

Approximating Nash Equilibria and Dense Subgraphs via an Approximate Version of Carathéodory's Theorem

Siddharth Barman*

Abstract

We present algorithmic applications of an approximate version of Carathéodory's theorem. The theorem states that given a set of vectors X in \mathbb{R}^d , for every vector in the convex hull of X there exists an ε -close (under the p -norm distance, for $2 \leq p < \infty$) vector that can be expressed as a convex combination of at most b vectors of X , where the bound b depends on ε and the norm p and is *independent* of the dimension d . This theorem can be derived by instantiating Maurey's lemma, early references to which can be found in the work of Pisier (1981) and Carl (1985). However, in this paper we present a self-contained proof of this result.

Using this theorem we establish that in a bimatrix game with $n \times n$ payoff matrices A, B , if the number of non-zero entries in any column of $A + B$ is at most s then an ε -Nash equilibrium of the game can be computed in time $n^{O(\frac{\log s}{\varepsilon^2})}$. This, in particular, gives us a polynomial-time approximation scheme for Nash equilibrium in games with fixed column sparsity s . Moreover, for arbitrary bimatrix games—since s can be at most n —the running time of our algorithm matches the best-known upper bound, which was obtained by Lipton, Markakis, and Mehta (2003).

The approximate Carathéodory's theorem also leads to an *additive* approximation algorithm for the normalized densest k -subgraph problem. Given a graph with n vertices and maximum degree d , the developed algorithm determines a subgraph with exactly k vertices with normalized density within ε (in the additive sense) of the optimal in time $n^{O(\frac{\log d}{\varepsilon^2})}$. Additionally, we show that a similar approximation result can be achieved for the problem of finding a $k \times k$ -bipartite subgraph of maximum normalized density.

1 Introduction

Carathéodory's theorem is a fundamental dimensionality result in convex geometry. It states that any vector in the convex hull of a set X in \mathbb{R}^d can be expressed as a convex combination of at most $d + 1$ vectors of X .¹ This paper considers a natural approximate version of Carathéodory's theorem where the goal is to seek convex combinations that are close enough to vectors in the convex hull.

*California Institute of Technology. barman@caltech.edu

¹This bound of $d + 1$ is tight.

Specifically, this approximate version establishes that given a set of vectors X in the p -unit ball² with norm $p \in [2, \infty)$, for every vector μ in the convex hull of X there exists an ε -close—under the p -norm distance—vector μ' that can be expressed as a convex combination of $\frac{4p}{\varepsilon^2}$ vectors of X . A notable aspect of this result is that the number of vectors of X that are required to express μ' , i.e., $\frac{4p}{\varepsilon^2}$, is independent of the underlying dimension d . This theorem can be derived by instantiating Maurey’s lemma, early references to which can be found in the work of Pisier [30] and Carl [10]. However, in this paper we present a self-contained proof of this result, which we proceed to outline below. The author was made aware of the connection with Maurey’s lemma after a preliminary version of this work had appeared.

To establish the approximate version of Carathéodory’s theorem we use the probabilistic method. Given a vector μ in the convex hull of a set $X \subset \mathbb{R}^d$, consider a convex combination of vectors of X that generates μ . The coefficients in this convex combination induce a probability distribution over X and the mean of this distribution is μ . The approach is to draw b independent and identically distributed (i.i.d.) samples from this distribution and show that with positive probability the sample mean, with an appropriate number of samples, is close to μ under the p -norm distance, for $p \in [2, \infty)$. Therefore, the probabilistic method implies that there exists a vector close to μ that can be expressed as a convex combination of at most b vectors, where b is the number of samples we drew.

Note that in this context applying the probabilistic method is a natural idea, but a direct application of this method will not work. Specifically, a dimension-free result is unlikely if we first try to prove that the i th component of the sample mean vector is close to the i th component of μ , for every $i \in [d]$; since this would entail a union bound over the number of components d . Bypassing such a component-wise analysis requires the use of atypical ideas. We are able to accomplish this task and, in particular, bound (in expectation) the p -norm distance between μ and the sample mean vector via an interesting application of Khintchine inequality (see Theorem 1).

Given the significance of Carathéodory’s theorem, this approximate version is interesting in its own right. The key contribution of the paper is to substantiate the algorithmic relevance of this approximate version by developing new algorithmic applications. Our applications include additive approximation algorithms for (i) Nash equilibria in two-player games, and (ii) the densest subgraph problem. These algorithmic results are outlined below.

Algorithmic Applications

Approximate Nash Equilibria. Nash equilibria are central constructs in game theory that are used to model likely outcomes of strategic interactions between self-interested entities, like human players. They denote distributions over actions of players under which no player can benefit, in expectation, by unilateral deviation. These solution concepts are arguably the most well-studied

²That is, X is contained in the set $\{v \in \mathbb{R}^d \mid \|v\|_p \leq 1\}$.

notions of rationality and questions about their computational complexity lie at the core of algorithmic game theory. In recent years, hardness results have been established for Nash equilibrium, even in two-player games [12, 14]. But, the question whether an *approximate* Nash equilibrium can be computed in polynomial time still remains open. Throughout this paper we will consider the standard additive notion of approximate Nash equilibria that are defined as follows: a pair distributions, one for each player, is said to be an ε -Nash equilibrium if any unilateral deviation increases utility by at most ε , in expectation.

We apply the approximate version of Carathéodory’s theorem to address this central open question. Specifically, we prove that in a bimatrix game with $n \times n$ payoff matrices A, B , i.e., a two-player game with n actions for each player, if the number of non-zero entries in any column of $A + B$ is at most s then an ε -Nash equilibrium of the game can be computed in time $n^{O\left(\frac{\log s}{\varepsilon^2}\right)}$. Our result, in particular, shows that games with fixed column sparsity s admit a polynomial-time approximation scheme (PTAS) for Nash equilibrium. Along the lines of zero-sum games (which model strict competition), games with fixed column sparsity capture settings in which, except for particular action profiles, the gains and losses of the two player balance out. In other words, such games are a natural generalization of zero-sum games; recall that zero-sum games admit efficient computation of Nash equilibrium (see, e.g., [29]).

It is also worth pointing out that for an arbitrary bimatrix game the running time of our algorithm is $n^{O\left(\frac{\log n}{\varepsilon^2}\right)}$, since s is at most n . Given that the best-known algorithm for computing ε -Nash equilibrium also runs in time $n^{O\left(\frac{\log n}{\varepsilon^2}\right)}$ [25], for general games the time complexity of our algorithm matches the best-known upper bound. Overall, this result provides a parameterized understanding of the complexity of computing approximate Nash equilibrium in terms of a very natural measure, the column sparsity s of the matrix $A + B$.

Our framework can address other notions of sparsity as well. Specifically, if there exist *constants* $\alpha, \beta \in \mathbb{R}_+$ and $\gamma \in \mathbb{R}$ such that the matrix $\alpha A + \beta B + \gamma \mathbb{1}_{n \times n}$ has column or row sparsity s , then our algorithm can be directly adopted to find an ε -Nash equilibrium of the game (A, B) in time $n^{O\left(\frac{\log s}{\varepsilon^2}\right)}$; here, $\mathbb{1}_{n \times n}$ is the all-ones $n \times n$ matrix.³ Additionally, the same running-time bound can be achieved for approximating Nash equilibrium in games wherein *both* matrices A and B have column or row sparsity s . Note that this case is not subsumed by the previous result; in particular, if the columns of matrix A and the rows of matrix B are sparse, then it is not necessary that $A + B$ has low column or row sparsity.

We also refine the following result of Daskalakis and Papadimitriou [17]: They develop a PTAS for bimatrix games that admit an equilibrium with small, specifically $O\left(\frac{1}{n}\right)$, probability values. This result is somewhat surprising, since such small-probability equilibria have large, $\Omega(n)$, support, and hence are not amenable to, say, exhaustive search. We show that if a game has an equilibrium with probability values $O\left(\frac{1}{m}\right)$, for $m \in [n]$, then an approximate equilibrium can be computed in time

³Note that given matrices A and B , parameters α , β , and γ can efficiently computed.

n^t , where $t = O\left(\frac{\log(s/m)}{\varepsilon^2}\right)$. Since $s \leq n$, we get the result of [17] as a special case.

Densest Subgraph. In the normalized densest k -subgraph problem (NDkS) we are given a simple graph and the objective is to find a size- k subgraph (i.e., a subgraph containing exactly k vertices) of maximum density; here, density is normalized to be at most one, i.e., for a subgraph with k vertices, it is defined to be the number of edges in the subgraph divided by k^2 . NDkS is simply a normalized version of the standard densest k -subgraph problem (see, e.g., [1] and references therein) wherein the goal is to find a subgraph with k vertices with the maximum possible number of edges in it. The densest k -subgraph problem (DkS) is computationally hard and it is shown in [1] that a constant-factor approximation for DkS is unlikely. This result implies that NDkS is hard to approximate (multiplicatively) within a constant factor as well.

In this paper we focus on an additive approximation for NDkS. In particular, our objective is to compute a size- k subgraph whose density is close (in the additive sense) to the optimal. The paper also presents additive approximations for the densest k -bipartite subgraph (DkBS) problem. DkBS is a natural variant of NDkS and the goal in this problem is to find size- k vertex subsets of maximum density. In the bipartite case, density of vertex subsets S and T is defined to be the number of edges between the two subsets divided by $|S||T|$.

Hardness of additively approximating DkBS was studied by Hazan and Krauthgamer [20]. Specifically, the reduction in [20] rules out an additive PTAS for DkBS, under complexity theoretic assumptions.⁴ In terms of upper bound, the result of Alon et al. [3] presents an algorithm for this problem that runs in time exponential in the rank of the adjacency matrix.

This paper develops the following complementary upper bounds: given a graph with n vertices and maximum degree d , an ε -additive approximation for NDkS can be computed in time $n^{O\left(\frac{\log d}{\varepsilon^2}\right)}$. This paper also presents an algorithm with the same time complexity for additively approximating DkBS.

1.1 Related Work

Approximate Version of Carathéodory's Theorem. In this paper we provide a self-contained proof of the approximate version of Carathéodory's theorem, employing the Khintchine inequality (see Theorem 1), and use the theorem to develop new approximation algorithms. As mentioned earlier, the approximate version of Carathéodory's theorem can also be obtained by instantiating Maurey's lemma, which, in particular, appears in the analysis and operator theory literatures; see, e.g., [30, 10, 8].

