Strong Low Degree Hardness for the Number Partitioning Problem

April 03, 2025

Abstract

Finding good solutions to the *number partitioning problem (NPP)* – that is, finding a partition of a set of N numbers to minimize the discrepancy between the sums of the two subsets – is a well-studied optimization problem, with applications to statistics, physics, and computer science. Along with having numerous practical applications in the design of randomized control trials and processor scheduling, the NPP is famous for posessing a *statistical-to-computational gap*: assuming the N numbers to be partitioned are i.i.d. standard Normal random variables, the optimal discrepancy is $2^{-\Theta(N)}$ with high probability, but the best polynomial-time algorithms only find solutions with a discrepancy of $2^{-\Theta(\log^2 N)}$. This gap is a common feature in optimization problems over random combinatorial structures, and indicates the need for a theory of computational complexity beyond worst-case analysis.

In this thesis, we prove a strong form of algorithmic hardness for the number partitioning problem, aiming to establish that this statistical-to-computational gap is an intrinsic feature of the NPP. We study *low degree algorithms*, as they provide both tunable stability behavior and are tractable models for a broad class of algorithms, under the widely successful *low degree heuristic*. Then, we establish a *brittleness property* on the geometry of the solution set, which ensures that stable algorithms are unable to efficiently traverse this random landscape. By combining these notions, we are able to show *strong low degree hardness*, in that low degree algorithms will fail to find good solutions with high probability. In addition, while we show that low degree polynomial algorithms are structurally ill-suited to the NPP, our results for the more general class of *low coordinate degree algorithms* demonstrate a sharp tradeoff between algorithmic runtime (vis-à-vis algorithmic complexity) and solution discrepancy, which our analysis suggests is optimal.

Finally, we establish a *repulsion property*, giving a precise tradeoff between the the discrepancy of a solution to a fixed instance and its distance to other solutions. We then leverage this to show that any algorithm fed through a truly randomized rounding scheme will fail to find solutions with high probability. This work demonstrates the effectiveness of using landscape properties to address questions about algorithmic hardness, and suggests interesting directions for future study.

Contents

1 Introduction

Suppose that we have N items, each with associated weights. How should we divide these items into two groups so that the sum of their weights is as close as possible? Or, is it possible to divide these items into two groups such that the absolute difference of the sums of their weights is below a certain threshold? This question is known as the *number partitioning problem (NPP)*, and has been a subject of fascination in statistics, physics, and computer science since its proposal in 1969 [41].

Formally, let $g_1, \ldots, g_N$ be N real numbers. The NPP is the problem of finding the subset A of $[N] := \{1, 2, \ldots, N\}$ which minimizes the discrepancy

$$\left| \sum_{i \in A} g_i - \sum_{i \notin A} g_i \right|.$$

Alternatively, identify the instance $g_1, \ldots, g_N$ with a point $g \in \mathbf{R}^N$. Then, choosing a subset $A \subseteq [N]$ is equivalent to choosing a point x in the N-dimensional binary hypercube $\Sigma_N := \{\pm 1\}^N$, where $x_i = +1$ is the same as including $i \in A$. The discrepancy of x is now $|\langle g, x \rangle|$, and solving the NPP means finding the x minimizing this discrepancy:

$$\min_{x \in \Sigma_N} |\langle g, x \rangle|. \tag{1.1}$$

Rephrased as a decision problem – whether there exists a subset $A \subseteq [N]$ (or a point $x \in \Sigma_N$) such that the discrepancy is zero, or sufficiently small – the NPP is NP-complete; this can be shown by reduction from the subset sum problem. In fact, the NPP is one of the six basic NP-complete problems of Garey and Johnson, and of those, the only one involving numbers [35, § 3.1].

Finding "good" solutions to this problem has a number of practical applications. For instance, the NPP and MWNPP[1] were first formulated by Graham, who considered it in the context of multi-

[1] That is, the *multiway number partitioning problem (MWNPP)*, in which we want to partition $g_1, \ldots, g_N$ into M subsets such that the within-subset sums are mutually close; what "mutually close" means precisely varies across the literature.

processor scheduling: dividing a group of tasks with known runtimes across a pool of processors so as to minimize one core being overworked while others stall [41]. Later work by Coffman, Garey, and Johnson, as well as Tsai, looked at utilizing algorithms for the NPP for designing multiprocessor schedulers or large integrated circuits [23], [94]. Coffman and Lueker also write on how the NPP can be applied as a framework for allocating material stocks, such as steel coils in factories, paintings in museums, or advertisements in newspapers [25].

One particularly important application of the NPP in statistics comes from the design of *randomized controlled trials*. Consider N individuals, each with a set of covariate information $g_i \in \mathbf{R}^d$. Then the problem is to divide them into a treatment group (denoted A_+) and a control group (denoted A_-), subject each to different conditions, and evaluate the responses. In order for such a trial to be accurate, it is necessary to ensure that the covariates across both groups are roughly the same. In our notation, this equates to finding an A_+ (with $A_- := [N] \setminus A_+$) minimizing

$$\min_{A_+ \subseteq [N]} \left\| \sum_{i \in A_+} g_i - \sum_{i \in A_-} g_i \right\|_\infty . \tag{1.2}$$

This extension of the NPP is often termed the *vector balancing problem (VBP)*, and many algorithms for solving the NPP/VBP come from designing such randomized controlled trials [59], [48].

On the other hand, in 1976, Merkle and Hellman devised one of the earliest public key cryptography schemes, deriving its hardness from their belief that a variant of the NPP was computationally difficult to solve – at the time, it was not yet known whether the NPP was NP-complete or not [81]. Their proposal was for the receiver, say Alice, to generate as a public key N natural numbers $(a_1, ..., a_N)$, with N typically around 100 and each a_i around 200 bits long. To encrypt a N-bit message $x = (x_1, ..., x_N) \in \{0,1\}^N$, the sender, say Bob, could compute

$$b := \sum_{i \in N} a_i x_i$$

and send the ciphertext b to Alice. Any eavesdropper would know $a_1, ..., a_N$, as well as b, and decrypting the message involved finding a subset of the a_i adding up to b. This is known as the *knapsack problem*, which is NP-complete, as can be shown by restriction to the NPP [35, 3.2.1(6)]. However, such NP-completeness is only a worst-case hardness guarantee; Merkle and Hellman's scheme involved Alice choosing $a_1, ..., a_N$ by cryptographically scrambling a sequence $(a'_1, ..., a'_N)$ for which solving the NPP was easy, enabling the receiver to practically decrypt the message x from the ciphertext b. In 1984, Shamir – one of the developers of the RSA cryptosystem still in use today – showed that one could exploit this public key generation process to reduce the "hard" knapsack problem to one which was solvable in polynomial time, rendering the Merkle-Hellman scheme insecure [91]. While today, Merkle-Hellman is but a footnote in the history of cryptography, it demonstrates the importance of looking beyond worst-case hardness and expanding complexity theory to describe the difficulty of the average problem instance.

Meow moe wmeow. Another major source of interest in the NPP, as well as potential explanations for when it is hard, come from statistical physics. In the 1980s, Derrida introduced the eponymous *random energy model (REM)*, a simplified example of a spin glass in which, unlike the Sherrington-Kirkpatrick or other p-spin glass models, the possible energy levels are independent of each other

[28], [29], [14]. Despite this simplicity, this model made possible heuristic analyses of the Parisi theory for mean field spin glasses, and it was suspected that arbitrary random discrete systems would locally behave like the REM [15], [62]. The NPP was the first system for which this local REM conjecture was shown [17], [18]. In addition, in the case when the g_i are independently chosen uniformly over $\{1, 2, ..., 2^M\}$, Gent and Walsh conjectured that the hardness of finding perfect partitions (i.e., with discrepancy zero if $\sum_i g_i$ is even, and one else) was controlled by the parameter $\kappa := \frac{m}{n}$ [44], [43]. Mertens soon gave a nonrigorous statistical mechanics argument suggesting the existence of a phase transition from $\kappa < 1$ to $\kappa > 1$; that is, while solutions exist in the low κ regime, they stop existing in the high κ regime [79]. It was also observed that this phase transition coincided with the empirical onset of computational hardness for typical algorithms, and Borgs, Chayes, and Pittel proved the existence of this phase transition soon after [49], [13].

1.1 The Statistical-to-Computational Gap

Many problems involving searches over random combinatorial structures (i.e., throughout high-dimensional statistics) exhibit a statistical-to-computational gap: the optimal values which are known to exist via non-constructive, probabilistic methods are far better than those achievable by state-of-the-art algorithms. In the pure optimization setting, examples such gaps are found in random constraint satisfaction [83], [1], [69], finding maximal independent sets in sparse random graphs [42], [22], the largest submatrix problem [40], [36], and the p-spin and diluted p-spin models [34], [24]. These gaps also arise in various "planted" models, such as matrix or tensor PCA [20], [73], [74], [55], [52], [3], high-dimensional linear regression [45], [46], or the infamously hard planted clique problem [56], [31], [84], [10], [47]. These indicate that these problems are "hard" in a way that goes beyond being NP; algorithms fail even on average-case instances.

The NPP is no exception: despite its apparent simplicity, its persistent importance in the random optimization literature comes from the shocking width of its associated statistical-to-computational gap. On the statistical side, the landmark result here is by Karmarkar et al., who showed that when the g_i are i.i.d.random variables, with distribution sufficiently nice,[2] then the minimum discrepancy of (1.1) is $\Theta\left(\sqrt{N}2^{-N}\right)$ with high probability as $N \to \infty$ [60]. Their result also extends to *even partitions*, where the sizes of each subset is equal (i.e., for N even), worsening only to $\Theta(N2^{-N})$. Yet the best known algorithms cannot achieve discrepancies close to this in polynomial time.

A first approach to the NPP, often termed the *greedy heuristic*, would be to sort the N inputs, place the largest in one subset, and place the subsequent largest numbers in the subset with the smaller total running sum. This takes $O(N \log N)$ time (due to the sorting step), but achieves a discrepancy of $O(N^{-1})$, extremely far off from the statistical optimum [80]. More recently, Krieger et al. developed an algorithm achieving a discrepancy of $O(N^{-2})$, but in exchange for this poor performance, their algorithm solves for a balanced partition, making it useful for randomized control trials [59].

The true breakthrough towards the statistical optimum came from Karmarkar and Karp, whose algorithm produced a discrepancy of $O\left(N^{-\alpha \log N}\right) = 2^{-O(\log^2 N)}$ with high probability. Their algorithm is rather complicated, involving randomization and a resampling step to make their analysis tractable, but their main contribution is the *differencing heuristic* [63]. The idea is as follows: if S is a list of items, then putting $g, g' \in S$ in opposite partitions is the same as removing g and g' and adding $|g - g'|$ back to S. Karmarkar and Karp propose two simpler algorithms based on this

[2] Specifically, having bounded density and finite fourth moment.

heuristic, the *partial differencing method (PDM)* and *largest differencing method (LDM)*, which they conjectured could also achieve discrepancies of $O\left(N^{-\alpha \log N}\right)$. In both, the items are first sorted, and the differencing is performed on the pairs of the largest and second largest items. However, in PDM, the remainders are ignored until all original numbers have been differenced, and then are resorted and repartitioned, while LDM reinserts the remainder in sorted order at each step; in any case, both algorithms are thus polynomial in N. Lueker soon disproved the claim that PDM could achieve the Karmarkar-Karp discrepancy, showing that when g_i were i.i.d. Uniform on $[0, 1]$, then the expected discrepancy was $\Theta(N^{-1})$, no better than the greedy algorithm [78]. However, for g_i i.i.d. Uniform or even Exponential, Yakir confirmed that LDM could achieve the performance of the original differencing algorithm, proving that its expected discrepancy was $N^{-\Theta(\log N)}$ [99]. The constant α was later estimated for LDM to be $\alpha = \frac{1}{2 \ln 2}$, via nonrigorous calculations [16].

Of course, at its most basic level, the NPP is a search problem over 2^N possible partitions, so given more and more time, an appropriate algorithm could keep improving its partition until it achieved the global optimum. To this degree, Korf developed alternatives known as the *complete greedy* and *complete Karmarkar-Karp* algorithms which, if run for exponentially long time, can find the globally optimal partition [67], [68]. This algorithm was later extended to multiway number partitioning [66]. See also Michiels et al. for extensions to balanced multiway partitioning [82].

For the multidimensional VBP case, Spencer showed in 1985 that the worse-case discrepancy of the VBP was at most $6\sqrt{N}$ for $d = N$ and $\|g_i\|_\infty \leq 1$ for all $1 \leq i \leq N$ [92]. However, his argument is an application of the probabilistic method, and does not construct such a solution. In the average case, Turner et al. proved that, under similar regularity assumptions on the g_i,[2] the minimum discrepancy is $\Theta\left(\sqrt{N} 2^{-N/d}\right)$ for all $d \leq o(N)$, with high probability [93]. For the regime $\delta = \Theta(N)$, Aubin et al. conjecture that there exists an explicit function $c(\delta)$ such that for $\delta > 0$, the discrepancy in the $d = \delta N$ regime is $c(\delta)\sqrt{N}$ with high probability [6]. To this end, Turner et al. also showed that for $d \leq \delta N$, one can achieve $O\left(\sqrt{N} 2^{-1/\delta}\right)$ with probability at least 99% [93]. On the algorithmic side, they generalized the Karmarkar-Karp algorithm to VBP, which, for $2 \leq d = O(\sqrt{\log N})$ finds partitions with discrepancy $2^{-\Theta(\log^2 N/d)}$, reproducing the gap of classical Karmarkar-Karp. On the other hand, in the superlinear regime $d \geq 2N$, this average-case discrepancy worsens to $\tilde{O}\left(\sqrt{N \log(2d/N)}\right)$ [27]. Yet, many proposed algorithms can achieve similar discrepancies, which is believed to be optimal for $d \geq N$ [92], [9], [75], [86].

1.2 Algorithmic Hardness and Landscape Obstructions

Classical algorithmic complexity theory – involving classes such as P, NP, etc. – is poorly suited to describing the hardness of random optimization problems, as these classes are based on the worst-case performance of available algorithms. In many cases, the statistically possible performance of solutions to random instances of these NP-complete problems will be far better than the worst-case analysis would suggest. How then, can we extend complexity theory to describe problems which, like the NPP, are hard on average? Along with the aforementioned statistical-to-computational gaps, the past two decades of research have shown that many methods can provide evidence of this average-case hardness, such as the failure of Markov chain algorithms [56], [39], [54], the failure of approximate message passing (AMP) algorithms [100], [19], or lower bounding performance against the sum-of-squares hierarchy or the statistical query model [55], [52], [87], [10], [61], [30], [32].

One particularly interesting approach is to prove average-case to worst-case reductions: if one shows that a polynomial-time algorithm for solving random instances could be used to design a polynomial-time algorithm for arbitrary instances, then assuming the problem was known to be in NP, it can be concluded that no such polynomial-time algorithm for the average case can exist [33]. This method has been used to show hardness for sparse PCA, detecting planted independent subgraphs, and more by reducing to the random planted clique problem [20], [11], [12]. To this extent, Hoberg et al. provided such evidence of hardness for the NPP by showing that a polynomial-time approximation oracle achieving discrepancies around $O\left(2^{\sqrt{N}}\right)$ could give polynomial-time approximations for Minkowski's problem, the latter of which is known to be hard [51]. More recently, Vafa and Vaikuntanathan showed that the Karmarkar-Karp algorithm's performance was nearly tight, assuming the worst-case hardness of the shortest vector problem on lattices [96]. Other conjectures suggested that the onset of algorithmic hardness was related to phase transitions in the solution landscapes, something which has been shown for random k-SAT, but this fails to describe hardness for optimization problems.

A more recent and widely successful approach is based on analyzing the geometry of the solution landscape. Many "hard" random optimization problems have a certain disconnectivity property, known as the *overlap gap property (OGP)* [33]. Roughly, this means there exist $0 \leq \nu_1 < \nu_2$ such that, for every two near-optimal states x, x' for a particular instance g of the problem either have $d(x, x') < \nu_1$ or $d(x, x') > \nu_2$. That is, pairs of solutions are either close to each other, or much further away – the condition that $\nu_1 < \nu_2$ ensures that the "diameter" of solution clusters is much smaller than the separation between these clusters.[3] Beyond ruling out the existence of pairs of near solutions with $d(x, x') \in [\nu_1, \nu_2]$, a common extension is the *multioverlap OGP (m-OGP)*: given an ensemble of m strongly correlated instances, there do not exist m-tuples of near solutions all equidistant from each other. This extension is often useful to lower the "threshold" at which the OGP starts to appear. Once established, the OGP and m-OGP, which is intrinsic to the problem, can be used to rule out the success of wide classes of stable algorithms [7], [1], [83], [42], [37], [88], [97].

For the NPP, it was expected for decades that the "brittleness" of the solution landscape would be a central barrier in finding successful algorithms to close the statistical-to-computational gap. Mertens wrote in 2001 that any local heuristics, which only looked at fractions of the domain, would fail to outperform random search [79, § 4.3]. This was backed up by the failure of many algorithms based on locally refining Karmarkar-Karp-optimal solutions, such as simulated annealing [2], [90], [57], [58], [4]. To that end, more recent approaches for algorithmic development are rooted in more global heuristics [64], [26], [89], and some of the most interesting directions in algorithmic development for the NPP comes from quantum computing: as this is outside our scope, we encourage the interested reader to consult [8], [98]. The main result to this effect comes from Gamarnik and Kızıldağ, who proved that for m of constant order, the m-OGP for NPP held for discrepancies of $2^{-\Theta(N)}$ (i.e., the statistical optimum), but was absent for smaller discrepancies of 2^{-E_N} with $\omega(1) \leq E_N \leq o(N)$ [39]. They do show, however, that the m-OGP in for $E_N \geq \omega(\sqrt{N \log N})$ could be recovered for m superconstant. This allowed them to prove that for $\varepsilon \in (0, 1/5)$, no stable algorithm (suitably defined) could find solutions with discrepancy 2^{-E_N} for $\omega\left(N \log^{-\frac{1}{5}+\varepsilon} N\right) \leq E_N \leq o(N)$ [39, Thm. 3.2]. These results point to the efficacy of using landscape obstructions to show algorithmic hardness for the NPP, which we will take advantage of in Section 4.

