

# Low-Degree Hardness of Random Optimization Problems

David Gamarnik\*, Aukosh Jagannath<sup>†</sup> and Alexander S. Wein<sup>‡</sup>

\*Sloan School of Management and Operations Research Center

Massachusetts Institute of Technology

Email: gamarnik@mit.edu

<sup>†</sup>Department of Statistics and Actuarial Science and Department of Applied Mathematics

University of Waterloo

Email: a.jagannath@uwaterloo.ca

<sup>‡</sup>Department of Mathematics, Courant Institute of Mathematical Sciences

New York University

Email: awein@cims.nyu.edu

**Abstract**—We consider the problem of finding nearly optimal solutions of optimization problems with random objective functions. Such problems arise widely in the theory of random graphs, theoretical computer science, and statistical physics. Two concrete problems we consider are (a) optimizing the Hamiltonian of a spherical or Ising  $p$ -spin glass model, and (b) finding a large independent set in a sparse Erdos-Renyi graph. Two families of algorithms are considered: (a) low-degree polynomials of the input—a general framework that captures methods such as approximate message passing and local algorithms on sparse graphs, among others; and (b) the Langevin dynamics algorithm, a canonical Monte Carlo analogue of the gradient descent algorithm (applicable only for the spherical  $p$ -spin glass Hamiltonian).

We show that neither family of algorithms can produce nearly optimal solutions with high probability. Our proof uses the fact that both models are known to exhibit a variant of the overlap gap property (OGP) of near-optimal solutions. Specifically, for both models, every two solutions whose objective values are above a certain threshold are either close or far from each other. The crux of our proof is the stability of both algorithms: a small perturbation of the input induces a small perturbation of the output. By an interpolation argument, such a stable algorithm cannot overcome the OGP barrier.

The stability of the Langevin dynamics is an immediate consequence of the well-posedness of stochastic differential equations. The stability of low-degree polynomials is established using concepts from Gaussian and Boolean Fourier analysis, including noise sensitivity, hypercontractivity, and total influence.

**Keywords**—optimization; overlap gap property; low-degree polynomials; Langevin dynamics;  $p$ -spin glass; independent set

## I. INTRODUCTION

In this paper, we study the problem of producing near-optimal solutions of random optimization problems by polynomials of low degree in the input data. We prove that no low-degree polynomial can succeed at achieving a certain objective value in two optimization problems: (a) optimizing the Hamiltonian of the (spherical or Ising)  $p$ -spin glass model, and (b) finding a large independent set in a sparse Erdős-Rényi graph, with high probability in the realization

of the problem. We rule out polynomials of degree as large as  $cn$  for the  $p$ -spin glass models and as large as  $cn/\log n$  for the independent set problem for some constant  $c$ , provided the algorithm is assumed to succeed modulo exponentially small in  $n$  probability, where  $n$  is the problem dimension. More generally, we provide a tradeoff between the degree of polynomials that we rule out and the success probability assumed. For the spherical  $p$ -spin model, we also give a lower bound against Langevin dynamics.

Our motivation for focusing on “low-degree” approximations is two-fold. Firstly, from an approximation theory perspective, producing near-optimal solutions by a polynomial in the input is very natural. Indeed, in many problems of interest the best known polynomial-time algorithms can be placed within the family of low-degree methods. For example, in the settings we consider here, the best known polynomial-time optimization results can be captured by the approximate message passing (AMP) framework [1], [2] (for the  $p$ -spin) and by the class of local algorithms on sparse graphs [3] (for the independent set problem), respectively. Both of these families of algorithms are captured by constant-degree polynomials; see the full version of this paper [4] for more details. For spherical  $p$ -spin glass models, earlier work of [5] introduced an algorithm which performs as well as AMP; we expect this algorithm to also fall into the family of low-degree methods, but verifying this is less clear. Secondly, a recent line of work [6], [7], [8], [9] on the *sum-of-squares hierarchy* has produced compelling evidence that the power of low-degree polynomials is a good proxy for the intrinsic computational complexity of a broad class of *hypothesis testing* problems. Below, we briefly review this theory of low-degree polynomials in hypothesis testing.

The low-degree framework was initiated in [7], [8], [9] to study computational hardness in hypothesis testing problems. Specifically, this line of work has focused on high-dimensional testing problems where the goal is to determine whether a given sample (e.g., an  $n$ -vertex graph) was drawn from the “null” distribution  $\mathbb{Q}_n$  (e.g., the Erdős-Rényi

model) or the “planted” distribution  $\mathbb{P}_n$  (e.g., a random graph with planted structure such as a large clique or a small cut). Through an explicit and relatively straightforward calculation, one can determine whether there exists a (multivariate) polynomial  $f$  (in the entries of the observed sample) of a given degree  $D = D(n)$  that can distinguish  $\mathbb{P}_n$  from  $\mathbb{Q}_n$  (in a particular sense) [7], [8], [9]. A conjecture of Hopkins [9] (inspired by [6], [7], [8]) postulates that for “natural” high-dimensional testing problems, if there is a polynomial-time algorithm to distinguish  $\mathbb{P}_n, \mathbb{Q}_n$  (with error probability  $o(1)$ ) then there is also an  $O(\log n)$ -degree polynomial that can distinguish  $\mathbb{P}_n, \mathbb{Q}_n$ . One justification for this conjecture is its deep connection with the *sum-of-squares (SoS) hierarchy*—a powerful class of meta-algorithms—and in particular the *pseudo-calibration* approach [6], which suggests that low-degree polynomials are as powerful as any SoS algorithm (see [8], [9], [10] for details). Another justification for the conjecture is that  $O(\log n)$ -degree polynomials can capture a very broad class of spectral methods (see [11, Theorem 4.4] for specifics), which in turn capture the best known algorithms for many high-dimensional testing problems (e.g., [12], [13], [8]). For many classical statistical tasks—planted clique, sparse PCA, community detection, tensor PCA, etc.—it has indeed been verified that  $O(\log n)$ -degree polynomials succeed (at testing) in the same parameter regime as the best known polynomial-time algorithms (e.g., [7], [8], [9], [14], [11], [15]). (Ofentimes, the hypothesis testing variants of these types of problems seem to be equally hard as the more standard task of recovering the planted signal.) Lower bounds against low-degree polynomials are one concrete form of evidence that the existing algorithms for these problems cannot be improved (at least without drastically new algorithmic techniques). For more details on the low-degree framework for hypothesis testing, we refer the reader to [9], [11].

One goal of the current work is to extend the low-degree framework to the setting of random optimization problems. This includes defining what it means for a low-degree polynomial to succeed at an optimization task, and giving techniques by which one can prove lower bounds against all low-degree polynomials. One difference between the optimization and testing settings is that many existing optimization algorithms can be represented as constant-degree polynomials (see the full version [4]), instead of the  $O(\log n)$ -degree required in the testing case. A substantial difficulty that we face in the optimization setting is that, in contrast to the testing setting, it does not seem possible to prove lower bounds against low-degree polynomials via a straightforward explicit calculation. To overcome this, our proofs take a more indirect route and leverage a certain structural property—the *overlap gap property (OGP)*—of the optimization landscape, combined with stability properties of low-degree polynomials. We also use similar techniques to give lower bounds against Langevin dynamics, a canonical

Monte Carlo analogue of gradient descent; while this is not a low-degree polynomial (due to its continuous-time nature), it is similar in spirit and has similar stability properties.

