A Random CSP with Connections to Discrepancy Theory and Randomized Trials

Eren C. Kızıldağ
Columbia University
Email: eck2170@columbia.edu

***Abstract*—We introduce a random constraint satisfaction problem (CSP) with non-uniform constraints that is closely related to the average-case discrepancy minimization problem in the non-proportional regime. Our proposal is particularly motivated by randomized controlled trials (RCTs) in statistics, involving different constraints. For the random CSP that we propose, we establish a sharp phase transition result regarding the existence of its solutions. We then precisely pinpoint the distance between the solution spaces corresponding to independent problem instances. In the context of RCTs, this quantifies the amount of reassignments needed if a similar RCT is to be repeated with an independent population and/or a potentially different set of constraints. We lastly study the solution space geometry, and show that, for certain values of constraints, the solutions are isolated singletons separated by linear Hamming distance.**

I. Introduction

Given vectors $Y_1, \dots, Y_n \in \mathbb{R}^d$, the vector balancing problem (VBP) seeks to find a *balanced* partition of these vectors, namely a $\boldsymbol{\sigma} \in \Sigma_n \triangleq \{-1, 1\}^n$ such that $\| \sum_{i \le d} \boldsymbol{\sigma}(i) Y_i \|_\infty$ is small (where $\|Y\|_\infty = \max_{1 \le j \le d} |Y(j)|$ for $Y \in \mathbb{R}^d$).

The VBP is of great practical and theoretical significance. In statistics, the VBP is closely related to the design of *randomized controlled trials* (RCTs)—often considered as the gold standard for clinical experiments. Consider an RCT involving n individuals, each characterized by covariate information $Y_i \in \mathbb{R}^d, 1 \le i \le n$, aimed at understanding the efficacy of an additive treatment effect (such as a new drug or vaccine).[1] These individuals are then split into two groups, i.e. *treatment* and *control*. For accurate inference for the treatment effect, a good *covariate balance* is essential. See [1]–[10] for a more elaborate discussion on RCTs and pointers to literature. Other applications of VBP include multiprocessor scheduling, design of VLSI circuits, and cryptography, see e.g. [11]–[13].

In addition to its practical relevance, the VBP is also widely studied in theoretical computer science, statistical physics, and discrepancy theory. Its one-dimensional version ($d = 1$), is known as the number partitioning problem (NPP); the NPP is among Karp's famous list of 21 NP-complete problems [14], as well as among Garey and Johnson's six basic NP-complete problems [15]. The NPP exhibits a certain *phase transition*, first conjectured in [16] using elegant yet non-rigorous tools of statistical mechanics, and subsequently confirmed later in [17]. Additionally, the NPP is also one of the first models in computer science for which the local REM conjecture from statistical physics is verified, see [18] for the original conjecture and [19], [20] for its proof.

The VBP is at the heart of combinatorics, in particular *discrepancy theory* [21]–[23]. Given $M \in \mathbb{R}^{d \times n}$, a canonical goal in discrepancy theory is to compute or bound its discrepancy $\mathcal{D}(M) \triangleq \min_{\boldsymbol{\sigma} \in \Sigma_n} \|M\boldsymbol{\sigma}\|_\infty$. Both *worst-case* as well as *average-case* settings were considered in the discrepancy literature. In the worst-case setting, a landmark result due to Spencer, dubbed as "six standard deviations suffice", asserts that if $\max_{i \le n} \|Y_i\|_\infty \le 1$, where $Y_i \in \mathbb{R}^d$ are the columns of $M \in \mathbb{R}^{d \times n}$, then $\mathcal{D}(M) \le 6\sqrt{n}$ [21]. While Spencer's guarantee is non-constructive, algorithmic guarantees were also sought in the discrepancy literature, see e.g. [24]–[27].

Of particular relevance to us is average-case discrepancy where $M \in \mathbb{R}^{d \times n}$ is random. In the special case where the entries of M are i.i.d. standard normal, $M_{ij} \sim \mathcal{N}(0, 1)$, $\mathcal{D}(M) = O(\sqrt{n} 2^{-n/d})$ w.h.p. as $n \to \infty$; case $d = 1$ is due to Karmarkar, Karp, Lueker, and Odlyzko [28], $d = O(1)$ is due to Costello [29], and $\omega(1) \le d \le o(n)$ is due to Turner, Meka, and Rigollet [3]. In the *proportional* regime where $d = \Theta(n)$, this model is closely related to the symmetric binary perceptron; Perkins and Xu [30] and Abbe, Li, and Sly [31] established independently that $\mathcal{D}(M) = (1 + o(1)) f(\alpha) \sqrt{n}$, where $f(\cdot)$ is an explicit function and $\alpha = d/n$ is held constant while $n \to \infty$. See next section for more details. These guarantees are non-constructive. As for the constructive guarantees, the best known algorithm for the VBP finds in polynomial time a $\boldsymbol{\sigma} \in \Sigma_n$ with $\|M\boldsymbol{\sigma}\|_\infty = 2^{-\Omega(\log^2 n/d)}$ w.h.p. as long as $d = O(\sqrt{\log n})$ [3]; no better algorithm that finds in polynomial time a $\boldsymbol{\sigma}$ with $\|M\boldsymbol{\sigma}\|_\infty = 2^{-\omega(\log^2 n/d)}$ w.h.p. is known. This is an instance of a *statistical-computational gap*—a striking gap between the existential and the best known algorithmic guarantee. Using insights from statistical physics, the limits of efficient algorithms were explored and rigorous hardness guarantees were obtained in [4], [5], [32]–[34].

In this paper, we propose and rigorously investigate a random constraint satisfaction problem (CSP) that is closely related to the VBP. Random CSPs are extensively studied in the literature through various lenses, ranging from existence of solutions and solution space geometry to algorithmic aspects; for further discussion, see [30] and the references therein. Our particular motivation comes from the design of RCTs that are subject to non-uniform constraints, as well as from a connection between discrepancy theory and random CSPs, see next section for details. The remainder of the paper is

[1] Vector Y_i carries information regarding individual i, such as their height, weight, blood sugar, and so on.

organized as follows. We introduce and motivate the model in Section II, present our main results in Section III, highlight some future research directions in Section IV, and finally, provide complete proofs in Section V.

Notation. Denote by Σ_n the set $\{-1,1\}^n$. Given $N \in \mathbb{N}$, $[N]$ denotes $\{1,2,\dots,N\}$. For any proposition E, denote its indicator by $\mathbb{1}\{E\} \in \{0,1\}$. Given a set S, denote by $|S|$ its cardinality. Given $\boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n$, let $\langle \boldsymbol{x},\boldsymbol{y}\rangle \triangleq \sum_{i\le n} \boldsymbol{x}_i\boldsymbol{y}_i$, $\|\boldsymbol{x}\|_\infty = \max_{i\in[n]} |\boldsymbol{x}_i|$ and $\|\boldsymbol{x}\|_1 = \sum_{i\in[n]} |\boldsymbol{x}_i|$. Given $\boldsymbol{\sigma},\boldsymbol{\sigma}' \in \Sigma_n$, $d_H(\boldsymbol{\sigma},\boldsymbol{\sigma}') \triangleq \sum_i \mathbb{1}\{\boldsymbol{\sigma}(i) \neq \boldsymbol{\sigma}'(i)\}$. For any $M \in \mathbb{R}^{d\times n}$, $\mathcal{D}(M) \triangleq \min_{\boldsymbol{\sigma}\in\Sigma_n} \|M\boldsymbol{\sigma}\|_\infty$. For any Σ, $\mathcal{N}(0,\Sigma)$ denotes the centered multivariate normal with covariance Σ. For $n \in \mathbb{N}$, I_n is the $n\times n$ identity matrix. For any $r>0$, $\log_r(\cdot)$ and $\exp_r(\cdot)$ denote respectively the logarithm and the exponential functions base r; for $r=e$, we omit the subscript. Given $p \in (0,1)$, $h(p) \triangleq -p\log_2 p - (1-p)\log_2(1-p)$ is the binary entropy. We omit all floor/ceiling operators. We use standard asymptotic notation, e.g. $\Theta(\cdot), O(\cdot), o(\cdot)$, and $\omega(\cdot)$, where the underlying asymptotics are taken w.r.t. $n\to\infty$.

II. Model and Motivation

In this section, we propose a random CSP with non-uniform constraints that is closely linked to the VBP.

Definition 1. *Fix a $\boldsymbol{c} = (c_1,\dots,c_d) \in \mathbb{R}^d_+$ and let $\Xi = \{X_1,\dots,X_d\} \subset \mathbb{R}^n$ be a collection of i.i.d. random vectors $X_i \sim \mathcal{N}(0,I_n), i\in[d]$. Define*

$$\mathcal{F}\big(\Xi,\boldsymbol{c}\big) = \bigcap_{i\le d} \big\{\boldsymbol{\sigma}\in\Sigma_n : |\langle \boldsymbol{\sigma}, X_i\rangle| \le \sqrt{n}2^{-c_i n}\big\}. \quad (1)$$

Several remarks are in order. First, $\mathcal{F}(\Xi,\boldsymbol{c})$ is the random set consisting of all $\boldsymbol{\sigma}\in\Sigma_n$ that satisfy $|\langle\boldsymbol{\sigma},X_i\rangle| \le \sqrt{n}2^{-c_i n}$ for all $i\in[d]$. For this reason, we refer to $\mathcal{F}(\Xi,\boldsymbol{c})$ as the solution space (corresponding to the underlying CSP). Next, note that the constraints are indeed non-uniform as c_i are potentially distinct. Throughout, we assume $c_i < 1$ for all $i\in[d]$. As we mentioned earlier, $\min_{\boldsymbol{\sigma}\in\Sigma_n}|\langle\boldsymbol{\sigma},X_i\rangle| = O(\sqrt{n}2^{-n})$ w.h.p. [28], so $\mathcal{F}\big(\Xi,\boldsymbol{c}\big) = \varnothing$ w.h.p. if $c_i > 1$ for some i. Our particular focus is on $\mathcal{F}(\Xi,\boldsymbol{c})$ when $n\to\infty$ while the number d of constraints remains constant in n, $d=O(1)$. We refer to this as the *non-proportional* regime; it is in contrast with the proportional regime where d scales linearly with n, $d=\Theta(n)$.

