Systèmes Dynamiques

HAMMOUCH Siham

1 Introduction

De nombreuses décisions économiques doivent intégrer la dimension temporelle : produire aujourd'hui peut avoir des conséquences demain. Cette problématique est devenue un enjeu majeur, notamment en matière d'environnement et de santé. Lorsqu'un agent cherche à optimiser son comportement dans le temps tout en tenant compte de l'évolution d'un système, on entre alors dans le domaine du contrôle optimal.

Ce projet s'inscrit dans ce cadre. Il s'appuie sur un article de Aart de Zeeuw (2024), qui étudie comment un agent (ou plusieurs) peuvent prendre des décisions optimales face à une contrainte dynamique, comme l'accumulation de pollution. L'article mobilise deux outils importants vus en cours : la méthode de Pontryagin et la programmation dynamique, chacun offrant une manière différente d'aborder les décisions dans le temps.

Le sujet devient encore plus intéressant et complexe lorsque plusieurs agents interagissent. Chacun agit de façon stratégique, tout en étant affecté par les choix des autres. Ces situations font appel aux jeux dynamiques, et permettent de comprendre des phénomènes économiques majeurs, comme la gestion collective des ressources ou la régulation des externalités.

Ce travail vise ainsi à mieux comprendre comment les mathématiques peuvent aider à représenter et analyser ces situations dynamiques et stratégiques, au croisement de l'économie, de l'optimisation et de la théorie des jeux.

Ce rapport s'organise en trois grands points. Je vais commencer par le cas d'un joueur unique, avant de passer au jeu à deux joueurs, puis je finirais sur le lien avec les Mean Field Games.

2 Cas d'un joueur unique : deux approches de contrôle optimal

2.1 Le cadre du problème d'optimisation dynamique

Je commence par poser les bases d'un problème classique de contrôle optimal en temps continu, avec un horizon infini. On se trouve dans ce type de situation lorsqu'un agent cherche à maximiser un gain cumulé (ou une utilité) sur une durée illimitée, tout en tenant compte d'une contrainte liée à l'évolution dans le temps d'un état économique.

C'est exactement le type de problème introduit dans l'article de Aart de Zeeuw (2024), où un agent (ou un joueur) cherche à maximiser une fonction objectif intertemporelle, soumise à une contrainte dynamique. Ce cadre, typique du contrôle optimal, est formulé mathématiquement comme suit :

$$\max_{u(\cdot)} \int_0^\infty e^{-rt} F(x(t), u(t))\, dt, \tag{1}$$

où :

- $x(t)$ représente l'état du système à l'instant t (par exemple : un stock de pollution),
- $u(t)$ est la variable de contrôle (par exemple : le niveau d'émission),
- $F(x, u)$ est la fonction de profit ou d'utilité instantanée,
- r est le taux d'actualisation.

Cette maximisation est soumise à une contrainte dynamique sur l'évolution de l'état:

$$\dot{x}(t) = f(x(t), u(t)), \quad x(0) = x_0, \tag{2}$$

où $\dot{x}(t)$ désigne la dérivée de l'état par rapport au temps. La fonction f modélise comment l'état évolue en fonction de ses propres valeurs et du contrôle.

Pour résoudre ce type de problème, j'utilise deux méthodes principales :

- la méthode de Pontryagin, ou principe du maximum, qui s'appuie sur un système d'équations différentielles mêlant l'état et une variable auxiliaire (le co-état),
- la méthode de Bellman, dite programmation dynamique, qui repose sur une équation fonctionnelle caractérisant la valeur optimale du problème.

Dans les deux approches, l'objectif est de trouver une stratégie optimale $u^*(t)$ qui, à partir de l'état initial $x(0) = x_0$, maximise la fonction objectif tout en respectant la dynamique imposée.

Les sections qui suivent présentent ces deux méthodes à travers un exemple concret, inspiré d'un modèle environnemental de gestion de la pollution, tel que présenté dans l'article étudié.

2.2 Méthode de Pontryagin

2.2.1 Présentation générale

La méthode de Pontryagin repose sur l'idée d'introduire une variable auxiliaire appelée co-état, qui joue le rôle de multiplicateur de Lagrange dans ce contexte dynamique. Cette méthode transforme le problème initial en un système d'équations différentielles couplées, qu'il faut résoudre pour obtenir la stratégie optimale.

Je prends un exemple simple : un agent choisit une intensité de production $y(t)$, qui génère des bénéfices, mais aussi de la pollution. Le stock de pollution, noté $s(t)$, augmente avec la production et décroît avec un taux naturel de dépollution, noté δ. L'agent cherche à maximiser son profit net, en tenant compte à la fois du coût de production et des dommages environnementaux.

Problème :

$$\max_{y(\cdot)} \int_0^\infty e^{-rt} \left[\alpha y(t) - \frac{1}{2} y(t)^2 - \frac{1}{2}\gamma s(t)^2\right] dt \tag{3}$$

sous la contrainte dynamique :

$$\dot{s}(t) = y(t) - \delta s(t), \quad s(0) = s_0. \tag{4}$$

Ce type de configuration est typique des modèles où la production a un effet négatif sur l'environnement. Ici, plus un agent produit, plus il gagne de l'argent, mais plus il pollue. Il doit donc faire un compromis entre produire beaucoup (pour maximiser ses profits) et limiter la pollution (pour ne pas avoir trop de dommages dans le futur).

La fonction objectif est construite de façon à refléter cet arbitrage :

- le terme $\alpha y(t)$ représente le gain de production : chaque unité produite rapporte α ;
- le terme $-\frac{1}{2}y(t)^2$ représente un coût de production croissant, ce qui empêche l'agent de produire à l'infini ;
- le terme $-\frac{1}{2}\gamma s(t)^2$ reflète les dommages causés par un stock de pollution élevé.

Je choisis une forme quadratique car elle est plus simple à manipuler mathématiquement et permet de représenter des coûts ou des dommages croissants de manière réaliste. Ce cadre est classique en économie de l'environnement, et permet d'utiliser les méthodes de contrôle optimal comme Pontryagin ou Bellman.

Cette forme canonique se retrouve dans l'article de Aart de Zeeuw (2024), bien que présentée de manière verbale. J'en propose ici une reformulation mathématique explicite, adaptée à mon cas d'étude.

Pour rendre ce cadre plus concret, on peut comparer l'expression générale du problème à sa spécification dans le modèle environnemental étudié. Le tableau suivant montre comment les différentes composantes s'adaptent :

Élément	**Formulation générale**	**Application au modèle de pollution**
État du système	$x(t)$	$s(t)$: stock de pollution
Contrôle	$u(t)$	$y(t)$: niveau de production
Objectif instantané	$F(x(t), u(t))$	$\alpha y(t) - \frac{1}{2}y(t)^2 - \frac{1}{2}\gamma s(t)^2$
Taux d'actualisation	r	r (inchangé)
Dynamique de l'état	$\dot{x}(t) = f(x(t), u(t))$	$\dot{s}(t) = y(t) - \delta s(t)$

Je vais maintenant résoudre ce problème à l'aide de la méthode de Pontryagin, en commençant par la définition du Hamiltonien.