While the OGP has been used to rule out various classes of other algorithms previously (see below), its usage in our current setting presents some substantial technical difficulties which we need to overcome. Roughly speaking, the property states that for every pair of nearly-optimal solutions  $x_1$  and  $x_2$ , their normalized overlap (normalized inner product) measured with respect to the ambient Hilbert space must lie in a disjoint union of intervals  $[0, \nu_1] \cup [\nu_2, 1]$ . This property extends to the case of families of instances as well in the sense that even if one considers a natural interpolation between two independent instances of the problem, for every two members of the interpolated family and every pair of solutions  $x_1, x_2$  which are near optimizers for these two members, respectively, it is still the case that the overlap of  $x_1$  and  $x_2$  belongs to  $[0, \nu_1] \cup [\nu_2, 1]$ . The main idea of the proof from OGP is based on the contradiction argument. If the result of the algorithm is known to be stable then, denoting by  $x(t)$  the result of the algorithm corresponding to the interpolation step  $t$ , it should be the case that the overlap between  $x(0)$  and  $x(t)$  changes “continuously”. At the same time we show separately that the starting solution  $x(0)$  and terminal solution  $x(1)$  have an overlap at most  $\nu_1$ , and thus at some point the overlap between  $x(0)$  and  $x(t)$  belongs to  $(\nu_1, \nu_2)$ , which is a contradiction.

Establishing stability for low-degree polynomials and Langevin dynamics is quite non-trivial and constitutes the key technical contribution of the paper. For the case of polynomials, these stability results harness results from Gaussian and Boolean Fourier analysis. We prove two separate variants of this stability result, depending on whether the random input is Gaussian- or Bernoulli-distributed. A key technical result in the Gaussian case is Theorem 3.1 in the full version [4] which informally states that if we have two  $\rho$ -correlated random instances  $X$  and  $Y$  of a random tensor, and  $f$  is a vector-valued low-degree polynomial defined on such tensors, then the distance  $\|f(X) - f(Y)\|_2$  is unlikely to exceed a certain value which depends continuously on  $\rho$ . In particular this distance is small when  $\rho \approx 1$ . Proving this result relies on a well-known consequence of hypercontractivity for low-degree polynomials, and basic properties of Hermite polynomials (the orthogonal polynomials of the Gaussian measure). In the case of Bernoulli-distributed inputs, we prove a related stability result (Theorem 4.2 in the full version [4]) which shows that when the input variables are resampled one at a time, the output of a vector-valued low-degree polynomial will never change significantly in one step, with nontrivial probability. The proof involves the notion of total influence from Boolean analysis, as well as a direct proof by induction on the dimension. The proof of stability for Langevin dynamics is based on the continuous dependence of stochastic differential equations

on their coefficients.

The OGP emerged for the first time in the context of spin glass theory and random constraint satisfaction problems. It was first proven implicitly in [16], [17], and [18]. These papers established that the set of satisfying assignments of a random K-SAT formula partitions into clusters above a certain clause-to-variables density. This was postulated as evidence of algorithmic hardness of finding satisfying assignments for such densities. Implicitly, the proof reveals that the overlaps of satisfying assignments exhibit the OGP, and clustering is inferred from this. It is worth noting that while OGP implies the existence of clusters, the converse is not necessarily the case, as one can easily construct a clustered space of solutions with overlaps spanning the entire interval  $[0, 1]$ . A direct algorithmic implication of the OGP was shown for the first time in [19], where OGP was proven to be a barrier for local algorithms—defined as the so-called *factors of i.i.d. (FIID)*—designed to find large independent sets in sparse Erdős-Rényi graphs. The OGP was used to show that, asymptotically, these algorithms cannot find independent sets larger than a multiplicative factor  $1/2 + 1/(2\sqrt{2}) \approx 0.85$  of optimal. The present paper recovers this result as a special case, since (as we discuss in the full version [4]) local algorithms can be captured by constant-degree polynomials. The lower bound against local algorithms was improved by [20] to a multiplicative factor of  $1/2$ . This is the best possible since  $1/2$ -optimal independent sets can be found by local algorithms; more precisely, this was shown in [3] for the case of random regular graphs, but a similar result is expected to hold for sparse Erdős-Rényi graphs as well (although we are not aware of any literature formally verifying this). It is not clear how to improve the multiplicative factor in the lower bound to  $1/2$  for low-degree polynomials, as [20] uses a more sophisticated variant of OGP than we use here. Several subsequent papers used OGP to rule out various classes of algorithms, including local algorithms for finding large cuts in random hypergraphs [21], random walk-based algorithms (WALKSAT) [22], and AMP-type algorithms for optimizing the Hamiltonian of the Ising  $p$ -spin model [23]. The current work draws inspiration from a key idea in [21], [23], namely that a particular variant of OGP—the same variant that we use in the current work—implies failure of any sufficiently “stable” algorithm.

We emphasize that the class of algorithms ruled out by the lower bounds in this paper (namely, low-degree polynomials) not only captures existing methods such as AMP and local algorithms, but contains a strictly larger (in a substantial way) class of algorithms than prior work on random optimization problems. We now illustrate this claim in the setting of the  $p$ -spin optimization problem. The best known polynomial-time algorithms for optimizing the  $p$ -spin Hamiltonian are captured by the AMP framework [1], [2]. Roughly speaking, AMP algorithms combine a linear

update step (tensor power iteration) with entry-wise nonlinear operations. For a fairly general class of  $p$ -spin optimization problems (including spherical and Ising mixed  $p$ -spin models), it is now known precisely what objective value can be reached by the best possible AMP algorithm [2]. While this may seem like the end of the story, we point out that for the related *tensor PCA* problem—which is a variant of the  $p$ -spin model with a planted rank-1 signal—AMP is known to be substantially sub-optimal compared to other polynomial-time algorithms [24]. None of the best known polynomial-time algorithms [24], [12], [13], [25], [26], [27] use the tensor power iteration step as in AMP, and there is evidence that this is fundamental [28]; instead, the optimal algorithms include spectral methods derived from different tensor operations such as *tensor unfolding* [24], [12] (which can be interpreted as a higher-order “lifting” of AMP [25]). These spectral methods are captured by  $O(\log n)$ -degree polynomials. With this in mind, we should *a priori* be concerned that AMP might also be sub-optimal for the (non-planted)  $p$ -spin optimization problem. This highlights the need for lower bounds that rule out not just AMP, but all low-degree polynomial algorithms. While the lower bounds in this paper do not achieve the precise optimal thresholds for objective value, they rule out quite a large class of algorithms compared to existing lower bounds for random optimization problems.

We refer the reader to the full version of this paper [4] for a more detailed discussion of how various optimization algorithms can be approximated by low-degree polynomials.

*Notation:* We use  $\|\cdot\|_2$  and  $\langle \cdot, \cdot \rangle$  to denote the standard  $\ell^2$  norm and inner product of vectors. We also use the same notation to denote the Frobenius norm and inner product of tensors. We use the term *polynomial* both to refer to (multivariate) polynomials  $\mathbb{R}^m \rightarrow \mathbb{R}$  in the usual sense, and to refer to vector-valued polynomials  $\mathbb{R}^m \rightarrow \mathbb{R}^n$  defined as in (3). We abuse notation and use the term *degree- $D$  polynomial* to mean a polynomial of degree at most  $D$ . A *random polynomial* has possibly-random coefficients, as defined in Section II-A1. We use  $A^c$  to denote the complement of an event  $A$ . Unless stated otherwise, asymptotic notation such as  $o(1)$  or  $\Omega(n)$  refers to the limit  $n \rightarrow \infty$  with all other parameters held fixed. In other words, this notation may hide constant factors depending on other parameters such as the degree  $d$  in the independent set problem.