[3] This is called the "overlap" gap property because, in the literature, this is often described in terms of the inner product of the solutions, as opposed to the distance between them.

1.3 Our Results

In this thesis, we use a variant of the OGP, which we term a *conditional landscape obstruction*, to prove low degree algorithmic hardness guarantees for the NPP at a range of discrepancy scales. Our obstruction is based on the observation that given a solution to one instance of the NPP, it is impossible to pin down the location of any solution to a strongly correlated instance, which prevents suitably stable algorithms from traversing the solution landscape. This is the "brittleness" of the NPP – even small changes in the instance drastically reshape the geometry of the solutions.

To start, let us formalize our terminology for the NPP.

Definition 1.1. Let $g \in \mathbf{R}^N$ be an instance of the NPP, and let $x \in \Sigma_N$. The *energy* of x is

$$E(x; g) := -\log_2 |\langle g, x \rangle|.$$

The *solution set* $S(E; g)$ is the set of all $x \in \Sigma_N$ that have energy at least E, i.e., that satisfy

$$|\langle g, x \rangle| \leq 2^{-E}.$$

Observe here that minimizing the discrepancy $|\langle g, x \rangle|$ corresponds to maximizing the energy $E(x; g)$. This terminology is motivated by the statistical physics literature, wherein random optimization problems are often reframed as energy maximization over a random landscape [79]. We further know that the *statistically optimal energy level* is $E = \Theta(N)$, while the best *computational energy level* (achievable in polynomial time) is $E = \Theta(\log^2 N)$.

For our purposes, an algorithm is a function $\mathcal{A}$: $\mathbf{R}^N \to \Sigma_N$ mapping instances g to partitions x. We will discuss extensions to randomized algorithms (which can depend on a random seed ω independent of g) and to $\mathbf{R}^N$-valued algorithms (which can be forced to give outputs on Σ_N via rounding) in later sections, but for our main analysis, considering deterministic Σ_N-valued algorithms will suffice. In particular, we consider the class of so-called *low degree algorithms*, given by either low degree polynomials or by functions with low *coordinate degree*. Compared to analytically-defined classes of stable algorithms (e.g. Lipschitz, etc.), these algorithms have an algebraic structure making them amenable to precise stability analysis. In addition, the *low degree heuristic* suggests that degree D algorithms (in either sense) are believed to serve as the simplest representatives for the class of $e^{\tilde{O}(D)}$-time algorithms [53]. This is a reasonable expectation for number partitioning, enabling us to translate our results into heuristic runtime bounds.

Our results show *strong low degree hardness* for the NPP at energy levels between the statistical and computational thresholds, in the sense of Huang and Sellke [54].

Definition 1.2 (Strong Low Degree Hardness [54, Def. 3])**.** A sequence of random search problems, that is, a N-indexed sequence of random input vectors

$$g_N \in \mathbf{R}^{d_N}$$

and random subsets

$$S_N = S_N(g_N) \subseteq \Sigma_N$$

exhibits *strong low degree hardness (SLDH) up to degree* $D \le o(D_N)$ if, for all sequences of degree $o(D_N)$ algorithms $\mathcal{A}_N\colon (g,\omega) \mapsto x$ with $\mathbf{E}\|\mathcal{A}(y_N)\|^2 \le O(N)$, we have

$$\mathbf{P}(\mathcal{A}_N(g_N,\omega) \in S_N) \le o(1).$$

There are two related notions of degree which we want to consider in Definition 1.2. The first is traditional polynomial degree, applicable for algorithms given in each coordinate by low degree polynomial functions of the inputs. In this case, we show

Theorem 1.3 (Results of Section 4.1)**.** *The NPP exhibits SLDH for degree D polynomial algorithms, for*

(a) $D \le o(\exp_2(\delta N/2))$ *when* $E = \delta N$ *for* $\delta > 0$*;*

(b) $D \le o(\exp_2(E/4))$ *when* $\omega(\log N) \le E \le o(N)$.

Under the low degree heuristic, this suggests that polynomial algorithms require double exponential time to achieve the statistical optimal discrepancy; given that brute-force search requires exponential time, this is strong evidence that polynomial algorithms are poor models for the NPP.

Thus, we turn to the second, more general notion of *coordinate degree*: a function $f\colon \mathbf{R}^N \to \mathbf{R}$ has coordinate degree D if it can be expressed as a linear combination of functions depending on combinations of no more than D coordinates. While related to polynomial degree, this enables us to consider a far broader class of algorithms, in which case we show

Theorem 1.4 (Results of Section 4.2)**.** *The NPP exhibits SLDH for coordinate degree D algorithms, for*

(a) $D \le o(N)$ *when* $E = \delta N$ *for* $\delta > 0$*;*

(b) $D \le o(E/\log^2 N)$ *when* $\omega(\log^2 N) \le E \le o(N)$.

These results are likely to be the best-possible under the low degree heuristic, which we discuss in Remark 4.15. In particular, the energy-degree tradeoff of $D \le \tilde{o}(E)$ implies finding solutions with energy E requires time $e^{\tilde{\Omega}(E)}$, and as we'll show, it is possible to achieve such discrepancies via a restricted exponential-time search. Given this, our method produces a sharp energy-runtime tradeoff, indicating there are no nontrivial algorithms that save more than a polylogarithmic factor in the runtime exponent over brute-force search. Overall, our approach towards Theorem 1.3 and Theorem 1.4 suggest that in the case of problems with brittle solution geometry, conditional landscape obstructions are an extremely powerful tool for proving algorithmic hardness.

The rest of the thesis is organized as follows. We review the low degree heuristic and work with low coordinate degree algorithms in Section 2. In particular, we provide a self-contained introduction to coordinate degree and related decompositions of L^2 functions in Section 2.1. Our main results then constitute Section 4; after giving an overview of our proof strategy, we prove Theorem 1.3 in Section 4.1, and likewise prove Theorem 1.4 in Section 4.2. We conclude in Section 4.3 by extending our results to the case of $\mathbf{R}^N$-valued algorithms and finish by discussing directions for future research.

1.4 Conventions and Fundamentals

We use the standard Bachmann-Landau notations $o(\cdot), O(\cdot), \omega(\cdot), \Omega(\cdot), \Theta(\cdot)$, in the limit $N \to \infty$. We abbreviate $f(N) \asymp g(N)$, $f(N) \ll g(N)$, or $f(N) \gg g(N)$ when $f(N) = \Theta(g(N))$, $f(N) = o(g(N))$, $f(N) = \omega(g(N))$, respectively. In addition, we write $f(N) \propto g(N)$, $f(N) \lesssim g(N)$, or

$f(N) \gtrsim g(N)$ when there exists an N-independent constant C such that $f(N) = Cg(N)$, $f(N) \leq Cg(N)$, or $f(N) \geq Cg(N)$ for all N, respectively.

We write $[N] \coloneqq \{1, ..., N\}$. If $S \subseteq [N]$, then we write $\overline{S} \coloneqq [N] \setminus S$ for the complimentary set of indices. If $x \in \mathbf{R}^N$ and $S \subseteq [N]$, then the *restriction of x to the coordinates in S* is the vector x_S with

$$(x_S)_i \coloneqq \begin{cases} x_i & i \in S, \\ 0 & \text{else.} \end{cases}$$

In particular, for $x, y \in \mathbf{R}^N$, $\langle x_S, y \rangle = \langle x, y_S \rangle = \langle x_S, y_S \rangle$.

On $\mathbf{R}^N$, we write $\|\cdot\|$ for the Euclidean norm, and $B_r(x) \coloneqq \{y \in \mathbf{R}^N : \|y - x\| < r\}$ for the Euclidean ball of radius r around x.

We use $\mathcal{N}(\mu, \sigma^2)$ to denote the scalar Normal distribution with given mean and variance. In addition, we write "i.i.d." to mean independently and identically distributed, and "r.v." to mean random variable (or random vector, if it is clear from context).

For $p \in [0, 1]$ and a pair (g, g') of standard Normal random vectors, we say (g, g') are *p-correlated* if g' is distributed as

$$g' = pg + \sqrt{1 - p^2}\tilde{g},$$

where $\tilde{g}$ is an independent copy of g.

all 1-dimensional projections $\langle g, x \rangle, \langle g', x \rangle$ have covariance matrix proportional to $\left(\begin{smallmatrix} 1 & p \\ p & 1 \end{smallmatrix}\right)$; we denote such a pair by $g' \sim_p g$. or, $g' \sim_p g$

We say (g, g') are *p-resampled* if g is a standard Normal random vector and g' is drawn as follows: for each $i \in [N]$ independently,

$$g'_i = \begin{cases} g_i & \text{with probability } p, \\ \text{drawn from } \mathcal{N}(0, 1) & \text{with probability } 1 - p. \end{cases}$$

We denote such a pair by $g' \sim \mathcal{L}_p(g)$.

In both cases, g and g' are marginally multivariate standard Normal and have entrywise correlation p.

— coordinate degree

Let γ_N be the N-dimensional standard Normal measure on $\mathbf{R}^N$. The *N-dimensional Gaussian space* is the space $L^2(\mathbf{R}^N, \gamma^N)$ of L^2 functions of N i.i.d. standard Normal r.v.s.

For $g \in \mathbf{R}^N$ and $S \subseteq [N]$.

For more details, see [70, § 1.3] or [85, § 8.3].

$$V_S \coloneqq \{f \in L^2(\gamma_N) : f(g) \text{ depends only on } g_S\},$$
$$V_{\leq D} \coloneqq \sum_{\substack{J \subseteq [N] \\ |J| \leq D}} V_T.$$

These subsets describe functions which only depend on some subset of coordinates, or on some bounded number of coordinates. Note that $V_{[N]} = V_{\leq N} = L^2(\mathbf{R}^N, \pi^{\otimes N})$.

The *coordinate degree* of a function $f \in L^2(\gamma_N)$ is defined as $\min\{D : f \in V_{\leq D}\}$.

Note that if f is a degree D polynomial, then it has coordinate degree at most D.

Moreover, we have

[85, Exer. 8.18]

$$p^D \mathbf{E}\|f(g)\|^2 \leq \mathbf{E}_{g' \sim \mathcal{L}_p(g)}[f(g) \cdot f(g')] \leq \mathbf{E}\|f(g)\|^2$$

Theorem 4.19(a)

— algorithms

A *(randomized) algorithm* is a measurable function $\mathcal{A}$: $(g, \omega) \mapsto x \in \Sigma_N$, where $\omega \in \Omega_N$ is an independent random variable. Such an $\mathcal{A}$ is *deterministic* if it does not depend on ω.

With the notions of low coordinate degree functions or low degree polynomials in hand, we can consider algorithms based on such functions.

A *polynomial algorithm* is an algorithm $\mathcal{A}(g, \omega)$ where each coordinate of $\mathcal{A}(g, \omega)$ is given by a polynomial in the N entries of g. If $\mathcal{A}$ is a polynomial algorithm, then it has degree D if each coordinate has degree at most D (with at least one equality).

Suppose an algorithm $\mathcal{A}(g, \omega)$ is such that each coordinate of $\mathcal{A}(-, \omega)$ is in $L^2(\mathbf{R}^N, \pi^{\otimes N})$. Then, the *coordinate degree* of $\mathcal{A}$ is the maximum coordinate degree of $\mathcal{A}(-, \omega)$.

With Theorem 2.8 and Theorem 2.15, we can derive the following algorithmic L^2 stability bound.

Throughout the remainder of this thesis, we will make use of the following general results:

Lemma 1.5 (Normal Small-Probability Estimate)**.** *Let $E, \sigma^2 > 0$, and suppose $Z \mid \mu \sim \mathcal{N}(\mu, \sigma^2)$. Then*

$$\mathbf{P}(|Z| \leq 2^{-E} \mid \mu) \leq \exp_2\left(-E - \frac{1}{2}\log_2(\sigma^2) + O(1)\right). \tag{1.3}$$

Proof: Observe that conditional on μ, the distribution of Z is bounded as

$$\varphi_{Z|\mu}(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \leq \left(2\pi\sigma^2\right)^{-1/2}.$$

Integrating over $|z| \leq 2^{-E}$ then gives (1.3), via

$$\mathbf{P}(|Z| \leq 2^{-E}) = \int_{|z| \leq 2^{-E}} \left(2\pi\sigma^2\right)^{-1/2} \mathrm{d}z \leq 2^{-E - \frac{1}{2}\log_2(2\pi\sigma^2) + 1}. \qquad \square$$

Note that (1.3) is a decreasing function of σ^2. Thus, if there exists γ with $\sigma^2 \geq \gamma > 0$, then (1.3) is bounded by $\exp_2(-E - \log_2(\gamma)/2 + O(1))$.

Lemma 1.6 (Chernoff-Hoeffding)**.** *Suppose that $K \leq N/2$, and let $h(x) = -x\log_2(x) - (1-x)\log_2(x)$ be the binary entropy function. Then, for $p := K/N$,*

$$\sum_{k \leq K} \binom{N}{k} \leq \exp_2(Nh(p)) \leq \exp_2\left(2Np\log_2\left(\frac{1}{p}\right)\right).$$

Proof: Consider a $\mathrm{Bin}(N,p)$ random variable S. Summing its PMF from 0 to K, we have

$$1 \geq \mathbf{P}(S \leq K) = \sum_{k \leq K} \binom{N}{k} p^k (1-p)^{N-k} \geq \sum_{k \leq K} \binom{N}{k} p^K (1-p)^{N-K}.$$

The last inequality follows by multiplying each term by $(p/(1-p))^{K-k} \leq 1$. Rearranging gives

$$\begin{aligned}
\sum_{k \leq K} \binom{N}{k} &\leq p^{-K}(1-p)^{-(N-K)} \\
&= \exp_2(-K\log_2(p) - (N-K)\log_2(1-p)) \\
&= \exp_2\left(N \cdot \left(-\frac{K}{N}\log_2(p) - \left(\frac{N-K}{N}\right)\log_2(1-p)\right)\right) \\
&= \exp_2(N \cdot (-p\log_2(p) - (1-p)\log_2(1-p))) = \exp_2(Nh(p)).
\end{aligned}$$

The final equality then follows from the bound $h(p) \leq 2p\log_2(1/p)$ for $p \leq 1/2$. □

1.5 Stability of Low Degree Algorithms

•

2 Low Degree Algorithms

For our purposes, an *algorithm* is a function which takes as input an instance $g \in \mathbf{R}^N$ and outputs some $x \in \Sigma_N$. This definition can be extended to functions giving outputs on $\mathbf{R}^N$ and rounding to a vertex on the hypercube Σ_N, which we consider in Section 4.3. Alternatively, we could consider *randomized algorithms* by taking as additional input some randomness ω independent of the problem instance. However, as this extension requires only minor changes, which we describe in Remark 4.8, most of our analysis will focus on the deterministic case.

The category of algorithms we consider are known as *low degree algorithms*. We treat two closely related notions of degree: first is *polynomial degree*, in which we assume our algorithms are given coordinate-wise by polynomials of some bounded degree. The second, more general notion is *coordinate degree*, which roughly counts how many coordinates can interact nonlinearly; this can be applied to arbitrary algorithms given by L^2 functions. While polynomial algorithms are widely known and studied, low coordinate degree algorithms were first introduced in Hopkins' thesis [53], and were later used by Brennan et al. [21] and Mossel et al. [65], [50] (although in the latter case, they were shown to be equivalent to polynomial algorithms). Compared to analytically-defined classes of algorithms (e.g. Lipschitz), these low degree algorithms have an algebraic structure that we can exploit to precisely control their stability properties.

As mentioned in the introduction, our goal is to show *strong low degree hardness*, meaning that low degree algorithms (either meaning low polynomial degree or low coordinate degree) fail to find solutions to the NPP with high probability. However, our proofs only use the low degree assumption to apply stability bounds: roughly, a stable algorithm cannot "overcome" the gaps between solutions for two closely-related instances of the NPP. Why, then, do we restrict to low degree algorithms specifically? The main reason is the *low degree heuristic*.

For nice random optimization problems, there exists a successful degree D algorithm if and only if there exists a successful algorithm running in time $e^{\tilde{O}(D)}$.

This heuristic was first proposed in Hopkins' thesis [53], and later expanded upon by Kunisky, Wein, and Bandeira [72], although this was primarily in the context of low degree polynomials for hypothesis testing. Kunisky later expanded these results when applying low coordinate degree methods towards hypothesis testing [70], [71]. Huang and Sellke then observed that strong low degree hardness up to degree $o(N)$ can be thought of as evidence of a random optimization problem requiring exponential $e^{\tilde{\Omega}(N)}$ time to find globally optimal solutions [54]. They prove strong low degree hardness for a variety of canonical problems: optimization of pure k-spin glasses, symmetric binary perceptrons, and random k-SAT, to name a few, most of which are optimal under the low degree heuristic. However, Huang and Mossel's work on broadcasting on trees, this heuristic breaks down: degree $D \leq O(\log N)$ algorithms fail despite there existing a linear-time algorithm known as Belief Propagation [50]. To this end, the authors suggest this discrepancy arises from the requirement of "depth" in the Belief Propagation algorithm – roughly, despite running in linear time, this algorithm still struggles in practice in the "hard" regime. As a takeaway, we can surmise that the low degree heuristic is reasonable for describing random search problems involving optimization of a "flat" structure, in which algorithmic complexity cannot hide behind N-independent factors. Thus, having an explicit handle on algorithm degree enables us to both control stability and extend our results to rule out general polynomial-time algorithms.

We start by introducing the theory of *Efron-Stein decompositions* and coordinate degree, and demonstrate how elementary Fourier analysis can give straightforward L^2 stability properties. We then review the theory of *Hermite polynomials*, which gives altered L^2 bounds for polynomial functions. This section then concludes with a discussion of our terminology for low polynomial degree and low coordinate degree algorithms, and we summarize our stability analysis in Proposition 2.19.

