JMLR: Workshop and Conference Proceedings 39:173–188, 2014 ACML 2014

Sparse Binary Zero-Sum Games

David Auger DAVID.AUGER@PRISM.UVSQ.FR
AlCAAP, Laboratoire PRiSM, Bât. Descartes, Université de Versailles Saint-Quentin-en-Yvelines, 45 avenue des États-Unis, F-78035 Versailles Cedex, France

Jialin Liu JIALIN.LIU@INRIA.FR
TAO, Lri, UMR CNRS 8623, Univ. Paris-Sud, F-91405 Orsay, France

Sylvie Ruette SYLVIE.RUETTE@MATH.U-PSUD.FR
Laboratoire de Mathématiques, CNRS UMR 8628, Bât. 425, Univ. Paris-Sud, F-91405 Orsay, France

David L. Saint-Pierre DAVIDLSTP@GMAIL.COM
Montefiore Institute, Université de Liège, Belgium

Olivier Teytaud OLIVIER.TEYTAUD@INRIA.FR
TAO, Lri, UMR CNRS 8623, Univ. Paris-Sud, F-91405 Orsay, France and OASE Lab, National Univ. of Tainan and AILab, National Dong Hwa Univ., Hualien, Taiwan

Editor: Dinh Phung and Hang Li

Abstract

Solving zero-sum matrix games is polynomial, because it boils down to linear programming. The approximate solving is sublinear by randomized algorithms on machines with random access memory. Algorithms working separately and independently on columns and rows have been proposed, with the same performance; these versions are compliant with matrix games with stochastic reward. (Flory and Teytaud, 2011) has proposed a new version, empirically performing better on sparse problems, i.e. cases in which the Nash equilibrium has small support. In this paper, we propose a variant, similar to their work, also dedicated to sparse problems, with provably better bounds than existing methods. We then experiment the method on a card game.

Keywords: Sparsity, bandit algorithms, zero-sum matrix games.

1. Introduction

Solving Nash equilibria (NE) of matrix games is important by itself, e.g. for some economical models; as a building block for Monte-Carlo Tree Search algorithms (Flory and Teytaud, 2011) when simultaneous actions are involved; and for robust stochastic optimization (Lambert III et al., 2005). Bandits algorithms (Lai and Robbins, 1985; Grigoriadis and Khachiyan, 1995; Auer et al., 1995; Audibert and Bubeck, 2009) are tools for handling exploration vs exploitation dilemma that are in particular useful for finding NE (Grigoriadis and Khachiyan, 1995; Audibert and Bubeck, 2009), with applications to two-player games (Kocsis and Szepesvari, 2006; Flory and Teytaud, 2011). To the best of our knowledge, the only paper in which sparsity in NE is used for accelerating bandits is (Flory and Teytaud, 2011).

Section 2 presents the framework of sparse bandits and related algorithms. Section 3 mathematically analyzes sparse bandits. Section 4 presents experiments on a card game and Section 5 concludes.

2. Algorithms for sparse bandits

In this section we define the problem and our proposed approach. Section 2.1 presents below the notion of NE in matrix games. Section 2.2 defines a classical bandit algorithm framework that aims at solving such a problem and Section 2.3 introduces two proposed algorithms adapted for sparse problems.

Throughout the paper, $[[a, b]]$ denotes $\{a, a+1, a+2, \dots, b-1, b\}$ and $\text{lcm}(x_1, \dots, x_n)$ denotes the least common multiple of $x_1, \dots, x_n$, where $a, b, x_1, \dots, x_n$ are integers.

2.1. Matrix games and Nash equilibria

Consider a matrix M of size $K \times K$ with values in $\{0, 1\}$ (we choose a square matrix for short notations, but the extension is straightforward). Player 1, the row player, chooses an action $i \in [[1, K]]$ and player 2, the column player, chooses an action $j \in [[1, K]]$; both actions are chosen simultaneously. Then, player 1 gets reward $M_{i,j}$ and player 2 gets reward $1 - M_{i,j}$. The game therefore sums to 1 (we consider games summing to 1 for convenience, but 0-sum games are equivalent).

A NE is a pair (x^*, y^*) (both in $[0, 1]^K$ and summing to 1) such that if i is distributed according to the distribution x^* (i.e. $i = k$ with probability x^*_k) and if j is distributed according to the distribution y^* (i.e. $j = k$ with probability y^*_k) then none of the players can expect a better average reward by changing unilaterally its strategy, i.e.

$$\forall x, y, \ (x^*)^T M y \geq (x^*)^T M y^* \geq x^T M y^*.$$

There might be several NE, but the value of the game M, given by $v = x^* M y^*$, remains the same. Moreover, we define an ϵ-NE as a pair (x, y) such that

$$\inf_{y'} x^T M y' > (x^*)^T M y^* - \epsilon \quad \text{and} \quad \sup_{x'} (x')^T M y < (x^*)^T M y^* + \epsilon.$$

2.2. Bandit algorithms

A state-of-the-art bandit algorithm to approximate a NE (which includes matrix games) is EXP3 (Auer et al., 1995). Here we present the version used in (Audibert and Bubeck, 2009).

At iteration $t \in [[1, T]]$, our version of EXP3 proceeds as follows (this is for one of the players; the same is done, independently, for the other player):

- At iteration 1, S is initialized as a null vector.
- Action i is chosen with probability $p(i) = \alpha_t / K + (1 - \alpha_t) \times \exp(\eta_t S_i) / \sum_j \exp(\eta_t S_j)$ for some sequences $(\alpha)_{t \geq 1}$ and $(\eta_t)_{t \geq 1}$, e.g. in (Bubeck and Cesa-Bianchi, 2012) $\alpha_t = 0$ and $\eta_t = \log(K)/(tK)$ or in (Audibert and Bubeck, 2009) $\eta_t = \min(\frac{4}{5}\sqrt{\frac{\log(K)}{TK}}, \frac{1}{K})$ and $\alpha_t = K\eta_t$.