## II. MAIN RESULTS

### A. Optimizing the $p$ -Spin Glass Hamiltonian

The first class of problems we consider here is optimization of the (pure)  $p$ -spin glass Hamiltonian, defined as follows. Fix an integer  $p \geq 2$  and let  $Y \in (\mathbb{R}^n)^{\otimes p}$  be a  $p$ -tensor with real coefficients. For  $x \in \mathbb{R}^n$ , consider the objective function

$$H_n(x; Y) = \frac{1}{n^{(p+1)/2}} \langle Y, x^{\otimes p} \rangle. \quad (1)$$

Note that all homogeneous polynomials of degree  $p$  (in the variables  $x$ ) can be written in this form for some  $Y$ . We focus on the case of a random coefficient tensor  $Y$ . In this setting, the function  $H_n$  is sometimes called the Hamiltonian for a  $p$ -spin glass model in the statistical physics literature. More precisely, for various choices of a (compact) domain  $\mathcal{X}_n \subset \mathbb{R}^n$ , we are interested in approximately solving the optimization problem

$$\max_{x \in \mathcal{X}_n} H_n(x; Y) \quad (2)$$

given a random realization of the coefficient tensor  $Y$  with i.i.d  $\mathcal{N}(0, 1)$  entries. Here and in the following we let  $\mathbb{P}_Y$  denote the law of  $Y$ . (When it is clear from context we omit the subscript  $Y$ .)

We begin first with a simple norm constraint, namely, we will take as domain  $\mathcal{S}_n = \{x \in \mathbb{R}^n : \|x\|_2 = \sqrt{n}\}$ , the sphere in  $\mathbb{R}^n$  of radius  $\sqrt{n}$ . We then turn to understanding a binary constraint, namely where the domain is the discrete hypercube  $\Sigma_n = \{+1, -1\}^n$ . Following the statistical physics literature, in the former setting, we call the objective the *spherical  $p$ -spin glass Hamiltonian* and the latter setting the *Ising  $p$ -spin glass Hamiltonian*.

In both settings, quite a lot is known about the maximum. It can be shown [29], [30] that the maximum value of  $H_n$  has an almost sure limit (as  $n \rightarrow \infty$  with  $p$  fixed), called the *ground state energy*, which we will denote by  $E_p(\mathcal{S})$  for the spherical setting and  $E_p(\Sigma)$  for the Ising setting. Explicit variational formulas are known for  $E_p(\mathcal{S})$  [31], [30], [32] and  $E_p(\Sigma)$  [33], [34].

Algorithmically, it is known how to find, in polynomial time, a solution of value  $E_p^\infty(\mathcal{S}) - \varepsilon$  or  $E_p^\infty(\Sigma_n) - \varepsilon$  (respectively for the spherical and Ising settings) for any constant  $\varepsilon > 0$  [5], [1], [2]. In both the spherical and Ising settings, these constants satisfy  $E_2^\infty = E_2$  and  $E_p^\infty < E_p$  for  $p \geq 3$ . In other words, it is known how to efficiently optimize arbitrarily close to the optimal value in the  $p = 2$  case, but not when  $p \geq 3$ .

1) *Low-Degree Polynomial Algorithms:* Our goal here is to understand how well one can optimize (2) via the output of a vector-valued low-degree polynomial in the coefficients  $Y$ . To simplify notation we will often abuse notation and refer to the space of  $p$ -tensors on  $\mathbb{R}^n$  by  $\mathbb{R}^m \cong (\mathbb{R}^n)^{\otimes p}$  where  $m = n^p$ .

We say that a function  $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$  is a polynomial of degree (at most)  $D$  if it may be written in the form

$$f(Y) = (f_1(Y), \dots, f_n(Y)), \quad (3)$$

where each  $f_i : \mathbb{R}^m \rightarrow \mathbb{R}$  is a polynomial of degree at most  $D$ .

We will also consider the case where  $f$  is allowed to have random coefficients, provided that these coefficients are independent of  $Y$ . That is, we will assume that there is some probability space  $(\Omega, \mathbb{P}_\omega)$  and that  $f : \mathbb{R}^m \times \Omega \rightarrow \mathbb{R}^n$

is such that  $f(\cdot, \omega)$  is a polynomial of degree at most  $D$  for each  $\omega \in \Omega$ . We will abuse notation and refer to this as a *random polynomial*  $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ .

Our precise notion of what it means for a polynomial to optimize  $H_n$  will depend somewhat on the domain  $\mathcal{X}_n$ . This is because it is too much to ask for the polynomial's output to lie in  $\mathcal{X}_n$  exactly, and so we fix a canonical rounding scheme that maps the polynomial's output to  $\mathcal{X}_n$ . We begin by defining this notion for the sphere:  $\mathcal{X}_n = \mathcal{S}_n$ .

*The spherical case:* We will round a polynomial's output to the sphere  $\mathcal{S}_n$  by normalizing it in the standard way. To this end, for a random polynomial  $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$  we define the random function  $g_f : \mathbb{R}^m \rightarrow \mathcal{S}_n \cup \{\infty\}$  by

$$g_f(Y, \omega) = \sqrt{n} \frac{f(Y, \omega)}{\|f(Y, \omega)\|_2},$$

with the convention  $g_f(Y, \omega) = \infty$  if  $f(Y, \omega) = 0$ .

**Definition II.1.** For parameters  $\mu \in \mathbb{R}$ ,  $\delta \in [0, 1]$ ,  $\gamma \in [0, 1]$ , and a random polynomial  $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ , we say that  $f$   $(\mu, \delta, \gamma)$ -optimizes the objective (1) on  $\mathcal{S}_n$  if the following are satisfied when  $(Y, \omega) \sim \mathbb{P}_Y \otimes \mathbb{P}_\omega$ :

- $\mathbb{E}_{Y, \omega} \|f(Y, \omega)\|_2^2 = n$  (normalization).
- With probability at least  $1 - \delta$  over  $Y$  and  $\omega$ , we have both  $H_n(g_f(Y, \omega); Y) \geq \mu$  and  $\|f(Y, \omega)\|_2 \geq \gamma\sqrt{n}$ .

Implicitly in this definition, the case  $f(Y, \omega) = 0$  must occur with probability at most  $\delta$ . The meaning of the parameters  $(\mu, \delta, \gamma)$  is as follows:  $\mu$  is the objective value attained after normalizing the polynomial's output to the sphere, and  $\delta$  is the algorithm's failure probability. Finally,  $\gamma$  is involved in the norm bound  $\|f(Y, \omega)\|_2 \geq \gamma\sqrt{n}$  that we need for technical reasons. Since the domain is  $\mathcal{S}_n$ ,  $f$  is "supposed to" output a vector of norm  $\sqrt{n}$ . While we do not require this to hold exactly (and have corrected for this by normalizing  $f$ 's output), we do need to require that  $f$  usually does not output a vector of norm too much smaller than  $\sqrt{n}$ . This norm bound is important for our proofs because it ensures that a small change in  $f(Y, \omega)$  can only induce a small change in  $g_f(Y, \omega)$ .

