Approximating Carathéodory’s Theorem and Nash Equilibria

Siddharth Barman*

Abstract

We prove an approximate version of Carathéodory’s theorem and present its algorithmic applications. In particular, given a set of vectors $X$ in $\mathbb{R}^d$, we show that for every vector in the convex hull of $X$ there exists a “nearby” vector that can be expressed as a convex combination of at most $b$ vectors of $X$, where the bound $b$ is independent of the dimension $d$.

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 approximate Nash equilibrium of the game can be computed in time $n^t$, where $t$ has a logarithmic dependence on $s$. Hence, our algorithm provides a novel understanding of the time required to compute an approximate Nash equilibrium in terms of the column sparsity $s$ of $A + B$. 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 is obtained in [Lipton, Markakis, and Mehta 2003].

The theorem also leads to an additive approximation algorithm for the densest $k$-bipartite subgraph problem. Given a graph with $n$ vertices and maximum degree $d$, the developed algorithm determines a $k \times k$ bipartite subgraph with near (in the additive sense) optimal density in time $n^t$, where $t$ has a logarithmic dependence on the degree $d$.

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 $X$. 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. Specifically, we prove that given a set of vectors $X$ in the $p$-unit ball $\mathbb{B}^p$ with norm $p \in [2, \infty)$, for every vector $\mu$ in the convex hull of $X$ there exists an $\varepsilon$ close—under the $p$-norm distance—vector $\mu'$ that can be expressed as a convex combination of at most $O(p/\varepsilon^2)$ vectors of $X$. A notable aspect

---

1California Institute of Technology. barman@caltech.edu

This bound of $d + 1$ is tight.

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

of this result is that the number of vectors of $X$ that are required to express $\mu'$, i.e., $O(p/\varepsilon^2)$, is independent of the underlying dimension $d$.

Given the significance of Carathéodory’s theorem, this approximate version is interesting in its own right. This paper further substantiates its relevance by showing how it leads to interesting algorithmic results. Our applications include additive approximation algorithms for (i) Nash equilibria in two-player games, and (ii) the densest bipartite subgraph problem. We outline these algorithmic results below.

**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 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 \[6, 8\]. But, the question whether an approximate Nash equilibrium can be computed in polynomial time still remains open. Thought this paper we will consider the standard additive notion of approximate Nash equilibria that are defined as follows: a product of independent distributions, one for each player, is said to be an $\varepsilon$-Nash equilibrium if unilateral deviation increases utility by at most $\varepsilon$, 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 $\varepsilon$-Nash equilibrium of the game can be computed in time $n \tilde{O}(\log s/\varepsilon^2)$. Our result, in particular, shows that games with fixed column sparsity $s$ admit a polynomial-time approximation scheme (PTAS) for Nash equilibrium. Note that such games are a natural generalization of zero-sum games; recall that zero-sum games admit efficient computation of Nash equilibrium (see, e.g., \[21\]). It is also worth pointing out that for an arbitrary bimatrix game the running time of our algorithm is $n \tilde{O}(\log n/\varepsilon^2)$, since $s$ is at most $n$. Given that the best-known algorithm for computing $\varepsilon$-Nash equilibrium also runs in time $n \tilde{O}(\log s/\varepsilon^2)$ \[18\], 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 $A + B$.

We also refine the following result of Daskalakis and Papadimitriou \[11\]: They develop a PTAS for bimatrix games that contain 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 $n^t$, where $t$ has a logarithmic dependence on $s/m$. Since $s \leq n$, we get the result of \[11\] as a
special case.

**Densest Bipartite Subgraph.** In the densest \( k \)-bipartite subgraph problem (DkBS) we are given a graph and the objective is to find size-\( k \) vertex subsets, \( S \) and \( T \), such that the number of edges (of the graph) between \( S \) and \( T \) is maximized. Here, the density of a bipartite subgraph, induced by vertex subsets \( S \) and \( T \), is defined to be the number of edges between the two subsets divided by \( |S||T| \). DkBS is a natural variant of the standard densest subgraph problem (see, e.g., [1] and references therein), in which the objective is to determine the densest subgraph of a specified size. This problem is computationally hard and it is shown in [1] that a constant-factor approximation for the densest subgraph problem (DS) is unlikely. This result indicates that DkBS is hard to approximate within a constant factor as well, since the optimal densities in these two problems (DS with size bound \( 2k \)) can differ by at most a factor of two.

In this paper we focus on additive approximations for DkBS, which entail determining a \( k \times k \) bipartite subgraph whose density is close (in the additive sense) to the optimal. A lower bound for additive approximations in general graphs follows from the work of Hazan and Krauthgamer [14]. Specifically, the reduction established in [14] rules out an additive PTAS for DkBS, under complexity theoretic assumptions. Complementing this result, we develop the following upper bound: given a graph with \( n \) vertices and maximum degree \( d \), a \( k \times k \) bipartite subgraph with density \( \varepsilon \) close to the optimal can be computed in time \( n^{O\left(\frac{\log d}{\varepsilon^2}\right)} \).

### 1.1 Related Work

The Euclidean case, i.e., \( p = 2 \), of the approximate version of Carathéodory’s theorem admits a direct proof (see, e.g., [26] or Appendix A). A key contribution of this paper is to establish the result for arbitrary norms \( p \in [2, \infty) \). Our generalization of this results to arbitrary norms \( p \in [2, \infty) \) requires new ideas, including an interesting application of the Kahane’s inequality (see Theorem 2), and is critical in proving our algorithmic applications. In particular, we prove the correctness of the proposed algorithms using the fact that the result holds for general norms between 2 and \( \infty \).

