

# Smoothed Complexity of 2-player Nash Equilibria

Shant Boodaghians\*, Joshua Brakensiek†, Samuel B. Hopkins‡, Aviad Rubinfeld†

\*University of Illinois at Urbana-Champaign  
boodagh2@illinois.edu

†Stanford University

{jbrakens,aviad}@cs.stanford.edu

‡University of California, Berkeley  
hopkins@berkeley.edu

**Abstract**—We prove that computing a Nash equilibrium of a two-player ( $n \times n$ ) game with payoffs in  $[-1, 1]$  is PPAD-hard (under randomized reductions) even in the smoothed analysis setting, smoothing with noise of constant magnitude. This gives a strong negative answer to conjectures of Spielman and Teng [ST06] and Cheng, Deng, and Teng [CDT09].

In contrast to prior work proving PPAD-hardness after smoothing by noise of magnitude  $1/\text{poly}(n)$  [CDT09], our smoothed complexity result is *not* proved via hardness of approximation for Nash equilibria. This is by necessity, since Nash equilibria can be approximated to constant error in quasi-polynomial time [LMM03]. Our results therefore separate smoothed complexity and hardness of approximation for Nash equilibria in two-player games.

The key ingredient in our reduction is the use of a random zero-sum game as a gadget to produce two-player games which remain hard even after smoothing. Our analysis crucially shows that all Nash equilibria of random zero-sum games are far from pure (with high probability), and that this remains true even after smoothing.

**Keywords**—Nash equilibrium; PPAD; zero-sum game; smoothed analysis; smoothed complexity; anticoncentration

## I. INTRODUCTION

Nash equilibrium is the central solution concept in game theory. Computational complexity results establishing the intractability of Nash equilibrium [SV06], [CDT09], [DGP09] suggest that players that are even mildly computationally bounded may not be able to converge to a Nash equilibrium in the worst case. However, the fragility of these intractable game constructions, together with the fact that random games are tractable [BVV07], have led experts to conjecture that Nash equilibrium should have smoothed polynomial time algorithms (e.g. [ST06] Conjecture 15, and [CDT09] Conjecture 2). If these conjectures were true, they could explain why players in realistic games can converge to equilibrium. In this paper, we prove that even with aggressive smoothing perturbations of constant magnitude, finding a Nash equilibrium continues to be PPAD-complete (under randomized reductions).

Shant is supported by NSF grant CCF-1750436.

Joshua is supported by an NSF Graduate Research Fellowship.

Samuel is supported by a Miller Postdoctoral Fellowship.

## Definition 1 ( $X$ -SMOOTHED-NASH).

For a distribution  $X$  on  $\mathbb{R}$  and problem size  $n$ , fix worst-case  $n \times n$  matrices  $W_A, W_B$  with entries in  $[-1, 1]$ , and let  $N_A, N_B$  be  $n \times n$  matrices whose entries are drawn i.i.d. from  $X$ .  $X$ -SMOOTHED-NASH is the problem of computing, with probability<sup>1</sup> at least  $1 - \frac{1}{n}$ , a Nash equilibrium of the game  $(W_A + N_A, W_B + N_B)$ .

## Theorem 1 (Main Theorem).

There exists a universal constant  $\epsilon > 0$ , such that for any probability distribution  $X$  on  $[-\epsilon, \epsilon]$ ,  $X$ -SMOOTHED-NASH is PPAD-hard under a randomized reduction.<sup>2</sup>

## A. Complexity context: smoothed analysis vs hardness of approximation

In their 2006 survey on smoothed analysis, Spielman and Teng posed the challenge ([ST06], Open Question 11) of exploring the connections between smoothed complexity and hardness of approximation. Concretely, they considered the example of two-player Nash equilibrium subject to  $\sigma$ -bounded perturbations: Given a hard game  $A, B \in [-1, 1]^{n \times n}$ , perturbing each entry independently gives rise to a new instance  $\hat{A}, \hat{B} \in [-1 - \sigma, 1 + \sigma]^{n \times n}$ ; any Nash equilibrium of  $\hat{A}, \hat{B}$  is an  $O(\sigma)$ -approximate-Nash equilibrium of the original game  $A, B$ . Hence, solving Nash equilibrium in the smoothed model is at least as hard as approximating Nash [ST06, Proposition 9.12].

More generally, any hard instance  $x$  of any computational

<sup>1</sup>Amplifying the success probability of smoothed algorithms is generally non-trivial (and sometimes impossible). We note that our hardness result continues to hold even for algorithms that are only required to succeed with probability  $o(1)$ .

<sup>2</sup>Formally, we assume that there is a (single-dimensional) distribution  $X'$  such that  $\Pr_{x \sim X, x' \sim X'}(|x - x'| > 1/\text{poly}(n)) \leq 1/\text{poly}(n)$  and  $X'$  can be sampled by randomized polynomial-time algorithm. This holds for any natural smoothing distribution – e.g. truncated Gaussian or uniform. For arbitrary  $X$  such  $X'$  can be sampled by a randomized algorithm which receives as input a  $\text{poly}(n)$ -size approximation of the CDF of  $X$ . For  $X$  where such advice is necessary, our arguments show that  $X$ -SMOOTHED-NASH is PPAD-hard under randomized reductions with polynomial-length advice.

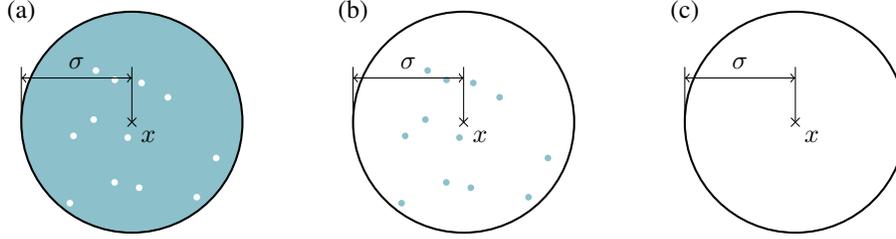


Figure 1. The  $\sigma$ -neighborhood of an intractable instance  $x$ . Tractable instances are colored. (1-a) **smoothed-algorithmica**: almost all instances in the  $\sigma$ -neighborhood are tractable, (1-b) **smoothed-complexity**: very few instances in the  $\sigma$ -neighborhood are tractable; (1-c) **hardness of approximation**: it is intractable to find a solution to any instance in the  $\sigma$ -neighborhood.

problem<sup>3</sup> (e.g.  $x = (A, B)$  in the case of Nash) can be in one of three states (as illustrated in Figure 1):

**Smoothed-algorithmica**<sup>4</sup>: Most instances in  $x$ 's neighborhood can be solved efficiently.

**Smoothed-complexity**: A small fraction of  $x$ 's neighborhood can be solved efficiently.

**Hardness-of-approximation**: Finding a solution for any instance in  $x$ 's neighborhood is intractable.<sup>5</sup>

Of course, as in Spielman and Teng's proposition, hardness-of-approximation immediately rules out efficient smoothed algorithms. But most interesting open problems in smoothed analysis admit approximation algorithms; this limits the applicability of using hardness-of-approximation to prove new smoothed-complexity results.

In contrast to the thriving literature on hardness of approximation and smoothed algorithms, smoothed complexity results are rare. In this paper, we make a small step toward establishing a theory of smoothed complexity, in the context of Spielman and Teng's original example: two-player Nash equilibrium subject to bounded perturbations<sup>6</sup>. While settling an open problem in equilibrium computation, we believe that our result is just the tip of the iceberg of the theory of smoothed complexity.

## B. Historical context

In 1928 Von Neumann [Neu28] proved that every (finite, perfect information) zero-sum game has an equilibrium; this result was extended to general games by Nash in

<sup>3</sup>Naturally, the correspondence between approximation algorithms and smoothed analysis requires matching the respective notions of approximation and smoothing perturbations.

<sup>4</sup>The name is "smoothed-algorithmica" is a cultural reference to Impagliazzo's five worlds [Imp95] and should not be interpreted to imply a technical connection. Formally speaking, the existence of smoothed-algorithmica yet intractable instances is in fact consistent with any of Impagliazzo's worlds *except* Algorithmica.

<sup>5</sup>Intuitively we would like to say that every instance in  $x$ 's neighborhood is intractable. Formally, however, this may be inaccurate. In fact in the case of Nash equilibrium it is provably false! Given game  $(A, B)$ , consider game  $(A', B)$ , where  $A' - A$  is a matrix whose entries are identically equal to some small  $\lambda$  which encodes a Nash equilibrium for  $(A, B)$  (and hence also for  $(A', B)$ ).

<sup>6</sup>Spielman and Teng discuss perturbing each entry by a uniform- $[-\sigma, \sigma]$  noise, but our result holds for any bounded i.i.d. perturbations

1951 [Nas51]. In 1947 Dantzig [Dan98] designed the simplex algorithm for solving linear programs (and thus also zero-sum games); in 1964 Lemke and Howson [LH64] gave a simplex-like algorithm for general games. Both are known to take exponential time in the worst case [KM72], [SV06], but are observed to perform much better in practice (e.g. [Sha87], [ARSvS10]).

For linear programming, Khachiyan [Kha79] gave the first polynomial time algorithm in 1979, and Spielman and Teng proved in 2004 [ST04] that the simplex algorithm has smoothed polynomial complexity. It was natural to hope (and in fact quite widely believed, e.g. [DGP05] and [ST06, Conjecture 9.51] respectively) that the last two results would again be extended to general games. Surprisingly, this was ruled out by Chen, Deng, and Teng [CDT09]. Specifically, they showed that  $1/\text{poly}(n)$ -approximate Nash equilibrium is hard, which by Spielman and Teng's proposition rules out any smoothed efficient algorithms for noise magnitude  $1/\text{poly}(n)$  (assuming PPAD is not contained in search-RP). Chen, Deng, and Teng nevertheless conjectured that for constant magnitude noise, two-player Nash equilibrium should have a polynomial time algorithm.

Progress on smoothed complexity of Nash with  $\varepsilon$ -noise (for small constant  $\varepsilon > 0$ ) was made by [Rub16] who proved the following hardness of approximation result: assuming the "Exponential Time Hypothesis for PPAD<sup>7</sup>", finding an  $\varepsilon$ -approximate Nash equilibrium requires quasipolynomial ( $\approx n^{\log(n)}$ ) time. By Spielman and Teng's proposition, this hardness of approximation result also implies an analogous quasipolynomial hardness in the smoothed setting. For hardness of approximation, the result of [Rub16] is essentially optimal due to a matching quasipolynomial time approximation algorithm [LMM03]. This quasipolynomial time algorithm does not extend to the smoothed case, and a large gap in the complexity of constant-smoothed Nash (quasipolynomial vs exponential) remained open.