2.1 Coordinate Degree and L^2 Stability

First, we consider a general class of putative algorithms, and construct the "coordinate decomposition" underlying the notion of coordinate degree. Given this notion, deriving stability bounds becomes a straightforward piece of functional analysis. To start, recall the notion of L^2 functions.

Definition 2.1. Let π be a probability distribution on $\mathbf{R}$. The L^2 space $L^2(\mathbf{R}^N, \pi^{\otimes N})$ is the space of functions $f\colon \mathbf{R}^N \to \mathbf{R}$ with finite L^2 norm:

$$\mathbf{E}[f^2] := \int_{\mathbf{R}^N} f(x)^2 \, \mathrm{d}\pi^{\otimes N}(x) < \infty.$$

Alternatively, this is the space of L^2 functions of N i.i.d. random variables x_i, distributed as π.

This is an extremely broad class of functions; for instance, all bounded functions are L^2. Given any function $f \in L^2(\mathbf{R}^N, \pi^{\otimes N})$, we can then consider how it depends on various subsets of the N input coordinates. In principle, everything about f should be reflected in how it acts on all possible such subsets. To formalize this intuition, define the following coordinate projection.

Definition 2.2. Let $f \in L^2(\mathbf{R}^N, \pi^{\otimes N})$ and $J \subseteq [N]$. The *projection of f onto J* is the function $f^{\subseteq J} \colon \mathbf{R}^N \to \mathbf{R}$ given by

$$f^{\subseteq J}(x) := \mathbf{E}[f(x_1, ..., x_n) \mid \{x_i, i \in J\}] = \mathbf{E}[f(x) \mid x_J].$$

Intuitively $f^{\subseteq J}$ is f with the $\overline{J}$ coordinates re-randomized, so $f^{\subseteq J}$ only depends on the coordinates in J. However, depending on how f accounts for higher-order interactions, it might be the case that $f^{\subseteq J}$ is fully described by some $f^{\subseteq J'}$, for $J' \subsetneq J$. What we really want is to decompose f as

$$f = \sum_{S \subseteq [N]} f^{=S}, \tag{2.1}$$

where each $f^{=S}$ only depends on the coordinates in S, but not any smaller subset. That is, if $T \subsetneq S$ and g depends only on the coordinates in T, then $\langle f^{=S}, g \rangle = 0$.

This decomposition, often called the *Efron-Stein*, *orthogonal*, or *Hoeffding* decomposition, does indeed exist. Its applications in statistics come from the fact that it provides a way of decomposing the total variance of a function into the components coming from specific sets of coordinates, a step which underlies the ANOVA methodology. These low coordinate degree decompositions have also been used in computational chemistry: see the review by Li et al. [76] for more details. The Efron-Stein decomposition exhibits the following combinatorial construction; our presentation largely follows [85, § 8.3], as well as the paper [70].

The motivating fact is that for any $J \subseteq [N]$, we should have

$$f^{\subseteq J} = \sum_{S \subseteq J} f^{=S}. \tag{2.2}$$

Intuitively, $f^{\subseteq J}$ captures everything about f depending on the coordinates in J, and each $f^{=S}$ captures precisely the interactions within each subset S of J. The construction of $f^{=S}$ proceeds by inverting this formula.

First, we consider the case $J = \emptyset$. It is clear that $f^{=\emptyset} = f^{\subseteq \emptyset}$, which, by Definition 2.2 is the constant function $\mathbf{E}[f]$. Next, if $J = \{j\}$ is a singleton, then (2.2) gives

$$f^{\subseteq \{j\}} = f^{=\emptyset} + f^{=\{j\}},$$

and as $f^{\subseteq \{j\}}(x) = \mathbf{E}\big[f \mid x_j\big]$, we get

$$f^{=\{j\}} = \mathbf{E}\big[f \mid x_j\big] - \mathbf{E}[f].$$

This function only depends on x_j; all other coordinates are averaged over, so this coordinate piece measures how the expectation of f changes given x_j.

Continuing on to sets of two coordinates, some brief manipulation gives, for $J = \{i, j\}$,

$$\begin{aligned} f^{\subseteq\{i,j\}} &= f^{=\emptyset} + f^{=\{i\}} + f^{=\{j\}} + f^{=\{i,j\}} \\ &= f^{\subseteq\emptyset} + \left(f^{\subseteq\{i\}} - f^{\subseteq\emptyset}\right) + \left(f^{\subseteq\{j\}} - f^{\subseteq\emptyset}\right) + f^{=\{i,j\}}, \\ \therefore f^{=\{i,j\}} &= f^{\subseteq\{i,j\}} - f^{\subseteq\{i\}} - f^{\subseteq\{j\}} + f^{\subseteq\emptyset}. \end{aligned}$$

We can imagine that this accounts for the two-way interaction of i and j, namely $f^{\subseteq\{i,j\}} = \mathbf{E}\big[f \mid x_i, x_j\big]$, while "correcting" for the one-way effects of x_i and x_j individually. Inductively, we can continue in this way and define all the $f^{=J}$ via inclusion-exclusion.

$$f^{=J} := \sum_{S\subseteq J} (-1)^{|J|-|S|} f^{\subseteq S} = \sum_{S\subseteq J} (-1)^{|J|-|S|} \mathbf{E}[f \mid x_S].$$

This construction, along with some direct calculations, leads to the following theorem.

Theorem 2.3 ([85, Thm 8.35])**.** *Each $f \in L^2(\mathbf{R}^N, \pi^{\otimes N})$ has a unique decomposition as*

$$f = \sum_{S\subseteq[N]} f^{=S},$$

known as the Efron-Stein decomposition*, where the functions $f^{=S} \in L^2(\mathbf{R}^N, \pi^{\otimes N})$ satisfy:*

(1) $f^{=S}$ depends only on the coordinates in S;

(2) if $T \subsetneq S$ and $g \in L^2(\mathbf{R}^N, \pi^{\otimes N})$ only depends on coordinates in T, then $\langle f^{=S}, g\rangle = 0$.

In addition, this decomposition has the following properties.

(3) Condition (2) holds whenever $S \not\subseteq T$.

(4) The decomposition is orthogonal: $\langle f^{=S}, f^{=T}\rangle = 0$ for $S \neq T$.

(5) $\sum_{S\subseteq T} f^{=S} = f^{\subseteq T}$.

(6) For each $S \subseteq [N]$, $f \mapsto f^{=S}$ is a linear operator.

In summary, this decomposition of $L^2(\mathbf{R}^N, \pi^{\otimes N})$ functions into their different interaction levels not only exists, but is orthogonal, enabling us to apply tools from elementary Fourier analysis.

Theorem 2.3 further implies that we can define subspaces of $L^2(\mathbf{R}^N, \pi^{\otimes N})$ (see also [70, § 1.3])

$$\begin{aligned} V_J &:= \{f \in L^2(\mathbf{R}^N, \pi^{\otimes N}) : f = f^{\subseteq J}\}, \\ V_{\leq D} &:= \sum_{\substack{J\subseteq[N] \\ |J|\leq D}} V_T. \end{aligned} \tag{2.3}$$

These capture functions which only depend on some subset of coordinates, or some bounded number of coordinates. Note that $V_{[N]} = V_{\leq N} = L^2(\mathbf{R}^N, \pi^{\otimes N})$.

Definition 2.4. The *coordinate degree* of a function $f \in L^2(\mathbf{R}^N, \pi^{\otimes N})$ is

$$\operatorname{cdeg}(f) := \max\big\{|S| : S \subseteq [N], f^{=S} \neq 0\big\} = \min\big\{D : f \in V_{\leq D}\big\}.$$

If $f = (f_1, \ldots, f_M)\colon \mathbf{R}^N \to \mathbf{R}^M$ is a multivariate function with each $f_i \in L^2(\mathbf{R}^N, \pi^{\otimes N})$, then

$$\operatorname{cdeg}(f) := \max_{i\in[M]} \operatorname{cdeg}(f_i).$$

Intuitively, the coordinate degree is the maximum size of (nonlinear) multivariate interaction that f accounts for. Of course, this degree is also bounded by N, very much unlike polynomial degree. Note as a special case that any multivariate polynomial of degree D has coordinate degree at most D. As an example, the function $x_1 + x_2$ has both polynomial degree and coordinate degree 1, while $x_1 + x_2^2$ has polynomial degree 2 and coordinate degree 1. We are especially interested in algorithms coming from functions in $V_{\leq D}$, which we term *low coordinate degree algorithms*.

As we are interested in how these algorithms behave for "close" instances, we are led to consider the following "noise operator," which measures the effect of small changes in the input on the Efron-Stein decomposition. We need the following notion of distance between instances.

Definition 2.5. For $p \in [0,1]$ and $x \in \mathbf{R}^N$, we say $y \in \mathbf{R}^N$ is *p-resampled from x*, denoted $y \sim \pi_p^{\otimes N}(x)$, if y is chosen as follows: for each $i \in [N]$, independently,

$$y_i = \begin{cases} x_i & \text{with probability } p, \\ \text{drawn from } \pi & \text{with probability } 1-p. \end{cases}$$

We say (x, y) are a *p-resampled pair*.

Note that being p-resampled and being p-correlated are rather different – for one, there is a nonzero probability that, for π a continuous probability distribution, $x = y$ when they are p-resampled, even though this almost surely never occurs if they were p-correlated.

Definition 2.6. For $p \in [0,1]$, the *noise operator* T_p is the linear operator on $L^2(\mathbf{R}^N, \pi^{\otimes N})$ defined by

$$T_p f(x) = \mathbf{E}_{y \sim \pi_p^{\otimes N}(x)}[f(y)].$$

In particular, $\langle f, T_p f \rangle = \mathbf{E}_{(x,y)\ p\text{-resampled}}[f(x) \cdot f(y)]$.

This noise operator changes the Efron-Stein decomposition, and hence the behavior of low coordinate degree functions, in a controlled way.

Lemma 2.7. *Let $p \in [0,1]$ and $f \in L^2(\mathbf{R}^N, \pi^{\otimes N})$ have decomposition $f = \sum_{S \subseteq [N]} f^{=S}$. Then*

$$T_p f(x) = \sum_{S \subseteq [N]} p^{|S|} f^{=S}.$$

Proof: Let J be a random subset formed by including each $i \in [N]$ independently with probability p. By definition, $T_p f(x) = \mathbf{E}_J[f^{\subseteq J}(x)]$ (i.e., pick a random subset of coordinates to fix, and re-randomize the rest). We know by Theorem 2.3 that $f^{\subseteq J} = \sum_{S \subseteq J} f^{=S}$, so

$$T_p f(x) = \mathbf{E}_J \left[\sum_{S \subseteq J} f^{=S} \right] = \sum_{S \subseteq [N]} \mathbf{E}_J[I(S \subseteq J)] \cdot f^{=S} = \sum_{S \subseteq [N]} p^{|S|} f^{=S},$$

since for a fixed $S \subseteq [N]$, the probability that $S \subseteq J$ is $p^{|S|}$. □

Thus, we can derive the following stability bound on low coordinate degree functions.

Theorem 2.8. *Let $p \in [0,1]$ and $f = (f_1, ..., f_M)$: $\mathbf{R}^N \to \mathbf{R}^M$ be a coordinate degree D multivariate function. Suppose that (x, y) are a p-resampled pair under $\pi^{\otimes N}$ and $\mathbf{E}\|f(x)\|^2 = 1$. Then*

$$\mathbf{E}\|f(x) - f(y)\|^2 \leq 2(1 - p^D) \leq 2(1-p)D. \tag{2.4}$$

Proof: Observe that

$$\begin{aligned} \mathbf{E}\|f(x) - f(y)\|^2 &= \mathbf{E}\|f(x)\|^2 + \mathbf{E}\|f(y)\|^2 - 2\mathbf{E}\langle f(x), f(y)\rangle \\ &= 2 - 2\left(\sum_i \mathbf{E}[f_i(x) f_i(y)]\right) \\ &= 2 - 2\left(\sum_i \langle f_i, T_p f_i\rangle\right). \end{aligned} \tag{2.5}$$

Here, we have for each $i \in [M]$ that

$$\langle f_i, T_p f_i\rangle = \left\langle \sum_{S\subseteq[N]} f_i^{=S}, \sum_{S\subseteq[N]} p^{|S|} f_i^{=S} \right\rangle = \sum_{S\subseteq[N]} p^{|S|} \left\| f_i^{=S} \right\|^2$$

by Lemma 2.7 and orthogonality. Now, as each f_i has coordinate degree at most D, the sum above can be taken only over $S \subseteq [N]$ with $0 \leq |S| \leq D$, giving the bound

$$p^D \mathbf{E}[f_i(x)^2] \leq \langle f_i, T_p f_i\rangle = \mathbf{E}\big[f_i(x) \cdot T_p f_i(x)\big] \leq \mathbf{E}[f_i(x)^2].$$

Summing up over i and using that $\mathbf{E}\|f(x)\|^2 = 1$ yields

$$p^D \leq \sum_i \langle f_i, T_p f_i\rangle = \mathbf{E}\langle f(x), f(y)\rangle \leq 1.$$

Finally, we can substitute into (2.5) to get[4]

$$\mathbf{E}\|f(x) - f(y)\|^2 \leq 2 - 2p^D = 2(1-p^D) \leq 2(1-p)D. \qquad \square$$

2.2 Hermite Polynomials

Alternatively, we can consider the much more restrictive (but more concrete) class of honest polynomials. When considered as functions of independent Normal variables, such functions admit a simple description in terms of *Hermite polynomials*, which enables us to prove bounds similar to Theorem 2.8. This theory is classical, and we encourage the reader to consult [85, §11] for details.

Definition 2.9. Let γ_N be the N-dimensional standard Normal measure on $\mathbf{R}^N$. The *N-dimensional Gaussian space* is the space $L^2(\mathbf{R}^N, \gamma^N)$ of L^2 functions of N i.i.d. standard Normal r.v.s.

Note that under the usual L^2 inner product $\langle f, g\rangle = \mathbf{E}[f \cdot g]$, this is a separable Hilbert space.

[4] The last inequality follows from $(1 - p^D) = (1-p)(1 + p + p^2 + ...p^{D-1})$; the bound is tight for $p \approx 1$.

It is a well-known fact that the monomials $1, z, z^2, \ldots$ form a complete basis for $L^2(\mathbf{R}, \gamma)$ [85, Thm 11.22]. However, these are far from an orthonormal "Fourier" basis; for instance, we know $\mathbf{E}[z^2] = 1$ for $z \sim \mathcal{N}(0,1)$. By the Gram-Schmidt process, these monomials can be converted into the *(normalized) Hermite polynomials* h_j for $j \geq 0$, given by

$$h_0(z) = 1, \qquad h_1(z) = z, \qquad h_2(z) = \frac{z^2 - 1}{\sqrt{2}}, \qquad h_3(z) = \frac{z^3 - 3z}{\sqrt{6}}, \qquad \ldots. \tag{2.6}$$

Note here that each h_j is a degree j polynomial. The following is well-known.

Theorem 2.10 ([85, Prop 11.30])**.** *The polynomials h_j form a complete orthonormal basis for $L^2(\mathbf{R}, \gamma)$.*

To extend this to $L^2(\mathbf{R}^N, \gamma^N)$, we can take products. For a multi-index $\alpha \in \mathbb{N}^N$, we define the multivariate Hermite polynomial $h_\alpha \colon \mathbf{R}^N \to \mathbf{R}$ as

$$h_\alpha(z) := \prod_{j=1}^{N} h_{\alpha_j}(z_j).$$

The degree of h_α is clearly $|\alpha| = \sum_j \alpha_j$.

Theorem 2.11. *The Hermite polynomials $(h_\alpha)_{\alpha \in \mathbb{N}^N}$ form a complete orthonormal basis for $L^2(\mathbf{R}^N, \gamma^N)$. In particular, every $f \in L^2(\mathbf{R}^N, \gamma^N)$ has a unique expansion (converging in the L^2 norm) as*

$$f(z) = \sum_{\alpha \in \mathbb{N}^N} \hat{f}(\alpha) h_\alpha(z).$$

As a consequence of the uniqueness of the expansion in Theorem 2.11, we see that polynomials are their own Hermite expansion. Namely, let $H^{\leq k} \subseteq L^2(\mathbf{R}^N, \gamma^N)$ be the subset of multivariate polynomials of degree at most k. Then, any $f \in H^{\leq k}$ can be Hermite expanded as

$$f(z) = \sum_{\alpha \in \mathbb{N}^N} \hat{f}(\alpha) h_\alpha(z) = \sum_{|\alpha| \leq k} \hat{f}(\alpha) h_\alpha(z).$$

Thus, $H^{\leq k}$ is the closed linear span of the set $\{h_\alpha : |\alpha| \leq k\}$.

When working with honest polynomials, the traditional notion of correlation is a much more natural measure of "distance" between inputs.

Definition 2.12. Let (x, y) be a pair of N-dimensional standard Normal vectors. We say (x, y) are *p-correlated* if (x_i, y_i) are p-correlated for each $i \in [N]$, and these pairs are mutually independent.

Analogously to the Efron-Stein setting, we can consider the resulting "noise operator" as a way of measuring the effect on a function of a small change in the input.

Definition 2.13. For $p \in [0, 1]$, the *Gaussian noise operator* T_p is the linear operator on $L^2(\mathbf{R}^N, \gamma^N)$:

This operator admits a more classical description in terms of the Ornstein-Uhlenbeck semigroup, but we will not need that connection here. As it happens, a straightforward computation with the Normal moment generating function gives the following lemma.

Lemma 2.14 ([85, Prop 11.37])**.** *Let $p \in [0,1]$ and $f \in L^2(\mathbf{R}^N, \gamma^N)$. Then, $T_p f$ has Hermite expansion*

$$T_p f = \sum_{\alpha \in \mathbb{N}^N} p^{|\alpha|} \hat{f}(\alpha) h_\alpha,$$

and in particular,

$$\langle f, T_p f \rangle = \sum_{\alpha \in \mathbb{N}^N} p^{|\alpha|} \hat{f}(\alpha)^2.$$

With this in hand, we can prove a similar stability bound to Theorem 2.8.

