1. Introduction

Overview of number partitioning problem.

Application: randomized control trials.

Other applications.

- Circuit design, etc.

Importance as a basic NP-complete problem.

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

1.3. Number Packing Problem

Let N be the dimensionality, and $\Sigma_N := \{\pm 1\}$ be the binary cube. Suppose we're given $g \sim \mathcal{N}(0, I_N)$. We want to find $x \in \Sigma_N$ such that we minimize $|\langle x, g\rangle|$.

Definition 1.1. Let $\delta > 0$. The *solution set* $S(\delta; g)$ is the set of all $x \in \Sigma_N$ that are δ-*optimal*, satisfying

$$\frac{1}{\sqrt{N}}|\langle g, x\rangle| \leq 2^{-\delta N}. \tag{1.1}$$

(1.1) Hi

1.4. Existing Results

1. $X_i, 1 \leq i \leq n$ i.i.d. uniform from $\{1, 2, ..., M := 2^m\}$, with $\kappa := \frac{m}{n}$, then phase transition going from $\kappa < 1$ to $\kappa > 1$.
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 $2^{-\Theta(n)} \leq 2^{-o(n)}$)

1.5. Glossary and conventions

Conventions:

1. log means log in base 2, exp is exp base 2 - natural log/exponent is $\ln / e^x$.
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 $x, y \in \mathbf{R}^N$,

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

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"

1.6. Literature Review

[AC08]

- S2C gap for random constraint satisfaction

[AR06]

-

[Add+17]

-

[Ali+05]

-

[AFG96]

-

[Asp+20]

-

[APZ19]

-

[BPW18]

-

[Ban10]

-

[Bar+16]

-
[BFM04]
-
[BGT13]
-
[BR13]
-
[BM08]
-
[BCP01]
-
[BB19]
-
[BBH19]
-
[CV13]
-
[Che+19]
-
[CGJ78]
-
[CL91]
-
[CE15]
-
[COY19]
-
[DM15]
-
[DKS17]
-
[Fel+16]
-
[FF98]
-
[GK21]
-
[Gam+22]
-

[GK21]

•

[GJW22]

•

[GZ19]

•

[GZ19]

•

[GS13]

•

[GJ19]

•

[Gam21]

•

[GJS21]

•

[GS17]

•

[GZ19]

•

[GJ79]

•

[GW98]

•

[GW00]

•

[Har+23]

•

[HTF09]

•

[HLS14]

•

[Hob+16]

•

[Hop+17]

•

[Hop18]

•

[HSS15]

- [HS25]
- [Jer92]
- [Joh+89]
- [Joh+91]
- [Kar+86]
- [KK83]
- [Kea98]
- [Kız23]
- [Koj10]
- [Kor95]
- [Kor98]
- [Kor09]
- [Kot+17]
- [KKS14]
- [KAK19]
- [KWB19]
- [LW07]
- [LRR17]
- [LM12]
-

[Lue87]

•

[MPW15]

•

[MH78]

•

[Mer03]

•

[Mer01]

•

[MMZ05]

•

[Mic+03]

•

[O'D21]

•

[RSS19]

•

[RV17]

•

[Rot16]

•

[SBD21]

•

[SFD96]

•

[Tsa92]

•

[TMR20]

•

[Wai19]

•

[Wei20]

•

[Wen+23]

•

[Yak96]

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

[ZK16]

•

1.7. Our Results

2. Low-Degree Algorithms

For our purposes, an *algorithm* is a function which takes as input a problem instance $g \sim \mathcal{N}(0, I_N)$ 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), or to taking as additional input some randomness ω, allowing for randomized algorithms. 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, which we term Efron-Stein degree, is a more flexible notion which can be applied to a much broader class of algorithms. As we will see in Section 3.3, these classes of algorithms exhibit quantitatively different behavior, in line with existing heuristics for the "brittleness" of NPP.

2.1. Efron-Stein Degree and L^2 Stability

First, we consider a very general class of putative algorithms, where the notion of "degree" corresponds to how complex the interactions between the input variables can get. 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 $L^2\left(\mathbf{R}^N, \pi^{\otimes N}\right)$ is the space of functions $f : \mathbf{R}^N \to \mathbf{R}$ with finite L^2 norm.

