HAL
open science

Contributions à l'étude des propriétés asymptotiques en contrôle optimal et en jeux répétés

Xiaoxi Li

To cite this version:

Xiaoxi Li. Contributions à l'étude des propriétés asymptotiques en contrôle optimal et en jeux répétés. Optimisation et contrôle [math.OC]. Université Pierre et Marie Curie - Paris VI, 2015. Français. NNT : 2015PA066231 . tel-01232379

HAL Id: tel-01232379

https://theses.hal.science/tel-01232379v1

Submitted on 23 Nov 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 Science Mathématiques de Paris Centre

Thèse de doctorat

Discipline : Mathématiques Appliquées

Contributions à l'étude des propriétés asymptotiques en contrôle optimal et en jeux répétés

---

présentée par

Xiaoxi LI

dirigée par Sylvain Sorin

Soutenue le 22 septembre 2015 devant le jury composé de :

Mme Hélène Frankowska	Université Pierre-et-Marie Curie	président
M. Rida Laraki	Université Paris Dauphine	rapporteur
M. Eilon Solan	Tel Aviv University	rapporteur
M. Sylvain Sorin	Université Pierre-et-Marie Curie	directeur
M. Marc Quincampoix	Université de Brest	examinateur
M. Jérôme Renault	Université Toulouse 1 Capitole	examinateur
M. Tristan Tomala	HEC Paris	examinateur

Institut de Mathématiques de Jussieu-
Paris Rive Gauche
Case 247, UPMC-Paris 6
4 place Jussieu
75252 Paris Cedex 5

UPMC
Ecole Doctorale de Sciences
Mathématiques de Paris Centre
4 place Jussieu
75252 Paris Cedex 05
Boite courrier 290

Remerciements

First of all, my many thanks go to my thesis advisor, Sylvain Sorin. I enjoyed very much Sylvain's profound and elegant lectures in OJME which gave me a general picture of mathematical game theory. Sylvain is generous in sharing research questions and his ideas with students like me. I benefit a lot from the discussions with Sylvain, whose insightful comments shed lights on my research several times when I was in difficulty. I would also like to thank Sylvain for spending so many hours in reading and re-reading my manuscripts, picking out typos, and giving numerous remarks. Thanks to Sylvain, I had the chance to discover the beauty of French life style, such as wine and oyster tasting. I am encouraged by him to speak French all the time. All these have largely enriched my life in France.

I would like to express my special gratitudes to Jérôme Renault, who introduced me to the French community of game theory and showed me his kindness and friendship during my stay in France. I appreciate his lectures on game theory in Toulouse and in Paris. Since my memoire supervised by Jérôme at TSE (2009/10), it is always enjoyable to work with him. Discussion with him is quite provoking. At the beginning of my thesis (2012), he invited me to visit TSE and has taken plenty of time to talk with me on interesting problems. In this dissertation (Chapter 3), it is a great pleasure to have a co-authored paper with Jérôme.

I would like to thank the referees Rida Laraki and Eilon Solan for taking time to read this dissertation and to give useful comments. During my thesis, support from Rida is always available. Eilon showed his interest in my research, and gave me the chance to visit Tel Aviv at the end of my thesis.

I am grateful to all the jury members for their participation. My thanks go to Hélène Frankowska for her course of optimal control, which helped me conduct the research in this area. I would like to express my thanks to Marc Quincampoix for inviting me to Brest (2014), where our joint project was initialled. Many fruitful discussions with Tristan Tomala on repeated games are acknowledged.

I would like to acknowledge the important contributions to this dissertation from my co-authors, Marc Quincampoix, Jérôme Renault and Xavier Venel.

During my thesis, discussions with Guillaume Vigeral, Fabien Gensbittel, Miquel Oliu Barton and Marco Mazzola are quite helpful. Proof-readings of several chapters by Xavier Venel, Bruno Ziliotto, Guillaume Vigeral and Joon Kown are gratefully acknowledged. I also benefited from conversations with many other colleagues as Pablo Maldonado, Daniel Hoehener, Marie Laclau, Gaetan Fournier, Cheng Wan, Mario Bravo, Vianney Perchet, Yannick Viossat, Tonon Daniela, Teresa Scarinci, Ihab Haidar *etc.* I would like to thank

them for their helps and also for their precious friendships.

Grateful acknowledgements go to IMJ-PRG (Equipe Combinatoire et Optimisation, director Jean Paul Allouche) for hosting me during the 4 years of my thesis, to Paris 1 (UFR 06, coordinator Thierry Lafay) for providing me an ATER allocation in the last year of my thesis, to TSE for their support of my application for the Eiffel Scholarship (coordinator Aude Schloesing) in the first year of my arrival in France, and to Hausdorff Research Institute for Mathematics (HIM) for their financial support during my stay in Bonn 2013. I would like to thank also the organizers of Paris Game Theory Seminar, Workshops at Roscoff and at Foljuif, and Summer School of Game Theory at Aussois, for providing me the opportunities of academic communications.

Finally, I would like to say thanks to my family. It is their love and constant support that make my dream today come true.

Xiaoxi Li, 18 septembre 2015 à Paris

Résumé

Cette thèse étudie des propriétés limites de problèmes de contrôle optimal (un joueur, en temps continu) et de jeux répétés à somme nulle (à deux joueurs, en temps discret) avec horizon tendant vers l'infini. Plus précisément, nous étudions la convergence de la fonction valeur lorsque la durée du problème de contrôle ou la répétition du jeu tend vers l'infini (analyse asymptotique), et l'existence de stratégies robustes, i.e. des stratégies ε-optimales pour guarantir la valeur limite dans tous les problèmes de contrôle de durée suffisamment longue ou dans tous les jeux répétés de répétition suffisamment large (analyse uniforme).

La partie sur le contrôle optimal est composée de trois chapitres.

Le chapitre 2 est un article de présentation de la littérature récente sur les propriétés à long terme dans divers modèles d'optimisation dynamique. Dans les deux chapitres suivants, nous nous concentrons sur les problèmes de contrôle optimal où le coût de la trajectoire est évalué par une mesure de probabilité générale sur $\mathbb{R}_+$, au lieu de la moyenne de T-horizon (moyenne de Cesàro) ou de la λ-escompté (moyenne d'Abel).

Dans le chapitre 3, nous introduisons une condition de régularité asymptotique pour une suite de mesures de probabilité sur $\mathbb{R}_+$ induisant un horizon tendant vers l'infini (en particulier, T tendant vers l'infini ou λ tendant vers zéro). Nous montrons que pour toute suite d'évaluations satisfaisant cette condition, la suite associée des valeurs du problème de contrôle converge uniformément si et seulement si cette suite est totalement bornée pour la norme uniforme. On en déduit que pour des problèmes de contrôle définis sur un domaine invariant compact et vérifiant une certaine condition de non-expansivité, la fonction valeur définie par une mesure de probabilité générale converge quand l'évaluation devient suffisamment régulière. En outre, nous prouvons dans le chapitre 4 que sous les mêmes conditions de compacité et de non-expansivité, il existe des contrôles ε-optimaux pour tous les problèmes où le coût de la trajectoire est évalués par une mesure de probbailité suffisamment régulières.

La partie sur les jeux répétés se compose de deux chapitres.

Le chapitre 5 est consacré à l'étude d'une sous-classe de jeux absorbants à information incomplète d'un côté. Le modèle que nous considérons est une généralisation du Big match à information incomplète d'un côté introduit par Sorin (1984). Nous démontrons l'existence de la valeur limite, du $Maxmin$, du $Minmax$, et l'égalité du $Maxmin$ et de la valeur limite.

Dans le chapitre 6, nous établissons plusieurs résultats concernant des jeux récursifs. Nous considérons d'abord les jeux récursifs avec un espace dénombrable d'états et prouvons que si la famille des fonctions valeur des jeux à n étapes est totalement bornée pour la norme uniforme, alors la valeur uniforme existe. En particulier, la convergence uniforme des valeurs des jeux à n étapes implique la convergence uniforme des valeurs des jeux escomptés. À l'aide d'un résultat dans Rosenberg et Vieille (2000), on en déduit un théorème taubérien uniforme pour les jeux récursifs. Deuxièmement, nous appliquons le résultat d'existence de la valeur uniforme à une classe des modèles général de jeux répétés et nous prouvons que la valeur limite et le $Maxmin$ existent et sont égaux. Ces jeux répétés sont des jeux récursifs avec signaux où le joueur 1 peut toujours déduire le signal du joueur 2 de son propre signal.

Mots-clefs

Optimisation dynamique ; contrôle optimal ; évaluation générale ; jeux répétés à somme nulle ; jeux stochastiques ; analyse asymptotique ; analyse uniforme ; jeux stochastiques à information incomplète d'un côté ; valeur limite ; valeur uniforme ; théorème taubérien.

Contributions to the analysis of asymptotic properties in optimal control and repeated games

Abstract

This dissertation studies limit properties in optimal control problems (one-player, in continuous time) and in zero-sum repeated games (two-player, in discrete time) with large horizons. More precisely, we investigate the convergence of the value function when the duration of the control problem or the repetition of the game tends to infinity (the asymptotic analysis), and the existence of robust strategies, i.e. ε-optimal strategies to guarantee the limit value in all control problems with sufficiently long durations or in all repeated games with sufficiently large repetitions (the uniform analysis).

The part on optimal control is composed of three chapters.

Chapter 2 is a survey article on recent literature of long-term properties in various models of dynamic optimization. In the following two chapters, we focus on optimal control problems where the running cost is evaluated by a general probability measure, instead of the usual T-horizon average (Cesàro mean) or the λ-discount (Abel mean).

In Chapter 3, we introduce an asymptotic regularity condition for a sequence of probability measures on positive real numbers which induces a horizon tending to infinity (in particular T tending to infinity or λ tending to zero) for the control problem. We prove that for any sequence of evaluations satisfying this condition, the associated sequence of value functions of the control problem converges uniformly if and only if this sequence is totally bounded for the uniform norm. We deduce that for control problems defined on a compact invarant domain and satisfying some nonexpansive condition, the value function defined by a general probability measure converges as the evaluation becomes sufficiently regular. Further, we prove in Chapter 4 that under the same compact and nonexpansive conditions, there exist ε-optimal controls for all problems where the running cost is evaluated by a sufficiently regular probability measure.

The part on repeated games consists of two chapters.

Chapter 5 is devoted to the study of a subclass of absorbing games with one-sided incomplete information. The model we consider is a generalization of Big match with one-sided incomplete information introduced by Sorin (1984). We prove the existence of the limit value, $Maxmin$, $Minmax$, and that $Maxmin$ is equal to the limit value.

In Chapter 6, we establish several results for recursive games. We first consider recursive games with a countable state space and prove that if the family of n-stage value functions is totally bounded for the uniform norm, the uniform value exists. In particular, the uniform convergence of n-stage values implies the uniform convergence of λ-discounted values. Combined with a result in Rosenberg and Vieille (2000), we deduce a uniform Tauberian theorem for recursive games. Second, we use the existence result of uniform value to a class of the generalized models of repeated games and prove that the limit value and $Maxmin$ both exist and are equal. This class of repeated games are recursive games with signals where player 1 can always deduce the signal of player 2 from his own along the play.

Keywords

Dynamic optimization; optimal control; general evaluation; zero-sum repeated games; stochastic games; asymptotic analysis; uniform analysis; stochastic games with incomplete information on one side; limit value; uniform value; Tauberian theorem.

Contents

1 Introduction **9**
1.1 The models . . . 15
1.2 Main research interests . . . 16
1.3 Related literature . . . 17
1.4 Organization and main results of this dissertation . . . 19

2 Propriétés à long terme en optimisation dynamique **25**
2.1 Introduction . . . 26
2.2 Model . . . 27
2.3 Between Abel mean and Cesàro mean . . . 29
2.4 Asymptotic analysis . . . 36
2.5 Uniform analysis . . . 39
2.6 TV-uniform value in compact nonexpansive case . . . 42

3 Valeur limite pour le problème de contrôle optimal avec évaluations générales **51**
3.1 Introduction . . . 52
3.2 Preliminaries . . . 55
3.3 On the long-term condition (LTC) . . . 56
3.4 Main Result . . . 60
3.5 Proof of main result: Theorem 3.4.1 . . . 64
3.6 Concluding discussions . . . 67
3.7 Appendix . . . 69

4 Valeur uniforme pour le probème de contrôle optimal dans le cadre "compact non expansif" avec évaluations générales **71**
4.1 Introduction . . . 72
4.2 Preliminaries . . . 74
4.3 Proof of Theorem 4.1.7 . . . 80

5 Big match généralisé à information incompète d'un côté **87**
5.1 Introduction . . . 88
5.2 The model and the results . . . 90
5.3 Asymptotic analysis . . . 92
5.4 Uniform analysis: $Maxmin = \Lambda(p)$ . . . 99
5.5 Uniform analysis: $Minmax$ . . . 119
5.6 Appendix . . . 122

6 Jeux récursifs: valeur uniforme, théorème Taubérian et la conjecture de Mertens **129**
6.1 Introduction 130
6.2 Preliminaries: model and notations 132
6.3 Main results 134
6.4 Proofs 136
6.5 Application to recursive games with signals 146
6.6 Appendix 158

Bibliography **161**

Chapter 1

Introduction

Introduction (version française)

1. Les modèles

Nous étudions dans cette thèse plusieurs modèles d'optimisation dynamique, dont les problèmes de *contrôle optimal* (à un joueur et en temps continu) et *jeux répétés à somme nulle* (à deux joueurs et en temps discret).

Modèles d'interaction stratégique: différentes sous-classes de jeux répétés

Quand un seul jeu (à somme nulle) est répété, il n'y a pas d'autres considérations stratégiques entre les joueurs [1]. Dans la littérature de jeux répétés, il y a deux thèmes principaux concernant la variable d'état: l'aspect *dynamique* (*jeux stochastiques*, Shapley [53]) et l'aspect *informationnel* (*jeux répétés à information incomplètes*, Aumann et Maschler [5]).

Dans un jeu stochastique, après avoir observé les dernières actions et l'état actuel (l'état stochastique), les joueurs choisissent leurs actions simultanément. Leurs actions conjointes, ainsi que l'état actuel, induisent un paiement courrant et une probabilité de transition pour l'état de l'étape suivante. Un *jeu absorbant* est un jeu stochastique où tous les états sauf un sont absorbants [2]. Un *jeu récursif* est un jeu stochastique où le paiement est toujours zéro avant l'absorption.

Dans le modèle de jeux répétés à information incomplète d'un côté, le jeu matrice (l'état de la nature) est choisi (et fixé) selon une loi de probabilité. La réalisation de l'état est communiquée à un seul joueur. Pendant le jeu, les joueurs n'observent pas les paiements mais seulement les actions.

Pour combiner à la fois l'aspect dynamique et l'aspect informationnel, on définit le modèle des *jeux stochastiques à information incomplète d'un côté*: un jeu stochastique est choisi selon une loi de probabilité, et est communiqué à un seul joueur. Les actions des joueurs et l'état stochastique sont publics, mais les paiements ne le sont pas.

Un *modèle général de jeux répétés* est proposé par Mertens [33] (cf. Mertens *et al.* [35]): à chaque étape, les joueurs choisissent leurs actions en fonction de leur informa-

1. Cela ne veut pas le cas si le jeu à jouer est à somme non nulle car répétitions permettent coopération potentielle (voir par exemple le jeu "dilemme du prisonnier").

2. Un état est absorbant dans un jeu stochastique s'il ne change plus dès lors qu'il est atteint.

tion privée, qui génèrent un paiement courant et une probabilité de transition pour l'état. Pendant le jeu, les joueurs reçoivent des signaux privés concernant les actions jouées et l'états, qui sont pas publics en général. Dans un *jeu répété avec un joueur plus informé que l'autre*, il y a un joueur (plus informé) qui peut, à chaque étape, déduire le signal de son adversaire (moins informé) de son propre signal. *"Un contrôleur informé"* se réfère au cas où le joueur moins informé n'a aucune influence sur la transition de l'état.

Modèles à un seul joueur: programmation dynamique et contrôle optimal

Quand il y a un seul joueur dans le problème d'optimisation dynamique, un jeu stochastique se révèle être un processus de décision markovien (ou une chaîne de Markov contrôlée). Un problème de programmation dynamique est un modèle d'optimisation dynamique à un joueur avec transitions déterministes et en temps discret. Dans un problème de contrôle optimal, il y a un seul joueur (le contrôleur) qui choisit ses actions en temps continu, et l'état évolue selon un système différentiel contrôlé.

2. Le centre d'intérêt de la thèse

Nous sommes intéressés par les propriétés à long terme de problèmes d'optimisation dynamiques en temps discret et en temps continu, i.e. l'existence et la caractérisation de la valeur limite lorsque l'horizon tend vers l'infini, ou lorsque le facteur escompté tend vers zéro (*analyse asymptotique*), et en outre, l'existence de stratégies/contrôles ε-optimaux pour les joueurs à garantir la valeur limite dans tous les problèmes avec suffisamment grands horizons (*analyse uniforme*).

1) Traditionnellement, le flux du paiement dans un jeu répété et le coût de la trajectoire d'un problème de contrôle optimal sont évalués par des moyennes de Cesàro ou des moyennes d'Abel. Comme une approche générale, le paiement ou le coût peuvent être évalué par une mesure de probabilité (une *évaluation*) au fil du temps (sur $\mathbb{N}^*$ les nombres entiers positifs en temps discret et $\mathbb{R}_+$ les nombres réels positifs en temps continu). La valeur limite et la valeur uniforme sont définies de la même façon pour une suite d'évaluations qui induisent une durée moyenne du problème tendant vers l'infini.

En temps discret, la notion suivante est utilisée pour définir la régularité d'une mesure de probabilité (évaluation) (cf. Sorin [59]):

Definition *La* ***variation totale*** *d'une mesure de probabilité* $\theta = (\theta_t)_{t\geq 1}$ *sur* $\mathbb{N}^*$ *est:* $TV(\theta) = \sum_{t\geq 1} |\theta_t - \theta_{t+1}|$.

Renault [46] a fourni des conditions suffisantes pour l'existence d'une valeur limite dans les problèmes de programmation dynamique avec évaluations générales dans le sens suivant: soit v_θ la valeur du problème associée à une évaluation θ, alors v_θ converge uniformément (vers certaine fonction) losque $TV(\theta)$ tend vers zéro. Renault et Venel [47] ont étudié l'existence d'une valeur uniforme dans des problèmes de programmation dynamique avec des évaluations générales tel qu'il existe des contrôles ε-optimaux pour le joueur à garantir la valeur limite dans tous les problèmes définis par une certaine évaluation θ avec $TV(\theta)$ suffisamment petite.

Un objectif de cette thèse est d'introduire une condition de régularité analogue pour les évaluations en temps continu, et l'utiliser enusuite pour l'analyse asymptotique et l'analyse uniforme dans les problèmes de contrôle optimal avec évaluations générales.

2) Quand il n'y a pas de valeur pour l'analyse uniforme (la valeur uniforme), on définit

et étudie l'existence du $Maxmin$ du jeu infinitement répété, qui est le montant maximal qui pouvent être uniformément garanti par le joueur 1 et est, en même temps, le montant minimal qui pouvent être uniformément défendu par le joueur 2. Le $Minmax$ est définie de façon duale.

Mertens ([33]) a conjecturé que dans un modéle général de jeux répétés avec un joueur (joueur 1 le maximiseur) plus informé que l'autre, $Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$, i.e. la valeur limite et le $Maxmin$ existent et sont égaux. Ce résultat a été établi pour
– certaines classes de jeux absorbants à information incomplète d'un côté: Sorin [55], Sorin [56], Sorin and Zamir [61];
– les jeux récursifs à information incomplète d'un côté: Rosenberg and Vieille [52];
– les jeux répétés avec un contrôleur informé: Rosenberg *et al.* [50], Renault [45], Gensbittel *et al.* [21].

Récemment, Ziliotto [68] a construit un contre-exemple (un jeu répété des signaux symétriques), ce qui prouve que la conjecture est fausse en général. Il est maintenant une tâche difficile, celle d'identifier les sous-classes de jeux répétés pour lesquels la conjecture de Mertens est vraie. Le deuxième objectif de cette thèse vise à étendre les résultats positifs existants à plusieurs sous-classes.

3. La littérature liée à cette problèmatique

Programmation dynamique et contrôle optimal: le cas à un joueur

Motivé par l'étude de jeux répétés avec un contrôleur informé (Renault [45]), Renault [44] a fourni des conditions suffisantes pour l'existence de la valeur limite et la valeur uniforme dans les problèmes de programmation dynamique avec un espace de l'états arbitraire. En corollaire, on obtient que dans un problème de programmation dynamique défini sur un espace d'état compact et dont la correspondance de transition satisfait une certaine condition de non-expansivité, la valeur uniforme existe.

Quincampoix et Renault [42] ont obtenu des résultats analogues à Renault [44] en temps continu: la valeur uniforme existe dans tout problème de contrôle optimal défini sur un domaine invariant compact et dont la dynamique contrôlée satisfait une certaine condition de non-expansivité.

Renault [46] a généralisé l'analyse asymptotique de Renault [44] aux problèmes avec évaluations générales. L'analyse uniforme pour plusieurs modèles d'optimisations dynamiques avec des évaluations générales et en temps discret sont obtenus dans Renault et Venel [47]. Pour les problèmes de programmation dynamiques définis sur un espace d'état compact et satisfaisant une certaine condition de non-expansivité, Renault et Venel [47] ont prouvé que: le joueur a des stratégies ε-optimales (éventuellement aléatoires) pour le joueur à garantir la valeur limite dans tous les problèmes où le flux du paiment est évalué par une évaluation θ avec $TV(\theta)$ assez petit.

Jeux absorbants à information incomplète d'un côté

Le Big match (désormais BM) est le premier exemple non trivial (à la Blackwell et Ferguson [13]) de jeux absorbants pour lesquels la valeur uniforme existe. Sorin [55] a étudié le BM à information incomplète d'un côté (type I), où le joueur 1 est le joueur informé et a une action en renforçant l'absorption et l'autre en conservant l'état non-absorbant. Pour cette classe de jeux, $Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$, et le $Minmax$

existe, mais la valeur uniforme ne existe pas [3].

Pour les jeux absorbants à information incomplète d'un côté, Rosenberg [48] a utilisé *l'approche d'opérateur* (cf. Rosenberg et Sorin [51]) pour prouver l'existence de la valeur limite, i.e. pour en déduire l'unicité du point d'accumulation de (v_λ) ou (v_n) défini par l'opérateur de Shapley. Aucun résultat général est encore établi pour l'analyse uniforme de cette classe de jeux.

Jeux récursifs à information incomplète d'un côté

Pour les jeux récursifs à information incomplète d'un côté, Rosenberg et Vieille [52] ont prouvé que $Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$. Pour obtenir ce résultat, ils ont montré que le joueur 1 garantit tout point d'accumulation w de (v_λ) et que, sachant la stratégie du joueur 1, joueur 2 peut calculer l'état auxiliaire (sa croyance sur l'état de la nature) pour défendre w.

Jeux répétés avec un joueur plus informé que l'autre

Dans un modèle général de jeux répétés, il est possible de définir un jeu stochastique auxiliaire avec les croyances des joueurs (ses types dans "l'espace universel des croyances", cf. Coulomb [17]) comme variable d'état. Mertens [33] a conjecturé que dans un jeu répété où le maximiseur est toujours plus informé que le minimiseur, si ce jeu stochastique auxiliaire associé a une valeur uniforme, alors le $Maxmin$ existe dans le jeu original et est égal à cette valeur.

Renault [45] a considéré une sous-classe de ce modèle général, les jeux répétés avec un contrôleur plus informé, i.e. le joueur 1 est toujours informé l'état et du signal du joueur 2 (y compris en particulier les coups du joueur 2) et l'évolution de l'état ne dépend pas de l'action du joueur 2. Pour ce modèle, Renault [45] a prouvé que la valeur uniforme existe, un résultat plus fort que la conjecture de Mertens.

Gensbittel *et al.* [21] ont étendu Renault [45] pour une configuration plus générale telle que la valeur uniforme existe si les information du joueur 1 est plus précis (mais peut contenir ni l'état ni les coups du joueur 2) que celui du joueur 2 et l'évolution de la croyance du joueur 2 est contrôlée uniquement par le joueur 1.

4. Plan et résultats principaux de cette thèse

Ce manuscrit est divisé en deux parties principales.
La première partie (contrôle optimal) contient trois chapitres:
– Le chapitre 2 est un chapitre de l'enquête sur l'optimisation dynamique;
– Le chapitre 3 étudie l'analyse asymptotique en contrôle optimal avec des évaluations générales;
– Le chapitre 4 concerne l'analyse uniforme en contrôle optimal avec des évaluations générales.

La deuxième partie (jeux répétés) contient deux chapitres:
– Le chapitre 5 est consacré à l'étude du Big match généralisé à information incomplète d'un côté;
– Le chapiter 6 contient plusieurs résultats sur les jeux récursifs.

3. Même si pour les deux classes, jeux stochastiques (Mertens et Neyman [34]) et jeux répétés à information incomplète d'un côté (Aumann et Mascheler [5]), la valeur uniforme existe.

Part I: Contrôle optimal

Le chapitre 2: Propriétés à long terme en optimisation dynamique

Ceci est une présentation (le chapitre 2) de la littérature récente sur les propriétés à long terme des problèmes d'optimisation dynamique. Nous nous concentrons sur les problèmes de programmation dynamique, le modèle d'un joueur avec transitions déterministes et en temps discret. Des commentaires détaillés sont effectués pour des extensions en temps continu (contrôle optimal) par une comparaison des techniques et des résultats dans le cadre en temps discret. Nous insistons sur l'approche générale, i.e. l'analyse asymptotique et l'analyse uniforme pour le probléme d'optimisation dynamique où les paiements sont évalués par des mesures de probabilité générales. Certaines applications aux processus de décision markovien (avec observation standard ou avec des observations partielles sur l'état), les jeux répétés avec un contrôleur informé sont également discutés.

Les chapitres 3-4: Contrôle optimal avec des évaluations générales

Dans les deux chapitres suivants, nous considérons le problème de contrôle optimal où le coût de la trajectoire est évalué par une mesure générale de probabilité sur $\mathbb{R}_+$. Pour étudier les propriétés de contrôle optimal à long terme, nous introduisons la condition de régularité asymptotique suivante pour une suite d'évaluations qui induisent une durée moyenne du problème tendant vers l'infini.

Definition *Soit $\theta \in \Delta(\mathbb{R}_+)$ une mesure de probabilité sur $\mathbb{R}_+$. Pour tout $s \geq 0$, sa* ***s-variation totale*** *est: $TV_s(\theta) = \max_{Q \in \mathcal{B}(\mathbb{R}_+)} |\theta(Q) - \theta(Q+s)|$. Une suite d'évaluations $(\theta^k)_{k\geq 1}$ satisfait la* ***condition à long terme*** *si: $\sup_{0\leq s\leq 1} TV_s(\theta^k) \longrightarrow_{k\to\infty} 0$.*

Le résultat principal du chapitre 3 est le suivant. Pour toute suite $(\theta^k)_{k\geq 1}$ satisfaisant la condition à long terme, soit $\{V_{\theta^k} : k \geq 1\}$ la famille associée des fonctions valeur pour le problème de contrôle. $(V_{\theta^k})_{k\geq 1}$ converge alors uniformément si et seulement si la famille $\{V_{\theta^k} : k \geq 1\}$ est totalement bornée pour la norme uniforme. En outre, il y a une unique fonction limite V^* en cas de convergence des différentes suites satisfaisant la condition à long terme; une caractérisation de la fonction valeur V^* est également fournie, qui dépend en général de l'état initial (contrairement à la plupart des résultats dans la littérature qui supposent conditions ergodiques, cf. Alvarez and Bardi [1], Arisawa [2], Arisawa and Lions [3], Bensoussan [9], Gaitsgory [20], etc.).

En corollaire, on en déduit que: dans un problème de contrôle optimal qui est "compact non expansif", i.e. le problème 1) est définie sur un domaine invariant compact; 2) a une fonction de coût qui est continu en la variable d'état et ne dépend pas de la variable de contrôle; 3) satisfait une certaine condition non expansif, $||V_\theta - V^*||_\infty$ tend vers zéro losque $\sup_{0\leq s\leq 1} TV_s(\theta)$ tend vers zéro. Cela généralise l'analyse asymptotique de Quincampoix et Renault [42] qui traite des moyennes de Cesàro.

Dans le chapitre 4, nous continuons avec l'analyse uniforme pour des problèmes de contrôle optimal "compact non expansif" où le coût de la trajectoire est évalué par des mesures de probabilité générales. On obtient l'existence de la valeur uniforme dans cette classe de problèmes, i.e. le contrôleur a des contrôles ε-optimaux (éventuellement aléatoires) pour le joueur à garantir V^* dans tous les problèmes de contrôle où le coût de la trajectoire est évalué par une mesure $\theta \in \Delta(\mathbb{R}_+)$ avec $\sup_{0\leq s\leq 1} TV_s(\theta)$ suffisamment petite. Cela généralise l'analyse uniforme dans Quincampoix et Renault [42] qui traite des

moyennes de Cesàro.

Partie II: Jeux répétés

Le chapitre 5: Big match généralisé à information incomplète d'un côté

Dans le chapitre 5, nous étudions une sous-classe de jeux absorbants, nommé Big match généralisé" (désormais GBM), qui prend la même forme que le Big match (type I), sauf que lorsque le joueur 1 joue l'action en renforçant l'absorption, la probabilité d'absorption est strictement positif, mais peut ne pas nécessairement être 1. Pour GBM à information incomplète d'un côté (type I), nous généralisons résultats de Sorin [55]. Le résultat $Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$ est obtenu et la valeur limite est caractérisée par la valeur d'un "jeu limite" auxiliaire. Nous obtenons aussi l'existence du $Minmax$ caractérisé par un second jeu auxiliaire.

Notre résultat dans l'analyse asymptotique n'est pas une conséquence de Rosenberg [48] pour deux raisons: nous considérons une probabilité de transition dépendant de l'état , ce qui n'est pas le cas dans le modèle étudié par Rosenberg [48]; nous considérons le flux du paiement du jeu répété évalué par des mesures de probabilité générales, et nous prouvons que la fonction valeur converge lorsque le poids maximal d'évaluation sur chaque étape tend vers zéro.

Le chapitre 6: Jeux récursifs

Ce chapitre contient plusieurs résultats concernant les jeux récursifs et est divisé en deux parties.

Le résultat principal dans la première partie est que pour un jeu récursif avec espace infini d'état, à condition que la famille des valeurs de jeux à n étapes soit totalement bornée pour la norme uniforme, la valeur uniforme existe. En particulier, la convergence uniforme de (v_n) implique la convergence uniforme de (v_λ). À d'aide d'un résultat inversé dans Rosenberg et Vieille [52], on en déduit un théorème taubérien pour les jeux récursifs.

Dans la deuxième partie, nous utilisons le résultat d'existence de la valeur uniforme pour le modèle général de jeux récursifs avec un joueur plus informé que l'autre.

– Pour ce faire, nous définissons d'abord un jeu récursif auxiliaire avec la croyance de second ordre (sur la croyance du joueur 1 sur l'état) du joueur 2 comme variable d'état.

– Ensuite, nous utilisons les résultats généraux établis dans Gensbittel *et al.* [21] pour déduire que la famille des fonctions valeur de jeux auxiliaires à n étapes est totalement bornée, elle a donc une valeur uniforme.

– Enfin, nous montrons que le joueur 1 peut garantir la valeur uniforme en imitant des stratégies ε-optimales dans le jeu auxiliaire; pour le joueur 2 à défendre la valeur uniforme, nous construisons un deuxième jeu auxiliaire qui a une même valeur uniforme que la première, et des stratégies ε-optimales dans ce jeu fournissant le joueur 2 réponses optimales uniformes dans le jeu répété original.

Introduction (English version)

1.1 The models

We study in this dissertation several models of dynamic optimization, including *optimal control problems* (one-player and in continuous time) and *zero-sum repeated games* (two-player and in discrete time).

Models of strategic interactions: different subclasses of repeated games

When a single stage game (zero-sum) is repeated, there are no further strategic considerations among players [4]. In the literature of repeated games, there are two main topics concerning the state variable: *dynamical* aspect (*stochastic games*, Shapley [53]) and the *informational* aspect (*repeated games with incomplete information*, Aumann and Mascheler [5]).

In a stochastic game, after observing the past actions and the current state (the stochastic state), players choose their actions simultaneously. Their joint actions, together with the current state, induce a current stage payoff and also a probability transition for the state in the next stage. An *absorbing game* is a stochastic game where all states but one are absorbing [5]. A *recursive game* is a stochastic game where the stage payoff is always zero before absorption.

In the model of repeated games with incomplete information on one side, the stage game (the state of the nature) is chosen (and kept fixed) according to some probability distribution. The realization of the state is communicated to one player only. During the play, players do not observe the stage payoffs but only the actions.

To combine both the dynamical aspect and the informational aspect, one defines the model of *stochastic games with incomplete information on one side*: one stochastic game among a family is chosen to be played according to some probability distribution, and is communicated to one player only. Both the stochastic state and players' actions are public, but not the stage payoffs.

A *general model of repeated games* is proposed by Mertens [33] (cf. Mertens *et al.* [35]): at each stage, players take actions according to their private information, which generate a stage payoff and also a probability transition for the state in the next stage. Along the play, players receive private signals concerning the actions and states, which are in general not public. In a *repeated game with one player more informed than the other*, there is one player (more informed) who can always deduce the opponent's (less informed) signal from his own at each stage. "*An informed controller*" refers to the case when the less informed player has no influence on the transition of the state.

Models with only one player: dynamic programming and optimal control

When there is a single player in the dynamic optimization problem, a stochastic game turns out to be a *Markovian decision process* (or *controlled Markovian chain*). *Dynamic programming* is a one-player dynamic optimization model with deterministic transition and

4. This is not the case if the stage game being played is non-zero sum since repetition enables potential cooperation (see for example the game of "Prisoner's Dilemma").

5. A state in a stochastic game is absorbing if it does not change any more once it is reached.

in discrete time. In an *optimal control problem*, there is a single player (the controller) who takes actions in continuous time, and the state evolves along a controlled differential system.

1.2 Main research interests

We are interested in the long-term properties in dynamic optimization problems both in discrete time and in continuous time, i.e. the existence and the characterization of the limit value when the horizon tends to infinity or the discount factor tends to zero (the *asymptotic analysis*), and further, the existence of approximately optimal strategies/controls for players to guarantee the limit value for all problems with sufficiently large horizon (the *uniform analysis*).

1) Traditionally, the payoff stream in a repeated game or the running cost in an optimal control problem is evaluated by Cesàro means or Abel means. As a general approach, the payoff stream or the running cost can be evaluated by any probability measure (*evaluation*) over time (on $\mathbb{N}^*$ the positive integers in discrete time and on $\mathbb{R}_+$ the positive real numbers in continuous time). The asymptotic value and the uniform value are defined in a similar way for a sequence of evaluations whose expected duration tends to infinity.

In the discrete time framework, the following notion is used to define the regularity of a probability measure (cf. Sorin [59]):

Definition 1.2.1. *The* ***total variation*** *of a probability measure* $\theta = (\theta_t)_{t\geq 1}$ *over* $\mathbb{N}^*$ *is:* $TV(\theta) = \sum_{t\geq 1} |\theta_t - \theta_{t+1}|$.

Renault [46] provided sufficient conditions for the existence of a limit value in dynamic progmming problems with general evaluations in the following sense: let v_θ be the value of the problem associated with any evaluation θ, then v_θ converges uniformly (to some value function) as $TV(\theta)$ tends to zero. Renault and Venel [47] studied the existence of a uniform value in dynamic progmming problems with general evaluations such that there exists approximately optimal controls to guarantee the limit value in all problems defined by some evaluation θ with $TV(\theta)$ sufficiently small.

One objective of this dissertation is to introduce an analogous regularity condition for evaluations in the continuous time framework, and then use it for the asymptotic analysis and the uniform analysis for optimal control problems with general evaluations.

2) When there is no value for the uniform analysis (the uniform value), one defines, and studies the existence [6] of, $Maxmin$ of the inifnitely repeated game, which is the maximal amount that can be uniformly guaranteed by player 1 and is at the same time the minimal amount that can be uniformly defended by player 2. $Minmax$ is defined in a dual way.

Mertens [33] conjectured that in a general model of repeated games with one player (player 1 the maximizer) more informed than the other, $Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$, i.e. both the limit value and $Maxmin$ exist and they are equal. This result has been established for

– some classes of absorbing games with one-sided incomplete information: Sorin [55], Sorin [56], Sorin and Zamir [61];
– recursive games with one-sided incomplete information: Rosenberg and Vieille [52];
– repeated games with an informed controller: Rosenberg *et al.* [50], Renault [45], Gensbittel *et al.* [21].

6. The existence of $Maxmin$ is not trival even though, to the best knowledge of the author, there has been no example yet explicitely proving that $Maxmin$ does not exist. The counterexample in Ziliotto [68] disproves Mertens' conjecture by a non existence of the limit value.

Recently, Ziliotto [68] constructed a conterexample (a repeated game with symmetric signals), proving that the conjecture is false in general. It is now a challenging task to identify the subclasses of repeated games for which Mertens's conjecture is true. The second object of this dissertation aims at extending the existing positive results to several larger subclasses.

1.3 Related literature

1.3.1 One-player case: dynamic programming and optimal control

Lehrer and Sorin [28] proved a Tauberian theorem in dynamic programming: (v_n) converges uniformly as n tends to infinity if and only if (v_λ) converges uniformly as λ tends to zero, and in case of convergence, both limits are the same. The analogous result in continuous time framework is obtaind in Oliu-Barton and Vigeral [40] for optimal control problems.

Motivated from the study of repeated games with an informed controller (Renault [45]), Renault [44] provided sufficient conditions for the existence of limit value and of uniform value in dynamic programming problems with an arbitrary state space. The main conditions are expressed as compactness of a family of value functions for some auxiliary problems. One corollary states that in a dynamic programming problem defined on a compact state space whose transition correspondence satisfies some nonexpansive condition, the uniform value exists.

