A. Additional Details on Algorithms

A.1. Additional Background of \( \alpha \)-rank

Given a \( n \)-player game, where each player \( i \in [n] \) has a finite set \( S_i \) of pure strategies. Let \( S = \Pi_i S_i \) denote the set of joint strategies. For each tuple \( s = (s_1, \ldots, s_n) \in S \) of pure strategies, the game specifies a joint payoffs \( M(s) \) of players. The vector of expected payoffs is denoted \( \bar{M}(s) = (M^1(s), \ldots, M^n(s)) \in \mathbb{R}^n \). \( \alpha \)-rank computes rankings following four steps: 1) construct payoff matrix for each player \( M^i, i \in [n] \); 2) construct transition matrix by Equation (2); 3) compute the stationary distribution of \( C \), as \( \pi \); 4) return the ranking of strategies according to probabilities in \( \pi \). Below is the computation of transition matrix \( C \):

\[
C_{i,\sigma} = \begin{cases} 
\frac{1 - \exp(-\alpha (M^i(\sigma) - M^i(s)))}{\eta} & \text{if } M^i(\sigma) \neq M^i(s) \\
0 & \text{otherwise}
\end{cases}
\]  

(2)

where the coefficient \( \eta \) is defined as \( \eta = (\sum_{i=1}^n (|S_i| - 1))^{-1} \), and \( \alpha > 0, \eta \in \mathbb{N} \) are hyperparameters. Let \( C_{\sigma, \tau} = 0 \) for all \( \tau \) that differ from \( \sigma \) in more than a single player’s strategy. \( C_{\sigma, \sigma} = 1 - \sum_{\tau \neq \sigma} C_{\sigma, \tau} \) ensures that transition distributions are valid.

Our two-player meta-games setting is the single population case of traditional \( \alpha \)-rank that two players have a shared pure strategies space \( S \), and the joint strategies space is defined as \( S \times S \). The payoffs of joint strategies are saved as a payoff matrix \( \bar{M} \), where \( M_{ij}, M_{ji} \) represents the payoffs of strategy \( S_i \) and strategy \( S_j \) respectively. Thus we could construct the transition matrix \( C \) between strategies in \( S \) by Equation (1) and get the ranking of strategies in \( S \) eventually.

A.2. Supporting Algorithms

Algorithm 3 gives the details of RG-UCB (Rowland et al., 2019) algorithm as a supplement of Algorithm 2. RG-UCB is composed by a sampling scheme \( S \) and a stopping condition \( C(\delta) \). It adopts Uniform-exhaustive (UE) as sampling scheme \( S \). At each time, it uniformly randoms a pair from all pairs need to be estimated to make a simulation. For the stopping condition \( C(\delta) \), Hoeffding (UCB) is considered as confidence-bound for stopping the evaluation of \( M_{ij} \). With \( \delta \) as confidence level and \( K \) as interaction times of \( M_{ij} \), we can get \( M_{ij} \) bounded in \( [\bar{M}_{ij} - \epsilon, \bar{M}_{ij} + \epsilon] \), where \( \bar{M}_{ij} \) is empirical estimation and \( \epsilon \) is a very small quantity calculated by the Hoeffding inequality and \( \epsilon < \sqrt{\frac{4M_{max}^2 \log(2/\delta)}{K}} \).

Algorithm 4 gives the OptSpace algorithm (Keshavan & Oh, 2009; Keshavan et al., 2009; 2010) as a supplement to Algorithm 1 and 2. OptSpace reconstructs a low rank matrix from a small subset of entries. Given incomplete observations \( M^\Omega \), OptSpace aims to find \( \bar{M} \), such that \( \bar{M} = U \Sigma V \), and \( ||M^\Omega - \bar{M}||_F \) is minimized. It relies on singular value decomposition for an initial guess and then adopts local manifold optimization. Two important steps are Trimming and Rank-\( r \) projection. Trimming eliminates over-represented rows and columns in \( M^\Omega \), which are those containing more than \( 2|\Omega|/n \) observed entries. Let \( \tilde{M}^\Omega \) denote the trimmed matrix. Rank-\( r \) projection is then applied to find the initialization of \( U_0, V_0 \). The singular value decomposition of the trimmed matrix \( \tilde{M}^\Omega \) is defined as: \( \tilde{M}^\Omega = \sum_{i=1}^n \Sigma_i U_i V_i^T \), where \( \Sigma_1 \geq \Sigma_2 \ldots \geq \Sigma_n \) are singular values. Then the rank-\( r \) projection of \( \tilde{M}^\Omega \) is defined as: \( P_r(\tilde{M}^\Omega) = \frac{n}{n^2} \sum_{i=1}^r \Sigma_i U_i V_i^T \). Then we get the reconstructed matrix \( \bar{M} \) through gradient descent on the Grassman manifold, with initial condition \( (U_0, V_0) \). For more detailed descriptions, see (Keshavan & Oh, 2009; Keshavan et al., 2009; 2010).
We first give the necessary lemmas and theorems for our proof.