In this work, we resolve the complexity of two-player Nash equilibrium with constant-magnitude smoothing, prov-

<sup>7</sup>The Exponential Time Hypothesis (ETH) for PPAD is a strengthening of PPAD  $\not\subseteq$  (search-)RP, which postulates that End-of-Line (the canonical PPAD-problem) requires  $2^{\Omega(n)}$  time.

ing that it is PPAD-complete (under randomized reductions). Compared to [Rub16], we rule out smoothed polynomial time algorithms under a much weaker assumption (PPAD  $\not\subseteq$  (search-)RP vs ETH for PPAD).<sup>8</sup> Alternatively, comparing both results under the same assumption, ETH for PPAD, we prove a much stronger lower bound on the running time ( $2^{\text{poly}(n)}$  vs  $\approx n^{\log(n)}$ ).<sup>9</sup> Finally, another advantage of our result compared to [Rub16] is that our proof is much simpler, and in particular does not require any PCP-like machinery.

### C. Intuition and roadmap

We will reduce  $1/\text{poly}(n)$ -approximate Nash to  $X$ -SMOOTHED-NASH. The starting point of our reduction is the following simple idea: for any mixed strategies  $(x, y)$  which are spread over a large number of actions, the noise from the smoothing averages out. In contrast, if we start with an off-the-shelf PPAD-hard game  $(P, Q)$  and amplify it by simple repetition (formally, tensor the payoff matrices  $P, Q$  with the all ones matrix  $J$ ), the signal from  $P, Q$  will remain strong even with respect to well-spread strategies. This means that given a well-spread (in a sense we make precise later) Nash equilibrium  $x, y$  for a tensored, smoothed game  $(P \otimes J + N_P, Q \otimes J + N_Q)$ , we can recover a  $1/\text{poly}(n)$ -approximate equilibrium for  $(P, Q)$ .

There is one major problem with the reduction suggested above: an oracle for  $X$ -SMOOTHED-NASH might not return a well-spread equilibrium  $(x, y)$ . Our goal henceforth is to modify this construction to create a game where no Nash equilibrium has strategies concentrated on a small number of actions. Note that pure or even small-support equilibria don't break only our proof approach: they can be found efficiently by brute-force enumeration, so such games cannot be hard.

Which games have no strategies concentrated on a small number of actions? At one extreme, if the entries of the payoff matrices are entirely i.i.d. (from any continuous distribution), a folklore result states that the game has a pure equilibrium with probability approaching  $1 - 1/e$ . This creates a significant problem: we have to work with games where the entries are smoothed with independent noise – if such games turn out also to have pure or small-support strategies, then they cannot be hard.

In contrast to i.i.d. random games, we observe that random zero-sum games tend to have only well spread equilibria [Rob06], [Jon04]. For example, they are exponentially unlikely to have a pure equilibrium; intuitively, if a pure strategy profile is exceptionally good for one player, it is likely exceptionally bad for the other. In the context of our proof approach, another advantage of random zero-sum

<sup>8</sup>As discussed above, this holds in the case that  $X$  is approximately polynomial-time sampleable – otherwise we require the assumption PPAD  $\not\subseteq$  (search-)RP/poly.

<sup>9</sup>In fact, under the plausible hypothesis that the true complexity of End-of-Line (the canonical PPAD-problem) is  $\approx 2^{n^\alpha}$  for some constant  $0 < \alpha \leq 1/2$ , our result implies the qualitatively-same strong lower bound, and the result of [Rub16] completely breaks.

games is that with respect to well-spread mixed strategies, they will also average out. That is, even if we add a random zero-sum game  $Z$ , we can still hope to recover a  $1/\text{poly}(n)$  Nash equilibrium for  $(P, Q)$  from a well-spread equilibrium for  $(P \otimes J + Z + N_P, Q \otimes J - Z + N_Q)$ . Our main technical task is to show that adding a random zero-sum game in this fashion produces a game with only well-spread Nash equilibria, even in the presence of the i.i.d. smoothing  $N_A, N_B$ .

*Our first step is to rule out all small support equilibria.* In Section IV we formalize the above intuition, showing that every equilibrium of a random zero-sum game has large supports, even when we add constant-magnitude perturbations. For technical reasons, our proof in this section works for random zero-sum games whose entries are drawn uniformly from discrete  $\{-1, 1\}$ .

*Our second step is to obtain a robust version of no-small-support.* Namely, building on the fact every equilibrium has large support, in Section V we prove that it must be well-spread (formally, the mixed strategies have small  $\|\cdot\|_2$  norm). For technical reasons, our proof in this section works for random zero-sum games whose entries are drawn uniformly from continuous  $[-1, 1]$ . Fortunately, we can make both of proofs work simultaneously by taking the sum of a  $\{-1, 1\}$  and a  $[-1, 1]$  zero-sum games.

*Putting it all together.* To summarize, our final construction of hard instance is given by:

$$\begin{aligned} A &:= P \otimes J + \underbrace{Z_{\{-1,1\}}}_{\text{large support}} + \underbrace{Z_{[-1,1]}}_{\text{well-spread}} \\ B &:= \underbrace{Q \otimes J}_{\text{PPAD-hard}} - \underbrace{Z_{\{-1,1\}}}_{\text{large support}} - \underbrace{Z_{[-1,1]}}_{\text{well-spread}}, \end{aligned}$$

where  $Z_{\{-1,1\}}, Z_{[-1,1]}$  are random matrices with i.i.d. entries uniformly sampled from  $\{-1, 1\}$  and  $[-1, 1]$  (respectively), and  $(P, Q)$  is a PPAD-hard bimatrix game, and  $J$  is an (appropriate-dimension) all-ones matrix.

In Section III, we show that when the Nash equilibrium strategies are well-spread, the random zero-sum games and random perturbations average out. Thanks to the amplification, the signal from  $(P, Q)$  remains sufficiently strong. Thus, we can map any Nash equilibrium of  $(A, B)$  to a  $1/\text{poly}(n)$ -approximate Nash equilibrium of  $(P, Q)$ . By [CDT09] this suffices to establish PPAD-hardness (under randomized reductions).

**Remark** (Inverse-polynomial signal-to-noise ratio).

Interestingly, the amplification of  $(P, Q)$  by repetition is so powerful that our proof would go through even if we were to multiply  $P$  and  $Q$  by an inverse-polynomial small scalar.<sup>10</sup> In this sense, we show that Nash remains intractable even subject to noise (zero-sum + i.i.d.) that is *polynomially larger* than the worst-case signal.

<sup>10</sup>We only informally state the result to prioritize simplicity, but it will be evident by the remarks in Section II-A.

#### D. Additional related work

Subsequent to the seminal works of [DGP09], [CDT09] which showed that Nash equilibrium is PPAD-complete, there has been an active line of work on algorithms with provable guarantees for exact or approximate equilibria in special cases including: sparse games [Bar18], low-rank games [KT10], [AGMS11], positive-semidefinite games [ALSV13], anonymous games [DP15], [CDS17], tree games [EGG06], [BLP15], [OI16]. Complexity limitations for most of these special cases known as well: sparse games [CDT06], [LS18], low-rank games [Meh14], anonymous games [CDO15], and tree games [DFS20].

More relevant to the topic of smoothed analysis, it is known that when equilibria do not fluctuate when the input is perturbed, finding equilibria can be done efficiently [BB17]. Furthermore, a game chosen at random is likely to have easy-to-find equilibria [BVV07].

Spielman and Teng [ST06, Open Question 11] ask whether there is a relation between approximation hardness and smoothed lower bounds: the former implies the latter, but little else is known regarding smoothed lower bounds. For the case of integer linear programs over the unit cube, Beier and Vöcking [BV06] show that a problem has polynomial smoothed complexity if and only if it admits a pseudo-polynomial algorithm. Note that a pseudo-polynomial algorithm can be used to approximate by truncating input numbers. For other problems, we are aware of a few papers that argue smoothed complexity lower bounds via approximation hardness, e.g. [CDT09], [HT07], [KN07].

## II. PRELIMINARIES

We formally define the problem here, and present some remarks. Let  $n$  be a positive integer. We let  $e_i \in \mathbb{R}^n$  be the  $i$ th indicator vector. Let  $A, B \in \mathbb{R}^{n \times n}$  be payoff matrices (corresponding to Alice and Bob). We define a Nash equilibrium to be vectors  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_{\geq 0}^n$ , called *mixed strategies*, such that  $\|\mathbf{x}\|_1 = \|\mathbf{y}\|_1 = 1$  we have that

$$\begin{aligned} \mathbf{x}^\top A \mathbf{y} &= \max_{i \in [n]} e_i^\top A \mathbf{y} \\ \mathbf{x}^\top B \mathbf{y} &= \max_{i \in [n]} \mathbf{x}^\top B e_i. \end{aligned}$$

We say that an equilibrium is  $\epsilon$ -approximate if

$$\begin{aligned} \mathbf{x}^\top A \mathbf{y} + \epsilon &\geq \max_{i \in [n]} e_i^\top A \mathbf{y} \\ \mathbf{x}^\top B \mathbf{y} + \epsilon &\geq \max_{i \in [n]} \mathbf{x}^\top B e_i. \end{aligned}$$

For a given equilibrium  $\mathbf{x}, \mathbf{y}$  (often clear from context), we let  $A^{00}$  and  $B^{00}$  be the restrictions of  $A$  and  $B$  to  $\text{supp}(\mathbf{x}) \times \text{supp}(\mathbf{y})$ , respectively. We let  $A^{10}$  and  $B^{10}$  be the restrictions to  $\text{supp}(\mathbf{x}) \times \text{supp}(\mathbf{y})$ , etc. Computing any Nash equilibrium, even  $n^{-O(1)}$ -approximate, is known to be PPAD-complete [CDT09]:

**Theorem 2** ([CDT09]). *For all  $c > 0$ , computing an  $n^{-c}$ -approximate Nash equilibrium of an  $n \times n$  bimatrix game with entries bounded in  $[0, 1]$  is PPAD-complete.*

#### A. Remarks on the reduction

The reduction, presented in Section III, will ultimately take a hard instance of Theorem 2 and transform it into an instance of  $X$ -SMOOTHED-NASH, for suitable distributions  $X$ . By the nature of the reduction, if one applies the same reduction with a wider hardness-of-approximation guarantee, one can deduce that for a suitable constant  $c > 0$ , it is PPAD-hard under a randomized reduction to find a  $n^{-c}$ -approximate equilibrium of  $X$ -SMOOTHED-NASH (see, e.g., Eq. 5). This has two interesting implications.

First, this means that if you truncate the output of the distribution  $X$ , as well as the uniform distribution sampled in the reduction, to  $O(\log n)$  bits, it is still PPAD-hard to find an (approximate) equilibrium for the resulting instance. In particular, the smoothed complexity result is robust to the underlying arithmetic representation of the payoffs.

Second, scaling down the hard instance of Theorem 2 by a small polynomial still maintains an  $n^{-O(1)}$  hardness-of-approximation guarantee. Thus, as mentioned in the introduction, the reduction implies that Nash remains intractable even subject to noise (zero-sum + i.i.d.) that is *polynomially* larger than the worst-case signal.

