SLDH For NPP

Strong Low Degree Hardness for the Number Partitioning Problem

Rushil Mallarapu[1]
Harvard University, Cambridge, MA 02138

Abstract. Meow Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet, ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum defuturum, quas natura non depravata desiderat. Et quem ad me accedis, saluto: 'chaere,' inquam, 'Tite!' lictores, turma omnis chorusque: 'chaere, Tite!' hinc hostis mi Albucius, hinc inimicus. Sed iure Mucius. Ego autem mirari satis non queo unde hoc sit tam insolens domesticarum rerum fastidium. Non est omnino hic docendi locus; sed ita prorsus existimo, neque eum Torquatum, qui hoc primus cognomen invenerit, aut torquem illum hosti detraxisse, ut aliquam ex eo est consecutus? – Laudem et caritatem, quae sunt vitae.

[1] Written under the joint supervision of Professor Mark Sellke and Professor Subhabrata Sen.

Acknowledgments

Meow Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet, ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum defuturum, quas natura non depravata desiderat. Et quem ad me accedis, saluto: 'chaere,' inquam, 'Tite!' lictores, turma omnis chorusque: 'chaere, Tite!' hinc hostis mi Albucius, hinc inimicus. Sed iure Mucius. Ego autem mirari satis non queo unde hoc sit tam insolens domesticarum rerum fastidium. Non est omnino hic docendi locus; sed ita prorsus existimo, neque eum Torquatum, qui hoc primus cognomen invenerit, aut torquem illum hosti detraxisse, ut aliquam ex eo est consecutus? – Laudem et caritatem, quae sunt vitae sine metu degendae praesidia firmissima. – Filium morte multavit. – Si sine causa, nollem me ab eo delectari, quod ista Platonis, Aristoteli, Theophrasti orationis ornamenta neglexerit. Nam illud quidem physici, credere aliquid esse minimum, quod profecto numquam putavisset, si a Polyaeno, familiari suo, geometrica discere maluisset quam illum etiam ipsum dedocere. Sol Democrito magnus videtur, quippe homini erudito in geometriaque perfecto, huic pedalis fortasse; tantum enim esse omnino in nostris poetis aut inertissimae segnitiae est aut fastidii delicatissimi. Mihi quidem videtur, inermis ac nudus est. Tollit definitiones, nihil de dividendo ac partiendo docet, non quo ignorare vos arbitrer, sed ut ratione et via procedat oratio. Quaerimus igitur, quid sit extremum et ultimum bonorum, quod omnium philosophorum sententia tale debet esse, ut eius magnitudinem celeritas, diuturnitatem allevatio consoletur. Ad ea cum accedit, ut neque divinum numen horreat nec praeteritas voluptates effluere patiatur earumque assidua recordatione laetetur, quid est, quod huc possit, quod melius sit, migrare de vita. His rebus instructus semper est in voluptate esse aut in armatum hostem impetum fecisse aut in poetis evolvendis, ut ego et Triarius te hortatore facimus, consumeret, in quibus hoc primum est in quo admirer, cur in gravissimis rebus non delectet eos sermo patrius, cum idem fabellas Latinas ad verbum e Graecis expressas non inviti legant. Quis enim tam inimicus paene nomini Romano est, qui Ennii Medeam aut Antiopam Pacuvii spernat aut reiciat, quod se isdem Euripidis fabulis delectari dicat, Latinas litteras oderit? Synephebos ego, inquit, potius Caecilii aut Andriam Terentii quam utramque Menandri legam? A quibus tantum dissentio, ut, cum Sophocles vel optime scripserit Electram, tamen male conversam Atilii mihi legendam putem, de quo Lucilius: 'ferreum scriptorem', verum, opinor, scriptorem tamen, ut legendus sit. Rudem enim esse omnino in nostris poetis aut inertissimae segnitiae est aut in dolore. Omnis autem privatione doloris putat Epicurus.

Contents

1 Introduction

Suppose we have N items, each with associated weights. How should we divide these items into two groups such that the sum of their weights is as close as possible? Alternatively, is it possible to divide these items into two groups such that the absolute difference of the sum of their weights is below a certain threshold? This question is known in statistics, physics, and computer science as the *number partitioning problem (NPP)*, and has been the subject of intense study since its proposal in 1969 [1].

Let $g_1, \ldots, g_N$ be N real numbers. The *number partitioning problem (NPP)* asks: what is the subset A of $[N] := \{1, 2, \ldots, N\}$ such that the sum of the g_i for $i \in A$ and the sum of the remaining g_i are as close as possible? More formally, the A we want to find is the one minimizing the discrepancy

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

When rephrased as a decision problem (i.e., whether there exists an A 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 [2].

The number partitioning problem can be rephrased in the following way. Let our instance $g_1, \ldots, g_N$ be identified with a point $g \in \mathbf{R}^N$. Then, a choice of $A \subseteq [N]$ 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:

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

()

A common extension to the NPP 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. While what "mutually close" precisely means varies across the literature, this problem has myriad practical applications which motivate the study of the NPP.

For instance, the NPP was first formulated by Graham, who considered it in the context of multiprocessor 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 [1]. Later work by Coffman, Garey, and Johnson, as well as by Tsai, looked at utilizing algorithms designed for the NPP for designing multiprocessor schedulers or large integrated circuits [3, 4]. 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 [5].

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 [6]. Their proposal was for the reciever, 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. Then, to encrypt a N-bit message, $x = (x_1, ..., x_N)$, with $x_i \in \{0, 1\}$, 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 [2]. 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 reciever 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 [7]. 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.

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 $\boldsymbol{g}_i \in \mathbf{R}^d$. 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 $A_- := [N] \setminus A_+$) to minimize

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

This multidimensional 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 [8, 9].

Another major source of interest in the NPP comes 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 indepedent of each other [10–12]. 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 [13, 14]. The NPP was the first system for which this local REM conjecture was shown [15, 16]. 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 $\kappa := \frac{m}{n}$ [17, 18]. 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 [19]. 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 [20, 21]. (concluding sentence meow)

Of more interest to us is the the typical optical discrepancy in the average-case, i.e., here we assume the instance inputs g_i are i.i.d. random variables.

The landmark result here is by Karmarkar et al., who showed that when the distribution of the g_i's is sufficiently nice,[2] then the minimum discrepancy of (1.1) is $\Theta\left(\sqrt{N}2^{-N}\right)$ as $N \to \infty$ with high probability as $N \to \infty$ [22]. Their result also extends to *even partitions*, where the sizes of each subset is equal (i.e., for N even), worsening only to $\Theta(N2^{-N})$.

On the algorithmic side, ()

A first approach, 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 [23]. 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, which makes it useful for randomized control trials applications [8].

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* [24]. 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 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 that PDM could achieve the KK 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 [25]. 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 $N^{-\Theta(\log N)}$ [26]. The constant α was later estimated for LDM to be $\alpha = \frac{1}{2 \ln 2}$, via non-rigorous calculations [27].

Of course, at its most basic, 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 acheived the global optimum. To this degree, Korf developed alternatives known as the *complete greedy* and

[2] Specifically, having bounded density, variance σ^2, and finite 4th moment.

complete Karmarkar-Karp algorithms which, if run for exponentially long time, can find the globally optimal partition [28, 29]. This algorithm was later extended to multiway number partitioning [30]. See also Michiels et al. for extensions to balaced multiway partitioning [31]. ()

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 $\|\boldsymbol{g}_i\|_\infty \leq 1$ for all i [32]. 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 $\boldsymbol{g}_i$,[2] the minimum discrepancy is $\Theta\left(\sqrt{N}2^{-N/d}\right)$ for all $d \leq o(N)$, with high probability [33]. For the regime $\delta = \Theta(N)$, Aubin et al. conjecture there exists an explicit function $c(\delta)$ such that for $\delta > 0$, then the discrepancy in the $d = \delta N$ regime is $c(\delta)\sqrt{N}$ with high probability [34]. 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% [33]. 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 $\widetilde{O}\left(\sqrt{N\log(2d/N)}\right)$ [35]. Yet, many proposed algorithms can achieve similar discrepancies, which is believed to be optimal for $d \geq N$ [32, 36–38]. (concluding sentence)

1.1 Hardness and Statistical-to-Computational Gaps

Many problems involving searches over random combinatorial structures (i.e., throughout high-dimensional statistics) are known to exhibit a statistical-to-computational gap. In the pure optimization setting, such gaps are exhibited in random constraint satisfaction [39–41], finding maximal independent sets in sparse random graphs [42, 43], the largest submatrix problem [44, 45], and the p-spin and diluted p-spin models [46–48]. These gaps also arise in various "planted" models, such as matrix or tensor PCA [49–54], high-dimensional linear regression [55, 56], or the infamously hard planted clique problem [57–61].

Evidence of hardness:

- Failure of MCMC: [57, 62]
- Failure of AMP: [63, 64]
- Reductions from planted clique - [49, 65, 66]
- Lower bounds agains Sum of Squares hierarchy: [52, 53, 60, 67]
- Lower bounds in statistical query model: [68–70]
- Low degree methods, and low degree likelihood ratio: [71, 72]

1.2 Overlap Gap Property

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?

One approach would be to prove random-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 concldued that no such polynomial-time algorithm for the average case can exist [73]. To this extent, Hoberg et al. provided such evidence of hardness for the NPP by showing that a polynomial-time approximation oracle that achieved 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 [74]. 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 [75]. Other conjectures suggested that the onset of algorithmic hardness was related to phase transitions in the solution landscapes, something which 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 landscapes. Many of the "hard" random optimization problems have a certain disconnectivity property, known as the *overlap gap property (OGP)* [73]. 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, every pair 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 itself, can then be leveraged to rule out the success of a wide class of stable algorithms [39, 40, 42, 76–79].

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 [19]. This was backed up by the failure of many algorithms based on locally refining KK-optimal solutions, such as simulated annealing [80–84]. To that end, more recent approaches for algorithmic development are rooted in more global heuristics [85–87], 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 [88, 89]. 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 $\omega(1) \leq E_N \leq o(N)$ [90]. 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 $\omega\left(n \log^{-\frac{1}{5}+\varepsilon} N\right) \leq E_N$ [90]. These results point to the efficacy of using landscape obstructions to show algorithmic hardness for the NPP, which we'll take advantage of in Section 3.

1.3 Our Results

Low degree heuristic: degree D algorithms are a proxy for the class of $e^{\widetilde{O}(D)}$-time algorithms.

[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.

