1. Chapter 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

2. Number Packing Problem

Let N be the dimensionality, and $\Sigma_N \coloneqq \{\pm 1\}$ 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 2.1. Let $\delta > 0$. The *solution set* $S(\delta; g)$ is the set of all $x \in \Sigma_N$ that are δ-*optimal*, satisfying

$$\frac{1}{\sqrt{N}} |\langle g, x \rangle| \leq 2^{-\delta N}. \tag{2.1}$$

(2.1) Hi

2.1. Existing Results

1. $X_i, 1 \leq i \leq n$ i.i.d. uniform from $\{1, 2, ..., M \coloneqq 2^m\}$, with $\kappa \coloneqq \frac{m}{n}$, 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 $2^{-\Theta(n)} \leq 2^{-o(n)}$)

3. Glossary and conventions

Conventions:

1. log means log in base 2, exp is exp base 2 - natural log/exponent is $\ln/e^x$.
2.

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"

4. Literature Review

[1]

[2]

[3]

[4]

[5]

[6]

[7]

[8]

[9]

[10]

[11]

[12]

[13]

[14]

[15]

[16]

[17]

[18]

[19]

[20]

[21]

[22]

[23]

[24]

[25]

[26]

[27]

[28]

[29]

[30]

[31]

[32]

[33]

[34]

[35]

[36]

[37]

[38]

[39]

[40]

[41]

[42]

[43]

[44]

[45]

[46]

[47]

[48]

[49]

[50]

[51]

[52]

[53]

[54]

[55]

[56]

[57]

[58]

[59]

[60]

[61]

[62]

[63]

[64]

[65]

[66]

[67]

[68]

[69]

[70]

[71]

[72]

[73]

[74]

[75]

[76]

[77]

[78]

[79]

[80]

[81]

[82]

[83]

[84]

[85]

[86]

[87]

[88]

5. Low-Degree Algorithms

What are algorithms of interest

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

Why study low-degree algorithms (poly time heuristic + simple)

Different notions of degree.

Overview of section

- Efron-Stein notion
- Hermite notion
- Algorithms and Stability Bounds

5.1. Efron-Stein Polynomials (TODO)

Definition 5.1. 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.

Motivation for studying decompositions of functions by projecting onto coordinates.

This section largely follows [76, § 8.3].

Want to decompose

$$f = \sum_{S \subseteq [n]} f^{=S} \tag{5.1}$$

Want $f^{=S}$ to only depend on the coordinates in S.

If $T \not\subseteq S$ and g depends only on the coordinates in T, then $\langle f^{=S}, g \rangle = 0$.

Definition 5.2. 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.

In particular, we should have that

$$f^{\subseteq J} = \sum_{S \subseteq J} f^{=S} \tag{5.2}$$

First, we consider the case $J = \varnothing$. It is clear that $f^{=\varnothing} = f^{\subseteq \varnothing}$, which is the constant function $\mathbf{E}[f]$.

Next, if $J = \{j\}$ is a singleton, (5.2) 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. It measures what difference in expectation of f is given x_j.

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

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

Inductively, all the $f^{=J}$ can be defined via the principle of inclusion-exclusion.

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

Theorem 5.3 ([76, 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.*

Definition 5.4. 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 $f = (f_1, \ldots, f_M) : \mathbf{R}^N \to \mathbf{R}^M$ 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.

As we are interested in how a function behaves under small changes in its input, we are led to consider the following "noise operator." First, we need the following notion of distance between problem instances:

Definition 5.5. 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 5.6. 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, $\langle f, T_p f\rangle = \mathbf{E}_{(x,y)\ p\text{-resampled}}[f(x) \cdot f(y)]$.