2.2.2 Étape 1 : Définir le Hamiltonien

Pour appliquer la méthode de Pontryagin, je commence par construire une fonction appelée "Hamiltonien", qui permet de combiner à la fois l'objectif instantané de l'agent et la contrainte dynamique sur l'évolution de l'état.

Dans mon cas, l'agent cherche à maximiser un profit net instantané donné par :

$$\alpha y(t) - \frac{1}{2}y(t)^2 - \frac{1}{2}\gamma s(t)^2,$$

Mais ce gain est soumis à une contrainte dynamique sur l'évolution du stock de pollution :

$$\dot{s}(t) = y(t) - \delta s(t).$$

Pour tenir compte de cette contrainte dans l'optimisation, j'introduis une variable auxiliaire $\lambda(t)$ appelée co-état, qui joue un rôle similaire à un multiplicateur de Lagrange dans les problèmes statiques. Cette variable mesure le coût marginal d'une variation du stock de pollution sur la fonction objectif.

Le Hamiltonien (en valeur courante) se définit alors comme la somme :

- du profit instantané net,
- et du produit du co-état λ par la contrainte dynamique $\dot{s}(t)$.

J'obtiens donc :

$$H(s, y, \lambda) = \alpha y - \frac{1}{2}y^2 - \frac{1}{2}\gamma s^2 + \lambda(y - \delta s).$$

$\Rightarrow$ *Construction d'un système dynamique (état + co-état).*

Cette fonction sera utilisée pour déterminer les conditions nécessaires à l'optimalité.

2.2.3 Étape 2 : Conditions d'optimalité (Pontryagin)

Une fois le Hamiltonien défini, la méthode de Pontryagin impose un ensemble de conditions nécessaires pour qu'un contrôle soit optimal. Ces conditions traduisent le fait que, le long de la trajectoire optimale, le Hamiltonien doit être maximisé à chaque instant.

Trois équations fondamentales en découlent :

- **Condition d'optimalité sur le contrôle** $y(t)$: Le contrôle optimal est celui qui maximise le Hamiltonien par rapport à y. Je dérive donc H par rapport à y :

$$\frac{\partial H}{\partial y} = \alpha - y + \lambda = 0 \quad \Rightarrow \quad y(t) = \alpha + \lambda(t) \tag{5}$$

Cette équation me dit que la production optimale dépend directement du co-état : plus la pollution est coûteuse à la marge (si λ est négatif), plus le niveau de production est réduit.

$\Rightarrow$ *Stratégie déterminée à l'instant initial (open-loop).*

- **Équation d'état (évolution de $s(t)$)** : Elle est donnée directement par la contrainte dynamique du problème :

$$\dot{s}(t) = y(t) - \delta s(t) \tag{6}$$

Elle indique que la pollution augmente avec la production $y(t)$ et décroît naturellement avec un taux de dépollution δ.

- **Équation de co-état (évolution de $\lambda(t)$)** :

Cette équation décrit comment évolue la variable de co-état $\lambda(t)$ au cours du temps. Elle se déduit en dérivant le Hamiltonien par rapport à l'état s, en tenant compte de tous les termes qui dépendent de s.

Le Hamiltonien contient deux termes où s intervient :

$$H(s, y, \lambda) = \cdots - \frac{1}{2}\gamma s^2 + \lambda(y - \delta s)$$

Je dérive donc ces deux termes :

$$\frac{\partial}{\partial s}\left(-\frac{1}{2}\gamma s^2\right) = -\gamma s, \quad \frac{\partial}{\partial s}\left(\lambda(-\delta s)\right) = -\delta\lambda$$

Donc :

$$\frac{\partial H}{\partial s} = -\gamma s - \delta\lambda$$

J'applique ensuite la condition dynamique de Pontryagin :

$$\dot{\lambda}(t) = r\lambda(t) - \frac{\partial H}{\partial s} = r\lambda(t) - (-\gamma s(t) - \delta\lambda(t)) \tag{7}$$

Ce qui donne :

$$\dot{\lambda}(t) = (r + \delta)\lambda(t) + \gamma s(t) \tag{8}$$

Cette équation indique que la dynamique du coût marginal de la pollution $\lambda(t)$ dépend à la fois de son niveau actuel et de l'état du système $s(t)$. Elle reflète le fait que plus la pollution augmente, plus le coût intertemporel associé s'intensifie.

Ce système d'équations différentielles détermine entièrement la trajectoire optimale de l'état $s(t)$ et de la variable de co-état $\lambda(t)$ dans le temps.

2.2.4 Étape 3 : Système dynamique et état stationnaire

En remplaçant l'expression optimale du contrôle $y(t) = \alpha + \lambda(t)$ obtenue à l'étape précédente dans l'équation d'état $\dot{s}(t) = y(t) - \delta s(t)$, j'obtiens un système de deux équations différentielles couplées reliant l'état $s(t)$ et le co-état $\lambda(t)$:

$$\begin{cases} \dot{s}(t) = \alpha + \lambda(t) - \delta s(t) \\ \dot{\lambda}(t) = (r+\delta)\lambda(t) + \gamma s(t) \end{cases} \tag{9}$$

Ce système est linéaire, ce qui facilite l'analyse de sa dynamique. Je m'intéresse notamment à l'état stationnaire, c'est-à-dire la situation vers laquelle le système converge lorsque $t \to \infty$ et que les dérivées s'annulent : $\dot{s}(t) = 0$ et $\dot{\lambda}(t) = 0$.

Je cherche donc à résoudre le système suivant :

$$\begin{cases} 0 = \alpha + \lambda^* - \delta s^* \\ 0 = (r+\delta)\lambda^* + \gamma s^* \end{cases}$$

À partir de la première équation, j'isole s^* :

$$s^* = \frac{\alpha + \lambda^*}{\delta}$$

Puis j'injecte cette expression dans la seconde équation :

$$0 = (r+\delta)\lambda^* + \gamma \left(\frac{\alpha + \lambda^*}{\delta} \right)$$

Je regroupe les termes en λ^* :

$$(r+\delta)\lambda^* + \frac{\gamma\alpha}{\delta} + \frac{\gamma\lambda^*}{\delta} = 0 \quad \Rightarrow \quad \lambda^* \left(r + \delta + \frac{\gamma}{\delta} \right) = -\frac{\gamma\alpha}{\delta}$$

Je simplifie cette expression :

$$\lambda^* = -\frac{\gamma\alpha}{\delta(r+\delta) + \gamma}$$

Et j'en déduis s^* :

$$s^* = \frac{\alpha + \lambda^*}{\delta} = \frac{\alpha}{\delta} - \frac{\gamma\alpha}{\delta(\delta(r+\delta)+\gamma)} = \frac{\alpha(r+\delta)}{\delta(r+\delta)+\gamma}$$

Finalement, les valeurs stationnaires sont :

$$s^* = \frac{\alpha(r+\delta)}{\delta(r+\delta)+\gamma}, \quad \lambda^* = -\frac{\gamma\alpha}{\delta(r+\delta)+\gamma} \tag{10}$$

Ces deux expressions me donnent la situation d'équilibre vers laquelle tendent la pollution et son coût marginal associé lorsque l'agent applique de manière continue la stratégie optimale.