#### B. Concentration for random bilinear forms

We introduce here the following concentration bound which is useful in our result.

**Definition 2** (Subgaussian random variable). A  $\mathbb{R}$ -valued random variable  $X$  is subgaussian with variance proxy  $s^2 > 0$  if for all  $t > 0$ ,  $\mathbb{E} \exp(tX) \leq \exp(s^2 t^2 / 2)$ . Note that if  $X \in [-b, b]$  for some  $b > 0$  with probability 1, then  $X$  is subgaussian with variance proxy  $b^2/4$ .

**Lemma 3.** *Let  $A$  be an  $n \times n$  matrix with independent subgaussian entries with variance proxy at most 1. For all  $u > 0$ , with probability at least  $1 - \exp(-u^2)$ , all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  with  $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1$  have*

$$\mathbf{x}^\top A \mathbf{y} \leq O(\sqrt{\log n} + u)(\|\mathbf{x}\|_1 + \|\mathbf{y}\|_1).$$

*As a corollary, with the same probability, all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  with  $\|\mathbf{x}\|_1, \|\mathbf{y}\|_1 \leq 1$  have*

$$\mathbf{x}^\top A \mathbf{y} \leq O(\sqrt{\log n} + u)(\|\mathbf{x}\|_2 + \|\mathbf{y}\|_2).$$

The proof of this lemma is deferred to the Appendix.

## III. THE REDUCTION, AND PROOF OF THEOREM 1

First, we show in Section III-A the reduction in the case that the noise distribution  $X$  is symmetric, i.e., the probability of sampling  $a$  and  $-a$  is identical for all  $a > 0$ . We then show in Section III-B a slight modification which works for any distribution  $X$ .

### A. The symmetric case

Let  $\epsilon > 0$  be a sufficiently small constant. Let  $X$  be any symmetric distribution on  $[-\epsilon, \epsilon]$ . Let  $n, b$  be positive integers such that  $b$  divides  $n$ ,  $b = n^{0.01}$ , and  $n$  is sufficiently large. We divide  $[n]$  into  $b$  blocks which we label  $I_i := \{(i-1)\frac{n}{b}+1, (i-1)\frac{n}{b}+2, \dots, i\frac{n}{b}\}$ . We let  $\ell := n/b = n^{0.99}$  denote the *block length*.

Let  $P, Q \in \mathbb{R}^{b \times b}$  be payoff matrices. Let  $J_\ell$  denote the  $\ell \times \ell$  all 1's matrix. Let  $Z_0$  be an  $n \times n$  matrix whose entries are sampled i.i.d. from the Rademacher distribution (i.e., the uniform distribution on  $\{-1, 1\}$ ). Let  $Z_1$  be an  $n \times n$  matrix whose entries are sampled i.i.d. from the uniform distribution on  $[-1, 1]$ . Let  $A_\epsilon, B_\epsilon$  be  $n \times n$  matrices whose entries are i.i.d. sampled from  $X$  (all distributions independent).<sup>11</sup>

$$\begin{aligned} A &:= P \otimes J_\ell + Z_0 + Z_1 + A_\epsilon \\ B &:= Q \otimes J_\ell - Z_0 - Z_1 + B_\epsilon, \end{aligned}$$

where  $P \otimes J_\ell$  denotes the  $n \times n$  matrix, where every entries in block  $I_i \times I_j$  is  $P_{i,j}$ .

We present here the final result of this paper. We will refer without proof to a bound on the norm of the equilibrium strategy vectors, and we defer its proof to the rest of the paper, namely Sections IV and V. This norm bound is the technical heart of this paper, and the present section illustrates its strength.

We seek to show that equilibria of the reduced game  $(A, B)$  can be used to efficiently produce approximate equilibria to the game  $(P, Q)$ , which we have assumed is hard to approximate. Let  $(\mathbf{x}, \mathbf{y})$  be an equilibrium of  $(A, B)$ . We will show in Section V that, with high probability,  $\|\mathbf{x}\|_2, \|\mathbf{y}\|_2 \leq n^{-0.2}$ , even when  $\epsilon$  is a constant. Note that  $b = n^{0.01}$  is the dimension of the input game  $(P, Q)$ . Define  $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$  to be distributions over  $[b]$  such that for all  $i \in [n]$

$$\hat{x}_i = \sum_{i' \in I_i} x_{i'}, \quad \hat{y}_i = \sum_{i' \in I_i} y_{i'}.$$

**Theorem 4.** *With probability  $1 - n^{-2}$ , we have that  $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$  is a  $b^{-19}$ -approximate equilibrium of  $(P, Q)$ .*

*Proof:* We claim that  $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$  is an  $b^{-19} = n^{-0.19}$ -approximate equilibrium of  $(P, Q)$  with high probability. Assume not, without loss of generality, Alice would benefit from deviating from  $\hat{\mathbf{x}}$ . That is, there exists  $i \in [b]$  such that

$$\hat{\mathbf{x}}^\top P \hat{\mathbf{y}} \leq e_i^\top P \hat{\mathbf{y}} - b^{-19}. \quad (1)$$

Define  $\mathbf{u}_S$  to be the uniform probability vector on support  $S$ , then, the above is equivalent to

$$\hat{\mathbf{x}}^\top (P \otimes J_\ell) \mathbf{y} \leq \mathbf{u}_{I_i}^\top (P \otimes J_\ell) \mathbf{y} - b^{-19}. \quad (2)$$

<sup>11</sup>The to meet the definition of  $X$ -SMOOTHED-NASH, which specifies that the hard game must have entries between  $[-1, 1]$ , we can scale the construction (and thus  $X$ ) by a factor of 3.

By Lemma 3, we may assume that the concentration inequality holds for  $\frac{1}{2+\epsilon}(Z_0 + Z_1 + A_\epsilon)$ , then we know that

$$|\mathbf{x}^\top (Z_0 + A_\epsilon) \mathbf{y}| \leq O(\sqrt{\log n} n^{-0.2}) \quad (3)$$

$$|\mathbf{u}_{I_i}^\top (Z_0 + A_\epsilon) \mathbf{y}| \leq O(\sqrt{\log n} n^{-0.2}) \quad (4)$$

Combining Eqs. 2,3, and 4 we get

$$\hat{\mathbf{x}}^\top \mathbf{A} \mathbf{y} \leq \mathbf{u}_{I_i}^\top \mathbf{A} \mathbf{y} - b^{-19} + O(\sqrt{\log n} n^{-0.2}) < \mathbf{u}_{I_i}^\top \mathbf{A} \mathbf{y}. \quad (5)$$

since  $b = n^{0.01}$ . This contradicts that  $(\mathbf{x}, \mathbf{y})$  is a Nash equilibrium of  $(A, B)$ .

By a similar argument, Bob does not wish to deviate with high probability. Therefore,  $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$  is a  $b^{-19}$ -approximate Nash equilibrium of  $(P, Q)$ . ■

Since finding a  $b^{-19}$ -approximate Nash equilibrium is PPAD-hard [CDT09] when  $P$  and  $Q$  have constant sized entries, finding the smoothed equilibrium of  $(A, B)$  is PPAD-hard. Since the proofs of Sections IV and V hold when  $X$  is supported on  $[-\epsilon, \epsilon]$  for  $\epsilon > 0$  constant, this is an instance of  $X$ -SMOOTHED NASH, and therefore concludes the proof of Theorem 1 when  $X$  is a symmetric distribution.

### B. General $X$

Let  $X$  be any distribution supported on  $[-\epsilon/2, \epsilon/2]$ . Let  $Y := X - X'$  be the distribution on  $[-\epsilon, \epsilon]$  which takes two i.i.d. samples from  $X$  and subtracts them. Note that  $Y$  is a symmetric distribution, so by the previous section we have that  $Y$ -SMOOTHED NASH is hard. In particular, it is hard to find an equilibrium from the distribution

$$\begin{aligned} A &:= P \otimes J_\ell + Z_0 + Z_1 + A_Y \\ B &:= Q \otimes J_\ell - Z_0 - Z_1 + B_Y, \end{aligned}$$

where  $A_Y$  and  $B_Y$  are matrix whose entries are i.i.d. samples from  $Y$ . We can rewrite  $A_Y = A_X - A'_X$  and  $B_Y = B_X - B'_X$ , where  $A_X, A'_X, B_X, B'_X$  are all i.i.d. matrix samples from  $X$ . Thus, the distribution can be rewritten as

$$\begin{aligned} A &:= (P \otimes J_\ell + Z_0 + Z_1 - A'_X) + A_X \\ B &:= (Q \otimes J_\ell - Z_0 - Z_1 - B'_X) + B_X, \end{aligned}$$

This is an instance of  $X$ -SMOOTHED NASH, and we conclude Theorem 1 for arbitrary  $X$ , losing a factor 2 on  $\epsilon$ .

## IV. EQUILIBRIA HAVE LARGE SUPPORT

In this section and the following, we will show the bound on  $\|\mathbf{x}\|_2, \|\mathbf{y}\|_2$  which was required in the proof of Theorem 1. We first show that the support of the equilibria is large with high probability. Then, in Section V, use this to argue that the weight must be sufficiently spread. The main result of this section is the following lemma.

**Lemma 5.** *With probability  $1 - n^{-3}$ , for every Nash equilibrium  $(\mathbf{x}, \mathbf{y})$  of  $(A, B)$ , we have that  $|\text{supp}(\mathbf{x})| = |\text{supp}(\mathbf{y})| > n^{0.96}$ .*

We prove this result using methods partially inspired by [Jon04]. Observe that a Nash equilibrium of  $(A, B)$  requires that

$$\begin{aligned} \mathbf{x}^\top A \mathbf{y} &\geq \mathbf{e}_i^\top A \mathbf{y} && \text{for all } i \in [n] \\ \mathbf{x}^\top B \mathbf{y} &\geq \mathbf{x}^\top B \mathbf{e}_j && \text{for all } j \in [n] \\ \implies \mathbf{x}^\top (A + B) \mathbf{y} &\geq \mathbf{e}_i^\top A \mathbf{y} + \mathbf{x}^\top B \mathbf{e}_j && \text{for all } i, j \in [n]. \end{aligned} \quad (6)$$

We seek to show that Eq. 6 cannot hold when the support  $\mathbf{x}, \mathbf{y}$  is sufficiently small.<sup>12</sup> To do that, we propose a “benchmark” to which both the LHS and the maximum value of the RHS of Eq. 6 are comparable to. To define this benchmark, we begin by introducing a notion of *robust partition* of the strategy vectors. Consider  $\mathbf{x} \in \mathbb{R}^n$  such that  $\|\mathbf{x}\|_1 = 1$ . Let  $L = \lceil \log_2 n \rceil / 2^{100}$ . Let  $D = 2^{2^{500}}$ . Let  $E_1, \dots, E_L$  be intervals such that  $E_i = (D^{-i}, D^{-(i-1)})$  for all  $i < L$  and  $E_L = [0, D^{-(L-1)}]$ . Let  $\mathbf{x} = \mathbf{x}^{(1)} + \dots + \mathbf{x}^{(L)}$  such that

$$\mathbf{x}_j^{(i)} = \begin{cases} \mathbf{x}_j & \mathbf{x}_j \in E_i \\ 0 & \text{otherwise} \end{cases}$$

We say that  $\mathbf{x}^{(i)}$  is *sparse* if it has at most  $L$  nonzero coordinates; otherwise we say  $\mathbf{x}^{(i)}$  is *dense*. Let  $\mathbf{x}_{\text{sparse}}$  be the sum of the sparse  $\mathbf{x}^{(i)}$ 's and  $\mathbf{x}_{\text{dense}}$  be the sum of the dense ones. Note that  $\mathbf{x} = \mathbf{x}_{\text{sparse}} + \mathbf{x}_{\text{dense}}$ . Now define the following quantity

$$\beta(\mathbf{x}) = \sqrt{\log n} \|\mathbf{x}_{\text{dense}}\|_2 + \|\mathbf{x}_{\text{sparse}}\|_1.$$

We call  $\beta(\mathbf{x})$  the *benchmark* for  $\mathbf{x}$ . This quantity will appear in a number of concentration/anti-concentration inequalities. First, we show a key anticoncentration inequality concerning this robust partition.