**Algorithm 3** ResponseGraphUCB($\delta, S, C(\delta)$)

1: Construct list $L$ of pairs of strategy profiles to compare;
2: Initialize tables $\tilde{M}, N$ to store empirical means and interaction counts while $L$ is not empty do;
3: while $L$ is not empty do
4: Select a strategy profile $s$ appearing in an edge in $L$ using sampling scheme $S$;
5: Simulate one interaction for $s$ and update $\tilde{M}, N$ accordingly;
6: Check whether any edges are resolved according to $C(\delta)$, remove them from $L$ if so return empirical table $\tilde{M}$.
7: end while

**Algorithm 4** OptSpace(Matrix completion of $M^{(i)}$)

Input: A chosen rank $r$, sampling operator $\Omega \in [n] \times [n]$
Output: The recovered matrix $\tilde{M}$
1: Trim $M^{(i)}$, and let $\tilde{M}^{(i)}$ be the output;
2: Compute the rank-$r$ projection of $M^{(i)}$, $\Pi_r(\tilde{M}^{(i)}) = U_0 \Sigma_0 V_0^T$;
3: Minimize $\hat{F}(U, V)$ through gradient descent, with initial condition $(U_0, V_0)$.
4: Return $\tilde{M} = U \Sigma V^T$

**B. Theories and Proofs**

**B.1. Details of definition and theorem for Proposition 1**

**Definition 1** (($\mu_0, \mu_1$)-Incoherence (Keshavan et al., 2009)). Let matrix $M \in \mathbb{R}^{n \times n}$ of rank $r$ and the singular value decomposition is $M = U \Sigma V^T$, $U, V \in \mathbb{R}^{n \times r}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{r \times r}$ is a diagonal matrix. In matrix $\Sigma$, $\Sigma_{min} = \Sigma_1 \leq \ldots \leq \Sigma_{r} = \Sigma_{max}$, and define $\kappa = (\Sigma_{max}/\Sigma_{min})$. If $M$ meet the following two conditions:

1. $\forall i \in [n]: \sum_{k=1}^{r} U^2_{ik} \leq \mu_0 r$, $\sum_{k=1}^{r} V^2_{ik} \leq \mu_0 r$
2. $\forall i, j \in [n]: |\sum_{k=1}^{r} U_{ik} (\frac{\Sigma_{kj}}{\Sigma_{jj}}) V_{jk}| \leq \mu_1 \sqrt{r}$

then $M$ is defined as ($\mu_0, \mu_1$)-incoherent.

This condition describes that one cannot expect to recover the payoff matrix if the meaningful payoffs are in the null space of the sampling operator. Let $\| \cdot \|_*$ denote the nuclear norm, which is a summation of all singular values. The following theorem supports the result in Proposition 1.

**Theorem 3.** (Keshavan et al., 2010) Assume $M \in \mathbb{R}^{n \times n}$ of rank $r$ that satisfies the incoherence conditions with $(\mu_0, \mu_1)$. Let $\mu = \max \{ \mu_0, \mu_1 \}$. Further, assume $\Sigma_{min} \leq \Sigma_1 \leq \ldots \leq \Sigma_r \leq \Sigma_{max}$ with $\Sigma_{min}, \Sigma_{max}$ bounded away from 0 and $\infty$. Then there exists a numerical constant $C$ such that, if

$$|\Omega| \geq Cnr\sqrt{\alpha} \left( \frac{\Sigma_{max}}{\Sigma_{min}} \right)^2 \max \left\{ \mu_0 \log n, \mu^2 r \sqrt{\alpha} \left( \frac{\Sigma_{max}}{\Sigma_{min}} \right)^4 \right\}$$

then the output of OptSpace $\tilde{M}$ converges, with high probability, to the matrix $M$.

The proof of Proposition 1 directly follows by applying Theorem 3 with $\alpha = 1$.