Quincampoix and Renault [42] obtained the result analogue to Renault [44] in continuous time framework: the uniform value exists in any optimal control problem defined on a compact invariant domain whose controlled dynamic satisfies some nonexpansive condition. Being different from most results in the literature of optimal control which assume certain ergodic condition (cf. Alvarez and Bardi [1], Arisawa [2], Arisawa and Lions [3], Bensoussan [9], Gaitsgory [20], etc.), Quincampoix and Renault [42] provided a characterization for the limit value function which may depend on the initial state.

Renault [46] generalized the asymptotic analysis in Renaul [44] to problems with general evaluations. The uniform analysis for several models of dynamic optimizations with general evaluations and in discrete time framework are obtained in Renault and Venel [47]. For dynamic programming problems defined on a compact state space and satisfying some nonexpansive condition, Renault and Venel[47] proved that: the decision-maker (player) has ε-optimal play (may be random) to guarantee the limit value in all problems in which the payoff stream is evaluated by some θ with $TV(\theta)$ small enough.

1.3.2 Absorbing games with one-sided incomplete information

Big match (henceforth BM) is the first non-trivial example (due to Blackwell and Ferguson [13]) of absorbing games for which the uniform value is proven to exist. Sorin [55] studied BM with one-sided incomplete information (type I), where player 1 is the informed player and has one action enforcing the absorption and the other one keeping the state non-absorbing. For this class of games, $Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$ is true and $Minmax$ exists, but the uniform value does not exist [7].

To prove the result $Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$, Sorin [55] introduced an auxiliary "limit game" played on $[0,1]$, where player 1 (the informed player) chooses a family

7. Even though for both stochastic games (Mertens and Neyman [34]) and repeated games of incomplete information on one side (Aumann and Mascheler [5]), uniform value is proven to exist.

of stopping times to hit the enforcing action (of absorption) and player 2 plays a history independent strategy in continuous time, such that the value of the "limit game" characterizes both $Maxmin$ and the limit value of the repeated game. The idea of introducing an auxiliary game with reduced but "statistically sufficient" strategy sets for players to mimic the (optimal) plays in the repeated game appeared as early as Mertens and Zamir [37], which deals with "repeated games without a recursive structure" (see also Waternaux [67], Sorin [58]).

The difficulty in the proof of Sorin [55] is the existence of a pair of "equalizing" strategies in the "limit game" such that the payoff is constant on $[0, 1]$ which is equal to its value; moreover, player 1 can adapt this pair to the empirical frequence of player 2's actions such that the average payoff in the repeated game is equal to the value of the "limit game" if player 2 follows his strategy in the pair and is less otherwise.

In another paper, Sorin [56] studied BM with one-sided incomplete information (type II), where the informed player is still player 1 while the absorption is enforced by (one of the two actions of) player 2. For these games, the difficulty is the existence of $Minmax$, which is related to the *approachability* (cf. Blackwell [11]) in stochastic games with vector payoffs. Techniques in Sorin [55] and Sorin [56] are used in Sorin [57] to solve a class of repeated games with state-dependent signalling functions.

For absorbing games with one-sided incomplete information, Rosenberg [48] used the *operator approach* (cf. Rosenberg and Sorin [51]) to prove the existence of limit value, i.e. to deduce the uniqueness of the accumulation point of (v_λ) or (v_n) defined by Shapley operator. No general result is available yet for the uniform analysis of this class of games.

1.3.3 Recursive games with one-sided incomplete information

For recursive games with one-sided incomplete information, Rosenberg and Vieille [52] proved that $Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$. To obtain the result, they showed that player 1 guarantees any accumulation point w of (v_λ), and more, knowing player 1's strategy, player 2 can compute the auxiliary state (his belief over the state of the nature) so as to defend w.

The ε-optimal strategy in Rosenberg and Vieille [52] is constructed as an alternation between two types of strategies (similar construction appeared also in Solan and Vieille [54]): one is optimal in the "projective game" ($\lambda = 0$ in the Shapley equation which defines v_λ as its unique fixed point) for w; the other one is optimal in some λ-discounted game with $||w - v_\lambda||_\infty$ small.

Moreover, they showed in one example that $Maxmin \neq Minmax$. The existence of $Minmax$ is in general unknown.

1.3.4 Repeated games with one player more informed than the other

Merten's conjecture (programme) In a general model of repeated games, it is possible to define an auxiliary stochastic game with player's beliefs (their types in the "universal belief space", cf. Coulomb [17]) as the state variable. Mertens [33] conjectured that in a repeated game with the maximizer more informed than the minimizer, if the associated auxiliary stochastic game has a uniform value, then $Maxmin$ in the original game exists and is equal to this value. Briefly, Mertens' programme (conjecture) for this class of repeated games consists of the following three steps:

1) Define an auxiliary stochastic game where players' beliefs are the state variables;

2) Prove that this auxiliary stochastic game with infinite state space has a uniform value;

3) Show that both players can mimic ε-optimal strategies in the auxiliary game so as to guarantee or to defend the value of the auxiliary game in the original repeated game.

Repeated games with an informed controller Renault [45] considered a subclass of this general model, repeated games with an informed controller, that is, player 1 is always informed the state and player 2's signal (including in particular player 2's action) and the evolution of the state does not depend on the action of player 2. For this model, Renault [45] proved that the uniform value exists, a result stronger than Mertens' conjecture.

The proof of Renault [45] can be seen as an implementation of Mertens' programme: first, a reduction of the repeated game to a one-player dynamic programming problem (as if player 2 is playing a best resply at each stage) is made; second, the result in Renault [44] was used to obtain the existence of uniform value for the reduced problem; finally, it is possible for player 1 to transform any ε-optimal play in the dynamic programming problem to an ε-optimal strategy in the original repeated game. As for player 2, using the fact that the transition is independent of his action, one deduces that the uniform value of the reduced problem can be guaranteed by him by splitting the play into blocks.

Gensbittel *et al.* [21] extended Renault [45] to a more general setup such that the uniform value exists if player 1's information is more accurate (but may contain neither the state nor the actions of player 2) than player 2's and the evolution of player 2's belief is uniquely controlled by player 1.

1.4 Organization and main results of this dissertation

This manuscript is divided into two main parts.
The first part (optimal control) contains three chapters:
– Chapter 2 is a survey chapter on one-player dynamic optimization;
– Chapter 3 studies the asymptotic analysis in optimal control with general evaluations;
– Chapter 4 concerns the uniform analysis in optimal control with general evaluations.
The second part (repeated games) contains two chapters:
– Chapter 5 is devoted to the study of generalized Big match with one-sided incomplete information;
– Chpater 6 contains several results on recursive games.

1.4.1 Part I: optimal control

Chapter 2: A survey article on long-term properties in dynamic optimization

This chapter contains a survey article of recent literature on the long-term properties in one-player dynamic optimization problems. We focus on the deterministic dynamic programming in discrete time. Extensive comments are also made on extensions to continuous time framework (optimal control) by a comparison of the techniques and results in discrete time framework. We emphasize the approach for the asymptotic analysis and for the uniform analysis defined by general probability measures. Some applications to Markovian decision process (with standard observation or with partial observations of the state), repeated games with an informed controller are also discussed.

Chapter 3-4: Optimal control with general evaluations

We consider in the following two chapters optimal control problems where the running cost is evaluated by a general probability measure over $\mathbb{R}_+$. To study the associated long-

term properties of the control problem, we introduce the following asymptotic regularity condition for a sequence of evaluations whose expected duration tends to infinity.

Definition 1.4.1. *Let $\theta \in \Delta(\mathbb{R}_+)$ be a probability measure over $\mathbb{R}_+$. For any $s \geq 0$, its* ***s-total variation*** *is: $TV_s(\theta) = \max_{Q \in \mathcal{B}(\mathbb{R}_+)} |\theta(Q) - \theta(Q+s)|$. A sequence of evaluations $(\theta^k)_{k\geq 1}$ satisfies the* ***long-term condition*** *if: $\sup_{0\leq s\leq 1} TV_s(\theta^k) \longrightarrow_{k\to\infty} 0$.*

Our main result in Chapter 3 is as follows. For any sequence $(\theta^k)_{k\geq 1}$ satisfying the long-term condition, let $\{V_{\theta^k} : k \geq 1\}$ be the associated family of value functions for the control problem. Then $(V_{\theta^k})_{k\geq 1}$ converges uniformly if and only if the family $\{V_{\theta^k} : k \geq 1\}$ is totally bounded for the uniform norm. Moreover, there is a unique limit function V^* in case of convergence for different sequences satisfying the long-term condition; a characterization of the value function V^* is also provided which in general depends on the initial state (being different from most results in the literature assuming ergodic condition).

As a corollary, we deduce that: in a *compact nonexpansive* optimal control problem, i.e. the problem 1) is defined on a compact invariant domain; 2) has a running cost function that is continuous in the state variable and does not dependent on the control variable; 3) satisfies certain nonexpansive condition, $||V_\theta - V^*||_\infty$ tends to zero as $\sup_{0\leq s\leq 1} TV_s(\theta)$ vanishes. This generalizes the asymptotic analysis in Quincampoix and Renault [42] which deals with Cesàro means.

In Chapter 4, we continue with the uniform analysis for compact nonexpansive optimal control problems associated to general evaluations. We obtain the existence of uniform value in this class of problems, that is, the controller has ε-optimal controls (may be random) to guatantee V^* for all control problems in which the running cost is evaluated by some $\theta \in \Delta(\mathbb{R}_+)$ with $\sup_{0\leq s\leq 1} TV_s(\theta)$ sufficiently small. This generalizes the uniform analysis in Quincampoix and Renault [42] which deals with Cesàro means.

1.4.2 Part II: Repeated games

Chapter 5: Generalized Big match with one-sided incomplete information

In Chapter 5, we study a subclass of absorbing games, named "generalized Big match" (hence forth GBM), which takes the same form of Big match (type I) except that when player 1 plays the enforcing action (of absorption), the absorbing probability is strictly positive but may not necessarily be one. For GBM with one-sided incomplete information (type I), we generalize results in Sorin [55]. The result $Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$ is obtained and the limit value is characterized by the value of an auxiliary "limit game". We obtain also the existence of $Minmax$ characterized by a second auxiliary game.

Even though the enforcing action does not induce a probability one of absorption as in BM, by playing a bounded number (denoted by M) of times this action, the absorption will occur with a probability close to one. In the asymptotic analysis, we define a sequence of auxiliary games (discretization of some "limit game") with reduced strategies such that this number M is a state variable in the auxiliary game. To obtain the result, we show that the optimal strategies in the auxiliary games give asymptotic optimal strategies for both players in the original repeated game.

In the uniform analysis of $Maxmin$, we first establish properties in the "limit game" played on $[0, 1]$ associated with some BM with one-sided incomplete information, and derive the corresponding ε-optimal strategies (in a generic form) in the repeated game. This is similar to Sorin [55], while the difference is that the BM used to define the "limit game" in our case takes a generalized form such that an absorption leads to some absorbing

state rather than some absorbing payoff. Next, we iterate M times the auxiliary ε-optimal strategies that are in generic form to define a behavior strategy in the repeated game.

To show that the iterated strategy is ε-optimal, we make an inductive analysis on the number M. Indeed, at the last time of playing the enforcing action, the expected probability of absorption is almost one, thus it corresponds to BM with absorbing payoffs; at each other time of playing the enforcing action, we are in the situation of BM with absorbing states.

Our result in the asymptotic analysis is not implied by Rosenberg [48] for two reasons: we consider the transition probability to be state-dependent, which is not the case in Rosenberg [48]; we consider the payoff stream of the repeated game evaluated by a general probability measure, and prove that the value function converges as the weight of the measure on each stage tends to zero.

Chapter 6: Recursive games

This chapter contains several results on recursive games and is divided into two parts.

The main result in the first part is that for a recursive game with infinite state space, provided that the family of the n-stage values is totally bounded for the uniform norm, the uniform value exists. In particular, the uniform convergence of (v_n) implies the uniform convergence of (v_λ). Together with a reversed result in Rosenberg and Vieille [52], we deduce a Tauberian theorem for recursive games.

To prove the existence of uniform value, we use v the point-wise limit superior function of (v_n) as the target function for player 1 (dual for player 2).

– We first prove that in a *positive-valued recursive game* (for some $M > 0$: at any non-absorbing state x, there is a uniformly bounded stage number $n(x)$ such that $v_{n(x)}(x) \geq M$), player 1 has strategies σ^ε to guarantee the value $x \mapsto v_{n(x)}(x)$ and to enforce an absorption in finite time (*uniformly terminating*).

– Next we make a reduction of any recursive game to an auxiliary positive-valued one by turning any non-absorbing state x to an absorbing one for which $v(x) < \varepsilon$.

– Then the ε-optimal strategy is constructed as the alternation between two types: to play σ^ε as long as $v(x_n) \geq 2\varepsilon$; and shift to play an optimal strategy in the "projective game" defined by v as long as $v(x_n)$ drops down to 0.

We use the uniformly terminating property of σ^ε and the fact that the stage payoff is zero on non-absorbing states to derive the result.

In the second part, we use the existence result of uniform value to implement Mertens' programme for a general model of recursive games with one player more informed than the other.

– For this aim, we first define an auxiliary recursive game with player 2's second order belief (over player 1's belief over the true state) as the state variable.

– Next, we use the general results established in Gensbittel *et al.* [21] to deduce that the n-stage values in the auxiliary game is totally bounded, thus it has a uniform value.

– Finally we show that player 1 can guarantee the uniform value by mimicking ε-optimal strategies in the auxiliary game; for player 2 to defend the uniform value, we construct a second auxiliary game which has a same uniform value as the first one, and the ε-optimal strategies in it provides player 2 uniform optimal replies in the original repeated game.

Première partie

Contrôle optimal

Chapter 2

Propriétés à long terme en optimisation dynamique

Ce chapitre est issu de l'article d'enquête *Long-term properties in dynamic optimization*.

Long-term properties in dynamic optimization

2.1 Introduction

Dynamic optimization models decision making in a dynamic environment: a decision-maker takes action time after time, which induces a current reward/payoff and also a transition for the next state. Depending on the dynamic system being *discrete* or *continuous*, and being *deterministic* or *stochastic*, different optimization models have been developped for study (Bellman [7], Bellman [8], Blackwell [12], Pontryagin *et al.* [41], etc.), including for example *(deterministic) dynamic programming*, *Markovian decision process*, *optimal control*, *stochastic optimal control*.

This chapter studies the long-term properties of dynamic optimization. We are interested in the asymptotic behavior of the values (the *asymptotic analysis*) or the existence of robust (approximately) optimal strategies (the *uniform analysis*) in different optimization models where the expected duration of the problem is large. Traditionally [1], we consider the optimization problem associated with an average payoff (Cesàro mean) for a fixed horizon and then let the horizon tend to infinity, or the optimization problem associated with a discounted payoff (Abel mean) for a fixed discount factor and then let the discount factor tend to zero. We also study the model where the payoff stream is evaluated by a family of probability measures on the positive integers such that the weight of the measure on each stage becomes negligible.

We focus on the model of deterministic dynamic programming in discrete time. We comment also extensively on extensions of the results to models with stochastic transition (Markovian decision process, gambling house) and models in continuous time (optimal control). Some applications to Markovian decision process with partial observations (POMDP) and zero-sum repeated games with an informed controller will also be discussed.

The organization of this chapter is as follows. Section 2 describes the model of dynamic programming, and introduces the value notions we are going to study. Section 3 compares different notions of asymptotic analysis (Tauberian theorem) or uniform analysis (uniform value and Blackwell optimality). Section 4 concerns the asymptotic analysis, and Section 5 contains sufficient conditions for the existence of uniform values. In the last section we focus on a specific model, the "compact nonexpansive" case, for which a very strong notion of uniform value is studied (associated with general probability measures). At the end, we look at the applications to POMDP and to repeated games. Extensions to continuous time framewrok (optimal control) will be also discussed in each section.

1. There are also other ways to evaluate the infinite flow payoff, for example, the limit inferior or the limit superior of the Cesàro means. These approaches are not discussed here.

2.2 Model

We consider dynamic programming (henceforth DP) problems with *arbitrary state space* and bounded rewards.

The formal description of the model is as follows. Let Z be a non empty set of states, $F : Z \rightrightarrows Z$ is a non empty valued transition correspondence, and $r : Z \to [0,1]$ is the reward function. Starting at the initial state z_0 in Z, the decision-maker chooses some new state $z_1 \in F(z_0)$, realizing a stage payoff $r(z_1)$. Then he chooses again some new state $z_2 \in F(z_1)$, realizing a stage payoff $r(z_2)$, *etc* ... The DP problem is summarized as $\Gamma = \langle Z, F, r\rangle$. We write $\Gamma(z_0)$ for a DP problem with an initial state z_0.

Definition 2.2.1. *A **play** at z_0 is a sequence $s = (z_1, ..., z_t, ...)$ in Z^∞ such that $z_{t+1} \in F(z_t)$ for each $t \geq 0$. Denote by $S(z_0)$ the set of plays at z_0.*

*A play $s = (z_t)_{t\geq 1} \in S(z_0)$ is **stationary** if there is a function $f : Z \to Z$ (a selection of F) such that $z_{t+1} = f(z_t)$ for all $t \geq 0$.*

Let $\Delta(\mathbb{N}^*)$ be the set of probability distributions over $\mathbb{N}^* = \{1, ..., t, ...\}$ the set of positive integers. Any $\theta = (\theta_t)_{t\geq 1} \in \Delta(\mathbb{N}^*)$ is an *evaluation* (for the payoff stream), where θ_t is the weight of stage t.

Definition 2.2.2. *Given any $\theta = (\theta_t)_{t\geq 1} \in \Delta(\mathbb{N}^*)$, the **$\theta$-value** at z_0 is*

$$v_\theta(z_0) = \sup_{s=(z_t)\in S(z_0)} \gamma_\theta(s), \text{ where } \gamma_\theta(s) = \sum_{t\geq 1} \theta_t r(z_t).$$

Particular cases of evaluations and their corresponding value functions include:

n-stage average evaluation (Cesàro mean): $\forall n \in \mathbb{N}^*$, $\bar{\theta}^n = (\frac{1}{n}, ..., \frac{1}{n}, 0, ...)$, and the n-stage value at z_0 is:

$$v_n(z_0) = \sup_{s\in S(z_0)} \gamma_n(s), \quad \text{where } \gamma_n(s) = \frac{1}{n}\sum_{t=1}^{n} r(z_t);$$

λ-discounted evaluation (Abel mean): $\forall \lambda \in (0,1]$, $\bar{\theta}^\lambda = \left(\lambda(1-\lambda)^{t-1}\right)_{t\geq 1}$, and the λ-discounted value at z_0 is:

$$v_\lambda(z_0) = \sup_{s\in S(z_0)} \gamma_\lambda(s), \quad \text{where } \gamma_\lambda(s) = \sum_{t\geq 1} \lambda(1-\lambda)^{t-1} r(z_t).$$

The asymptotic analysis (the value approach) concerns the asymptotic behavior of v_θ as the weight of each stage under θ becomes negligible. When $(\theta_t)_{t\geq 1}$ is non-increasing in t, it is natural to take θ_1 tending to zero. In particular, we study the convergence of (v_n) as n tends to infinity or the convergence of (v_λ) as λ tends to zero.

For evaluations not necessarily decreasing, the following example shows that the asymptotic analysis under the convergence condition "$\sup_{t\geq 1} \theta_t \longrightarrow 0$" is in general too weak to have the convergence of v_θ.

Example 1. *Consider an un-controlled deterministic process alternating beween two states, 0 and 1, with a payoff stream $(0, 1, 0, 1, ...)$. Let (μ^k) and (η^k) be two sequence of evaluations with $\mu^k = \frac{1}{k}\sum_{t=1}^{2k} \mathbb{1}_{t\in 2\mathbb{N}}$ and $\eta^k = \frac{1}{k}\sum_{t=1}^{2k} \mathbb{1}_{t\in 2\mathbb{N}+1}$, $\forall k \geq 1$. We have $v_{\mu^k} = 1$ and $v_{\eta^k} = 0$ for all $k \geq 1$, even though both sequences of evaluations have vanishing stage weights.*

For the asymptotic analysis under general evaluations, the following notion is introduced in the literature (cf. Sorin [59] p.105).

Definition 2.2.3. *The **total variation** of any evaluation $\theta \in \Delta(\mathbb{N}^*)$ is*

$$TV(\theta) = \sum_{t\geq 1} |\theta_{t+1} - \theta_t|.$$

Definition 2.2.4. *Γ has a **limit value** v if for all $z_0 \in Z$, $\lim_{n\to\infty} v_n(z_0) = \lim_{\lambda\to 0} v_\lambda(z_0) = v(z_0)$, and both convergences are uniform in z_0.*

*Γ has a **TV-limit value** v if for any sequence of evaluations (θ^k) with $TV(\theta^k) \longrightarrow_{k\to\infty} 0$, (v_{θ^k}) converges uniformly (on Z) to v as k tends to infinity.*

*More generally, let $\Theta \subseteq \Delta(\mathbb{N}^*)$ be a subset of evaluations and $\varphi : \Theta :\to \mathbb{R}$, Γ has a (Θ, φ)**-limit value** v if for any sequence (θ^k) in Θ with $\varphi(\theta^k) \longrightarrow_{k\to\infty} 0$, (v_{θ^k}) converges uniformly (on Z) to v as k tends to infinity.*

Below is an equivalent description of the (Θ, φ)-limit value.

Lemma 2.2.5. *For any (Θ, φ) be given. Γ has a (Θ, φ)-limit value v if and only if there is (Θ, φ)-uniform convergence of $\{v_\theta\}$ to v, i.e,*

$$\forall \varepsilon > 0, \exists \alpha > 0, \forall \theta \in \Theta, s.t.\ \varphi(\theta) \leq \alpha,\ |v_\theta(z_0) - v(z_0)| \leq \varepsilon,\ \forall z_0 \in Z.$$

Proof. One direction is evident. Suppose now that Γ has a (Θ, φ)-limit value v but there is no (Θ, φ)-uniform convergence of $\{v_\theta\}$ to v, i.e. $\exists \varepsilon_0 > 0, \forall \alpha > 0, \exists \theta \in \Delta(\Theta),\ s.t.\ \varphi(\theta) \leq \alpha$ and $\sup_{z_0\in Z} |v_\theta(z_0) - v(z_0)| > \varepsilon_0$. Fixing $\varepsilon_0 > 0$ as above, we consider a vanishing positive sequence (α_k). Then there exists a sequence of evaluations (θ^k) such that $\varphi(\theta^k) \leq \alpha_k \longrightarrow_{k\to\infty} 0$ while $\sup_{z_0\in Z} |v_{\theta^k}(z_0) - v(z_0)| \geq \varepsilon_0 > 0,\ \forall k \geq 1$. This contradicts the fact that v is the (Θ, φ)-limit value of Γ. □

The uniform analysis (the strategy approach) studies a stronger notion of values (than the asymptotic analysis). It asks for the existence of approximately optimal play for all n-stage problems with n large enough (or, for all θ-evaluated problems with $TV(\theta)$ small enough).

Definition 2.2.6. *Let v be the limit value of Γ. Then Γ has a **uniform value** if the decision-maker uniformly guarantees v, i.e.,*

$$\forall \varepsilon > 0, \exists n_0 > 0, \forall z_0 \in Z, \exists \sigma^\varepsilon \in S(z_0),\ s.t.\ \gamma_n(\sigma^\varepsilon) \geq v(z_0) - \varepsilon, \forall n \geq n_0.$$

*Let v be the TV-limit value of Γ. Then Γ has a **TV-uniform value** v if the decision-maker uniformly guarantees the TV-limit value v, i.e.,*

$$\forall \varepsilon > 0, \exists \alpha > 0, \forall z_0 \in Z, \exists \sigma^\varepsilon \in S(z_0),\ s.t.\ \gamma_\theta(\sigma^\varepsilon) \geq v(z_0) - \varepsilon, \forall \theta \in \Delta(\mathbb{N}^*) \text{ with } TV(\theta) \leq \alpha.$$

*More generally, let v be some (Θ, φ)-limit value. Then Γ has a (Θ, φ)**-uniform value** v if the decision-maker uniformly guarantees the (Θ, φ)–limit value v, i.e.,*

$$\forall \varepsilon > 0, \exists \alpha > 0, \forall z_0 \in Z, \exists \sigma^\varepsilon \in S(z_0),\ s.t.\ \gamma_\theta(\sigma^\varepsilon) \geq v(z_0) - \varepsilon, \forall \theta \in \Theta \text{ with } \varphi(\theta) \leq \alpha.$$

The play σ^ε appearing in the above definition is called *ε-optimal.*

Particular examples of (Θ, φ) include:

1. $\Theta = \{\bar{\theta}^n : n \in \mathbb{N}\}$ and $\varphi(\bar{\theta}^n) = \frac{1}{n}$.

2. $\Theta = \{\bar{\theta}^\lambda : 0 < \lambda \leq 1\}$ and $\varphi(\bar{\theta}^\lambda) = \lambda$.
3. $\Theta = \Delta(\mathbb{N})$ and $\varphi(\theta) = \sup_{t\geq 1} \theta_t$.
4. $\Theta = \Delta(\mathbb{N})$ and $\varphi(\theta) = TV(\theta)$.

We name *Abel-limit value* (resp. *Cesàro-limit value*) for the family of evaluations being the λ-discounting (resp. the n-stage averages). Samely we name *Abel-uniform value* and *Cesàro-uniform value.* The Cesàro-uniform value is by definition the uniform value.

For different models to be later studied, the notions of limit values and uniform values will be defined in a similar way.

2.3 Between Abel mean and Cesàro mean

The study of Abel-limit (uniform) value corresponds to *discounted dynamic programming* in the literature. The equivalence between the Abel-limit value and the Cesàro-limit value refers to Tauberian type results (Lehrer and Sorin [28]), which we discuss in the first and second subsections. The relation between the (Cesàro-)uniform value and the Abel-uniform value is also mentioned. In the last subsection, we present a counterexample (due to Lehrer and Monderer [27] and Monderer and Sorin [38]) showing that the existence of limit value does not imply the existence of uniform value.

2.3.1 A Tauberian theorem

When there is no decision-maker, the limits of Cesàro means and Abel means of a positive sequence $(a_t)_{t\geq 1}$ satisfy the following Abelian theorem (cf. Lippman [32]):

$$\limsup_{n\to\infty} \bar{a}_n \geq \limsup_{\lambda\to 0} \bar{a}_\lambda \geq \liminf_{\lambda\to 0} \bar{a}_\lambda \geq \liminf_{n\to\infty} \bar{a}_n.$$

Moreover, Hardy and Littlewood (cf. Lippman [32]) proved that the the convergence of $\bar{a}_\lambda$ as λ tends to zero implies the convergence of $\bar{a}_n$ as n tends to infinity (together with Abelian theorem to have obtained the so-called Tauberian theorem).

This result is generalized by Lehrer and Sorin [28] to the framework of dynamic programming.

Theorem 2.3.1 (Lehrer and Sorin 1992)**.** *In a dynamic programming problem Γ, the uniform convergence of v_λ as λ tends to zero is equivalent to the uniform convergence of v_n as n tends to infinity. Moreover, in case of convergence, both limits are the same.*

The following example underlines the fact that *uniform convergence* (rather than pointwise convergence) is essential for the result.

Example 2 (Lehrer and Sorin 1992)**.** *$Z = \mathbb{N} \times \mathbb{N}$. $F(x, 0) = \{(x, 1), (x + 1, 0)\}$ and $F(x, y) = \{(x, y + 1)\}, \forall x \geq 0, y \geq 1$. $\forall x \geq 0$: $r(x, y) = 0$ for $y = 0$ or $y \geq x$; $r(x, y) = 1$ otherwise. $\lim_{n\to\infty} v_n(0, 0) = \frac{1}{2} \neq \frac{1}{4} = \lim_{\lambda\to 0} v_\lambda(0, 0)$. Moreover, v_n does not converge uniformly: $\lim_{n\to\infty} v_n(x, 0) = \frac{1}{2}$ while $v_x(x, 0) = 1$ for all $x \geq 1$.*

The proof of Theorem 2.3.1 relies on the the following preliminary results (Lemma 2.3.2 to Lemma 2.3.6). The general idea is to write the λ-discounted average payoff as the convex combination of n-stage average payoffs (cf. Lemma 2.3.3). Some of the results are presented in a more general framework (families of decreasing evaluations) in the next subsection.

The first result shows that the limit value is decreasing on any play.

Lemma 2.3.2. *For any play $s = (z_m) \in S(z_0)$, we have: for all $m \geq 1$,*

$$\limsup_{n\to\infty} v_n(z_m) \leq \limsup_{n\to\infty} v_n(z_0) \text{ and } \limsup_{\lambda\to 0} v_\lambda(z_m) \leq \limsup_{\lambda\to 0} v_\lambda(z_0).$$

The following result states that the λ-discounted payoff can be written as a convex combination of n-stage payoffs.

Lemma 2.3.3. *For any $\lambda \in (0,1]$ and any play $s \in S(z_0)$,*

$$\sum_{t=1}^{n} \lambda(1-\lambda)^{t-1} r(z_t) = \lambda^2 \sum_{t=1}^{n-1} t(1-\lambda)^{t-1}\gamma_t(s) + \lambda(1-\lambda)^{n-1} n\gamma_n(s), \forall n \geq 1$$

and

$$\sum_{t\geq 1} \lambda(1-\lambda)^{t-1} r(z_t) = \lambda^2 \sum_{t\geq 1} t(1-\lambda)^{t-1}\gamma_t(s).$$

The following result (compare with the Abel theorem in the uncontrolled problem) is deduced from above.

Proposition 2.3.4. *For all $\varepsilon > 0$ and all $N > 1$, there exists $\lambda_0 > 0$ such that: for all $\lambda \in (0, \lambda_0]$ and for all $z_0 \in Z$, there is $n \geq N$ satisfying*

$$v_n(z_0) \geq v_\lambda(z_0) - \varepsilon.$$

This implies that $\limsup_{n\to\infty} v_n \geq \limsup_{\lambda\to 0} v_\lambda$.

A more precise formulation of equations in Lemma 2.3.3 is as follows. Define $M[\alpha, \beta; \lambda] = \lambda^2 \sum_{t=1\alpha}^{\beta} t(1-\lambda)^{t-1}$.

Lemma 2.3.5. *There exists N_0 and ε_0 such that for any $n \geq N_0$ and any $\varepsilon \leq \varepsilon_0$ one has:*

$$M\big[(1-\varepsilon)n, n; 1/n\big] \geq \varepsilon/2e.$$

For any $\delta > 0$, there exists $\varepsilon_0 > 0$ such that for any $\varepsilon \leq \varepsilon_0$ there is some N_0 such that $n \geq N_0$ implies:

$$M\big[\varepsilon n, (1-\varepsilon)n; 1/n\sqrt{\varepsilon}\big] \geq 1 - \delta.$$

Another general property is:

Lemma 2.3.6. *For any $\varepsilon > 0$, any $z_0 \in Z$, and any $n \geq 1$, there exists a play $s = (z_t) \in S(z_0)$ and a stage L such that*

$$\frac{1}{T}\sum_{\ell=1}^{T} r(z_{L+\ell}) \geq v_n(z_0), \text{ for all } 1 \leq T \leq \frac{n\varepsilon}{2}.$$

Remark 2.3.7. *Ziliotto [70] generalizes Theorem 2.3.1 to two-player stochastic games. The approach of Ziliotto [70] is different, which is through the study of the Shapley operator.*

2.3.2 Extensions

Stochastic dynamic programming (MDP)

The Tauberian theorem extends easily to the case with stochastic transitions. Consider for example the following model of Markovian decision process (MDP).

A standard model of MDP The MDP $\Psi = \langle K, A, q, g\rangle$ is defined by the following elements: a state space K, a nonempty set A of actions, a transition function q from $K \times A$ to $\Delta_f(K)$, the set of probability distributions over K with finite support, and a payoff function g from $K \times A$ to $[0,1]$. The MDP with an initial probability distribution $p_0 \in \Delta_f(K)$, denoted by $\Psi(p_0)$, is played as follows: an initial state k_1 is chosen according to p_0, and k_1 is communicated to the decision-maker. Then he takes an action a_1 in A, which induces, together with k_1, a stage payoff $g(k_1, a_1)$ and a transition probability $q(\cdot|k_1, a_1)$ for the new state k_2. The play moves then to the next stage: the decision-maker observes k_2 and then takes again an action a_2 in A, etc.

A behavior strategy for the decision-maker is a sequence $\sigma = (\sigma_t)_{t\geq 1}$, where for each t, $\sigma_t : K \times (A \times K)^{t-1} \to \Delta_f(A)$ specifies the mixed action to be played at stage t. The set of strategies is denoted by Σ. Any $\sigma \in \Sigma$ defines, together with p_0 and q, a unique probability distribution $\mathbb{P}_\sigma^{p_0}$ over $(S \times A)^\infty$ (for the product sigma-algebra). σ is a pure strategy if each σ_t takes value in Dirac measures.

For any $\theta \in \Delta(\mathbb{N}^*)$, $v_\theta(p_0) = \sup_{\sigma\in\Sigma} \mathbb{E}_\sigma^{p_0}[\gamma_\theta(h)]$, where $\gamma_\theta(h) = \sum_{t\geq 1} \theta_t g(k_t, a_t)$ for any deterministic play $h = (k_t, a_t)_{t\geq 1} \in (K \times A)^\infty$ and $\mathbb{E}_\sigma^{p_0}$ is the expectation w.r.t. $\mathbb{P}_\sigma^{p_0}$. The notions of limit value, TV-limit value, uniform value and TV-uniform value for $\Psi(p_0)$ are defined accordingly as in the DP. For a fixed evaluation θ, v_θ can be realized with pure strategies.

We define from $\Psi(p_0)$ an equivalent auxiliary dynamic programming problem $\Gamma(z_0)$, where
– the set of space is $Z = \Delta_f(K) \times [0,1]$;
– the initial state is $z_0 = (p_0, 0)$;
– the reward function is $r : Z \to [0,1]$ with $r(p,x) = x$ for all (p,x) in Z;
– the transition correspondence is $F : Z \rightrightarrows Z$ such that: for every $z = (p,x)$ in Z,

$$F(z) = \Big\{\Big(\sum_k p^k q(k,a_k), \sum_k p^k g(k,a_k)\Big) \mid a_k \in A, \forall k \in K\Big\}.$$

Let $\tilde{v}_\theta$ be the θ-value function of Γ. Since the transition mapping $F(\cdot)$ depends on z only through its first component, then for any $z = (p,x)$ and $z' = (p,x')$, $\tilde{v}_\theta(p,x) = \tilde{v}_\theta(p,x') := \tilde{v}_\theta(p)$. Moreover, anything that can be guaranteed by the decision-maker in $\Gamma(p,0)$ can be guaranteed in $\Psi(p)$. Thus $\tilde{v}_\theta(p) = v_\theta(p)$ for any $\theta \in \Delta(\mathbb{N}^*)$, and in particular, we have the same n-stage values and λ-discounted values in both models. A uniform Tauberian theorem for MDP then follows from Theorem 2.3.1.

Continuous time framework

Oliu-Barton and Vigeral [40] extended the uniform Tauberian theorem to (deterministic) continuous time framework. See also Buckdahn *et al.*[14] for the generalization to stochastic control systems.

A model of optimal control Consider an optimal control problem $\mathcal{J} = \langle U, f, g\rangle$ which is defined by the following elements: U is a metric space, $f : \mathbb{R}^d \times U \to \mathbb{R}^d$ is a Borel

measurable function describing the dynamic, $g : \mathbb{R}^d \times U \to [0;1]$ is Borel measurable function which defines the running cost. A control $\mathbf{u}$ is a measurable function from $\mathbb{R}_+$ to U, and the set of controls is denoted by $\mathcal{U}$. The controlled dynamic is:

$$\mathbf{y}'(\mathbf{s}) = f(\mathbf{y}(s), \mathbf{u}(s)),\ \mathbf{y}(0) = y_0 \in \mathbb{R}^d. \tag{2.3.1}$$

We assume suitable regularity conditions on f (i.e., uniformly Lipschitz in $y \in \mathbb{R}^d$ and bounded by a linear functional for all $u \in U$) such that given an initial state $y_0 \in \mathbb{R}^d$, any control $\mathbf{u} \in \mathcal{U}$ defines a unique absolutely continuous solution from $\mathbb{R}_+$ to $\mathbb{R}^d$ for the controlled dynamic (2.3.1). Denote this solution (the *trajectory*) associated with $(\mathbf{u}, y_0)$ by: $s \mapsto \mathbf{y}(s, \mathbf{u}, y_0)$.

Let $\theta \in \Delta(\mathbb{R}_+)$ be a Borel probability measure over $\mathbb{R}_+$ (call θ an *evaluation*). The θ-value of the optimal control problem $\mathcal{J}$ is

$$V_\theta(y_0) = \inf_{\mathbf{u}\in\mathcal{U}} \int_{[0,+\infty)} g\Big(\mathbf{y}(s, \mathbf{u}, y_0), \mathbf{u}(s)\Big)\mathrm{d}\theta(s).$$

Specific evaluations and their associated value functions are
Cesàro mean: $\forall t > 0$, $\overline{\theta}_t$ has a density $s \mapsto f_{\overline{\theta}_t}(s) = \frac{1}{t}\mathbb{1}_{[0,t]}(s)$, and the *t-horizon value* is

$$V_t(y_0) = \inf_{\mathbf{u}\in\mathcal{U}} \frac{1}{t}\int_0^t g\Big(\mathbf{y}(s, \mathbf{u}, y_0), \mathbf{u}(s)\Big)\mathrm{d}s$$

Abel mean: $\forall \lambda \in (0, 1]$, θ_λ has a density $s \mapsto f_{\theta_\lambda}(s) = \lambda e^{-\lambda s}$, and the *λ-discounted value* is

$$V_\lambda(y_0) = \inf_{\mathbf{u}\in\mathcal{U}} \int_0^{+\infty} \lambda e^{-\lambda s} g\Big(\mathbf{y}(s, \mathbf{u}, y_0), \mathbf{u}(s)\Big)\mathrm{d}s$$

Theorem 2.3.8 (Oliu-Barton and Vigeral 2013)**.** *Let V be a function defined on $\mathbb{R}^d$. In an optimal control problem $\mathcal{J}$,*

$$V_t \longrightarrow_{t\to\infty} V \iff V_\lambda \longrightarrow_{\lambda\to 0} V \quad (\textit{convergence uniform on } \mathbb{R}^d).$$

To apply [2] Theorem 2.3.1 for the above result, we follow here Oliu-Barton and Vigeral [40] to associate the equivalent dynamic programming problem $\Gamma = \langle Z, F, r\rangle$ (with an initial state $z_0 = (y_0, 0)$) with an optimal control problem $\mathcal{J}$ (with an initial state y_0):
– the state space is $Z = \mathbb{R}^d \times [0, 1]$;
– the transition correspondence is $F : Z \rightrightarrows Z$ such that: $\forall(\omega, x), (\omega', x') \in \mathbb{R}^d \times [0, 1]$,

$$(\omega', x') \in F(\omega, x) \iff \exists \mathbf{u} \in \mathcal{U}\ s.t.\ \mathbf{y}(1, \mathbf{u}, \omega) = \omega',\ \text{and}\ \int_0^1 g\Big(\mathbf{y}(s, \mathbf{u}, \omega), \mathbf{u}(s)\Big)\mathrm{d}s = x'.$$

– the reward function $r : Z \to [0, 1]$ is: $\forall(\omega, x) \in \mathbb{R}^d \times [0, 1]$, $r(\omega, x) = x$.