Definition 1.1 (Strong Low-Degree Hardness)**.** A random search problem, i.e. a N-indexed sequence of input vectors $y_N \in \mathbf{R}^{d_N}$ and random subsets $S_N = S_N(y_N) \subseteq \Sigma_N$, exhibits *strong low degree hardness (SLDH) up to degree* $D \leq o(D_N)$ if, for all sequences of degree $o(D_N)$ algorithms $(\mathcal{A}_N)$ with $\mathbf{E}\|\mathcal{A}(y_N)\|^2 \leq O(N)$, we have

$$\mathbf{P}(\mathcal{A}(y_N) \in S_N) \leq o(1).$$

There are two related notions of degree which we want to consider in Definition 1.1. The first is traditional polynomial degree, applicable for algorithms given in each coordinate by low degree polynomial functions of the inputs. The second uses the more general notion of *coordinate degree*: a function $f : \mathbf{R}^N \to \mathbf{R}$ has coordinate degree D if it can be expressed as a linear combination of functions depending on combinations of no more than D coordinates. ()

Our reasons for condisdering low degree algorithms are twofold.

Theorem 1.2 (Results of Section 3.1)**.** *The NPP exhibits SLDH for degree D polynomial algorithms, when*

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

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

Theorem 1.3 (Results of Section 3.2)**.** *The NPP exhibits SLDH for coordinate degree D algorithms, when*

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

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

1.4 Notation and Conventions

Definition 1.4. Let $x \in \Sigma_N$. The *energy* of x (with respect to the instance g) 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

$$|\langle g, x\rangle| \leq 2^{-E}. \tag{1.3}$$

- This terminology is motivated by the statistical physics literature, wherein random optimiztation problems are often reframed as energy maximization over a random landscape [19].
- Observe that minimizing the discrepancy $|\langle g, x\rangle|$ corresponds to maximizing the energy E.

1. "instance"/"disorder" - g, instance of the NPP problem
2. "discrepancy" - for a given g, value of $\min_{x \in \Sigma_N} |\langle g, x\rangle|$
3. "energy" - negative exponent of discrepancy, i.e. if discrepancy is 2^{-E}, then energy is E. Lower energy indicates "worse" discrepancy.
4. "near-ground state"/"approximate solution"

Conventions:

1. On $\mathbf{R}^N$ we write $\|\cdot\|_2 = \|\cdot\|$ for the Euclidean norm, and $\|\cdot\|_1$ for the ℓ^1 norm.
2. If $x \in \mathbf{R}^N$ and $S \subseteq [N]$, then x_S is vector with

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

In particular, for $x, y \in \mathbf{R}^N$,

$$\langle x_S, y\rangle = \langle x, y_S\rangle = \langle x_S, y_S\rangle.$$

3. meow
4. If $S \subseteq [N]$, then $\overline{S} := [N] \setminus S$ is the complementary set of indices.
5. $B(x, r) = \{y \in \mathbf{R}^N : \|y - x\| < r\}$ is ℓ^2 unit ball.
6. Recall by Jensen's inequality that for any real numbers $d_1, ..., d_n$, we have

$$\left(\sum_{i=1}^{n} d_i\right)^2 \leq n \sum_{i=1}^{n} d_i^2.$$

We will use this in the following way: suppose $x^{(1)}, ..., x^{(n)}, x^{(n+1)}$ are n vectors in $\mathbf{R}^N$. Then

$$\left\|x^{(1)} - x^{(n+1)}\right\|^2 \leq \left(\sum_{i=1}^{n}\left\|x^{(i)} - x^{(i+1)}\right\|\right)^2 \leq n \sum_{i=1}^{n} \left\|x^{(i)} - x^{(i+1)}\right\|^2 \tag{1.4}$$

Throughout we will make key use of the following lemma:

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

$$\mathbf{P}(|Z| \leq 2^{-E} \mid \mu) \leq \exp_2\left(-E - \frac{1}{2}\log_2(\sigma^2) + O(1)\right). \tag{1.5}$$

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

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

Integrating over $|z| \leq 2^{-E}$ then gives (1.5), via

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

Note that this is decreasing function of σ^2. Thus, for instance if there exists γ with $\sigma^2 \geq \gamma > 0$, then (1.5) is bounded by $\exp_2(-E - \log_2(\gamma)/2 + O(1))$.

Lemma 1.6. *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

$$\begin{aligned}\sum_{k\leq K}\binom{N}{k} &\leq p^{-K}(1-p)^{-(N-K)} \\ &= \exp_2(-K\log_2(p) - (N-K)\log_2(1-p))\end{aligned}$$

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

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

2 Low-Degree Algorithms

For our purposes, an *algorithm* is a function which takes as input a problem 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. Alternatively, we could consider *randomized algorithms* via taking as additional input some randomness ω independent of the problem instance. However, most of our analysis will focus on the deterministic case.

To further restrict the category of algorithms considered, we specifically restrict to *low degree algorithms*. Compared to analytically-defined classes of algorithms (e.g. Lipschitz), these algorithms have a regular algebraic structure that we can exploit to precisely control their stability properties. In particular, our goal is to show *strong low degree hardness*, in the sense of [62].

Definition 2.1 (Strong Low-Degree Hardness)**.** A random search problem, namely a N-indexed sequence of input vectors $y_N \in \mathbf{R}^{d_N}$ and random subsets $S_N = S_{N(y_N)} \subseteq \Sigma_N$, exhibits *strong low degree hardness up to degree $D \le o(D_N)$* if, for all sequences of degree $o(D_N)$ algorithms $(\mathcal{A}_N)$ with $\mathbf{E}\|\mathcal{A}(y_N)\|^2 \le O(N)$, we have

$$\mathbf{P}(\mathcal{A}(y_N) \in S_N) \le o(1).$$

Low degree heuristic: [91, 92] for extensions [93]

Sometimes it fails: [94]

In addition, degree D polynomials are a heuristic proxy for the class of $e^{\widetilde{O}(D)}$-time algorithms [41, 71]. Thus, strong low degree hardness up to $o(N)$ can be thought of as evidence of requiring exponential (i.e. $e^{\Omega(N)}$) time to find globally optimal solutions.

For the case of NPP, we consider two distinct notions of degree. One is traditional polynomial degree, which has an intuitive interpretation, but the other, known in the ltierature as "coordinate degree," is a more flexible notion which can be applied to a much broader class of algorithms. As we will see in Section 3, these classes of algorithms exhibit quantitatively different behavior, in line with existing heuristics for the "brittleness" of NPP.

2.1 Coordinate Degree and L^2 Stability

First, we consider a general class of putative algorithms, where the notion of "degree" corresponds to how many variables can interact nonlinearly with each other. Given this notion, deriving stability bounds becomes a straightforward piece of functional analysis. To start, recall the notion of L^2 functions:

Definition 2.2. 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 $f : \mathbf{R}^N \to \mathbf{R}$ 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.

Note that 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 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 2.3. Let $f \in L^2(\mathbf{R}^N, \pi^{\otimes N})$ and $J \subseteq [N]$, with $\overline{J} = [N] \setminus J$. The *projection of f onto J* is the function $f^{\subseteq J} : \mathbf{R}^N \to \mathbf{R}$ 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 $T \nsubseteq S$ 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, and exhibits the following combinatorial construction. Our presentation largely follows [95], as well as the paper [91].

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 $f^{\subseteq S}$ 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 2.3 is the constant function $\mathbf{E}[f]$. Next, if $J = \{j\}$ is a singleton, (2.2) gives

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

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

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

This function only depends on x_j; all other coordinates are averaged over, thus measuring how the expectation of f changes given x_j.

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

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

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

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

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

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

Theorem 2.4 ([95])**.** *Let $f \in L^2(\mathbf{R}^N, \pi^{\otimes N})$. Then f 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 any $L^2(\mathbf{R}^N, \pi^{\otimes N})$ function into it's different interaction levels not only uniquely exists, but is an orthogonal decomposition, enabling us to apply tools from elementary Fourier analysis.

Theorem 2.4 further implies that we can define subspaces of $L^2(\mathbf{R}^N, \pi^{\otimes N})$ (see also [91])

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

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

With this, we can define the notion of "coordinate degree":

Definition 2.5. The *coordinate degree* of a function $f \in L^2(\mathbf{R}^N, \pi^{\otimes N})$ is

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

If $f = (f_1, ..., f_M) : \mathbf{R}^N \to \mathbf{R}^M$ is a multivariate function, then

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

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 function behaves under small changes in its input, we are led to consider the following "noise operator," which lets us measures the effect of small changes in the input on the coordinate decomposition. First, we need the following notion of distance between problem instances:

Definition 2.6. 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 a.s. never occurs if they were p-correlated.

Definition 2.7. 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 2.8. *Let $p \in [0,1]$ and $f \in L^2(\mathbf{R}^N, \pi^{\otimes N})$ have Efron-Stein decomposition $f = \sum_{S \subseteq [N]} f^{=S}$. Then*

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

Proof: Let J denote a p-random subset of $[N]$, i.e. with J 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 2.4 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 2.9. *Let $p \in [0,1]$ and let $f = (f_1, \ldots, f_M) : \mathbf{R}^N \to \mathbf{R}^M$ be a multivariate function with coordinate degree D and each $f_i \in L^2(\mathbf{R}^N, \pi^{\otimes N})$. 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 2.8 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 $\mathbf{E}\|f(x)\|^2 = 1$, gives

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

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 similar bounds as Theorem 2.9. This theory is much more classical, so we encourage the interested reader to see [95] for details.

Definition 2.10. Let γ_N be the N-dimensional standard Normal measure on $\mathbf{R}^N$. Then 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.

It is a well-known fact that the monomials $1, z, z^2, \ldots$ form a complete basis for $L^2(\mathbf{R}, \gamma)$ [95]. However, these are far from an orthonormal "Fourier" basis; for instance, we know $\mathbf{E}[z^2] = 1$ for

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

$z \sim \mathcal{N}(0,1)$. By the Gram-Schmidt process, these monomials can be converted into the *(normalized) Hermite polynomials* h_j for $j \geq 0$, given as

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

Note here that each h_j is a degree j polynomial. With these, we have:

Theorem 2.11 ([95]). *The Hermite polynomials $\left(h_j\right)_{j\geq 0}$ 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 $h_\alpha : \mathbf{R}^N \to \mathbf{R}$ 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 2.12. *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 in 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 , 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 2.13. Let (x, y) be 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.

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

Definition 2.14. For $p \in [0,1]$, the *Gaussian noise operator* T_p is the linear operator on $L^2(\mathbf{R}^N, \gamma^N)$, given by

$$T_p f(x) = \mathbf{E}_{y \text{ } p\text{-correlated to } x}[f(y)] = \mathbf{E}_{y\sim\mathcal{N}(0,I_N)}\left[f\left(px + \sqrt{1-p^2}y\right)\right]$$

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 2.15 ([95]). *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 2.9.