**Lemma 6.** *Assume that  $X$  is the uniform distribution on  $\{-1, 1\}$  (i.e., the Rademacher distribution). There exists a universal constant  $c > 0$  with the following property: For all  $\mathbf{x} \in \mathbb{R}^n$  such that  $\|\mathbf{x}\|_1 = 1$ , with probability at least  $n^{-0.001}$  over  $\mathbf{v} \sim X^n$*

$$\langle \mathbf{v}, \mathbf{x} \rangle \geq c\beta(\mathbf{x}).$$

The proof of the above lemma is deferred to the Appendix. The following concentration bound will also be of use. For any distribution  $X$ , we let  $X^{n \times n}$  denote the distribution of  $n \times n$  matrices with entries i.i.d. samples from  $X$ .

**Claim 7.** *Let  $X$  be any distribution on  $[-1, 1]$ . There exists a universal constant  $C > 0$  such that for all  $n \geq 0$ , with probability  $1 - 1/n^4$  over  $M \sim X^{n \times n}$ , for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  such that  $\|\mathbf{x}\|_1 = \|\mathbf{y}\|_1 = 1$ , we have that*

$$|\mathbf{x}^\top M \mathbf{y}| \leq C \cdot (\beta(\mathbf{x}) + \beta(\mathbf{y})).$$

<sup>12</sup>In the case of [Jon04], which considers zero-sum games, the LHS of (6) is equal to 0, so it suffices to bound the probability that the RHS is positive for *some*  $i$  and  $j$ .

*Proof:* Apply Lemma 3 to  $M$  with  $u = \sqrt{3 \log n}$ . Then, there is a universal constant  $C'$  such that with probability  $1 - 1/n^3$ , for all  $\mathbf{x}, \mathbf{y}$  with  $\ell_1$  norm 1,

$$|\mathbf{x}_{\text{dense}}^\top M \mathbf{y}_{\text{dense}}| \leq C' \sqrt{\log n} (\|\mathbf{x}_{\text{dense}}\|_2 + \|\mathbf{y}_{\text{dense}}\|_2).$$

Thus, since the entries of  $M$  have absolute value at most 1,

$$\begin{aligned} |\mathbf{x}^\top M \mathbf{y}| &\leq |\mathbf{x}^\top M \mathbf{y}_{\text{sparse}}| + |\mathbf{x}_{\text{sparse}}^\top M \mathbf{y}_{\text{dense}}| \\ &\quad + |\mathbf{x}_{\text{dense}}^\top M \mathbf{y}_{\text{dense}}| \\ &\leq \|\mathbf{y}_{\text{sparse}}\|_1 + \|\mathbf{x}_{\text{sparse}}\|_1 \\ &\quad + C' \sqrt{\log n} (\|\mathbf{x}_{\text{dense}}\|_2 + \|\mathbf{y}_{\text{dense}}\|_2) \\ &\leq \max(C', 1) (\beta(\mathbf{x}) + \beta(\mathbf{y})). \end{aligned}$$

Thus, we can set  $C = \max(C', 1)$ .  $\blacksquare$

These lemmas will allow us to prove Lemma 5. We present first the following facts about equilibria in random games.

**Proposition 8.** *With probability 1, for nonempty  $S, T \subset [n]$  there is at most one Nash equilibrium  $(\mathbf{x}, \mathbf{y})$  of  $(A, B)$  with  $S = \text{supp}(\mathbf{x})$  and  $T = \text{supp}(\mathbf{y})$ . Further, with probability 1 all such equilibria have  $|S| = |T|$ .*

*Proof:* Fix nonempty  $S, T \subset [n]$ . Fix  $i_0 \in S$ . Assume without loss of generality that  $|S| \geq |T|$ . Denote  $A^{00}$  as the sub-matrix of  $A$  restricted to rows indexed by  $S$  and columns indexed by  $T$ . For any equilibrium  $(\mathbf{x}, \mathbf{y})$  with supports  $S$  and  $T$ , we have that  $\mathbf{x}^\top A \mathbf{y} = \mathbf{e}_{i_0}^\top A^{00} \mathbf{y}$  for all  $i \in S$ , when treating  $\mathbf{x}$  and  $\mathbf{y}$  as  $|S|$ - and  $|T|$ -dimensional vectors, respectively. Therefore,

$$(\mathbf{e}_i - \mathbf{e}_{i_0})^\top A^{00} \mathbf{y} = 0 \text{ for all } i \in S \setminus \{i_0\}. \quad (7)$$

Since all the entries of  $A^{00}$  are drawn independently from a continuous distribution, the null space of the linear system (7) has dimension  $\max(|T| - |S| + 1, 0) \leq 1$  with probability 1. Since  $\mathbf{y} \neq 0$  the null space must have dimension exactly 1. Thus,  $|T| - |S| + 1 \geq 1$ , which implies that  $|S| = |T|$  and the solution  $\mathbf{y}$  is unique, as there can be at most one vector in a 1-dimensional subspace with coordinates summing to one. By a similar argument  $\mathbf{x}$  is also unique.

Since there are only finitely many choices of  $S$  and  $T$ , with probability 1 the proposition holds for all Nash equilibria simultaneously.  $\blacksquare$

With probability 1, all equilibria of  $A$  and  $B$  will have the same support size, and further, for every pair of possible supports  $S \subset [n]$  and  $T \subset [n]$  there is at most one equilibrium. We let  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  denote the probability distributions of strategies in this equilibrium.

We can now prove Lemma 5

*Proof of Lemma 5.:* Assume (which happens with probability  $1 - n^{-4}$ ) that the event described in Claim 7 occurs for  $M = \frac{1}{2\epsilon}(A_\epsilon + B_\epsilon)$ . Fix  $S, T \subset [n]$  with  $|S|, |T| < \ell/10$ . We seek to show that with probability at most  $2^{-\ell}$ ,  $S$  and  $T$  can be the support of a Nash equilibrium. By Proposition 8, we can assume that  $|S| = |T|$ .

Also by Proposition 8, with probability 1, there is at most one equilibrium  $(\mathbf{x}, \mathbf{y})$  on the game  $(A^{00}, B^{00})$  with full support. Note that  $\mathbf{x}$  and  $\mathbf{y}$ , if they exist, are independent of the entries of  $A$  and  $B$  outside of  $S \times T$ . As mentioned earlier in the section, in order for the equilibrium to extend, the Ineq. 6 must hold:

$$\mathbf{x}^\top(A+B)\mathbf{y} \geq \mathbf{e}_i^\top A\mathbf{y} + \mathbf{x}^\top B\mathbf{e}_j \quad \text{for all } i, j \in [n].$$

Say that  $i \in [n] \setminus S$  is  $S$ -good if  $\mathbf{e}_i^\top(Z_0 + Z_1 + A_\epsilon)\mathbf{y} > c\beta(\mathbf{y})$ . By Lemma 6, we know that  $\mathbf{e}_i^\top Z_1\mathbf{y} > c\beta(\mathbf{y})$  with probability at least  $n^{-0.001}$ . Independently, we have that  $\mathbf{e}_i^\top(Z_0 + A_\epsilon)\mathbf{y} \geq 0$  with probability at least  $1/2$  (since  $Z_0 + A_\epsilon$  is a mean-zero matrix distribution). Therefore, both this event happens with probability at least  $n^{-0.001}/2 \geq n^{-0.01}$ .

Likewise, say that  $j \in [n] \setminus T$  is  $T$ -good if  $\mathbf{x}^\top(-Z_0 - Z_1 - B_\epsilon)\mathbf{e}_j > c\beta(\mathbf{x})$ . By the same argument, this also happens with probability at least  $n^{-0.01}$ . Furthermore, the  $S$ -good events and  $T$ -good events are independent of each other because each event is based on a disjoint subset of entries  $Z$  from  $Z_0$  and  $Z_1$ .

Since  $\mathbf{x}$  and  $\mathbf{y}$  are probability distributions, there exists  $i_0 \in S$  and  $j_0 \in T$  such that  $\mathbf{e}_{i_0}^\top(P \otimes J_\ell)\mathbf{y} \geq \mathbf{x}^\top(P \otimes J_\ell)\mathbf{y}$  and  $\mathbf{x}^\top(Q \otimes J_\ell)\mathbf{e}_{j_0} \geq \mathbf{x}^\top(Q \otimes J_\ell)\mathbf{y}$ . Let  $i', j' \in [b]$  be the indices of the blocks such that  $i_0 \in I_{i'}$  and  $j_0 \in I_{j'}$ . Since we assume that  $|S|, |T| \leq \ell/10$ , we have that  $I_{i'} \setminus S$  and  $I_{j'} \setminus T$  both have size at least  $9\ell/10$ .