We now provide some motivations for our model.

a) Design of RCTs: Let $M \in \mathbb{R}^{d\times n}$ have rows $X_1,\dots,X_d \in \mathbb{R}^n$ and columns $Y_1,\dots,Y_n\in\mathbb{R}^d$. In the context of RCTs, n individuals are participating in a study that aims to assess the efficacy of an additive treatment effect (such as a new drug or vaccine). In particular, Y_i is the vector of covariates associated with individual $i\in[n]$, and the regime $n \gg d$ is relevant as the number of participants is likely larger than that of covariates. Individuals i with $\boldsymbol{\sigma}(i)=1$ are assigned to the treatment group which gets the drug/vaccine, and those with $\boldsymbol{\sigma}(i)=-1$ to the control group which gets a placebo; the responses are evaluated. Now, fix a covariate $j\in[d]$ and consider the set $\{Y_i(j) : i\in[n]\}$, namely the values that covariate j takes across the population of n people. As mentioned earlier, a good covariate balance is essential for accurate inference, the main goal of the RCT. One way to interpret the covariate balance is that for any $j\in[d]$, the difference (regarding covariate j) between the treatment and the control group does not exceed a fixed threshold: for some predetermined t_j, $\mathcal{D}_j \triangleq |\sum_{i:\boldsymbol{\sigma}(i)=1} Y_i(j) - \sum_{i:\boldsymbol{\sigma}(i)=-1} Y_i(j)| < t_j$ for all $j\in[d]$. Naturally, the values t_j need not be equal since the range of covariate values need not be the same; therefore designs with such non-uniform constraints are of potential practical interest. Notice now that if $\boldsymbol{\sigma}\in\mathcal{F}(\Xi,\boldsymbol{c})$, then $\mathcal{D}_j \le t_j$ for all $j\in[d]$, where $t_j = \sqrt{n}2^{-c_j n}$. That is, any $\boldsymbol{\sigma}\in\mathcal{F}(\Xi,\boldsymbol{c})$ is a valid design. Hence, to design the RCT, it suffices to solve the random CSP arising in Definition 1. While a more compelling model would relax the distributional assumptions on Ξ and focus on t_i's that are not necessarily at an exponentially small scale, our model, incorporating non-uniform constraints into the design, should be viewed as a preliminary attempt towards studying such more sophisticated random designs. More complicated designs where Y_i have mixed (integer and continuous) entries or where X_i are not identically distributed are left for future work.

b) Connections between Discrepancy and Random CSPs: We explicate the following connection between discrepancy theory and random CSPs. Consider a random $M\in\mathbb{R}^{d\times n}$ in the proportional regime $d=\Theta(n)$, where $\alpha \triangleq d/n$ is fixed while $n\to\infty$. Given $\kappa>0$, the symmetric binary perceptron (SBP) model studies the set of solutions to $\|M\boldsymbol{\sigma}\|_\infty \le \kappa\sqrt{n}, \boldsymbol{\sigma}\in\Sigma_n$ [35]. The SBP is a toy neural network storing random patterns. Two fundamental questions about this model are: (a) what is the largest α for which such a $\boldsymbol{\sigma}$ exists, and (b) when do efficient search algorithms work? Note that this is the inverse of the discrepancy perspective, where one fixes $\alpha>0$ first and seeks the smallest $\kappa>0$ (i.e., the right hand side) for which a $\boldsymbol{\sigma}$ with $\|M\boldsymbol{\sigma}\|_\infty \le \kappa\sqrt{n}$ exists. Moreover, the best known algorithm for the SBP also comes from the discrepancy literature [36], further highlighting the connection between the two. The SBP received significant attention and was studied extensively, see e.g. [30]–[33], [37], [38]. The model we introduce in Definition 1 is similar, being the dual of the discrepancy view. We ask: for which fixed $\boldsymbol{c}\in\mathbb{R}^d_+$ (i.e., the right hand side) do solutions satisfying (1) exist? On the other hand, our model involves non-uniform constraints, and concerns the non-proportional regime where the number d of constraints remains constant in n, $d=O(1)$, whereas the SBP studies regime d growing linearly with n, $d=\Theta(n)$. We highlight that the exponential scaling in (1) is due to the fact that for $M\in\mathbb{R}^{d\times n}$ with random i.i.d. entries and $d=O(1)$, the discrepancy of M is also exponentially small, $\mathcal{D}(M) = \sqrt{n}2^{-\Omega(n)}$ w.h.p. [3], [28], [29]; this is in contrast with the SBP where the $\sqrt{n}$ scaling is needed since $\mathcal{D}(M) = O(\sqrt{n})$ w.h.p. for $d=\Theta(n)$, see [30], [31].

III. Main Results

A. Existence of Solutions: A Sharp Phase Transition

Given $\boldsymbol{c}\in\mathbb{R}^d_+$, the first natural question is: when is $\mathcal{F}(\Xi,\boldsymbol{c})$ per (1) non-empty, ideally w.h.p. as $n\to\infty$? (The probability

is taken w.r.t. Ξ.) Our first main result answers this question.

Theorem 1. *Fix $d \in \mathbb{N}$ and a $\boldsymbol{c} = (c_1, \dots, c_d) \in \mathbb{R}_+^d$. Then,*

$$\lim_{n\to\infty} \mathbb{P}\big[\mathcal{F}(\Xi, \boldsymbol{c}) \neq \varnothing\big] = \begin{cases} 0, & \text{if } \|\boldsymbol{c}\|_1 > 1 \\ 1, & \text{if } \|\boldsymbol{c}\|_1 < 1 \end{cases}.$$

That is, the event $\big\{\mathcal{F}(\Xi, \boldsymbol{c}) \neq \varnothing\big\}$ undergoes a *sharp phase transition* as $\|\boldsymbol{c}\|_1$ varies: as $n \to \infty$, $\mathcal{F}(\Xi, \boldsymbol{c})$ is w.h.p. non-empty (resp. empty) if $\|\boldsymbol{c}\|_1 < 1$ (resp. $\|\boldsymbol{c}\|_1 > 1$). Note that this is in agreement with Costello [29] who establishes that $\min_{\boldsymbol{\sigma}\in\Sigma_n} \|M\boldsymbol{\sigma}\|_\infty$ is of order $\sqrt{n}2^{-n/d}$ w.h.p. if $M \in \mathbb{R}^{d\times n}$ and $d = O(1)$. In some sense, Costello's result is closely related to a special case of Theorem 1 with $c_i \sim 1/d, \forall i \in [d]$.

Proof Idea The proof of Theorem 1 is based on the *moment method* [39]. Specifically, let T be the random variable counting the number of elements in $\mathcal{F}(\Xi, \boldsymbol{c})$: $T = |\mathcal{F}(\Xi, \boldsymbol{c})|$. When $\|\boldsymbol{c}\|_1 > 1$, we show that $\mathbb{E}[T] = \exp(-\Theta(n))$. Markov's inequality then gives $\mathbb{P}[T \geq 1] \leq \exp(-\Theta(n))$, yielding one part of Theorem 1. This is known as the first moment method. The case $\|\boldsymbol{c}\|_1 < 1$, on the other hand, is more delicate; it requires estimating both $\mathbb{E}[T]$ as well as $\mathbb{E}[T^2]$ and applying the Paley-Zygmund Inequality [39]: $\mathbb{P}[T > 0] \geq \mathbb{E}[T]^2/\mathbb{E}[T^2]$. Estimating $\mathbb{E}[T^2]$ is somewhat involved—it requires studying a sum running over all pairs $(\boldsymbol{\sigma}, \boldsymbol{\sigma}') \in \Sigma_n \times \Sigma_n$. Our argument shows that this sum is essentially dominated by pairs that are nearly orthogonal, i.e. $n^{-1}\langle \boldsymbol{\sigma}, \boldsymbol{\sigma}'\rangle \in [-\epsilon, \epsilon]$ for ϵ small. See Section V-B for the complete proof.

B. Independent Instances: Distance Between Solution Spaces

Our next focus is on solution spaces generated by independent instances. To set the stage, let $\Xi = \{X_1, \dots, X_d\} \subset \mathbb{R}^n$ be as in Definition 1 and Ξ' be an i.i.d. copy of Ξ. Fix $\boldsymbol{c} \in \mathbb{R}_+^d$ and $\boldsymbol{c}' \in \mathbb{R}_+^d$, not necessarily equal with $\|\boldsymbol{c}\|_1 < 1$ and $\|\boldsymbol{c}'\|_1 < 1$, and consider $\mathcal{F}(\Xi, \boldsymbol{c})$ and $\mathcal{F}(\Xi', \boldsymbol{c}')$ per (1). Two natural random geometrical questions are: (a) when is $\mathcal{F}(\Xi, \boldsymbol{c}) \cap \mathcal{F}(\Xi', \boldsymbol{c}')$ non-empty, and (b) when the intersection is empty, how far apart are the sets $\mathcal{F}(\Xi, \boldsymbol{c})$ and $\mathcal{F}(\Xi', \boldsymbol{c}')$? Before presenting our main result, we motivate these questions in the context of RCTs.

Motivation from RCTs Suppose we have a design $\boldsymbol{\sigma} \in \mathcal{F}(\Xi, \boldsymbol{c})$ involving a population Ξ and constraints (prescribed by) $\boldsymbol{c} \in \mathbb{R}_+^d$.[2] Suppose that we are to design a new RCT $\boldsymbol{\sigma}'$ involving a new population Ξ' and potentially different constraints $\boldsymbol{c}'$. One particular example would be repeating a similar RCT either at a different region (so the populations do not overlap) or many years later (where one might need new participants) with potentially different constraints. In this case, it is plausible to assume Ξ' is an i.i.d. copy of Ξ; the questions above simply ask whether one can use the existing design $\boldsymbol{\sigma}$ as is, and if not, how many changes (in the coordinates of $\boldsymbol{\sigma}$) are needed.[3]

[2]Recall that for M, whose rows are X_i, the elements of Ξ, its columns represent covariates corresponding to different individuals.

[3]We do not inquire on how to constructively find such a design, an interesting question left for future work.

