithm. But our argument seems inherently tied to the symmetry of f.

MMSE for AMP

In many high-dimensional inference problems, state-of-the-art efficient partial recovery is achieved by AMP algorithms. Our group aimed to show that AMP has the optimal mean squared error among the class of low-degree polynomials.

A prototypical inference problems for which AMP is believed to be optimal is the rank-1 estimation problem: given $Y = \lambda XX^T + W$, where $X \in \mathbb{R}^n$ has i.i.d. coordinates drawn from a distribution μ, and W is a GOE matrix. For λ greater than a threshold, it is known that for any $\varepsilon > 0$, there exists $\rho(\lambda) > 0$ and an AMP algorithm $\mathcal{A} : \mathbb{R}^{n\times n} \to \mathbb{R}^n$, such that w.h.p. as $n \to \infty$,

$$\frac{\langle X, \mathcal{A}(Y)\rangle}{\|X\|\|\mathcal{A}(Y)\|} \geq \rho(\lambda) - \varepsilon\,.$$

To that show AMP is LDP-optimal, we wish to prove the following conjecture: For any $\varepsilon > 0$, there exists a sequence of large integers D_n such that

$$\sup_f \frac{\mathbb{E}[\langle X, f(Y)\rangle]}{\sqrt{\mathbb{E}[X^2]\mathbb{E}[f(Y)^2]}} \leq \rho(\lambda) + \varepsilon\,, \tag{1}$$

where the supremum is taken over $f : \mathbb{R}^{n\times n} \to \mathbb{R}$ such that each coordinate is a polynomial that has degree no more than D_n. Here in the conjecture, for large integers D_n we mean at least $D_n \to \infty$ as $n \to \infty$, and ideally $D_n = \omega(\log n)$ or even $D_n = n^{1-o(1)}$.

Part of Conjecture has been resolved by Montanari and Wein in [MW-amp], where the authors establish the above result for a diverging sequence of D_n, under the assumption that μ has non-zero mean. Thus, our main focus was to either improve the bound for D_n to $\omega(\log n)$ or prove anything meaningful for the case where μ has zero mean.

Our group had some thoughts on these directions. We noted that in a recent work [sharp-est], Sohn and Wein introduced a new constructive method for upper-bounding (1). Roughly speaking, instead of upper-bounding the ratio in (1) for all polynomials, one only needs to find a "dual certificate" that (1) holds. Towards this goal, it seems to require a

good guess on the correct form of the dual certificate, and then construct it carefully by hand.

Our group agreed that, to make an insightful guess, a good understanding of the optimal polynomial f for the ratio would be helpful. On the other hand, building on the belief that AMP achieves low-degree optimality, we discussed an approach to *explicitly* calculate the optimal polynomial f. The strategy is as follows: First, recall that the AMP algorithm has the following form

$$m_0 \in \mathbb{R}^n \text{ is an initialization}, \quad m_{k+1} = Y f_k(m_k) - b_k m_{k-1},\ k = 1, 2, \dots,$$

where $f_k : \mathbb{R} \to \mathbb{R}$ are carefully chosen functions, b_k are appropriate constants associated with f_k and the output of AMP will be m_K for a large constant K. We have explicit expressions for the functions f_k of the optimal AMP, and we can approximate each f_k using a polynomial $\hat{f}_k$ by, e.g., Taylor expansion. Then we consider the algorithm

$$m_0 \in \mathbb{R}^n \text{ is an initialization}, \quad m_{k+1} = Y \hat{f}_k(m_k) - \hat{b}_k m_{k-1},\ k = 1, 2, \dots,$$

with output m_K (here $\hat{b}_k$ are appropriate constants). Note that now $m_K = f(Y)$ for some polynomial f. In principle, by taking good polynomial approximators $\hat{f}_k$, we get a polynomial that almost achieves optimality of the ratio in (1).

The strategy described above may seem daunting due to the complexity of the iteration process. However, our group made several observations that help simplify the computation. First, there is a strong symmetry among the variables, which means we do not need to track the coefficients for every term. Additionally, by performing the expansion, one can quickly observe that the majority of terms correspond to "tree-polynomials" introduced in [MW-amp], which we believe make the predominant contribution. Combining these two points, we concluded that it is sufficient to track only the coefficients of the tree-polynomials. Finally, by utilizing the form of the AMP iterate, we can derive fixed-point equations and recursive formulas for the coefficients of the tree-polynomials, making all computations explicit.

Our goal was to gain insights into the correct dual certificate through these calculations. While many challenges remain to be addressed, we believe that the approach outlined above represents a crucial first step that captures the nature of how AMP achieves the optimality of low-degree polynomials.