Theorem 2.15. *Let $p \in [0,1]$ and $f = (f_1, ..., f_M) \colon \mathbf{R}^N \to \mathbf{R}^M$ be a multivariate degree D polynomial. Suppose that (x, y) are a p-correlated pair of standard Normal vectors and $\mathbf{E}\|f(x)\|^2 = 1$. Then,*

$$\mathbf{E}\|f(x) - f(y)\|^2 \leq 2(1 - p^D) \leq 2(1-p)D. \tag{2.7}$$

Proof: The proof is almost identical to that of Theorem 2.8 (see also [38, Lem. 3.4]). The main modification is in realizing that for each f_i, having degree at most D implies that $\hat{f}_i(\alpha) = 0$ for $|\alpha| > D$. Thus, as $p^D \leq p^s \leq 1$ for all $s \leq D$, we can apply Lemma 2.14 to get

$$p^D \mathbf{E}[f_i(x)^2] \leq \langle f_i, T_p f_i \rangle = \sum_{\alpha \in \mathbb{N}^N : |\alpha| \leq D} p^{|\alpha|} \hat{f}_i(\alpha)^2 \leq \mathbf{E}[f_i(x)^2].$$

From here, the proof proceeds as before. □

As a comparison to the case for functions with coordinate degree D, notice that Theorem 2.15 gives, generically, a much looser bound. In exchange, being able to use p-correlation as a "metric" on the input domain will turn out to offer significant benefits in the arguments which follow, justifying equal consideration of both classes of functions.

2.3 Stability of Low Degree Algorithms

We now formalize our notion of "algorithm" from Section 1.3.

Definition 2.16. A *(randomized) algorithm* is a measurable function $\mathcal{A} \colon (g, \omega) \mapsto x \in \Sigma_N$, where $\omega \in \Omega_N$ is an independent random variable. Such an $\mathcal{A}$ is *deterministic* if it does not depend on ω.

With the notions of low coordinate degree functions or low degree polynomials in hand, we can consider algorithms based on such functions.

Definition 2.17. A *polynomial algorithm* is an algorithm $\mathcal{A}(g, \omega)$ where each coordinate of $\mathcal{A}(g, \omega)$ is given by a polynomial in the N entries of g. If $\mathcal{A}$ is a polynomial algorithm, then it has degree D if each coordinate has degree at most D (with at least one equality).

Definition 2.18. Suppose an algorithm $\mathcal{A}(g, \omega)$ is such that each coordinate of $\mathcal{A}(-, \omega)$ is in $L^2(\mathbf{R}^N, \pi^{\otimes N})$. Then, the *coordinate degree* of $\mathcal{A}$ is the maximum coordinate degree of $\mathcal{A}(-, \omega)$.

With Theorem 2.8 and Theorem 2.15, we can derive the following algorithmic L^2 stability bound.

Proposition 2.19 (Low Degree Stability – [54, Prop. 1.9])**.** *Suppose we have a deterministic algorithm $\mathcal{A}$ with degree (resp. coordinate degree) $\leq D$ and norm $\mathbf{E}\|\mathcal{A}(g)\|^2 \leq CN$. Then, for inputs g, g' which are $(1-\varepsilon)$-correlated (resp. $(1-\varepsilon)$-resampled),*

$$\mathbf{E}\|\mathcal{A}(g) - \mathcal{A}(g')\|^2 \leq 2CD\varepsilon N, \tag{2.8}$$

and thus

$$\mathbf{P}\left(\|\mathcal{A}(g) - \mathcal{A}(g')\| \geq 2\sqrt{\eta N}\right) \leq \frac{CD\varepsilon}{2\eta} \asymp \frac{D\varepsilon}{\eta}. \tag{2.9}$$

Proof: Let $C' := \mathbf{E}\|\mathcal{A}(g)\|^2$ and define the rescaling $\mathcal{A}' := \mathcal{A}/\sqrt{C'}$. Then, by Theorem 2.15 (or Theorem 2.8, in the low coordinate degree case), we have

$$\mathbf{E}\|\mathcal{A}'(g) - \mathcal{A}'(g')\|^2 = \frac{1}{C'}\mathbf{E}\|\mathcal{A}(g) - \mathcal{A}(g')\|^2 \leq 2D\varepsilon.$$

Multiplying by C' gives (2.8) (as $C' \leq CN$). Finally, (2.9) follows from Markov's inequality. □

Remark 2.20. Note that Proposition 2.19 also holds for randomized algorithms. Namely, if $\mathcal{A}(g, \omega)$ is a randomized algorithm with polynomial or coordinate degree D and $\mathbf{E}_{g,\omega}\|\mathcal{A}(g,\omega)\|^2 \leq CN$, then applying Markov's inequality to $\omega \mapsto \mathbf{E}[\|\mathcal{A}(g,\omega)\|^2 \mid \omega]$ allows us to reduce to the deterministic case, possibly after adjusting C.

3 Proofs of Strong Low Degree Hardness

4 Proofs of Strong Low Degree Hardness

In this section, we prove Theorem 1.3 and Theorem 1.4 – that is, we exhibit strong low degree hardness for both low polynomial degree and low coordinate degree algorithms.

Our argument utilizes what can be thought of as a "conditional" version of the overlap gap property. Traditionally, proofs of algorithmic hardness use the overlap gap property as a global obstruction: one shows that with high probability, there are no tuples of good solutions to a family of correlated instances which are all roughly the same distance apart. Here, however, we show a local obstruction; we condition on being able to solve a single instance and show that after a small change to the instance, we cannot guarantee any solutions will exist close to the first one. This is an instance of the "brittleness," so to speak, that makes NPP so frustrating to solve; even small changes in the instance break the landscape geometry, so that even if solutions exist, there is no way to know where they will end up.

This conditional landscape obstruction approach is partially inspired by Huang and Sellke's recent work on strong low degree hardness for finding optima in spin glasses [54]. However, a main reason for not appealing to an OGP-style result is Gamarnik and Kızıldağ's disproof of the m-OGP for sublinear energy levels [39, Thm. 2.5].

Our conditional obstruction (Proposition 4.5 in the low degree polynomial case, and Proposition 4.12 in the low coordinate degree case) is established by a first moment computation. That is, we show that given "correlated" instances g, g' and a point $x \in \Sigma_N$ such that g', x are conditionally independent given g, then any fixed point $x' \in \Sigma_N$ has low probability of solving g'; then, the same must hold for all x' in a suitably small neighborhood of x. This is similar to the proof of the OGP in the linear energy regime [39], but our method allows us to work with sublinear energy levels. Heuristically, this is because the cardinality of neighborhoods of x grows exponentially in N, which means that the number of m-tuples of such points grows much faster than any sublinearly small probability. In contrast, the disproof of the OGP in the sublinear energy regime of Gamarnik and Kızıldağ follows from a second moment computation: they show that the majority of pairs of m -tuples of solutions are nearly "uncorrelated," which again implies that globally, looking at large ensembles of solutions fails to capture the brittleness of the NPP for cardinality reasons.

The proof of Theorem 1.3, stated formally as Theorem 4.6 and Theorem 4.7, is as follows.[5] Let E be an energy level and D a maximum algorithm degree, both depending on N. We assume that D is bounded by a level depending on E and N, corresponding to the low degree regime in which we want to show hardness. We then choose parameters η (depending on E and N) and ε (depending on E, D, and N). As described in Section 2, assume $\mathcal{A}$ is a deterministic, Σ_N-valued algorithm with polynomial degree at most D. Our goal is to show that for our choices of η and ε,

$$\mathbf{P}(\mathcal{A}(g) \in S(E; g)) \to 0$$

as $N \to \infty$. This is done in the following steps.

(a) Consider a $(1-\varepsilon)$-correlated pair g, g' of NPP instances. These are N-dimensional standard Normal vectors which are p-correlated for $p = 1 - \varepsilon$ (when considering coordinate degree, we instead require them to be p-resampled).

(b) For ε small, g and g' have correlation close to 1. By Proposition 2.19, this implies that the outputs of a low degree polynomial algorithm $\mathcal{A}$ will be within distance $2\sqrt{\eta N}$ of each other with high probability.

(c) For η small and fixed $\mathcal{A}(g)$, Proposition 4.5 shows that conditional on g, g' has no solutions within distance $2\sqrt{\eta N}$ of $\mathcal{A}(g)$. This is the conditional landscape obstruction we described above.

(d) Put together, these points imply that it is unlikely for $\mathcal{A}$ to find solutions to *both* g and g' such that the stability guarantee of Proposition 2.19 holds. By the positive correlation statement in Lemma 4.2, we conclude that $\mathcal{A}(g) \notin S(E; g)$ with high probability.

We can summarize the parameters in our argument in the following table.

[5] The proof of Theorem 1.4 requires only minor modifications.

Parameter	Meaning	Desired Direction	Intuition
N	Dimension	-	Showing strong hardness *asymptotically*, want uniformly large.
E	Energy; want x such that $\lvert\langle g, x\rangle\rvert \le 2^{-E}$	Small	Smaller E rules out weaker solutions; know $\Omega(\log^2 N) \le E \le \Theta(N)$.
D	Algorithm degree	Large	Higher degree means more complex (i.e., longer time) algorithms fail.
ε	Distance between g and g'	Small	Want to show that small changes in instance lead to "breaking" of landscape.
η	Instability; $\lVert\mathcal{A}(g) - \mathcal{A}(g')\rVert \le 2\sqrt{\eta N}$, for g and g' close	Large (but bounded by E, N)	Large η indicates a more unstable algorithm; want to show that even weakly stable algorithms fail.

Table 1: Explanation of Parameters

For the remainder of this section, we first show strong low degree hardness for polynomial algorithms, and then for low coordinate degree algorithms. Throughout, we aim to keep constants as explicit as possible, to clarify the nature in which ε and η behave in the limit as $N \to \infty$. We end by interpreting our results through the lens of the low degree heuristic, as well as discuss the extensions needed to consider randomized Σ_N-valued algorithms.

4.1 Hardness for Low Degree Polynomial Algorithms

First, we consider the case where $\mathcal{A}$ is a polynomial algorithm with degree D. Let g, g' be $(1-\varepsilon)$-correlated standard Normal r.v.s, and suppose $x \in \Sigma_N$ depends only on g. Furthermore, let $\eta > 0$ be a parameter chosen in a manner specified later. We define the events

$$\begin{aligned} S_{\text{solve}} &:= \{\mathcal{A}(g) \in S(E; g), \mathcal{A}(g') \in S(E; g')\}, \\ S_{\text{stable}} &:= \left\{\lVert\mathcal{A}(g) - \mathcal{A}(g')\rVert \le 2\sqrt{\eta N}\right\}, \\ S_{\text{cond}}(x) &:= \begin{Bmatrix} \nexists\ x' \in S(E; g') \text{ such that} \\ \lVert x - x'\rVert \le 2\sqrt{\eta N} \end{Bmatrix}. \end{aligned} \tag{4.1}$$

Intuitively, the first two events ask that the algorithm solves both instances and is stable, respectively. The last event, which depends on x, corresponds to the conditional landscape obstruction: for an x depending only on g, there is no solution to g' which is close to x.

Lemma 4.1. *For $x := \mathcal{A}(g)$, we have $S_{\text{solve}} \cap S_{\text{stable}} \cap S_{\text{cond}}(x) = \emptyset$.*

Proof: Suppose that S_{solve} and S_{stable} both occur. Letting $x := \mathcal{A}(g)$ (which only depends on g) and $x' := \mathcal{A}(g')$, we know $x' \in S(E; g')$ and is within distance $2\sqrt{\eta N}$ of x, contradicting $S_{\text{cond}}(x)$. □

Now, define $p^{\mathrm{cor}}_{\mathrm{solve}}$ as the probability that the algorithm solves a single random instance:

$$p^{\mathrm{cor}}_{\mathrm{solve}} = \mathbf{P}(\mathcal{A}(g) \in S(E;g)). \tag{4.2}$$

We have the following positive correlation bound, which enables us to handle pairs of instances.

Lemma 4.2. *For g, g' being $(1-\varepsilon)$-correlated, we have*

$$\mathbf{P}(S_{\mathrm{solve}}) = \mathbf{P}(\mathcal{A}(g) \in S(E;g), \mathcal{A}(g') \in S(E;g')) \geq (p^{\mathrm{cor}}_{\mathrm{solve}})^2.$$

Proof: Let $\tilde{g}, g^{(0)}, g^{(1)}$ be three i.i.d. copies of g, and observe that g, g' are jointly representable as

$$g = \sqrt{1-\varepsilon}\tilde{g} + \sqrt{\varepsilon}g^{(0)}, \qquad\qquad g' = \sqrt{1-\varepsilon}\tilde{g} + \sqrt{\varepsilon}g^{(1)}.$$

Thus, since g, g' are conditionally independent given $\tilde{g}$, we have

$$\begin{aligned}
\mathbf{P}(\mathcal{A}(g) \in S(E;g), \mathcal{A}(g') \in S(E;g')) &= \mathbf{E}[\mathbf{P}(\mathcal{A}(g) \in S(E;g), \mathcal{A}(g') \in S(E;g') \mid \tilde{g})] \\
&= \mathbf{E}[\mathbf{P}(\mathcal{A}(g) \in S(E;g) \mid \tilde{g})^2] \\
&\geq \mathbf{E}[\mathbf{P}(\mathcal{A}(g) \in S(E;g) \mid \tilde{g})]^2 = (p^{\mathrm{cor}}_{\mathrm{solve}})^2,
\end{aligned}$$

where the last line follows by Jensen's inequality. □

Remark 4.3. Note that Lemma 4.2 also holds in the case where $\mathcal{A}(g,\omega)$ is randomized, in the sense of Definition 2.16. Namely, write

$$\begin{aligned}
p &= \mathbf{P}(\mathcal{A}(g,\omega) \in S(E;g)), & P &= \mathbf{P}(\mathcal{A}(g,\omega) \in S(E;g), \mathcal{A}(g',\omega) \in S(E;g')), \\
p(\omega) &= \mathbf{P}(\mathcal{A}(g,\omega) \in S(E;g) \mid \omega), & P(\omega) &= \mathbf{P}(\mathcal{A}(g,\omega) \in S(E;g), \mathcal{A}(g',\omega) \in S(E;g') \mid \omega).
\end{aligned}$$

Lemma 4.2 shows that for any $\omega \in \Omega_N$, $P(\omega) \geq p(\omega)^2$. Then, by Jensen's inequality,

$$P = \mathbf{E}[P(\omega)] \geq \mathbf{E}[p(\omega)^2] \geq \mathbf{E}[p(\omega)]^2 = p^2.$$

Thus, in combination with Remark 2.20, the remainder of the proof also applies when $\mathcal{A}$ depends on an independent random seed ω.

Meanwhile, define $p^{\mathrm{cor}}_{\mathrm{unstable}}, p^{\mathrm{cor}}_{\mathrm{cond}}(x)$, and $p^{\mathrm{cor}}_{\mathrm{cond}}$ by

$$p^{\mathrm{cor}}_{\mathrm{unstable}} := 1 - \mathbf{P}(S_{\mathrm{stable}}), \qquad p^{\mathrm{cor}}_{\mathrm{cond}}(x) := 1 - \mathbf{P}(S_{\mathrm{cond}}(x)), \qquad p^{\mathrm{cor}}_{\mathrm{cond}} := \max_{x \in \Sigma_N} p^{\mathrm{cor}}_{\mathrm{cond}}(x). \tag{4.3}$$

By Lemma 4.1, we know that for $x := \mathcal{A}(g)$,

$$\mathbf{P}(S_{\mathrm{solve}}) + \mathbf{P}(S_{\mathrm{stable}}) + \mathbf{P}(S_{\mathrm{cond}}(x)) \leq 2,$$

and rearranging yields

$$(p^{\mathrm{cor}}_{\mathrm{solve}})^2 \leq p^{\mathrm{cor}}_{\mathrm{unstable}} + p^{\mathrm{cor}}_{\mathrm{cond}}. \tag{4.4}$$

Our proof now follows from showing that, for appropriate choices of ε and η depending on D, E, and N, both $p^{\mathrm{cor}}_{\mathrm{unstable}}$ and $p^{\mathrm{cor}}_{\mathrm{cond}}$ are $o(1)$. The former is controlled by Proposition 2.19, so all that remains is to control the latter. To this end, we start by bounding the size of neighborhoods on Σ_N.

Proposition 4.4 (Hypercube Neighborhood Size)**.** *Fix $x \in \Sigma_N$, and let $\eta \leq 1/2$. Then the number of x' within distance $2\sqrt{\eta N}$ of x is bounded by*

$$\left|\left\{x' \in \Sigma_N : \|x - x'\| \leq 2\sqrt{\eta N}\right\}\right| \leq \exp_2(2\eta \log_2(1/\eta) N).$$

Proof: Let k be the number of coordinates which differ between x and x' (i.e., the Hamming distance). We have $\|x - x'\|^2 = 4k$, so $\|x - x'\| \leq 2\sqrt{\eta N}$ if and only if $k \leq N\eta$. Moreover, $k \leq N/2$ for $\eta \leq 1/2$. Thus, by Lemma 1.6,

$$\sum_{k \leq N\eta} \binom{N}{k} \leq \exp_2(2\eta \log_2(1/\eta) N). \qquad \square$$

Thus, within a small neighborhood of any $x \in \Sigma_N$, the number of nearby points is exponential in N, with a more nontrivial dependence on η. The question is then how many of these are solutions to the correlated instance g'. This forms the heart of our conditional landscape obstruction.