**B.2. Proof of Theorem 1**

We first give the necessary lemmas and theorems for our proof.

**Lemma 1.** (Rowland et al., 2019) Suppose there are $n$ strategies and all payoffs are bounded in the interval $[-M_{max}, M_{max}]$, and define $L(\alpha, M_{max}) = 2\alpha \exp(2\alpha M_{max})$, and $g(\alpha, n, p, M_{max}) = \eta \exp(2\alpha M_{max})$, where $\alpha, n, p$ are all hyperparameters in $\alpha$-rank. Let $\epsilon \in (0, 18 \times 2^{-n} \sum_{i=1}^{n-1} \binom{n}{i} n^i)$. If $\sup_{(i,j) \in [n] \times [n]} |\tilde{M}_{i,j} - M_{i,j}| \leq \frac{g(\alpha, n, p, M_{max})}{18L(\alpha, M_{max}) \sum_{i=1}^{n-1} \binom{n}{i} n^i}$, then we have $\max_{i \in [n]} |\tilde{\pi}(i) - \pi(i)| \leq \epsilon$.

**Theorem 4.** (Keshavan et al., 2009) Let $M \in \mathbb{R}^{n \times n}$ be a ($\mu_0, \mu_1$)-incoherent matrix of rank $r$ and the singular value decomposition is $M = U \Sigma V^T$, where $U, V \in \mathbb{R}^{n \times r}$ are orthogonal matrices and $\Sigma \in \mathbb{R}^{r \times r}$ is a diagonal matrix. In
matrix $\Sigma, \Sigma_{\min} = \Sigma_1 \leq \ldots \leq \Sigma_1 = \Sigma_{\max}$, and define $\kappa = (\Sigma_{\max}/\Sigma_{\min})$. Let $\tilde{M} = M + \mathbf{Z}$ be the observed matrix with noise $\mathbf{Z}$. Define $\Omega \subseteq [n] \times [n]$ is the sampling operator in which $m$ entries are randomly selected for observation from all $n^2$ entries. Therefore, the matrix with noise observed by the sampling operator $\Omega$ is $\tilde{M}^\Omega = M^\Omega + \mathbf{Z}^\Omega$. There exist constants $C, C'$ such that if the number of sampled entries satisfies

$$
|\Omega| \geq C\kappa^2 n \max(\mu_0 r \log(n), \mu_0^2 r^2 \kappa^4, \mu_1^2 r^2 \kappa^2)
$$

and get $\tilde{M}$ through performing matrix completion algorithm OptSpace (Keshavan et al., 2009) on $\tilde{M}^\Omega$ then we have

$$
\frac{1}{n} \| \tilde{M} - M \|_F \leq C' \kappa^2 n^2 \sqrt{\frac{2}{|\Omega|}} \| \mathbf{Z}^\Omega \|_2 + \| \mathbf{Z} \|_F
$$

with probability at least $1 - \frac{1}{m}$. The right hand side above is less than $\Sigma_{\min}$.

**Theorem 5.** (Keshavan et al., 2009) For any matrix $M \in \mathbb{R}^{n \times n}$ and any set $\Omega \subseteq [n] \times [n]$, \n
$$
\| M^\Omega \|_2 \leq \frac{2|\Omega|}{n} \max_{(i,j) \in \Omega} |M_{ij}|
$$

Now we are ready to provide the proof for Theorem 1.

**Proof of Theorem 1.** According to Theorem 4 and 5, we have

$$
\| \tilde{M} - \hat{M} \|_F \leq \| \tilde{M} - M \|_F + \| M - \hat{M} \|_F
$$

(3)

$$
\leq C' \kappa^2 n^2 \sqrt{\frac{2}{|\Omega|}} \| \mathbf{Z}^\Omega \|_2 + \| \mathbf{Z} \|_F
$$

(4)

$$
\leq C' \kappa^2 n^2 \sqrt{\frac{2}{|\Omega|}} \cdot \frac{2|\Omega|}{n} \max_{(i,j) \in \Omega} |Z_{ij}| + n \| \mathbf{Z} \|_{\max}
$$

(5)

$$
\leq (2C' \kappa^2 \sqrt{\tau} + 1)n \| \mathbf{Z} \|_{\max}.
$$

(6)