Lemma 5.7. *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 5.3 that $f^{\subseteq J} = \sum_{S \subseteq J} f^{=S}$, so

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

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

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

Theorem 5.8. *Let $p \in [0,1]$ and let $f = (f_1, \ldots, f_M) : \mathbf{R}^N \to \mathbf{R}^M$ be a multivariate function with 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*

$$\mathbf{E}\|f(x) - f(y)\|^2 \leq 2\left(1 - p^D\right) \leq 2(1-p)D. \tag{5.3}$$

Proof: Observe that

$$\begin{aligned} \mathbf{E}\|f(x) - f(y)\|^2 &= \mathbf{E}\|f(x)\|^2 + \mathbf{E}\|f(y)\|^2 - 2\mathbf{E}\langle f(x), f(y)\rangle \\ &= 2 - 2\left(\sum_i \mathbf{E}[f_i(x) f_i(y)]\right) \\ &= 2 - 2\left(\sum_i \langle f_i, T_p f_i\rangle\right). \end{aligned} \tag{5.4}$$

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 5.7 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 (5.4) to get

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

[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 $p \approx 1$.

as desired. $\square$

5.2. Hermite Polynomials (TODO)

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 5.8. This theory is much more classical, so we encourage the interested reader to see [76, § 11] for details.

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

Definition 5.9. 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 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 $L^2(\mathbf{R}, \gamma)$ [76, 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

$$h_0(z) = 1, \qquad h_1(z) = z, \qquad h_2(z) = \frac{z^2 - 1}{\sqrt{2}}, \qquad h_3(z) = \frac{z^3 - 3z}{\sqrt{6}}, \qquad \ldots \tag{5.5}$$

Note here that each h_j is a degree j polynomial.

It is then straightforward to show the following:

Theorem 5.10 ([76, 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

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

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

Theorem 5.11. *The Hermite polynomials $(h_\alpha)_{\alpha\in\mathbb{N}^N}$ form a complete orthonormal basis for $L^2(\mathbf{R}^N, \gamma^N)$. In particular, every $f \in L^2(\mathbf{R}^N, \gamma^N)$ has a unique expansion 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

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

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

Def. noise operator/Ornstein-Uhlenbeck operator

Compute effect of noise operator on Hermite polys

Thrm. Hermite polys form basis for 1D Gaussian space

Thrm. Products of Hermite polys form basis for N-dim Gaussian space

Noise operator on arbitrary function with given Hermite expansion

Definition 5.12. Let (g, g') be N-dimensional standard Normal vectors. We say (g, g') are *p-correlated* if (g_i, g_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 5.13. For $p \in [0, 1]$, the *Gaussian noise operator* T_p is the linear operator on $L^2(\mathbf{R}^N, \gamma^N)$, given by

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

In particular, a straightforward computation with the Normal moment generating function gives

Remark that degree D function can be expressed in terms of degree D and lower Hermite polynomials - gives a Hilbert basis which reflects the natural algebraic grading.

Thrm. Function stability for degree D polynomials.

5.3. Algorithms

Def. Randomized algorithm

Def. degree of algorithm is degree as multivariate function.

Discussion of how low-degree algs are approximate for class of Lipschitz algorithms?

Need for rounding function to land on Σ_N

Construction of randomized rounding function.

Constr. rounded algorithm.

Lemma. stability of rounding

Thrm. Stability of randomized algorithms (part 1 of Prop 1.9)

Show that Markov gives a useful bound on

Lemma 5.14. *Let $f : \mathbf{R}^N \to \mathbf{R}^N$, $p \in [0, 1]$, and X, Y marginally N-dimensional standard Normal vectors. Suppose that $\mathbf{E}\|f(X)\|_2^2 = 1$ and either of the following cases hold:*

I. *(X, Y) are a p-resampled pair, and f is a degree-D function.*

II. (X, Y) *are* p*-correlated, and* f *is a degree-*D *polynomial.*

Then

$$\mathbf{E}\|f(X) - f(Y)\|_2^2 \le 2\left(1 - p^D\right).$$

5.4. Algorithms

Definition 5.15. A *randomized algorithm* is a measurable function $\mathcal{A}^\circ : (g, \omega) \mapsto \boldsymbol{x} \in \mathbf{R}^N$, where $\omega \in \Omega_N$ is an independent random variable in some Polish space. Such an $\mathcal{A}^\circ$ is *deterministic* if it does not depend on ω.