Define the following value functions in Γ, $\forall\ (y_0, x) \in \mathbb{R}^d \times [0, 1]$: for any $n \in \mathbb{N}^*$,

$$\nu_n(y_0, x) = \inf_{(z_i)_i\in S(z_0)} \frac{1}{n}\sum_{i=1}^n r(z_i) = \inf_{\mathbf{u}\in\mathcal{U}} \frac{1}{n}\sum_{i=1}^n \left[\int_{i-1}^i g\Big(\mathbf{y}(s, \mathbf{u}, y_0), \mathbf{u}(s)\Big)\mathrm{d}s\right];$$

and for any $\lambda \in (0, 1]$,

$$\nu_\lambda(y_0, x) = \inf_{(z_i)_i\in S(z_0)} \sum_{i\geq 1} \lambda(1-\lambda)^{i-1} r(z_i) = \inf_{\mathbf{u}\in\mathcal{U}} \sum_{i\geq 1} \lambda(1-\lambda)^{i-1}\left[\int_{i-1}^i g\Big(\mathbf{y}(s, \mathbf{u}, y_0), \mathbf{u}(s)\Big)\mathrm{d}s\right].$$

2. Oliu-Barton and Vigeral [40] provided also a direct proof for the result, which takes the analogue form of Theorem 2.3.1.

Since the transition correspondence F depends on $z = (y_0, x)$ only through its first component, we obtain that $\nu_n(y_0, x) = \nu_n(y_0, 0) := \nu_n(y_0)$ and $\nu_\lambda(y_0, x) = \nu_\lambda(y_0, 0) := \nu_\lambda(y_0)$ for all x.

Finally, to obtain a uniform Tauberian theorem for $\mathcal{J}$, it remains to verify that for any y_0:

$$|\nu_{\lfloor t \rfloor}(y_0) - V_t(y_0)| \leq \sum_{i=1}^{\lfloor t \rfloor} \int_{i-1}^{i} \Big|1/\lfloor t \rfloor - 1/t\Big| \mathrm{d}s + \int_{\lfloor t \rfloor}^{t} \Big|1/t\Big| \mathrm{d}s \leq \frac{2}{t} \longrightarrow_{t\to\infty} 0$$

and

$$|\nu_\lambda(y_0) - V_\lambda(y_0)| \leq \lambda \int_0^{+\infty} \Big|(1-\lambda)^{\lfloor t \rfloor} - e^{-\lambda t}\Big| \mathrm{d}s \leq 2(e^\lambda - 1) \longrightarrow_{\lambda\to 0} 0.$$

See Section 4 for the analogue reduction of asymptotic analysis in continuous time to discrete time associated with general evaluations.

General evalutaions

Consider (θ^k) and (μ^k) any two sequences of evaluations satisfying certain regularity condition (ex. with vanishing stage weights, or with vanishing total variations), one may study the equivalence between the uniform convergence of (v_{θ^k}) and of (v_{μ^k}), and in particular, the equivalence to the uniform convergence of (v_n).

Decreasing evaluations The uniform Tauberian theorem is generalized by Monderer and Sorin [38] to families of decreasing evaluations satisfying some *extra conditions*. The main argument is the fact that for any decreasing evaluation, its average payoff can be written as a convex combination of the n-stage average payoffs, thus the approach in Lehrer and Sorin [28] may extend. Here the *extra conditions* refer to 1) certain regularity condition to define a convergence, and 2) properties analogue to Lemma 2.3.5.

The proof uses generalized versions of Lemma 2.3.2, Lemma 2.3.3, Proposition 2.3.4. Denote $\|\theta\| := \sup_{t\geq 1} \theta_t$ for any $\theta \in \Delta(\mathbb{N}^*)$.

Lemma 2.3.2' *For any play $s = (z_m) \in S(z_0)$, one has: for any $m \geq 1$,*

$$\limsup_{\|\theta\|\to 0} v_\theta(z_m) \leq \limsup_{\|\theta\|\to 0} v_\theta(z_0).$$

Proof. For any $\varepsilon > 0$, let $\mu \in \Delta(\mathbb{N}^*)$ with $\|\mu\| \leq \varepsilon$ and $v_\mu(z_1) \geq \limsup_{\|\theta\|\to 0} v_\theta(z_1) - \varepsilon/4$. Take $s' \in S(z_1)$ with $\gamma_\mu(s') \geq v_\mu(z_1) - \varepsilon/4$. Define now $\theta \in \Delta(\mathbb{N}^*)$ to be (satisfying $\|\theta\| \leq \varepsilon$):

$$\theta_1 = \varepsilon/2 \ \text{ and } \ \theta_t = (1-\varepsilon/2)\mu_{t-1}, \ \text{for any } \ t \geq 2.$$

By construction, $s'' = (z_1, s')$ is a play in $S(z_0)$, and the associated θ-evaluated payoff is

$$\gamma_\theta(s'') = \varepsilon/2 r(z_1) + (1-\varepsilon/2)\gamma_\mu(s') \geq \limsup_{\|\theta\|\to 0} v_\theta(z_1) - \varepsilon.$$

This proves $\limsup_{\|\theta\|\to 0} v_\theta(z_0) \geq \limsup_{\|\theta\|\to 0} v_\theta(z_1)$, and the result is obtained by iteration. □

Lemme 2.3.3' *For any evaluation $\theta \in \mathcal{D}$ and a play $s \in S(z_0)$,*

$$\sum_{t=1}^{n} \theta_t r(z_t) = \sum_{t=1}^{n-1} t(\theta_t - \theta_{t+1})\gamma_t(s) + n\theta_n \gamma_n(s), \ \forall n \geq 1$$

and

$$\sum_{t\geq 1}\theta_t r(z_t) = \sum_{t\geq 1} t(\theta_t - \theta_{t+1})\gamma_t(s).$$

Proposition 2.3.4' *For all $\varepsilon > 0$ and all $N > 1$, there exists $\eta_0 > 0$ such that: for all $\theta \in \mathcal{D}$ with $\theta_1 \leq \eta_0$, for all $z_0 \in Z$, there is $n \geq N$ satisfying*

$$v_n(z_0) \geq v_\theta(z_0) - \varepsilon.$$

This implies that

$$\limsup_{n\to\infty} v_n \geq \limsup_{\theta\in\mathcal{D}:\ \theta_1\to 0} v_\theta$$

Proof. Given $\varepsilon > 0$ and $N > 1$, let $\eta_0 = \varepsilon/2N$. Then $\sum_{t=1}^{N-1} t(\theta_t - \theta_{t+1}) \leq N\theta_1 < \varepsilon/2$ for any $\theta \in \mathcal{D}$ with $\theta_1 \leq \eta_0$. For any $z_0 \in Z$ and $\theta \in \mathcal{D}$ with $\theta_1 \leq \eta$, let $s = (z_t) \in S(z_0)$ an $\varepsilon/2$-optimal play for the θ-evaluated DP at z_0: $\gamma_\theta(s) \geq v_\theta(z_0) - \varepsilon/2$. By Lemma 2.3.3, we deduce $\sum_{t\geq N} t(\theta_t - \theta_{t+1})r(z_t) \geq v_\theta(z_0) - \varepsilon$. As $\sum_{t\geq N} t(\theta_t - \theta_{t+1}) \leq 1$ and the reward is between $[0, 1]$, there is some $n \geq N$ such that $\gamma_n(s) \geq \max\{0, v_\theta(z_0) - \varepsilon\} \geq v_\theta(z_0) - \varepsilon$. □

Families with vanishing total variations Consider (θ^k) and (μ^k) any two families that are not necessarily decreasing but with vanishing total varations. One may ask (extra conditions for) the equivalence between the uniform convergence of (v_{θ^k}) and of (v_{μ^k}), and in particular, the equivalence to the uniform convergence of the n-stage values.

The following example (simplified from Renault [46], Example 3.2) shows that for one direction, this is not true: for a particular family of evaluations with vanishing total variation, the associated value functions converge uniformly; while the n-stage values do not have uniform convergence (an analogous example in continuous time can be found in Li *et al.* [30]).

Example 3. *The state space is $Z = \mathbb{N} \times \{0, 1\}$. For any $x \in \mathbb{N}$: $F(x, 0) = \{(x + 1, 0), (x, 1)\}$; $F(x, 1) = \{(x, 0), (x - 1, 1)\}$ for $x \geq 1$ and $F(0, 1) = \{(0, 0)\}$. The reward function is: $r(x, 0) = 0$ and $r(x, 1) = 1$ for any x. Consider the family of evaluations: for any $k \geq 1$, $\theta^k = \sum_{t\geq 1} \frac{1}{k}\mathbb{1}_{\{k+1\leq t\leq 2k\}}(t)$. We have:*

- *v_{θ^k} converges uniformly to 1:*
  - *$v_{\theta^k}(x, 0) = 1$ for $k \geq 1$;*
  - *$v_{\theta^k}(x, 1) = 1$ for $k \leq x$ or $x \geq 2$;*
  - *$v_{\theta^k}(0, 1) = v_{\theta^k}(1, 1) = 1 - 1/k$, for any $k \geq 1$.*
- *$v_n(z)$ converges to $1/2$ for any $z \in \mathbb{N} \times \{0, 1\}$. However, the convegence is not uniform as $v_x(x, 0) = 1$ for any $x \in \mathbb{N}$.*

For the other direction, it is true for uncontrolled problems (zero-player), as shown by the following

Proposition 2.3.9 (Renault 2014)**.** *Let (θ^k) be a sequence of evaluations with $TV(\theta^k) \longrightarrow_{k\to\infty} 0$ and γ a function defined from Z to $[0, 1]$. For any $z_0 \in Z$, let $s \in S(z_0)$ be the unique play at z_0. Then:*

$$\gamma_n(s) \longrightarrow_{n\to\infty} \gamma(s),\ \textit{uniformly w.r.t. } z_0 \implies \gamma_{\theta^k}(s) \longrightarrow_{k\to\infty} \gamma(s),\ \textit{uniformly w.r.t. } z_0.$$

It is unknown for one-player problems:

Question: In a dynamic programming problem, does the existence of limit value imply the existence of TV-limit value?

See Proposition 6.1 in Li *et al.* [30] for the analogous result and related discussion in continuous time.

2.3.3 Blackwell optimality

When Abel-limit value exists, i.e. (v_λ) converges uniformly, the existence of Abel-uniform value is weaker than the so-called *Blackwell optimality* condition in the literature, which is defined as:

$$\exists \lambda_0 > 0, \forall z_0 \in Z, \exists \sigma \in \Sigma(z_0),\ s.t.\ \gamma_\lambda(\sigma) \geq v_\lambda(z_0), \forall \lambda \in (0, \lambda_0].$$

The play σ appearing in the above definition for Blackwell optimality is called 0-*optimal*.

Theorem 2.3.10 (Blackwell 1962)**.** *Let Z be a finite state space. Then Blackwell optimality exists in Γ with pure stationary plays*[3]*. The uniform value exists in pure stationary plays.*

Remark 2.3.11. *The proof relies on the fact that in any λ-discounted problem a pure stationary optimal strategy exists. One deduces that the value function v_λ can be expressed as a rational fraction of λ for λ close to zero.*

The existence of uniform value implies the existence of Abel-uniform value. On the other hand, the existence of Abel-uniform value does not imply [4] the existence of uniform value (cf. Renault [44], Lemma 5.4).

2.3.4 Limit value does not imply uniform value

By the uniform Tauberian theorem (Theorem 2.3.1), the existence of limit value is equivalent to the uniform convergence of the n-stage values. Monderer and Sorin [38] (see also Lehrer and Monderer [27]) provided a conterexample for which (v_n) converges uniformly, however the uniform value does not exist.

Example 4 (Monderer and Sorin 1993)**.** *Start introducing an infinite tree $T(x)$ with x its root for any $x \in [0,1]$. At x, $T[x]$ has a countable number of branches indexed by $x(n) \in \mathbb{N}^*$, and any node on each such branch has an outgoing degree one. Let $x(n;m)$ be the m-th node of the branch $x(n)$. Consider now two positive sequences (δ_n) and (ε_n) decreasing to zero. We define the directed graph by attaching any $x(n;m)$ with the root of the tree $T\big[\max\{x-\delta_m, 0\}\big]$ for any n.*
To finish defining a DP on this directed graph, we set
– $F(z)$ to be the set of successors of z for any node z;
– the reward function to be: for any x, n, $r\big(x(n;m)\big) = x - \delta_n$ for $\varepsilon_n n \leq m \leq n$ and $r\big(x(n;m)\big) = 0$ for $1 \leq m \leq \varepsilon_n n$.
By a specific choice of the sequences (δ_n) and (ε_n), one can prove that 1). $v_n(z)$ converges to $r(z)$ for any node z and the convergence is uniform in z; 2). $\sup_{s\in S(z)} \liminf_{n\to\infty} \gamma_n(s) = 0$ for each z. Thus the uniform value does not exist.

Consider for example the root of $T[1]$ the starting point. First, one observes that during any branch $x(n)$, it is always optimal to stay until the node $x(n;n)$ and then to leave from it. The decision turns to be a sequence of integers $(m_k)_{k=1}^\infty$ inducing the play: to stay in the m_1-th branch of $T[1]$ until the node m_1; then to stay in the m_2-th branch of $T\big[1-\delta_{m_1}\big]$ until the node m_2,..., to stay in the m_{k+1}-th branch of $T\big[1-\sum_{k'\leq k}\delta_{m_{k'}}\big]$

3. In two-player zero sum stochastic games, there is in general no stationary ε-optimal strategies for players to guarantee the limit value. See for example the Big match (Blackwell and Ferguson [13]).

4. The notion of uniform value (or Abel-uniform value) in Renault [44] is defined pointwise. Renault [44] presented an example such that at one initial state there is Abel-uniform value (Blackwell optimality) but no uniform value.

until the node m_k,.... In the induced reward sequence, at each "round" (corresponding to a different branch in one tree): a stream of (positive) rewards $(1-\sum_{k'\leq k}\delta_{m_{k'}})$ has a length $(1-\varepsilon_{m_{k+1}})m_{k+1}$ and it is after a stream of zeros of length $\varepsilon_{m_{k+1}}m_{k+1}$.

The balance of the decision is as follows. On one hand, m_k needs to be taken large so the decrease by δ_{m_k} is small, and the positive reward in the next "round" is close to 1; on the other hand, this might induce an increasing in $\varepsilon_{m_k}m_k$ the length of the waiting time for the positive rewards in the next "round". In general, the construction is to take (δ_n) converging to zero very slowly thus the sequence (m_k) goes to infinity very fast; and to take for example $\varepsilon_n = 1/\sqrt{n}$, thus $\varepsilon_{m_k}m_k$ tends to infinity and overweights $\sum_{k'<k} m_{k'}$.

For the asymptotic analysis: for any given large n and node z, an optimal play is to approximately go through zero for $\varepsilon_n n$ stages, and through a reward close to $r(z)-\delta_n$ for $(1-\varepsilon_n)n$ stages.

For the uniform analysis: there is no ending stage, thus, to obtain any positive rewards, the play has to go through a large number of zeros which will overweight all positive rewards of previous stages.

2.4 Asymptotic analysis

In dynamic programming with finite state space, the existence of limit value follows from the algebraicity of the function $\lambda \mapsto v_\lambda$ (valid also for two-player zero-sum stochastic games with finite state space and finite action spaces, cf. Bewley and Kohlberg [10]). Without the assumption of finite state space, further conditions are needed for the asymptotic analysis. Renault [44] provided conditions (compactness property of the family $\{v_n\}$) for the existence of limit value in DP with arbitrary state space. In Renault [46], this approach has been generalized to study the limit value associated with any family of evaluations with vanishing total variation. The result in Renault [46] implies the existence of TV-limit value in compact nonexpansive DP, and is extended to continuous time framework by Li *et al.*[30] (see Chapter 3).

Introduce the auxiliary value function $v_{m,\theta}$ as:

Definition 2.4.1. *For any $m \geq 0$, $\theta \in \Delta(\mathbb{N}^*)$ and a play $s = (z_t)_{t\geq 1} \in S(z_0)$, denote*

$$\gamma_{m,\theta}(s) = \sum_{t=1}^{\infty} \theta_t r(z_{m+t}) \text{ and } v_{m,\theta}(z_0) = \sup_{s\in S(z_0)} \gamma_{m,\theta}(s).$$

The explanation of $v_{m,\theta}(z_0)$ is as follows. Consider the auxiliary problem where starting from z_0, the decision-maker takes m steps to reach a "good" state , and the payoff stream from stage $m+1$ on is evalueted by θ, and $v_{m,\theta}(z_0)$ is the value of this problem.

The following function v^* characterizes the limit value in case of convergence:

Definition 2.4.2. *For any $z_0 \in Z$,*

$$v^*(z_0) = \inf_{\theta\in\Delta(\mathbb{N}^*)} \sup_{m\geq 0} v_{m,\theta^k}(z_0). \tag{2.4.1}$$

The main result of Renault [46] is the following:

Theorem 2.4.3 (Renault 2014)**.** *Let $(\theta^k)_{k\geq 1}$ be a sequence of evaluations with $TV(\theta^k) \xrightarrow[k\to\infty]{} 0$. Then the "inf" in Eq. (2.4.1) over $\Delta(\mathbb{N}^*)$ can be taken over $\{\theta^k\}$, i.e.,*

$$v^*(z_0) = \inf_{k\geq 1} \sup_{m\geq 0} v_{m,\theta^k}(z_0), \ \forall z_0 \in Z.$$

Moreover, any accumulation point of the sequence $(v_{\theta^k})_k$ for the uniform norm is v^*.

This implies that for any sequence $(\theta^k)_{k\geq 1}$ with $TV(\theta^k) \xrightarrow[k\to\infty]{} 0$, (V_{θ^k}) uniformly converges if and only if it is totally bounded for the uniform norm. And in case of convergence, the limit value is v^*.

A sketch of proof

Fix a sequence $(\theta^k)_k$ with vanishing total variation, and denote:

$$v^-(z_0) = \liminf_{k\to\infty} v_{\theta^k}(z_0) \quad \text{and} \quad v^+(z_0) = \limsup_{k\to\infty} v_{\theta^k}(z_0), \ \forall z_0 \in Z.$$

A first step is to bound v^- and v^+ in terms of the auxiliary value functions $\{v_{m,\theta^k}\}$:

Proposition 2.4.4. *For every state $z_0 \in Z$ and any $m_0 \geq 0$,*

$$\inf_{k\geq 1} \sup_{0\leq m\leq m_0} v_{m,\theta^k}(z_0) \leq v^-(z_0) \leq v^+(z_0) \leq \inf_{k\geq 1} \sup_{m\geq 0} v_{m,\theta^k}(z_0) = v^*(z_0).$$

Then the uniform convergence of (v_{θ^k}) to v^* can be deduced from the uniform convergence of $\inf_{k\geq 1}\sup_{0\leq m\leq m_0} v_{m,\theta^k}$ to $\inf_{k\geq 1}\sup_{m\geq 0} v_{m,\theta^k}$ as m_0 tends to infinity. This is the second step of the proof.

Using the reachable set of the transition correspondence F, the inequalities in Proposition 2.4.4 can be re-writen in an equivalent form.

Definition 2.4.5. *$F^0(z) = \{z\}$ for every state, and $F^{n+1} = F^n \circ F$ for every $n \geq 0$, where the composition is defined by $G \circ H(z) = \{z'' \in Z : z'' \in G(z') \text{ for some } z' \in H(z)\}$. Let $m_0 \geq 0$, write $G^{m_0}(z_0) = \cup_{n=0}^{m_0} F^n(z_0)$, which is the set of states that the decision-maker can reach by at most m_0 steps starting from the initial state $z_0 \in Z$, and $G^\infty(z_0) = \cup_{n=0}^{\infty} F^n(z_0)$, which is the set of the states that the decision-maker can reach by finite steps starting from z_0.*

Proposition 2.4.6. *For every state $z_0 \in Z$ and any $m_0 \geq 0$,*

$$\inf_{k\geq 1} \sup_{z'\in G^{m_0}(z_0)} v_{\theta^k}(z') \leq v^-(z_0) \leq v^+(z_0) \leq \inf_{k\geq 1} \sup_{z'\in G^{\infty}(z_0)} v_{\theta^k}(z').$$

Finally, the proof is achieved by the following arguments:
– **Step 1**: *Viewing Z as a totally bounded pseudometric space.* Define $\tilde{d}(z, z') = \sup_k |v_{\theta^k}(z) - v_{\theta^k}(z')|$ for all z, z' in Z. Then $(Z, \tilde{d})$ is a pseudometric space (and may not be Hausdorff).
– **Step 2**: *Convergence of reachable sets.* Fix $z_0 \in Z$. Use the fact that $(Z, \tilde{d})$ is totally bounded pseudometric to obtain the convergence of $(G^m(z_0))_{m\geq 1}$ to $G^\infty(z_0)$ in the sense that:

$$\forall \varepsilon > 0, \exists m_0 \geq 0, \ s.t. \ \forall z' \in G^\infty(z_0), \exists z'' \in G^{m_0}(z), \tilde{d}(z, z') \leq \varepsilon.$$

– **Step 3**: *Convergence of $(v_{\theta^k})_k$.* Deduce from the convergence condition in **Step 2** that

$$\inf_{k\geq 1} \sup_{z'\in G^{m_0}(z_0)} v_{\theta^k}(z') \geq \inf_{k\geq 1} \sup_{z'\in G^{\infty}(z_0)} v_{\theta^k}(z') - 2\varepsilon,$$

thus the point-wise convergence of (v_{θ^k}) to v^*. The uniform convergence follows from the fact that each v_{θ^k} is 1-Lipschitz for $\tilde{d}$ and $(Z, \tilde{d})$ is totally bounded. □

Remark 2.4.7. *Renault [44] proved that $(v_n)_{n\geq 1}$ converges uniformly if and only if the space $(\{v_n : n \geq 1\}, \|\cdot\|_\infty)$ is totally bounded. This corresponds to the particular case of Theorem 2.4.3 if one takes the family (θ^k) to be $(\bar{\theta}^n)$ the n-stage averages.*

The following corollary gives sufficient conditions relying on hypothesis directly expressed in terms of the basic data of the problem.

Definition 2.4.8. *The dynamic programming problem $\Gamma = \langle Z, r, F\rangle$ is called* ***compact nonexpansive*** *if the following conditions are satisfied:*
A.1) the space (Z, d) is metric precompact;
A.2) r is uniformly continuous on Z;
A.3) F is nonexpansive for d, i.e.,

$$\forall z, z' \in Z, \forall y \in F(z), \exists y' \in F(z'),\ s.t.\ \ d(y, y') \leq d(z, z').$$

Corollary 2.4.9. *Let $\Gamma = \langle Z, F, r\rangle$ be a compact nonexpansive dynamic programming problem. Then there is TV- uniform convergence of the value functions $\{v_\theta\}$ to v^*, i.e.,*

$$\forall \varepsilon > 0, \exists \alpha > 0, \forall \theta \in \Delta(\mathbb{N}^*),\ s.t.\ TV(\theta) \leq \alpha, ||v_\theta - v^*|| \leq \varepsilon.$$

In fact, when Γ is *compact nonexpansive*, the family $\{v_\theta\}$ is uniformly (in $z_0 \in Z$) equicontinuous, and according to Ascoli's theorem, the space $(\{v_\theta\}, \|\cdot\|_\infty)$ is totally bounded.

Extension to continuous time framework

Results in continuous time framework analogue to Theorem 2.4.3 are obtained in Li *et al.* [30] (see Chapter 3). Consider the optimal control problem $\mathcal{J} = \langle U, f, g\rangle$ described in Subsection 3.2.2.

The analogous notion of total variation for a probability measure and the analogous convergence condition as "vanishing total variation" are defined in the following way.

Definition 2.4.10. *A sequence of evaluations $(\theta_k)_{k\geq 1}$ in $\Delta(\mathbb{R}_+)$ satisfies the* ***long-term condition (LTC)*** *if*

$$\overline{TV}_S(\theta^k) \xrightarrow[k\to\infty]{} 0, \quad \textit{for all } S > 0 \ (\textit{or equivalently, for some } S > 0),$$

where for any $\theta \in \Delta(\mathbb{R}_+)$ and $S > 0$:

$$\overline{TV}_S(\theta) = \sup_{0\leq s\leq S} TV_s(\theta),\ \textit{and}\ \ TV_s(\theta) = \sup_{Q\in\mathcal{B}(\mathbb{R}_+)} |\theta(Q) - \theta(Q + s)|.$$

Remark 2.4.11. *1) Let $\theta \in \Delta(\mathbb{R}_+)$ be absolutely continuous with f_θ its density function, then:*

$$TV_s(\theta) = \frac{1}{2}\int_{[0,+\infty)} \big|f_\theta(s) - f_\theta(s + t)\big|\mathrm{d}t,\ \forall s \geq 0.$$

2) If $t \mapsto f_{\theta^k}(t)$ is decreasing ,$\forall k \geq 1$, then (θ^k) satisfies the LTC iff $\theta^k([0, M]) \longrightarrow_{k\to\infty} 0,\ \forall M > 0$.

$$\text{Define}:\ V^*(y_0) = \sup_{\theta\in\Delta(\mathbb{R}_+)} \inf_{s\in\mathbb{R}_+} \inf_{\mathbf{u}\in\mathcal{U}} \int_{[0,+\infty)} g\Big(\mathbf{y}(t + s, \mathbf{u}, y_0), \mathbf{u}(t + s)\Big)\mathrm{d}\theta(t),\quad \forall y_0 \in \mathbb{R}^d.$$

The following results are analogue to Theorem 2.4.3 in discrete time:

Theorem 2.4.12 (Li *et al.* 2015)**.** *Let $(\theta^k)_{k\geq 1}$ be a sequence of evaluations satisfying the long-term condition. Then:*

$$V^*(y_0) = \sup_{k\geq 1} \inf_{s\in\mathbb{R}_+} \inf_{\mathbf{u}\in\mathcal{U}} \int_{[0,+\infty)} g\Big(\mathbf{y}(t+s,\mathbf{u},y_0),\mathbf{u}(t+s)\Big)\mathrm{d}\theta^k(t), \quad \forall y_0 \in \mathbb{R}^d.$$

Any accumulation point of the sequence (V_{θ^k}) for the uniform norm is V^*.

In particular, for any (θ^k) satisfying the LTC, (V_{θ^k}) uniformly converges if and only if the space $(\{V_{\theta^k}, k\geq 1\}, \|\cdot\|_\infty)$ is totally bounded. And in case of convergence, the limit value is V^*.

Remark 2.4.13. *A different approach can be followed through a direct reduction of the problem in continuous time to a problem in discrete time. As in Subsection 2.3.2, let $\Gamma = \langle Z, F, r\rangle$ be the DP problem associated with the optimal control problem $\mathcal{J} = \langle U, f, g\rangle$.*

Fix $\theta \in \Delta(\mathbb{R}_+)$ and denote by $\mu := \mu(\theta) \in \Delta(\mathbb{N}^*)$ the measure: $\mu_i = \theta\big([i-1,i)\big)$, $\forall i \in \mathbb{N}^*$. Assume that θ is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}_+$, and let f_θ be its density function. Define $\widehat{TV}(\theta) = \sum_{i\geq 1}\int_{i-1}^{i}\Big|f_\theta(s)-\mu_i\Big|\mathrm{d}s + TV\big(\mu(\theta)\big)$. The $\mu(\theta)$-value of the DP $\Gamma(z_0)$ for $z_0 = (y_0, 0)$ is:

$$\nu_\mu(y_0) = \inf_{(z_i)_i\in S(z_0)} \sum_{i\geq 1}\mu_i r(z_i) = \inf_{\mathbf{u}\in\mathcal{U}} \sum_{i\geq 1}\mu_i\left[\int_{i-1}^{i} g\Big(\mathbf{y}(s,\mathbf{u},y_0),\mathbf{u}(s)\Big)\mathrm{d}s\right]$$

We write $V_\theta(y_0) = \inf_{\mathbf{u}\in\mathcal{U}}\sum_{i\geq 1}\int_{i-1}^{i} g\Big(\mathbf{y}(s,\mathbf{u},y_0),\mathbf{u}(s)\Big)f_\theta(s)ds$ and obtain :

$$\Big|\nu_\mu(y_0) - V_\theta(y_0)\Big| \leq \sum_{i\geq 1}\int_{i-1}^{i}\Big|f_\theta(s)-\mu_i\Big|\mathrm{d}s \leq \widehat{TV}(\theta). \tag{2.4.2}$$

Consider a sequence of evaluations (θ^k) in $\Delta(\mathbb{R}_+)$, and denote by $(\mu^k) := (\mu(\theta^k))$ for its corresponding sequence in $\Delta(\mathbb{N}^*)$. Then we have

Theorem 2.4.14. *Let $\mathcal{J}$ be an optimal control problem. (θ^k) is a sequence of evaluations with $\widehat{TV}(\theta^k) \longrightarrow_{k\to 0} 0$. Then V_{θ^k} converges uniformly iff ν_{μ^k} converges uniformly. In case of convergence, the limit is $\nu^*(y_0) = \sup_{\mu\in\Delta(\mathbb{N}^*)}\inf_{t\geq 0}\nu_{t,\mu}(y_0)$, where $\nu_{t,\mu}(y_0) = \inf_{s\in S(y_0,0)}\gamma_{t,\mu}(s), \forall y_0$.*

Question: 1) What is the relation between two limit functions ν^* and V^*? 2) One can prove that "(θ^k) satisfies the LTC $\Longrightarrow$ $\widehat{TV}(\theta^k) \longrightarrow_{k\to\infty} 0 \Longrightarrow (\theta^k)$ satisfies the LTC'", where the LTC' is defined as $TV_s(\theta^k) \longrightarrow_{k\to\infty} 0, \forall s \geq 0$. It is unknown whether the LTC is strictly stronger than the LTC'.

2.5 Uniform analysis

For dynamic programming problems having an infinite state space, the arguments for the existence of Blackwell optimality in finite case does not apply. In the context of two player zero-sum stochastic games, Mertens and Neyman [34] provided sufficient conditions for the existence of uniform value: the function $\lambda \to v_\lambda$ satisfies certain "bounded variation" property [5]. Renault [44] proved other conditions for the existence of uniform value in DP. The main condition is expressed as compactness of the family of a family of auxiliary value functions $\{w_{m,n}\}$ defined as follows.

5. This property is established in Bewley and Kohlberg [10] for stochastic games with a finite state space and finite action sets

Definition 2.5.1. *For any $m \geq 0$, $n \geq 1$ and $z_0 \in Z$,*

$$w_{m,n}(z_0) = \sup_{s \in S(z_0)} \min_{1 \leq t \leq n} \gamma_{m,t}(s), \quad \textit{where} \quad \gamma_{m,t}(s) = \frac{1}{n} \sum_{s=1}^{t} r(z_{m+s}).$$

Notice that $\gamma_{m,t}(s)$ is short for $\gamma_{m,\bar{\theta}^t}(s)$ where $\bar{\theta}^t = \frac{1}{t}\mathbb{1}_{[1,t]}$ is the average of the first t stages. So we write $v_{m,n}(z_0) = \sup_{s \in S(z_0)} \gamma_{m,n}(s)$ for $v_{m,\bar{\theta}^n}(z_0)$ (cf. Def. 2.4.1).

Remark 2.5.2. *The interpretation of the value function $w_{m,n}$ is that: the decision-maker takes m steps to reach a "good" initial state, but then his payoff is the minimum among the next following n average rewards (as if some adversary trying to mimimize the average rewards by choosing the length of the remaining game). This is related to the notion of uniform value, which requires the play to be approximately optimal for any problem with sufficiently large horizon.*

Preliminary results concerning $w_{m,n}$ are :

Proposition 2.5.3. *For all state $z_0 \in Z$,*

$$\begin{aligned}\sup_{m \geq 0} \inf_{n \geq 1} w_{m,n}(z_0) \leq \sup_{m \geq 0} \inf_{n \geq 1} v_{m,n}(z_0) =& v^-(z_0) \\ \leq& v^+(z_0) \leq v^*(z_0) = \inf_{n \geq 1} \sup_{m \geq 0} v_{m,n}(z_0) = \inf_{n \geq 1} \sup_{m \geq 0} w_{m,n}(z_0).\end{aligned}$$

The main result of Renault [44] is

Theorem 2.5.4 (Renault 2011)**.** *Assume that the space $(\{w_{m,n} : m \geq 0, n \geq 1\}, \|\cdot\|_\infty)$ is totally bounded. Then the dynamic programming problem Γ has a uniform value v^*, where*

$$v^*(z_0) = v^+(z_0) = v^-(z_0) = \sup_{m \geq 0} \inf_{n \geq 1} v_{m,n}(z_0) = \sup_{m \geq 0} \inf_{n \geq 1} w_{m,n}(z_0), \forall z_0 \in Z.$$

A sketch of proof of Theorem 2.5.4

The first part of the proof is to show that the operators "sup" and "inf" in $\sup_m \inf_n w_{m,n}$ commute under the assumptions. This helps to establish the convergence of $(v_n)_n$ to v^*. The proof is similar to Theorem 2.4.3. One defines $\tilde{d}(z, z') = \sup_{m,n} |w_{m,n}(z) - w_{m,n}(z')|$ for all $z, z' \in Z$ such that $(Z, \tilde{d})$ is a pseudometric space and is totally bounded, then the convergence results are obtained, first for reachable sets, and second for value functions.

The second part is to show that uniform ε-optimal plays exist for all problems with sufficient large horizons. When "sup" and "inf" commute, one deduces the following: within some finite steps (for some m), the decision-maker arrives at a "good" state z_m such that for all (large) n, there is a play $s \in S(z_m)$ such that $\gamma_t(s)$ is above $v^*(z_0)$ for any $t \leq n$.

The ε-optimal play is then constructed by blocks, and on each of them: first a finite m steps to reach a "good" position, and then a large n steps for an average payoff above v^*. The optimality is obtained if m can be taken uniformly bounded (such that the duration used to reach a good position is negligible) and the limit value $v^*(z_{m+n})$ after each block is nondecreasing. A precise result is as follows:

Lemma 2.5.5. $\forall \varepsilon > 0, \exists M \geq 0, \exists K \geq 1, \forall z_0 \in Z, \exists m \leq M, \forall n \geq K, \exists s = (z_t)_{t \geq 1} \in S(z_0)$ *such that*

$$\min_{1 \leq t \leq n} \gamma_{m,t}(s) \geq v^*(z_0) - \varepsilon/2 \textit{ and } v^*(z_{m+n}) \geq v^*(z_0) - \varepsilon.$$

□

Remark 2.5.6. *1) The first inequality "*$\min_{1\le t\le n}\gamma_{m,t}(s)\ge v^*(z_0)-\varepsilon/2$*" takes a similar form as Lemma 2.3.6 (Lehrer and Sorin 1992) and also as Proposition 2 in Rosenberg et al. 2002 for MDP with imperfect observations on the state.*
2) The proof for "$v^*(z_{m+n})\ge v^*(z_0)-\varepsilon$*", i.e. the target function* v^* *is nondecreasing after each block, relies on the first inequality, where the definition of* $w_{m,n}$ *comes into play.*

Comments [Comparison with Mertens and Neyman [34]]
Mertens and Neyman proved the existence of uniform value for stochastic games with infinite state space provided that $\lambda\mapsto v_\lambda$ *has bounded variation. Here, the function* $w_{m,n}$ *(whose family is totally bounded) plays the role of* v_λ *in constructing the* ε*-optimal strategies* σ^ε*.*
– On each block, σ^ε *follows an optimal strategy in some* λ*-discounted game in Mertens and Neymen [34]*; and here, σ^ε follows an optimal play for $w_{m,n}$ with some m,n.
– The discount factor λ *for each block is updated using the payoff stream during the past block so as to control the average payoff nearly above the limit value. At the same time, the total variation property of* v_λ *helps to establish an approximate submartingale inequality (nondecreasing) for* v_λ *(hence the limit value for* λ *small). These correspond to the two inequalities established in Lemma 2.5.5.*

In the compact nonexpansive case, the condition in Theorem 2.5.4 is satisfied, thus

Corollary 2.5.7. *Let* Γ *be a compact nonexpansive DP problem, then uniform value exists in* Γ*, and it is equal to* v^**.*

Extension to optimal control problems

Consider the optimal control problem $\mathcal{J}=\langle U,g,f\rangle$ described in Subsection 2.3.2.

Definition 2.5.8. $\mathcal{J}$ *is a* ***compact nonexpansive*** *if it satisfies the following three conditions:*
A.1) the control dynamic has a compact invariant set Y*:* $\mathbf{y}(t,\mathbf{u},y_0)\in Y,\ \forall t\ge 0,\forall\mathbf{u}\in\mathcal{U},\forall y_0\in Y$*.*
A.2) the running cost function g *does not depend on* u *and is continuous in* y*.*
A.3) the control dynamic is nonexpansive, i.e.,

$$\forall y_1,y_2\in\mathbb{R}^d,\ \sup_{a\in U}\inf_{b\in U}\langle y_1-y_2,f(y_1,a)-f(y_2,b)\rangle\le 0.$$

The nonexpansive condition is used to obtain the following useful result.