⁴They reduce the problem of determining a *planted clique* to that of computing an ε -additive approximation for DkBS, with a sufficiently small but constant ε .

Approximate Nash Equilibria. The computation of equilibria is an active area of research. Nash equilibria is known to be computationally hard [12, 14], and in light of these findings, a considerable effort has been directed towards understanding the complexity of *approximate* Nash equilibrium. Results in this direction include both upper bounds [25, 22, 15, 21, 16, 23, 18, 7, 33, 34, 3, 2] and lower bounds [20, 13, 9]. In particular, it is known that for a general bimatrix game an approximate Nash equilibrium can be computed in quasi-polynomial time [25]. Polynomial time algorithms have been developed for computing approximate Nash equilibria for fixed values of the approximation factor ε ; the best-known result of this type shows that a 0.3393-approximate Nash equilibrium can be computed in polynomial time [33]. In addition, several interesting classes of games have been identified that admit a PTAS [21, 17, 34, 3, 2]. For example, the result of Alon et al. [3] provides a PTAS for games in which the sum of the payoff matrices, i.e., $A + B$, has logarithmic rank. Our result is incomparable to such rank based results, since a sparse matrix can have high rank and a low-rank matrix can have high sparsity.

Chen et al. [11] considered sparsity in the context of games and showed that computing an exact Nash equilibrium is hard even if *both* the payoff matrices have a fixed number of non-zero entries in every row *and* column. It was observed in [17] that such games admit a trivial PTAS.⁵ Note that we study a strictly larger class of games and provide a PTAS for games in which the row *or* column sparsity of $A + B$ is fixed.

Densest Subgraph. The best-known (multiplicative) approximation ratio for the densest k -subgraph problem is $n^{(1/4+o(1))}$ [6]. But unlike this result, our work addresses additive approximations with normalized density as the maximization objective. In particular, we approximate NDkS by approximately solving a quadratic program, which is similar to the quadratic program used in the Motzkin-Straus theorem [28].

In addition, our approximation algorithm for DkBS is based on solving a bilinear program that was formulated by Alon et al. [3]. This bilinear program was used in [3] to develop an additive PTAS for DkBS in particular classes of graphs, including ones with low-rank adjacency matrices. This paper supplements prior work by developing an approximation algorithm whose running time is parametrized by the maximum degree of the given graph, and not by the rank of its adjacency matrix.

1.2 Techniques

Approximate Nash Equilibria. Our algorithm for computing an approximate Nash equilibrium relies on finding a near-optimal solution of a bilinear program (BP). The BP we consider was formulated by Mangasarian and Stone [26] and its optimal (near-optimal) solutions correspond to

⁵In particular, the product of uniform distributions over players' actions corresponds to an approximate Nash equilibrium in such games.

exact (approximate) Nash equilibria of the given game. Below we provide a sketch of our algorithm that determines a near-optimal solution of this BP.

The variables of the BP, x and y , correspond to probability distributions that are mixed strategies of the players and its objective is to maximize $x^T C y$, where C is the sum of the payoff matrices of the game.⁶ Suppose we knew the vector $u := C\hat{y}$, for some Nash equilibrium (\hat{x}, \hat{y}) . Then, a Nash equilibrium can be efficiently computed by solving a linear program (with variables x and y) that is obtained by modifying the BP as follows: replace $x^T C y$ by $x^T u$ as the objective and include the constraint $Cy = u$. Section 4 shows that this idea can be used to find an approximate Nash equilibrium, even if u is not exactly equal to $C\hat{y}$ but close to it. That is, to find an approximate Nash equilibrium it suffices to have a vector u for which $\|C\hat{y} - u\|_p$ is small.

To apply the approximate version of Carathéodory's theorem we observe that $C\hat{y}$ is a vector in the convex hull of the *columns* of C . Also, note that in the context of (additive) approximate Nash equilibria the payoff matrices are normalized, hence the absolute value of any entry of matrix C is no more than, say, 2. This entry-wise normalization implies that if no column of matrix C has more than s non-zero entries, then the log s norm of the columns is a fixed constant: $\|C^i\|_p \leq (s \cdot 2^p)^{1/p} = 2 \cdot 2^{\frac{\log s}{p}} \leq 4$, where C^i is the i th column of C and norm $p = \log s$. This is a simple but critical observation, since it implies that, modulo a small scaling factor, the columns of an C lie in the log s -unit ball. At this point we can apply the approximate version of Carathéodory's theorem to guarantee that close to $C\hat{y}$ there exists a vector u that can be expressed as a convex combination of about $p = \log s$ columns of C . We show in Section 4 that exhaustively searching for u takes $n^{O(\log s)}$ time, where n is the number of columns of C . Thus we can find a vector close to $C\hat{y}$ and hence determine a near-optimal solution of the bilinear program. This way we get an approximate Nash equilibrium and the running time of the algorithm is dominated by the exhaustive search.

Overall, this template for approximating Nash equilibria in sparse games is made possible by the approximate version of Carathéodory's theorem. It is notable that our algorithmic framework employs arbitrary norms $p \in [2, \infty)$, and in this sense it goes beyond standard ε -*net*-based results that typically use norms 1, 2, or ∞ .

Densest Subgraph. The algorithmic approach outlined above applies to any quadratic or bilinear program in which the objective matrix is column (or row) sparse and the feasible region is contained in the simplex. We use this observation to develop an additive approximations for NDkS and DkBS.

Specifically, we formulate a quadratic program, near-optimal solutions of which correspond to approximate-solutions of NDkS. The column sparsity of the objective matrix in the quadratic program is equal to the maximum degree of the underlying graph plus one. Hence, using the above mentioned observation, we obtain the approximation result for NDkS. The same template applies to DkBS; for this problem we employ a bilinear program from [3].

⁶We ignore the linear part of the objective for ease of presentation, see Section 4 for details.

1.3 Organization

We begin by setting up notation in Section 2. Then, in Section 3 we present the approximate version of Carathéodory’s theorem. Algorithmic applications of the theorem are developed in Sections 4 and 5. In Section 6 we consider convex hulls of matrices and also detail approximate versions of the colorful Carathéodory theorem and Tverberg’s theorem. Finally, Section 7 presents a lower bound proving showing that, in general, ε -close (under the p -norm distance with $p \in [2, \infty)$) vectors cannot be expressed as a convex combination of less than $\frac{1}{4 \varepsilon^{p/(p-1)}}$ vectors of the given set.

2 Notation

Write $\|x\|_p$ to denote the p -norm of a vector $x \in \mathbb{R}^d$. The Euclidean norm is denoted by $\|x\|$, i.e., we drop the subscript from $\|x\|_2$. The number of non-zero components of a vector x is specified via the ℓ_0 “norm”: $\|x\|_0 := |\{i \mid x_i \neq 0\}|$. Let Δ^n be the set of probability distributions over the set $[n]$. For $x \in \Delta^n$, we define $\text{Supp}(x) := \{i \mid x_i \neq 0\}$. Similarly, for a vector $v \in \mathbb{R}^n$ write $\text{Supp}(v)$ to denote the set $\{i \mid v_i \neq 0\}$.

Given a set $X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^d$, we use the standard abbreviation $\text{conv}(X)$ for the convex hull of X . A vector $y \in \text{conv}(X)$ is said to be k uniform with respect to X if there exists a size k multiset S of $[n]$ such that $y = \frac{1}{k} \sum_{i \in S} x_i$. In particular, if vector y is k uniform with respect to X then y can be expressed as a convex combination of at most k vectors from X . Throughout, the set X will be clear from context so we will simply say that a vector is k uniform and not explicitly mention the fact that uniformity is with respect to X .

3 Approximate Version of Carathéodory’s Theorem

A key technical ingredient in our proof is Khintchine inequality (see, e.g., [19] and [24]). The following form of the inequality is derived from a result stated in [31].

Theorem 1 (Khintchine Inequality). *Let r_1, r_2, \dots, r_m be a sequence of i.i.d. Rademacher ± 1 random variables, i.e., $\Pr(r_i = \pm 1) = \frac{1}{2}$ for all $i \in [m]$. In addition, let $u_1, u_2, \dots, u_m \in \mathbb{R}^d$ be a deterministic sequence of vectors. Then, for $2 \leq p < \infty$*

$$\mathbb{E} \left\| \sum_{i=1}^m r_i u_i \right\|_p \leq \sqrt{p} \left(\sum_{i=1}^m \|u_i\|_p^2 \right)^{\frac{1}{2}}. \quad (1)$$

Proof. Given vector $v \in \mathbb{R}^d$, write $\text{diag}(v)$ to denote the $d \times d$ diagonal matrix whose diagonal is equal to v . Note that if matrix $Q = \text{diag}(v)$ then $\|Q\|_{S_p} = \|v\|_p$, where $\|Q\|_{S_p}$ denotes the Schatten p -norm of Q , i.e., $\|Q\|_{S_p} = \|\sigma(Q)\|_p$, where $\sigma(Q)$ is the vector of singular values of Q . In addition, say we construct diagonal matrices for a sequence of vectors $u_1, u_2, \dots, u_m \in \mathbb{R}^d$, i.e.,

set $Q_i = \text{diag}(u_i)$ for all $i \in [m]$, then for any sequence of scalars $\xi_1, \xi_2, \dots, \xi_m \in \mathbb{R}$ we have $\sum_{i=1}^m \xi_i Q_i = \text{diag}(\sum_{i=1}^m \xi_i u_i)$.

In order to prove the theorem statement for vectors $u_1, u_2, \dots, u_m \in \mathbb{R}^d$, we simply use Theorem 2 of [31]. In particular, setting $Q_i = \text{diag}(u_i)$ for all $i \in [m]$ in Theorem 2 of [31] and using the above stated observations we get:

$$\mathbb{E} \left\| \sum_{i=1}^m r_i u_i \right\|_p^p \leq p^{p/2} \left(\sum_{i=1}^m \|u_i\|_p^2 \right)^{\frac{p}{2}}. \quad (2)$$

For $p \geq 2$ the p th root is a concave function; hence Jensen's inequality, applied to (2), gives us the desired result. □

We are ready to prove the main result of this section. Note that in the following theorem the scaling term γ is defined with respect to the p norm.