- Let r be the received reward.
- Update S_i: $S_i \leftarrow S_i + r/p(i)$ (and S_j for $j \neq i$ is not modified).

This algorithm, as well as its variants, converges to the NE as explained in (Audibert and Bubeck, 2009; Bubeck and Cesa-Bianchi, 2012) (see also (Grigoriadis and Khachiyan, 1995; Auer et al., 1995)).

2.3. Rounded bandits.

(Flory and Teytaud, 2011) proposes a generic way of adapting bandits to the sparse case. Basically, the bandit runs as usual, and then the solution is pruned. We here propose a variant of their algorithm, as explained in Alg. 1. Our definition of sparsity does not assume that the matrix M is sparse; rather, we assume that the two vectors of the NE x^* and y^* have support of moderate size k, i.e. $k << K$. This assumption certainly holds for many games.

Algorithm 1 RBANDIT, the rounded bandit algorithm. $\delta \in]0,1[$ is a parameter.

Input: A $K \times K$ matrix M, defined by a mapping $(i,j) \mapsto M_{i,j}$.
Sparsity assumption: We assume that the NE (x^*, y^*) is unique and sparse, i.e. x^* and y^* have support of size $\leq k$.
Define $p = 1/(4ck^k)$, where $c = \text{lcm}(1, 2, \dots \lfloor k^k/2 \rfloor)$.
Let q be such that on $K \times K$ matrices the bandit algorithm with q iterations provides a p-NE with probability $1 - \delta$.
Run the bandit algorithm for q iterations.
Let (x, y) be the resulting approximate NE.
Approximate by multiples of $1/c$ as follows:

$$\begin{aligned} x'_i &= \lfloor cx_i + \frac{1}{2} \rfloor / c \\ y'_i &= \lfloor cy_i + \frac{1}{2} \rfloor / c. \end{aligned}$$

Renormalize: $x'' = x' / \sum_i x'_i$; $y'' = y' / \sum_i y'_i$.
Output (x'', y'') as an approximate NE.

The version in (Flory and Teytaud, 2011) is based on truncations (components less than a given ϵ are removed), instead of roundings; such a version is presented in Alg. 2 (the exact solving after truncation was not in (Flory and Teytaud, 2011)). The drawback is that for an exact solving we need a second step of exact solving, but we will see that the resulting complexity is better than the complexity of the rounded bandit (Alg. 1) because in case of rounding we need a higher precision in the bandit part.

(Flory and Teytaud, 2011) gives no proof and does not provide any tool for choosing c. The aim of this paper is to show that our versions (Alg. 1 and 2, i.e. rounded/truncated bandits) find exact NE, faster than existing methods when the solution is sparse.

Algorithm 2 TBANDIT, the truncated bandit algorithm. $\delta \in]0,1[$ is a parameter.

Input: A $K \times K$ matrix M, defined by a mapping $(i,j) \mapsto M_{i,j}$.
Sparsity assumption: We assume that the Nash equilibrium (x^*, y^*) is unique and sparse, i.e. x^* and y^* have support of size $\leq k$.
Define $p = 1/(4ck^k)$, where $c = \lceil k^{k/2} \rceil$.
Let q be such that on $K \times K$ matrices the bandit algorithm with q iterations provides a p-NE with probability $1 - \delta$.
Run the bandit algorithm for q iterations.
Consider the game restricted to rows i such that $x_i > 1/(2c)$ and to columns j such that $y_j > 1/(2c)$.
Let (x', y') be the exact Nash equilibrium of this restricted game (computed in polynomial time by linear programming).
Output (x', y') as approximate Nash equilibrium of the complete game (complete with 0's for missing coordinates).

3. Analysis

This section is devoted to the mathematical analysis of sparse bandit algorithms. Section 3.1 introduces the useful notations. Section 3.2 shows some properties of supports of Nash equilibria. Section 3.3 gives some useful results on denominators of rational probabilities involved in Nash equilibria. Section 3.4 presents stability results (showing that in sparse bandits, good strategies are also close to the Nash equilibrium). Section 3.5 concludes by properties on sparse bandits algorithms.

3.1. Terminology

We use the classical terminology of game theory. We consider a Matrix Game M of size $K \times K$ as above. A pure strategy is an index $i \in [[1, K]]$. A mixed strategy is a vector of size K with non-negative coefficients summing to 1.

Let e_i be the vector $(0, 0, \ldots, 0, 1, 0 \ldots, 0)^T$ with a one only at the i^{th} position; by a slight abuse of notation we will use this notation independently of the dimension of the vector (i.e. e_i can be used both in $\mathbb{R}^{10}$ and $\mathbb{R}^{50}$).

Let Δ denote the set of probability vectors, that is, $\Delta = \{y : \sum_j y_j = 1 \text{ and } \forall j, y_j \geq 0\}$; this implicitly depends on the dimension of the vectors, we do not precise it since there will be no ambiguity. The *support* of a vector $y \in \Delta$ is the set of indices j such that $y_j > 0$. For short, $\sup_x$ (or $\sup_y$, $\inf_x$, $\inf_y$ equivalently) means supremum on $x \in \Delta$. The *value* of $y \in \Delta$ for M is $V(y) = \sup_x x^T M y$.

Recall that v denotes the value of the game M, that is, it satisfies

$$\forall x, y \in \Delta,\ x^T M y^* \leq v \leq (x^*)^T M y, \tag{1}$$

and $v = (x^*)^T M y^* = V(y^*)$ if (x^*, y^*) is a Nash equilibrium.