Lemma 2.5.9. *[Quincampoix and Renault 2011] Let the control dynamic be nonexpansive, then:*

$$\forall y_1,y_2\in\mathbb{R}^d,\forall\mathbf{u}\in\mathcal{U},\exists\mathbf{v}\in\mathcal{U}\ s.t.\ \|\mathbf{y}(t,\mathbf{u},y_1)-\mathbf{y}(t,\mathbf{v},y_2)\|\le\|y_1-y_2\|,\ \forall t\ge 0.$$

An analogous result to Theorem 2.5.4 in continuous time is obtain:

Theorem 2.5.10 (Quincampoix and Renault 2011)**.** *Let* $\mathcal{J}$ *be a compact nonexpansive optimal control problem, then uniform value exists, i.e.,*

$$\forall\varepsilon,\exists T_0>0,s.t.\ \forall y_0\in Y,\exists\mathbf{u}\in\mathcal{U}:\ \frac{1}{t}\int_0^t g(t,\mathbf{u},y_0)\mathrm{d}t\le V^*(y_0)+\varepsilon,\forall t\ge T_0.$$

The proof of Quincampoix and Renault [42] uses the assumptions (compact and nonexpansive) to obtain the result in continuous time similar to Lemma 2.5.5.

2.6 TV-uniform value in compact nonexpansive case

When the dynamic programming problem is compact nonexpansive, Renault and Venel [47] proved the existence of TV-uniform value when the decision-maker uses mixed strategies. Their results are applied to models of standard Markovian decision process with finite states (MDP), Markovian decision process with finite states and imperfect observations (POMDP) and zero-sum repeated games with an informed controller. They introduced a new distance d^* for the probability spaces (decision-maker or player's beliefs) such that the auxiliary problems are compact nonexpansive. Moreover, a new characterization for the limit value is provided *via* invariant measures.

2.6.1 Compact nonexpansive gambling house

A basic model in consideration is a (compact) stochastic dynamic programming problem named *gambling house*. $\Gamma = \langle Z, F, r\rangle$ is defined by: (Z, d) is a compact metric space, $r : Z \to [0, 1]$ is a continuous reward function, and the transition correspondence $F : Z \rightrightarrows \Delta_f(Z)$ is stochastic. Starting at $z_0 \in Z$, the decision-maker chooses some $u_1 \in F(z_0)$, then $z_1 \in Z$ is realized by the probability law u_1, and the stage reward is $r(z_1)$. At each stage $t \geq 1$, the decision-maker chooses $u_t \in F(z_{t-1})$, z_t is realized by u_t, and the stage reward is $r(z_t)$.

Definition 2.6.1. *The Kantorovich-Rubinstein distance d_{KR} on $\Delta(Z)$ is:*

$$\forall u, v \in \Delta(Z),\ d_{KR}(u, v) = \sup_{f\in E_1} \Big| \int_Z f(p)\mathrm{d}u(p) - \int_Z f(p)\mathrm{d}v(p)\Big|,$$

where E_1 is the set of 1-Lipschitz functions for (Z, d).

The gambling house Γ is *nonexpansive* if

$$\forall z, z' \in Z, \forall u \in F(z), \exists u' \in F(z'),\ s.t.\ d_{KR}(u, u') \leq d(z, z').$$

Conisider now the gambling house $\Gamma = \langle Z, r, F\rangle$ that is compact nonexpansive. The first main result of Renault and Venel [47] consists of two parts: first introduce some function w^* *via* invariant measures and prove that it is the TV- limit value; second prove the existence of TV-uniform value.

The mixed extension of F is the correspondence $\hat{F}$ from $\Delta_f(Z)$ to itself, defined as:

$$\hat{F}(u) = \Big\{ \sum_{x\in X} u(x)f(x) :\ s.t.\ f : Z \to \Delta_f(Z) \text{ and } f(x) \in \Delta_f(F(x)), \forall x \in Z\Big\},\ \forall u \in \Delta_f(Z).$$

Extend r to $\Delta(Z)$ linearly by $r(u) = \int_Z r(z)du(z), \forall u \in \Delta(Z)$. Define the function w^* to be:

$$\forall z \in Z,\ w^*(z) = \inf \big\{w(z)|w : \Delta(Z) \to [0, 1] \text{ affine continuous } s.t.$$
$$(1)\forall z' \in Z, w(z') \geq \sup_{u\in F(z')} w(u) \text{ and } (2)\forall u \in R, w(u) \geq r(u)\big\},$$

where $R = \big\{u \in \Delta(Z) : (u, u) \in \mathrm{cl}(Graph\hat{F})\big\}$ is the set of *invariant measures* of Γ.

Theorem 2.6.2 (Renault and Venel 2013)**.** *Γ has a TV-limit value which is w^*.*

A sketch of proof

The proof of Theorem 2.6.2 employs some "comparison principle". Take v any accumulation point of the family $\{v_\theta\}$ with $\|v_{\theta^k} - v\|_\infty \to_{k\to\infty} 0$ and $TV(\theta^k) \to_{k\to\infty} 0$ for some (θ^k), then:
A) v satisfies conditions (1) and (2), thus $v \geq w^*$. To prove that v satisfies condition (2), the nonexpansive property is used to construct for any $u \in R$ a play (for $\hat{F}$) staying around u;
B) any w satisfying conditions (1) and (2) is larger than v, thus $w^* \geq v$. For $p \in Z$, let $\sigma^k = (u_t^k)_{t\geq 1}$ be an ε-optimal play for $v_{\theta^k}(p)$, $\forall k$. Define $u(k) = \sum_{t\geq 1} \theta_t u_t^k$ and use "$TV(\theta^k) \to_{k\to\infty} 0$" to obtain a limit point u of $(u(k))$ which is an invariant measure. w satisfies conditions (1) and (2) thus $w(p) \geq w(u) \geq r(u) \geq v(p) - \varepsilon$. □

Comments

1) The existence of TV-limit value for Γ (and its equality to $v^*(z_0) = \inf_\theta \sup_m v_{m,\theta}(z_0), \forall z_0$) can be deduced from Corollary 2.4.9 of Theorem 2.4.3. Indeed, consider the deterministic gambling house (dynamic programming) $\hat{\Gamma} = \langle \Delta(Z), r, \hat{F}\rangle$: it is "equivalent" to $\Gamma = \langle Z, r, F\rangle$ and is compact nonexpansive for d_{KR}. This theorem provides a second characterzation of the TV-limit value.

2) The characterization of the limit value as a unique solution to certain functional (in)equalities is close to the so-called MZ operator (Mertens and Zamir [36]), which is used to study the limit value in repeated games with incomplete information on both sides. See Cardaliaguet et al.[16] for the use of "comparison principle" in the asymptotic analysis of other classes of repeated games.

Definition 2.6.3. *A **mixed play** at z_0 is a sequence $\sigma = (u_1, ..., u_t, ...)$ in $\Delta_f(Z)$ such that $u_1 \in \Delta_f(F(z_0))$ and $u_{t+1} \in \hat{F}(u_t)$ for each $t \geq 1$. Denote by $\Sigma(z_0)$ the set of mixed plays at z_0.*

Theorem 2.6.4 (Renault and Venel 2013)**.** *Γ has a TV-uniform value w^* in mixed plays, i.e.,*

$$\forall \varepsilon > 0, \exists \alpha > 0, \forall z_0 \in Z, \exists \sigma \in \Sigma(z_0),\ s.t.\ \gamma_\theta(\sigma) \geq v^*(z_0) - \varepsilon, \forall \theta \in \Delta(\mathbb{N}^*) \text{ with } TV(\theta) \leq \alpha.$$

A sketch of proof

It is equivalent to work on the deterministic problem $\hat{\Gamma} = \langle \hat{Z}, r, \hat{F}\rangle$ with $\hat{Z} := \Delta(Z)$.

1) A first step is to establish the following result: $\forall \varepsilon > 0, \exists n_0,\ s.t.$

$$\forall u_0 \in \hat{Z}, \forall T \geq 0, \exists \sigma^T = (u_t^T)_{t\geq 1} \in \Sigma(u_0), \gamma_{t,n_0}(\sigma^T) \geq v^*(u_0) - \varepsilon,\ \forall t \in \{1, ..., T\}. \tag{2.6.1}$$

To obtain 2.6.1, one can
– define and prove by minmax theorem the following (for some $\beta(\mu, n) \in \Delta(\mathbb{N}^*)$):

$$h_{T,n}(u_0) =_{def} \sup_{\sigma\in\Sigma(u_0)} \inf_{0\leq t\leq T} \gamma_{t,n}(\sigma) = \inf_{\mu\in\Delta\left([0,T]\right)} \sup_{\sigma\in\Sigma(u_0)} \sum_{0\leq t\leq T} \mu_t \gamma_{t,n}(\sigma) = \inf_{\mu\in\Delta\left([0,T]\right)} v_{\beta(\mu,n)}(u_0),$$

– show that $TV\big(\beta(\mu, n)\big) \to_{n\to\infty} 0$ uniformly in T and μ, thus $\inf_{T\geq 0} h_{T,n} \to_{n\to\infty} v^*$ uniformly.

2) The object is to obtain a play such that the average payoff on each (consecutive) block of length n_0 is above the limit value v^*. That is,

$$\forall \varepsilon > 0, \exists n_0,\ s.t. : \forall u_0 \in \hat{Z}, \exists \sigma' = (u'_t)_{t\geq 1} \in \Sigma(u_0), \gamma_{t,n_0}(\sigma') \geq v^*(u_0) - 2\varepsilon,\ \forall\ t \geq 0. \tag{2.6.2}$$

Let $(u^{T_k})_{k\geq 1}$ be a sequence of "optimal" plays defined as in (2.6.1). Consider its limit $\overline{u} = (\overline{u}_t)_{t\geq 1}$: for each $t \geq 1$, $\bar{u}_t$ is an accumulation point of $(u_t^{T_k})_k$. Moreover, the nonexpansive property helps to construct a play $\sigma' = (u'_t)$ that is ε-close to $\overline{u}$ along the play, which implies (2.6.2).

3) The last point is to show that σ' is 3ε-optimal. Here the same argument as in Proposition 2.3.9 is employed. The idea is that: on each block (consecutive) of length n_0, one compares the θ-valued payoff (normalized) with the Cesàro mean (which is above $v^*(u_0)$ by 2.6.2), and finds out that the total difference of the infinite blocks is controlled by $TV(\theta)$.

More precisely, we fix u_0 and σ', and consider the uncontrolled problem Γ' defined on the play. By (2.6.1), one obtains that the n-stage value (payoff) of Γ' converges (consider $\liminf$) uniformly (on the trajectory $\{u'_t\}$) to some φ which is above $v^*(u_0)-2\varepsilon$ everythere, thus by Proposition 2.3.9, the θ-value (payoff) of Γ' at u_0 is above $v^*(u_0) - 3\varepsilon$ for all θ with $TV(\theta)$ sufficiently small ($\leq \varepsilon$). □

Comments

The approach here for the uniform analysis is quite different from Theorem 2.5.4.

1) First, note that the auxiliary value function $h_{T,n}$ is defined differently from $w_{m,n}$: $h_{T,n}$ is defined as if an adversary chooses a bad starting stage $t \leq T$, while $w_{m,n}$ is defined as if the adversary chooses a bad averaging length $t \leq n$. Nevertheless, both quantities are linked with the uniform optimal play.

2) The construction of the ε-optimal play is not by blocks (even though a comparison of the average payoff is by blocks, to the Cesàro mean). Rather, the play is obtained as the "limit" of a sequence of optimal plays u^T in finite horizons, where each u^T is optimal for h_{T,n_0} for a sufficiently large n_0.

3) The use of mixed plays is to obtain some convexity such that minmax theorem applies for $h_{T,n}$. Indeed, the proof replies on the fact that the mixed extension $\hat{F}$ of F is affine, i.e. $\hat{F}(\alpha u + (1-\alpha)u') = \hat{F}(u) + (1-\alpha)\hat{F}(u')$, $\forall u, u' \in Z, \forall \alpha \in [0,1]$, thus $\Sigma(u_0) \subseteq \hat{Z}^\infty$ is convex.

Extension to continuous time framework

Consider an optimal control problem $\mathcal{J} = \langle U, f, g\rangle$ defined in Subsection 3.2.2, which is compact nonexpansive (cf. Def. 2.5.8). Li [29] (Chapter 4) proves that the TV-uniform value exists in $\mathcal{J}$, using a similar approach as Theorem 2.6.2, which is defferent from Theorem 2.5.10.

Definition 2.6.5. *A **random control** is a pair $((\Omega, \mathcal{B}(\Omega), \lambda), \mathbf{u})$, where $(\Omega, \mathcal{B}(\Omega), \lambda)$ is some standard Borel probability space and $\mathbf{u} : \Omega \times [0,\infty) \to U$ is a Borel measurable mapping.*

Theorem 2.6.6 (Li 2015)**.** *A compact nonexpansive optimal control problem $\mathcal{J}$ has a TV- uniform value V^*, i.e. for each $\varepsilon > 0$ there is some $\eta > 0$, $S > 0$ and a random control $((\Omega, \mathcal{B}(\Omega), \lambda), \mathbf{u})$ such that:*

$$\forall \theta \in \Delta(\mathbb{R}_+), \quad \Big(\sup_{0\leq s\leq S} TV_s(\theta) \leq \eta \Longrightarrow (\forall y_0 \in Y, \int_\Omega \gamma_\theta(y_0, \mathbf{u}(\omega,\cdot))\mathrm{d}\lambda(\omega) \leq V(y_0) + \varepsilon)\Big).$$

Applications to MDP with a finite set of states

Consider the standard model of Markovian decision process (MDP) $\Psi = \langle K, A, p_0, q, g\rangle$ defined in Subsection 3.2.1 with the state space K being finite. Let $\Gamma = \langle Z, r, F\rangle$ be the auxiliary (deterministic) dynamic programming problem that is defined equivalent to Ψ.

Set $d((p,x),(p',x')) = \max\{\|p-p'\|_1, |x-x'|\}$ for any $(p,x),(p,x') \in Z = \Delta(K)\times[0,1]$. Then the DP problem Γ is *compact nonexpansive*: (Z,d) is compact, r is continuous, and moreover F is nonexpansive for d. Previous results for compact nonexpansive DP's are applied for Γ to obtain (Theorem 2.6.4 and Corollary 2.5.7):

Theorem 2.6.7 (Renault 2011, Renault and Venel 2013)**.** *MDP with a finite set of states, played with behavior strategies, has a TV- uniform value. MDP with a finite set of states, played with pure strategies, has a uniform value.*

2.6.2 A distance on probability spaces and its applications to POMDP and repeated games with an informed controller

We consider some applications of the results of compact nonexpansive gambling house to POMDP and repeated games with an informed controller. In order to apply Theorem 2.6.4, one may define an associated auxiliary dynamic programming (a deterministic gambling house) for the problem with player's belief as the state variable. However, the auxiliary transition correspondence might not be nonexpansive for d_{KR}. For this aim, Renault and Venel [47] introduced a new distance d_* on the probability spaces such that the auxiliary problem is compact nonexpansive.

Let $(X, \|\cdot\|)$ be a compact set of some normed vector space.

Definition 2.6.8. *Define the distance d_* on $\Delta(X)$ to be: for any $u,v \in \Delta(X)$,*

$$d_*(u,v) = \sup_{f\in D_1} \Big| \int_Z f(p)\mathrm{d}u(p) - \int_Z f(p)\mathrm{d}v(p)\Big|,$$

where $\Big(\mathcal{C}(X)$ denotes the set of countinuous functions on $X\Big)$

$$D_1 = \Big\{ f \in \mathcal{C}(X) \mid \forall x,y, \in X, \forall\alpha, \forall\beta \geq 0,\ \alpha f(x) - \beta f(y) \leq \|\alpha x - \beta y\| \Big\}.$$

The distance d_* is of particular interest when K is a finite set and $X = \Delta(K)$ is the simplex. The following mapping appears in optimization problems with incomplete information on the state.

Definition 2.6.9. *For each finite S, define the* ***posterior mapping (disintegration)*** *$\psi_S : \Delta(K\times S) \to \Delta_f(X)$ by:*

$$\psi_S(\pi) = \sum_{s\in S} \pi(s)\delta_{\bar{p}(s)},$$

where for all s, $\pi(s) = \sum_k \pi(k,s)$ and $\bar{p}(s)$ is the posterior on K given s.

Theorem 2.6.10 (Renault and Venel 2013)**.** *The mapping $\psi_S : (\Delta(K\times S), \|\cdot\|_1) \to (\Delta(X), d_*)$ is 1-Lipschitz (nonexpansive). d_* metrizes the weak-$*$ topology on $\Delta(X)$.*

Comments

1) By definition $D_1 \subseteq E_1$, thus $d_* \leq d_{KR}$.

2) When $X = \Delta(K)$, the set D_1 in defining d_* can be replaced by

$$D_0 = \Big\{ u : X \to \mathbb{R} | \forall p \in X, u(p) = Val(\sum_k p^k G^k),\ \textit{for some matrices } (G^k) \textit{ with values in } [-1,+1] \Big\},$$

where Val is the value operator for a matrix game, and $p \mapsto u(p) = Val(\sum_k p^k G^k)$ defines the value of the "non revealing game" of the incomplete information game $(p, \{G^k\})$, i.e. $p \in \Delta(K)$ and k is chosen according to p; the realization of k is communicated to player 1 only and the matrice game G^k is then repeated.

The model of MDP's with partial observations (POMDP)

In a POMDP, the decision-maker does not perfectly observe the state, rather he receives a (random) signal at each stage, which depends on the current state and the action.

The set of finite state K, the set of actions A, and the reward function $g : K \times A \to [0,1]$ are given as before. Let S be a nonempty set of signals, and q be a transition function from $K \times A$ to $\Delta_f(S \times K)$. This POMDP $\Psi(p_0)$ is played as follows: k_1 is chosen according to the initial distribution $p_0 \in \Delta(K)$ but is not told to the decision-maker. At every stage $t \geq 1$, the decision-maker takes an action $a_t \in A$ which (together with the current state k_t), induces a (unobserved) stage payoff $g(k_t, a_t)$. Then the pair (s_t, k_{t+1}) is chosen according to the distribution $q(\cdot|k_t, a_t)$, and the signal s_t is communicated to the decision-maker. The new state is k_{t+1} and the play proceeds to the next stage.

A behavior strategy is a sequence $\sigma = (\sigma_t)_{t\geq 1}$ where for each t, $\sigma_t : (A \times S)^{t-1} \to \Delta_f(A)$. Any σ defines, together with p_0 and q, a unique probability distribution over $(K \times A \times S)^\infty$. Let $v_\theta(p_0)$ be the θ-value of $\Psi(p_0)$ for any evaluation θ.

As in the standard MDP, we define a dynamic programming problem equivalent to Ψ with the decision-maker's beliefs as the auxiliary state space. We write $X = \Delta(K)$. To introduce a deterministic DP, we need a larger state space than X. Indeed, consider the current state k following a distribution $p \in X$, and the decision-maker takes an action $a \in A$. Since q has finite support, this defines a vector of distributions $\big(\hat{q}^s(p,a)\big)_s$ in X, where $\hat{q}^s(p,a)\big)_s$ denotes the decision-maker's posterior belief of p over the new state $k' \in K$ after receiving the signal s. We write this probability as[6] $\big(P_{p,a}(s) := \sum_{k,k'\in K} p^k q(k', s|k, a)\big)$

$$\hat{q}(p,a) = \sum_{s\in S} P_{p,a}(s)\delta_{\hat{q}(p,a)} \in \Delta_f(X).$$

Define from $\Psi(p_0)$ the auxiliary DP $\Gamma(z_0)$ by:
– the set of space is $Z = \Delta_f(X) \times [0,1]$;
– the initial state is $z_0 = (\delta_{p_0}, 0)$;
– the reward function is $r : Z \to [0,1]$ with $r(u,x) = x$ for all (u,x) in Z;
– the transition correspondence $F : Z \rightrightarrows Z$ is: for every $z = (u,x)$ in Z,

$$F(z) = \Big\{\big(H(u,f), R(u,f)\big) \mid f : X \to \Delta_f(A)\Big\},$$

where

$$H(u,f) = \sum_{p\in X} u(p)\Big(\sum_{a\in A} f(p)(a)\hat{q}(p,a)\Big) \in \Delta_f(X),$$
$$\text{and} \quad R(u,f) = \sum_{p\in X} u(p)\Big(\sum_{k\in K, a\in A} p^k f(p)(a) g(k,a)\Big).$$

Let $\tilde{v}_\theta$ be the θ-value function of Γ. Since the transition mapping $F(\cdot)$ depends on z only through its first component, then for any $z = (u,x)$, $\tilde{v}_\theta(z) = v_\theta(u)$ (which is affine

6. $\hat{q}(p,a)$ can be written as $\psi_S(\pi^{p,a})$ for some $\pi^{p,a} \in \Delta(K \times S)$ defined as the joint probability distribution over $K \times S$ that is induced by p and a, i.e. $\forall (k,s) \in K \times S$, $\pi^{p,a}(k,s) = \int_X \Big(\big(p(u)\big)^k \sum_{k'} q(s,k'|k,a)\Big) du$.

on $\Delta_f(X)$). Moreover, anything that can be guaranteed by the decision-maker in $\Gamma(u,0)$ can be guaranteed in $\Psi(u)$.

In order to deduce the existence of TV- uniform value for Γ, thus for Ψ, the metric d on $Z = \Delta_f(X) \times [0,1]$ is introduced: $d((u,x),(u',x')) = \max\{d_*(u,u'), |x-x'|\}$ where the distance d_* on $\Delta_f(X)$ is defined as in Definition 2.6.8. Then applying[7] Theorem 2.6.10 (plus a duality formula for d_*), one obtains that F is affine (as a mixed extension) and nonexpansive for d.

Theorem 2.6.11 (Renault and Venel 2013)**.** *POMDP with a finite state of space, played with behavior strategies, have a TV- uniform value.*

Comment

1) The existence of uniform value for POMDP with a finite state space was established by Rosenberg et al. [49] for a finite action set (or a compact action set and with some continuity of g and q) and any signal set. An application of Theorem 2.5.4 proves the existence of uniform value for POMDP with a finite state space and any action set (Renaut [44]). This result generalizes both results to TV-uniform value.

2) The three proofs use $\Delta(\Delta(K))$ as the auxiliary space and all of them employ randomization in the construction of ε-optimal strategies. The use of lotteries are different. In both Renault [44] and the proof here, the convexity is needed to have the auxiliary correspondence F affine, thus a minmax theorem applies. In Rosenberg et al.[49], the use of lotteries is partially due to the difficulty to find a distance on $\Delta(\Delta(K))$ such that the transition correspondence in the auxiliary DP problem is 1-Lipschitz (non-expansive).

3) Now with d_*, the auxiliary DP problem satisfies the nonexpansive property, one may wonder whether the randomness is still needed for an ε-optimal play in POMPD with a finite state space. The recent article by Venel and Zillioto [62] solved this problem by constructing pure ones.

The model of repeated games with an informed controller

A general model $\mathcal{G} = \langle K, I, J, C, D, q, g\rangle$ of zero-sum repeated games (cf. Mertens *et al.* [35]) consists of[8]:
– a finite set of states K;
– two finite set of actions I and J, and two finite set of signals C and D;
– a transition function $q : K \times I \times J \to \Delta(K \times C \times D)$;
– a payoff function $g : K \times I \times J \to [0,1]$.

The game $\mathcal{G}(\pi)$ with an initial state $\pi \in \Delta(K \times C \times D)$ is played as follows. Initially, the triple (k_1, c_1, d_1) is drawn according to π. At stage 1: player 1 learns c_1 and player 2 learns d_1. Then simultaneously player 1 chooses an action $i_1 \in I$ and player 2 chooses an action $j_1 \in J$. The stage payoff is $g(k_1, i_1, j_1)$, and the new triple (k_2, c_2, d_2) is drawn according to $q(k_1, i_1, j_1)$. The game proceeds to stage 2: player 1 observes c_2, and player 2 observes d_2 etc...

A behavior strategy for player 1 is a sequence $\sigma = (\sigma_t)_{t\geq 1}$ where for each t, $\sigma_t : (I \times C)^{t-1} \times C \to \Delta(I)$. Similary for a behavior strategy τ for player 2. Let Σ be the set of behavior stratagies for player 1, and $\mathcal{T}$ for player 2. Given π, any strategy profile

7. The proof in Renault and Venel [47] for this result is not explicitly a direct application of Theorem 2.6.4. They transformed a POMDP Ψ into a standard MDP $\hat{\Psi}$ with $K = \Delta(X)$ the state space, and then proved the existence of TV- uniform value for $\hat{\Psi}$ using the distance d_*. In the proof of the later result for $\hat{\Psi}$, an equivalent deterministic dynamic programming Γ is constructed and then similar proof as Theorem 2.6.4 follows. Our same comment applies for the next application to repeated games.

8. The finiteness assumption on the data K, I, J, C, D is not necessary: indeed the result extends to a measurable setting as long as the value $v_\theta(\pi)$ exists for a fixed evaluation θ.

(σ, τ) defines a unique probability distribution $\mathbb{P}^{\pi}_{\sigma,\tau}$ over $(K \times C \times D \times I \times J)^{\infty}$. For any $\theta \in \Delta(\mathbb{N}^*)$, the θ-evaluated payoff of (σ, τ) is $\gamma_\theta(\pi, \sigma, \tau) = \mathbb{E}^{\pi}_{\sigma,\tau}\Big[\sum_{t\geq 1} \theta_t g(k_t, i_t, j_t)\Big]$. Let $v_\theta(\pi)$ be the θ-value of $\mathcal{G}(\pi)$ for any evaluation θ:

$$v_\theta(\pi) = \inf_{\tau\in\mathcal{T}} \sup_{\sigma\in\Sigma} \gamma_\theta(\pi, \sigma, \tau) = \sup_{\sigma\in\Sigma} \inf_{\tau\in\mathcal{T}} \gamma_\theta(\pi, \sigma, \tau),$$

which exists by Sion's minmax theorem. The limit value and TV-limit value are defined as in previous cases.

Definition 2.6.12. *The repeated game $\Gamma(\pi)$ has a TV-**limit value** if:*

$$\forall \varepsilon > 0, \exists \alpha > 0,\ s.t. \forall \theta \in \Delta(\mathbb{N}^*),\ TV(\theta) \leq \alpha,\ |v_\theta(\pi) - v_\theta(\pi)| \leq \varepsilon.$$

Definition 2.6.13. *The repeated game $\Gamma(\pi)$ has a TV-**uniform value** if it has a TV-limit value v^* and each player guarantees it, that is, for all $\varepsilon > 0$, there is some $\alpha > 0$ and $(\sigma^*, \tau^*) \in \Sigma \times \mathcal{T}$, s.t.: for all $\theta \in \Delta(\mathbb{N}^*)$ with $TV(\theta) \leq \alpha$,*

$$\gamma_\theta(\pi, \sigma^*, \tau^*) + \varepsilon \geq v^*(\pi) \geq \gamma_\theta(\pi, \sigma, \tau^*) - \varepsilon, \quad \forall (\sigma, \tau) \in \Sigma \times \mathcal{T}.$$

$\Gamma(\pi)$ is called a *repeated game with an informed controller* if the following two conditions are satisfied:

H.1 *Player 1 is informed*, in the sense that he can always deduce the state and player 2's signal from his own signal.

H.2 *Player 1 controls the transition*, in the sense that the marginal on $K \times D$ of the transition q does not depend on player 2's action.

The second assumption implies that player 2's action has no influence on his own information, thus he has no influence on his belief about the state or about player 1's belief about his belief.

Theorem 2.6.14 (Renault and Venel 2013)**.** *Let $\mathcal{G}$ be a zero-sum repeated game satisfying **H.1** and **H.2**, then $\mathcal{G}$ has a TV-uniform value.*

A sketch of proof

The main idea of the proof is to find an "equivalent" dynamic programming problem that has a TV-uniform value, and the ε-optimal plays are transformed to ε-optimal strategies for player 1 in the repeated game. Moreover, using the fact that player 2 has no influence on the transition, he can play by independent blocks of equal length, on each of them the average payoff is above the limit value.

Let $\pi^{K\times D}$ be the marginal of π on $K \times D$, and we write $\hat{\pi} = \psi_S(\pi^{K\times D}) \in \Delta_f(X)$ with $X = \Delta(K)$. $\hat{\pi}$ is thus the initial distribution of player 2's beliefs over K.

1) Define from $\mathcal{G}(\pi)$ a standard Markovian decision process (MDP) $\Psi(\hat{\pi})$ where the state space is $X = \Delta(K)$, player 2's beliefs over K, and the reward is the minimal expected payoff given player 1's actions (player 2 is playing a best response).

2) Prove by recursive formula that the θ-values are equal in both probems: $v_\theta(\pi) = \hat{v}_\theta(\hat{\pi})$.

3) Show that the MDP Ψ has a TV-uniform value. For this, one can define from Ψ an equivalent deterministic DP problem Γ (with a state space $Z = \Delta_f(X) \times [0, 1]$) which is compact nonexpansive (using the distance d_* defined on $\Delta(X)$). Then the proof as in Theorem 2.6.4 applies for Γ (thus Ψ) to have a TV-uniform value $\hat{v}^*$. This implies that $\mathcal{G}(\pi)$ has a TV-limit value $v^*(\pi) = \hat{v}^*(\hat{\pi})$.

4) Player 1 guarantees $v^*(\pi)$ in $\mathcal{G}(\pi)$ by mimicing ε-optimal plays in $\Gamma(\delta_{\hat{\pi}})$.

5) Player 2 guarantees $v^*(\pi)$ in $\mathcal{G}(\pi)$:
– First, $v^*(\pi)$ satisfies the equation (use Prop. 2.4.6) : $v^*(\pi) = \inf_n \sup_m v_{m,n}(\pi)$:
– Next, take n_0 with $v_{m,n}(\pi) \leq v^*(\pi) + \varepsilon$ for all $m \geq 0$. Consider player 2 playing by blocks of length n_0. The fact that his action has no infuence on the transition implies that there is a strategy guaranteeing him the n_0-stage average payoff on each block at most $v^*(\pi) - \varepsilon$.
– Finally, same argument as *Point 3)* in "Sketch of Proof of Theorem 2.6.4" applies. □

Comments

1) The model of ***repeated games with an informed controller*** *is introduced in Renault [45]. He proved the existence of uniform value by an application of Theorem 2.5.4, as the same approach here.*

2) Renault [45] unifies two models studied in the literature: Renault [43] for ***Markov chain games with lack of information on one side*** *and Rosenberg et al. [50] for* ***stochastic games with incomplete information and an informed controller***:
– in the first model, the initial state is chosen according to a probability distribution and then the sequence of states follow a (uncontrolled) Markov chain. At each stage the current state is learned by player 1, and player 2 knows only the initial distribution;
– in the second model[9]*, the state k is decomposed into two components (ℓ, ω): ℓ is the state of the nature, which is chosen (kept fixed) by some probability distribution and is communicated to player 1 only; ω is the public stochastic state, which follows a Markov process controlled by player 1.*

3) Being different from the approach in Renault [45] and Renault and Venel [47], neither Rosenberg et al. [50] nor Renault [43] employed a reduction of the repeated games to DP problems, rather, their proofs use basic tools in ***repeated games with incomplete information on one side*** *à la Aumann and Maschler [5]: non-revealing games and concavification operator, approachability, etc. The proof for player 2 to guarantee the limit value by blocks appeared already in Aumann and Maschler [5], and also in Renault [43].*

4) The approach is generalized by Gensbittel et al. [21] to a more general setup: player 1's first-order belief is more accurate than player 2's second-order belief, and player 1 controls the evolution of player 2's second-order beliefs.

9. Rosenberg *et al.* [50] considered also the other model where player 2–the uninformed player– controls the transition of the process. The uniform value does not exist and they characterized $Maxmin$ and $Minmax$ of the infinite game.

Chapter 3

Valeur limite pour le problème de contrôle optimal avec évaluations générales

Résumé Nous considérons le problème de contrôle optimal où le coût de la trajectoire est évalué par une mesure de probabilité sur $\mathbb{R}_+$. En cas particulier, on prend la moyenne de Cesàro du coût sur un horizon fix. La limite de la fonction valeur avec la moyenne de Cesàro lorsque l'horizon tend vers l'infini est largement étudié dans la littérature. Nous abordons la question plus générale de l'existence d'une limite pour les fonctions valeur définies par des évaluations générales qui satisfont certaines conditions à long terme.

Pour ce faire, nous introduisons une condition de régularité asymptotique pour une suite de mesures de probabilité sur $\mathbb{R}_+$. Notre résultat principal est que, pour toute suite de mesures de probabilité sur $\mathbb{R}_+$ vérifiant cette condition, les fonctions valeur associées convergent uniformément si et seulement si cette famille est totalement bornée pour la norme uniforme.

En corollaire, on obtient l'existence d'une valeur limite (pour les évaluations générales) pour les systèmes de contrôle définis sur un domaine invariant compact et satisfaisant une certaine condition de non-expansivité.

Mots-clés Contrôle optimal, valeur limite, la valeur moyenne à long temps, évaluation générale

---

Ce chapitre est issu de l'article *Limit value for optimal control with general evaluations* en collaboration avec Marc Quincampoix et Jérôme Renault, et il est accepté pour publication dans la revue *Discrete and Continuous Dynamical System - Series A*.

Limit value for optimal control with general evaluations

joint with Marc Quincampoix (Brest) and Jérôme Renault (Toulouse)

To appear in *Discrete and Continuous Dynamical System - Series A*

Abstract. We consider optimal control problems where the running cost of the trajectory is evaluated by a probability measure on $\mathbb{R}_+$. As a particular case, we take the Cesàro average of the running cost over a fixed horizon. The limit of the value with Cesàro average when the horizon tends to infinity is widely studied in the literature. We address the more general question of the existence of a limit for values defined by general evaluations satisfying certain long-term condition.

For this aim, we introduce an asymptotic regularity condition for a sequence of probability measures on $\mathbb{R}_+$. Our main result is that, for any sequence of probability measures on $\mathbb{R}_+$ satisfying this condition, the associated value functions converge uniformly if and only if this family is totally bounded for the uniform norm.

As a byproduct, we obtain the existence of a limit value (for general evaluations) for control systems defined on a compact invariant domain and satisfying suitable nonexpansive property.

Keywords Optimal control, limit value, long time average value, general means

3.1 Introduction

Let U be a metric space. We consider a control system defined on $\mathbb{R}^d$ whose dynamic is given by

$$\mathbf{y}'(t) = f(\mathbf{y}(t), \mathbf{u}(t)) \tag{3.1.1}$$

where $f : \mathbb{R}^d \times U \to \mathbb{R}^d$ and $\mathbf{u}$ is a measurable function – called the control – from $\mathbb{R}_+$ to U. We will make later on assumptions on (3.1.1) ensuring that for any initial condition $\mathbf{y}(0) = y_0$, and any control $\mathbf{u}$, the equation (3.1.1) has a unique solution $t \mapsto \mathbf{y}(t, \mathbf{u}, y_0)$ defined on $\mathbb{R}_+$.

Let $\theta \in \Delta(\mathbb{R}_+)$ be a Borel probability measure on $\mathbb{R}_+$. To any pair $(y_0, \mathbf{u})$, we associate a θ-evaluated cost

$$\gamma_\theta(y_0, \mathbf{u}) = \int_{[0,+\infty)} g(\mathbf{y}(t, \mathbf{u}, y_0), \mathbf{u}(t)) \mathrm{d}\theta(t),$$

where $g : \mathbb{R}^d \times U \to \mathbb{R}$ is Borel measurable bounded. We call θ an *evaluation* throughout the article.

We will refer to the previously described optimal control problem by the short notation $\mathcal{J} = \langle U, g, f \rangle$. Denote by $\mathcal{U}$ the set of controls. Given $\theta \in \Delta(\mathbb{R}_+)$, the θ-value function of $\mathcal{J}$ is:

$$V_\theta(y_0) = \inf_{\mathbf{u}\in\mathcal{U}} \gamma_\theta(y_0, \mathbf{u}) \tag{3.1.2}$$

Specific evaluations include
Cesàro mean: $\forall t > 0$, $\overline{\theta}_t$ has a density $s \mapsto f_{\overline{\theta}_t}(s) = \frac{1}{t}\mathbb{1}_{[0,t]}(s)$, and the t-horizon value is

$$V_{\overline{\theta}_t}(y_0) = \inf_{\mathbf{u}\in\mathcal{U}} \frac{1}{t}\int_0^t g(\mathbf{y}(s, \mathbf{u}, y_0), \mathbf{u}(s))\mathrm{d}s$$

Abel mean: $\forall \lambda \in (0, 1]$, θ_λ has a density $s \mapsto f_{\theta_\lambda}(s) = \lambda e^{-\lambda s}$, and the λ-discounted value is

$$V_{\theta_\lambda}(y_0) = \inf_{\mathbf{u}\in\mathcal{U}} \int_0^{+\infty} \lambda e^{-\lambda s} g(\mathbf{y}(s, \mathbf{u}, y_0), \mathbf{u}(s))\mathrm{d}s$$

The limit of the above value functions as t tends to infinity or as λ tends to zero are well investigated in the control literature (cf. Alvarez and Bardi [1], Arisawa and Lions [2], Arisawa [3], Bensoussan [9], Gaitsgory [20], Khasminskii [24] and the references therein), which are often called ergodic control.

The study of the relation between the two limits (Cesàro and Abel) refers to Tauberian-type results. Oliu-Barton and Vigeral [40] proved a uniform Tauberian theorem for optimal control problems (without ergodic condition, as opposed to Arisawa [2]), *i.e.* $V_{\overline{\theta}_t}(y_0)$ converges uniformly (in y_0) as t tends to infinity if and only if $V_{\theta_\lambda}(y_0)$ converges uniformly (in y_0) as λ tends to zero, and in case of uniform convergence, both limit functions are the same [1]. The establishment of Tauberian theorem ensures that the limit value (whenever it exists) does not depend on the particular chosen evaluations (to be Cesàro or Abel mean). One motivation of our research is to investigate a more general family of evaluations for which a possible unique "general" limit value is defined.

