1. Introduction

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 in an A such that the discrepancy is below a certain threshold, or even zero), the NPP is one of the six basic NP-complete problems of Garey and Johnson, and of those, the only one to deal with numbers [GJ79, § 3.1].

(talk about modifications and variants?)

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 $\Sigma_N := \{\pm 1\}^N$, and the discrepancy of x is now $|\langle g, x \rangle|$. The goal is now to find the x minimizing this discrepancy:

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

Definition 1.1. 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

$$|\langle g, x \rangle| \leq 2^{-E}. \tag{1.1}$$

- This terminology is motivated by the statistical physics literature, wherein random optimiztation problems are often reframed as energy maximization over a random landscape [Mer01].
- Observe that minimizing the discrepancy $|\langle g, x \rangle|$ corresponds to maximizing the energy E.

Overview of number partitioning problem.

Application: randomized control trials.

Other applications.

- Circuit design, etc.

Two questions of interest:

1. What is optimal solution.
2. How to find optimal solution.

1.1. Physical Interpretations

1.2. Statistical-to-Computational Gap

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

1.3. Existing Results

1. $X_i, 1 \leq i \leq n$ i.i.d. uniform from $\{1, 2, ..., M := 2^m\}$, with $\varkappa := \frac{m}{n}$, then phase transition going from $\varkappa < 1$ to $\varkappa > 1$.
2. Average case, X_i i.i.d. standard Normal.
3. Karmarkar [KKLO86] - NPP value is $\Theta\left(\sqrt{N} 2^{-N}\right)$ whp as $N \to \infty$ (doesn't need Normality).
4. Best polynomial-time algorithm: Karmarkar-Karp [KK82] - Discrepancy $O\left(N^{-\alpha \log N}\right) = 2^{-\Theta(\log^2 N)}$ whp as $N \to \infty$
5. PDM (paired differencing) heuristic - fails for i.i.d. uniform inputs with objective $\Theta\left(n^{-1}\right)$ (Lueker).
6. LDM (largest differencing) heuristic - works for i.i.d. Uniforms, with $n^{-\Theta(\log n)}$ (Yakir, with constant $\alpha = \frac{1}{2 \ln 2}$ calculated non-rigorously by Boettcher and Mertens).
7. Krieger - $O\left(n^{-2}\right)$ for balanced partition.
8. Hoberg [HHRY17] - computational hardness for worst-case discrepancy, as poly-time oracle that can get discrepancy to within $O\left(2^{\sqrt{n}}\right)$ would be oracle for Minkowski problem.
9. Gamarnik-Kizildag: Information-theoretic guarantee $E_n = n$, best computational guarantee $E_n = \Theta\left(\log^2 n\right)$.
10. Existence of m-OGP for $m = O(1)$ and $E_n = \Theta(n)$.
11. Absence for $\omega(1) \leq E_n = o(n)$
12. Existence for $\omega\left(\sqrt{n \log_2 n}\right) \leq E_n \leq o(n)$ for $m = \omega_{n(1)}$ (with changing η, β)
    1. While OGP not ruled out for $E_n \leq \omega\left(\sqrt{n \log_2 n}\right)$, argued that it is tight.
13. For $\varepsilon \in \left(0, \frac{1}{5}\right)$, no stable algorithm can solve $\omega\left(n \log^{-\frac{1}{5}+\varepsilon} n\right) \leq E_n \leq o(n)$
14. Possible to strengthen to $E_n = \Theta(n)$ (as $2^{-\Theta(n)} \leq 2^{-o(n)}$)

1.4. Our Results

1.5. Notation and Conventions

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. $B(x, r) = \left\{y \in \mathbf{R}^N : \|y - x\| < r\right\}$ is ℓ^2 unit ball.
5. 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

$$\left\|x^{(1)} - x^{(n+1)}\right\|^2 \le \left(\sum_{i=1}^{n} \left\|x^{(i)} - x^{(i+1)}\right\|\right)^2 \le n \sum_{i=1}^{n} \left\|x^{(i)} - x^{(i+1)}\right\|^2 \tag{1.2}$$

Throughout we will make key use of the following lemma:

Lemma 1.2 (Normal Small-Probability Estimate)**.** *Let $E, \sigma^2 > 0$, and μ, Z be random variables with $Z \mid \mu \sim \mathcal{N}(\mu, \sigma^2)$. for σ^2 a constant. Then*

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

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

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

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

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

□

Note that this is decreasing function of σ^2, e.g. it's bounded by $\exp_2\!\left(-E - \frac{1}{2}\log_2\left(\min \sigma^2\right)\right)$ (this bound is trivial unless $\sigma^2 \Rightarrow \gamma > 0$).

Lemma 1.3. *Suppose that $K \le N/2$, and let $h(x) = -x\log_2(x) - (1-x)\log_2(x)$ be the binary entropy function. Then, for $p \coloneqq K/N$,*

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

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

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

Here, the last inequality follows from the fact that $p \le (1-p)$, and we multiply each term by $\left(\frac{p}{1-p}\right)^{K-k} \le 1$. Now rearrange to get

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

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

1.5.1. Glossary:

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"

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 [HS25, Def. 3].

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 $D \leq o(D_N)$* 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

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

In addition, degree D polynomials are a heuristic proxy for the class of $e^{\tilde{O}(D)}$-time algorithms [Hop18, Kot+17]. 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.

$$\mathbf{E}\left[f^2\right] \coloneqq \int_{x=(x_1,\ldots,x_n)\in\mathbf{R}^N} f(x)^2 \, \mathrm{d}\pi^{\otimes N}(x) < \infty.$$

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 $f \in L^2\left(\mathbf{R}^N, \pi^{\otimes N}\right)$, 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 $f \in L^2\left(\mathbf{R}^N, \pi^{\otimes N}\right)$ and $J \subseteq [N]$, with $\bar{J} = [N] \setminus J$. The *projection of f onto J* is the function $f^{\subseteq J} : \mathbf{R}^N \to \mathbf{R}$ given by

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

Intuitively $f^{\subseteq J}$ is f with the $\bar{J}$ 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 $T \not\subseteq S$ and g depends only on the coordinates in T, then $\left\langle f^{=S}, g\right\rangle = 0$.

This decomposition, often called the *Efron-Stein*, *orthogonal*, or *Hoeffding* decomposition, does indeed exist, and exhibits the following combinatorial construction. Our presentation largely follows [O'D21, § 8.3], as well as the paper [Kun24].

