HAL
open science

Some links between discrete and continuous aspects in dynamic games

Juan Pablo Maldonado Lopez

To cite this version:

Juan Pablo Maldonado Lopez. Some links between discrete and continuous aspects in dynamic games. General Mathematics [math.GM]. Université Pierre et Marie Curie - Paris VI, 2014. English. NNT : 2014PA066271 . tel-01164995

HAL Id: tel-01164995

https://theses.hal.science/tel-01164995v1

Submitted on 18 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

École Doctorale de Sciences Mathématiques de Paris Centre

Thèse de doctorat

Discipline : Mathématiques Apliquées

présentée par

Juan Pablo MALDONADO LOPEZ

Some links between discrete and continuous aspects in dynamic games

---

dirigée par Sylvain SORIN

Soutenue le 4 novembre 2014 devant le jury composé de :

M. Sylvain SORIN	Université Pierre et Marie Curie	directeur
M. Rida LARAKI	Université Paris Dauphine	examinateur
M. Tristan TOMALA	HEC Paris	examinateur
M. Jean-Paul ALLOUCHE	Université Pierre et Marie Curie	examinateur
M. Stéphane GAUBERT	INRIA-Saclay	examinateur
M. Nicolas VIEILLE	HEC Paris	rapporteur
M. Martino BARDI	Università degli Studi di Padova	rapporteur
M. Pierre CARDALIAGUET	Université Paris Dauphine	président

Institut de Mathématiques de Jussieu
175, rue du Chevaleret
75 013 Paris

École Doctorale de Sciences Mathématiques de Paris Centre Case 188
4 place Jussieu
75 252 Paris cedex 05

Résumé

Résumé

Cette thèse étudie les liens entre a) les jeux en temps discret et continu, et b) les jeux à très grand nombre de joueurs identiques et les jeux avec un continuum de joueurs. Une motivation pour ces sujets ainsi que les contributions principales de cette thèse sont présentées dans le Chapitre 1. Le reste de la thèse est organisé en trois parties. La Partie I étudie les jeux différentiels à somme nulle et à deux joueurs. Nous décrivons dans le Chapitre 3 trois approches qui ont été proposées dans la littérature pour établir l'existence de la valeur dans les jeux différentiels à deux joueurs et à somme nulle, en soulignant les liens qui existent entre elles. Nous fournissons dans le Chapitre 4 une démonstration de l'existence de la valeur à l'aide d'une description explicite des stratégies ϵ-optimales. Le Chapitre 5 établit l'équivalence entre les solutions de minimax et les solutions de viscosité pour les équations de Hamilton-Jacobi-Isaacs. La Partie II porte sur les jeux à champ moyen en temps discret. L'espace d'action est supposé compact dans le Chapitre 6, et fini dans le Chapitre 7. Dans les deux cas, nous obtenons l'existence d'un ϵ- équilibre de Nash pour un jeu stochastique avec un nombre fini de joueurs identiques, où le terme d'approximation tend vers zéro lorsque le nombre de joueurs augmente. Nous obtenons dans le Chapitre 7 des bornes d'erreur explicites, ainsi que l'existence d'un ϵ-équilibre de Nash pour un jeu stochastique à durée d'étape évanescente et à un nombre fini de joueurs identiques. Dans ce cas, le terme d'approximation est fonction à la fois du nombre de joueurs et de la durée d'étape. Enfin, la Partie III porte sur les jeux stochastiques à durée d'étape évanescente, qui sont décrits dans le Chapitre 8. Il s'agit de jeux où un paramètre évolue selon une chaîne de Markov en temps continu, tandis que les joueurs choisissent leurs actions à des dates discrètes. La dynamique en temps continu dépend des actions des joueurs. Nous considérons trois évaluations différentes pour le paiement et deux structures d'information : dans un cas, les joueurs observent les actions passées et le paramètre, et dans l'autre, seules les actions passées sont observées.

Mots-clefs

Jeux dynamiques à somme nulle, jeux différentiels à somme nulle, jeux à champ moyen en temps discret, jeux stochastiques à étape evanescente

Some links between discrete and continuous in dynamic games

Abstract

In this thesis we describe some links between a) discrete and continuous time games and b) games with finitely many players and games with a continuum of players. A motivation to the subject and the main contributions are outlined in Chapter 2. The rest of the thesis is organized in three parts: Part I is devoted to differential games, describing the different approaches for establishing the existence of the value of two player, zero sum differential games in Chapter 3 and pointing out connections between them. In Chapter 4 we provide a proof of the existence of the value using an explicit description of ϵ-optimal strategies and a proof of the equivalence of minimax solutions and viscosity solutions for Hamilton-Jacobi-Isaacs equations in Chapter 5. Part II concerns discrete time mean field games. We study two models with different assumptions, in particular, in Chapter 6 we consider a compact action space while in Chapter 7 the action space is finite. In both cases we derive the existence of an ϵ-Nash equilibrium for a stochastic game with finitely many identical players, where the approximation error vanishes as the number of players increases. We obtain explicit error bounds in Chapter 7 where we also obtain the existence of an ϵ-Nash equilibrium for a stochastic game with short stage duration and finitely many identical players, with the approximation error depending both on the number of players and the duration of the stage. Part III is concerned with two player, zero sum stochastic games with short stage duration, described in Chapter 8. These are games where a parameter evolves following a continuous time Markov chain, while the players choose their actions at the nodes of a given partition of the positive real axis. The continuous time dynamics of the parameter depends on the actions of the players. We consider three different evaluations for the payoff and two different information structures: when players observe the past actions and the parameter and when players observe past actions but not the parameter.

Keywords

Zero sum dynamic games, zero sum differential games, discrete time mean field games, short stage stochastic games.

Contents

1 Introduction **1**
1.1 Thématiques abordées dans cette thèse . . . 1
1.2 Jeux différentiels . . . 1
1.3 Jeux à champ moyen en temps discret . . . 4
1.4 Jeux stochastiques à étape courte . . . 5
1.5 Nos contributions principales . . . 7

2 Introduction **9**
2.1 Motivation and examples . . . 9
2.2 Contents of this thesis . . . 10
2.3 Differential games . . . 12
2.4 Discrete time mean field games . . . 20
2.5 Zero-sum stochastic games with short-stage duration . . . 23
2.6 Short stage stochastic games . . . 25
2.7 Main contributions . . . 29

I Differential games **31**

3 Value of zero-sum differential games **33**
3.1 Introduction . . . 33
3.2 The discrete game approach . . . 38
3.3 The viscosity approach . . . 42
3.4 The strategic approach . . . 45
3.5 Some links between these approaches . . . 47

4 A strategy-based proof of the existence of the value in zero-sum differential games **51**
4.1 Comparison of trajectories . . . 51
4.2 Differential Games . . . 55
4.3 Appendix . . . 59

5 Generalized solutions of HJI equations **61**
5.1 The HJI equations . . . 61
5.2 Viscosity solutions . . . 63
5.3 Proximal solutions . . . 64
5.4 Minimax solutions . . . 65
5.5 The equivalence . . . 67

II Discrete time mean field games **71**

6 Discrete time mean field games **73**
6.1 Introduction 73
6.2 The mean field game equilibrium 74
6.3 The game with a continuum of players 76
6.4 The game with finitely many players 77

7 Discrete time mean field games: The short-stage limit **83**
7.1 Introduction 83
7.2 The discrete time model 84
7.3 A mean field game with frequent actions 90
7.4 Concluding remarks 93

III Stochastic games with frequent actions **97**

8 Stochastic zero-sum games with a continuous time dynamics **99**
8.1 Description of the general model 99
8.2 Evaluation of the payoff 101
8.3 No information on the state 104
8.4 Standard signalling 110
8.5 Some concluding remarks 115

Bibliography **123**

Chapitre 1

Introduction

1.1 Thématiques abordées dans cette thèse

Cette thèse porte principalement sur les jeux répétés (Partie I et III) à somme nulle et à deux joueurs. Dans ces jeux, les intérêts des joueurs sont opposés : le gain d'un joueur est la perte de l'autre.

Même dans ce cadre particulier, la théorie est assez riche et on voit intervenir des outils mathématiques très variés.

Une autre thématique qui nous intéresse, dans la Partie II, correspond aux jeux avec un très grand nombre des joueurs identiques, au sens où ils ont les mêmes fonctions de paiement et la même dynamique. Intuitivement, plus il y a de joueurs, plus l'analyse du jeu devient compliquée. Néanmoins, si les joueurs sont identiques, on peut controller cette complexité dans un terme dit de **champ moyen** qui sera défini plus tard.

Nous introduisons maintenant de façon plus précise les thématiques étudiés et les principales contributions.

1.2 Jeux différentiels

Soit $(t_1, x_1) \in [0,1] \times \mathbb{R}^n$. Soient U et V deux sous-ensembles compacts d'un espace euclidien.

On définit

$$\mathcal{U}(t_1) = \{\mathbf{u} : [t_1, 1] \to U, \text{ mesurable}\}, \quad \mathcal{V}(t_1) = \{\mathbf{v} : [t_1, 1] \to V, \text{ mesurable}\}.$$

Si $t_1 = 0$, ces ensembles seront notés $\mathcal{U}$ et $\mathcal{V}$, respectivement.

Les ensembles $\mathcal{U}(t_1), \mathcal{V}(t_1)$ sont les ensembles des **fonctions de contrôle**. Les éléments de U, V sont dits **contrôles** ou **actions**.

Soit $f : \mathbb{R}^n \times U \times V \to \mathbb{R}^n$ et $(\mathbf{u}, \mathbf{v}) \in \mathcal{U}(t_1) \times \mathcal{V}(t_1)$ une couple de fonctions de contrôle. On considère l'équation différentielle ordinaire (EDO) suivante

$$\mathbf{x}(t_1) = x_1, \quad \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t)) \quad \text{p.s. sur } [t_1, 1]. \tag{1.1}$$

On fait l'hypothèse suivante sur f, pour que la trajectoire de l'EDO soit bien définie :

Assumption 1.2.1. *On suppose que la fonction f est continue, bornée, et qu'il existe $c > 0$ tel que pour tout $(u, v) \in U \times V$ et $x, y \in \mathbb{R}^n$:*

$$\|f(x, u, v) - f(y, u, v)\| \leq c\|x - y\|.$$

On pose $\|f\| := \sup_{(x,u,v)} \|f(x, u, v)\| < +\infty$.

Avec cette hypothèse, on utilise le théorème de Carathéodory, [31, Chapter 2] pour déduire que l'EDO (1.1) possède une unique solution. L'évaluation de cette solution au temps s est noté par $\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](s)$ et est interprétée au sens étendu suivant : pour tout $t \in [t_1, 1]$,

$$\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](t) := x_1 + \int_{t_1}^{t} f(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](s), \mathbf{u}(s), \mathbf{v}(s))ds.$$

Cela définit la **dynamique**. On pourrait aussi considérer l'intervalle $[0, +\infty)$ pour définir la dynamique.

Pour bien spécifier un **jeu différentiel**, il faut en définir les objectifs et l'information et les stratégies de chaque joueur. Intuitivement, le joueur 1 choisit $\mathbf{u}$ et le joueur 2 $\mathbf{v}$ pour atteindre un **objectif quantitatif** ou un **objectif qualitatif**, qui nous allons spécifier tout de suite. Cette distinction entre objectifs quantitatifs et qualitatifs a déjà été faite par Isaacs [58], qui introduit les termes *games of kind* pour les jeux où l'objectif est qualitatif, et *games of degree* pour ceux où l'objectif est quantitatif.

On suppose que l'information est **complète**, ce qui veut dire que les joueurs connaissent touts les paramètres du jeu : état initial, dynamique, paiement et sa description.

Cas qualitatif

Pour le cas qualitatif, on considère le **jeu de cible** : l'objectif du joueur 1 est de faire que la variable d'état atteigne l'ensemble fermé M, dit cible, à la date $t = 1$, et l'objectif du joueur 2 est de l'en empêcher. On note ce jeu $\Gamma_M(t_1, x_1)$.

On se pose les questions suivantes :

Question 1. *1. Pour une condition initiale $(s, y) \in [t_1, 1] \times \mathbb{R}^n$ donnée, peut-on décider quel joueur a une stratégie gagnante ?*
2. Construire des stratégies spécifiques.

On reformule la première question de la manière suivante :

Problem 1. *Construire une partition de $[t_1, 1] \times \mathbb{R}^n$ en deux ensembles K_1, K_2, satisfaisant :*
i) Si $(1, x) \in K_1$, alors $x \in M$.
ii) Pour tout $(s, y) \in K_1$, il existe une stratégie du joueur 1 telle que la trajectoire induite reste sur K_1.
iii) Pour tout $(s, y) \in K_2$ il existe une stratégie du joueur 2 telle que la trajectoire induite n'atteigne pas l'ensemble cible M à la date $t = 1$.

Un théorème qui permet établir une telle caractérisation est un **théorème d'alternative**.

Nous n'étudions pas les jeux qualitatifs en détail, mais nous voudrions faire quelques remarques. Une complication importante dans les jeux en temps continu est qu'il n'existe pas une structure canonique d'information. Nos décrivons quelques exemples de structures d'information étudiées dans la littérature.

Un des premiers théorèmes d'alternative a été démontré par B.N. Pšeničnyj [79], qui a étudié le jeu de cible avec la classe de stratégies suivantes :

Definition 1.2.2. (ϵ-stratégies) On dit que les joueurs utilisent des **ϵ-stratégies** dans le jeu de cible si le jeu est joué de la façon suivante :
i) Les deux joueurs connaissent (t_1, x_1).
ii) Le joueur 2 choisit $\epsilon_1 > 0$ et une fonction de contrôle $\mathbf{v}_1$ qui sera joué dans l'intervalle $[t_1, t_1 + \epsilon_1]$ et informe le joueur 1 de son choix.

iii) A partir de cette information, le joueur 1 choisit sa fonction de contrôle.
iv) Au temps $t_1 + \epsilon_1$, le nouvel état est annoncé. La situation est répétée : le joueur 2 choisit ϵ_2, etc.

Avec cette structure d'information, plusieurs théorèmes d'alternative sont démontrés dans [79], sur des hypothèses différentes pour la dynamique et l'ensemble cible. Cependant, aucun lien avec le cas quantitatif n'est établi.

Krasovskii et Subbotin ont introduit la méthode d'**extremal aiming** [61] pour les jeux de cible. Cette méthode motive les résultats du Chapitre 4. Ils utilisent la notion de **stratégies positionnelles**, qui sont des limites de fonctions constantes par morceaux. En général, les fonctions de contrôle ainsi obtenues ne sont pas suffisamment régulières pour avoir une trajectoire bien définie, même au sens de Carathéodory. Donc, comme dans l'approache de Pšeničnyj, l'extremal aiming nous donne de l'information sur un jeu approximé.

Un théorème d'alternative plus récent a été proposé par Cardaliaguet [21] qui considère les stratégies non anticipatives, qui seront introduites dans le Chapitre 3. Ce résultat est important car il nous permet de résoudre le jeu de façon exacte.

Cas quantitatif

Soient $\ell : \mathbb{R}^n \times U \times V \to [0,1]$ et $g : \mathbb{R}^n \to [0,1]$ deux fonctions qui représentent respectivement un **paiement courant** et un **paiement terminal**. Pour le cas quantitatif, on peut considérer les évaluations de paiement suivantes :

1. Le jeu **escompté à l'horizon infini** : pour une histoire $(\mathbf{x}, \mathbf{u}, \mathbf{v})$, le paiement que le joueur 1 reçoit du joueur 2 est :
$$\int_{t_1}^{\infty} e^{-\rho s} \ell(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](s), \mathbf{u}(s), \mathbf{v}(s)) ds$$
avec $\rho > 0$.
2. Le jeu **à horizon fini** : à la date $t = 1$, le joueur 2 paie au joueur 1 :
$$\int_{t_1}^{1} \ell(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](s), \mathbf{u}(s), \mathbf{v}(s) ds + g(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](1)).$$

Ces jeux sont respectivement notés $\Gamma_\rho(t_1, x_1)$ et $\Gamma(t_1, x_1)$.

Pour résoudre un jeu quantitatif, il faut répondre aux questions suivantes :

Question 2. *1. Spécifier les conditions sur lesquelles on a existence et unicité de la fonction valeur et sa caractérisation.*
2. Donner des stratégies ϵ-optimales.

Comme dans le jeu de cible, on doit d'abord spécifier une structure d'information. Plusieurs structures d'information ont été proposées dans la littérature, voir Bardi et Capuzzo-Dolcetta [9, Chapter VIII].

Pour le jeu à l'horizon fini avec paiement courant $\ell \equiv 0$, on peut déduire de façon heuristique [58, Section 4.2] que le maxmin et le minmax sont des solutions des EDP suivantes

$$\frac{\partial w^-}{\partial t}(t,x) + \sup_{u \in U} \inf_{v \in V} \langle f(x,u,v), \nabla_x w^-(t,x) \rangle = 0 \tag{1.2a}$$

$$\frac{\partial w^+}{\partial t}(t,x) + \inf_{v \in V} \sup_{u \in U} \left\langle f(x,u,v), \nabla_x w^+(t,x) \right\rangle = 0 \tag{1.2b}$$

avec les conditions au bord $w^-(1,x) = w^+(1,x) = g(x)$.

Cette déduction heuristique a été menée par Isaacs [58, Section 4.2]. Le lien entre EDP et jeux différentiels a été explicité dans le cadre des solutions de viscosité [33] par Evans et Souganidis [36]. La notion de solution de viscosité a été introduite par Crandall et Lions [33], voir aussi le livre de Lions [67].

Si, de plus, la **condition d'Isaacs** est satisfaite, i.e. si on a l'égalité suivante,

$$\sup_{u\in U}\inf_{v\in V}\langle f(x,u,v),p\rangle = \inf_{v\in V}\sup_{u\in U}\langle f(x,u,v),p\rangle$$

pour tous $x,p\in\mathbb{R}^n$, on n'a qu'une seule équation, dite **l'équation d'Hamilton-Jacobi-Isaacs**.

En utilisant la méthode d'extremal aiming pour un certain jeu de cible, Krasovskii et Subbotin montrent l'existence et l'unicité de la valeur pour le jeu à horizon fini. Dans leur preuve on obtient une description explicite des stratégies ϵ-optimales. Le Chapitre 4 est inspiré de cette construction. Plus tard, Subbotin [97] propose une notion de solutions généralisées, les **solutions de minimax** qui permettent caractériser la valeur comme l'unique solution minimax de l'équation HJI. On montre l'équivalence des solutions de minimax avec les solutions de viscosité dans le Chapitre 5.

1.3 Jeux à champ moyen en temps discret

Les jeux à champ moyen en temps continu ont été introduits indépendamment par Huang, Caines et Malhamé [56, 57] et Lasry and Lions [64, 65, 66]. Le but de cette théorie est la modélisation de situations stratégiques avec un grand nombre des joueurs identiques et petits, au sens que l'influence d'un seul joueur sur les autres est négligeable.

Les jeux avec un continuum des joueurs ont déjà été étudies dans plusieurs contextes, notamment en économie, par Aumann [6], dans les jeux de congestion par Wardrop [106], et dans les jeux de population par Maynard Smith [73] et Maynard Smith et Price [74]. Ce qui est différent dans les jeux à champ moyen est l'aspect dynamique.

Les jeux à champ moyen ont une structure dite de **backward-forward**, qui est de façon intuitive l'idée suivante : chaque joueur "anticipe" un certain comportement moyen des autres dans un intervalle de temps et calcule son propre comportement optimal en prenant le comportement des autres comme un paramètre fixe. Donc, chaque joueur fait face à un problème de contrôle optimal. Si le comportement moyen des joueurs qui est induit par cette optimisation est le même que celui qui a été prédit, alors on dit que les joueurs sont dans un équilibre de champ moyen. On introduira des définitions précises dans le Chapitre 6.

Prenons l'exemple suivant, qu'on peut trouver dans les notes de Cardaliaguet [20] sur le cours de Lions au Collège de France.

Example 1.3.1. *On considère N joueurs dans $\mathbb{R}^d$. La position du joueur i a la date t est donnée par*

$$dX_t^i = \alpha_t^i dt + \sqrt{2}dB_t^i$$

Chaque joueur i minimise son coût :

$$\int_t^T \frac{1}{2}|\alpha_s^i|^2 + F(m_{s,N}^{-i})ds + g(x_T^i, m_{T,N}^{-i})$$

où $m_{s,N}^{-i} := \frac{1}{N-1}\sum_{j\neq i}\delta_{x_s^j}$.

De façon heuristique, si on prend la limite quand $N \to +\infty$, on obtient le système d'EDP suivant :

$$\begin{aligned}
-\frac{\partial u}{\partial t} - \Delta u + \frac{1}{2}|D_x u|^2 &= F(x,m) \ in \, \mathbb{R}^d \times (0,T); \\
\frac{\partial m}{\partial t} - \Delta m + \operatorname{div}(mDu) &= 0 \ in \, \mathbb{R}^d \times (0,T); \\
m(0) &= m_0; \\
u(x,T) &= g(x, m(T)) \ in \, \mathbb{R}^d.
\end{aligned}$$

Ici, dans la première équation, u dénote la valeur du problème de contrôle optimal pour un joueur quelconque si la distribution des joueurs est donnée par m. La deuxième est une équation de Kolmogorov qui décrit l'évolution de la distribution des joueurs dans $\mathbb{R}^d$.

Une motivation importante pour l'étude des jeux à champ moyen dans les applications est l'obtention des ϵ-équilibres de Nash dans les jeux à N joueurs, avec un terme d'approximation qui tend vers zero quand N tend vers l'infini.

Les jeux à champ moyen ont trouvé des applications, notamment dans certains problèmes en économie, voir Guéant, Lasry et Lions [49]. On fait aussi référence au survey de Gomes et Saude [47] pour une collection de résultats récents et au livre de Bensoussan, Frehse et Yam [12] pour les liens avec la théorie du contrôle optimal de champ moyen.

La plupart de la littérature étudie les jeux à champ moyen en temps continu. Une exception importante est l'article de Gomes, Mohr et Souza [48], qui étude le comportement asymptotique d'un jeu à l'horizon fini quand l'horizon tend vers l'infini d'un jeu avec un continuum de joueurs en temps discret.

Par contre, nous considérons un horizon fini fixe et nous proposons une construction d'un équilibre de Nash approximé pour un jeu à N joueurs. Le modèle que l'on étude dans le Chapitre 6 est l'analogue en temps fini du jeu étudié par Adlakha, Johari et Weintraub [2].

Nous nous intéressons aussi aux situations où les joueurs interagissent "fréquemment". Pour donner un sens mathématique à cette expression, il faut introduire un temps exogène, disons $\mathbb{R}_+$. Ici, chaque joueur observe et contrôle une chaîne de Markov en temps continu dont le générateur infinitésimal dépend du comportement moyen des autres. Les joueurs choisissent leurs actions aux instants de temps discrets, données par une partition de $\mathbb{R}_+$. Nous décrivons ces modèles dans le cas à deux joueurs et somme nulle dans la Section suivante. L'analogue pour les jeux à champ moyen est introduit dans le Chapitre 7.

1.4 Jeux stochastiques à étape courte

Dans les jeux stochastiques en temps discret, il n'existe pas de notion de "durée" des étapes du jeu. Pour en introduire une, on considère un temps exogène, qui sera représenté par les nombres réels positifs, $\mathbb{R}_+$.

Cela nous permet de donner une définition de "durée" de la façon suivante : Soit $\Pi = \{t_1, t_2, \ldots\}$ une partition de $\mathbb{R}_+$. Le nombre réel $\pi_k := t_{k+1} - t_k$ est la **durée** de la k-ème étape, qui commence à la date t_k.

Soit $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ une densité. Le **poids** de la k-ième étape est la quantité $\kappa(t_k)\pi_k$.

1.4.1 Dynamique

Soit Ω un ensemble fini, dit **espace de paramètres** et on note par A et B les **ensembles d'action** du joueur 1 et 2, respectivement. Soit $\gamma : \Omega \times A \times B$ une **fonction de paiement**

Le paramètre évolue en temps continu, en suivant une chaîne de Markov homogène avec **fonction de transition** $q : \Omega \times \Omega \times A \times B \to \mathbb{R}$, c'est a dire un fonction q qui satisfait, pour tout $(\omega, a, b) \in \Omega \times A \times B$:

$$0 \leq q(\omega, \omega', a, b) < +\infty, \;\; \omega' \neq \omega, \text{ et } \sum_{\omega' \in \Omega} q(\omega, \omega', a, b) = 0.$$

Pour $(a, b) \in A \times B$ fixé, la fonction de transition correspond à la vitesse avec laquelle le paramètre saute de ω a ω'. On note par $P(\cdot, \omega, a, b)$ le **semi-groupe de transition** correspondant, c'est a dire une famille de fonctions $P_\epsilon(\cdot, a, b) : \Omega \times \Omega \to [0, 1]$ tels que

$$\mathbb{P}(\omega_{t+\epsilon} = \omega' | \omega_t = \omega, a, b) = P_\epsilon(\omega, \omega', a, b) + o(\epsilon),$$

pour tous $t, \epsilon \geq 0$ et $\omega, \omega' \in \Omega$. L'application $t \mapsto P_t(\cdot, a, b)$ est solution de l'équation de Chapman-Kolmogorov

$$\dot{P}_t = Q^{a,b} P_t, \;\; P_0 = I,$$

où $Q^{a,b} := (q(\omega, \omega', a, b))_{\omega, \omega'}$ est le **générateur** de la chaîne de Markov avec semi groupe de transition $P(\cdot, a, b)$.

1.4.2 Information et stratégies

Le jeu se déroule de la façon suivante : à la date t_k, la valeur du paramètre est ω_k, que l'on suppose connu pour l'instant. Les joueurs choisissent leurs actions a_k, b_k. Puis, le paramètre suit la chaîne de Markov avec générateur Q^{a_k, b_k} pour une période de temps π_k. Le nouveau paramètre ω_{k+1} est observé à la date t_{k+1}. Sa loi est $P_{\pi_k}(\omega_k, \cdot, a_k, b_k)$.

Les actions restent constantes sur l'intervalle $[t_k, t_{k+1})$. Un **paiement instantané** $\gamma_s := \gamma(\omega_s, a_k, b_k)$ est payé pour le jour 2 au joueur 1 à la date $s \in [t_k, t_{k+1})$. A la date t_{k+1}, le **paiement d'étape** $\gamma_{\pi_k} := \int_{t_k}^{t_{k+1}} \kappa(s) \gamma_s ds$ a été reçu pour le joueur 1 et la situation se répète.

1.4.3 Evaluation du paiement

On considère les évaluations suivantes :

Modèle A : Le jeu à horizon infini

On considère d'abord le cas où la durée et le poids de l'étape sont égaux.

Soit Θ une probabilité décroissante sur $\mathbb{N}$ avec $\Theta = (\theta_1, \theta_2, \ldots)$ et $\theta_1 < 1/\|q\|$, où $\|q\| := \max_{(\omega, a, b) \in \Omega \times A \times B} |q(\omega, \omega, a, b)|$.

La $k-$ième étape commence à la date $s_k := \sum_{\ell < k} \theta_\ell$. La dynamique du jeu est celle décrite dans la Section 1.4.2, avec $\pi_k = \theta_k$.

Le paiement correspondant à l'histoire $h := \{\omega_1, a_1, b_1, \omega_2, a_2, b_2, \ldots\}$ est

$$\sum_{k=1}^{\infty} \theta_k \gamma_{\theta_k},$$

avec $\gamma_{\theta_k} := \theta_k \gamma(\omega_k, a_k, b_k)$. Ici, le poids du paiement à la k-ième étape est θ_k.

Modèle B : Le jeu stationnaire à étape courte

De façon intuitive, cet jeu est la discrétisation d'un jeu avec paiement

$$\int_0^{+\infty} \rho e^{-\rho s} \gamma_s ds,$$

avec $\rho > 0$. Soit $\Pi_\delta = \{0, \delta, 2\delta, \ldots\}$ une partition uniforme de $\mathbb{R}_+$, avec $0 < \delta < 1/\|q\|$. Soit $t_j^\delta := (j-1)\delta$ la date de la $j-$ième étape. Le jeu se déroule comme dans la Section 1.4.2.Le paramètre δ est ici la durée de l'étape.

Le paiement associé à l'histoire $h := \{\omega_1, a_1, b_1, \omega_2, a_2, b_2, \ldots\}$ est :

$$J_{\rho,\delta}(h) := \sum_{k=1}^{+\infty} \gamma_{\rho,j,\delta},$$

avec

$$\gamma_{\rho,j,\delta} := \int_{t_j^\delta}^{t_{j+1}^\delta} \rho e^{-\rho s} \gamma_s ds.$$

On s'intéresse ici au comportement limite lorsque δ et ρ tend vers zero.

Modèle C : Le jeu a étape courte et évaluation générale

On peut aussi considérer le paiement

$$\int_0^{+\infty} \kappa(s) \gamma_s ds,$$

où $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ est une densité sur $\mathbb{R}_+$. Dans le cas particulier $\kappa(s) := \rho e^{-\rho s}$ avec $\rho > 0$, on retombe sur le jeu précédent.

Soit $\kappa_{j,\delta} := \kappa(t_j^\delta)$. Le paiement correspondant à l'histoire $h := \{\omega_1, a_1, b_1, \omega_2, a_2, b_2, \ldots\}$ est

$$J_{\kappa,\delta}(h) := \sum_{j=1}^{+\infty} \gamma_{\kappa,j,\delta},$$

avec

$$\gamma_{\kappa,j,\delta} := \int_{t_j^\delta}^{t_{j+1}^\delta} \kappa(s) \gamma_s ds.$$

On suppose que les joueurs ont une mémoire parfaite.

Ces jeux ont une valeur par des arguments classiques. On s'intéresse au comportement asymptotique de la fonction valeur quand la durée de l'étape tend vers zéro et à sa caractérisation.

1.5 Nos contributions principales

Le Chapitre 3 est un survey qui décrit dans un cadre unifié trois approches différentes pour établir l'existence de la valeur d'un jeu différentiel à deux joueurs et à somme nulle :

i) L' **approche par discrétisation :** Cet approche a été étudiée par Fleming [38] et Friedman [41, 42]. On s'intéresse ici aux propriétés de la fonction valeur des jeux en temps discret qui approximent le jeu différentiel en temps continu.

ii) L'**approche EDP-solutions de viscosité**. Cet approche revient à Isaacs [58, p.67], qui déduit une équation aux dérivées partielles pour la valeur, dite équation d'Hamilton-Jacobi-Isaacs. Evans and Souganidis [36] ont formalisé cette idée dans le cadre des solutions de viscosité.

iii) L' **approche stratégique** de Krasovskii et Subbotin. On obtient ici l'existence de la valeur en utilisant des stratégies ϵ-optimales explicites.

Nous établions des liens entre ces approches.

Dans le Chapitre 4, on propose un preuve courte de l'existence de la valeur pour les jeux différentiels à somme nulle, horizon fini et paiement terminal, basé sur la construction de stratégies ϵ-optimales. Notre preuve est inspiré par Krasovskii et Subbotin [61]. Cet Chapitre est issu d'un travail en commun avec Miquel Oliu-Barton et accepté pour publication dans *Morfismos*.

Pour conclure la première partie, dans le Chapitre 5 on montre l'équivalence entre la définition des solutions de viscosité, introduites par Crandall et Lions [33] et la notion de solutions de minimax, introduites par Subbotin [96]. Notre preuve suit l'approche épigraphique de Frankowska [40]. A notre connaissance, l'équivalence entre solutions de viscosité et la définition "stratégique" des solutions de minimax n'a pas été explicité dans la littérature. Des idées similaires, mais dans un cadre plus général, avec un Hamiltonien mesurable en temps, ont été utilisés par Cardaliaguet et Plaskacz.

On introduit dans le Chapitre 6 un modèle pour les jeux à champ moyen en temps discret, inspiré par celui d'Adlakha, Johari et Weintraub [2]. Ce document fait partie d'un travail en cours avec S.C.P. Yam. On construit un ϵ-équilibre de Nash pour le jeu à N joueurs, où le terme d'erreur ϵ tend vers zéro lorsque N tend vers l'infini. On n'obtient pas ici de borne explicite en termes de N.

On developpe les résultats précédents dans le Chapitre 7. On propose ici une preuve alternative qui nous permet d'obtenir une borne explicite. On introduit aussi la notion de durée d'une étape dans cet Chapitre, ce qui nous permet d'obtenir un objet limite qui sert à construire un équilibre de Nash approximé pour le jeu à un nombre fini des joueurs, ou le terme d'erreur dépend du nombre de joueurs et de la durée de l'étape. Ce travail a été soumis pour publication.

Pour conclure, dans le Chapitre 8, on étudie les jeux stochastiques à durée d'étape evanescente (deux joueurs, somme nulle) dans plusieurs structures d'information :

i) Les deux joueurs observent les actions mais pas le paramètre : dans ce cas le jeu se réduit a un jeu différentiel. Sous certains hypothèses de régularité, on construit des stratégies ϵ-optimales, où ϵ dépend de la durée de l'étape.

ii) Signalisation standard : les joueurs observent le paramètre. De façon similaire au cas précédent, on obtient ici des objets limits pour construire stratégies ϵ-optimales, ce qui permet de démontrer la convergence de la suite des fonctions valeur.

Chapter 2

Introduction

2.1 Motivation and examples

The aim of game theory is to model the strategic interactions between self-interested agents, which are called **players** but that might be companies, populations, humans, computers, animals or simply mathematical objects. Such interaction is called a **game**. When the game is simple enough, it can be represented in matrix form as in the example below.

Example 2.1.1. *Let us consider the following game:*

	L	R
T	1	0
B	0	1

Player 1 is the row player, whose actions are Top or Bottom. Player 2, the column player, chooses among the actions Left or Right. A ***pure strategy*** *for the players is a function from their past information, i.e. their* ***private history****, to their action sets. In this particular situation, the game is played only once, so the set of histories is empty and a pure strategy is simply an indication of which action to play. The pure strategy sets for player 1 and player 2 are respectively $\mathcal{S}_1$ and $\mathcal{S}_2$. In this example, $\mathcal{S}_1 := \{T, B\}$ where T denotes the strategy "play Top", and B the strategy "play Bottom". Similarly, $\mathcal{S}_2 := \{L, R\}$ where L and R are the strategies "play Left" and "play Right".*

The numbers indicated on the matrix are the payoffs that player 1 receives from player 2. The situation pictured here is ***zero-sum*** *because one player's profit is at the others' expense. It is* ***one-shot*** *because players will meet only once to play this game. If strategies $\sigma \in \mathcal{S}_1, \tau \in \mathcal{S}_2$ are chosen, we denote the* ***payoff*** *by $\gamma(\sigma, \tau)$.*

Player 1 can choose his strategy optimally to ensure a payoff of at least

$$\underline{w} := \max_{\sigma \in \mathcal{S}_1} \min_{\tau \in \mathcal{S}_2} \gamma(\sigma, \tau) = 0.$$

In a similar way, player 2 can ensure that his payoff to player 1 is of at most

$$\overline{w} := \min_{\tau \in \mathcal{S}_2} \max_{\sigma \in \mathcal{S}_1} \gamma(\sigma, \tau) = 1.$$

When a game is described as above, with all the strategies available to the players and the corresponding payoffs, we say the game is in **normal form**. The quantities $\underline{w}$ and $\overline{w}$ introduced in the example above are the **maxmin** and **minmax** in pure strategies.

In the way we have specified the game in this example, there is really nothing to study. The outcome depends on who "goes first": if player 2 chooses his strategy after player 1, he can play a best reply and ensure a payoff of 0. The way out of this situation is to allow the players to choose their actions randomly. This enlargement of the strategy space is crucial for it allows players to "hide" their actions: if player 2 is not sure about what player 1 will do, he can not enforce a bad payoff for him. Denote by $\Sigma, \mathcal{T}$ the sets of mixed strategies of player 1 and 2. In this example, $\Sigma := \Delta(\{T, B\})$ and $\mathcal{T} := \Delta(\{L, R\})$, where, for a finite set $\mathcal{S}$, $\Delta(\mathcal{S})$ denotes the set of probability distributions over $\mathcal{S}$.

When mixed strategies are used, the following theorem holds:

Theorem 2.1.2. *(Minmax Theorem, von Neumann, 1928 [103]) For every two-player, zero-sum game with payoff function γ and finite action sets A, B there exist mixed strategies $\sigma^* \in \Sigma := \Delta(A)$ and $\tau^* \in \mathcal{T} := \Delta(B)$ for players 1 and 2, respectively, and a quantity v, called **value** such that, for all $(\sigma, \tau) \in \Sigma \times \mathcal{T}$:*

$$\gamma(\sigma^*, \tau) \geq v, \quad \textit{and} \quad \gamma(\sigma, \tau^*) \leq v.$$

This theorem is the cornerstone of game theory. A remark attributed to von Neumann is the following:

"As far as I can see, there could be no theory of games...without that theorem...I thought there was nothing worth publishing until the Minmax Theorem was proved."[28]

In the previous example, it is easy to see that the optimal strategies for each player are "play each action with probability 1/2" and the value is 1/2.

Of course, game theory has evolved far beyond the minmax theorem and constitutes an active area of research, comprising a large body of literature.

One important and particularly active area of research is **repeated games**. A repeated game is a game that is played more than once. This interaction may happen in **discrete time** or in **continuous time**. The repetition of a zero-sum game as the one above has no particular interest: playing i.i.d an optimal strategies each stage is optimal, and any normalized evaluation gives the value of the one-shot game. The interesting object to study are games where "something" changes with time. What exactly "something" means depends specifically on the model. The richness of the theory of repeated games comes from the fact that seemingly related models require very different tools, coming from many different branches of mathematics. Reciprocally, seemingly unrelated models can be studied with similar tools.

2.2 Contents of this thesis

This thesis concerns mostly two player, zero-sum repeated games (Part I and III). These are games where the players have opposite interests: one player's gain is at the other player's expense. Thus, players are in open competition.

Restricting to two player, zero-sum games is, admittedly, a simplification, but this by no means implies that the theory is trivial. We hope to convince the reader that the zero-sum case is already rich enough, covering different mathematical tools and ideas and leaving interesting questions unanswered.

Let us provide some motivating examples. This discussion is completely informal, proper definitions are introduced later.

Example 2.2.1. *(A game with two states)*

	L	R
T	$a_{11}^{+,p}$	a_{12}
B	a_{21}	a_{22}

Here, $a_{11}^{+,p}$ means that if (T, L) is played, player 1 receives a payoff of a_{11} and the game moves to the state $+$ with probability p. The payoff matrix of state $+$ is

	L	R
T	b_{11}	b_{12}
B	$b_{21}^{-,q}$	b_{22}

*Here $b_{21}^{-,q}$ means that if (B, L) is played, then player 1's payoff is b_{21} and the game returns with probability q to the state $-$, whose payoff matrix is the one above. Let us assume that the game is played infinitely often and denote by γ_k the **stage payoff**, that is, the payoff player 1 receives the $k-$th time the game is played, for $k = 1, 2, \ldots$. Let $\lambda \in (0, 1]$. The total payoff for player 1 is then:*

$$\sum_{k=1}^{\infty} \lambda(1-\lambda)^{k-1}\gamma_k.$$

Here, the value of the game depends on whether the initial state is $+$ or $-$.

In the example above, the factor $(1-\lambda)$ serves to represent the fact that the players are impatient and prefer current payoffs rather than future. An alternative interpretation, as provided in Shapley's [87] original paper is that of stopping probability: λ is the probability that the game stops, so that $\lambda(1-\lambda)^k$ is the probability that the game stops after $k+1$ stages.

The game above is played in discrete time. However, for many applications it is interesting to consider also games in continuous time, as motivated by the following example.

Example 2.2.2. *(Lion and Man) A lion and a man in a closed arena have equal maximum speed. What should the lion do to ensure his lunch?*

This example has been attributed to Rado by Littlewood [70, p.135] and remained a mathematical challenge for some time. It turns out that the lion can get as close as it wants to the man, but the man can avoid capture. We will not describe this here, but refer to Littlewood [70, p.135] for the original proof, attributed to Besicovitch.

The main difficulty in continuous time is that there is no canonical information pattern. This in turn implies that there is no canonical definition of strategies. Thus, the outcome of the game may depend on the information pattern adopted. This complication does not arise in discrete time, as we can unambiguously define the information available to the players at the beginning of each stage.

Several information patterns have been adopted in the differential games literature to handle this situation. For instance, in the framework of non anticipating strategies (defined in chapter 3), the interaction is of the form "strategy vs observed action", e.g. player 1 observing player 2's action before choosing his own. Thus, there is no ambiguity on the definition of the outcome, but the situation is no longer symmetric. One wishes to have a more symmetric "strategy vs strategy" interaction, that is, a normal form game.

A notion of strategies that allows to put the game in normal form, called **non anticipating strategies with delay**, was introduced by Buckdahn, Cardaliaguet and Rainer [17]. Their definition will be recalled in Chapter 3.

When the game is not in normal form, undesired phenomena may occur. For instance, it might happen that the outcome of the game is not uniquely defined or, in the example of the lion and the man, that both of them have a winning strategy, which is of course not desirable. For an amusing account of such paradoxes in the lion and man example, we refer to Bollobás, Leader and Walters [16].

Clearly, all of the examples discussed so far correspond to the two player, zero-sum case.

Another direction in which the theory is somewhat simpler are games with identical players, which we cover in Part II. It is intuitively clear that the larger the number of players is, the more sophisticated the analysis of the game becomes. However, when players are identical and influence each other by their average behavior and not individually, it is possible to wipe out this increasing complexity in a so-called **non-atomic** or **mean field** term, whose precise definition will be given later.

Let us provide a simple motivating example, borrowed from Guéant, Lasry and Lions [49, p.10].

Example 2.2.3. *Assume a continuum of agents, represented by the interval* $[0,1]$, *are attending a meeting. The meeting will not start unless a fraction* f *of the agents has arrived. Assume all the agents have a waiting cost (they do not want to spend time waiting for the meeting to start), plus a reputation and personal inconvenience costs for arriving later to the meeting. The agents want to choose their optimal arrival time.*

In this example, if one late-arriver and an early-arriver are switched, the remaining players are indifferent since the fraction of agents arriving early is the same.

Of course, in real life there is no such thing as a continuum of players, so from the point of view of applications it is interesting to know how well a non atomic game approximates an atomic game. We are interested in how to use the intuition of a non atomic game to construct ϵ-optimal strategies for a game with N players, where the approximation term ϵ goes to zero as N increases.

For the remaining of the introduction, let us describe more in detail the three parts of this thesis and highlight our main contributions.

2.3 Differential games

Let $(t_1, x_1) \in [0,1] \times \mathbb{R}^n$ and let U and V denote two compact sets of some euclidean spaces.

Define

$$\mathcal{U}(t_1) = \{\mathbf{u} : [t_1, 1] \to U, \text{ measurable}\}, \quad \mathcal{V}(t_1) = \{\mathbf{v} : [t_1, 1] \to V, \text{ measurable}\}.$$

Whenever $t_1 = 0$, we will use the more convenient notation $\mathcal{U}$ and $\mathcal{V}$ respectively.

The sets $\mathcal{U}(t_1), \mathcal{V}(t_1)$ are the sets of **control functions**. Elements of U, V are called **controls** or **actions**.

Let $f : \mathbb{R}^n \times U \times V \to \mathbb{R}^n$ and $(\mathbf{u}, \mathbf{v}) \in \mathcal{U}(t_1) \times \mathcal{V}(t_1)$ be a given pair of control functions. Consider a differential equation

$$\mathbf{x}(t_1) = x_1, \quad \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t)) \quad \text{a.e. on } [t_1, 1]. \tag{2.1}$$

We make the following assumption on f, to ensure that the trajectory of the above ODE is well defined:

Assumption 2.3.1. *Assume that the function f is jointly continuous and bounded and that there exists $c > 0$ such that for all $(u, v) \in U \times V$ and $x, y \in \mathbb{R}^n$:*

$$\|f(x, u, v) - f(y, u, v)\| \leq c\|x - y\|.$$

Let $\|f\| := \sup_{(x,u,v)} \|f(x, u, v)\| < +\infty$.

Under Assumption 2.3.1, it follows from Carathéodory's theorem, [31, Chapter 2] that (2.1) has a unique solution, whose value at time s is denoted by $\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](s)$, in the following extended sense: for any $t \in [t_1, +\infty)$,

$$\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](t) := x_1 + \int_{t_1}^{t} f(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](s), \mathbf{u}(s), \mathbf{v}(s))ds.$$

This defines the **dynamics**.

To correctly specify a **differential game** we need to define the objectives of the game and the information and strategies of the players. Informally, player 1 chooses $\mathbf{u}$ and player 2 $\mathbf{v}$ in order to achieve either a **quantitative objective** or a **qualitative objective**. As a quantitative objective, we will consider that player 1 wants to maximize a payoff depending on the trajectory, whereas for a qualitative objective we will focus on the case where player 1 wants the state variable to reach a target closed set M at time $t = 1$. This distinction was already made by Isaacs [58], who introduced the terms *games of kind* for games with a qualitative objective and *games of degree* for games with a quantitative objective.

We assume throughout this Section that players have **complete information**, that is, they know all the specifications of the game (initial state, dynamics, payoffs) as well as the past state variable and actions and the description of the game.

Qualitative case

For the qualitative case, let us consider the **target game**: player 1 aims to move the state variable to a terminal set M at time $t = 1$, while player 2 wants to prevent that. Let us denote the target game by $\Gamma_M(t_1, x_1)$.

As before, the natural questions one wants to answer in a target game are the following:

Question 3. *1. For a given initial condition $(s, y) \in [t_1, 1] \times \mathbb{R}^n$, is it possible to determine which player has a winning strategy?*
2. Provide explicit strategies (or at least $\epsilon-$optimal) for the players.

To answer the first question, let us note that it can be rewritten, informally, as the following:

Problem 2. *Construct a partition of $[t_1, 1] \times \mathbb{R}^n$ in two sets K_1, K_2, with the following properties:*
i) For any initial condition in K_1, player 1 can ensure victory (i.e. has a strategy that ensures the arrival to the target).
ii) For any initial condition in K_2, player 2 has a strategy that ensures him that the target is not reached at time 1.

A theorem that establishes such characterization is called an **alternative theorem**. Of course, alternative theorems depend on the class of strategies being considered. We will describe briefly some examples of alternative theorems in Section 2.3.3.

Quantitative case

Let $\ell : \mathbb{R}^n \times U \times V \to [0,1]$ and $g : \mathbb{R}^n \to [0,1]$. The payoffs in the quantitative case can be evaluated as follows:

1. The discounted **infinite horizon** game: for a given history of plays $(\mathbf{x}, \mathbf{u}, \mathbf{v})$ the payoff that player 1 receives from player 2 is
$$\int_{t_1}^{\infty} e^{-\rho s} \ell(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](s), \mathbf{u}(s), \mathbf{v}(s)) ds$$
where $\rho > 0$.
2. The **finite horizon** game: At time $t = 1$, player 2 gives to player 1 a payoff of
$$\int_{t_1}^{1} \ell(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](s), \mathbf{u}(s), \mathbf{v}(s)) ds + g(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](1)).$$

We denote these two games by $\Gamma_\rho(t_1, x_1)$ and $\Gamma(t_1, x_1)$ respectively. We do not cover the infinite horizon case, which has been extensively treated by Bardi and Capuzzo Dolcetta [9]. As we will show in Chapter 3, in the case of complete information, the analysis of the quantitative game with finite horizon described above can be reduced without loss of generality to the game with running payoff $\ell \equiv 0$, whenever ℓ satisfies the same regularity assumptions as the dynamics f.

Solving a quantitative game means to answer the following questions:

Question 4. *1. Give conditions for the existence and characterization of the value.*
2. Provide optimal (or ϵ-optimal) strategies.

Let us introduce the ideas to study zero-sum differential games by recalling first some results when zero and one players are present.

2.3.1 No players

In the absence of players, the dynamics is of the form

$$\dot{\mathbf{x}}(s) = f(\mathbf{x}(s)), \quad \mathbf{x}(t_1) = x_1, \tag{2.2}$$

where $x_1 \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous. In this case, the differential equation (2.2) admits a unique solution. As a quantitative objective, we can consider a payoff of the form described above.

In the quantitative case, there are no questions to be answered: the payoff associated to the trajectory is already determined. It still makes sense to consider the qualitative case.

Consider the extended dynamics

$$(\dot{t}(s), \dot{\mathbf{x}}(s)) = (1, f(\mathbf{x}(s))), \quad t(0) = 1, \ \mathbf{x}(0) = x_1, \tag{2.3}$$

Set $\bar{f} = (1, f)$. Let us recall some notions of viability theory that provide the framework to answer this question.

Definition 2.3.2. (Viability and invariance)

i) Let K be a closed subset of $[0,1] \times \mathbb{R}^n$ and $\bar{f} : [0,1] \times \mathbb{R}^n \to [0,1] \times \mathbb{R}^n$. The pair $(K, \bar{f})$ is **viable** by (2.3) if for any initial state $(t_1, x_1) \in K$, there exists a solution of (2.3) such that $(t, \mathbf{x}(t)) \in K$ for all $t \geq t_1$.

ii) We say $(K, \bar{f})$ is **invariant** if for every initial state $x_1 \in K$, all such solutions satisfy $(t, \mathbf{x}(t)) \in K$ for all $t \geq t_1$.

Since f is Lipschitz, (2.3) has a unique solution and thus the definitions of viability and invariance are equivalent. Let us point out that this will not be the case when one or two players are present, or in the non-Lipschitz case.

There are several characterization theorems for invariant and viable sets, starting with Nagumo's theorem [77]. Before stating this theorem, let us introduce some definitions.

Definition 2.3.3. Let K be a nonempty closed subset of $[0,1] \times \mathbb{R}^n$ and $z \in K$. Let $d_K(z') := \inf_{k \in K} \|z' - k\|$ denote the usual distance function.

— The **contingent cone** to K at z is the set

$$T_K(z) := \left\{ v \in \mathbb{R}^n \ : \ \liminf_{h \to 0^+} \frac{d_K(z + hv)}{h} = 0 \right\}.$$

— The **subnormal cone** to K at a point z that belongs to K is defined by

$$N_K^0(z) := \{ p \in [0,1] \times \mathbb{R}^n \ : \ \forall v \in T_K(z), \ \langle p, v \rangle \leq 0 \}.$$

Let us now state a version of Nagumo's theorem, which is general enough for our purposes.