Example. Let $\boldsymbol{U} = (U_1, \ldots, U_N)$ be i.i.d. Unif($[-1, 1]$). Then, we define the random coordinate-wise function

$$\mathrm{round}_{\boldsymbol{U}}(\boldsymbol{x}) = \left(\mathrm{round}_{U_1}(x_1), \ldots, \mathrm{round}_{U_N}(x_N)\right),$$

where we define

$$\mathrm{round}_U(x) = \begin{cases} 1 & x \ge U \\ -1 & x < U \end{cases}.$$

Example. Given a real-valued algorithm $\mathcal{A}^\circ$, we can convert it into a Σ_N-valued algorithm $\mathcal{A}$ via

$$\mathcal{A}(g, \omega, \boldsymbol{U}) := \mathtt{round}_{\boldsymbol{U}}\left(\mathcal{A}^\circ(g, \omega)\right).$$

Definition 5.16. Algorithm $\mathcal{A}$ is $(\varepsilon, \eta, p_{\text{unstable}})$-stable if, for g, g' $(1 - \varepsilon)$-correlated/resampled, we have

$$\mathbf{P}\left(\|\mathcal{A}(g) - \mathcal{A}(g')\| \le \eta\sqrt{N}\right) \ge 1 - p_{\text{unstable}}.$$

By the will of God (i.e. writeup pending), we have the following:

Lemma 5.17. *Algorithm* $\mathcal{A}$ *with degree* $\le D$ *and norm* $\mathbf{E}\|\mathcal{A}(g)\|^2 \le CN$ *has*

$$\mathbf{E}\|\mathcal{A}(g) - \mathcal{A}(g')\|^2 \le 2CN\varepsilon D,$$

and (because of randomized rounding)

$$\mathbf{E}\|\mathcal{A}(g) - \mathcal{A}(g')\|^4 \le 16CN^2\varepsilon D.$$

Thus,

$$\mathbf{P}\left(\|\mathcal{A}(g) - \mathcal{A}(g')\| \ge \eta\sqrt{N}\right) \le \frac{16CN^2\varepsilon D}{\eta^4 N^2} \asymp \frac{\varepsilon D}{\eta^4}.$$

As a consequence, a degree D algorithm $\mathcal{A}$ has $p_{\text{unstable}} = o_{N(1)}$ for $\eta^4 \gg \varepsilon D$.

[76]

[60]

6. Summary of Parameters

Parameter	Meaning	Desired Direction	Intuition
N	Dimension	Large	Showing hardness *asymptotically*, want "bad behavior" to pop up in low dimensions.
δ	Solution tightness; want to find x such that $\lvert\langle g, x\rangle\rvert \leq 2^{-\delta N}$	Small	Smaller δ implies weaker solutions, e.g. $\delta = 0$ is just finding solutions ≤ 1. Want to show even weak solutions are hard to find.
E	Solution tightness; "energy level"; 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 Karmarkar-Karp
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 $\lVert\mathcal{A}(g) - \mathcal{A}(g')\rVert \leq \eta\sqrt{N}$, for (g, g') $(1-\varepsilon)$-correlated.	Large	Large η indicates a more unstable algorithm; want to show that even weakly stable algorithms fail.

Table 1: Explanation of Parameters

7. Conditional Landscape Obstruction

Explain what the obstruction is.

We start with a bound on the geometry of the binary hypercube.

Lemma 7.1. *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

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

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

Lemma 7.2 (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 7.1, we get

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

Next, we can consider what this probability is in the case of correlated Normals.

Lemma 7.3. *Suppose (g, g') are $(1-\varepsilon)$-correlated 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

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

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

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

Note for instance that here 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.

