

Chapitre 7

Apprentissage des équilibres de Nash

Ce chapitre se consacre à des algorithmes d'apprentissage d'équilibres de Nash purs et à leur temps de convergence. La section 7.1 rappelle quelques résultats sur les dynamiques de meilleure réponse. Ensuite, la section 7.2 décrit un algorithme générique (en fait deux algorithmes) très général qui couvre ceux de la section 7.1. Elle présente des propriétés de convergence faible vers des équations différentielles. En paramétrant cet algorithme générique, on peut garantir que ces équations différentielles correspondent à des exemples classiques de dynamiques dans la théorie évolutionnaire des jeux. Cela est présenté dans la section 7.3. Nous discutons des liens qui existent entre le temps de convergence de l'algorithme et celui de l'équation différentielle ordinaire associée dans la section 7.4. Finalement, nous discutons de la dynamique de réplication sur les jeux de potentiel ordinal et exacts et nous discutons de leur temps de convergence dans la dernière section.

Les résultats discutés dans ce chapitre sont le fruit d'une collaboration avec Olivier Bournez, et correspondent essentiellement à [39] et [40].

Dans de nombreuses situations, certains participants peuvent tirer quelques avantages de l'utilisation de certaines ressources communes selon leur propre intérêt au détriment de l'intérêt collectif. Par exemple, les applications pair-à-pair avec des phénomènes tels que free-riding, fournissent un exemple des comportements problématiques dans des systèmes répartis à grande échelle. Le routage inter-domaine dans le réseau Internet, où certains intérêts économiques peuvent fournir des incitations pour altérer les performances globales, est un autre exemple.

Dans tous ces contextes, les participants adaptent souvent leur stratégie en fonction de leur connaissance locale du système par de petits ajustements afin d'améliorer leur propre profit. L'impact de chaque joueur sur le réseau est faible. Toutefois, comme le nombre d'acteurs est grand, une évolution globale du système peut se produire.

Ici, nous nous intéressons à comprendre sous quelles conditions le système peut converger vers des situations rationnelles, c'est-à-dire, vers des équilibres de Nash. Nous nous focalisons sur des dynamiques qui réalisent de petits ajustements stochastiques. Nous supposons ces ajustements complètement répartis puisque tous les joueurs participant aux jeux ont souvent une vue locale du système, sans vision globale du jeu.

Autrement dit, nous regardons le système évolué. A chaque instant, les joueurs choisissent les actions qu'ils vont réaliser. En fonction de l'état global du système, ils reçoivent une récompense, que l'on peut voir comme un coût. En fonction de leur passé, ils changent alors leur stratégie. Nous nous intéressons à l'évolution globale du système, et à comprendre sous quelles conditions on peut garantir que cette évolution converge vers des

équilibres de Nash.

7.1 Dynamique de meilleure réponse

Considérons tout d'abord les jeux de potentiel ordinal définis dans le chapitre 5. Rappelons que ces jeux possèdent toujours au moins un équilibre de Nash pur. La preuve de [Rosenthal, 1973] (qui correspond aussi au théorème 18 et à la discussion qui suit) peut se voir comme un algorithme d'apprentissage d'équilibres. En effet, elle se base sur le fait que si un joueur améliore son profit, alors la fonction de potentiel globale décroît aussi. La suite de profils correspondant à une suite de meilleures réponses individuelles correspond alors à une suite décroissante de la fonction de potentiel. Comme cette suite a un nombre fini de valeurs, une telle suite de profils doit être finie et donc doit atteindre un équilibre de Nash pur [Rosenthal, 1973]. Cette preuve de l'existence d'équilibres de Nash purs peut se transformer en la preuve que la dynamique de meilleure réponse converge vers un équilibre de Nash pur : si les joueurs jouent chacun leur tour et changent de ressources en améliorant systématiquement leur profit quand ils le peuvent, alors on aura convergence vers un équilibre de Nash pur.

Plus formellement :

Algorithme 0 (générique) de la dynamique de meilleure réponse :

1. A l'initialisation,
 - chaque joueur i a une stratégie pure.
2. A chaque étape t ,
 - Un joueur i est sélectionné en fonction d'une règle \mathcal{R}
 - Le joueur i choisit la (ou une s'il y en a plusieurs) stratégie qui minimise son coût (choix de la meilleure réponse).

La règle \mathcal{R} peut correspondre à

- la règle FIFO : “chaque joueur joue à son tour” ;
- la règle ALÉATOIRE : “le joueur est sélectionné aléatoirement uniformément” ;
- la règle MAX : “le joueur sélectionné est celui qui améliora le plus son coût” ;
- la règle MIN : “le joueur sélectionné est celui qui améliora le moins son coût” .

Plusieurs travaux sur les jeux de potentiel ordinal (comme les travaux [Goldberg, 2004], [Even-Dar et al., 2007], [Immorlica et al., 2009] pour les plus récents) ont calculé le temps de convergence selon la règle \mathcal{R} . Le calcul du temps de convergence se base alors sur la fonction de potentiel, en minorant à chaque mouvement la décroissance de la fonction de potentiel :

Théorème 29 ([Even-Dar et al., 2007]) *Pour un jeu d'ordonnement avec des machines identiques, l'algorithme atteint un ϵ -équilibre de Nash pur en au plus $O(W^2/\epsilon)$ étapes quelle que soit la politique \mathcal{R} de choix des joueurs, où W la somme des poids de toutes tâches.*

Ces bornes peuvent être améliorées selon la politique \mathcal{R} .

Théorème 30 ([Even-Dar et al., 2007]) *Un équilibre de Nash pur est atteint*

- en au plus $n(n+1)/2$ étapes avec la règle FIFO
 - en au plus n étapes avec la règle MAX
 - en au plus $(\frac{n}{K})^K/(2(K!))$ étapes avec la règle MIN
- avec n joueurs et $K = m - 1$.

Contrairement à ce qui se passe pour les jeux d'ordonnancement, les jeux de congestion, qui sont aussi des jeux de potentiel exact, possèdent des équilibres de Nash purs difficiles à calculer : le problème est PLS-complet [Fabrikant et al., 2004].

De ce fait, il est naturel de chercher à se restreindre à des instances avec de bonnes propriétés : [Chien and Sinclair, 2007] propose la condition du α -saut :

Définition 34 Une instance du jeu de congestion respecte la condition du α -saut, avec $\alpha \geq 1$, si la fonction de coût C_e sur chaque ressource e satisfait $C_e(k+1) \leq \alpha \times C_e(k)$ pour tout $k \geq 1$.