Proposition 4.5 (Fundamental Estimate – Correlated Case)**.** *Assume that (g, g') are $(1-\varepsilon)$-correlated standard Normal vectors. Then, for any x such that (g', x) are conditionally independent given g,*

$$\begin{aligned} p^{\mathrm{cor}}_{\mathrm{cond}}(x) &:= \mathbf{P}\begin{pmatrix} \exists\ x' \in S(E; g') \text{ such that} \\ \|x - x'\| \leq 2\sqrt{\eta N} \end{pmatrix} \\ &\leq \exp_2\left(-E - \frac{1}{2}\log_2(\varepsilon) + 2\eta \log_2\left(\frac{1}{\eta}\right) N + O(\log N)\right). \end{aligned} \tag{4.5}$$

Proof: For each x' within distance $2\sqrt{\eta N}$ of x, let

$$I_{x'} := I(x' \in S(E; g')) = I(|\langle g', x' \rangle| \leq 2^{-E}),$$

so that

$$p^{\mathrm{cor}}_{\mathrm{cond}}(x) = \mathbf{E}\left[\sum_{\|x - x'\| \leq 2\sqrt{\eta N}} \mathbf{E}[I_{x'} \mid g]\right] = \mathbf{E}\left[\sum_{\|x - x'\| \leq 2\sqrt{\eta N}} \mathbf{P}(|\langle g', x' \rangle| \leq 2^{-E} \mid g)\right]. \tag{4.6}$$

Note in particular that the range of this sum is independent of the inner probability, as g' and x are conditionally independent.

To bound the inner probability, let $\tilde{g}$ be a Normal vector independent to g, and set $p = 1 - \varepsilon$. Observe that g' can be represented as $pg + \sqrt{1-p^2}\tilde{g}$, so $\langle g', x' \rangle = p\langle g, x' \rangle + \sqrt{1-p^2}\langle \tilde{g}, x' \rangle$. We know $\langle \tilde{g}, x' \rangle \sim \mathcal{N}(0, N)$, so conditional on g, we have $\langle g', x' \rangle \mid g \sim \mathcal{N}(p\langle g, x' \rangle, (1-p^2)N)$. Note that $\langle g', x' \rangle$ is nondegenerate for $(1-p^2)N \geq \varepsilon N > 0$; thus by Lemma 1.5, we get

$$\mathbf{P}(|\langle g', x' \rangle| \leq 2^{-E} \mid g) \leq \exp_2\left(-E - \frac{1}{2}\log_2(\varepsilon) + O(\log N)\right). \tag{4.7}$$

Finally, by Proposition 4.4, the number of terms in the sum (4.6) is bounded by $\exp_2(2\eta \log_2(1/\eta)N)$, so given that (4.7) is independent of g, we deduce (4.5). □

With this obstruction in hand, we can turn to showing strong low degree hardness for polynomial algorithms. We start with hardness for linear energy levels, $E = \Theta(N)$; this corresponds to the statistically optimal regime, as per [60]. Our hardness result in this regime roughly corresponds to that of Gamarnik and Kızıldağ's Theorem 3.2, although their result applies to stable algorithms and does not show a low degree hardness-type statement [39, Thm. 3.2]. A key feature of considering polynomial algorithms is that in Proposition 4.5, we can let ε be exponentially small in E, which in the linear regime allows for it to be exponentially small in N. As we will see, this has rather extreme implications for the failure of polynomial algorithms under the low degree heuristic.

Theorem 4.6. *Let $\delta > 0$, $E := \delta N$, and g, g' be $(1-\varepsilon)$-correlated standard Normal r.v.s. Then, for any polynomial algorithm $\mathcal{A}$ with $\mathbf{E}\|\mathcal{A}(g)\|^2 \leq CN$ and degree $D \leq o(\exp_2(\delta N/2))$, there exist ε, η such that*

$$p^{\mathrm{cor}}_{\mathrm{solve}} = \mathbf{P}(\mathcal{A}(g) \in S(E; g)) = o(1).$$

Proof: Recall from (4.4) that it suffices to show that both $p^{\mathrm{cor}}_{\mathrm{cond}}$ and $p^{\mathrm{cor}}_{\mathrm{unstable}}$ vanish in the limit. Thus, first choose η sufficiently small, so that $2\eta \log_2(1/\eta) < \delta/4$; this results in η being independent of N. Next, choose $\varepsilon := \exp_2(-\delta N/2)$. By (4.3) and Proposition 4.5, these choices give

$$p^{\mathrm{cor}}_{\mathrm{cond}} \leq \exp_2\left(-\delta N - \frac{1}{2}\left(-\frac{\delta N}{2}\right) + \frac{\delta N}{4} + O(\log N)\right) = \exp_2\left(-\frac{\delta N}{2} + O(\log N)\right) = o(1).$$

We conclude by observing that for $D \leq o(\exp_2(\delta N/2))$, Proposition 2.19 gives

$$p^{\mathrm{cor}}_{\mathrm{unstable}} \leq \frac{CD\varepsilon}{2\eta} \asymp \frac{D\varepsilon}{\eta} \asymp D \cdot \exp_2\left(-\frac{\delta N}{2}\right) = o(1). \qquad \square$$

Next, we consider the sublinear energy regime $\omega(\log N) \leq E \leq o(N)$. This bridges the gap from the statistically optimal energy threshold down to the computational threshold. In particular, our method allows us to rule out degree $o\big(N^{O(N)}\big)$ polynomial algorithms even for achieving the same energy threshold as the Karmarkar-Karp algorithm; this is expected however, as neither the original Karmarkar-Karp algorithm nor the simplified LDM algorithm are polynomial.

Theorem 4.7. *Let $\omega(\log N) \leq E \leq o(N)$ and g, g' be $(1-\varepsilon)$-correlated standard Normal r.v.s. Then, for any polynomial algorithm $\mathcal{A}$ with $\mathbf{E}\|\mathcal{A}(g)\|^2 \leq CN$ and degree $D \leq o(\exp_2(E/4))$, there exist ε, η s.t.*

$$p^{\mathrm{cor}}_{\mathrm{solve}} = \mathbf{P}(\mathcal{A}(g) \in S(E; g)) = o(1).$$

Proof: As in Theorem 4.6, it suffices to show both $p^{\mathrm{cor}}_{\mathrm{cond}}$ and $p^{\mathrm{cor}}_{\mathrm{unstable}}$ are $o(1)$. To do this, we choose

$$\varepsilon = \exp_2\left(-\frac{E}{2}\right), \qquad \eta = \frac{E}{16N\log_2(N/E)}. \tag{4.8}$$

Then, simple analysis shows that for $\frac{E}{N} \ll 1$,

$$\frac{E}{4N} > 2\eta \log_2(1/\eta).$$

Thus, by Proposition 4.5, we get

$$\begin{aligned} p_{\text{cond}}^{\text{cor}} &\leq \exp_2\left(-E - \frac{1}{2}\log_2(\varepsilon) + 2\eta \log_2\left(\frac{1}{\eta}\right)N + O(\log N)\right) \\ &\leq \exp_2\left(-E + \frac{E}{4} + \frac{E}{4} + O(\log N)\right) = \exp_2\left(-\frac{E}{2} + O(\log N)\right) = o(1), \end{aligned}$$

using that $E \gg \log N$. By Proposition 2.19, the choice of $D = o(\exp_2(E/4))$ now gives

$$\begin{aligned} p_{\text{unstable}}^{\text{cor}} &\leq \frac{CD\varepsilon}{2\eta} \asymp \frac{D\varepsilon N \log_2(N/E)}{E} \\ &= \frac{D\exp_2(-E/2)N\log_2(N/E)}{E} \leq \frac{D\exp_2(-E/2)N\log_2(N)}{E} \\ &\leq D\exp_2\left(-\frac{E}{2} + \log_2(N) + \log_2\log_2(N) - \log_2(E)\right) \\ &\leq \exp_2\left(-\frac{E}{4} + \log_2(N) + \log_2\log_2(N) - \log_2(E)\right) = o(1), \end{aligned}$$

again following from $E \gg \log N$. Ergo, by (4.4), $(p_{\text{solve}}^{\text{cor}})^2 \leq p_{\text{unstable}}^{\text{cor}} + p_{\text{cond}}^{\text{cor}} = o(1)$. □

Holistically, these results imply that polynomial algorithms require degree exponential in the energy level to achieve solutions of the desired discrepancy. Under the low degree heuristic, this corresponds to requiring double exponential time – this is clearly beaten by brute force search in exponential time. In this case, strong low degree hardness of the NPP serves as evidence of polynomial algorithms being unsuited to these types of brittle random optimization problems.

Remark 4.8 (Extending to Randomized Algorithms)**.** As discussed in Remark 2.20 and Remark 4.3, if $\mathcal{A}(g, \omega)$ is a randomized Σ_N-valued low degree polynomial algorithm satisfying the averaged bound $\mathbf{E}\|\mathcal{A}(g,\omega)\|^2 \leq CN$, then for every ε, one can show Theorem 4.6 and Theorem 4.7 for $\mathcal{A}(-, \omega)$ for any fixed random seed. In particular, the conditional landscape obstruction Proposition 4.5 works without change when conditioning on ω throughout. Averaging these bounds then allows the proof to go through. We note that this extension to randomized algorithms also applies for low coordinate degree hardness.

4.2 Hardness for Low Coordinate Degree Algorithms

Next, let $\mathcal{A}$ have coordinate degree D. We now want g, g' to be $(1-\varepsilon)$-resampled standard Normal random variables, and we define the events

$$\begin{aligned} S_{\text{diff}} &:= \{g \neq g'\}, \\ S_{\text{solve}} &:= \{\mathcal{A}(g) \in S(E; g), \mathcal{A}(g') \in S(E; g')\}, \\ S_{\text{stable}} &:= \left\{\|\mathcal{A}(g) - \mathcal{A}(g')\| \leq 2\sqrt{\eta N}\right\}, \end{aligned} \tag{4.9}$$

$$S_{\text{cond}}(x) := \begin{Bmatrix} \nexists\ x' \in S(E; g') \text{ such that} \\ \|x - x'\| \le 2\sqrt{\eta N} \end{Bmatrix}. \tag{4.9}$$

Note that these are the same events as (4.1), plus the event that g' is nontrivially resampled from g.

Lemma 4.9. *For g, g' being $(1-\varepsilon)$-resampled, $\mathbf{P}(S_{\text{diff}}) = 1 - (1-\varepsilon)^N \le \varepsilon N$.*

Proof: This follows from the calculation

$$\mathbf{P}(g = g') = \prod_{i=1}^{N} \mathbf{P}(g_i = g_{i'}) = (1-\varepsilon)^N. \qquad \square$$

Lemma 4.10. *For $x := \mathcal{A}(g)$, we have $S_{\text{diff}} \cap S_{\text{solve}} \cap S_{\text{stable}} \cap S_{\text{cond}}(x) = \emptyset$.*

Proof: This follows from Lemma 4.1, noting that the proof did not use that $g \neq g'$ almost surely. $\square$

We can interpret this as saying $S_{\text{solve}}, S_{\text{stable}}, S_{\text{cond}}$ are all mutually exclusive, conditional on $g \neq g'$. The previous definition of $p^{\text{cor}}_{\text{solve}}$ in (4.2), which we now term $p^{\text{res}}_{\text{solve}}$, remains valid.

Lemma 4.11. *For g, g' being $(1-\varepsilon)$-resampled, we have*

$$\mathbf{P}(S_{\text{solve}}) = \mathbf{P}(\mathcal{A}(g) \in S(E; g), \mathcal{A}(g') \in S(E; g')) \ge (p^{\text{res}}_{\text{solve}})^2.$$

Proof: Let $\tilde{g}, g^{(0)}, g^{(1)}$ be three i.i.d. copies of g, and let J be a random subset of $[N]$, where each coordinate is included with probability $1-\varepsilon$. Then, g, g' are jointly representable as

$$g = \tilde{g}_J + g^{(0)}_{\bar{J}}, \qquad\qquad g' = \tilde{g}_J + g^{(1)}_{\bar{J}}.$$

Thus g and g' are conditionally independent given $(\tilde{g}, J)$, and the proof concludes as in Lemma 4.2. $\square$

Now, let us slightly redefine $p^{\text{res}}_{\text{unstable}}$ and $p^{\text{res}}_{\text{cond}}(x)$ to be

$$p^{\text{res}}_{\text{unstable}} := 1 - \mathbf{P}(S_{\text{stable}} \mid S_{\text{diff}}), \qquad\qquad p^{\text{res}}_{\text{cond}}(x) := 1 - \mathbf{P}(S_{\text{cond}}(x) \mid S_{\text{diff}}). \tag{4.10}$$

This is necessary as when $g = g'$, S_{stable} always holds and $S_{\text{cond}}(x)$ always fails. Note however that if we knew that $\mathbf{P}(S_{\text{diff}}) = 1$ (which is always the case for g, g' being $(1-\varepsilon)$-correlated), then these definitions agree with what we had in (4.4). Again, we can define $p^{\text{res}}_{\text{cond}}$ via (4.3).

Now, by Lemma 4.10, we know that for $x = \mathcal{A}(g)$, $\mathbf{P}(S_{\text{solve}}, S_{\text{stable}}, S_{\text{cond}}(x) \mid S_{\text{diff}}) = 0$, so

$$\mathbf{P}(S_{\text{solve}} | S_{\text{diff}}) + \mathbf{P}(S_{\text{stable}} | S_{\text{diff}}) + \mathbf{P}(S_{\text{cond}}(x) | S_{\text{diff}}) \le 2.$$

Thus, rearranging and multiplying by $\mathbf{P}(S_{\text{diff}})$ gives

$$\mathbf{P}(S_{\text{solve}}, S_{\text{diff}}) \le \mathbf{P}(S_{\text{diff}}) \cdot (p^{\text{res}}_{\text{unstable}} + p^{\text{res}}_{\text{cond}}) \le p^{\text{res}}_{\text{unstable}} + p^{\text{res}}_{\text{cond}}.$$

Adding $\mathbf{P}(S_{\text{solve}}, \neg S_{\text{diff}}) \le 1 - P(S_{\text{diff}})$ (so as to apply Lemma 4.11) now lets us conclude

$$(p^{\text{res}}_{\text{solve}})^2 \le \mathbf{P}(S_{\text{solve}}) \le p^{\text{res}}_{\text{unstable}} + p^{\text{res}}_{\text{cond}} + (1 - \mathbf{P}(S_{\text{diff}})). \tag{4.11}$$

As before, our proof follows from showing that for appropriate choices of ε and η (depending on D, E, and N), $p^{\text{res}}_{\text{unstable}}$ and $p^{\text{res}}_{\text{cond}}$ are $o(1)$. However, we are also required us to choose $\varepsilon \gg \frac{1}{N}$, so as to ensure that $g \neq g'$, as otherwise (a) $p^{\text{res}}_{\text{cond}}$ would be too large and (b) the $1 - \mathbf{P}(S_{\text{diff}})$ term would fail to vanish. This restriction on ε stops us from showing hardness for algorithms with degree larger than $o(N)$, as we will see shortly.

As before, we can establish a conditional landscape obstruction for resampled instances via a first moment computation. Here, we need to condition on the resampled instance being different, as otherwise the probability in question can be made to be 1 if x was chosen to solve g.

Proposition 4.12 (Fundamental Estimate – Resampled Case)**.** *Assume that (g, g') are $(1-\varepsilon)$-resampled standard Normal vectors. Then, for any x such that (g', x) are conditionally independent given g,*

$$\begin{aligned} p^{\text{res}}_{\text{cond}}(x) &= \mathbf{P}\left(\begin{matrix} \exists\ x' \in S(E; g') \text{ such that} \\ \|x - x'\| \le 2\sqrt{\eta N} \end{matrix} \middle| \, g \neq g' \right) \\ &\le \exp_2\left(-E + 2\eta \log_2\left(\frac{1}{\eta}\right) N + O(1) \right). \end{aligned} \tag{4.12}$$

Proof: We set up the proof as in Proposition 4.5. For each x' within distance $2\sqrt{\eta N}$ of x, let

$$I_{x'} := I\{x' \in S(E; g')\} = I\{|\langle g', x' \rangle| \le 2^{-E}\},$$

so that

$$\begin{aligned} p^{\text{res}}_{\text{cond}}(x) &= \mathbf{E}\left[\sum_{\|x - x'\| \le 2\sqrt{\eta N}} \mathbf{E}[I_{x'} \mid g, g \neq g'] \,\middle|\, g \neq g' \right] \\ &= \mathbf{E}\left[\sum_{\|x - x'\| \le 2\sqrt{\eta N}} \mathbf{P}(|\langle g', x' \rangle| \le 2^{-E} \mid g, g \neq g') \,\middle|\, g \neq g' \right]. \end{aligned} \tag{4.13}$$

Again, to bound the inner probability, let $\tilde{g}$ be a Normal vector independent of g. Let $J \subseteq [N]$ be a random subset where each $i \in J$ independently with probability $1 - \varepsilon$, so g' can be represented as $g' = g_J + \tilde{g}_{\overline{J}} \tilde{g}_{\overline{J}}$. For a fixed x' and conditional on (g, J), we know that $\langle \tilde{g}_{\overline{J}}, x' \rangle$ is $\mathcal{N}(0, N - |J|)$ and $\langle g_J, x' \rangle$ is deterministic. That is,

$$\langle g', x' \rangle \mid (g, J) \sim \mathcal{N}(\langle g_J, x' \rangle, N - |J|).$$

Conditioning on $g \neq g'$ is equivalent to conditioning on $|J| < N$, so $N - |J| \ge 1$. Thus, applying Lemma 1.5 and integrating over all valid choices of J gives

$$\mathbf{P}(|\langle g', x' \rangle| \le 2^{-E} \mid g, g \neq g') \le \exp_2(-E + O(1)). \tag{4.14}$$

By Proposition 4.4, the number of terms in the sum (4.13) is bounded by $\exp_2(2\eta \log_2(1/\eta) N)$, so summing (4.14) over these terms yields (4.12). □

Note that in contrast to Proposition 4.5, this bound does not involve ε explicitly, but the condition $g \neq g'$ requires $\varepsilon = \omega(1/N)$ to hold almost surely, by Lemma 4.9.

With this, we can show strong low degree hardness for low coordinate degree algorithms at linear energy levels $E = \Theta(N)$. As before, this corresponds to hardness at the statistically optimal energy regime, but now applies to an extremely broad category of algorithms.