Given $\theta \in \Delta(\mathbb{R}_+)$, the contribution of the interval $[T, +\infty)$ in the θ-evaluated cost vanishes as T becomes large. Thus the control problem is essentially interesting only on $[0, T_0]$ for certain T_0, which we roughly name the "duration" for the problem. In this article, we are interested in the long-term property of $\mathcal{J}$, *i.e.* the asymptotic behavior of the function $\theta \mapsto V_\theta$ when the "duration" of θ tends to infinity. In the particular examples of Cesàro mean and Abel mean, this corresponds to the convergence of $V_{\overline{\theta}_t}$ as t tends to infinity and of V_{θ_λ} as λ tends to zero. It is a priori unclear how to define the "duration" of a general evaluation θ over $\mathbb{R}_+$. If one simply assumes the expectation of θ to be large, we can obtain very different value functions, as is shown by the following

Example 5. *Consider the uncontrolled dynamic* $\mathbf{y}(t) = t$, *the running cost* $t \mapsto g(t) = \mathbb{1}_{\cup_{m=1}^{\infty}[2m-1,2m]}(t)$, *and two sequences of evaluations* $(\mu^k)_{k\geq 1}$ *and* $(\nu^k)_{k\geq 1}$ *with densities:* $f_{\mu^k} = \frac{1}{k}\mathbb{1}_{\cup_{m=1}^{k}[2m-1,2m]}$ *and* $f_{\nu^k} = \frac{1}{k}\mathbb{1}_{\cup_{m=1}^{k}[2m-2,2m-1]}$. *Clearly,* $V_{\mu^k} = 1$ *and* $V_{\nu^k} = 0$, $\forall k \geq 1$.

For this reason, we introduce an asymptotic regularity condition on evaluations, to express the "large duration" and the "asymptotic uniformity of distributions over $\mathbb{R}_+$", and we study the convergence of the value functions along a sequence of evaluations satisfying this condition.

1. See Buckdahn et al. [14] for a uniform Tauberian theorem in stochastic optimal control problems and Khlopin [25] for a uniform Tauberian theorem in differential games.

Definition 3.1.1. *For any $\theta \in \Delta(\mathbb{R}_+)$ and $s \geq 0$, we denote*

$$TV_s(\theta) = \sup_{Q \in \mathcal{B}(\mathbb{R}_+)} \big|\theta(Q) - \theta(Q+s)\big|,$$

where $\mathcal{B}(\mathbb{R}_+)$ is the set of Borel subsets of $\mathbb{R}_+$. A sequence of evaluations $(\theta^k)_k$ satisfies the long-term condition (LTC) if:

$$\forall S > 0, \quad \overline{TV}_S(\theta^k) \xrightarrow[k\to\infty]{} 0, \quad \textit{where} \quad \overline{TV}_S(\theta^k) = \sup_{0 \leq s \leq S} TV_s(\theta^k).$$

Our main result (Theorem 3.4.1) states that for any $(\theta^k)_k$ satisfying the LTC, $(V_{\theta^k})_k$ converges uniformly if and only if the family $\{V_{\theta^k}\}$ is totally bounded with respect to the uniform norm. Moreover, in this case, the limit is characterized by:

$$V^*(y_0) = \sup_{\theta \in \Delta(\mathbb{R}_+)} \inf_{s \in \mathbb{R}_+} \inf_{\mathbf{u} \in \mathcal{U}} \int_{[0,+\infty)} g\Big(\mathbf{y}(t+s, \mathbf{u}, y_0), \mathbf{u}(t+s)\Big) \mathrm{d}\theta(t), \quad \forall y_0 \in \mathbb{R}^d. \tag{3.1.3}$$

The above function V^* appears to be the unique possible long-term value function of the control problem.

The optimal control problem $\mathcal{J} = \langle U, g, f \rangle$ has a *general limit value* V^* if for any sequence $(\theta^k)_k$ satisfying the LTC, $(V_{\theta^k})_k$ converges uniformly to V^* as k tends to infinity.

As a byproduct of our main result, we obtain the existence of the general limit value for any control problem $\mathcal{J} = \langle U, g, f \rangle$ with a continuous running cost function g which does not depend on u and with a control dynamic (3.1.1) which is non-expansive and has a compact invariant set. This generalizes the already obtained result in Quincampoix and Renault [42] for optimal control with Cesàro mean.

Existing results in the literature are concerned mainly with the convergence of the t-horizon Cesàro mean values or the convergence of the λ-discounted Abel mean values. To the best of the authors' knowledge, this paper is the first to consider general long-term evaluations for optimal control problems [2].

Also it is worth pointing out that while many works (including cf. Alvarez and Bardi [1], Arisawa and Lions [2], Arisawa [3], Bensoussan [9], Gaitsgory [20], Khasminskii [24]) assume controllability or ergodicity conditions, the present approach does not reply on such conditions. This could be underlined by the fact that the limit value V^* may depend on the initial state y_0 (which does not occur under ergodic or controllability assumptions).

We also make here a link with the discrete time framework, in which an evaluation $\theta = (\theta_m)_{m \geq 1}$ is a probability measure over positive integers $\mathbb{N}^* = \mathbb{N} \backslash \{0\}$, and θ_t is the weight for the stage-t payoff. The analogous notion of *total variation* is defined for any $\theta \in \Delta(\mathbb{N}^*)$: $TV(\theta) = \sum_{m=1}^{\infty} |\theta_{m+1} - \theta_m|$ (cf. Sorin [59] and Renault [46]). Recently, the existence of the general limit value of dynamic optimization problems in several discrete time frameworks has been obtained in Renault [46], Renault and Venel [47] and Ziliotto [69]. Our work is partially inspired by Renault [46]. Similar tool within the proof appeared in Renault [44].

The article is organized as follows. Section 3.2 contains some preliminary notations and basic examples. The long-term condition is studied in Section 3.3. Section 3.4 contains our main result and its consequences. Two (counter)examples are also discussed. Section 3.5 is devoted to the proof of the main result. Some further discussions are given in Section 3.6.

2. In the context of stochastic optimal control, Goreac [23] obtained some Tauberian-type results for a family of evaluations satisfying some technical assumptions. In particular, only absolutely continuous evaluations w.r.t. the Lebesgue measures are considered. Bardi and Priuli [6] considered mean-fined games with the running cost evaluated by general probability distributions. However, their concern is not the asymptotic analysis.

3.2 Preliminaries

Consider the optimal control problem $\mathcal{J} = \langle U, g, f\rangle$. We make the following assumptions on g and f throughout the article:

$$\begin{cases} \text{the function } g : \mathbb{R}^d \times U \to \mathbb{R} \text{ is Borel measurable and bounded;} \\ \text{the function } f : \mathbb{R}^d \times U \to \mathbb{R}^d \text{ is Borel measurable, and satisfies:} \\ (*).\ \exists L \geq 0, \forall (y, \overline{y}) \in \mathbb{R}^{2d}, \forall u \in U,\ \|f(y,u) - f(\overline{y},u)\| \leq L\|y - \overline{y}\|, \\ (**).\ \exists a > 0, \forall (y,u) \in \mathbb{R}^d \times U,\ \|f(y,u)\| \leq a(1 + \|y\|). \end{cases} \tag{3.2.1}$$

Under (3.2.1), given any control $\mathbf{u}$ in $\mathcal{U}$ and any initial state $y_0 \in \mathbb{R}^d$, (3.1.1) has a unique absolutely continuous solution $t \mapsto \mathbf{y}(t, \mathbf{u}, y_0)$ defined on $[0, +\infty)$. As the running cost function $g : \mathbb{R}^d \times U \to \mathbb{R}$ is bounded, we can always assume that $g : \mathbb{R}^d \times U \to [0, 1]$ after some affine transformation.

Below we introduce several notations.

Reachable map R_t For any $y_0 \in \mathbb{R}^d$, the reachable map on $\mathbb{R}_+$, $t \mapsto R_t(y_0)$, is defined as:

$$R_t(y_0) = \Big\{ y \in \mathbb{R}^d | \exists\ \mathbf{u} \in \mathcal{U} :\ \mathbf{y}(t, \mathbf{u}, y_0) = y \Big\}. \tag{3.2.2}$$

$R_t(y_0)$ represents the set of states that the dynamic can reach via certain control at time t, starting from the initial state y_0 at time 0. We write $R^t(y_0) = \cup_{s=0}^t R_s(y_0)$ and $R(y_0) = \cup_{s=0}^\infty R_s(y_0)$. $R(y_0)$ is the set of states that can be reached at any finite time starting from y_0.

Image measure $\mathcal{T}_t\sharp\theta$ and the auxiliary value function $V_{\mathcal{T}_t\sharp\theta}$ Given $t \in \mathbb{R}$ and θ in $\Delta(\mathbb{R}_+)$, we use $\mathcal{T}_t\sharp\theta$ to denote the image (push-forward) measure of θ by the function $\mathcal{T}_t : s \mapsto s + t$, i.e.,

$$\mathcal{T}_t\sharp\theta(Q) = \theta\left(\mathcal{T}_t^{-1}(Q)\right) = \theta((Q - t) \cap \mathbb{R}_+),\ \forall Q \in \mathcal{B}(\mathbb{R}_+).$$

This leads us to write the t-shift θ-evaluated cost induced by a control $\mathbf{u}$ as follow:

$$\gamma_{\mathcal{T}_t\sharp\theta}(y_0, \mathbf{u}) = \int_{[0,+\infty)} g\Big(\mathbf{y}(s + t, \mathbf{u}, y_0), \mathbf{u}(s + t)\Big) \mathrm{d}\theta(s),\ \forall t \geq 0. \tag{3.2.3}$$

Taking on both sides of (3.2.3) the infimum over $\mathbf{u} \in \mathcal{U}$ and using the notation of reachable map R_t, we obtain the t-shift θ-value function

$$V_{\mathcal{T}_t\sharp\theta}(y_0) = \inf_{\mathbf{u}\in\mathcal{U}-} \gamma_{\mathcal{T}_t\sharp\theta}(y_0, \mathbf{u}) = \inf_{\overline{y}\in R_t(y_0)} V_\theta(\overline{y}). \tag{3.2.4}$$

In this article, we are concerned with the following notion of limit value for optimal control problems with general means.

Definition 3.2.1. *Let V be a function defined on $\mathbb{R}^d$. The optimal control problem $\mathcal{J}$ admits V as the* ***general limit value*** *if: for any sequence of evaluations $(\theta^k)_{k\geq 1}$ satisfying the LTC, $(V_{\theta^k})_k$ converges uniformly to V as k tends to infinity.*

There is ***general uniform convergence*** *of the value functions $\{V_\theta\}$ to V if:*

$$\forall \varepsilon > 0, \exists S > 0, \exists \eta > 0 \text{ s.t. } \forall \theta \in \Delta(\mathbb{R}_+), \text{with } \overline{TV}_S(\theta) \leq \eta,\ \|V_\theta - V\|_\infty \leq \varepsilon.$$

The two notions are indeed equivalent, as will be proven in Lemme 3.3.2.

Below are some basic examples of optimal control problems in which the general limit value exists.

Example 6. *y lies in $\mathcal{C} = \{z \in \mathbb{R}^2 : \|z\| \leq 1\}$ seen as the unit disk on the complex plane, there is no control, and the dynamic is given by $f(y) = i\ y$, where $i^2 = -1$. We have*

$$V_{\theta^k}(y_0) \xrightarrow[k\to\infty]{} \varphi(y_0) =_{def} \frac{1}{2\pi}\int_0^{2\pi} g(|y_0|e^{rit})\mathrm{d}t,\ \forall y_0 \in \mathcal{C}$$

for any sequence of evaluations $(\theta^k)_k$ satisfying the LTC.

To deduce the result, we employ two arguments. First, it is clear that the t-horizon value function V_t converges uniformly to the function φ as t tends to infinity. Second, for an uncontrolled problem, the uniform convergence of V_t to φ implies that φ is the general limit value. We leave the proof (cf. Proposition 3.6.1) and related discussions to Section 3.6.

Example 7. *y lies in the complex plane again, with $f(y, u) = i\ y\ u$, where $u \in U$ is a given bounded subset of $\mathbb{R}$, and g is any continuous function in y (which thus does not depend on u).*

Example 8. *$f(y, u) = -y + u$, where $u \in U$ a given bounded subset of $\mathbb{R}^d$, and g is any continuous function in y (which thus does not depend on u).*

We will show later (using Corollary 3.4.5) that the general limit value exists in both Example 7 and Example 8.

3.3 On the long-term condition (LTC)

In this section, we discuss the LTC. First, we give the following remarks.

Remark 3.3.1. *(a). By definition, one has*

$$\forall s \geq 0, \forall t \geq 0, \forall \theta \in \Delta(\mathbb{R}_+),\ TV_{s+t}(\theta) \leq TV_s(\theta) + TV_t(\theta).$$

This implies that $(\theta^k)_{k\geq 1}$ satisfies the LTC if and only if $\overline{TV}_1(\theta^k) \xrightarrow[k\to\infty]{} 0$.
(b). If one takes $Q = \mathbb{R}_+$ in definition of $TV_s(\theta^k)$ for each $s \geq 0$ and each $k \geq 1$, we deduce that if $(\theta^k)_{k\geq 1}$ satisfies the LTC, then $\theta^k([0, s]) \xrightarrow[k\to\infty]{} 0$ for any $s \geq 0$.

Lemma 3.3.2. *Let V be a function defined on $\mathbb{R}^d$. The optimal control problem $\mathcal{J}$ admits V as the general limit value if and only if there is general uniform convergence of the value functions $\{V_\theta\}$ to V.*

Proof. One direction is clear. Assume that V is the general limit value of $\mathcal{J}$, and we will prove general uniform convergence of $\{V_\theta\}$ to V. Suppose by contradiction that this is not true, *i.e.*, $\exists \varepsilon_0 > 0, \forall S > 0, \forall \eta^k > 0, \exists \theta^k \in \Delta(\mathbb{R}_+)$ with

$$\overline{TV}_S(\theta^k) \leq \eta^k \text{ and } \|V_{\theta^k} - V\|_\infty > \varepsilon_0,\ \forall k \geq 1.$$

Let $\varepsilon_0 > 0$ be fixed as above. Take $(\eta^k)_k$ a vanishing positive sequence, then there is a sequence of evaluations $(\theta^k)_k$ with $\overline{TV}_{S_0}(\theta^k) \leq \eta^k \xrightarrow[k\to\infty]{} 0$, and $\liminf_k \|V_{\theta^k} - V\|_\infty \geq \varepsilon_0$. According to Remark 3.3.1 (a), such $(\theta^k)_k$ satisfies the LTC, while $(V_{\theta^k})_k$ does not converges uniformly to V^*. This leads to a contradiction. □

Lemma 3.3.3. *Let θ be an evaluation absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}_+$, and f_θ be its density. For all $s \geq 0$, we write:*

$$I_s(\theta) = \int_{\mathbb{R}_+} |f_\theta(t+s) - f_\theta(t)| \mathrm{d}t.$$

Then we obtain that:

$$TV_s(\theta) \leq I_s(\theta) \leq 2TV_s(\theta).$$

Proof. First consider any Q in $\mathcal{B}(\mathbb{R}_+)$. $\theta(Q) - \theta(Q+s) = \int_{\mathbb{R}_+} (f_\theta(t) - f_\theta(t+s)) \mathbb{1}_{t \in Q} \mathrm{d}t$. So $TV_s(\theta) \leq I_s(\theta)$. Define now $Q = \{t \in \mathbb{R}_+ | f_\theta(t+s) \leq f_\theta(t)\}$, and $Q^c = \mathbb{R}_+ \backslash Q$. We have $I_s(\theta) = \big(\theta(Q) - \theta(Q+s)\big) + \big(\theta(Q^c+s) - \theta(Q^c)\big) \leq 2TV_s(\theta)$. □

Remark 3.3.4. *Let $(\theta^k)_{k\geq 1}$ be a sequence of evaluations with densities $(f_{\theta^k})_{k\geq 1}$, then:*
(a). following Lemma 3.3.3, (θ^k) satisfies the LTC if and only if $\sup_{0\leq s\leq 1} I_s(\theta^k) \xrightarrow[k\to\infty]{} 0$.
If moreover, $\forall k \geq 1$, $t \mapsto f_{\theta^k}(t)$ is non increasing on $\mathbb{R}_+$, then $(\theta^k)_{k\geq 1}$ satisfies the LTC if and only if $\forall s \geq 0$, $\theta^k([0,s]) = \int_{\mathbb{R}_+} f_\theta(t)\mathrm{d}t - \int_{\mathbb{R}_+} f_\theta(t+s)\mathrm{d}t \xrightarrow[k\to\infty]{} 0$.
(b). if (θ^k) satisfies the LTC, then $\int_{\mathbb{R}_+} t f_{\theta^k}(t)\mathrm{d}t \xrightarrow[k\to\infty]{} \infty$. Indeed, Chebychev's inequality gives that $\int_{\mathbb{R}_+} t f_{\theta^k}(t)\mathrm{d}t \geq M\Big(1 - \theta^k([0,M])\Big)$ for all $M > 0$.

Here we discuss several families of evaluations where the LTC condition is satisfied.

Example 9. *(Uniform distributions) Assume that for each k, θ^k is the uniform law over the interval $[a_k, b_k]$, with $0 \leq a_k \leq b_k$. For each k,*

- $\underline{s \geq b_k - a_k}$: $I_s(\theta^k) = \begin{cases} \frac{2}{b_k - a_k} & \text{if} \quad 0 < s < a_k \\ \frac{1+(b_k - s)}{b_k - a_k} & \text{if} \quad a_k < s < b_k \\ \frac{1}{b_k - a_k} & \text{if} \quad b_k < s \end{cases}$,
- $\underline{s < b_k - a_k}$: $I_s(\theta^k) = \begin{cases} \frac{2s}{b_k - a_k} & \text{if} \quad 0 < s < a_k \\ \frac{s + a_k}{b_k - a_k} & \text{if} \quad a_k < s < b_k \end{cases}$.

One can check easily that $(\theta^k)_k$ satisfies the LTC if and only if $b_k - a_k \xrightarrow[k\to\infty]{} \infty$. Indeed, by Remark 3.3.4 (a), it is sufficient to look at $I_s(\theta^k)$ for $s \in [0,1]$.

Example 10. *(Abel average) Assume that for each k, θ^k has density $s \mapsto f_{\theta^k}(s) = \lambda_k e^{-\lambda_k s} \mathbb{1}_{\mathbb{R}_+}(s)$, with $\lambda_k > 0$. Since $\forall k \geq 1$, $s \mapsto f_{\theta^k}(s)$ is non increasing, Remark 3.3.4 (a) implies that $(\theta^k)_k$ satisfies the LTC if and only if: $\forall T > 0$, $\theta^k([0,T]) = \int_{s=0}^T \lambda_k e^{-\lambda_k s}\mathrm{d}s = 1 - e^{-T\lambda_k} \xrightarrow[k\to\infty]{} 0$, which is again equivalent to $\lambda_k \xrightarrow[k\to\infty]{} 0$.*

Example 11. *(Folded normal distributions) Assume that for each k, θ^k is the distribution of a random variable $|X^k|$, where X^k follows a normal law $\mathcal{N}(m_k, \sigma_k^2)$. The density of θ^k is given by:*

$$\forall t \geq 0,\ f_{\theta^k}(t) = \frac{1}{\sigma_k\sqrt{2\pi}} \left[\exp\left(-\frac{1}{2}\left(\frac{t - m_k}{\sigma_k}\right)^2\right) + \exp\left(-\frac{1}{2}\left(\frac{t + m_k}{\sigma_k}\right)^2\right)\right].$$

Claim 3.3.5. *$(\theta^k)_k$ satisfies the LTC if and only if $\sigma_k \xrightarrow[k\to\infty]{} \infty$.*

Our argument relies on the following lemma, whose proof is put in the **Appendix**. Without loss of generality, we may assume that m_k is non-negative for each k.

Lemma 3.3.6. *Let θ be the distribution of X where $|X|$ follows the normal law $\mathbb{N}(m,\sigma)$ with $m,\sigma>0$. There exists some $t^*\in[0,m)$ such that $f'_\theta(t)>0$ for any $t\in(0,t^*)$ and $f'_\theta(t)<0$ for any $t\in(t^*,+\infty)$. Moreover, such t^* satisfies that: $(t^*)^2\geq m^2-\sigma^2$.*

Proof of Claim 3.3.5 We apply Lemma 3.3.6 to each evaluation θ_k to obtain some $t_k^*\in[0,m_k)$ such that: $f_{\theta^k}(\cdot)$ is increasing on $[0,t_k^*)$ and decreasing on $[t_k^*,\infty)$. This enables us to write:

$$\forall s\leq t_k^*,\quad I_s(\theta^k)=\int_{t_k^*-s}^{t_k^*}f_{\theta^k}(t)\mathrm{d}t+\int_{t_k^*-s}^{t_k^*}|f_{\theta^k}(t+s)-f_{\theta^k}(t)|\mathrm{d}t+\int_{t_k^*}^{t_k^*+s}f_{\theta^k}(t)\mathrm{d}t.$$

We deduce then $sf_{\theta^k}(t_k^*-s)\leq I_{\theta^k}(s)\leq 4sf_{\theta^k}(t_*^k)$ for $s\leq t_k^*$. Assume below $\hat{t}^*=_{def}\liminf_{k\to\infty}t_k^*>0$, and the analysis is analogue for $\hat{t}^*=0$, which we omit here.
(*). Suppose that $\sigma_k\to\infty$, then

$$f_{\theta^k}(t_k^*)=\frac{1}{\sigma_k\sqrt{2\pi}}\left[\exp\left(-\frac{1}{2}\left(\frac{t_k^*-m_k}{\sigma_k}\right)^2\right)+\exp\left(-\frac{1}{2}\left(\frac{t_k^*+m_k}{\sigma_k}\right)^2\right)\right]\leq\frac{2}{\sigma_k\sqrt{2\pi}}\xrightarrow[k\to\infty]{}0.$$

This implies that for $S=\hat{t}^*\wedge 1$, $\sup_{0\leq s\leq S}I_s(\theta^k)\xrightarrow[k\to\infty]{}0$.
(**). Conversely, suppose that $(\theta^k)_k$ satisfies the LTC. Then for any $s<\hat{t}^*$, $I_s(\theta^k)$ thus $f_{\theta^k}(t_k^*-s)$ vanishes as k tends to infinity. This implies that either $\sigma_k\to\infty$ or $(\sigma_k)_k$ is bounded and $\left(m_k-(t_k^*-s)\right)_k\to\infty$. Lemma 3.3.6 shows that the specified point t_k^* for the evaluation θ_k satisfies $(t_k^*)^2\geq m_k^2-\sigma_k^2$, thus $m_k-(t_k^*+s)\leq m_k-t_k^*\leq\frac{\sigma_k^2}{m_k+t_k^*}\leq\frac{\sigma_k^2}{m_k}$. If $(\sigma_k)_k$ is bounded, $(m_k-t_k^*)_k$ thus $(m_k)_k$ should tend to infinity, but this leads to a contradiction with $m_k-t_k^*\leq\frac{\sigma_k^2}{m_k}$. □

Now we link the LTC condition to the discrete time framework. In a discrete time dynamic optimization problem, a general evaluation on the payoff stream is a probability distribution over $\mathbb{N}^*=\mathbb{N}/\{0\}$ the set of postive integers. For any $\xi=(\xi_1,...,\xi_t,...)$ in $\Delta(\mathbb{N}^*)$, its "total variation" $TV(\xi)=\sum_{m=1}^\infty|\xi_{m+1}-\xi_m|$ is the stage by stage absolute difference between the measure ξ and its one-stage "shift" measure $\xi'=(\xi_2,...,\xi_{t+1},...)$. (cf. Sorin [59] or Renault [46]).

When the sequence of evaluations in continuous time admits step functions as densities, this link to discrete time framework is much clearer as seen by the following

Proposition 3.3.7. *Let $(\theta^k)_k$ be a sequence of absolutely continuous evaluations in $\Delta(\mathbb{R}_+)$, and their densities are given as: $\forall k\geq 1, f_{\theta^k}=\sum_{m=1}^\infty\xi_m^k\mathbb{1}_{[m-1,m)}$, where $\xi^k=(\xi_1^k,...,\xi_m^k,...,)\in\Delta(\mathbb{N}^*)$. Then $(\theta^k)_k$ satisfies the LTC if and only if $\sum_{m=1}^\infty|\xi_{m+1}^k-\xi_m^k|\xrightarrow[k\to\infty]{}0$.*

Proof: Fix $s\in[0,1]$. We shall write for each k,

$$I_s(\theta^k)=\sum_{m=1}^\infty\int_{[m-1,m)}\Big|f_{\theta^k}(t+s)-f_{\theta^k}(t)\Big|\mathrm{d}t.$$

For each $m=1,2,...$, we have

$$\begin{aligned}&\int_{[m-1,m)}\Big|f_{\theta^k}(t+s)-f_{\theta^k}(t)\Big|\mathrm{d}t\\=&\int_{[m-1,m-s)}\Big|f_{\theta^k}(t+s)-f_{\theta^k}(t)\Big|\mathrm{d}t+\int_{[m-s,m)}\Big(f_{\theta^k}(t+s)-f_{\theta^k}(t)\Big)\mathrm{d}t\\=&s\big|\xi_{m+1}^k-\xi_m^k\big|.\end{aligned}$$

As a consequence,

$$I_s(\theta^k) = s\sum_{m=1}^{\infty}|\xi_{m+1}^k - \xi_m^k| \leq \sum_{m=1}^{\infty}|\xi_{m+1}^k - \xi_m^k|,\ \forall s \in [0,1].$$

In view of Remark 3.3.4, $(\theta^k)_k$ satisfies the LTC if and only if $\sum_{m=1}^{\infty}|\xi_{m+1}^k - \xi_m^k| \xrightarrow[k\to\infty]{} 0$. □

We end this section by a preliminary lemma, which will be useful for later results.

Lemma 3.3.8. *Fix any $\theta \in \Delta(\mathbb{R}_+)$ and any $t \in \mathbb{R}_+$, we have*

$$\left|\int_{[0,+\infty)} h(s)\mathrm{d}\theta(s) - \int_{[0,+\infty)} h(s-t)\mathrm{d}\theta(s)\right| \leq TV_t(\theta)$$

and

$$\left|\int_{[0,+\infty)} h(s)\mathrm{d}\theta(s) - \int_{[0,+\infty)} h(s+t)\mathrm{d}\theta(s)\right| \leq 2TV_t(\theta)$$

for any $h(\cdot) \in \mathcal{M}(\mathbb{R}_+,[0,1])$,where $\mathcal{M}(\mathbb{R}_+,[0,1])$ is the set of Borel measurable functions defined from $\mathbb{R}_+$ to $[0,1]$.

Proof: We fix any $\theta \in \Delta(\mathbb{R}_+)$ and $t \in \mathbb{R}_+$. By definition of $\mathcal{T}_s\sharp\theta$, we have that for any $h(\cdot) \in \mathcal{M}(\mathbb{R}_+,[0,1])$:

$$\int_{[0,+\infty)} h(s)\mathrm{d}\theta(s) - \int_{[t,+\infty)} h(s-t)\mathrm{d}\theta(s) = \int_{[0,+\infty)} h(s)\mathrm{d}\theta(s) - \int_{[0,+\infty)} h(s)\mathrm{d}\mathcal{T}_{-t}\sharp\theta(s) \tag{3.3.1}$$

and

$$\int_{[0,+\infty)} h(s)\mathrm{d}\theta(s) - \int_{[0,+\infty)} h(s+t)\mathrm{d}\theta(s) = \int_{[0,+\infty)} h(s)\mathrm{d}\theta(s) - \int_{[0,+\infty)} h(s)\mathrm{d}\mathcal{T}_{t}\sharp\theta(s). \tag{3.3.2}$$

Since $\mathcal{T}_{-t}\sharp\theta$ and $\mathcal{T}_t\sharp\theta$ are both Borel measures on $\mathbb{R}_+$, "$\theta - \mathcal{T}_{-t}\sharp\theta$" and "$\theta - \mathcal{T}_t\sharp\theta$" are both signed measures. Hahn's decomposition theorem [3] implies that:

$$\sup_{h\in\mathcal{M}(\mathbb{R}_+,[0,1])}\left|\int_{[0,+\infty)} h(s)\mathrm{d}\theta(s) - \int_{[0,+\infty)} h(s)\mathrm{d}\mathcal{T}_{-t}\sharp\theta(s)\right| = \sup_{Q\in\mathcal{B}(\mathbb{R}_+)}\Big|\theta(Q) - \mathcal{T}_{-t}\sharp\theta(Q)\Big|.$$

and

$$\sup_{h\in\mathcal{M}(\mathbb{R}_+,[0,1])}\left|\int_{[0,+\infty)} h(s)\mathrm{d}\theta(s) - \int_{[0,+\infty)} h(s)\mathrm{d}\mathcal{T}_{t}\sharp\theta(s)\right| = \sup_{Q\in\mathcal{B}(\mathbb{R}_+)}\Big|\theta(Q) - \mathcal{T}_{t}\sharp\theta(Q)\Big|.$$

Combining with (3.3.1)-(3.3.2), we obtain:

$$\left|\int_{[0,+\infty)} h(s)\mathrm{d}\theta(s) - \int_{[t,+\infty)} h(s-t)\mathrm{d}\theta(s)\right| \leq \sup_{Q\in\mathcal{B}(\mathbb{R}_+)}\Big|\theta(Q) - \theta(Q+t)\Big| = TV_t(\theta)$$

and

$$\begin{aligned}\left|\int_{[0,+\infty)} h(s)\mathrm{d}\theta(s) - \int_{[0,+\infty)} h(s+t)\mathrm{d}\theta(s)\right| &\leq \sup_{Q\in\mathcal{B}(\mathbb{R}_+)}\Big|\theta(Q) - \theta(Q-t)\Big| \\ &\leq \theta\big([0,t)\big) + TV_t(\theta) \leq 2TV_t(\theta).\end{aligned}$$

The proof of the lemma is complete. □

3. The first author acknowledges Eilon Solan for the discussion on using Hahn's decomposition theorem.

3.4 Main Result

As will be shown in our main result, the function $V^*(y_0)$ defined in (3.1.3) characterizes the general limit value of the optimal control problem in case of convergence. We first rewrite it as

$$V^*(y_0) = \sup_{\mu\in\Delta(\mathbb{R}_+)} \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\mu}(y_0) = \sup_{\mu\in\Delta(\mathbb{R}_+)} \inf_{\overline{y}\in R(y_0)} V_\mu(\overline{y}).$$

We give the following interpretation: consider the auxiliary optimal control problem (game) where an adversary of the controller chooses an evaluation μ, and then knowing μ, the controller chooses some $\bar{y}$ in the reachable set $R(y_0)$ as the *initial* state. $V^*(y_0)$ is the maxmin of this problem.
Recall that a metric space X is *totally bounded* if for each $\varepsilon > 0$, X can be covered by finitely many balls of radius ε.

Theorem 3.4.1. *Let $(\theta^k)_{k\geq 1}$ be a sequence of evaluations satisfying the LTC. Then,*
(i). $V^* = \sup_{k\in\mathbb{N}} \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\theta^k}$.
(ii). The sequence $(V_{\theta^k})_k$ converges uniformly if and only if the space $(\{V_{\theta^k}\}, \|\cdot\|_\infty)$ is totally bounded, and the limit is equal to V^* in case of convergence.

Remark 3.4.2. *Let $(\theta^k)_k$ be a sequence of evaluations which contains a subsequence $(\theta^{\varphi_k})_k$ satisfying the LTC. Then Part (i) of Theorem 3.4.1 still holds true for $(\theta^k)_k$.*

A more precise convergence result is obtained if we suppose that there exists a compact set $Y \subseteq \mathbb{R}^d$ which is *invariant* for the dynamic (3.1.1), i.e. such that $\mathbf{y}(t, \mathbf{u}, y_0) \in Y$ for all $\mathbf{u} \in \mathcal{U}$, $t \geq 0$ and y_0 in Y.

Corollary 3.4.3. *Suppose that there is a compact set $Y \subseteq \mathbb{R}^d$ which is invariant for the dynamic (3.1.1), and that the family $\{V_\theta : \theta \in \Delta(\mathbb{R}_+)\}$ is uniformly equicontinuous on Y. Then there is general uniform convergence of the value functions $\{V_\theta\}$ to V^* on Y.*

Proof: By assumption, the family of value functions $\{V_\theta : \theta \in \Delta(\mathbb{R}_+)\}$ is both uniformly bounded and uniformly equicontinuous on the compact invariant set Y, so we can use Ascoli's theorem to deduce the total boundedness of the space $(\{V_\theta\}, \|\cdot\|_\infty)$. Theorem 3.4.1 implies that: for any $(\theta^k)_k$ satisfying the LTC, the corresponding sequence of value functions (V_{θ^k}) converges uniformly to V^* as k tends to infinity. Thus $\mathcal{J}$ has a general limit value given as V^*, and according to Lemma 3.3.2, there is uniform convergence of value functions $\{V_\theta\}$ to V^* on Y. □

We shall give an existence result of the general limit value under sufficient conditions expressed directly in terms of properties of the control dynamic (3.1.1) and of the running cost g.

Let us introduce the following *nonexpansive* condition (cf. Quincampoix and Renault [42]). The control dynamic (3.1.1) is non expansive if

$$\forall y_1, y_2 \in \mathbb{R}^d,\ \sup_{a\in U} \inf_{b\in U} \Big\langle y_1 - y_2, f(y_1, a) - f(y_2, b)\Big\rangle \leq 0.$$

Definition 3.4.4. *The optimal control problem $\mathcal{J} = \langle U, g, f\rangle$ is called* ***compact non expansive*** *if it satisfies the following three conditions:*
$(A.1)$ *there is a compact set $Y \subseteq \mathbb{R}^d$ which is the invariant for the dynamic (3.1.1);*
$(A.2)$ *the running cost function g does not depend on $u \in U$, and is continuous in $y \in \mathbb{R}^d$;*
$(A.3)$ *the control dynamic (3.1.1) is nonexpansive on Y.*

Corollary 3.4.5. *Assume (3.2.1) for the optimal control problem $\mathcal{J} = \langle U, g, f\rangle$. Suppose that that $\mathcal{J}$ is compact nonexpansive, then the general limit value exists in $\mathcal{J}$ and is given by V^* on Y.*

Proof: Under $(A.1)$ and $(A.3)$, Proposition 3.7 in Quincampoix and Renault [42] implies that:

$$\forall (y_1, y_2) \in Y^2, \forall \mathbf{u} \in \mathcal{U}, \exists \mathbf{v} \in \mathcal{U}, s.t.\ \forall t \geq 0,\ \|\mathbf{y}(t, \mathbf{u}, y_1) - \mathbf{y}(t, \mathbf{u}, y_2)\| \leq \|y_1 - y_2\|. \tag{3.4.1}$$

We claim that the family $(V_\theta)_{\theta\in\Delta(\mathbb{R}_+)}$ is uniformly equicontinuous on Y, thus Corollary 3.4.3 and Lemma 3.3.2 apply. Fix any $(y_1, y_2) \in Y^2$, $\theta \in \Delta(\mathbb{R}_+)$, and $\varepsilon > 0$. Let $\mathbf{u}$ be ε-optimal for $V_\theta(y_1)$:

$$V_\theta(y_1) \geq \int_{[0,+\infty)} g\Big(\mathbf{y}(s, \mathbf{u}, y_1)\Big) \mathrm{d}\theta(s) - \varepsilon.$$

According to the nonexpansive property, there exists $\mathbf{v}(\cdot)$ in $\mathcal{U}$ as in (3.4.1) such that

$$\|\mathbf{y}(s, \mathbf{u}, y_1) - \mathbf{y}(s, \mathbf{v}, y_2)\| \leq \|y_1 - y_2\|,\ \forall s \geq 0. \tag{3.4.2}$$

By definition, $V_\theta(y_2) \leq \int_{[0,+\infty)} g(\mathbf{y}(s, \mathbf{v}, y_2)) \mathrm{d}\theta(s)$, hence

$$V_\theta(y_2) - V_\theta(y_1) \leq \int_{[0,+\infty)} \Big[g(\mathbf{y}(s, \mathbf{v}, y_2)) - g(\mathbf{y}(s, \mathbf{u}, y_1))\Big] \mathrm{d}\theta(s) + \varepsilon.$$

Denoting ω_g the modulus of continuity of g, we obtain in view of (3.4.2):

$$V_\theta(y_2) - V_\theta(y_1) \quad \leq \quad \int_{[0,+\infty)} \Big[g(\mathbf{y}(s, \mathbf{v}, y_2)) - g(\mathbf{y}(s, \mathbf{u}, y_1))\Big] \mathrm{d}\theta(s) + \varepsilon \leq \omega_g(\|y_1 - y_2\|) + \varepsilon.$$

Interchanging y_1 and y_2 and taking into account of $\varepsilon > 0$ being arbitrary, we deduce that $(V_\theta)_{\theta\in\Delta(\mathbb{R}_+)}$ is uniformly equicontinuous on the invariant set Y. This finishes the proof. □

Remark 3.4.6. *Both Example 7 and Example 8 satisfy conditions of Corollary 3.4.5, so there is general uniform convergence of the value functions $\{V_\theta\}$ (the existence of the general limit value).*

Remark 3.4.7. *Our result generalizes Proposition 3.3 in Quincampoix and Renault [42] which proved the uniform convergence of the t-horizon values in compact nonexpansive optimal control problems.*

We end this section by presenting two (counter)examples, showing that the results in Theorem 3.4.1 do not hold if some of their conditions is not satisfied.

The first example is an uncontrolled dynamic. We show that if $(\theta^k)_k$ contains no subsequence satisfying the LTC, then the result in Part (i) of Theorem 3.4.1 does not hold, i.e. $\sup_{k\in\mathbb{N}^*} \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\theta^k}(y_0) < \sup_{\theta\in\Delta(\mathbb{R}_+)} \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\theta}(y_0)$ for some y_0 (cf. Remark 3.4.2).