**Approximate Nash Equilibria.** The computation of equilibria is an active area of research. Nash equilibria is known to be computationally hard [6, 8], 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 [18, 16, 9, 15, 10, 17, 12, 14, 24, 25, 2] and lower bounds [14, 7]. In particular, it is known that for a general bimatrix game an approximate Nash equilibrium can be computed in quasi-polynomial time [18]. Polynomial time algorithms have been developed for computing approximate Nash equilibria for fixed values of the approximation factor \( \varepsilon \); the best-known result of this type shows that a 0.3393-approximate Nash equilibrium can be computed in polynomial time [24]. In addition, several interesting classes of games have been

---

3 They reduce the problem of determining a planted clique to that of computing an \( \varepsilon \)-additive approximation for DkBS, with a sufficiently small but constant \( \varepsilon \).
identified that admit a PTAS \cite{15, 11, 25, 2}. For example, the result of Alon et al. \cite{2} provides a PTAS for games in which the sum of the payoff matrices, i.e., $A + B$, has logarithmic rank. Our paper complements such rank based results, since a sparse matrix can have high rank and a low-rank matrix can have high sparsity.

Chen et al. \cite{5} considered sparsity in the context of games and showed that computing a 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 \cite{11} that such games admit a trivial PTAS\footnote{In particular, the product of uniform distributions over players’ actions corresponds to an approximate Nash equilibrium in such games.}. Note that we study a strictly larger class of games and we provide a PTAS for games in which the row or column sparsity of $A + B$ is fixed.

**Densest Bipartite Subgraph.** The best-known (multiplicative) approximation ratio for densest subgraph (DS) is $n^{(1/4+\omega(1))}$ \cite{3}. But unlike this result, we address additive approximations in this paper. Our approximation algorithm for DkBS is based on solving a bilinear program that was formulated by Alon et al. \cite{2}. Specifically, this bilinear program was used in \cite{2} to develop an additive PTAS for DkBS for 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

We establish the approximate version of Carathéodory’s theorem via the probabilistic method. Given a vector $\mu$ in the convex hull of a set $X \subset \mathbb{R}^d$, we consider a convex combination of vectors of $X$ that generates $\mu$. The coefficients in this convex combination induce a probability distribution over $X$ and the mean of this distribution is $\mu$. Our 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 $\mu$ under the $p$-norm distance, for $p \in [2, \infty)$. Therefore, the probabilistic method implies that there exists a vector close to $\mu$ that can be expressed as a convex combination of at most $b$ vectors, where $b$ is the number of samples we drew. The primary goal is to show that $b$ is independent of the underlying dimension $d$. Note that 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 $\mu$, for every $i \in [d]$; since this would entail a union bound over the number of components $d$. Bypassing such a component-wise analysis, we are able to bound (in expectation) the $p$-norm distance between $\mu$ and the sample mean vector via an interesting application of Khintchine inequality (see Theorem \cite{2}).

**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 \cite{19} and its optimal (near-optimal) solutions correspond to 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.\footnote{We ignore the linear part of the objective for ease of presentation, see Section 4 for details.} Suppose we knew the vector $u := C \hat{y}$, for some Nash equilibrium $(\hat{x}, \hat{y})$. Then, a Nash equilibrium can be efficiently determined 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 $C y = u$. 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 our 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: $\|C^i\|_p \leq (s \cdot 2^p)^{1/p} = 2 \cdot 2^{\log_2 s} \leq 4$, 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 our 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.

**Densest Bipartite Subgraph.** Note that the algorithmic approach outlined above applies to any 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 approximation for DkBS. It was shown in \cite{2} that DkBS can be formulated as a bilinear program. In this program, the matrix in the objective (i.e., the bilinear form) is the adjacency matrix of the graph; hence, its column sparsity corresponds to the maximum degree of the graph. At a high level, our result for DkBS follows from these observations.

### 1.3 Organization

We begin by setting up notation in Section 2. Then, in Section 3 we prove the approximate version of Carathéodory’s theorem. Appendix A provides a somewhat elementary proof of our result for the Euclidean-norm case. Finally, the algorithmic applications of the theorem are presented in...
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 $\ell_0$ “norm”: $\|x\|_0 := |\{i \mid x_i \neq 0\}|$.

Given a set $X = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d$, 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

First we state McDiarmid’s inequality [20]. We will use this concentration bound in the proof of our theorem.

**Theorem 1** (McDiarmid’s inequality). Let $Z_1, Z_2, \ldots, Z_m \in \mathbb{Z}$ be independent random variables and $f : \mathbb{Z}^m \to \mathbb{R}$ be a function of $Z_1, Z_2, \ldots, Z_m$. If for all $i \in [m]$ and for all $z_1, z_2, \ldots, z_m, z'_i \in \mathbb{Z}$ the function $f$ satisfies

$$|f(z_1, \ldots, z_i, \ldots, z_m) - f(z_1, \ldots, z'_i, \ldots, z_m)| \leq c_i,$$

then for $t > 0$,

$$\Pr(|f - \mathbb{E}[f]| \geq t) \leq 2 \exp \left( \frac{-2t^2}{\sum_{i=1}^{m} c_i^2} \right).$$

A key technical ingredient in our proof is Kahane’s inequality (see, e.g., [13]), which can be considered as a vector version of Khintchine inequality.

The following form of the inequality can be derived from a result stated in [22] (see Theorem 2 in [22] and references therein).

**Theorem 2** (Kahane’s inequality [22, 23]). Let $r_1, r_2, \ldots, r_m$ be a sequence of i.i.d. Rademacher $\pm 1$ random variables, i.e., $\Pr(r_i = \pm 1) = \frac{1}{2}$ for all $i \in [m]$. In addition, let $u_1, u_2, \ldots, u_m \in \mathbb{R}^d$ be a deterministic sequence of vectors. Then, for $2 \leq p < \infty$

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

(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 $u_1, u_2, \ldots, u_m \in \mathbb{R}^d$, i.e., set $Q_i = \text{diag}(u_i)$ for all $i \in [m]$, then for any sequence of scalars $\xi_1, \xi_2, \ldots, \xi_m \in \mathbb{R}$ we have $\sum_{i=1}^m \xi_i Q_i = \text{diag}(\sum_{i=1}^m \xi u_i)$.