Having motivated the questions raised above, we define

$$d(\boldsymbol{c}, \boldsymbol{c}') = \min_{\boldsymbol{\sigma}\in\mathcal{F}(\Xi,\boldsymbol{c}), \boldsymbol{\sigma}'\in\mathcal{F}(\Xi',\boldsymbol{c}')} \frac{d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma}')}{n} \tag{2}$$

to be the normalized distance between $\mathcal{F}(\Xi, \boldsymbol{c})$ and $\mathcal{F}(\Xi', \boldsymbol{c}')$. Note that $d(\boldsymbol{c}, \boldsymbol{c}')$ is random and $d(\boldsymbol{c}, \boldsymbol{c}') \in [0, 1]$ almost surely for all $\boldsymbol{c}, \boldsymbol{c}'$. A direct corollary to Theorem 1 is as follows.

Corollary 1. *$\mathcal{F}(\Xi, \boldsymbol{c}) \cap \mathcal{F}(\Xi', \boldsymbol{c}') \neq \varnothing$ w.h.p. if $\|\boldsymbol{c}\|_1 + \|\boldsymbol{c}'\|_1 < 1$ and $\mathcal{F}(\Xi, \boldsymbol{c}) \cap \mathcal{F}(\Xi', \boldsymbol{c}') = \varnothing$ w.h.p. if $\|\boldsymbol{c}\|_1 + \|\boldsymbol{c}'\|_1 > 1$.*

In particular, $d(\boldsymbol{c}, \boldsymbol{c}') = 0$ w.h.p. if $\|\boldsymbol{c}\|_1 + \|\boldsymbol{c}'\|_1 < 1$. Corollary 1 follows immediately by applying Theorem 1 to the set $\mathcal{F}(\bar{\Xi}, \bar{\boldsymbol{c}})$ per (1) with $\bar{\Xi} = \Xi \cup \Xi'$ and $\bar{\boldsymbol{c}} = (c_1, \dots, c_d, c_1', \dots, c_d') \in \mathbb{R}_+^{2d}$, and noticing that $\mathcal{F}(\bar{\Xi}, \bar{\boldsymbol{c}}) = \mathcal{F}(\Xi, \boldsymbol{c}) \cap \mathcal{F}(\Xi', \boldsymbol{c}')$. This answers the first question, namely when the intersection is non-empty.

The second question, regarding the distance between $\mathcal{F}(\Xi, \boldsymbol{c})$ and $\mathcal{F}(\Xi', \boldsymbol{c}')$ when their intersection is empty, turns out more delicate, and is the subject of our next result.

Theorem 2. *Let $\boldsymbol{c}, \boldsymbol{c}' \in \mathbb{R}_+^d$ with the property that $\max\{\|\boldsymbol{c}\|_1, \|\boldsymbol{c}'\|_1\} < 1$ and $\|\boldsymbol{c}\|_1 + \|\boldsymbol{c}'\|_1 > 1$. Define by $\gamma^* \in (0, \frac{1}{2})$ the unique value*

$$h(\gamma^*) = \|\boldsymbol{c}\|_1 + \|\boldsymbol{c}'\|_1 - 1,$$

where $h(p) = -p\log_2 p - (1-p)\log_2(1-p)$, $p \in [0, 1]$, is the binary entropy function. Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\big[\big|d(\boldsymbol{c}, \boldsymbol{c}') - \gamma^*\big| \leq \epsilon\big] = 1.$$

That is, $d(\boldsymbol{c}, \boldsymbol{c}')$ converges in probability to γ^*, asserting that the distance between $\mathcal{F}(\Xi, \boldsymbol{c})$ and $\mathcal{F}(\Xi', \boldsymbol{c}')$ is (w.h.p.) of order $\Omega(n)$. Note that γ^* is well-defined as $h : [0, \frac{1}{2}] \to [0, 1]$ is a bijection (a well-known fact, see, e.g., [40]).

Proof Idea The proof of Theorem 2 is based on introducing certain auxiliary random variables and applying the moment method. The details of the second moment method, on the other hand, are quite involved; we need to study a sum running over all quadruples of form $(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_2')$ where $\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \mathcal{F}(\Xi, \boldsymbol{c})$, $\boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2' \in \mathcal{F}(\Xi', \boldsymbol{c}')$ and $d_H(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_i') \sim \gamma^*$, and understand their pairwise overlaps in order to study a certain covariance matrix arising in the probability calculation. Our argument shows that the second moment is dominated by nearly orthogonal quadruples, i.e. those with $\frac{1}{n}\langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2\rangle \in [-\delta, \delta]$ and $\frac{1}{n}\langle \boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2'\rangle \in [-\delta, \delta]$, where $\delta > 0$ is small. For full proof, see Section V-C.

C. Solution Space Geometry: Distance Between Solutions

Our last focus is on the *geometry* of the solution space $\mathcal{F}(\Xi, \boldsymbol{c})$ per (1). By inspecting (1), we observe that $\mathcal{F}(\Xi, \boldsymbol{c})$ shrinks as $\|\boldsymbol{c}\|_1$ gets larger. Intuitively, this suggests that solutions become 'more isolated' as $\|\boldsymbol{c}\|_1$ grows. Our next result confirms this.

Theorem 3. *Let $\mathcal{F}(\Xi, \boldsymbol{c})$ be as in* (1)*.*

(a) *Let $\|\boldsymbol{c}\|_1 > \frac{1}{2}$. Then, there exists a $\beta^* \triangleq \beta^*(\boldsymbol{c}) \in (0, 1)$ such that w.h.p. as $n \to \infty$,*

$$\min_{\boldsymbol{\sigma}, \boldsymbol{\sigma}'\in\mathcal{F}(\Xi,\boldsymbol{c}), \boldsymbol{\sigma}\neq\boldsymbol{\sigma}'} d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma}') \geq \beta^* n.$$

(b) *Let* $\|\boldsymbol{c}\|_1 < \frac{1}{2}$ *and* $\beta \in (0,1)$ *be arbitrary. Then,* $\mathbb{E}[N_\beta] = e^{\Theta(n)}$ *where*

$$N_\beta \triangleq \big|\big\{(\boldsymbol{\sigma},\boldsymbol{\sigma}') \in \mathcal{F}(\Xi,\boldsymbol{c})^2 : 1 \le d_H(\boldsymbol{\sigma},\boldsymbol{\sigma}') \le \beta n\big\}\big|.$$

Note that Part (a) of Theorem 3 asserts that if $\|\boldsymbol{c}\|_1 > \frac{1}{2}$, then the solutions are isolated: w.h.p. any pair of solutions $(\boldsymbol{\sigma},\boldsymbol{\sigma}')$ are $\Omega(n)$ apart. Part (b), on the other hand, gives a first moment 'evidence' towards the hypothesis that for small $\|\boldsymbol{c}\|_1$, there exist solution pairs at arbitrarily small distances. Proving this amounts to showing $N_\beta \ge 1$ w.h.p., for which one needs to study $\mathbb{E}[N_\beta^2]$. The second moment calculation, on the other hand, appears even more involved than Theorem 2; we leave it for future work. The proof of Theorem 3 is based on the first moment method and provided in Section V-D.

IV. Future Work

We close by outlining several future directions. It would be very interesting to show that for $\|\boldsymbol{c}\|_1 < \frac{1}{2}$ and any $\beta \in (0,1)$, there exist (w.h.p.) solution pairs at distance βn, which would strengthen Theorem 3(b). This can potentially be done through a second moment calculation. Another interesting question is whether the solution space exhibits the *shattering* property, see e.g. [41], [42]. While Theorem 1 yields the existence of solutions, our proof technique—the second moment method—is non-constructive. Are there polynomial-time algorithms to find $\boldsymbol{\sigma} \in \mathcal{F}(\Xi,\boldsymbol{c})$, and if so, what are their fundamental limits? These questions can be addressed using the Overlap Gap Property framework [43]. Lastly, it would be interesting to analyze a version of SBP with similar non-uniform constraints.

V. Proofs

We provide below the complete proofs of all of our results.

A. Auxiliary Results

We collect several useful auxilary results here. Our first result regards certain Gaussian probabilities.

Lemma 1. (a) *Let* $Z \sim \mathcal{N}(0,1)$ *and* $z = o_n(1)$. *Then,*

$$\mathbb{P}[|Z| \le z] = \sqrt{\frac{2}{\pi}} z(1 + o_n(1)).$$

(b) *Let* $Z, Z_\rho \sim \mathcal{N}(0,1)$ *with* $\mathbb{E}[ZZ_\rho] = \rho \in [0,1)$. *Suppose* z_1, z_2 *are such that* $(z_1^2 + z_2^2)/\sqrt{1-\rho^2} = o_n(1)$. *Then,*

$$\mathbb{P}\big[|Z| \le z_1, |Z_\rho| \le z_2\big] = \frac{2 z_1 z_2}{\pi\sqrt{1-\rho^2}}\big(1 + o_n(1)\big).$$

Lemma 1 is reproduced from [5, Lemma 5.6], see the proof therein. Our next auxiliary result regards binomial coefficients.

Lemma 2. (a) *Let* $\rho \in (0,1)$. *Then,* $\binom{n}{\rho n} = \exp_2\big(nh(\rho) + o(n)\big)$, *where* $h(\rho) = -\rho\log_2\rho - (1-\rho)\log_2(1-\rho)$.
(b) *Fix* $\alpha \le \frac{1}{2}$. *Then, for all* n,

$$\sum_{i \le \alpha n} \binom{n}{i} \le 2^{nh(\alpha)}.$$

See [44, Section 17.5] for the proof of Lemma 2(a) and [45, Theorem 3.1] for Lemma 2(b).

B. Proof of Theorem 1 (Compressed)

Fix a $\boldsymbol{c} \in \mathbb{R}_+^d$ and let $\mathcal{F}(\Xi,\boldsymbol{c})$ be as in (1). Define

$$T \triangleq \big|\mathcal{F}(\Xi,\boldsymbol{c})\big| = \sum_{\boldsymbol{\sigma}\in\Sigma_n} \mathbb{1}\big\{\boldsymbol{\sigma} \in \mathcal{F}(\Xi,\boldsymbol{c})\big\}. \tag{3}$$

Note that $Z_i = n^{-\frac{1}{2}}\langle \boldsymbol{\sigma}, X_i\rangle \sim \mathcal{N}(0,1), i \in [d]$ are i.i.d. So,

$$\mathbb{E}[T] = 2^n \mathbb{P}\big[\boldsymbol{\sigma} \in \mathcal{F}(\Xi,\boldsymbol{c})\big] = 2^n \prod_{1\le i\le d} \mathbb{P}\big[|Z_i| \le 2^{-c_i n}\big]. \tag{4}$$

Invoking Lemma 1(a), we obtain

$$\mathbb{P}\big[|Z_i| \le 2^{-c_i n}\big] = \sqrt{\frac{2}{\pi}} 2^{-c_i n}(1 + o_n(1)). \tag{5}$$

Combining (4) and (5), we arrive at

$$\mathbb{E}[T] = \left(\frac{2}{\pi}\right)^{\frac{d}{2}} 2^{n(1-\|\boldsymbol{c}\|_1)}\big(1 + o_n(1)\big). \tag{6}$$

Suppose $\|\boldsymbol{c}\|_1 > 1$. We then have $\mathbb{P}[T \ge 1] \le \mathbb{E}[T] \le \exp\big(-\Theta(n)\big)$, using Markov's inequality, (6), and the fact that $d = O(1)$. With this, we obtain that $\mathcal{F}(\Xi,\boldsymbol{c}) = \varnothing$ w.h.p.