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

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

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 we want to know about f should be reflected in how it acts on all possible such subsets. To formalize this intuition, we define the following coordinate projection:

Definition 2.3. Let $f \in L^2(\mathbf{R}^N, \pi^{\otimes N})$ and $J \subseteq [N]$, with $\overline{J} = [N] \setminus J$. The *projection of f onto J* is the function $f^{\subseteq J} : \mathbf{R}^N \to \mathbf{R}$ given by

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

This is f with the $\overline{J}$ coordinates re-randomized, so $f^{\subseteq J}$ only depends on x_J.

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

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

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

This decomposition, often called the *Efron-Stein decomposition*, does indeed exist, and exhibits the following combinatorial construction. Our presentation largely follows [O'D21, § 8.3] (who refers to this as the *orthogonal decomposition*).

The motivating fact is that we should expect 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 $f^{\subseteq \{j\}}(x) = \mathbf{E}[f \mid x_j]$, we get

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

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

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

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

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

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

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

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

In addition, this decomposition has the following properties:

3. *Condition 2. holds whenever $S \nsubseteq T$.*
4. *The decomposition is orthogonal: $\langle f^{=S}, f^{=T} \rangle = 0$ for $S \neq T$.*
5. *$\sum_{S \subseteq T} f^{=S} = f^{\subseteq T}$.*
6. *For each $S \subseteq [N]$, $f \mapsto f^{=S}$ is a linear operator.*

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

We can finally define the Efron-Stein notion of "degree":

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

$$\deg_{\mathrm{ES}}(f) = \max_{S \subseteq [N] \text{ s.t. } f^{=S} \neq 0} |S|.$$

If $f = (f_1, ..., f_M) : \mathbf{R}^N \to \mathbf{R}^M$ is a multivariate function, then the Efron-Stein degree of f is the maximum degree of the f_i.

Intuitively, the Efron-Stein degree is the maximum size of 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 Efron-Stein degree at most D.

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 Efron-Stein 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* 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) is 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.

Definition 2.7. For $p \in [0, 1]$, the *noise operator* is the linear operator T_p on $L^2(\mathbf{R}^N, \pi^{\otimes N})$, defined by, for y p-resampled from x

$$T_p f(x) = \mathbf{E}_{y\ p\text{-resampled from } x}[f(y)]$$

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

As claimed, we can compute how this operator changes the Efron-Stein decomposition:

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

Putting these facts together, we can derive the following stability bound on functions of bounded Efron-Stein degree.

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 Efron-Stein degree D and each $f_i \in L^2(\mathbf{R}^N, \pi^{\otimes N})$. Suppose that (x, y) are a p-resampled pair under $\pi^{\otimes N}$, and $\mathbf{E}\|f(x)\|^2 = 1$. Then*

$$\mathbf{E}\|f(x) - f(y)\|^2 \leq 2(1 - p^D) \leq 2(1 - p)D. \tag{2.3}$$

Proof: Observe that

$$\begin{aligned} \mathbf{E}\|f(x) - f(y)\|^2 &= \mathbf{E}\|f(x)\|^2 + \mathbf{E}\|f(y)\|^2 - 2\mathbf{E}\langle f(x), f(y)\rangle \\ &= 2 - 2\left(\sum_i \mathbf{E}[f_i(x) f_i(y)]\right) \\ &= 2 - 2\left(\sum_i \langle f_i, T_p f_i\rangle\right). \end{aligned} \tag{2.4}$$

Here, we have for each $i \in [N]$ 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 Efron-Stein 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 (2.4) to get

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

as desired. □

2.2. Hermite Polynomials

Alternatively, we can consider the much more restrictive (but more concrete) class of honest polynomials. When considered as functions of independent Normal variables, such functions admit a simple description in terms of *Hermite polynomials*, which enables us to prove similar bounds as Theorem 2.9. This theory is much more classical, so we encourage the interested reader to see [O'D21, § 11] for details.

To start, we consider the following space of L^2 functions:

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 $L^2\left(\mathbf{R}^N, \gamma^N\right)$ of L^2 functions of N i.i.d. standard Normal random variables.

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 $\mathbf{E}\left[z^2\right] = 1$ 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

$$h_0(z) = 1, \qquad h_1(z) = z, \qquad h_2(z) = \frac{z^2-1}{\sqrt{2}}, \qquad h_3(z) = \frac{z^3 - 3z}{\sqrt{6}}, \qquad \ldots \tag{2.5}$$

Note here that each h_j is a degree j polynomial.