Pour les jeux symétriques de congestion, (lorsque les joueurs ont même ensemble de stratégies), calculer un équilibre de Nash est PLS-complet [Fabrikant et al., 2004] et même si les jeux respectent la condition du 2-saut [Chien and Sinclair, 2007]. De ces résultats, la dynamique de meilleure réponse peut converger vers l'équilibre en un temps exponentiel. C'est pour cette raison que la dynamique de meilleure réponse est légèrement modifiée dans [Chien and Sinclair, 2007] en considérant que les changements d'un joueur ne se font plus en allant vers la meilleure réponse mais vers une ϵ -amélioration : voir [Chien and Sinclair, 2007] pour la définition formelle.

On a alors le résultat suivant :

Théorème 31 ([Chien and Sinclair, 2007]) Dans les jeux symétriques de congestion avec n joueurs satisfaisant la condition du α -saut, la dynamique de ϵ -amélioration évoquée ci-dessus converge vers un équilibre en $[n\alpha\epsilon^{-1}\log(nC)]$ où C est une borne supérieure des coûts de tous les joueurs.

Si tous les joueurs choisissent leurs stratégies en parallèle, plutôt que chacun leur tour, on peut considérer l'algorithme suivant, qui à notre connaissance, a été étudié uniquement pour les jeux d'ordonnancement :

1. A l'initialisation, chaque joueur i a une stratégie pure.
2. A chaque étape t , pour chaque joueur i , faire en parallèle,
 - Soit r_i la ressource courante de la tâche i .
 - Choisir une autre ressource de façon uniforme.
 - Si $C_{r_i}(t) \geq C_j(t) + (1 + \epsilon)w_i$ alors déplacer la tâche i vers la ressource j avec la probabilité $\epsilon(1 - \frac{C_j(t)}{C_{r_i}(t)})$.

Ce algorithme converge vers un équilibre en $O(\log \log m + n^4)$ étapes pour les tâches unitaires [Berenbrink et al., 2006], et en $O(mn\Delta^3(\epsilon^{-2}))$ étapes pour les tâches ayant des longueurs arbitraires [Berenbrink and Schulte, 2007] avec m machines et n tâches en moyenne. Cette preuve de convergence utilise les techniques de martingale reprises à la fin du chapitre.

Maintenant, nous allons définir des classes d'algorithmes d'apprentissages qui permettront de retomber de façon générique sur les dynamiques usuelles de la théorie des jeux évolutionnaires comme par exemple celle de réplication, celle de *logit*...

7.2 Classes d'algorithmes d'apprentissage.

Dans cette section, nous considérons des algorithmes génériques d'apprentissage, c'est-à-dire si l'on préfère, des familles d'algorithmes d'apprentissage.

Nous allons raisonner sur des familles d'algorithmes plutôt que sur un algorithme très précis, car il se trouve que l'étude de tous les algorithmes de la famille peut se faire uniformément, comme nous allons le voir.

Plus précisément, nous allons nous focaliser sur les familles qui correspondent à des algorithmes que l'on obtient en instanciant l'algorithme 1 ou l'algorithme 2 dans ce qui suit, pour une certaine fonction F , un paramètre b , et éventuellement une fonction σ .

Dans ceux-ci :

- le paramètre b vise à être un réel positif proche de 0 ;
- la fonction F est donnée par ses composantes, c'est-à-dire est donnée par les fonctions $F_i^b(r_i(t), s_i(t), q_i(t))$, que l'on veut aussi générique que possible : on impose seulement que le vecteur $q_i(t)$ reste toujours un vecteur de probabilité : c'est-à-dire, que $q_{i,\ell}(t) \in [0, 1]$ et $\sum_{\ell} q_{i,\ell}(t) = 1$ soient conservés.
- La fonction σ est le plus souvent l'identité, mais elle peut en différer pour obtenir certains comportements : dans le cas général, la fonction σ est donnée par ses composantes, les fonctions σ_i , que l'on impose seulement de satisfaire à $\sum_{\ell \in \mathcal{S}_i} \sigma_{i,\ell}(q_i) = 1$;

Les deux familles ont pour point commun qu'à chaque étape, tous les joueurs choisissent simultanément leur action en fonction de leur stratégie mixte courante et tel que plusieurs (voir algorithme 1) ou un joueur (voir algorithme 2) réactualise(nt) à chaque étape sa stratégie mixte.

Algorithme 1 : version synchrone

1. A l'initialisation (à l'étape 0),
 $q_i(0) \in \mathcal{K}_i$ est un vecteur arbitraire de probabilité pour tout i .
2. Pour chaque étape t ,
 - Chaque joueur i sélectionne une stratégie $s_i(t) \in \mathcal{S}_i$ selon la distribution de probabilité $\sigma_i(q_i(t))$.
 - Ces différents choix permettent d'obtenir un coût $r_i(t)$ pour chaque joueur i .
 - Chaque joueur i réactualise $q_i(t)$ de la façon suivante :

$$q_i(t+1) = q_i(t) + bF_i^b(r_i(t), s_i(t), q_i(t)). \quad (7.1)$$

Algorithme 2 : version asynchrone

1. A l'initialisation (à l'étape 0),
 $q_i(0) \in \mathcal{K}_i$ est un vecteur arbitraire de probabilité pour tout i .
2. Pour chaque étape t ,
 - Chaque joueur i sélectionne une stratégie $s_i(t) \in \mathcal{S}_i$ selon la distribution de probabilité $\sigma_i(q_i(t))$.
 - Ces différents choix permettent d'obtenir un coût $r_i(t)$ pour chaque joueur i .
 - Un joueur $i(t)$ est sélectionné de façon aléatoire :
 le joueur $i = i(t)$ est sélectionné avec la probabilité p_i ($\sum_{i=1}^n p_i = 1$).
 - Chaque joueur i réactualise son vecteur $q_i(t)$ selon la règle suivante :

$$q_i(t+1) = q_i(t) + \begin{cases} bF_i^b(r_i(t), s_i(t), q_i(t)) & \text{si } i = i(t) \\ 0 & \text{sinon} \end{cases}$$

Remarque : Ces deux algorithmes sont complètement répartis : à chaque étape, le joueur i a seulement besoin de connaître son coût r_i et sa stratégie mixte courante q_i pour réactualiser sa propre stratégie mixte courante q_i .