Theorem 2.16. *Let $p \in [0,1]$ and let $f = (f_1, \ldots, f_M) : \mathbf{R}^N \to \mathbf{R}^M$ be a multivariate polynomial with degree D. 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 2.9 (see also [96]). The main modification is to realize that for each f_i, having degree at most D implies that $\widehat{f_i}(\alpha) = 0$ for $|\alpha| > D$. Thus, as $p^D \leq p^s \leq 1$ for all $s \leq D$, we can apply Lemma 2.15 to get

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

From there, the proof proceeds as before. □

As a comparision to the case for functions with coordinate degree D, notice that Theorem 2.16 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 strengthenings in the arguments which follow, justifying equal consideration of both classes of functions.

2.3 Stability of Low-Degree Algorithms

With these notions of low degree functions/polynomials in hand, we can consider algorithms based on such functions.

Definition 2.17. 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.

In practice, we want to consider $\mathbf{R}^N$-valued algorithms as opposed to Σ_N-valued ones to avoid the resulting restrictions on the component functions. These can then be converted to Σ_N-valued algorithms by some rounding procedure. We discuss the necessary extensions to handling this rounding in Section 4.

Definition 2.18. 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, we say it has degree D if each coordinate has degree at most D (with at least one equality).

We can broaden the notion of polynomial algorithms (with their obvious notion of degree) to algorithms with a well-defined notion of coordinate degree:

Definition 2.19. 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 each of its coordinate functions.

By the low degree heuristic, these algorithms can be interpreted as a proxy for time N^D-algorithms, unlike classes based off of their stability properties, such as Lipschitz/Hölder continuous algorithms. Yet in addition to this interpretability, these algorithms also have accessible stability bounds:

Proposition 2.20 (Low-Degree Stability – [62])**.** *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 2.16 (or Theorem 2.9, 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. □

3 Proof of Strong Low-Degree Hardness

In this section, we prove strong low degree hardness for both low degree polynomial algorithms and algorithms with low Efron-Stein degree.

For now, we consider Σ_N-valued deterministic algorithms. We discuss the extension to $\mathbf{R}^N$-valued algorithms in Section 4. As outlined in Section 1.3, we show that TODO.

The key argument is as follows. Fix some energy levels E, depending on N. Suppose we have a Σ_N-valued, deterministic algorithm $\mathcal{A}$ given by a degree D polynomial (resp. an Efron-Stein degree D function), and we have two instances $g, g' \sim \mathcal{N}(0, I_N)$ which are $(1-\varepsilon)$-correlated (resp. $(1-\varepsilon)$-resampled), for $\varepsilon > 0$. Say $\mathcal{A}(g) = x \in \Sigma_N$ is a solution with energy at least E, i.e. it "solves" this NPP instance. For ε close to 0, $\mathcal{A}(g') = x'$ will be close to x, by low degree stability. However, by adjusting parameters carefully, we can make it so that with high probability (exponential in E), there are no solutions to g' close to x. By application of a correlation bound on the probability of solving any fixed instance, we can conclude that with high probability, $\mathcal{A}$ can't find solutions to NPP with energy E.

Our argument utilizes what can be thought of as a "conditional" version of the overlap gap property. Traditionally, the overlap gap property is a global obstruction: one shows that with high probability, one cannot find a tuple 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's no way to know where they'll end up.

First moment details meow.

We start with some setup which will apply, with minor modifications depending on the nature of the algorithm in consideration, to all of the energy regimes in discussion. After proving some preliminary estimates, we establish the existence of our conditional landscape obstruction, which is of independent interest. Finally, we conclude by establishing low degree hardness in both the linear and sublinear energy regimes.

Explain more meow.

3.1 Hardness for Low Degree Polynomial Algorithms

First, consider the case of $\mathcal{A}$ being a polynomial algorithm with degree D.

Let g, g' be $(1-\varepsilon)$-correlated standard Normal r.v.s, and let $x \in \Sigma_N$ depend only on g. Furthermore, let $\eta > 0$ be a parameter which will be chosen in a manner specified later. We define the following events:

$$S_{\text{solve}} = \{\mathcal{A}(g) \in S(E; g), \mathcal{A}(g') \in S(E; g')\} \tag{3.1}$$

$$
\begin{aligned}
S_{\mathrm{stable}} &= \left\{ \|\mathcal{A}(g) - \mathcal{A}(g')\| \le 2\sqrt{\eta N} \right\} \\
S_{\mathrm{cond}}(x) &= \left\{ \begin{array}{l} \nexists\ x' \in S(E; g') \text{ such that} \\ \|x - x'\| \le 2\sqrt{\eta N} \end{array} \right\}
\end{aligned}
\tag{3.1}
$$

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 3.1. *We have, for $x := \mathcal{A}(g)$, $S_{\mathrm{solve}} \cap S_{\mathrm{stable}} \cap S_{\mathrm{cond}}(x) = \emptyset$.*

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 have that $x' \in S(E; g')$ while also being within distance $2\sqrt{\eta N}$ of x. This contradicts $S_{\mathrm{cond}}(x)$, thus completing the proof. □

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

$$
p^{\mathrm{cor}}_{\mathrm{solve}} = \mathbf{P}(\mathcal{A}(g) \in S(E; g)). \tag{3.2}
$$

Then, we have the following correlation bound, which allows us to avoid union bounding over instances:

Lemma 3.2. *For g, g' being $(1-\varepsilon)$-correlated, we have*

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

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

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

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

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

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

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

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

In addition, define

$$
p^{\mathrm{cor}}_{\mathrm{cond}} := \max_{x \in \Sigma_N} p^{\mathrm{cor}}_{\mathrm{cond}}(x). \tag{3.3}
$$

By Lemma 3.1, we know that for $x := \mathcal{A}(g)$

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

and rearranging, we get that

$$
(p^{\mathrm{cor}}_{\mathrm{solve}})^2 \le p^{\mathrm{cor}}_{\mathrm{unstable}} + p^{\mathrm{cor}}_{\mathrm{cond}} \tag{3.4}
$$

Our proof follows by showing that, for appropriate choices of ε and η, depending on D, E, and N, we have $p^{\text{cor}}_{\text{unstable}}, p^{\text{cor}}_{\text{cond}} = o(1)$.

To this end, we start by bounding the size of neighborhoods on Σ_N.

Proposition 3.3 (Hypercube Neighborhood Size)**.** *Fix $x \in \Sigma_N$, and let $\eta \le 1/2$. Then the number of x' within distance $2\sqrt{\eta N}$ of x is*

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

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 $\|x - x'\| \le 2\sqrt{\eta N}$ iff $k \le N\eta$. Moreover, for $\eta \le \frac{1}{2}$, $k \le \frac{N}{2}$. Thus, by Lemma 1.6, we get

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

This shows that 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 how many of these are solutions to a correlated/resampled instance.

First, we consider the conditional probability of any fixed $x \in \Sigma_N$ solving a $(1 - \varepsilon)$-correlated problem instance g', given g:

Putting together these bounds, we conclude the following fundamental estimates of $p^{\text{cor}}_{\text{cond}}$, i.e. of the failure of our conditional landscape obstruction.

Proposition 3.4 (Fundamental Estimate – Correlated Case)**.** *Assume that (g, g') are $(1 - \varepsilon)$-correlated standard Normal vectors. Then, for any x only depending on g,*

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

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

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

so that

$$p^{\text{cor}}_{\text{cond}}(x) = \mathbf{E}\left[\sum_{\|x - x'\| \le 2\sqrt{\eta N}} \mathbf{E}[I_{x'} \mid g]\right] = \mathbf{E}\left[\sum_{\|x - x'\| \le 2\sqrt{\eta N}} \mathbf{P}(|\langle g', x'\rangle| \le 2^{-E} \mid g)\right] \tag{3.5}$$

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 $g' = pg + \sqrt{1 - p^2}\tilde{g}$, 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 $(1 - p^2)N \ge \varepsilon N > 0$; thus by Lemma 1.5, we get

$$\mathbf{P}(|\langle g', x'\rangle| \le 2^{-E} \mid g) \le \exp_2\left(-E - \frac{1}{2}\log_2(\varepsilon) + O(\log_2 N)\right). \tag{3.6}$$

Finally, by Proposition 3.3, the number of terms in the sum (3.5) is bounded by $\exp_2(2\eta\log_2(1/\eta)N)$, so given that (3.6) is independent of g, we conclude that

$$p^{\mathrm{cor}}_{\mathrm{cond}}(x) \le \exp_2\left(-E + -\frac{1}{2}\log_2(\varepsilon) + 2\eta\log_2\left(\frac{1}{\eta}\right)N + O(\log_2 N)\right). \qquad \square$$

Note for instance that ε can be exponentially small in E (e.g. $\varepsilon = \exp_2(-E/10)$), which for the case $E = \Theta(N)$ implies ε can be exponentially small in N.

Transition para meow.

Throughout this section, we let $E = \delta N$ for some $\delta > 0$, and aim to rule out the existence of low degree algorithms achieving these energy levels. This corresponds to the statistically optimal regime, as per [22]. These results roughly correspond to those in [90], although their result applies to stable algorithms more generally, and does not show a low degree hardness-type result.

Theorem 3.5. *Let $\delta > 0$ and $E = \delta N$, and let g, g' be $(1-\varepsilon)$-correlated standard Normal r.v.s. Then, for any degree $D \le o(\exp_2(\delta N/2))$ polynomial algorithm $\mathcal{A}$ (with $\mathbf{E}\|\mathcal{A}(g)\|^2 \le CN$), there exist $\varepsilon, \eta > 0$ such that $p^{\mathrm{cor}}_{\mathrm{solve}} = o(1)$.*

Proof: Recall from (3.4) that it suffices to show that both $p^{\mathrm{cor}}_{\mathrm{cond}}$ and $p^{\mathrm{cor}}_{\mathrm{unstable}}$ go to zero. Further, by (3.3) and Proposition 3.4, we have

$$p^{\mathrm{cor}}_{\mathrm{cond}} \le \exp_2\left(-E - \frac{1}{2}\log_2(\varepsilon) + 2\eta\log_2\left(\frac{1}{\eta}\right)N + O(\log_2 N)\right)$$

Thus, first choose η sufficiently small, such that $2\eta\log_2(1/\eta) < \delta/4$ – this results in η being independent of N. Next, choose $\varepsilon = \exp_2(-\delta N/2)$. This gives

$$p^{\mathrm{cor}}_{\mathrm{cond}} \le \exp_2\left(-\delta N - \frac{1}{2}\left(-\frac{\delta N}{2}\right) + \frac{\delta N}{4} + O(\log_2 N)\right) = \exp_2\left(-\frac{\delta N}{2} + O(\log_2 N)\right) = o(1).$$

Moreover, for $D \le o(\exp_2(\delta N/2))$, we get by Proposition 2.20 that

$$p^{\mathrm{cor}}_{\mathrm{unstable}} \le \frac{CD\varepsilon}{2\eta} \asymp \frac{D\varepsilon}{\eta} \asymp D\cdot\exp_2\left(-\frac{\delta N}{2}\right) \to 0.$$

By (3.4), we conclude that $(p^{\mathrm{cor}}_{\mathrm{solve}})^2 \le p^{\mathrm{cor}}_{\mathrm{unstable}} + p^{\mathrm{cor}}_{\mathrm{cond}} = o(1)$, thus completing the proof. $\square$

Remark that this implies poly algs are really bad, requiring double exponential time. meow.