In the remainder of the proof, we assume $\|\boldsymbol{c}\|_1 < 1$. Note first that $\mathbb{E}[T] = \exp\big(\Theta(n)\big)$ per (6). Fix $\epsilon > 0$ and define

$$\mathcal{T}_1 \triangleq \left\{(\boldsymbol{\sigma},\boldsymbol{\sigma}') : \frac{d_H(\boldsymbol{\sigma},\boldsymbol{\sigma}')}{n} \notin \left[\frac{1-\epsilon}{2}, \frac{1+\epsilon}{2}\right] \cup \{0,1\}\right\} \tag{7}$$

$$\mathcal{T}_2 \triangleq \left\{(\boldsymbol{\sigma},\boldsymbol{\sigma}') : \frac{d_H(\boldsymbol{\sigma},\boldsymbol{\sigma}')}{n} \in \left[\frac{1-\epsilon}{2}, \frac{1+\epsilon}{2}\right]\right\}. \tag{8}$$

Namely, any $(\boldsymbol{\sigma},\boldsymbol{\sigma}') \in \mathcal{T}_2$ is nearly orthogonal. We next estimate $\mathbb{E}[T^2]$. Using (3) and the linearity of expectation,

$$\mathbb{E}[T^2] = \sum_{(\boldsymbol{\sigma},\boldsymbol{\sigma}')\in\Sigma_n\times\Sigma_n} \mathbb{P}\big[\boldsymbol{\sigma} \in \mathcal{F}(\Xi,\boldsymbol{c}), \boldsymbol{\sigma}' \in \mathcal{F}(\Xi,\boldsymbol{c})\big]$$

$$= 2\sum_{\boldsymbol{\sigma}\in\Sigma_n} \mathbb{P}\big[\boldsymbol{\sigma} \in \mathcal{F}(\Xi,\boldsymbol{c})\big] + \sum_{(\boldsymbol{\sigma},\boldsymbol{\sigma}')\in\mathcal{T}_1} \mathbb{P}\big[\boldsymbol{\sigma},\boldsymbol{\sigma}' \in \mathcal{F}(\Xi,\boldsymbol{c})\big] \tag{9}$$

$$+ \sum_{(\boldsymbol{\sigma},\boldsymbol{\sigma}')\in\mathcal{T}_2} \mathbb{P}\big[\boldsymbol{\sigma},\boldsymbol{\sigma}' \in \mathcal{F}(\Xi,\boldsymbol{c})\big]. \tag{10}$$

We next control the terms in (9) and (10). To that end, recall $Z_i = n^{-\frac{1}{2}}\langle \boldsymbol{\sigma}, X_i\rangle$ and define $Z_i' \triangleq n^{-\frac{1}{2}}\langle \boldsymbol{\sigma}', X_i\rangle \sim \mathcal{N}(0,1)$.

Terms in (9)**:** We show these terms are negligible. First, $\sum_{\boldsymbol{\sigma}\in\Sigma_n} \mathbb{P}\big[\boldsymbol{\sigma} \in \mathcal{F}(\Xi,\boldsymbol{c})\big] = \mathbb{E}[T]$. As for the second term, fix $(\boldsymbol{\sigma},\boldsymbol{\sigma}') \in \mathcal{T}_1$, and notice that $\boldsymbol{\sigma} \ne \pm\boldsymbol{\sigma}'$. So, $d_H(\boldsymbol{\sigma},\boldsymbol{\sigma}') \in [1, n-1]$ and $\frac{1}{n}|\langle \boldsymbol{\sigma},\boldsymbol{\sigma}'\rangle| \le \frac{n-2}{n}$. We estimate $\mathbb{P}\big[\boldsymbol{\sigma},\boldsymbol{\sigma}' \in \mathcal{F}(\Xi,\boldsymbol{c})\big]$. To that end, observe that (Z_i, Z_i') is a bivariate normal with parameter ρ where $|\rho| = \mathbb{E}[Z_i Z_i'] = \frac{1}{n}|\langle \boldsymbol{\sigma},\boldsymbol{\sigma}'\rangle| \le 1 - \frac{2}{n}$. In particular, $1 - \rho^2 \ge \frac{2}{n}\big(2 - \frac{2}{n}\big) = \Omega(\frac{1}{n})$. So,

$$\mathbb{P}\big[\boldsymbol{\sigma},\boldsymbol{\sigma}' \in \mathcal{F}(\Xi,\boldsymbol{c})\big] = \prod_{1\le i\le d} \mathbb{P}\big[|Z_i| \le 2^{-c_i n}, |Z_i'| \le 2^{-c_i n}\big] \tag{11}$$

$$\le \left(\frac{2}{\pi}\right)^d 2^{-2\|\boldsymbol{c}\|_1 n} O\big(n^{d/2}\big) \tag{12}$$

where (11) uses independence of X_i arising in $\mathcal{F}(\Xi, \boldsymbol{c})$ and (12) uses the bivariate normal bound in Lemma 1(b). Next, we estimate $|\mathcal{T}_1|$ in (7). We have

$$|\mathcal{T}_1| \le 2\cdot 2^n \sum_{0\le k\le \frac{n(1-\epsilon)}{2}} \binom{n}{k} \le 2\cdot \exp_2\left(n + nh\left(\frac{1-\epsilon}{2}\right)\right), \tag{13}$$

where we used Lemma 2(b). Combining (12) and (13), we thus arrive at

$$\begin{aligned}
&\sum_{(\boldsymbol{\sigma},\boldsymbol{\sigma}')\in\mathcal{T}_1} \mathbb{P}\big[\boldsymbol{\sigma},\boldsymbol{\sigma}' \in \mathcal{F}(\Xi,\boldsymbol{c})\big] \\
&\le \left(\frac{2}{\pi}\right)^d O(n^{d/2}) \exp_2\left(n + nh\left(\frac{1-\epsilon}{2}\right) - 2n\|\boldsymbol{c}\|_1\right) \\
&\le \mathbb{E}[T]^2 O(n^{d/2}) \exp_2\left(-n + nh\left(\frac{1-\epsilon}{2}\right)\right) \qquad (14)\\
&= \mathbb{E}[T]^2 \exp\big(-\Theta(n)\big), \qquad (15)
\end{aligned}$$

where (14) uses $\mathbb{E}[T]$ per (6), and (15) uses the fact $d = O(1)$ and $h\left(\frac{1-\epsilon}{2}\right) < 1$.

Term in (10)**:** We now show $\mathbb{E}[T^2]$ is dominated by the sum in (10). Note first that $|\mathcal{T}_2| \le |\Sigma_n|^2 = 2^{2n}$, as $\mathcal{T}_2 \subset \Sigma_n \times \Sigma_n$. Next, fix a $(\boldsymbol{\sigma},\boldsymbol{\sigma}') \in \mathcal{T}_2$ and observe that $\frac{1}{n}\langle \boldsymbol{\sigma},\boldsymbol{\sigma}'\rangle \in [-\epsilon,\epsilon]$. Fix now an $i \in [d]$ and consider the bivariate normal (Z_i, Z_i') with parameter $\rho = \mathbb{E}[Z_i Z_i'] = \frac{1}{n}\langle \boldsymbol{\sigma},\boldsymbol{\sigma}'\rangle \in [-\epsilon,\epsilon]$. We have

$$\mathbb{P}\big[|Z_i| \le 2^{-c_i n}, |Z_i'| \le 2^{-c_i n}\big] \le \frac{2}{\pi\sqrt{1-\epsilon^2}} 2^{-2c_i n}, \tag{16}$$

using the anti-concentration bound per Lemma 1. Using (16) and the independence of X_i, $i \in [d]$,

$$\max_{(\boldsymbol{\sigma},\boldsymbol{\sigma}')\in\mathcal{T}_2} \mathbb{P}\big[\boldsymbol{\sigma},\boldsymbol{\sigma}' \in \mathcal{F}(\Xi,\boldsymbol{c})\big] \le \left(\frac{2}{\pi}\right)^d (1-\epsilon^2)^{-\frac{d}{2}} 2^{-2\|\boldsymbol{c}\|_1 n}. \tag{17}$$

Recalling $|\mathcal{T}_2| \le 2^{2n}$, we thus upper bound (10) by

$$\begin{aligned}
&\sum_{(\boldsymbol{\sigma},\boldsymbol{\sigma}')\in\mathcal{T}_2} \mathbb{P}\big[\boldsymbol{\sigma},\boldsymbol{\sigma}' \in \mathcal{F}(\Xi,\boldsymbol{c})\big] \\
&\le \left(\frac{2}{\pi}\right)^d 2^{2n(1-\|\boldsymbol{c}\|_1)} (1-\epsilon^2)^{-\frac{d}{2}} \le \mathbb{E}[T]^2 (1-\epsilon^2)^{-\frac{d}{2}}, \qquad (18)
\end{aligned}$$

where (18) follows by combining (6) and (17).