First, we consider the probability of a solution being optimal for a resampled instance.

Lemma 7.4. *Suppose* (g, g') *are* $(1-\varepsilon)$*-resampled Normal vectors, and let* $x \in \Sigma_N$*. Then,*

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

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

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

Let $\mu := \sum_{i \notin S} g_i x_i$. The conditional distribution of interest can now be expressed as $\mathbf{P}\left(|\langle g', x'\rangle| \le 2^{-E} \mid g, |S| \ge 1\right)$. Given $|S| \ge 1$, the quantity $\langle g', x'\rangle$ is a nondegenerate random variable, with density bounded above as

$$\varphi_{g,|S|}(z) = \frac{1}{\sqrt{2\pi|S|}} e^{-\frac{(z-\mu)^2}{2|S|}} \le \frac{1}{\sqrt{2\pi|S|}} \le \frac{1}{\sqrt{2\pi}}.$$

Hence, the quantity of interest can be bounded as

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

Note that in the resampled case, we can compute the probability that $g = g'$ as

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

which for $\varepsilon \ll 1$ is approximately $1 - N\varepsilon$. Thus, for $\varepsilon \gg \omega\left(\frac{1}{N}\right)$, we have

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

8. Proof of Low-Degree Hardness in 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 [54].

For now, let $\mathcal{A}$ denote a Σ_N-valued deterministic algorithm. We discuss the extension to random, $\mathbf{R}^N$-valued algorithms later on in (section ???).

8.1. Hermite Case

First, we consider

Assume for sake of contradiction that $p_{\text{solve}} \geq \Omega(1)$.

Let g, g' be $(1-\varepsilon)$-correlated instances. We define the following events:

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

In this case, set $\varepsilon = 2^{-\delta N} = o(1)$

Lemma 8.1. *There exists an $\eta > 0$ of constant order such that*

$$\mathbf{P}(S_{\text{cond}}) \geq 1 - e^{-cN}$$

for an arbitrary constant c.

$D = o(2^N)$.

$D\varepsilon = \frac{D}{2^N} * 2^{(1-\delta)N}$

Lemma 8.2. *For any $\omega(\log^2 N) \leq E \leq \Theta(N)$, there exist choices of ε, η (depending on N, E) such that* $\mathbf{P}(S_{\text{ogp}}) = o(1)$.

Proof: Observe that

$$\mathbf{P}(S_{\text{ogp}}) = \mathbf{E}[\mathbf{P}(S_{\text{ogp}} \mid g)]. \tag{8.1}$$

Conditional on g, we can compute $\mathbf{P}(S_{\text{ogp}} \mid g) = \mathbf{P}\left(\exists x' \in S(E; g'), \|x - x'\| \leq 2\eta\sqrt{N}\right)$ by setting $x = \mathcal{A}(g)$ (so x only depends on g), and union bounding Lemma 7.4 over the x' within $2\eta\sqrt{N}$ of x, as per Lemma 7.2:

$$\mathbf{P}(S_{\text{ogp}} \mid g) \leq \exp_2\left(-E + N\eta^2 \log_2(1/\eta^2) + O(1)\right).$$

We want to choose η such that

$$-E + N\eta^2 \log_2(1/\eta^2) = -\Omega(N)$$

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

Using the fact that $\sqrt{2x} \geq -x \log_2 x$, it suffices to pick η^2 with

$$\frac{E}{N} > 2\eta,$$

so $\eta^2 := \frac{E^2}{2N^2}$ is a valid choice.