Interprétation économique :

- Le niveau optimal de pollution s^* dépend du compromis entre les bénéfices de production (α) et les dommages environnementaux (γ).
- Le co-état λ^* est négatif : il reflète le coût marginal de la pollution.
- L'expression $y^* = \alpha + \lambda^*$ s'interprète comme un ajustement par une taxe optimale : l'agent réduirait sa production si une taxe environnementale de montant $-\lambda^*$ était imposée.

La programmation dynamique repose sur un principe fondamental formulé par Bellman, appelé principe d'optimalité. Il affirme qu'une stratégie est optimale si, à partir de chaque état que le système atteint, les décisions prévues pour la suite restent également optimales. En d'autres termes, les choix futurs dépendent uniquement de l'état courant, et non de l'histoire passée, ce qui permet de raisonner de manière récursive.

Jusqu'ici, j'ai résolu le problème à l'aide de la méthode de Pontryagin, en utilisant un système dynamique couplant l'état et une variable de co-état. Je vais maintenant présenter une autre approche classique du contrôle optimal : la programmation dynamique, qui repose sur une logique différente mais conduit, dans mon cas, au même résultat.

Contrairement à la méthode de Pontryagin qui travaille directement sur les trajectoires temporelles et le profit actualisé, la programmation dynamique repose sur une fonction de valeur $V(s)$, qui donne pour chaque état s le gain maximal qu'un agent peut espérer en optimisant ses décisions futures.

Au lieu de chercher une trajectoire optimale complète dès l'instant initial, cette méthode s'appuie sur la notion de fonction de valeur : pour chaque niveau initial de pollution s, je détermine le gain maximal que peut espérer l'agent à partir de cet état. Cette fonction, notée $V(s)$, résume l'ensemble du problème d'optimisation.

Le problème peut alors être reformulé de la manière suivante :

$$V(s) = \max_{y(\cdot)} \int_0^\infty e^{-rt} \left[\alpha y(t) - \frac{1}{2} y(t)^2 - \frac{1}{2} \gamma s(t)^2 \right] dt, \tag{11}$$

sous la contrainte dynamique :

$$\dot{s}(t) = y(t) - \delta s(t), \quad s(0) = s.$$

2.3 Méthode de Bellman

2.3.1 Présentation générale

L'élément central de la programmation dynamique est l'équation de Hamilton-Jacobi-Bellman (HJB), qui exprime la condition d'optimalité à chaque point de l'état. Dans mon cas stationnaire, elle prend la forme suivante :

$$rV(s) = \max_y \left[\alpha y - \frac{1}{2}y^2 - \frac{1}{2}\gamma s^2 + V'(s)(y - \delta s) \right]. \tag{12}$$

Le terme de droite correspond au gain immédiat, auquel s'ajoute l'effet anticipé d'un changement de l'état s. Quant au terme de gauche, $rV(s)$, il reflète le rendement actualisé de la fonction de valeur : autrement dit, le coût de renoncer à $V(s)$ aujourd'hui plutôt que de l'obtenir progressivement dans le temps.

2.3.2 Étape 1 : Résolution du problème intérieur

Pour résoudre l'équation HJB, je commence par maximiser l'expression entre crochets à droite de l'équation, c'est-à-dire le gain courant corrigé par le terme d'anticipation. Cette maximisation s'effectue par rapport à y, la variable de décision à chaque instant.

Je considère donc le problème statique suivant :

$$\max_y \left[\alpha y - \frac{1}{2}y^2 + V'(s)(y - \delta s) - \frac{1}{2}\gamma s^2 \right]$$

Comme s est donné, je maximise seulement par rapport à y. Les autres termes sont constants dans cette maximisation. Je calcule la dérivée par rapport à y :

$$\frac{\partial}{\partial y}\left(\alpha y - \frac{1}{2}y^2 + V'(s)(y - \delta s) - \frac{1}{2}\gamma s^2 \right) = 0 \quad \Rightarrow \quad y^*(s) = \alpha + V'(s)$$

J'obtiens ainsi une expression de la stratégie optimale en fonction de l'état courant s et de la dérivée de la fonction de valeur.

$\Rightarrow$ *Stratégie dépendante de l'état courant (feedback).*

2.3.3 Étape 2 : Substitution dans l'équation HJB

Je remplace maintenant cette expression $y^*(s)$ dans l'équation HJB pour obtenir une équation purement en fonction de s et de $V(s)$:

$$rV(s) = \alpha(\alpha + V') - \frac{1}{2}(\alpha + V')^2 - \frac{1}{2}\gamma s^2 + V'(\alpha + V' - \delta s)$$

Cette expression peut paraître difficile à manipuler directement. Pour simplifier les calculs, j'adopte une hypothèse fonctionnelle classique.

2.3.4 Étape 3 : Hypothèse quadratique et résolution

Je suppose que la fonction de valeur $V(s)$ est de forme quadratique :

$$V(s) = as^2 + bs + c \quad \Rightarrow \quad V'(s) = 2as + b$$

Cette hypothèse est cohérente avec la structure quadratique du problème (coûts et dommages), et permet de ramener l'équation HJB à une équation polynomiale. Le calcul est un peu long mais standard (on le retrouve dans l'article). J'identifie alors les coefficients du polynôme en regroupant les termes selon les puissances de s, ce qui permet de déduire les constantes a, b et c.

Le résultat est :

$$\begin{aligned} a &= \frac{r + 2\delta - \sqrt{(r+2\delta)^2 + 4\gamma}}{4} < 0 \\ b &= \frac{2a\alpha}{r + \delta - 2a} \\ c &= \frac{(\alpha + b)^2}{2r} \end{aligned}$$

2.3.5 Étape 4 : Stratégie optimale et trajectoire de l'état

Une fois a et b déterminés, je peux expliciter la stratégie optimale comme une fonction linéaire de l'état s :

$$y^*(s) = \alpha + V'(s) = \alpha + 2as + b$$

J'injecte cette expression dans la dynamique de l'état :

$$\dot{s}(t) = y^*(s) - \delta s = (2a - \delta)s + \alpha + b$$

Cette équation différentielle est linéaire du premier ordre. Elle est stable dès lors que le coefficient devant s est négatif, c'est-à-dire ici $2a - \delta < 0$, ce qui est garanti puisque $a < 0$ et $\delta > 0$. Cela implique que la trajectoire de $s(t)$ converge vers un état stationnaire.

Cette équation est de la forme classique :

$$\dot{s}(t) = \kappa s(t) + c,$$

avec $\kappa = 2a - \delta$ et $c = \alpha + b$. En posant $\dot{s}(t) = 0$, j'obtiens le point fixe vers lequel converge la trajectoire du système.

$\Rightarrow$ *À long terme, la pollution atteint un niveau stable, car l'agent adapte sa production selon l'état du système.*

J'en déduis que la pollution converge vers un niveau stationnaire donné par :

$$s^* = \frac{\alpha + b}{\delta - 2a}$$

Interprétation économique : Le résultat obtenu admet plusieurs interprétations économiques importantes, notamment sur le lien entre pollution et stratégie de production.