Second Moment Method. We apply Paley-Zygmund Inequality: for T, a non-negative integer-valued random variable

$$\mathbb{P}[T \ge 1] \ge \mathbb{E}[T]^2/\mathbb{E}[T^2]. \tag{19}$$

For a simple proof, see e.g. [39]. Combining (9), (10), (15) and (18), we arrive at

$$\mathbb{E}[T^2] \le \mathbb{E}[T] + \mathbb{E}[T]^2 2^{-\Theta(n)} + \mathbb{E}[T]^2 (1-\epsilon^2)^{-\frac{d}{2}}. \tag{20}$$

Observing that $\mathbb{E}[T]$ in (6) is of order $2^{\Theta(n)}$ for $\|\boldsymbol{c}\|_1 < 1$,

$$\begin{aligned}
\liminf_{n\to\infty} \mathbb{P}[T \ge 1] &\ge \liminf_{n\to\infty} \frac{1}{\mathbb{E}[T]^{-1} + 2^{-\Theta(n)} + (1-\epsilon^2)^{-\frac{d}{2}}} \qquad (21)\\
&= (1-\epsilon^2)^{\frac{d}{2}}, \qquad (22)
\end{aligned}$$

where (21) follows by combining (19) and (20) and recalling $d = O(1)$. Since $\epsilon > 0$ is arbitrary, we obtain by sending $\epsilon \to 0$ in (22) and using $\mathbb{P}[T \ge 1] \le 1$ that $\lim_{n\to\infty} \mathbb{P}[T \ge 1] = 1$ for $\|\boldsymbol{c}\|_1 < 1$, completing the proof of Theorem 1. □

C. Proof of Theorem 2

We prove Theorem 2 below. Clearly, it suffices to establish the result for $\epsilon > 0$ small enough. Fix $\boldsymbol{c}, \boldsymbol{c}' \in \mathbb{R}_+^d$ with $\|\boldsymbol{c}\|_1 + \|\boldsymbol{c}'\|_1 > 1$ and let $\gamma^* \in (0, \frac{1}{2})$ be the unique value such that

$$1 + h(\gamma^*) = \|\boldsymbol{c}\|_1 + \|\boldsymbol{c}'\|_1. \tag{23}$$

Fix $\epsilon > 0$ small and introduce the random variable $U_{\gamma^*,\epsilon}$:

$$U_{\gamma^*,\epsilon} = \Big|\Big\{(\boldsymbol{\sigma},\boldsymbol{\sigma}') \in \mathcal{F}(\Xi,\boldsymbol{c})\times\mathcal{F}(\Xi',\boldsymbol{c}') : \frac{d_H(\boldsymbol{\sigma},\boldsymbol{\sigma}')}{n} \le \gamma^* - \epsilon\Big\}\Big|.$$

We first show $\mathbb{E}[U_{\gamma^*,\epsilon}] = \exp(-\Theta(n))$. Note that

$$\begin{aligned}
\mathbb{E}[U_{\gamma^*,\epsilon}] &= \sum_{\substack{\boldsymbol{\sigma},\boldsymbol{\sigma}'\in\Sigma_n \\ d_H(\boldsymbol{\sigma},\boldsymbol{\sigma}')\le\gamma^*-\epsilon}} \mathbb{P}\big[\boldsymbol{\sigma} \in \mathcal{F}(\Xi,\boldsymbol{c})\big]\mathbb{P}\big[\boldsymbol{\sigma} \in \mathcal{F}(\Xi,\boldsymbol{c}')\big] \qquad (24)\\
&\sum_{0\le k\le n(\gamma^*-\epsilon)} 2^n \binom{n}{k} \mathbb{P}\big[\boldsymbol{\sigma} \in \mathcal{F}(\Xi,\boldsymbol{c})\big]\mathbb{P}\big[\boldsymbol{\sigma} \in \mathcal{F}(\Xi,\boldsymbol{c}')\big] \qquad (25)\\
&\le \exp_2\big(n\big(1 + h(\gamma^*-\epsilon) - \|\boldsymbol{c}\|_1 - \|\boldsymbol{c}'\|_1\big) + O(1)\big) \qquad (26)\\
&= \exp\big(-\Theta(n)\big), \qquad (27)
\end{aligned}$$

where (24) uses independence of Ξ and Ξ', (25) uses a simple counting argument, (26) is obtained via Lemma 2(b) and reasoning similar to (5), and finally, (27) follows from (23) and the fact $h(\gamma^* - \epsilon) < h(\gamma^*)$ as $\gamma^* < \frac{1}{2}$.

Next, $\mathbb{P}[U_{\gamma^*,\epsilon} \ge 1] \le \mathbb{E}[U_{\gamma^*,\epsilon}] \to 0$ by Markov's inequality. Noting that $\{U_{\gamma^*,\epsilon} = 0\} = \{d(\boldsymbol{\sigma},\boldsymbol{\sigma}') \ge \gamma^* - \epsilon\}$, we conclude

$$\lim_{n\to\infty} \mathbb{P}\big[d(\boldsymbol{\sigma},\boldsymbol{\sigma}') \ge \gamma^* - \epsilon\big] = 1. \tag{28}$$

To show $d(\boldsymbol{\sigma},\boldsymbol{\sigma}') \le \gamma^* + \epsilon$ w.h.p., introduce $\overline{U}_{\gamma^*,\epsilon}$:

$$\Big|\Big\{(\boldsymbol{\sigma},\boldsymbol{\sigma}') \in \mathcal{F}(\Xi,\boldsymbol{c}) \times \mathcal{F}(\Xi',\boldsymbol{c}') : \frac{d_H(\boldsymbol{\sigma},\boldsymbol{\sigma}')}{n} = \gamma^* + \epsilon\Big\}\Big|, \tag{29}$$

where ϵ is small enough, such that $\gamma^* < \gamma^* + \epsilon < \frac{1}{2}$.

Proposition 1. $\lim_{n\to\infty} \frac{\mathbb{E}[\overline{U}_{\gamma^*,\epsilon}]^2}{\mathbb{E}[\overline{U}^2_{\gamma^*,\epsilon}]} = 1.$

The proof Proposition 1 is involved, see Section V-E. Observe that $\{\overline{U}_{\gamma^*,\epsilon} \ge 1\} \subseteq \{d(\boldsymbol{c},\boldsymbol{c}') \le \gamma^* + \epsilon\}$, hence

$$\liminf_{n\to\infty} \mathbb{P}\big[d(\boldsymbol{c},\boldsymbol{c}') \le \gamma^* + \epsilon\big] \ge \liminf_{n\to\infty} \mathbb{P}\big[\overline{U}_{\gamma^*,\epsilon} \ge 1\big] = 1, \tag{30}$$

where we used Paley-Zygmund Inequality (19) and Proposition 1. Combining (30) with (28), we establish Theorem 2.

D. Proof of Theorem 3

We only prove Part (a); Part (b) follows similarly. The proof is based on the first moment method. Let $\|\boldsymbol{c}\|_1 > \frac{1}{2}$ and $\beta^* \in (0, \frac{1}{2})$ be any value such that $1 + h(\beta^*) < 2\|\boldsymbol{c}\|_1$. For N_{β^*} as in Theorem 3, we show $\mathbb{E}[N_{\beta^*}] = \exp(-\Theta(n))$. To that end, fix any $(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ with $1 \le d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma}') \le \beta^* n$. Note that $\frac{1}{n}|\langle \boldsymbol{\sigma}, \boldsymbol{\sigma}' \rangle| \le 1 - \frac{2}{n}$ as $\boldsymbol{\sigma} \ne \pm \boldsymbol{\sigma}'$. Setting $Z_i = n^{-\frac{1}{2}}\langle \boldsymbol{\sigma}, X_i \rangle$ and $Z_i' = n^{-\frac{1}{2}}\langle \boldsymbol{\sigma}', X_i \rangle$, we observe that $Z_i \sim \mathcal{N}(0,1)$ i.i.d., $Z_i' \sim \mathcal{N}(0,1)$ also i.i.d., and for any $i \in [d]$, (Z_i, Z_i') is a bivariate normal with density upper bounded by $1/(2\pi\sqrt{1 - (\langle \boldsymbol{\sigma}, \boldsymbol{\sigma}' \rangle/n)^2})$ which is $O(n)$. Thus,

$$\begin{aligned}\mathbb{P}\big[\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \mathcal{F}(\Xi, \boldsymbol{c})\big] &= \prod_{1 \le i \le d} \mathbb{P}\big[|Z_i| \le 2^{-c_i n}, |Z_i'| \le 2^{-c_i n}\big] \\ &\le \exp\big(-2\|\boldsymbol{c}\|_1 n + O(\log n)\big) \end{aligned} \tag{31}$$

As (31) is valid for all $(\boldsymbol{\sigma}, \boldsymbol{\sigma}')$ with $1 \le d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma}') \le \beta^* n$,

$$\begin{aligned}\mathbb{E}[N_{\beta^*}] &= \sum_{\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \Sigma_n : 1 \le d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma}') \le \beta^* n} \mathbb{P}\big[\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \mathcal{F}(\Xi, \boldsymbol{c})\big] \\ &\le 2^n \sum_{1 \le k \le \beta^* n} \binom{n}{k} \exp\big(-2\|\boldsymbol{c}\|_1 n + O(\log n)\big) \\ &\le \exp_2\big(n + nh(\beta^*) - 2n\|\boldsymbol{c}\|_1 + O(\log n)\big) = \exp(-\Theta(n)) \end{aligned} \tag{32}$$

where (32) uses Lemma 2(b). Thus $N_{\beta^*} = 0$ w.h.p. by Markov's inequality, which immediately yields that for $\|\boldsymbol{c}\|_1 > \frac{1}{2}$, w.h.p. $\min_{\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \mathcal{F}(\Xi, \boldsymbol{c}), \boldsymbol{\sigma} \ne \boldsymbol{\sigma}'} d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma}') \ge \beta^* n$.