Next, we let $\omega(\log_2 N) \le E \le o(N)$.

Theorem 3.6. *Let $\omega(\log_2^2 N) \le E \le o(N)$, and let g, g' be $(1-\varepsilon)$-correlated standard Normal r.v.s. Then, for any polynomial algorithm $\mathcal{A}$ with degree $D \le o(\exp_2(E/4))$ (and with $\mathbf{E}\|\mathcal{A}(g)\|^2 \le CN$), there exist $\varepsilon, \eta > 0$ such that $p^{\mathrm{cor}}_{\mathrm{solve}} = o(1)$.*

Proof: As in Theorem 3.5, it suffices to show that both $p^{\mathrm{cor}}_{\mathrm{cond}}$ and $p^{\mathrm{cor}}_{\mathrm{unstable}}$ go to zero. To do this, we choose

$$\varepsilon = \exp_2\left(-\frac{E}{2}\right), \qquad \eta = \frac{E}{16N\log_2(N/E)}. \tag{3.7}$$

With this choice of η, some simple analysis shows that for $\frac{E}{N} \ll 1$, we have that

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

Thus, by Proposition 3.4, we get

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

where the last equality follows as $E \gg \log_2 N$. Then, by Proposition 2.20, the choice of $D = o(\exp_2(E/4))$ 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, as $E \gg \log_2 N$. Ergo, by (3.4), $(p_{\text{solve}}^{\text{cor}})^2 \leq p_{\text{unstable}}^{\text{cor}} + p_{\text{cond}}^{\text{cor}} = o(1)$, as desired. □

3.2 Proof for Low Coordinate-Degree Algorithms

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

$$\begin{aligned} S_{\text{diff}} &= \{g \neq g'\} \\ S_{\text{solve}} &= \{\mathcal{A}(g) \in S(E;g), \mathcal{A}(g') \in S(E;g')\} \\ S_{\text{stable}} &= \left\{\|\mathcal{A}(g) - \mathcal{A}(g')\| \leq 2\sqrt{\eta N}\right\} \\ S_{\text{cond}}(x) &= \left\{\begin{matrix} \nexists\ x' \in S(E;g') \text{ such that} \\ \|x - x'\| \leq 2\sqrt{\eta N} \end{matrix}\right\} \end{aligned} \tag{3.8}$$

Note that these are the same events as (3.1), along with an event to ensure that g' is nontrivially resampled from g.

Lemma 3.7. *For g, g' being $(1-\varepsilon)$-resampled,* $\mathbf{P}(S_{\text{diff}}) = 1 - (1-\varepsilon)^N \leq \varepsilon N$.

Proof: Follows from calculation:

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

□

Lemma 3.8. *We have, for $x = \mathcal{A}(g)$, $S_{\text{diff}} \cap S_{\text{solve}} \cap S_{\text{stable}} \cap S_{\text{cond}}(x) = \emptyset$.*

Proof: This follows from Lemma 3.1, noting that the proof did not use that $g \neq g'$ almost surely. □

We should 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 $p_{\text{solve}}^{\text{cor}}$ in (3.2), which we now term $p_{\text{solve}}^{\text{res}}$, remains valid. In particular, we have

Lemma 3.9. *For g, g' being $(1-\varepsilon)$-resampled, we have*

$$\mathbf{P}(S_{\text{solve}}) = \mathbf{P}(\mathcal{A}(g) \in S(E; g), \mathcal{A}(g') \in S(E; g')) \geq \left(p_{\text{solve}}^{\text{res}}\right)^2$$

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

$$g = \tilde{g}_J + g^{(0)}_{[N]\setminus J}, \qquad g' = \tilde{g}_J + g^{(1)}_{[N]\setminus J}$$

where $\tilde{g}_J$ denotes the vector with coordinates $\tilde{g}_i$ if $i \in J$ and 0 else. Thus g and g' are conditionally independent, given $(\tilde{g}, J)$, and the proof concludes as in Lemma 3.2. □

Let us slightly redefine $p_{\text{unstable}}^{\text{res}}$ and $p_{\text{cond}}^{\text{res}}(x)$ by

$$p_{\text{unstable}}^{\text{res}} = 1 - \mathbf{P}(S_{\text{stable}} \mid S_{\text{diff}}), \qquad p_{\text{cond}}^{\text{res}}(x) = 1 - \mathbf{P}(S_{\text{cond}}(x) \mid S_{\text{diff}}). \tag{3.9}$$

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, these definitions agree with what we had in (3.4). Again, we can define $p_{\text{cond}}^{\text{res}}$ via (3.3), i.e. as the maximum of $p_{\text{cond}}^{\text{res}}(x)$ over Σ_N.

Now, by Lemma 3.8, 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

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

Thus, rearranging and multiplying by $\mathbf{P}(S_{\text{diff}})$ (so as to apply Lemma 3.9) gives

$$\left(p_{\text{solve}}^{\text{res}}\right)^2 \leq \mathbf{P}(S_{\text{diff}}) \cdot \left(p_{\text{unstable}}^{\text{res}} + p_{\text{cond}}^{\text{res}}\right) \tag{3.10}$$

As before, our proof follows by showing that, for appropriate choices of ε and η, depending on D, E, and N, that $p_{\text{unstable}}^{\text{res}}, p_{\text{cond}}^{\text{res}} = o(1)$. However, this also requires us to choose $\varepsilon \gg \frac{1}{N}$, so as to ensure that $g \neq g'$, as otherwise $p_{\text{unstable}}^{\text{res}}, p_{\text{cond}}^{\text{res}}$ would be too large. This restriction on ε effectively limits us from showing hardness for algorithms with degree larger than $o(N)$, as we will see shortly.

First, we bound the same probability of a fixed x solving a resampled instance. 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 3.10 (Fundamental Estimate – Resampled Case)**.** *Assume that (g, g') are $(1-\varepsilon)$-resampled standard Normal vectors. Then, for any x only depending on g,*

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

Proof: We follow the setup of proof of Proposition 3.4. 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

$$\begin{aligned} p_{\text{cond}}^{\text{res}}(x) &= \mathbf{E}\left[\sum_{\|x-x'\| \le 2\sqrt{\eta N}} \mathbf{E}[I_{x'} \mid g, g \ne g'] \right] \\ &= \mathbf{E}\left[\sum_{\|x-x'\| \le 2\sqrt{\eta N}} \mathbf{P}(|\langle g', x' \rangle| \le 2^{-E} \mid g, g \ne g') \,\middle|\, g \ne g' \right] \end{aligned} \tag{3.11}$$

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

$$\mathbf{P}(|\langle g', x' \rangle| \le 2^{-E} \mid g, g \ne g') \le \exp_2(-E + O(1)). \tag{3.12}$$

By Proposition 3.3, the number of terms in the sum (3.11) is bounded by $\exp_2(2\eta \log_2(1/\eta) N)$, so summing (3.12) allows us to conclude that

$$p_{\text{cond}}^{\text{res}}(x) \le \exp_2\left(-E + 2\eta \log_2\left(\frac{1}{\eta}\right) N + O(1) \right). \qquad \square$$

Note that in contrast to Proposition 3.4, this bound doesn't involve ε at all, but the condition $g \ne g'$ requires $\varepsilon = \omega(1/N)$ to hold almost surely, by Lemma 3.7.

With this, we can show strong low degree hardness for low coordinate degree algorithms at energy levels $E = \Theta(N)$.

Theorem 3.11. *Let $\delta > 0$ and $E = \delta N$, and let g, g' be $(1-\varepsilon)$-resampled standard Normal r.v.s. Then, for any algorithm $\mathcal{A}$ with coordinate degree $D \le o(N)$ and $\mathbf{E}\|\mathcal{A}(g)\|^2 \le CN$, there exist $\varepsilon, \eta > 0$ such that $p_{\text{solve}}^{\text{res}} = o(1)$.*

Proof: Recall from (3.10) that it suffices to show that both $p_{\text{cond}}^{\text{res}}$ and $p_{\text{unstable}}^{\text{res}}$ go to zero, while $\mathbf{P}(S_{\text{diff}}) \approx 1$. By Lemma 3.7, the latter condition is satisfied for $\varepsilon = \omega(1/N)$. Thus, pick

$$\varepsilon = \frac{\log_2(N/D)}{N}. \tag{3.13}$$

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. As the bound in Proposition 3.10 is independent of x, we get

$$p_{\mathrm{cond}}^{\mathrm{res}} \le \exp_2\left(-\delta N + \frac{\delta N}{4} + O(1)\right) = o(1).$$

Moreover, for $D \le o(N)$, Proposition 2.20 now gives

$$p_{\mathrm{unstable}}^{\mathrm{res}} \le \frac{CD\varepsilon}{2\eta} \asymp D \cdot \frac{\log_2(N/D)}{N} \to 0,$$

as $x \log_2(1/x) \to 0$ for $x \ll 1$. By (3.10), we conclude that $(p_{\mathrm{solve}}^{\mathrm{res}})^2 \le \mathbf{P}(S_{\mathrm{diff}}) \cdot (p_{\mathrm{unstable}}^{\mathrm{res}} + p_{\mathrm{cond}}^{\mathrm{res}}) = o(1)$, thus completing the proof. □

Sublinear case. We now consider sublinear energy levels, ranging from $(\log_2 N)^2 \ll E \ll N$. Note here that we have to increase our lower bound to $(\log_2 N)^2$ as opposed to $\log_2 N$ from Theorem 3.6, to address the requirement that $\varepsilon = \omega(1/N)$.

