La complexité des jeux à récompense moyenne

Jérémy Thibault à partir d'un article de Uri Zwick et Michael S. Paterson

22 février 2016

Résumé

Les jeux à récompense moyenne sont une famille de jeux introduite par Ehrenfeucht et Mycielski. Cet article décrit une méthode pseudo-polynomiale pour obtenir des stratégies optimales, le problème de décision étant dans $\mathsf{NP} \cap \mathsf{coNP}$. Est également décrite une réduction polynomiale vers les jeux simples stochastiques étudiés par Condon, jeux qui sont aussi connus pour être dans $\mathsf{NP} \cap \mathsf{coNP}$, mais pour lesquels aucun algorithme pseudo-polynomial n'est connu.

Mots-clés : Complexité ; Jeux ; Stratégies optimales ; Jeux simples stochastiques ;

Introduction

Soit $G = (V, E)$ un graphe orienté fini dans lequel chaque sommet possède au moins une arête, et soit $\omega : E \to \{-W, ..., 0, ..., W\}$ la fonction de poids associée à ce graphe.

Un jeu à récompense moyenne (JRM) sur ce graphe peut être défini ainsi : le jeu commence au sommet $a_0 \in V$. Le premier joueur choisit une arête $e_1 = (a_0, a_1)$, puis le deuxième joueur choisit une arête $e_2 = (a_1, a_2)$, et ainsi de suite. Le premier joueur souhaite maximiner $\liminf\limits_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \omega(e_i)$ tandis que le second joueur souhaite minimiser $\limsup\limits_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \omega(e_i)$.

L'étude d'un tel jeu revient à étudier une moyenne à long terme, qui peut modéliser une fréquence de visite dans un état. Elle permet ainsi d'étudier la part du temps passée dans un certain état pour un protocole. Toutefois (mais c'est aussi un avantage), l'étude de cette valeur néglige la phase transitoire (les premiers états).

Ehrenfeucht et Mycielski ont montré que pour chacun de ces jeux, il existe une valeur ν telle que le premier joueur a une stratégie telle que $\liminf\limits_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \omega(e_i) \geq \nu$ et telle que le second a une stratégie telle que $\limsup\limits_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \omega(e_i) \leq \nu$. De plus, cette stratégie ne dépend que du sommet en question (c'est une stratégie positionelle).
L'existence de cette valeur montre une stabilité de ce genre de jeux : qu'importe qui fixe sa stratégie et comment, la valeur optimale est la même.
Il existe également une version finie du jeu, qui se termine quand un cycle se forme. La valeur ν existe également dans le cas fini, et les stratégies positionnelles optimales sont les mêmes.

Cet article complète le travail de Ehrenfeucht et Mycielski en donnant un algorithme de complexité $O(|V|^3 \cdot |E| \cdot |W|)$ en temps pour trouver toutes les valeurs de tous les jeux joués sur un tel graphe. On donne également un algorithme polynomial en la taille du graphe (et pseudo-polynomial en les poids des arêtes) pour trouver les stratégies positionnelles optimales de chacun des joueurs.

On prouve également que si un joueur connaît la stratégie positionnelle de l'autre, une contre-stratégie optimale peut être trouvée en temps polynomial, ce qui implique que le problème de décision est à la fois dans NP et coNP.

Finalement, il est décrit une réduction des JRM vers les jeux simples stochastiques (JSS), d'abord en réduisant les JRM vers les jeux à récompense escomptée, puis en réduisant ces jeux aux JSS. Cette réduction montre que les jeux simples stochastiques, dont le problème de décision associé est également dans $\mathsf{NP}\cap\mathsf{coNP}$, sont au moins aussi difficiles que les JRM. Leur complexité exacte reste un problème ouvert, mais on suppose qu'ils sont strictement plus difficiles.

1 Jeux à récompense moyenne

1.1 Calcul des valeurs d'un jeu

Soient $G = (V, E)$ un graphe et $\omega : E \to -W, ..., 0, ..., W$ une fonction de poids sur ses arêtes. Posons $|V| = n$. On peut considérer que le graphe est bipartite avec V_1 l'ensemble des sommets du joueur 1 et V_2 ceux du joueur 2, éventuellement en dédoublant tous les sommets. Pour tout $a \in V$ on souhaite calculer $\nu(a)$ la valeur du jeu (fini ou infini) commençant en a.