In order to prove the theorem statement for vectors $u_1, u_2, \ldots, u_m \in \mathbb{R}^d$, we simply use Theorem 2 of [22]. In particular, setting $Q_i = \text{diag}(u_i)$ for all $i \in [m]$ in Theorem 2 of [22] and using the above stated observations we get:

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

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

Theorem 3. Given a set of vectors $X = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d$ and $\varepsilon > 0$. For every $\mu \in \text{conv}(X)$ and $2 \leq p < \infty$ there exists an $O \left( \frac{\varepsilon^2}{d^2} \right)$ 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 $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ corresponds to a probability distribution over vectors $x_1, x_2, \ldots, x_n$. That is, under probability distribution $\alpha$ vector $x_i$ is drawn with probability $\alpha_i$. The vector $\mu$ is the mean of this distribution. Specifically, the $j$th component of $\mu$ is the expected value of the random variable that takes value $x_{i,j}$ with probability $\alpha_i$, here $x_{i,j}$ is the $j$th component of vector $x_i$. We succinctly express these component-wise equalities as follows:

$$E_{v \sim \alpha}[v] = \mu.$$ \hfill (3)

Let $v_1, v_2, \ldots, v_m$ be $m$ i.i.d. draws from $\alpha$. 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 $\mu$.

$$g(v_1, v_2, \ldots, v_m) := \left\| \frac{1}{m} \sum_{i=1}^m v_i - \mu \right\|_p.$$ \hfill (4)
We will use McDiarmid’s inequality. In particular, we will establish that with positive probability the sample mean vector defined over \( m = \Theta \left( \frac{\nu^2}{\varepsilon^2} \right) \) draws is \( \varepsilon \) close to \( \mu \) in \( p \)-norm. Hence, the stated claim is implied by the probabilistic method.

For any \( m \) tuple \((v_1, \ldots, v_i, \ldots, v_m) \in \mathcal{X}^m\), and \( v'_i \in \mathcal{X} \) we show that \(|g(v_1, \ldots, v_i, \ldots, v_m) - g(v_1, \ldots, v'_i, \ldots, v_m)|\) is no more than \( 2\gamma/m \); recall that \( \gamma := \max_{x \in \mathcal{X}} \|x\|_p \). We can assume without loss of generality that \( g(v_1, \ldots, v_i, \ldots, v_m) \geq g(v_1, \ldots, v'_i, \ldots, v_m) \), since the other case is symmetric.

Setting \( u := \frac{1}{m} \sum_{j \neq i} v_j - \mu \) we have

\[
g(v_1, \ldots, v_i, \ldots, v_m) - g(v_1, \ldots, v'_i, \ldots, v_m) = \|u + \frac{1}{m} v_i\|_p - \|u + \frac{1}{m} v'_i\|_p \leq \|u\|_p + \frac{1}{m} \|v_i\|_p - \|u\|_p + \frac{1}{m} \|v'_i\|_p \leq \frac{1}{m} \|v_i\|_p + \frac{1}{m} \|v'_i\|_p \leq \frac{2\gamma}{m}\]  

Here the upper bound follows from triangle inequality: for vectors \( a \) and \( b \) we have \( \|a\|_p - \|b\|_p \leq \|a + b\|_p \leq \|a\|_p + \|b\|_p \). Inequality follows from the fact that \( v_i, v'_i \in \mathcal{X} \) for all \( i \), and hence their \( p \) norm is no more than \( \gamma \).

Given that \( g \) satisfies \(|g(v_1, \ldots, v_i, \ldots, v_m) - g(v_1, \ldots, v'_i, \ldots, v_m)| \leq 2\gamma/m \), we can apply McDiarmid’s inequality (see Theorem (1)), with \( c_i = 2\gamma/m \) for all \( i \in [m] \), to obtain

\[
\Pr(|g - \mathbb{E}[g]| \geq t) \leq 2 \exp \left( \frac{-2t^2m}{4\gamma^2} \right)
\]

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

\[
\mathbb{E}[g] \leq \frac{2\sqrt{p} \gamma}{\sqrt{m}}.
\]

Recall that in expectation the sampled mean is equal to \( \mu \), i.e., \( \mathbb{E}_{v_1, \ldots, v_m} \frac{1}{m} \sum_{i=1}^{m} v_i = \mu \). Hence, we have

\[
\mathbb{E}[g] = \mathbb{E}_{v_1, \ldots, v_m} \left\| \frac{1}{m} \sum_{i=1}^{m} v_i - \mu \right\|_p = \mathbb{E}_{v_1, \ldots, v_m} \left\| \frac{1}{m} \sum_{i=1}^{m} v_i - \mathbb{E}_{v'_1, \ldots, v'_m} \frac{1}{m} \sum_{i=1}^{m} v'_i \right\|_p = \mathbb{E}_{v_1, \ldots, v_m} \left\| \mathbb{E}_{v'_1, \ldots, v'_m} \left( \frac{1}{m} \sum_{i=1}^{m} v_i - \frac{1}{m} \sum_{i=1}^{m} v'_i \right) \right\|_p
\]
Note that \( \| \cdot \|_p \) is convex for \( p \geq 1 \). Therefore, Jensen’s inequality gives us:

\[
\mathbb{E}_{v_1, \ldots, v_m} \left[ \mathbb{E}_{v_1', \ldots, v'_m} \left( \frac{1}{m} \sum_{i=1}^{m} v_i - \frac{1}{m} \sum_{i=1}^{m} v'_i \right) \right]_p \leq \mathbb{E}_{v_1, \ldots, v_m} \mathbb{E}_{v_1', \ldots, v'_m} \left[ \left( \frac{1}{m} \sum_{i=1}^{m} v_i - \frac{1}{m} \sum_{i=1}^{m} v'_i \right) \right]_p
\]

(14)

\[
= \frac{1}{m} \mathbb{E}_{v_1, \ldots, v_m} \left( \sum_{i=1}^{m} (v_i - v'_i) \right)_p
\]