La dynamique de chaque instance de ces algorithmes est donnée par l'évolution du profil du jeu $Q(t) = (q_1(t), \dots, q_n(t)) \in \mathcal{K}$, à chaque étape t . Ce qui nous intéresse est le comportement asymptotique de $Q(t)$ et sa convergence possible vers des équilibres de Nash.

Nous notons $\mathbb{E}[Y | Q]$, l'espérance de la variable aléatoire Y lorsque les joueurs jouent selon le profil de stratégies mixtes donné par Q : c'est-à-dire lorsque le joueur j choisit l'action $\ell \in \mathcal{S}_j$ avec la probabilité $q_{j,\ell}$.

Plus généralement, nous noterons $\mathbb{E}_\sigma[Y | Q]$ l'espérance de la variable aléatoire Y lorsque les joueurs jouent selon le profil de stratégies mixtes donné par $\sigma(Q)$: c'est-à-dire lorsque le joueur j choisit l'action $\ell \in \mathcal{S}_j$ avec la probabilité $\sigma_{j,\ell}(q_{j,\ell})$.

Nous faisons l'hypothèse que la moyenne $\mathbb{E}_\sigma[F_i^b(r_i(t), s_i(t), q_i(t)) | Q(t)]$ est toujours définie, que $G_i(Q) = \lim_{b \rightarrow 0} \mathbb{E}_\sigma[F_i^b(r_i(t), s_i(t), q_i(t)) | Q]$ existe et que $G_i(Q)$ est une fonction continue de Q .

Quelque soit l'algorithme utilisé, il est facile de constater que $Q(t)$ est une chaîne de Markov homogène.

Définissons $\Delta Q(t)$ comme $\Delta Q(t) = Q(t+1) - Q(t)$, et $\Delta q_i(t)$ comme $q_i(t+1) - q_i(t)$.

Nous pouvons écrire

$$\mathbb{E}[\Delta q_i(t) | Q(t)] = b\tilde{p}_i \mathbb{E}_\sigma[F_i^b(r_i(t), s_i(t), q_i(t)) | Q(t)], \quad (7.2)$$

où \tilde{p}_i est 1 pour une instance de l'algorithme 1, et p_i pour une instance de l'algorithme 2.

Posons $G_i(Q) = \lim_{b \rightarrow 0} \tilde{p}_i \mathbb{E}_\sigma[F_i^b(r_i(t), s_i(t), q_i(t)) | Q(t)]$.

De façon informelle, $Q(t)$ converge vers la solution de l'équation différentielle ordinaire suivante :

$$\begin{cases} \frac{dX}{dt} & = G(X) \\ X(0) & = Q(0) \end{cases} \quad (7.3)$$

On peut s'y attendre, en raison du raisonnement informel suivant : supposons qu'on puisse remplacer $\mathbb{E}[\Delta q_i(t) | Q(t)]$ par $\Delta q_i(t)$ dans l'équation (7.2). Avec le changement

de variable $t \leftarrow tb$, on obtient

$$q_i(t+b) - q_i(t) = b\tilde{p}_i \mathbb{E}_\sigma [F_i^b(r_i(t), s_i(t), q_i(t)) | Q(t)].$$

En approximant $q_i(t+b) - q_i(t)$ par $b \frac{dq_i}{dt}(t)$ pour b petit, et le membre droit par $bG_i(Q)$, on s'attend à ce que le système se comporte comme l'équation différentielle (7.3) lorsque b est proche de 0.

Pour le prouver formellement, il faut introduire l'interpolation affine par morceaux (en anglais *piecewise-linear*) de $Q(t)$ notée $Q^b(\cdot)$ définie par

$$Q^b(t) = Q(\lfloor t/b \rfloor) + (t/b - \lfloor t/b \rfloor)(Q(\lfloor t/b \rfloor + 1) - Q(\lfloor t/b \rfloor)). \quad (7.4)$$

La fonction $Q^b(\cdot)$ appartient à l'espace de toutes les fonctions de \mathbb{R} dans \mathcal{K} qui sont continues droite et pourvues de limites gauche (fonctions *cad-lag*).

On peut alors considérer la suite $\{Q^b(\cdot) : b > 0\}$ et s'intéresser à la limite de cette suite notée $Q(\cdot)$ quand $b \rightarrow 0$.

Rappelons qu'une famille $(Y_t)_{t \in \mathbb{R}}$ de variables aléatoires converge *faiblement* vers une variable aléatoire Y , si $\mathbb{E}[h(X_t)]$ converge vers $\mathbb{E}[h(Y)]$ pour chaque fonction h bornée et continue.

On a alors le résultat formel suivant :

Théorème 32 *La suite des processus d'interpolation $\{Q^b(\cdot)\}$ converge faiblement, vers $Q(\cdot)$ quand $b \rightarrow 0$, vers la solution (unique et déterministe) de l'équation différentielle ordinaire (7.3).*

Ce théorème se prouve en utilisant des méthodes de preuve de convergence faible : formellement le théorème précédent se déduit de la preuve de [Stroock and Varadhan, 1979, Théorème 11.2.3].

En fait, nous voyons maintenant que le paramètre b correspond au pas de discrétisation de la dynamique continue donnée par (7.3).

Le résultat précédent est valable quelle que soit l'instance des familles d'algorithmes 1 et 2. Nous allons instancier dans la section 7.3 ces classes d'algorithmes pour obtenir différentes dynamiques continues, connues pour avoir leurs équilibres reliés aux équilibres de Nash des jeux.

7.3 Exemples de dynamiques

Nous allons instancier ces deux algorithmes sur trois types de dynamiques : des dynamiques de réplication de multi-population, *logit* et de la meilleure réponse seuillée.

7.3.1 Dynamique de réplication

Considérons le premier exemple en posant que

- les fonctions σ_i sont la fonction identité ;
- les fonctions $F_i^b(r_i(t), s_i(t), q_i(t))$ sont définis de la façon suivante :

$$F_i^b(r_i(t), s_i(t), q_i(t)) = \gamma(r_i(t))(e_{s_i(t)} - q_i(t)) \quad (7.5)$$

où $\gamma : \mathbb{R} \rightarrow [0, 1]$ est une fonction affine décroissante. En supposant que tous les coûts sont positifs, et par linéarité de la fonction d'espérance, nous pouvons noter que tous les coûts sont bornés par une constante M , et nous pouvons prendre $\gamma(x) = \frac{M-x}{M}$ par exemple.