3.2. Supports of Nash equilibria

Here we consider general matrices A, M with real coefficients. The following lemma is well known (see (Gale and Tucker, 1951, Lemma 1 page 318)).

Lemma 1 (Farkas Lemma) *There exists $y \geq 0$ satisfying $Ay = b$ if and only if there is no x such that $A^T x \geq 0$ and $x^T b < 0$.*

The following lemma is adapted from (Dantzig and Thapa, 2003). We give the proof for the sake of completeness.

Lemma 2 *Let I be the set of column indices i such that for all optimal solutions x^* of the row player we have $(x^*)^T M e_i = v$. Then there exists an optimal solution y^* for the column player whose support is exactly I.*

Proof First, we show that it is sufficient to prove that for any $i \in I$ there is an optimal solution $y^i \in \Delta$ such that $y^i_i > 0$. Then one considers any strictly convex combination of the y^i, for instance $y^* = \frac{1}{|I|} \sum_{i \in I} y^i$. The vector $y^* \in \Delta$ has support including I (because $y^i_i > 0$). On the other hand, y^* has support included in I (because by construction any optimal solution has support included in I).

So, it is sufficient to show that for any $i \in I$ there is an optimal y^i such that $y^i_i > 0$. We now prove this. Without loss of generality fix $i = 1 \in I$. Let us suppose that no optimal solution y of the column player has a positive coordinate $y_1 > 0$. In other words, the system

$$\begin{cases} \sum_i y_i & = & 1 \\ My & \leq & v\mathbb{1} \\ y_1 & > & 0 \\ y & \geq & 0 \end{cases}$$

has no solution, where $\mathbb{1}$ is the vector with all coefficients equal to 1. Equivalently, this means that the following system has no solution

$$\begin{cases} \sum_i y_i & = & 1 \\ (M - v\mathbb{1}_{K\times K})y & \leq & 0 \\ y_1 & > & 0 \\ y & \geq & 0 \end{cases}$$

where $\mathbb{1}_{K\times K}$ is the $K \times K$ matrix with all coefficients equal to 1. Introducing the variable $z = \frac{y}{y_1}$ and a slack variable w of size $K \times 1$, this is equivalent to saying that the system

$$\begin{cases} (M - v\mathbb{1}_{K\times K})z + w & = & 0 \\ z_1 & = & 1 \\ z, w & \geq & 0 \end{cases}$$

has no solution. By Lemma 1, applied with the concatenation (z, w) as y, $b = (0, 0, \ldots, 0, 1)$ and

$$A = \begin{pmatrix} M - v\mathbb{1}_{K\times K} & Id_K \\ 1\ 0\ \cdots \quad 0 & 0 \end{pmatrix},$$

we deduce the existence of a vector x and a real number ϵ such that

$$\begin{cases} x^T M_{.1} - v\sum_i x_i + \epsilon & \geq & 0 \\ x^T(M - v\mathbb{1}) & \geq & 0 \\ x & \geq & 0 \\ \epsilon & < & 0 \end{cases}$$

where $M_{.1}$ denotes the first column of M.

By the first equation above, x is not zero. Thus we can normalize x to get a vector in Δ, and we infer the existence of an optimal strategy $x^* \in \Delta$ for the row player such that

$$(x^*)^T M_{.1} > v;$$

this implies that we cannot have $1 \in I$, a contradiction. ■

Corollary 3 *If M admits a unique Nash Equilibrium (x^*, y^*) with support $J \times I$ then:*

- *for all $i \notin I$ and $j \notin J$ we have*

$$(x^*)^T M e_i > v \text{ and } e_j^T M y^* < v; \tag{2}$$

- *the submatrix M' of M with rows and columns respectively in J and I has a unique Nash Equilibrium which is the projection of x^*, y^* on $J \times I$.*

Proof The first part is a consequence of Lemma 2. Indeed, if $(x^*)^T M e_i = v$ with $i \notin I$, then there exists an optimal solution y' whose support contains i, which contradicts the uniqueness of y^*. Thus $(x^*)^T M e_i > v$ by Eq. (1). The statement for $j \notin J$ is symmetric.

For the second part, the projection is clearly a Nash equilibrium. Then, suppose that there is another Nash equilibrium for M' and let (x', y') be the only $2K$ vector whose projection on $J \times I$ is equal to this other equilibrium (in other words add zero coordinates for $i \notin I$ and $j \notin J$).

Consider now $(1-t)y^* + ty'$ with $t > 0$ and a row index j; then if $j \in J$

$$e_j^T M((1-t)y^* + ty') = (1-t)v + tv = v$$

and if $j \notin J$, then by the first part of this corollary the left part is at most v for t small enough. Since we have a finite number of rows this implies that by choosing t small enough we obtain a vector $(1-t)y^* + ty'$ which is another optimal solution for the column player in M, which contradicts the uniqueness. ■

3.3. Denominators of Nash equilibria

Consider a matrix M, with coefficients in $\{0, 1\}$, of size $k_1 \times k_2$ with $k_1 \geq 2$ and $k_2 \geq 2$.

Lemma 4 *Assume that the Nash equilibrium (x^*, y^*) is unique and that $\forall i, j,\ x_i^* > 0$ and $y_j^* > 0$.*
a) Then $k_1 = k_2$ and x^* and y^* are rational vectors which can be written with common denominator at most $k^{k/2}$ with $k = k_1 = k_2$.
b) Moreover, for all $x, y \in \Delta$, $x^T M y^* = (x^*)^T M y = (x^*)^T M y^* = v$.