Theorem 2.3.4. *(Nagumo) Let $\bar{f}$ be a continuous function. Then the following are equivalent:*

i) $(K, \bar{f})$ is invariant (or viable).

ii)

$$\forall z \in K, \ \bar{f}(z) \in T_K(z).$$

Of course, by definition, *ii)* is equivalent to

iii) For all $x \in K$, $\forall p \in N_K^0(x)$, $\left\langle p, \bar{f}(x) \right\rangle \leq 0$.

We omit the proof but refer to Aubin [4, Theorem 1.2.1]. The important fact here is that we can reduce the question of finding the points that reach the target to a geometrical property of the contingent cone and the dynamics.

2.3.2 One player

Let $\mathbf{u} \in \mathcal{U}(t_1)$ be a given control function. Consider a differential equation

$$\dot{\mathbf{x}}(s) = f(\mathbf{x}(s), \mathbf{u}(s)), \quad x(t_1) = x_1. \tag{2.4}$$

We assume that f satisfies the Assumptions 2.3.1, but omitting the second player.

In this situation it makes sense of distinguishing qualitative from quantitative objectives. In the quantitative case, the player wants to choose $\mathbf{u}$ in order to maximize

$$\int_{t_1}^{1} \ell(\mathbf{x}[t_1, x_1, \mathbf{u}](s), \mathbf{u}(s)) ds + g(\mathbf{x}[t_1, x_1, \mathbf{u}](1)).$$

Similarly, in the qualitative case, the player wants to choose $\mathbf{u}$ to ensure that the state reaches a target set M at time $t = 1$.

We will describe first the qualitative case as it turns out to be helpful for the quantitative case as well.

Qualitative case

Here, the trajectory of (2.4) depends on the choice of the control function $\mathbf{u}$. To obtain an alternative theorem in this case, one asks instead if for a given initial condition x_1 there exists a measurable control $\mathbf{u}$ such that the corresponding trajectory starting at x_1 reaches M at time $t = 1$.

Let us replace the differential equation (2.4) by the differential inclusion

$$\dot{\mathbf{x}}(s) \in F(\mathbf{x}(s)) := \cup_{u \in U} f(\mathbf{x}(s), u). \tag{2.5}$$

Clearly, any solution of (2.4) is a solution of (2.5). Conversely, when f is continuous with respect to the first variable and measurable with respect to the second, Filippov's measurable selection theorem [102, Theorem 2.3.13] ensures that for any trajectory $\mathbf{x}(\cdot)$ of (2.5) we can find a measurable control function $\mathbf{u}(\cdot)$ such that (2.4) holds.

We introduce now the central notion of this Section.

Definition 2.3.5. A pair $(K, \bar{F})$ where $K \subset [0,1] \times \mathbb{R}^n$ is closed and $\bar{F} : [0,1] \times \mathbb{R}^n \rightsquigarrow [0,1] \times \mathbb{R}^n$ is a set valued map is **viable** if for all $z_1 := (t_1, x_1) \in K$ there exists a solution of the differential inclusion

$$\dot{\mathbf{z}}(t) \in \bar{F}(\mathbf{z}(t)), \quad z(t_1) = z_1$$

that remains in K, i.e. $\mathbf{z}(t) \in K$ for all $t > t_1$.

In our case, $\bar{F} := (1, F)$. Let $\mathcal{B}$ denote the euclidean unit ball in $\mathbb{R}^{n+1}$.

Definition 2.3.6. A set valued map $\bar{F} : [0,1] \times \mathbb{R}^n \rightsquigarrow [0,1] \times \mathbb{R}^n$ is **Marchaud** if

a) For all $z \in [0,1] \times \mathbb{R}^n$, $\bar{F}(z)$ is a non-empty compact convex set.

b) $\bar{F}$ is upper semi-continuous, that is, $\forall z \in [0,1] \times \mathbb{R}^n$ and $\forall \epsilon > 0 \; \exists \; \delta > 0$ such that

$$\|z' - z\| < \delta \implies \bar{F}(z') \subset \bar{F}(z) + \epsilon\mathcal{B}.$$

c) F has linear growth in z, i.e. $\forall z \in [0,1] \times \mathbb{R}^n$ there exist constants γ and c such that

$$v \in \bar{F}(z) \implies \|v\| \leq \gamma\|z\| + c.$$

We are now ready to state an analogous of Theorem 2.3.4.

Theorem 2.3.7. *(Viability theorem) Let $\bar{F}$ be a Marchaud set valued map. Then the following are equivalent:*

i) $(K, \bar{F})$ is viable.

ii) $\bar{F}(z) \cap T_K(z) \neq \emptyset$ for all $z \in K$.

iii) For all $z \in K$, $\forall p \in N_K^0(z)$, $\exists v \in \bar{F}(z)$ s.t. $\langle p, v \rangle \leq 0$.

The viability theorem shows us that we can single out a trajectory that remains in K if we can do it pointwise, and conversely, thus extending Nagumo's theorem (Theorem 2.3.4) to differential inclusions. For the proof we refer to Aubin [Theorem 3.3.5][4]. In our case, *iii*) reads as

$$\forall z = (s, y) \in K, \; \forall p \in N_K^0(z), \; \exists u \in U \text{ s.t. } \langle p, (s, f(y, u)) \rangle \leq 0.$$

This suggest a way to find a control that allows the player to force the dynamics to stay in K, or close to it. Assume that the initial condition is in K and that the player will update his choice of control at discrete times $s_1, s_2, \ldots s_N = 1$. If at time s_m the state

$z_m := (s_m, y_m)$ is in K, choose any u. If z_m is outside of K, let w_m denote a closest point of z_m into K and choose u_m such that

$$\langle z_m - w_m, (s_m, f(y_m, u_m))\rangle \leq 0.$$

With this procedure one obtains the following.

Proposition 2.3.8. *Consider a partition of* $[0,1]$*, denoted* $\Pi := \{0 \leq t_1, \ldots, t_N = 1\}$ *and* $\|\Pi\|$ *its mesh, i.e.* $\|\Pi\| := \max_{m<N} t_{m+1} - t_m$*. Let* $(K, \bar{F})$ *be viable, with* F *as in* (2.5) *and* f *Lipschitz and bounded. Let* $(t_1, x_1) \in K$*.*

There exists a piecewise constant control $\mathbf{u}_\Pi(t)$ *such that*

$$\dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}_\Pi(t)), \quad \mathbf{x}(t_1) = x_1$$

satisfies

$$d_K^2(\mathbf{x}[t_1, x_1, \mathbf{u}_\Pi](1)) \leq C\|\Pi\|,$$

for a positive constant C *independent of the* Π*.*

We omit the proof as it is a corollary of the more general case with two players described in Chapter 4.

The construction described above is essentially the **extremal aiming** of Krasovskii and Subbotin [61] where player 2 is absent. It is also reminiscent of the construction in discrete time in the framework of Blackwell's approachability, see Blackwell [14].

Quantitative case

Let us consider the case of a terminal payoff at time $t = 1$, i.e., an objective of the form $g(\mathbf{x}(1))$.

For every initial condition $(t_1, x_1) \in [0,1] \times \mathbb{R}^n$, and a given $\mathbf{u}$, we have a unique trajectory, hence the following **value function**

$$\mathbf{V}(t_1, x_1) := \sup_{\mathbf{u} \in \mathcal{U}(t_1)} g(\mathbf{x}[t_1, x_1, \mathbf{u}](1))$$

is well defined.

The value function inherits the regularity of the payoff function. In particular, if g is Lipschitz, so is $\mathbf{V}(t, \cdot)$, for all t. We refer to Bardi and Capuzzo-Dolcetta [9, Chapter III, Prop. 3.1] for the proof.

The value function also satisfies the following crucial property.

Theorem 2.3.9. *(Dynamic programming principle) Assume* ℓ, g *are Lipschitz. Then, for all* $(t, x) \in [0,1] \times \mathbb{R}^n$ *and for all* $h > 0$ *:*

$$\mathbf{V}(t, x) = \max_{\mathbf{u} \in \mathcal{U}} \{\mathbf{V}(t + h, \mathbf{x}[t, x, \mathbf{u}](t + h))\}.$$

At least heuristically, by a Taylor series expansion, one can deduce from the dynamic programming principle stated above that the value function should solve the following PDE:

$$\frac{\partial \mathbf{V}}{\partial t} + \max_{u \in U} \langle \nabla_x \mathbf{V}, f(x, u)\rangle = 0 \tag{2.6}$$

with boundary condition

$$\mathbf{V}(1, x) = g(x).$$

The partial differential equation (PDE) (2.6) is called **Hamilton-Jacobi-Bellman** equation. This PDE is important in optimal control theory because it provides necessary and sufficient optimality conditions. We refer to Bardi and Capuzzo-Dolcetta [9, Chapter III, Section 3] for a detailed description and proofs.

An important problem is that (2.6) fails to have solutions in the classical sense, even if the data of the problem (dynamics and payoff functions) is smooth. A significant breakthrough was achieved by Crandall and Lions [33], who introduced the definition of **viscosity solutions**, which will be recalled in Chapter 3.

Relation between the quantitative and qualitative case

Let us point out an important connection between qualitative and quantitative problems.

Define $v_1 := \mathbf{V}(t_1, x_1)$. Consider the target game with dynamics

$$(\dot{t}(s), \dot{\mathbf{x}}(s)) \in (1, F(\mathbf{x})), \quad (t(0), \mathbf{x}(0)) = (t_1, x_1), \quad s \in [0, 1 - t_1] \tag{2.7}$$

and target set $M^* \subset [t_1, 1] \times \mathbb{R}^n$ defined by:

$$M^* := \{(1, y) \in [0, 1] \times \mathbb{R}^n \mid g(y) \geq v_1\}.$$

The link between the quantitative and qualitative game comes from the following proposition:

Proposition 2.3.10. *The set*

$$\mathcal{L}(v_1) := \{(s, y) \in [t_1, 1] \times \mathbb{R}^n \mid \mathbf{V}(s, y) \geq v_1\}$$

is viable under the dynamics (2.7).

The intuition here is that the value function is constant along optimal trajectories: since there is no running payoff, it depends only on the terminal state, that is, the state at time $t = 1$. Hence, for any initial conditions in $\mathcal{L}(v_1)$, there exists at least one trajectory that remains there (e.g. an optimal trajectory) and leads to a terminal state in M^*.

An important consequence of this fact is that solving the target problem described above and using the construction of Proposition 2.3.8 we obtain an explicit method to derive $\epsilon-$optimal strategies. We omit this construction here, but in Chapter 4 we describe in more detail the extension of this approach to the case of two players, zero sum games.

2.3.3 Two players

Qualitative case

We do not cover target games in detail in this thesis, but let us make some remarks. As it was pointed out before, the main complication in continuous time games comes from the fact that there is no canonical information structure. Let us describe briefly some of the information structures that have been proposed in the literature.

As an early example of alternative theorems for two players, let us mention the work of B.N. Pšeničnyj [79], who studies the target game with a different class of strategies

Definition 2.3.11. (ϵ-strategies) We say the players are using **ϵ-strategies** in the target game if the game is played as follows:

i) Both players know (t_1, x_1).

ii) player 2 chooses $\epsilon_1 > 0$ and informs player 1 of the control function $\mathbf{v}_1$ that he will use in the interval $[t_1, t_1 + \epsilon_1]$.

iii) Using this information, player 1 chooses his control function.

iv) At time $t_1 + \epsilon_1$, the new state is announced and the situation is repeated, with player 2 choosing ϵ_2.

Under this information structure, several examples of alternative theorems are proposed in [79], under different assumptions on the target set and the dynamics. However, in this theory no explicit connection with the quantitative case is made.

Krasovskii and Subbotin introduce the extremal aiming method [61] for target games. The description of this method motivates the work in Chapter 4. They use **positional strategies**, which are limits of piecewise constant motions. In general, the control functions originating from this procedure are not regular enough to obtain solutions in the Carathéodory sense. Thus, as in Pšeničnyj's approach, their construction provides information of an approximated game only.

In order to solve the target game exactly, Cardaliaguet [21] considers instead non anticipating strategies which are defined later in Chapter 3 and establishes an alternative theorem. Note that this is an important advance with respect to the other approaches, since it allows us to solve the target game exactly, instead of an approximate version.

Quantitative case

As we did for the target game, we need first to specify which information is available to the players and how are they allowed to interact. Different information patterns and strategies have been proposed in the literature, we refer to Bardi and Capuzzo-Dolcetta [9, Chapter VIII].

A common feature of these different notions is that they allow to reduce the problem to study a certain system of PDE's. For the finite horizon game with $\ell \equiv 0$, these PDE's are:

$$\frac{\partial w^-}{\partial t}(t,x) + \sup_{u\in U}\inf_{v\in V}\langle f(x,u,v), \nabla_x w^-(t,x)\rangle = 0 \tag{2.8a}$$

$$\frac{\partial w^+}{\partial t}(t,x) + \inf_{v\in V}\sup_{u\in U}\left\langle f(x,u,v), \nabla_x w^+(t,x)\right\rangle = 0 \tag{2.8b}$$

with boundary conditions $w^-(1,x) = w^+(1,x) = g(x)$.

Although the above relations were heuristically derived by Isaacs' [58, Section 4.2], the connection between PDE's and differential games was first made explicit in the framework of viscosity solutions [33] by Evans and Souganidis [36]. As pointed out by Lions in the introduction of the book [67], the study of these equations was a motivation to introduce the definition of viscosity solutions.

Note that under the following condition, called **Isaacs' condition**

$$\sup_{u\in U}\inf_{v\in V}\langle f(x,u,v), p\rangle = \inf_{v\in V}\sup_{u\in U}\langle f(x,u,v), p\rangle$$

holds for all $x, p \in \mathbb{R}^n$, there is only one equation, called **Hamilton-Jacobi-Isaacs** equation.

Let us remark that this condition is conceptually very strong. It imposes the existence of the value of a family of games in pure strategies, which denies the important role of randomization in game theory, as pointed out in von Neumann's remark above. In Chapter 3 we propose a way to avoid Isaacs' condition, inspired from Fleming [38]. A different approach to introduce randomized strategies in differential games and hence avoid Isaacs' condition has been proposed by Buckdahn, Li and Quincampoix [18].

Krasovskii and Subbotin applied their method to a suitable target set to establish the existence of the value for differential games with finite horizon and terminal payoff using an explicit description of $\epsilon-$optimal strategies. The material of Chapter 4 is inspired from their construction, which allows to establish the existence and characterization of the value function. Later, Subbotin [97] proposes a notion of generalized solutions, called **minimax solutions** and characterizes the value function as the unique minimax solution of the HJI equation. We establish the equivalence of minimax solutions for HJI equations with the standard machinery of viscosity solutions in Chapter 5.

2.4 Discrete time mean field games

Let us briefly comment on some related (and important) previous work on games with a continuum of players before moving on to the framework of mean field games, to which our contributions are more closely related.

2.4.1 Games with a continuum of players

As pointed out in Section 2.2, we are interested in modelling situations with a "large" number of identical agents. This statement is ill-defined, but it could mean either of the three following situations:

1. Games with a continuum of agents *per se*.
2. Convergence, in a suitable sense, of a sequence of games with atomic players to a non atomic limit game.
3. Use the limit non atomic game to compute $\epsilon-$optimal equilibria for the atomic game.

On the first situation, let us mention the pioneering work of Aumann [6]. One motivation for introducing games with a continuum of agents is an important concept in economics, *perfect competition*. The essential idea of this notion is that there are many agents whose individual influence on the economy (e.g. for affecting prices) is negligible. Thus, following [6], the natural way to study economies with perfect competition is to consider non atomic agents. The introduction of this idea allows Aumann [6] to solve a long standing conjecture in economics.

Later, Mas-Colell [72] introduced the notion of distributional equilibrium for a one shot game with a continuum of players, building on results of Schmeidler [86]. The definition of distributional equilibrium is recalled later in Chapter 6. Let us refer also to Milgrom and Weber [76] were several existence results of Nash equilibrium are established for games with incomplete information where the set of types is a continuum.

The references cited above concern one shot games only. In dynamic games, let us mention the extension of the model of Mas-Colell by Jovanovic and Rosenthal [59] for discrete time stochastic games, which is also introduced in Chapter 6. Later, Lasry and Lions [64, 65, 66] and Huang, Caines and Malhamé [56, 57] introduced the **mean field games** theory, which studies non atomic dynamic games in continuous time.

The idea of using a continuum of players is also present in the literature on congestion games, which goes back to Wardrop [106] and Smith [89]. We refer to Wan [104, 105] for an extensive survey of this literature. Games with a continuum of players have also been introduced in the framework of population games by Hamilton [51] and Maynard Smith [73].

As for the convergence of equilibria of games with finitely many players to an equilibrium of a non atomic game, an early example is the work of Haurie and Marcotte [52] in the framework of congestion games and Sandholm [85] for potential games.

In the framework of mean field games the convergence of the sequence of Nash equilibria of the N player games has been established by Lasry and Lions [64] for games with an ergodic payoff (see also Feleqi [37] for a detailed proof) and by Bardi [8] for linear quadratic mean field games.

Results of a similar flavour, based on stochastic approximation techniques, have been obtained by Benaïm and Weibull [10] for population games and by Gast, Gaujal and Le Boudec [44] for games with a centralized controller. In the stochastic approximation framework, the idea is to approximate the path of a Markov chain by a deterministic trajectory given by a suitable ordinary differential equation. The assumption these models have in common is that the probability that the relevant state variable in the N player game changes between two consecutive stages of the game goes to zero as N goes to infinity.

Our interest is more on the third situation: constructing an approximate Nash equilibria via suitably defined limit objects. In this sense, our contribution is closer to the work of Huang, Caines and Malhamé [57].

2.4.2 Continuous time mean field games

Mean field games have been introduced independently by Huang, Caines and Malhamé [56, 57] and by Lasry and Lions [64, 65, 66] and have received considerable attention in the literature. The aim of mean field games theory is to model situations with a large number of identical agents. Their distinctive feature is the **backward-forward** structure: each player anticipates a certain behavior of the other players and computes his own optimal behavior; if the observed aggregate behavior is consistent with the prediction, the players are said to be in a mean field game equilibrium. Precise definitions will be given in Chapter 6.

The following example is borrowed from Cardaliaguet's notes on mean field games [20, p.2]:

Example 2.4.1. *Let us consider first N players in $\mathbb{R}^d$. The random position of player i at time t is given by*

$$dX_t^i = \alpha_t^i dt + \sqrt{2} dB_t^i$$

Each player i aims to minimize the cost:

$$\int_t^T \frac{1}{2}|\alpha_s^i|^2 + F(m_{s,N}^{-i})ds + g(x_T^i, m_{T,N}^{-i})$$

where $m_{s,N}^{-i} := \frac{1}{N-1}\sum_{j\neq i}\delta_{x_s^j}$.

Heuristically, if one takes the limit as $N \to +\infty$, one obtains the following backward-forward system of coupled PDE's:

$$
\begin{aligned}
-\frac{\partial u}{\partial t} - \Delta u + \frac{1}{2}|D_x u|^2 &= F(x,m) \ in \, \mathbb{R}^d \times (0,T) \\
\frac{\partial m}{\partial t} - \Delta m + \operatorname{div}(mDu) &= 0 \ in \, \mathbb{R}^d \times (0,T) \\
m(0) &= m_0, \\
u(x,T) &= g(x, m(T)) \ in \, \mathbb{R}^d.
\end{aligned}
$$

One important motivation for the mean field games approach in applications is that it allows to construct approximate Nash equilibria of games with a large number of players, which is not computationally feasible when the number of players is large. The limit system consists of a Hamilton-Jacobi equation running backward in time and a Kolmogorov equation describing the aggregate evolution, running forward in time.

Mean field games have found applications in many areas, notably in economics, see Guéant, Lasry and Lions [49]. The lecture notes of Cardaliaguet [20], based on Lions' lectures at the Collège de France, provide a detailed account from the mathematical point of view. For a brief and more recent account of the continuous time theory, we refer to the survey by Gomes and Saude [47] and to the book of Bensoussan, Frehse and Yam [12] for connections with the theory of mean field type control problems.

However, a conceptual problem arises in continuous time. As we mentioned earlier in the introduction, randomization of the actions is a crucial concept in game theory. Even if we give up on choosing actions randomly, it turns out that in continuous time, for the Kolmogorov equation to be well defined, the solution of the Hamilton-Jacobi equation needs to be differentiable. This implies that at each time t, each player has a unique optimal choice. However, one can easily imagine situations where this does not hold. This can be sorted out if we consider a discrete time model, as we show in Chapter 7. Other conceptual problems, which are present also in discrete time, are addressed in Chapter 7.

With few exceptions, the mean field games literature has largely focused on mean field games in continuous time.

An important exception is the work of Gomes, Mohr and Souza [48]. They study the asymptotic behaviour as the time horizon goes to infinity, of a finite horizon, discrete time, finite state dynamic game with a continuum of players, and provide conditions for the convergence to a stationary solution.

We consider instead a fixed time horizon and provide a way to construct an approximate Nash equilibrium for the N player game. The model we introduce in Chapter 6 is the finite horizon version of the model introduced by Adlakha, Johari and Weintraub [2].

In many applications, it is desirable to consider "frequent" interactions between the players and a random dynamics that depends on both the individual and the aggregate state. For instance, one can think of competing agents in an online auction. To give a sense to "frequent", we need an exogenous time, that runs independently of the stages of the game.

With this in mind, we introduce as an exogenous time the positive real axis $\mathbb{R}_+$ and allow the players to receive information at discrete points. In this game, the dynamics of the state corresponds to a continuous time Markov chain. These games are described in more detail in Section 2.6. For now, let us point out that these games are, informally, discretizations of an underlying stochastic game in continuous time, which are analogue to the discretizations introduced by Fleming [38] for differential games.

Incorporating these ideas, we obtain a limit object that provides an approximate Nash equilibrium for games with sufficiently many players and sufficiently frequent interactions. We develop this model in detail in Chapter 7.

2.5 Zero-sum stochastic games with short-stage duration

2.5.1 General model of zero-sum repeated games

Let us introduce an abstract mathematical model for zero-sum dynamic games that extends the simple motivating examples we proposed. The model presented here is borrowed from Cardaliaguet Laraki and Sorin [23], see also Mertens, Sorin and Zamir [75, Section IV].

Let Ω, A, B, R, S be arbitrary sets and consider a function $\gamma : \Omega \times A \times B \to [0, 1]$. The set Ω is the **parameter space**. The sets A, B are the **sets of actions** of player 1 and 2 respectively.

The game is played as follows:

— An initial parameter ω_1 and signals r_1, s_1 are chosen randomly according to an initial distribution $\pi \in \Delta(R \times S \times \Omega)$. Player 1 receives signal r_1 and player 2 receives the signal s_1 and this is all the information they get.

— After player 1 (respectively, player 2) learns his signal r_1 (resp. s_1), player 1 (resp. 2) chooses an action a_1 (resp. b_1). The **stage payoff** $\gamma_1 := \gamma(\omega_1, a_1, b_1)$ is allocated to player 1 and is not necessarily observed by the players. The actions are chosen simultaneously and independently.

— A new value of the parameter and the signals is chosen according to a **transition function** $Q : \Omega \times A \times B \to \Delta(R \times S \times \Omega)$.

The situation is then repeated: (r_2, s_2, ω_2) are chosen according to $Q(\omega_1, a_1, b_1)$; knowing r_2(resp. s_2), an action a_2 (resp. b_2) is chosen, player 1 receives a payoff γ_2 from player 2 and so on.

A **(pure)behavioural strategy** σ for player 1 is a map from his private history $\mathcal{H}^1 := (r_1, a_1, r_2, a_2, \ldots)$ to $(A)\Delta(A)$. A pure or behavioural strategy τ for player 2 is defined similarly. Player 2's private history is denoted by $\mathcal{H}^2$. A **play** is a sequence $(\omega_1, r_1, s_1, a_1, b_1, \omega_2, r_2, s_2, a_2, b_2, \ldots)$. Let $H_\infty := (\Omega \times R \times S \times A \times B)^\infty$ denote the set of plays. A **mixed strategy** for player 1 is a probability distribution over his set of pure strategies, with an analogous definition holding for player 2.

By Kolmogorov's extension theorem, a couple of behavioural strategies defines a unique probability distribution over H_∞. Let us assume that the players have **perfect recall**, that is, that players remember the full history of the game. In this case, Kuhn's theorem [62] applies, which ensures that the games played in mixed or behavioural strategies are equivalent. Thus, we can consider without loss of generality that the game is played in mixed strategies. Note that the set of mixed strategies for each player is compact and convex.

A couple of strategies, along with π and Q, generates a unique probability distribution on the plays, the corresponding expectation is denoted by $\mathbb{E}_{\sigma\tau}$.

Two important classes captured by this model are **stochastic games** as introduced by Shapley [87], which correspond to public signals including the realization of the parameter, and **incomplete information** games as studied by Aumann and Maschler [7] which correspond to an absorbing transition of the parameter, which remains fixed, and no further information after the initial one on its value. By **complete information** we mean that the current state, payoffs and past actions are observed.

2.5.2 Payoff evaluations

The previous model determines a sequence of stage payoffs $\{\gamma_k\}_{k\in\mathbb{N}^+}$. The stream of stage payoffs can be evaluated in different ways.

Consider a probability distribution $\Theta = \{\theta_k\}_{k\in\mathbb{N}^+} \in \Delta(\mathbb{N}^+)$ and a history $h := \{\omega_1, a_1, b_1, \omega_2, a_2, b_2, \ldots\}$. The game Γ_Θ is the game with **evaluation function**:

$$J_\Theta(h) := \sum_k \theta_k \gamma_k.$$

For a couple of behavioral strategies (σ, τ) we define the payoff as the expectation with respect to their induced probability distribution over H_∞. From Kuhn's theorem, we can consider this game as being played in mixed strategies, since they induce the same probability distributions over H_∞. Thus, the value of Γ_Θ exists by Sion's minmax theorem [88] and is denoted v_Θ.

Some classical choices of Θ are the uniform partition $\Theta := (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$ so that the payoff becomes

$$\frac{1}{n}\sum_{k=1}^{n} \gamma_k$$

or the $\lambda-$ discounted evaluation

$$\sum_{k=1}^{\infty} \lambda(1-\lambda)^{k-1}\gamma_k$$

where $\lambda \in (0, 1]$. Let us denote by v_n the value of the $n-$stage repeated game and v_λ the value of the λ-discounted game.

2.5.3 Recursive structure and long games

Shapley [87] established the following recursive formula for stochastic games with complete information. Let $X = \Delta(A), Y = \Delta(B)$. Then:

$$v_\lambda(\omega) = \mathrm{val}_{x,y}\mathbb{E}_{xy}\left\{\lambda\gamma(\omega, x, y) + (1-\lambda)\sum_{\omega'\in\Omega} v_\lambda(\omega')Q(\omega, x, y)(\omega')\right\}. \tag{2.9}$$

A similar formula holds for the finite n-stage game, namely:

$$v_n(\omega) = \mathrm{val}_{x,y}\mathbb{E}_{xy}\left\{\frac{1}{n}\gamma(\omega, x, y) + \frac{n-1}{n}\sum_{\omega'\in\Omega} v_{n-1}(\omega')Q(\omega, x, y)(\omega')\right\}. \tag{2.10}$$

These formulae express the value of the game as a weighted average between today's payoff and the expected payoff from tomorrow on. An important consequence of these formulae is that the players have **stationary strategies** in the discounted case, that is, strategies that are functions of the current state only, and **Markovian strategies** in the finite case, that is, strategies that depend on the stage and the current state. In particular, the players do not need to know the move of the opponent.

Note also that (2.9) has a fixed point structure, which motivates the following definition.

Definition 2.5.1. (Shapley operator) Let $\mathcal{F}$ the set of functions from Ω to $\mathbb{R}$. The **Shapley operator** $\mathbf{T} : \mathcal{F} \to \mathcal{F}$ is defined as

$$\mathbf{T}[f](\omega) := \mathrm{val}_{x,y}\mathbb{E}_{xy}\left\{\lambda\gamma(\omega, x, y) + (1-\lambda)\sum_{\omega'\in\Omega} f(\omega')Q(\omega, x, y)(\omega')\right\}. \tag{2.11}$$

Introducing the following auxiliary operator $\mathbf{\Phi}$

$$\mathbf{\Phi}(\epsilon, f) := \epsilon \mathbf{T}\left(\frac{1-\epsilon}{\epsilon} f\right)$$

one obtains

$$v_\lambda = \mathbf{\Phi}(\lambda, v_\lambda), \quad v_n = \mathbf{\Phi}\left(\frac{1}{n}, v_{n-1}\right). \tag{2.12}$$

A similar recursive structure holds for games with incomplete information and for a general evaluation Θ.

Mertens, Sorin and Zamir [75, Section IV.3] associate to each repeated game an auxiliary stochastic game whose value functions satisfy a recursive equation of the type (2.9), in a suitably enlarged state space, called the **universal belief space**.

Let us describe this more precisely in the case where the signals of the players include each others' actions but not the parameter. Assume that the initial parameter is chosen with a commonly known lottery $\xi \in \Delta(\Omega)$, but remains unobserved. We assume that the players observe the actions, so that they can compute the posterior law of the parameter. In this case, the auxiliary state variable is the law of the parameter.

Observe that the fixed point characterization of v_λ does not hold for v_n. To recover the same function on both sides of the equation, Cardaliaguet, Laraki and Sorin [23] add a time variable, as described below, which represents the past fraction of the game.

For a given Θ, consider the induced partition $\Pi_\Theta = \{t_1, t_2, \ldots\}$ of $[0,1]$ where $t_1 = 0$, and $t_n = \sum_{m=1}^{n} \theta_m$ for $n > 1$. The repeated game is naturally represented as a game played between times 0 and 1 where the actions are constant on each subinterval $[t_{n-1}, t_n)$ of length θ_n. Let $V_\Theta(t_n, \cdot)$ denote the value of the game starting at t_n.

By definition, $V_\Theta(1, \cdot) = 0$ and

$$V_\Theta(t_n, \xi) = \mathrm{val}_{xy} \mathbb{E}^Q_{xy}\{\theta_{n+1}\gamma_n + V(t_{n+1}, \xi')\}. \tag{2.13}$$

By linear interpolation, V_Θ is extended to a function on $[0,1]$. Thus, asymptotic properties of v_Θ translate into asymptotic properties of V_Θ.

The study of the asymptotic properties of v_Θ is the so-called **asymptotic analysis** and the questions here are the existence and characterization of the limit.

The **variational approach**, initiated by Laraki in his PhD thesis and revisited in Cardaliaguet, Laraki and Sorin [23] consists on the study of asymptotic properties of v_Θ via suitable variational inequalities satisfied by the accumulation points of V_Θ.

The variational approach allowed to unify the proofs of existence of the asymptotic value for games with incomplete information, splitting games and absorbing games.

2.6 Short stage stochastic games

Let us introduce now a different family of stochastic games that enjoys nice asymptotic properties, although of a different nature, so that our results are not directly comparable with the classical framework. The model we study is in some sense more regular than classical stochastic games. The dynamic consists of a Markov chain, controlled by the players and evolving in continuous time, while the players are allowed to update their actions in discrete time. Here one is interested in the limit when the time between consecutive stages goes to zero. These are called **short stage** games.

Games where a payoff relevant parameter follows a continuous time Markov chain have been introduced by Zachrisson [107] under the name *Markov games*. However, his model

does not incorporate a notion of stage duration. They have also been studied by Tanaka and Wakuta [100, 99] and Tanaka and Lai [98]. These authors assume first the existence of Markovian strategies, from which they derive an equation for the value function that they use to construct stationary strategies.

Markov games with one player, also known as continuous time Markov decision processes have received more attention in the literature. We refer to the recent book of Hernández-Lerma and Prieto-Rumeau [54], where also some classes of Markov games are treated.

Short stage games have been introduced by Neyman [78]. A similar, but conceptually different model has been studied by Cardaliaguet, Rainer, Rosenberg and Vieille [26] in the framework of incomplete information. While Neyman considers a sequence of games and defines conditions for its convergence, in a suitable sense, to a limit game, Cardaliaguet, Rainer, Rosenberg and Vieille study a sort of Fleming [38] discretization of an underlying continuous time game. We follow this approach.

In discrete time repeated games, there is no exogenous notion of time. We can distinguish between two stages, but we can not speak a priori of the "duration" of each state.

Let us now consider a model with an underlying notion of time, represented by the positive real axis $\mathbb{R}_+$. This allows us to define a notion of "duration" of each stage as follows: Let $\Pi = \{t_1, t_2, \ldots\}$ denote a partition of $\mathbb{R}_+$, called **sequence of stages** and let $\pi_k := t_{k+1} - t_k$ denote the **duration** of the k-th stage, that begins at time t_k.

Let $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ denote an integrable function. The **weight** of the k-th stage is the quantity $\kappa(t_k)\pi_k$, which is an approximation of $\int_{t_k}^{t_{k+1}} \kappa(s)ds$.

2.6.1 Basic dynamics

Let Ω be a finite set, called the **parameter space** and let A and B denote the finite **action** sets of players 1 and 2 respectively. Let $\gamma : \Omega \times A \times B$ denote a **running payoff**.

The parameter evolves in continuous time, following an homogeneous Markov chain with **transition rate function** $q : \Omega \times \Omega \times A \times B \to \mathbb{R}$, i.e. a function that satisfies, for all $(\omega, a, b) \in \Omega \times A \times B$:

$$0 \leq q(\omega, \omega', a, b) < +\infty, \;\; \omega' \neq \omega, \text{ and } \sum_{\omega' \in \Omega} q(\omega, \omega', a, b) = 0.$$

For fixed $(a, b) \in A \times B$, the transition function corresponds to the speed with which the parameter jumps from ω to ω'. To these actions corresponds a **transition semigroup** $P(\cdot, \omega, a, b)$, which is a collection of maps $P_\epsilon(\cdot, a, b) : \Omega \times \Omega \to [0, 1]$ such that

$$\mathbb{P}(\omega_{t+\epsilon} = \omega' | \omega_t = \omega, a, b) = P_\epsilon(\omega, \omega', a, b) + o(\epsilon)$$

for all $t, \epsilon \geq 0$ and $\omega, \omega' \in \Omega$. The map $t \mapsto P_t(\cdot, a, b)$ is the solution of the Chapman-Kolmogorov equation

$$\dot{P}_t = Q^{a,b} P_t, \;\; P_0 = I$$

where the matrix $Q^{a,b} := (q(\omega, \omega', a, b))_{\omega,\omega'}$ is the **generator** of the Markov chain with transition semigroup $P(\cdot, a, b)$.

2.6.2 Information and strategies

We describe now how players influence the dynamics introduced in Section 2.6. The game is essentially the same as the general model introduced in Section 2.5.1, except that

the dynamics of ω_t is now in continuous time.

The game is played as follows: at time t_k, the value of the parameter is ω_k. For simplicity, assume that the players are informed about the current state ω_k and the past actions. Both players choose their pure actions a_k, b_k, possibly by randomization. Once the actions are chosen, the parameter follows the Markov chain with generator Q^{a_k,b_k} for a time π_k. The new parameter ω_{k+1} is observed at date t_{k+1}. Its law is $P_{\pi_k}(\omega_k, \cdot, a_k, b_k)$.

The actions are held constant in the interval $[t_k, t_{k+1})$. An **instantaneous payoff** $\gamma_s := \gamma(\omega_s, a_k, b_k)$ is allocated at time $s \in [t_k, t_{k+1})$. At time t_{k+1}, the **stage payoff** $\int_{t_k}^{t_{k+1}} \kappa(s)\gamma_s ds$ has been allocated and the situation is repeated.

2.6.3 Payoff evaluation

To evaluate the payoffs, we consider three different scenarios:

Model A: The game in $[0, 1]$

Let us consider first the case when the duration and the weight of the stage are equal.

Let Θ denote a decreasing probability measure over $\mathbb{N}$ with $\Theta = (\theta_1, \theta_2, \ldots)$ and $\theta_1 < 1/\|q\|$, where $\|q\| := \max_{(\omega,a,b)\in\Omega\times A\times B} |q(\omega, \omega, a, b)|$.

The $k-$th stage takes place at time $s_k := \sum_{\ell<k} \theta_\ell$.

The dynamics of the play at the $k-$th stage corresponds to the play at time s_k as described in Section 2.6.2, with $\pi_k = \theta_k$.

The payoff for a history $h := \{\omega_1, a_1, b_1, \omega_2, a_2, b_2, \ldots\}$ is

$$\sum_{k=1}^{\infty} \theta_k \gamma_{\theta_k}$$

where $\gamma_{\theta_k} := \theta_k \gamma(\omega_k, a_k, b_k)$ is the stage payoff. Here, the weight of the payoff at stage k is the constant θ_k.

Except for the dynamics, this game is exactly the general model of Section 2.5.1. This difference is crucial and we will elaborate on this later.

Model B: The stationary game with short stage

Here we consider a situation where the duration of the stage and the weight are no longer identical.

Informally, the game we describe is a discretization of an infinite horizon game with continuous time payoff:

$$\int_0^{+\infty} \rho e^{-\rho s} \gamma_s ds.$$

A common interpretation of ρ is as the **patience** of the players: the smaller ρ is, the players are more patient.

Let $\Pi_\delta = \{0, \delta, 2\delta, \ldots\}$ denote a uniform partition of $\mathbb{R}_+$, where $0 < \delta < 1/\|q\|$. Denote with $t_j^\delta := (j-1)\delta$ the instant where the $j-$th play takes place. The game is played as in Section 2.6.2. Here, the parameter δ is the stage duration, or, alternatively, $1/\delta$ is the **action frequency**.

The payoff corresponding to a history $h := \{\omega_1, a_1, b_1, \omega_2, a_2, b_2, \ldots\}$ is:

$$J_{\rho,\delta}(h) := \sum_{k=1}^{+\infty} \gamma_{\rho,j,\delta}$$

where

$$\gamma_{\rho,j,\delta} := \int_{t_j^\delta}^{t_{j+1}^\delta} \rho e^{-\rho s} \gamma_s ds.$$

We refer to this game as the (normalized) $\rho-$**discounted game with action frequency** $1/\delta$. Within this framework, it is natural to study the limit as δ and ρ go to zero and whether these limits commute.

Model C: The short stage game with arbitrary evaluation

We can extend the previous model to a discretization of an infinite horizon game with continuous time payoff:

$$\int_0^{+\infty} \kappa(s)\gamma_s ds$$

where $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ is an density function on $\mathbb{R}_+$. Some choices for κ might be, for instance, a uniform distribution with support on a compact interval or the exponential density $\kappa(s) := \rho e^{-\rho s}$ for a positive constant ρ, as in the previous Section.

Set $\kappa_{j,\delta} := \kappa(t_j^\delta)$. For a history $h := \{\omega_1, a_1, b_1, \omega_2, a_2, b_2, \ldots\}$, the corresponding payoff is

$$J_{\kappa,\delta}(h) := \sum_{j=1}^{+\infty} \gamma_{\kappa,j,\delta}.$$

with

$$\gamma_{\kappa,j,\delta} := \int_{t_j^\delta}^{t_{j+1}^\delta} \kappa(s)\gamma_s ds.$$

From the arguments in Section 2.5.1, if we assume perfect recall, then all these games have a value by Sion's minmax theorem.

Comparison of the evaluations

In model A, as well as in the classical framework, one studies the sequence of value functions for decreasing evaluations. In both cases the weight of each stage on the payoff is the same. The crucial difference is in the dynamics: in the classical framework the transition probability between two consecutive stages is independent of the weight of the stage payoff, while in our framework it goes to zero. This helps to avoid the oscillation phenomena that arise in the classical framework (see Ziliotto [108] and Sorin and Vigeral [93]).

In model A, we consider a decreasing sequence of evaluations, while in model C we consider a discretization of a fixed evaluation. This has the following consequence: let us suppose for a moment that κ has support in $[0, 2T]$, for $T > 0$. At half of the game, that is, at time $s = \frac{1}{2}$ if we are in model A or time $t = T$ in model C, the asymptotic accumulated payoff in model A, when the stage vanishes is:

$$\lim_{\|\Theta\|\to 0} \sum_{\ell=1}^{\min\{k:s_k<1/2\}} \theta_\ell \to 0$$

where $\|\Theta\| := \sup_{k\in\mathbb{N}} \theta_k = \theta_1$.

Whereas for model C, as the duration of the stage vanishes, the accumulated payoff at half of the game is:

$$\lim_{\delta \to 0} \sum_{j=1}^{\lfloor T/\delta \rfloor} \kappa_{j,\delta} \to \int_0^T \kappa(s) ds.$$

Clearly, model B is a particular case of the evaluation of model C. The interest of studying it separately relies on the stationarity of the value function (model B), which in particular allows us to study separately asymptotic properties in two time scales: with respect to the patience of the players and the frequency of play.

2.7 Main contributions

Let us highlight the main contributions of this thesis.

Chapter 3 is essentially a survey where we explain in a unified framework three approaches for establishing the existence and characterization of the value function of a two player, zero sum differential game, which are conceptually very different:

i) The **discrete game approach:** this idea goes back to Fleming [38] and Friedman [41, 42]. This approach consists in studying properties of the value functions of suitable discrete approximations of a differential game. The first results of convergence of the sequence of discrete value functions and characterization of the limit go back to Fleming [38] under strong regularity assumptions, which we can relax thanks to the machinery of viscosity solutions.

ii) The **viscosity approach**. This is initiated by the intuition of Isaacs [58, p.67], who guessed that the value function should be, whenever smooth, a classical solution of the HJI equation, under the Isaacs condition. This was formalized later by Evans and Souganidis [36] in the framework of viscosity solutions.

iii) The **strategic approach** of Krasovskii and Subbotin. Here one obtains the existence of the value function via an explicit construction of ϵ-optimal strategies. One recovers also characterization of the value function by introducing a notion of generalized solutions for the HJI equation, called minimax solutions [96].

We point out several connections between them and motivate the exposition on the next two chapters, which are devoted to the strategic approach and its connection with the viscosity solution approach.

In Chapter 4, we propose a short and self-contained proof of the existence of the value function in differential games with a terminal payoff, based on the construction of approximately optimal strategies. Our construction is inspired from the extremal aiming method of Krasovskii and Subbotin [61] and corresponds to what is called "strategic approach" in Chapter 3. This document is a joint work with Miquel Oliu-Barton and has been accepted for publication in *Morfismos*.

To close the first part, in Chapter 5 we establish the equivalence between the notion of viscosity solutions as defined by Crandall and Lions [33] and the more geometrical notion of minimax solutions. Minimax solutions arise in the theory of Krasovskii and Subbotin [61] as a generalized solution concept for the Hamilton-Jacobi-Isaacs equation of a differential game with terminal payoff. Our proof is inspired of the epigraphical approach introduced by Frankowska [40] and relies on an intermediate solution concept (proximal solutions) introduced by Clarke and Ledyaev [29]. To the best of our knowledge, the equivalence between viscosity solutions and the "strategic" definition of minimax solutions in terms of viable sets had not been made explicit for differential games, which motivated us to fill this small gap in the literature. In the more general case of Hamilton-Jacobi-Isaacs with

time-measurable hamiltonians, similar geometrical ideas can be found in Cardaliaguet and Plaskacz [24].

In Chapter 6 we introduce a model for discrete time mean field games with finite horizon, based on the model of Adlakha, Johari and Weintraub [2]. This document is part of some work in progress with S.C.P. Yam. We construct an ϵ- Nash equilibrium for the game with N players, where the error term ϵ goes to zero as N goes to infinity. However, we do not obtain here an explicit relation between ϵ and N.

An improvement with respect to the results of Chapter 6 is found in Chapter 7. Here we propose a different proof for a similar model, which allows us to obtain an error term depending explicitly on the number of players and the time horizon, as well as other constants of the game. In Chapter 7 we also explore a different asymptotic result, in terms of the stage duration, similar to the model of Neyman [78]. We provide a construction of an approximate Nash equilibrium for the game with N players, where the error term depends on both the number of players and the duration of the stage. This work has been submitted for publication.

In Chapter 8, we study stochastic games with short stage duration, as described in the previous Section. We consider different information scenarios:

i) Both players observe the actions but not the state: in this case, the game reduces to a differential game. We deduce limit equations for the value function as the duration of the stage goes to zero for models A,B,C. Under suitable regularity assumptions on the value function, we construct $\epsilon-$optimal strategies, where the approximation term ϵ depends on the duration of the stage.

ii) Standard signalling: In this case, we derive similar results to those for the case i), namely, deducing a suitable limit object and use it to generate almost optimal strategies. Since the state space and actions are finite, we can exploit the semi algebraic aspect of the model [1] and obtain asymptotic results in a double time scale, for the game with discounted payoffs: patience (or discount rate) and frequency (or stage duration), as in Neyman [78].

To conclude this Introduction, let us point out that a unifying thread of this thesis is the search for a *limit object* that helps play "almost optimally" in a given game. Here "almost optimally" depends on the particular game: in Chapter 6, the error of the strategy derived from the limit object vanishes as the number of players increases; in the results of Chapter 4 and 8 it is related to the duration of the game, while we derive an approximation term in terms of both in Chapter 7.

1. In the finite case, the value and the optimal strategies satisfy a system of polynomial inequalities, see Sorin [92].

Part I

Differential games

Chapter 3

Value of zero-sum differential games

Abstract: In this chapter we review and compare three different approaches for establishing the existence and characterization of the value function in differential games.

3.1 Introduction

Let $(t_1, x_1) \in [0,1] \times \mathbb{R}^n$ and let U and V denote two compact sets of some euclidean spaces. Let us define

$$\mathcal{U}(t_1) = \{\mathbf{u} : [t_1, 1] \to U, \text{ measurable}\}, \quad \mathcal{V}(t_1) = \{\mathbf{v} : [t_1, 1] \to V, \text{ measurable}\}.$$

Whenever $t_1 = 0$, we will use the more convenient notation $\mathcal{U}$ and $\mathcal{V}$ respectively.

The sets $\mathcal{U}(t_1), \mathcal{V}(t_1)$ are the sets of **control functions**. An element $\mathbf{u} \in \mathcal{U}(t_1)$ is called a control function for player 1, while an element $\mathbf{v} \in \mathcal{V}(t_1)$ is a control function for player 2. Elements of U, V are called **controls** or **actions**.

Let $f : \mathbb{R} \times U \times V \to \mathbb{R}^n$ and $(\mathbf{u}, \mathbf{v}) \in \mathcal{U}(t_1) \times \mathcal{V}(t_1)$ be a fixed pair of control functions. Consider a differential equation

$$\mathbf{x}(t_1) = x_1, \quad \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t)) \quad \text{a.e. on } [t_1, 1]. \tag{3.1}$$

To ensure the existence of solutions of this equation, we make the following Assumption, which holds for the rest of this chapter:

Assumption 3.1.1. *Assume that the function f is jointly continuous and bounded and that there exists $c > 0$ such that for all $(u, v) \in U \times V$ and $x, y \in \mathbb{R}^n$:*

$$\|f(x, u, v) - f(y, u, v)\| \leq c\|x - y\|.$$

Let $\|f\| := \sup_{(x,u,v)} \|f(x, u, v)\| < +\infty$.

Under Assumption 3.1.1, a consequence of Carathéodory's theorem, [31, Chapter 2] is the following:

Lemma 3.1.2. *For $(\mathbf{u}, \mathbf{v}) \in \mathcal{U} \times \mathcal{V}$ fixed, (3.1) has a unique solution, whose evaluation at time s is denoted by $\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](s) =: \mathbf{x}(s)$, in the following extended sense: for any $t \in [t_1, 1]$,*

$$\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](t) := x_1 + \int_{t_1}^{t} f(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](s), \mathbf{u}(s), \mathbf{v}(s))ds.$$

A **play** is a triplet $(\mathbf{x}, \mathbf{u}, \mathbf{v})$ where $(\mathbf{u}, \mathbf{v}) \in \mathcal{U}(t_1) \times \mathcal{V}(t_1)$ and $\mathbf{x}$ is a solution of (3.1) corresponding to $(\mathbf{u}, \mathbf{v})$.

Let us consider two functions $\ell : \mathbb{R}^n \times U \times V \to \mathbb{R}$ and $g : \mathbb{R} \to [0, 1]$. The **payoffs** associated to the play $(\mathbf{x}, \mathbf{u}, \mathbf{v})$ can be evaluated as follows:

1. The discounted **infinite horizon** game: for a given history of plays $(\mathbf{x}, \mathbf{u}, \mathbf{v})$ the payoff that player 1 receives from player 2 is
$$\int_{t_1}^{\infty} e^{-\rho s} \ell(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](s), \mathbf{u}(s), \mathbf{v}(s)) ds$$
where $\rho > 0$.
2. The **finite horizon** game: At time $t = 1$, player 2 gives to player 1 a payoff of
$$\int_{t_1}^{1} \ell(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](s), \mathbf{u}(s), \mathbf{v}(s) ds + g(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](1)).$$

We assume from now on that the payoff functions satisfy:

Assumption 3.1.3. *$\ell : \mathbb{R}^n \times U \times V \to \mathbb{R}$ satisfies Assumption 3.1.1 with constant L_ℓ and $g : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz continuous with Lipschitz constant L_g and bounded.*

Player 1 wants to choose $\mathbf{u}$ in order maximize the payoff he receives, while player 2 chooses $\mathbf{v}$ in order to minimize the payoff he gives to player 1. This situation is a **zero-sum differential games**. When the payoff is evaluated as an infinite horizon payoff, the game is denoted by $\Gamma_\rho(t_1, x_1)$. If the payoff is evaluated as a finite horitzon payoff, the game is denoted $\Gamma(t_1, x_1)$ respectively. We discuss here the finite horizon case only.

Up to a change of variables, one can assume that $\ell \equiv 0$. To see this, we consider, instead of $\mathbf{x}$, the enlarged state $\mathbf{y} := (\mathbf{x}, \mathbf{z})$ given by:

$$\mathbf{x}(t_1) = x_1, \quad \dot{\mathbf{x}}(t) = f(\mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t)) \quad \text{a.e. on } [t_1, 1] \tag{3.2}$$
$$\mathbf{z}(t_1) = 0, \quad \dot{\mathbf{z}}(t) = \ell(\mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t)) \quad \text{a.e. on } [t_1, 1]. \tag{3.3}$$

The payoff on this new game is of the form $g'(\mathbf{y}[t_1, y_1, \mathbf{u}, \mathbf{v}](1)) := \int_{t_1}^{1} \ell(\mathbf{x}[t_1, y_1, \mathbf{u}, \mathbf{v}](s), \mathbf{u}(s), \mathbf{v}(s) ds + g(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](1))$.