(15)

Let \( r_1, r_2, \ldots, r_m \) be a sequence of i.i.d. Rademacher \( \pm 1 \) 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

\[
\frac{1}{m} \mathbb{E}_{v_i, v'_i} \left( \sum_{i=1}^{m} (v_i - v'_i) \right)_p = \frac{1}{m} \mathbb{E}_{v_i, v'_i, r_i} \left( \sum_{i=1}^{m} r_i (v_i - v'_i) \right)_p
\]

(Triangle inequality)

\[
\leq \frac{1}{m} \mathbb{E}_{v_i, v'_i, r_i} \left[ \sum_{i=1}^{m} r_i v_i \right]_p + \sum_{i=1}^{m} r_i v'_i \right) \left( \sum_{i=1}^{m} r_i v'_i \right)_p\right]
\]

(Tower property)

\[
= \mathbb{E}_{v_i, v'_i} \left( \sum_{i=1}^{m} r_i v_i \right)_p + \mathbb{E}_{v'_i} \left( \sum_{i=1}^{m} r_i v'_i \right)_p
\]

\[
= \mathbb{E}_{v_i, r_i} \left( \sum_{i=1}^{m} r_i v_i \right)_p
\]

(16)

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

\[
\mathbb{E}_{v_i} \left( \sum_{i=1}^{m} r_i v_i \right)_p \left( r_1, \ldots, r_m \right) = \mathbb{E}_{v'_i} \left( \sum_{i=1}^{m} r_i v'_i \right)_p \left( r_1, \ldots, r_m \right)
\]

(17)

Overall, inequalities (13), (15), and (16) imply

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

where \( r_1, r_2, \ldots, r_m \) is a sequence of i.i.d. Rademacher \( \pm 1 \) random variables.
At this point we can apply Theorem 2 (Kahane’s inequality) with \( u_i = \frac{v_i}{m} \) to obtain
\[
\mathbb{E}_{v_i,r_i} \left\| \sum_{i=1}^{m} \frac{r_i v_i}{m} \right\|_p = \mathbb{E}_{v_i} \left[ \mathbb{E}_{r_i} \left[ \left\| \sum_{i=1}^{m} \frac{r_i v_i}{m} \right\|_p \right| v_1, \ldots, v_m \right] 
\]
\[
\leq \mathbb{E}_{v_i} \left[ \sqrt{p} \left( \sum_{i=1}^{m} \left\| \frac{v_i}{m} \right\|_p^2 \right)^{1/2} \right] 
\]
\[
\leq \mathbb{E}_{v_i} \left[ \sqrt{p} \left( \sum_{i=1}^{m} \frac{\gamma^2}{m^2} \right)^{1/2} \right] 
\]
\[
= \sqrt{p} \frac{\gamma}{\sqrt{m}} 
\]  
(22)

Inequality (21) uses the fact that random vectors \( v_i \) are supported over \( X \), so \( \left\| v_i \right\|_p \leq \gamma^2 \).

Using (18) and (22) we now have
\[
\mathbb{E}[g] \leq \frac{2\sqrt{p} \gamma}{\sqrt{m}} 
\]  
(23)

For sample size \( m = O \left( \frac{\gamma^2}{\varepsilon^2} \right) \) we have \( \mathbb{E}[g] \leq \varepsilon/2 \). Setting \( t = \frac{\varepsilon}{2} \) in inequality (9) gives us:
\[
\Pr (g \geq \varepsilon) \leq 2e^{-t} = 2e^{-2}. 
\]  
(24)

Therefore, with positive probability \( \left\| \frac{1}{m} \sum_{i=1}^{m} v_i - \mu \right\|_p \leq \varepsilon \). By the probabilistic method we get that there exists a vector \( \mu' \) that is \( \varepsilon \) close to \( \mu \) in \( p \) norm and can be expressed as a convex combination of at most \( m = O \left( \frac{\gamma^2}{\varepsilon^2} \right) \) vectors of \( X \).

We end this section by stating an \( \infty \)-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 \( \infty \)-norm.

**Theorem 4.** Given a set of vectors \( X = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d \), with \( \max_{x \in X} \|x\|_\infty \leq 1 \), and \( \varepsilon > 0 \). For every \( \mu \in \text{conv}(X) \) there exists an \( \tilde{O} \left( \frac{\log n}{\varepsilon^2} \right) \) 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 $[n] = \{1, 2, \ldots, n\}$, 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 $i, j \in [n]$.

Let $\Delta^n$ be the set of probability distributions over the set of pure strategies $[n]$. For $x \in \Delta^n$, we define $\text{Supp}(x) := \{i \mid x_i > 0\}$. Similarly, for a vector $v \in \mathbb{R}^n$ write $\text{Supp}(v)$ to denote the set $\{i \mid v_i > 0\}$. Further, $e_i \in \mathbb{R}^n$ is 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 $\Delta^n$, 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:

\begin{align}
  x^T A y & \geq e_i^T A y & \forall i \in [n] \quad \text{and} \quad (25) \\
  x^T B y & \geq x^T B e_j & \forall j \in [n] \quad (26) \\
  x^T A y & \geq e_i^T A y & \forall i \in [n] \quad (27)
\end{align}

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 $\varepsilon$-Nash equilibrium is no player can benefit more than $\varepsilon$, in expectation, by unilateral deviation. Formally,

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

\begin{align}
  x^T A y & \geq e_i^T A y - \varepsilon & \forall i \in [n] \quad \text{and} \quad (28) \\
  x^T B y & \geq x^T B e_j - \varepsilon & \forall j \in [n] \quad (29) \\
  x^T A y & \geq e_i^T A y & \forall i \in [n] \quad (30)
\end{align}

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 $\max_i \|C^i\|_0$ 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 \(C = A + B\).

Next we state the quantitative connection that we establish between the sparsity of a game and the time it take to compute an \(\varepsilon\)-Nash equilibrium.