By the choice of $\eta^2 \ll (h^{-1})(\delta) \asymp 1$, this bound gives $\mathbf{P}(S_{\text{ogp}}|g) \leq \exp_2(-O(N)) = o(1)$. Integrating over g gives the overall bound. □

When $CD\varepsilon N^2 = \omega_{N(1)}$ (i.e. goes to infinity),

$$\begin{aligned}\mathbf{P}(S_{\text{stable}}) &\leq \frac{16CD\varepsilon N^2}{16\eta^4 N^2} \\ &= \frac{CD\varepsilon}{\eta^4} = \frac{4CD\varepsilon N^4}{E^4}\end{aligned}$$

$D\varepsilon \to 0$ same as $D = o\left(\frac{1}{\varepsilon}\right) = o(N)$.

Lemma 8.3. $\mathbf{P}(S_{\text{solve}}, S_{\text{stable}}) \leq \mathbf{P}(S_{\text{ogp}}) = o(1)$.

Proof: The first inequality follows from definition, with $x = \mathcal{A}(g)$ and $x' = \mathcal{A}(g')$. For the second, observe that

$$\mathbf{P}(S_{\text{ogp}}) = \mathbf{E}[\mathbf{P}(S_{\text{ogp}} \mid g)].$$

Now, let $M = \delta N$, we can compute $\mathbf{P}(S_{\text{ogp}} \mid g) = \mathbf{P}\left(\exists x' \in S(\delta; g'), \|x - x'\| \leq \eta\sqrt{N}\right)$ by setting $x = \mathcal{A}(g)$ (so x only depends on g), and union bounding Lemma 7.4 over the x' within $\eta\sqrt{N}$ of x, as per Lemma 7.2:

$$\mathbf{P}(S_{\text{ogp}} \mid g) \leq \exp_2\left(-\delta N + Nh\left(\frac{\eta^2}{4}\right) + O(\log_2 N)\right).$$

By the choice of $\eta^2 \ll (h^{-1})(\delta) \asymp 1$, this bound gives $\mathbf{P}(S_{\text{ogp}}|g) \leq \exp_2(-O(N)) = o(1)$. Integrating over g gives the overall bound. □

However, by the choice of parameters above, we also have

$$\begin{aligned}\mathbf{P}(S_{\text{solve}}, S_{\text{stable}}) &\geq \mathbf{P}(S_{\text{solve}}) + \mathbf{P}(S_{\text{stable}}) - 1 \\ &\geq p_{\text{solve}}^4 + p_{\text{unstable}} \geq \Omega(1) - o(1) = \Omega(1),\end{aligned} \tag{8.2}$$

which is a contradiction.

9. 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`$\{N\}\{k\}l = O(N^k)$ possibilities, each of which is centered Gaussian with variance at least 1, so the smallest is typically of order $\Omega\left(N^{\{-k\}}\right)$.

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 $p_1, ..., p_N$ 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.

[76]

Bibliography

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

[2] 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

[3] 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

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

[5] 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.

[6] 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.

[7] 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.

[8] 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

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

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

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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

[17] 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

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

[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.

[20] 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.

[21] 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.

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

[23] 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.

[24] 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

[25] 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

[26] 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

[27] 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.

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

[29] 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

[30] 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. 2021, doi: 10.1214/20-AAP1625.

[31] 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

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

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

[34] 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

[35] 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

[36] 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.

[37] 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.

[38] 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.

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

[40] 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.

[41] 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.

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

[43] 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

[44] 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.

[45] 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.

[46] 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

[47] 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

[48] S. Hopkins, "Statistical Inference and the Sum of Squares Method," 2018.

[49] 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

[50] 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

[51] 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.

[52] 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

[53] 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

[54] 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.

[55] 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

[56] 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.

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

[58] 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.

[59] 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.

[60] 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.

[61] 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.

[62] 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

[63] 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.

[64] 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.

[65] 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

[66] 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.

[67] 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

[68] 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

[69] 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.

[70] 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.

[71] 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.

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

[73] 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.

[74] 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.

[75] 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.

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

[77] 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

[78] 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.

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

[80] 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.

[81] 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.

[82] 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.

[83] 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

[84] 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.

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

[86] 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.

[87] 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.

[88] 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.