In this chapter, we will consider the finite horizon problem with $\ell \equiv 0$ only. By **solving** the game $\Gamma(t_1, x_1)$, we mean to give an answer to the two following questions

1. *What* is the "best" payoff Player 1 can get?
2. *How* does Player 1 need to play to get such payoff?

which are formulated in terms of player 1, for simplicity, but completely analogous questions are posed for player 2. Despite being intuitive questions, there is no canonical way to answer them. In particular, it depends on how are players allowed to interact. So far, we have specified the dynamics and payoff for a given couple of control functions, but we have not detailed how these control functions are generated.

Strategies

Let us introduce the following definitions of strategies:

Definition 3.1.4. (Non anticipating strategies)

i) A **non anticipating strategy (NA)** for player 1 is a map $\alpha : \mathcal{V} \to \mathcal{U}$ such that, for $t \in [t_1, 1]$:
$$\mathbf{v}_1 \equiv \mathbf{v}_2 \text{ a.e. on } [t_1, t] \implies \alpha(\mathbf{v}_1) \equiv \alpha(\mathbf{v}_2) \text{ a.e. on } [t_1, t].$$

ii) Analogously, a **non anticipating strategy (NA)** for player 2 is a map $\beta : \mathcal{U} \to \mathcal{V}$ such that, for $t \in [t_1, 1]$:

$$\mathbf{u}_1 \equiv \mathbf{u}_2 \text{ a.e. on } [t_1, t] \implies \beta(\mathbf{u}_1) \equiv \beta(\mathbf{u}_2) \text{ a.e. on } [t_1, t].$$

These strategies are also sometimes [9] called VREK strategies, as an acronym for Varaiya [101], Roxin [84], Elliot and Kalton [35].

However, let us remark that these are *not* strategies in the sense of game theory. In particular, a pair of NA strategies may fail to give a well defined play, as the following example shows.

Example 3.1.5. *Let $U = V = [-1, 1]$. Consider the pair of NA strategies (α, β) defined by:*

$$\alpha(\mathbf{v}) = -\mathbf{v}, \;\; \beta(\mathbf{u}) = sgn(\mathbf{u}).$$

where $sgn(a) = 1$ if $a \geq 0$ and -1 otherwise. Suppose there exist $(\mathbf{u}, \mathbf{v})$ such that

$$\alpha(\mathbf{v}) = \mathbf{u} \;\; \beta(\mathbf{u}) = \mathbf{v}$$

hence, $sgn(-\mathbf{v}) = \mathbf{v}$, a contradiction.

To overcome this difficulty, Buckdahn, Cardaliaguet and Rainer [17] introduced the following notion of strategies.

Definition 3.1.6. (Non anticipating strategies with delay, first definition)

i) A **non anticipating strategy with delay (NAD)** for player 1 is a map $\alpha : \mathcal{V} \to \mathcal{U}$ such that, for some finite partition $t_1 < \cdots < t_N = 1$ of $[t_1, 1]$, for all $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{U}$ and $1 \leq m < N$:

$$\mathbf{v}_1 \equiv \mathbf{v}_2 \text{ a.e. on } [t_1, t_m] \implies \alpha(\mathbf{v}_1) \equiv \alpha(\mathbf{v}_2) \text{ a.e. on } [t_1, t_{m+1}].$$

ii) Similarly, a **non anticipating strategy with delay (NAD)** for player 2 is a map $\beta : \mathcal{U} \to \mathcal{V}$ such that, for some finite partition $t_1 < \cdots < t_N = 1$ of $[t_1, 1]$, for all $\mathbf{u}_1, \mathbf{u}_2 \in \mathcal{U}$ and $1 \leq m < N$:

$$\mathbf{u}_1 \equiv \mathbf{u}_2 \text{ a.e. on } [t_1, t_m] \implies \beta(\mathbf{u}_1) \equiv \beta(\mathbf{u}_2) \text{ a.e. on } [t_1, t_{m+1}].$$

Note that NAD strategies are a subset of NA strategies. Let $\mathcal{A}(t_1)$ and $\mathcal{B}(t_1)$ denote respectively the sets of non anticipating strategies for player 1 and 2. With this notion of strategies, one has the following Lemma:

Lemma 3.1.7. *([17, Lemma 2.4])Let $(\alpha, \beta) \in \mathcal{A}(t_1) \times \mathcal{B}(t_1)$ where at least one of the strategies is NAD. Then there exist unique controls $(\mathbf{u}, \mathbf{v}) \in \mathcal{U}(t_1) \times \mathcal{V}(t_1)$ such that:*

$$\alpha(\mathbf{v}) = \mathbf{u}, \; \beta(\mathbf{u}) = \mathbf{v}, \; \textit{a.e. in } [t_1, 1].$$

Denote by $\mathbf{x}[t, x, \alpha, \beta](s)$ the trajectory corresponding to the couple of controls $(\mathbf{u}, \mathbf{v})$ associated to the strategies (α, β).

An equivalent definition of non anticipating strategies is the following.

Definition 3.1.8. (Non anticipating strategies with delay, second definition)

i) A **non anticipating strategy with delay (NAD)** for player 1 is a map $\alpha : \mathcal{V} \to \mathcal{U}$ for which there is a delay $\tau > 0$ such that for all $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$ and for all t:

$$\mathbf{v}_1 \equiv \mathbf{v}_2 \text{ a.e. on } [t_1, t] \implies \alpha(\mathbf{v}_1) \equiv \alpha(\mathbf{v}_2) \text{ a.e. on } [t_1, t + \tau].$$

ii) Analogously, a **non anticipating strategy with delay (NAD)** for player 2 is a map $\beta : \mathcal{U} \to \mathcal{V}$ for which there is a delay $\tau > 0$ such that for all $\mathbf{u}_1, \mathbf{u}_2 \in \mathcal{U}$ and for all t:

$$\mathbf{u}_1 \equiv \mathbf{u}_2 \text{ a.e. on } [t_1, t] \implies \beta(\mathbf{u}_1) \equiv \beta(\mathbf{u}_2) \text{ a.e. on } [t_1, t+\tau].$$

Lemma 3.1.9. *Definition 3.1.6 and Definition 3.1.8 are equivalent.*

Proof. Let us note that any strategy satisfying $i)$ in Definition 3.1.6 has the property $i)$ in Definition 3.1.8. To see this, let $t \in [0,1]$ and k such that $t \in [t_k, t_{k+1})$ and set $\tau_t := t_{k+1} - t$. Note that the family of intervals $([t, \tau_t])_{t \in [t_1,1]}$ forms a cover of the compact interval $[t_1, 1]$, from which we extract a finite subcover $[t_1, \tau_{t_1}], [t_2, \tau_{t_2}], \dots, [t_M, \tau_{t_M}]$ and set $\tau := \min_{1 \le m \le M} \tau_{t_m} - t_m$. For the converse, let α and τ as in $i)$, Definition 3.1.8. Let k such that $k\tau \le 1 - t_1 < (k+1)\tau$. Then the partition $t_1, t_1 + \tau, \dots t_1 + k\tau$ satisfies $i)$ in Definition 3.1.6.

□

We denote by $\mathcal{A}_d, \mathcal{B}_d$ the sets of NAD strategies for player 1 and 2 respectively.

Non anticipating strategies capture the minimal requirements of a strategy: they forbid players to see the future. This class is however not satisfactory as the discussion above shows. We conclude this Section with another important difference:

Remark. Let us point out an important difference between NA and NAD strategies. The reaction of, say, player 1 at time t using a NA strategies against the control of player 2 may use the value of the control of player 2 at time t, whereas in the case of NAD strategies this is not the case. NA strategies correspond to alternate moves at time t, where the player using the strategies moves after the player using the controls, while NAD strategies correspond to simultaneous moves.

3.1.1 Hamilton-Jacobi-Isaacs equations

Depending on the class of strategies, we can define the appropriate value functions. Let us first consider the case where we allow the players to use non anticipating strategies.

Definition 3.1.10. (Definitions of the value functions, NA strategies)

i) The **lower value** function is defined by

$$\mathbf{V}_L(t,x) := \inf_{\beta \in \mathcal{B}} \sup_{\mathbf{u} \in \mathcal{U}(t)} g(\mathbf{x}[t,x,\mathbf{u},\beta(\mathbf{u})](1)).$$

ii) The **upper value** function is defined by

$$\mathbf{V}_U(t,x) := \sup_{\alpha \in \mathcal{A}} \inf_{\mathbf{v} \in \mathcal{V}(t)} g(\mathbf{x}[t,x,\alpha(\mathbf{v}),\mathbf{v}](1)).$$

When we restrict the players to use NAD strategies, since the game can be defined in normal form, it makes sense to define the maxmin and minmax as in classical game theory:

Definition 3.1.11. (Definitions of the value functions, NAD strategies)

i) The **maxmin** is defined by

$$\mathbf{V}^-(t,x) := \sup_{\alpha \in \mathcal{A}_d} \inf_{\beta \in \mathcal{B}_d} g(\mathbf{x}[t,x,\alpha,\beta](1)).$$

ii) The **minmax** is defined by

$$\mathbf{V}^+(t,x) := \inf_{\beta\in\mathcal{B}_d}\sup_{\alpha\in\mathcal{A}_d} g(\mathbf{x}[t,x,\alpha,\beta](1)).$$

One common feature of these value functions is that they are related to the following partial differential equations (PDE):

$$\frac{\partial w^-}{\partial t}(t,x) + \sup_{u\in U}\inf_{v\in V}\left\langle f(x,u,v), \nabla_x w^-(t,x)\right\rangle = 0 \tag{3.4a}$$

$$\frac{\partial w^+}{\partial t}(t,x) + \inf_{v\in V}\sup_{u\in U}\left\langle f(x,u,v), \nabla_x w^+(t,x)\right\rangle = 0 \tag{3.4b}$$

with boundary conditions $w^-(1,x) = w^+(1,x) = g(x)$.

More precisely, the lower value and the maxmin are viscosity solutions as introduced by Crandall and Lions [33]) and whose definition will be recalled later, of (3.4a) while the upper value and the minmax are viscosity solutions of (3.4b).

Although heuristically derived by Isaacs' [58, Section 4.2], the connection between PDE's and differential games was first made explicit in the framework of viscosity solutions [33] by Evans and Souganidis [36]. The PDE's (3.4a) and (3.4b) are respectively called **lower and upper Hamilton-Jacobi-Isaacs equations** (HJI equations).

If the terms in the above equations involving the sup and the inf coincide, the resulting PDE has a unique viscosity solution which is the **value** of the game. In particular, the value is independent of the class of strategies used (NA or NAD).

Let $\mathcal{G}(x,p)$ denote the **local game** with payoff $\langle f(x,u,v),p\rangle$, where player 1 chooses $u\in U$ and player 2 chooses $v\in V$. Under the following Assumption, the above equations reduce to one:

Assumption 3.1.12. *(Isaacs' condition) We assume that the local game $\mathcal{G}(x,p)$ has a value, for all $x,p\in\mathbb{R}^n$. Explicitely,*

$$\sup_{u\in U}\inf_{v\in V}\langle f(x,u,v),p\rangle = \inf_{v\in V}\sup_{u\in U}\langle f(x,u,v),p\rangle \tag{3.5}$$

holds for all $x,p\in\mathbb{R}^n$.

So far we have described the general problem and introduced two auxiliary PDE's which will play a role in the sequel. Let us now briefly sketch three approaches to solve the game $\Gamma(t_1,x_1)$. The rest of this chapter is devoted to a more detailed description of them.

In Section 3.2, we describe the **discrete game approach:** this idea goes back to Fleming [38] and Friedman [41, 42]. Here, one studies a game in discrete time where players are allowed to choose their actions on the nodes of a time partition. Under Assumptions 3.1.1 and 3.1.3, the upper and lower value functions of the finite games are uniformly Lipschitz with respect to the partition and hence the set of accumulation points as the mesh of the partition goes to zero is nonempty. The upper and lower value functions of the finite games satisfy a recursive formula, from which one can guess the limit equations (3.4a) and (3.4b). One then uses this limit equation to prove uniqueness of the accumulation points, which can be done in several ways, as we will see. Note however that this approach is not really concerned with defining an interaction in continuous time. The interaction is in discrete time and instead one deduces properties of the value functions of the finite games via the solutions of the HJI equations.

The next two approaches rely on defining the value functions of the game in continuous time game. The value functions satisfy recursive formulas analogous to those appearing in the discrete game approach and turn out to be solutions of a PDE, in a generalized (but equivalent) sense. The main difference between these two approaches is in the way the value functions are used to construct strategies.

In Section 3.3 we turn our attention to the **viscosity approach**. This is initiated by the intuition of Isaacs [58, p.67], who guessed that the value function should be, whenever smooth, a classical solution of the HJI equation, assuming (3.5). Evans and Souganidis [36] proved that the value functions defined using non anticipating strategies are indeed viscosity solution of the PDE's guessed by Isaacs. The machinery of viscosity solutions helps to appropriately answer to Question 1). We can use the solution of the HJI equation to answer Question 2) as well: when the value function is smooth, one player will steer the state in the direction of the gradient of the value, while his adversary will steer on the opposite direction. In the non-smooth case, one can still do this by replacing the gradient of $\mathbf{V}$ at a singular point x by the gradient of $\mathbf{V}$ at a neighboring regular point [97, Sections 3 and 5]. This however requires a precise knowledge of the value function: one can easily think of examples of real valued functions which are close in the uniform norm but their derivatives are very different.

The third approach, described in Section 3.4, is the **strategic approach**. This was initiated by Krasovskii and Subbotin [61, 97]. Following their ideas, one obtains the existence and characterization of the value function by introducing a notion of generalized solutions for the HJI equation (minimax solutions [96]), which are equivalent to viscosity solutions. Both solution concepts rely on very different techniques and have different motivations. The strategic approach provides an explicit construction of $\epsilon-$optimal strategies. An important advantage of these strategies is that they are more robust than those that can be obtained by the PDE approach with respect to measurement errors or imprecise knowledge of the value function.

Finally, in Section 3.5 we point out some connections between these approaches.

3.2 The discrete game approach

We will describe three different ways of associating a discrete game to the differential game $G(t,x)$. The common feature is that decisions are taken at discrete times only. What differs is either the sequence in which the actions are chosen (simultaneous or alternate moves), the dynamics and the way the players are allowed to update their actions. Let $\Pi_n = \{t = t_1 < t_2 \ldots < t_{n+1} = 1\}$ denote a finite partition. Let $\pi_{k+1} := t_{k+1} - t_k$ and $\|\Pi_n\| = \max_{k=1,\ldots n} \pi_k$ denote its **mesh** or **norm**.

3.2.1 The simultaneous Fleming value.

The approach we describe here goes back to [38]. We define the game $G_\Pi(t,x)$ starting at $t_1 = t$, $x_1 := x$ and repeated n times as follows: at time t_k, $k = 1, \ldots n$, both players remember the **history** $H_k := \{u_1, v_1, u_2, v_2, \ldots u_{k-1}, v_{k-1}\}$ of past actions and thus they know the current state x_k. They choose simultaneously and independently actions u_k and v_k using the lotteries $\sigma(t_k, x_k) \in \Delta_f(U)$ and $\tau(t_k, x_k) \in \Delta_f(V)$, where $\Delta_f(U)$ denotes the set of probability measures on U with finite support and σ, τ are behavioural strategies for player 1 and 2, that is, $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_n), \tau = (\tau_1, \tau_2, \ldots, \tau_n)$ where $\sigma_k : H_k \to \Delta_f(U)$ and $\tau_k : H_k \to \Delta_f(V)$. Denote by $\Sigma_\pi, \mathcal{T}_\pi$ the strategy sets of player 1 and player 2.

The state evolves according to

$$x_{k+1} = x_k + \pi_{k+1} f(x_k, u_k, v_k). \tag{3.6}$$

At time t_{k+1}, the players observe the actions chosen at time t_k, and thus they know the new state x_{k+1}. The game is repeated until $k = n$. After the n-th move, player 1 receives from player 2 a payoff $g(x_{n+1})$. We introduce the minmax and the maxmin:

$$\begin{aligned} \mathbf{V}^+_\Pi(t_k, x) &= \inf_{\tau \in \mathcal{T}_\Pi} \sup_{\sigma \in \Sigma_\Pi} \mathbb{E}_{\sigma\tau} g(x_{n+1}), \\ \mathbf{V}^-_\Pi(t_k, x) &= \sup_{\sigma \in \Sigma_\Pi} \inf_{\tau \in \mathcal{T}_\Pi} \mathbb{E}_{\sigma\tau} g(x_{n+1}). \end{aligned}$$

The first result is the existence of the value of the game $G_\Pi(t, x)$.

Proposition 3.2.1. *([38, Theorem 1])For every interval $[t, 1]$, every initial position x and every partition $\Pi = \{t = t_1 < t_2 \ldots < t_{n+1} = 1\}$, the finitely repeated game $G_\Pi(t, x)$ has a value, denoted $\mathbf{V}_\Pi(t, x)$, which is a Lipschitz function of x. Moreover, for $k = 1, \ldots, n-1$ the following recursive formula holds:*

$$\mathbf{V}_\Pi(t_k, x) = \mathrm{val}_{(\sigma,\tau) \in \Delta_f(U) \times \Delta_f(V)} \mathbb{E}_{\sigma,\tau} \left[\mathbf{V}_\Pi(t_{k+1}, x + \pi_{k+1} f(x, \tilde{u}, \tilde{v})) \right] \tag{3.7}$$

Proof. We prove first the existence of the value. We proceed by induction on the number of nodes of Π. If $n = 1$, the game becomes a one-shot game with payoff

$$g(x + (1 - t) f(x, u, v))$$

which is continuous on u and v and hence it has a value on mixed strategies with finite support [81, Theorem, p. 750]. Note that the value of this game is Lipschitz in x by the Lipschitz assumption of g. This and the recursive formula (3.7) imply the Lipschitz continuity of the value function with respect to x, with the same constant L_g, so it suffices to prove the recursive formula. Assume (3.7) holds for all partitions with $n \leq m$. Let $\Pi_{m+1} = \{t = t_1 < t_2 \ldots < t_{m+2} = 1\}$. Consider the game starting at $t = t_2$, which has m stages, and let $\Sigma_m, \mathcal{T}_m$ denote the respective strategy sets. By induction, the value exists and is continuous in x and thus the one-shot game with payoff $\mathbf{V}_{\Pi_{m+1}}(t_2, x + \pi_2 f(x, u, v))$ has a value $\tilde{\mathbf{V}}$. Let $\mathbf{V}^+_{\Pi_{m+1}}(t_k, x)$ and $\mathbf{V}^-_{\Pi_{m+1}}(t_k, x)$ the upper and lower values. Explicitely,

$$\begin{aligned} \mathbf{V}^+_{\Pi_{m+1}}(t_k, x) &= \inf_{\tau \in \mathcal{T}_m} \sup_{\sigma \in \Sigma_m} \mathbb{E}_{\sigma\tau} g(x_{m+2}), \\ \mathbf{V}^-_{\Pi_{m+1}}(t_k, x) &= \sup_{\sigma \in \Sigma_m} \inf_{\tau \in \mathcal{T}_m} \mathbb{E}_{\sigma\tau} g(x_{m+2}). \end{aligned}$$

It is easy to see that $\mathbf{V}^+_{\Pi_{m+1}} \geq \mathbf{V}^-_{\Pi_{m+1}}$. We will briefly sketch the proof for $\mathbf{V}^-_{\Pi_{m+1}} \geq \tilde{\mathbf{V}}$. Let μ be $\epsilon-$optimal in $\tilde{\mathbf{V}}$. Since f is continuous and bounded, the state x_2 belongs to a compact set $\mathcal{C}$ regardless of players' choices since their action spaces are also compact. Let $\{C_i\}_{i \in I}$ denote a finite partition of $\mathcal{C}$ with diameter ϵ and take points $y_i \in C_i$. Observe that, for $y \in C_i$, any $\epsilon-$ optimal strategy on the game $G_{\Pi_m}(t_2, y_i)$ is 2ϵ-optimal in $G_{\Pi_m}(t_2, y)$. Define the strategy σ as follows: play first μ and then the optimal strategy of $G_{\Pi_m}(t_2, y_i)$, if $x_2 \in C_i$. This strategy ensures a payoff of at least $\tilde{\mathbf{V}} + 3\epsilon$. Reversing the roles of the players, $\mathbf{V}^+_{\Pi_{m+1}} \leq \tilde{\mathbf{V}}$. This concludes the proof. □

We extend $\mathbf{V}_\Pi$ to $[t, 1]$ by linear interpolation and we still denote $\mathbf{V}_\Pi$ this extension. We have the following property:

Proposition 3.2.2. *Let* $\Pi = \{t = t_1 < t_2 \ldots < t_{n+1} = 1\}$ *be a partition. There exists a constant* K *independent of* Π *such that*

$$|\mathbf{V}_\Pi(t,x) - \mathbf{V}_\Pi(s,x)| \leq K|t-s|.$$

for all (t,s,x).

Proof. First observe that it suffices to prove the above formula for t, s consecutive nodes of the partition Π. Then from the recursive formula (3.7) and the Lipschitz continuity of $\mathbf{V}_\Pi(t,x)$ with respect to x we obtain:

$$\begin{aligned} |\mathbf{V}(t_k,x) - \mathbf{V}(t_{k+1},x)| &\leq |\mathbf{V}(t_{k+1}, x + \pi_{k+1}\|f\|) - \mathbf{V}(t_{k+1},x)| \\ &\|f\| \cdot L_g. \end{aligned}$$

The result now follows with $K = \|f\| \cdot L_g$. □

From Arzelà-Ascoli's theorem, it follows that there exist a subsequence Π_m of partitions such that $\mathbf{V}_{\Pi_m}$ converges uniformly in compact sets to a function $\mathbf{V}$. We will characterize this limit function in Section 3.5.1

3.2.2 The Fleming value

By allowing the players to randomize their moves as in Section 3.2.1, one gets automatically existence of the value. But this is not necessary for some asymptotic properties. Let us consider first the upper and lower values in NA strategies. This corresponds to games with alternating moves.

The players here play piecewise constant actions. Abusing the notation, let $\vec{u}, \vec{v}$ denote the piecewise constant functions corresponding to the vectors $\vec{u} = (u_1, u_2, \ldots, u_n)$ and $\vec{v} = (v_1, v_2, \ldots, v_n)$.

Discrete lower and upper values

Let us consider a game with the same dynamics as in (3.6) played on the nodes of the partition Π. Instead of choosing their actions simultaneously and independently, players will take turns. We will consider two auxiliary games, one in which player 1 knows the move of player 2 before choosing his own and another game in which now player 2 knows the move of player 1 in advance. As we discussed in Section 3.1, this can be done with non anticipating strategies, whose definition for finite games is as follows:

Definition 3.2.3. (NA strategies, discrete time). Let $\Pi = \{t_1 < t_2, \ldots t_{n+1} = 1\}$ be a fixed partition.

i) A **non anticipating strategy** α_Π for player 1 is a function $\alpha_\Pi : V^n \to U^n$ such that, for all $k = 1, 2, \ldots n$

$$\vec{v}_1 \equiv \vec{v}_2 \text{ a.e. on } [t_1, t_k] \implies \alpha_\Pi(\vec{v}_1) \equiv \alpha_\Pi(\vec{v}_2) \text{ a.e. on } [t_1, t_k].$$

ii) A **non anticipating strategy** β_Π for player 2 is a function $\beta_\Pi : U^n \to V^n$ such that, for all $k = 1, 2, \ldots n$

$$\vec{u}_1 \equiv \vec{u}_2 \text{ a.e. on } [t_1, t_k] \implies \beta_\Pi(\vec{u}_1) \equiv \beta_\Pi(\vec{u}_2) \text{ a.e. on } [t_1, t_k].$$

Let $\mathcal{A}_\Pi, \mathcal{B}_\Pi$ denote the strategy sets of player 1 and player 2 respectively. This allows us to define the value functions:

$$\begin{aligned}\mathbf{W}^-_\Pi(t_k, x) &:= \inf_{\beta\in\mathcal{B}_\Pi} \sup_{\mathbf{u}\in U^n} g(x_{n+1})\\ \mathbf{W}^+_\Pi(t_k, x) &:= \sup_{\alpha\in\mathcal{A}_\Pi} \inf_{\mathbf{v}\in V^n} g(x_{n+1}).\end{aligned}$$

These are the values of the **minorant game** $G^-_\Pi(t, x)$ in which player 1 chooses his action before player 2 and the **majorant game** $G^+_\Pi(t, x)$ in which the opposite happens.

Proposition 3.2.4. *The value function of the minorant game and majorant games satisfy respectively the following recursive formulae:*

$$\mathbf{W}^-_\Pi(t_k, x) := \max_{u\in U} \min_{v\in V} \mathbf{W}^-_\Pi(t_{k+1}, x + \pi_{k+1} f(x, u, v)). \tag{3.8a}$$

$$\mathbf{W}^+_\Pi(t_k, x) := \min_{v\in V} \max_{u\in U} \mathbf{W}^+_\Pi(t_{k+1}, x + \pi_{k+1} f(x, u, v)).. \tag{3.8b}$$

By similar arguments as in Section 3.2.1 one can prove the following:

Proposition 3.2.5. *The value functions of the minorant and majorant game given by* (3.8a) *and* (3.8b) *are uniformly Lipschitz and hence they have an accumulation point.*

We will postpone the characterization of the accumulation points to Section 3.5.1.

Discrete maxmin and minmax

Let us consider here the discrete maxmin and minmax.
We introduce first the corresponding definition of NAD strategies in discrete time.

Definition 3.2.6. (NAD strategies, discrete time). Let $\Pi = \{t_1 < t_2, \dots t_{n+1} = 1\}$ be a fixed partition.

i) A **non anticipating strategy with delay** α_Π for player 1 is a function $\alpha_\Pi : V^n \to U^n$ such that, for all $k = 1, 2, \dots n$

$$\vec{v}_1 \equiv \vec{v}_2 \text{ a.e. on } [t_1, t_k] \implies \alpha_\Pi(\vec{v}_1) \equiv \alpha_\Pi(\vec{v}_2) \text{ a.e. on } [t_1, t_{k+1}].$$

ii) A **non anticipating strategy with delay** β_Π for player 2 is a function $\beta_\Pi : U^n \to V^n$ such that, for all $k = 1, 2, \dots n$

$$\vec{u}_1 \equiv \vec{u}_2 \text{ a.e. on } [t_1, t_k] \implies \beta_\Pi(\vec{u}_1) \equiv \beta_\Pi(\vec{u}_2) \text{ a.e. on } [t_1, t_{k+1}].$$

Let $\mathcal{A}_{\Pi,d}, \mathcal{B}_{\Pi,d}$ denote the strategy sets of player 1 and player 2 respectively. This allows us to define the discrete maxmin and minmax:

$$\mathbf{W}^+_{\Pi,d}(t_k, x) := \inf_{\beta\in\mathcal{B}_{\Pi,d}} \sup_{\vec{u}\in U^n} g(x_{n+1}) \tag{3.9a}$$

$$\mathbf{W}^-_{\Pi,d}(t_k, x) := \sup_{\alpha\in\mathcal{A}_{\Pi,d}} \inf_{\vec{v}\in V^n} g(x_{n+1}) \tag{3.9b}$$

We will postpone the characterization of the accumulation points to Section 3.5.2.

3.2.3 The Friedman value

In Friedman's discretization [41], the majorant and minorant games are played as follows: in the minorant game, at time t_k both players observe the state and player 1 chooses first the measurable control function he will use on the interval $[t_k, t_{k+1}]$. His choice is announced to player 2, which in turn chooses his measurable control function. The state evolves according to (3.1) on the interval $[t_k, t_{k+1}]$ and the situation is repeated. Let $\mathcal{U}_k$ and $\mathcal{V}_k$ denote the following sets:

$$\begin{aligned} \mathcal{U}_k &:= \{\mathbf{u} \circ \phi_k \,:\, \mathbf{u} \in \mathcal{U}\} \\ \mathcal{V}_k &:= \{\mathbf{v} \circ \phi_k \,:\, \mathbf{v} \in \mathcal{V}\} \end{aligned}$$

where $\phi_k : [t_k, t_{k+1}] \to [0, 1]$ is the coordinate mapping $t \mapsto \frac{t-t_k}{\pi_k}$.

The value functions for the minorant and majorant games are

$$\begin{aligned} \mathbf{W}^-_{F,\Pi}(t_k, x) &:= \inf_{\beta \in \mathcal{B}_\Pi} \sup_{\mathbf{u} \in \mathcal{U}} g(x_{n+1}) \\ \mathbf{W}^+_{F,\Pi}(t_k, x) &:= \sup_{\alpha \in \mathcal{A}_\Pi} \inf_{\mathbf{v} \in \mathcal{V}} g(x_{n+1}). \end{aligned}$$

In a similar way, the value functions satisfy a dynamic programming principle and analogous regularity properties.

Proposition 3.2.7. *The recursive formula for the value of the minorant game and majorant game are respectively:*

$$\mathbf{W}^-_{F,\Pi}(t_k, x) := \max_{\mathbf{u} \in \mathcal{U}_k} \min_{\mathbf{v} \in \mathcal{V}_k} \mathbf{W}^-_{F,\Pi}(t_{k+1}, x + \int_{t_k}^{t_{k+1}} f(\mathbf{x}[t_k, x, \mathbf{u}, \mathbf{v}](s), \mathbf{u}(s), \mathbf{v}(s))ds) \tag{3.10a}$$

$$\mathbf{W}^+_{F,\Pi}(t_k, x) := \min_{\mathbf{v} \in \mathcal{V}_k} \max_{\mathbf{u} \in \mathcal{U}_k} \mathbf{W}^+_{F,\Pi}(t_{k+1}, x + \int_{t_k}^{t_{k+1}} f(\mathbf{x}[t_k, x, \mathbf{u}, \mathbf{v}](s), \mathbf{u}(s), \mathbf{v}(s))ds) \tag{3.10b}$$

Moreover, these value functions are uniformly Lipschitz and hence they have an accumulation point.

3.3 The viscosity approach

Let us recall the definitions of the lower and upper value functions, in the class of NA strategies:

$$\begin{aligned} \mathbf{V}^-_L(t_1, x_1) &:= \inf_{\beta \in \mathcal{B}} \sup_{\mathbf{u} \in \mathcal{U}} g\big(\mathbf{x}[t_1, x_1, \mathbf{u}, \beta[\mathbf{u}]](1)\big), \\ \mathbf{V}^+_U(t_1, x_1) &:= \sup_{\alpha \in \mathcal{A}} \inf_{\mathbf{v} \in \mathcal{V}} g\big(\mathbf{x}[t_1, x_1, \alpha[\mathbf{v}], \mathbf{v}](1)\big). \end{aligned}$$

As before, when we consider NAD strategies, the maxmin and the minmax are:

$$\mathbf{V}^-(t_1, x_1) := \sup_{\alpha \in \mathcal{A}_d} \inf_{\beta \in \mathcal{B}_d} g\big(\mathbf{x}[t_1, x_1, \alpha, \beta](1)\big), \tag{3.12}$$

$$\mathbf{V}^+(t_1, x_1) := \inf_{\beta \in \mathcal{B}_d} \sup_{\alpha \in \mathcal{A}_d} g\big(\mathbf{x}[t_1, x_1, \alpha, \beta](1)\big). \tag{3.13}$$

Naturally, $\mathbf{V}^-(1,x) = \mathbf{V}^+(1,x) = g(x)$. The inequality $\mathbf{V}^- \leq \mathbf{V}^+$ holds everywhere. If $\mathbf{V}^-(t_1,x_1) = \mathbf{V}^+(t_1,x_1)$, the game $\Gamma(t_1,x_1)$ has a value, denoted by $\mathbf{V}(t_1,x_1)$.

Under Assumptions 3.1.3 and 3.1.1 and from the familiar arguments [25, Proposition 1 and 2] and [22, Lemma 3.3], we collect the following properties of the value functions in the next proposition:

Proposition 3.3.1. *(Dynamic programming) The maxmin and the minmax are Lipschitz continuous and they satisfy the following dynamic programming principle: for all $(t,x) \in [0,1] \times \mathbb{R}^n$ and all $t' \in [t,1]$,*

$$\mathbf{V}^-(t,x) = \sup_{\alpha\in\mathcal{A}_d} \inf_{\beta\in\mathcal{B}_d} \mathbf{V}^-(t',\mathbf{x}[t,x,\alpha,\beta](t')) \tag{3.14a}$$

$$\mathbf{V}^+(t,x) = \inf_{\beta\in\mathcal{B}_d} \sup_{\alpha\in\mathcal{A}_d} \mathbf{V}^+(t',\mathbf{x}[t,x,\alpha,\beta](t')). \tag{3.14b}$$

The above dynamic programming principle is equivalent to

$$\mathbf{V}^-(t,x) = \sup_{\alpha\in\mathcal{A}_d} \inf_{\mathbf{v}\in\mathcal{V}} \mathbf{V}^-(t',\mathbf{x}[t,x,\alpha[\mathbf{v}],\mathbf{v}](t')) \tag{3.15a}$$

$$\mathbf{V}^+(t,x) = \inf_{\beta\in\mathcal{B}_d} \sup_{\mathbf{u}\in\mathcal{U}} \mathbf{V}^+(t',\mathbf{x}[t,x,\mathbf{u},\beta[\mathbf{u}]](t')). \tag{3.15b}$$

Naturally, a similar dynamic programming principle holds for the upper and lower value functions. For the rest of the chapter, we will focus on NAD strategies only.

From the dynamic programming equations and a Taylor series expansion around (t,x), one can deduce heuristically the following partial differential equations satisfied by the value functions

$$\frac{\partial \mathbf{V}^-}{\partial t}(t,x) + H^-(x,\nabla_x \mathbf{V}^-(t,x)) = 0 \tag{3.16a}$$

$$\frac{\partial \mathbf{V}^+}{\partial t}(t,x) + H^+(x,\nabla_x \mathbf{V}^+(t,x)) = 0 \tag{3.16b}$$

with boundary condition $\mathbf{V}^-(1,x) = \mathbf{V}^+(1,x) = g(x)$, where

$$H^-(x,p) := \sup_{u\in U} \inf_{v\in V} \langle f(x,u,v),p\rangle \tag{3.17a}$$

$$H^+(x,p) := \inf_{v\in V} \sup_{u\in U} \langle f(x,u,v),p\rangle \tag{3.17b}$$

are called the **lower and upper hamiltonians** respectively.

The equations (3.16a) and (3.16b) were heuristically derived by Isaacs [58, p.67]. In particular, if Isaacs' condition holds (c.f. (3.5))

$$H^-(x,p) = H^+(x,p) =: H(x,p) \ \ \forall (x,p) \in \mathbb{R}^n \times \mathbb{R}^n$$

and if the PDE

$$\frac{\partial \mathbf{V}}{\partial t}(t,x) + H(x,\nabla_x \mathbf{V}(t,x)) = 0 \tag{3.18}$$

with boundary condition $\mathbf{V}(1,x) = g(x)$ has a unique solution, this characterizes the value function.

However, there are two problems here: the first one is that the solution is seldom regular enough, regardless of the smoothness of the boundary condition. So one needs to come out with an appropriate notion of solution for the PDE (3.18). Evans and Souganidis [36] proved that the correct interpretation of solution for this PDE is the notion of viscosity solutions:

Definition 3.3.2. (Viscosity solutions, [33])

— A lower semicontinuous function $w : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ is a **viscosity supersolution** of (3.16a) if for any $(t,x) \in [0,1]\times\mathbb{R}^n$ and a C^1 test function $\phi : [0,1]\times\mathbb{R}^n \to \mathbb{R}$ such that $\phi(t,x) = w(t,x)$ and $\phi(s,y) \leq w(s,y)$ for (s,y) on a neighbourhood of (t,x), then

$$\frac{\partial \phi}{\partial s}(t,x) + H^-(x, \nabla_y \phi(t,x)) \leq 0. \tag{3.19}$$

— An upper semicontinuous function $w : [0,1]\times\mathbb{R}^n \to \mathbb{R}$ is a **viscosity subsolution** of (3.16b) if for any $(t,x) \in [0,1] \times \mathbb{R}^n$ and a C^1 test function $\phi : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ such that $\phi(t,x) = w(t,x)$ and $\phi(s,y) \geq w(s,y)$ for (s,y) on a neighbourhood of (t,x), then

$$\frac{\partial \phi}{\partial s}(t,x) + H^+(x, \nabla_y \phi(t,x)) \geq 0. \tag{3.20}$$

— A continuous function is a **viscosity solution** if it is both a viscosity super and subsolution.

Note that this definition of viscosity solutions has the signs reversed in the inequalities with respect to that in [33] since we are dealing with problems with terminal conditions instead of initial conditions. Evans and Souganidis [36] proved that the upper and lower value functions (3.13) and (3.13) are solutions of (3.16a) and (3.16b) in the viscosity sense and hence if Isaacs' condition (3.5) holds, the solution is unique.

3.3.1 Strategies in the PDE approach: the smooth case

We will describe now how the players can use the value function to construct the strategies and to prove that the proposed construction is indeed optimal. This is called a *verification theorem*. The purpose of this theorem is to illustrate the interest of computing the value function and how answering Question 1) helps to answer Question 2). The assumptions of this theorem are however rarely satisfied, its main purpose is being pedagogical. A first version of Verification Theorem is also due to Isaacs, see [58, Theorem 4.4.1].

Theorem 3.3.3. *Assume that* (3.18) *has a* C^1 *solution. Let us furthermore assume that there exist measurable maps* $u^* : [0,1] \times \mathbb{R}^n \to U$, $v^* : [0,1] \times \mathbb{R}^n \to V$ *such that:*

$$\mathbf{u}^*(t,x) \in \operatorname{argmax}_{u\in U} \left\{ \min_{v\in V} \langle \nabla_x \mathbf{V}(t,x), f(x,u,v)\rangle \right\} \tag{3.21a}$$

$$\mathbf{v}^*(t,x) \in \operatorname{argmin}_{v\in V} \left\{ \max_{u\in U} \langle \nabla_x \mathbf{V}(t,x), f(x,u,v)\rangle \right\}. \tag{3.21b}$$

Then, for all $(t_1,x_1) \in [0,1] \times \mathbb{R}^n$,

$$\mathbf{V}(t_1,x_1) = \inf_{\mathbf{v}\in\mathcal{V}} g\left(\mathbf{x}[t_1,x_1,\mathbf{u}^*,\mathbf{v}](1)\right) = \sup_{\mathbf{u}\in\mathcal{U}} g\left(\mathbf{x}[t_1,x_1,\mathbf{u},\mathbf{v}^*](1)\right) = g\left(\mathbf{x}[t_1,x_1,\mathbf{u}^*,\mathbf{v}^*](1)\right). \tag{3.22}$$

Proof. Take $\mathbf{v} \in \mathcal{V}$ and denote by $\mathbf{x}^1(t) := \mathbf{x}[t_1, x_1, \mathbf{u}^*, \mathbf{v}](t)$ and let $\mathbf{u}^*(t) := \mathbf{u}^*(t, \mathbf{x}^1(t))$ and $\mathbf{v}(t) := \mathbf{v}(t, \mathbf{x}^1(t))$ for simplicity. Then we have that:

$$\begin{aligned}
\frac{d}{dt}\mathbf{V}(t, \mathbf{x}^1(t)) &= \frac{\partial \mathbf{V}}{\partial t}(t, x) + \left\langle \nabla_x \mathbf{V}(t, \mathbf{x}^1(t)), f(\mathbf{x}^1(t), \mathbf{u}^*(t), \mathbf{v}(t)) \right\rangle \\
&\geq \frac{\partial \mathbf{V}}{\partial t}(t, x) + \inf_{v \in V} \left\langle \nabla_x \mathbf{V}(t, \mathbf{x}^1(t)), f(\mathbf{x}^1(t), \mathbf{u}^*(t), v) \right\rangle \\
&= 0.
\end{aligned}$$

Integrating from t_1 to 1, we get that:

$$\mathbf{V}(1, \mathbf{x}^1(1)) = g(\mathbf{x}^1(1)) \geq \mathbf{V}(t_1, x_1).$$

We conclude from this that $\mathbf{V}^-(t_1, x_1) \geq \mathbf{V}(t_1, x_1)$. In a completely analogous way, one proves that $\mathbf{V}^+(t_1, x_1) \leq \mathbf{V}(t_1, x_1)$ and hence $\mathbf{V}^-(t_1, x_1) = \mathbf{V}^+(t_1, x_1) = \mathbf{V}(t_1, x_1)$. □

Remark. This theorem is rarely useful, as typically the value functions are not smooth. However, it is possible to construct $\epsilon-$optimal strategies in a similar way. Instead of using the derivatives of the value function, one can consider derivatives of a sequence of smooth functions that converge to the value function. We will not detail this here, but we refer to [97, Chapter 1, Section 5] for a detailed description.

3.4 The strategic approach

We describe here the strategic approach, introduced by Krasovskii and Subbotin [61]. We reformulate their ideas in a modern language and clarify their proofs. For the sake of brevity, we will not do all detailed proofs here, but rather motivate the main ideas and refer the reader to Chapter 4 in this thesis or to our paper [71] for the complete proofs.

To prove the existence of the value, one needs to prove the inequality $\mathbf{V}^+ \leq \mathbf{V}^-$. One achieves this by showing that for every ϵ, player 2 has a NAD strategy such that he can ensure that his payoff is below $\mathbf{V}^- + \epsilon$. From the dynamic programming equation (3.14a) it is easy to prove that $\mathbf{V}^-$ satisfies the following inequality, for $t_1 \leq t' \leq 1$:

$$\forall (t_1, x_1) \quad \mathbf{V}^-(t, x) \geq \sup_{u \in U} \inf_{\mathbf{v} \in \mathcal{V}} \mathbf{V}^-(t', \mathbf{x}[t_1, x_1, u, \mathbf{v}](t')). \tag{3.23}$$

If S is a closed subset of $\mathbb{R}^n$ and $y \in \mathbb{R}^n$ denote with $d(y, S) := \inf_{p \in S} \|y - p\|$ the usual distance from a point to a set.

Denote by

$$\mathcal{W}^- := \{(t, x) \in [t_1, 1] \times \mathbb{R}^n \mid \mathbf{V}^-(t, x) \leq \mathbf{V}^-(t_1, x_1)\}$$

the $\mathbf{V}^-(t_1, x_1)-$level set of $\mathbf{V}^-$ and let

$$\mathcal{W}^-(t) := \left\{x \in \mathbb{R}^n \mid (t, x) \in \mathcal{W}^-\right\},$$

which is closed and non empty for all t (see Chapter 4).

One can prove that for every $\epsilon > 0$, player 2 is able to construct a strategy β_ϵ such that for every $\mathbf{u} \in \mathcal{U}$

$$d(\mathbf{x}[t_1, x_1, \mathbf{u}, \beta_\epsilon(\mathbf{u})](1), \mathcal{W}^-(1)) \leq \epsilon / L_g.$$

This, together with the Lipschitz continuity of g allows us to conclude. To see this, let z denote a projection of the point $y := \mathbf{x}[t_1, x_1, \mathbf{u}, \beta_\epsilon(\mathbf{u})](1)$ in $\mathcal{W}^-(1)$. Then,

$$g(y) \leq g(z) + L\|y - z\| = \mathbf{V}^-(1, z) + \epsilon \ .$$

Observe that the inequality (3.23) implies (see [71, Lemma 2.1] for the proof) that, for all $(t, x) \in \mathcal{W}^-$ and all $t' \in [t, 1]$:

$$\sup_{u \in U} \inf_{\mathbf{v} \in \mathcal{V}} d(\mathbf{x}[t, x, u, \mathbf{v}](t'), \mathcal{W}^-(t')) = 0. \tag{3.24}$$

This property helps player 2 to build a strategy in the following way: let (t, x) denote some initial conditions for (3.1) and let w be a point in the projection of x in the set $\mathcal{W}^-(t)$. Consider the local game $\mathcal{G}(x, \xi)$, where $\xi := x - w$. Let us assume that this game has a value in pure strategies or, equivalently, that Isaacs' condition holds (Assumption 3.1.12). Now let (u^*, v^*) be optimal actions in $\mathcal{G}(x, \xi)$ and $(\mathbf{u}, \mathbf{v}) \in \mathcal{U} \times \mathcal{V}$ arbitrary. Consider the trajectories:

$$\mathbf{x}(t) = x, \quad \dot{\mathbf{x}}(s) = f(\mathbf{x}(s), \mathbf{u}(s), v^*) \tag{3.25a}$$

$$\mathbf{w}(t) = w, \quad \dot{\mathbf{w}}(s) = f(\mathbf{w}(s), u^*, \mathbf{v}(s)) \tag{3.25b}$$

This situation is pictured in Figure 3.4.1. The crucial property that these two trajectories satisfy is enclosed in the following Lemma, which is [71, Lemma 1.1] and is inspired by [61, Lemma 2.3.1].

Lemma 3.4.1. *There exist $A, B > 0$ such that for all $s \in [t, 1]$, and for all $\mathbf{u}, \mathbf{v}$:*

$$\|\mathbf{x}[t, x, \mathbf{u}, \mathbf{v}](s) - \mathbf{w}[t, x, \mathbf{u}, \mathbf{v}](s)\|^2 \leq (1 + (s - t)A)\|x - w\|^2 + B(s - t)^2.$$

Note that this estimate is independent of $\mathbf{u}$ and $\mathbf{v}$. In particular, it holds if one considers $\mathbf{v}_\epsilon$ that realizes the inf in (3.24) for u^*. Hence, this estimate provides an useful upper bound for the distance from $\mathbf{x}(s)$ to $\mathcal{W}^-(s)$ if player 2 plays the constant control v^* in the interval $[t, s]$. We will show, with the help of this estimate, how to construct an $\epsilon-$optimal strategy for player 2. Let $\Pi_n = \{t_1 < t_2 < \ldots t_{n+1} = 1\}$ denote a partition of $[t_1, 1]$ and let π_1 and π_2 be two selection rules defined by $\gamma_1 : [0, 1] \times \mathbb{R}^n \to \mathbb{R}^n$ which assigns to each (t, x) a closest point to x in $\mathcal{W}^-(t)$; and $\gamma_2 : [0, 1] \times \mathbb{R}^n \times \mathbb{R}^n \to V$ which assigns to each (t, x, ξ) an optimal action for player 2 in the local game $\mathcal{G}(t, x, \xi)$. Finally, let:

$$\gamma : [0, 1] \times \mathbb{R}^n \to V, \quad (t, x) \mapsto \gamma_2(t, x, x - \gamma_1(t, x)).$$

Definition 3.4.2. An **extremal strategy** $\beta = \beta(\Pi, \gamma) : \mathcal{U} \to \mathcal{V}$ is defined inductively as follows: suppose that β is already defined on $[t_1, t_m]$ for some $0 \leq m < n$, and let $x_m := \mathbf{x}[t_1, x_1, \mathbf{u}, \beta(\mathbf{u})](t_m)$. Then set $\beta(\mathbf{u}) \equiv \gamma(t_m, x_m)$ on $[t_m, t_{m+1}]$.

These strategies are inspired by the *extremal aiming* method of Krasovskii and Subbotin [61, Section 2.4]. They are obtained applying recursively the construction of Lemma 3.4.1. Using the estimates there, one obtains the following

Proposition 3.4.3. *For any $\mathbf{u} \in \mathcal{U}$, $d^2(\mathbf{x}[t_1, x_1, \mathbf{u}, \beta(\mathbf{u})](1), \mathcal{W}^-(1)) \leq e^A B\|\Pi\|$ where $A, B > 0$ are constants independent of Π.*

Hence, we obtain the desired inequality $\mathbf{V}^+ \leq \mathbf{V}^-$ and thus the existence of the value function $\mathbf{V} := \mathbf{V}^+ = \mathbf{V}^-$.

One can also obtain a characterization of the value function as a minimax solution. We will elaborate on this point in Section 3.5.3.

Local Game: $\mathcal{G}(x, \xi)$

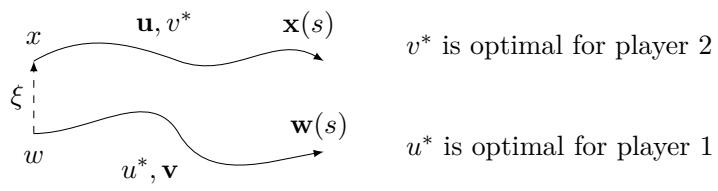

Figure 3.4.1 – Construction of two trajectories using the local game.

3.5 Some links between these approaches

Let us show in this Section how the three previously described approaches are related.

3.5.1 Convergence of the discretized values

Note that the sequence of Fleming and Friedman values satisfy very similar recursive equations. Thanks to a result on approximation schemes by Souganidis [94], it is possible to provide a unified proof of their convergence to the unique viscosity solution of the HJI PDE. Let us consider a function $H^* : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ that satisfies

Assumption 3.5.1. *(Regularity of H^*)*
a) $\forall x \in \mathbb{R}^n,\ |H^*(x,p) - H^*(x,q)| \leq C|p-q|$.
b) $\forall p \in \mathbb{R}^n,\ |H^*(x,p) - H^*(y,p)| \leq C|x-y|(1+|p|)$
for some positive constant C.

Under Assumption 3.5.1, the following result holds.

Theorem 3.5.2. *Under Assumption 3.5.1, the partial differential equation*

$$\frac{\partial u}{\partial t}(t,x) + H^*(x, \nabla_x u(t,x)) = 0 \tag{3.26}$$

with boundary condition $u(1,x) = g(x)$ has a unique viscosity solution.

For the proof, we refer to Crandall and Lions [33].

Now let $BUC(\mathbb{R}^n)$ denote the set of bounded and uniformly continuous functions on $\mathbb{R}^n$ and consider, for every $\rho > 0$ an operator $S_\rho : BUC(\mathbb{R}^n) \to BUC(\mathbb{R}^n)$ that satisfies:

i) $S_\rho(p) \geq S_\rho(q)$ if $p \geq q$.
ii) $S_\rho(p+k) = S_\rho(q) + k$, for $k \in \mathbb{R}$.
iii) $\lim_{\rho \to 0} \frac{\phi - S_\rho \phi}{\rho} = H^*(x, \nabla\phi)$, for all $\phi \in C^\infty$.

Consider the function $W_\Pi : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ defined recursively by

$$W_\Pi(1,x) = g(x),\ W_\Pi(t,x) = S_{t-t_{k+1}} W_\Pi(t_{k+1}, \cdot)(x),\ t \in [t_k, t_{k+1}].$$

Then the following theorem, due to Souganidis, holds (see [94] or Section 3 in [95] for a simplified version of the proof):

Theorem 3.5.3. *Let $H^* : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ that satisfies Assumption 3.5.1. Let V denote the unique viscosity solution of* (3.26).
Then $\|V - W_\Pi\|_\infty \to 0$ as $|\Pi| \to 0$.