Now, for any good  $i \in I_{i'} \setminus S$  and good  $j \in I_{j'} \setminus T$ , we have

$$\begin{aligned} & \mathbf{x}^\top(A+B)\mathbf{y} \\ &= \mathbf{x}^\top(P \otimes J_\ell)\mathbf{y} + \mathbf{x}^\top(Q \otimes J_\ell)\mathbf{y} + \mathbf{x}^\top(A_\epsilon + B_\epsilon)\mathbf{y} \\ &\leq \mathbf{e}_{i_0}^\top(P \otimes J_\ell)\mathbf{y} + \mathbf{x}^\top(Q \otimes J_\ell)\mathbf{e}_{j_0} + 2C\epsilon(\beta(\mathbf{x}) + \beta(\mathbf{y})) \\ &= \mathbf{e}_i^\top(P \otimes J_\ell)\mathbf{y} + \mathbf{x}^\top(Q \otimes J_\ell)\mathbf{e}_j + 2C\epsilon(\beta(\mathbf{x}) + \beta(\mathbf{y})) \\ &< \mathbf{e}_i^\top(P \otimes J_\ell)\mathbf{y} + \mathbf{x}^\top(Q \otimes J_\ell)\mathbf{e}_j + c(\beta(\mathbf{x}) + \beta(\mathbf{y})) \\ &\quad \text{(because } \epsilon < 2c/C) \\ &< \mathbf{e}_i^\top(P \otimes J_\ell + Z_0 + Z_1 + A_\epsilon)\mathbf{y} \\ &\quad + \mathbf{x}^\top(Q \otimes J_\ell - Z_0 - Z_1 + B_\epsilon)\mathbf{e}_j \\ &= \mathbf{e}_i^\top A\mathbf{y} + \mathbf{x}^\top B\mathbf{y}, \end{aligned}$$

which contradicts Ineq. 6. Thus, there must either be no good  $i \in I_{i'} \setminus S$  or there is no good  $j \in I_{j'} \setminus T$ . This happens with probability at most

$$2(1 - n^{-0.01})^{9\ell/10} \leq 2e^{-(0.9)\ell/n^{0.01}} \leq e^{-n^{0.97}},$$

where we use in the last inequality that  $n$  is sufficiently large. The number of pairs  $S, T$  with support at most  $n^{0.96}$  is at most

$$\binom{n}{\leq n^{0.96}}^2 \leq n^{2n^{0.96}}.$$

Note that for  $n$  sufficiently large,  $n^{2n^{0.96}}e^{-n^{0.97}} \ll n^{-4}$ . Thus, all equilibria have support size greater than  $n^{0.96}$  with probability at least  $1 - 2n^{-4} \geq 1 - n^{-3}$ . ■

## V. EQUILIBRIA HAVE SMALL $\ell_2$ NORM

Towards showing the missing bound in the proof of Theorem 1, the previous section showed that with high probability, any equilibrium must have polynomially large support. We complete here the proof of the norm bound, which in turn completes the proof of Theorem 1.

**Lemma 9.** *With probability  $1 - 20n^{-3}$ , for every Nash equilibrium  $(\mathbf{x}, \mathbf{y})$  of  $(A, B)$ , we have that  $\|\mathbf{x}\|_2, \|\mathbf{y}\|_2 \leq n^{-0.2}$ .*

We must, however, begin this section with a few technical results. We will need the following theorem, which is derived from the fact that the VC-dimension of the set of halfspaces in  $\mathbb{R}^d$  has VC-dimension at most  $d + 1$  – that is, the VC-dimension of  $\{\mathbf{x} \mapsto 1[\langle \mathbf{x}, \mathbf{v} \rangle + t \geq 0] : \mathbf{v} \in \mathbb{R}^d, t \in \mathbb{R}\}$  is at most  $d + 1$ . (See e.g. [Wai19], Example 4.21.)

**Theorem 10** (Multivariate Glivenko-Cantelli). *Let  $X$  be a random vector in  $\mathbb{R}^d$  and let  $X_1, \dots, X_n$  be independent copies of  $X$ . For all  $\delta \in [0, 1]$ , with probability  $1 - \delta$ ,*

$$\begin{aligned} & \sup_{\mathbf{v} \in \mathbb{R}^d, t \in \mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^n 1[\langle X_i, \mathbf{v} \rangle \geq t] - \Pr_X(\langle X, \mathbf{v} \rangle \geq t) \right| \\ & \leq O\left(\sqrt{\frac{d}{n}} + \sqrt{\frac{\log(1/\delta)}{n}}\right). \end{aligned}$$

We also need the following Littlewood-Offord-type theorem.

**Theorem 11** ([RV15], Theorem 1.2). *Let  $X_1, \dots, X_n$  be real-valued independent random variables with densities almost everywhere bounded by  $K$ . Let  $a_1, \dots, a_n \in \mathbb{R}$  with  $\sum_{i \leq n} a_i^2 = 1$ . Then the density of  $\sum_{i \leq n} a_i X_i$  is bounded by  $\sqrt{2}K$  almost everywhere.*

The following lemma, which we obtain as a corollary of these two theorems, allows us to argue that the entries of a product of a random matrix with a fixed vector are relatively spread out.

**Lemma 12.** *Let  $n, d$  be positive integers. Let  $X$  be an  $\mathbb{R}$ -valued random variable with density bounded by  $K$ . Let  $\mathbf{g}_1, \dots, \mathbf{g}_n$  be independent random vectors in  $\mathbb{R}^d$  whose coordinates are independent copies of  $X$ . With probability  $1 - \delta$ , for all unit vectors  $\mathbf{v} \in \mathbb{R}^d$  and all intervals  $[a, b] \subset \mathbb{R}$ ,*

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n 1[\langle \mathbf{g}_i, \mathbf{v} \rangle \in [a, b]] \\ & \leq \sqrt{2}K|a - b| + O\left(\sqrt{\frac{d}{n}} + \sqrt{\frac{\log(1/\delta)}{n}}\right). \end{aligned}$$

*Proof:* By Theorem 10, with probability at least  $1 - \delta$ , the CDFs of  $\langle \mathbf{g}, \mathbf{v} \rangle$  and the empirical distribution of

$\langle \mathbf{g}_i, \mathbf{v} \rangle$  have distance at most  $O\left(\sqrt{\frac{d}{n}} + \sqrt{\frac{\log(1/\delta)}{n}}\right)$ , for all  $\mathbf{v} \in \mathbb{R}^d$ . So it suffices to show that for every unit  $\mathbf{v} \in \mathbb{R}^d$ ,  $\Pr_{\mathbf{g}}(\langle \mathbf{g}, \mathbf{v} \rangle \in [a, b]) \leq \sqrt{2K}|a - b|$ . This follows immediately from Theorem 11. ■

Finally, this lemma allows us to prove the following claim.

**Claim 13.** *Let  $X$  be a distribution on  $[-1, 1]$  whose probability density is at most 100 everywhere. Let  $M \sim X^{n \times n}$ . With probability  $1 - n^{-4}$ , for every  $S, T \subset [n]$  with  $|S| \geq n^{0.95}$  and  $|T| \leq n^{0.85}$ , there exists disjoint  $S_1, S_2 \subset S$  of size at least  $n^{0.94}$  each such that for all unit vectors  $\mathbf{y} \in \mathbb{R}^n$  with support in  $T$  there exists  $r \in \mathbb{R}$  such that*

$$\begin{aligned} \mathbf{e}_{i_1}^\top M \mathbf{y} &\geq r + n^{-0.07} && \text{for all } i_1 \in S_1 \\ \mathbf{e}_{i_2}^\top M \mathbf{y} &\leq r && \text{for all } i_2 \in S_2. \end{aligned}$$

*Proof:* For every  $T \subset [n]$  of size at most  $n^{0.85}$ , apply Lemma 12 to the rows of  $M$  restricted to the columns of  $T$  (so  $d = |T| \leq n^{0.85}$ ) with  $\delta = e^{-n^{0.86}}$ . Thus, with probability  $1 - e^{-n^{0.86}}$ , for every unit vector  $\mathbf{y} \in \mathbb{R}^d$  supported on  $T$  and every interval  $[a, b]$  of length  $n^{-0.06}/10$ , the number of  $i \in [n]$  such that  $\mathbf{e}_i^\top M \mathbf{y} \in [a, b]$  is at most

$$n \left[ 100\sqrt{2}|a - b| + O\left(\sqrt{\frac{d}{n}} + \sqrt{\frac{\log(1/\delta)}{n}}\right) \right] = O(n^{0.94})$$

choices of  $i \in [n]$  for which  $\mathbf{e}_i^\top M \mathbf{y}$  falls in that interval. Since  $|S| \geq n^{0.95}$ , this implies there exist  $r \in \mathbb{R}$ , and disjoint  $S_1, S_2 \subset S$  of size at least  $n^{0.94}$  such that

$$\begin{aligned} \mathbf{e}_{i_1}^\top M \mathbf{y} &\geq r + \frac{n^{-0.06}}{10} \geq r + n^{-0.07} && \text{for all } i_1 \in S_1 \\ \mathbf{e}_{i_2}^\top M \mathbf{y} &\leq r && \text{for all } i_2 \in S_2. \end{aligned}$$

Taking the union bound over all choices of  $T$  we get this all happens with probability at most

$$1 - \binom{n}{\leq n^{0.85}} e^{-n^{0.86}} \geq 1 - e^{-n^{0.85}} \geq 1 - n^{-4}. \quad \blacksquare$$

We can now prove Lemma 9.

*Proof of Lemma 9:* With probability  $1 - n^{-3}$ , by Lemma 5, for every equilibrium  $(\mathbf{x}, \mathbf{y})$  of  $(A, B)$  with support  $S$  and  $T$ , respectively, we have that  $|S| = |T| \geq n^{0.96}$ . Since there are  $n^{0.01}$  blocks. By the pigeonhole principle there exists  $i_0, j_0 \in [b]$  such that  $|S \cap I_{i_0}|, |T \cap I_{j_0}| \geq n^{0.95}$ .

With probability  $1 - 2n^{-4}$ , Claim 13 holds with for both  $M = \frac{1}{2+\epsilon}(Z_0 + Z_1 + A_\epsilon)$  and  $M = \frac{1}{2+\epsilon}(-Z_0 - Z_1 + B_\epsilon)$ . Further, with probability at least  $1 - 2n^{-3}$ , Lemma 3 holds for  $M = \frac{1}{2+\epsilon}(Z_0 + Z_1 + A_\epsilon)$  and  $M = \frac{1}{2+\epsilon}(-Z_0 - Z_1 + B_\epsilon)$  with  $u = \sqrt{3} \log n$ .