Theorem 2. *Given a set of vectors $X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^d$ and $\varepsilon > 0$. For every $\mu \in \text{conv}(X)$ and $2 \leq p < \infty$ there exists an $\frac{4p\gamma^2}{\varepsilon^2}$ uniform vector $\mu' \in \text{conv}(X)$ such that $\|\mu - \mu'\|_p \leq \varepsilon$. Here, $\gamma := \max_{x \in X} \|x\|_p$.*

Proof. Express $\mu \in \text{conv}(X)$ as a convex combination of x_i s: $\mu = \sum_{i=1}^n \alpha_i x_i$ where $\alpha_i \geq 0$, for all $i \in [n]$, and $\sum_{i=1}^n \alpha_i = 1$. Note that $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$ corresponds to a probability distribution over vectors x_1, x_2, \dots, x_n . That is, under probability distribution α vector x_i is drawn with probability α_i . The vector μ is the mean of this distribution. Specifically, the j th component of μ is the expected value of the random variable that takes value $x_{i,j}$ with probability α_i , here $x_{i,j}$ is the j th component of vector x_i . We succinctly express these component-wise equalities as follows:

$$\mathbb{E}_{v \sim \alpha}[v] = \mu. \quad (3)$$

Let v_1, v_2, \dots, v_m be m i.i.d. draws from α . The sample mean vector is defined to be $\frac{1}{m} \sum_{i=1}^m v_i$. Below we specify function $g : X^m \rightarrow \mathbb{R}$ to quantify the p -norm distance between the sample mean vector and the μ .

$$g(v_1, v_2, \dots, v_m) := \left\| \frac{1}{m} \sum_{i=1}^m v_i - \mu \right\|_p. \quad (4)$$

The key technical part of the remainder of the proof is to show that

$$\mathbb{E}[g] \leq \frac{2\sqrt{p}\gamma}{\sqrt{m}}. \quad (5)$$

For $m = \frac{4p\gamma^2}{\varepsilon^2}$ this inequality reduces to $\mathbb{E}[g] \leq \varepsilon$. Therefore, when the number of samples $m = \frac{4p\gamma^2}{\varepsilon^2}$ we have $\Pr(g \leq \varepsilon) > 0$, i.e., with positive probability the sample mean vector is ε close to μ in the p -norm. Overall, the stated claim is implied by the probabilistic method.

Recall that in expectation the sampled mean is equal to μ , i.e., $\mathbb{E}_{v'_1, \dots, v'_m \sim \alpha} \frac{1}{m} \sum_{i=1}^m v'_i = \mu$. Hence, we have

$$\mathbb{E}[g] = \mathbb{E}_{v_1, \dots, v_m} \left\| \frac{1}{m} \sum_{i=1}^m v_i - \mu \right\|_p \quad (6)$$

$$= \mathbb{E}_{v_1, \dots, v_m} \left\| \frac{1}{m} \sum_{i=1}^m v_i - \mathbb{E}_{v'_1, \dots, v'_m} \frac{1}{m} \sum_{i=1}^m v'_i \right\|_p \quad (7)$$

$$= \mathbb{E}_{v_1, \dots, v_m} \left\| \mathbb{E}_{v'_1, \dots, v'_m} \left(\frac{1}{m} \sum_{i=1}^m v_i - \frac{1}{m} \sum_{i=1}^m v'_i \right) \right\|_p. \quad (8)$$

Note that $\|\cdot\|_p$ is convex for $p \geq 1$. Therefore, Jensen's inequality gives us:

$$\mathbb{E}_{v_1, \dots, v_m} \left\| \mathbb{E}_{v'_1, \dots, v'_m} \left(\frac{1}{m} \sum_{i=1}^m v_i - \frac{1}{m} \sum_{i=1}^m v'_i \right) \right\|_p \leq \mathbb{E}_{v_1, \dots, v_m} \mathbb{E}_{v'_1, \dots, v'_m} \left\| \left(\frac{1}{m} \sum_{i=1}^m v_i - \frac{1}{m} \sum_{i=1}^m v'_i \right) \right\|_p \quad (9)$$

$$= \frac{1}{m} \mathbb{E}_{v_1, \dots, v_m, v'_1, \dots, v'_m} \left\| \left(\sum_{i=1}^m (v_i - v'_i) \right) \right\|_p. \quad (10)$$

Let r_1, r_2, \dots, r_m be a sequence of i.i.d. Rademacher ± 1 random variables, i.e., $\Pr(r_i = \pm 1) = \frac{1}{2}$ for all $i \in [m]$. Since, for all $i \in [m]$, v_i and v'_i are i.i.d. copies we can write

$$\begin{aligned} \frac{1}{m} \mathbb{E}_{v_i, v'_i} \left\| \left(\sum_{i=1}^m (v_i - v'_i) \right) \right\|_p &= \frac{1}{m} \mathbb{E}_{v_i, v'_i, r_i} \left\| \left(\sum_{i=1}^m r_i (v_i - v'_i) \right) \right\|_p \\ &\leq \frac{1}{m} \mathbb{E}_{v_i, v'_i, r_i} \left[\left\| \sum_{i=1}^m r_i v_i \right\|_p + \left\| \sum_{i=1}^m r_i v'_i \right\|_p \right] \quad (\text{Triangle inequality}) \\ &= \frac{1}{m} \mathbb{E}_{r_i} \left[\mathbb{E}_{v_i, v'_i} \left(\left\| \sum_{i=1}^m r_i v_i \right\|_p + \left\| \sum_{i=1}^m r_i v'_i \right\|_p \mid r_1, \dots, r_m \right) \right] \quad (\text{Tower property}) \\ &= \frac{1}{m} \mathbb{E}_{r_i} \left[\mathbb{E}_{v_i} \left(\left\| \sum_{i=1}^m r_i v_i \right\|_p \mid r_1, \dots, r_m \right) + \mathbb{E}_{v'_i} \left(\left\| \sum_{i=1}^m r_i v'_i \right\|_p \mid r_1, \dots, r_m \right) \right] \\ &= \frac{1}{m} \mathbb{E}_{r_i} \left[2 \mathbb{E}_{v_i} \left(\left\| \sum_{i=1}^m r_i v_i \right\|_p \mid r_1, \dots, r_m \right) \right] \\ &= 2 \mathbb{E}_{v_i, r_i} \left\| \sum_{i=1}^m r_i \frac{v_i}{m} \right\|_p. \quad (11) \end{aligned}$$

The penultimate equality follows from the following (v_i s and v'_i s are i.i.d. copies)

$$\mathbb{E}_{v_i} \left(\left\| \sum_{i=1}^m r_i v_i \right\|_p \middle| r_1, \dots, r_m \right) = \mathbb{E}_{v'_i} \left(\left\| \sum_{i=1}^m r_i v'_i \right\|_p \middle| r_1, \dots, r_m \right). \quad (12)$$

Overall, inequalities (8), (10), and (11) imply

$$\mathbb{E}[g] \leq 2 \mathbb{E}_{v_i, r_i} \left\| \sum_{i=1}^m r_i \frac{v_i}{m} \right\|_p, \quad (13)$$

where r_1, r_2, \dots, r_m is a sequence of i.i.d. Rademacher ± 1 random variables.

At this point we can apply Theorem 1 (Khinchine inequality) with $u_i = \frac{v_i}{m}$ to obtain

$$\mathbb{E}_{v_i, r_i} \left\| \sum_{i=1}^m r_i \frac{v_i}{m} \right\|_p = \mathbb{E}_{v_i} \left[\mathbb{E}_{r_i} \left[\left\| \sum_{i=1}^m r_i \frac{v_i}{m} \right\|_p \middle| v_1, \dots, v_m \right] \right] \quad (14)$$

$$\leq \mathbb{E}_{v_i} \left[\sqrt{p} \left(\sum_{i=1}^m \left\| \frac{v_i}{m} \right\|_p^2 \right)^{1/2} \right] \quad (15)$$

$$\leq \mathbb{E}_{v_i} \left[\sqrt{p} \left(\sum_{i=1}^m \frac{\gamma^2}{m^2} \right)^{1/2} \right] \quad (16)$$

$$= \sqrt{p} \frac{\gamma}{\sqrt{m}}. \quad (17)$$

Inequality (16) uses the fact that random vectors v_i are supported over X , so $\|v_i\|_p^2 \leq \gamma^2$.

Using (13) and (17) we get

$$\mathbb{E}[g] \leq \frac{2\sqrt{p} \gamma}{\sqrt{m}}. \quad (18)$$

This completes the proof. □

We end this section by stating an ∞ -norm variant of our result. This theorem follows directly from Hoeffding's inequality. Note that in the following theorem the scaling of vectors in X is with respect to the ∞ -norm.

Theorem 3. *Given a set of vectors $X = \{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^d$, with $\max_{x \in X} \|x\|_\infty \leq 1$, and $\varepsilon > 0$. For every $\mu \in \text{conv}(X)$ there exists an $O\left(\frac{\log n}{\varepsilon^2}\right)$ uniform vector $\mu' \in \text{conv}(X)$ such that $\|\mu - \mu'\|_\infty \leq \varepsilon$.*

Proof. Apply Hoeffding's inequality component wise and take union bound. □

4 Computing Approximate Nash Equilibrium

Bimatrix Games. Bimatrix games are two player games in normal form. Such games are specified by a pair of $n \times n$ matrices (A, B) , which are termed the payoff matrices for the players. The first player, also called the row player, has payoff matrix A , and the second player, or the column player, has payoff matrix B . The strategy set for each player is $[n] = \{1, 2, \dots, n\}$, and, if the row player plays strategy i and column player plays strategy j , then the payoffs of the two players are A_{ij} and B_{ij} respectively. The payoffs of the players are normalized between -1 and 1 , so $A_{ij}, B_{ij} \in [-1, 1]$ for all $i, j \in [n]$.

Recall that Δ^n is the set of probability distributions over the set of pure strategies $[n]$. We use $e_i \in \mathbb{R}^n$ to denote the vector with 1 in the i th coordinate and 0's elsewhere. The players can randomize over their strategies by selecting any probability distribution in Δ^n , called a mixed strategy. When the row and column players play mixed strategies x and y respectively, the expected payoff of the row player is $x^T A y$ and the expected payoff of the column player is $x^T B y$.