We now state our main result on low-degree hardness of the spherical  $p$ -spin model, with the proof deferred to the full version [4].

**Theorem II.2.** For any even integer  $p \geq 4$  there exist constants  $\mu < E_p(\mathcal{S})$ ,  $n^* \in \mathbb{N}$ , and  $\delta^* > 0$  such that the following holds. For any  $n \geq n^*$ , any  $D \in \mathbb{N}$ , any  $\delta \leq \min\{\delta^*, \frac{1}{4} \exp(-2D)\}$ , and any  $\gamma \geq (2/3)^D$ , there is no random degree- $D$  polynomial that  $(\mu, \delta, \gamma)$ -optimizes (1) on  $\mathcal{S}_n$ .

A number of remarks are in order. First, this result exhibits a tradeoff between the degree  $D$  of polynomials that we can rule out and the failure probability  $\delta$  that we need to assume. In order to rule out polynomials of any constant degree,

we need only the mild assumption  $\delta = o(1)$ . On the other hand, if we are willing to restrict to algorithms of failure probability  $\delta = \exp(-cn)$  (which we believe is reasonable to expect in this setting), we can rule out all polynomials of degree  $D \leq c'n$  for a constant  $c' = c'(c)$ . It has been observed in various hypothesis testing problems that the class of degree- $n^\delta$  polynomials is at least as powerful as all known  $\exp(n^{\delta-o(1)})$ -time algorithms [9], [11], [15]. This suggests that optimizing arbitrarily close to the optimal value in the spherical  $p$ -spin (for  $p \geq 4$  even) requires fully exponential time  $\exp(n^{1-o(1)})$ .

The best known results for polynomial-time optimization of the spherical  $p$ -spin were first proved by [5] but can also be recovered via the AMP framework of [2]. As discussed in the full version [4], these AMP algorithms can be captured by constant-degree polynomials. Furthermore, the output of such an algorithm concentrates tightly around  $\sqrt{n}$  and thus easily satisfies the norm bound with  $\gamma = (2/3)^D$  required by our result. We also expect that these AMP algorithms have failure probability  $\delta = \exp(-\Omega(n))$ ; while this has not been established formally, a similar result on concentration of AMP-type algorithms has been shown in [23].

Our results are limited to the case where  $p \geq 4$  is even and  $\mu$  is a constant slightly smaller than the optimal value  $E_p(\mathcal{S})$ . These restrictions are in place because the OGP property used in our proof is only known to hold for these values of  $p$  and  $\mu$ . If the OGP were proven for other values of  $p$  or for a lower threshold  $\mu$ , our results would immediately extend to give low-degree hardness for these parameters (see Theorem 3.6 in the full version [4]). Note that we cannot hope for the result to hold when  $p = 2$  because this is a simple eigenvector problem with no computational hardness: there is a constant-degree algorithm to optimize arbitrarily close to the maximum (see the full version [4]).

*The Ising case:* We now turn to low-degree hardness in the Ising setting, where the domain is the hypercube:  $\mathcal{X}_n = \Sigma_n$ . In this case, we round a polynomial's output to the hypercube by applying the sign function. For  $x \in \mathbb{R}$ , let

$$\text{sgn}(x) = \begin{cases} +1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0, \end{cases}$$

and for a vector  $x \in \mathbb{R}^n$  let  $\text{sgn}(x)$  denote entry-wise application of  $\text{sgn}(\cdot)$ . We now define our notion of near optimality for a low-degree polynomial.

**Definition II.3.** *For parameters  $\mu \in \mathbb{R}$ ,  $\delta \in [0, 1]$ ,  $\gamma \in [0, 1]$ ,  $\eta \in [0, 1]$ , and a random polynomial  $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ , we say that  $f$   $(\mu, \delta, \gamma, \eta)$ -optimizes the objective (1) on  $\Sigma_n$  if the following are satisfied.*

- $\mathbb{E}_{Y, \omega} \|f(Y, \omega)\|_2^2 = n$  (normalization).
- With probability at least  $1 - \delta$  over  $Y$  and  $\omega$ , we have both  $H_n(\text{sgn}(f(Y, \omega)); Y) \geq \mu$  and  $|\{i \in [n] : |f_i(Y, \omega)| \geq \gamma\}| \geq (1 - \eta)n$ .

The interpretation of these parameters is similar to the spherical case, with the addition of  $\eta$  to take into account issues related to rounding. More precisely, as in the spherical case,  $\mu$  is the objective value attained after rounding the polynomial's output to the hypercube, and  $\delta$  is the failure probability. The parameters  $\gamma, \eta$  are involved in an additional technical condition, which requires  $f$ 's output not to be too “small” in a particular sense. Specifically, all but an  $\eta$ -fraction of the coordinates of  $f$ 's output must exceed  $\gamma$  in magnitude. The need for this condition in our proof arises in order to prevent a small change in  $f(Y, \omega)$  from inducing a large change in  $\text{sgn}(f(Y, \omega))$ .

We have the following result on low-degree hardness in the Ising setting. The proof is deferred to the full version [4].

**Theorem II.4.** *For any even integer  $p \geq 4$  there exist constants  $\mu < E_p(\Sigma)$ ,  $n^* \in \mathbb{N}$ ,  $\delta^* > 0$ , and  $\eta > 0$  such that the following holds. For any  $n \geq n^*$ , any  $D \in \mathbb{N}$ , any  $\delta \leq \min\{\delta^*, \frac{1}{4} \exp(-2D)\}$ , and any  $\gamma \geq (2/3)^D$ , there is no random degree- $D$  polynomial that  $(\mu, \delta, \gamma, \eta)$ -optimizes (1) on  $\Sigma_n$ .*

This result is very similar to the spherical case, and the discussion following Theorem II.2 also applies here. The best known algorithms for the Ising case also fall into the AMP framework [1], [2] and are thus captured by constant-degree polynomials. These polynomials output a solution “close” to the hypercube in a way that satisfies our technical condition involving  $\gamma, \eta$ . As in the spherical case, the case  $p = 2$  is computationally tractable; here it is not a simple eigenvector problem but can nonetheless be solved by the AMP algorithm of [1], [2].

2) *Langevin Dynamics and Gradient Descent:* One natural motivation for understanding low-degree hardness is to investigate the performance of natural iterative schemes, such as power iteration or gradient descent. In the spherical  $p$ -spin model, the natural analogue of these algorithms (in continuous time) are *Langevin dynamics* and *gradient flow*. While these are not directly low-degree methods, the overlap gap property can still be seen to imply hardness for these results in a fairly transparent manner.

To make this precise, let us introduce the following. Let  $B_t$  denote spherical Brownian motion. (For a textbook introduction to spherical Brownian motion see, e.g., [35].) For any variance  $\sigma \geq 0$ , we introduce *Langevin dynamics* for  $H_n$  to be the strong solution to the stochastic differential equation

$$dX_t = \sigma dB_t + \nabla H_n(X_t; Y) dt,$$

with  $X_0 = x$ , where here  $\nabla$  denotes the spherical gradient. Note that since  $H_n(x; Y)$  is a polynomial in  $x$ ,  $H_n$  is (surely) smooth and consequently the solution is well-defined in the strong sense [35]. The case  $\sigma = 0$  is referred to as *gradient flow* on the sphere.