We seek to show that any large-support equilibrium also has small  $\ell_2$  norm. Assume for sake of contradiction (and without loss of generality) that  $\|\mathbf{y}\|_2 \geq n^{-0.2}$ . Let  $S' = S \cap I_{i_0}$  and  $T'$  be the set of coordinates of  $\mathbf{y}$  which are greater than  $n^{-0.85}$ . Clearly  $|T'| \leq n^{0.85}$ . Let  $\mathbf{y}_{T'}$  be the

coordinates of  $\mathbf{y}$  supported on  $T'$  and  $\bar{\mathbf{y}}_{T'}$  be the remaining coordinates. Observe that

$$\|\bar{\mathbf{y}}_{T'}\|_2^2 \leq n \cdot (n^{-0.85})^2 = n^{-0.7} \leq \frac{\|\mathbf{y}\|_2^2}{2} \quad (8)$$

$$\|\mathbf{y}_{T'}\|_2^2 = \|\mathbf{y}\|_2^2 - \|\bar{\mathbf{y}}_{T'}\|_2^2 \geq \frac{\|\mathbf{y}\|_2^2}{2}. \quad (9)$$

Applying Claim 13 for  $M = \frac{1}{2+\epsilon}(Z_0 + Z_1 + A_\epsilon)$  and the sets  $S', T'$  and the vector  $\mathbf{y}' := \frac{\mathbf{y}_{T'}}{\|\mathbf{y}_{T'}\|_2}$ , there exists  $S'_1, S'_2 \subset S'$  and  $r \in \mathbb{R}$  such that (scaling by  $2 + \epsilon \geq 1$ )

$$\begin{aligned} \mathbf{e}_{i_1}^\top (Z_0 + Z_1 + A_\epsilon) \mathbf{y}' &\geq r + n^{-0.07} && \text{for all } i_1 \in S'_1 \\ \mathbf{e}_{i_2}^\top (Z_0 + Z_1 + A_\epsilon) \mathbf{y}' &\leq r && \text{for all } i_2 \in S'_2. \end{aligned}$$

Thus,

$$\begin{aligned} \mathbf{u}_{S'_1}^\top (Z_0 + Z_1 + A_\epsilon) \mathbf{y}' &\geq r + n^{-0.07} \\ \mathbf{u}_{S'_2}^\top (Z_0 + Z_1 + A_\epsilon) \mathbf{y}' &\leq r \\ \implies (\mathbf{u}_{S'_1} - \mathbf{u}_{S'_2})^\top (Z_0 + Z_1 + A_\epsilon) \mathbf{y}' &\geq n^{-0.07}. \end{aligned}$$

Applying (9),

$$(\mathbf{u}_{S'_1} - \mathbf{u}_{S'_2})^\top (Z_0 + Z_1 + A_\epsilon) \mathbf{y}_{T'} \geq n^{-0.07} \|\mathbf{y}\|_2 / 2 \geq n^{-0.28}.$$

Since Lemma 3 holds for  $M = \frac{1}{2+\epsilon}(Z_0 + Z_1 + A_\epsilon)$ , we have that

$$\begin{aligned} (\mathbf{u}_{S'_1} - \mathbf{u}_{S'_2})^\top (Z_0 + Z_1 + A_\epsilon) \bar{\mathbf{y}}_{T'} &\geq -(2 + \epsilon) C' \sqrt{\log n} (\|\mathbf{u}_{S'_1} - \mathbf{u}_{S'_2}\|_2 + \|\bar{\mathbf{y}}_{T'}\|_2) \\ &\geq -n^{0.01} \max(\sqrt{2} n^{-0.94/2}, n^{-0.7/2}) \\ &\geq -n^{-0.34}. \end{aligned}$$

Therefore, since  $\mathbf{y} = \mathbf{y}_{T'} + \bar{\mathbf{y}}_{T'}$

$$(\mathbf{u}_{S'_1} - \mathbf{u}_{S'_2})^\top (Z_0 + A_\epsilon) \mathbf{y} \geq n^{-0.28} - n^{-0.34} \geq n^{-0.29}.$$

Since  $S'_1$  and  $S'_2$  are subsets of the same block, we have that  $\mathbf{u}_{S'_1}(P \otimes J_\ell) = \mathbf{u}_{S'_2}(P \otimes J_\ell)$ . Therefore,

$$(\mathbf{u}_{S'_1} - \mathbf{u}_{S'_2})^\top A \mathbf{y} \geq n^{-0.29}.$$

But, since  $S'_1$  and  $S'_2$  are subsets of the support of  $\mathbf{x}$ , we know that

$$(\mathbf{u}_{S'_1} - \mathbf{u}_{S'_2})^\top A \mathbf{y} = 0,$$

thus we have a contradiction. Therefore,  $\|\mathbf{y}\|_2 \leq n^{-0.2}$ . By a similar argument (also with probability  $1 - 5n^{-3}$ ,  $\|\mathbf{x}\|_2 \leq n^{-0.2}$ , as desired. By the union bound, the total probability of success is at least  $1 - 20n^{-3} \geq 1 - n^{-2}$ . ■

#### ACKNOWLEDGEMENTS

We thank anonymous reviewers for helpful suggestions which improved this manuscript.

## REFERENCES

- [AGMS11] Bharat Adsul, Jugal Garg, Ruta Mehta, and Milind Sohoni. Rank-1 bimatrix games: a homeomorphism and a polynomial time algorithm. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 195–204, 2011.
- [ALSV13] Noga Alon, Troy Lee, Adi Shraibman, and Santosh S. Vempala. The approximate rank of a matrix and its algorithmic applications: approximate rank. In *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 675–684, 2013.
- [ARSvS10] David Avis, Gabriel D. Rosenberg, Rahul Savani, and Bernhard von Stengel. Enumeration of Nash equilibria for two-player games. *Economic Theory*, 42:9–37, 2010.
- [Ash90] Robert B. Ash. *Information Theory*. Dover Publications, Inc., 1990.
- [Bar18] Siddharth Barman. Approximating nash equilibria and dense subgraphs via an approximate version of carathéodory’s theorem. *SIAM J. Comput.*, 47(3):960–981, 2018.
- [BB17] Maria-Florina Balcan and Mark Braverman. Nash equilibria in perturbation-stable games. *Theory of Computing*, 13(1):1–31, 2017.
- [BLP15] Siddharth Barman, Katrina Ligett, and Georgios Piliouras. Approximating nash equilibria in tree polymatrix games. In *Algorithmic Game Theory - 8th International Symposium, SAGT 2015, Saarbrücken, Germany, September 28-30, 2015, Proceedings*, pages 285–296, 2015.
- [BV06] René Beier and Berthold Vöcking. Typical properties of winners and losers in discrete optimization. *SIAM Journal on Computing*, 35(4):855–881, 2006.
- [BVV07] Imre Bárány, Santosh Vempala, and Adrian Vetta. Nash equilibria in random games. *Random Structures & Algorithms*, 31(4):391–405, 2007.
- [CDO15] Xi Chen, David Durfee, and Anthi Orfanou. On the complexity of nash equilibria in anonymous games. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015, Portland, OR, USA, June 14-17, 2015*, pages 381–390, 2015.
- [CDS17] Yu Cheng, Ilias Diakonikolas, and Alistair Stewart. Playing anonymous games using simple strategies. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 616–631, 2017.
- [CDT06] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Sparse games are hard. In *International Workshop on Internet and Network Economics*, pages 262–273. Springer, 2006.
- [CDT09] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Settling the complexity of computing two-player nash equilibria. *Journal of the ACM (JACM)*, 56(3):14, 2009.
- [Dan98] George Bernard Dantzig. *Linear programming and extensions*, volume 48. Princeton university press, 1998.
- [DFS20] Argyrios Deligkas, John Fearnley, and Rahul Savani. Tree polymatrix games are ppad-hard. *CoRR*, abs/2002.12119, 2020.
- [DGP05] Konstantinos Daskalakis, Paul W. Goldberg, and Christos H. Papadimitriou. The complexity of computing a nash equilibrium. *Electronic Colloquium on Computational Complexity (ECCC)*, (115), 2005. Earlier version of [DGP09].
- [DGP09] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. The complexity of computing a nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.
- [DP15] Constantinos Daskalakis and Christos H. Papadimitriou. Approximate nash equilibria in anonymous games. *J. Econ. Theory*, 156:207–245, 2015.
- [Dzi14] Dainius Dzindzalieta. *Tight Bernoulli tail probability bounds*. PhD thesis, Vilnius University, 2014.
- [EGG06] Edith Elkind, Leslie Ann Goldberg, and Paul W. Goldberg. Nash equilibria in graphical games on trees revisited. In *Proceedings 7th ACM Conference on Electronic Commerce (EC-2006), Ann Arbor, Michigan, USA, June 11-15, 2006*, pages 100–109, 2006.
- [Erd45] Paul Erdős. On a lemma of littlewood and of ford. *Bulletin of the American Mathematical Society*, 51(12):898–902, 1945.
- [HT07] Li-Sha Huang and Shang-Hua Teng. On the approximation and smoothed complexity of leontief market equilibria. In *International Workshop on Frontiers in Algorithmics*, pages 96–107. Springer, 2007.
- [Imp95] Russell Impagliazzo. A personal view of average-case complexity. In *Proceedings of the Tenth Annual Structure in Complexity Theory Conference, Minneapolis, Minnesota, USA, June 19-22, 1995*, pages 134–147, 1995.
- [Jon04] Johan Jonasson. On the optimal strategy in a random game. *Electronic Communications in Probability*, 9:132–139, 2004.
- [Kha79] Leonid Genrikhovich Khachiyan. A polynomial algorithm in linear programming. In *Doklady Akademii Nauk*, volume 244, pages 1093–1096. Russian Academy of Sciences, 1979.
- [KM72] Victor Klee and George J. Minty. How good is the simplex algorithm? In *Inequalities III (Proceedings of the Third Symposium on Inequalities held at the University of California, Los Angeles, Calif., September 1–9, 1969, dedicated to the memory of Theodore S. Motzkin)*, 1972.

- [KN07] Jonathan A Kelner and Evdokia Nikolova. On the hardness and smoothed complexity of quasi-concave minimization. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 472–482. IEEE, 2007.
- [KT10] Ravi Kannan and Thorsten Theobald. Games of fixed rank: A hierarchy of bimatrix games. *Economic Theory*, 42(1):157–173, 2010.
- [LH64] Carlton E Lemke and Joseph T Howson, Jr. Equilibrium points of bimatrix games. *Journal of the Society for industrial and Applied Mathematics*, 12(2):413–423, 1964.
- [LMM03] Richard J Lipton, Evangelos Markakis, and Aranyak Mehta. Playing large games using simple strategies. In *Proceedings of the 4th ACM conference on Electronic commerce*, pages 36–41, 2003.
- [LS18] Zhengyang Liu and Ying Sheng. On the approximation of nash equilibria in sparse win-lose games. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.
- [Meh14] Ruta Mehta. Constant rank bimatrix games are ppad-hard. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 545–554, 2014.
- [Nas51] John Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.
- [Neu28] J v Neumann. Zur theorie der gesellschaftsspiele. *Mathematische annalen*, 100(1):295–320, 1928.
- [OI16] Luis E. Ortiz and Mohammad Tanvir Irfan. FPTAS for mixed-strategy nash equilibria in tree graphical games and their generalizations. *CoRR*, abs/1602.05237, 2016.
- [Rob06] David P Roberts. Nash equilibria of cauchy-random zero-sum and coordination matrix games. *International Journal of Game Theory*, 34(2):167–184, 2006.
- [Rub16] Aviad Rubinfeld. Settling the complexity of computing approximate two-player nash equilibria. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 258–265. IEEE, 2016.
- [RV15] Mark Rudelson and Roman Vershynin. Small ball probabilities for linear images of high-dimensional distributions. *International Mathematics Research Notices*, 2015(19):9594–9617, 2015.
- [Sha87] Ron Shamir. The efficiency of the simplex method: A survey. *Management Science*, 33(3):301–334, 1987.
- [ST04] Daniel A Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.
- [ST06] Daniel A Spielman and S-H Teng. Smoothed analysis of algorithms and heuristics. *LONDON MATHEMATICAL SOCIETY LECTURE NOTE SERIES*, 1(331):274–342, 2006.
- [SV06] Rahul Savani and Bernhard Von Stengel. Hard-to-solve bimatrix games. *Econometrica*, 74(2):397–429, 2006.
- [Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.
- [Wai19] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.