- La stratégie optimale obtenue dépend directement de l'état du système : plus le stock de pollution s est élevé, plus l'agent réduit son niveau de production y. Cela reflète une logique intuitive de précaution, qui limite l'accumulation de pollution à long terme.

- Cette méthode fournit naturellement une stratégie en feedback : les décisions sont ajustées à chaque instant en fonction de la situation actuelle. L'agent réagit donc en temps réel à l'évolution de la pollution, sans suivre une trajectoire figée à l'avance.

- Enfin, je retrouve une correspondance directe avec la méthode de Pontryagin : le co-état $\lambda(t)$ peut être interprété comme la dérivée de la fonction de valeur, c'est-à-dire $\lambda(t) = V'(s(t))$. Les deux méthodes aboutissent donc à la même logique économique, bien qu'elles partent de raisonnements différents : l'une repose sur un système dynamique, l'autre sur une équation fonctionnelle.

J'ai donc résolu ce problème à l'aide de deux approches différentes. La méthode de Bellman m'a permis d'identifier une stratégie optimale qui réagit en temps réel à l'état du système, en s'appuyant sur la fonction de valeur $V(s)$. Il est désormais intéressant de comparer cette méthode à celle de Pontryagin, tant sur le plan des résultats que des raisonnements sous-jacents.

2.4 Comparaison des deux méthodes et discussion

Bien qu'elles reposent sur des approches mathématiques différentes, les méthodes de Pontryagin et Bellman aboutissent ici à la même solution optimale. Cette équivalence est précieuse à comprendre, mais elle ne doit pas masquer les différences profondes entre les deux démarches.

2.4.1 Un résultat final identique

Dans les deux cas, la stratégie optimale s'écrit sous la forme :

$$y^*(t) = \alpha + \lambda(t) \quad \text{(Pontryagin)}, \qquad y^*(s) = \alpha + V'(s) \quad \text{(Bellman)}$$

Si j'identifie $\lambda(t) = V'(s(t))$, je retrouve exactement la même politique optimale. Les trajectoires de l'état $s(t)$ et les niveaux stationnaires sont également identiques :

$$s^* = \frac{\alpha + b}{\delta - 2a}, \qquad \lambda^* = V'(s^*) = 2as^* + b$$

2.4.2 Deux logiques différentes

La principale différence tient à la manière dont les méthodes abordent le problème :

- La méthode de Pontryagin cherche une trajectoire optimale complète dès le départ, à partir d'un état initial donné. Elle repose sur la résolution d'un système d'équations différentielles (état + co-état).
- La méthode de Bellman, au contraire, raisonne point par point : elle cherche quelle décision est optimale à chaque instant, en fonction de l'état courant. J'obtiens alors une équation fonctionnelle (HJB) pour la fonction de valeur.

Je peux dire que Pontryagin adopte une stratégie open-loop (la trajectoire est fixée une fois pour toutes), tandis que Bellman fournit naturellement une stratégie en feedback (elle s'adapte à l'état au fil du temps).

2.4.3 Rôle de l'information

Cette différence entre stratégie open-loop et stratégie en feedback a des implications économiques importantes :

- Une stratégie open-loop suppose que l'agent suit un plan fixé à l'avance, sans ajustement possible : elle ne réagit pas aux imprévus ou aux chocs.
- Une stratégie en feedback permet au contraire de réagir en temps réel à l'état du système. Elle est donc plus réaliste dans les contextes dynamiques, par exemple si le niveau de pollution évolue de manière inattendue.

Dans mon exemple, la méthode de Bellman fournit ainsi une stratégie plus souple et plus robuste face à l'incertitude ou aux perturbations.

2.4.4 Choix en pratique

Le choix entre les deux méthodes dépend des objectifs et du contexte :

- Pontryagin est plus simple à mettre en œuvre lorsque le problème est linéaire-quadratique. Il permet souvent d'obtenir une solution analytique rapidement.
- Bellman est plus puissant pour modéliser des stratégies réactives, intégrer de l'incertitude, ou résoudre des problèmes de manière numérique.

2.4.5 Conclusion de la comparaison

Dans un cadre bien posé comme celui-ci, les deux méthodes sont cohérentes et donnent la même stratégie. Pontryagin apporte une bonne intuition dynamique et se prête bien aux calculs explicites. Bellman, plus exigeant, permet une approche plus souple et plus générale. Leur complémentarité est précieuse pour comprendre les comportements intertemporels en économie.

3 Cas à plusieurs joueurs : du contrôle optimal au jeu différentiel

3.1 Présentation du jeu dynamique à deux joueurs

Après avoir étudié le cas d'un joueur unique, je passe maintenant à une situation plus riche : celle où deux agents interagissent dans un environnement commun. Cette configuration, centrale dans l'article de Aart de Zeeuw (2024), change profondément la nature du problème : il ne s'agit plus de maximiser une fonction objectif individuelle face à une contrainte dynamique donnée, mais de le faire en tenant compte des actions d'un autre joueur.

On entre ainsi dans le domaine des jeux différentiels : chaque joueur cherche à optimiser sa propre fonction objectif intertemporelle, tout en influençant et étant influencé par les décisions de l'autre. Contrairement au cas précédent, il n'y a plus une unique variable de contrôle, mais deux contrôles simultanés, et l'état du système évolue désormais sous l'effet combiné de ces deux actions.

$\Rightarrow$ *Je passe d'un problème de contrôle optimal à un problème d'équilibre stratégique.*

Structure du jeu Je considère ici deux joueurs $i = 1, 2$, chacun choisissant un contrôle $u_i(t)$ dans le but de maximiser son propre objectif. Le système est décrit par un unique état dynamique $x(t)$ (par exemple, le stock de pollution), affecté par les deux contrôles.

Autrement dit, les décisions de chacun ont des effets croisés, et l'état du système est partagé.

Chaque joueur choisit donc une trajectoire de contrôle $u_i(\cdot)$, tout en anticipant celle de l'autre, pour maximiser une fonction objectif de la forme:

$$J_i(u_1(\cdot), u_2(\cdot)) = \int_0^\infty e^{-rt} F_i(x(t), u_1(t), u_2(t))\, dt, \tag{13}$$

où :

- F_i représente le gain instantané du joueur i, qui dépend à la fois de l'état du système $x(t)$ et des deux actions $u_1(t)$ et $u_2(t)$,
- r est un taux d'actualisation commun aux deux joueurs.

sous la contrainte dynamique commune :

$$\dot{x}(t) = f(x(t), u_1(t), u_2(t)), \quad x(0) = x_0. \tag{14}$$

Chaque agent prend en compte l'impact de ses propres décisions ***et*** de celles de l'autre sur l'évolution de l'état. L'objectif n'est plus de trouver un contrôle optimal

unique, mais un équilibre entre les deux stratégies, tel qu'aucun joueur n'ait intérêt à dévier seul de sa stratégie.
$\Rightarrow$ *Je cherche un équilibre de Nash intertemporel, c'est-à-dire une situation dans laquelle chaque joueur maximise son gain intertemporel en tenant compte des décisions dynamiques de l'autre.*

Je vais maintenant détailler comment se formule ce problème dans l'article étudié, et quelles sont les conditions d'équilibre dans un tel jeu dynamique.