In this setting, it is natural to study the performance with

random starts which are independent of  $Y$ , e.g., a uniform at random start. In this case, if the initial distribution is given by  $X_0 \sim \nu$  for some  $\nu \in \mathcal{M}_1(\mathcal{S}_n)$ , the space of probability measures on  $\mathcal{S}_n$ , we will denote the law by  $Q_\nu$ . In this setting we have the following result which is, again, a consequence of the overlap gap property.

**Theorem II.5.** *Let  $p \geq 4$  be even. There exists  $\mu < E_p(\mathcal{S})$  and  $c > 0$  such that for any  $\sigma \geq 0$ ,  $T \geq 0$  fixed,  $n$  sufficiently large, and  $\nu \in \mathcal{M}_1(\mathcal{S}_n)$ , if  $X_t$  denotes Langevin dynamics for  $H_n(\cdot; Y)$  with variance  $\sigma$  and initial data  $\nu$ , then*

$$\mathbb{P}_Y \otimes Q_\nu(H_n(X_T; Y) \leq \mu) \geq 1 - \exp(-cn).$$

*In particular, the result holds for  $\nu_n = \text{Unif}(\mathcal{S}_n)$ , the uniform measure on  $\mathcal{S}_n$ .*

The proof can be found in the full version [4]. To our knowledge, this is the first proof that neither Langevin dynamics nor gradient descent reach the ground state from the uniform distribution as a random start. We note furthermore, that the above applies even to  $T \leq c' \log n$  for some  $c' > 0$  sufficiently small.

There has been a tremendous amount of attention paid to the Langevin dynamics of spherical  $p$ -spin glass models. It is impossible here to provide a complete reference though we point the reader here to the surveys [36], [37], [38], [39]. To date, much of the analysis of the dynamics in the *non-activated* regime considered here ( $n \rightarrow \infty$  and then  $t \rightarrow \infty$ ) has concentrated on the Crisanti–Horner–Sommers–Cugliandolo–Kurchan (CHSCK) equations approach [40], [41]. This approach centers around the analysis of a system of integro-differential equations which are satisfied by the scaling limit of natural observables of the underlying system. While this property of the scaling limit has now been shown rigorously [42], [43], there is limited rigorous understanding of the solutions of the CHSCK equations beyond the case when  $p = 2$ . A far richer picture is expected here related to the phenomenon of *aging* [38], [44].

More recently a new, differential inequality–based approach to understanding this regime was introduced in [45], which provides upper and lower bounds on the energy level reached for a given initial data. That being said, this upper bound is nontrivial only for  $\sigma$  sufficiently large.

We end by noting that overlap gap–like properties, namely “free energy barriers” have been used to develop spectral gap estimates for Langevin dynamics which control the corresponding  $L^2$ -mixing time [46], [47]. In [47], it was shown that exponentially-small spectral gaps are connected to the existence of free energy barriers for the overlap, which at very low temperatures can be shown to be equivalent to a variant of the overlap gap property in this setting. To our knowledge, however, this work is the first approach to connect the behavior of Langevin dynamics in the non-activated regime ( $n \rightarrow \infty$  and then  $t \rightarrow \infty$ ) that utilizes the overlap distribution. Finally we note here that the overlap

gap property has been connected to the spectral gap for local, reversible dynamics of Ising spin glass models in [47] as well as to gradient descent and approximate message passing schemes in [23].

### B. Maximum Independent Set Problem in Sparse Random Graphs

We now consider the problem of finding a large independent set in a sparse random graph. Here, we are given the adjacency matrix of an  $n$ -vertex graph, represented as  $Y \in \{0, 1\}^m$  where  $m = \binom{n}{2}$ . We write  $Y \sim G(n, d/n)$  to denote an Erdős–Rényi graph on  $n$  nodes with edge probability  $d/n$ , i.e., every possible edge occurs independently with probability  $d/n$ . We are interested in the regime where first  $n \rightarrow \infty$  (with  $d$  fixed) and then  $d \rightarrow \infty$ . A subset of nodes  $S \subseteq [n]$  is an *independent set* if it spans no edges, i.e., for every  $i, j \in S$ ,  $(i, j)$  is not an edge. Letting  $\mathcal{I}(Y)$  denote the set of all independent sets of the graph  $Y$ , consider the optimization problem

$$\max_{S \in \mathcal{I}(Y)} |S| \tag{4}$$

where  $Y \sim G(n, d/n)$ .

As  $n \rightarrow \infty$  with  $d$  fixed, the rescaled optimum value of (4) is known to converge to some limit with high probability:

$$\frac{1}{n} \max_{S \in \mathcal{I}(Y)} |S| \rightarrow \alpha_d,$$

as shown in [48]. The limit  $\alpha_d$  is known to have the following asymptotic behavior as  $d \rightarrow \infty$ :

$$\alpha_d = (1 + o_d(1)) \frac{2 \log d}{d},$$

as is known since the work of Frieze [49]. The best known polynomial-time algorithm for this problem is achieved by a straightforward greedy algorithm which constructs a  $1/2$ -optimal independent set, i.e., an independent set of size  $\frac{\log d}{d} n$  asymptotically as  $n \rightarrow \infty$  and then  $d \rightarrow \infty$ .

We will study the ability of low-degree polynomials to find a large independent set. It is too much to ask for a polynomial to exactly output the indicator vector of an independent set, so we fix the following rounding scheme that takes a polynomial’s output and returns an independent set. Recall the terminology for random polynomials defined in Section II-A1.

**Definition II.6.** *Let  $f : \{0, 1\}^m \rightarrow \mathbb{R}^n$  be a random polynomial. For  $Y \in \{0, 1\}^m$ , and  $\eta > 0$ , let  $V_f^\eta(Y, \omega) \in \mathcal{I}(Y)$  be the independent set obtained by the following procedure. Let*

$$A = \{i \in [n] : f_i(Y, \omega) \geq 1\},$$

$$\tilde{A} = \{i \in A : i \text{ has no neighbors in } A \text{ in the graph } Y\},$$

and

$$B = \{i \in [n] : f_i(Y, \omega) \in (1/2, 1)\}.$$

Let

$$V_f^\eta(Y, \omega) = \begin{cases} \tilde{A} & \text{if } |A \setminus \tilde{A}| + |B| \leq \eta n, \\ \emptyset & \text{otherwise.} \end{cases}$$

In other words,  $f$  should output a value  $\geq 1$  to indicate that a vertex is in the independent set and should output a value  $\leq 1/2$  to indicate that it is not. It is allowed to make up to  $\eta n$  “errors”, each of which can either be a vertex for which the output value lies in  $(1/2, 1)$ , or a vertex that violates the independent set constraint. Vertices that violate the independent set constraint are thrown out, and if too many errors are made then the empty set  $\emptyset$  is returned. For our proofs it is crucial that this definition of  $V_f^\eta$  ensures that a small change in  $f(Y, \omega)$  cannot induce a large change in the resulting independent set  $V_f^\eta(Y, \omega)$  (without encountering the failure event  $\emptyset$ ).

We now formally define what it means for a polynomial to find a large independent set.

**Definition II.7.** For parameters  $k \in \mathbb{N}$ ,  $\delta \in [0, 1]$ ,  $\gamma \geq 1$ ,  $\eta > 0$ , and a random polynomial  $f : \{0, 1\}^m \rightarrow \mathbb{R}^n$ , we say that  $f$  ( $k, \delta, \gamma, \eta$ )-optimizes (4) if the following are satisfied.