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 $J = \varnothing$. It is clear that $f^{=\varnothing} = f^{\subseteq\varnothing}$, which, by Definition 2.3 is the constant function $\mathbf{E}[f]$. Next, if $J = \{j\}$ is a singleton, (2.2) gives

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

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 $J = \{i, j\}$,

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

We can imagine that this accounts for the two-way interaction of i and j, namely $f^{\subseteq\{i,j\}} = \mathbf{E}\left[f \mid x_i, x_j\right]$, 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

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

This construction, along with some direct calculations, leads to the following theorem on Efron-Stein decompositions:

Theorem 2.4 ([O'D21, Thm 8.35])**.** *Let $f \in L^2\left(\mathbf{R}^N, \pi^{\otimes N}\right)$. Then f has a unique* Efron-Stein decomposition *as*

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

where the functions $f^{=S} \in L^2\left(\mathbf{R}^N, \pi^{\otimes N}\right)$ satisfy

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

In addition, this decomposition has the following properties:

3. *Condition 2. holds whenever $S \not\subseteq T$.*
4. *The decomposition is orthogonal: $\left\langle f^{=S}, f^{=T}\right\rangle = 0$ 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 $L^2\left(\mathbf{R}^N, \pi^{\otimes N}\right)$ 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 $L^2\left(\mathbf{R}^N, \pi^{\otimes N}\right)$ (see also [Kun24, § 1.3])

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

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

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

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

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

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

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

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

As we are interested in how these 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 $f = (f_1, ..., f_M) : \mathbf{R}^N \to \mathbf{R}^M$ 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*

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

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

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

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[1]

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

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

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

as Theorem 2.9. This theory is much more classical, so we encourage the interested reader to see [O'D21, § 11] 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 $L^2(\mathbf{R}, \gamma)$ [O'D21, Thm 11.22]. However, these are far from an orthonormal "Fourier" basis; for instance, we know $\mathbf{E}[z^2] = 1$ for $z \sim \mathcal{N}(0, 1)$. By the Gram-Schmidt process, these monomials can be converted into the *(normalized) Hermite polynomials* h_j for $j \geq 0$, given as

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

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

Theorem 2.11 ([O'D21, Prop 11.30])**.** *The Hermite polynomials $(h_j)_{j\geq 0}$ 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

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

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 ([O'D21, Prop 11.37])**.** *Let $p \in [0,1]$ and $f \in L^2(\mathbf{R}^N, \gamma^N)$. Then $T_p f$ has Hermite expansion*

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

and in particular,

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

With this in hand, we can prove a similar stability bound to Theorem 2.9.

Theorem 2.16. *Let $p \in [0,1]$ and let $f = (f_1, ..., f_M) : \mathbf{R}^N \to \mathbf{R}^M$ 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*

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

Proof: The proof is almost identical to that of Theorem 2.9 (see also [GJW22, Lem. 3.4]). The main modification is to realize that for each f_i, having degree at most D implies that $\hat{f}_i(\alpha) = 0$ for $|\alpha| > D$. Thus, as $p^D \leq p^s \leq 1$ for all $s \leq D$, we can apply Lemma 2.15 to get

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

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 – [HS25, Prop. 1.9])**.** *Suppose we have a deterministic algorithm $\mathcal{A}$ with degree (resp. coordinate degree) $\leq D$ and norm $\mathbf{E}\|\mathcal{A}(g)\|^2 \leq CN$. Then, for inputs g, g' which are $(1-\varepsilon)$-correlated (resp. $(1-\varepsilon)$-resampled),*

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

and thus

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

Proof: Let $C' := \mathbf{E}\|\mathcal{A}(g)\|^2$, and define the rescaling $\mathcal{A}' := \mathcal{A}/\sqrt{C'}$. Then, by Theorem 2.16 (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 1.4, 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:

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

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 $x := \mathcal{A}(g)$, $S_{\text{solve}} \cap S_{\text{stable}} \cap S_{\text{cond}}(x) = \varnothing$.*

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 $S_{\text{cond}}(x)$, thus completing the proof. □

First, define p_{solve} as the probability that the algorithm solves a single random instance:

$$p_{\text{solve}} = \mathbf{P}(\mathcal{A}(g) \in S(E; g)). \tag{3.2}$$

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*

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

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

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

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

Moreover, let us define p_{unstable} and $p_{\text{cond}}(x)$ by

$$p_{\text{unstable}} = 1 - \mathbf{P}(S_{\text{stable}}), \qquad p_{\text{cond}}(x) = 1 - \mathbf{P}(S_{\text{cond}}(x)).$$

In addition, define

$$p_{\text{cond}} := \max_{x \in \Sigma_N} p_{\text{cond}}(x). \tag{3.3}$$

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

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

and rearranging, we get that

$$p_{\text{solve}}^2 \leq p_{\text{unstable}} + p_{\text{cond}} \tag{3.4}$$

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

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 $\eta \leq \frac{1}{2}$. Then the number of x' within distance $2\sqrt{\eta N}$ of x is*

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

Proof: Let k be the number of coordinates which differ between x and x' (i.e. the Hamming distance). We have $\|x - x'\|^2 = 4k$, so $\|x - x'\| \leq 2\sqrt{\eta N}$ iff $k \leq N\eta$. Moreover, for $\eta \leq \frac{1}{2}$, $k \leq \frac{N}{2}$. Thus, by Lemma 1.3, we get

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

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_{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*,*

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

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

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

so that

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

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 $\langle g', x' \rangle \mid g \sim \mathcal{N}\left(p\langle g, x' \rangle, \left(1 - p^2\right)N\right)$. Note that $\langle g', x' \rangle$ is nondegenerate for $\left(1 - p^2\right)N \geq \varepsilon N > 0$; thus by Lemma 1.2, we get

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

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

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

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 [Kar+86]. These results roughly correspond to those in [GK21, Thm. 3.2], 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 $p_{\text{solve}} = o(1)$.*

Proof: Recall from (3.4) that it suffices to show that both p_{cond} and p_{unstable} go to zero. Further, by (3.3) and Proposition 3.4, we have

$$p_{\text{cond}} \le \exp_2\left(-E - \frac{1}{2}\log_2(\varepsilon) + 2\eta\log_2\left(\frac{1}{\eta}\right)N + O(\log_2 N)\right)$$

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

$$p_{\text{cond}} \le \exp_2\left(-\delta N - \frac{1}{2}\left(-\frac{\delta N}{2}\right) + \frac{\delta N}{4} + O(\log_2 N)\right) = \exp_2\left(-\frac{\delta N}{2} + O(\log_2 N)\right) = o(1).$$

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

$$p_{\text{unstable}} \le \frac{CD\varepsilon}{2\eta} \asymp \frac{D\varepsilon}{\eta} \asymp D \cdot \exp_2\left(-\frac{\delta N}{2}\right) \to 0.$$

By (3.4), we conclude that $p_{\text{solve}}^2 \le p_{\text{unstable}} + p_{\text{cond}} = o(1)$, thus completing the proof. □

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 $p_{\text{solve}} = o(1)$.*

Proof: As in Theorem 3.5, it suffices to show that both p_{cond} and p_{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

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

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

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

again, as $E \gg \log_2 N$. Ergo, by (3.4), $p_{\text{solve}}^2 \le p_{\text{unstable}} + p_{\text{cond}} = o(1)$, 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.

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

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,* $\mathbf{P}(S_{\text{diff}}) = 1 - (1-\varepsilon)^N \le \varepsilon N$.

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 $x = \mathcal{A}(g)$,* $S_{\text{diff}} \cap S_{\text{solve}} \cap S_{\text{stable}} \cap S_{\text{cond}}(x) = \varnothing$.