Proof The Nash equilibrium y^* verifies the following properties:

- the sum of the probabilities of all strategies for the "column" player is 1, i.e.

$$\sum_i y_i^* = 1. \tag{3}$$

- the expected reward for the "row" player playing strategy i against y^* is independent of i, i.e.

$$\forall i, \sum_j M_{i,j} y_j^* = \sum_j M_{1,j} y_j^*. \tag{4}$$

If there is another solution $y \neq y^*$ to Eqs. (3), (4), then $y' = y^* - \alpha(y - y^*)$ is another strategy for the column player. If α is small enough, then $y' \geq 0$ and $\sum y'_i = 1$; it is a correct mixed strategy, and its value is $x^* M y^* - \alpha x^* M(y - y^*)$ which is less than or equal to $x^* M y^*$ if α has the same sign as $x^* M(y - y^*)$. This contradicts the uniqueness of y^* as a Nash strategy for the column player.

As a consequence, Eqs. (3) and (4) are a characterization of the unique Nash equilibrium. Thus y^* can be computed by solving Eqs. (3) and (4); this is a linear system $Zy^* = (1, 0, 0, \ldots, 0)$ with one single solution and Z a matrix with k_1 rows and k_2 columns, and where Z has values in $\{-1, 0, 1\}$. The solution is unique, therefore $k_1 = k_2$ and Z is invertible.

Z^{-1} can be computed as

$$Z^{-1} = \frac{1}{\det(Z)} (\text{cofactor}(Z))^T \quad \text{where} \quad (\text{cofactor}(Z))_{ij} = (-1)^{i+j} \det\left((Z_{i'j'})_{i' \neq i, j' \neq j}\right).$$

The matrix Z has coefficients in $\{-1, 0, 1\}$; therefore, by Hadamard's maximum determinant problem: $|\det(Z)| \leq k^{k/2}$ ((Hadamard, 1893), see e.g. (Brenner and Cummings, 1972, p 626)). Moreover, the matrix cofactor(Z) has integer coefficients. This concludes the proof of the fact that y^* is rational with denominator $D = |\det(Z)|$ at most $k^{k/2}$, where $k = k_1 = k_2$. The same arguments using x^* instead of y^* show that x^* can also be written as a rational with the same denominator D.

To show b), notice that Eq. (4) can be rewritten as follows: $\forall i, (My^*)_i = (My^*)_1$. If $x \in \Delta$, then

$$x^T M y^* = \sum_i x_i (My^*)_i = \sum_i x_i (My^*)_1 = (My^*)_1 \text{ because } \sum_i x_i = 1.$$

Thus $x^T M y^*$ is independent of $x \in \Delta$, which implies that $x^T M y^* = (x^*)^T M y^* = v$. By symmetry, one also has: $\forall y \in \Delta$, $(x^*)^T M y = (x^*)^T M y^* = v$. ■

Please note that $k^{k/2}$ is known nearly optimal for matrices with coefficients in $\{0, 1\}$ (by Hadamard's work) for any k of the form 2^m. Also there are examples of matrices for which $V(y) = V(y^*) + \frac{1}{|\det(Z)|}\|y - y^*\|$ for y arbitrarily close to y^*.

3.4. Stability of Nash equilibria

In the general case of a zero-sum game, two mixed strategies can be far from each other, whenever both of them are very close, in terms of performance, to the performance of the

(assumed unique) Nash equilibrium. However, with a matrix M with values in $\{0,1\}$, this is not true anymore, as explained by the two lemmas below.

Lemma 5 *Let $k \geq 2$. Consider a $k \times k$ matrix M with elements in $\{0,1\}$ such that the Nash equilibrium (x^*, y^*) is unique and no pure strategy has a null weight. Then for all $y \in \Delta$ we have*

$$V(y) \geq V(y^*) + \frac{1}{k^{k/2}} \|y - y^*\|_\infty. \tag{5}$$

Proof By convexity of V, it is sufficient to prove that

$$\min_{u;\, \sum_i u_i = 0,\, \|u\|_\infty = 1} \lim_{t \to 0,\, t > 0} \frac{V(y^* + tu) - V(y^*)}{t} \geq 1/k^{k/2}.$$

This is equivalent to

$$\min_{u;\, \sum_i u_i = 0,\, \|u\|_\infty = 1} \max_i M_i.u \geq 1/k^{k/2}, \tag{6}$$

with M_i the i^{th} row of M as previously.

Let $\widetilde{u}$ be a vector in which the minimum is reached (it exists by compactness). Since $\|\widetilde{u}\|_\infty = 1$, there exists i_0 such that $|\widetilde{u}_{i_0}| = 1$. Let us assume without loss of generality, that $\widetilde{u}_{i_0} = 1$. The proof is the same if $\widetilde{u}_{i_0} = -1$. Thus, in Eq. (6), we can restrict to the vectors u such that $u_{i_0} = 1$, $\sum_i u_i = 0$ and $\forall i \in \{1, \ldots, k\}$, $-1 \leq u_i \leq 1$.

This is indeed a linear programming problem, as follows:

$$\min_{u \in \mathbb{R}^k,\, w \in \mathbb{R}} w$$

under constraints

$$\begin{aligned}
\forall i \in \{1, \ldots, k\}, \quad & -1 \leq u_i \leq 1, \\
\forall i \in \{1, \ldots, k\}, \quad & M_i.u \leq w, \\
& u_{i_0} = 1, \\
& \sum_{1 \leq i \leq k} u_i = 0.
\end{aligned}$$

It is known that when a linear problem in dimension $k+1$ has a non infinite optimum, then there is a solution (u, w) with $k+1$ linearly independent active constraints.

Let us pick such a solution $\overline{u}$. It is solution of $k+1$ linearly independent equations of the form either $u_i = 1$, or $u_i = -1$, or $M_i.u = w$, or $\sum_i u_i = 0$. Let us note the system

$$\begin{aligned}
\forall i \in P,\ u_i &= 1, \\
\forall i \in N,\ u_i &= -1, \\
\forall i \in H,\ M_i.u &= w, \\
\sum_{1 \leq i \leq k} u_i &= 0,
\end{aligned}$$

where P, N, H are the subsets of $\{1, \ldots, k\}$ where the corresponding constraints are active.