Definition 1 (Nash Equilibrium). *A mixed strategy pair (x, y) , $x, y \in \Delta^n$, is said to be a Nash equilibrium if and only if:*

$$x^T A y \geq e_i^T A y \quad \forall i \in [n] \quad \text{and} \quad (19)$$

$$x^T B y \geq x^T B e_j \quad \forall j \in [n]. \quad (20)$$

By definition, if (x, y) is a Nash equilibrium neither the row player nor the column player can benefit, in expectation, by unilaterally deviating to some other strategy. We say that a mixed strategy pair is an ε -Nash equilibrium if no player can benefit more than ε , in expectation, by unilateral deviation. Formally,

Definition 2 (ε -Nash Equilibrium). *A mixed strategy pair (x, y) , $x, y \in \Delta^n$, is said to be an ε -Nash equilibrium if and only if:*

$$x^T A y \geq e_i^T A y - \varepsilon \quad \forall i \in [n] \quad \text{and} \quad (21)$$

$$x^T B y \geq x^T B e_j - \varepsilon \quad \forall j \in [n]. \quad (22)$$

Throughout, we will write C to denote the sum of payoff matrices, $C := A + B$. We will denote the i th column of C by C^i , for $i \in [n]$. Note that $\|C^i\|_0$ is equal to the number of non-zero entries in the i th column of C .

In the following definition we ensure that the sparsity parameter s is at least 4 for ease of presentation. In particular, the running time of our algorithm depends on the log of the number of non-zero entries in the columns of C , i.e., log of the sparsity of the columns of matrix C . Setting $s \geq 4$ gives us $\log s \geq 2$. This allows us to state a single running-time bound, which holds even for corner cases wherein the column sparsity is, say, zero (and hence $\log(\max_i \|C^i\|_0)$ is undefined).

Definition 3 (*s*-Sparse Games). *The sparsity of a game (A, B) is defined to be*

$$s := \max\{\max_i \|C^i\|_0, 4\},$$

where matrix $C = A + B$.

The quantitative connection between the sparsity of a game and the time it takes to compute an ε -Nash equilibrium is stated below.

Theorem 4. *Let $A, B \in [-1, 1]^{n \times n}$ be the payoff matrices of an s -sparse bimatrix game. Then, an ε -Nash equilibrium of (A, B) can be computed in time*

$$n^{O\left(\frac{\log s}{\varepsilon^2}\right)}.$$

Our algorithm for computing ε -Nash equilibrium relies on the following bilinear program, which was formulated by Mangasarian and Stone [26]. As formally specified in Lemma 1 below, approximate solutions of this bilinear program correspond to approximate Nash equilibria.

$$\begin{aligned} & \max_{x, y, \pi_1, \pi_2} && x^T C y - \pi_1 - \pi_2 \\ \text{subject to} &&& x^T B \leq \mathbb{1}^T \pi_2 \\ &&& A y \leq \mathbb{1} \pi_1 \\ &&& x, y \in \Delta^n \\ &&& \pi_1, \pi_2 \in [-1, 1]. \end{aligned} \tag{BP}$$

Here $\mathbb{1}$ denotes the all-ones vector. Using the definition of Nash equilibrium one can show that the optimal solutions of this bilinear program correspond to Nash equilibria of the game (A, B) . Formally, we have

Theorem 5 (Equivalence Theorem [26]). *Mixed strategy pair (\hat{x}, \hat{y}) is a Nash equilibrium of the game (A, B) if and only if \hat{x} , \hat{y} , $\hat{\pi}_1$, and $\hat{\pi}_2$ form an optimal solution of the bilinear program (BP), for some scalars $\hat{\pi}_1$ and $\hat{\pi}_2$. In addition, the optimal value achieved by (BP) is equal to zero and the payoffs of the row and column player at this equilibrium are $\hat{\pi}_1$ and $\hat{\pi}_2$ respectively.*

A relevant observation is that that an approximate solution of the bilinear program corresponds to an ε -Nash equilibrium.

Lemma 1. *Let $x, y \in \Delta^n$ along with scalars π_1 and π_2 form a feasible solution of (BP) that achieves an objective function value more than $-\varepsilon$, i.e., $x^T C y \geq \pi_1 + \pi_2 - \varepsilon$. Then, (x, y) is an ε -Nash equilibrium of the game (A, B) .*

Proof. The feasibility of x, y implies that $\max_j x^T B e_j \leq \pi_2$ and $\max_i e_i^T A y \leq \pi_1$. Since the objective function value achieved by x, y is at least $-\varepsilon$ we have $x^T A y + x^T B y - \pi_1 - \pi_2 \geq -\varepsilon$. But, $x^T A y$ is at most π_1 and $x^T B y$ is at most π_2 . So, the following inequalities must hold

$$x^T A y \geq \pi_1 - \varepsilon \quad \text{and} \quad (23)$$

$$x^T B y \geq \pi_2 - \varepsilon. \quad (24)$$

Overall, we get $x^T A y \geq \max_i e_i^T A y - \varepsilon$ and $x^T B y \geq \max_j x^T B e_j - \varepsilon$. Hence (x, y) satisfies the definition of an ε -Nash equilibrium. \square

The proposed algorithm (see Algorithm 1) solves the following p -norm minimization problem, with $p \geq 2$. The program $\text{CP}(u)$ is parametrized by vector $u \in \mathbb{R}^n$ and it can be solved in polynomial time.⁷

$$\begin{aligned} \min_{x, y, \pi_1, \pi_2} \quad & \|Cy - u\|_p && (\text{CP}(u)) \\ \text{subject to} \quad & x^T u \geq \pi_1 + \pi_2 - \varepsilon/2 \\ & Ay \leq \mathbb{1}\pi_1 \\ & x^T B \leq \mathbb{1}^T \pi_2 \\ & x, y \in \Delta^n \\ & \pi_1, \pi_2 \in [-1, 1]. \end{aligned}$$

Proof of Theorem 4. Algorithm 1 iterates at most $n^{O(\frac{p}{\varepsilon^2})}$ times, since this is an upper bound on the number of multisets of size $O(\frac{p}{\varepsilon^2})$. Furthermore, in each iteration the algorithm solves convex program $\text{CP}(u)$, this takes polynomial time. Given that $p = \log s$, these observations establish the desired running-time bound.

Now, in order to prove the theorem we need to show that Alogorithm 1 is (i) Sound: any mixed strategy pair, (x, y) , returned by the algorithm is an approximate Nash equilibrium; (ii) Complete: the algorithm always returns a mixed-strategy pair.

Soundness: Lemma 1 implies that any mixed-strategy pair returned by the algorithm is guaranteed to be an ε -Nash equilibrium. Specifically, say for some u the “if” condition in Step 7 is met. In addition, let x and y be the returned optimal solution of $\text{CP}(u)$. Then,

$$|x^T(Cy - u)| \leq \|x\|_q \|Cy - u\|_p \quad (\text{Hölder's inequality}) \quad (25)$$

$$\leq 1 \times \varepsilon/2. \quad (26)$$

⁷Note that for fixed u , $\text{CP}(u)$ is a convex program. Specifically, given $u \in \mathbb{R}^n$ and matrix $C \in \mathbb{R}^{n \times n}$, for $p \geq 1$, the function $f(x) := \|Cx - u\|_p$ is convex.

Algorithm 1 Algorithm for computing ε -Nash equilibrium in s -sparse games

Given payoff matrices $A, B \in [-1, 1]^{n \times n}$ and $\varepsilon > 0$; Return: ε -Nash equilibrium of (A, B)

- 1: Write s to denote the sparsity of the game (A, B) and let $p = \log s$.
{Note that, by definition, $s \geq 4$; hence, $p \geq 2$.}
 - 2: Let \mathcal{U} be the collection of all multisets of $\{1, 2, \dots, n\}$ of cardinality at most $\frac{\kappa p}{\varepsilon^2}$, where κ is a fixed constant.
 - 3: Write C^i to denote the i th column of matrix $C = A + B$, for $i \in [n]$.
 - 4: **for all** multisets $S \in \mathcal{U}$ **do**
 - 5: Set $u = \frac{1}{|S|} \sum_{i \in S} C^i$.
 { u is an $|S|$ -uniform vector in the convex hull of the columns of C .}
 - 6: Solve convex program $\text{CP}(u)$.
 - 7: **if** the objective function value of $\text{CP}(u)$ is *less than* $\varepsilon/2$ **then**
 - 8: Return (x, y) , where x and y form an optimal solution of $\text{CP}(u)$.
 - 9: **end if**
 - 10: **end for**
-

Here $q = p/(p-1) \geq 1$, since $p \geq 2$. The second inequality follows from the fact that the objective function value of $\text{CP}(u)$ is no more than $\varepsilon/2$ and $\|x\|_q \leq \|x\|_1 = 1$. Since the returned x satisfies the feasibility constraints in $\text{CP}(u)$ we have $x^T u \geq \pi_1 + \pi_2 - \varepsilon/2$. Therefore, $x^T C y \geq x^T u - \varepsilon/2 \geq \pi_1 + \pi_2 - \varepsilon$. Overall, x and y satisfy the conditions in Lemma 1, and hence form an ε -Nash equilibrium.

Completeness: It remains to show that the “if” condition in Step 7 is satisfied at least once (and hence the algorithm successfully returns a mixed strategy pair (x, y)). Next we accomplish this task.

Write (\hat{x}, \hat{y}) to denote a Nash equilibrium of the given game and let $\hat{\pi}_1$ ($\hat{\pi}_2$) be the payoff of the row (column) player under this equilibrium. Note that $C\hat{y}$ lies in the convex hull of the columns of C . Furthermore, since the sparsity of the game is s , for $p = \log s$, we have

$$\|C^i\|_p \leq 4 \quad \forall i \in [n]. \quad (27)$$

This follows from the fact that the entries of matrix C lie between -2 and 2 (recall that the payoffs are normalized between -1 and 1); hence considering the p norm of any column i we get: $\|C^i\|_p \leq (2^p s)^{1/p} = 2(s)^{1/p} = 2(2^{\log s})^{1/p} = 4$.

Therefore, for $p = \log s$, we can apply Theorem 2 over the convex hull $\text{conv}(\{C^i\}_i)$ with $\gamma \leq 4$. In particular, for $\mu = C\hat{y}$, Theorem 2 implies that there exists a $O(\frac{p}{\varepsilon^2})$ uniform vector μ' such that $\|C\hat{y} - \mu'\|_p \leq \varepsilon/2$

Since μ' is $O(\frac{p}{\varepsilon^2})$ uniform, at some point during its execution the algorithm (with an appropriate

value of κ) will set $u = \mu'$. Therefore, at least once the algorithm will consider a u that satisfies

$$\|C\hat{y} - u\|_p \leq \varepsilon/2. \quad (28)$$

We show that in this case \hat{x} , \hat{y} , $\hat{\pi}_1$ and $\hat{\pi}_2$ form a feasible solution of $\text{CP}(u)$ that achieves an objective function value of no more than $\varepsilon/2$. That is, the “if” condition in Step 7 is satisfied for this choice of u .