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

We should interpret this as saying $S_{\text{solve}}, S_{\text{stable}}, S_{\text{cond}}$ are all mutually exclusive, conditional on $g \ne g'$.

The previous definition of p_{solve} in (3.2) remains valid. In particular, we have

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

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

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

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

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_{unstable} and $p_{\text{cond}}(x)$ by

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

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_{cond} via (3.3), i.e. as the maximum of $p_{\text{cond}}(x)$ 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

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

As before, our proof follows by showing that, for appropriate choices of ε and η, depending on D, E, and N, that $p_{\text{unstable}}, p_{\text{cond}} = o(1)$. However, this also requires us to choose $\varepsilon \gg \frac{1}{N}$, so as to ensure that $g \neq g'$, as otherwise $p_{\text{unstable}}, p_{\text{cond}}$ 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*,*

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

Proof: We follow the setup of proof of Proposition 3.4. For each x' within distance $2\sqrt{\eta N}$ of x, let

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

so that

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

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 $g \neq g'$ is equivalent to conditioning on $|J| < N$, so $N - |J| \geq 1$. Thus, applying Lemma 1.2 and integrating over all valid choices of J gives

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

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

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

Note that in contrast to Proposition 3.4, this bound doesn't involve ε at all, but the condition $g \neq g'$ 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 $D \leq o(N)$ and $\mathbf{E}\|\mathcal{A}(g)\|^2 \leq CN$, there exist $\varepsilon, \eta > 0$ such that $p_{\mathrm{solve}} = o(1)$.*

Proof: Recall from (3.10) that it suffices to show that both p_{cond} and p_{unstable} go to zero, while $\mathbf{P}(S_{\mathrm{diff}}) \approx 1$. 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

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

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

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

as $x \log_2(1/x) \to 0$ for $x \ll 1$. By (3.10), we conclude that $p_{\text{solve}}^2 \le \mathbf{P}(S_{\text{diff}}) \cdot (p_{\text{unstable}} + p_{\text{cond}}) = o(1)$, 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 $\omega\left((\log_2 N)^2\right) \le E \le o(N)$, and let g, g' be $(1-\varepsilon)$-resampled standard Normal r.v.s. Then, for any algorithm $\mathcal{A}$ with coordinate degree $D \le o\left(E/(\log_2 N)^2\right)$ and $\mathbf{E}\|\mathcal{A}(g)\|^2 \le CN$, there exist $\varepsilon, \eta > 0$ such that $p_{\text{solve}} = o(1)$.*

Proof: As in Theorem 3.11, choose ε as in (3.13), so that $\varepsilon = \omega(1/N)$ and $\mathbf{P}(S_{\text{diff}}) \approx 1$. 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

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

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

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

By (3.10), $p_{\text{solve}}^2 \le \mathbf{P}(S_{\text{diff}}) \cdot (p_{\text{unstable}} + p_{\text{cond}}) = o(1)$, 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 $E > \left(\log^2 N\right)$ by [KK83]
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 $\tilde{\mathcal{A}}$ coming from randomly rounding the output of $\mathcal{A}$, which changes an $\omega(1)$ number of coordinates, can find a solution with nonvanishing probability.

In principle, the first possibility fails via the same analysis as in Section 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 [HS25]. 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 1.2 and the following remark, it follows that

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

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 [Ver18, Exer. 0.0.5], 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

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

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

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:

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

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 $\hat{\mathcal{A}}_r$ to be the Σ_N-valued algorithm defined by

$$\hat{\mathcal{A}}_r(g) := \operatorname*{argmin}_{x' \in B(\mathcal{A}(g), r) \cap \Sigma_N} |\langle g, x' \rangle|. \qquad (4.2)$$

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

Observe that $S_{\mathrm{close}(r)}$ occuring is the same as $\hat{\mathcal{A}}_r$ finding a solution for g. In addition, note that practically speaking, computing $\hat{\mathcal{A}}_r$ requires additionally computing the energy of $O(1)$-many points on Σ_N. This requires 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 $\hat{\mathcal{A}}_r$ of $\mathcal{A}$ also are stable, with slightly modified bounds.

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

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

In particular, we have

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

Proof: Observe by the triangle inequality, as per (1.2), that

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

By Proposition 2.20, we know $\mathbf{E}\|\mathcal{A}(g) - \mathcal{A}(g')\|^2 \le 6CD\varepsilon N$. Moreover, we know that $\left\|\hat{\mathcal{A}}_r(g) - \mathcal{A}(g)\right\| \le r$ 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 $\hat{\mathcal{A}}_r$ 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 $\hat{\mathcal{A}}_r$. 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:

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

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

Moreover, we can define

$$p_{\text{solve}} = \mathbf{P}\big(\hat{\mathcal{A}}_r(g) \in S(E;g)\big) = \mathbf{P}(S_{\text{close}}(r)), \tag{4.5}$$

as well as

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

along with $p_{\text{cond}} \coloneqq \max_{x \in \Sigma_N} p_{\text{cond}}(x)$, echoing (3.9).

Observe that as p_{cond} makes no reference to any algorithm, the bound in Proposition 3.10 holds without change. Moreover, Lemma 4.4 lets us control p_{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*

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

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

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

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

$$p_{\text{solve}} = \mathbf{P}\big(\hat{\mathcal{A}}_r(g) \in S(E;g)\big) = o(1).$$

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

$$p_{\text{solve}}^2 \leq \mathbf{P}(S_{\text{diff}}) \cdot (p_{\text{unstable}} + p_{\text{cond}}). \tag{4.6}$$

To ensure $\mathbf{P}(S_{\text{diff}}) \approx 1$, 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

$$p_{\text{cond}} \leq \exp_2\!\left(-\frac{3E}{4} + O(1)\right) = o(1).$$

Finally, note that in the linear case, when $\eta = O(1)$, $\frac{r^2}{\eta N} = o(1)$ 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 $p_{\text{unstable}} = o(1)$. By (4.6), we conclude that $p_{\text{solve}} = o(1)$, as desired. □

Talk about implications meow.

4.3. No solve case – rounding is truly random.

$\langle g, x \rangle \sim \mathcal{N}(0, N)$

$$\mathbf{P}\left(|\langle g, x \rangle| \leq 2^{-E}\right) \leq \frac{2^{-E+1}}{\sqrt{2\pi N}} = \exp_2\left(-E - \frac{1}{2}\log_2(N) + O(1)\right)$$

Follows by Lemma 1.2. i.e., for $E \gg \log_2 N$, any fixed x is not solution to random instance whp. By conditioning, this implies that if x is random and independent from g, then it's a solution with $o(1)$ probability. Thus, if you truly had a random point, then it's almost certainly not a solution; that is, if your randomized rounding destroys your algorithms output, then whp you fail to find a solution.

Note: we should assume $\log_2^2 N \ll E \leq N$. Also, getting algorithms for polynomial discrepancy (n^-1, etc.) is basically trivial.

Let $x := \mathcal{A}(g)$. We write x^* for the coordinate-wise signs of x, i.e.

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

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.

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.