Counter-example 1. *Consider the uncontrolled dynamic on $\mathbb{R}$: $\mathbf{y}(0) = y_0$ and $\mathbf{y}'(t) = -(\mathbf{y}(t) - 1), \forall t \geq 0$. The trajectory is then $\mathbf{y}(t) = 1 + (y_0 - 1)e^{-t}$. The running cost function $g : \mathbb{R} \to [0, 1]$ is given by:*

$$g(y) = \begin{cases} 0 & \textit{if} \quad y < 0 \\ y & \textit{if} \quad 0 \leq y \leq 1 \\ 1 & \textit{if} \quad y > 1 \end{cases}$$

We have that $V^*(y_0) = \sup_{\theta\in\Delta(\mathbb{R}_+)} \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\theta}(y_0) = 1,\ \forall y_0 \in \mathbb{R}$. *Indeed, let* y_0 *be given and fix any* $\varepsilon > 0$, *there is some* $T_\varepsilon > 0$ *such that* $|\mathbf{y}(T) - 1| \le \varepsilon$ *for all* $T \ge T_\varepsilon$. *Take an evaluation* θ *in* $\Delta(\mathbb{R}_+)$ *with* $\theta([0, T_\varepsilon]) = 0$. *This enables us to deduce that: for all* $t \ge 0$,

$$V_{\mathcal{T}_t\sharp\theta}(y_0) = \int_{[T_\varepsilon,+\infty)} g(\mathbf{y}(s+t))\mathrm{d}\theta(s) \ge \int_{[T_\varepsilon,+\infty)} g(\mathbf{y}(T_\varepsilon))\mathrm{d}\theta(s) \ge (1-\mathbf{y}(T_\varepsilon))\theta([T_\varepsilon, +\infty]) \ge 1-\varepsilon.$$

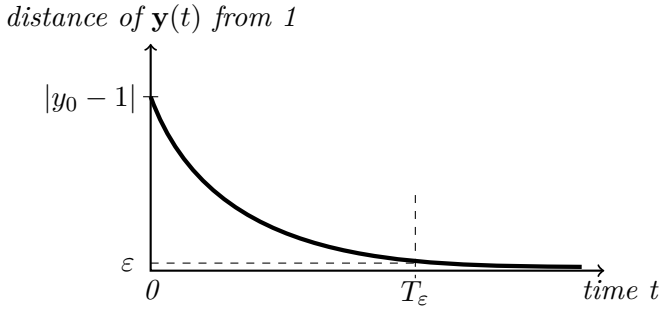


Figure 1*: The solution* $\boldsymbol{y}(t) = 1 + (y_0 - 1)e^{-t}$ *to the dynamic is represented by the thick curve. For given* $\varepsilon > 0$, $T_\varepsilon > 0$ *is chosen such that* $|\boldsymbol{y}(T_\varepsilon) - 1| = \varepsilon$.

Consider now any sequence of evaluations $(\theta^k)_k$ *which does not contain any subsequence satisfying the LTC. Under the assumption that the density* f_{θ^k} *for each evaluation* θ^k *is non increasing, we show that Part (i) of Theorem 3.4.1 is not valid:* $V^* \ne \sup_{k\in\mathbb{N}} \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\theta^k}$.

Indeed, let us take any $y_0 < 1$ *and suppose that* $\sup_{k\in\mathbb{N}} \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\theta^k}(y_0) = V^*(y_0)$, *which is equal to* 1 *as was proved. Let* $\varphi(k)$ *be a subsequence such that* $\lim_{k\to\infty} \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\theta^{\varphi(k)}}(y_0) = 1$. $(\theta^{\varphi(k)})_k$ *does not satisfy the LTC by assumption, so Remark 3.3.4 (a) implies that there exists some* $T > 0$ *with* $\theta^{\varphi(k)}([0, T]) \nrightarrow 0$. *Let* φ_m *be the subsequence of* φ *and* $\eta > 0$ *such that* $\theta^{\varphi_m(k)}([0, T]) \xrightarrow[k\to\infty]{} \eta$. *We obtain for any* $k \ge 1$,

$$\begin{aligned}\inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\theta^{\varphi_m(k)}}(y_0) \le V_{\theta^{\varphi_m(k)}}(y_0) &= \int_{[0,T]} g\left(\mathbf{y}(t)\right) \mathrm{d}\theta^{\varphi_m(k)}(t) + \int_{[T,+\infty]} g\left(\mathbf{y}(t)\right) \mathrm{d}\theta^{\varphi_m(k)}(t) \\ &\le \mathbf{y}(T)\theta^{\varphi_k(m)}([0, T]) + \theta^{\varphi_k(m)}([T, +\infty]).\end{aligned}$$

This implies that for such fixed $y_0 < 1$,

$$\lim_k \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\theta^{\varphi_m(k)}}(y_0) \le \mathbf{y}(T)\eta + (1 - \eta) < 1.$$

This contradicts the assumption that $\sup_{k\in\mathbb{N}} \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\theta^k}(y_0) = 1$, *and our claim is proved.*

In the second example, we study the convergence of the value functions of a control problem along two different sequences of evaluations satisfying the LTC. Along the first sequence, the value functions converge uniformly to V^*; while along the second, the value functions point-wisely converge, but not uniformly (thus the family of value functions is not totally bounded for the uniform norm), to a limit function which is different from V^*.

Counter-example 2. *Consider the control problem on the state space* $\mathbb{R} = (-\infty, +\infty)$, *where the control set is* $U = \{+1, -1\}$; *the dynamic is*[4]:

$$f(y, u) = u \text{ for all } (y, u) \in \mathbb{R}_+ \times U \text{ and } f(y, u) = -1 \text{ for all } (y, u) \in \mathbb{R}_-^* \times U,$$

4. Notice that the dynamic is discontinuous at $y = 0$ when $u = +1$. To get the desired asymptotic result under the Liptchitz regularity, one can slightly modify dynamic to set $f(y, +1) = y$ for $y \in [0, 1]$ and others unchanged.

where $\mathbb{R}_-^* = \mathbb{R}_-/\{0\}$*; and the running cost function is:*

$$g(y,u) = \begin{cases} +1 & \text{if} \quad u=+1,\ y \geq 0 \\ 0 & \text{if} \quad u=-1,\ y \geq 0 \\ +K & \text{if} \quad y<0 \end{cases}$$

Suppose that $K > 1$ *large enough, so the cost on* $\mathbb{R}_-$ *is positive and high. Whenever the state reaches* $y = 0$*, it is optimal to choose control* $u = +1$ *and this drives the state back to* $\mathbb{R}_+$*; on* $\mathbb{R}_-^*$*, the dynamic is* $f = -1$*, independent of control and state.* $V_\theta(y_0) = K$ *for all* y_0 *in* $\mathbb{R}_-^*$ *and* θ *in* $\Delta(\mathbb{R}_+)$*, so the reduced state space is* $\mathbb{R}_+$*, and we consider value functions defined on it.*

$V^*(y_0) = \sup_\theta \inf_{t\geq 0} V_{\mathcal{T}_t\sharp\theta}(y_0) = 0$ *for any* $y_0 \geq 0$*. Fix any* $y_0 \geq 0$*. For any* $\theta \in \Delta(\mathbb{R}_+)$ *and* $\varepsilon > 0$*, let* $t^\varepsilon \geq 0$ *such that* $\theta([0,t^\varepsilon]) \geq 1-\varepsilon$*. Define now the control* $\mathbf{u}^\varepsilon$ *to be:* $\mathbf{u}^\varepsilon(t) = +1$*, if* $t \in [0,t^\varepsilon]$ *and* $\mathbf{u}^\varepsilon(t) = -1$ *if* $t \in (t^\varepsilon, \infty)$*, which gives:* $\gamma_{\mathcal{T}_{t^\varepsilon}\sharp\theta}(y_0, \mathbf{u}^\varepsilon) \leq \varepsilon K$*.*

Consider $(\theta^k)_k$ *the sequence of evaluations with density* $f_{\theta^k}(s) = \frac{1}{k}\mathbb{1}_{[k,2k]}(s)$ *for each* k*, and* $(\bar{\theta}^k)_k$ *the sequence of* k*-horizon evaluations with density* $f_{\bar{\theta}^k}(s) = \frac{1}{k}\mathbb{1}_{[0,k]}(s)$ *for each* k*. We show that:*

$(\{V_{\theta^k}\}, \|\cdot\|_\infty)$ *is totally bounded and* (V_{θ^k}) *converges uniformly to* V^**; while* $(\{V_{\bar{\theta}^k}\}, \|\cdot\|_\infty)$ *is not totally bounded and* $(V_{\bar{\theta}^k})$ *does not converge to* V^**.*

Let $y_0 \geq 0$*, we have that:*

1. $V_{\theta^k}(y_0) = 0$*, for all* $k \geq 1$*. Indeed, one optimal control for* $V_{\theta^k}(y_0)$ *can be taken as:* $\mathbf{u}^*(t) = +1,\ t \in [0,k]$ *and* $\mathbf{u}^*(t) = -1,\ t \in (k, 2k]$*;*
2. $V_{\bar{\theta}^k}(y_0) = 0$ *if* $k \leq y_0$ *and* $V_{\bar{\theta}^k}(y_0) = \frac{1}{2} - \frac{y_0}{2k}$ *if* $k > y_0$*. Indeed, for* $k \leq y_0$*, one optimal control for* $V_{\bar{\theta}^k}(y_0)$ *can be taken as:* $\mathbf{u}^*(t) = -1,\ t \in [0,k]$*; for* $k > y_0$*, one optimal control for* $V_{\bar{\theta}^k}(y_0)$ *can be taken as:* $\mathbf{u}^*(t) = +1,\ t \in [0, \frac{k-y_0}{2}]$ *and* $\mathbf{u}^*(t) = -1,\ t \in (\frac{k-y_0}{2}, k]$*, so* $\gamma_{\bar{\theta}^k}(y_0, \mathbf{u}^*) = \frac{(k-y_0)/2}{k} = \frac{1}{2} - \frac{y_0}{2k}$*.*

See the following two pictures for illustration.

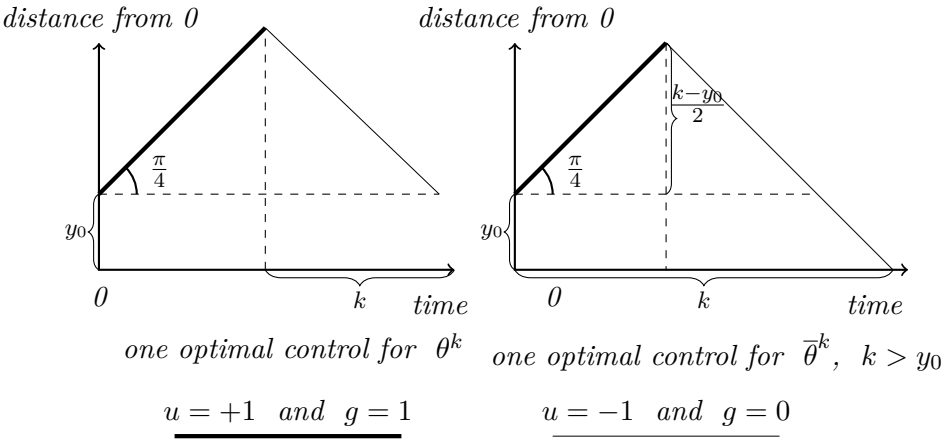


Figure 2*: The left figure describes the dynamic of one optimal control for the evaluation* θ^k*, which is* $\mathbf{u}^* = +1$ *on* $[0,k]$ *and* $\mathbf{u}^* = -1$ *on* $(k, 2k]$*; the right figure describes the dynamic of one optimal control for the evaluation* $\bar{\theta}^k$ *with* $k > y_0$*, which is* $\mathbf{u}^* = +1$ *on* $[0, \frac{k-y_0}{2}]$ *and* $\mathbf{u}^* = -1$ *on* $(\frac{k-y_0}{2}, k]$*. Here, the vertical axis represents the distance of* $\mathbf{y}(t)$

from zero and the thick trajectory (resp. thin trajectory) corresponds to state on which $u = +1$ and $g = 1$ (resp. $u = -1$ and $g = 0$).

We deduce that $(V_{\theta^k}(y_0))_k$ converges uniformly to $V^*(y_0) = 0$ on $\mathbb{R}_+$; and that $V_{\overline{\theta}^k}(y_0) \xrightarrow[k\to\infty]{} \frac{1}{2}$, while the convergence is not uniform in $y_0 \in \mathbb{R}_+$: indeed, for all $k \geq 1$, $V_{\overline{\theta}^k}(k) = 0$.

3.5 Proof of main result: Theorem 3.4.1

Consider in this section a sequence of evaluations $(\theta^k)_k$ that satisfies the LTC. As the proof is long, we divide it into two main parts:

- in Subsection 3.5.1, we present the first preliminary result, Proposition 3.5.1. It is used in two ways: first, we obtain an immediate consequence for later use, which bounds $\liminf_k V_{\theta^k}$ from below in terms of the auxiliary value functions $\{V_{\mathcal{T}_t\sharp\theta^k} : k \in \mathbb{N}^*, t \in \mathbb{R}_+\}$; second, we deduce from it in Corollary 3.5.2 the proof for Part (i) of Theorem 3.4.1.
- In Subsection 3.5.2, we prove Part (ii) of Theorem 3.4.1. Lemma 3.5.3 gives an upper bound of $\limsup_k V_{\theta^k}$ in terms of the auxiliary value functions $\{V_{\mathcal{T}_t\sharp\theta^k} : k \in \mathbb{N}^*, t \in \mathbb{R}_+\}$, which is used, together with the result from Proposition 3.5.1, to prove the convergence of (V_{θ^k}).

3.5.1 A first preliminary result and proof for Part (i)

Proposition 3.5.1. *For any μ in $\Delta(\mathbb{R}_+)$, and any initial state y_0 in $\mathbb{R}^d$,*

$$\inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\mu}(y_0) \leq \liminf_k V_{\theta^k}(y_0).$$

In particular, we have for all y_0 in $\mathbb{R}^d$,

$$\sup_{k\in\mathbb{N}^*} \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\theta^k}(y_0) \leq \liminf_k V_{\theta^k}(y_0).$$

Proof: Fixing y_0 and μ, we set $\beta =_{def} \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\mu}(y_0)$. For any $\varepsilon > 0$ fixed, there exists some $T_0 > 0$ such that $\mu([T_0, +\infty)) < \varepsilon$. Take any control $\mathbf{u}$ in $\mathcal{U}$. By definition of β, we have that

$$\forall T \geq 0, \int_{[0,+\infty)} g\Big(\mathbf{y}(t+T, \mathbf{u}, y_0), \mathbf{u}(t+T)\Big)\mathrm{d}\mu(t) \geq \beta,$$

thus

$$\forall T \geq 0, \int_{[0,T_0]} g\Big(\mathbf{y}(t+T, \mathbf{u}, y_0), \mathbf{u}(t+T)\Big)\mathrm{d}\mu(t) \geq \beta - \varepsilon. \tag{3.5.1}$$

For each $k \geq 1$, integrating both sides of (3.5.1) over $T \in [0, +\infty)$ w.r.t. the evaluation θ^k, we obtain

$$\int_{[0,+\infty)} \int_{[0,T_0]} g\Big(\mathbf{y}(t+T, \mathbf{u}, y_0), \mathbf{u}(t+T)\Big)\mathrm{d}\mu(t)\mathrm{d}\theta^k(T) \geq \beta - \varepsilon. \tag{3.5.2}$$

Applying Fubini's Theorem to (3.5.2) yields

$$\beta - \varepsilon \leq \int_{[0,T_0]} \left[\int_{[0,+\infty)} g\Big(\mathbf{y}(t+T, \mathbf{u}, y_0), \mathbf{u}(t+T)\Big)\mathrm{d}\theta^k(T)\right] \mathrm{d}\mu(t) = \int_{[0,T_0]} \Big[\gamma_{\mathcal{T}_t\sharp\theta^k}(y_0, \mathbf{u})\Big] \mathrm{d}\mu(t), \tag{3.5.3}$$

where $\gamma_{\mathcal{T}_t\sharp\theta^k}(y_0,\mathbf{u}) = \int_{[0,+\infty)} g(\mathbf{y}(t+T,\mathbf{u},y_0),\mathbf{u}(t+T))\mathrm{d}\theta^k(T)$. According to Lemma 3.3.8, we have $|\gamma_{\theta^k}(y_0,\mathbf{u}) - \gamma_{\mathcal{T}_t\sharp\theta^k}(y_0,\mathbf{u})| \leq 2TV_t(\theta^k)$. This enables us to rewrite (3.5.3) as:

$$\begin{aligned}
\beta - \varepsilon &\leq \int_{[0,T_0]} \Big(\gamma_{\theta^k}(y_0,\mathbf{u}) + 2TV_t(\theta^k)\Big)\,\mathrm{d}\mu(t) \\
&\leq \Big(\gamma_{\theta^k}(y_0,\mathbf{u}) + 2\overline{TV}_{T_0}(\theta^k)\Big)\,\mu([0,T_0]) \\
&\leq \gamma_{\theta^k}(y_0,\mathbf{u}) + 2\overline{TV}_{T_0}(\theta^k).
\end{aligned}$$

The control $\mathbf{u} \in \mathcal{U}$ being taken arbitrarily, we deduce that

$$\beta - \varepsilon \leq V_{\theta^k}(y_0) + 2\overline{TV}_{T_0}(\theta^k).$$

Since $(\theta^k)_k$ satisfies the LTC, $\overline{TV}_{T_0}(\theta^k)$ vanishes as k tends to infinity. The proof is achieved. □

We end the proof for Part (i) of Theorem 3.4.1 by the following corollary of Proposition 3.5.1.

Corollary 3.5.2. *[Proof for Part (i) of Theorem 3.4.1]*

$$V^*(y_0) = \sup_{k\geq 1} \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\theta^k}(y_0),\ \forall y_0 \in \mathbb{R}^d.$$

Proof: Fix $y_0 \in \mathbb{R}^d$, and denote $\varrho = \sup_{k\geq 1}\inf_{t\geq 0} V_{\mathcal{T}_t\sharp\theta^k}(y_0)$.We prove that $\varrho \geq V^*(y_0)$. For each $k \geq 1$, there exists $m(k)$ in $\mathbb{R}_+$ such that $V_{\mathcal{T}_{m(k)}\sharp\theta^k}(y_0) \leq \varrho + 1/k$. Since $\pi^k := \mathcal{T}_{m(k)}\sharp\theta^k$ is also an evaluation on $\mathbb{R}_+$, we have: for any $s \geq 0$,

$$TV_s(\pi^k) = \sup_{Q\in\mathcal{B}(\mathbb{R}_+)} \Big|\theta^k((Q-m(k))\cap\mathbb{R}_+) - \theta^k((Q-m(k)+s)\cap\mathbb{R}_+)\Big| \leq TV_s(\theta^k) + \theta^k([0,s]).$$

We deduce that $(\pi^k)_k$ satisfies the LTC whenever $(\theta^k)_k$ does so. According to Proposition 3.5.1,

$$\forall\mu \in \Delta(\mathbb{R}_+),\ \inf_{t\in\mathbb{R}_+} V_{\mathcal{T}_t\sharp\mu}(y_0) \leq \liminf_k V_{\mathcal{T}_{m(k)}\sharp\theta^k}(y_0) \leq \varrho,$$

thus $V^*(y_0) \leq \varrho$. The proof is complete. □

3.5.2 Proof for Part (ii)

We first give an upper bound on "$\limsup_k V_{\theta^k}$" in terms of the auxiliary value functions $\{V_{\mathcal{T}_t\sharp\theta^k} : k \in \mathbb{N}^*, t \in \mathbb{R}_+\}$.

Lemma 3.5.3. *For all $T_0 \geq 0$ and any y_0 in $\mathbb{R}^d$,*

$$\limsup_k V_{\theta^k}(y_0) = \limsup_k \inf_{t\leq T_0} V_{\mathcal{T}_t\sharp\theta^k}(y_0).$$

In particular, for all $T_0 \geq 0$ and any y_0 in $\mathbb{R}^d$,

$$\limsup_k V_{\theta^k}(y_0) \leq \sup_{k\geq 1}\inf_{t\leq T_0} V_{\mathcal{T}_t\sharp\theta^k}(y_0).$$

Proof: Fix $T_0 \geq 0$ and $y_0 \in \mathbb{R}^d$. For all $k \geq 1$ and $t \leq T_0$, we obtain as a direct consequence of Lemma 3.3.8 that

$$\gamma_{\theta^k}(y_0) \leq \gamma_{\mathcal{T}_t \sharp \theta^k}(y_0) + 2TV_t(\mu),$$

thus

$$V_{\theta^k}(y_0) \leq V_{\mathcal{T}_t \sharp \theta^k}(y_0) + 2TV_t(\theta^k) \leq \inf_{0 \leq t \leq T_0} V_{\mathcal{T}_t \sharp \theta^k}(y_0) + 2\overline{TV}_{T_0}(\theta^k).$$

Since $(\theta^k)_k$ satisfies the LTC, $\overline{TV}_{T_0}(\theta^k)$ vanishes as k tends to infinity. By taking "$\limsup_k$" on both sides of above inequality, the proof of the lemma is achieved. □

We summarize results from Proposition 3.5.1, Corollary 3.5.2 and Lemma 3.5.3 in the following chain form, which implies that the uniform convergence of $(V_{\theta^k})_k$ to V^* is obtained once we prove the uniform convergence of "$\sup_{k \geq 1} \inf_{\bar{y} \in R^{T_0}(y_0)} V_{\theta^k}(\bar{y})$" to "$\sup_{k \geq 1} \inf_{\bar{y} \in R(y_0)} V_{\theta^k}(\bar{y}) = V^*(y_0)$" as T_0 tends to infinity.

Corollary 3.5.4. *For all $T_0 \geq 0$ and y_0 in $\mathbb{R}^d$,*

$$\sup_{k \geq 1} \inf_{\bar{y} \in R^{T_0}(y_0)} V_{\theta^k}(\bar{y}) \geq \limsup_k V_{\theta^k}(y_0) \geq \liminf_k V_{\theta^k}(y_0) \geq \sup_{k \geq 1} \inf_{\bar{y} \in R(y_0)} V_{\theta^k}(\bar{y}) = V^*(y_0).$$

For any states y and $\bar{y}$ in $\mathbb{R}^d$, let us define $\tilde{d}(y, \bar{y}) = \sup_{k \geq 1} |V_{\theta^k}(y) - V_{\theta^k}(\bar{y})|$. The space $(\mathbb{R}^d, \tilde{d})$ is now a *pseudometric* space (may not be Hausdorff).

The following is similar to the proof of Theorem 2.5 in Renault [46], and is also similar to the proof of Theorem 3.10 in Renault [44]. We rewrite it here for sake of completeness. Roughly speaking, we shall use the total boundedness of the space $(\{V_{\theta^k}\}, \|\cdot\|_\infty)$ so as to deduce that the state space $(\mathbb{R}^d, \tilde{d})$ is totally bounded for the pseudometric $\tilde{d}$. This allows us to prove the convergence for $\tilde{d}$ of the reachable set R^T to R in bounded time. We are then able to prove the uniform convergence of "$\sup_{k \geq 1} \inf_{\bar{y} \in R^{T_0}(y_0)} V_{\theta^k}(\bar{y})$" to "$\sup_{k \geq 1} \inf_{\bar{y} \in R(y_0)} V_{\theta^k}(\bar{y}) = V^*(y_0)$" as T_0 tends to infinity.

Proof for Part (ii) of Theorem 3.4.1. One direction is clear: the uniform convergence of (V_{θ^k}) implies the total boundedness of the space $(\{V_{\theta^k}\}, \|\cdot\|_\infty)$. Suppose that $(\{V_{\theta^k}\}, \|\cdot\|_\infty)$ is totally bounded, we are going to show that (V_{θ^k}) converges uniformly to V^*.

Fix any $\varepsilon > 0$. By assumption, there exists a finite set of indices I such that for all $k \geq 1$, there exists $i \in I$ satisfying

$$\|V_{\theta^k} - V_{\theta^i}\|_\infty \leq \varepsilon/3.$$

$\{(V_{\theta^i}(y)), y \in \mathbb{R}^d\}$ is a subset of the compact metric space $([0,1]^I, \|\cdot\|_\infty)$, thus it is itself totally bounded, so there exists a finite subset X of $\mathbb{R}^d$ such that

$$\forall y \in \mathbb{R}^d, \exists x \in X, \forall i \in I, |V_{\theta^i}(y) - V_{\theta^i}(x)| \leq \varepsilon/3.$$

We have obtained that for each $\varepsilon > 0$, there exists a finite subset X of $\mathbb{R}^d$ such that for every $y \in \mathbb{R}^d$, there is $x \in X$ satisfying: for any $k \geq 1$ there is some $i \in I$ with

$$|V_{\theta^k}(y) - V_{\theta^k}(x)| \leq |V_{\theta^k}(y) - V_{\theta^i}(y)| + |V_{\theta^i}(y) - V_{\theta^i}(x)| + |V_{\theta^i}(x) - V_{\theta^k}(x)| \leq \varepsilon,$$

thus $\tilde{d}(y, x) \leq \varepsilon$. This implies that the pseudometric space $(\mathbb{R}^d, \tilde{d})$ is itself totally bounded.

Fix now y_0 in $\mathbb{R}^d$. It is by definition that

$$\forall T, S \in \mathbb{R}_+, S \geq T,\ R^T(y_0) \subset R^S(y_0) \subset R(y_0),$$

and

$$\forall \bar{y} \in R(y_0),\ \exists \bar{T} > 0\ s.t.\ \bar{y} \in R^{\bar{T}}(y_0).$$

From the total boundedness of $(\mathbb{R}^d, \tilde{d})$, we show that R^T converges to R in the following sense

$$\forall \varepsilon > 0,\ \exists T \geq 0 :\ \forall \overline{y} \in R(y_0),\ \exists \tilde{y} \in R^T(y_0),\ \tilde{d}(\overline{y}, \tilde{y}) \leq \varepsilon. \tag{3.5.4}$$

Indeed, let us first take $\{y_\ell\}$ a finite ε-cover of $R(y_0)$ for $\tilde{d}$. For each y_ℓ, let $T_\ell > 0$ with $y_\ell \in R^{T_\ell}(y_0)$. We then take $T = \max T_\ell$. Now for any $\bar{y} \in R(y_0)$, there is some y_ℓ with $\tilde{d}(\bar{y}, y_\ell) \leq \varepsilon$. Moreover, $y_\ell \in R^{T_\ell}(y_0) \subset R^T(y_0)$. This proves (3.5.4).

Consider $k \geq 1$ and $T \geq 0$ given by assertion (3.5.4) for the fixed $\varepsilon > 0$. Let $\overline{y} \in R(y_0)$ be such that $V_{\theta^k}(\overline{y}_0) \leq \inf_{\overline{y} \in R(y)} V_{\theta^k}(\overline{y}) + \varepsilon$, and then $\tilde{y}$ in $R^T(y_0)$ be such that $\tilde{d}(\overline{y}, \tilde{y}) \leq \varepsilon$. Since V_{θ^k} is clearly 1-Lipschitz for $\tilde{d}$, we obtain $V_{\theta^k}(\tilde{y}) \leq \inf_{\overline{y} \in R(y_0)} V_{\theta^k}(\overline{y}) + 2\varepsilon$. Consequently, $\inf_{\overline{y} \in R^T(y_0)} V_{\theta^k}(\overline{y}) \leq \inf_{\overline{y} \in R(y_0)} V_{\theta^k}(\overline{y}) + 2\varepsilon$ for all k, so

$$\sup_{k \geq 1} \inf_{\overline{y} \in R^T(y_0)} V_{\theta^k}(\overline{y}) \leq \sup_{k \geq 1} \inf_{\overline{y} \in R(y_0)} V_{\theta^k}(\overline{y}) + 2\varepsilon = V^*(y_0) + 2\varepsilon.$$

One obtains that $\limsup_k V_{\theta^k}(y_0) \leq \liminf_k V_{\theta^k}(y_0) + 2\varepsilon$, and so $\left(V_{\theta^k}(y_0)\right)_k$ converges to V^*. Since $(\mathbb{R}^d, \tilde{d})$ is totally bounded and all V_{θ^k} are 1-Liptschitz, the convergence is uniform. □

3.6 Concluding discussions

We comment on two unknown questions.

The first question concerns a weaker form of the long-term condition (LTC) defined as:

Long-term condition' (LTC') A sequence of evaluations $(\theta^k)_{k \geq 1}$ satisfies the LTC' if:

$$\forall s > 0,\quad TV_s(\theta^k) \xrightarrow[k \to \infty]{} 0.$$

Question 1. Is the LTC' strictly weaker than the LTC ?

One might want to construct an example of $(\theta^k)_k$ such that $TV_s(\theta^k) \xrightarrow[k \to \infty]{} 0$ for all $s > 0$ while $\overline{TV}_{s_0}(\theta^k) \xrightarrow[k \to \infty]{} \alpha > 0$ for some $s_0 > 0$ and $\alpha > 0$. The following example shows that this is possible if we consider only s being rational numbers. In general, the question is still open.

Example 12. *Given a positive integer k, consider the density θ^k with support included in $[0, k]$ by dividing $[0, k]$ into k^2 consecutive small intervals of length $1/k$, and θ^k is uniform over the union of all small odd intervals and puts no weight on small even intervals. Define the support*

$$S_k = \bigcup_{l \in \mathbb{N}, l \leq \frac{k^2-1}{2},} \left[\frac{2l}{k}, \frac{2l+1}{k}\right).$$

θ^k has density:

$$f_k(x) = \frac{2}{k}\,\mathbb{1}_{x\in S_k} = \frac{2}{k}\,\mathbb{1}_{x\in[0,k],E(kx)\in 2\mathbb{N}}$$

(where $2\mathbb{N}$ is the set of even numbers in $\mathbb{N}$, $E(x)$ is the integer part of x).

For each k, we have (consider $s = 1/k$):

$$\sup_{0\le s\le 1}\int_{x\ge 0}|f_k(x+s) - f_k(x)|\mathrm{d}x \ge 2 - 1/k$$

Consider now only k of the form $n!$, and we define the density $g_n = f_{n!}$ for each n in $\mathbb{N}$. For all $x \ge 0$,

$$g_n(x+s) - g_n(x) = \frac{2}{n!}\left(\mathbb{1}_{E(n!(x+s))\in 2\mathbb{N},x+s\le n!} - \mathbb{1}_{E(n!x)\in 2\mathbb{N},x\le n!}\right).$$

Assume s is a rational number. Then for n large enough, $n!s$ is an even integer, so for all x such that $0 \le x \le n! - s$, we have $g_n(x+s) - g_n(x) = 0$. Consequently,

$$\int_{x\ge 0}|g_n(x+s) - g_n(x)|\mathrm{d}x \xrightarrow[n\to\infty]{} 0.$$

The second question asks whether the existence of general limit value is strictly stronger than the uniform convergence of t-horizon values.

From Oliu-Barton and Vigeral [40], we know that a uniform Tauberian theorem holds in optimal control problems: the uniform convergence of V_t as t tends to infinity is equivalent to the uniform convergence of V_λ as λ tends to zero, and in case of convergence, both limits are the same. In Example 2, there is no uniform convergence of V_t (or equivalently of the V_λ) but uniform convergence of value functions for a particular sequence of evaluations satisfying the LTC. This leads us to ask the following

Question 2. Does the uniform convergence of V_t as t tends to infinity imply the existence of general limit value ?

We show that this is at least the case for uncontrolled problems (compare with Proposition 5.1 in Renault [46]). This provides also explanation for Example 6.

Proposition 3.6.1. *For an uncontrolled problem, the uniform convergence of V_t to some function φ as t tends to infinity implies general uniform convergence of the value functions $\{V_\theta\}$ to φ.*

Proof. Fix $\varepsilon > 0$. By the uniform convergence of V_t to φ, there is $S > 0$ such that

$$|V_S(y_0) - \varphi(y_0)| \le \varepsilon/3,\ \forall y_0 \in \mathbb{R}^d.$$

Consider any $\theta \in \Delta(\mathbb{R}^+)$ and $y_0 \in \mathbb{R}^d$. Denote by y_t the state reached by the uncontrolled dynamic at time t and starting from y_0. We have:

$$V_\theta(y_0) = \int_{[0,+\infty)} g(y_s)\mathrm{d}\theta(s) \ \text{ and } \ \varphi(y_s) = \varphi(y_0),\ \forall s \ge 0.$$

We write $V_{t,S}(y_0) = \frac{1}{S}\int_{[t,t+S)} g(y_s)\mathrm{d}s$ for each $t \ge 0$. Indeed, we have $V_{t,S}(y_0) = V_S(y_s)$, thus

$$|V_{t,S}(y_0) - \varphi(y_t)| = |V_{t,S}(y_0) - \varphi(y_0)| \le \varepsilon/3 \text{ for all } t \ge 0. \tag{3.6.1}$$

Integrating $V_{t,S}(y_0)$ over $t \geq 0$ w.r.t. θ, we obtain (using Fubini's theorem):

$$\int_{[0,+\infty)} V_{t,S}(y_0)\mathrm{d}\theta(t) = \int_{[0,+\infty)} \left(\frac{1}{S}\int_{[t,t+S)} g(y_s)\mathrm{d}s\right)\mathrm{d}\theta(t) = \int_{[0,+\infty)} \beta_s(\theta,S)g(y_s)\mathrm{d}s = V_{\varsigma(\theta,S)}(y_0),$$

where $\beta_s(\theta,S) = \frac{1}{S}\int_{[\max\{0,s-S\},s)} \mathrm{d}\theta(t)$, $\forall s \geq 0$, and $\varsigma(\theta,S)$ is an evaluation with $s \mapsto \beta_s(\theta,S)$ its density. This, together with (3.6.1), implies that $\left|V_{\varsigma(\theta,S)}(y_0) - \varphi(y_0)\right| \leq \varepsilon/3$. Next, we prove

$$\left|V_\theta(y_0) - V_{\varsigma(\theta,S)}(y_0)\right| \leq \sup_{Q\in\mathcal{B}(\mathbb{R}_+)} |\theta(Q) - \varsigma(\theta,S)(Q)| \leq 2TV_S(\theta). \tag{3.6.2}$$

The first inequality follows from Hahn's decomposition theorem applied to the sign measure "$\theta - \varsigma(\theta,S)$" (cf. Lemma 3.3.8). For the second one, we consider $Q \in \mathbb{R}_+$. Write $\beta_s(\theta,S) = \frac{1}{S}\int_{[s-S,s)} \mathrm{d}\theta(t)$ for all $s \geq 0$ by extending θ to $[-S,0)\cup\mathbb{R}_+$ with null on $[-S,0)$. This gives us:

$$\begin{aligned}\varsigma(\theta,S)(Q) &= \int_Q \beta_s(\theta,S)\mathrm{d}s = \frac{1}{S}\int_Q \left(\int_{[s-S,s)} \mathrm{d}\theta(t)\right)\mathrm{d}s \\ &= \int_{Q-S}\left(\frac{1}{S}\int_{[t,t+S)} \mathrm{d}s\right)\mathrm{d}\theta(t) \\ &= \theta(Q-S).\end{aligned}$$

We deduce then $|\theta(Q) - \varsigma(\theta,S)(Q)| \leq \theta([0,S)) + TV_S(\theta) \leq 2TV_S(\theta)$. This proves (3.6.2), thus

$$\left|V_\theta(y_0) - \varphi(y_0)\right| \leq \left|V_\theta(y_0) - V_{\varsigma(\theta,S)}(y_0)\right| + \left|V_{\varsigma(\theta,S)}(y_0) - \varphi(y_0)\right| \leq 2TV_S(\theta) + \varepsilon/3, \ \forall y_0.$$

This implies general uniform convergence of $\{V_\theta\}$ to φ by considering all θ with $\overline{TV}_S(\theta) \leq \varepsilon/3$. □

3.7 Appendix

Proof for Lemma 3.3.6: The following computation of $f'_\theta(t)$ is straightforward:

$$\forall t > 0, \quad f'_\theta(t) = \frac{1}{\sigma\sqrt{2\pi}}\Big[\exp\left(-\frac{1}{2}\Big(\frac{t-m}{\sigma}\Big)^2\right)\frac{m-t}{\sigma^2} - \exp\left(-\frac{1}{2}\Big(\frac{t+m}{\sigma}\Big)^2\right)\frac{m+t}{\sigma^2}\Big],$$

thus

$$\begin{aligned}&f'_\theta(t) > 0 \ \ (resp. < 0) \\ \iff &(m-t)\exp\left(-\frac{1}{2}\Big(\frac{t-m}{\sigma}\Big)^2\right) - (m+t)\exp\left(-\frac{1}{2}\Big(\frac{t+m}{\sigma}\Big)^2\right) > 0 \ \ (resp. < 0).\end{aligned}$$

As a consequence, one obtains that $f'_\theta(t) < 0$, $\forall t \geq m$. Now we look at $t \in (0,m)$. Denote $H(t) =_{def} \exp\left(\frac{2mt}{\sigma^2}\right) - \frac{m+t}{m-t}$, which enables us to write:

$$f'_\theta(t) > 0 \ (resp. < 0) \iff H(t) > 0 \ (resp. < 0), \quad \forall t \in (0,m).$$

From the above analysis, we deduce that the proof of the lemma is reduced to the proof for

Claim *There is some $t^* \in [0,m)$ such that $H(t) < 0$ for $t \in (0,t^*)$ and $H(t) > 0$ for $t \in (t^*,m)$. Moreover, such t^* satisfies $(t^*)^2 \geq m^2 - \sigma^2$.*

In order to prove the claim, we compute

– the values at the end point: $H(0) = 0$ and $\lim_{t\to m^-} H(t) = -\infty$;
– the first-order derivative at any $t \in [0, m)$:

$$H'(t) = \exp\left(\frac{2mt}{\sigma^2}\right)\frac{2m}{\sigma^2} - \frac{2m}{(m-t)^2} \tag{3.7.1}$$

– at any rest point $t^e \in [0, m)$ (i.e., $H(t^e) = 0$):

$$\exp\left(\frac{2mt^e}{\sigma^2}\right) = \frac{m+t^e}{m-t^e},$$

which is substituted back into (3.7.1), to yield

$$H'(t^e) > 0 \ (\ \textit{resp.}\ H'(t^e) < 0\ ) \iff (t^e)^2 < m^2 - \sigma^2 \ (\ \textit{resp.}\ (t^e)^2 > m^2 - \sigma^2\ ). \tag{3.7.2}$$

Next, it is easy for us to prove the following result:

Let $t_1^e \in [0, m)$ be a rest point for $H(\cdot)$, and suppose that $t_2^e \in (t_1^e, m)$ is the smallest rest point after t_1^e. Then $H'(t_1^e)H'(t_2^e) \le 0$ and if $H'(t_1^e) \le 0$, such t_2^e does not exist.