Rappelons que $e_{s_i(t)}$ correspond au vecteur unitaire avec l'élément $s_i(t)$ égal à 1.

Maintenant, nous allons définir vers quelles équations différentielles ordinaires, les deux algorithmes convergent alors faiblement.

En supposant qu'à l'étape t , le joueur i choisit la stratégie $s_i(t)$ et s'il réactualise sa stratégie courante :

$$\Delta q_{i,\ell}(t) = q_{i,\ell}(t+1) - q_{i,\ell}(t) = \begin{cases} -b\gamma(r_i(t))q_{i,\ell}(t) & \text{si } s_i(t) \neq \ell \\ -b\gamma(r_i(t))q_{i,\ell}(t) + b(\gamma(r_i(t))) & \text{si } s_i(t) = \ell. \end{cases}$$

Alors, nous avons

$$\begin{aligned} G_{i,\ell}(Q) &= \lim_{b \rightarrow 0} \frac{1}{b} \tilde{p}_i \mathbb{E}[\Delta q_{i,\ell}(t) | Q(t)] \\ &= q_{i,\ell} \tilde{p}_i (\mathbb{E}[\gamma(r_i(t)) | Q(t), s_i(t) = \ell] - \mathbb{E}[\gamma(r_i(t)) | Q(t)]). \end{aligned}$$

En d'autres termes, selon le théorème 32, il y a convergence faible vers les solutions de l'équation (7.3), qui correspondent ici aux solutions de l'équation différentielle ordinaire (en supposant que γ est une fonction décroissante dont le domaine d'image est entre 0 et 1) :

$$\frac{dq_{i,\ell}}{dt} = -q_{i,\ell} \tilde{p}_i (c_i(e_\ell, Q_{-i}) - c_i(q_i, Q_{-i})), \quad (7.6)$$

Autrement dit, l'algorithme stochastique se comporte comme une équation différentielle ordinaire correspondant à une dynamique de réplication (de multi-population).

Les propriétés de telles dynamiques sont bien connus : voir les théorèmes *de tout le monde* de la théorie des jeux évolutionnaires.

Proposition 1 *Toutes les assertions suivantes sont vraies pour les solutions de l'équation (7.6) : (i) Tous les équilibres de Nash sont des points stationnaires. (ii) Tous les équilibres de Nash stricts sont asymptotiquement stables. (iii) Tous les points stationnaires stables sont des équilibres de Nash.*

7.3.2 Dynamique de *logit*

Pour obtenir notre second exemple, inspiré de [Cominetti et al., 2008], il suffit de définir les fonctions σ_i de la façon suivante :

$$\sigma_{i\ell}(q_i) = \frac{\exp(q_{i,\ell}/\kappa)}{\sum_j \exp(q_{i,j}/\kappa)}, \quad (7.7)$$

c'est-à-dire une dynamique "*logit*", où κ est une constante positive. Lorsque κ tend vers 0, cette dynamique converge vers la dynamique de la meilleure réponse, tandis que lorsque κ tend vers l'infini, cette dynamique se comporte comme un choix uniforme [Hofbauer and Sigmund, 2003].

Considérons $F_i^b(r_i(t), s_i(t), q_i(t))$ défini comme suit

$$F_i^b(r_i(t), s_i(t), q_i(t)) = (\gamma(r_i(t)) - q_{i,s_i(t)})e_{s_i(t)} \quad (7.8)$$

Dans ce cas, nous avons :

$$\begin{aligned} G_{i,\ell}(Q) &= \lim_{b \rightarrow 0} \frac{1}{b} \tilde{p}_i \mathbb{E}_\sigma [\Delta q_{i,\ell}(t) | Q(t)] \\ &= \sigma_{i,\ell}(q_{i,\ell}(t)) \tilde{p}_i (\mathbb{E}_\sigma [\gamma(r_i(t)) | Q(t), s_i(t) = \ell] - q_{i,\ell}). \end{aligned}$$

En d'autres termes, selon le théorème 32, il y a convergence faible vers les solutions de l'équation (7.3), qui correspondent ici aux solutions de l'équation différentielle ordinaire :

$$\frac{dq_{i,\ell}}{dt} = \sigma_{i,\ell}(q_{i,\ell}(t)) (\mathbb{E}_\sigma [\gamma(r_i(t)) | Q(t), s_i(t) = \ell] - q_{i,\ell}). \quad (7.9)$$

Les points limites de ces équations différentielles ordinaires correspondent aux équilibres de Nash d'un jeu dont les coûts sont perturbés par un terme additif tendant vers 0 lorsque κ tend vers 0 :

Proposition 2 ([Hofbauer and Sigmund, 2003],[Cominetti et al., 2008]) *Les points stationnaires des dynamiques (7.9) correspondent aux équilibres de Nash des jeux dont les utilités sont données par $v_i(Q) = c_i(Q) - \kappa \sum_\ell q_{i,\ell} \log q_{i,\ell}$.*

Une fois de plus, pour des jeux en général, nous obtenons que la limite pour $b \rightarrow 0$ des dynamiques des algorithmes stochastiques est une équation différentielle ordinaire dont les points limites stables (si ils existent) sont des équilibres de Nash lorsque $t \rightarrow \infty$. S'il y a convergence de l'équation différentielle ordinaire alors, on peut espérer que ces algorithmes stochastiques apprennent des équilibres.

7.3.3 Dynamique de la meilleure réponse seuillée

Pour notre troisième exemple, nous prenons des fonctions σ_i comme la fonction identité, et les fonctions $F_i^b(r_i(t), s_i(t), q_i(t))$ comme définies de la façon suivante :

1. choisir une autre stratégie pure $e_j(t)$ de façon uniforme : la stratégie j est choisie avec une probabilité $\frac{1}{m_i}$.
2. jouer la stratégie $e_j(t)$. Cela permet au joueur i d'obtenir un nouveau coût $r_i(t)$.
3. considérer alors

$$F_i^b(r_i(t), s_i(t), q_i(t)) = \nu(r_i(t) - r_j(t))(e_j - q_i(t)),$$

où ν est une fonction croissante qui retourne 0 pour des arguments négatifs et 1 pour des arguments plus grand que $\epsilon > 0$.

Cela mène à des dynamiques proches des dynamiques considérées par les auteurs de [Berenbrink and Schulte, 2007]. Aussi longtemps que l'état courant $Q(t)$ n'est pas un ϵ -équilibre de Nash, certains joueurs chercheront à améliorer leur réponse. Ainsi, les points stationnaires doivent correspondre à des ϵ -équilibres de Nash.