Recall that, $\tau = \frac{2\epsilon g(\alpha, n, p, M_{\max})}{18L(\alpha, M_{\max}) \sum_{i=1}^{n-1} \binom{n}{i} n (2C' \kappa^2 \sqrt{\tau} + 1)n}$, Thus we have

$$
\sup_{(i,j) \in [n] \times [n]} |\tilde{M}_{ij} - \hat{M}_{ij}| \leq \| \tilde{M} - \hat{M} \|_F \leq \frac{\epsilon g(\alpha, \eta, p, M_{\max})}{18L(\alpha, M_{\max}) \sum_{i=1}^{n-1} \binom{n}{i} n i n}. \quad (7)
$$

By applying Lemma 1, we have $\max_{i \in [n]} |\tilde{\pi}(i) - \hat{\pi}(i)| \leq \epsilon$. Thus the proof of Theorem 1 is completed.

**B.3. Proofs of Theorem 2**

Now we are ready to prove Theorem 2.

**Proof of Theorem 2.** Define $\mathbf{Z} = \hat{M} - M$. Let $\tau = \frac{\epsilon g(\alpha, n, p, M_{\max}) |\Omega|}{18L(\alpha, M_{\max}) \sum_{i=1}^{n-1} \binom{n}{i} n i (2C' \kappa^2 \sqrt{\tau} + 1)n}$. Denote $\tilde{M}_{ij} = \frac{1}{K} \sum_{k=1}^{K} \tilde{M}_{ij}^k$,
then we have:

\[
P(\|Z^\Omega\|_2 > \tau) \leq P\left( \frac{2|\Omega| \max_{(i,j) \in \Omega}|Z_{ij}|}{n} > \tau \right) \quad \text{by Theorem 5}
\]

\[
= P\left( \max_{(i,j) \in \Omega} |Z_{ij}| > \frac{\tau n}{2|\Omega|} \right)
\]

\[
= P\left( \exists (i, j) \in \Omega : |\hat{M}_{ij} - M_{ij}| > \frac{\tau n}{2|\Omega|} \right)
\]

\[
\leq \sum_{(i,j) \in \Omega} P(|\hat{M}_{ij} - M_{ij}| > \frac{\tau n}{2|\Omega|})
\]

\[
\leq \sum_{i,j \in \Omega} \frac{1}{mn^3} \quad \text{(since } K > \frac{8M_{max}^2 \log(2mn^3)m^2}{\tau^2n^2})
\]

\[
= \frac{1}{n^3}
\]

Here (8) holds because of union bound theorem (Shalev-Shwartz & Ben-David, 2014). (9) holds because of Hoeffding’s Inequality: let \(X_1, X_2, \ldots, X_n\) be i.i.d random variables bounded in \([a, b]\), then for any \(\epsilon > 0\), \(P\left( \left| \frac{1}{K} \sum_{i=1}^K X_i - \mathbb{E}(X_i) \right| > \epsilon \right) \leq 2e^{-2K\epsilon^2/(b-a)^2}\). So we get that with probability at least \(1 - \frac{1}{n^3}\),

\[
\|Z^\Omega\|_2 \leq \frac{\epsilon g(\alpha, \eta, p, M_{\max})|\Omega|}{18L(\alpha, M_{\max}) \sum_{i=1}^{n-1} (\frac{\eta}{i}) i^n C' \kappa^2 n^2 \sqrt{r}}
\]

Thus, combined with Theorem 4 and the union bound, the probability that the first inequality (in Theorem 4) is true is \(1 - 1/n^3\), the probability that the second inequality (above) is true is \(1 - 1/n^3\), we can get with probability at least \(1 - \frac{2}{n^3}\), that:

\[
\|\hat{M} - M\|_F \leq \frac{\epsilon g(\alpha, \eta, p, M_{\max})}{18L(\alpha, M_{\max}) \sum_{i=1}^{n-1} (\frac{\eta}{i}) i^n}
\]

Obviously, \(\sup_{(i,j) \in [n] \times [n]} |\hat{M}_{i,j} - M_{i,j}| \leq \|\hat{M} - M\|_F\). By applying Lemma 1, the proof of Theorem 2 is completed. \(\square\)

### C. Further Experiments

**Additional results** Figure 7 and 8 show the results with \(\alpha = 0.001\) and \(\delta \in \{0.01, 0.1, 0.2\}\) on Bern(100) and soccer meta-game, as a supplement for Figure 5. Similarly, Figure 9 and 10 show the results with \(\alpha = 0.01\) and \(\delta \in \{0.01, 0.1, 0.2\}\) on Bern(100) and soccer meta-game, as a supplement for Figure 5. The results show that, across different choices of \(\alpha\)-rank parameters, our algorithm can estimate \(\alpha\)-rank with much fewer sample of pairs.