Given such a randomized rounding function, we can describe it in the following way. Let $p_1, ..., p_N$ be the probabilities of $\mathsf{round}(x)_i \neq (x^*)_i$. We assume without loss of generality that each $p_i \leq \frac{1}{2}$.

Lemma 4.8. *Draw N coin flips $I_i \sim \mathrm{Bern}(2p_i)$ and NNN signs $S_i \sim \mathrm{Unif}\{\pm 1\}$, all mutually independent, and define the random variable $\tilde{x} \in \Sigma_N$ by*

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

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

Proof: Conditioning on I_i, we can check that

$$\begin{aligned} \mathbf{P}(\tilde{x}_i \neq x_i) &= 2p_i \cdot \mathbf{P}(\tilde{x}_i = x_i \mid I_i = 1) + (1 - 2p_i) \cdot \mathbf{P}(\tilde{x}_i \neq x_i \mid I_i = 0) \\ &= 2p_i \cdot \frac{1}{2} + 0 = p_i. \end{aligned}$$

Thus, $\tilde{x}$ has the same probability of equalling x^* in each coordinate as $\operatorname{round}(x)$ does, as claimed. □

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

By Lemma 4.8, it makes sense to define $\tilde{\mathcal{A}}(g) := \operatorname{round}(\mathcal{A}(g))$, 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 $x = \mathcal{A}(g)$).

TODO: explain why we want to consider $\tilde{\mathcal{A}}(g) = \operatorname{round}(\mathcal{A}(g))$

Definition 4.9. Given $\mathcal{A}$, we can define two Σ_N-valued algorithms. Let $x := \mathcal{A}(g)$. Then

$$\mathcal{A}^*(g)_i := 2I(x_i > 0) - 1 \qquad \text{and} \qquad \tilde{\mathcal{A}}(g) =: \operatorname{round}(\mathcal{A}(g)).$$

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.

However, we still want to show that $\tilde{x} := \tilde{\mathcal{A}}(g)$ fails to solve g with high probability. Intuitively, the landscape of solutions is so fractured that any rounding procedure which produces results different from x^* will effectively be selecting a random point, and because any fixed point has such a low probability of being a solution, hardness still follows.

Lemma 4.10. *Suppose $p_1, \ldots, p_N$ are the probabilities of $\tilde{x}$ and x^* differing in each coordinate. Assume $\sum_i p_i = \omega(1)$. Then $\tilde{x} \neq x^*$ with high probability.*

Proof: Observe that as each coordinate is rounded independently, we can compute

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

For $\sum_i p_i = \omega(1)$, we get $\mathbf{P}(\tilde{x} = x^*) \leq e^{-\omega(1)} = o(1)$, as claimed. □

- Flip coin with prob $2p_i$
- If heads, randomize $\tilde{x}$ with probability $\frac{1}{2}$; if tails keep coord.
- Then,

$$\mathbf{P}(\tilde{x}_i = x_i^*) = 2p_i * \frac{1}{2} + (1 - 2p_i) * 0 = p_i.$$

Let K be a large constant, and let $S \subseteq [N]$ denote the coordinates of the first K coordinates to be randomized. Then, we can condition on $x_{[N]\setminus S}$, given which $\tilde{x}$ is a uniformly random point within a K-dimensional subcube of Σ_N. By Theorem 4.2, at most one of these points is in $S(E; g)$, so the probability of $\tilde{x}$ being a solution is at most 2^{-K}.

$$\mathbf{P}\left(|\langle g, \tilde{x} \rangle| \leq 2^{-E} \mid g, x_{[N]\setminus S}\right) \leq \exp_2\left(-E - \frac{1}{2} \log_2 |S| + O(1)\right).$$

First, assume $\neg S_{\text{solve}}$. In that case, $x := \mathcal{A}(g)$ is far from any solution, and randomized rounding fails with high probability. That is, $\mathbf{P}(\tilde{x} \in S(E; g)) = o(1)$

To see this, let x^* be the point on Σ_N closest to x (in principle, this is the vector which is coordinatewise ± 1 depending on whether each coordinate of x is positive or negative).

Let $p_1, \ldots, p_N$ be the probability of $\tilde{x}$ disagreeing with x_* on each coordinate.

- Require that no $p_i = 0$ (i.e. all coordinates have a chance to disagree)
- Then, for $x \in [0, 1)$, exists universal constant C such that $-\log(1 - x) \leq Cx$.
- Probability that $\tilde{x} = x_*$ is

$$\prod (1 - p_i) = \exp_2\left(\sum \log(1 - p_i)\right) \leq \exp_2\left(-C \sum p_i\right).$$

- If we assume that randomized rounding changes solution, then that requires this probability to go to zero, i.e. $\sum p_i = \omega(1)$.

In this case, consider following construction. For each $1 \leq i \leq N$, flip an independent coin H_i which lands heads with probability $2p_i$, and keep all the heads.

- By Second Borel-Cantelli lemma, $E_i = \{H_i \text{ heads}\}$, the E_i are independent, and

$$\sum_{i \geq 0} \mathbf{P}(E_i) = \sum 2p_i = \infty,$$

  so $\mathbf{P}(\limsup E_i) = 1$, i.e., get heads infinitely often.
- That is, number of heads is $\omega(1)$.
- For every coin with a head, round x^* by changing coord with probability $\frac{1}{2}$; if tails keep coord.
- That is, randomized rounding done by choosing random set of $\omega(1)$ coordinates and resampling them as iid Uniform in $\{-1, 1\}$.

Because number of coordinates being changed is $\omega(1)$, can pick large constant K such that whp there are at least K coordinates being changed.

- Only randomize first K heads, condition on the others. Thus, $\tilde{x}$ has K i.i.d., random coordinates.
- $\tilde{x}$ is random point in K-dimensional subcube, but by Proposition 4.1, only one out of the 2^K such points is a good solution.

Thus, probability for rounding to give a good solution is

- Randomized rounding in artificially difficult way. (I.e. this multistage procedure accomplishes the same thing as randomized rounding.)
- Now, randomized rounding is done by choosing a random set of $\omega(1)$ coordinates, and making those iid Uniform in $\{-1, 1\}$.
- Pick a large constant (e.g. 100), and only randomize the first 100 heads, and condition on the others (i.e. choose the others arbitrarily). Note that since $100 \geq \omega(1)$, there are at least 100 heads whp.
- Now rounded point is random point in 100 dimensional subcube, but at most one of them is a good solution by the claim at the top of the page.
- Combining, the probability for rounding to give a good solution is at most $o(1) + 2^{\{-100\}}$. Since 100 is arbitrary, this is $o(1)$ by sending parameters to 0 and/or infinity in the right order.

Let $\tilde{x}$ be the point on Σ_N after randomized rounding.

Moreover, let $\tilde{x}$ be the point

consider the case where

What could go wrong? It could be that all deterministic Σ_N algorithms fail, but an algorithm which is allowed to output a continuous point and then round it (potentially in a randomized way) could succeed. Such an algorithm would have to do more than just deterministically round, because

Let p_{solve} be probability that $\mathcal{A}$ outputs a point x which is k close in L^1 to a vertex and a good solution x^* exists in nbhd of that corner. Because solutions repel, such x^* is unique, so only hope is that x gets rounded to x^* with reasonable probability.