3.2 Un exemple : deux firmes polluantes en interaction

Pour illustrer le cadre d'un jeu différentiel, je reprends le modèle de pollution introduit précédemment, mais cette fois avec deux producteurs. Chacun choisit son niveau d'émission $y_i(t)$ (où $i = 1, 2$), ce qui affecte à la fois ses bénéfices et le niveau de pollution global, $s(t)$.

Le stock de pollution évolue désormais sous l'effet combiné des deux firmes :

$$\dot{s}(t) = y_1(t) + y_2(t) - \delta s(t), \quad s(0) = s_0 \tag{15}$$

Chaque firme maximise ses bénéfices nets, en tenant compte de ses propres émissions et du coût que représente la pollution globale. Je suppose ici que les fonctions objectifs sont symétriques pour simplifier.

Fonction objectif de chaque joueur :

$$J_i(y_1(\cdot), y_2(\cdot)) = \int_0^{\infty} e^{-rt} \left[\alpha y_i(t) - \frac{1}{2} y_i(t)^2 - \frac{1}{2} \gamma s(t)^2 \right] dt \tag{16}$$

Les termes sont interprétés de la même façon qu'en section précédente :

- $\alpha y_i(t)$ représente les bénéfices générés par la production de la firme i,
- $\frac{1}{2} y_i(t)^2$ est le coût de production,
- $\frac{1}{2} \gamma s(t)^2$ est un dommage environnemental commun, supporté également par chaque joueur.

$\Rightarrow$ *Les deux joueurs sont en concurrence dans un environnement partagé. Leur décision influence à la fois leur propre gain et celui de l'autre.*

L'objectif est maintenant de caractériser un équilibre dans ce jeu : une paire de stratégies $(y_1^*(\cdot), y_2^*(\cdot))$ telle que, pour chaque joueur, il n'existe pas d'autre stratégie qui lui permettrait d'améliorer son gain, en tenant l'autre constante.
$\Rightarrow$ *Je cherche un équilibre de Nash intertemporel.*

Dans les sections suivantes, je vais voir comment formuler les conditions d'équilibre à l'aide de la méthode de Pontryagin dans ce contexte à deux joueurs.

3.2.1 Méthode de Pontryagin dans un jeu différentiel

Dans un jeu dynamique, chaque joueur maximise son propre gain intertemporel, mais contrairement au cas d'un joueur unique, il doit ici tenir compte :

- de l'impact de ses propres décisions sur l'évolution de l'état du système (ici le stock de pollution $s(t)$),
- et du fait que cet état est aussi influencé par les décisions d'un autre joueur.

Autrement dit, chaque agent cherche à optimiser sa trajectoire, tout en intégrant la présence d'un autre acteur rationnel, qui agit lui aussi de manière stratégique. C'est ce cadre qui donne naissance à un jeu différentiel, dans lequel les interactions sont dynamiques et continues dans le temps.

$\Rightarrow$ *Chaque joueur résout un problème de contrôle optimal, mais dans un cadre concurrentiel où les décisions de l'autre jouent un rôle.*

Hamiltonien du joueur i : Comme dans le cas précédent, chaque joueur introduit un Hamiltonien pour formaliser son problème d'optimisation dynamique. Je note $\lambda_i(t)$ la variable de co-état associée au joueur i, qui mesure le coût marginal pour lui d'un changement dans le stock de pollution $s(t)$.

Le Hamiltonien du joueur i s'écrit alors :

$$H_i(s, y_1, y_2, \lambda_i) = \alpha y_i - \frac{1}{2} y_i^2 - \frac{1}{2}\gamma s^2 + \lambda_i(y_1 + y_2 - \delta s)$$

Ce Hamiltonien combine :

- le bénéfice net instantané du joueur i (premiers termes),
- et l'effet marginal de ses décisions sur l'évolution future du système, via le co-état λ_i.

$\Rightarrow$ *Chaque joueur introduit son propre Hamiltonien, car il perçoit différemment l'impact de la pollution sur sa trajectoire optimale.*

Conditions d'optimalité de Pontryagin pour chaque joueur En appliquant le principe du maximum à ce jeu, j'obtiens, pour chaque joueur $i = 1, 2$:

- **Optimisation du contrôle (maximisation du Hamiltonien)** :

  $$\frac{\partial H_i}{\partial y_i} = \alpha - y_i + \lambda_i = 0 \quad \Rightarrow \quad y_i(t) = \alpha + \lambda_i(t)$$

  Chaque joueur ajuste donc son niveau de production en fonction de son propre co-état. Ce contrôle est influencé par la manière dont il valorise (ou pénalise) l'impact de la pollution sur ses profits futurs.

- **Équation d'état (commune aux deux joueurs)** :

$$\dot{s}(t) = y_1(t) + y_2(t) - \delta s(t)$$

Cette équation reflète la dynamique du stock de pollution, alimenté par les émissions de chaque joueur, et atténué par le processus naturel de dépollution.

- **Équation de co-état (pour chaque joueur)** :

Je commence par dériver le Hamiltonien H_i par rapport à l'état s :

$$\frac{\partial H_i}{\partial s} = -\gamma s - \delta \lambda_i$$

J'applique ensuite la condition de Pontryagin :

$$\dot{\lambda}_i(t) = r\lambda_i(t) - \frac{\partial H_i}{\partial s} = r\lambda_i(t) + \gamma s(t) + \delta \lambda_i(t) = (r+\delta)\lambda_i(t) + \gamma s(t)$$

Cette équation traduit l'évolution de la valeur marginale du stock de pollution pour chaque joueur : plus le stock de pollution est élevé, plus le "prix futur" de cette pollution s'intensifie pour lui.

$\Rightarrow$ *Chaque joueur suit sa propre dynamique de co-état, mais tous partagent une même dynamique de l'état $s(t)$, ce qui crée une interdépendance stratégique.*

Système global à trois équations En combinant les conditions optimales précédentes, j'obtiens un système dynamique couplé décrivant l'évolution conjointe de l'état $s(t)$ et des co-états des deux joueurs.

En remplaçant les expressions optimales $y_i = \alpha + \lambda_i$ dans la dynamique de l'état, j'obtiens le système complet suivant :

$$\begin{cases} \dot{s}(t) = \alpha + \lambda_1(t) + \alpha + \lambda_2(t) - \delta s(t) \\ \dot{\lambda}_1(t) = (r+\delta)\lambda_1(t) + \gamma s(t) \\ \dot{\lambda}_2(t) = (r+\delta)\lambda_2(t) + \gamma s(t) \end{cases} \tag{17}$$

Ce système à trois équations couplées permet de suivre la dynamique conjointe :

- du stock de pollution $s(t)$, influencé par les deux joueurs,
- et des deux co-états $\lambda_1(t)$ et $\lambda_2(t)$, qui évoluent selon leur perception individuelle du coût marginal de la pollution.