- $\mathbb{E}_{Y, \omega} \|f(Y, \omega)\|_2^2 \leq \gamma k$ .
- With probability at least  $1 - \delta$  over  $Y$  and  $\omega$ , we have  $|V_f^\eta(Y, \omega)| \geq k$ .

The parameter  $k$  denotes the objective value attained (after rounding), i.e., the size of the independent set. For us,  $k$  will be a fixed multiple of  $\frac{\log d}{d}n$ , since this is the scale of the optimum. The parameter  $\delta$  is the algorithm’s failure probability. Note that if  $f$  were to “perfectly” output the  $\{0, 1\}$ -valued indicator vector of a size- $k$  independent set, then we would have  $\|f(Y, \omega)\|_2^2 = k$ . The parameter  $\gamma$  controls the degree to which this can be violated. Finally,  $\eta$  is the fraction of “errors” tolerated by the rounding process  $V_f^\eta$ .

We now state our main result of low-degree hardness of maximum independent set, with the proof deferred to the full version [4].

**Theorem II.8.** For any  $\alpha > 1 + 1/\sqrt{2}$  there exists  $d^* > 0$  such that for any  $d \geq d^*$  there exist  $n^* > 0$ ,  $\eta > 0$ , and  $C_1, C_2 > 0$  such that the following holds. Let  $n \geq n^*$ ,  $\gamma \geq 1$ , and  $D \leq \frac{C_2 n}{\gamma \log n}$ , and suppose  $\delta \geq 0$  satisfies

$$\delta < \exp(-C_1 \gamma D \log n).$$

Then for  $k = \alpha \frac{\log d}{d}n$ , there is no random degree- $D$  polynomial that ( $k, \delta, \gamma, \eta$ )-optimizes (4).

This shows that low-degree polynomials cannot find an independent set of size (asymptotically) exceeding  $(1 + 1/\sqrt{2}) \frac{\log d}{d}n$ , which is roughly 85% of the optimum. This is the threshold above which OGP can be shown using a first moment argument as in [19].

If  $\gamma$  is a constant, Theorem II.8 gives a similar tradeoff between  $D$  and  $\delta$  as our results for the  $p$ -spin model,

although here there is an extra factor of  $\log n$ . If we are willing to restrict to algorithms of failure probability  $\delta = \exp(-cn)$  then we can rule out all polynomials of degree  $D \leq c'n/\log n$  for a constant  $c' = c'(c)$ . As in the  $p$ -spin model, this suggests that exponential time  $\exp(n^{1-o(1)})$  is needed in order to find an independent set larger than  $(1 + 1/\sqrt{2}) \frac{\log d}{d}n$ .

As discussed in the introduction, the best known polynomial-time algorithm can find an independent set  $1/2$  as large as the optimum (asymptotically), and we expect this can also be achieved by a local algorithm (although this has only been shown rigorously for regular graphs). Any such local algorithm can be represented as a constant-degree polynomial (see the full version [4]). We expect that this polynomial satisfies our technical assumptions with parameters  $k = (1 + o_d(1)) \frac{\log d}{d}n$ ,  $\gamma = O(1)$ ,  $\delta = \exp(-\Omega(n))$ , and any constant  $\eta > 0$  (although we have not included a formal proof of this).

### C. The Overlap Gap Property

As discussed in the introduction, the preceding results will follow due to certain geometric properties of the super-level sets of the objectives. The main property is called the *overlap gap property (OGP)*. Let us begin by defining this formally in a general setting.

**Definition II.9.** We say that a family of real-valued functions  $\mathcal{F}$  with common domain  $\mathcal{X} \subset \mathbb{R}^n$  satisfies the overlap gap property for an overlap  $R : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$  with parameters  $\mu \in \mathbb{R}$  and  $0 \leq \nu_1 < \nu_2 \leq 1$  if for every  $f_1, f_2 \in \mathcal{F}$  and every  $x_1, x_2 \in \mathcal{X}$  satisfying  $f_k(x_k) \geq \mu$  for  $k = 1, 2$ , we have that  $R(x_1, x_2) \in [0, \nu_1] \cup [\nu_2, 1]$ .

For ease of notation, when this holds, we simply say that  $\mathcal{F}$  satisfies the  $(\mu, \nu_1, \nu_2)$ -OGP for  $R$  on  $\mathcal{X}$ . Furthermore, as it is often clear from context, we omit the dependence of the above on  $R$ .

While the definition above might be satisfied for trivial reasons and thus not be informative, it will be used in this paper in the setting where  $\|x\|_2^2 \leq n$  for every  $x \in \mathcal{X}$ ,  $R(x_1, x_2) = |\langle x_1, x_2 \rangle|/n$ , and with parameters chosen so that with high probability  $\mu < \sup_{x \in \mathcal{X}} H(x)$  for every  $H \in \mathcal{F}$ . Thus, in particular  $R(x_1, x_2) \leq 1$  for every  $x_1, x_2 \in \mathcal{X}$ , and  $\mu$  measures some proximity from optimal values for each objective function  $H$ . The definition says informally that for every two  $\mu$ -optimal solutions with respect to any two choices of objective functions, their normalized inner product is either at least  $\nu_2$  or at most  $\nu_1$ .

In the following, we require one other property of functions, namely separation of their superlevel sets.

**Definition II.10.** We say that two real-valued functions  $f, g$  with common domain  $\mathcal{X}$  are  $\nu$ -separated above  $\mu$  with respect to the overlap  $R : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$  if for any  $x, y \in \mathcal{X}$  with  $f(x) \geq \mu$  and  $g(y) \geq \mu$ , we have that  $R(x, y) \leq \nu$ .

This property can be thought of a strengthening of OGP for two distinct functions. In particular, the parameter  $\nu$  will typically equal the parameter  $\nu_1$  in the definition of OGP.

Let us now turn to stating the precise results regarding these properties in the settings we consider here. It can be shown that the overlap gap property holds for  $p$ -spin glass Hamiltonians in both the spherical and Ising settings with respect to the overlap  $R(x, y) = \frac{1}{n}|\langle x, y \rangle|$ . More precisely, let  $Y$  be i.i.d.  $\mathcal{N}(0, 1)$  and let  $Y'$  denote an independent copy of  $Y$ . Consider the corresponding family of real-valued functions

$$\begin{aligned} \mathcal{A}(Y, Y') & \\ &= \{\cos(\tau)H_n(\cdot; Y) + \sin(\tau)H_n(\cdot; Y') : \tau \in [0, \pi/2]\}. \end{aligned} \quad (5)$$

We then have the following, which will follow by combining bounds from [32], [33]. The second result is a restatement of [23, Theorem 3.4]. The proof can be found in the full version [4].