First of all the fact that the objective function value is no more than $\varepsilon/2$ follows directly from (28).

Since (\hat{x}, \hat{y}) is a Nash equilibrium, using Theorem 5 we get

$$\hat{x}^T C\hat{y} = \hat{\pi}_1 + \hat{\pi}_2. \quad (29)$$

Next we show that $\hat{x}^T u \geq \pi_1 + \pi_2 - \varepsilon/2$. Consider the following bound:

$$|\hat{x}^T(C\hat{y} - u)| \leq \|\hat{x}\|_q \|C\hat{y} - u\|_p \quad (\text{H\"older's inequality}) \quad (30)$$

$$\leq 1 \times \varepsilon/2. \quad (31)$$

Again, $q = p/(p-1) \geq 1$ and we have $\|\hat{x}\|_q \leq 1$. Here, the second inequality now follows from our choice of u . Since, $\hat{x}^T C\hat{y} = \hat{\pi}_1 + \hat{\pi}_2$, we have $\hat{x}^T u \geq \hat{\pi}_1 + \hat{\pi}_2 - \varepsilon/2$.

The remaining feasibility constraints of $\text{CP}(u)$ are satisfied as well. This simply follows from the fact that \hat{x} , \hat{y} , $\hat{\pi}_1$, and $\hat{\pi}_2$ form a feasible (in fact optimal) solution of (BP), see Theorem 5.

Overall, we get that the “if” condition in Step 7 will be satisfied at least once and this completes the proof. □

Ideas developed in this section can address other notions of sparsity as well. Specifically, if there exist $\alpha, \beta \in \mathbb{R}_+$ and $\gamma \in \mathbb{R}$ such that the matrix $\alpha A + \beta B + \gamma \mathbb{1}_{n \times n}$ has column or row sparsity s , then our algorithm can be used to find an ε -Nash equilibrium of the game (A, B) in time $n^{O\left(\frac{\lambda^2 \log s}{\varepsilon^2}\right)}$; here, $\mathbb{1}_{n \times n}$ is the all-ones $n \times n$ matrix and $\lambda := \max\{\alpha, \beta, 1/\alpha, 1/\beta\}$. This follows from the fact that a $(\min\{\alpha, \beta\} \varepsilon)$ -Nash equilibrium of the game $(\alpha A, \beta B + \gamma \mathbb{1}_{n \times n})$ is an ε -Nash equilibrium of the game (A, B) .

Furthermore, in time $n^{O\left(\frac{\log s}{\varepsilon^2}\right)}$, we can compute ε -Nash equilibria of games in which *both* matrices A and B have column or row sparsity s . Note that this case is not a direct corollary of Theorem 4. In particular, if the columns of matrix A and the rows of matrix B are s sparse, then it is not necessary that $A + B$ has low column or row sparsity. But, an approximate equilibrium of such a game can be computed by exhaustively searching for vectors, v and w , that are

$\varepsilon/4$ -close (under the $\log s$ -norm distance) to $A\hat{y}$ and $\hat{x}^T B$ respectively, here (\hat{x}, \hat{y}) is a Nash equilibrium of (A, B) . In this case, instead of $CP(u)$, we need to solve a convex program that minimizes $\|Ay - v\|_{\log s} + \|B^T x - w\|_{\log s}$ and has the following constraint $x^T v + w^T y \geq \pi_1 + \pi_2 - \varepsilon/2$ along with the constraints present in (BP). A direct extension of the proof of Theorem 4 shows that this program finds an ε -Nash equilibrium of the game (A, B) .

Remark 1. Consider the class of games in which the p norm of the columns of matrix C is a fixed constant. A simple modification of the arguments mentioned above shows that for such games an ε -Nash equilibrium can be computed in time $n^{O(\frac{p}{\varepsilon^2})}$.

Remark 2. Algorithm 1 can be adopted to find an approximate Nash equilibrium with large social welfare (the total payoffs of the players). Specifically, in order to determine whether there exists an approximate Nash equilibrium with social welfare more than $\alpha - \varepsilon$, we include the constraint $\pi_1 + \pi_2 \geq \alpha$ in $CP(u)$. The time complexity of the algorithm stays the same, and then via a binary search over α we can find an approximate Nash equilibrium with near-optimal social welfare.

Remark 3. In Algorithm 1, instead of the convex program $CP(u)$, we can solve the linear program with objective $\min \|Cy - u\|_\infty$ and constraints identical to $CP(u)$. Algorithm 1 still finds an approximate Nash, since $\|Cy - u\|_\infty \leq \|Cy - u\|_p$ and we have $|x^T(Cy - u)| \leq \|x\|_1 \|Cy - u\|_\infty \leq 1 \times \varepsilon/2$ (1 and ∞ are Hölder conjugates of each other).

Solving a linear program, in place of a convex program, would lead to a polynomial improvement in the running time of the algorithm. But, minimizing the p norm of $Cy - u$ remains useful in specific cases; in particular, it provides a better running-time bound when the game is guaranteed to have a “small probability” equilibrium. We detail this result in the following Section.

4.1 Small Probability Games

Daskalakis and Papadimitriou [17] showed that there exists a PTAS for games that contain an equilibrium with small—specifically, $O(\frac{1}{n})$ —probability values. This result is somewhat surprisingly, since such small-probability equilibria have large— $\Omega(n)$ —support, and hence are not amenable to, say, exhaustive search. This section shows that if a game has an equilibrium with probability values $O(\frac{1}{m})$, for some $1 \leq m \leq n$, then an approximate equilibrium can be computed in time $n^{O(k/\varepsilon^2)}$, where k has a logarithmic dependence on s/m . Since column sparsity s is no more than n , we get back the result of [17] as a special case.

Definition 4 (Small Probability Equilibrium). A Nash equilibrium (x, y) is said to be m -small probability if all the entries of x and y are at most $\frac{1}{m}$.

In [17] a PTAS is given for games that have an equilibrium with probability values at most $\frac{1}{\delta n}$, for some fixed constant $\delta \in (0, 1]$. Hence, the setting of [17] corresponds to games that have δn -small probability equilibrium. Next we prove a result for general m -small probability equilibria.

Theorem 6. Let $A, B \in [-1, 1]^{n \times n}$ be the payoff matrices of an s -sparse bimatrix game. If (A, B) contains an m -small probability Nash equilibrium, then an ε -Nash equilibrium of the game can be computed in time

$$n^{O\left(\frac{t}{\varepsilon^2}\right)},$$

where $t = \max\left\{2 \log\left(\frac{s}{m}\right), 2\right\}$.

Proof. Let norm $p = \max\left\{2 \log\left(\frac{s}{m}\right), 2\right\}$ and write q to denote the Hölder conjugate of p , i.e., q satisfies $\frac{1}{p} + \frac{1}{q} = 1$. To obtain this result Algorithm 1 is modified as follows: (i) use the updated value of p instead of the one specified in Step 1 of the algorithm; (ii) include convex constraint $\|x\|_q \leq m^{-1/p}$ in $\text{CP}(u)$; (iii) In Step 7 use $\frac{\varepsilon m^{1/p}}{2}$, instead of $\frac{\varepsilon}{2}$, as the threshold for returning a solution.

In order to establish the theorem we prove that the modified algorithm is (i) Sound: any mixed strategy pair, (x, y) , returned by it is an approximate Nash equilibrium; (ii) Complete: the algorithm always returns a mixed-strategy strategy pair.

Soundness: Below we show that if x and y are returned by this modified algorithm then they form a near-optimal solution of the bilinear program (BP) and, in particular, satisfy $x^T C y \geq \pi_1 + \pi_2 - \varepsilon$. Hence, Lemma 1 shows that any returned solution (x, y) is an ε -Nash equilibrium.

Say the algorithm returns (x, y) while considering vector u . Applying Hölder's inequality gives us:

$$|x^T(Cy - u)| \leq \|x\|_q \|Cy - u\|_p \tag{32}$$

$$\leq m^{-1/p} \frac{\varepsilon m^{1/p}}{2} \tag{33}$$

$$= \varepsilon/2. \tag{34}$$

Here the second inequality uses the fact that x satisfies the feasibility constraint $\|x\|_q \leq m^{-1/p}$ and the objective function value of $\text{CP}(u)$ is at most $\frac{\varepsilon m^{1/p}}{2}$. Since x is a feasible solution of $\text{CP}(u)$, it satisfies $x^T u \geq \pi_1 + \pi_2 - \varepsilon/2$. Therefore, using inequality (34) we get $x^T C y \geq \pi_1 + \pi_2 - \varepsilon$, as required.

Completeness: It remains to show that the “if” condition in Step 7 is satisfied at least once. To achieve this we prove that, for a particular u , an m -small probability Nash equilibrium forms a feasible solution of $\text{CP}(u)$ and achieves an objective function value of at most $\frac{\varepsilon m^{1/p}}{2}$. Therefore, for this u the “if” condition in Step 7 will be met.

Recall that the columns of C are s sparse and its entries are at most 2. Hence the p norm of any column C^i satisfies $\|C^i\|_p \leq (s2^p)^{1/p} = 2s^{1/p}$.

Let (\hat{x}, \hat{y}) be an m -small probability Nash equilibrium of the game. Theorem 2, applied over the convex hull $\text{conv}(\{C^i\}_i)$ with $\gamma \leq 2s^{1/p}$, guarantees that there exists a $O\left(\frac{p s^{2/p}}{\varepsilon^2 m^{2/p}}\right)$ uniform

vector that is $\frac{\varepsilon m^{1/p}}{2}$ close to $C\hat{y}$. Since $p \geq 2 \log\left(\frac{s}{m}\right)$, we have

$$\frac{s^{2/p}}{m^{2/p}} = \left(\frac{s}{m}\right)^{2/p} \quad (35)$$

$$\leq 2. \quad (36)$$

Therefore, there exists a $O\left(\frac{p}{\varepsilon^2}\right)$ uniform vector that is $\frac{\varepsilon m^{1/p}}{2}$ close to $C\hat{y}$. Such a vector, say u , will be selected by the algorithm in Step 4 at some point of time. Below we show that, for u , the mixed strategies of the Nash equilibrium \hat{x} and \hat{y} are feasible solutions that achieve the desired objective function value.