$\Rightarrow$ *Ce système reflète l'interaction stratégique continue entre les deux agents dans un environnement commun.*

3.2.2 État stationnaire du système dynamique

Je cherche maintenant à déterminer la situation d'équilibre vers laquelle le système converge lorsque le temps devient très grand (autrement dit, lorsque $\dot{s}(t) = \dot{\lambda}_1(t) = \dot{\lambda}_2(t) = 0$).

Étape 1 : Poser les équations stationnaires Je pars du système :

$$\begin{cases} 0 = 2\alpha + \lambda_1^* + \lambda_2^* - \delta s^* \\ 0 = (r+\delta)\lambda_1^* + \gamma s^* \\ 0 = (r+\delta)\lambda_2^* + \gamma s^* \end{cases}$$

Étape 2 : Identifier les co-états Les deux dernières équations sont symétriques. Elles donnent :

$$\lambda_1^* = \lambda_2^* = \lambda^* = -\frac{\gamma s^*}{r+\delta}$$

Étape 3 : Injecter dans la première équation Je remplace λ_1^* et λ_2^* dans la première équation :

$$0 = 2\alpha + 2\lambda^* - \delta s^* \quad \Rightarrow \quad 2\lambda^* = \delta s^* - 2\alpha$$

Puis je remplace λ^* par son expression :

$$2\left(-\frac{\gamma s^*}{r+\delta}\right) = \delta s^* - 2\alpha$$

Étape 4 : Résolution de l'équation Je regroupe les termes en s^* :

$$-\frac{2\gamma s^*}{r+\delta} - \delta s^* = -2\alpha \quad \Rightarrow \quad s^*\left(\delta + \frac{2\gamma}{r+\delta}\right) = 2\alpha$$

Je simplifie :

$$s^* = \frac{2\alpha}{\delta + \frac{2\gamma}{r+\delta}}$$

Pour rendre cette expression plus lisible, je mets tout au même dénominateur :

$$s^* = \frac{2\alpha(r+\delta)}{\delta(r+\delta) + 2\gamma}$$

Étape 5 : Expression de la variable de co-état En remplaçant s^* dans l'expression de λ^* :

$$\lambda^* = -\frac{\gamma s^*}{r+\delta} = -\frac{\gamma}{r+\delta} \cdot \frac{2\alpha(r+\delta)}{\delta(r+\delta)+2\gamma} = -\frac{2\alpha\gamma}{\delta(r+\delta)+2\gamma}$$

Conclusion L'état stationnaire du jeu est donné par :

$$s^* = \frac{2\alpha(r+\delta)}{\delta(r+\delta)+2\gamma}, \quad \lambda_1^* = \lambda_2^* = -\frac{2\alpha\gamma}{\delta(r+\delta)+2\gamma} \tag{18}$$

⇒ *L'équilibre dépend du compromis entre bénéfices de production (via α) et dommages environnementaux (via γ), comme dans le cas à un joueur, mais les effets sont ici partagés entre deux agents.*

Interprétation économique :

- Toutes choses égales par ailleurs, le stock de pollution stationnaire s^* est plus élevé que dans le cas d'un seul joueur. En effet, dans un cadre décentralisé, chaque joueur internalise uniquement une partie du coût environnemental global, ce qui aboutit à une pollution plus importante.
- Chaque co-état λ_i^* est négatif et identique ($\lambda_1^* = \lambda_2^*$), ce qui indique que chaque joueur reconnaît un coût marginal associé à la pollution. Toutefois, ce coût est plus faible que dans le cas centralisé, car il ne tient pas compte de l'effet global.
- L'intensité de production optimale de chaque joueur est :

  $$y_i^* = \alpha + \lambda_i^* = \alpha - \frac{2\alpha\gamma}{\delta(r+\delta)+2\gamma}$$

  Cette quantité est plus élevée que dans le cas centralisé, ce qui montre qu'un jeu à plusieurs agents mène à une surproduction et donc à une surexploitation de l'environnement.
- Ce résultat illustre bien le problème classique des externalités environnementales : chaque joueur maximise son propre gain en négligeant l'impact total de ses émissions sur l'environnement, ce qui conduit à un niveau de pollution sous-optimal pour la collectivité.

⇒ *Le jeu dynamique entre les deux agents conduit à un équilibre de Nash sous-optimal du point de vue collectif.*

3.3 Méthode de Bellman dans un jeu à deux joueurs

La méthode de Bellman peut aussi être utilisée dans le cadre d'un jeu dynamique à plusieurs joueurs. L'idée est de construire, pour chaque joueur i, une fonction de valeur $V_i(s)$ représentant le gain maximal qu'il peut espérer, compte tenu du comportement de l'autre joueur.

Chaque joueur choisit sa propre stratégie $y_i(t)$, en tenant compte de l'évolution du stock de pollution $s(t)$, influencé par les deux stratégies. Le système est donc toujours régi par la même dynamique :

$$\dot{s}(t) = y_1(t) + y_2(t) - \delta s(t)$$

mais ici, $y_1(t)$ et $y_2(t)$ sont choisis indépendamment, chacun pour maximiser un objectif individuel.

Je suppose que chaque joueur a une fonction de valeur de la forme $V_i(s)$, et raisonne comme s'il maîtrisait seul l'évolution de $s(t)$, l'action de l'autre étant prise comme donnée. L'équation HJB associée au joueur i est donc :

$$rV_i(s) = \max_{y_i} \left[\alpha y_i - \frac{1}{2} y_i^2 - \frac{1}{2}\gamma s^2 + V_i'(s)(y_i + y_j - \delta s) \right] \tag{19}$$

où j désigne l'autre joueur (donc $j \neq i$), et y_j est considéré comme exogène (variable fixée, non contrôlée par le joueur i) dans la maximisation de V_i.

3.3.1 Étape 1 : Optimisation individuelle

Chaque joueur maximise l'expression à droite par rapport à sa propre variable de contrôle y_i. Je calcule la dérivée première :

$$\frac{\partial}{\partial y_i} \left(\alpha y_i - \frac{1}{2} y_i^2 + V_i'(s)(y_i + y_j - \delta s) \right) = \alpha - y_i + V_i'(s)$$

J'en déduis :

$$y_i^*(s) = \alpha + V_i'(s)$$

3.3.2 Étape 2 : Équation HJB résolue pour chaque joueur

En remplaçant $y_i^*(s)$ dans (19), j'obtiens :

$$rV_i(s) = \alpha(\alpha + V_i') - \frac{1}{2}(\alpha + V_i')^2 - \frac{1}{2}\gamma s^2 + V_i'(s)\left[(\alpha + V_i'(s)) + y_j(s) - \delta s\right]$$

Cette équation dépend encore de $y_j(s)$, la stratégie de l'autre joueur. Je suis donc face à un système couplé : pour chaque joueur, la fonction de valeur optimale dépend du comportement de l'autre.