Proof. Observe that the sequence W_Π has a subsequence that converges locally uniformly to a function W. Let $\phi : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ a C^1 function such that $W(\bar{t}, \bar{x}) = \phi(\bar{t}, \bar{x})$ and $W(s,y) \leq \phi(s,y)$ for $(s,y) \in [0,1] \times \mathbb{R}^n$. Now let W_{Π_n} be a convergent subsequence. Here, $\Pi_n = \{t_1 = t, t_2^n, \dots t_{n+1}^n = 1\}$. Let k such that $t_k^n \leq \bar{t} < t_{k+1}^n$ and $(\bar{t}_n, \bar{x}_n)$ such that $W_{\Pi_n} - \phi$ has a maximum at $(\bar{t}_n, \bar{x}_n)$, when restricted to $[t_k^n, t_{k+1}^n]$. It follows that, for all $x \in \mathbb{R}^n$,

$$W_{\Pi_n}(t_{k+1}^n, x) \leq \phi(t_{k+1}^n, x) + W_{\Pi_n}(\bar{t}_n, \bar{x}_n) - \phi(\bar{t}_n, \bar{x}_n).$$

Applying the operator $S_{\bar{t}_n - t_{k+1}^n}$ to both sides of this inequality and using the fact that this operator is monotone and homogeneously additive (properties $i)$ and $ii)$, we obtain:

$$W_{\Pi_n}(\bar{t}_n, \bar{x}_n) \leq S_{\bar{t}_n - t_{k+1}^n} \phi(t_k^n, x) + W_{\Pi_n}(\bar{t}_n, \bar{x}_n) - \phi(\bar{t}_n, \bar{x}_n).$$

Rearranging these terms and using property $iii)$ gives

$$0 \leq \frac{\phi(t_{k+1}^n, x) - S_{\bar{t_n} - t_{k+1}^n} \phi(t_{k+1}^n, x) + \phi(\bar{t_n}, \bar{x_n}) - \phi(t_{k+1}^n, x)}{\bar{t_n} - t_{k+1}^n}.$$

As $n \to +\infty$ we get:

$$0 \leq \frac{\partial \phi}{\partial t}(t,x) + H^*(x, \nabla_x \phi(t,x)).$$

This proves that the function W is a subsolution. The proof for supersolutions is completely analogous. From Theorem 3.5.2, the viscosity solution is unique, hence $V = W$. □

Observe that the value functions of the simultaneous Fleming value, the Fleming values and the Friedman values satisfy a similar recursive structure that the one described above. One can write the appropriate operator S_ρ which are on each case are given by the right hand sides of (3.7), (3.8a), (3.8b), (3.10a), and (3.10b). Explicitly, for $s \in [0,1]$ and $x \in \mathbb{R}^n$, if $h \in BUC(\mathbb{R}^n)$:

$$S_\rho h(s,x) = \begin{cases} \mathrm{val}_{(\mu,\nu) \in \Delta_f(U) \times \Delta_f(V)} h(s+\rho, x + \rho f(x,u,v)) & \text{in (3.7)}, \\ \max_{u \in U} \min_{v \in V} h(s+\rho, x + \rho f(x,u,v)) & \text{in (3.8a)}, \\ \min_{v \in V} \max_{u \in U} h(s+\rho, x + \rho f(x,u,v)) & \text{in (3.8b)}, \\ \max_{\mathbf{u} \in \mathcal{U}_{s,\rho}} \min_{\mathbf{v} \in \mathcal{V}_{s,\rho}} h(s+\rho, x + \int_s^{s+\rho} f(x(\omega), \mathbf{u}(\omega), \mathbf{v}(\omega)) d\omega) & \text{in (3.10a)}, \\ \min_{\mathbf{v} \in \mathcal{V}_{s,\rho}} \max_{\mathbf{u} \in \mathcal{U}_{s,\rho}} h(s+\rho, x + \int_s^{s+\rho} f(x(\omega), \mathbf{u}(\omega), \mathbf{v}(\omega)) d\omega), & \text{in (3.10b)}. \end{cases}$$

3.5.2 From the viscosity approach to the discrete game approach

As we already saw in Section 3.3.1, it is possible in the smooth case to use the value function to construct $\epsilon-$optimal strategies of the continuous time game. In this Section, we will show that one can use these limit strategies to prove the uniqueness of the accumulation points of the families of the maxmin and minmax of the discrete time games, in the smooth case. We have the following:

Proposition 3.5.4. *Let* $\mathbf{V}^- : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ *denote a viscosity supersolution of* (3.16b) *and let* $\Pi_0 = \{0 = s_1 < s_2, \dots, s_N = 1\}$ *be a partition such that* $\mathbf{V}^-$ *is continuously differentiable in every interval* (s_k, s_{k+1}). *The family of discrete maxmin, as defined in* (3.9b), *converge to* $\mathbf{V}^-$ *as* $|\Pi| \to 0$, *for all* Π *refinement of* Π_0.

Proof. Denote with $(x_k^*)_k$ the sequence of states in the game (as will be defined below), with $x_1^* = x$.

Let α^* be the following strategy: at time t_k, observe x_k^* and choose u_k^* such that

$$u_k^* \in \text{argmax}_{u\in U}\left\{\min_{v\in V}\langle f(x_k^*,u,v),\nabla\phi_k(t_k,x_k^*)\rangle\right\}$$

Let $\vec{v} = (v_1, v_2, \dots, v_n)$ be an arbitrary sequence of actions of player 2 and let $x_1^* = x, x_2^*, \dots x_{n+1}^*$ denote the trajectory induced by $\vec{v}$ and α^*. We have that, if we write $\Delta\mathbf{V}_k := \frac{\mathbf{V}^-(t_{k+1},x_{k+1}^*)-\mathbf{V}^-(t_k,x_k^*)}{\pi_{k+1}}$,

$$\begin{aligned}\Delta\mathbf{V}_k &= \frac{\partial\mathbf{V}^-}{\partial t}(t_k,x_k^*) + \langle f(x_k^*,u_k^*,v_k),\nabla_x\mathbf{V}^-(t_k,x_k^*)\rangle + o(\pi_{k+1})\\ &\geq \frac{\partial\mathbf{V}^-}{\partial t}(t_k,x_k^*) + \min_{v\in V}\langle f(x_k,u_k^*,v),\nabla_x\mathbf{V}^-(t_k,x_k^*)\rangle + o(\pi_{k+1})\end{aligned}$$

Note that

$$\frac{\partial\mathbf{V}^-}{\partial t}(t_k,x_k^*) + \min_{v\in V}\langle f(x_k,u_k^*,v),\nabla_x\mathbf{V}^-(t_k,x_k^*)\rangle$$

is equal to

$$\frac{\partial\mathbf{V}^-}{\partial t}(t_k,x_k^*) + \max_{u\in U}\min_{v\in V}\langle f(x_k^*,u,v),\nabla_x\mathbf{V}^-(t_k,x_k^*)\rangle = 0.$$

Integrating from t_1 to 1 one obtains:

$$g(x_{n+1}^*) \geq \mathbf{V}^-(t_1,x_1^*) + o(|\Pi|)$$

which concludes the proof. □

When the state space is finite, the idea of Proposition 3.5.4, namely to use the limit equation to generate an $\epsilon-$optimal strategy and then prove the convergence of the discrete value functions has been used in the context of stochastic games of short stage duration by Neyman. We refer to [78, Theorem 1].

The same idea, i.e. to use a limit object to construct approximate strategies, is central to this thesis and has proven fruitful in game theory. One important example is the Big Match with incomplete information on one side, due to Sorin [90, 91], where a limit continuous time game is deduced and discretizations of optimal strategies of the limit game are used to construct approximate strategies of discrete time games.

3.5.3 Equivalence of minimax and viscosity solutions

Before introducing the definition of minimax solutions, due to Subbotin [96], we recall some notions of viability theory. Let $F : \mathbb{R}^N \rightsquigarrow \mathbb{R}^N$ be a multiple valued map. A **trajectory** of the differential inclusion

$$\dot{y}(t) \in F(y(t)),\ y(t_1) = y_1, \tag{3.27}$$

is an absolutely continuous function $y : [t_1, 1] \to \mathbb{R}^N$ that satisfies (3.27). A pair $(\mathcal{S}, F)$ of a closed set $\mathcal{S} \subset \mathbb{R}^N$ and a multiple valued map $F : \mathbb{R}^N \rightsquigarrow \mathbb{R}^N$ is called **viable** or **weakly invariant** if for every initial condition $(t_1, y_1) \in [0, 1] \times \mathcal{S}$ there exists a trajectory of (3.27) such that $y(t) \in S$ for all $t_1 \leq t \leq 1$.

For every $u \in U$, we will now consider the following differential inclusion

$$\begin{cases} \dot{s} & = 1 \\ \dot{x} & \in f(x, u, V) \\ \dot{z} & = 0 \end{cases} \tag{3.28}$$

In a similar way, we also consider, for every $v \in V$, the differential inclusion

$$\begin{cases} \dot{s} & = 1 \\ \dot{x} & \in f(x, U, v) \\ \dot{z} & = 0 \end{cases} \tag{3.29}$$

Definition 3.5.5. (Minimax solutions)

— A lower semicontinuous function $w : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ is a **minimax supersolution** of (3.16a) if for any $u \in U$, $(\text{epi}w, E_u^-)$ is viable.

— An upper semicontinuous function $w : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ is a **minimax subsolution** of (3.16b) if for any $v \in V$, $(\text{hyp}w, E_v^+)$ is viable.

— A continuous function is a **minimax solution** if it is both a minimax super and subsolution.

The relationship with games is the following: using the subdynamic programming principle (3.23), one can prove the viability of the epigraph of $\mathbf{V}^-$, which is a slightly stronger condition that (3.24). Then, via the strategic approach, which relies on this viability property, one obtains the existence of the value.

The equivalence between minimax and viscosity solutions was first established by Subbotin. We refer to Subbotin [96] for a detailed account of this theory. The ideas behind these two solution concepts are different: while Subbotin's approach relies more on the geometrical ideas of weak invariance (his terminology for viability), while Crandall and Lions' viscosity solutions rely on analytical techniques (i.e. the vanishing viscosity method, see Crandall and Lions [33]). In control theory, the relationship between viability theory and viscosity solutions has been established by Frankowska [40].

A different approach to prove this equivalence has been explored by Lions and Souganidis [68, 69]. As we have seen, the dynamic programming principle, together with the regularity assumptions on the data of the problem allow to characterize the value functions as viscosity sub and super solutions. In [68], Lions and Souganidis go in the opposite direction: they show that sub and super solutions satisfy suitable dynamic programming inequalities, which are in turn used to prove the equivalence between Subbotin's definition of minimax solutions using directional derivatives.

A different approach based on proximal calculus and viability is described in Chapter 5 of this thesis.

Chapter 4

A strategy-based proof of the existence of the value in zero-sum differential games

Abstract: The value of a zero-sum differential games is known to exist, under Isaacs' condition, and it is the unique viscosity solution of a Hamilton-Jacobi-Isaacs equation. This approach, in spite of being very effective, does not provide information about the strategies the players should use. In this note we provide a self-contained proof of the existence of the value based on the construction of ϵ-optimal strategies, which is inspired by the extremal aiming method from [61]. This Chapter is based on joint work with Miquel Oliu-Barton and has been accepted for publication in *Morfismos*.

4.1 Comparison of trajectories

Let U and V be compact subsets of some euclidean space, let $\|\cdot\|$ be the euclidean norm in $\mathbb{R}^n$, and let $f : [0,1] \times \mathbb{R}^n \times U \times V \to \mathbb{R}^n$. For each $x \in \mathbb{R}^n$ and $\mathcal{Z} \subset \mathbb{R}^n$, let

$$D(x, \mathcal{Z}) := \inf_{z \in \mathcal{Z}} \|x - z\|$$

be the usual distance from x to the set $\mathcal{Z}$.

Assumption 4.1.1. *f is uniformly bounded, continuous and there exists $c \geq 0$ such that for all $(u, v) \in U \times V$, $(s, t) \in [0,1]^2$ and $x, y \in \mathbb{R}^n$:*

$$\|f(t, x, u, v) - f(s, y, u, v)\| \leq c(|t - s| + \|x - y\|).$$

Let $\|f\| := \sup_{(t,x,u,v)} \|f(t, x, u, v)\| < +\infty$.

The local game. For each $(t, x), \xi \in [0,1] \times \mathbb{R}^n \times \mathbb{R}^n$, the **local game** $\Gamma(t, x, \xi)$ is a one-shot game with action sets U and V and payoff function:

$$(u, v) \mapsto \langle \xi, f(t, x, u, v) \rangle.$$

Let $H^-(t, x, \xi)$ and $H^+(t, x, \xi)$ be its maxmin and minmax respectively:

$$\begin{aligned} H^-(t, x, \xi) &:= \max_{u \in U} \min_{v \in V} \langle \xi, f(t, x, u, v) \rangle, \\ H^+(t, x, \xi) &:= \min_{v \in V} \max_{u \in U} \langle \xi, f(t, x, u, v) \rangle. \end{aligned}$$

Local Game: $\Gamma(t_1, x_1, \xi_1)$

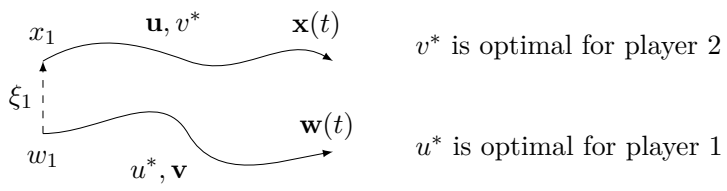

Figure 4.1.1 – Construction of two trajectories using the local game.

These functions satisfy $H^- \leq H^+$. If the equality $H^+(t, x, \xi) = H^-(t, x, \xi)$ holds, the game $\Gamma(t, x, \xi)$ has a value, denoted by $H(t, x, \xi)$.

Assumption 4.1.2. *$\Gamma(t, x, \xi)$ has a value for all $(t, x, \xi) \in [0, 1] \times \mathbb{R}^n \times \mathbb{R}^n$.*

Assumptions 4.1.1 and 4.1.2 hold in the rest of the paper.

4.1.1 A key Lemma

Introduce the sets of controls:

$$\mathcal{U} = \{\mathbf{u} : [0, 1] \to U, \text{ measurable}\}, \quad \mathcal{V} = \{\mathbf{v} : [0, 1] \to V, \text{ measurable}\}.$$

Consider the following dynamical system, where $t_1 \in [0, 1]$, $z_1 \in \mathbb{R}^n$ and $(\mathbf{u}, \mathbf{v}) \in \mathcal{U} \times \mathcal{V}$:

$$\mathbf{z}(t_1) = z_1, \quad \dot{\mathbf{z}}(t) = f(t, \mathbf{z}(t), \mathbf{u}(t), \mathbf{v}(t)) \tag{4.1}$$

The Assumption 4.1.1 ensures the existence of a unique solution to (4.1), which is denoted by $\mathbf{z}[t_1, z_1, \mathbf{u}, \mathbf{v}]$, in the following extended sense: for any $t \in [t_0, 1]$,

$$\mathbf{z}[t_1, z_1, \mathbf{u}, \mathbf{v}](t) := z_1 + \int_{t_1}^{t} f(s, \mathbf{z}[t_1, z_1, \mathbf{u}, \mathbf{v}](s), \mathbf{u}(s), \mathbf{v}(s))ds.$$

This result is due to Carathéodory and can be found in [31, Chapter 2]. Elements of U and V are identified with constant controls.

The purpose of this section is to bound the distance between two trajectories: one starting from x_1 and controlled by $(\mathbf{u}, v)$, and another one starting from w_1 and controlled by $(u, \mathbf{v})$. The appropriate pair (u, v) is obtained using the existence of the value and of optimal actions in the local game: let u^* (resp. v^*) be optimal for player 1 (resp. 2) in $\Gamma(t_1, x_1, \xi_1)$, where $\xi_1 := x_1 - w_1$.

Let $\mathbf{x} := \mathbf{x}[t_1, x_1, \mathbf{u}, v^*]$ and $\mathbf{w} := \mathbf{w}[t_1, w_1, u^*, \mathbf{v}]$ (see Figure 4.1.1). The following lemma is inspired by [61, Lemma 2.3.1].

Lemma 4.1.3. *There exist $A, B \in \mathbb{R}_+$ such that for all $t \in [t_1, 1]$:*

$$\|\mathbf{x}(t) - \mathbf{w}(t)\|^2 \leq (1 + (t - t_1)A)\|x_1 - w_1\|^2 + B(t - t_1)^2.$$

Furthermore, A and B are independent of the controls.

Proof. Let $d_1 := \|x_1 - w_1\|$ and $\mathbf{d}(t) := \|\mathbf{x}(t) - \mathbf{w}(t)\|$. Then:

$$\mathbf{d}^2(t) = \left\|\xi_1 + \int_{t_1}^t [f(s, \mathbf{x}(s), \mathbf{u}(s), v^*) - f(s, \mathbf{w}(s), u^*, \mathbf{v}(s))]ds\right\|^2. \tag{4.2}$$

The boundedness of f implies that:

$$\left\|\int_{t_1}^t [f(s, \mathbf{x}(s), \mathbf{u}(s), v^*) - f(s, \mathbf{w}(s), u^*, \mathbf{v}(s))]ds\right\|^2 \leq 4\|f\|^2(t - t_1)^2. \tag{4.3}$$

Claim: For all $s \in [t_1, 1]$, and for all $(u, v) \in U \times V$:

$$\langle \xi_1, f(s, \mathbf{x}(s), u, v^*) - f(s, \mathbf{w}(s), u^*, v)\rangle \leq 2C(s)d_1 + cd_1^2, \tag{4.4}$$

where $C(s) := c(1 + \|f\|)(s - t_1)$.
Proof of the claim. Assumption 4.1.1 implies $\|\mathbf{x}(s) - x_1\| \leq (s - t_1)\|f\|$, and then:

$$\|f(s, \mathbf{x}(s), u, v^*) - f(t_1, x_1, u, v^*)\| \leq c\big((s - t_1) + \|f\|(s - t_1)\big) = C(s).$$

From the Cauchy-Schwartz inequality and the optimality of v^* one gets:

$$\begin{aligned} \langle \xi_1, f(s, \mathbf{x}(s), u, v^*)\rangle &\leq \langle \xi_1, f(t_1, x_1, u, v^*)\rangle + C(s)d_1, & (4.5)\\ &\leq H^+(t_1, x_1, \xi_0) + C(s)d_1. & (4.6)\end{aligned}$$

Similarly, Assumption 4.1.1 implies $\|\mathbf{w}(s) - x_1\| \leq d_1 + (s - t_1)\|f\|$, and then:

$$\|f(s, \mathbf{w}(s), u^*, v) - f(t_1, x_1, u^*, v)\| \leq C(s) + cd_1.$$

Using the Cauchy-Schwartz inequality, and the optimality of u^*:

$$\begin{aligned} \langle \xi_1, f(s, \mathbf{w}(s), u^*, v)\rangle &\geq \langle \xi_1, f(t_1, x_1, u^*, v)\rangle - (C(s) + cd_1)d_1, & (4.7)\\ &\geq H^-(t_1, x_1, \xi_1) - C(s)d_1 - cd_1^2. & (4.8)\end{aligned}$$

The claim follows by substracting the inequalities (4.6) and (4.8) and using Assumption 4.1.2 to cancel $(H^+ - H^-)(t_1, x_1, \xi_1)$.
In particular, (4.4) holds for $(u, v) = (\mathbf{u}(s), \mathbf{v}(s))$. Note that $\int_{t_1}^t 2C(s)ds \leq (t - t_1)C(t)$. Thus, integrating (4.4) over $[t_1, t]$ yields:

$$\int_{t_1}^t \langle \xi_1, f(s, \mathbf{x}(s), \mathbf{u}(s), v^*) - f(s, \mathbf{w}(s), u^*, \mathbf{v}(s))\rangle ds \leq (t - t_1)(2C(t)d_1 + cd_1^2). \tag{4.9}$$

Using the estimates (4.3) and (4.9) in (4.2) we obtain:

$$\mathbf{d}^2(t) \leq d_1^2 + 4\|f\|^2(t - t_1)^2 + 2(t - t_1)C(t)d_1 + c(t - t_1)d_1^2.$$

Finally, using the relations $d_1 \leq 1 + d_1^2$ and $(t - t_1)C(t) = c(1 + \|f\|)(t - t_1)^2$, the result follows with $A := 3c + 2\|f\|$ and $B := 4\|f\|^2 + 2c(1 + \|f\|)$. □

4.1.2 Consequences

We give here three direct consequences of Lemma 4.1.3. First, we use a set of times $\Pi = \{0 = t_1 < \cdots < t_N = 1\}$ in $[0, 1]$ to construct two trajectories on $[t_1, t_N]$ inductively. Applying Lemma 4.1.3 to the intervals $[t_m, t_{m+1}]$ for $m = 1 \ldots, N - 1$, we obtain a bound for the distance between the two at time t_N. In particular, if the two trajectories start from the same state then their distance at time t_N vanishes as $\|\Pi\| := \max_{1\leq m\leq N} t_m - t_{m-1}$ tends to 0. Later, we replace the distance between two trajectories by the distance between a trajectory and a set. Finally, we combine the two aspects; the result obtained therein is used in Section 4.2 to prove the existence of the value of zero-sum differential games with terminal payoff.

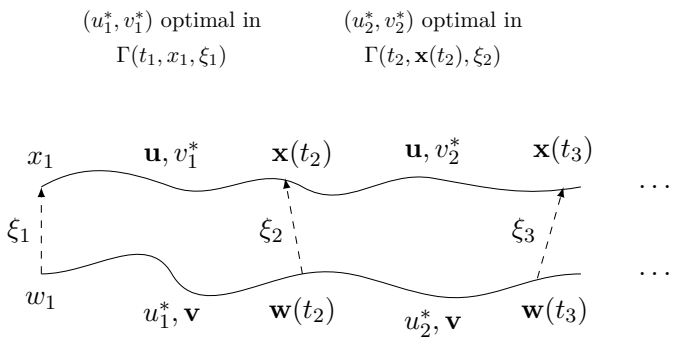

Figure 4.1.2 – Iterative construction of the two trajectories.

Induction

Let $(\mathbf{u}, \mathbf{v}) \in \mathcal{U} \times \mathcal{V}$ be a pair of controls. Define the trajectories $\mathbf{x}$ and $\mathbf{w}$ on $[t_1, t_N]$ inductively: let $\mathbf{x}(t_1) = x_1$ and $\mathbf{w}(t_1) = w_1$ and suppose that $\mathbf{x}(t)$ and $\mathbf{w}(t)$ are defined on $[t_1, t_m]$ for some $m = 1, \dots, N-1$. Consider the local game $\Gamma(t_m, \mathbf{x}(t_m), \xi_m)$, where $\xi_m := \mathbf{x}(t_m) - \mathbf{w}(t_m)$, and let $u_m^* \in U$ and $v_m^* \in V$ be optimal actions for player 1 and 2 respectively.

For $t \in [t_m, t_{m+1}]$, set $\mathbf{x}(t) := \mathbf{x}[t_m, \mathbf{x}(t_m), \mathbf{u}, v_m^*](t)$ and $\mathbf{w}(t) := \mathbf{w}[t_m, \mathbf{w}(t_m), u_m^*, \mathbf{v}](t)$ (see Figure 4.1.2).

Corollary 4.1.4. $\|\mathbf{x}(t_N) - \mathbf{w}(t_N)\|^2 \leq e^A(\|x_1 - w_1\|^2 + B\|\Pi\|)$.

Proof. For any $1 \leq m \leq N$, put $d_m := \|\mathbf{x}(t_m) - \mathbf{w}(t_m)\|$. By Lemma 4.1.3, one has:

$$d_m^2 \leq (1 + (t_m - t_{m-1})A)d_{m-1}^2 + B(t_m - t_{m-1})^2.$$

By induction, one obtains:

$$d_N^2 \leq \exp\left(A \sum_{m=1}^{N} (t_m - t_{m-1})\right)\left(d_1^2 + B \sum_{m=1}^{N} (t_m - t_{m-1})^2\right).$$

The result follows, since $\sum_{m=1}^{N}(t_m - t_{m-1}) \leq 1$ and $\sum_{m=1}^{N}(t_m - t_{m-1})^2 \leq \|\Pi\|$. □

Distance to a set

Let $\mathcal{W} \subset [t_1, 1] \times \mathbb{R}^n$ be a set satisfying the following properties:

- **P1:** For any $t \in [t_1, 1]$, $\mathcal{W}(t) := \{x \in \mathbb{R}^n \mid (t, x) \in \mathcal{W}\}$ is closed and nonempty.
- **P2:** For any $(t, x) \in \mathcal{W}$ and any $t' \in [t, 1]$:

$$\sup_{u \in U} \inf_{\mathbf{v} \in \mathcal{V}} D(\mathbf{x}[t, x, u, \mathbf{v}](t'), \mathcal{W}(t')) = 0.$$

Equivalent formulations of **P2** were introduced by Aubin [3], although our formulation is inspired by the notion of stable bridge in [61].

Let $x_1 \in \mathbb{R}^n$, let $w_1 \in \operatorname{argmin}_{\mathcal{W}(t_1)} \|x_1 - w_1\|$ be some closest point to x_1 in $\mathcal{W}(t_1)$ and let v^* be optimal for player 2 in the local game $\Gamma(t_1, x_1, x_1 - w_1)$.

Corollary 4.1.5. *For every $t \in [t_1, 1]$ and $\mathbf{u} \in \mathcal{U}$:*

$$D^2(\mathbf{x}[t_1, x_1, \mathbf{u}, v^*](t), \mathcal{W}(t)) \leq (1 + (t - t_1)A)D^2(x_1, \mathcal{W}(t_1)) + B(t - t_1)^2.$$

Proof. Let $\mathbf{u} \in \mathcal{U}$ be fixed and let u^* be optimal in $\Gamma(t_1, x_1, x_1 - w_1)$. By **P2**, for all $\epsilon > 0$ there exists $\mathbf{v}_{(\epsilon,u^*)} \in \mathcal{V}$ such that the point $\mathbf{w}_\epsilon(t) := \mathbf{x}[t_1, w_1, u^*, \mathbf{v}_{(\epsilon,u^*)}](t)$ satisfies $D(\mathbf{w}_\epsilon(t), \mathcal{W}(t)) \leq \epsilon$ (see Figure 4.1.3). We use the following abbreviation: $\mathbf{x_u}(t) := \mathbf{x}[t_1, x_1, \mathbf{u}, v^*](t)$. The triangular inequality gives $D(\mathbf{x_u}(t), \mathcal{W}(t)) \leq \|\mathbf{x_u}(t) - \mathbf{w}_\epsilon(t)\| + \epsilon$. Taking the limit, as $\epsilon \to 0$, one has that:

$$D^2(\mathbf{x_u}(t), \mathcal{W}(t)) \leq \lim_{\epsilon \to 0} \|\mathbf{x_u}(t) - \mathbf{w}_\epsilon(t)\|^2.$$

By Lemma 4.1.3, $\|\mathbf{x_u}(t) - \mathbf{w}_\epsilon(t)\|^2 \leq (1 + (t - t_1)A)\|x_1 - w_1\|^2 + B(t - t_1)^2$ for all $\epsilon > 0$. The result follows by the choice of w_1. □

A key Corollary

For any $\mathbf{u} \in \mathcal{U}$, define a trajectory $\mathbf{x_u}$ on $[t_1, t_N]$ inductively, as follows: let $\mathbf{x_u}(t_1) = x_1$ and suppose that $\mathbf{x_u}$ is defined on $[t_1, t_m]$ for some $m = 1, \ldots, N-1$. Let $w_m \in \operatorname{argmin}_{w \in \mathcal{W}(t_m)} \|\mathbf{x_u}(t_m) - w\|$ be a closest point to $\mathbf{x_u}(t_m)$ in $\mathcal{W}(t_m)$, and let v_m^* be optimal for player 2 in the local game $\Gamma(t_m, \mathbf{x_u}(t_m), \mathbf{x_u}(t_m) - w_m)$.

Implicitly, we are using two selection rules π_1 and π_2 defined as follows: $\pi_1 : [0,1] \times \mathbb{R}^n \to \mathbb{R}^n$ assigns to each (t, x) a closest point to x in $\mathcal{W}(t)$; $\pi_2 : [0,1] \times \mathbb{R}^n \times \mathbb{R}^n \to V$ assigns to each (t, x, ξ) an optimal action for player 2 in the local game $\Gamma(t, x, \xi)$. Thus,

$$v_m^* = \pi_2(t_m, \mathbf{x_u}(t_m), \mathbf{x_u}(t_m) - \pi_1(\mathbf{x_u}(t_m))).$$

For $t \in [t_m, t_{m+1}]$, put $\mathbf{x_u}(t) := \mathbf{x}[t_m, \mathbf{x_u}(t_m), \mathbf{u}, v_m^*](t)$. Define a control $\beta(\mathbf{u}) \in \mathcal{V}$ inductively by setting $\beta(\mathbf{u}) \equiv v_m^*$ on $[t_m, t_{m+1}]$ for all $1 \leq m < N$, so that $\mathbf{x_u}(t) = \mathbf{x}[t_1, x_1, \mathbf{u}, \beta(\mathbf{u})](t)$, for all $t \in [t_1, t_N]$.

Note that the action v_m^* used in the interval $[t_m, t_{m+1}]$ depends only on the current position $\mathbf{x_u}(t_m)$ and on the set $\mathcal{W}(t_m)$. Moreover, the current position depends only on $v_1^*, \ldots, v_{m-1}^*$ and on the restriction of $\mathbf{u}$ to the interval $[t_1, t_m]$. In particular, the control $\beta(\mathbf{u})$ is piecewise constant and depends on the set of times Π. Finally, note that for $\mathbf{u}_1, \mathbf{u}_2 \in \mathcal{U}$ such that $\mathbf{u}_1 \equiv \mathbf{u}_2$ on $[t_1, t_m]$ for some $0 \leq m < N$, the construction described above gives $\beta(\mathbf{u}_1) \equiv \beta(\mathbf{u}_2)$ on $[t_1, t_{m+1}]$. In this sense, $\beta : \mathcal{U} \to \mathcal{V}$ is non anticipating with delay with respect to the set of times Π.

Putting Corollaries 4.1.4 and 4.1.5 together and choosing $x_1 \in \mathcal{W}(t_1)$ yields a useful bound.

Corollary 4.1.6. *For any* $\mathbf{u} \in \mathcal{U}$, $D^2(\mathbf{x}[t_1, x_1, \mathbf{u}, \beta(\mathbf{u})](t_N), \mathcal{W}(t_N)) \leq e^A B \|\Pi\|$.

This result can be interpreted as follows: under **P1-P2** for any control $\mathbf{u} \in \mathcal{U}$ there exists a "reply" $\beta(\mathbf{u}) \in \mathcal{V}$ (which is non anticipating with delay, and piecewise constant along Π) which keeps a trajectory starting from $\mathcal{W}(t_1)$ at time t_1 arbitrarily close to $\mathcal{W}(t_N)$ at time t_N.

4.2 Differential Games

Consider now the zero-sum differential game $\mathcal{G}(t_1, x_1)$ played in $[t_1, 1]$ and with the following dynamics in $\mathbb{R}^n$:

$$\mathbf{x}(t_1) = x_1, \quad \dot{\mathbf{x}}(t) = f(t, \mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t)) \quad \text{(a.e. on } [t_1, 1]).$$

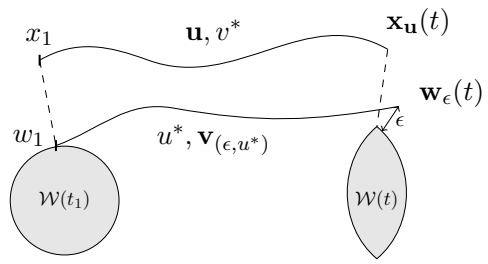

Figure 4.1.3 – Distance to a set $\mathcal{W} \subset [t_0, 1] \times \mathbb{R}^n$ satisfying **P1** and **P2**.

Definition 4.2.1. A strategy for player 2 is a map $\beta : \mathcal{U} \to \mathcal{V}$ such that, for some finite partition $s_1 < \cdots < s_N$ of $[t_1, 1]$, for all $\mathbf{u}_1, \mathbf{u}_2 \in \mathcal{U}$ and $1 \leq m < N$:

$$\mathbf{u}_1 \equiv \mathbf{u}_2 \text{ a.e. on } [s_1, s_m] \implies \beta(\mathbf{u}_1) \equiv \beta(\mathbf{u}_2) \text{ a.e. on } [s_1, s_{m+1}].$$

These strategies are called nonanticipative strategies with delay (NAD) [25, Section 2.2] in contrast to the classical nonanticipative strategies. The strategies for player 1 are defined in a dual manner. Let $\mathcal{A}$ (resp. $\mathcal{B}$) the set of strategies for player 1 (resp. 2). For any pair of strategies $(\alpha, \beta) \in \mathcal{A} \times \mathcal{B}$, there exists a unique pair $(\bar{\mathbf{u}}, \bar{\mathbf{v}}) \in \mathcal{U} \times \mathcal{V}$ such that $\alpha(\bar{\mathbf{v}}) = \bar{\mathbf{u}}$ and $\beta(\bar{\mathbf{u}}) = \bar{\mathbf{v}}$ [25, Lemma 1]. This fact is crucial for it allows to define $\mathbf{x}[t_1, x_1, \alpha, \beta] := \mathbf{x}[t_1, x_1, \bar{\mathbf{u}}, \bar{\mathbf{v}}]$ in a unique manner.

The payoff function has two components: a running payoff $\gamma : [0, 1] \times \mathbb{R}^n \times U \times V \to \mathbb{R}$ and a terminal payoff $g : \mathbb{R}^n \to \mathbb{R}$. We assume that the running payoff γ satisfies the same regularity assumptions as the dynamics f. In this case, we apply the classical transformation of a Bolza problem into a Mayer problem, to get rid of the running payoff: enlarge the state space from $\mathbb{R}^n$ to $\mathbb{R}^{n+1}$, where the last coordinate represents the accumulated payoff; define an auxiliary terminal payoff function $\widetilde{g} : \mathbb{R}^{n+1} \to \mathbb{R}$ as $\widetilde{g}(x, y) = g(x) + y$; we thus obtain an equivalent differential game with no running payoff and dynamic $\widetilde{f} = (f, \gamma)$. Consequently, we can assume without loss of generality that $\gamma \equiv 0$.

Assumption 4.2.2. *g is Lipschitz continuous.*

Assumption 4.2.2 holds in the rest of the paper. Introduce the lower and upper value functions:

$$\begin{aligned} \mathbf{V}^-(t_1, x_1) &:= \sup_{\alpha \in \mathcal{A}} \inf_{\beta \in \mathbb{B}} g\big(\mathbf{x}[t_1, x_1, \alpha, \beta](1)\big), \\ \mathbf{V}^+(t_1, x_1) &:= \inf_{\beta \in \mathcal{B}} \sup_{\alpha \in \mathcal{A}} g\big(\mathbf{x}[t_1, x_1, \alpha, \beta](1)\big). \end{aligned}$$

The inequality $\mathbf{V}^- \leq \mathbf{V}^+$ holds everywhere. If $\mathbf{V}^-(t_1, x_1) = \mathbf{V}^+(t_1, x_1)$, the game $\mathcal{G}(t_1, x_1)$ has a value, denoted by $\mathbf{V}(t_1, x_1)$. Under Assumption 4.1.2, usually known as Isaacs' condition, the value exists as the unique viscosity solution of some Hamilton-Jacobi-Isaacs equation with a boundary condition [36]. The functional approach is very effective for it yields the existence and a characterization of the value function. However, it does not tell us much about the strategies the players should use. In this note we focus on the strategies, as in [61], and prove the existence of the value using an explicit construction of ϵ-optimal strategies. Let us end this section by stating the dynamic programming principle [25, Proposition 2] satisfied by $\mathbf{V}^-$: for all $(t, x) \in [0, 1] \times \mathbb{R}^n$ and all $t' \in [t, 1]$,

$$\mathbf{V}^-(t, x) = \sup_{\alpha \in \mathcal{A}} \inf_{\beta \in \mathcal{B}} \mathbf{V}^-\big(t', \mathbf{x}[t, x, \alpha, \beta](t')\big). \tag{4.10}$$

The dynamic programming principle consists in two inequalities: the $\geq$ (resp. $\leq$) inequality is the superoptimality (resp. suboptimality) programming principle.

4.2.1 Existence of the value

Let $\phi : [t_1, 1] \times \mathbb{R}^n \to \mathbb{R}$ be a real function satisfying the following properties:

(i) ϕ is lower semi continuous.

(ii) For all $(t, x) \in [t_1, 1] \times \mathbb{R}^n$ and $t' \in [t, 1]$:

$$\phi(t, x) \geq \sup_{u \in U} \inf_{\mathbf{v} \in \mathcal{V}} \phi(t', \mathbf{x}[t, x, u, \mathbf{v}](t'));$$

(iii) $\phi(1, x) \geq g(x)$, for all $x \in \mathbb{R}^n$.

Definition 4.2.3. For any $\ell \in \mathbb{R}$, define the **ℓ-level set of** ϕ by:

$$\mathcal{W}_\ell^\phi = \{(t, x) \in [t_1, 1] \times \mathbb{R}^n \mid \phi(t, x) \leq \ell\}$$

and

$$\mathcal{W}_\ell^\phi(t) = \{x \in \mathbb{R}^n \mid \phi(t, x) \leq \ell\}$$

.

Lemma 4.2.4. *For any $\ell \geq \phi(t_1, x_1)$, the ℓ-level set of ϕ satisfies* **P1** *and* **P2**.

Proof. $x_1 \in \mathcal{W}_\ell^\phi(t_1)$ so that $\mathcal{W}_\ell^\phi(t_1)$ is non empty. By (i), $\mathcal{W}_\ell^\phi(t)$ is a closed set for all $t \in [0, 1]$. The property (ii) implies that for any $t \in [t_1, 1]$, $u \in U$ and $n \in \mathbb{N}^*$ there exists $\mathbf{v}_n \in \mathcal{V}$ such that:

$$\ell \geq \phi(t_1, x_1) \geq \phi(t, \mathbf{x}[t_1, x_1, u, \mathbf{v}_n](t)) - \frac{1}{n}. \tag{4.11}$$

The boundedness of f implies that $x_n := \mathbf{x}[t_1, x_1, u, \mathbf{v}_n](t)$ belongs to some compact set. Consider a subsequence $(x_n)_n$ such that $\lim \phi(t, x_n) = \liminf_{n\to\infty} \phi(t, x_n)$, and such that $(x_n)_n$ converges to $\bar{x} \in \mathbb{R}^n$. Take the limit, as $n \to \infty$, in (4.11). Then by (i) one has:

$$\ell \geq \phi(t_1, x_1) \geq \phi(t, \bar{x}).$$

Consequently, $\bar{x} \in \mathcal{W}_\ell^\phi(t) \neq \emptyset$ and $\inf_{n\in\mathbb{N}^*} \mathbf{d}(\mathbf{x}[t_1, x_1, u, \mathbf{v}_n](t), \mathcal{W}_\ell^\phi(t)) = 0$. The proof of these two properties still holds by replacing (t_1, x_1) and $t \in [t_1, 1]$ by any $(t, x) \in \mathcal{W}_\ell^\phi$ and $t' \in [t, 1]$, so that $\mathcal{W}_\ell^\phi$ satisfies **P1** and **P2**. □

Extremal strategies in $\mathcal{G}(t_1, x_1)$

Let $\mathcal{W}^\phi \subset [t_0, 1] \times \mathbb{R}^n$ be the $\phi(t_1, x_1)$-level set of ϕ, i.e.:

$$\mathcal{W}^\phi := \{(t, x) \in [t_0, 1] \times \mathbb{R}^n \mid \phi(t, x) \leq \phi(t_1, x_1)\}.$$

As in Section 4.1.2, let π_1 and π_2 be two selection rules defined as follows: $\pi_1 : [0, 1]\times\mathbb{R}^n \to \mathbb{R}^n$ assigns to each (t, x) a closest point to x in $\mathcal{W}^\phi(t)$; $\pi_2 : [0, 1]\times\mathbb{R}^n\times\mathbb{R}^n \to V$ assigns to each (t, x, ξ) an optimal action for player 2 in the local game $\Gamma(t, x, \xi)$. Finally, let:

$$\pi : [0, 1] \times \mathbb{R}^n \to V, \quad (t, x) \mapsto \pi_2(t, x, x - \pi_1(t, x)).$$

Definition 4.2.5. An **extremal strategy** $\beta = \beta(\phi, \Pi, \pi) : \mathcal{U} \to \mathcal{V}$ is defined inductively as follows: suppose that β is already defined on $[t_1, t_m]$ for some $1 \leq m < N$, and let $x_m := \mathbf{x}[t_1, x_1, \mathbf{u}, \beta(\mathbf{u})](t_m)$. Then set $\beta(\mathbf{u}) \equiv \pi(t_m, x_m)$ on $[t_m, t_{m+1}]$.

These strategies are inspired by the *extremal aiming* method of Krasovskii and Subbotin [61, Section 2.4].

Proposition 4.2.6. *For some $C \in \mathbb{R}_+$, and for any extremal strategy $\beta = \beta(\phi, \Pi, \pi)$:*

$$g(\mathbf{x}[t_1, x_1, \mathbf{u}, \beta(\mathbf{u})](1)) \leq \phi(t_1, x_1) + C\sqrt{\|\Pi\|}, \quad \forall \mathbf{u} \in \mathcal{U}.$$

Proof. Without loss of generality, $t_N = 1$ so that $x_N = \mathbf{x}[t_1, x_1, \mathbf{u}, \beta(\mathbf{u})](1)$. By Lemma 4.2.4, $\mathcal{W}^\phi$ satisfies **P1** and **P2**. Thus, by Corollary 4.1.6:

$$D^2(x_N, \mathcal{W}^\phi(1)) \leq e^A B\|\Pi\|. \tag{4.12}$$

Using (iii) one obtains that:

$$\mathcal{W}^\phi(1) = \{x \in \mathbb{R}^n \mid \phi(1, x) \leq \phi(t_0, x_0)\} \subset \{x \in \mathbb{R}^n \mid g(x) \leq \phi(t_1, x_1)\}.$$

Let w_N be a closest point to x_N in $\mathcal{W}(1)$ and let κ be the Lipschitz constant of g. Then:

$$\begin{aligned} g(x_N) &\leq g(w_N) + \kappa\|x_N - w_N\|, \\ &\leq \phi(t_1, x_1) + \kappa D(x_N, \mathcal{W}^\phi(1)). \end{aligned}$$

The result follows from (4.12). □

Theorem 4.2.7. *The differential game $\mathcal{G}(t_1, x_1)$ has a value $\mathbf{V}$. Moreover, the extremal strategy $\beta(\mathbf{V}, \Pi, \pi)$ is asymptotically optimal for player 2, as $\|\Pi\| \to 0$.*

Proof. We claim that $\mathbf{V}^-$ satisfies (i), (ii) and (iii) and refer to the Appendix for a proof: $\mathbf{V}^-(1, x) = g(x)$, for all $x \in \mathbb{R}^n$, so that (iii) holds; (ii) can be easily deduced from the superdynamic programming principle (4.10) (Claim 1) or proved directly (Claim 3); Assumption 4.1.1 and 4.2.2 imply, using Gronwall's lemma, that the map $x \mapsto \mathbf{V}^-(t, x)$ is Lipschitz continuous for all $t \in [t_1, 1]$, so that (i) holds (Claim 2). Thus, by Proposition 4.2.6:

$$\mathbf{V}^+(t_1, x_1) \leq \sup_{\mathbf{u} \in \mathcal{U}} g(\mathbf{x}[t_1, x_1, \mathbf{u}, \beta(\mathbf{u})](1)) \leq \mathbf{V}^-(t_1, x_1) + C\sqrt{\|\Pi\|}.$$

The existence of the value follows by letting $\|\Pi\|$ tend to 0. Fix now the extremal strategy $\beta = \beta(\mathbf{V}, \Pi, \pi)$ of player 2. Then, to every strategy $\alpha \in \mathcal{A}$ of player 1 corresponds a unique control $\mathbf{u} \in \mathcal{U}$ so that, by Proposition 4.2.6:

$$\sup_{\alpha \in \mathcal{A}} g(\mathbf{x}[t_1, x_1, \alpha, \beta](1)) = \sup_{\mathbf{u} \in \mathcal{U}} g(\mathbf{x}[t_1, x_1, \mathbf{u}, \beta(\mathbf{u})](1)), \tag{4.13}$$

$$\leq \mathbf{V}(t_1, x_1) + C\sqrt{\|\Pi\|}. \tag{4.14}$$

Consequently, for any $\epsilon > 0$, the strategy $\beta(\mathbf{V}, \Pi, \pi)$ is ϵ-optimal for sufficiently small $\|\Pi\|$. □

4.3 Appendix

Claim 1. *The super dynamic programming principle* (4.10) *implies that* $\mathbf{V}^-$ *satisfies* (ii).

Proof. Identify every $u \in U$ with a strategy that plays u on $[t_0, 1]$ regardless of $\mathbf{v}$. Then:

$$\begin{aligned}\sup_{\alpha\in\mathcal{A}}\inf_{\beta\in\mathcal{B}} \mathbf{V}^-(t', \mathbf{x}[t_1, x_1, \alpha, \beta](t')) &\geq \sup_{u\in U}\inf_{\beta\in\mathcal{B}} \mathbf{V}^-(t', \mathbf{x}[t_1, x_1, u, \beta(u)](t')) \\ &\geq \sup_{u\in U}\inf_{\mathbf{v}\in\mathcal{V}} \mathbf{V}^-(t', \mathbf{x}[t_1, x_1, u, \mathbf{v}](t')).\end{aligned}$$

The first inequality is clear because $U \subset \mathcal{A}$; the second comes from the fact that $\beta(u) \in \mathcal{V}$ for all $u \in U$. □

Claim 2. $\mathbf{V}^-$ *satisfies* (i).

Proof. Using Assumption 4.1.1 and Gronwall's lemma one obtains that, for all $t \in [t_1, 1]$, $(\mathbf{u}, \mathbf{v}) \in \mathcal{U} \times \mathcal{V}$, and $x, y \in \mathbb{R}^n$:

$$\|\mathbf{x}[t_1, x, \mathbf{u}, \mathbf{v}](t) - \mathbf{x}[t_1, y, \mathbf{u}, \mathbf{v}](t)\| \leq e^{c(t-t_1)}\|x - y\|.$$

Let κ be a Lipschitz constant for g. Then, for all $(\mathbf{u}, \mathbf{v}) \in \mathcal{U} \times \mathcal{V}$, and for all $x, y \in \mathbb{R}^n$:

$$\big|g(\mathbf{x}[t_1, x, \mathbf{u}, \mathbf{v}](1)) - g(\mathbf{x}[t_1, y, \mathbf{u}, \mathbf{v}](1))\big| \leq \kappa e^c\|x - y\|.$$

Consequently, the map $x \mapsto \mathbf{V}^-(t, x)$ is κe^c-Lipschitz continuous for all $t \in [t_1, 1]$, which is a stronger requirement than (i). □

For the sake of completeness, let us end this note by proving that $\mathbf{V}^-$ satisfies (ii) directly. The super dynamic programming principle (4.10) can be proved in the same way.

Claim 3. $\mathbf{V}^-$ *satisfies* (ii).

Proof. Let $(t, x) \in [t_1, 1] \times \mathbb{R}^n$, let $t' \in [t, 1]$ and let $\epsilon > 0$ be fixed. An ϵ-optimal strategy for player 1 in $\mathcal{G}(t, x)$ is a strategy $\alpha \in \mathcal{A}$ such that:

$$\sup_{\mathbf{v}\in\mathcal{V}} g(\mathbf{x}[t, x, \alpha(\mathbf{v}), \mathbf{v}](1)) \geq \mathbf{V}^-(t, x) - \epsilon.$$

The Lipschitz continuity of $z \mapsto \mathbf{V}^-(t', z)$ implies the existence of some $\delta > 0$ such that any ϵ-optimal strategy in $\mathcal{G}(t', x')$ remains 2ϵ-optimal in $\mathcal{G}(t', z)$, for all $z \in B(x', \delta)$ (the euclidean ball of radius δ and center x'). By compactness, $B(x, \|f\|)$ can be covered by some finite family $(E_i)_{i\in I}$ of pairwise disjoint sets such that $E_i \subset B(x_i, \delta)$ for some $x_i \in B(x, \|f\|)$ $(i \in I)$. Let $\alpha_i \in \mathcal{A}$ $(i \in I)$ be an ϵ-optimal strategy for player 1 in $\mathcal{G}(t', x_i)$. For any $u \in U$ and $\mathbf{v} \in \mathcal{V}$, put $\mathbf{x}_{u,\mathbf{v}} := \mathbf{x}[x, t, u, \mathbf{v}]$. Note that $\mathbf{x}_{u,\mathbf{v}}(t')$ depends only on the restriction of $\mathbf{v}$ to $[t, t']$. The definition of α_i and E_i $(i \in I)$ ensures that, for all $\mathbf{v}' \in \mathcal{V}$:

$$g(\mathbf{x}[t', \mathbf{x}_{u,\mathbf{v}}(t'), \alpha_i, \mathbf{v}'](1))1_{\{\mathbf{x}_{u,\mathbf{v}}(t')\in E_i\}} \geq \mathbf{V}^-(t', \mathbf{x}_{u,\mathbf{v}}(t'))1_{\{\mathbf{x}_{u,\mathbf{v}}(t')\in E_i\}} - 2\epsilon.$$

For each $u \in U$, define a strategy $\alpha_u \in \mathcal{A}$ for player 1 in $\mathcal{G}(t, x)$ as follows. For all $\mathbf{v}' \in \mathcal{V}$:

$$\alpha_u(\mathbf{v}')(s) = \begin{cases} u & \text{if } s \in [t, t'), \\ \alpha_i(\mathbf{v}')(s) & \text{if } s \in [t', 1] \text{ and } \mathbf{x}_{u,\mathbf{v}}(t') \in E_i. \end{cases}$$

First, let us check that α_u is a strategy in $\mathcal{G}(t, x)$. Indeed, let $s_1 < \cdots < s_N$ be a common partition of $[t', 1]$ for the strategies $(\alpha_i)_i$ – this is possible because the family is finite.