7.4 Propriétés de convergence.

7.4.1 Bornes sur la vitesse de convergence

Le défaut des résultats de convergence faible dans l'esprit du théorème 32 est qu'ils ne nous donnent aucun renseignement sur la vitesse de convergence. En effet, il n'est pas possible de déduire, étant donné $\epsilon > 0$, quel est le temps T , pour que les états de notre système soient à une distance inférieure à ϵ de l'équation différentielle ordinaire et quel est l'ordre de grandeur de b pour avoir une telle propriété.

Il est en fait possible d'obtenir des bornes sur la vitesse de convergence.

Observons pour cela que la dynamique (7.1) peut s'écrire

$$Q(t+1) = Q(t) + bG(Q) + bD(Q) + bR(Q) \quad (7.10)$$

où

$$D(Q) = \mathbb{E}_\sigma[F_i^b(r_i(t), s_i(t), q_i(t)) | Q(t)] - G_i(Q)$$

peut être vu comme une perturbation déterministe et

$$R(Q) = F_i^b(r_i(t), s_i(t), q_i(t)) - \mathbb{E}_\sigma[F_i^b(r_i(t), s_i(t), q_i(t)) | Q(t)]$$

comme une perturbation aléatoire.

Par hypothèses, $D(Q)$ est en $\mathcal{O}(b)$ et $\mathbb{E}[R(Q) | Q] = 0$.

En l'absence de ces perturbations, la dynamique (7.10) se comporte comme un schéma de discrétisation d'Euler $q_i(t+1) = q_i(t) + bG_i(Q)$ pour résoudre l'équation différentielle ordinaire (7.3).

L'idée est de généraliser les résultats connus sur les bornes sur les erreurs commises par les méthodes d'Euler pour résoudre les équations différentielles en prenant en compte les perturbations.

Cela donne les résultats suivants dont les preuves sont inspirées par les constructions de [Benaim, 1999].

Théorème 33 ([40]) *Supposons que l'équation différentielle ordinaire converge vers un équilibre de Nash Q^* pour certaines conditions initiales $Q(0)$. Soit $T(\epsilon)$ le temps de nécessaire pour converger vers des points à distance ϵ de Q^* .*

Alors pour tout $\epsilon > 0$, l'algorithme stochastique partant de $Q(0)$ va converger avec une forte probabilité vers un état à distance 2ϵ de Q^ . Cela se produira en un nombre d'étapes de l'ordre de*

$$\frac{1}{\epsilon} T(\epsilon) \exp(\Lambda T(\epsilon)),$$

si b est de l'ordre de $\epsilon \exp(-\Lambda T(\epsilon))$, et si Λ est une constante telle que G est Λ -Lipschtizienne.

Autrement dit, le temps requis par l'algorithme stochastique pour converger est formellement relié au temps de convergence de l'équation différentielle ordinaire correspondante.

En d'autres termes,

Corollaire 33.1 ([40]) *Borner le temps de convergence de l'algorithme stochastique correspond exactement à borner le temps de convergence de l'équation différentielle ordinaire associée (7.3).*

7.4.2 Exemple : le jeu symétrique à deux joueurs

Dans cette section, nous considérons les dynamiques décrites dans la section 7.3.1, qui tendent vers une l'équation différentielle ordinaire de la forme (7.6). Nous faisons leur étude exhaustive sur les jeux génériques à deux joueurs symétriques.

Les dynamiques de réplication ne changent pas même si les modifications de la matrice de coût décrites dans la section 5.1.3 sont réalisées. Les dynamiques (7.6) peuvent ainsi être formulées de la façon suivante :

$$\begin{aligned} \frac{dq_{1,1}}{dt} &= (a_1 q_{2,1} - a_2 q_{2,2}) q_{1,1} q_{1,2} & \frac{dq_{2,1}}{dt} &= (a_1 q_{1,1} - a_2 q_{1,2}) q_{2,1} q_{2,2} \\ \frac{dq_{1,2}}{dt} &= -\frac{dq_{1,1}}{dt} & \frac{dq_{2,2}}{dt} &= -\frac{dq_{2,1}}{dt} \end{aligned}$$

Dans le chapitre 5, nous avons rappelé la classification de [Weibull, 1995], des jeux symétriques à deux joueurs ayant deux stratégies. Les comportements peuvent se ranger en quatre catégories.

Catégorie I avec $a_1 a_2 < 0$, $a_1 < 0$ et $a_2 > 0$. Nous allons étudier le comportement de la trajectoire de la dynamique de réplication. La figure 7.1 représente le portrait de phase¹ dans le jeu où $a_1 = -1$ et $a_2 = 1$.

En constatant que $q_{2,1} + q_{2,2} = 1$, il est facile d'observer que $(a_1 q_{2,1} - a_2 q_{2,2}) < -\delta$ où $\delta = \min(-a_1, a_2)$. Comme $\frac{dq_{1,1}}{dt} = (a_1 q_{2,1} - a_2 q_{2,2}) q_{1,1} (1 - q_{1,1})$, nous avons $\frac{dq_{1,1}}{dt} < \delta(1 - q_{1,1}) q_{1,1}$ où $\delta = \min(-a_1, a_2)$. Par conséquent, nous obtenons

- $q_{1,1}(t) < e^{-\delta t} q_{1,1}(0)$ dès que $q_{1,1}(0) \neq 0$,
- $q_{2,1}(t) < e^{-\delta t} q_{2,1}(0)$ dès que $q_{2,1}(0) \neq 0$ (de façon symétrique).

Cela implique que toute solution de l'équation différentielle ordinaire du système va converger vers l'unique équilibre de Nash pur dont le profil est $q_{1,1} = q_{2,1} = 0$. Par conséquent, le temps de convergence $T(\epsilon)$ de l'équation différentielle pour atteindre cet équilibre est de l'ordre de $\frac{1}{\delta} \ln(\frac{1}{\epsilon})$.