Theorem 3.12. *Let $\omega\big((\log_2 N)^2\big) \le E \le o(N)$, and let g, g' be $(1-\varepsilon)$-resampled standard Normal r.v.s. Then, for any algorithm $\mathcal{A}$ with coordinate degree $D \le o\big(E/(\log_2 N)^2\big)$ and $\mathbf{E}\|\mathcal{A}(g)\|^2 \le CN$, there exist $\varepsilon, \eta > 0$ such that $p_{\mathrm{solve}}^{\mathrm{res}} = o(1)$.*

Proof: As in Theorem 3.11, choose ε as in (3.13), so that $\varepsilon = \omega(1/N)$ and $\mathbf{P}(S_{\mathrm{diff}}) \approx 1$. However, to account for $E \le o(N)$, we need to adjust η as $N \to \infty$. Thus, choose η as in (3.7): this ensures that $\varepsilon = \omega(1/N)$ and that $2\eta \log_2(1/\eta) < E/4N$ for $E \ll N$. By Proposition 3.10, this guarantees that

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

The low coordinate degree requirement $D \le o\big(E/(\log_2 N)^2\big)$ plus Proposition 2.20 now gives

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

By (3.10), $(p_{\mathrm{solve}}^{\mathrm{res}})^2 \le \mathbf{P}(S_{\mathrm{diff}}) \cdot (p_{\mathrm{unstable}}^{\mathrm{res}} + p_{\mathrm{cond}}^{\mathrm{res}}) = o(1)$, thus completing the proof. □

3.3 Summary of Parameters

Parameter	Meaning	Desired Direction	Intuition
N	Dimension	Large	Showing hardness *asymptotically*, want "bad behavior" to pop up in low dimensions.
E	Solution energy; want to find x such that $\lvert\langle g, x\rangle\rvert \leq 2^{-E}$	Small	Smaller E implies weaker solutions, and can consider full range of $1 \ll E \ll N$. Know that $E > (\log^2 N)$ by [24]
D	Algorithm degree (in either Efron-Stein sense or usual polynomial sense.)	Large	Higher degree means more complexity. Want to show even complex algorithms fail.
ε	Complement of correlation/ resample probability; (g,g') are $(1-\varepsilon)$-correlated.	Small	ε is "distance" between g, g'. Want to show that small changes in disorder lead to "breaking" of landscape.
η	Algorithm instability; $\mathcal{A}$ is stable if $\lVert\mathcal{A}(g) - \mathcal{A}(g')\rVert \leq 2\sqrt{\eta N}$, for (g, g') close.	Large	Large η indicates a more unstable algorithm; want to show that even weakly stable algorithms fail.

Table 1: Explanation of Parameters

4 Extensions to Real-Valued Algorithms

With Section 3, we have established strong low degree hardness for both low degree polynomial algorithms 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 an $\mathbf{R}^N$-valued algorithm. We want to show that

I. No low degree $\mathcal{A}$ can reliably output points *close* – within constant distance – to a solution,

II. No Σ_N-valued algorithm $\widetilde{\mathcal{A}}$ 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 3, while the second fails because 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, the more common way to convert a $\mathbf{R}^N$-valued algorithm into a Σ_N-valued one is by rounding the outputs, as in [62]. Doing this directly can lead to difficulties in performing the stability analysis. In our case, though, if we know 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 are those which flip $\omega(1)$ many coordinates.

We first describe a landscape obstruction to finding multiple solutions at the same energy level for a random NPP instance. Then, we show hardness in both of the aforementioned cases. meow.

4.1 Solutions repel meow

Introduce section meow.

No two adjacent points on Σ_N (or pairs within $k = O(1)$ distance) which are both good solutions to the same problem.

Proposition 4.1. *Fix distinct points $x, x' \in \Sigma_N$ and let $g \sim \mathcal{N}(0, I_N)$ be a random instance. Then,*

$$\mathbf{P}(x, x' \in S(E; g)) \leq \exp_2(-E + O(1)) = \exp_2(-E + O(1)).$$

Proof: For $x \neq x'$, let $J \subseteq [N]$ denote the subset of coordinates in which x, x' differ, i.e. $x_J \neq x'_J$. In particular, we can write

$$x = x_{[N]\setminus J} + x_J, \qquad x' = x_{[N]\setminus J} - x_J.$$

Thus, for a fixed pair (x, x'), if $-2^{-E} \leq \langle g, x\rangle, \langle g, x'\rangle \leq 2^{-E}$, we can expand this into

$$-2^{-E} \leq \left\langle g, x_{[N]\setminus J}\right\rangle + \langle g, x_J\rangle \leq 2^{-E},$$
$$-2^{-E} \leq \left\langle g, x_{[N]\setminus J}\right\rangle - \langle g, x_J\rangle \leq 2^{-E}.$$

Multiplying the lower equation by -1 and adding the resulting inequalities gives $|\langle g, x_J\rangle| \leq 2^{-E}$. Note that $\langle g, x_J\rangle \sim \mathcal{N}(0, |J|)$ (and is nondegenerate, as $|J| > 0$). By Lemma 1.5 and the following remark, it follows that

$$\mathbf{P}(x, x' \in S(E; g)) \leq \mathbf{P}(|\langle g, x_J\rangle| \leq 2^{-E}) \leq \exp_2(-E + O(1)). \qquad \square$$

Remarks on theorem below meow.

Theorem 4.2 (Solutions Can't Be Close)**.** *Consider any distances $k = \Omega(1)$ and energy levels $E \gg k\log_2 N$. Then for any instance g, there are no pairs of distinct solutions $x, x' \in S(E; g)$ with $\|x - x'\| \leq 2\sqrt{k}$ (i.e. within k coordinate flips of each other) with high probability.*

Proof: Observe that by Proposition 4.1, finding a pair of distinct solutions within distance $2\sqrt{k}$ implies finding some subset of at most k coordinates $J \subset [N]$ of g and $|J|$ signs x_J such that $|\langle g_J, x_J\rangle|$ is small. For any g, there are at most 2^k choices of signs and, by [97], there are

$$\sum_{1 \leq k' \leq k} \binom{N}{k'} \leq \left(\frac{eN}{k}\right)^k \leq (eN)^k = 2^{O(k\log_2 N)}$$

choices of such subsets. Union bounding Proposition 4.1 over these $\exp_2 O(k\log_2 N)$ choices, we get

$$\mathbf{P}\begin{pmatrix} \exists \text{ x},x' \text{ s.t.} \\ \text{(a) } \|x - x'\| \leq 2\sqrt{k}, \\ \text{(b) } \text{x},x' \in S(E;g) \end{pmatrix} \leq \mathbf{P}\begin{pmatrix} \exists \text{ J} \subset [N],\ x_J \in \{\pm 1\}^{|J|} \text{ s.t.} \\ \text{(a) } |J| \leq \text{k}, \\ \text{(b) } |\langle g_J, x_J\rangle| \leq \exp_2\ (-E) \end{pmatrix} \leq \exp_2(-E + O(k\log_2 N)) = o(1). \tag{4.1}$$

Note that the last equality holds as $E \gg k\log_2 N$. $\square$

4.2 Proof of Hardness for Close Algorithms

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

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

Note that as r is fixed (potentially depending on $\mathcal{A}$, but independent of N or g), we can convert $\mathcal{A}$ into a Σ_N-valued algorithm by considering the corners of Σ_N within constant distance of $\mathcal{A}(g)$.

Definition 4.3. Let $r > 0$ and $\mathcal{A}$ be an $\mathbf{R}^N$-valued algorithm. Define $\widehat{\mathcal{A}_r}$ to be the Σ_N-valued algorithm defined by

$$\widehat{\mathcal{A}_r}(g) := \underset{x' \in B(\mathcal{A}(g), r) \cap \Sigma_N}{\operatorname{argmin}} |\langle g, x'\rangle|. \tag{4.2}$$

If $B(\mathcal{A}(g), r) \cap \Sigma_N = \emptyset$, then set $\widehat{\mathcal{A}_r}(g) := (1/g_1, 0, \ldots)$, which always has energy 0.

Observe that $S_{\mathrm{close}(r)}$ occuring is the same as $\widehat{\mathcal{A}_r}$ finding a solution for g. In addition, note that practically speaking, computing $\widehat{\mathcal{A}_r}$ requires additionally computing 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 $\mathcal{A}$ is low degree (or low coordinate degree) then we can derive useful stability bounds for its outputs. Luckily, this modification $\widehat{\mathcal{A}_r}$ of $\mathcal{A}$ also are stable, with slightly modified bounds.

Lemma 4.4. *Suppose that* $\mathbf{E}\|\mathcal{A}(g)\|^2 \leq CN$ *and that* $\mathcal{A}$ *has degree* $\leq D$ *(resp. coordinate degree* $\leq D$*), and let* (g, g') *be* $(1-\varepsilon)$*-correlated (resp.* $(1-\varepsilon)$*-resampled). Then* $\widehat{\mathcal{A}_r}$ *as defined above has*

$$\mathbf{E}\left\|\widehat{\mathcal{A}_r}(g) - \widehat{\mathcal{A}_r}(g')\right\|^2 \leq 6CD\varepsilon N + 6r^2.$$

In particular, we have

$$\mathbf{P}\left(\left\|\widehat{\mathcal{A}_r}(g) - \widehat{\mathcal{A}_r}(g')\right\| \geq 2\sqrt{\eta N}\right) \leq \frac{3CD\varepsilon}{2\eta} + \frac{3r^2}{2\eta N}. \tag{4.3}$$

Proof: Observe by the triangle inequality, as per (1.4), that

$$\left\|\widehat{\mathcal{A}_r}(g) - \widehat{\mathcal{A}_r}(g')\right\|^2 \leq 3\left(\left\|\widehat{\mathcal{A}_r}(g) - \mathcal{A}(g)\right\|^2 + \|\mathcal{A}(g) - \mathcal{A}(g')\|^2 + \left\|\mathcal{A}(g') - \widehat{\mathcal{A}_r}(g')\right\|^2\right).$$

By Proposition 2.20, we know $\mathbf{E}\|\mathcal{A}(g) - \mathcal{A}(g')\|^2 \leq 6CD\varepsilon N$. Moreover, we know that $\left\|\widehat{\mathcal{A}_r}(g) - \mathcal{A}(g)\right\| \leq r$ by definition, so the remaining terms can be bounded by $3r^2$ each deterministically. Finally, (4.2) follows from Markov's inequality. □

Of course, computing $\widehat{\mathcal{A}_r}$ is certainly never polynomial, and does not preserve any low coordinate degree assumptions in a controllable way. Thus, we cannot directly hope for Theorem 3.5, Theorem 3.6, Theorem 3.11, or Theorem 3.12 to hold meow

We show for $\mathcal{A}$ being a $\mathbf{R}^N$-valued, low coordinate degree algorithm and any $r = O(1)$, low degree hardness still holds for $\widehat{\mathcal{A}_r}$. Note that by a similar argument, we can show hardness in the case that $\mathcal{A}$ is a low degree polynomial algorithm, but we omit the proof meow.

We recall the setup from Section 3.2. Let g, g' be $(1-\varepsilon)$-resampled standard Normal vectors. Define the following events:

$$\begin{aligned}
S_{\mathrm{diff}} &= \{g \neq g'\} \\
S_{\mathrm{solve}} &= \left\{\widehat{\mathcal{A}_r}(g) \in S(E; g), \widehat{\mathcal{A}_r}(g') \in S(E; g')\right\} \\
S_{\mathrm{stable}} &= \left\{\left\|\widehat{\mathcal{A}_r}(g) - \widehat{\mathcal{A}_r}(g')\right\| \leq 2\sqrt{\eta N}\right\} \\
S_{\mathrm{cond}}(x) &= \left\{\begin{array}{l}\nexists\, x' \in S(E; g') \text{ such that} \\ \|x - x'\| \leq 2\sqrt{\eta N}\end{array}\right\}
\end{aligned} \tag{4.4}$$

These are the same events as in (3.8), just adapted to $\widehat{\mathcal{A}_r}$. In particular, Lemma 3.8 holds unchanged.