**Theorem 5.** Let \(A, B \in [-1, 1]^{n \times n}\) be the payoff matrices of an \(s\)-sparse bimatrix game. Then, an \(\varepsilon\)-Nash equilibrium of \((A, B)\) can be computed in time

\[
nO\left(\frac{\log s}{\varepsilon^2}\right).
\]

Our algorithm for computing \(\varepsilon\)-Nash equilibrium relies on the following bilinear program, which was formulated by Mangasarian and Stone [19]. As formally specified in Lemma 1 below, approximate solutions of this bilinear program correspond to approximate Nash equilibria.

\[
\max_{x, y, \pi_1, \pi_2} \quad x^T C y - \pi_1 - \pi_2 \\
\text{subject to} \quad x^T B \leq \mathbb{1}^T \pi_2 \\
\quad A^T y \leq \mathbb{1} \pi_1 \\
\quad x, y \in \Delta^n \\
\quad \pi_1, \pi_2 \in [-1, 1].
\]

(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 6** (Equivalence Theorem [19]). 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 is \(\hat{\pi}_1\) and \(\hat{\pi}_2\) respectively.

A relevant observation is that that an approximate solution of the bilinear program corresponds to \(\varepsilon\)-Nash equilibrium.

**Lemma 1.** Let \(x, y \in \Delta^n\) along with scalars \(\pi_1\) and \(\pi_2\) form a feasible solution of (BP) that achieves an objective function value of more than \(-\varepsilon\), i.e., \(x^T C y \geq \pi_1 + \pi_2 - \varepsilon\). Then, \((x, y)\) is an \(\varepsilon\)-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 \(\pi_1\) and \(x^T B y\) is at most \(\pi_2\). So, the following inequalities must hold:

\[
x^T A y \geq \pi_1 - \varepsilon \quad \text{and} \quad x^T B y \geq \pi_2 - \varepsilon
\]

(31)

(32)

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 \(\varepsilon\)-Nash equilibrium. \(\square\)

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

\[
\min_{x,y,\pi_1,\pi_2} \|Cy - u\|_p
\]

subject to

\[
x^T u \geq \pi_1 + \pi_2 - \varepsilon/2
\]

\[
Ay \leq 1\pi_1
\]

\[
x^T B \leq 1^T \pi_2
\]

\[
x, y \in \Delta^n
\]

\[
\pi_1, \pi_2 \in [-1, 1].
\]

Proof of Theorem 5. Algorithm 1 iterates at most \(n^{O\left(\frac{1}{p^2}\right)}\) times, since this is an upper bound on the number of multisets of size \(O\left(\frac{1}{\varepsilon^2}\right)\). Furthermore, in each iteration the algorithm solves convex program \(CP(u)\), 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 Algorithm 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 \(\varepsilon\)-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 \(CP(u)\). Then,

\[
|x^T(Cy - u)| \leq \|x\|_q\|Cy - u\|_p \quad \text{(Hölder’s inequality)}
\]

\[
\leq 1 \times \varepsilon/2
\]

(33)

(34)

Here \(q = p/(p - 1) \geq 1\), since \(p \geq 2\). The second inequality follows from the fact that the objective function value of \(CP(u)\) is no more than \(\varepsilon/2\) and \(\|x\|_q \leq \|x\|_1 = 1\). Since the returned \(x\) satisfies

6Note that for fixed \(u\), \(CP(u)\) 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 $\varepsilon$-Nash equilibrium in $s$-sparse games

Given payoff matrices $A, B \in [-1, 1]^{n \times n}$ and $\varepsilon > 0$; Return: $\varepsilon$-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 $\{1, 2, \ldots, n\}$ of cardinality at most $\frac{s^p}{2}$, where $\kappa$ is a fixed constant.
3: for all multisets $S \in \mathcal{U}$ do
   4: Set $u = \frac{1}{|S|} \sum_{i \in S} C^i$.
   {Note that $u$ is a $|S|$-uniform vector in the convex hull of the columns of $C$.}
5: Solve convex program CP$(u)$.
6: if the objective function value of CP$(u)$ is less than $\varepsilon/2$ then
7:  Return $(x, y)$, where $x$ and $y$ form an optimal solution of CP$(u)$.
8: end if
9: end for

the feasibility constraints in 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 $\varepsilon$-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

$$\|C^i\|_p \leq 4 \quad \forall i \in [n]. \quad (35)$$

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:

$$\|C^i\|_p \leq (2p s)^{1/p} = 2(s)^{1/p} = 2(2^{\log s})^{1/p} = 4.$$ 

Therefore, for $p = \log s$, we can apply Theorem 3 over the convex hull $\text{conv} \{C^i\}$ with $\gamma \leq 4$. In particular, for $\mu = C \hat{y}$, Theorem 3 implies that there exists a $O(\frac{p}{\varepsilon})$ uniform vector $\mu' \in \text{conv}(\{C^i\})$ such that

$$\|C \hat{y} - \mu'\|_p \leq \varepsilon/2.$$ 

Since $\mu'$ is $O(\frac{p}{\varepsilon})$ uniform, at some point during its execution the algorithm (with an appropriate value of $\kappa$) will set $u = \mu'$. Therefore, at least once the algorithm will consider a $u$ that satisfies:

$$\|C \hat{y} - u\|_p \leq \varepsilon/2 \quad (36)$$
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 (36).

Since $(\hat{x}, \hat{y})$ is a Nash equilibrium, using Theorem 6 we get

$$\hat{x}^T C \hat{y} = \hat{\pi}_1 + \hat{\pi}_2. \quad (37)$$

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

$$|\hat{x}^T (C\hat{y} - u)| \leq \|\hat{x}\|_q \|C\hat{y} - u\|_p \quad \text{(Hölder’s inequality)} \quad (38)$$

$$\leq 1 \times \varepsilon/2. \quad (39)$$

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 6.