We can remove w by setting $M_j.u = M_i.u$ for some fixed $i \in H$ and all $j \in H \setminus \{i\}$. Then, $\overline{u}$ is solution of a system of k equations in dimension k, with coefficients in $\{-1, 0, 1\}$.

We use the same trick as in the proof of Lemma 4; $\overline{u}$ is solution of a system of k linear equations with all coefficients in $\{-1, 0, 1\}$. Therefore all coordinates of $\overline{u}$ are rational numbers with a common denominator $D \leq k^{k/2}$. Then $M_i.\widetilde{u} = M_i.\overline{u}$ has a denominator $D \leq k^{k/2}$ and is positive; therefore $M_i.\widetilde{u} \geq 1/k^{k/2}$. This proves the expected result.

■

Lemma 6 *Consider M a $K \times K$ matrix with coefficients in $\{0, 1\}$ and assume that the Nash equilibrium (x^*, y^*) is unique. Let J be the support of y^* and $k = \#J$. Then*

$$\forall j \notin J,\ (x^*)^T M e_j \geq v + \frac{1}{k^{\frac{k}{2}}}.$$

Proof According to Lemma 4, v and the coefficients of x^*, y^* are multiple of some constant $c \geq \frac{1}{k^{\frac{k}{2}}}$. Thus, for every j, $(x^*)^T M e_j$ is also a multiple of c. Fix $j \notin J$. By Corollary 3, $(x^*)^T M e_j > v$, which implies that

$$(x^*)^T M e_j \geq v + c.$$

■

Combining Lemmas 5 and 6 yields the following

Theorem 7 *Consider a matrix M of size $K \times K$ and with coefficients in $\{0, 1\}$. Assume that there is a unique Nash equilibrium (x^*, y^*). Let k be the size of the supports of x^*, y^*. Then*

$$\forall y \in \Delta, V(y) - V(y^*) \geq \frac{1}{2k^k} \|y - y^*\|_\infty. \tag{7}$$

Proof Define $c = \frac{1}{k^{\frac{k}{2}}}$. Let J be the support of y^*. For every $y \in \Delta$, one can write $y = ay' + by''$, with $a = \sum_{j \in J} y_i \in [0, 1]$, $a + b = 1$ and $y', y'' \in \Delta$ satisfying: $\forall j \notin J$, $y'_j = 0$ and $\forall j \in J$, $y''_j = 0$. For every index i, one has

$$y_i - y^*_i = \begin{cases} ay'_i - y^*_i = a(y'_i - y^*_i) - by^*_i & \text{if } i \in J \\ by''_i & \text{if } i \notin J \end{cases}$$

Thus

$$\|y - y^*\|_\infty = \max\{\|a(y' - y^*) - by^*\|_\infty, b\|y''\|_\infty\}. \tag{8}$$

Then, define $\delta = \|y - y^*\|_\infty$, $\delta_1 = \|y' - y^*\|_\infty$ and $\delta_2 = \|y''\|_\infty$. One has

$$V(y) \geq (x^*)^T M y = a(x^*)^T M y' + b \sum_{j \notin J} (x^*)^T M (y''_j e_j).$$

By Lemma 4(b), $(x^*)^T M y' = v$; and by Lemma 6, $(x^*)^T M e_j \geq v + c$ for all $j \notin J$. thus

$$V(y) \geq av + b(v + c) \sum_{j \notin J} y''_j,$$

that is,

$$V(y) - v \geq cb. \tag{9}$$

By Eq. (8), either $\delta = b\delta_2 \leq b$, or $\delta = \|a(y' - y^*) - by^*\|_\infty$. If $\delta \leq b$, the result is given by Eq. (9) because $c < 1$ and hence $c^2 \leq c$. From now on, assume that $\delta = \|a(y' - y^*) - by^*\|_\infty$. Then, $\delta \leq a\delta_1 + b\|y^*\|_\infty \leq a\delta_1 + b$. Equivalently,

$$a\delta_1 \geq \delta - b. \tag{10}$$

We split the end of the proof into two cases.

Case 1: $b \geq c\delta/2$. Then

$$\begin{aligned} V(y) - v &\geq cb \quad \text{by Eq. (9)} \\ &\geq c^2\delta/2 \quad \text{by assumption on } b \end{aligned}$$

which gives the expected result in case 1.

Case 2: $b < c\delta/2$. Since y' has the same support as y^*, there exists $x' \in \Delta$ with the same support as x^* such that $x' M y' - v \geq c\delta_1$ by Lemma 5. Moreover,

$$V(y) - v \geq x' M y - v = a(x' M y' - v) + b(x' M y'' - v).$$

Hence,

$$\begin{aligned} V(y) - v &\geq ac\delta_1 - vb \quad (\text{because } x' M y'' \geq 0) \\ &\geq c\delta(a\delta_1/\delta) - vb \\ &\geq c\delta\left(1 - \frac{b}{\delta}\right) - vb \quad (\text{using Eq. (10)}) \\ &\geq c\delta\left(1 - \frac{c}{2} - \frac{v}{2}\right) \quad (\text{using } b < c\delta/2) \\ &\geq c\delta(1 - c)/2 \quad (\text{using } v \leq 1) \end{aligned}$$

Since $1 - c \geq c$, we get the expected result in case 2. ■

3.5. Application to sparse bandit algorithms

Consider a matrix M as in Theorem 7. By Lemma 4, (x^*, y^*) can be written with a common denominator at most $k^{k/2}$. Define $C = \text{lcm}(1, 2, 3, \ldots, \lfloor k^{k/2} \rfloor)$. By the prime number theorem, it is known [1] that $C = O(\exp(k^{k/2}(1 + o(1))))$.