Theorem 4.13. *Let $\delta > 0$, $E := \delta N$, and g, g' be $(1-\varepsilon)$-resampled standard Normal r.v.s. Then, for any algorithm $\mathcal{A}$ with $\mathbf{E}\|\mathcal{A}(g)\|^2 \leq CN$ and coordinate degree $D \leq o(N)$, there exist ε, η such that*

$$p^{\text{res}}_{\text{solve}} = \mathbf{P}(\mathcal{A}(g) \in S(E; g)) = o(1).$$

Proof: Recall from (4.11) that it suffices to show that both $p^{\text{res}}_{\text{cond}}$ and $p^{\text{res}}_{\text{unstable}}$ vanish in the limit, while $\mathbf{P}(S_{\text{diff}}) \to 1$. By Lemma 4.9, the latter condition is satisfied for $\varepsilon = \omega(1/N)$. Thus, pick

$$\varepsilon = \frac{\log_2(N/D)}{N}. \tag{4.15}$$

Note that this satisfies $N\varepsilon = \log_2(N/D) \gg 1$ for $D = o(N)$. Next, choose η such that $2\eta \log_2(1/\eta) < \delta/4$; again, this results in η being independent of N. By Proposition 4.12, we get

$$p^{\text{res}}_{\text{cond}} \leq \exp_2\left(-\delta N + \frac{\delta N}{4} + O(1)\right) = o(1).$$

Moreover, for $D \leq o(N)$, Proposition 2.19 now gives

$$p^{\text{res}}_{\text{unstable}} \leq \frac{CD\varepsilon}{2\eta} \asymp D \cdot \frac{\log_2(N/D)}{N} = o(1).$$

By (4.11), we conclude that $\left(p^{\text{res}}_{\text{solve}}\right)^2 \leq p^{\text{res}}_{\text{unstable}} + p^{\text{res}}_{\text{cond}} + (1 - P(S_{\text{diff}})) = o(1)$. □

Finally, combining the ideas behind Theorem 4.7 and our conditional landscape obstruction for $(1-\varepsilon)$-resampled Normal random variables, we can show hardness for algorithms with low coordinate degree at sublinear energy levels, ranging from $\log^2 N \ll E \ll N$. Here we have to increase our lower bound to $\log^2 N$ as opposed to $\log N$ from Theorem 4.7, to address the requirement that $\varepsilon = \omega(1/N)$, but this still enables us to "close" the statistical-to-computational gap by proving hardness in this range. Note also that our method also allows us to derive a clear tradeoff between solution energy and algorithm degree.

Theorem 4.14. *Let $\omega(\log^2 N) \leq E \leq o(N)$, and g, g' be $(1-\varepsilon)$-resampled standard Normal r.v.s. Then, for any algorithm $\mathcal{A}$ with $\mathbf{E}\|\mathcal{A}(g)\|^2 \leq CN$ and coordinate degree $D \leq o(E/\log^2 N)$, there exist ε, η s.t.*

$$p^{\text{res}}_{\text{solve}} = \mathbf{P}(\mathcal{A}(g) \in S(E; g)) = o(1).$$

Proof: The strategy is the same as in Theorem 4.13. Start by choosing ε as in (4.15), so that $\varepsilon = \omega(1/N)$ and $\mathbf{P}(S_{\text{diff}}) \to 1$. To account for $E \leq o(N)$, choose η as in (4.8); this ensures $2\eta \log_2(1/\eta) < E/4N$ for $E \ll N$. By Proposition 4.12, this then guarantees that

$$p_{\text{cond}}^{\text{res}} \leq \exp_2\left(-E + 2\eta \log_2\left(\frac{1}{\eta}\right)N + O(1)\right) \leq \exp_2\left(-\frac{3E}{4} + O(1)\right) = o(1).$$

The low coordinate degree requirement $D \leq o(E/\log^2 N)$ plus Proposition 2.19 now gives

$$\begin{aligned} p_{\text{unstable}}^{\text{res}} &\leq \frac{CD\varepsilon}{2\eta} \asymp \frac{D\varepsilon N \log_2(N/E)}{E} \\ &= \frac{D\log_2(N/D)\log_2(N/E)}{E} \leq \frac{D(\log_2 N)^2}{E} = o(1). \end{aligned}$$

We are now done by (4.11). □

Remark 4.15 (Tightness of Coordinate Degree Bounds)**.** For any $E \leq \Theta(N)$, there is an easy method to achieve a discrepancy of 2^{-E} in $e^{O(E)}$ time.

(a) Pick a subset $J \subseteq [N]$ of E coordinates.

(b) Run LDM on the restricted NPP $g_{\overline{J}}$ to find a partition $x_{\overline{J}}$ with $\langle g_{\overline{J}}, x_{\overline{J}}\rangle \leq 1$.

(c) If we fix the values of $x_{\overline{J}}$, the NPP given by g turns into finding x_J minimizing $|\langle g, x\rangle| = |\langle g_J, x_J\rangle + \langle g_{\overline{J}}, x_{\overline{J}}\rangle|$. Note here that $\langle g, x\rangle | (g_{\overline{J}}, x_{\overline{J}}) \sim \mathcal{N}(\mu, E)$, for $\mu = \langle g_{\overline{J}}, x_{\overline{J}}\rangle$.

(d) Given the statistical energy threshold is $\Theta(N)$, we know g has a solution with energy E with high probability. Moreover, by the proof of Lemma 1.5, the probability of any x_J solving g_J is independent of $O(1)$ constant shifts to the instance, so we can conclude that this restricted NPP also has an energy E solution.

(e) Thus, at this stage, we can brute force search over the remaining J coordinates, which gives a solution with energy E with high probability, in $e^{O(E)}$ time.

In particular, this suggests that our results Theorem 4.13 and Theorem 4.14 are optimal under the low degree heuristic. Namely, low degree hardness of finding solutions with energy E holds up to degree $\tilde{o}(E)$, which implies finding such solutions requires at least time $e^{\tilde{\Omega}(E)}$. This restricted brute force strategy shows that it is indeed possible to find these solutions in time $e^{\tilde{O}(E)}$, implying that our method gives the optimal energy-runtime tradeoff.

It is worthwhile asking whether the low degree heuristic is truly appropriate in our brittle setting. For instance, in most cases where it has been applied to a random optimization problem (e.g., by Huang and Sellke [54]), the objective under consideration has been fairly stable. However, the NPP has a very one-dimensional landscape, lacking the "depth" which foils the low degree heuristic for, e.g., broadcasting on trees [50]. Moreover, the sharp energy-runtime tradeoff in Remark 4.15 is suggestive of the strength of this heuristic in this case.

As a final remark, consider that an algorithm with coordinate degree $\Omega(N)$ (equivalently, $\Theta(N)$) is one which considers nonlinear interactions between some constant fraction of all the coordinates as N gets large. Intuitively, such an algorithm is forced to look at how a large number of instance elements balance against each other, giving further evidence to the claim that any sufficiently local algorithm for the NPP will be no better than random search. The good algorithms must be global, which reflects recent developments in heuristics for computing solutions to the NPP [64], [26], [89].

4.3 Extensions to Real-Valued Algorithms

In Section 4, we have established strong low degree hardness for both low degree polynomial and low coordinate degree algorithms. However, our stability analysis assumed that the algorithms in question were Σ_N-valued. In this section, we show that this assumption is not in fact as restrictive as it might appear.

Throughout, let $\mathcal{A}$ denote a $\mathbf{R}^N$-valued algorithm. We want to show that:

(a) no low degree $\mathcal{A}$ can reliably output points *close* – within constant distance – to a solution;

(b) no Σ_N-valued algorithm $\tilde{\mathcal{A}}$ coming from randomly rounding the output of $\mathcal{A}$, which changes an $\omega(1)$ number of coordinates, can find a solution with nonvanishing probability.

In principle, the first possibility fails via the same analysis as in Section 4, while the second fails because the landscape of solutions to any given NPP instance is sparse.

Why are these the only two possibilities? For $\mathcal{A}$ to provide a way to actually solve the NPP, we must be able to turn its outputs on $\mathbf{R}^N$ into points on Σ_N. If $\mathcal{A}$ could output points within an constant distance (independent of the instance) of a solution, then we could convert $\mathcal{A}$ into a Σ_N-valued algorithm by manually computing the energy of all points close to its output and returning the energy-maximizing point.

However, another common way to convert a $\mathbf{R}^N$-valued algorithm into a Σ_N-valued one is by rounding the outputs, as in [5], [54]. Doing this directly can lead to difficulties in performing the stability analysis. In our case, if we know that no $\mathcal{A}$ can reliably output points within constant distance of a solution, then any rounding scheme which only flips $O(1)$ many coordinates will assuredly fail. Thus, the only rounding schemes worth considering must flip $\omega(1)$ many coordinates.

We first show that no low degree $\mathcal{A}$ can find points within constant distance of a solution, effectively by reproducing the argument of Section 4.2. We then turn to describing a landscape obstruction to randomized rounding, relying on what we term the *solution repulsion property*: solutions to any NPP instance are far away from each other, with this distance tradeoff controlled by the energy level of the solution set in consideration. This can then be leveraged to show that any sufficiently randomized rounding scheme will always fail to find solutions at energies higher than the computational threshold.

4.4 Hardness for Close Algorithms

Throughout this section, fix a distance $r = O(1)$. Consider the event that the $\mathbf{R}^N$-valued algorithm $\mathcal{A}$ outputs a point close to a solution for an instance g:

$$S_{\text{close}}(r) = \left\{ \begin{matrix} \exists\ \hat{x} \in S(E; g) \text{ s.t.} \\ \mathcal{A}(g) \in B_r(\hat{x}) \end{matrix} \right\} = \{B_r(\mathcal{A}(g)) \cap S(E; g) \neq \emptyset\}.$$

Note that since r is of constant order, we can convert $\mathcal{A}$ into a Σ_N-valued algorithm by first rounding the $\mathcal{A}(g)$ into the solid binary hypercube and then picking the best corner of Σ_N within constant distance of this output.

Let clip: $\mathbf{R}^N \to [-1, 1]^N$ be the function which rounds $x \in \mathbf{R}^N$ into the cube $[-1, 1]^N$:

$$\mathsf{clip}(x)_i = \begin{cases} -1 & x_i \le -1, \\ x_i & -1 < x_i < 1, \\ 1 & x_i \ge 1. \end{cases}$$

Note that clip is 1-Lipschitz with respect to the Euclidean norm.

Definition 4.16. Let $r > 0$ and $\mathcal{A}$ be an algorithm. Define the $[-1,1]^N$-valued algorithm $\hat{\mathcal{A}}_r$ by

$$\hat{\mathcal{A}}_r(g) := \operatorname*{argmin}_{x' \in B_r(\mathsf{clip}(\mathcal{A}(g))) \cap \Sigma_N} |\langle g, x' \rangle|. \tag{4.16}$$

If $B_r(\mathsf{clip}(\mathcal{A}(g))) \cap \Sigma_N = \emptyset$, then set $\hat{\mathcal{A}}_r(g) := \mathsf{clip}(\mathcal{A}(g))$, which is necessarily not in Σ_N.

Observe that $S_{\mathrm{close}}(r)$ occurring implies $\hat{\mathcal{A}}_r$ finds a solution for g. In addition, computing $\hat{\mathcal{A}}_r$ in practice requires additionally calculating the energy of $O(1)$-many points on Σ_N. This requires only an additional $O(N)$ operations.

Recall from Section 2.3 that if $\mathcal{A}$ has low polynomial or coordinate degree, then we can derive useful stability bounds for its outputs. Adjusting the bounds, this modification $\hat{\mathcal{A}}_r$ of $\mathcal{A}$ is also stable.

Lemma 4.17. *Suppose that $\mathbf{E}\|\mathcal{A}(g)\|^2 \le CN$ and $\mathcal{A}$ has degree $\le D$ (resp. coordinate degree $\le D$). Let (g, g') be $(1-\varepsilon)$-correlated (resp. $(1-\varepsilon)$-resampled). Then, $\hat{\mathcal{A}}_r$ as defined above satisfies*

$$\mathbf{E}\left\|\hat{\mathcal{A}}_r(g) - \hat{\mathcal{A}}_r(g')\right\|^2 \le 4CD\varepsilon N + 8r^2. \tag{4.17}$$

In particular,

$$\mathbf{P}\left(\left\|\hat{\mathcal{A}}_r(g) - \hat{\mathcal{A}}_r(g')\right\| \ge 2\sqrt{\eta N}\right) \le \frac{CD\varepsilon}{\eta} + \frac{2r^2}{\eta N}. \tag{4.18}$$

Proof: Observe that by the triangle inequality, $\left\|\hat{\mathcal{A}}_r(g) - \hat{\mathcal{A}}_r(g')\right\|$ is bounded by

$$\begin{aligned} &\left\|\hat{\mathcal{A}}_r(g) - \mathsf{clip}(\mathcal{A}(g))\right\| + \|\mathsf{clip}(\mathcal{A}(g)) - \mathsf{clip}(\mathcal{A}(g'))\| + \left\|\mathsf{clip}(\mathcal{A}(g')) - \hat{\mathcal{A}}_r(g')\right\| \\ &\qquad \le 2r + \|\mathcal{A}(g) - \mathcal{A}(g')\|. \end{aligned}$$

This follows as clip is 1-Lipschitz and the corner-picking step in (4.16) only moves $\hat{\mathcal{A}}_r(g)$ from $\mathsf{clip}(\mathcal{A}(r))$ by at most r. By Jensen's inequality, squaring this gives

$$\left\|\hat{\mathcal{A}}_r(g) - \hat{\mathcal{A}}_r(g')\right\|^2 \le 2\left(4r^2 + \|\mathcal{A}(g) - \mathcal{A}(g')\|^2\right).$$

Combining this with Proposition 2.19 gives (4.17), and (4.18) follows from Markov's inequality. □

Of course, our construction of $\hat{\mathcal{A}}_r$ is certainly never polynomial and does not preserve coordinate degree in a controllable way. Thus, we cannot directly hope for Theorem 4.6, Theorem 4.7, Theorem 4.13, or Theorem 4.14 to hold. However, because this rounding does not drastically alter the stability analysis, we are still able to show that for any $\mathbf{R}^N$-valued low coordinate degree algorithm $\mathcal{A}$ and $r = O(1)$, strong low degree hardness holds for $\hat{\mathcal{A}}_r$. The same argument proves hardness when $\mathcal{A}$ is a low degree polynomial algorithm; this is omitted for brevity.

We recall the setup from Section 4.2. Let g, g' be $(1-\varepsilon)$-resampled standard Normal vectors and define the events

$$\begin{aligned} S_{\text{diff}} &:= \{g \neq g'\}, \\ S_{\text{solve}} &:= \left\{\hat{\mathcal{A}}_r(g) \in S(E;g), \hat{\mathcal{A}}_r(g') \in S(E;g')\right\}, \\ S_{\text{stable}} &:= \left\{\left\|\hat{\mathcal{A}}_r(g) - \hat{\mathcal{A}}_r(g')\right\| \leq 2\sqrt{\eta N}\right\}, \\ S_{\text{cond}}(x) &:= \left\{\begin{array}{l} \nexists\ x' \in S(E;g') \text{ such that} \\ \|x - x'\| \leq 2\sqrt{\eta N} \end{array}\right\}. \end{aligned} \tag{4.19}$$

These are the same events as in (4.9), just adapted to $\hat{\mathcal{A}}_r$. In particular, Lemma 4.10 holds unchanged.

Moreover, we define

$$\hat{p}^{\text{cor}}_{\text{solve}} := \mathbf{P}\left(\hat{\mathcal{A}}_r(g) \in S(E;g)\right) \geq \mathbf{P}(S_{\text{close}}(r)), \tag{4.20}$$

$$\hat{p}^{\text{cor}}_{\text{unstable}} := 1 - \mathbf{P}(S_{\text{stable}} \mid S_{\text{diff}}), \qquad \hat{p}^{\text{cor}}_{\text{cond}}(x) := 1 - \mathbf{P}(S_{\text{cond}}(x) \mid S_{\text{diff}}),$$

along with $\hat{p}^{\text{cor}}_{\text{cond}} := \max_{x \in \Sigma_N} \hat{p}^{\text{cor}}_{\text{cond}}(x)$, echoing (4.10). Observe that since $\hat{p}^{\text{cor}}_{\text{cond}}$ makes no reference to any algorithm, the bound in Proposition 4.12 holds without change. Moreover, Lemma 4.17 lets us control $\hat{p}^{\text{cor}}_{\text{unstable}}$. The final piece needed is an appropriate analog of Lemma 4.11.

Lemma 4.18. *For g, g' being $(1-\varepsilon)$-resampled, we have*

$$\mathbf{P}(S_{\text{solve}}) = \mathbf{P}\left(\hat{\mathcal{A}}_r(g) \in S(E;g), \hat{\mathcal{A}}_r(g') \in S(E;g')\right) \geq (\hat{p}^{\text{cor}}_{\text{solve}})^2.$$

Proof: Observe that, letting $+$ denote Minkowski sum, we have

$$\left\{\hat{\mathcal{A}}_r(g) \in S(E;g)\right\} = \{\mathsf{clip}(\mathcal{A}(g)) \in S(E;g) + B_r(0)\}.$$

Expanding $S(E;g)$, the proof proceeds as in Lemma 4.11. □

Theorem 4.19. *Let $\omega(\log^2 N) \leq E \leq \Theta(N)$, and let g, g' be $(1-\varepsilon)$-resampled standard Normal r.v.s. Consider any $r = O(1)$ and $\mathbf{R}^N$-valued $\mathcal{A}$ with $\mathbf{E}\|\mathcal{A}(g)\|^2 \leq CN$, and assume in addition that*

(a) if $E = \delta N = \Theta(N)$ for $\delta > 0$, then $\mathcal{A}$ has coordinate degree $D \leq o(N)$;

(b) if $\log^2 N \ll E \ll N$, then $\mathcal{A}$ has coordinate degree $D \leq o(E/\log^2 N)$.