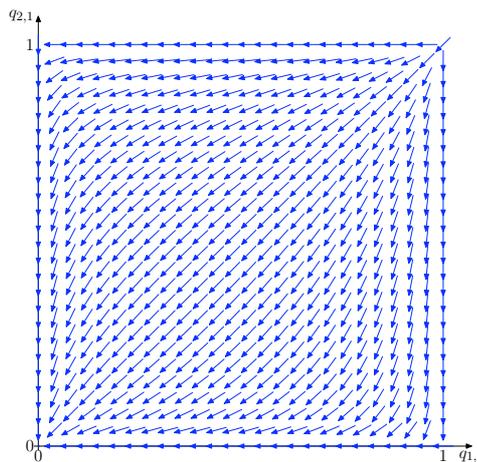


FIG. 7.1 – Le portrait de phase de l'équation de réplication pour un jeu de catégorie I

¹Un portrait de phase est une représentation géométrique d'un système dynamique : la direction de la trajectoire du système est représenté pour chaque point.

Catégorie IV avec $a_1 a_2 < 0$, $a_1 > 0$ et $a_2 < 0$: Le raisonnement et les résultats sont identiques à ceux de la catégorie I. Des calculs similaires permettent de constater que $\frac{dq_{1,2}}{dt} < \delta(1 - q_{1,2})q_{1,2}$, et que $\frac{dq_{2,2}}{dt} < \delta(1 - q_{2,2})q_{2,2}$ où $\delta = \min(a_1, -a_2)$. Le temps de convergence $T(\epsilon)$ vers l'unique équilibre de Nash pur est de l'ordre de $\frac{1}{\delta} \ln(\frac{1}{\epsilon})$.

Catégorie II avec $a_1 a_2 > 0$ et a_1, a_2 les deux sont positifs : La figure 7.2(a) représente le portrait de phase dans le jeu où $a_1 = 1$ et $a_2 = 2$. Pour étudier le comportement de la dynamique, nous posons $\lambda = \frac{a_2}{a_1 + a_2} - q_{2,1}$ et $\gamma = \frac{a_1}{a_1 + a_2} - q_{1,1}$.

Par calcul, observons $\frac{dq_{1,1}}{dt} = -\lambda(a_1 + a_2)q_{1,1}(1 - q_{1,1})$. Si $\lambda > 0$, alors $\frac{dq_{1,1}}{dt} < -(a_1 + a_2)\lambda q_{1,1} < 0$. Si $\lambda < 0$, alors $\frac{dq_{1,1}}{dt} > -(a_1 + a_2)\lambda q_{1,1} > 0$, ce qui entraîne $\frac{d(1 - q_{1,1})}{dt} < (a_1 + a_2)\lambda(1 - q_{1,1}) < 0$.

De façon symétrique, nous obtenons

$$- \frac{dq_{2,1}}{dt} < -(a_1 + a_2)\gamma q_{2,1} < 0 \text{ si } \gamma > 0$$

$$- \frac{d(1 - q_{2,1})}{dt} < (a_1 + a_2)\gamma(1 - q_{2,1}) < 0 \text{ si } \gamma < 0$$

Par conséquence, lorsque $\lambda > 0$, $q_{1,1}$ décroît, tandis que pour $\lambda < 0$, $q_{1,1}$ augmente. De façon symétrique pour $q_{2,1}$. L'équilibre de Nash mixte dont le profil correspond à $q_{1,1} = \frac{a_2}{a_1 + a_2}$ et à $q_{2,1} = \frac{a_1}{a_1 + a_2}$, est un point de selle. A l'exception de ce point, quelque soit le point initial, la dynamique converge vers un de ces deux équilibres de Nash pur (dont les profils sont $q_{1,1} = q_{2,1} = 0$ et $q_{1,1} = q_{2,1} = 1$).

L'espace peut se diviser en quatre régions dont chaque région a pour coin l'équilibre de Nash mixte.

Pour les deux régions diagonales, en fonction des calculs précédents, les dynamiques convergent vers un équilibre de Nash pur sans quitter sa région avec la vitesse de l'ordre de $\exp(-\delta t)$, pour $\delta = \min(|\lambda|, |\gamma|)$. Par conséquence, le temps $T(\epsilon)$ de convergence est de l'ordre de $\frac{1}{\delta} \ln(\frac{1}{\epsilon})$.

Pour les deux autres régions, la dynamique quitte sa région en temps de l'ordre de $\frac{1}{\delta} \ln(\frac{1}{\epsilon'})$ avec $\delta = \min(|\lambda|, |\gamma|)$ et avec ϵ' la distance entre ce point et l'équilibre de Nash mixte.

Catégorie III où $a_1 a_2 > 0$ et où a_1, a_2 sont tous les deux négatifs. La figure 7.2(b) représente le portrait de phase dans le jeu où $a_1 = -1$ et $a_2 = -2$.

Rappelons qu'il existe un unique équilibre de Nash mixte et deux équilibres de Nash purs. L'équilibre de Nash mixte a le profil suivant : $q_{1,1} = \frac{a_2}{a_1 + a_2}$, $q_{2,1} = \frac{a_1}{a_1 + a_2}$. Le premier équilibre de Nash pur a le profil suivant : $q_{1,1} = 1$, $q_{2,1} = 0$. Le second équilibre de Nash pur a le profil suivant : $q_{1,1} = 0$, $q_{2,1} = 1$.

Les dynamiques peuvent être étudiées comme pour celle de la catégorie II en partageant l'espace en quatre régions. La différence est maintenant que les régions diagonales sont quittées en un temps fini et que dans les régions anti-diagonales, les dynamiques convergent vers les équilibres de Nash purs.

La discussion précédente permet d'affirmer la convergence vers un point à distance plus petite que ϵ d'un équilibre de Nash en un temps polynomial en $\frac{1}{\epsilon}$, si b est de l'ordre de $\frac{1}{\epsilon}$.

Pour les jeux généraux, on peut espérer des informations sur la convergence en étudiant les valeurs propres du jacobien en les équilibres des dynamiques en s'inspirant de la théorie classique des systèmes dynamiques [Hirsch et al., 2003].