It is then straightforward to show the following:

Theorem 2.11 ([O'D21, Prop 11.30])**.** *The Hermite polynomials $\left(h_j\right)_{j\geq 0}$ form a complete orthonormal basis for $L^2(\mathbf{R}, \gamma)$.*

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

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*

$$\mathbf{E}\|f(x) - f(y)\|^2 \le 2\left(1 - p^D\right) \le 2(1-p)D. \tag{2.6}$$

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 $p^D \le p^s \le 1$ for all $s \le D$, we can apply Lemma 2.15 to get

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

From there, the proof proceeds as before. □

As a comparision to the case for functions with Efron-Stein degree D, notice that Theorem 2.16 gives, generically, a much looser bound. For instance, the function $f(x) = x_1^2 x_2^4$ has Efron-Stein degree 2, but polynomial degree 6. In exchange, being able to use p-correlation as a "metric" on the input domain will turn out to offer significant benefits in the arguments which follow, justifying equal consideration of both classes of functions.

2.3. Stability of Low-Degree Algorithms

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 ???).

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 Efron-Stein degree:

Definition 2.19. Suppose an algorithm $\mathcal{A}(g, \omega)$ is such that each coordinate of $\mathcal{A}(-, \omega)$ is in $L^2\left(\mathbf{R}^N, \pi^{\otimes N}\right)$. Then, the *Efron-Stein degree* of $\mathcal{A}$ is the maximum Efron-Stein 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 (or Efron-Stein degree) $\le D$ and norm $\mathbf{E}\|\mathcal{A}(g)\|^2 \le CN$. Then, for inputs g, g' which are $(1-\varepsilon)$-correlated,*

$$\mathbf{E}\|\mathcal{A}(g) - \mathcal{A}(g')\|^2 \leq 2CD\varepsilon N, \tag{2.7}$$

and thus

$$\mathbf{P}\Big(\|\mathcal{A}(g) - \mathcal{A}(g')\| \geq 2\sqrt{\eta N}\Big) \leq \frac{CD\varepsilon}{2\eta} \asymp \frac{D\varepsilon}{\eta} \tag{2.8}$$

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 Efron-Stein 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', and using that $C' \leq CN$, we get (2.7). Finally, (2.8) follows immediately 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 random, $\mathbf{R}^N$-valued algorithms later on in (section ???). As outlined in Section 1.7,

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.

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.

3.1. Proof Outline and Preliminary Estimates

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:

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

Intuitively, the first two events ask that the algorithm solves both instances and is stable, respectively. The last event 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* $S_{\text{solve}} \cap S_{\text{stable}} \cap S_{\text{cond}} = \emptyset$.

Proof: Suppose that S_{solve} and S_{stable} both occur. Letting $x := \mathcal{A}(g)$ (which only depends on g) and $x' := \mathcal{A}(g')$, we have that $x' \in S(E; g')$ while also being within distance $2\sqrt{\eta N}$ of x. This contradicts S_{cond}, thus completing the proof. □

Define p_{solve} as the probability that the algorithm solves a single instance:

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

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

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

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

Moreover, let us define p_{unstable} and p_{cond} by

$$\mathbf{P}(S_{\text{stable}}) = 1 - p_{\text{unstable}}$$

and

$$\mathbf{P}(S_{\text{cond}}) = 1 - p_{\text{cond}}.$$

By Lemma 3.1, we know that

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

and rearranging, we get that

$$p_{\text{solve}}^2 \leq p_{\text{unstable}} + p_{\text{cond}} \tag{3.2}$$

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

Next, let $\mathcal{A}$ be given by a function with Efron-Stein degree D. We now want g, g' to be $(1-\varepsilon)$-resampled standard Normals. We define the following events.

$$\begin{aligned} S_{\text{diff}} &= \{g \neq g'\} \\ S_{\text{solve}} &= \{\mathcal{A}(g) \in S(E; g), \mathcal{A}(g') \in S(E; g')\} \\ S_{\text{stable}} &= \left\{\|\mathcal{A}(g) - \mathcal{A}(g')\| \leq 2\sqrt{\eta N}\right\} \\ S_{\text{cond}} &= \left\{\begin{matrix} \nexists\, x' \in S(E; g') \text{ such that} \\ \|x - x'\| \leq 2\sqrt{\eta N} \end{matrix}\right\} \end{aligned} \tag{3.3}$$