## APPENDIX

### A. Proof of Lemma 3

To prove this lemma, we rely on the following powerful comparison inequality of Talagrand.

**Theorem 14** (Talagrand’s comparison inequality, high-probability version. [Ver18], Exercise 8.6.5). *Suppose that  $\{X_s\}_{s \in S}$  is a collection of  $\mathbb{R}$ -valued random variables, indexed by some  $S \subseteq \mathbb{R}^n$ ,  $0 \notin S$ . Suppose that for all  $s, t \in S$ ,  $X_s - X_t$  is subgaussian with variance proxy at most  $\|s - t\|_2$ . There is a universal constant  $C > 0$  such that for all  $u > 0$ , with probability at least  $1 - \exp(-u^2)$ ,*

$$\sup_{s \in S} X_s \leq C \left( \mathbb{E}_{g \sim \mathcal{N}(0, I)} \sup_{s \in S} \langle g, s \rangle + u \cdot \sup_{s \in S} \|s\|_2 \right).$$

Now we can prove Lemma 3.

**Lemma 3.** *Let  $A$  be an  $n \times n$  matrix with independent subgaussian entries with variance proxy at most 1. For all  $u > 0$ , with probability at least  $1 - \exp(-u^2)$ , all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  with  $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1$  have*

$$\mathbf{x}^\top A \mathbf{y} \leq O(\sqrt{\log n} + u)(\|\mathbf{x}\|_1 + \|\mathbf{y}\|_1).$$

*As a corollary, with the same probability, all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  with  $\|\mathbf{x}\|_1, \|\mathbf{y}\|_1 \leq 1$  have*

$$\mathbf{x}^\top A \mathbf{y} \leq O(\sqrt{\log n} + u)(\|\mathbf{x}\|_2 + \|\mathbf{y}\|_2).$$

*Proof:* Consider for each  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  the random variable  $\mathbf{x}^\top A \mathbf{y} / (\|\mathbf{x}\|_1 + \|\mathbf{y}\|_1)$ . Since the entries of  $A$  are subgaussian with variance proxy 1, there is a universal  $C > 0$  such that  $\langle U, A \rangle$  is subgaussian with variance proxy  $C \|U\|_F^2$ , where  $\|\cdot\|_F$  is the Frobenius norm, for any  $n \times n$  matrix  $U$ . Hence, for  $\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}' \in \mathbb{R}^n$ ,

$$\frac{\mathbf{x}^\top A \mathbf{y}}{\|\mathbf{x}\|_1 + \|\mathbf{y}\|_1} - \frac{(\mathbf{x}')^\top A \mathbf{y}'}{\|\mathbf{x}'\|_1 + \|\mathbf{y}'\|_1}$$

is subgaussian with variance proxy  $C\|\mathbf{x}\mathbf{y}^\top/(\|\mathbf{x}\|_1+\|\mathbf{y}\|_1)-(\mathbf{x}')(\mathbf{y}')^\top/(\|\mathbf{x}'\|_1+\|\mathbf{y}'\|_1)\|_F^2$ . We claim that

$$\begin{aligned} & \left\| \frac{\mathbf{x}\mathbf{y}^\top}{\|\mathbf{x}\|_1+\|\mathbf{y}\|_1} - \frac{(\mathbf{x}')(\mathbf{y}')^\top}{\|\mathbf{x}'\|_1+\|\mathbf{y}'\|_1} \right\|_F^2 \\ & \leq \left\| \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{x}\|_1+\|\mathbf{y}\|_1} - \frac{(\mathbf{x}', \mathbf{y}')}{\|\mathbf{x}'\|_1+\|\mathbf{y}'\|_1} \right\|_2^2, \end{aligned}$$

where  $(\mathbf{x}, \mathbf{y})$  denotes the concatenation of  $\mathbf{x}$  and  $\mathbf{y}$  to a  $2n$ -length vector. To see this, recalling that  $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = \|\mathbf{x}'\|_2 = \|\mathbf{y}'\|_2 = 1$ , let  $m = \|\mathbf{x}\|_1 + \|\mathbf{y}\|_1$  and  $m' = \|\mathbf{x}'\|_1 + \|\mathbf{y}'\|_1$  and expand both sides, it is equivalent to prove

$$\begin{aligned} & \frac{m^2 + (m')^2 - 2m(m')\langle \mathbf{x}, \mathbf{x}' \rangle \langle \mathbf{y}, \mathbf{y}' \rangle}{m^2(m')^2} \\ & \leq \frac{2m^2 + 2(m')^2 - 2m(m')\langle \mathbf{x}, \mathbf{x}' \rangle - 2m(m')\langle \mathbf{y}, \mathbf{y}' \rangle}{m^2(m')^2}. \end{aligned}$$

This is equivalent to

$$\frac{1}{m^2} + \frac{1}{(m')^2} - 2\frac{\langle \mathbf{x}, \mathbf{x}' \rangle + \langle \mathbf{y}, \mathbf{y}' \rangle - \langle \mathbf{x}, \mathbf{x}' \rangle \langle \mathbf{y}, \mathbf{y}' \rangle}{mm'} \geq 0.$$

Dividing by  $2/mm'$  and using  $1/m^2 + 1/(m')^2 \geq 2/mm'$ , it is enough to show

$$1 - \langle \mathbf{x}, \mathbf{x}' \rangle - \langle \mathbf{y}, \mathbf{y}' \rangle - \langle \mathbf{x}, \mathbf{x}' \rangle \langle \mathbf{y}, \mathbf{y}' \rangle \geq 0.$$

This factors as  $(1 - \langle \mathbf{x}, \mathbf{x}' \rangle)(1 - \langle \mathbf{y}, \mathbf{y}' \rangle) \geq 0$  since we assumed  $\mathbf{x}, \mathbf{x}', \mathbf{y}, \mathbf{y}'$  were unit vectors.

Now we can apply Theorem 14 to see that with probability at least  $1 - \exp(-u^2)$ ,

$$\begin{aligned} & \sup_{\substack{\mathbf{x}, \mathbf{y} \\ \|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1}} \frac{\mathbf{x}^\top A \mathbf{y}}{\|\mathbf{x}\|_1 + \|\mathbf{y}\|_1} \\ & \leq C \left( \mathbb{E}_{\mathbf{g} \sim \mathcal{N}(0, I)} \sup_{\substack{\mathbf{x}, \mathbf{y} \\ \|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1}} \frac{\langle (\mathbf{x}, \mathbf{y}), \mathbf{g} \rangle}{\|\mathbf{x}\|_1 + \|\mathbf{y}\|_1} + u \right) \end{aligned}$$

where  $C$  is a universal constant,  $\mathbf{g}$  is a length  $2n$  Gaussian vector with independent coordinates, and we have used that  $\|(\mathbf{x}, \mathbf{y})\|_2 \leq \|(\mathbf{x}, \mathbf{y})\|_1 = \|\mathbf{x}\|_1 + \|\mathbf{y}\|_1$ . To finish the argument, observe that

$$\begin{aligned} & \mathbb{E}_{\mathbf{g} \sim \mathcal{N}(0, I)} \sup_{\substack{\mathbf{x}, \mathbf{y} \\ \|\mathbf{x}\|_2 = \|\mathbf{y}\|_2 = 1}} \frac{\langle (\mathbf{x}, \mathbf{y}), \mathbf{g} \rangle}{\|\mathbf{x}\|_1 + \|\mathbf{y}\|_1} \\ & = \mathbb{E}_{\mathbf{g} \sim \mathcal{N}(0, I)} \|\mathbf{g}\|_\infty \leq O(\sqrt{\log n}). \end{aligned}$$

Finally, to prove the corollary, note that we just showed that with probability at least  $1 - \exp(-u^2)$ , all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  with  $\|\mathbf{x}\|_1 = \|\mathbf{y}\|_1 = 1$  have  $\mathbf{x}^\top A \mathbf{y} / \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \leq O(\sqrt{\log n} + u) \cdot (1/\|\mathbf{x}\|_2 + 1/\|\mathbf{y}\|_2)$ . Multiplying by  $\|\mathbf{x}\|_2 \|\mathbf{y}\|_2$  implies the corollary.  $\blacksquare$

## B. Proof of Lemma 6

1) *Facts about the binomial distribution:* In our result, we need the following bound of Erdős [Erd45].

**Theorem 15** ([Erd45], variant of [Dzi14]). *Let  $a_1, \dots, a_n \geq 1$  be real numbers and  $\epsilon_1, \dots, \epsilon_n$  be Rademacher random variables (uniform distribution on  $\{-1, 1\}$ ) then for all integers  $k \geq 1$ ,*

$$\Pr[a_1 \epsilon_1 + \dots + a_n \epsilon_n \geq k - 1] \geq \Pr[\epsilon_1 + \dots + \epsilon_n \geq k].$$

Furthermore, the following binomial inequality will be useful:

**Lemma 16** ([Ash90]). *For all  $k$  and  $n$ ,*

$$\binom{n}{k} \geq \frac{2^{nH(k/n)}}{\sqrt{8n}},$$

where  $H(\cdot)$  is the binary entropy function.

Note that when  $k = \frac{n}{2}(1 + \delta)$ , then

$$\begin{aligned} H(k/n) & := -\frac{1+\delta}{2} \log_2\left(\frac{1}{2}(1+\delta)\right) - \frac{1-\delta}{2} \log_2\left(\frac{1}{2}(1-\delta)\right) \\ & \geq 1 - \frac{1}{\ln 2} \left( \frac{1+\delta}{2} \cdot \delta + \frac{1-\delta}{2} \cdot (-\delta) \right) \\ & = 1 - (\log_2 e) \delta^2. \end{aligned}$$

Combining with the above inequality gives

$$\frac{1}{2^n} \binom{n}{k} \geq \frac{1}{\sqrt{8n}} e^{-n\delta^2}.$$

This allows us to show the following:

**Claim 17.** *For all integers  $n \geq k \geq 0$  with  $n$  sufficiently large*

$$\frac{1}{2^n} \sum_{i=\frac{n+k}{2}}^n \binom{n}{i} \geq \frac{1}{10000} \exp\left(-\frac{10k^2}{n}\right). \quad (10)$$

*Proof:* Note that here,  $\delta = \frac{k}{n}$ . If  $k \geq n - 2\sqrt{n}$ , then

$$-10k^2/n \leq -10n + 40\sqrt{n} - 400 \leq -9n$$

for  $n$  sufficiently large. Note that the LHS of 10 is at least  $2^{-n} > e^{-9n}$ , and thus is at least the RHS.