Indeed, $H'(t_1^e)H'(t_2^e) \le 0$ can be derived from the continuity of $H(\cdot)$; suppose that $H'(t_1^e) \le 0$, we have from (3.7.2) that $(t_1^e)^2 \ge m^2 - \sigma^2$ and $H'(t_2^e) \ge 0$, thus $(t_2^e)^2 \le m^2 - \sigma^2$. However, this leads to a controdiction to $t_2^e > t_1^e$, so t_2^e does not exist whenever $H'(t_1^e) \le 0$.

Finally, remark that $H(0) = 0$, thus $t = 0$ is a rest point. We discuss the following two cases:

Case 1. $m^2 - \sigma^2 \le 0$, thus $H'(0) \le 0$.

This implies that no rest point exists after 0. Since $\lim_{t\to m^-} H(t) = -\infty$, we deduce that $H(t) < 0,\ \forall t \in (0, m)$. The claim is proved for $t^* = 0$.

Case 2. $m^2 - \sigma^2 > 0$, thus $H'(0) > 0$.

$\lim_{t\to m^-} H(t) = -\infty$ implies that some rest point exists in $(0, m)$. Take t^e the closest to 0, implying that $H(t) > 0,\ \forall t \in (0, t^e)$. Further, we obtain that $H'(t^e) \le 0$ by the continuity of $H(\cdot)$. Again, there exists no other rest point after t^e. Since $\lim_{t\to m^-} H(t) = -\infty$, we deduce that $H(t) < 0,\ \forall t \in (t^e, m)$. The claim is proved for $t^* = t^e$.

To conclude, we see that in both cases such t^* exists and satisfies $(t^*)^2 \ge m^2 - \sigma^2$, thus the claim is proved. This finishes our proof for the lemma. □

Acknowledgements The authors thank Sylvain Sorin for numerous helpful comments. Part of this research was taken under the grant ANR-10-BLAN 0112.

Chapter 4

Valeur uniforme pour le probème de contrôle optimal dans le cadre "compact non expansif" avec évaluations générales

Résumé Nous considérons le problème de contrôle optimal avec le coût de la trajectoire évalué par une mesure de probabilité sur $\mathbb{R}_+$. Nous utilisons la notion de s- variation totale introduite dans Li *et al.* [30] pour définir une condition de régularité asymptotique pour une suite des mesures de probabilité. Dans des cas particuliers des moyennes de Cesàro ou des moyennes d'Abel, cette condition exige que l'horizon tende vers l'infini ou que le facteur d'escompté tende vers zéro. Pour le système de contrôle satisfaisant une certaine condition de non-expansivité et définie sur un domaine invariant compact, nous prouvons l'existence des contrôles ε-optimaux dans tous les problèmes où le coût est évalué par des mesures de probabilité suffisamment régulières. Ce résultat généralise celui de Quincampoix et Renault [42], qui ont établi l'existence de la valeur uniforme pour les problèmes de contrôle où le coût de la trajectoire est évalué par les moyennes de Cesàro.

Keywords Contrôle optimal, valeur uniforme, la valeur moyenne à long temps, évaluation générale

---

Ce chapitre est issu de l'article *Uniform value for some nonexpansive optimal control problems with general evaluations.*

Uniform value for some nonexpansive optimal control problems with general evaluations

Abstract We consider optimal control problems with the running cost evaluated by a probability measure over $\mathbb{R}_+$. To study limit properties with respect to the evaluation, we use the notion of s-total variation introduced in Li *et al.* Li *et al.* [30] to define an asymptotic regularity condition for a sequence of probability measures. In particular case of Cesàro means or Abel means, this condition asks for the horizon T, on which the cost is averaged, to tend to infinity or for the discount factor λ to tend to zero. For the control system defined on a compact domain and satisfying some nonexpansive condition, we prove the existence of ε-optimal control for all control problems where the cost is evaluated by sufficiently regular probability measures. This generalizes result in Quincampoix and Renault [42], which proved the existence of uniform value for the running cost evaluated by Cesàro means.

Keywords Optimal control, uniform value, long time average value, general evaluation

4.1 Introduction

Let U be a compact subset of a separable metric space. A *control* $\mathbf{u}$ is a measurable function from $\mathbb{R}_+$ to U. Denote by $\mathcal{U}$ the set of all controls. We consider the following control system:

$$\mathbf{y}'(t) = f\big(\mathbf{y}(t), \mathbf{u}(t)\big),\ \mathbf{y}(0) = y_0. \tag{4.1.1}$$

where $f : \mathbb{R}^d \times U \to \mathbb{R}^d$, and $y_0 \in \mathbb{R}^d$ is the initial state. Let $g : \mathbb{R}^d \times U \to \mathbb{R}$ be the running cost function. We make the following assumptions on f and g:

$$\begin{cases} \text{the function } g : \mathbb{R}^d \times U \to \mathbb{R} \text{ is Borel measurable and bounded;} \\ \text{the function } f : \mathbb{R}^d \times U \to \mathbb{R}^d \text{ is Borel measurable, and satisfies:} \\ (*).\ \exists L \geq 0, \forall (y, \overline{y}) \in \mathbb{R}^{2d}, \forall u \in U, ||f(y,u) - f(\overline{y}, u)|| \leq L||y - \overline{y}||, \\ (**).\ \exists a > 0, \forall (y,u) \in \mathbb{R}^d \times U, ||f(y,u)|| \leq a(1 + ||y||). \end{cases} \tag{4.1.2}$$

Then, given $y_0 \in \mathbb{R}^d$, any $\mathbf{u} \in \mathcal{U}$ defines a unique absolutely continuous solution to (4.1.1) on $\mathbb{R}_+$, denoted $\mathbf{y}(t, u, y_0)$.

Denote by $\mathcal{J} = \langle U, g, f \rangle$ the optimal control problem described above. $\Delta(\mathbb{R}_+)$ is the set of probability measures on $\mathbb{R}_+$ and any $\theta \in \Delta(\mathbb{R}_+)$ is called an *evaluation*. The θ-value of the control problem is defined as:

$$V_\theta(y_0) = \inf_{\mathbf{u} \in \mathcal{U}} \gamma_\theta(y_0, \mathbf{u}), \quad \text{with} \quad \gamma_\theta(y_0, \mathbf{u}) = \int_0^{+\infty} g\big(\mathbf{y}(t, \mathbf{u}, y_0), \mathbf{u}(t)\big) \mathrm{d}\theta(t). \tag{4.1.3}$$

The following notion is introduced in Li *et al.* [30] to define a regularity of any evaluation θ, and is used to study the asymptotic behavior of V_θ as θ becomes more and more regular.

Definition 4.1.1. *For any $s \geq 0$, the* ***s-total variation*** *of an evaluation θ is:*

$$TV_s(\theta) = \max_{Q \in \mathcal{B}(\mathbb{R}_+)} |\theta(Q) - \theta(Q+s)|.$$

Definition 4.1.2. *The optimal control problem $\mathcal{J} = \langle U, g, f \rangle$ has a* ***general limit value*** *given by some function V defined on $\mathbb{R}^d$ if: for each $\varepsilon > 0$ there is some $\eta > 0$ and $S > 0$ such that:*

$$\forall \theta \in \Delta(\mathbb{R}_+), \quad \Big(\sup_{0 \leq s \leq S} TV_s(\theta) \leq \eta \Longrightarrow (\forall y_0 \in \mathbb{R}^d, \ |V_\theta(y_0) - V(y_0)| \leq \varepsilon) \Big).$$

When we consider the usual Cesàro mean (T-horizon average) or Abel mean (λ-discounted) of the running cost, the condition "vanishing s-total variation" corresponds to that "the horizon T increasing to infinity" or that "the discount factor λ decreasing to zero" (cf. Li *et al.* [30]).

We restrict ourselves to the study of the following class of control problems.

Definition 4.1.3. *The optimal control problem $\mathcal{J} = \langle U, f, g \rangle$ is* ***compact nonexpansive*** *if:*
A.1) the control dynamic (4.1.1) has a compact invariant set $Y \subseteq \mathbb{R}^d$, i.e. $\mathbf{y}(t, \mathbf{u}, y_0) \in Y$, $\forall t \geq 0$ for all $\mathbf{u} \in \mathcal{U}$ and $y_0 \in Y$;
A.2) the running cost g does not depend on u and is continuous in $y \in Y$;
A.3) the control dynamic (4.1.1) is nonexpansive, i.e.,

$$\forall y_1, y_2 \in \mathbb{R}^d, \ \sup_{a \in U} \inf_{b \in U} \langle y_1 - y_2, f(y_1, a) - f(y_2, b) \rangle \leq 0. \tag{4.1.4}$$

Consider a control problem $\mathcal{J} = \langle U, g, f \rangle$ compact nonexpansive with invariant set Y. Li *et al.* [30] (see Corollary 4.5) proved that $\mathcal{J}$ has a general limit value characterized by the following function:

$$V^*(y_0) = \sup_{\theta \in \Delta(\mathbb{R}_+)} \inf_{s \in \mathbb{R}_+} \inf_{\mathbf{u} \in \mathcal{U}} \int_0^\infty g\Big(\mathbf{y}(t+s, \mathbf{u}, y_0)\Big) \mathrm{d}\theta(t), \quad \forall y_0 \in Y. \tag{4.1.5}$$

Proposition 4.1.4. *[Li et al. 2015] Let $\mathcal{J} = \langle U, f, g \rangle$ be compact nonexpansive with invariant set Y. Then V^* is the general limit value of $\mathcal{J}$.*

We study here a notion of value which is stronger than the general limit value, namely the general uniform value, which asks for the existence of approximately optimal control in all control problems where the running cost is evaluated by any θ with $\sup_{0 \leq s \leq S} TV_s(\theta)$ sufficiently small for some fixed $S > 0$. In order to give a formal definition, we first introduce the notion of *random control*.

Definition 4.1.5. *A* ***random control*** *is a pair $\big((\Omega, \mathcal{B}(\Omega), \lambda), \mathbf{u}\big)$, where $(\Omega, \mathcal{B}(\Omega), \lambda)$ is some standard Borel probability space and $\mathbf{u} : Y \times \Omega \times [0, +\infty) \to U$ is a Borel measurable mapping*[1].

Denote by $\tilde{\mathcal{U}}$ the set of all random controls, which is convex in the following sense. For any $\mathbf{u}_1, \mathbf{u}_2 \in \tilde{\mathcal{U}}$ and $\alpha \in [0, 1]$, define $\mathbf{u} = \alpha \mathbf{u}_1 + (1 - \alpha)\mathbf{u}_2$ to take the value of $\mathbf{u}_1$ with probability α and of $\mathbf{u}_2$ with probability $(1 - \alpha)$. It is easy to construct a product probability space such that $\mathbf{u}$ is in $\tilde{\mathcal{U}}$.

1. We have extended the definition of a random control to dependent on the initial state for later use.

We might shortly write Ω for the triple $(\Omega, \mathcal{B}(\Omega), \lambda)$. Let $(\Omega, \mathbf{u})$ be a random control, then for any initial point $y_0 \in Y$ and any $\omega \in \Omega$, $\mathbf{u}_\omega(y_0, \cdot) := \mathbf{u}(y_0, \omega, \cdot)$ defined from $[0, +\infty)$ to U is a (pure) control in $\mathcal{U}$, which we denote by $\mathbf{u}_\omega(y_0)$. The Borel probability space Ω serves as a random device for the controller to choose a pure control in $\mathcal{U}$.

The expected θ-evaluated payoff induced by any random control $(\Omega, \mathbf{u})$ and initial point y_0 is denoted by

$$\gamma_\theta(y_0, \mathbf{u}) = \int_\Omega \gamma_\theta(y_0, \mathbf{u}_\omega(y_0)) \mathrm{d}\lambda(\omega) = \int_\Omega \int_{[0,+\infty)} g\Big(\mathbf{y}(t, \mathbf{u}_\omega(y_0), y_0)\Big) \mathrm{d}\theta(t) \mathrm{d}\lambda(\omega),$$

and the expected θ-value in random controls is $\tilde{V}_\theta(y_0) = \inf_{\mathbf{u} \in \tilde{\mathcal{U}}} \gamma_\theta(y_0, \mathbf{u})$. The payoff function $\gamma_\theta(y_0, \cdot)$ is affine [2] in $\mathbf{u}$, thus the value function in random controls is the same as that in pure controls, that is, $\tilde{V}_\theta(y_0) = V_\theta(y_0)$ for all $y_0 \in Y$ and $\theta \in \Delta(\mathbb{R}_+)$.

Definition 4.1.6. *The optimal control problem $\mathcal{J} = \langle U, f, g \rangle$ has a* ***general uniform value*** *if for each $\varepsilon > 0$ there is some $\eta > 0$, $S > 0$ and a random control $\mathbf{u} \in \tilde{\mathcal{U}}$ such that:*

$$\forall \theta \in \Delta(\mathbb{R}_+), \quad \Big(\sup_{0 \le s \le S} TV_s(\theta) \le \eta \Longrightarrow \big(\forall y_0 \in Y,\ \gamma_\theta(y_0, \mathbf{u}) \le V^*(y_0) + \varepsilon\big)\Big),$$

where V^* is defined as in (4.1.5).

The random control $\mathbf{u}$ appearing in the above definition is called an ε-optimal control for the control problem $\mathcal{J}$.

Our main result is the following:

Theorem 4.1.7. *Assume that the optimal control problem $\mathcal{J} = \langle U, f, g \rangle$ is compact nonexpansive. Then it has a general uniform value.*

Quincampoix and Renault [42] proved the existence of (pure) ε-optimal control for the compact nonexpansive control problems when the running cost is evaluated by Cesàro means. Our result generalizes it. However, it is unknown whether pure ε-optimal control exists with general evaluations. Our proof is partially inspired by Renault [47], which established analogous results in discrete time framework.

4.2 Preliminaries

We introduce several further notations concerning random controls.

Our use of random strategies/control in continuous time game/control problem follows Cardaliaguet [15] (cf. Aumann [4] for the introduction of randomized strategies for infinite games). We are going to work on a set $\mathcal{S}$ of probability spaces, which is stable by countable product. To fix the ideas, choose

$$\mathcal{S} = \{([0,1]^n, \mathcal{B}([0,1]^n), \lambda^n)\,,\ \text{for some } n \in \mathbb{N}^* \cup \{\infty\}\},$$

where $\mathcal{B}([0,1]^n)$ is the class of Borel sets of $[0,1]^n$, λ^n is the Lebesgue measure on $\mathbb{R}^n$ for $n < \infty$ and $\mathcal{B}([0,1]^\infty)$ is the product Borel-field, λ^∞ is the product measure for $n = \infty$.

2. To see this point, consider for example the Borel probability space to be $\big([0,1], \mathcal{B}([0,1]), \lambda\big)$ where λ is the Lebesgue measure. We take any $\mathbf{u}^1, \mathbf{u}^2 : Y \times [0,1] \times [0, +\infty) \to U$ two random controls, any $\alpha \in [0,1]$, and let $\mathbf{u}^3 : Y \times [0,1] \times [0, +\infty) \to U$ be the random control as a convex combination of $\mathbf{u}^1$ and $\mathbf{u}^2$ with coefficient α: for any $y_0 \in Y$, $\mathbf{u}^3(y_0, \omega, t) = \mathbf{u}^1(y_0, \frac{\omega}{\alpha}, t)$ for $\omega \in [0, \alpha]$ and $\mathbf{u}^3(y_0, \omega, t) = \mathbf{u}^2(y_0, \frac{\omega - \alpha}{1-\alpha}, t)$ for $\omega \in (\alpha, 1]$. Using the change of variables, we obtain $\gamma_\theta(y_0, \mathbf{u}^3) = \alpha \cdot \gamma_\theta(y_0, \mathbf{u}^1) + (1-\alpha) \cdot \gamma_\theta(y_0, \mathbf{u}^2)$.

We might write simply $\mathbf{u} \in \tilde{\mathcal{U}}$ without explicitly mentioning the underlying probability space.

For any $(y_0, \mathbf{u}) \in Y \times \tilde{\mathcal{U}}$ and $t \geq 0$, we denote $\tilde{\mathbf{y}}(t, \mathbf{u}, y_0) = \int_\Omega \delta_{\mathbf{y}(t,\mathbf{u}_\omega(y_0),y_0)} \mathrm{d}\lambda(\omega)$ for the distribution of the state at time t. Any $z \in \Delta(Y)$ is a probability distribution over Y. Let z be the distribution of the initial state, then $t \mapsto \tilde{\mathbf{y}}(t, \mathbf{u}, z) = \int_Y \int_\Omega \delta_{\mathbf{y}(t,\mathbf{u}_\omega(y_0),y_0)} \mathrm{d}\lambda(\omega) \mathrm{d}z(y_0)$ is the expected trajectory in $\Delta(Y)$ w.r.t. z. The above notations are consistent when a point $y_0 \in Y$ is identified with the Dirac measure $\delta_{y_0} \in \Delta(Y)$.

For any $\tilde{\mathbf{y}} \in \Delta(Y)$, let $g(\tilde{\mathbf{y}}) = \int_{p \in Y} g(p) \mathrm{d}\tilde{\mathbf{y}}(p)$ be the affine extension of g. Using Fubini's theorem, we have: for any $y_0 \in Y$ and $\mathbf{u} \in \tilde{\mathcal{U}}$,

$$\gamma_\theta(y_0, \mathbf{u}) = \int_{[0,+\infty)} \int_\Omega g\Big(\mathbf{y}(t, \mathbf{u}_\omega(y_0), y_0)\Big) \mathrm{d}\lambda(\omega) \mathrm{d}\theta(t) = \int_{[0,+\infty)} g\Big(\tilde{\mathbf{y}}(t, \mathbf{u}, y_0)\Big) \mathrm{d}\theta(t).$$

Before proceeding to the proof of Theorem 4.1.7, we establish in this subsection several preliminary results concerning the nonexpansive property of the dynamic in $\Delta(Y)$.

Let d_{KR} be the *Kantorovich-Rubinstein distance* on $\Delta(Y)$:

$$\forall z, z' \in \Delta(Y), \quad d_{KR}(z, z') = \sup_{h \in Lip(1)} \left| \int_Y h \mathrm{d}z - \int_Y h \mathrm{d}z' \right|,$$

where $Lip(1)$ denotes the set of bounded 1-Lipschitz functions defined on Y.

Lemma 4.2.1. *For any z_1, z_2 in $\Delta(Y)$ and $\mathbf{u} : Y \times \Omega \times [0, +\infty) \to U$ a random control, there exists some random control $\mathbf{v} : Y \times \hat{\Omega} \times \Omega \times [0, +\infty) \to U$ such that*

$$d_{KR}\Big(\tilde{\mathbf{y}}(t, \mathbf{u}, z_1), \tilde{\mathbf{y}}(t, \mathbf{v}, z_2)\Big) \leq d_{KR}(z_1, z_2), \ \forall t \geq 0.$$

Proof: We fix $\mathbf{u} : Y \times \Omega \times [0, +\infty) \to U$ a random control defined on the probability space $(\Omega, \mathcal{B}(\Omega), \lambda)$. We first show the result for z_1 and z_2 being Dirac measures. Let p, q in Y. $\mathbf{u}_\omega(p) : [0, +\infty) \to U$ is a pure control for any $\omega \in \Omega$. By the nonexpansive assumption, Proposition 3.6 in Quincampoix and Renault [42] implies that: for given $\omega \in \Omega$, there is some pure control $\mathbf{u}^p(q, \omega, \cdot) : [0, +\infty) \to U$, which we denote by $\mathbf{u}^p_\omega(q) := \mathbf{u}^p(q, \omega, \cdot)$, such that

$$\Big\| \mathbf{y}(t, \mathbf{u}_\omega(p), p) - \mathbf{y}(t, \mathbf{u}^p_\omega(q), q) \Big\| \leq \|p - q\|, \ \forall t \geq 0. \tag{4.2.1}$$

Moreover, one can take the mapping $(q, t) \mapsto \mathbf{u}^p_\omega(q, t) := \mathbf{u}^p(q, \omega, t)$ jointly measurable[3]. Letting $\mathbf{u}^p(q, t) = \big(\mathbf{u}^p_\omega(q, t)\big)_{\omega \in \Omega}$ for all $t \geq 0$, this defines a random control:

$$\begin{aligned} \mathbf{u}^p : Y \times \Omega \times [0, +\infty) &\mapsto U \\ (q, \omega, t) &\mapsto \mathbf{u}^p_\omega(q, t). \end{aligned}$$

Moreover, we deduce from (4.2.1) that: for any $t \geq 0$,

$$\begin{aligned} d_{KR}\Big(\tilde{\mathbf{y}}(t, \mathbf{u}, p), \tilde{\mathbf{y}}(t, \mathbf{u}^p, q)\Big) &= d_{KR}\Big(\int_\Omega \delta_{\mathbf{y}(t,\mathbf{u}_\omega(p),p)} \mathrm{d}\lambda(\omega), \int_\Omega \delta_{\mathbf{y}(t,\mathbf{u}^p_\omega(q),q)} \mathrm{d}\lambda(\omega) \Big) \\ &\leq \int_\Omega \Big\| \mathbf{y}(t, \mathbf{u}_\omega(p), p) - \mathbf{y}(t, \mathbf{u}^p_\omega(q), q) \Big\| \mathrm{d}\lambda(\omega) \\ &\leq ||p - q||. \end{aligned} \tag{4.2.2}$$

3. Indeed, consider for (4.1.4): we fix any $y_2 \in Y$ and $a^{\cdot} : x \mapsto a^x \in U$ measurable. Then we can apply the measurable selection theorem (cf. Theorem 9.1 in Wagner [66]) for the optimization problem $\inf_{b \in U} \big\langle y_1 - y_2, f(y_1, a^{y_1}) - f(y_2, b) \big\rangle$ to obtain the existence of a measurable mapping $b^{\cdot} : x \mapsto b^x \in U$ satisfying $\big\langle y_1 - y_2, f(y_1, a^{y_1}) - f(y_2, b^{y_1}) \big\rangle \leq 0, \ \forall y_1 \in Y$. The same augument as in the proof of Prop. 3.6 in Quincampoix and Renault [42] implies the result.

Consider now z_1, z_2 in $\Delta(Y)$. By the Kantorovich-Rubinstein duality formula (cf. Theorem 5.10 in Villani [64]), there is a coupling $\xi(\cdot,\cdot) \in \Delta(Y \times Y)$ with first marginal z_1 and second marginal z_2, satisfying:

$$d_{KR}(z_1, z_2) = \int_{Y\times Y} \|p - q\| \mathrm{d}\xi(p, q). \tag{4.2.3}$$

Let $q \mapsto \xi(\cdot|q) \in \Delta(Y)$ be the conditional distribution of ξ on its first marginal. Consider now the Borel isomorphic mapping (between two standard Borel spaces)

$$h_q : \hat{\Omega} = \Big([0,1], \mathcal{B}([0,1]), \lambda\Big) \mapsto \Big(Y, \mathcal{B}(Y), \xi(\cdot|q)\Big)$$

where

$$\forall B \in \mathcal{B}(Y),\ \xi(B|q) = \lambda\left(h_q^{-1}(B)\right).$$

Define now

$$\begin{array}{rcl} \mathbf{v} : Y \times \hat{\Omega} \times \Omega \times [0, +\infty) & \mapsto & U \\ (q, \hat{\omega}, \omega, t) & \mapsto & \mathbf{v}_{(\hat{\omega},\omega)}(q, t) = \mathbf{u}_{\omega}^{h_q(\hat{\omega})}(q, t), \end{array}$$

which is jointly measurable as the composition function of $\mathbf{u}_{\omega}^{\cdot}(q, t)$ and $h_q(\hat{\omega})$. $(\Omega_{\mathbf{v}}, \mathbf{v})$ is then a random control defined on the product Borel probability space $(\hat{\Omega}\otimes\Omega, \mathcal{B}(\hat{\Omega}\otimes\Omega), \lambda^2)$. The interpretation of the random control $\mathbf{v}$ is that: at each initial point $q \in Y$, $\mathbf{v}(q, \cdot)$ randomly takes the value $\mathbf{u}^p(q, \cdot)$ according to the probability law $\mathrm{d}\xi(p|q)$.

Next, we check that $\mathbf{v}$ satisfies the nonexpansive condition. For any $t \geq 0$ and $q \in supp(z_2)$, we use Fubini's theorem (first on $\hat{\Omega} \times \Omega$ and then on $Y \times Y$) and the change of variable "$p = h_q(\hat{\omega})$" to obtain:

$$\begin{aligned} \tilde{\mathbf{y}}(t, \mathbf{v}, z_2) = \int_Y \tilde{\mathbf{y}}(t, \mathbf{v}, q)\mathrm{d}z_2(q) &= \int_Y \int_{\hat{\Omega}\times\Omega} \delta_{\mathbf{y}\left(t, \mathbf{u}_{\omega}^{h_q(\hat{\omega})}(q), q\right)} \mathrm{d}\lambda^2(\hat{\omega}, \omega)\mathrm{d}z_2(q) \\ &= \int_Y \int_{\hat{\Omega}} \left[\int_{\Omega} \delta_{\mathbf{y}\left(t, \mathbf{u}_{\omega}^{h_q(\hat{\omega})}(q), q\right)} \mathrm{d}\lambda(\hat{\omega})\right] \mathrm{d}\lambda(\omega)\mathrm{d}z_2(q) \\ &= \int_Y \int_Y \left[\int_{\Omega} \delta_{\mathbf{y}\left(t, \mathbf{u}_{\omega}^{p}(q), q\right)} \mathrm{d}\lambda(\omega)\right])\mathrm{d}\xi(p|q)\mathrm{d}z_2(q) \\ &= \int_{Y\times Y} \tilde{\mathbf{y}}(t, \mathbf{u}^p, q)\mathrm{d}\xi(p, q). \end{aligned}$$

We then deduce that:

$$\begin{aligned} d_{KR}\Big(\tilde{\mathbf{y}}(t, \mathbf{u}, z_1), \tilde{\mathbf{y}}(t, \mathbf{v}, z_2)\Big) &= d_{KR}\Big(\int_Y \tilde{\mathbf{y}}(t, \mathbf{u}, p)\mathrm{d}z_1(p),\ \int_{Y\times Y} \tilde{\mathbf{y}}(t, \mathbf{u}^p, q)\mathrm{d}\xi(p, q)\Big) \\ &= d_{KR}(\int_{Y\times Y} \tilde{\mathbf{y}}(t, \mathbf{u}, p)\mathrm{d}\xi(p, q),\ \int_{Y\times Y} \tilde{\mathbf{y}}(t, \mathbf{u}^p, q)\mathrm{d}\xi(p, q)) \\ &\leq \int_{Y\times Y} d_{KR}\Big(\tilde{\mathbf{y}}(t, \mathbf{u}, p), \tilde{\mathbf{y}}(t, \mathbf{u}^p, q)\Big)\mathrm{d}\xi(p, q) \\ \Big(\text{by (4.2.2)}\Big) &\leq \int_{Y\times Y} ||p - q||\mathrm{d}\xi(p, q) \\ \Big(\text{by (4.2.3)}\Big) &= d_{KR}(z_1, z_2). \end{aligned}$$

This completes our proof for the lemma. □

The proof of the theorem involves some compact property of the set of random controls for compact non-expansive control problem, as we will show in Lemma 4.2.7: the "limit trajectory" in $\Delta(Y)$ (cf. Definition 4.2.6) of a sequence of random controls can be arbitrarily approximated by the trajectory induced by one random control. Here the random control that we are going to construct is defined as concatenations of certain sequence of random controls, namely *behavior control*. Formal definitions are as follows.

Definition 4.2.2. *For any two random controls $(\Omega_{\mathbf{u}}, \mathbf{u})$ and $(\Omega_{\mathbf{v}}, \mathbf{v})$ and a time $T > 0$. The **concatenation** of $\mathbf{u}$ and $\mathbf{v}$ at time T is defined as the random control $(\Omega_{\mathbf{u}} \times \Omega_{\mathbf{v}}, \mathbf{u} \oplus_T \mathbf{v})$ with: $\forall (y_0, (\omega_1, \omega_2), t) \in Y \times (\Omega_{\mathbf{u}} \times \Omega_{\mathbf{v}}) \times [0, +\infty)$,*

$$[\mathbf{u} \oplus_T \mathbf{v}]_{(\omega_1,\omega_2)}(y_0, t) = \mathbb{1}_{\{t<T\}} \mathbf{u}_{\omega_1}(y_0, t) + \mathbb{1}_{\{t \geq T\}} \mathbf{v}_{\omega_2}\Big(\mathbf{y}(y_0, \mathbf{u}_{\omega_1}(y_0), T), t - T\Big).$$

Definition 4.2.3. *Fix $0 = t_0 < t_1 \cdots < t_m < \cdots$ a partition of $[0, +\infty)$, and $(\Omega_m, \mathbf{u}^m)_{m \geq 1}$ a sequence of random controls. Let $\big(\otimes_{m'=1}^{m} \Omega_{m'}, \mathbf{u}^{[m]} \big)_{m \geq 1}$ be a sequence of random controls defined inductively as: $\mathbf{u}^{[1]} = \mathbf{u}^1$ and $\mathbf{u}^{[m+1]} = \mathbf{u}^{[m]} \oplus_{t_m} \mathbf{u}^{m+1}$ for any $m \geq 1$. The **behavior control***

$$(\Omega_{[\infty]}, \mathbf{u}^{[\infty]}) := \big(\otimes_{m \geq 1} \Omega_m, \mathbf{u}^1 \oplus_{t_1} \cdots \mathbf{u}^t \oplus_{t_m} \cdots \big)$$

is defined as the concatenations of $(\mathbf{u}^m)$ sequentially at points (t_m):

$$\forall (y_0, (\omega_m)_{m \geq 1}, t) \in Y \times \Omega_{[\infty]} \times [0, +\infty), \quad \mathbf{u}^{[\infty]}_{(\omega_m)_{m \geq 1}}(y_0, t) = \sum_{m \geq 1} \mathbb{1}_{\{t_{m-1} \leq t < t_m\}} \mathbf{u}^{[m]}_{\omega^m}(y_0, t),$$

where $\omega^m := (\omega_1, ..., \omega_m)$.

Remark 4.2.4. *The behavior control $(\Omega_{[\infty]}, \mathbf{u}^{[\infty]})$ is also a random control with the product Borel space $\Omega_{[\infty]} = \otimes_{m \geq 1} \Omega_m$, which, as a countable union, is still in $\mathcal{S}$.*

Remark 4.2.5. *On the other hand, from Kuhn's theorem (cf. Aumann [4], Sec. 5): for any random control, one can construct a behavior control such that the trajectories in $\Delta(Y)$ induced by them are the same. More precisely, we fix $(t_m)_{m \geq 0}$ a partition of $[0, +\infty)$, $(\Omega_{\mathbf{u}}, \mathbf{u})$ a random control and $y_0 \in Y$ an initial state. Then, there exists some behavior control $\big(\otimes_{m \geq 1} \Omega_m, \overline{\mathbf{u}} \big)$ as concatenations of some sequence of random controls $\big(\Omega_m, \overline{\mathbf{u}}^m\big)_{m \geq 1}$ at points (t_m) such that starting from y_0, the trajectories in $\Delta(Y)$ generated by both $\mathbf{u}$ and $\overline{\mathbf{u}}$ are the same, i.e. $\tilde{\mathbf{y}}(t, \mathbf{u}, y_0) = \tilde{\mathbf{y}}(t, \overline{\mathbf{u}}, y_0)$, $\forall t \geq 0, a.e.$*

Fix any y_0 in Y. $\big(\tilde{\mathbf{y}}(\cdot, \mathbf{u}^k, y_0)\big)_{k \geq 1}$ is the sequence of trajectories in $\Delta(Y)$ generated by a sequence of random controls $(\mathbf{u}^k)_{k \geq 1}$ with the same initial point y_0.

Definition 4.2.6. *A measurable mapping $t \mapsto \overline{\mathbf{y}}(t)$ defined from $\mathbb{R}_+$ to $(\Delta(Y), d_{KR})$ is a **limit trajectory** of the sequence $\big(\tilde{\mathbf{y}}(\cdot, \mathbf{u}^k, y_0)\big)_{k \geq 1}$ if there is a subsequence $\big(\tilde{\mathbf{y}}(\cdot, \mathbf{u}^{\psi(k)}, y_0)\big)_{k \geq 1}$ such that for any $m \geq 1$, $\tilde{\mathbf{y}}(\cdot, \mathbf{u}^{\psi(k)}, y_0)$ converges (for d_{KR}) to $\overline{\mathbf{y}}(\cdot)$ uniformly on $[0, m]$ as k tends to infinity.*

We first show that the limit trajectory exists for any sequence.

Lemma 4.2.7. *$\big(\tilde{\mathbf{y}}(\cdot, \mathbf{u}^k, y_0)\big)_{k \geq 1}$ has a limit trajectory in $\Delta(Y)$.*

Proof. Fix an $m \geq 0$, we look at the restriction of each $\tilde{\mathbf{y}}(\cdot, \mathbf{u}^k, y_0)$ on the compact interval $[m, m+1]$. Then the family $\{\tilde{\mathbf{y}}(\cdot, \mathbf{u}^k, y_0) : k \geq 1\}$ are continuous mappings from $[m, m+1]$ to the compact domain $(\Delta(Y), d_{KR})$. One can use Ascoli's theorem to deduce the existence of a uniform convergent subsequence $\big(\tilde{\mathbf{y}}(\cdot, \mathbf{u}^{\psi(k)}, y_0)\big)_{k\geq 1}$ on $[m, m+1]$. To obtain this, it is sufficient for us to prove that the family $\{\tilde{\mathbf{y}}(\cdot, \mathbf{u}^k, y_0) : k \geq 1\}$ (restricted on $[m, m+1]$) is equicontinuous.

We fix any $k \geq 1$ and $s, t \in [m, m+1]$. Then by the definition of d_{KR}, we deduce that

$$\begin{aligned} d_{KR}\Big(\tilde{\mathbf{y}}(t, \mathbf{u}^k, y_0), \tilde{\mathbf{y}}(s, \mathbf{u}^k, y_0)\Big) &\leq \int_\Omega \Big\|\mathbf{y}(t, \mathbf{u}^k_\omega(y_0), y_0) - \mathbf{y}(s, \mathbf{u}^k_\omega(y_0), y_0)\Big\| \mathrm{d}\lambda(\omega) \\ &\leq a(1 + \sup_{y\in Y} \|y\|)|t - s|, \end{aligned} \tag{4.2.4}$$

where we have used in the last inequality the fact that the trajectory $t \mapsto \mathbf{y}(t, \mathbf{u}^k_\omega(y_0), y_0)$ is absolutely continuous (cf. assumptions in (4.1.2)) . As $Y \subseteq \mathbb{R}^d$ is compact, (4.2.4) proves that the family $\{\tilde{\mathbf{y}}(\cdot, \mathbf{u}^k, y_0) : k \geq 1\}$ (restricted on $[m, m+1]$) is equicontinuous. By extracting subsequences for each m, we obtain the existence of a limit trajectory of $\big(\tilde{\mathbf{y}}(\cdot, \mathbf{u}^k, y_0)\big)_{k\geq 1}$. □

Lemma 4.2.8. *Let $\overline{\mathbf{y}}(\cdot) : t \mapsto \overline{\mathbf{y}}(t)$ be a limit trajectory of $\big(\tilde{\mathbf{y}}(\cdot, \mathbf{u}^k, y_0)\big)_k$. Then for any $\varepsilon > 0$, there is some behavior control $\mathbf{u}^*$ whose trajectory in $\Delta(Y)$ is ε-close to $\overline{\mathbf{y}}(\cdot)$ along time, i.e.,*

$$\forall \varepsilon > 0,\ \exists \mathbf{u}^* \in \tilde{\mathcal{U}},\ s.t.\ d_{KR}\Big(\tilde{\mathbf{y}}(t, \mathbf{u}^*, y_0), \overline{\mathbf{y}}(t)\Big) \leq \varepsilon, \quad \forall t \geq 0.$$

Proof: The idea is to construct a behavior control $\mathbf{u}^*$ by consecutive intervals, such that on each of them, $\mathbf{u}^*$ follows one random control in the family $\{\mathbf{u}^k\}$ whose trajectory is close to the limit $\overline{\mathbf{y}}$. The proof relies on the nonexpansive property established in Lemma 4.2.1, which ensures that by iteration, the trajectory generated by $\mathbf{u}^*$ is close to $\overline{\mathbf{y}}$ on the whole $\mathbb{R}_+$.

Let $\varepsilon > 0$ be fixed. The behavior control $\mathbf{u}^*$ will be constructed as concatenations of a sequence of random controls $(\hat{\mathbf{u}}^{K_m})$ (to be specified later on) at points $\{1, 2, 3, ...\}$.

By definition, $t \mapsto \overline{\mathbf{y}}(t)$ is a limit trajectory of $\big(\tilde{\mathbf{y}}(\cdot, \mathbf{u}^k, y_0)\big)_k$ in $\Delta(Y)$ for d_{KR}, so for each $m \geq 0$, there exists some $K_{m+1} > 0$ such that:

$$d_{KR}\Big(\tilde{\mathbf{y}}(t, \mathbf{u}^{K_{m+1}}, y_0), \overline{\mathbf{y}}(t)\Big) \leq \varepsilon^{m+1},\ \forall t \in [m, m+1]. \tag{4.2.5}$$

Following Remark 4.2.5, we could have assumed that each $\mathbf{u}^{K_{m+1}}$ is a behavior control, and let $\overline{\mathbf{u}}^{K_{m+1}} : Y \times \Omega \times [0, +\infty) \to U$ be the component of $\mathbf{u}^{K_{m+1}}$ on interval $[m, m+1]$.

In order to define the behavior control $\mathbf{u}^*$, it is sufficient to construct by induction a sequence of random controls $(\hat{\mathbf{u}}^{K_m})_{m\geq 1}$ such that for all $m \geq 1$,

$$d_{KR}\Big(\tilde{\mathbf{y}}(t, \mathbf{u}^{[m]}, y_0), \overline{\mathbf{y}}(t)\Big) \leq 2\sum_{\ell=1}^{m} \varepsilon^\ell, \quad \forall t \in [0, m],$$

where

$$\mathbf{u}^{[1]} = \hat{\mathbf{u}}^{K_1} \quad \text{and} \quad \mathbf{u}^{[m]} = \hat{\mathbf{u}}^{K_1} \oplus_1 \cdots \oplus_{m-1} \hat{\mathbf{u}}^{K_m}, \quad m \geq 2.$$

For $m = 1$, let $\hat{\mathbf{u}}^{K_1} = \overline{\mathbf{u}}^{K_1}$, then from (4.2.5), $d_{KR}\Big(\tilde{\mathbf{y}}(t_1, \mathbf{u}^{[1]}, y_0), \overline{\mathbf{y}}(t)\Big) \leq \varepsilon,\ \forall t \in [0, 1]$. This initializes our induction.