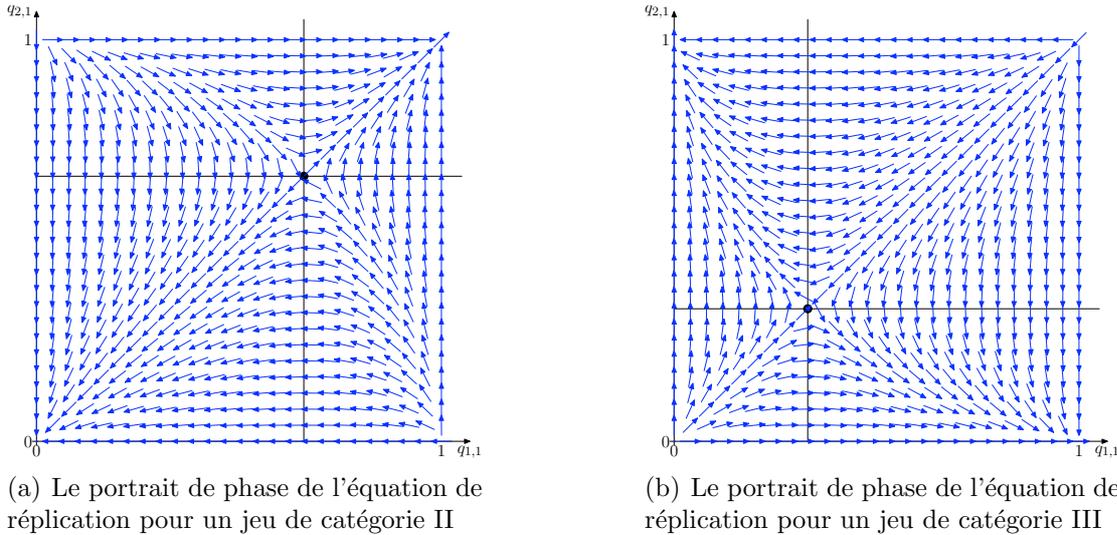


FIG. 7.2 – Le portrait de phase de l'équation de réplication pour un jeu de catégorie III. Le point correspond à l'équilibre de Nash mixte.

7.5 Algorithmes d'apprentissage pour les jeux de potentiel ordinal

Nous allons maintenant étudier la convergence pour les jeux de potentiel ordinal en utilisant l'algorithme 2.

On se focalise sur des dynamiques de *réplication perturbée* de la forme suivante :

- les fonctions σ_i sont la fonction identité ;
- les fonctions $F_i^b(r_i(t), s_i(t), q_i(t))$ sont définies de la façon suivante :

$$F_i^b(r_i(t), s_i(t), q_i(t)) = \mathcal{O}(b) + \begin{cases} \gamma(r_i(t))(e_{s_i(t)} - q_i(t)) & \text{avec probabilité } \alpha \\ b(e_{s_j} - q_i(t)) & \text{avec probabilité } 1 - \alpha, \\ & \text{où } j \in \{1, \dots, m_i\} \\ & \text{est choisit uniformément,} \end{cases} \quad (7.11)$$

où

- le paramètre α est une constante telle que $0 < \alpha < 1$;
- $\gamma : \mathbb{R} \rightarrow [0, 1]$ est une fonction *décroissante*. Les même remarques sur la fonction γ que celles faites lorsque nous avons discuté de la dynamique de réplication s'appliquent dans ce contexte.

Le terme $\mathcal{O}(b)$ désigne une perturbation qui reste de l'ordre du paramètre b , lorsque b tend vers 0. Cette perturbation peut être aléatoire (stochastique).

En utilisant les mêmes arguments que précédemment (par le calcul de $\Delta q_{i,\ell}(t)$ et de $G_i(Q) = \lim_{b \rightarrow 0} \frac{1}{b p_i} \mathbb{E}[\Delta q_{i,\ell}(t) | Q(t)]$), nous pouvons affirmer que de telles dynamiques correspondent aux dynamiques de réplication de multi-population : en d'autres termes, selon le théorème 32, il y a convergence faible vers les solutions de l'équation (7.3), qui correspondent ici aux solutions de l'équation différentielle ordinaire (7.6).

Nous insistons sur le fait que le raisonnement à venir ne peut pas s'appliquer (au moins directement) sur l'algorithme 1. En effet, la preuve du temps de convergence basée sur les arguments de martingale, ne fonctionne pas ici lorsque les joueurs réactualisent leur

stratégies mixtes simultanément.

7.5.1 Caractérisation via des fonctions de Lyapunov

Définition 35 (Jeu de Lyapunov [39]) *Un jeu est un jeu de Lyapunov s'il possède une fonction de Lyapunov, c'est-à-dire s'il existe une fonction $F : \mathcal{K} \rightarrow \mathbb{R}$ positive ou nulle, de classe \mathcal{C}^1 , telle que pour tout i, ℓ et Q , quand $G(Q) \neq 0$,*

$$\sum_{i,\ell} p_i \frac{\partial F}{\partial q_{i,\ell}}(Q) G_{i,\ell}(Q) < 0, \quad (7.12)$$

où $G_{i,\ell}$ correspond à la dynamique (7.6) sur l'espace des stratégies mixtes \mathcal{K} .

La classe des jeux de Lyapunov contient celle des jeux de potentiel ordinal et par conséquence celle des jeux de potentiel exact. De plus, on dira qu'une fonction de Lyapunov $F : \mathcal{K} \rightarrow \mathbb{R}$ est *multiaffine*, si et seulement si elle peut être définie comme un polynôme en toutes ses variables, qui est de degré 1 en chaque variable, et tel que chacun de ces monomes ne contient aucun facteur de la forme $q_{i,\ell} q_{i,\ell'}$.

Théorème 34 ([39]) *Un jeu de potentiel ordinal est un jeu de Lyapunov. De plus, les fonctions de Lyapunov sont multiaffines.*

Plus précisément, si la fonction ϕ est la fonction de potentiel du jeu de potentiel ordinal, alors la moyenne

$$F(Q) = \mathbb{E}[\phi(Q)] = \mathbb{E}[\phi(Q) \mid \text{les joueurs sélectionnent les stratégies pures selon } Q]$$

est une fonction de Lyapunov.

Définition 36 (Jeu de potentiel [Sandholm, 2001]) *Un jeu est un jeu de potentiel continu s'il existe une fonction de classe \mathcal{C}^1 $F : K \rightarrow \mathbb{R}$ telle que pour tout i, ℓ et Q ,*

$$\frac{\partial F}{\partial q_{i,\ell}}(Q) = u_i(e_\ell, Q). \quad (7.13)$$

Proposition 3 ([39]) *Un jeu de potentiel continu est un jeu de Lyapunov. De plus, les fonctions de Lyapunov sont multiaffines.*

Rappelons que les jeux de potentiel exacts sont définis à la page 45.