Thus, α_u is a strategy with respect to the set of times $t < t' < s_2 < \cdots < s_N$. For any $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$, let $\mathbf{v}_1 \circ_t \mathbf{v}_2 \in \mathcal{V}$ be the concatenation of the two controls at time t, i.e. $(\mathbf{v}_1 \circ_t \mathbf{v}_2)(s) = \mathbf{v}_1(s)$ if $s \in [0,t]$ and $(\mathbf{v}_1 \circ_t \mathbf{v}_2)(s) = \mathbf{v}_2(s)$ if $s \in [t,1]$. Then, for any $\mathbf{v}'' = \mathbf{v} \circ_{t'} \mathbf{v}' \in \mathcal{V}$:

$$\begin{aligned} g\big(\mathbf{x}[t,x,\alpha_u,\mathbf{v}''](1)\big) &= \textstyle\sum_{i\in I} g\big(\mathbf{x}[t',\mathbf{x}_{u,\mathbf{v}}(t'),\alpha_i,\mathbf{v}'](1)\big)1_{\{\mathbf{x}_{u,\mathbf{v}}(t')\in E_i\}}, \\ &\geq \textstyle\sum_{i\in I} \mathbf{V}^-\big(t',\mathbf{x}_{u,\mathbf{v}}(t')\big)1_{\{\mathbf{x}_{u,\mathbf{v}}(t')\in E_i\}} - 2\epsilon, \\ &= \mathbf{V}^-\big(t',\mathbf{x}_{u,\mathbf{v}}(t')\big) - 2\epsilon. \end{aligned}$$

Taking the infimum in $\mathcal{V}$ and the supremum in U yields the desired result:

$$\begin{aligned} \mathbf{V}^-(t,x) &\geq \sup_{u\in U} \inf_{\mathbf{v}''\in\mathcal{V}} g\big(\mathbf{x}[t,x,\alpha_u,\mathbf{v}''](1)\big), \\ &\geq \sup_{u\in U} \inf_{\mathbf{v}\in\mathcal{V}} \mathbf{V}^-\big(t',\mathbf{x}_{u,\mathbf{v}}(t')\big) - 2\epsilon. \end{aligned}$$

Conclude by letting ϵ tend to 0. □

Chapter 5

Generalized solutions of HJI equations

Abstract: The purpose of this Chapter is to give a simple proof of the equivalence of solution concepts for HJI equations.

In this note we are interested in the following partial differential equation,

$$\frac{\partial w}{\partial t} + H(t, x, \nabla w) = 0$$

where $H : [0,1] \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a continuous function. These partial differential equations are usually called Hamilton-Jacobi-Isaacs equations and arise from two player, zero sum differential games. These equations usually do not admit classical solutions, i.e. continuously differentiable everywhere. Instead, one has to look for alternative definitions of solutions. We refer to each of these alternative definitions as a **solution concept** throughout this note. We are interested in proving the equivalence of three solution concepts: minimax solutions, due to Subbotin [96], proximal solutions, introduced by Clarke and Ledyaev [29] and viscosity solutions, introduced by Crandall and Lions [33]. The proof of the equivalence of these three solution concepts was already done in [29], but with different tools and in particular, using another equivalent definition of minimax solutions in terms of directional derivatives. Our approach is more in the spirit of viability and proximal calculus.

5.1 The HJI equations

We introduce the model where the HJI equations we are interested in arise. In this Section we describe the context that motivated us to look at the equivalence of these solution concepts, namely the connection between the geometrical approaches of Subbotin and Clarke with the viscosity solution approach for establishing the characterization of the value function of a two player, zero sum differential game. This Section is not essential for the rest of the Chapter and might be skipped by a reader who is only interested in the equivalence but not the motivation of the study of this particular equation.

The model in this Section is partially borrowed from Cardaliaguet and Quincampoix [25]. Let us assume we are given the following:

(i) U and V are compact subsets of some finite dimensional euclidean spaces.

(ii) $f : [0,1] \times \mathbb{R}^n \times U \times V \to \mathbb{R}^n$ is continuous, uniformly bounded and Lipschitz continuous with respect to the second variable.

(iii) f has linear growth in (t,x), i.e. $\exists \gamma, c > 0$ such that

$$\|f(t,x,u,v)\| \leq \gamma \|(t,x)\| + c, \ \forall (t,x,u,v) \in [0,1] \times \mathbb{R}^n \times U \times V.$$

(iv) $f(t,x,u,V) := \cup_{v \in V} f(t,x,u,v)$ and $f(t,x,U,v) := \cup_{u \in U} f(t,x,u,v)$ are convex for all (t,x,u) and (t,x,v), respectively.

(v) $g : \mathbb{R}^n \to \mathbb{R}$ is Lipschitz.

Denote by $\mathcal{U}(t_1)$ and $\mathcal{V}(t_1)$ the sets

$$\begin{aligned} \mathcal{U}(t_1) &:= \{\mathbf{u} : [t_1,1] \to U : \ \mathbf{u} \text{ is Lebesgue measurable } \}, \\ \mathcal{V}(t_1) &:= \{\mathbf{v} : [t_1,1] \to V : \ \mathbf{v} \text{ is Lebesgue measurable } \}. \end{aligned}$$

Consider a two player, zero sum differential game with dynamics

$$\dot{\mathbf{x}}(t) = f(t, \mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t)), \tag{5.1}$$

with initial data $(t_1, x_1) \in [0,1] \times \mathbb{R}^n$.

Under these Assumptions, a unique absolutely continuous solution of (5.1) exists, for a given pair $(\mathbf{u}, \mathbf{v}) \in \mathcal{U}(t_1) \times \mathcal{V}(t_1)$. The evaluation of this solution at time t is denoted by $\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](t)$. We will use the shorter notation $\mathbf{x}(t)$ when no confusion arises. At time $t = 1$, a terminal payoff $g(\mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](1))$, that player 1 receives from player 2.

To define a real strategic interaction we need to specify how are players allowed to react to their adversary. For that let us introduce the following notion of strategies:

Definition 5.1.1. A **non anticipating strategy with delay** (NAD) for player 1 is a function $\alpha : \mathcal{V}(t_1) \to \mathcal{U}(t_1)$ with the following property: there exists a partition $s_1 < \ldots < s_N = 1$ of $[t_1, 1]$ such that

$$\mathbf{v}_1 \equiv \mathbf{v}_2 \ \text{ in } \ [s_1, s_k] \ \implies \alpha[\mathbf{v}_1] \equiv \alpha[\mathbf{v}_2] \ \text{ in } \ [s_1, s_{k+1}], \ k = 1, \ldots N-1.$$

Non anticipating strategies with delay are defined symmetrically for player 2. The sets of non anticipating strategies with delay are denoted by $\mathcal{A}_d(t_1)$ and $\mathcal{B}_d(t_1)$, for player 1 and player 2, respectively. The main reason for using NAD's is that the game can be put in normal form:

Lemma 5.1.2. *Let $\alpha \in \mathcal{A}_d(t_1)$, $\beta \in \mathcal{B}_d(t_1)$. There exist a unique pair of controls $(\mathbf{u}(t), \mathbf{v}(t)) \in \mathcal{U}(t_1) \times \mathcal{V}(t_1)$ such that*

$$\alpha[\mathbf{v}](t) = \mathbf{u}(t), \quad \beta[\mathbf{u}](t) = \mathbf{v}(t).$$

The proof is by induction on the number of nodes of the partition. We refer to [17, Lemma 2.4] for the details.

For a pair of strategies $(\alpha, \beta) \in \mathcal{A}_d(t) \times \mathcal{B}_d(t)$, we denote by $\mathbf{x}[t, x, \alpha, \beta](s)$ the function $\mathbf{x}[t, x, \mathbf{u}, \mathbf{v}](s)$, where $(\mathbf{u}, \mathbf{v}) \in \mathcal{U}(t) \times \mathcal{V}(t)$ are the controls associated to the pair (α, β) as in Lemma 5.1.2. The **upper** and **lower value functions** are given by:

$$\begin{aligned} w^+(t,x) &:= \inf_{\beta \in \mathcal{B}_d} \sup_{\alpha \in \mathcal{A}_d} g(\mathbf{x}[t,x,\alpha\beta](1)), \\ w^-(t,x) &:= \sup_{\alpha \in \mathcal{A}_d} \inf_{\beta \in \mathcal{B}_d} g(\mathbf{x}[t,x,\alpha,\beta](1)). \end{aligned}$$

Moreover, the value functions are Lipschitz [25, Proposition 1] and they satisfy the following dynamic programming property [25, Proposition 2]: for all $(t,x) \in [0,1] \times \mathbb{R}^n$, and for all $s > t$,

$$w^+(t,x) = \inf_{\beta\in\mathcal{B}_d} \sup_{\alpha\in\mathcal{A}_d} w^+(s,\mathbf{x}[t,x,\alpha,\beta](s)), \tag{5.2a}$$

$$w^-(t,x) = \sup_{\alpha\in\mathcal{A}_d} \inf_{\beta\in\mathcal{B}_d} w^-(s,\mathbf{x}[t,x,\alpha,\beta](s)). \tag{5.2b}$$

If the value functions where smooth, the dynamic programming property and a Taylor series expansion would easily imply that w^+, w^- are solutions of

$$\frac{\partial w^+}{\partial t} + H^+(t,x,\nabla w^+) = 0, \tag{5.3a}$$

$$\frac{\partial w^-}{\partial t} + H^-(t,x,\nabla w^-) = 0. \tag{5.3b}$$

where

$$\begin{aligned} H^+(t,x,p) &:= \min_{v\in V}\max_{u\in U} \langle f(t,x,u,v),p\rangle\,, \\ H^-(t,x,p) &:= \max_{u\in U}\min_{v\in V} \langle f(t,x,u,v),p\rangle \end{aligned}$$

However, the value functions are not smooth in general. We now proceed to recall the three solution concepts that where mentioned in the introduction. For the rest of the note, we denote by $\mathcal{B}$ the unit ball in the corresponding euclidean space. For a C^1 function $\phi : [0,1]\times\mathbb{R}^n \to \mathbb{R}$ we denote with $\frac{\partial\phi}{\partial s}(t,x)$ its time derivative and by $\nabla_y\phi(t,x)$ its space derivative, when they are evaluated at the point (t,x).

5.2 Viscosity solutions

We recall first the now classical notion of viscosity solution of Crandall and Lions [33], applied to our framework.

Definition 5.2.1. (Viscosity solutions)

— A lower semicontinuous function $w : [0,1]\times\mathbb{R}^n \to \mathbb{R}$ is a **viscosity supersolution** of (5.3b) if for any $(t,x)\in[0,1]\times\mathbb{R}^n$ and a C^1 test function $\phi : [0,1]\times\mathbb{R}^n \to \mathbb{R}$ such that $\phi(t,x) = w(t,x)$ and $\phi(s,y) \le w(s,y)$ for (s,y) on a neighborhood of (t,x), then

$$\frac{\partial\phi}{\partial s}(t,x) + H^-(t,x,\nabla_y\phi(t,x)) \le 0. \tag{5.4}$$

— An upper semicontinuous function $w : [0,1]\times\mathbb{R}^n \to \mathbb{R}$ is a **viscosity subsolution** of (5.3a) if for any $(t,x)\in[0,1]\times\mathbb{R}^n$ and a C^1 test function $\phi : [0,1]\times\mathbb{R}^n \to \mathbb{R}$ such that $\phi(t,x) = w(t,x)$ and $\phi(s,y) \ge w(s,y)$ for (s,y) on a neighborhood of (t,x), then

$$\frac{\partial\phi}{\partial s}(t,x) + H^+(t,x,\nabla_y\phi(t,x)) \ge 0. \tag{5.5}$$

— A continuous function is a **viscosity solution** if it is both a viscosity super and subsolution.

5.3 Proximal solutions

Before the definition of proximal solutions, due to Clarke and Ledyaev [29], we recall first some concepts of proximal calculus.

Let S be a closed subset of $\mathbb{R}^n$ and consider a point x outside S. Denote the distance from a point $x \in \mathbb{R}^n$ to a set S by $d_S(x) := \min_{s\in S} \|x - s\|$.

Definition 5.3.1. The **proximal normal cone** at $s \in S$ is defined by

$$N_S^P(s) := \{\xi \in \mathbb{R}^n : \exists \lambda > 0 \text{ such that } d_S(s + \lambda\xi) = \lambda\|\xi\|\}.$$

It is easy to see that this definition is equivalent to

$$\{\xi \in \mathbb{R}^n : \ \exists \sigma = \sigma(\xi, s) \geq 0 \text{ s.t. } \langle \xi, s' - s\rangle \leq \sigma\|s' - s\|^2, \ \forall s' \in S\}. \tag{5.6}$$

Let us assume now that S is the epigraph of a lower semicontinuous function $h : [0, 1] \times \mathbb{R}^n \to \mathbb{R}$, that is,

$$S = \text{epi}h = \{(t, x, z) \in [0, 1] \times \mathbb{R}^n \times \mathbb{R} \ : \ z \geq h(t, x)\}.$$

Definition 5.3.2. A **proximal subgradient** of h at a point (t, x) is a vector $\xi = (\xi_t, \xi_x) \in \mathbb{R} \times \mathbb{R}^n$ such that

$$(\xi, -1) \in N^P_{\text{epi}h}(t, x, h(t, x)).$$

The (possibly empty) set of proximal subgradients, called **proximal subdifferential**, is denoted by $\partial^\pi h(t, x)$.

Observe that, since a cone is involved in the definition of proximal subdifferential, if $\alpha > 0$,

$$(\xi, -\alpha) \in N^P_{\text{epi}h}(t, x, h(t, x)) \implies \xi/\alpha \in \partial^\pi h(t, x).$$

The superdifferential, denoted $\partial_\pi h(t, x)$ is defined in a completely analogous way, by considering now the hypograph of an upper semicontinuous function.

Definition 5.3.3. (Proximal solutions)

— A lower semicontinuous function $w : [0, 1] \times \mathbb{R}^n \to \mathbb{R}$ is a **proximal supersolution** of (5.3b) if for all (t, x) and any $\xi = (\xi_t, \xi_x) \in \partial^\pi w(t, x)$,

$$\xi_t + H^-(t, x, \xi_x) \leq 0. \tag{5.7}$$

— An upper semicontinuous function $w : [0, 1] \times \mathbb{R}^n \to \mathbb{R}$ is a **proximal subsolution** of (5.3a) if for all (t, x) and any $\xi = (\xi_t, \xi_x) \in \partial_\pi w(t, x)$

$$\xi_t + H^+(t, x, \xi_x) \geq 0. \tag{5.8}$$

— A continuous function is a **proximal solution** if it is both a proximal super and subsolution.

5.3.1 Some results of proximal calculus

We close this section with three results concerning proximal normals and subgradients, that will be needed later.

Lemma 5.3.4. *Let $h : \mathbb{R}^N \to \mathbb{R}$ be lower semicontinuous and $(\xi, -\alpha) \in N^P_{\mathrm{epi}h}(y, h(y))$, where $\xi \in \mathbb{R}^N, \alpha \in \mathbb{R}$. Then $\alpha \geq 0$.*

Proof. By contradiction. Assume $\alpha < 0$. Then $(y, h(y) - \alpha) \in \mathrm{epi}h$ which gives

$$\exists t > 0 \text{ s.t. } \|t\xi\|^2 + t^2\alpha^2 = d^2_{\mathrm{epi}h}(y + t\xi, h(y) - t\alpha) \leq \|t\xi\|^2$$

which is a contradiction. □

The following result [80, Theorem 2.4], derived originally in the proof of [83, Theorem 1] tells us that we can approximate horizontal normals to the epigraph of a lower semicontinuous function by a sequence of non degenerate normals.

Theorem 5.3.5. *Let $h : \mathbb{R}^N \to \mathbb{R}$ be lower semi continuous and $(\psi, 0) \in N^P_{\mathrm{epi}h}(y, h(y))$, with $\psi \neq 0$. For every $\epsilon > 0$, there exist $y' \in y + \epsilon\mathcal{B}, \lambda \in (0, \epsilon)$ and $\xi \in \psi + \epsilon\mathcal{B}$ such that*

$$(\xi, -\lambda) \in N^P_{\mathrm{epi}h}(y', h(y')).$$

The last theorem gives a local characterization of proximal subgradients.

Theorem 5.3.6. *Let $h : [0, 1] \times \mathbb{R}^n \to \mathbb{R}$ be a lower semicontinuous function. Then $\xi \in \partial^\pi(h(t, x))$ if and only if there exist positive numbers σ, η such that*

$$h(s, y) \geq h(t, x) + \langle \xi, (s, y) - (t, x)\rangle - \sigma\|(s, y) - (t, x)\|^2 \tag{5.9}$$

for all $(s, y) \in (t, x) + \eta\mathcal{B}$.

The proof is a bit technical and thus we refer to [30, Theorem 2.5] for the details.

5.4 Minimax solutions

Before introducing the definition of minimax solutions, due to Subbotin [96], we recall some notions of viability theory. Let $F : \mathbb{R}^N \rightsquigarrow \mathbb{R}^N$ be a multiple valued map. A **trajectory** of the differential inclusion

$$\dot{y}(t) \in F(y(t)),\ y(t_1) = y_1, \tag{5.10}$$

is an absolutely continuous function $y : [t_1, 1] \to \mathbb{R}^N$ that satisfies (5.10). A pair (S, F) of a closed set $S \subset \mathbb{R}^N$ and a multiple valued map $F : \mathbb{R}^N \rightsquigarrow \mathbb{R}^N$ is called **viable** or **weakly invariant** if for every initial condition $(t_1, y_1) \in [0, 1] \times \mathcal{S}$ there exists a trajectory of (5.10) such that $y(t) \in S$ for all $t > t_1$.

For every $u \in U$, we will now consider the following set valued map $E^-_u : [0, 1] \times \mathbb{R}^n \times \mathbb{R} \rightsquigarrow \{1\} \times \mathbb{R}^n \times \{0\}$ given by

$$E^-_u(s, x, z) := \{1\} \times f(t, x, u, V) \times \{0\}$$

and the differential inclusion

$$(\dot{s}, \dot{x}, \dot{z}) \in E^-_u(s, x, z). \tag{5.11}$$

In a similar way, we also consider, for every $v \in V$, the set valued map, $E_v^+ : [0,1] \times \mathbb{R}^n \times \mathbb{R} \rightsquigarrow \{1\} \times \mathbb{R}^n \times \{0\}$ given by

$$E_v^+(s,x,z) := \{1\} \times f(t,x,U,v) \times \{0\} \tag{5.12}$$

and the differential inclusion

$$(\dot{s}, \dot{x}, \dot{z}) \in E_v^+(s,x,z).$$

Definition 5.4.1. (Minimax solutions)

— A lower semicontinuous function $w : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ is a **minimax supersolution** of (5.3b) if for any $u \in U$, $(\mathrm{epi} w, E_u^-)$ is viable.

— An upper semicontinuous function $w : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ is a **minimax subsolution** of (5.3a) if for any $v \in V$, $(\mathrm{hyp} w, E_v^+)$ is viable.

— A continuous function is a **minimax solution** if it is both a minimax super and subsolution.

An important tool for the sequel is the viability theorem, which we introduce now.

Definition 5.4.2. A set valued map $F : \mathbb{R}^N \rightsquigarrow \mathbb{R}^N$ is **Marchaud** if

a) For all y, $F(y)$ is a nonempty compact convex set.

b) F is upper semicontinuous, that is, $\forall y$ and $\forall \epsilon > 0 \exists \delta > 0$ such that

$$\|y' - y\| < \delta \implies F(y') \subset F(y) + \epsilon \mathcal{B}.$$

c) F has linear growth in y, i.e. there exist positive constants γ and c such that, for all $y \in \mathbb{R}^N$,

$$z \in F(y) \implies \|z\| \leq \gamma \|y\| + c.$$

Theorem 5.4.3. *(Viability Theorem) Let F be a Marchaud set valued map. Then the following are equivalent:*

i) (S, F) is viable.

ii) For all $y \in S$, $\forall p \in N_S^P(y)$, $\exists z \in F(y)$ s.t. $\langle p, z \rangle \leq 0$.

For the proof we refer to [4, Theorem 3.3.6].

A last useful result from viability theory is the following theorem, due to Filippov.

Theorem 5.4.4 (Filippov)**.** *Consider $\mathcal{Z}$ be a compact subset of a complete separable metric space. Let $h : \mathbb{R}^N \times \mathcal{Z} \to \mathbb{R}^N$ be a continuous function. If $q : \mathbb{R}^N \to \mathbb{R}^N$ is a measurable function that satisfies*

$$q(y) \in h(y, \mathcal{Z}), \text{ for almost all } y,$$

there is a measurable selection $\tilde{z}$, i.e. a function $\tilde{z} : \mathbb{R}^N \to \mathcal{Z}$ such that

$$q(y) = h(y, \tilde{z}(y)).$$

For the proof, see [5, Theorem 8.2.10].

Filippov's theorem allows us to parametrize trajectories of differential inclusions by a measurable control, as we show now.

Corollary 5.4.5. *Consider $\mathcal{Z}$ be a compact subset of a complete separable metric space. Let $h : \mathbb{R}^N \times \mathcal{Z} \to \mathbb{R}^N$ be a continuous function. If $y : [t_1, 1] \to \mathbb{R}^N$ is a trajectory of*

$$\dot{y} \in h(y, \mathcal{Z}),\ y(t_1) = y_1,$$

there exists a measurable control $\tilde{\mathbf{z}} : [0, 1] \to \mathcal{Z}$ such that

$$\dot{\mathbf{y}}(t) = h(\mathbf{y}(t), \tilde{\mathbf{z}}(t)),\ y(t_1) = y_1,$$

has a unique absolutely continuous solution.

Proof. From Filippov's theorem, one obtains easily that there exists a measurable control $\mathbf{z}_1 : \mathbb{R}^N \to \mathcal{Z}$ such that

$$\dot{\mathbf{y}}(t) = h(\mathbf{y}(t), \mathbf{z}_1(\mathbf{y}(t))),\ y(t_1) = y_1.$$

Since by assumption y is absolutely continuous, taking $\tilde{\mathbf{z}} = \mathbf{z}_1 \circ \mathbf{y}$ gives the desired measurable control. The uniqueness follows from Carathéodory's existence theorem [31, Theorem 1.1, chapter 2]. □

Before we proceed, let us clarify the interest in this differential inclusion informally. We focus on w^-. From the dynamic programming principle (5.2b), we see that, if the game starts at (t_1, x_1) and both players play optimally, they generate a trajectory on which w^- remains constant. When player 1 plays a constant control u, the viability of epiw^- with respect to (5.11) implies, by Corollary 5.4.5 that player 2 can find a control $\tilde{\mathbf{v}} \in \mathcal{V}(t_1)$ such that

$$w^-\left(s, \mathbf{x}[t_1, x_1, u, \tilde{\mathbf{v}}](s)\right) \leq w^-(t_1, x_1),\ \forall s \in [t_1, 1].$$

5.5 The equivalence

We will prove in this section the equivalence between the three solution concepts. Actually, we prove a stronger result, since we will prove the equivalence of supersolution concepts (the proofs for subsolutions being analogous).

Proposition 5.5.1. *The following are equivalent:*

(a) w^- is a viscosity supersolution of (5.3b).
(b) w^- is a proximal supersolution of (5.3b).
(c) w^- is a minimax supersolution of (5.3b).

Proof. We prove $(a) \implies (b) \implies (c) \implies (a)$.

— (a) ⇒ (b):
Let w^- be a viscosity supersolution and $\xi \in \partial^\pi w^-(t, x)$. Let σ, η as in Theorem 5.3.6. Define the following test function:

$$\phi(s, y) := w^-(t, x) + \langle \xi, (s, y) - (t, x) \rangle - \sigma \|(s, y) - (t, x)\|^2$$

for $(s, y) \in (t, x) + \eta\mathcal{B}$.
Observe that $\phi(t, x) = w^-(t, x)$ and $\phi(s, y) \leq w^-(s, y)$ by (5.9). Since $\frac{\partial \phi}{\partial s}(t, x) = \xi_t$ and $\nabla_y \phi(t, x) = \xi_x$, substituting these last two terms in in (5.4) gives (5.7).

— (b) $\Rightarrow$ (c):
Let $u \in U$ fixed. Let $(\psi_t, \psi_x, -\alpha) \in N^P_{\text{epi}w^-}(t, x, w^-(t,x))$. By Lemma 5.3.4, we only need to consider $\alpha \geq 0$. If $\alpha > 0$, then $\xi := \psi/\alpha \in \partial^\pi w^-(t,x)$ and since w is a proximal supersolution,

$$\xi_t + \min_{v \in V} \langle f(t,x,u,v), \xi_x \rangle \leq 0.$$

Since the above minimum is attained, there exists $\bar{v} \in V$ such that

$$\langle (\xi_t, \xi_x, -1), (1, f(t,x,u,\bar{v}), 0) \rangle \leq 0$$

which implies (c) by Theorem 5.4.3. It remains to look at the case $\alpha = 0$.
Fix $\epsilon > 0$ and $\{\epsilon_n\}_{n\in\mathbb{N}}$ a sequence of positive numbers that converges to zero. Applying Theorem 5.3.5 for each ϵ_n we obtain that there exists $(t_n, x_n, \psi_{t_n}, \psi_{x_n}, -\lambda_n)$ such that

$$(\psi_{t_n}, \psi_{x_n}, -\lambda_n) \in \partial^\pi w^-(t_n, x_n).$$

and $\lambda_n > 0$. Note that the sequence $\{(t_n, x_n, \psi_{t_n}, \psi_{x_n}, -\lambda_n)\}_{n\in\mathbb{N}}$ converges to $(t, x, \psi_t, \psi_x, 0)$ as n goes to infinity.
Now let $v_n \in V$ such that

$$\psi_{t_n} + \langle f(t_n, x_n, u, v_n), \psi_{x_n} \rangle \leq 0.$$

It follows that, for $n \geq N_\epsilon$,

$$\begin{aligned} \psi_t + \min_{v\in V} \langle f(t,x,u,v), \psi_x \rangle &\leq \psi_t + \langle f(t,x,u,v_n), \psi_x \rangle \\ &\leq \psi_{t_n} + \langle f(t_n,x_n,u,v_n), \psi_{x_n} \rangle + \epsilon \\ &\leq \epsilon. \end{aligned}$$

The second inequality follows from the convergence of $\{(t_n, x_n, \psi_{t_n}, \psi_{x_n}, -\lambda_n)\}_{n\in\mathbb{N}}$ and the third from the choice of v_n.
— (c) $\Rightarrow$ (a):
Take $\phi : [0,1] \times \mathbb{R}^n \to \mathbb{R}$ a C^1 test function such that

$$\phi(t,x) = w^-(t,x), \quad \phi(s,y) \leq w^-(s,y)$$

for (s,y) in a neighborhood $\mathcal{N}_{(t,x)}$ of (t,x).
Now fix $u \in U$. We have that, by Corollary 5.4.5, there exists $\tilde{\mathbf{v}}$ such that

$$w^-\left(s, \mathbf{x}[t,x,u,\tilde{\mathbf{v}}](s)\right) \leq w^-(t,x).$$

For s small enough, $(s, \mathbf{x}[t,x,u,\tilde{\mathbf{v}}](s))$ belongs to $\mathcal{N}_{(t,x)}$. From the definition of ϕ,

$$\phi\left(s, \mathbf{x}[t,x,u,\tilde{\mathbf{v}}](s)\right) \leq \phi(t,x).$$

Let $y := x[t,x,u,\tilde{\mathbf{v}}](s)$. Performing a Taylor series expansion around (t,x) on the left hand side,

$$\phi(t,x) + \frac{\partial \phi}{\partial s}(t,x)(s-t) + \langle y - x, \nabla_y \phi(t,x) \rangle + O(\|(s,y) - (t,x)\|^2) \leq \phi(t,x).$$

This leads to

$$\frac{\partial \phi}{\partial s}(t,x)(s-t)+\left\langle \int_t^s f(\tau, \mathbf{x}[t,x,u,\tilde{\mathbf{v}}](\tau), \mathbf{x}(\tau), u, \tilde{\mathbf{v}}(\tau))d\tau, \nabla_y \phi(t,x)\right\rangle + O(\|(s,y)-(t,x)\|^2) \leq 0.$$

Lebesgue's differentiation theorem tells us that

$$\frac{1}{|s-t|}\int_t^s f(\tau, \mathbf{x}[t,x,u,\tilde{\mathbf{v}}](\tau), u, \tilde{v}(\tau))d\tau \to f(t, \mathbf{x}(t), u, \tilde{\mathbf{v}}(t)), \quad \text{as} \quad s \to t,$$

for almost every t. Thus, dividing by $(s-t)$ and taking the limit as s goes to t,

$$\frac{\partial \phi}{\partial s}(t,x) + \langle f(t,x,u,\tilde{v}(t)), \nabla_y \phi(t,x)\rangle \leq 0,$$

which clearly implies

$$\frac{\partial \phi}{\partial s}(t,x) + \min_{v \in V} \langle f(t,x,u,v), \nabla_y \phi(t,x)\rangle \leq 0.$$

Since u is arbitrary, from our compacity and continuity assumptions we can take the maximum with respect to u and conclude the proof.

□

Part II

Discrete time mean field games

Chapter 6

Discrete time mean field games

Abstract: We study a discrete time, finite horizon game with a continuum of identical players. Our work is motivated by the theory of mean field games, recently introduced by Lasry and Lions and by Huang, Caines and Malhamé. We prove that the discrete analogue of a mean field equilibrium is a distributional equilibrium in the sense of Jovanovic and Rosenthal of the game with a continuum of players and an approximate Nash equilibrium of the game with finitely many players, where the approximation error goes to zero as the number of players tends to infinity. This chapter is based on some work in progress with S.C.P. Yam, Chinese University of Hong Kong.

6.1 Introduction

In this paper we study a mean field model for a discrete time, discrete state space finitely repeated stochastic game. Mean field games have been introduced independently by Huang, Caines and Malhamé [56] and by Lasry and Lions [64, 65, 66] and have received considerable attention in the literature. For a comprehensive introduction see for instance Guéant, Lasry and Lions [49] or the lecture notes of Cardaliaguet [20], based on Lions' lectures at the Collège de France, as well as the book by Bensoussan, Frehse and Yam [12].

One important motivation for the mean field approach in applications is that it allows to construct approximate Nash equilibria of games with a large number of players. These equilibria are "simple" in the sense that one replaces a complicated stochastic process (the behaviour of our adversaries) by a classical Markov decision process via an averaging argument.

A common feature of the above mentioned literature is that they study the continuous time framework. We believe that the discrete time setting has independent interest and might provide a powerful tool for many applications. An important exception is the work of Gomes, Mohr and Souza [48], in which a model for a finite horizon, discrete time, finite state dynamic game with infinitely many players is studied and its asymptotic behaviour as the time horizon tends to infinity is investigated. The authors obtain exponential convergence to a stationary solution. Our asymptotic result is concerned instead with the number of players and in our framework the time horizon remains fixed.

Our work is closer to Adlakha, Johari and Weintraub [2]. The authors consider an infinite horizon, discrete time discounted game with a discrete but unbounded state space and discounted payoff. We consider instead a finite horizon game.

The approach of the model we present here is morally in the spirit of Huang, Caines and Malhamé [57]: construct a *limit object* and use it to approximate Nash equilibria of

the finite player game. The complementary approach of studying the limit behaviour of N player games as $N \to +\infty$ has been developed by Lasry and Lions [64] for an ergodic payoff (see Feleqi [37] for a detailed derivation) and by Bardi [8] for the linear-quadratic case.

Our paper is organized as follows: we introduce the notion of mean field equilibrium in Section 6.2 and prove its existence. In Section 6.3 we prove that the mean field equilibrium is a distributional equilibrium as defined by Jovanovic and Rosenthal [59] of the game with a continuum of players. Finally, in Section 6.4 we study the game with finitely many players and provide the proof of the main result, namely the approximation of a Nash equilibrium of the finite player game as the number of players goes to infinity.

6.2 The mean field game equilibrium

Let A denote a compact subset of a metric space, which will be referred to as the **action** set. Denote by $\mathcal{P}(A)$ the set of Borel probability measures on A. Recall that $\mathcal{P}(A)$ is compact with the weak-* topology and is metrizable by the Kantorovich-Rubinstein distance.

Denote by $\mathcal{X}$ the **state** space, which, unless otherwise stated, we assume to be a finite set. The set of probability distributions on $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$, which is naturally embedded in $\mathbb{R}^{|\mathcal{X}|}$ when $\mathcal{X}$ is finite and thus equipped with the euclidean metric. The current and terminal **payoff** functions

$$\ell : \mathcal{X} \times A \times \mathcal{P}(\mathcal{X}) \to [0,1], \quad g : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \to [0,1]$$

are uniformly bounded, jointly continuous and Lipschitz continuous in the last variable uniformly with respect to the remaining variables, with Lipschitz constants L_ℓ, L_g respectively.

The **transition function**

$$Q : \mathcal{X} \times A \times \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$$

is jointly continuous and bounded and satisfies, for all $(x, a, b, m, m') \in \mathcal{X} \times A \times A \times \mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{X})$ the following Lipschitz conditions:

$$\begin{aligned} \|Q(x,a,m) - Q(x,b,m)\|_\infty &\leq L_q \|a-b\| \\ \|Q(x,a,m) - Q(x,a,m')\|_\infty &\leq \|m-m'\|_\infty. \end{aligned}$$

These assumptions will hold during the rest of the Chapter.

Denote with $\mathcal{T}$ the set $\{0, 1, 2 \ldots, T-1\}$ which represents the set of **stages** of the game and let $\mathbf{m} = (m_t)_{t=0}^T \in \mathcal{P}(\mathcal{X})^{T+1}$ be an external and fixed parameter.

Consider the following **one player** game $\Gamma_{\mathbf{m}}$: at stage $t \in \mathcal{T}$, the player observes his own state x_t and chooses an action a_t. Once the action is chosen, he receives the payoff $\ell(x_t, a_t, m_t)$. The new state x_{t+1} is chosen randomly using the transition function $Q(x_t, a_t, m_t)$ and the situation is repeated. At stage $t = T$ a final payoff $g(x_T, m_T)$ is allocated. The initial state of the player is chosen using the distribution m_0. We assume also that $\mathbf{m}$ is known.

Let $\mathcal{H}_t = (\mathcal{X} \times A)^t$ denote the history up to stage t and let $\mathcal{H} = \cup_{t \in \mathcal{T}} \mathcal{H}_t$ the set of possible histories. A **pure strategy** is a function $\sigma : \mathcal{H} \to A$ where $\sigma = (\sigma_t)_{t \in \mathcal{T}}$. Denote by Σ the set of strategies.

The **total payoff** functional for the player, when he uses the strategy σ, is denoted by $J_1 : \mathcal{X} \times \Sigma \times \mathcal{P}(\mathcal{X})^{T+1} \to [0,2]$ and is given by

$$J_1(x,\sigma,\mathbf{m}) := \mathbb{E}^Q\left[\sum_{t\in\mathcal{T}} \ell(x_t^i,\sigma_t(x_t),m_t) + g(x_T,m_T) \mid x_0 = x\right].$$

We introduce the following value function for a one player game whose state is $x \in \mathcal{X}$ at time $s = 0,\dots T$:

$$V(s,x,\mathbf{m}) := \sup_{\sigma\in\Sigma} \mathbb{E}^Q\left[\sum_{t\geq s}^{T-1} \ell(x_t,\sigma_t(x_t),m_t) + g(x_T,m_T) \mid x_s = x\right].$$

Here, $\mathbb{E}^Q$ denotes the expectation with respect to the (random) transition function Q. From the familiar arguments, see for instance Hernández-Lerma and Lasserre [53, Section 3.2] we obtain the following dynamic programming equation:

$$V(s,x,\mathbf{m}) \quad = \quad \max_{a\in A}\left\{\ell(x,a,m_s) + \sum_{y\in\mathcal{X}} Q(x,a,m_s)(y)\cdot V(s+1,y,\mathbf{m})\right\} \tag{6.1}$$

with terminal condition $V(T,x,\mathbf{m}) = g(x,m_T)$.

Now let $\sigma \in \Sigma$ and let $m_0^\sigma := m_0$. We define, for $t \geq 0$:

$$m_{t+1}^\sigma(x) := \sum_{y\in\mathcal{X}} Q(y,\sigma_t(y),m_t^\sigma)(x)\cdot m_t^\sigma(y). \tag{6.2}$$

An important consequence of the dynamic programming equation (6.1) is the existence of optimal **Markovian** pure strategies, i.e. functions of the form $\sigma = (\sigma_t)_{t\in\mathcal{T}}$ where $\sigma_t : \mathcal{X} \to A$. Denote by Σ_M the set of Markovian strategies.

We are ready to introduce the main concept of this section.

Definition 6.2.1. A **mean field equilibrium** is a pair $(\sigma,\mathbf{m}) \in \Sigma\times\mathcal{P}(\mathcal{X})^{T+1}$ such that:

1. σ is an optimal strategy in the one player game $\Gamma_{\mathbf{m}}$, computed using the dynamic programming equation (6.1) .
2. $\mathbf{m}$ is the trajectory followed by m_0 according to (6.2) for the strategy σ.

We will provide an interpretation of the mean field equilibrium in Section 6.3. Let us first establish its existence in the remaining of this Section. The crucial assumption we need for that is the following.

Assumption 6.2.2. *Let $\mathcal{F}$ denote the set of functions $f : \mathcal{T}\times\mathcal{X}\times\mathcal{P}(\mathcal{X})^{T+1} \to [0,1]$. Assume that for all $(s,x,\mathbf{m},f) \in \mathcal{T}\times\mathcal{X}\times\mathcal{P}(\mathcal{X})^{T+1}\times\mathcal{F}$, the quantity*

$$\left\{\ell(x,a,m_s) + \sum_{y\in\mathcal{X}} Q(x,a,m_s)(y)\cdot f(s+1,y,\mathbf{m})\right\}$$

viewed as a function of a, reaches it maximum at a unique point.

Assumption 6.2.2 holds in particular in the following case.

Assumption 6.2.3. *Assume A is convex and that for all $(x,m) \in \mathcal{X}\times\mathcal{P}(\mathcal{X})$, $\ell(x,\cdot,m)$ and $Q(x,\cdot,m)$ are affine functions.*

Proposition 6.2.4. *Let m_0 denote the initial distribution of the players in the state space $\mathcal{X}$. Under Assumption 6.2.2, there exists a mean field equilibrium.*

Proof. Let $\Phi : \mathcal{P}(\mathcal{X})^{T+1} \to \Sigma_M$ the map that sends $\mathbf{m}$ to the optimal Markovian strategy $\sigma^{\mathbf{m}} = (\sigma_t^{\mathbf{m}})_{t\in\mathcal{T}}$ in the game $\Gamma_{\mathbf{m}}$. Note that $\sigma^{\mathbf{m}}$ is unique from Assumption 6.2.2. From the continuity assumptions and the dynamic programming equation it easily follows by induction that Φ is continuous. The map $\Psi : \Sigma_M \to \mathcal{P}(\mathcal{X})^{T+1}$ that sends a Markovian strategy $(\sigma_t)_{t\in\mathcal{T}}$ to the vector $\mathbf{m}^\sigma$ computed by (6.2) is also easily verified to be continuous. Since $\Psi \circ \Phi$ is the composition of continuous maps and $\mathcal{P}(\mathcal{X})^{T+1}$ is compact and convex, then it has a fixed point by Schauder's fixed point theorem. □

Note that the above existence result also holds in the case where $\mathcal{X}$ is a compact metric space.

6.3 The game with a continuum of players

Let us provide an interpretation of the mean field equilibrium as an equilibrium for a game with a continuum of players in a suitable sense, called distributional equilibrium.

The notion of distributional equilibrium for games with a continuum of identical agents was introduced by Mas-Colell [72] (under the name Cournot-Nash equilibrium) and later extended to stochastic games by Jovanovic and Rosenthal [59]. Distributional equilibria have also been studied in the framework of one shot games with finitely many players and incomplete information, see Milgrom and Weber [76]. For this Section, we allow $\mathcal{X}$ to be a compact metric space.

Let I be a continuum of players, for instance take $I = [0, 1]$. Let $m_0 \in \mathcal{P}(\mathcal{X})$ denote the initial distribution of the players. The game is played as follows: at stage $t = 0, 1 \dots T-1$, player i observes his own state x_t^i and the state distribution of the players m_t and chooses an action a_t^i. Actions are chosen simultaneously and independently. Once the actions are chosen, player i receives the payoff $\ell(x_t^i, a_t^i, m_t)$ and (ignoring any potential measurability problems for the moment) the corresponding distribution on the state-action space, $\Delta_t \in \mathcal{P}(\mathcal{X} \times A)$, is announced. The marginal distribution $\Delta_{t,\mathcal{X}}$ on $\mathcal{X}$ satisfies $\Delta_{t,\mathcal{X}} = m_t$. The new state distribution is given by

$$F\#\Delta_t(\cdot) := \int_{\mathcal{X}\times A} Q(y, a, m_t)(\cdot)\Delta_t(dy \times da).$$

and the situation is repeated.

Assume a vector of state-action distributions $\Delta := (\Delta_t)_{t\in\mathcal{T}}$, where $\Delta_t \in \mathcal{P}(\mathcal{X} \times A)$ is fixed. Define the following sequence

$$V_s(x, \Delta) := \max_{a\in A} \left\{ \ell(x, a, \Delta_{s,\mathcal{X}}) + \int_{\mathcal{X}} V_{s+1}(x', \Delta) dF\#\Delta_t(x') \right\} \tag{6.3}$$

with $V_T(x, \Delta) = g(x, F\#\Delta_{T-1})$. This represents the optimal expected payoff a player at state x in time s would get if the sequence of state-action distributions Δ is fixed.

Definition 6.3.1. Let m_0 fixed. The sequence of state-action distributions Δ is a **distributional equilibrium** if:

1. $\Delta_{0,\mathcal{X}} = m_0$, and $\Delta_{t+1,\mathcal{X}} = F\#\Delta_t$
2. $\forall t \in \mathcal{T}$, $\Delta_t \left(\{(x, a) \in \mathcal{X} \times A : J_\infty(x, a, \Delta_t) \geq J_\infty(x, a', \Delta_t), \forall a' \in A\}\right) = 1.$

where $(V_s)_{s\in\mathcal{T}}$ is defined by (6.3) and

$$J_\infty(x, a, \Delta_s) := \ell(x, a, \Delta_{s,\mathcal{X}}) + \int_{\mathcal{X}} V_{s+1}(x', \Delta) dF\#\Delta_t(x').$$

Note that the way in which distributional equilibrium is defined helps us to avoid measurability problems: intuitively, we are looking at the sequence of state-action distributions *after* it happened.

The existence of distributional equilibrium for stochastic games with a continuum of anonymous players has been established by Jovanovic and Rosenthal [59, Theorem 1].

Proposition 6.3.2. *The mean field equilibrium (σ, m) is a distributional equilibrium.*

Proof. Let (σ, m) denote a mean field equilibrium and for $t = 0, \dots T-1$ let $\mu_t := m_t \otimes \sigma_t$. Observe that $\mu_t(x, a) > 0 \iff a = \sigma_t(x)$. The marginal distributions satisfy $\mu_{t,\mathcal{X}} = m_t$ and from (6.2) we obtain that $m_{t+1} = F\#m_t$. Thus, the first condition for a distributional equilibrium is satisfied. The second condition is also satisfied from the optimality of σ. □

6.4 The game with finitely many players

Let I denote the set of players and assume $|I| = N$. Let $X^j_{t,N}$ be a random variable that describes the position of player j at time t. The states of the players at time $t = 0$ are chosen i.i.d. We reserve capital letters for random variables and lower case letters for their realizations.

The N-player game is played as follows: at stage 0, the state of each player is chosen using the lottery m_0, which gives an average state distribution denoted by $m_{0,N}$. At stage $t = 1 \dots T-1$, player i observes his own state $x^i_{t,N}$ and the average state distribution $m_{t,N}$ and chooses an action $a^i_{t,N}$. Actions are chosen simultaneously and independently. Once player i has chosen his action, he receives the payoff $\ell(x^i_{t,N}, a^i_{t,N}, m_{t,N})$. The new state $X^i_{t+1,N}$ is chosen randomly using the transition function $Q(x^i_{t,N}, a^i_{t,N}, m_{t,N})$. The random average state distribution is denoted by $M_{t+1,N} := \frac{1}{N}\sum_{j\in I} \delta_{X^j_{t+1,N}}$.

At the beginning of stage $t+1$, the realization of $X^i_{t+1,N}$ and $M_{t+1,N}$, denoted $x^i_{t+1,N}$ and $m_{t+1,N}$ respectively, are observed, and the situation is repeated. At stage $t = T$ a final payoff $g(x^i_{T,N}, m_{T,N})$ is allocated.

A **behavioral strategy** for player i is a vector $\pi^i = (\pi^i_t)^T_{t=1}$ where $\pi^i_t : \tilde{\mathcal{H}}_t \to \mathcal{P}(A)$ and $\tilde{\mathcal{H}}_t = (\mathcal{X} \times A \times \mathcal{P}(\mathcal{X}))^t$ is the set of all possible histories up to date t. Denote by Σ the set of behavioral strategies for each player and note that $\Sigma_M \subset \Sigma$, where Σ_M denotes the set of Markovian strategies, that is, the set of functions of the form $\sigma : [1, T] \times \mathcal{X} \to \mathcal{P}(A)$.

A **strategy profile** is a vector $\pi = (\pi^i)_{i\in I}$, where π^i is a behavioral strategy of player i. The payoff of player i, when using the strategy π^i and when his adversaries use the strategy profile $\pi^{-i} \in \Sigma^{N-1}$ is

$$J^i_N(x, m_0, \pi^i, \pi^{-i}) := \mathbb{E}^Q_\pi \left\{ \sum_{t\in\mathcal{T}} \ell(x^i_{t,N}, a^i_{t,N}, m_{t,N}) + g(x^i_{T,N}, m_{T,N}) \right\}$$

Definition 6.4.1. An $\epsilon-$**Nash equilibrium** where $\epsilon > 0$, is a strategy profile $(\pi^i)_{i\in I}$ such that, for all player i and all behavioural strategy τ^i,

$$J^i_N(x, m_0, \tau^i, \pi^{-i}) - \epsilon \le J^i_N(x, m_0, \pi^i, \pi^{-i}).$$

Our main result is the following:

Theorem 6.4.2. *Let $(\sigma, \mathbf{m})$ be a mean field equilibrium. For all $\epsilon > 0$ there exists N_ϵ such that, if $N \ge N_\epsilon$ the Markovian strategy σ is an ϵ-Nash equilibrium in the N player game.*

Moreover, if $J_N^i(x,\sigma,\mathbf{m})$ denotes the payoff of player i in the N-player game when all players follow σ and player $i's$ initial state is x, then:

$$|J_N^i(x,\sigma,\mathbf{m}) - J_1(x,\sigma,\mathbf{m}| \to 0$$

almost surely as $N \to +\infty$.

Before proceeding to the proof, we collect some technical results, starting with a refinement of the law of large numbers for Bernoulli variables due to Adlakha, Johari and Weintraub [2, Lemma 11].

Lemma 6.4.3. *Suppose $0 \le p_N \le 1$ for all N, and that $p_N \to p$ as $N \to +\infty$. For each N, let $Z_{1,N}, \dots Z_{N,N}$ be i.i.d Bernoulli random variables with parameter p_N. Then almost surely:*

$$\lim_{N\to+\infty} \frac{1}{N}\sum_{k=1}^{N} Z_{k,N} = p.$$

Proof. Let $\epsilon > 0$. Using Hoeffding's inequality,[55]

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{k=1}^{N} Z_{k,N} - p_N\right| > \epsilon\right) \le 2e^{-2N\epsilon^2}.$$

Then, by the Borel-Cantelli lemma, the event on the left hand side occurs almost surely for only finitely many N. The result now follows. □

The following Lemma is the crucial ingredient for our main result.

Lemma 6.4.4. *Let x be a fixed initial state, $(\sigma,\mathbf{m})$ a mean field equilibrium. Assume all the players $j \neq i$ in the N-player game follow the mean field equilibrium strategy σ and that player i follows the strategy τ. Then, for all $t \in \mathcal{T}$,*

$$\|m_{t,N} - m_t\|_\infty \to 0$$

almost surely as $N \to +\infty$.

Proof. Denote by $M_{t+1,\sigma,N}^{-i} := \frac{1}{N-1}\sum_{j\neq i}\delta_{X_{t+1,N}^j}$ the average state distribution of the players using σ.

We proceed by induction. First observe that the case $t = 0$ is clear by the strong law of large numbers.

Now let $\left(Y_{t,\sigma,N}^{y,k}\right)_{k=1}^{(N-1)\cdot m_{t,\sigma,N}(y)}$ be a sequence of Bernoulli i.i.d. random variables such that

$$Y_{t,\sigma,N}^{y,k} = \begin{cases} 1, \text{ with probability } Q(y,\sigma_t(y),m_{t,N})(x) \\ 0, \text{ with probability } 1 - Q(y,\sigma_t(y),m_{t,N})(x). \end{cases}$$

Hence,

$$\begin{aligned}(N-1)\cdot M_{t+1,\sigma,N}^{-i}(x) &= \sum_{y\in\mathcal{X}}\left(\sum_{k=1}^{(N-1)\cdot m_{t,N}(y)} Y_{t,\sigma,N}^{y,k}\right) \\ &= \sum_{y\in\mathcal{X}}(N-1)\cdot m_{t,N}(y)\left(\frac{1}{(N-1)\cdot m_{t,N}(y)}\sum_{k=1}^{(N-1)\cdot m_{t,N}(y)} Y_{t,\sigma,N}^{y,k}\right).\end{aligned}$$

In the above equations, we interpret the term on the parenthesis as zero if $m_{t,N}(y) = 0$. Since we are interested in large values of N, we only have to consider those states y such that $m_t(y) > 0$ by induction hypothesis.

For such states y, the term on parenthesis converges to $Q(y, \sigma_t(y), m_t)(x)$ almost surely from Lemma 6.4.3. Since

$$M_{t+1,N} = \frac{1}{N}\left(X^i_{t+1,N} + (N-1)M^{-i}_{t+1,\sigma,N}\right)$$

the result now follows. □

Lemma 6.4.5. *Let x be a fixed initial state, (σ, m) a mean field equilibrium and $(a^i_{t,N})_{t\in\mathcal{T}}$ an arbitrary sequence of actions of player i.*

Consider the following two trajectories:

1. *The trajectory $(x^i_{t,N})_{t=0}^T$ of player i in the N- player game where all his adversaries follow a mean field equilibrium strategy σ, defined by*

$$x^i_{t+1,N} \sim Q(x^i_{t,N}, a^i_{t,N}, m_{t,N}).$$

2. *The trajectory generated by*

$$x^i_{t+1} \sim Q(x^i_t, a^i_{t,N}, m_t).$$

Then, for all $y \in \mathcal{X}$ and for all $t \in \mathcal{T}$,

$$|\mathbb{P}(x^i_{t,N} = y) - \mathbb{P}(x^i_t = y)| \to 0$$

almost surely as $N \to +\infty$.