3.3.3 Étape 3 : Équilibre de Nash stationnaire

À l'équilibre, chaque joueur anticipe correctement le comportement de l'autre. Je cherche donc une situation où les deux joueurs adoptent la même stratégie fonctionnelle :

$$V_1'(s) = V_2'(s) = V'(s), \quad y_1(s) = y_2(s) = y(s)$$

et donc :

$$y(s) = \alpha + V'(s)$$

Le stock évolue alors selon :

$$\dot{s}(t) = 2y(s) - \delta s = 2(\alpha + V'(s)) - \delta s$$

3.3.4 Étape 4 : Forme quadratique de la fonction de valeur

Comme précédemment, je suppose que :

$$V(s) = as^2 + bs + c \quad \Rightarrow \quad V'(s) = 2as + b$$

J'injecte cette forme dans l'équation HJB résolue, en utilisant $y_j(s) = y(s) = \alpha + V'(s)$ pour obtenir une équation polynomiale, puis j'identifie les coefficients.

Les calculs (détaillés dans l'article) donnent les valeurs suivantes :

$$\begin{aligned} a &= \frac{r + 2\delta - \sqrt{(r+2\delta)^2 + 8\gamma}}{4} \\ b &= \frac{2a\alpha}{r + \delta - 2a} \\ c &= \frac{(\alpha + b)^2}{2r} \end{aligned}$$

3.3.5 Étape 5 : Stratégie optimale et trajectoire

La stratégie individuelle devient :

$$y_i^*(s) = \alpha + 2as + b$$

Et la dynamique du stock suit :

$$\dot{s}(t) = 2(\alpha + 2as + b) - \delta s = (4a - \delta)s + 2(\alpha + b)$$

qui converge vers un état stationnaire :

$$s^* = \frac{2(\alpha + b)}{\delta - 4a}$$

Interprétation économique :

- La stratégie optimale reste une fonction croissante de s : plus la pollution est forte, plus chaque joueur réduit sa production.
- Comme dans la méthode de Pontryagin, cette approche aboutit à une stratégie en feedback, adaptée à l'état courant.

- L'équilibre de Nash ici correspond au point où chaque joueur adopte une politique optimale, compte tenu du comportement anticipé de l'autre.
- On observe encore une pollution excessive par rapport au cas centralisé, illustrant le manque de coordination entre les agents.

$\Rightarrow$ *La programmation dynamique permet d'identifier un équilibre de Nash stationnaire en feedback, cohérent avec la logique stratégique du jeu à deux joueurs.*

Comparaison avec le cas coopératif Quand un seul agent ou une autorité centrale contrôle la pollution, il prend en compte tous les effets de ses décisions, ce qui permet d'atteindre un niveau de pollution plus faible et mieux pour la société. En revanche, quand il y a plusieurs agents, chacun cherche à maximiser son propre intérêt sans vraiment se soucier des conséquences pour l'autre. Cela conduit à une pollution plus importante que dans le cas centralisé.

Économiquement, l'équilibre entre les deux joueurs n'est pas efficace : il serait possible de faire mieux pour tout le monde en coopérant et en réduisant les émissions. Mais chacun a intérêt à produire un peu plus que ce qui serait "idéal", car il supporte seulement une partie du coût de la pollution totale.

Même si les deux agents décidaient de coopérer, il serait difficile de respecter cet accord sur la durée. Dans un jeu dynamique sans autorité pour surveiller, chacun a toujours une tentation de tricher pour obtenir plus de gains immédiats. C'est ce qu'on appelle un problème de crédibilité : sans mécanisme pour forcer le respect de l'accord, la coopération n'est pas stable à long terme.

4 Lien avec les Mean Field Games

La section précédente a permis de passer d'un problème de contrôle optimal avec un seul agent à un jeu dynamique à deux joueurs. Mais dans de nombreux contextes économiques, les interactions stratégiques ne se limitent pas à deux individus : elles impliquent un grand nombre de participants. C'est là qu'interviennent les Mean Field Games (MFG), ou "jeux à champ moyen".

4.1 Pourquoi introduire les Mean Field Games ?

Les jeux différentiels classiques deviennent très complexes dès qu'on augmente le nombre de joueurs. Lorsque chaque agent influence et est influencé par tous les autres, il devient pratiquement impossible de résoudre le système en explicitant chaque interaction.

Pour contourner cette difficulté, les Mean Field Games proposent une idée simple mais puissante : lorsqu'il y a une infinité de joueurs identiques et symétriques, chacun réagit à l'état moyen du groupe (le champ moyen), plutôt qu'à chaque individu pris isolément.

$\Rightarrow$ *Je passe d'un jeu à N joueurs à une analyse en population continue.*

Autrement dit, les Mean Field Games peuvent être vus comme la limite d'un jeu différentiel classique lorsque le nombre de joueurs N devient très grand ($N \to \infty$). Cette approche permet de simplifier les interactions tout en capturant l'effet agrégé du comportement individuel.

Ce cadre permet de modéliser des interactions stratégiques dans de grands ensembles (par exemple : firmes en concurrence, agents économiques, consommateurs), tout en gardant une structure mathématique exploitable.

4.2 Logique d'un MFG en lien avec mon modèle

Dans mon exemple, chaque joueur choisit un niveau d'émission $y_i(t)$ qui affecte un stock global de pollution $s(t)$, partagé par tous. Lorsque le nombre de joueurs devient très grand (tendance vers l'infini), l'influence individuelle d'un agent donné sur l'évolution du système devient négligeable. Toutefois, la somme des comportements individuels continue de produire un effet collectif significatif sur l'état du système.

Dans ce contexte, je modélise alors le comportement d'un agent représentatif, en raisonnant comme s'il interagissait non plus avec des agents identifiables, mais avec une moyenne agrégée de la population. Cette approche donne naissance à ce qu'on appelle un mean field game (MFG), ou jeu à champ moyen.

$\Rightarrow$ *Chaque agent prend comme donnée une trajectoire anticipée du champ moyen (ici, la pollution moyenne $s(t)$) et optimise son propre comportement en réponse à cette évolution.*

Concrètement, chaque agent résout un problème de contrôle optimal classique (comme dans la méthode de Pontryagin ou Bellman), mais dans lequel l'état du système dépend d'une trajectoire moyenne exogène $s(t)$. Une fois l'ensemble des stratégies individuelles déterminées, je vérifie que la pollution moyenne résultante coïncide bien avec la trajectoire anticipée : c'est la condition d'équilibre d'un MFG.
$\Rightarrow$ *Un MFG correspond donc à une situation d'équilibre où les anticipations agrégées des agents sont vérifiées ex post par le comportement collectif réel.*

4.3 Structure mathématique d'un Mean Field Game

Le cadre des jeux à champ moyen (Mean Field Games), bien qu'au-delà du périmètre strict de l'article étudié, constitue une extension naturelle du raisonnement stratégique à un grand nombre d'agents. Je peux alors décrire l'équilibre à l'aide d'un système couplé d'équations de Hamilton-Jacobi-Bellman et de Kolmogorov, comme proposé dans Cardaliaguet et al. (2019)[1]. Un mean field game est formellement défini par un système couplé de deux équations :

[1]Cardaliaguet, P., Delarue, F., Lasry, J.-M., & Lions, P.-L. (2019). *The Master Equation and the Convergence Problem in Mean Field Games.* Princeton University Press.