Table 3 shows the statistics of real world games that is used in Figure 1. Table 4 shows results of twelve real world games with \(\alpha\)-conv metric, as a supplement of Table 2, which demonstrates that higher rank will lead to lower approximation error on payoff matrices and better convergence to \(\alpha\)-rank.
Table 3. Statistics of payoffs from real world games from (Czarnecki et al., 2020). $k$ denote the number of dominant singular values such that $\sum_{i}^{k} \Sigma_{i} / \sum_{i}^{n} \Sigma_{i} \geq 80\%$.

<table>
<thead>
<tr>
<th>Game</th>
<th># policies</th>
<th>rank</th>
<th>$k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10,3-Blotto</td>
<td>66</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>10,4-Blotto</td>
<td>286</td>
<td>40</td>
<td>14</td>
</tr>
<tr>
<td>10,5-Blotto</td>
<td>1001</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td>3-move parity game 2</td>
<td>160</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>5,3-Blotto</td>
<td>21</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>5,4-Blotto</td>
<td>56</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>5,5-Blotto</td>
<td>126</td>
<td>20</td>
<td>10</td>
</tr>
<tr>
<td>AlphaStar</td>
<td>888</td>
<td>888</td>
<td>238</td>
</tr>
<tr>
<td>Blotto</td>
<td>1001</td>
<td>50</td>
<td>16</td>
</tr>
<tr>
<td>Disc game</td>
<td>1000</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Elo game + noise=0.1</td>
<td>1000</td>
<td>1000</td>
<td>396</td>
</tr>
<tr>
<td>Elo game + noise=0.5</td>
<td>1000</td>
<td>1000</td>
<td>507</td>
</tr>
<tr>
<td>Elo game + noise=1.0</td>
<td>1000</td>
<td>1000</td>
<td>524</td>
</tr>
<tr>
<td>Elo game</td>
<td>1000</td>
<td>38</td>
<td>2</td>
</tr>
<tr>
<td>Kuhn-poker</td>
<td>64</td>
<td>64</td>
<td>24</td>
</tr>
<tr>
<td>Normal Bernoulli game</td>
<td>1000</td>
<td>1000</td>
<td>499</td>
</tr>
<tr>
<td>Rock-Paper-Scissors</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Random game of skill</td>
<td>1000</td>
<td>1000</td>
<td>515</td>
</tr>
<tr>
<td>Transitive game</td>
<td>1000</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Triangular game</td>
<td>1000</td>
<td>1000</td>
<td>137</td>
</tr>
<tr>
<td>connect_four</td>
<td>1470</td>
<td>1464</td>
<td>297</td>
</tr>
<tr>
<td>go(board_size=3,komi=6.5)</td>
<td>1933</td>
<td>1924</td>
<td>516</td>
</tr>
<tr>
<td>go(board_size=4,komi=6.5)</td>
<td>1679</td>
<td>1668</td>
<td>380</td>
</tr>
<tr>
<td>hex(board_size=3)</td>
<td>766</td>
<td>764</td>
<td>232</td>
</tr>
<tr>
<td>misere(game=tic_tac_toe())</td>
<td>926</td>
<td>926</td>
<td>295</td>
</tr>
<tr>
<td>quoridor(board_size=3)</td>
<td>1404</td>
<td>1306</td>
<td>244</td>
</tr>
<tr>
<td>quoridor(board_size=4)</td>
<td>1540</td>
<td>1464</td>
<td>343</td>
</tr>
<tr>
<td>tic_tac_toe</td>
<td>880</td>
<td>880</td>
<td>285</td>
</tr>
</tbody>
</table>

Figure 7. Bernoulli game with $n = 100, r = 10, \alpha = 0.001$ with noisy evaluations.
Figure 8. Soccer meta-game with $n = 200, r = 10, \alpha = 0.001$ with noisy evaluations.

Figure 9. Bernoulli game with $n = 100, r = 10, \alpha = 0.01$ with noisy evaluations.

Figure 10. Soccer meta-game with $n = 200, r = 10, \alpha = 0.01$ with noisy evaluations.
Table 4. Results on twelve real world games with noise free evaluations. (Left of plot) Recovery error on the payoff matrices. (Right of the plot) $\alpha$-conv error showing the convergence to $\alpha$-rank.