On the other hand, if  $k \leq n - 2\sqrt{n}$ , then by Lemma 16, the sum of the first  $\sqrt{n}$  terms is at least

$$\begin{aligned} \sqrt{n} \binom{n}{\frac{n+k}{2} + \sqrt{n}} & \geq \sqrt{n} \frac{1}{\sqrt{8n}} \exp\left(-n \cdot \left(\frac{k + 2\sqrt{n}}{n}\right)^2\right) \\ & = \frac{1}{\sqrt{8}} \exp\left(-\frac{k^2 + 4k\sqrt{n} + 4n}{n}\right) \\ & \geq \frac{1}{\sqrt{8}} \exp\left(-\frac{5k^2 + 5n}{n}\right) \\ & = \frac{1}{e^5 \sqrt{8}} \exp\left(-\frac{5k^2}{n}\right), \end{aligned}$$

which implies the claim.  $\blacksquare$

We recall here the statement of Lemma 6, and give a proof with the above results:

**Lemma 6.** *Assume that  $X$  is the uniform distribution on  $\{-1, 1\}$  (i.e., the Rademacher distribution). There exists a universal constant  $c > 0$  with the following property: For all  $\mathbf{x} \in \mathbb{R}^n$  such that  $\|\mathbf{x}\|_1 = 1$ , with probability at least  $n^{-0.001}$  over  $\mathbf{v} \sim X^n$*

$$\langle \mathbf{v}, \mathbf{x} \rangle \geq c\beta(\mathbf{x}).$$

*Proof:* Recall, we have defined the following: let  $L = \lceil \log_2 n / 2^{100} \rceil$ . Let  $D = 2^{2^{500}}$  and let  $E_1, \dots, E_L$  be intervals such that  $E_i = (D^{-i}, D^{-(i-1)}]$  for all  $i < L$  and  $E_L = [0, D^{-(L-1)}]$ . Let  $\mathbf{x} = \mathbf{x}^{(1)} + \dots + \mathbf{x}^{(L)}$  such that  $\mathbf{x}_j^{(i)} = \mathbf{x}_j \cdot \mathbf{1}[\mathbf{x}_j \in E_i]$ . We say  $\mathbf{x}^{(i)}$  is *sparse* if it has at most  $L$  nonzero coordinates; otherwise it is *dense*. Let  $F \subset \{1, 2, \dots, L\}$  be the set of dense indices. Let  $\mathbf{x}_{\text{sparse}}$  be the sum of the sparse  $\mathbf{x}^{(i)}$ 's and  $\mathbf{x}_{\text{dense}}$  be the sum of the dense ones, and define

$$\beta(\mathbf{x}) = \sqrt{\log n} \|\mathbf{x}_{\text{dense}}\|_2 + \|\mathbf{x}_{\text{sparse}}\|_1.$$

Note that if we drop  $\mathbf{x}^{(L)}$ ,  $\beta$  changes by at most  $\sqrt{\log n} \|\mathbf{x}^{(L)}\|_1 \leq n \sqrt{\log n} \cdot n^{-2^{400}+1}$ , a negligibly small term. Thus, we can without loss of generality assume that  $\mathbf{x}^{(L)} = 0$ .

Since  $\beta(\mathbf{x}) = \|\mathbf{x}_{\text{sparse}}\|_1 + \sqrt{\log n} \|\mathbf{x}_{\text{dense}}\|_2$ , we have for any  $\mathbf{x}$ , at least one of  $\|\mathbf{x}_{\text{sparse}}\|_1$  or  $\sqrt{\log n} \|\mathbf{x}_{\text{dense}}\|_2$  is at least  $\frac{1}{2}\beta(\mathbf{x})$ . Assume we know that with probability at least  $2n^{-0.001}$ ,  $\langle \mathbf{v}, \mathbf{x}_{\text{sparse}} \rangle = \Omega(\|\mathbf{x}_{\text{sparse}}\|_1)$ ; and with probability at least  $2n^{-0.001}$ ,  $\langle \mathbf{v}, \mathbf{x}_{\text{dense}} \rangle = \Omega(\sqrt{\log n} \|\mathbf{x}_{\text{dense}}\|_2)$ . Then, we know with probability at least  $n^{-0.001}$ , one of  $\langle \mathbf{v}, \mathbf{x}_{\text{sparse}} \rangle$  and  $\langle \mathbf{v}, \mathbf{x}_{\text{dense}} \rangle$  is at least  $\Omega(\beta(\mathbf{x}))$  and the other is at least 0 and thus their sum is at least  $\beta(\mathbf{x})$ . We split the remainder of the proof into two parts.

*Part 1,  $\langle \mathbf{v}, \mathbf{x}_{\text{sparse}} \rangle = \Omega(\|\mathbf{x}_{\text{sparse}}\|_1)$ :* Let  $\mathbf{x}'$  be the  $2L$  largest coordinates of  $\mathbf{x}_{\text{sparse}}$ . Note that  $\|\mathbf{x}'\|_1$  is at least  $D^2$  times the sum of the next  $2L$  largest coordinates of  $\mathbf{x}_{\text{sparse}}$  and at least  $D^4$  times the sum of the next  $2L$  largest coordinates after that, etc. Thus,  $\|\mathbf{x}'\|_1 \geq \frac{1}{2}\|\mathbf{x}_{\text{sparse}}\|_1$ .

Now with probability  $1/2^{2L}$ , because  $\mathbf{v}$  has i.i.d. Rademacher entries,  $\langle \mathbf{v}, \mathbf{x}' \rangle = \|\mathbf{x}'\|_1$ , and with probability at least  $1/2$ ,  $\langle \mathbf{v}, \mathbf{x}_{\text{sparse}} - \mathbf{x}' \rangle \geq 0$ . Thus, with probability at least  $1/2^{2L+1} \geq 2n^{-0.001}$ ,  $\langle \mathbf{v}, \mathbf{x}_{\text{sparse}} \rangle \geq \frac{1}{2}\|\mathbf{x}_{\text{sparse}}\|_1$ .

*Part 2,  $\langle \mathbf{v}, \mathbf{x}_{\text{dense}} \rangle = \Omega(\sqrt{\log n} \|\mathbf{x}_{\text{dense}}\|_2)$ :* Since  $\mathbf{x}_{\text{dense}} = \sum_{i \in F} \mathbf{x}^{(i)}$ , we have that

$$\Pr \left[ \langle \mathbf{v}, \mathbf{x}_{\text{dense}} \rangle \geq \frac{\sqrt{\log n}}{1000D} \|\mathbf{x}_{\text{dense}}\|_2 \right] \quad (11)$$

$$\geq \prod_{i \in F} \Pr \left[ \langle \mathbf{v}, \mathbf{x}^{(i)} \rangle \geq \frac{\sqrt{\log n} \cdot \|\mathbf{x}^{(i)}\|_2}{1000D \|\mathbf{x}_{\text{dense}}\|_2} \right] \quad (12)$$

Consider  $i \in F$ , and let  $m_i \geq L + 1$  be the support size of  $\mathbf{x}^{(i)}$ . Since  $\mathbf{x}_j^{(i)} D^i \geq 1$  for all  $j$  in the support of  $\mathbf{x}^{(i)}$ , we have by Theorem 15 and Claim 17, that for any integer  $k \in [0, m_i]$

$$\begin{aligned} \Pr \left[ \langle \mathbf{v}, \mathbf{x}^{(i)} \rangle \geq \frac{k}{D^i} \right] &\geq \sum_{i=\frac{m_i+k}{2}+1}^{m_i} \binom{m_i}{i} \\ &\geq \frac{1}{10000} \exp \left( -10m_i \left( \frac{k+2}{m_i} \right)^2 \right). \end{aligned}$$

Observe that  $\|\mathbf{x}^{(i)}\|_2 \leq \sqrt{m_i} \|\mathbf{x}^{(i)}\|_\infty \leq \sqrt{m_i} D^{-(i-1)}$ . Thus,

$$\begin{aligned} \Pr \left[ \langle \mathbf{v}, \mathbf{x}^{(i)} \rangle \geq \frac{k}{D \sqrt{m_i}} \|\mathbf{x}^{(i)}\|_2 \right] \\ \geq \frac{1}{10000} \exp \left( -10m_i \left( \frac{k+2}{m_i} \right)^2 \right). \end{aligned}$$

Let

$$k = \left\lceil \frac{1}{1000} \sqrt{m_i \log n} \cdot \frac{\|\mathbf{x}^{(i)}\|_2}{\|\mathbf{x}_{\text{dense}}\|_2} \right\rceil.$$

Then, note that

$$\frac{k+2}{m_i} \leq \frac{3}{m_i} + \frac{1}{1000} \sqrt{\frac{\log n}{m_i}} \cdot \frac{\|\mathbf{x}^{(i)}\|_2}{\|\mathbf{x}_{\text{dense}}\|_2}$$

Thus, since  $(a+b)^2 \leq 2a^2 + 2b^2$ ,

$$-10m_i \cdot \left( \frac{k+2}{m_i} \right)^2 \geq -\frac{180}{m_i} - \frac{\log n}{5 \cdot 10^4} \cdot \frac{\|\mathbf{x}^{(i)}\|_2^2}{\|\mathbf{x}_{\text{dense}}\|_2^2}.$$

Therefore,

$$\begin{aligned} \Pr \left[ \langle \mathbf{v}, \mathbf{x}^{(i)} \rangle \geq \frac{\sqrt{\log n}}{1000D} \cdot \frac{\|\mathbf{x}^{(i)}\|_2^2}{\|\mathbf{x}_{\text{dense}}\|_2} \right] \\ \geq \frac{1}{10^4} \exp \left( -\frac{180}{m_i} - \frac{\log n}{5 \cdot 10^4} \cdot \frac{\|\mathbf{x}^{(i)}\|_2^2}{\|\mathbf{x}_{\text{dense}}\|_2^2} \right). \end{aligned}$$

Applying Eq. 12, and noting that each  $m_i \geq L \geq |F|$ .

$$\begin{aligned} \Pr \left[ \langle \mathbf{v}, \mathbf{x}_{\text{dense}} \rangle \geq \frac{\sqrt{\log n}}{1000D} \|\mathbf{x}_{\text{dense}}\|_2 \right] \\ \geq \frac{1}{10^{4L}} \exp \left( -\sum_{i \in F} \frac{180}{m_i} - \frac{\log n}{5 \cdot 10^4} \right) \\ = \frac{1}{10^{4L} e^{180}} n^{-10^{-5}} \\ \geq n^{-2^{-90}} n^{-10^{-5}} \\ \geq n^{-0.001}, \end{aligned}$$

For  $n$  sufficiently large. This concludes the proof.  $\blacksquare$