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

Lemma 3.3. *We have $S_{\text{diff}} \cap S_{\text{solve}} \cap S_{\text{stable}} \cap S_{\text{cond}} = \varnothing$.*

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

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

Lemma 3.4. *For g, g' being $(1-\varepsilon)$-resampled, $\mathbf{P}(S_{\text{diff}}) = 1 - (1-\varepsilon)^N \leq \varepsilon N$.*

Proof: Follows from calculation:

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

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

Lemma 3.5. *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: First, the statement is trivial if $g = g'$, as $p_{\text{solve}} \leq 1$, so assume that $g \neq g'$. 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_{cond} by

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

and

$$\mathbf{P}(S_{\text{cond}} | S_{\text{diff}}) = 1 - p_{\text{cond}}.$$

This is necessary as $p_{\text{unstable}}, p_{\text{cond}} = 1$ given $g = g'$. Note however that for $\mathbf{P}(S_{\text{diff}}) = 1$, as is the case for g, g' being $(1 - \varepsilon)$-correlated, these definitions agree with what we had in (3.2).

Now, by Lemma 3.3, we know that $\mathbf{P}(S_{\text{solve}}, S_{\text{stable}}, S_{\text{cond}} | S_{\text{diff}}) = 0$, so

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

Thus, rearranging and multiplying by $\mathbf{P}(S_{\text{diff}})$ (so as to apply Lemma 3.5) gives

$$p_{\text{solve}}^2 \leq \mathbf{P}(S_{\text{diff}}) \cdot (p_{\text{unstable}} + p_{\text{cond}}) \tag{3.4}$$

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.

3.2. 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 $\vert\langle g, x\rangle\vert \leq 2^{-E}$	Small	Smaller E implies weaker solutions, and can consider full range of $1 \ll E \ll N$. Know that $E > (\log^2 N)$ by [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 $\Vert\mathcal{A}(g) - \mathcal{A}(g')\Vert \leq 2\sqrt{\eta N}$, 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

3.3. Conditional Landscape Obstruction

Our goal is to show that in both cases, whether we consider g' correlated to or resampled from g,

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

(Of course, the condition $g \neq g'$ is vacuously true for (g, g') $(1-\varepsilon)$-correlated.

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

Lemma 3.6. *Suppose that $K \leq N/2$, and let $h(x) = -x\log(x) - (1-x)\log(x)$ be the binary entropy function. Then, for $p := K/N$,*

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

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

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

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

$$\sum_{k\leq K}\binom{N}{k} \leq p^{-K}(1-p)^{-(N-K)}$$
$$= \exp(-K\log(p) - (N-K)\log(1-p))$$
$$= \exp\left(N\cdot\left(-\frac{K}{N}\log(p) - \left(\frac{N-K}{N}\right)\log(1-p)\right)\right)$$
$$= \exp(N\cdot(-p\log(p) - (1-p)\log(1-p))) = \exp(Nh(p)).$$

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

Proposition 3.7 (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*

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

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 3.6, we get

$$\sum_{k\leq N\eta}\binom{N}{k} \leq \exp(Nh(\eta)) \leq \exp(2\eta\log(1/\eta)N).$$ □

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:

Lemma 3.8. *Suppose (g, g') are $(1-\varepsilon)$-correlated standard Normal vectors, and let $x \in \Sigma_N$. Then*

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

Proof: Let $\tilde{g}$ be an independent Normal vector to g, and observe that g' can be represented as $g' = pg + \sqrt{1-p^2}\tilde{g}$, for $p = 1-\varepsilon$. Thus, $\langle g', x\rangle = p\langle g, x\rangle + \sqrt{1-p^2}\langle \tilde{g}, x\rangle$. Observe $\langle g, x\rangle$ is constant given g, and $\langle \tilde{g}, x\rangle$ is a Normal r.v. with mean 0 and variance N, so $\langle g', x\rangle \sim \mathcal{N}\left(p\langle g, x\rangle, \left(1-p^2\right)N\right)$. This random variable is nondegenerate for $\left(1-p^2\right)N > 0$, with density bounded above as

$$\varphi_g(z) = \frac{1}{\sqrt{2\pi(1-p^2)N}} e^{-\frac{(z-p\langle g,x\rangle)^2}{2(1-p^2)N}} \leq \frac{1}{\sqrt{2\pi(1-p^2)N}}$$
$$\leq \frac{1}{\sqrt{2\pi\varepsilon N}} = \exp\left(-\frac{1}{2}\log(\varepsilon) + O(\log N)\right)$$