(Weaker than traditional solution case).

Then, either $\tilde{\mathcal{A}}$ finds this good solution with reasonable probability, or

Argument:

- Algorithm $\mathcal{A}$ which is deterministic $\mathbf{R}^N \to \mathbf{R}^N$. Suppose $\tilde{\mathcal{A}} : \mathbf{R}^N \to \Sigma^N$ is $\mathcal{A}$ passed through any nontrivial rounding procedure.
- Say $\mathcal{A}(g) = x$. Let $x^* \in \Sigma_N$ be closest point to x, and $\tilde{x} = \tilde{\mathcal{A}}(g)$ be the rounding of x.
- If $x^* = \tilde{x}$, we're done.
- Else, we know that only one of x^* and $\tilde{g}$ are a good solution, by Theorem 4.2. It's x^* with probability p_{solve}.
  - Here, we're assuming randomized rounding changes at most some $O(1)$ amount of coordinates.
-

Thus, rounding would destroy the solution.

5. Literature Review

5.1. Applications of NPP

[Tsa92]

- Application of NPP to process scheduling

[KAK19]

- NPP for randomized contorl testing

5.2. Algorithms for solving real-world cases

[Joh+89]

- Overview of simulated annealing

[Joh+91]

- Failure of sim annealing for NPP

[SFD96]

- Several order of magnitude improvement over annealing, but with greater commputation time, by modifying differencing heuristic.

[Koj10]

- Using linear programing solver for NPP.

[SBD21]

- Memetic algebraic diffeq for NPP
- Evolutionary algorithm
- Experimental calculation

5.2.1. Quantum algorithms

[Asp+20]

- Quantum hardware for solving NPP.

[Wen+23]

- Experimental solution using quantum computing.

5.3. Algorithms for average time case

[Kor95]

- Initial work on CKK

[Kor98]

- Extend KK to complete algorithm; will get better

[Lue87]

- PDM heuristic fails

[Yak96]

- Showed LDM achieves $n^{\log(n)}$ performance despite being a simple heuristic, for uniform instance.

5.4. Statistical to Computational Gaps

5.5. OGP Examples

5.5.1. Hardness Examples

[Jer92]

- MCMC can’t find cliques; algorithm failure

[ZK16]

- Inference using algorithms - overview of pedagogy using statistical physics framework.

5.6. Low-Degree Heuristic

[KWB19]

- Kunisky, Wein, Banderia - discussion on low degree heuristic for hypothesis testing.

————————-

[AC08]

- Phase transitions for random constraint satisfaction
- S2C gap for random constraint satisfaction

[AR06]

- Random constraint satisfaction

[Add+17]

- Local algorithms for SK.

[Ali+05]

- NPP as unconstrained quadratic binary problem, and efficient metaheuristic algorithm.

[AFG96]

- Randomized differencing heuristic for NPP; computational simulations.

[APZ19]

- OGP for SBPs.

[BPW18]

- Computational gaps in terms of signal-to-noise and S2C for Bayesian inference.

[Ban10]

- Generalized version of NPP with multiple sets

[Bar+16]

- Sum of squares bound

[BFM04]

- REM approach to NPP (Derrida model)

[BGT13]

- S2C for random graphs

[BR13]

- S2C for sparce PCA

[Bis+24]
- Generalization of NPP allowing some numbers to be split up

[BM08]
- Fix constant α in KK algorithm discrepancy

[BCP01]
- Phase transitions for integral NPP

[BB19]
- Strong hardness for sparse PCA

[BBH19]
- S2C in sparce problems via planted clique
- Spiritually similar to conditional landscape obstructions, in that you fix one instance and study how it changes??

[CV13]
- Random polytopes

[Che+19]
- Local algorithms fail for max-cut

[CGJ78]
- Motivation for bin packing application to multiprocessor scheduling

[CL91]
- Book summarizing results of Karmarkar-Karp

[CE15]
- Independent sets in random graphs

[COY19]
- Evolutionary algorithms for NP hard NPP

[DM15]
- Sum of squares bounds

[DKS17]
- Estimation of Gaussian mixtures

[Fel+16]
- Planted clique detection

[FF98]
- Physics perspective for uniform instances

[GK21]
- prove OGP and stable hardness for NPP

[Gam+22]
- Barriers in Symmetric Binary Perceptron

[GK21]

- Average hardness of computing SK partition function

[GJW22]

- GJW22, low degree poly algorithms for Boolean circuits
- Lemma 3.4!

[GZ19]

- Phase transition in high-dim regression with binary coeffs

[GZ19]

- Planted clique: OGP for dense subgraphs

[GS13]

- Original OGP paper with Gamarnik-Sudan

[GJ19]

- OGP and AMP

[Gam21]

- Overview/summary of OGP

[GJS21]

- Principal submatrix recovery

[GS17]

- Local algorithms for NAE-k-SAT

[GZ19]

- Local search for sparse high-dim regression

[GJ79]

- Garey-Johnson book on NP hardness

[GW98]

- Phase transitions for NPP: performance of algorithms depends on their constrainedness.
- i.e. number of their solutions, e.g. if on state space of 2^N states, this parameter is > 1, you're screwed

$$\varkappa := 1 - \frac{\log(\#\text{ of solns})}{N}$$

[GW00]

- Phase transitions in simulated annealing

[Har+23]

- Application of NPP to radomized control testing

[HLS14]

- Local-global study of sparse graphs

[Hob+16]

- Hardness of number balancing (diff from NPP) by reduction to Minkowski/shortest vector.

[Hop+17]

- Signal recovery using sum-of-squares semidefinite programming
- Early suggestion of low degree heuristic

[Hop18]

- Hopkins thesis - intoduced low degree hypothesis

[HSS15]

- Tensor PCA via sum of squares

[HS25]

- SLDH paper

[Kar+86]

- original analysis of hardness

[KK83]

- KK algorithm - time $O(N \log N)$

[Kea98]

- Classification and learning in presence of noise

[Kız23]

- Planted version of NPP, with explicit analysis + hardness results

[Kor09]

- CKK for larger sets

[Kot+17]

- Sum of squares for constraint satisfaction.

[KKS14]

- Heuristics for multidimensional NPP

[LW07]

- Independent sets in regular graphs of girth

[LRR17]

- Discrepancy coloring - poly time algorithm

[LM12]

- Constructive proof of discrepancy minimizing coloring

[MPW15]

- Sum of squares in planted case

[MH78]

- Using NPP for cryptography

[Mer03]

- Phase transition and overview of NPP

[Mer01]

- Physics notation as applied to NPP

- "Any heuristic that exploits a fraction of the domain, generating and evaluating a series of feasible configurations, cannot be significantly better than random search." section 4.3

[MMZ05]

- Random k-SAT/CSP clustering

[Mic+03]

- Worst case performance of KK algorithm when attempting balanced Number Partitioning

[O'D21]

- Textbook on Boolean functions

[RSS19]

- Hidh dimensional estimation for SoS - more SoS stuff

[RV17]

- Failure of local algorithm for independent sets in graphs

[Rot16]

- Partial coloring of sets (discrepancy min)

[TMR20]

- Multidimensional NPP - poly time algorithm achieving $e^{-\Omega(\log^2 \frac{N}{m})}$, for $m = O\left(\sqrt{\log n}\right)$.