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

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

\text{Remark 2.} & \quad \text{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, \text{ we include the constraint } \pi_1 + \pi_2 \geq \alpha \text{ in } \text{CP}(u). \text{ The time complexity of the algorithm stays the same, and then via a binary search over } \alpha \text{ we can find an approximate Nash equilibrium with near-optimal social welfare.} \\

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

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 [11] showed that there exists a PTAS for games that contain an equilibrium with small—specifically, $O\left(\frac{1}{n}\right)$—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 $O\left(\frac{1}{m}\right)$, 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 [11] 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 [11] 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 [11] corresponds to games that have $\delta n$-small probability equilibrium. Next we prove a result for general $m$-small probability equilibrium.

Theorem 7. 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 $\varepsilon$-Nash equilibrium of the game can be computed in time

$$n^{O\left(\frac{1}{t^2}\right)},$$

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 [11] is modified as follows: (i) use the updated value of $p$ instead of the one specified in Step [11] of the algorithm; (ii) include convex constraint $\|x\|_q \leq m^{-1/p}$ in 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 [11] shows that any returned solution $(x, y)$ is an $\varepsilon$-Nash equilibrium.

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

$$|x^T (C y - u)| \leq \|x\|_q \|C y - u\|_p$$

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

$$= \varepsilon/2$$

16
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 CP$(u)$ is at most $\frac{\epsilon m^{1/p}}{2}$. Since $x$ is a feasible solution of CP$(u)$, it satisfies $x^T u \geq \pi_1 + \pi_2 - \epsilon/2$. Therefore, using inequality (12) we get $x^T C y \geq \pi_1 + \pi_2 - \epsilon$, 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 CP$(u)$ and achieves an objective function value of at most $\frac{\epsilon m^{1/p}}{2}$. 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 $\|C^i\|_p \leq (s 2^p)^{1/p} = 2s^{1/p}$.

Let $(\hat{x}, \hat{y})$ be an $m$-small probability Nash equilibrium of the game. Theorem 3, applied over the convex hull conv$(\{C^i\})$ with $\gamma \leq 2s^{1/p}$, guarantees that there exists a $O\left(\frac{p s^{2/p}}{m^{2/p}}\right)$ uniform vector that is $\frac{\epsilon m^{1/p}}{2}$ close to $C\hat{y}$. Since $p \geq 2 \log \left(\frac{s}{m}\right)$, we have

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

Therefore, there exists a $O\left(\frac{p s}{m}\right)$ uniform vector that is $\frac{\epsilon m^{1/p}}{2}$ close to $C\hat{y}$. Such a vector, say $u$, will be selected by the algorithm in Step 8 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:

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

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 6). 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 $\hat{x}^T u \geq \hat{\pi}_1 + \hat{\pi}_2 - \epsilon/2$. This follows from Hölder’s inequality:

$$|\hat{x}^T (C\hat{y} - u)| \leq \|\hat{x}\|_q \|C\hat{y} - u\|_p \leq m^{-1/p} \epsilon m^{1/p} \leq \epsilon/2.$$
Overall, for the above specified \( u \), the “if” condition in Step 4 will be satisfied. This shows that the algorithms successfully returns an \( \varepsilon \)-Nash equilibrium.

The algorithm iterates at most \( n^{O(\frac{1}{\varepsilon^2})} \) times, since this is an upper bound on the number of multisets of size \( O(\frac{1}{\varepsilon^2}) \). 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.

5 Densest Bipartite Subgraph

This section presents an additive approximation for the densest \( k \)-bipartite subgraph (DkBS) problem. 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:

\[
\rho(S, T) := \frac{|E(S, T)|}{|S||T|},
\]

(50)

here \( E(S, T) \) denotes the set of edges that connect \( S \) and \( T \). We use \( \rho_k^*(G) \) to denote the optimal bipartite density of the graph, i.e., \( \rho_k^*(G) := \max_{S,T \subseteq V : |S|, |T| = k} \rho(S, T) \). Note that the subsets \( S \) and \( T \) are not required to be disjoint.

Next we state a bilinear program from [2] to approximate DkBS. Here, \( A \) denotes the adjacency matrix of the given graph \( G \) and \( n = |V| \).

\[
\begin{align*}
\max_{x, y} \quad & x^T Ay \\
\text{subject to} \quad & x, y \in \Delta^n \\
& x_i, y_i \leq \frac{1}{k} \quad \forall i \in [n].
\end{align*}
\]

(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 \( \rho_k^*(G) \). In addition, given an additive \( \varepsilon \)-approximate solution of (BP-DkBS), \((x', y')\), we can efficiently determine an \( \varepsilon \)-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 \( \rho(S', T') \geq \rho_k^*(G) - \varepsilon \). In other words, in order to determine an approximate solution of DkBS it suffices to compute an approximate solution of (BP-DkBS). To complete the argument, next we show that (BP-DkBS) can be efficiently approximated when the column sparsity of the adjacency matrix (i.e., the degree of the graph) is low.
If the maximum degree of the graph \( G \) is \( d \) then the number of non-zero components in any column of \( A \) is no more than \( d \). This implies that, for \( i \in [n] \), we have \( \|A^i\|_p \leq d^{1/p} \). Here, \( A^i \) denotes the \( i \)th column of \( A \).

Now for \( p = \log d \) the following bound holds for all \( i \in [n] \): \( \|A^i\|_p \leq 2 \). Therefore, along the lines of Algorithm 1 we can enumerate over all \( O\left(\frac{d}{\varepsilon^2}\right) \)-uniform vectors in the convex hull of the columns of \( A \) and find an \( \varepsilon \)-approximate solution of \( \text{(BP-DkBS)} \). This establishes the following theorem.

**Theorem 8.** 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 \( \rho^*_k(G) - \varepsilon \).

**Acknowledgements**

The author thanks Federico Echenique and Anthony Man-Cho So for helpful discussions. 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 \( \kappa \)-subgraph from average case hardness, 2011.

[2] 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.

[3] 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.

[4] 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.

[5] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Sparse games are hard. In *Internet and Network Economics*, pages 262–273. Springer, 2006.