Moreover, we can define

$$\hat{p}^{\mathrm{cor}}_{\mathrm{solve}} = \mathbf{P}\left(\widehat{\mathcal{A}_r}(g) \in S(E; g)\right) = \mathbf{P}(S_{\mathrm{close}}(r)), \tag{4.5}$$

as well as

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

along with $\hat{p}^{\mathrm{cor}}_{\mathrm{cond}} := \max_{x \in \Sigma_N} \hat{p}^{\mathrm{cor}}_{\mathrm{cond}}(x)$, echoing (3.9).

Observe that as $\hat{p}^{\mathrm{cor}}_{\mathrm{cond}}$ makes no reference to any algorithm, the bound in Proposition 3.10 holds without change. Moreover, Lemma 4.4 lets us control $\hat{p}^{\mathrm{cor}}_{\mathrm{unstable}}$. The final piece needed is an appropriate analog of Lemma 3.9.

Lemma 4.5. *For g, g' being $(1-\varepsilon)$-resampled, we have*

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

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

$$\Big\{\widehat{\mathcal{A}}_r(g) \in S(E; g)\Big\} = \{\mathcal{A}(g) \in S(E; g) + B(0, r)\}.$$

Expanding $S(E; g)$, the proof concludes as in Lemma 3.9. □

Theorem 4.6. *Let $\omega\Big((\log_2 N)^2\Big) \leq E \leq \Theta(N)$, 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 $(\log_2 N)^2 \ll E \ll N$, then $\mathcal{A}$ has coordinate degree $D \leq o\Big(E/(\log_2 N)^2\Big)$.

Let $\widehat{\mathcal{A}}_r$ be defined as in Definition 4.3. Then there exist $\varepsilon, \eta > 0$ such that

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

Proof: First, by Lemma 3.8, the appropriate adjustment of (3.10) holds, namely that

$$(\hat{p}^{\mathrm{cor}}_{\mathrm{solve}})^2 \leq \mathbf{P}(S_{\mathrm{diff}}) \cdot (\hat{p}^{\mathrm{cor}}_{\mathrm{unstable}} + \hat{p}^{\mathrm{cor}}_{\mathrm{cond}}). \tag{4.6}$$

To ensure $\mathbf{P}(S_{\mathrm{diff}}) \approx 1$, we begin by following (3.13) and choosing $\varepsilon = \log_2(N/D)/N$. Moreover, following the proof of Theorem 3.11 and Theorem 3.12, we know that choosing

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

in conjunction with Proposition 3.10, guarantees that

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

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

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

as $E \gg (\log_2 N)^2$. Thus, applying the properly modified Lemma 4.4 with these choices of ε, η, we see that $\hat{p}^{\mathrm{cor}}_{\mathrm{unstable}} = o(1)$. By (4.6), we conclude that $\hat{p}^{\mathrm{cor}}_{\mathrm{solve}} = o(1)$, as desired. □

Talk about implications meow.

4.3 Truly Random Rounding

At this point, one might wonder whether, while deterministic algorithms fail, perhaps a randomized rounding scheme could save us, maybe by assiging small values to coordinates which it was less confident in. However, this approach is blunted by the same brittleness of the NPP landscape that established the conditional obstruction of Proposition 3.4 and Proposition 3.10. In particular, by Theorem 4.2, if you have a subcube of Σ_N nonconstant but bounded dimension, then with high probability at most 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 $(\log_2 N)^2 \ll E \leq N$; note that for lower values of E, algorithms like [24] already achieve discrepancies of $N^{O(\log_2 N)}$ 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 define the modified alg $\mathcal{A}^*(g) := \mathcal{A}(g)^*$.

Remark 4.7. meow $\mathcal{A}^*$ fails and is still degree D lcdf, even if it stops being a polynomial. Bounds on D worsen, but only to what you'd expect. Note that if $\mathcal{A}$ has coordinate degree D, then $\mathcal{A}^*$ also has coordinate degree D. As a deterministic Σ_N-valued algorithm, strong low degree hardness as proved in the previous section applies.

In contrast to deterministically taking signs of the outputs of $\mathcal{A}$ (which corresponds to deterministically rounding the outputs of $\mathcal{A}$ to Σ_N), 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. Given such a randomized rounding function, we can describe it in the following way. Let $p_1(x), ..., p_N(x)$ be defined by

$$p_i(x) := \max\left(\mathbf{P}(\mathsf{round}(x)_i \neq x_i^*), \frac{1}{2}\right). \tag{4.7}$$

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

Lemma 4.8. *Fix $x \in \mathbf{R}^N$, and draw N coin flips $I_{x,i} \sim \mathrm{Bern}(2p_i(x))$ and N signs $S_i \sim \mathrm{Unif}\{\pm 1\}$, all mutually independent, and define the random variable $\tilde{x} \in \Sigma_N$ by*

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

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

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

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

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

By Lemma 4.8, we can redefine $\mathsf{round}(x)$ to be $\tilde{x}$ as constructed without loss of generality.

It thus makes sense to define $\widetilde{\mathcal{A}}(g) := \mathsf{round}(\mathcal{A}(g))$, 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 doesn't depend directly on g, only the output $x = \mathcal{A}(g)$). We should expect that if $\widetilde{\mathcal{A}} = \mathcal{A}^*$ (e.g., if $\mathcal{A}$ outputs values far outside the cube $[-1, 1]^N$) with high probability, then low degree hardness will still apply, as $\mathcal{A}^*$ is deterministic. However, in general, any $\widetilde{\mathcal{A}}$ 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 be effectively 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 4.8 which a.s. differs from picking the closest point deterministically will resample a diverging number of coordinates.

Lemma 4.9. *Fix $x \in \mathbf{R}^N$, and let $p_1(x), ..., p_N(x)$ be defined as in (4.7). Then $\tilde{x} \neq x^*$ with high probability iff $\sum_i p_i(x) = \omega(1)$. 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

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

Thus, $\mathbf{P}(\tilde{x} = x^*) = o(1)$ iff $\sum_i p_i(x) = \omega(1)$, as claimed.

Next, following the construction of $\tilde{x}$ in Lemma 4.8, 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 $\tilde{x}_i$ is resampled infinitely often with probability 1, thus completing the proof. □

With this, we can apply Theorem 4.2, which shows that solutions resist clustering at a rate related to their energy level (i.e. higher energy solutions are push each other further apart), to conclude that any $\widetilde{\mathcal{A}}$ which is not equal to $\mathcal{A}^*$ with high probability fails to find solutions.

Theorem 4.10. *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 $E \geq \omega\left((\log_2 N)^2\right)$, we have*

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

Proof: Following the characterization of $\tilde{x}$ in Lemma 4.8, let $K := \max\left(\log_2 N, \sum_i I_{x,i}\right)$. By assumption on $\sum_i p_i(x)$ and Lemma 4.9, 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 $S \subseteq [N]$ denote the set of the first K coordinates to be resampled, so that $K = |S|$. Consider now

$$\mathbf{P}\left(\tilde{x} \in S(E; g) \mid \tilde{x}_{[N] \setminus S}\right),$$

where we fix the coordinates outside of S 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 4.2, the probability there is more than one such $\tilde{x} \in S(E; g)$ is bounded by

$$\exp_2(-E + O(K \log_2 N)) \leq \exp_2\Big(-E + O\Big((\log_2 N)^2\Big)\Big) = o(1),$$

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

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

meow discuss possible extensions of randomization schemes and whether you expect those to work instead.

5 Conclusion

Meow

5.1 Future Work

Bibliography

1. Graham, R.L.: Bounds on Multiprocessing Timing Anomalies. SIAM Journal on Applied Mathematics. 17, 416–429 (1969). https://doi.org/10.1137/0117039.
2. Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman, New York (1979).
3. Coffman, E.G., Jr., Garey, M.R., Johnson, D.S.: An Application of Bin-Packing to Multiprocessor Scheduling. SIAM Journal on Computing. 7, 1–17 (1978). https://doi.org/10.1137/0207001.
4. Tsai, L.-H.: Asymptotic Analysis of an Algorithm for Balanced Parallel Processor Scheduling. SIAM Journal on Computing. 21, 59–64 (1992). https://doi.org/10.1137/0221007.
5. Coffman, E.G., Lueker, G.S.: Probabilistic Analysis of Packing and Partitioning Algorithms. J. Wiley & sons, New York (1991).
6. Merkle, R., Hellman, M.: Hiding Information and Signatures in Trapdoor Knapsacks. IEEE Transactions on Information Theory. 24, 525–530 (1978). https://doi.org/10.1109/TIT.1978.1055927.
7. Shamir, A.: A Polynomial Time Algorithm for Breaking the Basic Merkle-Hellman Cryptosystem. In: 23rd Annual Symposium on Foundations of Computer Science (Sfcs 1982). pp. 145–152 (1982). https://doi.org/10.1109/SFCS.1982.5.
8. Krieger, A.M., Azriel, D., Kapelner, A.: Nearly Random Designs with Greatly Improved Balance. Biometrika. 106, 695–701 (2019). https://doi.org/10.1093/biomet/asz026.
9. Harshaw, C., Sävje, F., Spielman, D., Zhang, P.: Balancing Covariates in Randomized Experiments with the Gram-Schmidt Walk Design, http://arxiv.org/abs/1911.03071, last accessed 2025/03/15. https://doi.org/10.48550/arXiv.1911.03071.
10. Derrida, B.: Random-Energy Model: Limit of a Family of Disordered Models. Physical Review Letters. 45, 79–82 (1980). https://doi.org/10.1103/PhysRevLett.45.79.
11. Derrida, B.: Random-Energy Model: An Exactly Solvable Model of Disordered Systems. Physical Review B. 24, 2613–2626 (1981). https://doi.org/10.1103/PhysRevB.24.2613.
12. Bauke, H., Franz, S., Mertens, S.: Number Partitioning as a Random Energy Model. Journal of Statistical Mechanics: Theory and Experiment. 2004, P4003 (2004). https://doi.org/10.1088/1742-5468/2004/04/P04003.
13. Bauke, H., Mertens, S.: Universality in the Level Statistics of Disordered Systems. Physical Review E. 70, 25102 (2004). https://doi.org/10.1103/PhysRevE.70.025102.
14. Kistler, N.: Derrida's Random Energy Models. From Spin Glasses to the Extremes of Correlated Random Fields, http://arxiv.org/abs/1412.0958, last accessed 2025/03/30. https://doi.org/10.48550/arXiv.1412.0958.