Let $\hat{\mathcal{A}}_r$ be as in Definition 4.16. Then there exist $\varepsilon, \eta > 0$ such that

$$\hat{p}^{\text{cor}}_{\text{solve}} = \mathbf{P}\left(\hat{\mathcal{A}}_r(g) \in S(E;g)\right) = o(1).$$

Proof: First, by Lemma 4.10, the appropriate adjustment of (4.11) holds, namely that

$$(\hat{p}^{\text{cor}}_{\text{solve}})^2 \leq \hat{p}^{\text{cor}}_{\text{unstable}} + \hat{p}^{\text{cor}}_{\text{cond}} + (1 - \mathbf{P}(S_{\text{diff}})). \tag{4.21}$$

To ensure $\mathbf{P}(S_{\text{diff}}) \to 1$, we begin by following (4.15) and choosing $\varepsilon = \log_2(N/D)/N$. Moreover, following the proof of Theorem 4.13 and Theorem 4.14, we know that choosing

$$\eta = \begin{cases} O(1) \text{ s.t. } 2\eta \log_2(1/\eta) < \delta/4 & E = \delta N, \\ \frac{E}{16N \log_2(N/E)} & E \leq o(N), \end{cases}$$

in conjunction with Proposition 4.12, guarantees that

$$\hat{p}^{\text{cor}}_{\text{cond}} \leq \exp_2\left(-\frac{3E}{4} + O(1)\right) = o(1).$$

Finally, note that in the linear case, when $\eta = O(1)$, we trivially have $\frac{r^2}{\eta N} = o(1)$. In the sublinear case, for $\eta = E/(16N \log_2(N/E))$, we instead get

$$\eta N = \frac{E}{16 \log_2(N/E)} \geq \frac{E}{16 \log_2 N} = \omega(1),$$

since $E \gg \log^2 N$. Thus, applying the properly modified Lemma 4.17 with these choices of ε and η, we see that $\hat{p}^{\text{cor}}_{\text{unstable}} = o(1)$. By (4.21), we conclude that $\hat{p}^{\text{cor}}_{\text{solve}} = o(1)$. □

Note that as $\hat{p}^{\text{cor}}_{\text{solve}}$ upper bounds $\mathbf{P}(S_{\text{close}}(r))$, this argument shows algorithmic hardness for low degree $\mathbf{R}^N$-valued algorithms aiming to output points within constant distance of a solution.

5 Truly Random Rounding

While deterministic algorithms fail to get close to NPP solutions, perhaps a randomized rounding scheme could work instead. As discussed above, the failure of algorithms finding outputs within a constant distance of a solution motivates considering rounding schemes which are "truly random," in that they change a superconstant number of coordinates. However, this approach is blunted by the same brittleness of the NPP landscape that established the conditional obstruction of Proposition 4.5 and Proposition 4.12. In particular, Theorem 5.4 shows that if one has a subcube of Σ_N with dimension growing slowly with N, then at most only one of those points will be a solution.

For this section, again let $\mathcal{A}$ be a deterministic $\mathbf{R}^N$-valued algorithm. Moreover, assume we are searching for solutions with energy between $\log^2 N \ll E \leq N$; note that for lower values, the Karmarkar-Karp algorithm can already achieve discrepancies of $N^{-\Theta(\log N)}$ energy in polynomial time.

To start, for any $x \in \mathbf{R}^N$, we write x^* for the coordinate-wise signs of x, i.e.,

$$x_i^* := \begin{cases} +1 & x_i > 0, \\ -1 & x_i \leq 0. \end{cases}$$

We can then define the deterministically rounded algorithm $\mathcal{A}^*(g) := \mathcal{A}(g)^*$.

Remark 5.1. Observe that if $\mathcal{A}$ was a low coordinate degree algorithm, then $\mathcal{A}^*$ has the same coordinate degree, so strong low degree hardness as proved in Section 4.2 still applies. On the other hand, if $\mathcal{A}$ was a low polynomial degree algorithm, then $\mathcal{A}^*$ will not be polynomial, but as coordinate

degree bounds polynomial degree, we can recover strong low degree hardness, albeit with worse bounds on D.

In contrast to deterministically rounding of the outputs of $\mathcal{A}$ by taking signs, we can consider passing the output of $\mathcal{A}$ through a randomized rounding scheme. Let $\mathsf{round}(x, \omega)$: $\mathbf{R}^N \times \Omega \to \Sigma_N$ denote any randomized rounding function, with randomness ω independent of the input. We will often suppress the ω in the notation, and treat $\mathsf{round}(x)$ as a Σ_N-valued random variable. We can describe such a randomized rounding function in the following way. Define $p_1(x), ..., p_N(x)$ by

$$p_i(x) := \max\left(\mathbf{P}(\mathsf{round}(x)_i \neq x_i^*), \frac{1}{2}\right). \tag{5.1}$$

We need to guarantee that each $p_i(x) \leq 1/2$ for the following alternative description of $\mathsf{round}(x)$.

Lemma 5.2. *Fix $x \in \mathbf{R}^N$. Draw N coin flips $I_{x,i} \sim \mathrm{Bern}(2p_i(x))$ as well as N signs $S_i \sim \mathrm{Unif}\{\pm 1\}$, all mutually independent; define the random variable $\tilde{x} \in \Sigma_N$ by*

$$\tilde{x}_i := S_i I_{x,i} + \big(1 - I_{x,i}\big) x_i^*.$$

Then $\tilde{x} \sim \mathsf{round}(x)$.

Proof: Conditioning on $I_{x,i}$, we can check that

$$\mathbf{P}(\tilde{x}_i \neq x_i) = 2p_i(x) \cdot \mathbf{P}\big(\tilde{x}_i = x_i \mid I_{x,i} = 1\big) + (1 - 2p_i(x)) \cdot \mathbf{P}\big(\tilde{x}_i \neq x_i \mid I_{x,i} = 0\big) = p_i(x).$$

Thus, $\mathbf{P}(\tilde{x}_i = x_i^*) = \mathbf{P}(\mathsf{round}(x)_i = x_i^*)$. □

By Lemma 5.2, we can redefine $\mathsf{round}(x)$ to be $\tilde{x}$ as constructed above without loss of generality.

It thus makes sense to define $\tilde{\mathcal{A}}(g) := \mathsf{round}(\mathcal{A}(g))$, which is now (a) Σ_N-valued and (b) randomized only in the transition from $\mathbf{R}^N$ to Σ_N (i.e., the rounding does not depend directly on g, but only on the output $x = \mathcal{A}(g)$). We should expect that if $\tilde{\mathcal{A}} = \mathcal{A}^*$ (e.g., if $\mathcal{A}$ outputs values far outside the cube $[-1, 1]^N$) with high probability, then low degree hardness will still apply, since $\mathcal{A}^*$ is deterministic. However, in general, any $\tilde{\mathcal{A}}$ which differs from $\mathcal{A}^*$ will fail to solve g with high probability. This is independent of any assumptions on $\mathcal{A}$: any rounding scheme will introduce so much randomness that $\tilde{x}$ will effectively be a random point, which has a vanishing probability of being a solution because of how sparse and disconnected the NPP landscape is.

To see this, we first show that any randomized rounding scheme as in Lemma 5.2 which differs almost surely from simply picking the signs will resample a diverging number of coordinates.

Lemma 5.3. *Fix $x \in \mathbf{R}^N$, and let $p_1(x), ..., p_N(x)$ be defined as in (5.1). Then $\tilde{x} \neq x^*$ with high probability if and only if $\sum_i p_i(x) = \omega(1)$. Moreover, assuming that $\sum_i p_i(x) = \omega(1)$, the number of coordinates in which $\tilde{x}$ is resampled diverges almost surely.*

Proof: Recall that for $x \in [0, 1/2]$, $\log_2(1 - x) = \Theta(x)$. Thus, as each coordinate of x is rounded independently, we can compute

$$\mathbf{P}(\tilde{x} = x^*) = \prod_i (1 - p_i(x)) = \exp_2\left(\sum_i \log_2(1 - p_i(x))\right) \le \exp_2\left(-\Theta\left(\sum_i p_i(x)\right)\right).$$

Thus, $\mathbf{P}(\tilde{x} = x^*) = o(1)$ if and only if $\sum_i p_i(x) = \omega(1)$.

Next, following the construction of $\tilde{x}$ in Lemma 5.2, let $E_i = \{I_{x,i} = 1\}$ be the event that $\tilde{x}_i$ is resampled from $\mathrm{Unif}\{\pm 1\}$, independently of x_i^*. The E_i are independent, so by Borel-Cantelli, $\sum_i \mathbf{P}(E_i) = 2\sum_i p_i(x) = \omega(1)$ implies that $\tilde{x}_i$ is resampled infinitely often with probability 1. □

To take advantage of the above construction, we show Theorem 5.4: this is a landscape obstruction for single instances of the NPP which shows that solutions resist clustering at a rate related to their energy level (i.e., higher energy solutions push each other further apart). This will let us conclude that any $\tilde{\mathcal{A}}$ which is not equal to $\mathcal{A}^*$ with high probability fails to find solutions.

Theorem 5.4 (Solutions Repel)**.** *Consider any distances $k = \Omega(1)$ and energy levels $E \gg k \log N$. With high probability, there are no pairs of distinct solutions $x, x' \in S(E; g)$ to an instance g with $\|x - x'\| \le 2\sqrt{k}$ (i.e., within k sign flips of each other):*

$$\mathbf{P}\begin{pmatrix}\exists\ (x, x') \in S(E; g) \text{ s.t.} \\ \|x - x'\| \le 2\sqrt{k}.\end{pmatrix} \le \exp_2(-E + O(k \log N)) = o(1). \tag{5.2}$$

Proof: Consider any $x \neq x'$, and let $J \subseteq [N]$ denote the coordinates in which x, x' differ. Then

$$x = x_{\overline{J}} + x_J, \qquad\qquad x' = x_{\overline{J}} - x_J.$$

Assuming both $x, x' \in S(E; g)$, we can expand the inequalities $-2^{-E} \le \langle g, x\rangle, \langle g, x'\rangle \le 2^{-E}$ into

$$-2^{-E} \le \langle g, x_{\overline{J}}\rangle + \langle g, x_J\rangle \le 2^{-E},$$
$$-2^{-E} \le \langle g, x_{\overline{J}}\rangle - \langle g, x_J\rangle \le 2^{-E}.$$

Multiplying the lower equation by -1 and adding the resulting inequalities gives $|\langle g, x_J\rangle| \le 2^{-E}$.

Thus, finding pairs of distinct solutions within distance $2\sqrt{k}$ implies finding a subset $J \subseteq [N]$ of at most k coordinates and $|J|$ signs x_J such that $|\langle g_J, x_J\rangle| \le 2^{-E}$. By [95, Exer. 0.0.5], there are

$$\sum_{1 \le k' \le k} \binom{N}{k'} \le \left(\frac{eN}{k}\right)^k \le (eN)^k = 2^{O(k \log N)}$$

choices of such subsets, and at most 2^k choices of signs. Now, $\langle g_J, x_J\rangle \sim \mathcal{N}(0, |J|)$, and as $|J| \ge 1$, Lemma 1.5 and the following remark implies $\mathbf{P}(|\langle g_J, x_J\rangle| \le 2^{-E}) \le \exp_2(-E + O(1))$. Union bounding this over the $2^{O(k \log N)}$ possibilities gives (5.2). □

Here, our technique of converting pairs of solutions into subvectors of g which must have small sum enables us to reduce the size of the set we union bound over from $2^{O(N)}$ to $2^{O(k \log N)}$. Moreover, observe that this proof can be adapted to show that for a fixed $x \in S(E; g)$, there are no other solutions within k sign flips with high probability.

Finally, we exhibit strong hardness for truly randomized algorithms. Roughly, this holds because if enough coordinates are resampled, the resulting point is random within a subcube of dimension growing slowly with Σ_N, which overwhelms the brittleness in Theorem 5.4.

Theorem 5.5. *Let $x = \mathcal{A}(g)$, and define $x^*, \tilde{x}$, etc., as previously. Moreover, assume that for any x in the possible outputs of $\mathcal{A}$, we have $\sum_i p_i(x) = \omega(1)$. Then, for any $E \geq \omega(\log^2 N)$, we have*

$$\mathbf{P}\big(\tilde{\mathcal{A}}(g) \in S(E; g)\big) = \mathbf{P}(\tilde{x} \in S(E; g)) \leq o(1).$$

Proof: Following the characterization of $\tilde{x}$ in Lemma 5.2, let $K := \max\big(\log_2 N, \sum_i I_{x,i}\big)$. By the assumptions on $\sum_i p_i(x)$ and Lemma 5.3, we know K, which is at least the number of coordinates which are resampled, is bounded as $1 \ll K \leq \log_2 N$, for any possible $x = \mathcal{A}(g)$. Now, let $J \subseteq [N]$ denote the set of the first K coordinates to be resampled, so that $K = |J|$, and consider

$$\mathbf{P}(\tilde{x} \in S(E; g) \mid \tilde{x}_{\overline{J}}),$$

where we fix the coordinates outside of J and let $\tilde{x}$ be uniformly sampled from a K-dimensional subcube of Σ_N. All such $\tilde{x}$ are within distance $2\sqrt{K}$ of each other, so by Theorem 5.4, the probability that there is more than one such $\tilde{x} \in S(E; g)$ is bounded by

$$\exp_2(-E + O(K \log N)) \leq \exp_2(-E + O(\log^2 N)) = o(1),$$

by assumption on E. Thus, the probability that any of the $\tilde{x}$ is in $S(E; g)$ is bounded by 2^{-K}, whence

$$\mathbf{P}(\tilde{x} \in S(E; g)) = \mathbf{E}[\mathbf{P}(\tilde{x} \in S(E; g) \mid \tilde{x}_{\overline{J}})] \leq 2^{-K} \leq o(1). \qquad \square$$

While this rules out many possible randomized rounding schemes, there is still the potential for rounding in a way that depends on both g and x to find solutions with nonvanishing probability. More generally, recent work by Li and Schramm has pointed out that the presence of an OGP or a conditional landscape obstruction is itself evidence of the brittleness of a random optimization problem [77]. Thus, stable algorithms (e.g., Lipschitz, smooth, etc.) are intrinsically ill-suited for such tasks. In light of this, low (coordinate) degree algorithms, which can be stable but are not required to be continuous or smooth, provide better intrinsic models. Given that, new approaches on algorithms for the NPP could focus on non-stable algorithms, such as linear or semidefinite programming. We invite these as interesting directions for potential future work.

References

[1] D. Achlioptas and A. Coja-Oghlan, "Algorithmic Barriers from Phase Transitions," in *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, Oct. 2008, pp. 793–802. doi: 10.1109/FOCS.2008.11.

[2] M. F. Argüello, T. A. Feo, and O. Goldschmidt, "Randomized Methods for the Number Partitioning Problem," *Computers & Operations Research*, vol. 23, no. 2, pp. 103–111, Feb. 1996, doi: 10.1016/0305-0548(95)E0020-L.

[3] G. B. Arous, R. Gheissari, and A. Jagannath, "Algorithmic Thresholds for Tensor PCA," *The Annals of Probability*, vol. 48, no. 4, pp. 2052–2087, Jul. 2020, doi: 10.1214/19-AOP1415.

[4] B. Alidaee, F. Glover, G. A. Kochenberger, and C. Rego, "A New Modeling and Solution Approach for the Number Partitioning Problem," *Journal of Applied Mathematics and Decision Sciences*, vol. 2005, no. 2, pp. 113–121, Jan. 2005, doi: 10.1155/JAMDS.2005.113.

[5] A. E. Alaoui, A. Montanari, and M. Sellke, "Sampling from the Sherrington-Kirkpatrick Gibbs Measure via Algorithmic Stochastic Localization." Accessed: Mar. 31, 2025. [Online]. Available: http://arxiv.org/abs/2203.05093

[6] B. Aubin, W. Perkins, and L. Zdeborová, "Storage Capacity in Symmetric Binary Perceptrons," *Journal of Physics A: Mathematical and Theoretical*, vol. 52, no. 29, p. 294003, Jul. 2019, doi: 10.1088/1751-8121/ab227a.

[7] D. Achlioptas and F. Ricci-Tersenghi, "On the Solution-Space Geometry of Random Constraint Satisfaction Problems." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/cs/0611052

[8] L. Asproni, D. Caputo, B. Silva, G. Fazzi, and M. Magagnini, "Accuracy and Minor Embedding in Subqubo Decomposition with Fully Connected Large Problems: A Case Study about the Number Partitioning Problem," *Quantum Machine Intelligence*, vol. 2, no. 1, p. 4, Jun. 2020, doi: 10.1007/s42484-020-00014-w.

[9] N. Bansal, "Constructive Algorithms for Discrepancy Minimization." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1002.2259

[10] B. Barak, S. B. Hopkins, J. Kelner, P. K. Kothari, A. Moitra, and A. Potechin, "A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1604.03084

[11] M. Brennan and G. Bresler, "Optimal Average-Case Reductions to Sparse PCA: From Weak Assumptions to Strong Hardness." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1902.07380

[12] M. Brennan, G. Bresler, and W. Huleihel, "Reducibility and Computational Lower Bounds for Problems with Planted Sparse Structure." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1806.07508

[13] C. Borgs, J. Chayes, and B. Pittel, "Phase Transition and Finite-size Scaling for the Integer Partitioning Problem," *Random Structures & Algorithms*, vol. 19, no. 3–4, pp. 247–288, Oct. 2001, doi: 10.1002/rsa.10004.

[14] H. Bauke, S. Franz, and S. Mertens, "Number Partitioning as a Random Energy Model," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2004, no. 4, p. P4003, Apr. 2004, doi: 10.1088/1742-5468/2004/04/P04003.

[15] H. Bauke and S. Mertens, "Universality in the Level Statistics of Disordered Systems," *Physical Review E*, vol. 70, no. 2, p. 25102, Aug. 2004, doi: 10.1103/PhysRevE.70.025102.