Integrating this bound over the interval $|z| \leq 2^{-E}$, we conclude that

$$\mathbf{P}\left(|\langle g', x\rangle| \leq 2^{-E} \mid g\right) = \int_{|z|\leq -2^{-E}} \varphi_{g,|S|}(z)\,\mathrm{d}z \leq \exp\left(-E - \frac{1}{2}\log(\varepsilon) + O(\log N)\right).$$ □

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

Next, 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.

Lemma 3.9. *Suppose* (g, g') *are* $(1-\varepsilon)$*-resampled standard Normal vectors, and let* $x \in \Sigma_N$*. Then,*

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

Proof: Let $S = \{i \in [N] : g_i \ne g_i'\}$ be the set of indices where g and g′ differ. We can express

$$\langle g', x\rangle = \sum_{i\in[N]} g_i' x_i = \sum_{i\notin S} g_i x_i + \sum_{i\in S} g_{i'} x_i \sim \mathcal{N}\left(\sum_{i\notin S} g_i x_i, |S|\right).$$

The conditional distribution of interest can now be expressed as $\mathbf{P}\big(|\langle g', x\rangle| \le 2^{-E} \mid g, |S| \ge 1\big)$. Given $|S| \ge 1$, the quantity $\langle g', x'\rangle$ is a nondegenerate random variable, with density bounded above as

$$\varphi_{g,|S|}(z) = \frac{1}{\sqrt{2\pi|S|}} e^{-\frac{(z - \sum_{i\notin S} g_i x_i)^2}{2|S|}} \le \frac{1}{\sqrt{2\pi|S|}} \le \frac{1}{\sqrt{2\pi}}.$$

Hence, the quantity of interest can be bounded as

$$\mathbf{P}\big(|\langle g', x\rangle| \le 2^{-E} \mid g, g \ne g'\big) \le \int_{|z|\le -2^{-E}} \varphi_{g,|S|}(z)\,\mathrm{d}z \le \sqrt{\frac{2}{\pi}} \exp(-E) = \exp(-E + O(1)).$$ □

Note that in contrast to Lemma 3.9, this bound doesn't involve ε at all, but the condition $g \ne g'$ requires $\varepsilon = \omega(1/N)$ to hold w.p. 1.

Putting together these bounds, we conclude the following fundamental estimates of p_{cond}, i.e. of the failure of our conditional landscape obstruction.

Proposition 3.10 (Fundamental Estimate – Correlated Case)**.** *Assume that* (g, g') *are* $(1-\varepsilon)$*-correlated standard Normal vectors. Then, for any* x *only depending on* g*,*

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

Proof: Observe that

$$p_{\text{cond}} = \mathbf{E}\left[\mathbf{P}\left(\left\{\begin{matrix} \exists\, x' \text{ s.t.} \\ \text{(I)}\quad |\langle g', x'\rangle| \le \exp(-E) \\ \text{(II)}\ \|x - x'\| \le 2\sqrt{\eta N} \end{matrix}\right\} \middle|\, g \right)\right].$$

Then, for fixed x, we know there are $\exp(2\eta\log(1/\eta)N)$-many x' satisfying condition (II), with each having an exponentially small probability of satisfying condition (I). Thus, we conclude by union bounding Lemma 3.8 (which is independent of g) over Proposition 3.7. □

By the same proof, using Lemma 3.9 instead of Lemma 3.8, we show:

Proposition 3.11 (Fundamental Estimate – Resampled Case)**.** *Assume that* (g, g') *are* $(1-\varepsilon)$*-resampled standard Normal vectors. Then, for any* x *only depending on* g,

$$p_{\text{cond}} = \mathbf{P}\left(\left\{ \begin{array}{l} \exists\, x' \in S(E; g') \text{ such that} \\ \|x - x'\| \le 2\sqrt{\eta N} \end{array} \right\} \middle|\, g \ne g' \right) \le \exp\!\left(-E + 2\eta \log\!\left(\frac{1}{\eta}\right) N + O(1) \right).$$

3.4. Hardness in the Linear Energy Regime

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 3.12. *Let* $\delta > 0$ *and* $E = \delta N$*, and let* g, g' *be* $(1-\varepsilon)$*-correlated standard Normal r.v.s. Then, for any degree* $D \le o(\exp(\delta N/2))$ *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 (3.2) that it suffices to show that both p_{cond} and p_{unstable} go to zero. Further, by Proposition 3.10, we have

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

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

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

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

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

By (3.2), 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.