**Theorem II.11.** *Take as overlap  $R(x, y) = \frac{1}{n}|\langle x, y \rangle|$  and let  $Y$  and  $Y'$  be independent  $p$ -tensors with i.i.d.  $\mathcal{N}(0, 1)$  entries. For every even  $p \geq 4$  there exists an  $\varepsilon > 0$  such that the following holds:*

- 1) *For the domain  $\mathcal{S}_n$ , there are some  $0 \leq \nu_1 < \nu_2 \leq 1$  and some  $c > 0$  such that the following holds with probability at least  $1 - \exp(-cn)$ :*
  - $\mathcal{A}(Y, Y')$  has the overlap gap property for  $R$  with parameters  $(E_p(\mathcal{S}) - \varepsilon, \nu_1, \nu_2)$ .
  - $H_n(\cdot; Y)$  and  $H_n(\cdot; Y')$  are  $\nu_1$ -separated above  $E_p(\mathcal{S}) - \varepsilon$  with respect to  $R$ .
- 2) *For the domain  $\Sigma_n$ , there are some  $0 \leq \nu_1 < \nu_2 \leq 1$  and some  $c > 0$  such that the following holds with probability at least  $1 - \exp(-cn)$ :*
  - $\mathcal{A}(Y, Y')$  has the overlap gap property for  $R$  with parameters  $(E_p(\Sigma) - \varepsilon, \nu_1, \nu_2)$ .
  - $H_n(\cdot; Y)$  and  $H_n(\cdot; Y')$  are  $\nu_1$ -separated above  $E_p(\Sigma) - \varepsilon$  with respect to  $R$ .

Let us now turn to the maximum independent set problem. Let us begin by first observing that we may place this family of optimization problem on a common domain. To this end, consider as domain, the Boolean hypercube  $\mathcal{B}_n = \{0, 1\}^n$ . Note that by viewing a vector  $x$  as the indicator function of the set  $S = S(x) := \{i : x_i = 1\}$ , we have a correspondence between the points  $x \in \mathcal{B}_n$  and subsets of the vertex set  $[n]$ . Let  $m = \binom{n}{2}$ , let  $Y \in \{0, 1\}^m$  denote the adjacency matrix of some graph on  $[n]$  vertices, and consider the function  $F(x; Y)$  given by

$$F(x; Y) = |S(x)| \cdot \mathbb{1}\{S(x) \in \mathcal{I}(Y)\}.$$

The maximum independent set problem for  $Y$  can then be written in the form

$$\max_{x \in \mathcal{B}_n} F(x; Y).$$

Let us now construct the analogue of the family  $\mathcal{A}(Y, Y')$  from (5) in this setting.

**Definition II.12.** *For  $Y, Y' \in \{0, 1\}^m$ , the path from  $Y$  to  $Y'$  is  $Y = Z_0 \rightarrow Z_1 \rightarrow \dots \rightarrow Z_m = Y'$  where  $(Z_i)_j = Y_j$  for  $j > i$  and  $(Z_i)_j = Y'_j$  otherwise. The path is denoted by  $Y \mapsto Y'$ .*

Here (and throughout) we have fixed an arbitrary order by which to index the edges of a graph (the coordinates of  $Y$ ).

Now let  $Y, Y' \in \{0, 1\}^m$  be (the adjacency matrices of) independent  $G(n, d/n)$  random graphs. We can then consider the family of functions

$$\mathcal{F}(Y, Y') = \{F(\cdot; Z) : Z \text{ is on the path } Y \mapsto Y'\}. \quad (6)$$

We can now state the relevant overlap gap property.

**Theorem II.13.** *For any  $\alpha > 1 + 1/\sqrt{2}$  there exist constants  $0 \leq \tilde{\nu}_1 < \tilde{\nu}_2 \leq 1$  and  $d^* > 0$  such that for any constant  $d \geq d^*$ , the following holds. If  $Y, Y' \sim G(n, d/n)$  independently, the following holds with probability at least  $1 - \exp(-\Omega(n))$ .*

- *The family of functions  $\mathcal{F}$  from (6) with domain  $\mathcal{X} = \mathcal{B}_n$  satisfies the overlap gap property with overlap  $R(x_1, x_2) = \frac{1}{n}|\langle x_1, x_2 \rangle|$  and parameters  $\mu = k := \alpha \frac{\log d}{d}n$ ,  $\nu_1 = \tilde{\nu}_1 \frac{k}{n}$ ,  $\nu_2 = \tilde{\nu}_2 \frac{k}{n}$  with probability at least  $1 - \exp(-\Omega(n))$ .*
- *Furthermore, the functions  $F(\cdot; Y)$  and  $F(\cdot; Y')$  are  $\nu_1$ -separated above  $\mu$ .*

Above (and throughout),  $\Omega(n)$  pertains to the limit  $n \rightarrow \infty$  with  $\alpha, d$  fixed, i.e., it hides a constant factor depending on  $\alpha, d$ . Note that here the overlap is simply the (normalized) cardinality of the intersection of the two sets:  $R(x_1, x_2) = \frac{1}{n}|S(x_1) \cap S(x_2)|$ .

The proof of Theorem II.13—which is deferred to the full version [4]—is an adaptation of the first moment argument of [19]: we compute the expected number of pairs of independent sets whose overlap lies in the “forbidden” region, and show that this is exponentially small.

#### ACKNOWLEDGMENT

D.G. supported by ONR grant N00014-17-1-2790. A.J. supported by NSERC grants RGPIN-2020-04597 and DGEGR-2020-00199. A.S.W. partially supported by NSF grant DMS-1712730 and by the Simons Collaboration on Algorithms and Geometry.

For helpful discussions, A.S.W. is grateful to Léo Molane, Eliran Subag, Jean-Christophe Mourrat, and Ahmed El Alaoui.

#### REFERENCES

- [1] A. Montanari, “Optimization of the Sherrington-Kirkpatrick hamiltonian,” in *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2019, pp. 1417–1433.