[VV25]

- Assuming hardness of shortest vector on lattice, derived polynomial-time hardness for NPP;
- Prove KK is tight; no poly time algorithm achieves energy of $\Omega(\log^3 N)$

[Wei20]

- Low degree polynomial hardness for max independent set.

Bibliography

[AC08] D. Achlioptas and A. Coja-Oghlan, "Algorithmic Barriers from Phase Transitions," in *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, Oct. 2008, pp. 793–802. doi: 10.1109/FOCS.2008.11.

[Add+17] L. Addario-Berry, L. Devroye, G. Lugosi, and R. I. Oliveira, "Local Optima of the Sherrington-Kirkpatrick Hamiltonian." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1712.07775

[AFG96] M. F. Argüello, T. A. Feo, and O. Goldschmidt, "Randomized Methods for the Number Partitioning Problem," *Computers & Operations Research*, vol. 23, no. 2, pp. 103–111, Feb. 1996, doi: 10.1016/0305-0548(95)E0020-L.

[Ali+05] B. Alidaee, F. Glover, G. A. Kochenberger, and C. Rego, "A New Modeling and Solution Approach for the Number Partitioning Problem," *Journal of Applied Mathematics and Decision Sciences*, vol. 2005, no. 2, pp. 113–121, Jan. 2005, doi: 10.1155/JAMDS.2005.113.

[APZ19] B. Aubin, W. Perkins, and L. Zdeborová, "Storage Capacity in Symmetric Binary Perceptrons," *Journal of Physics A: Mathematical and Theoretical*, vol. 52, no. 29, p. 294003, Jul. 2019, doi: 10.1088/1751-8121/ab227a.

[AR06] D. Achlioptas and F. Ricci-Tersenghi, “On the Solution-Space Geometry of Random Constraint Satisfaction Problems.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/cs/0611052

[Asp+20] L. Asproni, D. Caputo, B. Silva, G. Fazzi, and M. Magagnini, “Accuracy and Minor Embedding in Subqubo Decomposition with Fully Connected Large Problems: A Case Study about the Number Partitioning Problem,” *Quantum Machine Intelligence*, vol. 2, no. 1, p. 4, Jun. 2020, doi: 10.1007/s42484-020-00014-w.

[Ban10] N. Bansal, “Constructive Algorithms for Discrepancy Minimization.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1002.2259

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

[BB19] M. Brennan and G. Bresler, “Optimal Average-Case Reductions to Sparse PCA: From Weak Assumptions to Strong Hardness.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1902.07380

[BBH19] M. Brennan, G. Bresler, and W. Huleihel, “Reducibility and Computational Lower Bounds for Problems with Planted Sparse Structure.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1806.07508

[BCP01] C. Borgs, J. Chayes, and B. Pittel, “Phase Transition and Finite-size Scaling for the Integer Partitioning Problem,” *Random Structures & Algorithms*, vol. 19, no. 3–4, pp. 247–288, Oct. 2001, doi: 10.1002/rsa.10004.

[BFM04] H. Bauke, S. Franz, and S. Mertens, “Number Partitioning as a Random Energy Model,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2004, no. 4, p. P4003, Apr. 2004, doi: 10.1088/1742-5468/2004/04/P04003.

[BGT13] M. Bayati, D. Gamarnik, and P. Tetali, “Combinatorial Approach to the Interpolation Method and Scaling Limits in Sparse Random Graphs,” *The Annals of Probability*, vol. 41, no. 6, Nov. 2013, doi: 10.1214/12-AOP816.

[Bis+24] S. Bismuth, V. Makarov, E. Segal-Halevi, and D. Shapira, “Partitioning Problems with Splittings and Interval Targets.” Accessed: Mar. 20, 2025. [Online]. Available: http://arxiv.org/abs/2204.11753

[BM08] S. Boettcher and S. Mertens, “Analysis of the Karmarkar-Karp Differencing Algorithm,” *The European Physical Journal B*, vol. 65, no. 1, pp. 131–140, Sep. 2008, doi: 10.1140/epjb/e2008-00320-9.

[BPW18] A. S. Bandeira, A. Perry, and A. S. Wein, “Notes on Computational-to-Statistical Gaps: Predictions Using Statistical Physics.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1803.11132

[BR13] Q. Berthet and P. Rigollet, “Computational Lower Bounds for Sparse PCA.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1304.0828

[CE15] A. Coja-Oghlan and C. Efthymiou, “On Independent Sets in Random Graphs,” *Random Structures & Algorithms*, vol. 47, no. 3, pp. 436–486, Oct. 2015, doi: 10.1002/rsa.20550.

[CGJ78] E. G. Coffman Jr., M. R. Garey, and D. S. Johnson, “An Application of Bin-Packing to Multiprocessor Scheduling,” *SIAM Journal on Computing*, vol. 7, no. 1, pp. 1–17, Feb. 1978, doi: 10.1137/0207001.

[Che+19] W.-K. Chen, D. Gamarnik, D. Panchenko, and M. Rahman, “Suboptimality of Local Algorithms for a Class of Max-Cut Problems,” *The Annals of Probability*, vol. 47, no. 3, May 2019, doi: 10.1214/18-AOP1291.

[CL91] E. G. Coffman and G. S. Lueker, *Probabilistic Analysis of Packing and Partitioning Algorithms*. in Wiley-Interscience Series in Discrete Mathematics and Optimization. New York: J. Wiley & sons, 1991.

[COY19] D. Corus, P. S. Oliveto, and D. Yazdani, “Artificial Immune Systems Can Find Arbitrarily Good Approximations for the NP-hard Number Partitioning Problem,” *Artificial Intelligence*, vol. 274, pp. 180–196, Sep. 2019, doi: 10.1016/j.artint.2019.03.001.

[CV13] K. Chandrasekaran and S. Vempala, “Integer Feasibility of Random Polytopes.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1111.4649

[DKS17] I. Diakonikolas, D. M. Kane, and A. Stewart, “Statistical Query Lower Bounds for Robust Estimation of High-dimensional Gaussians and Gaussian Mixtures.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1611.03473

[DM15] Y. Deshpande and A. Montanari, “Improved Sum-of-Squares Lower Bounds for Hidden Clique and Hidden Submatrix Problems.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1502.06590

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

[FF98] F. F. Ferreira and J. F. Fontanari, “Probabilistic Analysis of the Number Partitioning Problem,” *Journal of Physics A: Mathematical and General*, vol. 31, no. 15, p. 3417, Apr. 1998, doi: 10.1088/0305-4470/31/15/007.