E. Proof of Proposition 1

Theorem 2 is based on Proposition 1, which we prove now. In what follows, suppose that $\epsilon > 0$ is small, so that $\gamma^* + \epsilon < \frac{1}{2}$. In particular,

$$h(\gamma^* + \epsilon) > h(\gamma^*),$$

and therefore

$$h(\gamma^* + \epsilon) + 1 - \|\boldsymbol{c}\|_1 - \|\boldsymbol{c}'\|_1 > 1. \tag{33}$$

Estimating $\mathbb{E}[\overline{U}_{\gamma^*, \epsilon}]$. Fix $\boldsymbol{\sigma} \in \Sigma_n$. Set $Z_i \triangleq n^{-\frac{1}{2}}\langle \boldsymbol{\sigma}, X_i \rangle \sim \mathcal{N}(0,1), i \in [d]$ which are i.i.d., and observe that

$$\begin{aligned}\mathbb{P}\big[\boldsymbol{\sigma} \in \mathcal{F}(\Xi, \boldsymbol{c})\big] &= \prod_{1 \le i \le d} \mathbb{P}\big[|Z_i| \le 2^{-c_i n}\big] && (34) \\ &= \prod_{1 \le i \le d} \left(\frac{2}{\sqrt{2\pi}} 2^{-c_i n}(1 + o_n(1))\right) && (35) \\ &= \left(\frac{2}{\pi}\right)^{d/2} 2^{-\|\boldsymbol{c}\|_1 n}\big(1 + o_n(1)\big), && (36) \end{aligned}$$

where (34) uses independence of X_i, (35) uses Lemma 1(a), and (36) uses the fact $d = O(1)$ as $n \to \infty$. Hence,

$$\begin{aligned}\mathbb{E}\big[\overline{U}_{\gamma^*, \epsilon}\big] &= \sum_{\substack{\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \Sigma_n \\ d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = n(\gamma^* + \epsilon)}} \mathbb{P}\big[\boldsymbol{\sigma} \in \mathcal{F}(\Xi, \boldsymbol{c}), \boldsymbol{\sigma}' \in \mathcal{F}(\Xi', \boldsymbol{c}')\big] \\ &= \sum_{\substack{\boldsymbol{\sigma}, \boldsymbol{\sigma}' \in \Sigma_n \\ d_H(\boldsymbol{\sigma}, \boldsymbol{\sigma}') = n(\gamma^* + \epsilon)}} \mathbb{P}\big[\boldsymbol{\sigma} \in \mathcal{F}(\Xi, \boldsymbol{c})\big]\mathbb{P}\big[\boldsymbol{\sigma}' \in \mathcal{F}(\Xi', \boldsymbol{c}')\big] && (37) \\ &= 2^n \binom{n}{n(\gamma^* + \epsilon)} \left(\frac{2}{\pi}\right)^d e^{-n(\|\boldsymbol{c}\|_1 + \|\boldsymbol{c}'\|_1)}(1 + o_n(1)), && (38) \end{aligned}$$

where (37) uses the independence of Ξ and Ξ' and (38) uses (36). We also record the following esimate: recalling

$$\binom{n}{n(\gamma^* + \epsilon)} = \exp_2\big(h(\gamma^* + \epsilon) + o(n)\big)$$

per Lemma 2(a), we have by using (33) that

$$\mathbb{E}[\overline{U}_{\gamma^*, \epsilon}] = e^{\Theta(n)}. \tag{39}$$

Estimating $\mathbb{E}\big[\overline{U}^2_{\gamma^*, \epsilon}\big]$ Estimating the second moment is more complicated. We first show the following lemma.

Lemma 3. *Let $\mathcal{O} \in (-1, 1)$. Then,*

$$\begin{aligned}&\sup_{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \Sigma_n : \frac{1}{n}\langle \boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \rangle = \mathcal{O}} \mathbb{P}\big[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \mathcal{F}(\Xi, \boldsymbol{c})\big] \\ &\le \left(\frac{2}{\pi}\right)^d (1 - \mathcal{O}^2)^{-\frac{d}{2}} 2^{-2\|\boldsymbol{c}\|_1 n}(1 + o_n(1)). \end{aligned}$$

Proof of Lemma 3. Let $Z_i = n^{-\frac{1}{2}}\langle \boldsymbol{\sigma}_1, X_i \rangle$ and $C_i = n^{-\frac{1}{2}}\langle \boldsymbol{\sigma}_2, X_i \rangle$. Note that $Z_i \sim \mathcal{N}(0,1)$ i.i.d., $C_i \sim \mathcal{N}(0,1)$ also i.i.d., and (Z_i, C_i) is a bivariate normal with parameter $\mathcal{O}$ for every $i \in [d]$. Hence,

$$\begin{aligned}\mathbb{P}\big[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \mathcal{F}(\Xi, \boldsymbol{c})\big] &= \prod_{1 \le i \le d} \mathbb{P}\big[|Z_i| \le 2^{-c_i n}, |C_i| \le 2^{-c_i n}\big] && (40) \\ &= \prod_{1 \le i \le d} \left(\frac{2 \cdot 2^{-2c_i n}}{\pi\sqrt{1 - \mathcal{O}^2}}(1 + o_n(1))\right) && (41) \\ &= \left(\frac{2}{\pi}\right)^d (1 - \mathcal{O}^2)^{-\frac{d}{2}} 2^{-2\|\boldsymbol{c}\|_1 n}(1 + o_n(1)), && (42) \end{aligned}$$

where (40) uses independence of $X_i, i \in [d]$ and (41) uses Lemma 1(b) for the bivariate normal (Z_i, C_i) with $z_1 = z_2 = 2^{-c_i n}$ and (42) uses the fact $d = O(1)$. □

Having shown Lemma 3, we now estimate the second moment. Notice first that

$$\begin{aligned}&\mathbb{E}\big[\overline{U}^2_{\gamma^*, \epsilon}\big] \\ &= \sum_{\substack{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_2' \in \Sigma_n \\ d_H(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_1') = n(\gamma^* + \epsilon) \\ d_H(\boldsymbol{\sigma}_2, \boldsymbol{\sigma}_2') = n(\gamma^* + \epsilon)}} \mathbb{P}\big[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \mathcal{F}(\Xi, \boldsymbol{c}), \boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2' \in \mathcal{F}(\Xi', \boldsymbol{c}')\big]. \end{aligned} \tag{43}$$

Let $\mathcal{Q}$ be a shorthand for the quadruple $(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2, \boldsymbol{\sigma}_2')$ and

$$\mathcal{S} \triangleq \big\{\mathcal{Q} : d_H(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_1') = d_H(\boldsymbol{\sigma}_2, \boldsymbol{\sigma}_2') = n(\gamma^* + \epsilon)\big\} \subset \Sigma_n^4.$$

Fix $\delta > 0$ small and introduce the sets

$$\mathcal{J}(\delta) = \left[\frac{n(1-\delta)}{2}, \frac{n(1+\delta)}{2}\right] \tag{44}$$

$$\mathcal{J}'(\delta) = [1, n-1] \setminus \left[\frac{n(1-\delta)}{2}, \frac{n(1+\delta)}{2}\right]. \tag{45}$$

Partitioning Quadruples. Define the sets of quadruples

$$\mathcal{T}_0 \triangleq \big\{\mathcal{Q} \in \mathcal{S} : (\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_1') = (\pm\boldsymbol{\sigma}_2, \pm\boldsymbol{\sigma}_2')\big\} \tag{46}$$

$$\mathcal{T}_1 \triangleq \big\{\mathcal{Q} \in \mathcal{S} : \boldsymbol{\sigma}_1 = \pm\boldsymbol{\sigma}_2, \boldsymbol{\sigma}_1' \neq \pm\boldsymbol{\sigma}_2'\big\} \tag{47}$$

$$\mathcal{T}_2 \triangleq \big\{\mathcal{Q} \in \mathcal{S} : \boldsymbol{\sigma}_1' = \pm\boldsymbol{\sigma}_2', \boldsymbol{\sigma}_1 \neq \pm\boldsymbol{\sigma}_2\big\}. \tag{48}$$

Let

$$\mathcal{S}' \triangleq \mathcal{S} \setminus \big(\mathcal{T}_0 \cup \mathcal{T}_1 \cup \mathcal{T}_2\big).$$

Note that if $\mathcal{Q} \in \mathcal{S}'$ then $\boldsymbol{\sigma}_1 \neq \pm\boldsymbol{\sigma}_2$ and $\boldsymbol{\sigma}_1' \neq \pm\boldsymbol{\sigma}_2'$. Next, define

$$\mathcal{T}_3 \triangleq \big\{\mathcal{Q} \in \mathcal{S}' : d_H(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \in \mathcal{J}'(\delta) \text{ or } d_H(\boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2') \in \mathcal{J}'(\delta)\big\} \tag{49}$$

$$\mathcal{T}_4 \triangleq \big\{\mathcal{Q} \in \mathcal{S}' : d_H(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2) \in \mathcal{J}(\delta) \text{ and } d_H(\boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2') \in \mathcal{J}(\delta)\big\} \tag{50}$$

where $\mathcal{J}'(\delta)$ and $\mathcal{J}(\delta)$ appear, respectively, in (45) and (44). Note that $\mathcal{T}_i$, $i = 0, \ldots, 4$ partition the set of all quadruples that we investigate, i.e., $\mathcal{S}$. Furthermore, (43) yields

$$\mathbb{E}\big[\overline{U}^2_{\gamma^*,\epsilon}\big] = \sum_{i=0}^{4} \sum_{\mathcal{Q}\in\mathcal{T}_i} \mathbb{P}\big[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \mathcal{F}(\Xi, \boldsymbol{c}), \boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2' \in \mathcal{F}(\Xi', \boldsymbol{c}')\big]. \tag{51}$$

In what follows, we essentially show that the dominant contribution comes from $\mathcal{T}_4$. To that end, we first study $\mathcal{T}_0$.