Proof. We proceed by induction. Observe first that

$$\begin{aligned}
\mathbb{P}(x^i_{t+1,N} = y) &= \sum_{z\in\mathcal{X}} \mathbb{P}(x^i_{t,N} = z)\mathbb{P}(y \,|\, z, a^i_{t,N}, m_{t,N}) \\
\mathbb{P}(x^i_{t+1} = y) &= \sum_{z\in\mathcal{X}} \mathbb{P}(x^i_t = z)\mathbb{P}(y \,|\, z, a^i_{t,N}, m_t)
\end{aligned}$$

The case $t = 0$ is easy since all the terms in the above sums are zero except the one involving the initial state x, hence

$$\begin{aligned}
|\mathbb{P}(x^i_{1,N} = y) - \mathbb{P}(x^i_1 = y)| &= |Q(x, a^i_{0,N}, m_{0,N})(y) - Q(x, a^i_{0,N}, m_0)(y)| \\
&\leq L_Q \left\|m_{0,N} - m_0\right\|_\infty
\end{aligned}$$

which converges to zero almost surely by the strong law of large numbers. Now observe that

$$\mathbb{P}(x^i_{t+1,N} = y) - \sum_{z\in\mathcal{X}} \mathbb{P}(x^i_{t,N} = z)\mathbb{P}(y \,|\, z, a^i_{t,N}, m_t) \leq L_Q \left\|m_{t,N} - m_t\right\| \tag{6.4}$$

and

$$\sum_{z\in\mathcal{X}} \mathbb{P}(x^i_{t,N} = z)\mathbb{P}(y|z, a^i_{t,N}, m_t) - \mathbb{P}(x^i_{t+1} = y) \tag{6.5}$$

is equal to

$$\sum_{z\in\mathcal{X}} \Big(\mathbb{P}(x^i_{t,N}=z)-\mathbb{P}(x^i_t=z)\Big)\cdot \mathbb{P}(y|z,a^i_{t,N},m_t). \tag{6.6}$$

Adding (6.5) to the left hand side of (6.4) and (6.6) to the right hand side, the remaining quantity in the right hand side converges to zero almost surely by the induction hypothesis, and Lemma 6.4.4. □

Lemma 6.4.6. *Consider the trajectories in Lemma 6.4.5. We have that, for all* $t = 0,\dots T-1$,

$$\limsup_{N\to+\infty} \mathbb{E}\left[\ell(x^i_{t,N},a^i_{t,N},m_{t,N})-\ell(x^i_t,a^i_{t,N},m_t)\right]\le 0.$$

Proof. Let

$$E^i_{t,N}=\mathbb{E}\left[\ell(x^i_{t,N},a^i_{t,N},m_{t,N})-\ell(x^i_t,a^i_{t,N},m_t)\right].$$

We have that

$$\begin{aligned} E^i_{t,N} &= \mathbb{E}\left[\ell(x^i_{t,N},a^i_{t,N},m_{t,N})-\ell(x^i_{t,N},a^i_{t,N},m_t)\right]\\ &+ \mathbb{E}\left[\ell(x^i_{t,N},a^i_{t,N},m_t)-\ell(x^i_t,a^i_{t,N},m_t)\right]\\ &=: A_1+A_2. \end{aligned}$$

Observe that the term A_1 is easily bounded since

$$\begin{aligned} A_1 &\le \sum_{y\in\mathcal{X}} \mathbb{P}(x^i_{t,N}=y)\max_{a\in A}|\ell(y,a,m_{t,N})-\ell(y,a,m_t)|\\ &\le L_\ell\,\|m_{t,N}-m_t\|, \end{aligned}$$

which, by Lemma 6.4.4, tends to zero almost surely. For the term A_2, observe that

$$A_2\le \sum_{y\in\mathcal{X}} |\mathbb{P}(x^i_{t,N}=y)-\mathbb{P}(x^i_t=y)|\cdot\|\ell\|$$

which also tends to zero as $N\to+\infty$. □

We are ready to prove Theorem 6.4.2.

Proof. Let τ^i be any strategy of player i in the game with N players and let $(\sigma,\mathbf{m})$ be a mean field equilibrium. If player i uses the strategy τ^i and everyone else uses the strategy σ, the corresponding payoff is

$$J^i_N(x,m_0,\tau^i,\sigma^{-i})=\mathbb{E}^Q_{\tau^i,\sigma^{-i}}\left[\sum_{t\in\mathcal{T}}\ell(x,\tau^i(x^i_{t,N},m^{-i}_{t,N}),m_{t,N})+g(x^i_{T,N},m_{T,N})\right]. \tag{6.7}$$

Recall that the payoff for strategy τ^i in the one player game $\Gamma_{\mathbf{m}}$ is:

$$J_1(x,\tau^i,\mathbf{m}):=\mathbb{E}^Q_{\tau^i,\sigma^{-i}}\left[\sum_{t\in\mathcal{T}}\ell(x^i_{t,N},\tau^i(x^i_{t,N},m^{-i}_{t,N}),m_t)+g(x^i_{T,N},m_T)\right]. \tag{6.8}$$

Since

$$\sup_{\tau^i \in \Sigma} J_1(x, \tau^i, \mathbf{m}) = J_1(x, \sigma, \mathbf{m}).$$

We have that, for all $\tau^i \in \Sigma$,

$$\begin{aligned} D &= J_N^i(x, m_0, \tau^i, \sigma^{-i}) - J_1(x, \sigma, \mathbf{m}) + J_1(x, \sigma, \mathbf{m}) - J_N^i(x, m_0, \sigma^i, \sigma^{-i}) \\ &\leq \Big[J_N^i(x, m_0, \tau^i, \sigma^{-i}) - J_1(x, \tau^i, \mathbf{m})\Big] + \Big[J_1(x, \sigma, \mathbf{m}) - J_N^i(x, m_0, \sigma^i, \sigma^{-i})\Big] \\ &\leq D_1 + D_2. \end{aligned}$$

Here,

$$D_1 = |J_N^i(x, m_0, \tau^i, \sigma^{-i}) - J_1(x, \tau^i, \mathbf{m})|$$

and

$$D_2 = |J_1(x, \sigma, \mathbf{m}) - J_N^i(x, m_0, \sigma, \sigma^{-i})|$$

The result now follows if we apply Lemma 6.4.6 separately to each term on the sums appearing in D_1 and D_2. □

Chapter 7

Discrete time mean field games: The short-stage limit

Abstract: In this note we provide a model for discrete time mean field games. Our main contributions are an explicit approximation in the discounted case and an approximation result for a mean field game with short-stage duration.

7.1 Introduction

In this paper we study a model for a discrete time, discrete state space, finitely repeated stochastic games where the transition and the payoff of the players depend on the position in space and the actions of the adversaries, but not on their identities. We assume all the players have the same dynamics and the same payoff, thus, for each player, we can consider the influence of the adversaries only through the empirical distribution of the state-action pair.

Mean field games have been introduced independently by Huang, Caines and Malhamé [56] and by Lasry and Lions [64, 65, 66] and have received considerable attention in the literature. The aim of the mean field games paradigm is to describe situations with many interacting agents whose preferences and dynamics depend on the aggregate effect of the other agents. Mean field game models are composed by two parts: a *backward component*, where each agent considers the aggregate behavior as an external parameter and computes myopically his own optimal behavior and a *forward component*, which is the evolution of the initial distribution in the state space under a common strategy. Mean field games have found applications in many different areas, we refer to [49] and the references therein for examples.

Most of the models studied in the literature so far are in continuous time, while the discrete time case has received less attention. The discrete time case has not only independent interest, but also allows to model more general transitions, contrary to the assumption usually made in continuous time mean field games that the *noise* in the dynamics of the players is independent of their actions. In discrete time, we can also allow the players to choose their actions randomly, as in classical game theory. However, for some applications it might be relevant to consider frequent interactions between the players. This motivates the study of a limit model as the duration of each stage tends to zero, which we pursue in Section 7.3.

The main novelty of our work with respect to the previous work on discrete time mean field games is the short-stage version. Short-stage games have been recently introduced in [78]. The aim of this theory is to study games where players are allowed to interact

more frequently. Incorporating this machinery, we obtain a *limit object* that provides an approximate Nash equilibrium for games with sufficiently many players and sufficiently frequent interactions.

In [48], a discrete time, finite state mean field game with a continuum of players is studied. The authors study a finite horizon game and prove the exponential convergence of the finite-horizon mean field equilibrium to a stationary solution. There are two significant differences with our work. First, we consider a *fixed* time horizon. Second, we are interested in constructing approximate equilibria for games with large numbers of players, while in [48] a continuum of players is considered. An anonymous referee pointed us to the recent paper [34] where a model for linear quadratic mean field games in discrete time is studied.

Our work is closer to [2], where a similar notion is studied for an infinite horizon, discounted stochastic game. While we restrict our framework to a finite state space (in [2] an unbounded state space is considered), we provide explicit approximation estimates in terms of the basic parameters of the game. Our estimate is of the same order as the one in [57] in continuous time.

Let us remark also that we study a discrete-time game by itself, and not the discretization of a continuous-time mean field game for numerical solution purposes. Numerical methods have been initially developed in [1]. A semi-Lagrangian scheme has been proposed in [19] for the deterministic, finite-horizon case. The full discretization has been studied in [27].

The paper is organized as follows, In Section 7.2, we describe the model and some results on the existence of mean field equilibrium, as well as the approximation results with explicit convergence rates. In Section 7.3 we introduce a short-stage version of the discounted stochastic mean field game. In the Appendix, we prove an approximation lemma which allows to prove the results we present in Section 7.2.3.

7.2 The discrete time model

7.2.1 Mean field equilibrium

Let Ω and A denote respectively the state and action sets. We assume both to be finite. Let $\mathcal{Z} := \Delta(\Omega \times A)$, where, for a finite set $\mathcal{S}$, $\Delta(\mathcal{S})$ denotes the set of probability distributions over $\mathcal{S}$. Consider a bounded payoff function $g : \Omega \times A \times \mathcal{Z} \to [0,1]$ and a transition function $Q : \Omega \times A \times \mathcal{Z} \to \Delta(\Omega)$. Let n be a fixed positive integer. Let us define a family of auxiliary one-player games, parameterized by a vector $\mathbf{z} = (z_1, z_2, \dots, z_n) \in \mathcal{Z}^n$. The one-player dynamic programming problem $\Gamma^n_{\mathbf{z}}$ is defined as follows: at stage k, the player observes the state $\omega_k \in \Omega$ and chooses the action $a_k \in A$ from which he receives a payoff $g(\omega_k, a_k, z_k)$ and the new state is chosen according to the law $Q(\omega_k, a_k, z_k)$. A **pure behavior strategy** (resp. **mixed behavior strategy**) is a sequence of functions $\sigma = (\sigma_1, \dots, \sigma_n)$ where $\sigma_k : \mathcal{H}^1_k \to A$ (resp. $\sigma_k : \mathcal{H}^1_k \to \Delta(A)$). Here, $\mathcal{H}^1_k = (\Omega \times A)^{k-1} \cup \Omega \times \Delta(\Omega)$ denotes the set of **histories** up to time k, for $k = 1, \dots, n$. Let Σ_n denote the set of pure strategies. The player knows $\mathbf{z}$ and observes the payoff. We introduce the value function $V_n : \Omega \times \mathcal{Z} \to \mathbb{R}_+$ for the game $\Gamma^n_{\mathbf{z}}$:

$$V_n(\omega, \mathbf{z}) := \max_{\sigma \in \Sigma_n} \mathbb{E}^Q \left(\sum_{k=1}^{n} g(\omega_k, a_k, z_k) \mid \omega_1 = \omega \right)$$

One can also consider an infinitely repeated game Γ^λ_z with parameters $z \in \mathcal{Z}$ and $\lambda \in (0,1]$, played as before but where the payoff is evaluated by

$$V_\lambda(\omega, z) := \max_{\sigma\in\Sigma} \mathbb{E}^Q \left(\sum_{k=1}^{\infty} (1-\lambda)^{k-1} g(\omega_k, a_k, z) \mid \omega_1 = \omega \right).$$

From the familiar arguments[1], one can prove that the value functions satisfy the following recursive formulae (dynamic programming principle):

$$V_n(\omega, \mathbf{z}) = \max_{a\in A} \left\{ g(\omega, a, z_1) + \sum_{\omega'\in\Omega} V_{n-1}(\omega', \mathbf{z}^+) Q(\omega, a, z_1)(\omega') \right\} \tag{7.1}$$

and

$$V_\lambda(\omega, z) = \max_{a\in A} \left\{ g(\omega, a, z) + (1-\lambda) \sum_{\omega'\in\Omega} V_\lambda(\omega', z) Q(\omega, a, z)(\omega') \right\}. \tag{7.2}$$

In (7.1), if $\mathbf{z} = (z_1, z_2, \dots z_n)$, then $\mathbf{z}^+$ denotes the vector $(z_2, z_3, \dots z_n)$.

The dynamic programming principle (7.1) also tells us that the player can restrict his attention to the set of **Markovian strategies** $\Sigma_n^M \subset \Sigma_n$, which consists of all the functions $\sigma = (\sigma_1, \dots, \sigma_n)$ such that $\sigma_k : \Omega \to \Delta(A)$.

Let $m_1 \in \Delta(\Omega)$ given and let $\mathcal{Z}_1^n := \{\mathbf{z} \in \mathcal{Z}^n : \ z_1|_\Omega = m_1\}$. For the rest of the paper, $z_k|_\Omega$ denotes the marginal distribution of $z_k \in \mathcal{Z}$ on the set Ω. Define $\Psi_n : \mathcal{Z}_1^n \rightrightarrows \Sigma_n^M$ as the set valued map that associates to every $\mathbf{z} \in \mathcal{Z}_1^n$ the set of optimal Markovian strategies in $\Gamma_{\mathbf{z}}^n$.

Let $\Phi_n : \Sigma_n^M \to \mathcal{Z}_1^n$ defined by $\sigma \mapsto z^\sigma$ where the sequence z^σ is recursively defined by setting $z_1^\sigma(\omega, a) := m_1(\omega) \cdot \sigma_1[\omega](a)$ and

$$z_{k+1}^\sigma(\omega, a) := \sum_{(\omega', a')\in\Omega\times A} Q(\omega', a', z_k^\sigma)(\omega) \cdot z_k^\sigma(\omega', a') \cdot \sigma_{k+1}[\omega](a). \tag{7.3}$$

We are interested in the fixed points of $\Phi_n \circ \Psi_n$. In order to apply fixed point theorems, one needs to ensure certain continuity and convexity properties, which will hold under the following assumptions.

Assumption 7.2.1. *(Lipschitz continuity) There exists positive real numbers L_Q, L_g such that for all $(\omega, a, y, z) \in \Omega \times A \times \mathcal{Z} \times \mathcal{Z}$,*

$$\|Q(\omega, a, y) - Q(\omega, a, z)\|_\infty \le L_Q \|y - z\|_1$$

and

$$\|g(\omega, a, y) - g(\omega, a, z)\|_\infty \le L_g \|y - z\|_\infty .$$

For the existence results in these sections, this continuity assumption can be relaxed, however, for our main approximation results we need Lipschitz continuity. One way to ensure convexity properties is the following:

Assumption 7.2.2. *(Independent transitions) For all $(\omega, a, y, z) \in \Omega \times A \times \mathcal{Z} \times \mathcal{Z}$,*

$$Q(\omega, a, y) = Q(\omega, a, z) =: Q(\omega, a).$$

In order to avoid this assumption, one needs to impose a different assumption so that a convexity property can still be preserved.

1. This is a one-player Markov decision process, which are well understood. See for instance [53].

Assumption 7.2.3. *(Uniqueness of the maximizer) The right hand side of equations* (7.1), (7.2) *admits a unique maximizer, i.e. there exists a unique pure Markovian optimal strategy.*

It is possible to provide conditions on the basic model data that ensure that Assumption 7.2.3 holds, see for example Assumption 2 and 3 in [2] or Assumptions 1-3 in [48]. As uniqueness of the maximizer might hold under other circumstances, we prefer not to write down explicit conditions.

A straightforward application of Brouwer's fixed point theorem yields the following result.

Proposition 7.2.4. *If Q satisfies Assumptions 7.2.1 and 7.2.3, then $\Phi_n \circ \Psi_n$ has a fixed point.*

Definition 7.2.5. A **mean field equilibrium with initial mass** m_1 for the $n-$stage game is a pair $(\sigma, \mathbf{z}) \in \Sigma_n^M \times \mathcal{Z}_1^n$ such that $\mathbf{z}$ is a fixed point of $\Phi_n \circ \Psi_n$ and $\sigma = \Psi_n(\mathbf{z})$.

For the discounted case, from (7.2) one obtains that there exist optimal **stationary strategies**, i.e. functions of the form $\sigma : \Omega \to \Delta(A)$. Let Σ denote the set of stationary strategies. Define $\Psi_\lambda : \mathcal{Z} \to \mathcal{Z} \times \Sigma$ as the set valued map that associates to every $z \in \mathcal{Z}$ the pair (z, S_z^λ), where S_z^λ is the set of optimal stationary strategies in Γ_z^λ.

We will make the following ergodicity assumption throughout the paper.

Assumption 7.2.6. *(Ergodicity) For all $y \in \mathcal{Z}$, $\sigma \in \Sigma$, and $(\omega', a'), (\omega, a) \in \Omega \times A$, the Markov chain $(Z_k)_{k\in\mathbb{N}}$ over $\Omega \times A$ with transition probability*

$$\mathbb{P}\left(Z_{k+1} = (\omega, a) \mid Z_k = (\omega', a')\right) = Q(\omega', a', y)(\omega) \cdot \sigma[\omega](a) =: \mathbf{Q}[y, \sigma]_{(\omega',a'),(\omega,a)} \tag{7.4}$$

has a stationary distribution $z = z[y, \sigma]$. In other words, for y, σ given, there exists z such that the following holds

$$\mathbf{Q}[y, \sigma]z = z. \tag{7.5}$$

Let $\Phi_\lambda : \mathcal{Z} \times \Sigma \to \mathcal{Z}$ be defined by $(y, \sigma) \mapsto z[y, \sigma]$ where $z[y, \sigma] \in \mathcal{Z}$ is defined by

$$z[y, \sigma](\omega, a) := \sum_{(\omega',a')\in\Omega\times A} Q(\omega', a', y)(\omega) \cdot \sigma[\omega](a) \cdot z[y, \sigma](\omega', a').$$

Under Assumption 7.2.6, Φ_λ is well defined. One obtains analogous results to Proposition 7.2.4.

Proposition 7.2.7. *Under Assumptions 7.2.1 and 7.2.6, $\Phi_\lambda \circ \Psi_\lambda$ has a fixed point.*

Proof. The upper semicontinuity follows easily from the assumptions. For the convexity, let $z \in \mathcal{Z}$ and $z_1, z_2 \in \Phi_\lambda \circ \Psi_\lambda(z)$. Let $z_\theta := \theta z_1 + (1 - \theta)z_2$. Consider two stationary strategies $\sigma_1, \sigma_2 \in \Psi_\lambda(z)$ such that $\Phi_\lambda(\sigma_1) = z_1, \Phi_\lambda(\sigma_2) = z_2$. Let σ_θ be the following strategy:

$$\sigma_\theta[\omega](a) := \frac{\theta \cdot z_1(\omega, a) \cdot \sigma_1[\omega](a) + (1 - \theta) \cdot z_2(\omega, a) \cdot \sigma_2[\omega](a)}{z_\theta(\omega, a)}.$$

Observe that σ_θ is optimal for Γ_z^λ from the optimality of σ_1, σ_2. We make the convention that when $z_\theta(\omega, a) = 0$, $\sigma_\theta[\omega](a) = 0$. Note also that

$$z_\theta(\omega, a) = 0 \iff z_1(\omega, a), z_2(\omega, a) = 0.$$

We have, up to excluding the above trivial cases,

$$\begin{aligned} z_\theta(\omega, a) &= \theta z_1(\omega, a) + (1-\theta) z_2(\omega, a) \\ &= \theta \sum_{\omega', a'} Q(\omega', a', z)(\omega) \cdot \sigma_1[\omega](a) \cdot z_1(\omega, a) + \\ &+ (1-\theta) \sum_{\omega', a'} Q(\omega', a', z)(\omega) \cdot \sigma_2[\omega](a) \cdot z_2(\omega, a) \\ &= \sum_{\omega', a'} Q(\omega', a', z)(\omega) \cdot \sigma_\theta[\omega](a) \cdot z_\theta(\omega, a) \end{aligned}$$

Hence, $z_\theta \in \Phi_\lambda \circ \Psi_\lambda(z)$.

□

In the case of pure strategies, one has the following:

Proposition 7.2.8. *Under Assumptions 7.2.1, 7.2.3 and 7.2.6, $\Phi_\lambda \circ \Psi_\lambda$ has a fixed point.*

The proof of this proposition is a straightforward application of Brouwer's fixed point theorem.

Definition 7.2.9. A **stationary mean field equilibrium** is a pair $(\sigma, z) \in \Sigma \times \mathcal{Z}$ such that z is a fixed point of $\Phi_\lambda \circ \Psi_\lambda$ and σ is the strategy associated to $\Psi_\lambda(z)$.

7.2.2 The N-player game

We consider a $n-$stage stochastic game $\Gamma_{n,N}[m_1]$ with $N+1$ identical players, i.e. with common state space Ω, action set A, stage payoff $g : \Omega \times A \times \mathcal{Z} \to [0,1]$ and transition function $Q : \Omega \times A \times \mathcal{Z} \to \Delta(\Omega)$ played as follows: at stage k, for $k = 1, \ldots, n$, each player i observes his own state ω_k^i and the state of each of the adversaries and chooses his action a_k^i. The initial state of each player is sampled i.i.d using the lottery m_1. The actions of the players are chosen simultaneously and independently. After the actions were chosen, each player has a state-action pair $z_{k,N}^i := (\omega_{k,N}^i, a_k^i)$. The payoff for player i is $g(\omega_{k,N}^i, a_k^i, z_{k,N})$ where

$$z_{k,N}(\omega, a) := \frac{1}{N} \sum_{j=1}^{N} 1_{\{z_{k,N}^j = (\omega, a)\}}$$

denotes the empirical distribution of the state-action pairs of the players after the play at stage k. The new state for player i, $\omega_{k+1,N}^i$, is chosen according to the law $Q(\omega_{k,N}^i, a_k^i, z_{k,N})$ and the situation is repeated.

A **behavioral strategy** for player i is a vector $\pi^i = (\pi_k^i)_{k=1}^n$ where $\pi_k^i : \mathcal{H}_k \to \Delta(A)$ and $\mathcal{H}_k = (\Omega \times A \times \mathcal{Z})^{k-1} \cup (\Omega \times \Omega^{N-1})$ is the set of all possible histories up to stage k. Denote by $\Sigma_{n,N}$ the set of behavioral strategies for each player and note that $\Sigma_n^M \subset \Sigma_{n,N}$. A **strategy profile** is a vector $\pi = (\pi^i)_{i=1,\ldots,N}$, where π^i is a behavioral strategy of player i. The average payoff of player i, when using the strategy π^i and when his adversaries use the strategy profile $\pi^{-i} \in (\Sigma_{n,N})^{N-1}$ is

$$J_N^i(\omega, m_1, \pi^i, \pi^{-i}) := \mathbb{E}^Q \left[\sum_{k=1}^{n} g(\omega_{k,N}^i, a_{k,N}^i, z_{k,N}) \mid \omega_{1,N}^i = \omega \right].$$

One can also consider a game $\Gamma_{\lambda,N}[m_1]$ with infinite horizon and payoff

$$J^i_\lambda(\omega, m_1, \pi^i, \pi^{-i}) := \mathbb{E}^Q \left[\sum_{k=1}^{\infty} (1-\lambda)^k g(\omega^i_{k,N}, a^i_{k,N}, z_{k,N}) \mid \omega^i_{1,N} = \omega \right]. \tag{7.6}$$

where $\lambda \in (0, 1]$.

Definition 7.2.10. An $\epsilon-$**Nash equilibrium** for the average payoff, where $\epsilon > 0$, is a strategy profile $(\pi^i)_{i=1,\dots,N}$ such that, for all player i and all behavioral strategy τ^i,

$$J^i_N(\omega, m_1, \tau, \pi^{-i}) - \epsilon \leq J^i_N(\omega, m_1, \pi^i, \pi^{-i}).$$

Analogously, an $\epsilon-$**Nash equilibrium** for the λ- discounted payoff is a strategy profile $(\pi^i)_{i=1,\dots,N}$ such that, for all player i and all behavioral strategy τ^i,

$$J^i_\lambda(\omega, m_1, \tau, \pi^{-i}) - \epsilon \leq J^i_\lambda(\omega, m_1, \pi^i, \pi^{-i}).$$

7.2.3 Approximation results

We are ready to state our first main result in the finite horizon case, which is an easy consequence of the approximation lemma in Section 7.4. This result is an estimate of the maximal deviation between the trajectories followed by a player if the observed aggregate state-action of his adversaries affects his own transition functions and an independent game in which he plays the same action, but the transition function takes as argument the corresponding aggregate state-action of the mean field equilibrium.

Throughout this Section Assumption 7.2.1 holds.

Proposition 7.2.11. *Let $(\sigma, \mathbf{z})$ be a mean field equilibrium. Assume player i's adversaries play the mean field equilibrium strategy, whereas player i's action at time s, a^i_s, is arbitrary.*

- *Let Z^i_{s+1} denote the state-action pair of player i at time $s+1$ when his transitions are influenced by the mean field term z_s , i.e. $Z^i_{s+1} \sim Q(\omega^i_s, a^i_s, z_s)$.*
- *$Z^i_{s+1,N}$ denote the state-action pair of player i at time $s+1$ when the empirical state-action pair of the adversaries influences his transitions, i.e. $Z^i_{s+1,N} \sim Q(\omega^i_s, a^i_s, z_{s,N})$.*

Then we have:

$$\mathbb{E}\left(\max_{s \leq n} \|Z^i_s - Z^i_{s,N}\|_\infty \right) \leq \frac{L_Q n |\Omega \times A| \exp(n(\|Q\|_\infty + L_Q))}{\sqrt{N}}$$

Proof. Let $\mathcal{S} := \Omega \times A$ in Lemma 7.4.3. In this lemma, we take $P(\cdot) := Q(a^i_s, \cdot)$.

□

Once the difference of the trajectories of player i in the N player game and in the one-player game where the state-action term enters as a parameter is bounded, we obtain the following result:

Theorem 7.2.12. *Let $\epsilon > 0$ be given. In a finite horizon, N player game, there exists N_0 such that for $N > N_0$ the mean field equilibrium is an ϵ-equilibrium, where the approximation error ϵ is given by:*

$$\frac{L_Q(L_g + \|g\|_\infty) n |\Omega \times A| \exp(n(\|Q\|_\infty + L_Q))}{\sqrt{N}}.$$

Proof. Consider a game with N players and let us focus on the payoff function of player i. Let $(\sigma, \mathbf{z})$ denote a mean field equilibrium for a given initial distribution m_1 and let $\omega_1 = \omega$. We have that, for all behavior strategy τ:

$$\begin{aligned}
D &:= J_N^i(\omega, m_1, \tau, \sigma^{-i}) - J_N^i(\omega, m_1, \sigma, \sigma^{-i}) \\
&= J_N^i(\omega, m_1, \tau, \sigma^{-i}) - \mathbb{E}^Q \sum_{k=1}^n g(\omega_k, \sigma(\omega_k), z_k) + \\
&+ \mathbb{E}^Q \sum_{k=1}^n g(\omega_k, \sigma(\omega_k), z_k) - J_N^i(\omega, m_1, \sigma, \sigma^{-i}) \\
&\leq \left[J_N^i(\omega, m_1, \tau, \sigma^{-i}) - \mathbb{E}^Q \sum_{k=1}^n g(\omega_k, \tau(\omega_k), z_k) \right] + \\
&+ \left[\mathbb{E}^Q \sum_{k=1}^n g(\omega_k, \sigma(\omega_k), z_k) - J_N^i(\omega, m_1, \sigma, \sigma^{-i}) \right]
\end{aligned}$$

The above inequality comes from the optimality of σ. The result now follows applying Lemma 7.4.4 to each of the terms in brackets. □

Remark. Note that the discounted case can be reduced to the finite horizon case: indeed, it suffices to find $K \in \mathbb{N}$ large enough so that

$$\lambda(1-\lambda)^K \|g\|_\infty < \epsilon/2$$

and consider the N_0 for $\epsilon/2$ in Theorem 7.2.12. However, this may not be appropriate when λ is small because the number of stages will be too large. For small λ, it makes more sense to consider the construction we proposed, as it does not require to consider a large number of stages. This will be the case for instance in Section 7.3.

Remark. Our bound suggests that the number of players should be much larger than the length of the game. This seems intuitive, since one would expect that if there are not enough players and they play for many stages, it could happen that the empirical distribution at early stages of the game is too far from the predicted distribution and this error would be propagated.

Remark. Our result is on the spirit of [57]: construct a *limit object* that induces ϵ-Nash equilibria in games with large players. Our limit object corresponds heuristically to a game with a continuum of players. The complementary approach of studying the limits of a sequence of Nash equilibria of games with finitely many players has been explored in some cases, see for instance [8, 64] but the general case remains open. We will illustrate this remark in Example 7.2.14.

We conclude this Section with two illustrative examples:

Example 7.2.13. *As an application let us consider the following example, adapted from learning by doing [43]. Consider the industry of online hotel booking, where many firms offer accommodation. In this case, the state space Ω is the reputation of the firm, the action set A is the capital to be invested. Assume each firm aims to improve their payoff, which is a function of their reputation and their investments ℓ, by making investments and/or adjusting their offers. The reputation changes according to the transition Q. Note that in this context makes sense to consider independent transitions, since one would expect that each firms' present reputation depends exclusively on their past reputation and their*

investment, as in Assumption 7.2.2. The firms interact with each other through the payoff function g, which represents their market share. For instance, if all the firms have similar reputation, customers might be indifferent and the utilities will be shared evenly, whereas if there are few firms with outstanding reputation, they may have higher revenues.

Example 7.2.14. *Consider a game with N players, where each player chooses whether to drive on the left or right side of the street. Assume that the payoff for driving on the same side as everyone else is* 1 *and zero otherwise. Observe that 'everyone drives left' and 'everyone drives right' are Nash equilibria. However, this game has more equilibria which are sensible to the number of players present in the game, for instance 'everyone drives right if N is even'. In this case it does not make sense to consider limits of Nash equilibria, it is rather desirable to have equilibria that are independent of the number of players.*

7.3 A mean field game with frequent actions

The aim of this section is to study a model for mean field games where players are allowed to play more frequently.

7.3.1 The one-player game

Let $\delta > 0$ and $z \in \mathcal{Z}$. In the spirit of [78], we consider a family of discrete time repeated games parameterized by δ as follows: let $\mu : \Omega \times \Omega \times A \times \mathcal{Z} \to \mathbb{R}^+$ bounded and such that, for all $(\omega, a) \in \Omega \times A$,

$$\mu(\omega, \omega, a, z) = -\sum_{\omega' \neq \omega} \mu(\omega, \omega', a, z)$$

That is, $\mu(\omega, \omega', a, z)$ defines the escape velocity from ω to ω'.

For δ small enough, the function

$$Q_\delta(\omega, a, z)(\omega') := \delta\mu(\omega', \omega, a, z), \text{ for } \omega' \neq \omega, \quad Q_\delta(\omega, a, z)(\omega) = 1 + \delta\mu(\omega, \omega, a, z) \tag{7.7}$$

defines transition probabilities. Introduce the notation $g_\delta := \delta g$ and let $\Gamma_z^{\lambda,\delta}$ denote the one player game with stage payoff g_δ and transition function Q_δ. For a fixed δ, this is exactly a discounted one-player game as introduced in Section 7.2 to define a stationary mean field equilibrium. The stationary mean field equilibrium defined through these games enjoys, for a fixed δ, identical approximation properties as in Section 7.2.3 in terms of the number of players. Our goal in this section is to provide a limit object that provides simultaneously good approximations for a large enough population of players and for a short enough time between plays.

Let $0 < \rho < 1$. Informally, our aim is to approximate a mean field equilibrium for the stochastic game in continuous time with payoff $\int_0^\infty e^{-\rho t} g(\omega_t, a_t, z)dt$ via mean field equilibria of the discrete time games $\Gamma_z^{\lambda,\delta}$. The discount factor needs to be adjusted so that the accumulated payoff at the fraction t of the continuous time game is indeed the limit of the accumulated payoffs during the first $\lfloor \frac{t}{\delta} \rfloor$ stages of the discrete time game. This is achieved by taking the discount factor $\lambda = \rho\delta$. Denote by V_λ^δ the value function of the game $\Gamma_z^{\lambda,\delta}$. Taking $\lambda = \rho\delta$ in (7.2) and dividing by δ yields

$$\rho V_{\rho\delta}^\delta(\omega, z) = \max_{a \in A} \left\{ g(\omega, a, z) + (1 - \rho\delta) \sum_{\omega' \in \Omega} \mu(\omega, \omega', a, z) V_{\rho\delta}^\delta(\omega', z) \right\} \tag{7.8}$$

which suggests that if f is an accumulation point of $\left(V^{\delta}_{\rho\delta}\right)_{\delta>0}$, then it should satisfy

$$\rho f(\omega,z) = \max_{a\in A}\left\{g(\omega,a,z) + \sum_{\omega'\in\Omega}\mu(\omega,\omega',a,z)f(\omega',z)\right\}. \tag{7.9}$$

Let us provide a proof of this result. The proof is inspired from the proof of Theorem 1 in [78].

Proposition 7.3.1. *The equation* (7.9) *has a unique solution, denoted V_ρ. Moreover, $V^{\delta}_{\rho\delta} \to V_\rho$ uniformly as $\delta \to 0$.*

Proof. Let $C(\Omega\times\mathcal{Z})$ denote the set of continuous real-valued functions over $\Omega\times\mathcal{Z}$. Let $T : C(\Omega\times\mathcal{Z}) \to C(\Omega\times\mathcal{Z})$ be the following operator:

$$Tf(\omega,z) = \frac{1}{\|\mu\|_\infty+\rho}\max_{a\in A}\left(g(\omega,a,z) + \sum\nolimits_{\omega'\in\Omega}\mu(\omega,\omega',a,z)f(\omega',z) + \|\mu\|_\infty f(\omega,z)\right).$$

Note that $T(f+c\mathbf{1}) = Tf + c\frac{\|\mu\|_\infty}{\|\mu\|_\infty+\rho}$, and that T is monotone, i.e. $Tf \geq Tg$ whenever $f \geq g$. Here, $\mathbf{1}$ denotes the constant function 1. Consequently, T is a $\frac{\|\mu\|}{\|\mu\|_\infty+\rho}$-contraction and has a unique fixed point V_ρ. Besides, note that $Tv = v$ if and only if v is a solution to the following implicit equation:

$$\rho v(\omega,z) = \max_{a\in A}\left(g(\omega,a,z) + \sum\nolimits_{\omega'\in\Omega}\mu(\omega,\omega',a,z)v(\omega',z)\right) \tag{7.10}$$

Denote with V_ρ the unique solution of (7.10) and let σ^ρ_z be an optimal stationary strategy in (7.10). Consider the stochastic one-player game $\Gamma^{\rho\delta,\delta}_z$ with initial state ω and let $Y_m := \mathbb{E}\left[\delta g(\omega_m,\sigma^\rho_z(\omega_m),z) + (1-\rho\delta)V_\rho(\omega_{m+1},z) \mid \mathcal{H}^1_m\right]$. Then

$$\begin{aligned}
Y_m &= \delta g(\omega_m,\sigma^\rho_z(\omega_m),z) + (1-\rho\delta)\sum_{\omega'\in\Omega}Q_\delta(\omega_m,a,z)(\omega')V_\rho(\omega',z)\\
&= \delta g(\omega_m,\sigma^\rho_z(\omega_m),z) + (1-\rho\delta)\sum_{\omega'\in\Omega}\delta\mu(\omega_m,\omega',a,z)V_\rho(\omega',m_1) + (1-\rho\delta)V_\rho(\omega_m,z)\\
&\geq V_\rho(\omega_m,z) - \rho\delta^2\sum_{\omega'\in\Omega}\mu(\omega,\omega',a,z)V_\rho(\omega',m_1)\\
&\geq V_\rho(\omega_m,z) - 2\delta^2\|\mu\|_\infty\|g\|_\infty.
\end{aligned}$$

The last inequality follows directly from (7.10) and using the fact that $V_\rho < \|\|g\|_\infty$ and $\rho < 1$.

Hence,

$$\begin{aligned}
\mathbb{E}^{Q_\delta}(1-\rho\delta)^m\delta g(\omega_m,\sigma^\rho(\omega_m),z) &\geq \mathbb{E}^{Q_\delta}(1-\rho\delta)^m V_\rho(\omega_m,z) -\\
&- \mathbb{E}^{Q_\delta}(1-\rho\delta)^{m+1}V_\rho(\omega_{m+1},z) - (1-\rho\delta)^m 2\delta^2\|\mu\|\|g\|.
\end{aligned}$$

Summing over m (since we consider the payoff as in (7.6)) we obtain that $V^{\delta}_{\rho\delta}(\omega,m_1) \geq V_\rho(\omega,z) - \frac{2\|\mu\|_\infty\|g\|_\infty}{\rho}\delta$, and σ^ρ_z is asymptotically optimal in $\Gamma^{\rho\delta,\delta}_z$, as $\delta\to 0$.

The proof of the opposite inequality follows from (7.10) and considering any arbitrary strategy instead of an optimal one in the above proof. □

7.3.2 The evolution of the state-action pair

Let $y \in \mathcal{Z}$ and σ_δ an optimal stationary strategy in the game $\Gamma_y^{\rho\delta,\delta}$. Let Q_δ be a transition as defined before and $\mathbf{Q}_\delta$ its associated transition matrix, as defined in (7.4). Denote by z_δ the invariant state-action pair under the transition Q_δ, as introduced in (7.7), and the strategy σ_δ. Hence, by (7.5) in Assumption 7.2.6 it should satisfy:

$$\mathbf{Q}_\delta z_\delta = z_\delta.$$

Since $\mathbf{Q}_\delta = I + \delta L[\sigma_\delta, y]$, where $L[\sigma_\delta, y]_{(\omega,a),(\omega',a')} := \mu(\omega, \omega', \sigma_\delta(\omega), y) \cdot \sigma_\delta[\omega'](a')$,

$$L[\sigma_\delta, y] z_\delta = 0.$$

Heuristically, the limit state-action pair as the stage duration goes to zero, corresponding to the limit strategy σ_y^ρ should solve:

$$L[\sigma_y^\rho, y] z = 0.$$

Assumption 7.3.2. *For every $y \in \mathcal{Z}$ and $\sigma \in \Sigma$, optimal stationary strategy given by* (7.9) *with $z = y$, the equation*

$$L[\sigma, y] z = 0. \tag{7.11}$$

has a solution, denoted $z[\sigma, y]$.

Let $\Psi^\rho : \mathcal{Z} \to \mathcal{Z} \times \Sigma$ be defined by $y \mapsto (y, \sigma_y^\rho)$ and $\Phi^\rho : \mathcal{Z} \times \Sigma \to \mathcal{Z}$ be defined by $(y, \sigma) \mapsto z[\sigma, y]$. Under this assumption, let us introduce the following definition:

Definition 7.3.3. A **limit stationary mean field equlibrium** is a pair $(\sigma, z) \in \Sigma \times \mathcal{Z}$ such that z is a fixed point of $\Phi^\rho \circ \Psi^\rho$ and σ is the strategy associated to $\Psi^\rho(z)$.

One can prove existence of the limit stationary mean field equilibrium under the appropriate uniqueness assumption of the optimal action, analogous to Assumption 7.2.3.

Assumption 7.3.4. *The right hand side of* (7.9) *has a unique maximizer.*

By a straightforward application of Brouwer's fixed point theorem, one has:

Proposition 7.3.5. *Under Assumptions 7.3.2 and 7.3.4, the operator $T^\rho := \Phi^\rho \circ \Psi^\rho$ has a fixed point.*

7.3.3 The approximation of the N player game with frequent actions

Let us state now our second approximation result, namely the approximation of the game with sufficiently short stage and sufficiently many players.

Theorem 7.3.6. *For every $\epsilon > 0$ there exist $\delta_0 > 0$ and $N_0 \in \mathbb{N}$ such that for all $\delta < \delta_0$ and $N > N_0$ the strategy provided by the limit stationary mean field equilibrium is an 2ϵ-Nash equilibrium of the discounted mean field game with discount factor $\lambda = \rho\delta$ and N players.*

Proof. Let $\epsilon > 0$ fixed and consider a limit stationary mean field equilibrium (σ, z). Observe from the proof of Proposition 7.3.1 that one can choose δ_0 small enough so that σ is $\epsilon-$optimal for the one-player discounted game with discount $\rho\delta$, for all $\delta < \delta_0$. Let K_0 such that

$$\rho\delta_0(1 - \rho\delta_0)^{K_0} \|g\|_\infty < \epsilon/2.$$

Finally, let us take the N_0 given by Theorem 7.2.12 for the game of $n = K_0$ stages and error $\epsilon/2$. □

To conclude this Section, let us provide an example of a possible application of our model.

Example 7.3.7. *As an example of application, let us revisit the example of the online booking industry (Example 7.2.13). We consider again the state space as the reputation of the firm but restrict the action set to the offers the firm can post online. By monitoring each other actions, firms can frequently update their offers and promotions (with the help perhaps of automated software) to change their reputation levels.*

7.4 Concluding remarks

An interesting feature of the mean field game models from the point of view of applications is the simplification it entails: on the equilibrium, each player has at his disposal an extremely simple strategy that depends only on his current state and he does not need to keep track of the other players, provided the number of players is large enough. This is because the aggregate state-action of the other players is regarded as a parameter, which deviates from the actual realization of the aggregated state-action with very small probability.

However, this nice feature is also its curse. One problem is that the mean field equilibrium need not be unique. If there is a coordinator of the game that informs the players which mean field equilibrium should be played, there are no problems. In applications, this will typically not be the case. One way around would be to provide the players with an adaptation mechanism. To explain this point, let us revisit the example of the driving game:

Example 7.4.1. *Consider the driving game of Example 7.2.14 with N players. The only equilibria that do not depend on N are everyone on the left and everyone on the right. Consider the following adaptation mechanism: each player chooses left or right with probability $\frac{1}{2}$ on the first stage. On the second stage, observing the realizations of the first stage, each player looks at everyone's choice (and recalls its own) and imitates the choice of the majority. Thus, from stage three, the players will be on an equilibrium path if N is odd. If N is even, there is positive probability that none of the equilibria is reached.*

A proper study of adaptation mechanisms for mean field games in the general case is clearly an interesting direction of future research.

Acknowledgments: I am grateful to Martino Bardi, who motivated this research during a very pleasant stay at the University of Padova. I am also grateful to my PhD advisor, Sylvain Sorin, for his careful reading and to an anonymous referee for useful comments. This work was partially supported by the Commission of the European Communities under the 7th Framework Programme Marie Curie Initial Training Network (FP7-PEOPLE-2010-ITN), project SADCO, contract number 264735.

Appendix: An approximation Lemma

Let $\mathcal{S}$ denote a finite set. We identify the set $\mathcal{S}$ with the canonical basis of $\mathbb{R}^{|\mathcal{S}|}$. Denote by $\mathcal{M}$ the subset of $\mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$ consisting of transition matrices for Markov chains over $\mathcal{S}$.

Let $P : \Delta(\mathcal{S}) \to \mathcal{M}$ denote a Lipschitz continuous function with respect to the L_1 norm with Lipschitz constant L_P. Since $\mathcal{S}$ is finite, we have that the total variation distance

in $\Delta(\mathcal{S})$, defined as $\|\mu-\nu\|_\infty := \max_{A\subset\mathcal{S}} |\mu(A)-\nu(A)|$ is related to the L_1 distance by $\|\mu-\nu\|_\infty = \frac{1}{2}\|\mu-\nu\|_1$, so that a L_P-Lipschitz function in the L_1 norm is $2L_P$ in the total variation norm.

Let $T > 1$ be an integer, representing the number of stages. For $i = 1,\dots,N$ and $k = 0, 1\dots, T-1$ define the following

$$\begin{aligned} X^i_{k+1} &= P(m_k)X^i_k \\ X^i_{k+1,N} &= P(m_{k,N})X^i_{k,N} \end{aligned}$$

where $X^i_0 = X^i_{0,N}$ is a random variable with law m_0 and $X^i_{0,N}$ are sampled i.i.d with probability m_0. Here, m_k denotes the law of X^i_k and $m_{k,N} := \frac{1}{N}\sum_{i=1}^N X^i_{k,N}$.

Observe that

$$\begin{aligned} X^i_{k+1} &= X^i_0 + \sum_{\ell=0}^k X^i_{\ell+1} - X^i_\ell \\ X^i_{k+1,N} &= X^i_{0,N} + \sum_{\ell=0}^k X^i_{\ell+1} - X^i_{\ell,N} \end{aligned}$$

so that

$$X^i_{k+1} - X^i_{k+1,N} = \sum_{\ell=0}^k (P(m_\ell)-I)X^i_\ell + \sum_{\ell=0}^k (P(m_{\ell,N})-I)X^i_{\ell,N}.$$

Before we proceed to the approximation lemma, let us introduce $\xi_{k,N} := \frac{1}{N}\sum_{i=1}^N X^i_k$. Observe that $m_k = \mathbb{E}\xi_{k,N}$. We have the following

Proposition 7.4.2. *The following estimate holds:*

$$\mathbb{E}\|\xi_{k,N} - m_k\|_1 \le |\mathcal{S}|/2\sqrt{N}.$$

Proof. For every $s \in \mathcal{S}$, the random variable $\xi_{k,N}(s)$ is the average of N independent Bernoulli variables. Hence, by definition of the variance and Jensen's inequality,

$$\mathbb{E}|\xi_{k,N}(s) - m_k(s)| \le \sqrt{\mathrm{var}(\xi_{k,N}(s))} \le \frac{1}{2\sqrt{N}}$$

Summing over s gives the result. □

Let $\mathcal{F}_k$ denote the filtration generated by the observed history up to stage k. We are ready to prove the following approximation lemma.

Lemma 7.4.3. *The following estimate holds: for all $i = 1,\dots,N$,*

$$\mathbb{E}\left(\max_{s\le T-1} \|X^i_{s+1} - X^i_{s+1,N}\|_\infty\right) \le \frac{L_P T|\mathcal{S}|\exp(T(\|P\|_\infty + L_P))}{\sqrt{N}}$$

Proof. Let $D^i_k := \mathbb{E}\left(\max_{s\le k}\|X^i_{s+1} - X^i_{s+1,N}\|_\infty \mid \mathcal{F}_k\right)$ and $D_k := \max_{1\le i\le N} D^i_k$. Now observe that, for any i:

$$
\begin{aligned}
D_{k+1}^i &\leq \sum_{\ell=0}^{k} \mathbb{E}\|P(m_\ell)X_\ell^i - P(\xi_{\ell,N})X_\ell^i\|_\infty - && (7.12)\\
&+ \mathbb{E}\|P(\xi_{\ell,N})X_\ell^i - P(m_{\ell,N})X_{\ell,N}^i\|_\infty + \mathbb{E}\|X_\ell^i - X_{\ell,N}^i\|_\infty && (7.13)\\
&\leq \sum_{\ell=0}^{k} L_P \mathbb{E}\|m_\ell - \xi_{\ell,N}\|_1 + (\|P\|_\infty + L_P)D_\ell. && (7.14)
\end{aligned}
$$

From here it follows that

$$
D_{k+1} \leq \frac{L_P T|\mathcal{S}|}{\sqrt{N}} + (\|P\|_\infty + L_P)\sum_{\ell=0}^{k} D_\ell.
$$

The first inequality ((7.12) and (7.13)) follows from the triangle inequality. For (7.14), we use the fact that

$$
\begin{aligned}
\mathbb{E}\left(\|P(m_\ell)X_\ell^i - P(\xi_{\ell,N})X_\ell^i\|_\infty\right) &\leq \mathbb{E}\left(\|P(m_\ell) - P(\xi_{\ell,N})\|_\infty \|X_\ell^i\|_1\right)\\
&\leq L_P \mathbb{E}\|m_\ell - \xi_{\ell,N}\|_1.
\end{aligned}
$$

and that

$$
\begin{aligned}
\mathbb{E}\|P(\xi_{\ell,N})X_\ell^i - P(m_{\ell,N})X_{\ell,N}^i\|_\infty &\leq \mathbb{E}\|P(\xi_{\ell,N})(X_\ell^i - X_{\ell,N}^i)\|_\infty + \mathbb{E}\|P(\xi_{\ell,N})X_{\ell,N}^i - P(m_{\ell,N})X_{\ell,N}^i\|_\infty\\
&\leq \|P\|_\infty \mathbb{E}\|X_\ell^i - X_{\ell,N}^i\|_\infty + L_P D_\ell
\end{aligned}
$$

The conclusion follows from induction. □

Let us prove a useful lemma.

Lemma 7.4.4. *Let $f : \mathcal{S} \times \Delta(\mathcal{S}) \to \mathbb{R}$ be a bounded and L_f-Lipschitz continuous function with the second variable respect to the L_1 norm. Then we have*

$$
\mathbb{E}\left(|f(X_{k+1,N}^i, m_{k+1,N}) - f(X_{k+1}^i, m_{k+1})| \mid \mathcal{F}_k\right) \leq (L_f + \|f\|_\infty)D_{k+1}
$$

Proof. First observe that

$$
\begin{aligned}
\mathbb{E}\left(|f(X_{k+1,N}^i, m_{k+1,N}) - f(X_{k+1,N}^i, m_{k+1})| \mid \mathcal{F}_k\right) &\leq L_f \mathbb{E}\left(\|m_{k+1,N} - m_{k+1}\|_1 \mid \mathcal{F}_k\right)\\
&\leq L_f D_{k+1}
\end{aligned}
$$

We also have that

$$
\mathbb{E}\left(|f(X_{k+1,N}^i, m_{k+1}) - f(X_{k+1}^i, m_{k+1})| \mid \mathcal{F}_k\right) \leq \|f\|_\infty D_{k+1}
$$

Combining these two inequalities yields the result. □

Part III

Stochastic games with frequent actions

Chapter 8

Stochastic zero-sum games with a continuous time dynamics

Abstract: We present in a unified framework some results concerning a family of stochastic games where a payoff-relevant parameter evolves following a continuous time Markov chain. The chain is jointly controlled by two players that choose their actions in discrete time. We characterize the asymptotic value as the time between consecutive actions goes to zero under different assumptions on the information the players receive and on the payoff evaluations.