On construit une version intermédiaire du jeu qui s'arête au bout de k étapes. On appelle $\nu_k(a)$ la valeur du jeu commençant en $a \in V$ par le joueur 1 si $a \in V_1$ ou le joueur 2 si $a \in V_2$, cette valeur correspondant à la somme des poids. Cette version finie sera utilisée comme étape intermédiaire pour obtenir la valeur du jeu.

Théorème 1. *Les valeurs $\nu_k(a)$, pour tout $a \in V$ peuvent être calculées en un temps $O(k \cdot |E|)$. On a la relation de récurrence :* $\nu_k(a) = \begin{cases} \max_{(a,b)\in E}\{\omega(a,b) + \nu_{k-1}(b)\} & \text{si } a \in V_1 \\ \min_{(a,b)\in E}\{\omega(a,b) + \nu_{k-1}(b)\} & \text{si } a \in V_2 \end{cases}$ *avec la condition initiale* $v_0(a) = 0$

Il semble alors clair que $\lim_{k\to\infty} \frac{\nu_k(a)}{k} = \nu(a)$.

Théorème 2. *Pour tout $a \in V$, on a :* $k \cdot \nu(a) - 2nW \leq \nu_k(a) \leq k \cdot \nu(a) + 2nW$.

Il est maintenant possible de calculer les valeurs exactes pour ces jeux finis ou infinis.

Lemme 1. *Soit $G = (V, E)$ un graphe orienté et $\omega : E \to \{-W, ..., 0, ..., W\}$ une fonction de poids sur ses arêtes. Alors la valeur $\nu(a)$ pour tous les $a \in V$ est un rationnel, de dénominateur pair valant au plus n. De plus, la distance entre deux valeurs possibles de $\nu(a)$ vaut au moins* $\frac{2}{n(n-1)}$

ν_k fournit une valeur approchée de ν. ν étant un rationnel avec des coefficients bornés, on va choisir k suffisamment grand pour que l'intervalle ne contienne qu'un seul rationnel possible.

Théorème 3. *Soit $G = (V, E)$ un graphe orienté et $\omega : E \to \{-W, ..., 0, ..., W\}$ une fonction de poids sur ses arêtes. La valeur $\nu(a)$ pour tous les $a \in V$ peut être calculée en temps $O(|V|^3 \cdot |E| \cdot W)$.*

Démonstration. Pour $k = 2n^3W$, on peut calculer les valeurs $\nu_k(a)$ pour $a \in V$ en temps $O(|V|^3 \cdot |E| \cdot W)$ d'après le théorème 1. D'après le théorème 2, on peut encadrer $\nu(a)$ qui est un rationnel (cas fini). Son dénominateur est au plus n, et par une minoration de la distance entre deux valeurs possibles de $\nu(a)$ on a l'existence d'un unique rationnel $\nu(a)$ que l'on peut donc calculer. $\square$

Le problème de décision associé peut se résoudre plus efficacement, puisqu'on a besoin de moins de précision.

Théorème 4. *Soit G et ω semblable au théorème précédent, et soit T un seuil entier. Le problème de décision consistant à savoir si $\nu(a) < T$, $\nu(a) = T$ ou $\nu(a) > T$ pour tout $a \in V$ peut être résolu en temps $O(|V|^2 \cdot |E| \cdot W)$*

Ainsi, on vient de trouver une méthode de calcul des valeurs optimales ν des JRM. On peut maintenant s'intéresser à la recherche de stratégies optimales.

1.2 Recherche de stratégies optimales

On recherche des stratégies positionnelles optimales.

Théorème 5. *Soit $G = (V, E)$ un graphe orienté et $\omega : E \to -W, ..., 0, ..., W$ une fonction de poids sur ses arêtes. Les stratégies positionnelles optimales peuvent être trouvées en temps $O(|V|^4 \cdot |E| \cdot \log(\frac{|E|}{|V|}) \cdot W)$*

Démonstration. Décrivons l'algorithme utilisé.
Tout d'abord, on calcule les valeurs $\nu(a)$ pour tous les $a \in V$.
Si toutes les sommets de V_1 sont de degré sortant 1, alors le joueur 1 a une stratégie unique positionnelle et optimale.
Sinon, considérons un sommet $a \in V_1$ de degré sortant $d > 1$. Supprimons la moitié des arêtes sortantes de a, et recalculons la valeur de a, $\nu'(a)$. Si les valeurs sont identiques alors il existe une stratégie positionnelle optimale qui n'utilise pas les arêtes enlevées ; sinon, il en existe une qui en utilise une. On peut alors se restreindre au sous-graphe correspondant au cas qui nous intéresse et ainsi déterminer une stratégie positionnelle optimale pour ce joueur. $\square$

Il est également intéressant de se pencher sur le cas où l'on connaît la stratégie adverse. Celà correspond au cas où il y a une seule arête sortante de chaque sommet de V_2. Un algorithme de complexité $O(|V| \cdot |E|)$ proposé par Karp permettant de trouver un cycle de poids moyen minimum ou maximum d'un graphe fournit une approche efficace de ces problèmes particuliers.