To establish the feasibility of \hat{x} , we first upper bound its q norm. The fact that its entries at most $1/m$ and $q = \frac{p}{p-1} \geq 1$ implies:

$$\|\hat{x}\|_q \leq \left(m \frac{1}{m^q}\right)^{1/q} \quad (37)$$

$$= m^{-1/p}. \quad (38)$$

Since (\hat{x}, \hat{y}) is a Nash equilibrium, there exists payoffs $\hat{\pi}_1$ and $\hat{\pi}_2$ such $\hat{x}^T C \hat{y} = \hat{\pi}_1 + \hat{\pi}_2$ (Theorem 5). Also, \hat{x} , \hat{y} , $\hat{\pi}_1$, and $\hat{\pi}_2$ are feasible with respect to (BP). Hence, the only constraint of CP(u) that we still need to verify is $\hat{x}^T u \geq \hat{\pi}_1 + \hat{\pi}_2 - \varepsilon/2$. This follows from Hölder's inequality:

$$|\hat{x}^T (C\hat{y} - u)| \leq \|\hat{x}\|_q \|C\hat{y} - u\|_p \quad (39)$$

$$\leq m^{-1/p} \frac{\varepsilon m^{1/p}}{2} \quad (40)$$

$$= \varepsilon/2. \quad (41)$$

Overall, for the above specified u , the “if” condition in Step 7 will be satisfied. This shows that the algorithms successfully returns an ε -Nash equilibrium.

The algorithm iterates at most $n^{O\left(\frac{p}{\varepsilon^2}\right)}$ times, since this is an upper bound on the number of multisets of size $O\left(\frac{p}{\varepsilon^2}\right)$. Furthermore, in each iteration the algorithm solves a convex program, this takes polynomial time. These observations establish the desired running-time bound and complete the proof. \square

5 Densest Subgraph

This section and the next one present additive approximations for the normalized densest k -subgraph problem (NDkS) and the densest k -bipartite subgraph problem (DkBS) respectively.

In NDkS we are given a simple graph $G = (V, E)$ along with a size parameter $k \leq |V|$ and the goal is to find a maximum density subgraph containing exactly k vertices. Here, density of a size- k

subgraph $S = (V_S, E_S)$ is defined to be $\rho(S) := |E_S|/k^2$. Note that in NDkS density is normalized to be at most one.

Next, we present a quadratic program, an ε -additive approximate solution of which can be used to efficiently find an ε -additive approximate solution of NDkS. Write A to denote the adjacency matrix of the given graph G and let n be the number of vertices in G . Define matrix $C := \frac{1}{2}A + I$, here I is the $n \times n$ identity matrix.

$$\begin{aligned} & \max_x && x^T C x && \text{(QP)} \\ \text{subject to} &&& x_i \leq \frac{1}{k} && \forall i \in [n] \\ &&& x \in \Delta^n \end{aligned}$$

Write S^* to denote an optimal solution of the given NDkS instance and let z^* denote the optimal value of the quadratic program (QP). Given a solution x of (QP) that achieves an objective function value of $z^* - \varepsilon$ (i.e., x satisfies $x^T C x \geq z^* - \varepsilon$), we show how to efficiently find a subgraph S that is an ε -additive approximate solution of NDkS, i.e., find a size- k subgraph S that satisfies $\rho(S) \geq \rho(S^*) - \varepsilon$. Towards this end, the following lemma serves as a useful tool.

Lemma 2. *Given a feasible solution, y , of the quadratic program (QP) we can find in polynomial time a feasible solution z that satisfies $z^T C z \geq y^T C y$ and, moreover, every component of z is either 0 or $\frac{1}{k}$, i.e., $z_i \in \{0, \frac{1}{k}\}$ for each $i \in [n]$.*

Proof. Given feasible solution y write $M(y) := \{i \in [n] \mid 0 < y_i < 1/k\}$. We iteratively update y to decrease the cardinality of $M(y)$, and at the same time ensure that the objective function value (i.e., $y^T C y$) does not decrease. Since $|M(y)| \leq n$ we iterate at most n times. This will overall establish the lemma.

Note that each $i \in [n]$ indices both a component of y and a vertex of the graph G . Write γ_i to denote the total y value of vertex i and its neighbors in G , i.e., $\gamma_i := y_i + \sum_{j:(i,j) \in E} y_j$. Since $y \in \Delta^n$ and integer $k > 1$, the cardinality of set $M(y)$ is either 0 or strictly greater than one. If $M(y) = \emptyset$ then the stated claims follows simply by setting $z = y$. Otherwise, if $|M(y)| \geq 2$ then we select two vertices $i, j \in M(y)$ and change y_i and y_j such that the size of $M(y)$ decreases by at least one. We select $i, j \in M(y)$ as follows:

- If there exist vertices $i, j \in M(y)$ that are not connected by an edge in G , then without loss of generality we assume that $\gamma_i + y_i \geq \gamma_j + y_j$.
- If every pair of vertices $i, j \in M(y)$ is connected by an edge, then without loss of generality we assume that $\gamma_i \geq \gamma_j$.

Let $\delta := \min \{y_j, \frac{1}{k} - y_i\}$. We update $y_i \leftarrow y_i + \delta$ and $y_j \leftarrow y_j - \delta$, this update ensures that either y_j goes down to zero or y_i becomes equal to $1/k$; hence, the size of $M(y)$ decreases by at least one.

Next we show this change in y does not decrease the objective function value $y^T C y$. Thereby we get the stated claim.

Consider the case in which i and j are not connected via an edge. The other case in which $(i, j) \in E$ (and we have $s_i \geq s_j$) follows along the same lines.

Before the update the following inequality holds, $\gamma_i + y_i \geq \gamma_j + y_j$. For any δ , the change in objective function value is equal to $(y_i + \delta)(\gamma_i + \delta) + (y_j - \delta)(\gamma_j - \delta) - (y_i \gamma_i + y_j \gamma_j)$. This quality is equal to $\delta(y_i + \gamma_i - y_j - \gamma_j) + \delta^2$. Since $y_i + \gamma_i - y_j - \gamma_j \geq 0$ and δ is nonnegative, we get that the update in y does not decrease the objective function value. This completes the proof. \square

Recall that $\rho(S^*)$ and z^* denote the optimal values of the given NDkS instance and (QP) respectively. Using Lemma 2 we get the following proposition.

Proposition 1. *The optimal value of (QP) is equal to the optimal value of the NDkS instance plus $1/k$, i.e., $z^* = \rho(S^*) + 1/k$.*

Proof. We use S^* to obtain a feasible solution, \hat{x} , for (QP) as follows: for each $i \in V(S^*)$ set $\hat{x}_i = 1/k$ and for every $j \notin V(S^*)$ set $\hat{x}_j = 0$. Here, $V(S^*)$ denotes the set of vertices of the subgraph S^* .

By the definition of matrix C we get that $\hat{x}^T C \hat{x} = \rho(S^*) + 1/k$. This implies that

$$z^* \geq \rho(S^*) + 1/k. \quad (42)$$

Using Lemma 2 we can obtain an optimal solution x' of the QP such that the components of x' are either 0 or $1/k$. By definition, $(x')^T C x' = z^*$. Write S' to denote the subgraph induced by the vertex subset $\text{Supp}(x')$. Note that $z^* = (x')^T C x' = \rho(S') + 1/k$. Given that S^* is an optimal solution of the NDkS instance, we have $\rho(S^*) \geq \rho(S')$. Hence, the following inequality holds

$$\rho(S^*) + 1/k \geq z^*. \quad (43)$$

Inequalities (42) and (43) imply the stated claim. \square

Finally, we establish the connection between (QP) and NDkS.

Theorem 7. *Given an ε -additive approximate solution of (QP) we can find an ε -additive approximate solution of NDkS in polynomial time.*

Proof. Given an ε -additive approximate solution of (QP), via Lemma 2, we can find an ε -additive approximation \hat{x} whose components are either 0 or $1/k$. Write \hat{S} to denote the subgraph induced

by vertex subset $\text{Supp}(\hat{x})$; recall that, each $i \in [n]$ indices both a component of \hat{x} and a vertex of the graph G .

Note that $\hat{x}^T C \hat{x} = \rho(\hat{S}) + 1/k$. Since $\hat{x}^T C \hat{x} \geq z^* - \varepsilon$, the following inequality holds $\rho(\hat{S}) + 1/k \geq z^* - \varepsilon$. Using Propostion 1 we get that \hat{S} is an ε -additive approximate solution of NDkS, $\rho(\hat{S}) \geq \rho(S^*) - \varepsilon$. \square

In other words, in order to determine an approximate solution of NDkS it suffices to compute an approximate solution of (QP).

Note that, if the maximum degree of the graph G is d then the number of non-zero components in any column of the objective matrix C is no more than $d + 1$. This implies that, for $i \in [n]$, we have $\|C^i\|_p \leq (d + 1)^{1/p}$. Here, C^i denotes the i th column of C . Now for $p = \log(d + 1)$ the following bound holds for all $i \in [n]$: $\|C^i\|_p \leq 2$. Therefore, as in Algorithm 1, enumerating all $O\left(\frac{p}{\varepsilon^2}\right)$ -uniform vectors in the convex hull of the columns of C , we can find an ε -approximate solution of (QP). This establishes the following theorem.

Theorem 8. *Let G be a graph with n vertices and maximum degree d . Then, an ε -additive approximation of NDkS over G can be determined in time*

$$n^{O\left(\frac{\log d}{\varepsilon^2}\right)}.$$

In general, this gives us an additive approximation algorithm for NDkS that runs in quasi-polynomial time.

5.1 Densest Bipartite Suubgraph

In DkBS we are given a graph $G = (V, E)$ along with a size parameter $k \leq |V|$ and the goal is to find size- k vertex subsets, S and T , such that the density of edges between S and T is maximized. Specifically, the (bipartite) density of vertex subsets S and T is defined as follows:

$$\rho(S, T) := \frac{|E(S, T)|}{|S||T|}, \quad (44)$$

here $E(S, T)$ denotes the set of edges that connect S and T .