$$\begin{aligned}
&\sum_{\mathcal{Q}\in\mathcal{T}_0} \mathbb{P}\big[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \mathcal{F}(\Xi, \boldsymbol{c}), \boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2' \in \mathcal{F}(\Xi', \boldsymbol{c}')\big] \\
&= 4 \sum_{\substack{\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_1' \\ d_H(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_1') = n(\gamma^* + \epsilon)}} \mathbb{P}\big[\boldsymbol{\sigma}_1 \in \mathcal{F}(\Xi, \boldsymbol{c}), \boldsymbol{\sigma}_1' \in \mathcal{F}(\Xi', \boldsymbol{c}')\big] \\
&= 4\mathbb{E}[\overline{U}_{\gamma^*,\epsilon}] \leq \mathbb{E}[\overline{U}_{\gamma^*,\epsilon}]^2 e^{-\Theta(n)},
\end{aligned} \tag{52}$$

where the last step uses (39). We next study $\mathcal{T}_1$ (and $\mathcal{T}_2$, from symmetry). Notice that

$$|\mathcal{T}_1| \leq 2 \cdot 2^n \binom{n}{n(\gamma^* + \epsilon)}^2, \tag{53}$$

since there are 2^n choices for $\boldsymbol{\sigma}_1$, 2 choices for $\boldsymbol{\sigma}_2$ (having fixed $\boldsymbol{\sigma}_1$) and at most $\binom{n}{n(\gamma^*+\epsilon)}^2$ for $(\boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2')$. Next, fix any $\mathcal{Q} \in \mathcal{T}_1$. Note that

$$\begin{aligned}
&\max_{\mathcal{Q}\in\mathcal{T}_1} \mathbb{P}\big[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \mathcal{F}(\Xi, \boldsymbol{c}), \boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2' \in \mathcal{F}(\Xi', \boldsymbol{c}')\big] \\
&\leq \mathbb{P}[\boldsymbol{\sigma}_1 \in \mathcal{F}(\Xi, \boldsymbol{c})] \max_{\boldsymbol{\sigma}_1' \neq \pm\boldsymbol{\sigma}_2'} \mathbb{P}\big[\boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2' \in \mathcal{F}(\Xi', \boldsymbol{c}')\big] \qquad (54) \\
&\leq \left(\frac{2}{\pi}\right)^{\frac{3d}{2}} 2^{-\|\boldsymbol{c}\|_1 n} \cdot O(n^{\frac{d}{2}}) 2^{-2\|\boldsymbol{c}'\|_1 n}(1 + o_n(1)) \qquad (55)
\end{aligned}$$

where (54) uses $\boldsymbol{\sigma}_1 = \pm\boldsymbol{\sigma}_2$, and (55) is obtained by combining (36) and Lemma 3 together with the fact $\boldsymbol{\sigma}_1' \neq \pm\boldsymbol{\sigma}_2'$ so that $\frac{1}{n}|\langle\boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2'\rangle| \leq 1 - \frac{1}{n}$. Thus,

$$\begin{aligned}
&\sum_{\mathcal{Q}\in\mathcal{T}_1} \mathbb{P}\big[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \mathcal{F}(\Xi, \boldsymbol{c}), \boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2' \in \mathcal{F}(\Xi', \boldsymbol{c}')\big] \\
&\leq 2^n \binom{n}{n(\gamma^* + \epsilon)}^2 \left(\frac{2}{\pi}\right)^{\frac{3d}{2}} 2^{-\|\boldsymbol{c}\|_1 - 2\|\boldsymbol{c}'\|n} \cdot O(n^{\frac{d}{2}})(1 + o_n(1)) \qquad (56) \\
&\leq \mathbb{E}\big[\overline{U}^2_{\gamma^*,\epsilon}\big] \exp_2\big(-n + n\|\boldsymbol{c}\|_1 + O(\log n)\big) \qquad (57)
\end{aligned}$$

where (56) follows by combining (53) and (55), and (57) follows by recalling (38). As $\|\boldsymbol{c}\|_1 < 1$ by assumption, we thus conclude

$$\begin{aligned}
&\sum_{\mathcal{Q}\in\mathcal{T}_1} \mathbb{P}\big[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \mathcal{F}(\Xi, \boldsymbol{c}), \boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2' \in \mathcal{F}(\Xi', \boldsymbol{c}')\big] \\
&\leq \mathbb{E}\big[\overline{U}^2_{\gamma^*,\epsilon}\big] e^{-\Theta(n)}.
\end{aligned} \tag{58}$$

A similar reasoning also yields

$$\begin{aligned}
&\sum_{\mathcal{Q}\in\mathcal{T}_2} \mathbb{P}\big[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \mathcal{F}(\Xi, \boldsymbol{c}), \boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2' \in \mathcal{F}(\Xi', \boldsymbol{c}')\big] \\
&\leq \mathbb{E}\big[\overline{U}^2_{\gamma^*,\epsilon}\big] e^{-\Theta(n)}.
\end{aligned} \tag{59}$$

We now focus on the sum over $\mathcal{T}_3$ per (49). We first upper bound the cardinality of $\mathcal{T}_3$. Recall $\mathcal{J}'(\delta)$ from (45). Note that there are 2^n choices for $\boldsymbol{\sigma}_1$, and having fixed a $\boldsymbol{\sigma}_1$, there are

$$\sum_{k\in\mathcal{J}'(\delta)} \binom{n}{k} \leq \exp_2\left(nh\left(\frac{1-\delta}{2}\right)\right) \tag{60}$$

choices for $\boldsymbol{\sigma}_2$, where we used Lemma 2(b). Having fixed $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$, there are at most

$$\binom{n}{n(\gamma^* + \epsilon)}^2 \tag{61}$$

choices for $(\boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2')$ subject to $d_H(\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_1') = d_H(\boldsymbol{\sigma}_2, \boldsymbol{\sigma}_2') = n(\gamma^* + \epsilon)$. Combining (60) and (61), we thus obtain

$$|\mathcal{T}_3| \leq 2 \cdot \exp_2\left(n + nh\left(\frac{1-\delta}{2}\right)\right) \binom{n}{n(\gamma^* + \epsilon)}^2, \tag{62}$$

where the extra factor follows by repeating the same calculation for those quadruples with $(\boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2') \in \mathcal{J}'(\delta)$. We next control the probability term.

$$\begin{aligned}
&\max_{\mathcal{Q}\in\mathcal{T}_3} \mathbb{P}\big[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \mathcal{F}(\Xi, \boldsymbol{c}), \boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2' \in \mathcal{F}(\Xi', \boldsymbol{c}')\big] \\
&\leq \max_{\boldsymbol{\sigma}_1 \neq \pm\boldsymbol{\sigma}_2} \mathbb{P}\big[\boldsymbol{\sigma}_1, \boldsymbol{\sigma}_2 \in \mathcal{F}(\Xi, \boldsymbol{c})\big] \max_{\boldsymbol{\sigma}_1' \neq \pm\boldsymbol{\sigma}_2'} \mathbb{P}\big[\boldsymbol{\sigma}_1', \boldsymbol{\sigma}_2' \in \mathcal{F}(\Xi', \boldsymbol{c}')\big] \qquad (63) \\
&\leq \left(\frac{2}{\pi}\right)^{2d} O(n^d) \cdot 2^{-2n(\|\boldsymbol{c}\|_1 + \|\boldsymbol{c}'\|_1)} \cdot \big(1 + o_n(1)\big), \qquad (64)
\end{aligned}$$

where (63) uses the fact Ξ and Ξ' are independent and (64) follows by using Lemma 3 and using identical reasoning outlined after (55) above. Combining (62) and (64), we obtain

$$\begin{aligned}&\sum_{\mathcal{Q}\in\mathcal{T}_3}\mathbb{P}\big[\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\mathcal{F}(\Xi,\boldsymbol{c}),\boldsymbol{\sigma}_1',\boldsymbol{\sigma}_2'\in\mathcal{F}(\Xi',\boldsymbol{c}')\big]\\&\leq\binom{n}{n(\gamma^*+\epsilon)}^2\times\\&\exp_2\left(n\left(1+h\left(\frac{1-\delta}{2}\right)-\|\boldsymbol{c}\|_1-\|\boldsymbol{c}'\|_1\right)+O(\log n)\right),\end{aligned}\tag{65}$$

using the fact $d=O(1)$. Using (38), we thus conclude that

$$\begin{aligned}&\sum_{\mathcal{Q}\in\mathcal{T}_3}\mathbb{P}\big[\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\mathcal{F}(\Xi,\boldsymbol{c}),\boldsymbol{\sigma}_1',\boldsymbol{\sigma}_2'\in\mathcal{F}(\Xi',\boldsymbol{c}')\big]\\&\leq\mathbb{E}\big[\overline{U}^2_{\gamma^*,\epsilon}\big]e^{-\Theta(n)}.\end{aligned}\tag{66}$$

We lastly study $\mathcal{T}_4$ in (50) and show that the quadruples from this set brings the dominant contribution to the second moment. We first bound

$$|\mathcal{T}_4|\leq 2^{2n}\binom{n}{n(\gamma^*+\epsilon)}^2,\tag{67}$$

where there are 2^n choices for $\boldsymbol{\sigma}_1$ and 2^n choices for $\boldsymbol{\sigma}_2$; having fixed $\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2$, there are $\binom{n}{n(\gamma^*+\epsilon)}$ choices for each of $\boldsymbol{\sigma}_1'$ and $\boldsymbol{\sigma}_2'$. This bound is crude, but suffices for our purposes. We next fix

$$\mathcal{Q}=(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_1',\boldsymbol{\sigma}_2,\boldsymbol{\sigma}_2')\in\mathcal{T}_4.$$

Using $\mathcal{T}_4$ from (50) as well as the set $\mathcal{J}(\delta)$ per (44), we have

$$d_H(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2),d_H(\boldsymbol{\sigma}_1',\boldsymbol{\sigma}_2')\in\left[\frac{n(1-\delta)}{2},\frac{n(1+\delta)}{2}\right].$$

Since $\frac{1}{n}\langle\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\rangle=\frac{1}{n}(n-2d_H(\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2))$, we obtain

$$\frac{1}{n}\langle\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\rangle\in[-\delta,\delta]\quad\text{and}\quad\frac{1}{n}\langle\boldsymbol{\sigma}_1',\boldsymbol{\sigma}_2'\rangle\in[-\delta,\delta].\tag{68}$$

Next, observe by using Lemma 3 and (68) that

$$\begin{aligned}&\max_{\mathcal{Q}\in\mathcal{T}_4}\mathbb{P}\big[\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\mathcal{F}(\Xi,\boldsymbol{c}),\boldsymbol{\sigma}_1',\boldsymbol{\sigma}_2'\in\mathcal{F}(\Xi',\boldsymbol{c}')\big]\\&\leq\max_{\mathcal{Q}\in\mathcal{T}_4}\big(\mathbb{P}\big[\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\mathcal{F}(\Xi,\boldsymbol{c})\big]\mathbb{P}\big[\boldsymbol{\sigma}_1',\boldsymbol{\sigma}_2'\in\mathcal{F}(\Xi',\boldsymbol{c}')\big]\big)\\&\leq\left(\frac{2}{\pi}\right)^{2d}\big(1-\delta^2\big)^{-d}2^{-2n(\|\boldsymbol{c}\|_1+\|\boldsymbol{c}'\|_1)}\big(1+o_n(1)\big).\end{aligned}\tag{69}$$

Combining (67) and (69), we arrive at

$$\begin{aligned}&\sum_{\mathcal{Q}\in\mathcal{T}_4}\mathbb{P}\big[\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\mathcal{F}(\Xi,\boldsymbol{c}),\boldsymbol{\sigma}_1',\boldsymbol{\sigma}_2'\in\mathcal{F}(\Xi',\boldsymbol{c}')\big]\\&\leq\left(\frac{2}{\pi}\right)^{2d}\big(1-\delta^2\big)^{-d}\binom{n}{n(\gamma^*+\epsilon)}^2(1+o_n(1))\\&\times\exp_2\big(2n\big(1-\|\boldsymbol{c}\|_1-\|\boldsymbol{c}'\|_1\big)\big).\end{aligned}\tag{70}$$