8.1 Description of the general model

8.1.1 Basic dynamics

Let Ω be a finite set, called the **parameter space** and let A and B denote the finite **action** sets of players 1 and 2 respectively.

The parameter evolves in continuous time, following an homogeneous Markov chain with **transition rate function** $q : \Omega \times \Omega \times A \times B \to \mathbb{R}$, i.e. a function that satisfies, for all $(\omega, a, b) \in \Omega \times A \times B$:

$$0 \leq q(\omega, \omega', a, b) < +\infty, \;\; \omega' \neq \omega, \text{ and } \sum_{\omega' \in \Omega} q(\omega, \omega', a, b) = 0.$$

For fixed $(a, b) \in A \times B$, the transition function corresponds to the speed with which the parameter jumps from ω to ω'. To these actions corresponds a **transition semigroup** $P(\cdot, a, b)$, which is a collection of maps $P_\epsilon(\cdot, a, b) : \Omega \times \Omega \to [0, 1]$ such that

$$\mathbb{P}(\omega_{t+\epsilon} = \omega' | \omega_t = \omega, a, b) = P_\epsilon(\omega, \omega', a, b) + o(\epsilon)$$

for all $t, \epsilon \geq 0$ and $\omega, \omega' \in \Omega$. The map $t \mapsto P_t(\cdot, a, b)$ is the solution of the Chapman-Kolmogorov equation

$$\dot{P}_t = Q^{a,b} P_t, \;\; P_0 = I \tag{8.1}$$

where the matrix $Q^{a,b} := (q(\omega, \omega', a, b))_{\omega,\omega'}$ is the **generator** of the Markov chain with transition semigroup $P(\cdot, a, b)$.

Let us introduce some notation: let $\|q\| := \max_{\omega,a,b} |q(\omega, \omega, a, b)|$ denote the largest transition rate. Denote by $\|\gamma\| := \max_{\omega,a,b} \gamma(\omega, a, b)$ the uniform norm of the payoff function and for a positive constant δ, let $\gamma_\delta := \gamma\delta$.

8.1.2 Information and strategies

We describe now how players influence the dynamics introduced in Section 8.1.1.

Consider a partition $\Pi = \{t_1 = 0, t_2, \dots\}$ of $\mathbb{R}_+$ and set $\pi_k := t_{k+1} - t_k$. On the discrete time game we describe below, we assume that the k-th stage of the game takes place at time t_k and its **duration** is π_k.

Let R, S be two sets, called the **signal sets** and Ω, A, B as in Section 8.1.1. Let $\gamma : \Omega \times A \times B \to [0,1]$ denote the **running payoff**. We consider also a **transition function** $Q : \Omega \times A \times B \to \Delta(R \times S)$.

Let us begin with a general model adapted from Mertens, Sorin and Zamir [75, Section IV].

— An initial parameter ω_1 and signals r_1, s_1 are chosen randomly according to an initial distribution $Q_0 \in \Delta(\Omega \times R \times S)$.

— After learning his signal r_1(resp. s_1), player 1 (resp. 2) chooses an action a_1 (resp. b_1). The **stage payoff** $\gamma_1 := \pi_1 \gamma(\omega_1, a_1, b_1)$ is allocated to player 1. The actions are chosen independently.

— The parameter follows the Markov chain with generator Q^{a_1,b_1} in the time interval $[t_1, t_2]$. Hence, at time $t = t_2$, the parameter ω_2 has law $P_{\pi_1}(\omega_1, \cdot, a_1, b_1)$.

— Similarly, at time t_k, for $k \geq 2$, which corresponds to the $k-th$ stage of the game, the value of the parameter is ω_k. The signals r_k, s_k for players 1 and 2 respectively are chosen according to $Q(\omega_k, a_{k-1}, b_{k-1})$. Each player learns his signal and choose their actions a_k,b_k, and a payoff $\gamma_k := \gamma(\omega_k, a_k, b_k)$ is assigned. The parameter ω_k follows the Markov chain with generator Q^{a_k,b_k} on the time interval $[t_k, t_{k+1}]$. At time $t = t_{k+1}$, the parameter ω_{k+1} has law $P_{\pi_k}(\omega_k, \cdot, a_k, b_k)$ and the situation is repeated.

This description is known by both players, including all the relevant parameters of the game: $Q_0, Q, q, \Pi, \gamma, \Omega, A, B, R, S$.

Let $\mathcal{H}^1_j$ denote the information available to player 1 at stage j, that is, the set of sequences $(r_1, a_1, r_2, a_2, \dots, r_j)$. A similar definition holds for $\mathcal{H}^2_j$. A **behavioural strategy** σ for player 1 is a map from his private history $\mathcal{H}^1 := \cup_{j \geq 1} \mathcal{H}^1_j$ to $\Delta(A)$. A behavioural strategy τ for player 2 is defined similarly. The set of behavioural strategies are denoted by Σ and $\mathcal{T}$ for player 1 and 2, respectively. Let $H^\infty := \mathcal{H}^1 \cup \mathcal{H}^2$.

By Kolmogorov's extension theorem, a couple of behavioural strategies (σ, τ), together with Q_0, q, Q, defines a unique probability distribution $\mathbb{P}_{\sigma\tau}$ over the set of all plays, $(\Omega \times A \times B)^{\mathbb{N}}$, endowed with the cylinder σ-algebra. The corresponding expectation is denoted by $\mathbb{E}_{\sigma\tau}$. Let us assume that the players have **perfect recall**, that is, that players remember the full history of the game. In this case, Kuhn's theorem [62] applies, which ensures that the games played in mixed or behavioural strategies are equivalent. Thus, we can consider without loss of generality that the game is played in mixed strategies. Note that the set of mixed strategies for each player is compact and convex.

This model is inspired from the general model of Mertens, Sorin and Zamir [75, Section IV]. The crucial difference is that here the parameter evolves continuously and the transition probability depends on the duration of the stage. Thus, our results are not directly comparable to those of the classical framework of discrete time repeated games.

Games where a payoff relevant parameter follows a continuous time Markov chain have been introduced by Zachrisson [107] under the name *Markov games*. They have also been studied by Tanaka and Wakuta [100, 99] and Tanaka and Lai [98]. These authors assume first the existence of Markovian strategies, from which they derive an equation for the value function that they use to construct stationary strategies. A generalization for unbounded payoff and transition rates has been analysed by Guo and Hernández-Lerma

[50]. We refer to Hernández-Lerma and Prieto-Rumeau [54] for a recent account of zero-sum Markov games. Note however than in the references mentioned above the players choose their actions in continuous time and do not introduce a notion of stage duration.

The notion of stage duration has been introduced by Nyeman [78]. The model presented here is closer to Cardaliaguet, Rainer, Rosenberg and Vieille [26], although the information structure in our examples is symmetric (both players know the same), while they consider a model with asymmetric information. We will elaborate on this point later in Section 8.5.2.

8.2 Evaluation of the payoff

8.2.1 Model A: Game in [0,1]

Let Θ denote a decreasing probability measure over $\mathbb{N}$ with $\Theta = (\theta_1, \theta_2, \ldots)$ and $\theta_1 < 1/\|q\|$, where $\|q\| := \max_{(\omega,a,b)\in\Omega\times A\times B} |q(\omega,\omega,a,b)|$.

Let $s_k := \sum_{\ell<k} \theta_\ell$ and consider the partition $\Pi_\Theta := \{0 = s_1, s_2, s_3, \ldots\}$ of $[0,1]$. The dynamics of the play at the $k-$th stage corresponds to the play at time s_k as described in Section 8.1.2.

The payoff for a history $h := \{\omega_1, r_1, s_1, a_1, b_1, \omega_2, r_2, s_2, a_2, b_2, \ldots\}$ is

$$\sum_{k=1}^{\infty} \theta_k \gamma_k.$$

Some choices of Θ are the uniform partition $\Theta := (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$. In this case, we speak of a $n-$stage game with **average payoff**

$$J_n(h) := \frac{1}{n} \sum_{k=1}^{n} \gamma_k.$$

One can also consider the $\lambda-$ **discounted evaluation**

$$J_\lambda(h) := \lambda \sum_{k=1}^{\infty} (1-\lambda)^{k-1} \gamma_k$$

where $\lambda \in (0,1]$.

8.2.2 Model B: The stationary game with short stage duration

Informally, the game we study is a discretization of an infinite horizon game with continuous time payoff:

$$\int_0^{+\infty} \rho e^{-\rho s} \gamma_s ds. \tag{8.2}$$

Here, $\rho > 0$ is a positive constant, known to both players, and $\gamma_s := \gamma(\omega_s, a_k, b_k)$, for $s \in [t_k, t_{k+1})$ denotes the **instantaneous payoff**.

A common interpretation of ρ is as the **patience** of the players: the smaller ρ is, the players are more patient. In economic applications, it often represents the interest rate.

Let $\Pi_\delta = \{0, \delta, 2\delta, \ldots\}$ denote a uniform partition of $\mathbb{R}_+$, with $0 < \delta < 1/\|q\|$. Denote by $t_j^\delta := (j-1)\delta$ the instant where the $j-$th play takes place. The game is played as in Section 8.1.2. Here, the parameter δ is the stage duration. Alternatively, $1/\delta$ is the **action frequency**.

We consider an approximation by a Riemann sum of the integral in (8.2), so that we account for the payoff only at the nodes of Π_δ and ignore the payoff between stages, where the parameter may evolve. The payoff corresponding to a history $h := \{\omega_1, r_1, s_1, a_1, b_1, \omega_2, r_2, s_2, a_2, b_2, \ldots\}$ is:

$$J_{\rho,\delta}(h) := \sum_{k=1}^{+\infty} \gamma_{\rho,j,\delta}$$

where

$$\gamma_{\rho,j,\delta} := \rho\delta e^{-\rho j\delta}\gamma_j.$$

We refer to this game as the $\rho-$**discounted game with action frequency** $1/\delta$. Within this framework, it is natural to study the existence of the limit as δ and ρ go to zero and to investigate whether these limits commute.

8.2.3 Model C: The short stage game with arbitrary evaluation

We can extend the previous model to a discretization of an infinite horizon game with continuous time payoff:

$$\int_0^{+\infty} \kappa(s)\gamma_s ds.$$

where $\kappa : \mathbb{R}_+ \to \mathbb{R}_+$ is a strictly decreasing density function. When κ is an exponential density, we recover model B.

As before, we consider a Riemann sum to approximate the integral payoff and ignore the payoff between stages.

Set $\kappa_{j,\delta} := \kappa(t_j^\delta)$. For a history $h := \{\omega_1, a_1, b_1, \omega_2, a_2, b_2, \ldots\}$, the corresponding payoff is

$$J_{\kappa,\delta}(h) := \sum_{j=1}^{+\infty} \gamma_{\kappa,j,\delta}.$$

with

$$\gamma_{\kappa,j,\delta} := \delta\kappa_{j,\delta}\gamma_j.$$

8.2.4 Comparison of the evaluations

Let us compare the models we propose with each other and with the classical framework of discrete time repeated games.

While in both model A and the compact game associated to a discrete time repeated game, as described in Sorin [92, Chapter 1] one studies the sequence of value functions for decreasing evaluations, the crucial difference is in the dynamics: in the classical framework the transition probability between two consecutive stages is independent of the weight of the stage payoff, while in our framework it goes to zero. This helps to avoid the oscillation phenomena that arise in the classical framework (see Ziliotto [108] and Sorin and Vigeral [93]).

We show now the difference between models A and C. Let us assume that $\gamma \equiv 1$. Take any $s \in [0, 1]$ and recall that $s_k = \sum_{\ell<k} \theta_\ell$. Observe that

$$\lim_{\|\Theta\|\to 0} \sum_{k=1}^{\min\{k:s_k\leq s\}} \theta_k \to 0$$

where $\|\Theta\| := \sup_{k\in\mathbb{N}} \theta_k$. Intuitively, what happens in model A is that at each stage in model A, the weight and the duration vanish at the same rate. Thus, the more often the play, the longer it takes to accumulate a prescribed fraction of the payoff.

Now take any $t \in \mathbb{R}_+$. Observe that:

$$\lim_{\delta\to 0} \sum_{j=1}^{\lfloor t/\delta \rfloor} \kappa_{j,\delta} \to \int_0^t \kappa(s)ds.$$

Informally, in model C the vanishing rates of the weight and the stage duration are different: as the stage becomes sufficiently small, the payoff starts to accumulate. This is by construction, since the payoff of model C is a discretization of an underlying continuous time payoff.

While model B is a particular case of model C, it has independent interest for several reasons, which are explained later. Essentially, the advantage is that since the sequence of value functions of the discrete games are stationary, then so is the limit object. As we will see, this allows in the case of perfect information, to obtain asymptotic results in two scales: frequency of the game and patience of the players.

8.2.5 Strategic and compact approach

The approach we follow to establish asymptotic results for the sequences of value functions in the above models is what we call a **strategic approach**. Roughly speaking, it consists of the following:

i) Deduce heuristically a limit object.
ii) Use a limit object to construct ϵ-optimal strategies of the approximating game.

This approach is morally similar to the idea of Isaacs, as described in Chapter 3, or the Krasovskii-Subbotin approach for differential games, discussed in Chapter 4. One deduces a limit PDE for the value function. Once the solution of this PDE is obtained, the players construct ϵ-optimal strategies using its level sets.

Another example of limit object is due to Sorin [90, 91] for a famous stochastic game, called the Big Match and introduced by Blackwell and Ferguson [15]. Sorin deduces a limit game in continuous time. The discretization of optimal strategies in this limit game in continuous time gives ϵ-optimal strategies, where the approximation error decreases as the horizon of the repeated game increases. An interesting feature here is that the limit of optimal strategies in the Big Match in discrete time is not an optimal strategy of the limit game in continuous time. We refer to the book of Sorin [92, Section 5.3] for a more detailed discussion.

Other examples include the results on discrete time mean field games we have established in Chapters 6 and 7, the formulas derived by Laraki for absorbing games [63] and a recent result of Neyman [78, Theorem 1], which we revisit in Section 8.4.2.

An alternative approach is the **compact approach**, which consists, roughly speaking, of the following:

i) Prove that the sequence of value functions has an accumulation point.
ii) Establish variational properties for the set of accumulation points.
iii) Prove uniqueness of the accumulation points.

This approach is morally closer to the variational approach of Cardaliaguet, Laraki and Sorin [23].

An advantage of the strategic approach is that it provides an explicit description of ϵ-optimal strategies, which may not be easy to obtain in the compact approach. A disadvantage of the strategic approach is that the limit object may require strong regularity properties, which are not easy to guarantee in general.

8.3 No information on the state

Let us study first the case when both players observe each others' actions but not the parameter. The information available to each player is symmetric, i.e. $\mathcal{H}_k^1 = \mathcal{H}_k^2 =: \mathcal{H}_k$ for all k. In this case, the state variable is the law of the parameter, which is updated at each stage conditional on the observed actions. This defines a dynamics on the state space $\Delta(\Omega)$ which we describe now.

8.3.1 Dynamics in $\Delta(\Omega)$

Let us consider a partition Π as in Section 8.1.2.

Assume that the initial value of the parameter ω_1 is chosen at time t_1 with an initial probability distribution $\xi =: \xi_1 \in \Delta(\Omega)$. The signals received by the players are $r_1 = s_1 = \{\xi\}$ and $r_k = s_k = \{a_{k-1}, b_{k-1}\}$ for $k \geq 1$. Thus, at stage $k+1$, players know $\{\xi, a_1, b_1, \dots a_{k-1}, b_{k-1}, a_k, b_k\}$. Hence, they are able to calculate the law $\xi_{k+1} \in \Delta(\Omega)$ of ω_{k+1} before choosing their actions a_{k+1}, b_{k+1}. Let us explain briefly how this is done.

Once the actions are chosen at time t_k, consider the generator induced by the actions

$$Q^{a_k,b_k}(\omega,\omega') = q(\omega,\omega',a_k,b_k),\ \omega,\omega' \in \Omega. \tag{8.3}$$

The matrix Q^{a_k,b_k} induces a continuous time Markov chain on Ω and the map $t \mapsto P_{t-t_k}(\cdot, a_k, b_k)$ solves the ODE

$$\dot{P}_t = Q^{a_k,b_k} \cdot P_t,\ \ P_{t_k} = I,\ t_k \leq t < t_{k+1}. \tag{8.4}$$

Hence, we have that $P_{\pi_k} = \exp(\pi_k Q^{a_k,b_k})$ and thus the law at stage $k+1$ is given by

$$\xi_{k+1} = \exp(\pi_k Q^{a_k,b_k}) \cdot \xi_k.$$

When the players observe each others' actions, they know the generator (8.3), thus they can compute, using (8.4), the transition probabilities for the time interval where this generator will act on the Markov chain (the time interval is also known). In fact they can resume these computations in one, as we now show.

Lemma 8.3.1. *Let $\xi_s \in \Delta(\Omega)$ denote the law of the state at time s and $(a,b) \in A \times B$ a fixed pair of actions. Then the map $s \mapsto \xi_s$ is differentiable and satisfies:*

$$\dot{\xi}_s = f(\xi_s, a, b),\ \ \xi_t = \xi. \tag{8.5}$$

where $f(\xi, a, b) := Q^{a,b} \cdot \xi$.

Proof. Let $\xi_{t+\epsilon}$ denote the law at time $t+\epsilon$, for $\epsilon > 0$ and the fixed pair of actions a, b. Since $\xi_{t+\epsilon} = \exp(\epsilon Q^{a,b}) \cdot \xi_t$, we have that

$$\frac{\xi_{t+\epsilon} - \xi_t}{\epsilon} = \frac{(\exp(\epsilon Q^{a,b}) - I)\xi_t}{\epsilon} = Q^{a,b}\xi_t + o(\epsilon).$$

□

8.3.2 The associated differential game and the HJI equation

Let us consider the differential game

$$\dot{\xi}(s) = f(\xi(s), \mathbf{a}(s), \mathbf{b}(s)), \;\; \xi(0) = \xi \tag{8.6}$$

where $\mathbf{a} : [0,1] \to A,\ \mathbf{b} : [0,1] \to B$ are measurable functions.

We are interested in the special case $f(\xi, a, b) := Q^{a,b} \cdot \xi$. In particular, Carathéodory's theorem [31, Chapter 2] applies, which ensures that (8.6) has a unique absolutely continuous solution for any initial condition $(t, \xi) \in [0,1] \times \mathbb{R}^{|\Omega|}$. We denote the evaluation of this solution at time s by $\xi[0, \xi, \mathbf{a}, \mathbf{b}](s)$.

Although this differential game and our model problem have different state space, since their dynamics is defined in all $\mathbb{R}^{|\Omega|}$: for our model problem, the dynamics (8.5) is defined in the whole state space $\mathbb{R}^{|\Omega|}$; it only makes sense in the context on the problem where the initial condition belongs to the simplex $\Delta(\Omega)$, as this dynamics leaves the simplex invariant.

The differential games we associate to our model problems have the same dynamics but different evaluations. The strategies are defined as in the simultaneous Fleming value, introduced in Chapter 3: the players choose their actions simultaneously and independently at the nodes of a given partition Π of $\mathbb{R}_+$, they hold their actions constant on each interval $[t_k, t_{k+1})$ while the state evolves according to (8.6). It follows from the results of Chapter 3 that, for all the evaluations we consider (models A,B,C) the simultaneous Fleming value exists. Moreover, as the mesh of Π goes to zero, the sequence of value functions converges to suitable Hamilton-Jacobi-Isaacs PDE's, depending on the evaluation considered. These PDE's are the limit objects we consider to construct ϵ-optimal strategies.

However, strictly speaking, the games we study are not differential games. While the dynamics is deterministic, as in classical differential games, in our case the dynamics is randomly chosen. In this sense the games we study are closer to discretizations of differential games as in the simultaneous Fleming value [38], already studied in Chapter 3.

Another difference is that the relevant information for the players is the realized actions, not the trajectory of the parameter, while in differential games with complete information the relevant information is the state.

Finally, let us introduce the notation $U := \Delta(A), V := \Delta(B)$.

8.3.3 Model A

Let us begin with the payoff evaluation of model A, as introduced in Section 8.2.1.

We recall some results on discretization of differential games. Later, we use the value function of the corresponding differential game to characterize the limit of the sequence of value functions $V_{S,\Theta}$ as $\|\Theta\| \to 0$.

The limit object

Let us consider the differential game with dynamics (8.6) and payoff

$$\int_0^1 \gamma(\xi[0, \xi, \mathbf{a}, \mathbf{b}](s), \mathbf{a}(s), \mathbf{b}(s)) ds. \tag{8.7}$$

Let us consider the discrete game for the simultaneous Fleming value associated to Θ as follows: we allow the players to choose their actions (possibly randomly) at the nodes

of Θ.[1]

Let $W_{F,\Theta}(s_k,\xi)$ denote the simultaneous Fleming value as defined in Chapter 3 with respect to the partition Π_Θ. For all $\xi \in \Delta(\Omega)$ consider the extension of $W_{F,\Theta}(\cdot,\xi)$ to $[0,1]$ by linear interpolation. From the results of Souganidis [94], which have already been described in Chapter 3, we have:

Proposition 8.3.2. *Let W_F be an accumulation point of the family $\{W_{F,\Theta}\}_\Theta$. Then W_F is the unique viscosity solution of*

$$\frac{\partial w}{\partial t}(t,\xi) + H(\xi, \nabla w(t,\xi)) = 0 \tag{8.8}$$

with terminal condition

$$w(1,\xi) = 0$$

where

$$\begin{aligned} H(\xi,p) &:= \max_{u\in\Delta(A)} \min_{v\in\Delta(B)} \mathbb{E}_{uv}\left\{\left\langle Q^{a,b}\cdot\xi, p\right\rangle + \gamma(\xi,a,b)\right\} \\ &= \min_{v\in\Delta(B)} \max_{u\in\Delta(A)} \mathbb{E}_{uv}\left\{\left\langle Q^{a,b}\cdot\xi, p\right\rangle + \gamma(\xi,a,b)\right\}. \end{aligned}$$

Convergence to the limit object

Let $V_{S,\Theta}(s_k,\xi)$ denote the value of the game starting at $s_k \in \Pi_\Theta$ and where the initial law of the parameter is ξ, that is:

$$\begin{aligned} V_{S,\Theta}(s_k,\xi) &= \sup_{\sigma\in\Sigma} \inf_{\tau\in\mathcal{T}} \mathbb{E}_{\sigma\tau}\left\{\sum_{m=k}^{+\infty} \theta_m\gamma_m\right\} \\ &= \inf_{\tau\in\mathcal{T}} \sup_{\sigma\in\Sigma} \mathbb{E}_{\sigma\tau}\left\{\sum_{m=k}^{+\infty} \theta_m\gamma_m\right\}. \end{aligned}$$

Let us prove the following dynamic programming principle.

Proposition 8.3.3. *We have that, for all $(s_k,\xi) \in \Pi_\Theta \times \Delta(\Omega)$:*

$$V_{S,\Theta}(s_k,\xi) = \max_{u\in U} \min_{v\in V} \mathbb{E}_{u,v}\left\{\theta_k\gamma(\xi,a,b) + V_{S,\Theta}\left(s_{k+1}, \exp\left(\theta_k Q^{a,b}\right)\cdot\xi\right)\right\} \tag{8.9}$$

Proof. Let $w(\xi)$ denote the right hand side of (8.9). We will prove first that $V_{S,\Theta}(s_k,\xi) \geq w(\xi)$. Consider an optimal strategy for player 1 in $w(\xi)$, that is, a strategy u^* such that:

$$w(\xi) = \min_{v\in V} \mathbb{E}_{u^*,v}\left\{\theta_k\gamma(\xi,a,b) + V_{S,\Theta}\left(s_{k+1}, \exp\left(\theta_k Q^{a,b}\right)\cdot\xi\right)\right\}. \tag{8.10}$$

Now let $\sigma^+(a,b)$ denote an optimal strategy in the game starting at time s_{k+1} after actions (a,b) where played at time s_k. Finally, let $\sigma^* = (u^*,\sigma^+)$. Let $\tau \in \mathcal{T}$ denote a strategy of player 2 and write $\tau = (v,\tau^+)$ where $\tau^+ \in \mathcal{T}$ and $v \in V$. Thus,

1. Note that this discretization is not the same as in [39], where alternating moves of the players are considered.

$$\begin{aligned}\mathbb{E}_{\sigma^*\tau}\left\{\sum_{m=k}^{+\infty}\theta_m\gamma_m\right\} &= \mathbb{E}_{u^*,v}\left\{\theta_k\gamma_k+\mathbb{E}_{\sigma^+,\tau^+}\left\{\sum_{m=k+1}^{+\infty}\theta_m\gamma_m\,|\,a,b\right\}\right\}\\ &\geq w(\xi).\end{aligned}$$

To prove that $V_{S,\Theta}(s_k,\xi)\leq w(\xi)$ let σ^* denote an optimal strategy in $V_{S,\Theta}(s_k,\xi)$ and let $\sigma_1\in U$ denote its first component, that is, the strategy at date s_k. Let v^* denote a best reply to σ_1 in (8.9). By playing v^*, player 2 ensures that the payoff for player 1 is less than $w(\xi)$, hence the result follows.

□

From the arguments of Chapter 3, $W_{F,\Theta}$ satisfies the same recurrence formula as $V_{S,\Theta}$. Hence we have the following.

Proposition 8.3.4. *The unique accumulation point of $V_{S,\Theta}$ is the unique viscosity solution of* (8.8).

Let us prove the convergence of $V_{S,\Theta}$ in a different way, assuming that (8.8) has a sufficiently regular solution. This helps us to produce $\epsilon-$optimal strategies for the sequence of discrete games.

Proposition 8.3.5. *Assume that* (8.8) *has a solution V_S such that:*

i) For all $t\in[0,1]$, the function $V_S(t,\cdot)$ is continuously differentiable.

ii) There exists a partition $\Theta_0=\{0=r_1<r_2,\dots,r_N=1\}$ such that, for all $\xi\in\Delta(\Omega)$, $V_S(\cdot,\xi)$ is continuously differentiable in every interval (r_k,r_{k+1}).

Then, the family $\{V_{S,\Theta}\}_\Theta$, converge to V_S as $\|\Theta\|\to 0$, for all Θ refinement of Θ_0.

Proof. Let $\Theta=\{0=s_1<s_2,\dots,s_N=1\}$ be a refinement of Θ_0. Denote with $(\xi_k^*)_k$ the sequence of states in the game (as will be defined below), with $\xi_1^*=\xi$.

Let α^* be the following strategy: at time s_k, observe ξ_k^* and choose u_k^* such that

$$u_k^*\in\operatorname{argmax}_{u\in U}\left\{\min_{v\in V}\langle f(\xi_k^*,u,v),\nabla V_S(s_k,\xi_k^*)\rangle\right\}$$

Let $\vec{v}=(v_1,v_2,\dots,v_n)$ be an arbitrary sequence of actions of player 2 and let $\xi_1^*=\xi,\xi_2^*,\dots\xi_{n+1}^*$ denote the trajectory induced by $\vec{v}$ and α^*.

Let $\Delta V_k:=\frac{V_S(s_{k+1},\xi_{k+1}^*)-V_S(s_k,\xi_k^*)}{\theta_k}$. From the dynamic programming equation (8.9), and doing a first order expansion in Taylor series, we get that:

$$\begin{aligned}\Delta V_k &= \frac{\partial V_S}{\partial s}(s_k,\xi_k^*)+\mathbb{E}_{u_k^*v_k}\left\{\left\langle Q^{a,b}\cdot\xi_k^*,\nabla_\xi V_S(s_k,\xi_k^*)\right\rangle+\gamma(\xi_k^*,a,b)\right\}+o(\theta_k)\\ &\geq \frac{\partial V_S}{\partial s}(s_k,\xi_k^*)+\min_{v\in V}\mathbb{E}_{u_k^*v}\left\{\left\langle Q^{a,b}\cdot\xi_k^*,\nabla_\xi V_S(s_k,\xi_k^*)\right\rangle+\gamma(\xi_k^*,a,b)\right\}+o(\theta_k)\end{aligned}$$

Note that

$$\frac{\partial V_S}{\partial s}(s_k,\xi_k^*)+\min_{v\in V}\mathbb{E}_{u_k^*v}\left\{\left\langle Q^{a,b}\cdot\xi_k^*,\nabla_\xi V_S(s_k,\xi_k^*)\right\rangle+\gamma(\xi_k^*,a,b)\right\}$$

is equal to

$$\frac{\partial V_S}{\partial s}(s_k,\xi_k^*)+\max_{u\in U}\min_{v\in V}\mathbb{E}_{uv}\left\{\left\langle Q^{a,b}\cdot\xi_k^*,\nabla_\xi V_S(s_k,\xi_k^*)\right\rangle+\gamma(\xi_k^*,a,b)\right\}=0.$$

Integrating the above equation from t_1 to 1 we obtain:

$$\sum_{k=1}^{N} \mathbb{E}_{u_k^*, v_k} \theta_k \gamma(\xi_k^*, a, b) \quad \geq \quad V_S(s_1, \xi_1^*) + o(\|\Theta\|).$$

Since $\vec{v}$ is arbitrary, this implies that the payoff of strategy α^* is ϵ optimal, where $\epsilon = \epsilon(\|\Theta\|)$. □

8.3.4 Model B

Let us consider a slightly modified version of model B. Instead of the payoff $\gamma_{\rho,j,\delta}$ defined in Section 8.2.2, we divide by the normalization constant ρ. Denote its value by $V_{S,\rho,\delta}$.

The differential game

To prove the convergence of $V_{S,\rho,\delta}$ as $\delta \to 0$, we proceed as before and study an associated differential game. The dynamics of the auxiliary differential game is given again by (8.6), but the payoff is now

$$\int_0^\infty e^{-\rho s} \gamma(\xi[0, \xi, \mathbf{a}, \mathbf{b}](s), \mathbf{a}(s), \mathbf{b}(s)) ds. \tag{8.11}$$

Differential games with discounted payoffs are treated extensively by Bardi and Capuzzo-Dolcetta [9, Chapter VIII]. Again, we point out that our framework slightly differs from Bardi and Capuzzo-Dolcetta [9, Theorem 3.19, Chapter VIII] in that the dynamics is randomly chosen, albeit deterministic. This difference is, however, conceptual, and becomes irrelevant as the limit equation (8.12) is the same in both cases.

Let $W_{F,\rho,\delta}$ denote the value of the differential game where the players play piecewise constant actions (randomly chosen) in the nodes of Π_δ, as in the simultaneous Fleming value in Chapter 3, with the dynamics and the payoff described by (8.6) and (8.11) respectively. Again, the following theorem follows from Souganidis' results as described in Chapter 3.

Theorem 8.3.6. *The accumulation point of $\{W_{F,\rho,\delta}\}_{\delta>0}$ is the unique viscosity solution of*

$$\rho w(\xi) - H(\xi, \nabla w(\xi)) = 0 \tag{8.12}$$

where

$$\begin{aligned} H(\xi, p) \quad &:= \quad \max_{u \in \Delta(A)} \min_{v \in \Delta(B)} \mathbb{E}_{uv} \left\{ \left\langle Q^{a,b} \cdot \xi, p \right\rangle + \gamma(\xi, a, b) \right\} \\ &= \quad \min_{v \in \Delta(B)} \max_{u \in \Delta(A)} \mathbb{E}_{uv} \left\{ \left\langle Q^{a,b} \cdot \xi, p \right\rangle + \gamma(\xi, a, b) \right\}. \end{aligned}$$

To prove the convergence of $V_{S,\rho,\delta}$ it suffices as before to show that both $V_{S,\rho,\delta}$ and $W_{F,\rho,\delta}$ satisfy the same dynamic programming equation.

This gives us a candidate for a limit object, namely, a sufficiently regular solution of (8.12). We omit the proof for the moment, since in the next Section we show, in a more general case, how to use the limit object to derive ϵ-optimal strategies.

We will derive later in Section 8.4.2 a similar equation to (8.12).

8.3.5 Model C

Let us describe the general case, with κ as in Section 8.2.3. We obtain the game with payoff

$$\sum_{j=1}^{\infty} \gamma_{\kappa,j,\delta}$$

where

$$\gamma_{\kappa,j,\delta} = \delta\kappa(t_j^\delta)\gamma(\omega_j, a_j, b_j).$$

Again it follows from Sion's minmax theorem that this game has a value, denoted by $V_{S,\kappa,\delta}$.

Moreover, from similar arguments as in Chapter 3, the value function satisfies the following dynamic programming principle.

Proposition 8.3.7. *The value function $V_{S,\kappa,\delta}$ satisfies:*

$$V_{S,\kappa,\delta}(t_j^\delta, \xi) = \mathrm{val}_{uv}\mathbb{E}_{uv}\left\{\delta\kappa(t_j^\delta)\gamma(\xi, a, b) + V_{S,\kappa,\delta}\left(t_{j+1}^\delta, \exp(\delta Q^{a,b}\cdot\xi)\right)\right\}. \tag{8.13}$$

Following the ideas of the preceding Sections, one would expect to associate to $V_{S,\kappa,\delta}$ the corresponding simultaneous Fleming value, which in this case would be the simultaneous Fleming value of a game with dynamics (8.6) and payoff:

$$\int_0^{+\infty} \kappa(s)\gamma(\xi[0,\xi,\mathbf{a},\mathbf{b}](s), \mathbf{a}(s), \mathbf{b}(s))ds.$$

The reason we do not proceed in the same way relies on the limit equation. Assuming that $V_{S,\kappa,\delta}$ is regular enough and performing a first order Taylor expansion in (8.13), we deduce heuristically the following limit equation, for $w : \mathbb{R}_+ \times \xi \to \mathbb{R}_+$:

$$\frac{\partial w}{\partial t} + \mathrm{val}_{uv}\mathbb{E}_{uv}\left\{\kappa(t)\gamma(\xi, a, b) + \left\langle \nabla_\xi w, Q^{a,b}\cdot\xi\right\rangle\right\} = 0 \tag{8.14}$$

The problem is that Souganidis' approach relies on the uniqueness of viscosity solutions of the limit PDE. For the moment we do not know under which conditions would this hold.

We provide an alternative proof of convergence of the family $V_{S,\kappa,\delta}$, assuming (8.14) satisfies suitable regularity assumptions.

Theorem 8.3.8. *Assume that* (8.14) *has a C^2 solution $V_{S,\kappa}$ such that, for all $(t,\xi) \in [0,1]\times\Delta(\Omega)$:*

$$\|D^2 V_{S,\kappa}(t,\xi)\| \le \kappa(t). \tag{8.15}$$

Then $\lim_{\delta\to 0} V_{S,\kappa,\delta} = V_{S,\kappa}$.

Proof. For a couple $(t,\xi) \in \mathbb{R}_+ \times \Delta(\Omega)$ denote by x_t an optimal stationary strategy for player 1 in the local game with payoff:

$$\kappa(t)\gamma(\xi, a, b) + \left\langle \nabla_\xi V_{S,\kappa}(t,\xi), Q^{a,b}\cdot\xi\right\rangle.$$

Denote for simplicity $x_j := x_{t_j^\delta}$ and $\kappa_j := \kappa(t_j^\delta)$. We will show that the strategy $x := (x_j)_{j\in\mathbb{N}}$ is ϵ-optimal. For this, let $\tau := (\tau_1, \tau_2, \ldots)$ be an arbitrary strategy of player 2. Let us introduce:

$$Y_j := \mathbb{E}_{x_j,\tau_j}\left\{\delta\kappa_j\gamma_j + V_{S,\kappa}\left(t^\delta_{j+1}, \exp(\delta Q^{a,b})\cdot\xi\right) \mid \mathcal{H}_j\right\}.$$

First, observe that $V_{S,\kappa}\left(t^\delta_{j+1}, \exp(\delta Q^{a,b})\cdot\xi\right)$ is, by Taylor expansion and (8.15), greater or equal than:

$$V_{S,\kappa}(t^\delta_j,\xi) + \delta\frac{\partial V_{S,\kappa}}{\partial t}(t^\delta_j,\xi) + \delta\left\langle\nabla_\xi V_{S,\kappa}(t^\delta_j,\xi), Q^{a,b}\cdot\xi\right\rangle + \kappa_j\delta^2/2$$

Using the optimality of x_j gives:

$$Y_j \geq V_{S,\kappa}(t^\delta_j,\xi) - \kappa_j\delta^2/2$$

Finally, let $P_j := \mathbb{E}_{x_j,\tau_j}\{\delta\kappa_j\gamma_j \mid \mathcal{H}_j\}$ denote the conditional expectation of the stage payoff. We have that

$$P_j \geq V_{S,\kappa}(t^\delta_j,\xi) - \mathbb{E}_{x_j,\tau_j}\left\{V_{S,\kappa}\left(t^\delta_{j+1}, \exp(\delta Q^{a,b})\cdot\xi\right) \mid \mathcal{H}_j\right\} - \kappa_j\delta^2/2$$

Summing over j, we obtain on the left hand side the payoff of strategy x, while on the right hand side, after cancellations, and since $\sum_j \kappa_j\delta = 1$, we obtain $\delta/2$ as error term.

□

If we do the change of variable $w(t,\xi) \mapsto e^{-\rho t}V(\xi)$ in (8.14) we recover an equation of the form (8.12).

Let us point out here an interesting feature of the above models: in the absence of information about the state, we can represent a dynamic, random, discrete game with an unknown payoff-relevant parameter into a dynamic game in continuous time where the dynamics is deterministic but randomly chosen.

8.4 Standard signalling

Let us assume in this Section that players have **standard signalling**, that is, they observe the state and each others' actions. We use here the notation from Section 8.2.

Recall that in Section 8.3, the players had no information of the parameter, but they constructed an auxiliary state function using the observed actions. For the results stated here, the parameter is also available, so that the parameter itself becomes the natural state variable.

Stochastic games with standard signalling have been introduced by Shapley [87]. The difference with our model is that in our model the dynamics of the state is in continuous time, and that the transition probability depends on the duration of the stage.

The model we describe in this Section is inspired from the model of Neyman [78]. The difference is conceptual: while Neyman defines a notion of convergence for a family of games, we study instead a Fleming discretization of an underlying continuous time process. The difference is irrelevant in practice and we recover similar results as those obtained by Neyman, with similar techniques.

8.4.1 Model A

Let us introduce the value function:

$$V_{C,\Theta}(s_k,\omega) := \sup_{\sigma\in\Sigma}\inf_{\tau\in\mathcal{T}} \mathbb{E}^q_{\sigma\tau}\left\{\sum_{m=k}^{\infty}\theta_m\gamma_m\right\} = \inf_{\tau\in\mathcal{T}}\sup_{\sigma\in\Sigma} \mathbb{E}^q_{\sigma\tau}\left\{\sum_{m=k}^{\infty}\theta_m\gamma_m\right\}.$$

We consider its extension by linear interpolation to $[0,1]\times\Omega$, for which we keep the same notation.

The following dynamic programming principle holds (see Sorin [92, Chapter 5]):

$$V_{C,\Theta}(s_k,\omega) = \mathrm{val}_{u,v}\mathbb{E}^q_{u,v}\left\{\theta_k\gamma(\omega,a,b) + \sum_{\omega'\in\Omega} V_{C,\Theta}(s_{k+1},\omega')\exp(\theta_k Q^{a,b})(\omega,\omega')\right\}.$$

Heuristically, if we do a first order expansion of the exponential term above and divide by θ_k we obtain an equation of the form:

$$\frac{dz}{ds} + \mathrm{val}_{u,v}\mathbb{E}^q_{u,v}\left\{\gamma(\omega,a,b) + \sum_{\omega'\in\Omega} z(s,\omega')q(\omega,\omega',a,b)\right\} = 0. \tag{8.16}$$

We are ready to establish our first asymptotic result in this Section.

Note here that the derivative of the limit object depends only on the first variable.

Theorem 8.4.1. *If* (8.16) *has a unique* C^1 *solution, denoted* V_C*, then* $\lim_{|\Theta|\to 0} V_{C,\Theta} = V_C$.

Proof. Let $x_k(\omega)$ denote an optimal strategy in the game with stage payoff

$$\gamma(\omega,a,b) + \sum_{\omega'\in\Omega} V(s_k,\omega')q(\omega,\omega',a,b).$$

Denote by $x_k = (x_k(\omega))_{\omega\in\Omega}$. We will prove that the strategy $x := (x_k)_{k\in\mathbb{N}}$ is ϵ-optimal. Let $\tau = (\tau_1,\tau_2,\ldots)$ denote an arbitrary strategy of player 2.

Introduce

$$Y_k := \mathbb{E}^q_{x_k,\tau_k}\left\{\theta_k\gamma(\omega,a,b) + \sum_{\omega'\in\Omega} V_C(s_{k+1},\omega')\exp(\theta_k Q^{a,b})(\omega,\omega') \mid \mathcal{H}_k\right\}.$$

First observe that

$$A := \sum_{\omega'\in\Omega} V_C(s_{k+1},\omega')\exp(\theta_k Q^{a,b})(\omega,\omega') = \sum_{\omega'\in\Omega} V_C(s_{k+1},\omega')\theta_k q(\omega,\omega',a,b) + V_C(s_{k+1},\omega)$$

Doing a first order approximation gives

$$V_C(s_{k+1},\omega') = V_C(s_k,\omega') + \theta_k\frac{dV_C}{ds}(s_k,\omega') + O(\theta_k^2)$$

Altogether we get that A equals:

$$\sum_{\omega'\in\Omega}\theta_k V_C(s_k,\omega')q(\omega,\omega',a,b) + V_C(s_k,\omega) + \theta_k\frac{dV_C}{ds}(s_k,\omega) + \sum_{\omega'\in\Omega}\theta_k^2\frac{dV_C}{ds}(s_k,\omega')q(\omega,\omega',a,b) + O(\theta_k^2)$$

Since the evaluation is strictly decreasing, $\frac{dV_C}{ds}$ is negative.

Taking the conditional expectation and using the optimality of x_k gives:

$$Y_k \geq V_C(s_k,\omega) - 2\theta_k^2\|\gamma\|\cdot\|q\|$$

We conclude as in the proof of Theorem 8.3.8.

□

The equation (8.16) may not have a global solution, but we can apply this result along the refinements of a partition where the solution exists locally, as in Proposition 8.3.5.

8.4.2 Model B

When a discounted payoff is considered (model B), the value function of the finite games is stationary, hence in the limit equation no longer appears a derivative, as in Theorem 8.4.1. This allows us to use the limit object to construct strategies without any regularity assumptions.

Moreover, in the finite case (states and actions) we are able to establish asymptotic properties in two time scales: patience of the players and duration of the stage.

For this, we follow Neyman [78] to recover a characterization similar to 8.3.2, in terms of the limit equation:

$$\rho f(\omega) = \mathrm{val}_{u,v}\mathbb{E}_{uv}\left\{\gamma(\omega,a,b) + \sum\nolimits_{\omega'\in\Omega} q(\omega,\omega',a,b)f(\omega')\right\}. \tag{8.17}$$

Proposition 8.4.2. *(Remark 8 in [78]). The equation* (8.17) *has a unique solution, denoted* $V_{C,\rho}$.

Proof. Observe that any solution of (8.17) is a fixed point of

$$\Psi f(\omega) := \frac{1}{\|q\|+\rho}\mathrm{val}_{u,v}\mathbb{E}_{uv}\left\{\gamma(\omega,a,b) + \sum\nolimits_{\omega'\in\Omega} q(\omega,\omega',a,b)f(\omega') + \|q\|f(\omega)\right\}.$$

Note that $\Psi(f+c\mathbf{1}) = \Psi f + c\frac{\|q\|}{\|q\|+\rho}$, and that Ψ is monotonic. Consequently, Ψ is a $\frac{\|q\|}{\|q\|+\rho}$-contraction and has a unique fixed point $V_{C,\rho}$. □

Let $V_{C,\rho,\delta}$ denote the value function of the game with **unnormalized** payoff

$$\sum_{m\geq 0} e^{-\rho\delta}\delta\gamma_m.$$

This payoff is slightly different from (8.2). For simplicity, we drop off the normalization constant ρ and consider a Riemann sum instead of the integral. Denote this game by $\Gamma^{\delta}_{\rho\delta}$.

The proof of the following result is directly adapted from Neyman[78, Theorem 1]. The difference between our approach and his is conceptual: Neyman studies conditions for a family of discrete time stochastic games to converge, while we study the discretization of an underlying stochastic game. However, the methods are very similar.

Theorem 8.4.3. *(Theorem 1 in [78]). The sequence* $(V_{C,\rho,\delta})_{\delta>0}$ *converges to* $V_{C,\rho}$.

Proof. Let $x(\rho) \in \Delta(A)^{\Omega}$ be an optimal stationary strategy in (8.17). Consider the stochastic game $\Gamma^{\delta}_{\rho\delta}$ with initial state ω. Let us prove that, for any strategy $\tau = (\tau_1,\tau_2,\ldots)$ of player 2, player 1 guarantees V_ρ up to a vanishing error.

Let $Y_m := \mathbb{E}\left\{\gamma_\delta(\omega_m, x(\rho)_m, \tau_m) + e^{-\rho\delta}V_{C,\rho}(\omega_{m+1}) \mid h_m\right\}$. Then

$$\begin{aligned}
Y_m &= \gamma_\delta(\omega_m, x(\rho)_m, \tau_m) + e^{-\rho\delta}\sum_{\omega'\in\Omega} Q_\delta(\omega',\omega_m,x(\rho)_m,\tau_m)V_{C,\rho}(\omega') \\
&= \gamma_\delta(\omega_m, x(\rho)_m, \tau_m) + e^{-\rho\delta}\sum_{\omega'\in\Omega} \delta q(\omega',\omega_m,x(\rho)_m,\tau_m)V_{\rho}(\omega') + e^{-\rho\delta}V_{C,\rho}(\omega_m) \\
&\geq V_{C,\rho}(\omega_m) - \rho\delta^2\sum_{\omega'\in\Omega} q(\omega',\omega_m,x(\rho)_m,\tau_m)V_{C,\rho}(\omega') \\
&\geq V_{C,\rho}(\omega_m) - 2\delta^2\|q\|\|\gamma\|.
\end{aligned}$$

Hence,

$$\begin{aligned}\mathbb{E}e^{-\rho\delta m}\gamma_\delta(\omega_m, x(\rho)_m, \tau_m) &\geq \mathbb{E}e^{-\rho\delta m}V_{C,\rho}(\omega_m) - \\ &- \mathbb{E}e^{-\rho\delta(m+1)}V_{C,\rho}(\omega_{m+1}) \\ &- 2e^{-\rho\delta m}\delta^2\|q\|\|\gamma\|.\end{aligned}$$

The result now follows summing over m. Thus, $V_{C,\rho,\delta}(\omega) \geq V_{C,\rho}(\omega) - \frac{2\|q\|\|\gamma\|}{\rho}\delta$, and $x(\rho)$ is asymptotically optimal in the stochastic game with discount λ_δ, payoff γ_δ and transition Q_δ, as $\delta \to 0$. A dual result holds for player 2, so that

$$\|V_{C,\rho,\delta} - V_{C,\rho}\| \leq \frac{2\|q\|\|\gamma\|}{\rho}\delta. \tag{8.18}$$

□

Note here that, thanks to the stationarity, we do not require any regularity on the limit object, unlike Theorem 8.4.1, but the crucial idea is the same: derive approximately optimal strategies from the limit object to obtain convergence of the sequence of value functions.

Asymptotic results in two scales

As noted before, the fact that we have two scales in our model gives us two kind of asymptotic results: with respect to the duration of the stage and with respect to the discount factor.

The asymptotic results in the stationary case with perfect information can be summarized by the following diagram:

$$\begin{array}{ccc} V_{C,\rho,\delta} & \xrightarrow{\lambda=\rho\delta,\ \delta\to 0} & V_{C,\rho} \\ {\scriptstyle \lambda=\rho\delta\to 0}\downarrow & & \downarrow{\scriptstyle \rho\to 0} \\ v_C^\delta & \xrightarrow{\delta\to 0} & v_C \end{array}$$

We have already defined $V_{C,\rho,\delta}$ and $V_{C,\rho}$. The quantity v_C^δ corresponds to the asymptotic value of a classical, finite stochastic game in the sense of Shapley [87]. The existence of the asymptotic value is ensured by Bewley and Kohlberg [13].

Proposition 8.4.4. *(Neyman [78, Theorem 2] The sequence $\rho V_{C,\rho}$ converges uniformly, as $\rho \to 0$.*

Proof. Let $x(\rho) \in \Delta(A)^\Omega$ and $y(\rho) \in \Delta(B)^\Omega$ be optimal in (8.17), which defines $V_{C,\rho}$. Then, $(\rho, x(\rho), y(\rho), V_{C,\rho})$ is a semialgebraic set, because it is defined by finitely many polynomial equations. By Tarski-Seidenberg elimination theorem, (Benedetti and Risler,[11, Theorem 2.21, p.54]) $\rho \mapsto V_\rho$ is a semialgebraic function. Clearly, $\rho V_{C,\rho,\delta} \leq \|\gamma\|$, for all δ and ρ. Consequently, $\rho \mapsto \rho V_{C,\rho}$ is a bounded, semialgebraic function. The existence of $v := \lim_{\rho\to 0} \rho V_{C,\rho}$ follows. □

Finally, it follows from (8.18) that:

Corollary 8.4.5. *The sequence $\rho V_{C,\rho,\delta}$ converges, as $\delta \to 0$, uniformly in ρ.*

Note that we used here the fact that Ω, A, B are finite to obtain convergence. It remains an open question whether similar "diagrams", i.e. study of asymptotics in duration and patience, can be established in the other models discussed here, although this seems unlikely since in the other models the state space is no longer finite .

8.4.3 Model C

In this case, the value function satisfies the following dynamic programming principle.