15. Borgs, C., Chayes, J., Mertens, S., Nair, C.: Proof of the Local REM Conjecture for Number Partitioning. I: Constant Energy Scales. Random Structures & Algorithms. 34, 217–240 (2009). https://doi.org/10.1002/rsa.20255.

16. Borgs, C., Chayes, J., Mertens, S., Nair, C.: Proof of the Local REM Conjecture for Number Partitioning. II. Growing Energy Scales. Random Structures & Algorithms. 34, 241–284 (2009). https://doi.org/10.1002/rsa.20256.

17. Gent, I.P., Walsh, T.: Analysis of Heuristics for Number Partitioning. Computational Intelligence. 14, 430–451 (1998). https://doi.org/10.1111/0824-7935.00069.

18. Gent, I., Walsh, T.: Phase Transitions and Annealed Theories: Number Partitioning as a Case Study. Instituto Cultura. (2000).

19. Mertens, S.: A Physicist's Approach to Number Partitioning. Theoretical Computer Science. 265, 79–108 (2001). https://doi.org/10.1016/S0304-3975(01)00153-0.

20. Hayes, B.: The Easiest Hard Problem. American Scientist. 90, 113–117 (2002).

21. Borgs, C., Chayes, J., Pittel, B.: Phase Transition and Finite-size Scaling for the Integer Partitioning Problem. Random Structures & Algorithms. 19, 247–288 (2001). https://doi.org/10.1002/rsa.10004.

22. Karmarkar, N., Karp, R.M., Lueker, G.S., Odlyzko, A.M.: Probabilistic Analysis of Optimum Partitioning. Journal of Applied Probability. 23, 626–645 (1986). https://doi.org/10.2307/3214002.

23. Mertens, S.: The Easiest Hard Problem: Number Partitioning, http://arxiv.org/abs/cond-mat/0310317, last accessed 2025/03/15. https://doi.org/10.48550/arXiv.cond-mat/0310317.

24. Karmarkar, N., Karp, R.M.: The Differencing Method of Set Partitioning. (1983).

25. Lueker, G.S.: A Note on the Average-Case Behavior of a Simple Differencing Method for Partitioning. Operations Research Letters. 6, 285–287 (1987). https://doi.org/10.1016/0167-6377(87)90044-7.

26. Yakir, B.: The Differencing Algorithm LDM for Partitioning: A Proof of a Conjecture of Karmarkar and Karp. Mathematics of Operations Research. 21, 85–99 (1996). https://doi.org/10.1287/moor.21.1.85.

27. Boettcher, S., Mertens, S.: Analysis of the Karmarkar-Karp Differencing Algorithm. The European Physical Journal B. 65, 131–140 (2008). https://doi.org/10.1140/epjb/e2008-00320-9.

28. Korf, R.E.: From Approximate to Optimal Solutions: A Case Study of Number Partitioning. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1. pp. 266–272. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1995).

29. Korf, R.E.: A Complete Anytime Algorithm for Number Partitioning. Artificial Intelligence. 106, 181–203 (1998). https://doi.org/10.1016/S0004-3702(98)00086-1.

30. Korf, R.E.: Multi-Way Number Partitioning. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence. pp. 538–543. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2009).

31. Michiels, W., Korst, J., Aarts, E., Van Leeuwen, J.: Performance Ratios for the Differencing Method Applied to the Balanced Number Partitioning Problem, http://link.springer.com/10.1007/3-540-36494-3_51, (2003). https://doi.org/10.1007/3-540-36494-3_51.

32. Spencer, J.: Six Standard Deviations Suffice. Transactions of the American Mathematical Society. 289, 679–706 (1985). https://doi.org/10.1090/S0002-9947-1985-0784009-0.

33. Turner, P., Meka, R., Rigollet, P.: Balancing Gaussian Vectors in High Dimension, http://arxiv.org/abs/1910.13972, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1910.13972.

34. Aubin, B., Perkins, W., Zdeborová, L.: Storage Capacity in Symmetric Binary Perceptrons. Journal of Physics A: Mathematical and Theoretical. 52, 294003 (2019). https://doi.org/10.1088/1751-8121/ab227a.

35. Chandrasekaran, K., Vempala, S.: Integer Feasibility of Random Polytopes, http://arxiv.org/abs/1111.4649, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1111.4649.

36. Bansal, N.: Constructive Algorithms for Discrepancy Minimization, http://arxiv.org/abs/1002.2259, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1002.2259.

37. Lovett, S., Meka, R.: Constructive Discrepancy Minimization by Walking on The Edges, http://arxiv.org/abs/1203.5747, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1203.5747.

38. Rothvoss, T.: Constructive Discrepancy Minimization for Convex Sets, http://arxiv.org/abs/1404.0339, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1404.0339.

39. Mézard, M., Mora, T., Zecchina, R.: Clustering of Solutions in the Random Satisfiability Problem. Physical Review Letters. 94, 197205 (2005). https://doi.org/10.1103/PhysRevLett.94.197205.

40. Achlioptas, D., Coja-Oghlan, A.: Algorithmic Barriers from Phase Transitions. In: 2008 49th Annual IEEE Symposium on Foundations of Computer Science. pp. 793–802 (2008). https://doi.org/10.1109/FOCS.2008.11.

41. Kothari, P.K., Mori, R., O'Donnell, R., Witmer, D.: Sum of Squares Lower Bounds for Refuting Any CSP, http://arxiv.org/abs/1701.04521, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1701.04521.

42. Gamarnik, D., Sudan, M.: Limits of Local Algorithms over Sparse Random Graphs. In: Proceedings of the 5th Conference on Innovations in Theoretical Computer Science. pp. 369–376. Association for Computing Machinery, New York, NY, USA (2014). https://doi.org/10.1145/2554797.2554831.

43. Coja-Oghlan, A., Efthymiou, C.: On Independent Sets in Random Graphs. Random Structures & Algorithms. 47, 436–486 (2015). https://doi.org/10.1002/rsa.20550.

44. Gamarnik, D., Li, Q.: Finding a Large Submatrix of a Gaussian Random Matrix, http://arxiv.org/abs/1602.08529, last accessed 2025/03/29. https://doi.org/10.48550/arXiv.1602.08529.

45. Gamarnik, D., Jagannath, A., Sen, S.: The Overlap Gap Property in Principal Submatrix Recovery. Probability Theory and Related Fields. 181, 757–814 (2021). https://doi.org/10.1007/s00440-021-01089-7.

46. Gamarnik, D., Jagannath, A.: The Overlap Gap Property and Approximate Message Passing Algorithms for p-Spin Models, http://arxiv.org/abs/1911.06943, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1911.06943.

47. Montanari, A.: Optimization of the Sherrington-Kirkpatrick Hamiltonian, http://arxiv.org/abs/1812.10897, last accessed 2025/03/29. https://doi.org/10.48550/arXiv.1812.10897.

48. Chen, W.-K., Gamarnik, D., Panchenko, D., Rahman, M.: Suboptimality of Local Algorithms for a Class of Max-Cut Problems. The Annals of Probability. 47, (2019). https://doi.org/10.1214/18-AOP1291.

49. Berthet, Q., Rigollet, P.: Computational Lower Bounds for Sparse PCA, http://arxiv.org/abs/1304.0828, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1304.0828.

50. Lesieur, T., Krzakala, F., Zdeborová, L.: 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). pp. 680–687 (2015). https://doi.org/10.1109/ALLERTON.2015.7447070.

51. Lesieur, T., Krzakala, F., Zdeborova, L.: Phase Transitions in Sparse PCA. In: 2015 IEEE International Symposium on Information Theory (ISIT). pp. 1635–1639 (2015). https://doi.org/10.1109/ISIT.2015.7282733.

52. Hopkins, S.B., Shi, J., Steurer, D.: Tensor Principal Component Analysis via Sum-of-Squares Proofs, http://arxiv.org/abs/1507.03269, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1507.03269.

53. Hopkins, S.B., Kothari, P.K., Potechin, A., Raghavendra, P., Schramm, T., Steurer, D.: The Power of Sum-of-Squares for Detecting Hidden Structures, http://arxiv.org/abs/1710.05017, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1710.05017.

54. Arous, G.B., Gheissari, R., Jagannath, A.: Algorithmic Thresholds for Tensor PCA. The Annals of Probability. 48, 2052–2087 (2020). https://doi.org/10.1214/19-AOP1415.

55. Gamarnik, D., Zadik, I.: Sparse High-Dimensional Linear Regression. Algorithmic Barriers and a Local Search Algorithm, http://arxiv.org/abs/1711.04952, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1711.04952.

56. Gamarnik, D., Zadik, I.: High-Dimensional Regression with Binary Coefficients. Estimating Squared Error and a Phase Transition, http://arxiv.org/abs/1701.04455, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1701.04455.

57. Jerrum, M.: Large Cliques Elude the Metropolis Process. Random Structures & Algorithms. 3, 347–359 (1992). https://doi.org/10.1002/rsa.3240030402.

58. Deshpande, Y., Montanari, A.: Improved Sum-of-Squares Lower Bounds for Hidden Clique and Hidden Submatrix Problems, http://arxiv.org/abs/1502.06590, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1502.06590.

59. Meka, R., Potechin, A., Wigderson, A.: Sum-of-Squares Lower Bounds for Planted Clique. In: Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing. pp. 87–96. ACM, Portland Oregon USA (2015). https://doi.org/10.1145/2746539.2746600.

60. Barak, B., Hopkins, S.B., Kelner, J., Kothari, P.K., Moitra, A., Potechin, A.: A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem, http://arxiv.org/abs/1604.03084, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1604.03084.

61. Gamarnik, D., Zadik, I.: The Landscape of the Planted Clique Problem: Dense Subgraphs and the Overlap Gap Property, http://arxiv.org/abs/1904.07174, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1904.07174.

62. Huang, B., Sellke, M.: Strong Low Degree Hardness for Stable Local Optima in Spin Glasses, http://arxiv.org/abs/2501.06427, last accessed 2025/01/30. https://doi.org/10.48550/arXiv.2501.06427.

63. Zdeborová, L., Krzakala, F.: Statistical Physics of Inference: Thresholds and Algorithms. Advances in Physics. 65, 453–552 (2016). https://doi.org/10.1080/00018732.2016.1211393.

64. Bandeira, A.S., Perry, A., Wein, A.S.: Notes on Computational-to-Statistical Gaps: Predictions Using Statistical Physics, http://arxiv.org/abs/1803.11132, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1803.11132.

65. Brennan, M., Bresler, G.: Optimal Average-Case Reductions to Sparse PCA: From Weak Assumptions to Strong Hardness, http://arxiv.org/abs/1902.07380, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1902.07380.

66. Brennan, M., Bresler, G., Huleihel, W.: Reducibility and Computational Lower Bounds for Problems with Planted Sparse Structure, http://arxiv.org/abs/1806.07508, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1806.07508.