Proposition 4 ([39]) *Un jeu de potentiel exact ayant ϕ comme fonction de potentiel est un jeu de potentiel continu avec $F(Q) = \mathbb{E}[\phi(Q)]$ comme fonction de potentiel. Et en restreignant F aux fonctions de classe \mathcal{C}^2 , le jeu de potentiel continu est aussi un jeu de potentiel exact.*

Un jeu de Lyapunov peut à priori avoir une fonction de potentiel non multiaffine, et par conséquence il peut exister des jeux de Lyapunov qui ne sont pas des jeux de potentiel ordinal.

L'intérêt de fonctions de Lyapunov est qu'elles fournissent un outil pour prouver des propriétés de convergence. Rappelons que l'ensemble limite $\omega(Q_0)$ du point Q_0 est l'ensemble des points d'accumulation des trajectoires partant du point Q_0 : en considérant une trajectoire partant de Q_0 , cet ensemble contient Q^* avec $Q^* = \lim_{n \rightarrow \infty} Q(t_n)$, pour toute suite croissante $(t_n)_{n \geq 0} \in \mathbb{R}$.

Proposition 5 ([39]) *Pour tout jeu de Lyapunov, les solutions de l'équation différentielle ordinaire (7.3) ont leur ensemble des points limites non vide, compact, composé uniquement de points stationnaires de cette dynamique. Et sur ces points limites, F est constant.*

Observons que toutes les classes précédentes sont des jeux de Lyapunov selon les dynamiques (7.6). Et par conséquent,

Corollaire 34.1 ([39]) *Dans un jeu de Lyapunov, quelque soit la condition initiale, les solutions de l'équation différentielles ordinaires (7.6) vont converger. Les points limites stables sont des équilibres de Nash.*

7.5.2 Techniques de Martingale

Il est en fait possible de dire plus de choses sur le temps de convergence pour les types de jeux précédemment cités (ordinaire, de potentiel exact, de potentiel continu, d'ordonnancement, de congestion, d'allocation de tâches) par l'utilisation de techniques de martingales.

En effet, il est possible dans ce cas de travailler directement sur l'algorithme stochastique en évitant le passage à l'équation différentielle ordinaire (7.3) et à la double limite $b \rightarrow 0, t \rightarrow \infty$.

Le lemme suivant est fondamental, et à la base du raisonnement.

Lemme 1 *Lorsque F est une fonction de Lyapunov multiaffine,*

$$\mathbb{E}[\Delta F(Q(t+1)) | Q(t)] = \sum_{i=1}^n \sum_{\ell=1}^{m_i} \frac{\partial F}{\partial q_{i,\ell}}(Q(t)) \mathbb{E}[\Delta q_{i,\ell} | Q(t)], \quad (7.14)$$

où $\Delta F(t) = F(Q(t+1)) - F(Q(t))$.

Lorsque l'on considère une dynamique de réplication en utilisant le fait que $G_i(Q) = \lim_{b \rightarrow 0} \tilde{F}_i^b(Q)$ le coté droit de l'équation (7.14) est

$$b \sum_{i=1}^n \sum_{\ell=1}^{m_i} p_i \frac{\partial F}{\partial q_{i,\ell}}(Q) G_{i,\ell}(Q) + \mathcal{O}(b^2). \quad (7.15)$$

L'espérance de cette quantité est négative par l'équation (7.12) lorsque $G(Q) \neq 0$ et b suffisamment petit.

Lorsque b est suffisamment petit, $(F(Q(t)))_t$ sera une super-martingale jusqu'à attendre le point où (7.15) est proche de 0.

De plus, pour les dynamiques de réplication, l'équation (7.15) peut se réécrire

$$-b \frac{1}{4} \sum_i p_i \sum_{\ell \neq \ell'} q_{i,\ell} q_{i,\ell'} (u_i(e_\ell, Q_{-i}) - u_i(e_{\ell'}, Q_{-i}))^2 + \mathcal{O}(b^2).$$

Lorsque l'on considère les dynamiques stochastiques perturbées en utilisant des arguments de la théorie des martingales, on peut obtenir des résultats sur la stabilité : nous notons $L(\mu)$ pour le sous ensemble des états Q tel que $F(Q) \leq \mu$.

Proposition 6 ([39]) *Soit $\lambda > 1$. Soit $Q(t_0)$ un état. Soit b un réel suffisamment petit pour que (7.14) soit négatif ou nul en dehors de $L(F(Q(0)))$. Alors $Q(t)$ sera tel que $w(Q(t) \in L(\lambda F(Q(t_0))))$ à jamais après un temps $t \geq t_0$ avec une probabilité plus grande que $1 - \frac{1}{\lambda}$.*

Remarquons que pour un jeu de Lyapunov, avec une fonction F de Lyapunov multiaffine, les points Q^* réalisant un minimum local F^* de F sur le compact \mathcal{K} doivent correspondre aux équilibres de Nash.

Une dynamique est dite *perturbée*, si pour tout $Q \in K$, pour n'importe quel voisinage V de Q dans sa fermeture, la probabilité que $Q(t+1) \in V$ quand $Q(t) = Q$ est positive.

Si la dynamique est perturbée, alors la chaîne de Markov associée est ergodique. Ce implique que tout voisinage sera visité avec une probabilité positive.

Ainsi, si dans un voisinage d'un tel point, nous pouvons appliquer la proposition précédente, on obtiendrait que presque sûrement, après un certain temps, $Q(t)$ sera proche d'un équilibre de Nash pour toujours avec une forte probabilité. Le défaut de cette approche est le fait que nous donnons aucune borne sur le temps pour atteindre un tel voisinage.

Il est en fait possible d'obtenir des bornes sur le temps espéré de convergence, en utilisant les arguments précédents (F est une super-martingale) :

Théorème 35 ([39]) *Considérons un jeu de Lyapunov avec une fonction F de Lyapunov multiaffine. Supposons $b = \mathcal{O}(\epsilon)$. Quel que soit l'état initial de l'algorithme, l'algorithme atteindra presque sûrement un ϵ -équilibre de Nash. De plus, le temps espéré $T(\epsilon)$ est majoré par*

$$T(\epsilon) \leq \mathcal{O}\left(\frac{F(Q(0))}{\epsilon}\right).$$

Nous croyons que ces bornes sont ténues pour des jeux de potentiel ordinal génériques. En effet, dans ce type de jeu, il n'est pas nécessaire d'avoir une relation entre les variations des utilités et de celle de la fonction de potentiel : seul le signe de la variation doit être identique.

Nous pensons qu'il existe de meilleures bornes pour les jeux de congestion sous certaines hypothèses, comme par exemple la condition du α -saut. Nous travaillons actuellement sur ce point.