Théorème 6. *Soit $G = (V, E)$ un graphe bipartite orienté ainsi qu'une fonction de poids réelle $\omega : E \to R$. Supposons que le degré sortant de chacun des sommets $v_2 \in V_2$ vaut 1. Alors les valeurs de tous les sommets et une stratégie positionnelle optimale pour le joueur 1 peut être trouvée en temps $O(|V| \cdot |E|)$.*

On a directement le corollaire suivant :

Théorème 7. *Le problème de décision correspondant aux JRM (étant donné un jeu G et un nombre ν, est-ce que la valeur de G est supérieure à ν ?) est dans* $\mathsf{NP} \cap \mathsf{coNP}$

Il suffit de déterminer une stratégie pour le joueur 1 (pour NP) ou pour le joueur 2 (pour coNP) et de les vérifier, en temps polynomial.

2 Réduction aux jeux simples stochastiques

On va montrer que les jeux à récompense moyenne se réduisent à une autre famille de jeux, les jeux simples stochastiques (JSS). Cette réduction montre que ces jeux sont au moins aussi difficiles que les jeux simples stochastiques, mais on suppose que les JRM sont strictement plus simple. Actuellement, aucun algorithme polynomial n'est connu pour ces JSS, mais un algorithme sous-exponentiel l'est.

2.1 Jeux à récompense escomptée

On décrit un nouveau type de jeux, les jeux à récompense escomptée (JRE).

Soit $0 < \lambda < 1$ un nombre réel appelé taux d'escompte. Cette fois ci, le poids de la i-ième arête est multiplié par $(1-\lambda)\lambda^i$. Le résultat du jeu devient $(1-\lambda)\sum_{i=0}^{+\infty}\lambda^i\omega(e_i)$, et les objectifs des joueurs sont toujours de maximiser ou minimiser cette valeur.

Soit $G = (V_1, V_2, E)$ un graphe orienté bipartite où $V_1 = \{u_1, ..., u_{n_1}\}$ et $V_2 = \{v_1, ..., v_{n_2}\}$. On note $a_{i,j}$ le poids de (u_i, v_j) (éventuellement $-\infty$ si l'arête n'est pas dans E) et de même $b_{j,i}$ le poids de (v_j, u_i) (éventuellement $+\infty$ si l'arête n'est pas dans E).
On note également $x_i = x_i(\lambda)$ la valeur d'un tel jeu qui commence en u_i et y_j celle d'un jeu qui commence en v_j.

Théorème 8. *Les vecteurs $(x_1, ..., x_{n_1})$ et $(y_1, ..., y_{n_1})$ sont les solutions des équations*
$x_i = \max_{1\le j\le n_2}\{(1-\lambda)a_{i,j} + \lambda y_j\}$ *pour* $1 \le i \le n_1$
$y_j = \max_{1\le i\le n_1}\{(1-\lambda)b_{j,i} + \lambda x_i\}$ *pour* $1 \le j \le n_2$

Si $\nu(\lambda)$ est la valeur d'un tel jeu, on s'attend à ce que cette valeur tende vers ν quand λ tend vers 1. C'est bien le cas.

On va effectuer une réduction des JRM vers les JRE.

Théorème 9. *Dans les mêmes conditions que précédemment, et avec les mêmes notations, on a :* $\nu - 2n(1-\lambda)W \le \nu(\lambda) \le \nu + 2n(1-\lambda)W$.

En particulier, si $\lambda = 1 - \frac{1}{2n^3W}$, on peut obtenir ν d'une manière similaire au théorème 3. On vient d'obtenir une réduction des JRM aux JRE.

2.2 Jeux simples stochastiques

On souhaite maintenant s'intéresser à un second type de jeux, les jeux simple stochastiques. On va montrer que les JRE se réduisent à ces JSS, et ainsi que les JRM se réduisent aux JSS.

Ces jeux se jouent sur un graphe $G = (V, E)$ tel que V est l'union disjointe de V_{max}, V_{min} et $V_{average}$ et qui contient également un sommet de départ et deux sommets spéciaux, le 0-puits et le 1-puits.