- L'**équation de Hamilton-Jacobi-Bellman (HJB)**, qui caractérise le comportement optimal de l'agent représentatif face à une dynamique moyenne supposée donnée.

- L'**équation de Kolmogorov (ou Fokker-Planck)**, qui décrit comment évolue dans le temps la distribution des agents dans l'espace des états, selon leurs décisions optimales.

Ce système s'écrit généralement comme :

$$\text{(HJB)} \quad \begin{cases} rV(s,t) = \max_y \left[\alpha y - \frac{1}{2}y^2 - \frac{1}{2}\gamma s^2 + \partial_s V(s,t)(y - \delta s)\right] \\ \quad \text{avec } y^*(s,t) = \alpha + \partial_s V(s,t) \end{cases} \tag{20}$$

$$\text{(Kolmogorov)} \quad \partial_t m(s,t) + \partial_s \left[m(s,t)(y^*(s,t) - \delta s)\right] = 0 \tag{21}$$

où :

- $V(s,t)$ est la fonction de valeur à l'instant t pour un agent dans l'état s ;

- $m(s,t)$ est la densité d'agents dans l'état s à l'instant t ;

- $y^*(s,t)$ est la stratégie optimale des agents, dépendant de l'état et du temps.

⇒ *L'équilibre d'un MFG est atteint lorsque la stratégie optimale y^* (issue de l'HJB) entraîne une évolution $m(s,t)$ (via Kolmogorov) cohérente avec ce qui est anticipé.*

4.4 Application à mon modèle de pollution

Dans mon modèle, un agent unique choisit une trajectoire d'émission $y(t)$, en réponse à une trajectoire anticipée du stock moyen de pollution $s(t)$. Je suppose que :

- La fonction objectif de chaque agent est identique à celle vue précédemment : maximisation d'un profit actualisé.

- L'évolution de la pollution est déterminée par la somme des émissions des agents, soit par la distribution moyenne $m(s,t)$.

Le problème se reformule alors comme suit :

$$\max_{y(\cdot)} \int_0^\infty e^{-rt} \left[\alpha y(t) - \frac{1}{2}y(t)^2 - \frac{1}{2}\gamma s(t)^2\right] dt \tag{22}$$

avec la dynamique anticipée :

$$\dot{s}(t) = \bar{y}(t) - \delta s(t) \quad \text{où } \bar{y}(t) = \int y^*(s,t)m(s,t)ds \tag{23}$$

$\Rightarrow$ *Chaque agent anticipe une pollution moyenne $s(t)$ et optimise son émission $y(t)$ en conséquence. La trajectoire réelle du champ moyen doit coïncider avec la moyenne agrégée des comportements optimaux.*

C'est cette correspondance entre les anticipations individuelles et le comportement collectif qui définit l'équilibre d'un Mean Field Game : les trajectoires anticipées de pollution s'auto-réalisent.

4.5 Conclusion sur le lien avec les MFG

Le modèle de pollution que j'ai étudié peut être vu comme un cas particulier d'un mean field game lorsque le nombre d'agents devient grand. Les outils mathématiques utilisés (HJB, stratégies feedback, trajectoires dynamiques) restent similaires, mais s'appliquent désormais à une population continue d'agents.

Les mean field games permettent ainsi de :

- généraliser les jeux différentiels à $N \to \infty$ joueurs,
- modéliser des externalités globales comme la pollution ou la congestion,
- conserver une résolution analytique ou numérique en exploitant des hypothèses de symétrie.

Remarque : Bien que mon modèle reste déterministe, les Mean Field Games peuvent également intégrer de l'incertitude via des équations de Fokker–Planck stochastiques, ce qui les rend adaptés à des contextes plus réalistes impliquant du bruit ou des chocs aléatoires.
$\Rightarrow$ *Les MFG offrent une approche robuste pour analyser des dynamiques collectives issues de comportements stratégiques individuels, comme dans mon problème de pollution à grande échelle.*

5 Conclusion

Pour conclure, ce travail illustre l'apport des méthodes mathématiques dans l'analyse de comportements économiques dynamiques. Que ce soit pour une décision individuelle, une interaction stratégique ou une dynamique collective, les outils de contrôle optimal et de théorie des jeux permettent de formaliser des raisonnements économiques complexes, tout en mettant en évidence les effets d'interdépendance et les enjeux d'optimisation dans le temps.

Dans le cas d'un agent unique, j'ai mis en évidence que les méthodes de Pontryagin et de Bellman conduisent, malgré des approches différentes, à une même stratégie optimale. La pollution tend vers un niveau stationnaire qui reflète un arbitrage entre bénéfices de production et dommages environnementaux, et l'agent adapte sa production en fonction de l'état du système.

Dans le cas à deux joueurs, le passage à un jeu différentiel révèle une dynamique stratégique plus complexe. Chaque joueur optimise son comportement en tenant compte de l'autre, ce qui conduit à un équilibre de Nash intertemporel. Ce nouvel équilibre entraîne une pollution plus élevée qu'en situation centralisée, mettant en évidence le problème d'externalité négative et de coordination.

Enfin, les Mean Field Games permettent de prolonger ce raisonnement à une infinité d'agents. Ils capturent l'effet collectif agrégé tout en restant analysables grâce à une modélisation continue. On y retrouve les mêmes logiques : stratégie en feedback, équilibre stationnaire, et influence de l'anticipation sur les comportements individuels.

Cela dit, ce travail repose sur plusieurs hypothèses fortes : agents identiques et parfaitement rationnels, dynamiques déterministes, absence d'incertitude ou d'asymétrie d'information. Ces simplifications facilitent l'analyse mais peuvent limiter l'applicabilité directe à des situations réelles, où les comportements sont plus hétérogènes et les chocs plus imprévisibles.

De plus, les modèles étudiés supposent que les agents peuvent calculer et suivre des stratégies optimales complexes sans difficulté, ce qui peut être irréaliste en pratique. En réalité, les limites cognitives et les imperfections d'information rendent souvent difficile l'application précise de telles stratégies.

En pratique, certains problèmes de contrôle optimal ou de jeu différentiel deviennent aussi très complexes à résoudre numériquement, surtout dans des cadres plus généraux (non linéaires ou stochastiques). Cela rend nécessaire le recours à des méthodes d'approximation ou à des simulations.

Ces éléments ouvrent des perspectives d'extension intéressantes : par exemple, intégrer de l'incertitude explicite dans la dynamique (modèles stochastiques), ou introduire des mécanismes d'apprentissage par renforcement pour représenter des agents qui adaptent progressivement leur comportement sans résoudre exactement leur problème optimal.

References

[1] de Zeeuw, A. (2024). *Dynamic Game Models in Environmental Economics*. Document de travail, version de janvier 2024.

[2] Cardaliaguet, P., Delarue, F., Lasry, J.-M., & Lions, P.-L. (2019). *The Master Equation and the Convergence Problem in Mean Field Games*. Princeton University Press.