67. Raghavendra, P., Schramm, T., Steurer, D.: High-Dimensional Estimation via Sum-of-Squares Proofs, http://arxiv.org/abs/1807.11419, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1807.11419.

68. Kearns, M.: Efficient Noise-Tolerant Learning from Statistical Queries. Journal of the ACM. 45, 983–1006 (1998). https://doi.org/10.1145/293347.293351.

69. Diakonikolas, I., Kane, D.M., Stewart, A.: Statistical Query Lower Bounds for Robust Estimation of High-dimensional Gaussians and Gaussian Mixtures, http://arxiv.org/abs/1611.03473, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1611.03473.

70. Feldman, V., Grigorescu, E., Reyzin, L., Vempala, S., Xiao, Y.: Statistical Algorithms and a Lower Bound for Detecting Planted Clique, http://arxiv.org/abs/1201.1214, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1201.1214.

71. Hopkins, S.: Statistical Inference and the Sum of Squares Method, (2018).

72. Kunisky, D., Wein, A.S., Bandeira, A.S.: Notes on Computational Hardness of Hypothesis Testing: Predictions Using the Low-Degree Likelihood Ratio, http://arxiv.org/abs/1907.11636, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1907.11636.

73. Gamarnik, D.: The Overlap Gap Property: A Geometric Barrier to Optimizing over Random Structures. Proceedings of the National Academy of Sciences. 118, e2108492118 (2021). https://doi.org/10.1073/pnas.2108492118.

74. Hoberg, R., Ramadas, H., Rothvoss, T., Yang, X.: Number Balancing Is as Hard as Minkowski's Theorem and Shortest Vector, http://arxiv.org/abs/1611.08757, last accessed 2025/03/15. https://doi.org/10.48550/arXiv.1611.08757.

75. Vafa, N., Vaikuntanathan, V.: Symmetric Perceptrons, Number Partitioning and Lattices, http://arxiv.org/abs/2501.16517, last accessed 2025/03/20. https://doi.org/10.48550/arXiv.2501.16517.

76. Achlioptas, D., Ricci-Tersenghi, F.: On the Solution-Space Geometry of Random Constraint Satisfaction Problems, http://arxiv.org/abs/cs/0611052, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.cs/0611052.
77. Gamarnik, D., Jagannath, A., Wein, A.S.: Low-Degree Hardness of Random Optimization Problems. In: 2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS). pp. 131–140 (2020). https://doi.org/10.1109/FOCS46700.2020.00021.
78. Rahman, M., Virag, B.: Local Algorithms for Independent Sets Are Half-Optimal. The Annals of Probability. 45, (2017). https://doi.org/10.1214/16-AOP1094.
79. Wein, A.S.: Optimal Low-Degree Hardness of Maximum Independent Set, http://arxiv.org/abs/2010.06563, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.2010.06563.
80. Argüello, M.F., Feo, T.A., Goldschmidt, O.: Randomized Methods for the Number Partitioning Problem. Computers & Operations Research. 23, 103–111 (1996). https://doi.org/10.1016/0305-0548(95)E0020-L.
81. Storer, R.H., Flanders, S.W., David Wu, S.: Problem Space Local Search for Number Partitioning. Annals of Operations Research. 63, 463–487 (1996). https://doi.org/10.1007/BF02156630.
82. Johnson, D.S., Aragon, C.R., McGeoch, L.A., Schevon, C.: Optimization by Simulated Annealing: An Experimental Evaluation; Part I, Graph Partitioning. Operations Research. 37, 865–892 (1989).
83. Johnson, D.S., Aragon, C.R., McGeoch, L.A., Schevon, C.: Optimization by Simulated Annealing: An Experimental Evaluation; Part II, Graph Coloring and Number Partitioning. Operations Research. 39, 378–406 (1991).
84. Alidaee, B., Glover, F., Kochenberger, G.A., Rego, C.: A New Modeling and Solution Approach for the Number Partitioning Problem. Journal of Applied Mathematics and Decision Sciences. 2005, 113–121 (2005). https://doi.org/10.1155/JAMDS.2005.113.
85. Kratica, J., Kojić, J., Savić, A.: Two Metaheuristic Approaches for Solving Multidimensional Two-Way Number Partitioning Problem. Computers & Operations Research. 46, 59–68 (2014). https://doi.org/10.1016/j.cor.2014.01.003.
86. Corus, D., Oliveto, P.S., Yazdani, D.: Artificial Immune Systems Can Find Arbitrarily Good Approximations for the NP-hard Number Partitioning Problem. Artificial Intelligence. 274, 180–196 (2019). https://doi.org/10.1016/j.artint.2019.03.001.
87. Santucci, V., Baioletti, M., Di Bari, G.: An Improved Memetic Algebraic Differential Evolution for Solving the Multidimensional Two-Way Number Partitioning Problem. Expert Systems with Applications. 178, 114938 (2021). https://doi.org/10.1016/j.eswa.2021.114938.
88. Asproni, L., Caputo, D., Silva, B., Fazzi, G., Magagnini, M.: Accuracy and Minor Embedding in Subqubo Decomposition with Fully Connected Large Problems: A Case Study about the Number Partitioning Problem. Quantum Machine Intelligence. 2, 4 (2020). https://doi.org/10.1007/s42484-020-00014-w.
89. Wen, J., Wang, Z., Huang, Z., Cai, D., Jia, B., Cao, C., Ma, Y., Wei, H., Wen, K., Qian, L.: Optical Experimental Solution for the Multiway Number Partitioning Problem and Its Application to

Computing Power Scheduling. Science China Physics, Mechanics & Astronomy. 66, 290313 (2023). https://doi.org/10.1007/s11433-023-2147-3.

90. Gamarnik, D., Kızıldağ, E.C.: Algorithmic Obstructions in the Random Number Partitioning Problem, http://arxiv.org/abs/2103.01369, last accessed 2025/03/15. https://doi.org/10.48550/arXiv.2103.01369.

91. Kunisky, D.: Low Coordinate Degree Algorithms I: Universality of Computational Thresholds for Hypothesis Testing, http://arxiv.org/abs/2403.07862, last accessed 2025/03/26. https://doi.org/10.48550/arXiv.2403.07862.

92. Kunisky, D.: Low Coordinate Degree Algorithms II: Categorical Signals and Generalized Stochastic Block Models, http://arxiv.org/abs/2412.21155, last accessed 2025/03/26. https://doi.org/10.48550/arXiv.2412.21155.

93. Montanari, A., Wein, A.S.: Equivalence of Approximate Message Passing and Low-Degree Polynomials in Rank-One Matrix Estimation, http://arxiv.org/abs/2212.06996, last accessed 2025/03/26. https://doi.org/10.48550/arXiv.2212.06996.

94. Huang, H., Mossel, E.: Optimal Low Degree Hardness for Broadcasting on Trees, http://arxiv.org/abs/2502.04861, last accessed 2025/03/26. https://doi.org/10.48550/arXiv.2502.04861.

95. O'Donnell, R.: Analysis of Boolean Functions, http://arxiv.org/abs/2105.10386, last accessed 2025/03/15. https://doi.org/10.48550/arXiv.2105.10386.

96. Gamarnik, D., Jagannath, A., Wein, A.S.: Hardness of Random Optimization Problems for Boolean Circuits, Low-Degree Polynomials, and Langevin Dynamics, http://arxiv.org/abs/2004.12063, last accessed 2025/03/15. https://doi.org/10.48550/arXiv.2004.12063.

97. Vershynin, R.: High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge University Press, New York, NY (2018).

98. Addario-Berry, L., Devroye, L., Lugosi, G., Oliveira, R.I.: Local Optima of the Sherrington-Kirkpatrick Hamiltonian, http://arxiv.org/abs/1712.07775, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1712.07775.

99. Bayati, M., Gamarnik, D., Tetali, P.: Combinatorial Approach to the Interpolation Method and Scaling Limits in Sparse Random Graphs. The Annals of Probability. 41, (2013). https://doi.org/10.1214/12-AOP816.

100. Bismuth, S., Makarov, V., Segal-Halevi, E., Shapira, D.: Partitioning Problems with Splittings and Interval Targets, http://arxiv.org/abs/2204.11753, last accessed 2025/03/20. https://doi.org/10.48550/arXiv.2204.11753.

101. Ferreira, F.F., Fontanari, J.F.: Probabilistic Analysis of the Number Partitioning Problem. Journal of Physics A: Mathematical and General. 31, 3417 (1998). https://doi.org/10.1088/0305-4470/31/15/007.

102. Gamarnik, D., Kızıldağ, E.C., Perkins, W., Xu, C.: Algorithms and Barriers in the Symmetric Binary Perceptron Model, http://arxiv.org/abs/2203.15667, last accessed 2025/03/15. https://doi.org/10.48550/arXiv.2203.15667.

103. Gamarnik, D., Kizildag, E.: Computing the Partition Function of the Sherrington-Kirkpatrick Model Is Hard on Average. The Annals of Applied Probability. 31, (2021). https://doi.org/10.1214/20-AAP1625.

104. Gamarnik, D., Sudan, M.: Performance of Sequential Local Algorithms for the Random NAE-K-SAT Problem. SIAM Journal on Computing. 46, 590–619 (2017). https://doi.org/10.1137/140989728.

105. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer New York, New York, NY (2009). https://doi.org/10.1007/978-0-387-84858-7.

106. Hatami, H., Lovász, L., Szegedy, B.: Limits of Locally–Globally Convergent Graph Sequences. Geometric and Functional Analysis. 24, 269–296 (2014). https://doi.org/10.1007/s00039-014-0258-7.

107. Kızıldağ, E.C.: Planted Random Number Partitioning Problem, http://arxiv.org/abs/2309.15115, last accessed 2025/03/15. https://doi.org/10.48550/arXiv.2309.15115.

108. Kojić, J.: Integer Linear Programming Model for Multidimensional Two-Way Number Partitioning Problem. Computers & Mathematics with Applications. 60, 2302–2308 (2010). https://doi.org/10.1016/j.camwa.2010.08.024.

109. Lauer, J., Wormald, N.: Large Independent Sets in Regular Graphs of Large Girth. Journal of Combinatorial Theory, Series B. 97, 999–1009 (2007). https://doi.org/10.1016/j.jctb.2007.02.006.

110. Levy, A., Ramadas, H., Rothvoss, T.: Deterministic Discrepancy Minimization via the Multiplicative Weight Update Method, http://arxiv.org/abs/1611.08752, last accessed 2025/03/16. https://doi.org/10.48550/arXiv.1611.08752.

111. Wainwright, M.J.: High-Dimensional Statistics: A Non-Asymptotic Viewpoint. Cambridge University Press, Cambridge (2019). https://doi.org/10.1017/9781108627771.