À chaque étape du jeu, on se déplace sur le graphe selon trois règles :

1. À un sommet max, le joueur 1 choisit l'arête empruntée.
2. À un sommet min, le joueur 2 choisit l'arête empruntée.
3. À un sommet moyen, l'arête empruntée est choisie au hasard selon une probabilité attachée à chaque arête.

Le jeu se termine quand on atteint un sommet puit et la valeur de ce jeu est la probabilité d'atteindre le 1-puits (c'est à dire que le joueur 1 gagne) si les deux joueurs jouent de manière optimale. Des stratégies positionelles optimales existent. Condon a montré que ces jeux sont dans $\mathsf{NP} \cap \mathsf{coNP}$.

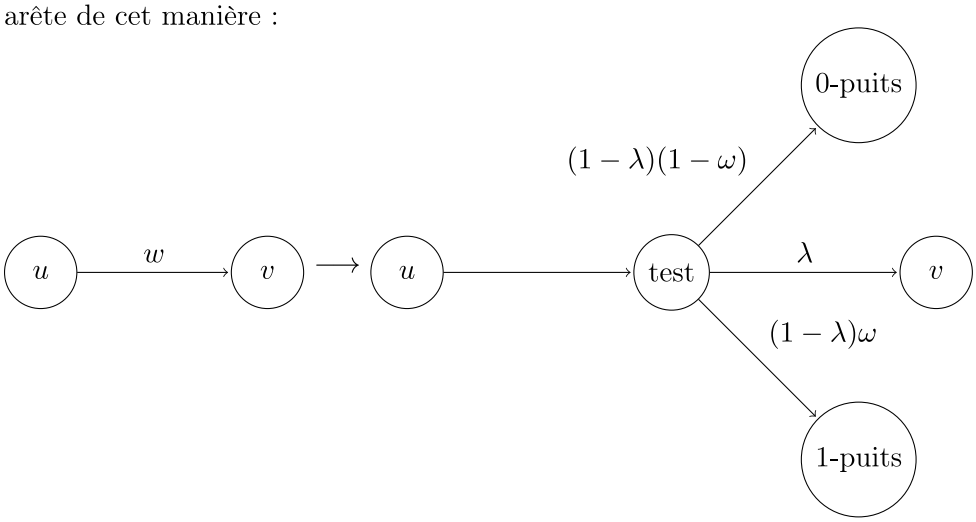
On s'intéresse en particulier à des jeux où chaque arête a une probabilité 0.5. On peut montrer que tout JSS peut être réduit à un jeu de ce type. Condon a montré le théorème suivant, semblable au théorème 8.

Théorème 10. *Soit $G = (V, E)$ un JSS qui termine avec une probabilité 1. Soit $p(u, v)$ la probabilité associée à l'arête (u, v). Les valeurs $\nu(u)$ des sommets G vérifient :*

$$\nu(u) = \begin{cases} \max_{(u,v)\in E}\{\nu(v)\} \text{ si } u \text{ est un sommet max} \\ \min_{(u,v)\in E}\{\nu(v)\} \text{ si } u \text{ est un sommet min} \\ \sum_{(u,v)\in E}\{p(u,v)\cdot\nu(v)\} \text{ si } u \text{ est un sommet moyen} \end{cases}$$

avec les conditions ν(0-puits) $= 0$ et ν(1-puits) $= 1$

Par une transformation linéaire on peut modifier les poids d'un JRE pour qu'ils soient dans l'intervalle $[0, 1]$ Il est possible de construire un JSS à partir d'un tel JRE, en transformant toute arête de cet manière :



On obtient alors un JSS qui possède les mêmes propriétés (et valeurs optimales) que le JRE original, d'où la réduction.

3 Conclusion

L'étude des jeux à récompenses moyennes effectuée par Ehrenfeucht et Mycielski, qui modélisent des phénomènes susceptibles de passer plusieurs fois dans les mêmes états, a permis d'exhiber l'existence de stratégies optimales positionelles pour chaque joueur, permettant d'atteindre une valeur optimale. Cet article exhibe ainsi des algorithmes permettant dans un premier temps de calculer ces valeurs optimales, puis d'en déduire des stratégies optimales. On en déduit que le problème est dans $\mathsf{NP} \cap \mathsf{coNP}$.

Ensuite, on a réduit ces jeux aux jeux simples stochastiques, qui restent un problème ouvert en terme de complexité. Ces jeux simples stochastiques sont donc au moins aussi durs que les jeux à récompense escomptée.