We discuss in parallel the truncated bandit algorithm (Alg. 2) and the rounded bandit algorithm (Alg. 1), as follows.

By construction of the algorithms, with probability $1 - \delta$, the bandit algorithm finds a u-Nash equilibrium (x, y), for

1. See details in http://mathworld.wolfram.com/LeastCommonMultiple.html.

- $u < 1/(4k^{k/2}k^k)$ for the TBANDIT algorithm.
- $u < 1/(4Ck^k)$ for the RBANDIT algorithm.

By Theorem 7, this implies that

- $\|x - x^*\|_\infty \leq 2uk^k < 1/(2k^{k/2})$ (idem for $\|y - y^*\|_\infty$) for TBANDIT;
- $\|x - x^*\|_\infty \leq 1/(2C)$ (idem for $\|y - y^*\|_\infty$) for RBANDIT.

Then:

- Truncated algorithm: all non-zero coordinates of x^* are at least $1/k^{k/2}$ and $|x_i^* - x_i| < 1/(2k^{k/2})$ with probability $\geq 1 - \delta$ (and the same for y^*, y); so with probability $1 - \delta$ the Nash equilibrium (x', y') of the reduced game is the solution (x^*, y^*) (after filling missing coordinates with 0).
- Rounded algorithm: the denominator of the coordinates of x^*, y^* is a divisor of C, so with probability $1 - \delta$,

$$\|Cx - Cx^*\|_\infty < 1/2, \ \|Cy - Cy^*\|_\infty < 1/2,$$

and Cx^* and Cy^* are integers. So $x^* = \lfloor x + \frac{1}{2} \rfloor / C$; RBANDIT finds the exact solution with probability $\geq 1 - \delta$.

For example, if using the Grigoriadis & Khachiyan algorithm (Grigoriadis and Khachiyan, 1995), or variants of EXP3 (Audibert and Bubeck, 2009), one can ensure precision u with fixed probability in time

- $O\left(K \log K \frac{1}{u}\right) = O\left(K(\log K)(k^{3k})\right)$ for the truncated version;
- $O\left(K \log K \frac{1}{u}\right) = O(K(\log K)k^{2k} \exp(2k^{k/2}(1 + o(1))))$ for the rounded version.

Then, we get, after rounding (rounded version) or after truncating and polynomial-time solving (truncated version), the exact solution y^* with fixed probability and time

- $O\left(K \log K \cdot k^{3k} + poly(k)\right)$ for the truncated algorithm (Alg. 2);
- $O\left(K \log K \cdot k^{2k} \exp(2k^{k/2}(1 + o(1)))\right)$ for the rounded algorithm (Alg. 1).

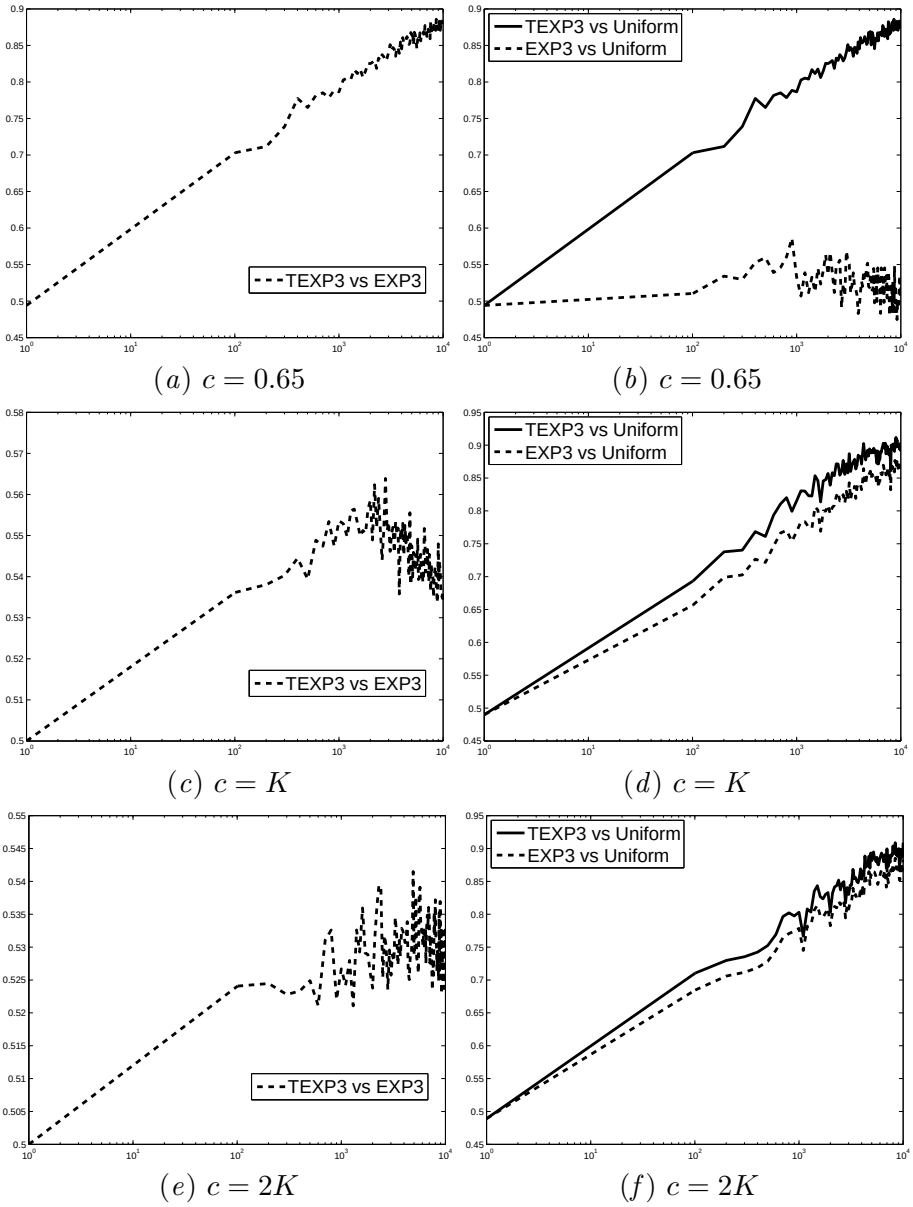
The truncated algorithm (Alg. 2) is therefore better.