- [2] A. El Alaoui, A. Montanari, and M. Sellke, “Optimization of mean-field spin glasses,” *arXiv preprint arXiv:2001.00904*, 2020.
- [3] J. Lauer and N. Wormald, “Large independent sets in regular graphs of large girth.” *Journal of Combinatorial Theory, Series B*, vol. 97, pp. 999–1009, 2007.
- [4] D. Gamarnik, A. Jagannath, and A. S. Wein, “Low-degree hardness of random optimization problems,” *arXiv preprint arXiv:2004.12063*, 2020.
- [5] E. Subag, “Following the ground-states of full-RSB spherical spin glasses,” *arXiv preprint arXiv:1812.04588*, 2018.
- [6] B. Barak, S. Hopkins, J. Kelner, P. K. Kothari, A. Moitra, and A. Potechin, “A nearly tight sum-of-squares lower bound for the planted clique problem,” *SIAM Journal on Computing*, vol. 48, no. 2, pp. 687–735, 2019.
- [7] S. B. Hopkins and D. Steurer, “Efficient bayesian estimation from few samples: community detection and related problems,” in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2017, pp. 379–390.
- [8] S. B. Hopkins, P. K. Kothari, A. Potechin, P. Raghavendra, T. Schramm, and D. Steurer, “The power of sum-of-squares for detecting hidden structures,” in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2017, pp. 720–731.
- [9] S. Hopkins, “Statistical inference and the sum of squares method,” Ph.D. dissertation, Cornell University, 2018.
- [10] P. Raghavendra, T. Schramm, and D. Steurer, “High-dimensional estimation via sum-of-squares proofs,” *arXiv preprint arXiv:1807.11419*, vol. 6, 2018.
- [11] D. Kunisky, A. S. Wein, and A. S. Bandeira, “Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio,” *arXiv preprint arXiv:1907.11636*, 2019.
- [12] S. B. Hopkins, J. Shi, and D. Steurer, “Tensor principal component analysis via sum-of-square proofs,” in *Conference on Learning Theory*, 2015, pp. 956–1006.
- [13] S. B. Hopkins, T. Schramm, J. Shi, and D. Steurer, “Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors,” in *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, 2016, pp. 178–191.
- [14] A. S. Bandeira, D. Kunisky, and A. S. Wein, “Computational hardness of certifying bounds on constrained PCA problems,” *arXiv preprint arXiv:1902.07324*, 2019.
- [15] Y. Ding, D. Kunisky, A. S. Wein, and A. S. Bandeira, “Subexponential-time algorithms for sparse PCA,” *arXiv preprint arXiv:1907.11635*, 2019.
- [16] D. Achlioptas and A. Coja-Oghlan, “Algorithmic barriers from phase transitions,” in *2008 IEEE 49th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2008, pp. 793–802.
- [17] D. Achlioptas, A. Coja-Oghlan, and F. Ricci-Tersenghi, “On the solution space geometry of random formulas,” *Random Structures and Algorithms*, vol. 38, pp. 251–268, 2011.
- [18] M. Mézard, T. Mora, and R. Zecchina, “Clustering of solutions in the random satisfiability problem,” *Physical Review Letters*, vol. 94, no. 19, p. 197205, 2005.
- [19] D. Gamarnik and M. Sudan, “Limits of local algorithms over sparse random graphs,” *Annals of Probability*, vol. 45, pp. 2353–2376, 2017.
- [20] M. Rahman and B. Virag, “Local algorithms for independent sets are half-optimal,” *The Annals of Probability*, vol. 45, no. 3, pp. 1543–1577, 2017.
- [21] W.-K. Chen, D. Gamarnik, D. Panchenko, and M. Rahman, “Suboptimality of local algorithms for a class of max-cut problems,” *The Annals of Probability*, vol. 47, no. 3, pp. 1587–1618, 2019.
- [22] A. Coja-Oghlan, A. Haqshenas, and S. Hetterich, “Walksat stalls well below satisfiability,” *SIAM Journal on Discrete Mathematics*, vol. 31, no. 2, pp. 1160–1173, 2017.
- [23] D. Gamarnik and A. Jagannath, “The overlap gap property and approximate message passing algorithms for  $p$ -spin models,” *arXiv preprint arXiv:1911.06943*, 2019.
- [24] E. Richard and A. Montanari, “A statistical model for tensor PCA,” in *Advances in Neural Information Processing Systems*, 2014, pp. 2897–2905.
- [25] A. S. Wein, A. El Alaoui, and C. Moore, “The Kikuchi hierarchy and tensor PCA,” in *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2019, pp. 1446–1468.
- [26] M. B. Hastings, “Classical and quantum algorithms for tensor principal component analysis,” *Quantum*, vol. 4, p. 237, 2020.
- [27] G. Biroli, C. Cammarota, and F. Ricci-Tersenghi, “How to iron out rough landscapes and get optimal performances: Averaged gradient descent and its application to tensor PCA,” *Journal of Physics A: Mathematical and Theoretical*, 2020.
- [28] G. Ben Arous, R. Gheissari, and A. Jagannath, “Algorithmic thresholds for tensor PCA,” *arXiv preprint arXiv:1808.00921*, 2018.
- [29] F. Guerra and F. L. Toninelli, “The thermodynamic limit in mean field spin glass models,” *Comm. Math. Phys.*, vol. 230, no. 1, pp. 71–79, 2002.
- [30] A. Jagannath and I. Tobasco, “Low temperature asymptotics of spherical mean field spin glasses,” *Comm. Math. Phys.*, vol. 352, no. 3, pp. 979–1017, 2017.
- [31] A. Auffinger, G. Ben Arous, and J. Černý, “Random matrices and complexity of spin glasses,” *Comm. Pure Appl. Math.*, vol. 66, no. 2, pp. 165–201, 2013.
- [32] W.-K. Chen and A. Sen, “Parisi formula, disorder chaos and fluctuation for the ground state energy in the spherical mixed  $p$ -spin models,” *Communications in Mathematical Physics*, vol. 350, no. 1, pp. 129–173, 2017.

- [33] A. Auffinger and W.-K. Chen, “On the energy landscape of spherical spin glasses,” *Advances in Mathematics*, vol. 330, pp. 553–588, 2018.
- [34] A. Jagannath and S. Sen, “On the unbalanced cut problem and the generalized Sherrington-Kirkpatrick model,” *arXiv preprint arXiv:1707.09042*, 2017.
- [35] E. P. Hsu, *Stochastic analysis on manifolds*, ser. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, 2002, vol. 38.
- [36] J.-P. Bouchaud, L. F. Cugliandolo, J. Kurchan, and M. Mézard, “Out of equilibrium dynamics in spin-glasses and other glassy systems,” *Spin glasses and random fields*, pp. 161–223, 1998.
- [37] L. F. Cugliandolo, “Course 7: Dynamics of glassy systems,” in *Slow Relaxations and nonequilibrium dynamics in condensed matter*. Springer, 2003, pp. 367–521.
- [38] A. Guionnet, “Dynamics for spherical models of spin-glass and aging,” in *Spin glasses*, ser. Lecture Notes in Math. Springer, Berlin, 2007, vol. 1900, pp. 117–144.
- [39] A. Jagannath, “Dynamics of mean field spin glasses on short and long timescales,” *Journal of Mathematical Physics*, vol. 60, no. 8, p. 083305, 2019.
- [40] A. Crisanti, H. Horner, and H.-J. Sommers, “The spherical p-spin interaction spin-glass model,” *Zeitschrift für Physik B Condensed Matter*, vol. 92, no. 2, pp. 257–271, 1993.
- [41] L. F. Cugliandolo and J. Kurchan, “Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model,” *Phys. Rev. Lett.*, vol. 71, pp. 173–176, Jul 1993.
- [42] G. Ben Arous, A. Dembo, and A. Guionnet, “Aging of spherical spin glasses,” *Probab. Theory Related Fields*, vol. 120, no. 1, pp. 1–67, 2001.
- [43] —, “Cugliandolo-Kurchan equations for dynamics of spin-glasses,” *Probab. Theory Related Fields*, vol. 136, no. 4, pp. 619–660, 2006.
- [44] G. Ben Arous, “Aging and spin-glass dynamics,” in *Proceedings of the International Congress of Mathematicians, Vol. III (Beijing, 2002)*. Higher Ed. Press, Beijing, 2002, pp. 3–14.
- [45] G. Ben Arous, R. Gheissari, and A. Jagannath, “Bounding flows for spherical spin glass dynamics,” *Communications in Mathematical Physics*, vol. 373, no. 3, pp. 1011–1048, 2020.
- [46] R. Gheissari and A. Jagannath, “On the spectral gap of spherical spin glass dynamics,” *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 55, no. 2, pp. 756–776, 05 2019.
- [47] G. Ben Arous and A. Jagannath, “Spectral gap estimates in mean field spin glasses,” *Communications in Mathematical Physics*, vol. 361, no. 1, pp. 1–52, 2018.
- [48] M. Bayati, D. Gamarnik, and P. Tetali, “Combinatorial approach to the interpolation method and scaling limits in sparse random graphs,” *Annals of Probability. (Conference version in Proc. 42nd Ann. Symposium on the Theory of Computing (STOC) 2010)*, vol. 41, pp. 4080–4115, 2013.
- [49] A. Frieze, “On the independence number of random graphs,” *Discrete Mathematics*, vol. 81, pp. 171–175, 1990.