Let vertex subsets S^* and T^* form an optimal solution of the given DkBS instance. Hence, $\rho(S^*, T^*)$ is equal to the optimal density. Next we state a bilinear program from [3] to approximate DkBS. Here, A denotes the adjacency matrix of the given graph G and $n = |V|$.

$$\begin{aligned} & \max_{x, y} \quad x^T A y \\ & \text{subject to} \quad x, y \in \Delta^n \\ & \quad \quad \quad x_i, y_i \leq \frac{1}{k} \quad \forall i \in [n]. \end{aligned} \quad (\text{BP-DkBS})$$

Note that optimizing (BP-DkBS) over x , with a fixed y , corresponds to solving a linear program. Therefore, for any fixed y there exists an optimal *basic* feasible x , and vice versa. In other words, for any feasible pair (x_0, y_0) we can find (x, y) , such that $x_0^T A y_0 \leq x^T A y$ and the all the components of x and y are either 0 or $1/k$. This observation implies that the optimal value of (BP-DkBS) is equal to $\rho(S^*, T^*)$. In addition, given an additive ε -approximate solution of (BP-DkBS), (x', y') , we can efficiently determine an ε -approximate solution of DkBS. Specifically, we can assume without loss of generality that x' and y' are basic, and then for $S' := \text{Supp}(x')$ and $T' := \text{Supp}(y')$ we have $\rho(S', T') \geq \rho(S^*, T^*) - \varepsilon$. In other words, in order to determine an approximate solution of DkBS it suffices to compute an approximate solution of (BP-DkBS).

Note that the column sparsity of the objective matrix in (BP-DkBS)—i.e, the column sparsity of the adjacency matrix A —is equal to the maximum degree of the given graph. Therefore, as outlined in the previous section, we can modify Algorithm 1 to obtain the following theorem.

Theorem 9. *Let G be a graph with n vertices and maximum degree d . Then, there exists an algorithm that runs in time $n^{O\left(\frac{\log d}{\varepsilon^2}\right)}$ and computes a $k \times k$ -bipartite subgraph of density at least $\rho(S^*, T^*) - \varepsilon$.*

6 Extensions

This section extends Theorem 2 to address convex hulls of matrices. Here we also detail approximate versions of certain generalizations of the Carathéodory's theorem; specifically, we focus on the colorful Carathéodory theorem and Tverberg's theorem.

First of all, note that a $d \times d$ matrix can be considered as a vector in \mathbb{R}^{d^2} , and hence directly applying Theorem 2 to vectors in \mathbb{R}^{d^2} we get results for entrywise p -norms. Recall that for a $d \times d$ matrix Y the entrywise p -norm is defined as follows $\|Y\|_p := \left(\sum_{i=1}^d \sum_{j=1}^d |Y_{i,j}|^p\right)^{1/p}$.

In particular, we get the an approximate version of the Birkhoff-von Neumann theorem by considering $d \times d$ matrices as vector of size d^2 and applying Theorem 2 with norm $p = \log d$. We know via the Birkhoff-von Neumann theorem that any $d \times d$ doubly stochastic matrix can be expressed as a convex combination of $d \times d$ permutation matrices (see, e.g., [5]). The following corollary shows that for every doubly stochastic matrix D there exists an ε -close (in the entrywise $\log d$ -norm, and hence in the entrywise ∞ -norm) doubly stochastic matrix D' that can be expressed as a convex combination of $O\left(\frac{\log d}{\varepsilon^2}\right)$ permutation matrices.

We say that a doubly stochastic matrix D' is k uniform if there exists a size k multiset Π of permutation matrices such that $D' = \frac{1}{k} \sum_{P \in \Pi} P$.

Corollary 1. *For every $d \times d$ doubly stochastic matrix D there exists an $O\left(\frac{\log d}{\varepsilon^2}\right)$ uniform doubly stochastic matrix D' such that $\max_{i,j} |D_{i,j} - D'_{i,j}| \leq \varepsilon$.*

In addition to entrywise norms, we can establish an approximate version of Carathéodory's theorem for matrices under the Schatten p -norm. Write $\|Y\|_{S_p}$ to denote that Schatten p -norm of

a $d \times d'$ matrix Y , i.e., $\|Y\|_{S_p} := \|\sigma(Y)\|_p$, where $\sigma(Y)$ is the vector of singular values of Y . Given a set of matrices $\mathcal{Y} = \{Y_1, Y_2, \dots, Y_n\} \subset \mathbb{R}^{d \times d'}$, we say that a matrix $M' \in \text{conv}(\mathcal{Y})$ is k uniform (implicitly with respect to \mathcal{Y}) if there exists a size k multiset S of $[n]$ such that $M' = \frac{1}{k} \sum_{i \in S} Y_i$.

We can directly adopt the proof of Theorem 2 and, in particular, use the matrix version of Khintchine inequality [31, 32] (instead of Theorem 1) to obtain the following result.

Theorem 10. *Given a set of matrices $\mathcal{Y} = \{Y_1, Y_2, \dots, Y_n\} \subset \mathbb{R}^{d \times d'}$ and $\varepsilon > 0$. For every matrix $M \in \text{conv}(\mathcal{Y})$ and $2 \leq p < \infty$ there exists an $O\left(\frac{p\gamma^2}{\varepsilon^2}\right)$ uniform matrix $M' \in \text{conv}(\mathcal{Y})$ such that $\|M - M'\|_{S_p} \leq \varepsilon$. Here, $\gamma := \max_{Y \in \mathcal{Y}} \|Y\|_{S_p}$.*

6.1 Approximating Colorful Carathéodory and Tverberg's Theorem

The colorful Carathéodory theorem (see [4] and [27]) asserts that if the convex hulls of $d+1$ sets (the color classes) $X_1, X_2, \dots, X_{d+1} \subset \mathbb{R}^d$ intersect, then every vector in the intersection $\mu \in \cap_i \text{conv}(X_i)$ can be expressed as a convex combination of vectors, each of which has a different color.

Theorem 11 (Colorful Carathéodory Theorem). *Let X_1, X_2, \dots, X_{d+1} be $d+1$ sets in \mathbb{R}^d and vector $\mu \in \cap_i \text{conv}(X_i)$. Then, there exists $d+1$ vectors x_1, x_2, \dots, x_{d+1} such that $x_i \in X_i$ for each i and $\mu \in \text{conv}(\{x_1, x_2, \dots, x_{d+1}\})$.*

For a given collection of $d+1$ sets (the color classes) $X_1, X_2, \dots, X_{d+1} \subset \mathbb{R}^d$, a set $R \in \mathbb{R}^d$ is called a rainbow if $|R \cap X_i| = 1$ for all $i \in [d+1]$.

The colorful Carathéodory theorem guarantees the existence of a rainbow $R = \{x_1, x_2, \dots, x_{d+1}\}$ whose convex hull contains the given vector μ . But, determining a rainbow R with $\mu \in \text{conv}(R)$ in polynomial time remains an interesting open problem; see, e.g., [27]. Here we consider a natural approximate version of this question wherein the goal is to find a rainbow R' whose convex hull is ε close to μ in the following sense $\inf_{v \in \text{conv}(R')} \|\mu - v\|_p \leq \varepsilon$. Below we show that, for appropriately scaled vectors, Theorem 2 can be used to efficiently determine such a rainbow R' .

Theorem 2 implies that there exists a vector μ' that satisfies $\|\mu - \mu'\|_p \leq \varepsilon$ and can be expressed as a convex combination of $t = O\left(\frac{p\gamma^2}{\varepsilon^2}\right)$ vectors of R , for $p \geq 2$ and $\gamma := \max_{x \in \cup_i X_i} \|x\|_p$. Let R' be a size t subset of R such that $\mu' \in \text{conv}(R')$. Write $n = \sum_{i=1}^{d+1} |X_i|$ and note that we can exhaustively search for R' in time $\binom{d+1}{t} n^{O(t)}$. For a guessed R' we can test if its convex hull is close to μ (i.e., $\inf_{v \in \text{conv}(R')} \|\mu - v\|_p \leq \varepsilon$) via a convex program. Moreover, after finding an appropriate subset R' we can extend it to obtain a rainbow that is ε close to μ .

Another generalization of Carathéodory's theorem is Tverberg's Theorem (see [35] and [27]), which is stated below.

Theorem 12 (Tverberg's Theorem). *Any set of $(r-1)(d+1)+1$ vectors $X \subset \mathbb{R}^d$ can be partitioned into r pairwise disjoint subsets $X_1, X_2, \dots, X_r \subseteq X$ such that their convex hulls intersect:*

$$\bigcap_{i=1}^r \text{conv}(X_i) \neq \phi.$$

Note that here the special case of $r = 2$ corresponds to Radon’s theorem (see, e.g., [27]).

Next we consider an approximate version of Tverberg’s theorem. At a high level, our result guarantees the existence of pairwise disjoint subsets, X'_1, \dots, X'_r , that have small cardinality and whose convex hulls, $\text{conv}(X'_1), \dots, \text{conv}(X'_r)$, are “concurrently close” to each other. Formally:

Definition 5 (Concurrently ε close). *Sets $V_1, V_2, \dots, V_r \subset \mathbb{R}^d$ are said to be concurrently ε close under the p -norm distance if there exists a vector $\mu \in \mathbb{R}^d$ such that $\inf_{v \in \text{conv}(V_i)} \|\mu - v\|_p \leq \varepsilon$, for all $i \in [r]$.*

Tverberg’s theorem together with Theorem 2 establish the following result.

Theorem 13. *Let norm $p \in [2, \infty)$ and parameter $\varepsilon > 0$. Then, any set of $(r - 1)(d + 1) + 1$ vectors $X \subset \mathbb{R}^d$ can be partitioned into r pairwise disjoint subsets $X'_1, X'_2, \dots, X'_r \subseteq X$ that are concurrently ε close under the p -norm distance and satisfy $|X'_i| = O\left(\frac{p\gamma^2}{\varepsilon^2}\right)$, for all $i \in [r]$. Here, $\gamma := \max_{x \in X} \|x\|_p$.*

Tverberg’s theorem guarantees the existence of a partition with subsets X_1, X_2, \dots, X_r whose convex hulls intersect. It is an interesting open problem whether such a partition can be determined in polynomial time; see [27]. But, note that Theorem 13 implies that for fixed r we can efficiently determine an approximate solution for this problem, i.e., determine a partition consisting of subsets that are concurrently ε close. In particular, with $t = O\left(\frac{p\gamma^2}{\varepsilon^2}\right)$, we can exhaustively search for size t subsets $X'_1, X'_2, \dots, X'_r \subseteq X$ that are concurrently ε close under the p -norm distance.⁸ This search runs in time $n^{O(rt)}$. Finally, we can extend the disjoint subsets $X'_1, X'_2, \dots, X'_r \subseteq X$ —by arbitrarily assigning the vectors in $X \setminus (\cup_i X'_i)$ —to obtain a partition of X that is concurrently ε close.