4. Experiments

We work on the Pokemon card game. More precisely, we work on the *metagaming* part, i.e. the choice of the deck; the *ingaming* is then handled by a simulator with exact solving. The source code is freely available at `http://www.lri.fr/~teytaud/games.html`.

At first a normal EXP3 is executed using our empirically tuned formula

$$\begin{aligned} p(i) &= \left(1 + \frac{c-1}{\sqrt{t}}\right)^{-1} \\ &\times \left(1/(c \times \sqrt{t}) + (1 - 1/\sqrt{t}) \times \exp(S_i/\sqrt{t}) / \sum_j \exp(S_j/\sqrt{t})\right) \end{aligned} \tag{11}$$



(a) $c = 0.65$ (b) $c = 0.65$

(c) $c = K$ (d) $c = K$

(e) $c = 2K$ (f) $c = 2K$

Figure 1: Performance (%) in terms of budget T for the game of Pokemon using 2 cards. The left column shows TEXP3 playing against EXP3 for different values of c. The right column shows EXP3 and TEXP3 playing against the random uniform baseline. We tested a wide range of values for c and TEXP3 performs better than EXP3 regardless of c.

to compute the probability of arm i and we normalize the probabilities if needs be. After T iterations, the TEXP3 takes the decision whether an arm is part of the NE is based upon

a threshold ζ as explained in Alg. 3. Alg. 3 is based on Alg. 2, with constants adapted empirically and without the exact solving at the end.

Algorithm 3 TEXP3, an algorithm proposed in (Flory and Teytaud, 2011) used in these experiments.

Input: A $K \times K$ matrix M, defined by a mapping $(i,j) \mapsto M_{i,j}$. A number T of rounds.

Run EXP3 with Eq. (11), which provides an approximation (x,y) of the Nash equilibrium.

Define:

$$\begin{aligned} \zeta &= \max_{a \in K} \frac{(Tx_a)^{0.7}}{T} \\ x'_i &= x_i \text{ if } x_i \geq \zeta \text{ and } x'_i = 0 \text{ otherwise.} \\ x''_i &= x'_i / \sum_{j \in \{1,\dots,K\}} x'_j. \end{aligned}$$

Define:

$$\begin{aligned} \zeta' &= \max_{a \in K} \frac{(Ty_a)^{0.7}}{T} \\ y'_i &= y_i \text{ if } y_i \geq \zeta' \text{ and } y'_i = 0 \text{ otherwise.} \\ y''_i &= y'_i / \sum_{j \in \{1,\dots,K\}} y'_j. \end{aligned}$$

Output x'' and y'' as approximate Nash equilibrium of the complete game.

Figures 1(*a*), 1(*c*) and 1(*e*) show the performance of TEXP3 playing against EXP3 for different values of c and Figures 1(*b*), 1(*d*) and 1(*f*) present the performance of TEXP3 and EXP3 when playing against the random uniform baseline; the probability distribution obtained by EXP3 and TEXP3 after T iterations of EXP3 and (for TEXP3) after truncation and renormalization are used against random. To ensure that no player gains from being the first player, we make them play as both the row player and the column player and we display the result. Each point in the Figure consists in the means from 100 independent runs.

In all Figures, TEXP3 provides a consistent improvement over EXP3. Even in Figure 1(*b*), where EXP3 seems relatively weak against the random baseline, TEXP3 manages to maintain a performance similar to the ones in Figure 1(*d*) and 1(*f*).

5. Conclusion

(Grigoriadis and Khachiyan, 1995; Auer et al., 1995; Audibert and Bubeck, 2009) are great steps forward in zero-sum matrix games, and beyond. They provide algorithms solving $K \times K$ matrix games, with precision ϵ and for a fixed confidence, in time $O(K \log K/\epsilon^2)$.

As noticed in (Grigoriadis and Khachiyan, 1995), this has the surprising property that the complexity is sublinear in the size of the matrix, with a fixed risk.

We here show that, with coefficients in $\{0, 1\}$, if there is a unique sparse Nash equilibrium with support of size k for each player, then this bound can be reduced to $K \log K \cdot k^{3k}$, with no precision parameter (we provide an exact solution), with a fixed confidence $1 - \delta$:

- the dependency in K is the same as in (Grigoriadis and Khachiyan, 1995);
- there is no dependency in ϵ.

Practical relevance of this work

We want here to discuss the practical relevance of our results; two aspects are (i) the existence of very sparse problems, and (ii) the possible implementation of real world algorithms inspired by our results.

The first point is the existence of very sparse problems. We have seen that the sparsity level that we need, for our algorithm to outperform the state of the art for exact solutions, is $k << \log(K)/\log(\log(K))$. Obviously, one can design arbitrarily sparse zero-sum matrix games. Also, in real games, the quantity of moves worth being analyzed is usually much smaller than the number of legal moves; it is difficult to quantify this numerically for existing games as finding the meaningful strategies requires a lot of expertise and is difficult to quantify. For theoretical games, the classical centipede (Rosenthal, 1981) game does not have a unique Nash equilibrium; but if we remove strategies which make no sense (i.e. we do not consider the variants of a strategy after a move which finishes the game) the centipede game is highly sparse - only one strategy is in the Nash equilibrium, the one which immediately defects. The centipede game is not a zero-sum game, but zero-sum variants exist with the same sparsity property (Fey et al., 1996).