[6] 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.

[7] Constantinos Daskalakis. On the complexity of approximating a nash equilibrium. *ACM Transactions on Algorithms (TALG)*, 9(3):23, 2013.
[8] 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.

[9] Constantinos Daskalakis, Aranyak Mehta, and Christos Papadimitriou. A note on approximate nash equilibria. In *Internet and Network Economics*, pages 297–306. Springer, 2006.

[10] 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.

[11] 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.

[12] 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.

[13] D.J.H. Garling. *Inequalities: a journey into linear analysis*, volume 19. Cambridge University Press Cambridge, 2007.

[14] Elad Hazan and Robert Krauthgamer. How hard is it to approximate the best nash equilibrium? *SIAM Journal on Computing*, 40(1):79–91, 2011.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.
A Proof of the Euclidean Case

Here we give an elementary proof of the approximate version of Carathéodory’s theorem specifically for the Euclidean case.

Theorem 9. Given a set of vectors $X = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d$, with $\gamma := \max_{x \in X} \|x\|$, and $\varepsilon > 0$. For every $\mu \in \text{conv}(X)$ there exists a $\frac{16\gamma^2}{\varepsilon^2}$ uniform vector $\mu' \in \text{conv}(X)$ such that $\|\mu - \mu'\| \leq \varepsilon$.

Proof. Write $\mu \in \text{conv}(X)$ 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 $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ corresponds to a probability distribution over vectors $x_1, x_2, \ldots, x_n$. The vector $\mu$ is the mean of this distribution; specifically, the $j$th component of $\mu$ is the expected value of the random variable that takes value $x_{i,j}$ with probability $\alpha_i$; here $x_{i,j}$ is the $j$th component of vector $x_i$. We succinctly express these component-wise equalities as follows:

$$E_{\alpha}[x] = \mu, \quad (51)$$

here $x$ is a random vector that is equal to vector $x_i$ with probability $\alpha_i$.

We establish the existence of vector $\mu'$ via the probabilistic method. We draw $m = \frac{16\gamma^2}{\varepsilon^2}$ i.i.d. vectors from the distribution $\alpha$ and set $\mu'$ to be the sample mean vector. Using McDiarmid’s
inequality we show that with positive probability the following holds: \( \| \mu - \mu' \| \leq \varepsilon \). This proves the existence of vector \( \mu' \) that satisfies the desired properties.

Let \( v_1, v_2, \ldots, v_m \in X \) be \( m \) independent draws from \( \alpha \). The sample mean vector (associated with the \( m \) independently drawn vectors) is \( \frac{1}{m} \sum_{i=1}^{m} v_i \). Note that, for all \( i \in [m] \), \( v_i \in X \) and hence the sample mean vector is a convex combination of at most \( m \) vectors from \( X \).

Applying linearity of expectation (component wise) we get
\[
\mathbb{E}_{v_1, \ldots, v_m \sim \alpha} \left( \frac{1}{m} \sum_{i=1}^{m} v_i \right) = \mu. \tag{52}
\]

In order to bound the Euclidean distance between the sampled mean vector and \( \mu \) (the actual mean) we will consider the following function \( f : X^m \rightarrow \mathbb{R} \),
\[
f(v_1, v_2, \ldots, v_m) := \left\| \frac{1}{m} \sum_{i=1}^{m} v_i - \mu \right\| \tag{53}
\]

For any \( m \) tuple \( (v_1, \ldots, v_i, \ldots, v_m) \in X^m \), and \( v'_i \in X \) we show that \( |f(v_1, \ldots, v_i, \ldots, v_m) - f(v_1, \ldots, v'_i, \ldots, v_m)| \) is no more than \( 2\gamma/m \); recall that \( \gamma := \max_{x \in X} \|x\| \). We can assume without loss of generality that \( f(v_1, \ldots, v_i, \ldots, v_m) \geq f(v_1, \ldots, v'_i, \ldots, v_m) \), since the other case is symmetric.

Setting \( u := \frac{1}{m} \sum_{j \neq i} v_j - \mu \) we have
\[
f(v_1, \ldots, v_i, \ldots, v_m) - f(v_1, \ldots, v'_i, \ldots, v_m) = \left\| u + \frac{1}{m} v_i \right\| \right. - \left. \left\| u + \frac{1}{m} v'_i \right\| \tag{54}
\]
\[
\leq \|u\| + \frac{1}{m} \|v_i\| - \|u\| + \frac{1}{m} \|v'_i\| \tag{55}
\]
\[
\leq \frac{1}{m} \|v_i\| + \frac{1}{m} \|v'_i\| \tag{56}
\]
\[
\leq \frac{2\gamma}{m}. \tag{57}
\]

Here the upper bound (55) follows from triangle inequality: for vectors \( a, b \) and \( b' \) we have \( \|a + b\| \leq \|a\| + \|b\| \) and \( \|a + b'\| \geq \|a\| - \|b'\| \).

Given that \( f \) satisfies \( |f(v_1, \ldots, v_i, \ldots, v_m) - f(v_1, \ldots, v'_i, \ldots, v_m)| \leq 2\gamma/m \), we can apply Mc-Diarmid’s inequality (see Theorem 1), with \( c_i = 2\gamma/m \) for all \( i \in [m] \), to obtain
\[
\Pr(\{|f - \mathbb{E}[f]| \geq t\}) \leq 2 \exp \left( \frac{-2t^2m}{4\gamma^2} \right) \tag{58}
\]

The remainder of the proof entails showing that \( \mathbb{E}[f] \) (i.e., the expected Euclidean distance between the sample mean vector and the underlying mean \( \mu \)) is no more than \( 2\gamma/\sqrt{m} \). We prove this bound next.