[Gam+22] D. Gamarnik, E. C. Kızıldağ, W. Perkins, and C. Xu, “Algorithms and Barriers in the Symmetric Binary Perceptron Model.” Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/2203.15667

[Gam21] D. Gamarnik, “The Overlap Gap Property: A Geometric Barrier to Optimizing over Random Structures,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 41, p. e2108492118, Oct. 2021, doi: 10.1073/pnas.2108492118.

[GJ19] D. Gamarnik and A. Jagannath, “The Overlap Gap Property and Approximate Message Passing Algorithms for p-Spin Models.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1911.06943

[GJ79] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. in A Series of Books in the Mathematical Sciences. New York: W. H. Freeman, 1979.

[GJS21] D. Gamarnik, A. Jagannath, and S. Sen, “The Overlap Gap Property in Principal Submatrix Recovery,” *Probability Theory and Related Fields*, vol. 181, no. 4, pp. 757–814, Dec. 2021, doi: 10.1007/s00440-021-01089-7.

[GJW22] D. Gamarnik, A. Jagannath, and A. S. Wein, “Hardness of Random Optimization Problems for Boolean Circuits, Low-Degree Polynomials, and Langevin Dynamics.” Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/2004.12063

[GK21] D. Gamarnik and E. C. Kızıldağ, “Algorithmic Obstructions in the Random Number Partitioning Problem.” Accessed: Mar. 15, 2025a. [Online]. Available: http://arxiv.org/abs/2103.01369

[GK21] D. Gamarnik and E. Kizildag, “Computing the Partition Function of the Sherrington-Kirkpatrick Model Is Hard on Average,” *The Annals of Applied Probability*, vol. 31, no. 3, Jun. 2021b, doi: 10.1214/20-AAP1625.

[GS13] D. Gamarnik and M. Sudan, “Limits of Local Algorithms over Sparse Random Graphs.” Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/1304.1831

[GS17] D. Gamarnik and M. Sudan, “Performance of Sequential Local Algorithms for the Random NAE-K-SAT Problem,” *SIAM Journal on Computing*, vol. 46, no. 2, pp. 590–619, Jan. 2017, doi: 10.1137/140989728.

[GW00] I. Gent and T. Walsh, “Phase Transitions and Annealed Theories: Number Partitioning as a Case Study,” *Instituto Cultura*, Jun. 2000.

[GW98] I. P. Gent and T. Walsh, “Analysis of Heuristics for Number Partitioning,” *Computational Intelligence*, vol. 14, no. 3, pp. 430–451, 1998, doi: 10.1111/0824-7935.00069.

[GZ19] D. Gamarnik and I. Zadik, “High-Dimensional Regression with Binary Coefficients. Estimating Squared Error and a Phase Transition.” Accessed: Mar. 16, 2025a. [Online]. Available: http://arxiv.org/abs/1701.04455

[GZ19] D. Gamarnik and I. Zadik, “The Landscape of the Planted Clique Problem: Dense Subgraphs and the Overlap Gap Property.” Accessed: Mar. 16, 2025b. [Online]. Available: http://arxiv.org/abs/1904.07174

[GZ19] D. Gamarnik and I. Zadik, “Sparse High-Dimensional Linear Regression. Algorithmic Barriers and a Local Search Algorithm.” Accessed: Mar. 16, 2025c. [Online]. Available: http://arxiv.org/abs/1711.04952

[Har+23] C. Harshaw, F. Sävje, D. Spielman, and P. Zhang, “Balancing Covariates in Randomized Experiments with the Gram-Schmidt Walk Design.” Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/1911.03071

[HLS14] H. Hatami, L. Lovász, and B. Szegedy, “Limits of Locally–Globally Convergent Graph Sequences,” *Geometric and Functional Analysis*, vol. 24, no. 1, pp. 269–296, Feb. 2014, doi: 10.1007/s00039-014-0258-7.

[Hob+16] R. Hoberg, H. Ramadas, T. Rothvoss, and X. Yang, “Number Balancing Is as Hard as Minkowski's Theorem and Shortest Vector.” Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/1611.08757

[Hop+17] S. B. Hopkins, P. K. Kothari, A. Potechin, P. Raghavendra, T. Schramm, and D. Steurer, “The Power of Sum-of-Squares for Detecting Hidden Structures.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1710.05017

[Hop18] S. Hopkins, “Statistical Inference and the Sum of Squares Method,” 2018.

[HS25] B. Huang and M. Sellke, “Strong Low Degree Hardness for Stable Local Optima in Spin Glasses.” Accessed: Jan. 30, 2025. [Online]. Available: http://arxiv.org/abs/2501.06427

[HSS15] S. B. Hopkins, J. Shi, and D. Steurer, “Tensor Principal Component Analysis via Sum-of-Squares Proofs.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1507.03269

[Jer92] M. Jerrum, “Large Cliques Elude the Metropolis Process,” *Random Structures & Algorithms*, vol. 3, no. 4, pp. 347–359, Jan. 1992, doi: 10.1002/rsa.3240030402.

[Joh+89] D. S. Johnson, C. R. Aragon, L. A. McGeoch, and C. Schevon, “Optimization by Simulated Annealing: An Experimental Evaluation; Part I, Graph Partitioning,” *Operations Research*, vol. 37, no. 6, pp. 865–892, 1989, Accessed: Mar. 15, 2025. [Online]. Available: http://www.jstor.org/stable/171470

[Joh+91] D. S. Johnson, C. R. Aragon, L. A. McGeoch, and C. Schevon, “Optimization by Simulated Annealing: An Experimental Evaluation; Part II, Graph Coloring and Number Partitioning,” *Operations Research*, vol. 39, no. 3, pp. 378–406, 1991, Accessed: Mar. 15, 2025. [Online]. Available: http://www.jstor.org/stable/171393

[KAK19] A. M. Krieger, D. Azriel, and A. Kapelner, “Nearly Random Designs with Greatly Improved Balance,” *Biometrika*, vol. 106, no. 3, pp. 695–701, Sep. 2019, doi: 10.1093/biomet/asz026.

[Kar+86] N. Karmarkar, R. M. Karp, G. S. Lueker, and A. M. Odlyzko, “Probabilistic Analysis of Optimum Partitioning,” *Journal of Applied Probability*, vol. 23, no. 3, pp. 626–645, 1986, doi: 10.2307/3214002.

[Kea98] M. Kearns, “Efficient Noise-Tolerant Learning from Statistical Queries,” *Journal of the ACM*, vol. 45, no. 6, pp. 983–1006, Nov. 1998, doi: 10.1145/293347.293351.

[KK83] N. Karmarker and R. M. Karp, “The Differencing Method of Set Partitioning,” 1983. Accessed: Mar. 15, 2025. [Online]. Available: https://www2.eecs.berkeley.edu/Pubs/TechRpts/1983/6353.html

[KKS14] J. Kratica, J. Kojić, and A. Savić, “Two Metaheuristic Approaches for Solving Multidimensional Two-Way Number Partitioning Problem,” *Computers & Operations Research*, vol. 46, pp. 59–68, Jun. 2014, doi: 10.1016/j.cor.2014.01.003.