We also provide experimental results which show the relevance of this work in a real-world case. In particular, the TEXP3 modification, with its default parametrization from (Flory and Teytaud, 2011), performs better than the EXP3 counterpart for all tested values of the parameters. TEXP3 is not new, but the formal proof given here is new.

Extensions

The algorithm that we propose is based on the rounding of the solution given by EXP3 (Auer et al., 1995), or Grigoriadis' algorithm (Grigoriadis and Khachiyan, 1995), or INF (Audibert and Bubeck, 2009). Our algorithm works thanks to Theorem 7; any matrix game such that Theorem 7 is valid and such that Grigoriadis' algorithm, or EXP3 or INF works properly can be tackled similarly. In particular:

- Our bound in Theorem 7 (based on the constant in Lemma 4) can be adapted to the case of rational coefficients in M (instead of just 0 and 1);
- EXP3 and INF have no problem for stochastic cases as discussed below.

This leads to the following extensions:

- A first natural extension is the case of rational coefficients with a given denominator. The bound can be adapted to this case.

- Another extension is the case of stochastic versions of matrix games: i.e. if player 1 plays i and player 2 plays j, then the reward is a random variable $\widetilde{M}_{i,j}$, with expectation $M_{i,j}$, independently sampled at each trial. This does not change the result provided that $M_{i,j}$ verifies a condition as above, i.e. a common denominator bounded by some known integer.

Second, we point out that in the real world, using sparsity provides substantial benefit. For example, (Flory and Teytaud, 2011; Chou et al., 2012) get benefits on a real-world internet card game using sparsity; their algorithms are similar to ours (for finding approximate results), but without the final step for finding an exact solution. We reproduce here their experiments in yet another game (Section 4) and get improvements far better than the theoretical bound. This suggests that, beyond the guaranteed improvement, there is a large improvement in many practical cases.

Further work

K is usually huge; an algorithm linear in K might be not practical. Algorithms with complexity $O(\sqrt{K})$, $O(\log(K))$ might be useful; we know that such algorithms can not do any approximation of a Nash equilibrium without additional assumptions, and maybe some regularity assumptions on the strategies of the row player and on the strategies of the column player are required for this. Such a framework (K huge but regularity assumptions on row strategies and regularity assumptions on column strategies) looks like a relevant framework for applications.

A distinct further work is the case in which we have no upper bound on the sparsity parameter k.

Finally, two assumptions might be partially removed: uniqueness of the Nash-equilibrium, and 0-sum nature of the game. In particular, extending to the case of a unique subgame perfect equilibrium looks like a promising direction.

Acknowledgment

We are grateful to the BIRS seminar on Combinatorial Game Theory, to the Dagstuhl seminar on the Theory of Evolutionary Algorithms, and to the Bielefeld seminar on Search Methods, to National Science Council (Taiwan) for NSC grants 99-2923-E-024-003-MY3 and NSC 100-2811-E-024-001. We are grateful to C. Teytaud for implementing the Pokemon deck.

References

J.-Y. Audibert and S. Bubeck. Minimax policies for adversarial and stochastic bandits. In *22th annual conference on learning theory*, Montreal, Jun 2009.

Peter Auer, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire. Gambling in a rigged casino: the adversarial multi-armed bandit problem. In *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, pages 322–331. IEEE Computer Society Press, Los Alamitos, CA, 1995.

Joel Brenner and Larry Cummings. The Hadamard maximum determinant problem. *Amer. Math. Monthly*, 79:626–630, 1972. ISSN 0002-9890.

Sébastien Bubeck and Nicolò Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

Cheng-Wei Chou, Ping-Chiang Chou, Chang-Shing Lee, David Lupien Saint-Pierre, Olivier Teytaud, Mei-Hui Wang, Li-Wen Wu, and Shi-Jim Yen. Strategic choices: Small budgets and simple regret. In *Proceedings of TAAI 2012*, page 6 pages, 2012.

George Dantzig and Mukund Thapa. *Linear Programming 2: Theory and Extensions*. Springer, 2003.

Mark Fey, Richard D. McKelvey, and Thomas R. Palfrey. An experimental study of constant-sum centipede games. *International Journal of Game Theory*, 25(3):269–287, 1996.

Sébastien Flory and Olivier Teytaud. Upper confidence trees with short term partial information. In *Procedings of EvoGames 2011*, page accepted. Springer, 2011.

Kuhn Gale and Tucker. Linear programming and the theory of games. In Koopmans, editor, *Activity Analysis of Production and Allocation*, chapter XII. Wiley, 1951.

Michael D. Grigoriadis and Leonid G. Khachiyan. A sublinear-time randomized approximation algorithm for matrix games. *Operations Research Letters*, 18(2):53–58, Sep 1995.

Jacques Hadamard. Résolution d'une question relative aux déterminants. *Bull. Sci. Math.*, 17:240–246, 1893.

L Kocsis and Cs Szepesvari. Bandit based Monte-Carlo planning. In *15th European Conference on Machine Learning (ECML)*, pages 282–293, 2006.

T.L. Lai and H. Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6:4–22, 1985.

Theodore J. Lambert III, Marina A. Epelman, and Robert L. Smith. A fictitious play approach to large-scale optimization. *Oper. Res.*, 53(3):477–489, 2005. ISSN 0030-364X. doi: http://dx.doi.org/10.1287/opre.1040.0178.

R. Rosenthal. Games of perfect information, predatory pricing, and the chain store. *Journal of Economic Theory*, 25:92–100, 1981.