**Claim 1.** \( \mathbb{E}[f] \leq 2\gamma/\sqrt{m} \)
Proof. We establish this claim by induction over $m$. In particular, we prove that

$$
\mathbb{E}_{\nu_1, \nu_2, \ldots, \nu_m \sim \alpha} \left[ \left\| \frac{1}{m} \sum_i \nu_i - \mu \right\|^2 \right] \leq \frac{4\gamma^2}{m} 
$$

(59)

We get the stated claim, $\mathbb{E}_{\nu_1, \nu_2, \ldots, \nu_m \sim \alpha} \left[ \left\| \frac{1}{m} \sum_i \nu_i - \mu \right\| \right] \leq \frac{2\gamma}{\sqrt{m}}$ by applying Jensen’s inequality to (59).

For the base case, i.e., $m = 1$ we have

$$
\mathbb{E}_{\nu \sim \alpha} [\|v - \mu\|^2] \leq \mathbb{E}_{\nu \sim \alpha} [(\|v\| + \|\mu\|)^2] \quad \text{(triangle inequality)}
$$

(60)

$$
\leq 4\gamma^2.
$$

(61)

Here we use the fact that all the vectors in the support of $\alpha$ have Euclidean norm no more than $\gamma$ and $\|\mu\| \leq \gamma$, this follows from the convexity of the Euclidean norm: $\|\mu\| = \| \sum_i \alpha_i x_i \| \leq \sum_i \alpha_i \|x_i\| \leq \gamma$.

Assuming, by the induction hypothesis, the stated claim for sample size $m - 1$, we show that it holds for $m$ samples as well.

$$
\mathbb{E}_{\nu_1 \sim \alpha} \left[ \left\| \frac{1}{m} \sum_i \nu_i - \mu \right\|^2 \right] = \mathbb{E}_{\nu_1 \sim \alpha} \left[ \left\| \frac{m - 1}{m} \left( \frac{1}{m - 1} \sum_{i=1}^{m-1} \nu_i - \mu \right) + \frac{1}{m} (\nu_m - \mu) \right\|^2 \right]
$$

$$
= \frac{(m - 1)^2}{m^2} \mathbb{E}_{\nu_1, \ldots, \nu_{m-1} \sim \alpha} \left[ \left\| \frac{1}{m - 1} \sum_{i=1}^{m-1} \nu_i - \mu \right\|^2 \right] + \frac{1}{m^2} \mathbb{E}_{\nu_m \sim \alpha} \left[ \|\nu_m - \mu\|^2 \right] + \frac{2(m - 1)}{m^2} \mathbb{E}_{\nu_1 \sim \alpha} \left[ (\nu_m - \mu)^T \left( \frac{1}{m - 1} \sum_{i=1}^{m-1} \nu_i - \mu \right) \right]
$$

$$
< \frac{(m - 1)^2}{m^2} \frac{4\gamma^2}{m - 1} + \frac{1}{m^2} \mathbb{E}_{\nu_m \sim \alpha} \left[ \|\nu_m - \mu\|^2 \right] + \frac{2(m - 1)}{m^2} \mathbb{E}_{\nu_1 \sim \alpha} \left[ (\nu_m - \mu)^T \left( \frac{1}{m - 1} \sum_{i=1}^{m-1} \nu_i - \mu \right) \right]
$$

$$
\leq \frac{(m - 1)^2}{m^2} \frac{4\gamma^2}{m - 1} + \frac{4\gamma^2}{m^2} + \frac{2(m - 1)}{m^2} \mathbb{E}_{\nu_1 \sim \alpha} \left[ (\nu_m - \mu)^T \left( \frac{1}{m - 1} \sum_{i=1}^{m-1} \nu_i - \mu \right) \right]
$$

$$
= \frac{4\gamma^2}{m} + \frac{2(m - 1)}{m^2} \mathbb{E}_{\nu_1 \sim \alpha} \left[ (\nu_m - \mu)^T \left( \frac{1}{m - 1} \sum_{i=1}^{m-1} \nu_i - \mu \right) \right]
$$

Here the first inequality follows from the induction hypothesis and the second from (61). Finally, we show that $\mathbb{E}_{\nu_1 \sim \alpha} \left[ (\nu_m - \mu)^T \left( \frac{1}{m - 1} \sum_{i=1}^{m-1} \nu_i - \mu \right) \right] = 0$ thus proving the theorem.

Note that $(\nu_m - \mu)$ and $(\frac{1}{m - 1} \sum_{i=1}^{m-1} \nu_i - \mu)$ are independent (component wise):

$$
\mathbb{E}_{\nu_1 \sim \alpha} \left[ (\nu_m - \mu)^T \left( \frac{1}{m - 1} \sum_{i=1}^{m-1} \nu_i - \mu \right) \right] = \mathbb{E}[(\nu_m - \mu)^T] \mathbb{E} \left[ \left( \frac{1}{m - 1} \sum_{i=1}^{m-1} \nu_i - \mu \right) \right].
$$
Since \((v_m - \mu)\) is a mean-zero vector (and so is \(\left(\frac{1}{m-1} \sum_{i=1}^{m-1} v_i - \mu\right)\)), i.e., \(\mathbb{E}[(v_m - \mu)] = 0\), the desired claim follows.

Using the above claim and the fact that \(m = \frac{16\gamma^2}{\varepsilon^2}\), we get \(\mathbb{E}[f] \leq \frac{\varepsilon}{2}\). Overall, setting \(t = \frac{\varepsilon}{2}\) in inequality (58) we get

\[
\Pr(f \geq \varepsilon) \leq e^{-2}.
\] (62)

Therefore, with positive probability \(\|\frac{1}{m} \sum_{i=1}^{m} v_i - \mu\| \leq \varepsilon\), i.e., the sample mean vector \(\frac{1}{m} \sum_{i=1}^{m} v_i\) is \(\varepsilon\) close (in Euclidean distance) to \(\mu\). By the probabilistic method we get that there exists a vector \(\mu'\) that is \(\varepsilon\) close to \(\mu\) and can be expressed as a convex combination of at most \(m = \frac{16\gamma^2}{\varepsilon^2}\) vectors in \(X\).