Theorem 3.13. *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 Efron-Stein degree* $D \le o(N)$ *(and 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, while $\mathbf{P}(S_{\text{diff}}) \approx 1$. By Lemma 3.4, the latter condition is satisfied for $\varepsilon = \omega(1/N)$. Thus, pick $\varepsilon = \frac{\log(N/D)}{N}$: note that this satisfies $N\varepsilon = \log(N/D) \gg 1$, for $D = o(N)$. Next, choose η such that $2\eta\log(1/\eta) < \delta/4$ – again, this results in η being independent of N. By Proposition 3.11 we get

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

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

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

as $x \log(1/x) \to 0$ for $x \ll 1$. By (3.4), we conclude that $p_{\text{solve}}^2 \leq \mathbf{P}(S_{\text{diff}}) \cdot (p_{\text{unstable}} + p_{\text{cond}}) = o(1)$, thus completing the proof. □

3.5. Hardness in the Sublinear Regime

In this section, we let $\omega\left((\log N)^2\right) \leq E \leq o(N)$.

Theorem 3.14. *Let $\omega\left(\log^2 N\right) \leq E \leq o(N)$, and let g, g' be $(1-\varepsilon)$-correlated standard Normal r.v.s. Then, for any polynomial algorithm $\mathcal{A}$ with degree $D \leq o(\exp(E/4))$ (and with $\mathbf{E}\|\mathcal{A}(g)\|^2 \leq CN$), there exist $\varepsilon, \eta > 0$ such that $p_{\text{solve}} = o(1)$.*

Proof: As in Theorem 3.12, it suffices to show that both p_{cond} and p_{unstable} go to zero. To do this, we choose

$$\varepsilon = \exp\left(-\frac{E}{2}\right), \qquad \eta = \frac{E}{16N\log(N/E)}.$$

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

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

Thus, by Proposition 3.10, we get

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

where the last equality follows as $E \gg \log N$. Then, by Proposition 2.20, the choice of $D = o(\exp(E/4))$ gives

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

again, as $E \gg \log N$. Ergo, by (3.2), $p_{\text{solve}}^2 \leq p_{\text{unstable}} + p_{\text{cond}} = o(1)$, as desired. □

Theorem 3.15. *Let $\omega\left(\log^2 N\right) \leq E \leq o(N)$, and let g, g' be $(1-\varepsilon)$-resampled standard Normal r.v.s. Then, for any algorithm $\mathcal{A}$ with Efron-Stein degree $D \leq o\left(E/\log^2 N\right)$ (and with $\mathbf{E}\|\mathcal{A}(g)\|^2 \leq CN$), there exist $\varepsilon, \eta > 0$ such that $p_{\text{solve}} = o(1)$.*

Proof: As in Theorem 3.13, it suffices to show that both p_{cond} and p_{unstable} go to zero, while $\mathbf{P}(S_{\text{diff}}) \approx 1$ (i.e., with $\varepsilon = \omega(1/N)$. As before, pick $\varepsilon = \log(N/D)/N$, ensuring that $N\varepsilon = \log(N/D) \gg 1$ (for $D = o(N)$, which holds as $E \leq o(N)$). Now recall that by Proposition 3.11 we have

$$p_{\text{cond}} \leq \exp(-E + 2\eta \log(1/\eta)N + O(1)).$$

In particular, if we choose

$$\eta = \frac{E}{16N \log(N/E)},$$

we have that

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

for $E/N \ll 1$, thus ensuring $p_{\text{cond}} \leq \exp(-3E/4 + O(1)) = o(1)$ (as $E \gg \log N$). Finally, the choice of $D \leq o\left(E/\log^2 N\right)$ combined with Proposition 2.20 now gives

$$\begin{aligned} p_{\text{unstable}} &\leq \frac{CD\varepsilon}{2\eta} \asymp \frac{D\varepsilon N \log(N/E)}{E} \\ &= \frac{D \log(N/D) \log(N/E)}{E} \leq \frac{D \log^2 N}{E} \to 0 \end{aligned}$$