Proposition 8.4.6. *The value function $V_{C,\kappa,\delta}$ satisfies:*

$$V_{C,\kappa,\delta}(t_j^\delta,\omega) = \mathrm{val}_{uv}\mathbb{E}_{uv}\left\{\delta\kappa_j\gamma_j + \sum_{\omega'\in\Omega} V_{C,\kappa,\delta}(t_{j+1}^\delta,\omega')\cdot\exp(\delta Q^{a,b})(\omega,\omega')\right\}.$$

We can deduce heuristically, by Taylor expansion with respect to time, that the limit equation should be a solution of

$$\frac{dw}{dt}(t,\omega) = \mathrm{val}_{uv}\mathbb{E}_{uv}\left\{\kappa(t)\gamma(\omega,a,b) + \sum_{\omega'\in\Omega} w(t,\omega')q(\omega,\omega',a,b)\right\}. \tag{8.19}$$

The proof of the following result follows in a similar way as the proof of Theorem 8.4.1.

Theorem 8.4.7. *Assume* (8.19) *has a C^1 solution. Then* $\lim_{\delta\to 0} V_{C,\kappa,\delta} = V_{C,\kappa}$.

Proof. Let x_t denote an optimal strategy for player 1 in the local game

$$\kappa(t)\gamma(\omega,a,b) + \sum_{\omega'\in\Omega} V_{C,\kappa}(t,\omega')q(\omega,\omega',a,b)$$

Denote by $x_j := x_{\rho(t_j^\delta)}$ and $\kappa_j := \kappa(t_j^\delta)$ for simplicity. We will prove that the strategy $x := (x_j)_{j\in\mathbb{N}}$ is ϵ-optimal.

To see this, let $\tau := (\tau_1,\tau_2,\ldots)$ denote an arbitrary strategy of player 2. Let us introduce:

$$Y_j := \mathbb{E}_{x_j,\tau_j}\left\{\delta\kappa_j\gamma(\omega_j,a,b) + \sum_{\omega'\in\Omega} V_{C,\kappa}(t_{j+1}^\delta,\omega')\cdot\exp(\delta Q^{a,b})(\omega_j,\omega') \mid \mathcal{H}_j\right\}.$$

Now observe that

$$V_{C,\kappa}(t_{j+1}^\delta,\omega') = V_{C,\kappa}(t_j^\delta,\omega') + \delta\frac{dV_{C,\kappa}}{dt}(t',\omega')$$

for some $t'\in[t_j^\delta,t_{j+1}^\delta]$. From (8.19) we deduce

$$\frac{dV_{C,\kappa}}{dt}(t',\omega') \le \|\gamma\|\cdot\kappa(t')\cdot(1+\|q\|)$$

We also have

$$\exp(\delta Q^{a,b}) = I + \delta Q^{a,b} + e_1(\delta)$$

where $e_1(\delta)\le\|q\|\delta^2/2$.

We get that $\sum_{\omega'\in\Omega} V_{C,\kappa}(t_{j+1}^\delta,\omega')\cdot\exp(\delta Q^{a,b})(\omega_j,\omega')$ equals

$$\begin{aligned}\textstyle\sum_{\omega'\in\Omega} &\left\{\delta V_{C,\kappa}(t_j^\delta,\omega')q(\omega_j,\omega',a,b) + \delta^2\frac{dV_{C,\kappa}}{dt}(t',\omega')q(\omega_j,\omega',a,b) + V_{C,\kappa}(t_j^\delta,\omega')e_1(\delta)\right\} + \\ + &\quad V_{C,\kappa}(t_j^\delta,\omega_j) + \delta\frac{dV_{C,\kappa}}{dt}(t_j^\delta,\omega_j) + o(\delta^2)\end{aligned}$$

From the definition of x_j and using the upper bounds we obtain that there exists a constant C, independent of δ, such that:

$$\mathbb{E}_{x_j,\tau_j}\left\{\kappa_j\delta\gamma(\omega_j,a,b)\mid\mathcal{H}_j\right\}\geq V_{C,\kappa}(t_j^\delta,\omega_j)-C\cdot\delta^2\kappa_j.$$

The result now follows after summing over j, as in Theorem 8.3.8. □

8.5 Some concluding remarks

Let us conclude this Chapter by making some remarks on two related models and possible extensions. We describe first the game studied by Cardaliaguet and Quincampoix [25], whose information structure is similar as the model in Section 8.3. Later, we make some remarks on the asymmetric information case and the possibility of extending the variational approach of Cardaliaguet, Laraki and Sorin [23] to that framework.

8.5.1 A differential game with blind players

Let us begin with the game studied by Cardaliaguet and Quincampoix in [25].

Let $(t_1,x_1)\in[0,1]\times\mathbb{R}^n$. For $(\mathbf{u},\mathbf{v})\in\mathcal{U}(t_1)\times\mathcal{V}(t_1)$ fixed, consider the differential equation

$$\mathbf{x}(t_1)=x_1,\quad\dot{\mathbf{x}}(t)=f(\mathbf{x}(t),\mathbf{u}(t),\mathbf{v}(t))\quad\text{a.e. on }[t_1,1].\tag{8.20}$$

Assume that the function f is jointly continuous and bounded and that there exists $c>0$ such that for all $(u,v)\in U\times V$ and $x,y\in\mathbb{R}^n$:

$$\|f(x,u,v)-f(y,u,v)\|\leq c\|x-y\|.$$

In this case, it follows from Carathéodory's theorem, [31, Chapter 2] that the above equation has a unique solution, which we denote by $\mathbf{x}[t_1,x_1,\mathbf{u},\mathbf{v}]$, in the following extended sense: for any $t\in[t_1,1]$,

$$\mathbf{x}[t_1,x_1,\mathbf{u},\mathbf{v}](t):=x_1+\int_{t_1}^t f(\mathbf{x}[t_1,x_1,\mathbf{u},\mathbf{v}](s),\mathbf{u}(s),\mathbf{v}(s))ds.$$

In this model, $\mathbf{x}[t_1,x_1,\mathbf{u},\mathbf{v}](t)$ plays the role of the randomly evolving parameter in the basic model in Section 8.1.1.

Finally, consider a Lipschitz continuous terminal payoff $g:\mathbb{R}^n\to[0,1]$. The goal of the game is for player 1 to maximize the quantity $g(\mathbf{x}[t_1,x_1,\mathbf{u},\mathbf{v}](1))$, which player 2 wants to minimize.

Whenever x_1 is known by both players, this game is well understood, see for instance Evans and Souganidis [36]. The problem studied by Cardaliaguet and Quincampoix [25] is the existence of the value when x_1 is not known, but chosen randomly according to a commonly known probability distribution $\mu_1\in\mathcal{M}(\mathbb{R}^n)$, where $\mathcal{M}(\mathbb{R}^n)$ denotes the set of Borel probability measures with finite second moment. The players observe each others' actions, but receive no further information on the parameter.

Note that in this model the parameter follows a deterministic trajectory and evolves on a continuous state space, while in Section 8.3 the parameter evolves randomly and takes only finitely many values. Another important difference is that in Section 8.3 the interaction happens at discrete times, so that their actions are piecewise constant, while in the present model the players interact continuously. Note also that here the state space is no longer compact, which leads to important technical difficulties.

For a pair of NAD strategies $(\alpha, \beta) \in \mathcal{A}_d \times \mathcal{B}_d$ and $t, s \in [0,1], x \in \mathbb{R}^n$ with $t < s$, we denote by $\mathbf{x}[t, x, \alpha, \beta](\cdot)$ the trajectory $\mathbf{x}[t, x, \mathbf{u}, \mathbf{v}](\cdot)$ associated to the corresponding couple of controls $(\mathbf{u}, \mathbf{v})$ given by Lemma 2.4 in [17] and was recalled in the Introduction of Chapter 3

As in Section 8.3, the natural state space to consider is the state of probability measures over the set of values of the parameter.

Let us introduce the value functions of this game:

Definition 8.5.1. For $(t, \mu) \in [0,1] \times \mathcal{M}(\mathbb{R}^n)$:

i) The **maxmin** is defined by

$$\mathbf{V}^-(t, \mu) := \sup_{\alpha \in \mathcal{A}_d} \inf_{\beta \in \mathcal{B}_d} \int_{\mathbb{R}^n} g(\mathbf{x}[t, x, \alpha, \beta](1)) d\mu(x).$$

ii) The **minmax** is defined by

$$\mathbf{V}^+(t, \mu) := \inf_{\beta \in \mathcal{B}_d} \sup_{\alpha \in \mathcal{A}_d} \int_{\mathbb{R}^n} g(\mathbf{x}[t, x, \alpha, \beta](1)) d\mu(x).$$

While the auxiliary state space in Section 8.3 is a finite dimensional polyhedron, the state space in the model described here is not even finite dimensional nor normed. It has a suitable metric, which we define below.

Definition 8.5.2. Let $\mu, \nu \in \mathcal{M}$. The **Wasserstein distance** $\mathbf{d}(\mu, \nu)$ is defined by:

$$\mathbf{d}(\mu, \nu) := \inf_{\eta} \left\{ \left(\int_{\mathbb{R}^{2n}} |x - y|^2 d\eta(x, y) \right)^{\frac{1}{2}} \right\}$$

where the infimum is taken over the set of all probability measures in $\mathbb{R}^{2n}$ that satisfy:

$$\pi_1 \# \eta = \mu \text{ and } \pi_2 \# \eta = \nu$$

where π_1, π_2 are the projections over the first and second coordinate, respectively, and $\pi_1 \# \eta$ denotes the **push-forward** of η by π_1, i.e., the measure in $\mathbb{R}^n$ that satisfies

$$\pi_1 \# \eta(A) = \eta(\pi_1^{-1}(A)), \ \ \forall A \subset \mathbb{R}^n \text{ measurable}$$

An optimal measure that reaches the infimum is called an **optimal plan** from μ to ν.

The Wasserstein distance enjoys two properties: first, it is the metric in which the value functions defined above are Lipschitz, see Cardaliaguet and Quincampoix [25, Lemma 3 and Proposition 1]. Second, the optimal plans allow to define the appropriate sub and super differentials, which in turn are needed to introduce a suitable definition of viscosity solutions in the infinite dimensional space $\mathcal{M}$, see Cardaliaguet [25, Lemma 4 and Definition 1]. Note that these complications do not arise in the model described in Section 8.3.

Let us suppose that the support of μ_1 is the compact set $K_1 \subset \mathbb{R}^n$ and that for all $(\mathbf{u}, \mathbf{v})$, the corresponding dynamics defined by (8.20) leaves K_1 invariant, i.e. is such that, for all $t \geq t_1$:

$$x_1 \in K_1 \implies \mathbf{x}[t_1, x_1, \mathbf{u}, \mathbf{v}](t) \in K_1.$$

Now, if we consider a finite discretization of K_1 with diameter $h > 0$, denoted $K_{1,h}$, then we have that the game with finite state space $K_{1,h}$ as in Section 8.3 is an "approximation" of the game of Cardaliaguet and Quincampoix. Thus, we could, at least informally, approximate a game in the Wasserstein space by a sequence of differential games.

We leave this question unanswered, but let us refer to Kloeckner [60] and the references therein which could help to establish such approximation.

8.5.2 Private information, privately controlled parameter

We describe in this Section a generalization of the model of Aumann-Maschler [7] of repeated games with incomplete information on both sides. We consider a situation were each player observes and controls the evolution of its own continuous time Markov chain and the payoff depends on both. The model described here is a generalization of the Aumann-Maschler model since here the private information of each player changes, while in their model it remains fixed.

Several dynamic extensions of the classical model of Aumann-Maschler of incomplete information games have been proposed in the literature. A first extension is due to Renault [82], who studies a game where one player observes a Markov chain in discrete time while his adversary observes the actions only. This has been later extended by Gensbittel and Renault [46] to the case when each player observes his own Markov chain and the actions of the other player only. However, both papers consider discrete time games only, with no underlying continuous time dynamics.

A different information structure is considered in Gensbittel [45]. In that paper, the players receive different information about an exogenous continuous time process with two coordinates: player 1 observes both coordinates while player 2 observes only the second coordinate. As in [26], the limit value when the stage duration goes to zero is studied and different characterization results obtained.

The model described in this Section is borrowed from Cardaliaguet, Rainer, Rosenberg and Vieille [26].

Let L, M denote two finite sets such that $\Omega = L \times M$. Let $q_L : L \times L \times A \to \mathbb{R}$ and $q_M : M \times M \times B \to \mathbb{R}$ denote two transition rate functions with associated generators Q_L^a, Q_M^b given by:

$$Q_L^a := \left(q_L(\ell, \ell', a)\right)_{\ell,\ell'} \quad \text{and} \quad Q_M^b := \left(q_M(m, m', b)\right)_{m,m'} .$$

A couple $(\ell_1, m_1) \in L \times M$ is chosen with the commonly known law $\xi \otimes \eta \in \Delta(L) \times \Delta(M)$. Player 1 observes the value of ℓ_1 and player 2 of m_1 and choose their actions a_1, b_1.

The game is played as follows: at stage $k \geq 2$, player 1 receives a signal $r_k = \{\ell_k, b_{k-1}\}$ and player 2 receives a signal $s_k = \{m_k, a_{k-1}\}$. The players choose their actions, a_k, b_k respectively, which depend on the information they receive. Once the actions are chosen, the parameters ℓ_k and m_k follow the Markov chain with generators $Q_L^{a_k}$ and $Q_M^{b_k}$ respectively, on the interval $[t_k, t_{k+1}]$ and the situation is repeated. Hence this model is a particular case of the general model introduced in Section 8.1.1.

Recall from Section 8.1.2 that the history for player 1, denoted $\mathcal{H}^1$ is the set of sequences of the form $(r_1, a_1, r_2, a_2, \ldots)$, with a similar definition for player 2. A **behavioural strategy** for player 1 is a function $\sigma : \mathcal{H}^1 \to \Delta(A)^L$. Behavioural strategies for player 2 are defined in a completely analogous way. Note that the strategies here depend on the private information of the players.

A strategy μ of player 1 is **non-revealing** if it is of the form $\mu : \mathcal{H}^1 \to \Delta(A)$. Similarly, a non-revealing strategy of player 2 is a function $\nu : \mathcal{H}^2 \to \Delta(B)$. That is, non-revealing strategies are strategies in which players ignore their information about the parameter.

The natural state space in this setting is the law of the unobserved parameter. The players can compute this law as follows:

Player 2 can update his beliefs on player 1's parameter conditional on player 1's actions if he knows how player 1 generated his actions. More precisely, assume player 2's belief is $\xi \in \Delta(L)$ and that he knows that player 1 used the strategy $x \in \Delta(A)^L$. If player 2 observes the action a, his conditional belief is given by:

$$\hat{\xi}(x,a) := \left(\frac{\xi_r x_r(a)}{x(\xi)(a)}\right)_{r\in L}$$

where

$$x(\xi)(a) := \sum_{r\in L} \xi_r x_r(a).$$

The computation for player 2 is completely analogous: let $y \in \Delta(B)^M$ and $\eta \in \Delta(M)$. The conditional distribution of player 2's parameter from player 1's point of view, when strategy y is used and the initial distribution is η is:

$$\hat{\eta}(y,b) := \left(\frac{\eta_r y_r(b)}{y(\eta)(b)}\right)_{r\in M}$$

where

$$y(\eta)(b) := \sum_{r\in M} \eta_r y_r(b).$$

At time $k+1$, the transition law of the variable ℓ_{k+1} is :

$$\xi_{k+1}^{a_k} := \exp\left(\pi_k Q_L^{a_k}\right) \cdot \hat{\xi}_k(x, a_k). \tag{8.21}$$

Similarly, the transition law of the variable m_{k+1} is:

$$\eta_{k+1}^{b_k} := \exp\left(\pi_k Q_M^{b_k}\right) \cdot \hat{\eta}_k(y, b_k). \tag{8.22}$$

To obtain an auxiliary game with a recursive structure as in the general model of Mertens, Sorin and Zamir [75, Section IV] recalled in the Introduction of this thesis, we consider as state variable the belief of each player on its unobserved parameter. In this case, the computation of the state variable is very different from the game with no information on the parameter. In particular, in this auxiliary game, each player needs to know the strategy of his adversary to calculate the law of his adversary's parameter.

Once the auxiliary state variables and the dynamics of the game have been specified, one can define the corresponding payoffs as in model A, B, C and their value functions.

Let us describe a possible extension of the variational approach when the payoff is evaluated as in model A. The crucial difficulty relies on the comparison principle.

The case of model B has been treated by Cardaliaguet, Rainer, Rosenberg and Vieille [26]. Their main results include the existence and characterization of the asymptotic value as $\delta \to 0$ as the unique viscosity solution of a limit PDE, with a modified notion of viscosity solutions. In the case of only one uninformed player, a different characterization of the asymptotic value is obtained as a martingale maximization problem.

8.5.3 Model A

As in Section 8.1.1, let Θ denote a decreasing evaluation. The Θ-evaluation of the payoff associated to a history $h = \{\ell_1, m_1, a_1, b_1, \ell_2, m_2, a_2, b_2, \ldots\}$ is

$$J_{A,\Theta}(h) := \mathbb{E}\left\{\sum_{k=1}^{\infty} \theta_k \gamma\left((\ell_k, m_k), a_k, b_k\right)\right\}$$

where the expectation is taken on each interval $[s_k, s_{k+1})$ with respect to the probability induced by the corresponding generators.

Let $\Gamma_{A,\Theta}(s_k,\xi,\eta)$ denote the game starting at s_k and $V_{A,\Theta}(s_k,\xi,\eta)$ denote its value, i.e.,

$$\begin{aligned} V_{A,\Theta}(s_k,\xi,\eta) &:= \max_{\sigma\in\Sigma}\min_{\tau\in\mathcal{T}} \mathbb{E}_{\sigma\tau}\left\{\sum_{r=k}^{\infty}\theta_r\gamma\left((\ell_r,m_r),a_r,b_r\right)\right\} \\ &:= \min_{\tau\in\mathcal{T}}\max_{\sigma\in\Sigma} \mathbb{E}_{\sigma\tau}\left\{\sum_{r=k}^{\infty}\theta_r\gamma\left((\ell_r,m_r),a_r,b_r\right)\right\}. \end{aligned}$$

The value of this game exists by Sion's minmax theorem.

We study this game directly as a game played on the nodes of a partition of $[0,1]$. First let us collect some properties of the value function in the next proposition.

Proposition 8.5.3. *Let $(s_k,\xi,\eta)\in\Pi_\Theta\times\Delta(L)\times\Delta(M)$. The value function $V_{A,\Theta}(s_k,\xi,\nu)$ satisfies the following properties:*

a) $V_{A,\Theta}(s_k,\cdot,\nu)$ is concave.

b) $V_{A,\Theta}(s_k,\xi,\cdot)$ is convex.

c) The following recursive formula holds:

$$\begin{aligned} V_{A,\Theta}(s_k,\xi,\eta) &= \max_{x\in\Delta(A)^L}\min_{y\in\Delta(B)^M}\mathbb{E}_{xy}\left\{\theta_k\gamma(\xi\otimes\eta,a,b)+V_{A,\Theta}(s_{k+1},\xi_k^a,\eta_k^b)\right\} \\ &= \min_{y\in\Delta(B)^M}\max_{x\in\Delta(A)^L}\mathbb{E}_{xy}\left\{\theta_k\gamma(\xi\otimes\eta,a,b)+V_{A,\Theta}(s_{k+1},\xi_k^a,\eta_k^b)\right\} \end{aligned}$$

with $\xi_k^a := \exp\left(\theta_k Q_L^a\right)\cdot\hat{\xi}(x,a)$ and $\eta_k^b := \exp\left(\theta_k Q_M^b\right)\cdot\hat{\eta}(y,b)$.

d) Let $V_{A,\Theta}$ denote the linear extension to $\mathbb{R}_+$ of the above value functions. The family $\{V_{A,\Theta}\}_\Theta$ is uniformly Lipschitz.

Proof. Let us do the proof of Property a), since the proof of Property b) is completely analogous. For this, we follow Sorin [92, Lemma 2.2]. Let $\xi := \alpha\xi_1+(1-\xi_2)$, where $\xi_1,\xi_2\in\Delta(L)$ and $\alpha\in[0,1]$. Now let us consider two auxiliary games, with an extra stage. In the first game, $i\in\{1,2\}$ is chosen with probability $(\alpha,1-\alpha)$, player 2 is informed of i and then the game $\Gamma_{A,\Theta}(s_k,\xi_i,\eta)$ is played. In the second game, player 2 is not informed. Note that the knowledge of i is irrelevant for player 1, since he knows the parameter ℓ_1. The first situation is better for player 2. Since the first game is equivalent to playing $\Gamma_{A,\Theta}(s_k,\xi_1,\eta)$ with probability α and $\Gamma_{A,\Theta}(t_k,\xi_2,\eta)$ with probability $1-\alpha$, then:

$$\alpha V_{A,\Theta}(s_k,\xi_1,\eta)+(1-\alpha)V_{A,\Theta}(s_k,\xi_2,\eta)\le V_{A,\Theta}(s_k,\xi,\eta).$$

Property c) follows from similar arguments as in Sorin [92, Proposition 4.21].

For Property d), let us prove first the Lipschitz continuity of $V_{A,\Theta}(s_k,\cdot,\cdot)$. For a couple of behavioural strategies (σ,τ) let

$$J_{A,\Theta}(\sigma,\tau,s_k,\xi,\eta) := \mathbb{E}_{\sigma\tau}\left\{\sum_{r=k}^{\infty}\theta_r\gamma\left((\ell_r,m_r),a_r,b_r\right)\right\}.$$

If $(\xi_1,\eta_1)\in\Delta(L)\times\Delta(M)$ and $(\xi_2,\eta_2)\in\Delta(L)\times\Delta(M)$ denote two couples of beliefs of each player at time s_ℓ and for $i=1,2$ $(\xi_i^a,\eta_i^b)\in\Delta(L)\times\Delta(M)$ denote the beliefs at time $s_{\ell+1}$ after observing actions $(a,b)\in A\times B$, then, it follows from (8.21) and (8.22) that:

$$\|(\xi_1^a,\eta_1^b)-(\xi_2^a,\eta_2^b)\|\le\exp\left(\theta_\ell\|q\|\right)\cdot\|(\xi_1,\eta_1)-(\xi_2,\eta_2)\|.$$

It follows that, if $J_i := J_{A,\Theta}(\sigma, \tau, s_k, \xi_i, \eta_i)$ for simplicity, we obtain:

$$\begin{aligned} |J_1 - J_2| &\leq \sum_{r=k}^{\infty} \theta_r \exp\left(\theta_\ell \|q\|\right) \|\gamma\| \cdot \|(\xi_1, \eta_1) - (\xi_2, \eta_2)\| && (8.23) \\ &\leq \sum_{r=k}^{\infty} \theta_r \cdot (1 + \theta_\ell \|q\|) \|\gamma\| \cdot \|(\xi_1, \eta_1) - (\xi_2, \eta_2)\| && (8.24) \\ &\leq (1 + \|q\|) \cdot \|\gamma\| \cdot \|(\xi_1, \eta_1) - (\xi_2, \eta_2)\|. && (8.25) \end{aligned}$$

The Lipschitz continuity follows from the recursive formula in Property c) and the Lipschitz continuity of $V_{A,\Theta}(s_k, \cdot, \cdot)$. □

From Property d) and Arzelà-Ascoli's theorem, the set of accumulation points of $\{V_{A,\Theta}\}_\Theta$ is non-empty. Finally, let us introduce the **non-revealing game**. This is the game where players use the non-revealing strategies introduced above. Since this is a particular case of the game studied in Section 8.3.3, we know its value exists and is denoted by U_Θ. The value of the non-revealing game satisfies:

$$\begin{aligned} U_\Theta(t_k, \xi, \nu) &= \max_{x \in \Delta(A)} \min_{y \in \Delta(B)} \mathbb{E}_{xy} \left\{ \theta_k \gamma(\xi \otimes \eta, a, b) + U_\Theta(t_{k+1}, \xi_k^a, \eta_k^b) \right\} \\ &= \min_{y \in \Delta(B)} \max_{x \in \Delta(A)} \mathbb{E}_{xy} \left\{ \theta_k \gamma(\xi \otimes \eta, a, b) + U_\Theta(t_{k+1}, \xi_k^a, \eta_k^b) \right\} \end{aligned}$$

As a straightforward consequence of Proposition 8.3.2, we have:

Corollary 8.5.4. *The family $\{U_\Theta\}_\Theta$ has an accumulation point which is the unique viscosity solution of*

$$\frac{\partial w}{\partial t}(t, \xi, \eta) + H^*(\xi, \eta, \nabla w(t, \xi, \eta)) = 0 \tag{8.26}$$

with terminal condition

$$w(1, \xi, \eta) = 0$$

where

$$\begin{aligned} H^*(\xi, \eta, p_1, p_2) &:= \max_{x \in \Delta(A)} \min_{y \in \Delta(B)} \mathbb{E}_{xy} \left\{ \langle Q_L^a \cdot \xi, p_1 \rangle + \left\langle Q_M^b \cdot \eta, p_2 \right\rangle + \gamma(\xi \otimes \eta, a, b) \right\} \\ &= \min_{y \in \Delta(B)} \max_{x \in \Delta(A)} \mathbb{E}_{xy} \left\{ \langle Q_L^a \cdot \xi, p_1 \rangle + \left\langle Q_M^b \cdot \eta, p_2 \right\rangle + \gamma(\xi \otimes \eta, a, b) \right\}. \end{aligned}$$

8.5.4 The variational approach

An important tool for establishing asymptotic properties of the value function is the **variational approach** introduced in Cardaliaguet, Laraki and Sorin [23]. Their main results are recalled in this Section and the proofs adapted whenever needed.

Denote by $\mathcal{F}_S$ the set of functions $f : \Delta(L) \times \Delta(M) \to [0, 1]$ such that, for all $(\xi, \eta) \in \Delta(L) \times \Delta(M)$, $f(\cdot, \eta)$ is concave and $f(\xi, \cdot)$ is convex. Let us introduce the operator $\mathbf{T} : \mathcal{F}_S \to \mathcal{F}_S$ which is defined by:

$$\mathbf{T}[f](\xi, \eta) := \mathrm{val}_{x,y} \mathbb{E}_{xy} \left\{ f(\hat{\xi}(x, a), \hat{\eta}(y, b)) \right\}$$

This is the **projective operator**, see Sorin [92, Appendix C]. Note that the projective operator is continuous. One important property of the projective operator is the following:

Lemma 8.5.5. *(Lemma 4.26 in [92, Appendix C]). For all $f \in \mathcal{F}_S$ and $(\xi, \eta) \in \Delta(L) \times \Delta(M)$, we have*

$$\mathbf{T}[f](\xi, \eta) = f(\xi, \eta).$$

Let $\mathcal{F}$ denote the set of functions $W : [0,1] \times \Delta(L) \times \Delta(M) \to [0,1]$ such that, for all $t \in [0,1]$, $W(t, \cdot, \cdot) \in \mathcal{F}_S$. Let $\mathcal{F}_0$ denote the set of accumulation points of $\{V_{A,\Theta}\}_\Theta$, which is non-empty by Property d) in Proposition 8.5.3. From the dynamic programming formula, i.e. Property c) in Proposition 8.5.3, we get that $\mathcal{F}_0 \subset \mathcal{F}$.

Fix $W \in \mathcal{F}$ and denote by $\mathbf{X}(t, \xi, \eta, W)$ the set of optimal strategies of player 1 in $\mathbf{T}[W(t, \cdot, \cdot)](\xi, \eta)$. The set $\mathbf{Y}(t, \xi, \eta, W)$ is defined for player 2 in a completely analogous way. A set of strategies is **non revealing** if all its elements are non revealing strategies.

We introduce now two variational properties, adapted from those proposed by Cardaliaguet, Laraki and Sorin [23, Properties P1 and P2].

Properties 8.5.6. *(Variational properties) Let $\phi : [0,1] \times \Delta(L) \times \Delta(M) \to \mathbb{R}$ denote a C^1 test function and $W \in \mathcal{F}$.*

- ***P1:*** *If $(t, \xi, \eta) \in [0,1] \times \Delta(L) \times \Delta(M)$ are such that $\mathbf{X}(t, \xi, \eta, W)$ is non-revealing, and $W - \phi$ has a local maximum at (t, ξ, η) then*

$$\frac{\partial \phi}{\partial t}(t, \xi, \eta) + H^*(\xi, \eta, \nabla \phi(t, \xi, \eta)) \geq 0. \tag{8.27}$$

- ***P2:*** *If $(t, \xi, \eta) \in [0,1] \times \Delta(L) \times \Delta(M)$ are such that $\mathbf{Y}(t, \xi, \eta, W)$ is non-revealing, and $W - \phi$ has a local minimum at (t, ξ, η) then*

$$\frac{\partial \phi}{\partial t}(t, \xi, \eta) + H^*(\xi, \eta, \nabla \phi(t, \xi, \eta)) \leq 0. \tag{8.28}$$

The following Lemma follows from Property c) in Proposition 8.5.3. The proof is adapted from Lemma 8 in Cardaliaguet, Laraki and Sorin [23].

Lemma 8.5.7. *Any $W \in \mathcal{F}_0$ satisfies **P1** and **P2**.*

Proof. Let W satisfying **P1** at (t, ξ, η), and without loss of generality assume the maximum is strict. Let $(\Theta_i)_{i \in \mathbb{N}}$ be a subsequence of partitions of $[0,1]$ such that V_{A,Θ_i} converges uniformly to W as $i \to \infty$. Take a fixed $\epsilon > 0$ and choose i_0 such that

$$\|V_{A,\Theta_i} - W\| < \epsilon, \quad i > i_0$$

and set $\Theta := \Theta_{i_0}$. Denote by $\Pi_\Theta = \{t_k\}_{k \in \mathbb{N}}$ the partition induced by Θ. Finally, let k such that the restriction of $V_{A,\Theta} - \phi$ to Π_Θ has a strict maximum at t_k. Since t is a strict maximum, $t_k \to t$ as $i \to \infty$.

Recall that

$$V_{A,\Theta}(t_k, \xi, \eta) = \max_{x \in \Delta(A)^L} \min_{y \in \Delta(B)^M} \mathbb{E}_{xy} \left\{ \theta_k \gamma(\xi \otimes \eta, a, b) + V_{A,\Theta}(t_{k+1}, \xi_k^a, \eta_k^b) \right\}.$$

Let x_k denote an optimal strategy for player 1 in the right hand side of the above equation and b any action of player 2. We have that:

$$\begin{aligned} V_{A,\Theta}(t_k, \xi, \eta) &\leq \sum_{a \in A} \left\{ \theta_k \gamma(\xi \otimes \eta, x_k, b) + x_k(a) V_{A,\Theta}(t_{k+1}, \exp(\theta_k Q_L^a) \cdot \hat{\xi}(x_k, a), \exp(\theta_k Q_M^b) \cdot \eta) \right\} \\ &\leq \theta_k \gamma(\xi \otimes \eta, x_k, b) + V_{A,\Theta}(t_{k+1}, \exp(\theta_k Q_L^a) \cdot \xi, \exp(\theta_k Q_M^b) \cdot \eta) \end{aligned}$$

where the second inequality comes from concavity.

Since (t_k, ξ, η) is a strict maximum of $W - \phi$ on $\Pi_\Theta \times \Delta(L) \times \Delta(M)$, we have that

$$V_{A,\Theta}(t_{k+1}, \exp(\theta_k Q_L^a) \cdot \xi, \exp(\theta_k Q_M^b) \cdot \eta) - V_{A,\Theta}(t_k, \xi, \eta)$$

is less or equal than

$$\phi(t_{k+1}, \exp(\theta_k Q_L^a) \cdot \xi, \exp(\theta_k Q_M^b) \cdot \eta) - \phi(t, \xi, \eta)$$

from which we obtain

$$0 \leq \gamma(\xi \otimes \eta, x_k, b) + \frac{\phi(t_{k+1}, \exp(\theta_k Q_L^a) \cdot \xi, \exp(\theta_k Q_M^b) \cdot \eta) - \phi(t, \xi, \eta)}{\theta_k}.$$

Since $\Delta(A)^L$ is compact, we have that x_k converges to a point x, which, by upper semi continuity of the argmax correspondence and uniform convergence of V_{A,Θ_i} gives $x \in \mathbf{X}(t, \xi, \eta, W)$. By assumption, x is non revealing. Thus, we get

$$0 \leq \frac{\partial \phi}{\partial t} + \min_{b \in \Delta(B)} \mathbb{E}_{xy} \left\{ \langle Q_L^a \cdot \xi, \nabla_\xi \phi \rangle + \left\langle Q_M^b \cdot \eta, \nabla_\eta \phi \right\rangle + \gamma(\xi \otimes \eta, a, b) \right\}$$

for all x non revealing, which concludes the proof. □

Following the variational approach of Cardaliaguet, Laraki and Sorin [23], the next step would be to establish a sort of comparison principle, which is crucial to obtain a unique accumulation point for the family $V_{A,\Theta}$.

Conjecture 1. *Let W_1 and W_2 in $\mathcal{F}$ satisfying respectively **P1, P2** and*

- ***P3:*** $W(1, \xi, \eta) \leq W_2(1, \xi, \eta)$, $\forall (\xi, \eta) \in \Delta(L) \times \Delta(M)$.

Then, for all $t \in [0, 1]$,

$$W_1(t, \xi, \eta) \leq W_2(t, \xi, \eta).$$

To prove this, the idea is to assume that

$$\max_{t, \xi, \eta} W_1(t, \xi, \eta) - W_2(t, \xi, \eta) = \delta > 0. \tag{8.29}$$

and deduce, by suitable penalization arguments, the existence of a point (t', ξ', η') such that:

— $t' < 1$

— At the point (t', ξ', η'), W_1 satisfies **P1**.

— At the point (t', ξ', η'), W_2 satisfies **P2**.

Thus, finding such point contradicts (8.29).

However, deducing the existence of such point requires very subtle and technically involved arguments and is the main difficulty to extend the variational approach to this framework.

A similar difficulty arises in Cardaliaguet, Rainer, Rosenberg and Vieille [26]. The authors establish a comparison principle for the limiting partial differential equation corresponding to model B using the techniques originally introduced by Crandall, Ishi and Lions [32]. We conjecture that similar tools could be used in this framework.

Bibliography

[1] Y. Achdou and I. Capuzzo-Dolcetta, *Mean field games: Numerical methods*, SIAM J. Numer. Anal. **48** (2010), 1136–1162.

[2] S. Adlakha, R. Johari, and G.Y. Weintraub, *Equilibria of dynamic games with many players: Existence, approximation, and market structure*, arXiv:1011.5537, 2011.

[3] J.P. Aubin, *Victory and defeat in differential games*, Modelling and Control of Systems, Lecture Notes in Control and Inform. Sci., Springer-Verlag, 1989, pp. 337–347.

[4] ———, *Viability Theory*, Birkhäuser Boston Inc., 1991.

[5] J.P. Aubin and H. Frankowska, *Set-Valued Analysis*, Birkhäuser Boston Inc., 1990.

[6] R. J. Aumann, *Markets with a continuum of traders*, Econometrica **32** (1964), 39–50.

[7] R. J. Aumann and M. B. Maschler, *Repeated Games with Incomplete Information*, MIT Press, 1995, With the collaboration of Richard E. Stearns.

[8] M. Bardi, *Explicit solutions of some linear-quadratic mean field games*, Networks and heterogeneous media **7** (2012), 243–261.

[9] M. Bardi and I. Capuzzo-Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhäuser Boston Inc., 1997, With appendices by M. Falcone and P. Soravia.

[10] M. Benaïm and J. Weibull, *Deterministic approximation of stochastic evolution in games*, Econometrica **71** (2003), 873–903.

[11] R. Benedetti and J.J. Risler, *Real Algebraic and Semi-Algebraic Sets*, Hermann, 1990.

[12] A. Bensoussan, J. Frehse, and S.C.P. Yam, *Mean Field Games and Mean Field Type Control Theory*, Springer-Verlag, 2013.

[13] T. Bewley and E. Kohlberg, *The asymptotic theory of stochastic games*, Math. Oper. Res. **1** (1976), 197–208.

[14] D. Blackwell, *An analog of the minimax theorem for vector payoffs*, Pacific J. Math. **6** (1956), 1–8.

[15] D. Blackwell and T. Ferguson, *The "Big Match"*, Annals of Math. Stat. **39** (1968), 159–163.

[16] B. Bollobas, I. Leader, and M. Walters, *Lion and man - can both win?*, Israel J. Math. **189** (2012), 267–286.

[17] R. Buckdahn, P. Cardaliaguet, and C. Rainer, *Nash equilibrium payoffs for non-zero sum stochastic differential games*, SIAM J. Control Optim. **43** (2004), 624–642.

[18] R. Buckdahn, J. Li, and M. Quincampoix, *Value function of differential games without isaacs conditions. an approach with nonanticipative mixed strategies*, Internat. J. Game Theory (2012), 1–32.

[19] F. Camilli and F. Silva, *A semi-discrete approximation for a first order mean field game problem*, Networks and heterogeneous media **7** (2012), 263–277.

[20] P. Cardaliaguet, *Notes on mean field games (from P.L. Lions' lectures at Collège de France)*, 2012.

[21] _____, *A differential game with two players and one target*, SIAM J. Control Optim. **34** (1996), 1441–1460.

[22] _____, *Introduction to Differential Games*, Lecture notes, 2010.

[23] P. Cardaliaguet, R. Laraki, and S. Sorin, *A continuous time approach for the asymptotic value in two-person zero-sum repeated games*, SIAM J. Control Optim. **50** (2012), 1573–1596.

[24] P. Cardaliaguet and S. Plaskacz, *Invariant solutions of differential games and hamilton-jacobi-isaacs equations for time measurable hamiltonians*, SIAM J. Control Optim. **38** (2000), 1501–1520.

[25] P. Cardaliaguet and M. Quincampoix, *Deterministic differential games under probability knowledge of initial condition*, Int. Game Theory Rev. **10** (2008), 1–16.

[26] P. Cardaliaguet, C. Rainer, D. Rosenberg, and N. Vieille, *Markov games with frequent actions and incomplete information*, arXiv:1307.3365, 2013.

[27] E. Carlini and F. Silva, *A fully-discrete semi-Lagrangian approximation for a first order mean field game problem*, arXiv 1212.4757, 2013.

[28] J. Casti, *Five Golden Rules: Great Theories of 20th–Century Mathematics - And Why They Matter,*, ch. The minimax theorem, John Wiley & Sons, Inc., 1996.

[29] F. H. Clarke and Yu. S. Ledyaev, *Mean value inequalities in Hilbert space*, Trans. Amer. Math. Soc. **344** (1994), 307–324.

[30] F. H. Clarke, Yu. S. Ledyaev, R. J. Stern, and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, Graduate Texts in Mathematics, vol. 178, Springer-Verlag, New York, 1998.

[31] E. A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*, McGraw-Hill Book Company, Inc., 1955.

[32] M. G. Crandall, H. Ishii, and P.L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. (N.S.) **27** (1992), no. 1, 1–67.

[33] M. G. Crandall and P.L. Lions, *Viscosity solutions of Hamilton-Jacobi equations*, Trans. Amer. Math. Soc. **277** (1983), 1–42.

[34] R. Elliot, Xun Li, and Yuan-Hua Ni, *Discrete time mean-field stochastic linear-quadratic optimal control problems*, Automatica **49** (2013), 3222–3233.

[35] R.J. Elliot and N.J. Kalton, *The existence of value in differential games*, vol. 126, American Mathematical Society, 1972.

[36] L.C. Evans and P.E. Souganidis, *Differential games and representation formulas for solutions of Hamilton-Jacobi-Isaacs equations*, Indiana Univ. Math. J. **33** (1984), 773–797.

[37] E. Feleqi, *The derivation of ergodic mean field game equations for several populations of players*, Dyn. Games Appl. **3** (2013), 523–536.

[38] W. H. Fleming, *A note on differential games of prescribed duration*, Contributions to the Theory of Games (M. Dresher, L.S. Shapley, and A.W. Tucker, eds.), vol. III, Annals of Math. Studies, no. 39, Princeton University Press, 1957, pp. 407–412.

[39] ———, *The convergence problem for differential games II*, Advances in Game Theory (K. Dresher, A.W. Tucker, and P. Wolfe, eds.), Annals of Math. Studies, no. 52, Princeton University Press, 1964, pp. 195–210.

[40] H. Frankowska, *Lower semicontinuous solutions of Hamilton-Jacobi-Bellman equations*, SIAM J. Control Optim. **31** (1993), 257–272.

[41] A. Friedman, *Differential Games*, Wiley New York, 1971.

[42] ———, *Differential Games*, CBMS 18 Amer. Math. Soc., 1974.

[43] D. Fudenberg and J. Tirole, *Learning-by-doing and market performance*, The Bell J. of Economics **14** (1983), 522–530.

[44] N. Gast, B. Gaujal, and J.Y. Le-Boudec, *Mean field for Markov decision processes: from discrete to continuous optimization*, IEEE Transactions on Automatic Control **57** (2012), 2266–2280.

[45] F. Gensbittel, *Continuous-time limit of dynamic games with incomplete information and a more informed player.*, HAL–00910970.

[46] F. Gensbittel and J. Renault, *The value of Markov chain games with incomplete information on both sides*, HAL-00745986, 2012.

[47] D. Gomes and J. Saude, *Mean field games models: a brief survey*, Dyn. Games Appl. **4** (2013), 110–154.

[48] D.A. Gomes, J. Mohr, and R.R. Souza, *Discrete time, finite state space mean field games*, J. Math. Pures et Ap. **93** (2010), 308–328.

[49] O. Guéant, J.M. Lasry, and P.L. Lions, *Mean Field Games and Applications*, Paris Princeton Lectures on Mathematical Finance 2010, Lecture Notes in Math., vol. 2003, Springer-Verlag, 2011, pp. 205–266.

[50] X. Guo and O. Hernández-Lerma, *Zero-sum continuous-time Markov games with unbounded transition and discounted payoff rates*, Bernoulli **11** (2005), 1009–1029.

[51] W.D. Hamilton, *Extraordinary sex ratios*, Science **156** (1967), 477–488.

[52] A. Haurie and P. Marcotte, *On the relationship between Nash-Cournot and Wardrop equilibria*, Networks **15** (1984), 295–308.

[53] O. Hernández-Lerma and J.B. Lasserre, *Discrete-Time Markov Control Problems*, Springer-Verlag, 1996.

[54] O. Hernández-Lerma and T. Prieto-Rumeau, *Selected Topics on Continuous-Time Controlled Markov Chains and Markov Games*, Imperial College Press, 2012.

[55] W. Hoeffding, *Probability inequalities for the sum in sampling without replacement*, Annals of Statistics **2** (1963), 39–48.

[56] M. Huang, P.E. Caines, and R.P. Malhamé, *Individual and mass behavior in large population stochastic wireless power control problems: centralized and nash equilibrium solutions.*, Proceedings of the 42nd IEEE Conference on Decision and Control, 2003.

[57] ———, *Large population stochastic dynamic games: closed-loop Mc Kean-Vlasov systems and the Nash certainty equivalence principle*, Communications in Information and Systems **6** (2006), 221–252.

[58] R. Isaacs, *Differential Games: A Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization*, Dover Books on Mathematics, 1999.

[59] B. Jovanovic and R. Rosenthal, *Anonymous sequential games*, Journal of Mathematical Economics **17** (1988), 77–87.

[60] B. Kloeckner, *Approximation by finitely supported measures*, arXiv:1003.1035v2.

[61] N. N. Krasovskii and A. Subbotin, *Game-Theoretical Control Problems*, Springer-Verlag, New York, 1988.

[62] H.W. Kuhn, *Extensive games and the problem of information*, Contribution to the Theory of Games (H.W. Kuhn and A.W. Tucker, eds.), vol. II, Annals of Mathematical Studies, no. 28, Princeton University Press, 1953, pp. 193–216.

[63] R. Laraki, *Explicit formulas for repeated games with absorbing states*, Internat. J. Game Theory **39** (2010), no. 1-2, 53–69.

[64] J.M. Lasry and P.L. Lions, *Jeux à champ moyen. i. Le cas stationnaire*, C.R. Math. Acad. Sci. Paris **343** (2006), 619–625.

[65] J.M Lasry and P.L. Lions, *Jeux à champ moyen. ii. Horizon fini et controle optimal.*, C.R. Math. Acad. Sci. Paris **343** (2006), 679–684.

[66] J.M. Lasry and P.L. Lions, *Mean field games*, Japanese Journal of Mathematics **2** (2007), 229–260.

[67] P.L. Lions, *Generalized Solutions of Hamilton-Jacobi Equations*, Pitman, 1982.

[68] P.L. Lions and P.E. Souganidis, *Differential games, optimal control and directional derivatives of viscosity solutions of Bellman's and Isaacs' equations*, SIAM J. Control Optim. **23** (1985), 506–583.

[69] ———, *Differential games and directional derivatives of viscosity solutions of Isaacs' equations II.*, SIAM J. Control Optim. **24** (1986), 1086–1089.

[70] J.E. Littlewood, *A Mathematicians' Miscellany*, Methuen, 1953.

[71] J.P. Maldonado Lopez and M. Oliu-Barton, *A strategy-based proof of the existence of the value in zero-sum differential games*, to appear in Morfismos, 2014.

[72] A. Mas-Colell, *On a theorem of Schmeidler*, Journal of Mathematical Economics **13** (1984), 201–206.

[73] J. Maynard Smith, *Game theory and the evolution of fighting*, On Evolution, Edinburgh University Press, 1972.

[74] J. Maynard Smith and G.R. Price, *The logic of animal conflict*, Nature **246** (1973), 15–18.

[75] J.F. Mertens, S. Sorin, and S. Zamir, *Repeated Games*, Core Discussion Papers 9420-9422, 1994.

[76] P.R. Milgrom and R.J. Weber, *Distributional strategies for games with incomplete information*, Math. Oper. Res. **10** (1985), 619–632.

[77] M. Nagumo, *Über die Lage der Integralkurven Gewöhnlicher Differentialgleichungen*, Proc. Phys.-Math. Soc. Japan (3) **24** (1942), 551–559.

[78] A. Neyman, *Stochastic games with short-stage duration*, Dyn. Games Appl. **3** (2013), 236–278.

[79] B.N. Pšeničnyj, *Leçons sur les jeux différentiels*, Contrôle optimal et jeux différentiels, Institute de Recherche d'Informatique et d'Automatique, 1971.

[80] M. L. Radulescu and F. H. Clarke, *Geometric approximation of proximal normals*, J. Convex Anal. **4** (1997), 373–379.

[81] T. E. S. Raghavan, *Zero-sum two-person games*, Handbook of game theory with economic applications, Vol. II, Handbooks in Econom., vol. 11, North-Holland, Amsterdam, 1994, With an appendix by T. Radzik and Raghavan, pp. 735–768.

[82] J. Renault, *The value of Markov chain games with lack of information on one side*, Math. Oper. Res. **31** (2006), 490–512.

[83] R. T. Rockafellar, *Proximal subgradients, marginal values, and augmented Lagrangians in nonconvex optimization*, Math. Oper. Res. **6** (1981), 424–436.

[84] E. Roxin, *The axiomatic approach in differential games*, J. Optim. Theory Appl. **3** (1969), 153–163.

[85] W. Sandholm, *Potential games with continuous player sets*, Journal of Economic Theory **97** (2001), 81–108.

[86] D. Schmeidler, *Equilibrium points of nonatomic games*, Journal of Statistical Phys. **7** (1973), 295–300.

[87] L. S. Shapley, *Stochastic games*, Proc. Nat. Acad. Sci. U.S.A. **39** (1953), 1095–1100.

[88] M. Sion, *On general minimax theorems*, Pacific J. Math. **8** (1958), 171–176.

[89] M.J. Smith, *The existence, uniqueness and stability of traffic equilibria*, Transp. Res. Part B **13** (1979), 295–304.

[90] S. Sorin, *Big Match with lack of information on one side (Part I)*, Internat. J. Game Theory **13** (1984), 201–255.

[91] _____, *Big Match with lack of information on one side (Part II)*, Internat. J. Game Theory **14** (1985), 173–204.

[92] _____, *A First Course on Zero-Sum Repeated Games*, vol. 37, Springer-Verlag, 2002.

[93] S. Sorin and G. Vigeral, *Reversibility and oscillations in zero-sum discounted stochastic games*, HAL-00869656, 2013.

[94] P.E. Souganidis, *Approximation schemes for the viscosity solutions of Hamilton-Jacobi equations*, Journal of Diff. Eq. **59** (1985), 1–43.

[95] _____, *Two-player, zero-sum differential games and viscosity solutions*, Stochastic and differential games (M. Bardi, T.E.S. Raghavan, and T. Parthasarathy, eds.), Ann. Internat. Soc. Dynam. Games, vol. 4, Birkhäuser Boston, 1999, pp. 69–104.

[96] A. Subbotin, *Generalized Solutions of First-Order PDEs*, Birkhäuser Boston Inc., 1995.

[97] _____, *Constructive theory of positional differential games and generalized solutions to Hamilton-Jacobi equations*, Stochastic and differential games (M. Bardi, T.E.S. Raghavan, and T. Parthasarathy, eds.), Ann. Internat. Soc. Dynam. Games, vol. 4, Birkhäuser Boston, 1999, pp. 3–67.

[98] K. Tanaka and H.C. Lai, *A two-person zero-sum Markov game with a stopped set*, Journal of Math. Analysis and Applications **86** (1982), 54–68.

[99] K. Tanaka and K. Wakuta, *On continuous time Markov games with the expected average reward criterion*, Sci. Rep. Niigata Univ. Ser. A (1977), 15–24.

[100] _____, *On continuous time Markov games with countable state space*, Journal of the Operations Research Soc. of Japan **21** (1978), 17–26.

[101] P. Varaiya, *The existence of solutions to a differential game*, SIAM J. Control Optim. **5** (1967), 153–162.

[102] R. Vinter, *Optimal Control*, Birkhäuser Boston Inc., 2010.

[103] J. von Neumann, *Zur Theorie der Gesellschatsspiele*, Mathematische Annalen **100** (1928), 295–320.

[104] C. Wan, *Jeux de congestion dans les réseaux: modèles et proprietés, partie I*, Technique et Science Informatiques **32** (2013), 951–980.

[105] ———, *Jeux de congestion dans les réseaux: modèles et proprietés, partie II*, Technique et Science Informatiques **32** (2013), 981–1005.

[106] J.G. Wardrop, *Some theoretical aspects of road traffic research*, Proc. Inst. Civil Eng. Part II **1** (1952), 325–378.

[107] L.E. Zachrisson, *Markov games*, Advances in Game Theory (M. Dresher, A.W. Tucker, and P. Wolfe, eds.), Annals of Mathematics Studies, no. 52, Princeton University Press, 1964, pp. 211–254.

[108] B. Ziliotto, *Zero-sum repeated games: counterexamples to the existence of the asymptotic value and the conjecture maxmin* $= limv_n$, HAL-00824039, 2013.