[Koj10] J. Kojić, “Integer Linear Programming Model for Multidimensional Two-Way Number Partitioning Problem,” *Computers & Mathematics with Applications*, vol. 60, no. 8, pp. 2302–2308, Oct. 2010, doi: 10.1016/j.camwa.2010.08.024.

[Kor09] R. E. Korf, “Multi-Way Number Partitioning,” in *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, in IJCAI'09. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Jul. 2009, pp. 538–543.

[Kor95] R. E. Korf, “From Approximate to Optimal Solutions: A Case Study of Number Partitioning,” in *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1*, in IJCAI'95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Aug. 1995, pp. 266–272.

[Kor98] R. E. Korf, “A Complete Anytime Algorithm for Number Partitioning,” *Artificial Intelligence*, vol. 106, no. 2, pp. 181–203, Dec. 1998, doi: 10.1016/S0004-3702(98)00086-1.

[Kot+17] P. K. Kothari, R. Mori, R. O'Donnell, and D. Witmer, “Sum of Squares Lower Bounds for Refuting Any CSP.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1701.04521

[Kun24] D. Kunisky, “Low Coordinate Degree Algorithms I: Universality of Computational Thresholds for Hypothesis Testing.” Accessed: Mar. 26, 2025. [Online]. Available: http://arxiv.org/abs/2403.07862

[KWB19] D. Kunisky, A. S. Wein, and A. S. Bandeira, “Notes on Computational Hardness of Hypothesis Testing: Predictions Using the Low-Degree Likelihood Ratio.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1907.11636

[Kız23] E. C. Kızıldağ, “Planted Random Number Partitioning Problem.” Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/2309.15115

[LM12] S. Lovett and R. Meka, “Constructive Discrepancy Minimization by Walking on The Edges.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1203.5747

[LRR17] A. Levy, H. Ramadas, and T. Rothvoss, “Deterministic Discrepancy Minimization via the Multiplicative Weight Update Method.” Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1611.08752

[Lue87] G. S. Lueker, “A Note on the Average-Case Behavior of a Simple Differencing Method for Partitioning,” *Operations Research Letters*, vol. 6, no. 6, pp. 285–287, Dec. 1987, doi: 10.1016/0167-6377(87)90044-7.

[LW07] J. Lauer and N. Wormald, “Large Independent Sets in Regular Graphs of Large Girth,” *Journal of Combinatorial Theory, Series B*, vol. 97, no. 6, pp. 999–1009, Nov. 2007, doi: 10.1016/j.jctb.2007.02.006.

[Mer01] S. Mertens, "A Physicist's Approach to Number Partitioning," *Theoretical Computer Science*, vol. 265, no. 1, pp. 79–108, Aug. 2001, doi: 10.1016/S0304-3975(01)00153-0.

[Mer03] S. Mertens, "The Easiest Hard Problem: Number Partitioning." Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/cond-mat/0310317

[MH78] R. Merkle and M. Hellman, "Hiding Information and Signatures in Trapdoor Knapsacks," *IEEE Transactions on Information Theory*, vol. 24, no. 5, pp. 525–530, Sep. 1978, doi: 10.1109/TIT.1978.1055927.

[Mic+03] W. Michiels, J. Korst, E. Aarts, and J. Van Leeuwen, "Performance Ratios for the Differencing Method Applied to the Balanced Number Partitioning Problem," *STACS 2003*, vol. 2607. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 583–595, 2003. doi: 10.1007/3-540-36494-3_51.

[MMZ05] M. Mézard, T. Mora, and R. Zecchina, "Clustering of Solutions in the Random Satisfiability Problem," *Physical Review Letters*, vol. 94, no. 19, p. 197205, May 2005, doi: 10.1103/PhysRevLett.94.197205.

[MPW15] R. Meka, A. Potechin, and A. Wigderson, "Sum-of-Squares Lower Bounds for Planted Clique," in *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, Portland Oregon USA: ACM, Jun. 2015, pp. 87–96. doi: 10.1145/2746539.2746600.

[O'D21] R. O'Donnell, "Analysis of Boolean Functions." Accessed: Mar. 15, 2025. [Online]. Available: http://arxiv.org/abs/2105.10386

[Rot16] T. Rothvoss, "Constructive Discrepancy Minimization for Convex Sets." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1404.0339

[RSS19] P. Raghavendra, T. Schramm, and D. Steurer, "High-Dimensional Estimation via Sum-of-Squares Proofs." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1807.11419

[RV17] M. Rahman and B. Virag, "Local Algorithms for Independent Sets Are Half-Optimal," *The Annals of Probability*, vol. 45, no. 3, May 2017, doi: 10.1214/16-AOP1094.

[SBD21] V. Santucci, M. Baioletti, and G. Di Bari, "An Improved Memetic Algebraic Differential Evolution for Solving the Multidimensional Two-Way Number Partitioning Problem," *Expert Systems with Applications*, vol. 178, p. 114938, Sep. 2021, doi: 10.1016/j.eswa.2021.114938.

[SFD96] R. H. Storer, S. W. Flanders, and S. David Wu, "Problem Space Local Search for Number Partitioning," *Annals of Operations Research*, vol. 63, no. 4, pp. 463–487, Aug. 1996, doi: 10.1007/BF02156630.

[TMR20] P. Turner, R. Meka, and P. Rigollet, "Balancing Gaussian Vectors in High Dimension." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/1910.13972

[Tsa92] L.-H. Tsai, "Asymptotic Analysis of an Algorithm for Balanced Parallel Processor Scheduling," *SIAM Journal on Computing*, vol. 21, no. 1, pp. 59–64, Feb. 1992, doi: 10.1137/0221007.

[Ver18] R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science*, 1st ed. in Cambridge Series in Statistical and Probabilistic Mathematics. New York, NY: Cambridge University Press, 2018.

[VV25] N. Vafa and V. Vaikuntanathan, "Symmetric Perceptrons, Number Partitioning and Lattices." Accessed: Mar. 20, 2025. [Online]. Available: http://arxiv.org/abs/2501.16517

[Wei20] A. S. Wein, "Optimal Low-Degree Hardness of Maximum Independent Set." Accessed: Mar. 16, 2025. [Online]. Available: http://arxiv.org/abs/2010.06563

[Wen+23] J. Wen *et al.*, "Optical Experimental Solution for the Multiway Number Partitioning Problem and Its Application to Computing Power Scheduling," *Science China Physics, Mechanics & Astronomy*, vol. 66, no. 9, p. 290313, Sep. 2023, doi: 10.1007/s11433-023-2147-3.

[Yak96] B. Yakir, "The Differencing Algorithm LDM for Partitioning: A Proof of a Conjecture of Karmarkar and Karp," *Mathematics of Operations Research*, vol. 21, no. 1, pp. 85–99, Feb. 1996, doi: 10.1287/moor.21.1.85.

[ZK16] L. Zdeborová and F. Krzakala, "Statistical Physics of Inference: Thresholds and Algorithms," *Advances in Physics*, vol. 65, no. 5, pp. 453–552, Sep. 2016, doi: 10.1080/00018732.2016.1211393.