By (3.4), $p_{\text{solve}}^2 \leq \mathbf{P}(S_{\text{diff}}) \cdot (p_{\text{unstable}} + p_{\text{cond}}) = o(1)$, thus completing the proof. □

4. Randomized Rounding Things

Claim: no two adjacent points on Σ_N (or pairs within $k = O(1)$ distance) which are both good solutions to the same problem. The reason is that this would require a subset of k signed coordinates $\pm g_{\{i_1\}}, \ldots, \pm g_{\{i_k\}}$ to have small sum, and there are only 2^k `binom`$\{N\}\{k\}l = O\left(N^k\right)$ possibilities, each of which is centered Gaussian with variance at least 1, so the smallest is typically of order $\Omega\left(N^{\{-k\}}\right)$.

Proposition 4.1. *Fix distinct points $x, x' \in \Sigma_N$ with $\|x - x'\| \leq 2\sqrt{k}$ (i.e. x, x' differ by k sign flips), for $k = O(1)$, and let g be any instance. Then,*

$$\mathbf{P}(x \in S(E; g) \text{ and } x' \in S(E; g)) \leq \exp(-E + O(1)).$$

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

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

Thus, for a fixed x, x', if

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

we can expand this into

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

Multiplying the lower equation by -1 and adding the resulting inequalities gives

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

where $\langle g, x_J\rangle$ is a $\mathcal{N}(0, k)$ r.v. (note that $k > 0$ so it is nondegenerate). Moreover, as $k = O(1)$, we get by the logic in Lemma 3.8 that

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

Theorem 4.2 (Solutions Can't Be Close)**.** *Let $k = O(1)$ and $E \gg \log N$. Then for any instance g, with high probability there are no pairs of distinct solutions $x, x' \in S(E; g)$ with $\|x - x'\| \leq 2\sqrt{k}$.*

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 $2^k = O(1)$ choices of signs and, by [Ver18, Exer. 0.0.5],

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

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

$$\mathbf{P}\begin{pmatrix} \exists\ \text{x},x' \text{ s.t.} \\ \text{(I)} \quad \|x - x'\| \leq 2\sqrt{k}, \\ \text{(II)} \quad \text{x},x' \in S(E; g) \end{pmatrix} \leq \mathbf{P}\begin{pmatrix} \exists\ \text{J} \subset [N], x_J \in \{\pm 1\}^k \text{ s.t.} \\ \text{(I)} \quad |J| \leq \text{k}, \\ \text{(II)} \quad |\langle g_J, x_J\rangle| \leq \exp(-E) \end{pmatrix} \leq \exp(-E + O(\log N)) = o(1). \qquad \square$$

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.
-

Meow meow Thus, rounding would destroy the solution.

- Say we're in the case where rounding changes the solution. (i.e. rounding moves x to point that is not the closest point x_*, whp.)
- Let $p_1, \ldots, p_N$ be the probabilities of disagreeing with x_* on each coordinate.
  - We know that $\sum p_i$ diverges (this is equivalent to the statement that rounding will changes the solution whp).
  - Reason: for each coord, chance of staying at that coordinate is $e^{-\Theta(p_i)}$.
- For each i, flip coin with heads prob $2p_i$, and keep all the heads.
  - By Borel-Cantelli type argument, typical number of heads will be $\omega(1)$.
- For every coin with a head, change coord with prob 50%, if tails, keep coord.
  - 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 $100l = \omega(1)$, 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.

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.

[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

[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

[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.

[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.

[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.

[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.

[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

[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

[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.

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

[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.

[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.

[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

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

[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.

[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.

[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.

[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.

[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.

[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

[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

[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.

[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

[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

[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.

[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

[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

[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

[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

[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.

[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.

[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.

[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

[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.

[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.

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

[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

[HTF09] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. in Springer Series in Statistics. New York, NY: Springer New York, 2009. doi: 10.1007/978-0-387-84858-7.

[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.

[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

[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

[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

[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.

[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

[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.

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

[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.

[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.

[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.

[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

[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.

[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.

[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

[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.

[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

[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

[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.

[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.

[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.

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

[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.

[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.

[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.

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

[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.

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

[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.

[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.

[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

[Wai19] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. in Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press, 2019. doi: 10.1017/9781108627771.

[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.

[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.

[90] 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

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