7 Lower Bound

This section presents a lower bound that proves that in general ε -close (under the p -norm distance with $p \in [2, \infty)$) vectors cannot be expressed as a convex combination of less than $\frac{1}{4\varepsilon^{p/(p-1)}}$ vectors of the given set. This, in particular, implies that the $1/\varepsilon^2$ dependence established in Theorem 2 is tight for the Euclidean case ($p = 2$).

To prove the lower bound we consider the convex hull of the standard basis vectors $e_1, e_2, \dots, e_d \in \mathbb{R}^d$, i.e., consider the simplex Δ^d . Write $\bar{e} \in \Delta^d$ to denote the uniform convex combination of e_1, e_2, \dots, e_d : $\bar{e} := (1/d, 1/d, \dots, 1/d)^T$. Next we show that any vector that can be expressed as a

⁸A polynomial-time algorithm exists for testing whether given subsets X'_1, X'_2, \dots, X'_r are concurrently ε close under the p -norm distance. In particular, we can run the ellipsoid method with a separation oracle that solves the convex optimization problem $\min_{v \in \text{conv}(X'_i)} \|u - v\|_p$ to test if there exists a separating hyperplane for proposed ellipsoid center u .

convex combination of less than $\frac{1}{4 \varepsilon^{p/(p-1)}}$ standard basis vectors is at least ε away from \bar{e} in the p -norm distance. This establishes the desired lower bound.

Note that (trivially) any vector in Δ^d can be expressed as a convex combination of $\frac{1}{\varepsilon^{p/(p-1)}}$ standard basis vectors if, say, $\frac{1}{\varepsilon^{p/(p-1)}} = d$. Hence, to obtain a meaningful lower bound we need to assume that $\frac{1}{\varepsilon^{p/(p-1)}}$ is less than d .

Proposition 2. *Let $\frac{1}{\varepsilon^{p/(p-1)}} < d$ and norm $p \geq 2$. Then, any vector $u \in \Delta^d$ that can be expressed as a convex combination of less than $\frac{1}{4 \varepsilon^{p/(p-1)}}$ standard basis vectors must satisfy $\|\bar{e} - u\|_p > \varepsilon$.*

Proof. Write $k = \frac{1}{4 \varepsilon^{p/(p-1)}}$. Since \bar{e} is symmetric, we can assume without loss of generality that u is a convex combination of vectors e_1, e_2, \dots, e_k . The optimal value of the following convex program lower bounds the p -norm distance between u and \bar{e} :

$$\begin{aligned} \min_{\alpha_1, \dots, \alpha_k} \quad & \left\| \bar{e} - \sum_{i=1}^k \alpha_i e_i \right\|_p \\ \text{subject to} \quad & \sum_{i=1}^k \alpha_i = 1 \\ & \alpha_i \geq 0 \quad \forall i \in [k]. \end{aligned}$$

K.K.T. conditions imply that the optimal is achieved by setting $\alpha_i = 1/k$ for all $i \in [k]$. Hence, the p -norm distance between \bar{e} and u is at least $(k(\frac{1}{k} - \frac{1}{d})^p)^{1/p}$. The choice of k and the assumption that $\frac{1}{\varepsilon^{p/(p-1)}} < d$ guarantee that $\|\bar{e} - u\|_p > \frac{3\varepsilon}{4^{1/p}}$. Since $p \geq 2$, we get the desired claim. \square

Acknowledgements

The author thanks Federico Echenique, Katrina Ligett, Assaf Naor, Aviad Rubinfeld, Anthony Man-Cho So, and Joel Tropp for helpful discussions and references. This research was supported by NSF grants CNS-0846025 and CCF-1101470, along with a Linde/SISL postdoctoral fellowship.

References

- [1] Noga Alon, Sanjeev Arora, Rajsekar Manokaran, Dana Moshkovitz, and Omri Weinstein. Inapproximability of densest κ -subgraph from average case hardness. <http://www.tau.ac.il/~nogaa/PDFS/dks8.pdf>, 2011.
- [2] Noga Alon, Troy Lee, and Adi Shraibman. The cover number of a matrix and its algorithmic applications. *Approximation, Randomization, and Combinatorial Optimization (APPROX/RANDOM)*, 2014.

- [3] Noga Alon, Troy Lee, Adi Shraibman, and Santosh Vempala. The approximate rank of a matrix and its algorithmic applications: approximate rank. In *Proceedings of the 45th annual ACM symposium on Symposium on theory of computing*, pages 675–684. ACM, 2013.
- [4] Imre Bárány. A generalization of carathéodory’s theorem. *Discrete Mathematics*, 40(2):141–152, 1982.
- [5] Alexander Barvinok. *A course in convexity*, volume 54. American Mathematical Soc., 2002.
- [6] Aditya Bhaskara, Moses Charikar, Eden Chlamtac, Uriel Feige, and Aravindan Vijayaraghavan. Detecting high log-densities: an $O(n^{1/4})$ approximation for densest k-subgraph. In *Proceedings of the 42nd ACM symposium on Theory of computing*, pages 201–210. ACM, 2010.
- [7] Hartwig Bosse, Jaroslaw Byrka, and Evangelos Markakis. New algorithms for approximate Nash equilibria in bimatrix games. In *Internet and Network Economics*, pages 17–29. Springer, 2007.
- [8] Jean Bourgain, Alain Pajor, SJ Szarek, and N Tomczak-Jaegermann. On the duality problem for entropy numbers of operators. In *Geometric aspects of functional analysis*, pages 50–63. Springer, 1989.
- [9] Mark Braverman, Young Kun Ko, and Omri Weinstein. Approximating the best Nash equilibrium in $n^{o(\log n)}$ -time breaks the exponential time hypothesis. *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2015.
- [10] Bernd Carl. Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in Banach spaces. In *Annales de l’institut Fourier*, volume 35, pages 79–118. Institut Fourier, 1985.
- [11] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Sparse games are hard. In *Internet and Network Economics*, pages 262–273. Springer, 2006.
- [12] 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.
- [13] Constantinos Daskalakis. On the complexity of approximating a Nash equilibrium. *ACM Transactions on Algorithms (TALG)*, 9(3):23, 2013.
- [14] 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.
- [15] Constantinos Daskalakis, Aranyak Mehta, and Christos Papadimitriou. A note on approximate Nash equilibria. In *Internet and Network Economics*, pages 297–306. Springer, 2006.

- [16] Constantinos Daskalakis, Aranyak Mehta, and Christos Papadimitriou. Progress in approximate Nash equilibria. In *Proceedings of the 8th ACM conference on Electronic commerce*, pages 355–358. ACM, 2007.
- [17] Constantinos Daskalakis and Christos H Papadimitriou. On oblivious ptas’s for Nash equilibrium. In *Proceedings of the 41st annual ACM symposium on Theory of computing*, pages 75–84. ACM, 2009.
- [18] Tomas Feder, Hamid Nazerzadeh, and Amin Saberi. Approximating Nash equilibria using small-support strategies. In *Proceedings of the 8th ACM conference on Electronic commerce*, pages 352–354. ACM, 2007.
- [19] D.J.H. Garling. *Inequalities: a journey into linear analysis*, volume 19. Cambridge University Press Cambridge, 2007.
- [20] Elad Hazan and Robert Krauthgamer. How hard is it to approximate the best Nash equilibrium? *SIAM Journal on Computing*, 40(1):79–91, 2011.
- [21] Ravi Kannan and Thorsten Theobald. Games of fixed rank: A hierarchy of bimatrix games. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1124–1132. Society for Industrial and Applied Mathematics, 2007.
- [22] Spyros C Kontogiannis, Panagiota N Panagopoulou, and Paul G Spirakis. Polynomial algorithms for approximating Nash equilibria of bimatrix games. In *Internet and Network Economics*, pages 286–296. Springer, 2006.
- [23] Spyros C Kontogiannis and Paul G Spirakis. Efficient algorithms for constant well supported approximate equilibria in bimatrix games. In *Automata, Languages and Programming*, pages 595–606. Springer, 2007.
- [24] Joram Lindenstrauss and Lior Tzafriri. *Classical Banach spaces*. Springer Berlin-Heidelberg-New York, 1973.
- [25] 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. ACM, 2003.
- [26] Olvi L. Mangasarian and H. Stone. Two-person nonzero-sum games and quadratic programming. *Journal of Mathematical Analysis and Applications*, 9(3):348–355, 1964.
- [27] Jiří Matoušek. *Lectures on discrete geometry*, volume 108. Springer New York, 2002.
- [28] T. S. Motzkin and E. G. Straus. Maxima for graphs and a new proof of a theorem of Turán. *Canadian Journal of Mathematics*, 17:533–540, 1965.

- [29] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. Algorithmic game theory. 2007.
- [30] Gilles Pisier. Remarques sur un résultat non publié de B. Maurey (French) [Remarks on an unpublished result of B. Maurey]. *Séminaire Analyse fonctionnelle (dit)*, pages 1–12, 1981.
- [31] Anthony Man-Cho So. Moment inequalities for sums of random matrices and their applications in optimization. *Mathematical programming*, 130(1):125–151, 2011.
- [32] Nicole Tomczak-Jaegermann. The moduli of smoothness and convexity and the Rademacher averages of the trace classes S_p ($1 \leq p < \infty$). *Studia Mathematica*, 50(2):163–182, 1974.
- [33] Haralampos Tsaknakis and Paul G Spirakis. An optimization approach for approximate Nash equilibria. In *Internet and Network Economics*, pages 42–56. Springer, 2007.
- [34] Haralampos Tsaknakis and Paul G Spirakis. Practical and efficient approximations of Nash equilibria for win-lose games based on graph spectra. In *Internet and Network Economics*, pages 378–390. Springer, 2010.
- [35] Helge Tverberg. A generalization of Radon’s theorem. *J. London Math. Soc.*, 41(1):123–128, 1966.