[16] S. Boettcher and S. Mertens, "Analysis of the Karmarkar-Karp Differencing Algorithm," *The European Physical Journal B*, vol. 65, no. 1, pp. 131–140, Sep. 2008, doi: 10.1140/epjb/e2008-00320-9.

[17] C. Borgs, J. Chayes, S. Mertens, and C. Nair, "Proof of the Local REM Conjecture for Number Partitioning. I: Constant Energy Scales," *Random Structures & Algorithms*, vol. 34, no. 2, pp. 217–240, 2009, doi: 10.1002/rsa.20255.

[18] C. Borgs, J. Chayes, S. Mertens, and C. Nair, "Proof of the Local REM Conjecture for Number Partitioning. II. Growing Energy Scales," *Random Structures & Algorithms*, vol. 34, no. 2, pp. 241–284, 2009, doi: 10.1002/rsa.20256.

[19] A. S. Bandeira, A. Perry, and A. S. Wein, "Notes on Computational-to-Statistical Gaps: Predictions Using Statistical Physics." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1803.11132

[20] Q. Berthet and P. Rigollet, "Computational Lower Bounds for Sparse PCA." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1304.0828

[21] M. Brennan, G. Bresler, S. B. Hopkins, J. Li, and T. Schramm, "Statistical Query Algorithms and Low-Degree Tests Are Almost Equivalent." Accessed: Mar. 31, 2025. [Online]. Available: http://arxiv.org/abs/2009.06107

[22] A. Coja-Oghlan and C. Efthymiou, "On Independent Sets in Random Graphs," *Random Structures & Algorithms*, vol. 47, no. 3, pp. 436–486, Oct. 2015, doi: 10.1002/rsa.20550.

[23] E. G. Coffman Jr., M. R. Garey, and D. S. Johnson, "An Application of Bin-Packing to Multiprocessor Scheduling," *SIAM Journal on Computing*, vol. 7, no. 1, pp. 1–17, Feb. 1978, doi: 10.1137/0207001.

[24] 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, May 2019, doi: 10.1214/18-AOP1291.

[25] E. G. Coffman and G. S. Lueker, *Probabilistic Analysis of Packing and Partitioning Algorithms*. in Wiley-Interscience Series in Discrete Mathematics and Optimization. New York: J. Wiley & sons, 1991.

[26] D. Corus, P. S. Oliveto, and D. Yazdani, "Artificial Immune Systems Can Find Arbitrarily Good Approximations for the NP-hard Number Partitioning Problem," *Artificial Intelligence*, vol. 274, pp. 180–196, Sep. 2019, doi: 10.1016/j.artint.2019.03.001.

[27] K. Chandrasekaran and S. Vempala, "Integer Feasibility of Random Polytopes." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1111.4649

[28] B. Derrida, "Random-Energy Model: Limit of a Family of Disordered Models," *Physical Review Letters*, vol. 45, no. 2, pp. 79–82, Jul. 1980, doi: 10.1103/PhysRevLett.45.79.

[29] B. Derrida, "Random-Energy Model: An Exactly Solvable Model of Disordered Systems," *Physical Review B*, vol. 24, no. 5, pp. 2613–2626, Sep. 1981, doi: 10.1103/PhysRevB.24.2613.

[30] I. Diakonikolas, D. M. Kane, and A. Stewart, "Statistical Query Lower Bounds for Robust Estimation of High-dimensional Gaussians and Gaussian Mixtures." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1611.03473

[31] Y. Deshpande and A. Montanari, “Improved Sum-of-Squares Lower Bounds for Hidden Clique and Hidden Submatrix Problems.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1502.06590

[32] V. Feldman, E. Grigorescu, L. Reyzin, S. Vempala, and Y. Xiao, “Statistical Algorithms and a Lower Bound for Detecting Planted Clique.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1201.1214

[33] D. Gamarnik, “The Overlap Gap Property: A Geometric Barrier to Optimizing over Random Structures,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 41, p. e2108492118, Oct. 2021, doi: 10.1073/pnas.2108492118.

[34] D. Gamarnik and A. Jagannath, “The Overlap Gap Property and Approximate Message Passing Algorithms for p-Spin Models.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1911.06943

[35] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. in A Series of Books in the Mathematical Sciences. New York: W. H. Freeman, 1979.

[36] D. Gamarnik, A. Jagannath, and S. Sen, “The Overlap Gap Property in Principal Submatrix Recovery,” *Probability Theory and Related Fields*, vol. 181, no. 4, pp. 757–814, Dec. 2021, doi: 10.1007/s00440-021-01089-7.

[37] D. Gamarnik, A. Jagannath, and A. S. Wein, “Low-Degree Hardness of Random Optimization Problems,” in *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, Nov. 2020, pp. 131–140. doi: 10.1109/FOCS46700.2020.00021.

[38] D. Gamarnik, A. Jagannath, and A. S. Wein, “Hardness of Random Optimization Problems for Boolean Circuits, Low-Degree Polynomials, and Langevin Dynamics.” Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/2004.12063

[39] D. Gamarnik and E. C. Kızıldağ, “Algorithmic Obstructions in the Random Number Partitioning Problem.” Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/2103.01369

[40] D. Gamarnik and Q. Li, “Finding a Large Submatrix of a Gaussian Random Matrix.” Accessed: Mar. 29, 2025. [Online]. Available: http://arxiv.org/abs/1602.08529

[41] R. L. Graham, “Bounds on Multiprocessing Timing Anomalies,” *SIAM Journal on Applied Mathematics*, vol. 17, no. 2, pp. 416–429, Mar. 1969, doi: 10.1137/0117039.

[42] D. Gamarnik and M. Sudan, “Limits of Local Algorithms over Sparse Random Graphs,” in *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*, in ITCS '14. New York, NY, USA: Association for Computing Machinery, Jan. 2014, pp. 369–376. doi: 10.1145/2554797.2554831.

[43] I. Gent and T. Walsh, “Phase Transitions and Annealed Theories: Number Partitioning as a Case Study,” *Instituto Cultura*, Jun. 2000.

[44] I. P. Gent and T. Walsh, “Analysis of Heuristics for Number Partitioning,” *Computational Intelligence*, vol. 14, no. 3, pp. 430–451, 1998, doi: 10.1111/0824-7935.00069.

[45] D. Gamarnik and I. Zadik, "Sparse High-Dimensional Linear Regression. Algorithmic Barriers and a Local Search Algorithm." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1711.04952

[46] D. Gamarnik and I. Zadik, "High-Dimensional Regression with Binary Coefficients. Estimating Squared Error and a Phase Transition." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1701.04455

[47] D. Gamarnik and I. Zadik, "The Landscape of the Planted Clique Problem: Dense Subgraphs and the Overlap Gap Property." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1904.07174

[48] C. Harshaw, F. Sävje, D. Spielman, and P. Zhang, "Balancing Covariates in Randomized Experiments with the Gram-Schmidt Walk Design." Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/1911.03071

[49] B. Hayes, "The Easiest Hard Problem," *American Scientist*, vol. 90, no. 2, pp. 113–117, Apr. 2002.

[50] H. Huang and E. Mossel, "Optimal Low Degree Hardness for Broadcasting on Trees." Accessed: Mar. 26, 2025. [Online]. Available: http://arxiv.org/abs/2502.04861

[51] R. Hoberg, H. Ramadas, T. Rothvoss, and X. Yang, "Number Balancing Is as Hard as Minkowski's Theorem and Shortest Vector." Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/1611.08757

[52] 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." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1710.05017

[53] S. Hopkins, "Statistical Inference and the Sum of Squares Method," 2018. [Online]. Available: https://www.samuelbhopkins.com/thesis.pdf

[54] B. Huang and M. Sellke, "Strong Low Degree Hardness for Stable Local Optima in Spin Glasses." Accessed: Jan. 30, 2025. [Online]. Available: http://arxiv.org/abs/2501.06427

[55] S. B. Hopkins, J. Shi, and D. Steurer, "Tensor Principal Component Analysis via Sum-of-Squares Proofs." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1507.03269

[56] M. Jerrum, "Large Cliques Elude the Metropolis Process," *Random Structures & Algorithms*, vol. 3, no. 4, pp. 347–359, Jan. 1992, doi: 10.1002/rsa.3240030402.

[57] D. S. Johnson, C. R. Aragon, L. A. McGeoch, and C. Schevon, "Optimization by Simulated Annealing: An Experimental Evaluation; Part I, Graph Partitioning," *Operations Research*, vol. 37, no. 6, pp. 865–892, 1989, Accessed: Mar. 15, 2025. [Online]. Available: http://www.jstor.org/stable/171470

[58] D. S. Johnson, C. R. Aragon, L. A. McGeoch, and C. Schevon, "Optimization by Simulated Annealing: An Experimental Evaluation; Part II, Graph Coloring and Number Partitioning," *Operations Research*, vol. 39, no. 3, pp. 378–406, 1991, Accessed: Mar. 15, 2025. [Online]. Available: http://www.jstor.org/stable/171393

[59] A. M. Krieger, D. Azriel, and A. Kapelner, "Nearly Random Designs with Greatly Improved Balance," *Biometrika*, vol. 106, no. 3, pp. 695–701, Sep. 2019, doi: 10.1093/biomet/asz026.

[60] N. Karmarkar, R. M. Karp, G. S. Lueker, and A. M. Odlyzko, "Probabilistic Analysis of Optimum Partitioning," *Journal of Applied Probability*, vol. 23, no. 3, pp. 626–645, 1986, doi: 10.2307/3214002.

[61] M. Kearns, "Efficient Noise-Tolerant Learning from Statistical Queries," *Journal of the ACM*, vol. 45, no. 6, pp. 983–1006, Nov. 1998, doi: 10.1145/293347.293351.

[62] N. Kistler, "Derrida's Random Energy Models. From Spin Glasses to the Extremes of Correlated Random Fields." Accessed: Mar. 30, 2025. [Online]. Available: http://arxiv.org/abs/1412.0958

[63] N. Karmarkar and R. M. Karp, "The Differencing Method of Set Partitioning," 1983. Accessed: Mar. 15, 2025. [Online]. Available: https://www2.eecs.berkeley.edu/Pubs/TechRpts/1983/6353.html

[64] J. Kratica, J. Kojić, and A. Savić, "Two Metaheuristic Approaches for Solving Multidimensional Two-Way Number Partitioning Problem," *Computers & Operations Research*, vol. 46, pp. 59–68, Jun. 2014, doi: 10.1016/j.cor.2014.01.003.

[65] F. Koehler and E. Mossel, "Reconstruction on Trees and Low-Degree Polynomials." Accessed: Mar. 31, 2025. [Online]. Available: http://arxiv.org/abs/2109.06915

[66] R. E. Korf, "Multi-Way Number Partitioning," in *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, in IJCAI'09. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Jul. 2009, pp. 538–543.

[67] R. E. Korf, "From Approximate to Optimal Solutions: A Case Study of Number Partitioning," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1*, in IJCAI'95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Aug. 1995, pp. 266–272.

[68] R. E. Korf, "A Complete Anytime Algorithm for Number Partitioning," *Artificial Intelligence*, vol. 106, no. 2, pp. 181–203, Dec. 1998, doi: 10.1016/S0004-3702(98)00086-1.

[69] P. K. Kothari, R. Mori, R. O'Donnell, and D. Witmer, "Sum of Squares Lower Bounds for Refuting Any CSP." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1701.04521

[70] D. Kunisky, "Low Coordinate Degree Algorithms I: Universality of Computational Thresholds for Hypothesis Testing." Accessed: Mar. 26, 2025. [Online]. Available: http://arxiv.org/abs/2403.07862

[71] D. Kunisky, "Low Coordinate Degree Algorithms II: Categorical Signals and Generalized Stochastic Block Models." Accessed: Mar. 26, 2025. [Online]. Available: http://arxiv.org/abs/2412.21155

[72] D. Kunisky, A. S. Wein, and A. S. Bandeira, "Notes on Computational Hardness of Hypothesis Testing: Predictions Using the Low-Degree Likelihood Ratio." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1907.11636

[73] T. Lesieur, F. Krzakala, and L. Zdeborová, "MMSE of Probabilistic Low-Rank Matrix Estimation: Universality with Respect to the Output Channel," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2015, pp. 680–687. doi: 10.1109/ALLERTON.2015.7447070.

[74] T. Lesieur, F. Krzakala, and L. Zdeborova, "Phase Transitions in Sparse PCA," in *2015 IEEE International Symposium on Information Theory (ISIT)*, Jun. 2015, pp. 1635–1639. doi: 10.1109/ISIT.2015.7282733.

[75] S. Lovett and R. Meka, "Constructive Discrepancy Minimization by Walking on The Edges." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1203.5747

[76] G. Li, C. Rosenthal, and H. Rabitz, "High Dimensional Model Representations," *The Journal of Physical Chemistry A*, vol. 105, no. 33, pp. 7765–7777, Aug. 2001, doi: 10.1021/jp010450t.

[77] S. Li and T. Schramm, "Some Easy Optimization Problems Have the Overlap-Gap Property." Accessed: Mar. 31, 2025. [Online]. Available: http://arxiv.org/abs/2411.01836

[78] G. S. Lueker, "A Note on the Average-Case Behavior of a Simple Differencing Method for Partitioning," *Operations Research Letters*, vol. 6, no. 6, pp. 285–287, Dec. 1987, doi: 10.1016/0167-6377(87)90044-7.

[79] S. Mertens, "A Physicist's Approach to Number Partitioning," *Theoretical Computer Science*, vol. 265, no. 1, pp. 79–108, Aug. 2001, doi: 10.1016/S0304-3975(01)00153-0.

[80] S. Mertens, "The Easiest Hard Problem: Number Partitioning." Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/cond-mat/0310317

[81] R. Merkle and M. Hellman, "Hiding Information and Signatures in Trapdoor Knapsacks," *IEEE Transactions on Information Theory*, vol. 24, no. 5, pp. 525–530, Sep. 1978, doi: 10.1109/TIT.1978.1055927.

[82] W. Michiels, J. Korst, E. Aarts, and J. Van Leeuwen, "Performance Ratios for the Differencing Method Applied to the Balanced Number Partitioning Problem," *STACS 2003*, vol. 2607. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 583–595, 2003. doi: 10.1007/3-540-36494-3_51.

[83] M. Mézard, T. Mora, and R. Zecchina, "Clustering of Solutions in the Random Satisfiability Problem," *Physical Review Letters*, vol. 94, no. 19, p. 197205, May 2005, doi: 10.1103/PhysRevLett.94.197205.

[84] R. Meka, A. Potechin, and A. Wigderson, "Sum-of-Squares Lower Bounds for Planted Clique," in *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, Portland Oregon USA: ACM, Jun. 2015, pp. 87–96. doi: 10.1145/2746539.2746600.

[85] R. O'Donnell, "Analysis of Boolean Functions." Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/2105.10386

[86] T. Rothvoss, "Constructive Discrepancy Minimization for Convex Sets." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1404.0339

[87] P. Raghavendra, T. Schramm, and D. Steurer, "High-Dimensional Estimation via Sum-of-Squares Proofs." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1807.11419

[88] M. Rahman and B. Virag, "Local Algorithms for Independent Sets Are Half-Optimal," *The Annals of Probability*, vol. 45, no. 3, May 2017, doi: 10.1214/16-AOP1094.

[89] V. Santucci, M. Baioletti, and G. Di Bari, "An Improved Memetic Algebraic Differential Evolution for Solving the Multidimensional Two-Way Number Partitioning Problem," *Expert Systems with Applications*, vol. 178, p. 114938, Sep. 2021, doi: 10.1016/j.eswa.2021.114938.

[90] R. H. Storer, S. W. Flanders, and S. David Wu, "Problem Space Local Search for Number Partitioning," *Annals of Operations Research*, vol. 63, no. 4, pp. 463–487, Aug. 1996, doi: 10.1007/BF02156630.

[91] A. Shamir, "A Polynomial Time Algorithm for Breaking the Basic Merkle-Hellman Cryptosystem," in *23rd Annual Symposium on Foundations of Computer Science (Sfcs 1982)*, Nov. 1982, pp. 145–152. doi: 10.1109/SFCS.1982.5.

[92] J. Spencer, "Six Standard Deviations Suffice," *Transactions of the American Mathematical Society*, vol. 289, no. 2, pp. 679–706, 1985, doi: 10.1090/S0002-9947-1985-0784009-0.

[93] P. Turner, R. Meka, and P. Rigollet, "Balancing Gaussian Vectors in High Dimension." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1910.13972

[94] L.-H. Tsai, "Asymptotic Analysis of an Algorithm for Balanced Parallel Processor Scheduling," *SIAM Journal on Computing*, vol. 21, no. 1, pp. 59–64, Feb. 1992, doi: 10.1137/0221007.

[95] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, 1st ed. in Cambridge Series in Statistical and Probabilistic Mathematics. New York, NY: Cambridge University Press, 2018.

[96] N. Vafa and V. Vaikuntanathan, "Symmetric Perceptrons, Number Partitioning and Lattices." Accessed: Mar. 20, 2025. [Online]. Available: http://arxiv.org/abs/2501.16517

[97] A. S. Wein, "Optimal Low-Degree Hardness of Maximum Independent Set." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/2010.06563

[98] J. Wen *et al.*, "Optical Experimental Solution for the Multiway Number Partitioning Problem and Its Application to Computing Power Scheduling," *Science China Physics, Mechanics & Astronomy*, vol. 66, no. 9, p. 290313, Sep. 2023, doi: 10.1007/s11433-023-2147-3.

[99] B. Yakir, "The Differencing Algorithm LDM for Partitioning: A Proof of a Conjecture of Karmarkar and Karp," *Mathematics of Operations Research*, vol. 21, no. 1, pp. 85–99, Feb. 1996, doi: 10.1287/moor.21.1.85.

[100] L. Zdeborová and F. Krzakala, "Statistical Physics of Inference: Thresholds and Algorithms," *Advances in Physics*, vol. 65, no. 5, pp. 453–552, Sep. 2016, doi: 10.1080/00018732.2016.1211393.