Combining (70) with the expression for $\mathbb{E}\big[\overline{U}^2_{\gamma^*,\epsilon}\big]$ per (38), we thus obtain

$$\begin{aligned}&\sum_{\mathcal{Q}\in\mathcal{T}_4}\mathbb{P}\big[\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\mathcal{F}(\Xi,\boldsymbol{c}),\boldsymbol{\sigma}_1',\boldsymbol{\sigma}_2'\in\mathcal{F}(\Xi',\boldsymbol{c}')\big]\\&\leq\mathbb{E}\big[\overline{U}^2_{\gamma^*,\epsilon}\big]\big(1-\delta^2\big)^{-d}\end{aligned}\tag{71}$$

We are now ready to conclude the proof of Proposition 1. Observe that

$$\frac{\mathbb{E}[\overline{U}_{\gamma^*,\epsilon}]^2}{\mathbb{E}[\overline{U}^2_{\gamma^*,\epsilon}]}\geq\frac{1}{e^{-\Theta(n)}+(1-\delta^2)^{-d}}$$

by combining (51) with (52), (58), (59), (66) and (71). Sending $n\to\infty$ (while keeping $\delta>0$ a constant)

$$\liminf_{n\to\infty}\frac{\mathbb{E}[\overline{U}_{\gamma^*,\epsilon}]^2}{\mathbb{E}[\overline{U}^2_{\gamma^*,\epsilon}]}\geq\big(1-\delta^2\big)^d.\tag{72}$$

Since $\delta>0$ above is arbitrary and the left hand side is independent of δ, we thus have, upon sending $\delta\to0$, that

$$\liminf_{n\to\infty}\frac{\mathbb{E}[\overline{U}_{\gamma^*,\epsilon}]^2}{\mathbb{E}[\overline{U}^2_{\gamma^*,\epsilon}]}\geq1.$$

As

$$\frac{\mathbb{E}[\overline{U}_{\gamma^*,\epsilon}]^2}{\mathbb{E}[\overline{U}^2_{\gamma^*,\epsilon}]}\leq1$$

trivially by Jensen's inequality, we also have

$$\limsup_{n\to\infty}\frac{\mathbb{E}[\overline{U}_{\gamma^*,\epsilon}]^2}{\mathbb{E}[\overline{U}^2_{\gamma^*,\epsilon}]}\leq1,$$

and therefore

$$\lim_{n\to\infty}\frac{\mathbb{E}[\overline{U}_{\gamma^*,\epsilon}]^2}{\mathbb{E}[\overline{U}^2_{\gamma^*,\epsilon}]}=1,$$

establishing Proposition 1.

REFERENCES

[1] A. M. Krieger, D. Azriel, and A. Kapelner, "Nearly random designs with greatly improved balance," *Biometrika*, vol. 106, no. 3, pp. 695–701, 2019.

[2] C. Harshaw, F. Sävje, D. Spielman, and P. Zhang, "Balancing covariates in randomized experiments using the gram-schmidt walk," *arXiv preprint arXiv:1911.03071*, 2019.

[3] P. Turner, R. Meka, and P. Rigollet, "Balancing Gaussian vectors in high dimension," in *Conference on Learning Theory*. PMLR, 2020, pp. 3455–3486.

[4] D. Gamarnik and E. C. Kızıldağ, "Algorithmic obstructions in the random number partitioning problem," *The Annals of Applied Probability*, vol. 33, no. 6B, pp. 5497–5563, 2023.

[5] E. C. Kızıldağ, "Planted random number partitioning problem," *arXiv preprint arXiv:2309.15115*, 2023.

[6] Y. Wang and X. Li, "Rerandomization with diminishing covariate imbalance and diverging number of covariates," *The Annals of Statistics*, vol. 50, no. 6, pp. 3439–3465, 2022.

[7] A. Kapelner, A. M. Krieger, M. Sklar, U. Shalit, and D. Azriel, "Harmonizing optimized designs with classic randomization in experiments," *The American Statistician*, vol. 75, no. 2, pp. 195–206, 2021.

[8] A. Kapelner, A. M. Krieger, M. Sklar, and D. Azriel, "Optimal rerandomization designs via a criterion that provides insurance against failed experiments," *Journal of Statistical Planning and Inference*, vol. 219, pp. 63–84, 2022.

[9] N. Kallus, "Optimal a priori balance in the design of controlled experiments," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 80, no. 1, pp. 85–112, 2018.

[10] D. Bertsimas, M. Johnson, and N. Kallus, "The power of optimization over randomization in designing experiments involving small samples," *Operations Research*, vol. 63, no. 4, pp. 868–876, 2015.

[11] R. Merkle and M. Hellman, "Hiding information and signatures in trapdoor knapsacks," *IEEE transactions on Information Theory*, vol. 24, no. 5, pp. 525–530, 1978.

[12] E. G. Coffman and G. S. Lueker, *Probabilistic analysis of packing and partitioning algorithms*. Wiley-Interscience, 1991.

[13] L.-H. Tsai, "Asymptotic analysis of an algorithm for balanced parallel processor scheduling," *SIAM Journal on Computing*, vol. 21, no. 1, pp. 59–64, 1992.

[14] R. M. Karp, *Reducibility among combinatorial problems*. Springer, 2010.

[15] M. R. Garey and D. S. Johnson, *Computers and Intractability; A Guide to the Theory of NP-Completeness*. USA: W. H. Freeman & Co., 1990.

[16] S. Mertens, "Phase transition in the number partitioning problem," *Physical Review Letters*, vol. 81, no. 20, p. 4281, 1998.

[17] 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, 2001.

[18] H. Bauke and S. Mertens, "Universality in the level statistics of disordered systems," *Physical Review E*, vol. 70, no. 2, p. 025102, 2004.

[19] C. Borgs, J. Chayes, S. Mertens, and C. Nair, "Proof of the local rem conjecture for number partitioning. i: Constant energy scales," *Random Structures & Algorithms*, vol. 34, no. 2, pp. 217–240, 2009.

[20] ——, "Proof of the local rem conjecture for number partitioning. ii. growing energy scales," *Random Structures & Algorithms*, vol. 34, no. 2, pp. 241–284, 2009.

[21] J. Spencer, "Six standard deviations suffice," *Transactions of the American mathematical society*, vol. 289, no. 2, pp. 679–706, 1985.

[22] J. Matousek, *Geometric discrepancy: An illustrated guide*. Springer Science & Business Media, 1999, vol. 18.

[23] B. Chazelle, *The Discrepancy Method: Randomness and Complexity*. Cambridge University Press, 2000.

[24] N. Bansal, "Constructive algorithms for discrepancy minimization," in *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. IEEE, 2010, pp. 3–10.

[25] S. Lovett and R. Meka, "Constructive discrepancy minimization by walking on the edges," *SIAM Journal on Computing*, vol. 44, no. 5, pp. 1573–1582, 2015.

[26] A. Levy, H. Ramadas, and T. Rothvoss, "Deterministic discrepancy minimization via the multiplicative weight update method," in *International Conference on Integer Programming and Combinatorial Optimization*. Springer, 2017, pp. 380–391.

[27] T. Rothvoss, "Constructive discrepancy minimization for convex sets," *SIAM Journal on Computing*, vol. 46, no. 1, pp. 224–234, 2017.

[28] 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.

[29] K. P. Costello, "Balancing gaussian vectors," *Israel Journal of Mathematics*, vol. 172, no. 1, pp. 145–156, 2009.

[30] W. Perkins and C. Xu, "Frozen 1-RSB structure of the symmetric Ising perceptron," in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, 2021, pp. 1579–1588.

[31] E. Abbe, S. Li, and A. Sly, "Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron," *arXiv preprint arXiv:2102.13069*, 2021.

[32] D. Gamarnik, E. C. Kızıldağ, W. Perkins, and C. Xu, "Algorithms and barriers in the symmetric binary perceptron model," in *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2022, pp. 576–587.

[33] D. Gamarnik, E. C. Kizildağ, W. Perkins, and C. Xu, "Geometric barriers for stable and online algorithms for discrepancy minimization," in *The Thirty Sixth Annual Conference on Learning Theory*. PMLR, 2023, pp. 3231–3263.

[34] E. C. Kızıldağ, "Algorithms and algorithmic barriers in high-dimensional statistics and random combinatorial structures," Ph.D. dissertation, Massachusetts Institute of Technology, 2022.

[35] 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, 2019.

[36] N. Bansal and J. Spencer, "On-line balancing of random inputs," *Random Structures and Algorithms*, vol. 57, no. 4, pp. 879–891, Dec. 2020.

[37] E. Abbe, S. Li, and A. Sly, "Binary perceptron: efficient algorithms can find solutions in a rare well-connected cluster," *arXiv preprint arXiv:2111.03084*, 2021.

[38] E. C. Kızıldağ and T. Wakhare, "Symmetric perceptron with random labels," in *2023 International Conference on Sampling Theory and Applications (SampTA)*. IEEE, 2023, pp. 1–5.

[39] N. Alon and J. H. Spencer, *The probabilistic method*. John Wiley & Sons, 2016.

[40] Y. Polyanskiy and Y. Wu, "Lecture notes on information theory," *Lecture Notes for ECE563 (UIUC) and 6.441 (MIT)*, pp. 2012–2016, 2016.

[41] D. Achlioptas and A. Coja-Oghlan, "Algorithmic barriers from phase transitions," in *2008 49th Annual IEEE Symposium on Foundations of Computer Science*. IEEE, 2008, pp. 793–802.

[42] D. Gamarnik, A. Jagannath, and E. C. Kızıldağ, "Shattering in the ising pure p-spin model," *arXiv preprint arXiv:2307.07461*, 2023.

[43] D. Gamarnik, "The overlap gap property: A topological barrier to optimizing over random structures," *Proceedings of the National Academy of Sciences*, vol. 118, no. 41, 2021.

[44] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006.

[45] D. Galvin, "Three tutorial lectures on entropy and counting," *arXiv preprint arXiv:1406.7872*, 2014.