Assume that $\hat{\mathbf{u}}^{K_1}, ..., \hat{\mathbf{u}}^{K_m}$ are defined and let $\mathbf{u}^{[m]} = \hat{\mathbf{u}}^{K_1} \oplus_1 \cdots \oplus_{m-1} \hat{\mathbf{u}}^{K_m}$ ($\mathbf{u}^{[1]} = \hat{\mathbf{u}}^{K_1}$ for $m = 1$) satisfy:

$$d_{KR}\Big(\tilde{\mathbf{y}}(t, \mathbf{u}^{[m]}, y_0), \overline{\mathbf{y}}(t)\Big) \leq 2\sum_{l=1}^{m} \varepsilon^l, \ \forall t \in [0, m]. \tag{4.2.6}$$

Next we construct the control $\hat{\mathbf{u}}^{K_{m+1}}$ thus complete the definition of $\mathbf{u}^{[m+1]}$ on $[m, m+1)$. To do this, we consider the two distributions $\tilde{\mathbf{y}}(m, \mathbf{u}^{[m]}, y_0)$ and $\tilde{\mathbf{y}}(m, \mathbf{u}^{K_{m+1}}, y_0)$. Take $t = m$ in (4.2.5) and in (4.2.6), we use the triangle inequality obtain a bound on the distance between them:

$$d_{KR}\Big(\tilde{\mathbf{y}}(m, \mathbf{u}^{[m]}, y_0), \tilde{\mathbf{y}}(m, \mathbf{u}^{K_{m+1}}, y_0)\Big) \leq \quad 2\sum_{l=1}^{m} \varepsilon^l + \varepsilon^{m+1}. \tag{4.2.7}$$

We consider then the random control $\overline{\mathbf{u}}^{K_{m+1}}$ on the starting distribution $\tilde{\mathbf{y}}(m, \mathbf{u}^{K_{m+1}}, y_0)$, and apply Lemma 4.2.1 to deduce the existence of some random control $\hat{\mathbf{u}}^{K_{m+1}}$ on the starting distribution $\tilde{\mathbf{y}}(m, \mathbf{u}^{[m]}, y_0)$ such that:

$$\begin{aligned} & d_{KR}\Big(\tilde{\mathbf{y}}(\Delta, \hat{\mathbf{u}}^{K_{m+1}}, \tilde{\mathbf{y}}(m, \mathbf{u}^{[m]}, y_0)), \ \tilde{\mathbf{y}}(\Delta, \bar{\mathbf{u}}^{K_{m+1}}, \tilde{\mathbf{y}}(m, \mathbf{u}^{K_{m+1}}, y_0)\Big) \\ \leq\ & d_{KR}\Big(\tilde{\mathbf{y}}(m, \mathbf{u}^{[m]}, y_0), \ \tilde{\mathbf{y}}(m, \mathbf{u}^{K_{m+1}}, y_0)\Big) \\ \leq\ & 2\sum_{l=1}^{m} \varepsilon^l + \varepsilon^{m+1}, \ \forall \Delta \in [0, 1]. \end{aligned} \tag{4.2.8}$$

By definition of $\overline{\mathbf{u}}^{K_{m+1}}$, we have that for all $\Delta \in [0, 1]$,

$$\tilde{\mathbf{y}}(\Delta, \overline{\mathbf{u}}^{K_{m+1}}, \tilde{\mathbf{y}}(m, \mathbf{u}^{K_{m+1}}, y_0)) = \tilde{\mathbf{y}}(\Delta + m, \mathbf{u}^{K_{m+1}}, y_0). \tag{4.2.9}$$

Define now $\hat{\mathbf{u}}^{[m+1]} = \hat{\mathbf{u}}^{[m]} \oplus_m \hat{\mathbf{u}}^{K_m+1}$. This gives us:

$$\tilde{\mathbf{y}}(\Delta, \hat{\mathbf{u}}^{K_{m+1}}, \tilde{\mathbf{y}}(m, \mathbf{u}^{[m]}, y_0)) = \tilde{\mathbf{y}}(\Delta + m, \mathbf{u}^{[m+1]}, y_0). \tag{4.2.10}$$

We substitute (4.2.9) and (4.2.10) back into (4.2.8) and use the change the variable to obtain that:

$$\forall t \in [m, m+1], \quad d_{KR}\Big(\tilde{\mathbf{y}}(t, \mathbf{u}^{[m+1]}, y_0), \ \tilde{\mathbf{y}}(t, \mathbf{u}^{K_{m+1}}, y_0)\Big) \leq 2\sum_{l=1}^{m} \varepsilon^l + \varepsilon^{m+1}.$$

Finally, with the help of definition of K_{m+1} in (4.2.5), we obtain that: for any $t \in [m, m+1]$,

$$d_{KR}\Big(\tilde{\mathbf{y}}(t, \mathbf{u}^{[m+1]}, y_0), \ \bar{\mathbf{y}}(t)\Big) \leq 2\sum_{l=1}^{m} \varepsilon^l + \varepsilon^{m+1} + \varepsilon^{m+1} = 2\sum_{l=1}^{m+1} \varepsilon^l.$$

This finishes the inductive definition of the sequence $\left(\hat{\mathbf{u}}^{K_m}\right)_{m \geq 1}$.

To conclude, we set the behavior control

$$\mathbf{u}^* = \hat{\mathbf{u}}^{K_1} \oplus_1 \cdots \hat{\mathbf{u}}^{K_m} \oplus_m \cdots,$$

as concatenations of $\left(\hat{\mathbf{u}}^{K_m}\right)_{m \geq 1}$ at points $\{m \geq 1\}$, and by our inductive construction:

$$\forall t \geq 0, \ d_{KR}\Big(\tilde{\mathbf{y}}(t, \mathbf{u}^*, y_0), \ \bar{\mathbf{y}}(t)\Big) \leq 2\sum_{\ell=1}^{\infty} \varepsilon^\ell = \frac{2\varepsilon}{1-\varepsilon} \leq 3\varepsilon,$$

as long as $\varepsilon \in (0, \frac{1}{3}]$. This completes our proof for the lemma by considering $\varepsilon' = \varepsilon/3$. $\square$

4.3 Proof of Theorem 4.1.7

This section is devoted for the proof of Theorem 4.1.7, which is divided into three parts.

Part A aims at establishing certain optimality properties for a sequence of controls (Lemma 4.3.1); In **Part B**, we use the compact nonexpansive property (Lemma 4.2.8) to obtain a "limit" control of the above sequence ensuring that the average cost on each (consecutive) block of fixed length is no more than V^* (Eq. (4.3.8)); **Part C** concludes the proof through a comparison of the (normalized) θ-evaluated payoff to the average cost by blocks.

Part A. For any $t \geq 0$, $S > 0$, $y_0 \in Y$ and $\mathbf{u} \in \tilde{\mathcal{U}}$, denote

$$\gamma_{t,S}(y_0, \mathbf{u}) = \frac{1}{S} \int_{[t,t+S]} g(\tilde{\mathbf{y}}(s, \mathbf{u}, y_0)) \mathrm{d}s, \ \forall y_0 \in Y.$$

For $T \geq 0$, we put

$$\varphi_{T,S}(y_0) = \inf_{\mathbf{u} \in \tilde{\mathcal{U}}} \sup_{\mu \in \Delta([0,T])} \int_{[0,T]} \gamma_{t,S}(y_0, \mathbf{u}) \mathrm{d}\mu(t).$$

Fix any T, S, y_0. We first prove a minmax result for $\varphi_{T,S}(y_0)$. We denote for each $s \geq 0$:

$$\beta_s(\mu, S) = \frac{1}{S} \int_{\max\{0,\ s-S\}}^{\min\{T,s\}} \mathrm{d}\mu(t) = \frac{1}{S} \mu\Big([0,T] \cap [s-S, s]\Big).$$

Then from the definition of $\gamma_{t,S}(y_0, \mathbf{u})$, we obtain that

$$\begin{aligned} \int_{[0,T]} \gamma_{t,S}(y_0, \mathbf{u}) \mathrm{d}\mu(t) &= \int_{[0,T]} \left(\frac{1}{S} \int_{[t,t+S]} g(\tilde{\mathbf{y}}(s, \mathbf{u}, y_0)) ds \right) \mathrm{d}\mu(t) \\ (\text{"Fubini's theorem"}) &= \int_{[0,T+S]} \beta_s(\mu, S) g\left(\tilde{\mathbf{y}}(s, \mathbf{u}, y_0)\right) \mathrm{d}s. \end{aligned}$$

Note that for each fixed $S > 0$ and $\mu \in \Delta([0,T])$, the mapping $t \mapsto \beta_t(\mu, S)$ defines a density function of some evaluation over $\mathbb{R}_+$, which we denote by $\varsigma(\mu, S)$. This enables us to write

$$\varphi_{T,S}(y_0) = \inf_{\mathbf{u} \in \tilde{\mathcal{U}}} \sup_{\mu \in \Delta([0,T])} \gamma_{\varsigma(\mu,S)}(y_0, \mathbf{u}).$$

Next, we use Sion's minmax theorem (cf. Appendix A.3 in Sorin [59]) to show that the operators "inf" and "sup" of the above equation commute. Indeed, $\tilde{\mathcal{U}}$ is convex; $\Delta([0,T])$ is convex and weak-* compact and the payoff function $(\mu, \mathbf{u}) \mapsto \gamma_{\varsigma(\mu,S)}(y_0, \mathbf{u})$ is affine in both μ and $\mathbf{u}$; moreover the function $\gamma_{t,S}(y_0, \mathbf{u})$ is continuous in t for given $\mathbf{u}$ (g is continuous in y and each trajectory is absolutely continuous), and so is $\mathbf{u} \mapsto \gamma_{\varsigma(\mu,S)}(y_0, \mathbf{u})$. Then we obtain:

$$\varphi_{T,S}(y_0) = \sup_{\mu \in \Delta([0,T])} \inf_{\mathbf{u} \in \tilde{\mathcal{U}}} \gamma_{\varsigma(\mu,S)}(y_0, \mathbf{u}) = \sup_{\mu \in \Delta([0,T])} V_{\varsigma(\mu,S)}(y_0). \tag{4.3.1}$$

Lemma 4.3.1. *For any $\varepsilon > 0$, there is some $S_0 > 0$ such that*

$$\forall T \geq 0, \exists \mathbf{u}^T \in \tilde{\mathcal{U}} : \forall y_0 \in Y, \ \gamma_{t,S_0}(y_0, \mathbf{u}^T) \leq V^*(y_0) + \varepsilon, \forall t \leq T.$$

Proof for Lemma 4.3.1: For any $\theta \in \Delta(\mathbb{R}_+)$ and $s \geq 0$, denote $I_s(\theta) = \int_{[0,+\infty)} \big|f_\theta(t+s) - f_\theta(t)\big|\mathrm{d}t$. Following Li *et al.* [30] (cf. Lemma 3.3), for any evaluation θ that is absolutely continuous w.r.t. the Lebesgue measure thus admitting a density function $t \mapsto f_\theta(t)$, we have that:

$$\frac{I_s(\theta)}{2} \leq TV_s(\theta) \leq I_s(\theta),\ \forall s \geq 0.$$

For any $S > 0$, $T \geq 0$ and $\mu \in \big([0,T]\big)$, we apply the above expression for $\varsigma(\mu, S)$ so as to obtain the following bound: $\forall s \in [0, S]$,

$$\begin{aligned} I_s(\varsigma(\mu,S)) =& \frac{1}{S}\int_{[0,T+S]} \Big[\mu\big([t-S, t-S+s] \cap [0,T]\big) + \mu\big([t, t+s] \cap [0,T]\big)\Big]\mathrm{d}t \\ =& \int_{[0,T+S]} \int_{[t-S,t-S+s]\cap[0,T]} \mathrm{d}\mu(s')\mathrm{d}t + \int_{[0,T+S]} \int_{[t,t+s]\cap[0,T]} \mathrm{d}\mu(s')\mathrm{d}t \\ (\text{'' Fubini's theorem''}) =& \frac{s}{S} \cdot \Big(\mu\big([-S, T+s] \cap [0,T]\big) + \mu\big([0, s+T+S] \cap [0,T]\big)\Big) \\ \leq& \frac{2s}{S}. \end{aligned} \tag{4.3.2}$$

According to Proposition 4.1.4, the general limit value exists and is equal to V^*, i.e. for any $\varepsilon > 0$, we take $\eta > 0$ and $S' > 0$ such that:

$$\forall \theta \in \Delta(\mathbb{R}_+),\quad \left(\sup_{0 \leq s \leq S'} TV(\theta) \leq \eta' \Longrightarrow (\forall y_0 \in Y,\ |V_\theta(y_0) - V^*(y_0)| \leq \varepsilon/2)\right). \tag{4.3.3}$$

Take $S_0 = \max\{\frac{2S'}{\eta'}, S'\}$. From (4.3.2), we obtain that: $\forall s \in [0, S'], S \geq S_0, T > 0, \mu \in \Delta\big([0,T]\big)$,

$$TV_s\left(\varsigma(\mu,S)\right) \leq \frac{2s}{S} \leq \frac{2S'}{S_0} \leq \eta',\ \text{thus by (4.3.3)}\ :\ \forall y_0 \in Y,\ |V_{\varsigma(\mu,S)}(y_0) - V^*(y_0)| \leq \varepsilon/2.$$

Next, from Eq. (4.3.1), $\varphi_{T,S}(y_0) = \sup_{\mu \in \Delta([0,T])} V_{\varsigma(\mu,S)}(y_0)$, we deduce that

$$\forall \varepsilon > 0, \exists S_0 > 0 : \forall S \geq S_0, \forall T \geq 0, \forall y_0 \in Y,\ |\varphi_{T,S}(y_0) - V^*(y_0)| \leq \varepsilon/2. \tag{4.3.4}$$

Finally, according to the definition $\varphi_{T,S}(y_0) = \inf_{\mathbf{u}} \sup_\mu \gamma_{\varsigma(\mu,S)}(y_0, \mathbf{u})$, we have

$$\forall T > 0, \exists \mathbf{u}^T \in \tilde{\mathcal{U}} : \forall y_0 \in Y,\ \gamma_{t,S_0}(y_0, \mathbf{u}^T) \leq \varphi_{T,S_0}(y_0) + \varepsilon/2, \forall t \in [0,T].$$

Together with (4.3.4), one obtains

$$\forall \varepsilon > 0, \exists S_0 > 0 : \forall T \geq 0, \exists \mathbf{u}^T \in \tilde{\mathcal{U}} : \forall y_0 \in Y,\ \gamma_{t,S_0}(y_0, \mathbf{u}^T) \leq V^*(y_0) + \varepsilon, \forall t \leq T. \tag{4.3.5}$$

The proof of the lemma is then complete. □

Part B. Fix now any $\varepsilon > 0$, and consider $S_0 > 0$ and the random control $\mathbf{u}^T \in \tilde{\mathcal{U}}$ for any $T > 0$ given as in Lemma 4.3.1. We take an increasing sequence $(T_k)_{k \geq 1}$ in $\mathbb{R}_+$ and fix any $y_0 \in Y$. For each $k \geq 1$, $t \mapsto \tilde{\mathbf{y}}(t, \mathbf{u}^{T_k}, y_0)$ is the trajectory of $\mathbf{u}^{T_k}$ in $\Delta(Y)$. Thus from (4.3.5), we obtain:

$$\gamma_{t,S_0}(y_0, \mathbf{u}^{T_k}) = \frac{1}{S_0}\int_{[t,t+S_0]} g\left(\tilde{\mathbf{y}}(t, \mathbf{u}^{T_k}, y_0)\right)\mathrm{d}s \leq V^*(y_0) + \varepsilon \text{ for all } t \leq T_k. \tag{4.3.6}$$

Let $\bar{\mathbf{y}}(\cdot) : t \mapsto \bar{\mathbf{y}}(t)$ be a limit trajectory of the sequence $\big(\tilde{\mathbf{y}}(\cdot, \mathbf{u}^{T_k}, y_0)\big)_{k\geq 1}$ i.e. there is a subsequence $\psi(k)$ such that $\tilde{\mathbf{y}}(\cdot, \mathbf{u}^{T_{\psi(k)}}, y_0)$ converges uniformly to $\bar{\mathbf{y}}(\cdot)$ on each $[m, m+1]$. Since g is continuous on the compact invariant set Y, and $\Delta(Y)$ is weak-* compact for the topology induced by the distance d_{KR} (cf. Theorem 6.9 in Villani [64]), we let k tend to infinity (along the subsequence $\psi(k)$) in (4.3.6) to get

$$\frac{1}{S_0}\int_{[t,t+S_0]} g\big(\bar{\mathbf{y}}(s)\big)\mathrm{d}s \leq V^*(y_0) + \varepsilon,\ \forall t \geq 0. \tag{4.3.7}$$

Now we apply Lemma 4.2.8 for the sequence $\big(\tilde{\mathbf{y}}(\cdot, \mathbf{u}^{T_k}, y_0)\big)_{k\geq 1}$ and its limit trajectory $\bar{\mathbf{y}}(\cdot)$ to obtain the existence of some behavior control $\mathbf{u}^*$ such that:

$$d_{KR}\big(\tilde{\mathbf{y}}(t, \mathbf{u}^*, y_0), \bar{\mathbf{y}}(t)\big) \leq \varepsilon,\ \forall t \geq 0.$$

Together with (4.3.7), we obtain:

$$\gamma_{t,S_0}(y_0, \mathbf{u}^*) = \frac{1}{S_0}\int_{[t,t+S_0]} g\big(\tilde{\mathbf{y}}(s, \mathbf{u}^*, y_0)\big)\mathrm{d}s \leq V^*(y_0) + 2\varepsilon,\ \forall t \geq 0. \tag{4.3.8}$$

Part C. The computation below is analog to the proof of Proposition 6.1 in Li *et al.* [30].

Let $\theta \in \Delta(\mathbb{R}_+)$ be any evaluation. We integrate (4.3.8) over $t \geq 0$ w.r.t. θ, to obtain:

$$\begin{aligned} V^*(y_0) + 2\varepsilon \geq \int_{[0,+\infty)} \gamma_{t,S_0}(y_0, \mathbf{u}^*)\mathrm{d}\theta(t) &= \int_{[0,+\infty)} \left(\frac{1}{S_0}\int_{[t,t+S_0]} g\left(\tilde{\mathbf{y}}(s, \mathbf{u}^*, y_0)\right)\mathrm{d}s\right)\mathrm{d}\theta(t) \\ (''\text{Fubini's theorem}'') &= \int_{[0,+\infty)} \beta_s(\theta, S_0) g\left(\tilde{\mathbf{y}}(s, \mathbf{u}^*, y_0)\right)\mathrm{d}s \\ &= \gamma_{\varsigma(\theta,S_0)}(y_0, \mathbf{u}^*), \end{aligned} \tag{4.3.9}$$

where $\beta_s(\theta, S_0) = \frac{1}{S_0}\int_{\max\{0,s-S_0\}}^{s} \mathrm{d}\theta(t),\ \forall s \geq 0$, and $\varsigma(\theta, S_0)$ is the evaluation with $s \mapsto \beta_s(\theta, S_0)$ its density function.

Next, we show that

$$|\gamma_\theta(y_0, \mathbf{u}^*) - \gamma_{\varsigma(\theta,S_0)}(y_0, \mathbf{u}^*)| \leq \sup_{Q\in\mathcal{B}(\mathbb{R}_+)} |\theta(Q) - \varsigma(\theta, S_0)(Q)| \leq 2TV_{S_0}(\theta). \tag{4.3.10}$$

Indeed, the first inequality follows from Hahn's decomposition theorem applied to the sign measure "$\theta - \varsigma(\theta, S_0)$" (cf. Lemma 3.7 in Li *et al.* [30]). Let Q be any Borel set on $\mathbb{R}_+$. We write $\beta_s(\theta, S_0) = \frac{1}{S_0}\int_{s-S_0}^{s} \mathrm{d}\theta(t)$ for all $s \geq 0$ by considering θ as a probability measure over $[-S_0, 0) \cup \mathbb{R}_+$ null on $[-S_0, 0)$. We have

$$\begin{aligned} \varsigma(\theta, S_0)(Q) = \int_{s\in Q} \beta_s(\theta, S_0)\mathrm{d}s &= \frac{1}{S_0}\int_{s\in Q}\left(\int_{t\in[s-S_0,s]} \mathrm{d}\theta(t)\right)\mathrm{d}s \\ (''\text{Fubini's theorem}'') &= \int_{t\in Q-S_0}\left(\frac{1}{S_0}\int_{s\in[t,t+S_0]} \mathrm{d}s\right)\mathrm{d}\theta(t) \\ &= \theta(Q - S_0). \end{aligned}$$

Thus we have $|\theta(Q) - \varsigma(\theta, S_0)(Q)| = |\theta(Q) - \theta(Q - S_0)| \leq \theta([0, S_0)) + TV_{S_0}(\theta) \leq 2TV_{S_0}(\theta)$. This proves (4.3.10) by taking the supremum over $Q \in \mathcal{B}(\mathbb{R}_+)$.

Finally, we substitute (4.3.9) into (4.3.10), to obtain:

$$\gamma_\theta(y_0, \mathbf{u}^*) \leq V^*(y_0) + 2TV_{S_0}(\theta) \leq V^*(y_0) + 3\varepsilon,$$

for all $\theta \in \Delta(\mathbb{R}^+)$ with $\sup_{0\leq s\leq S_0} TV_s(\theta) \leq \varepsilon$.

To conclude, we have obtained that: $\forall \varepsilon, \exists S_0 > 0, \exists \mathbf{u}^* \in \tilde{U}$,

$$\forall \theta \in \Delta(\mathbb{R}^+),\ \Big(\sup_{0\leq s\leq S_0} TV_s(\theta) \leq \varepsilon \Longrightarrow \gamma_\theta(y_0, \mathbf{u}^*) \leq V^*(y_0) + 3\varepsilon,\ \forall y_0 \in Y \Big).$$

As $\varepsilon > 0$ is arbitrary, this proves Theorem 4.1.7 by taking "$\eta = \varepsilon$" and "$S = S_0$".

Acknowledgment This article was written during the course of my PhD thesis. I wish to thank my supervisor Sylvain Sorin for numerous helpful comments. I thank also Marc Quincampoix, Fabien Gensbittel and Marco Mazzola for useful discussions.

Deuxième partie

Jeux répétés

Chapter 5

Big match généralisé à information incompète d'un côté

Résumé Nous étudions une sous-classe de jeux absorbants comme une généralisation du Big match à la Blackwell et Ferguson [13]. Pour le Big match généralisé à information incomplète d'un côté, nous prouvons l'existence de la valeur asymptotique, du $Maxmin$ et du $Minmax$, et que la valeur asymptotique est égale au $Maxmin$.

Nos résultats généralisent ceux de Sorin [55]. Le résultat de l'existence de la valeur asymptotique n'est pas une conséquence de Rosenberg [48] pour deux raisons: d'abord, nous considérons le flux du paiement évalué par des mesures de probabilité générales sur des nombres entiers positifs et nous prouvons que la fonction valeur associée converge quand le poids maximal de la mesure sur chaque étape tend vers zéro; deuxièmement, nous ne supposons pas que la probabilité d'absorption soit indépendante de l'état comme dans Rosenberg [48].

Mots-clés Jeux stochastiques, jeux absorbants à information incomplète d'un côté, Big match, valeur asymptotique, valeur uniforme, Maxmin, Minmax

---

Ce chapitre est issu de l'article *Generalized Big match with one-sided incomplete information.*

Generalized Big match with one-sided incomplete information

Abstract We study one subclass of absorbing games as a generalization of Big match due to Blackwell and Ferguson [13]. For "generalized Big match" with one-sided incomplete information, we prove the existence of the asymptotic value, $Maxmin$ and $Minmax$, and that the asymptotic value is equal to $Maxmin$.

Our results generalize that of Sorin [55]. The existence result of the asymptotic value is not a consequence of Rosenberg [48] for two reasons: first, we consider the payoff stream evaluated by general probability measures on the positive integers and prove that the associated value function converges as the maximal weight of the measure on stages tends to zero; second, we do not assume the absorbing probability to be state-independent as in Rosenberg [48].

Keywords Stochastic games, absorbing games with incomplete information on one side, Big match, asymptotic value, uniform value, Maxmin, Minmax

5.1 Introduction

Absorbing games are stochastic games where all states but one are absorbing. One example is the *Big match* (henceforth BM) introduced by Gillette [22], represented by the following 2×2 matrix:

	L	R
T	1^*	0^*
B	0	1

.

At each stage, player 1 chooses an action in $\{Top, Bottom\}$ and player 2 chooses an action in $\{Left, Right\}$. The state remains non-absorbing as long as player 1 chooses $Bottom$; playing the action Top induces an absorption, which is either on 1^* with an absorbing payoff 1 or on 0^* with an absorbing payoff 0.

Blackwell and Ferguson [13] solved BM by proving the existence of the uniform value (cf. Def. 5.2.2). This result is extended by Kohlberg [26] to absorbing games. For stochastic games, Bewley and Kohlberg [10] proved the existence of the asymptotic value (the convergence of the n-stage value v_n and of the λ-discounted value v_λ, and both to the same limit, cf. Def. 5.2.1), and later on, Mertens and Neyman [34] proved the existence of the uniform value.

In a model of stochastic games with one-sided incomplete information, one game among a family is chosen according to a given probability distribution, and the selected game is communicated to player 1 only. For these games, v_n or v_λ satisfies the Shapley equation (cf. Eq. 5.3.1), which defines an auxiliary stochastic game with player 2's *posterior* belief entering the state variable. This reduction involves an infinite auxiliary state space, hence arguments in [10] or [34] for finite stochastic games do not apply.

Sorin [55] studied BM with one-sided incomplete information (type I), and proved the existence of the asymptotic value, *Maxmin* and *Minmax*. Moreover $Maxmin \neq Minmax$ thus uniform value does not exist, and $Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$.

It has been conjectured by Mertens [33] (cf. Coulomb [17]) that in a general model of repeated games where player 1 is always more informed than player 2, both $Maxmin$ and the asymptotic value exist and $Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$. Since Sorin [55], several positive results of type "$Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to\infty} v_\lambda$" have been established: Sorin [56] for "BM" with one-sided incomplete information (type II), Rosenberg *et al.* [50], Renault [45] and Gensbittel *et al.* [21] for an informed controller, Rosenberg and Vieille [52] and Li and Venel [31] for recursive games with one-sided information information.

Recently, a counterexample is constructed by Ziliotto [68] to disprove this conjecture for a general model. It becomes now a challenging problem to identify the subclass of repeated games for this result to hold true.

We aim at extending the results in Sorin [55] into the model of *generalized Big match (henceforth GBM) with one-sided incomplete information*. GBM is a generalized model of BM in the sense that, whenever Top is played, the absorbing probability is strictly positive yet not necessarily one. Nevertheless, after playing the "exceptional" move Top (which enforces the absorption whatever is played by player 2) for a bounded number of times, the state is absorbed with a probability almost one. We then introduce a *counting number* for these moves, which either is defined as an auxiliary state variable in the asymptotic analysis, or helps establish an induction in uniform analysis.

Our main idea of the proof, analog to Sorin [55], is to construct an auxiliary "limit game" played $[0, 1]$, which has a value and it characterizes both the asymptotic value and $Maxmin$ of the original repeated game.

In the asymptotic analysis, we approximate the repeated game by the "limit game" on $[0, 1]$ such that players can mimic the optimal plays in the auxiliary game, which will give them asymptotic optimal strategies in the original game.

- Being different from Sorin [55], we are not defining the "limit game" directly on $[0, 1]$, rather, we consider a sequence of its discretizations with vanishing mesh.
- Our method is similar to Sorin [58], which studies the asymptotic value of repeated games with symmetric incomplete information where at at each stage the signal is either non-revealing or completely revealing (*repeated games without a recursive structure*). There, the completely revealing of the state corresponds to an absorption and the "exceptional" moves are those action pairs inducing a strictly positive probability of revealing.

Our study of $Maxmin$ relies on three aspects: first, the results in Sorin [55] for BM with incomplete information (where an absorption leads to an absorbing payoff); second, the generalization of the results in Sorin [55] to BM with incomplete information and with "absorbing state", i.e. playing the action Top may not lead to an absorbing payoff but will lead to a different state (i.e., revealing the information and increasing the probability of absorption); third, induction analysis on the finite number of playing Top.

For each play of the move Top, we specify some auxiliary absorbing payoff for an "absorbing state" (which contains the *posterior* belief on the state of the uninformed player and a number of times for Top to be played in the remaining game), to be the amount that can be guaranteed by player 1 or be defended by player 2 in the remaining infinitely repeated game in the new auxiliary "absorbing state". Our induction analysis starts from

the last move of *Top*, which, with an expected probability almost one, corresponds to a BM with incomplete information. So the results in Sorin [55] apply.

The inductive analysis is similar to Neyman and Sorin [39], which studies the equilibrium in repeated games with symmetric incomplete information and with random symmetric signals. There, the auxiliary state variable is players' common *posterior* belief on the state, and the induction is on the number of significant "jumps" of the *posterior* martingale. Such number is bounded due to the martingale convergence theorem, and the initialization of their induction relies on the existence result of equilibrium in absorbing games established in Vrieze and Thuijsman [65].

Let us mention also Rosenberg [48] which proves the existence of asymptotic value ($\lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$) for absorbing games with one-sided incomplete information, under the assumption that the transition probability be independent across different games. The analysis in Rosenberg [48] is through the *operator approach* (see also Rosenberg and Sorin [51]), which studies the asymptotic behavior of v_n or v_λ *via* Shapley equation. Unlike our construction, this approach does not provide explicitly asymptotic optimal strategies.

Our asymptotic result is not implied by Rosenberg [48] in the following sense: first, we consider the payoff streams evaluated by general probability measures and prove the convergence of the value function as the maximal weight of the measure on each stage tends to zero (the *sup-asymptotic value*, see Def. 5.2.1); second, in our model the transition probability is state-dependent.

The existence of sup-asymptotic value seems specific for absorbing games. On one hand, for absorbing games with complete information, Cardaliaguet *et al.* [16] proved the existence of sup-asymptotic value (see also Ziliotto [69] for the generalization of the result to absorbing games with infinite actions). On the other hand, this is not true for general stochstic games. We refer to Ziliotto [69] for a systematic study.

The organization of our paper is as follows. Section 2 describes the model and our main results. Section 3 concerns the asymptotic analysis. We prove the existence of $Maxmin$ and its equality to the asymptotic value in Section 4. The study of $Minmax$ is in Section 5.

5.2 The model and the results

The game An absorbing game with one-sided incomplete information Γ_∞ is described as follows. The state of world $k \in K$ is chosen by nature according to some probability distribution $p \in \Delta(K)$, and is communicated to player 1 only. Ω is the set of stochastic states with only one state ω^o that is non-absorbing. $\omega_1 = \omega^o$ and at each stage $t \geq 1$, after observing the previous moves of both players, simultaneously, player 1 chooses an action $i_t \in I$ and player 2 chooses an action $j_t \in J$. $g^k(\omega_t, i_t, j_t)$ is the stage payoff, and $q^k(\cdot|\omega_t, i_t, j_t) \in \Delta(\Omega)$ is the transition probability function for ω_{t+1} satisfying: $q^k(\omega^*|\omega^*, i, j) = 1$ for all $(\omega^*, i, j) \in \Omega/\{\omega^o\} \times I \times J$.

We consider the following specific model of Γ_∞, namely the *generalized Big match (GBM) with one-sided incomplete information*:

- $I = \{Top, Bottom\}$ ($\{T, B\}$ for short);
- $\Omega/\{\omega^o\} = \{\omega^{j*}|j \in J\}$;
- Let $(\chi^k(j))_{k,j}$ in $(0,1]^{K\times J}$. $q^k(\cdot)$ satisfies: $\forall j \in J$,
  - $q^k(\omega^o|\omega^o, B, j) = 1$,
  - $q^k(\omega^{j*}|\omega^o, T, j) = \chi^k(j)$ and $q^k(\omega^o|\omega^o, T, j) = 1 - \chi^k(j)$;

– Let $\left(a^k(j)\right)_{k,j}$ and $\left(a^{k*}(j)\right)_{k,j}$ in $\in \mathbb{R}^{K\times J}$. $g^k(\cdot)$ satisfies: $\forall (i,j,j') \in \{T,B\}\times J^2$

$$g^k(\omega^o, i, j) = a^k(j) \text{ and } g^k(\omega^{j'*}, i, j) = a^{k*}(j').$$

Γ_∞ is represented by the following matrix:

$$\forall k \in K, \quad \Gamma^k = \begin{array}{|c|c|c|c|} \hline & ... & j \in J & ... \\ \hline T & ... & \chi^k(j), a^{k*}(j) & ... \\ \hline B & ... & a^k(j) & ... \\ \hline \end{array}.$$

When $\chi^k(j) = 1, \forall (j,k) \in J\times K, \Gamma_\infty$ corresponds to the model of Big match with one-sided incomplete information (Type I) studied in Sorin [55].

We set $H_t = (\Omega \times I \times J)^{t-1}$ for each $t \geq 1$ and $H_\infty = (\Omega \times I \times J)^{\mathbb{N}}$. Any element $h = (k, \omega_1, i_1, j_1, ..., \omega_t, i_t, j_t, ...,)$ in $K \times H_\infty$ is called a *play* of the game. Each $h_t \in H_t$ is identified with a cylinder set of H_∞, and we denote by $\mathcal{H}_t^2$ the σ-algebra induced by H_t over H_∞, and by $\mathcal{H}_t^1$ the σ-algebra induced by $K \times H_t$ over $K \times H_\infty$. $\mathcal{H}_t^\ell$ is the information available for player ℓ at stage t, $\ell = 1, 2$. We endow $K \times H_\infty$ with the product σ-filed $\mathcal{H}_\infty = \sigma(\mathcal{H}_t^1, t \geq 1)$.

We assume perfect recall for both players , therefore Kuhn's theorem applies and it is without loss of generality for us to consider only behavior strategies. A behavior strategy $\sigma = (\sigma_t)_{t\geq 1}$ for player 1 is a sequence of measurable mappings with $\sigma_t : (K \times H_t, \mathcal{H}_t^1) \to \Delta(I)$, $\forall t \geq 1$. Similarly, a behavior strategy $\tau = (\tau_t)_{t\geq 1}$ for player 2 is a sequence of measurable mappings with $\tau_t : (H_t, \mathcal{H}_t^2) \to \Delta(J)$, $\forall t \geq 1$. Denote by Σ the set of behavior strategies for player 1 and by $\mathcal{T}$ for player 2. By Kolmogrov's extension theorem, together with p and q, any strategy profile $(\sigma, \tau) \in \Sigma \times \mathcal{T}$, induces a unique probability distribution over $(K \times H_\infty, \mathcal{H}_\infty)$. $\mathbb{E}^p_{\sigma,\tau}[\cdot]$ denotes the expectation of this probability, and $\mathbb{E}^k_{\sigma,\tau}[\cdot]$ its conditional expectation for given k.

Evaluations and values A play $h \in K \times H_\infty$ induces a stream of stage payoffs $(g^k(\omega_1, i_1, j_1), ..., g^k(\omega_t, i_t, j_t), ...)$. We consider general means of this stream by probability distributions over $N^* = \mathbb{N}/\{0\}$. For any $\xi = (\xi_t)_{t\geq 1} \in \Delta(\mathbb{N}^*)$, the expected ξ-evaluated payoff associated with $(\sigma, \tau) \in \Sigma \times \mathcal{T}$ is

$$\gamma^p_\xi(\sigma,\tau) = \mathbb{E}^p_{\sigma,\tau}\left[\sum_{t=1}^{\infty} \xi_t g^k(\omega_t, i_t, j_t)\right] = \sum_k p^k \gamma^k_\xi(\sigma,\tau), \text{ where } \gamma^k_\xi(\sigma,\tau) = \mathbb{E}^k_{\sigma,\tau}\left[\sum_{t=1}^{\infty} \xi_t g^k(\omega_t, i_t, j_t)\right].$$

As particular cases, the n-stage average payoff is $\gamma^p_n(\sigma,\tau) = \mathbb{E}^p_{\sigma,\tau}\left[\frac{1}{n}\sum_{t=1}^n g^k(\omega_t, i_t, j_t)\right]$ and the λ-discounted payoff is $\gamma^p_\lambda(\sigma,\tau) = \mathbb{E}^p_{\sigma,\tau}\left[\sum_{t=1}^\infty \lambda(1-\lambda)^{t-1} g^k(\omega_t, i_t, j_t)\right]$ for any $n \in \mathbb{N}$ and $\lambda \in (0, 1]$.

Let Γ_ξ be the ξ-evaluated game where player 1 aims at maximizing $\gamma^p_\xi(\sigma,\tau)$ while player 2 aims at minimizing it. By Sion's Minmax theorem (cf. Sorin [59], A.3), Γ_ξ has a value, denoted by $v_\xi(p)$. Let $v_n(p)$ be the value of n-stage game and $v_\lambda(p)$ be the value of the λ-discounted game.

We are interested in the asymptotic behavior of $v_\xi(p)$ for ξ defining a large "duration" of the game. In the particular cases, we study the convergence of $v_n(p)$ as n tends to infinity or of $v_\lambda(p)$ as λ tends to zero, and the equality of both limits in case of convergence.

Definition 5.2.1. Γ_∞ *has an* **asymptotic value** v if $\lim_{n\to\infty} v_n(p) = \lim_{\lambda\to 0}(p) = v$. More generally, Γ_∞ has a **sup-asymptotic value** v if:

$$\forall \varepsilon > 0, \exists \eta > 0, \text{ s.t. for all } \xi = (\xi_t) \in \Delta(\mathbb{N}^*) \text{ with } \sup_{t\geq 1} \xi_t \leq \eta, \text{ we have } |v_\xi(p) - v(p)| \leq \varepsilon.$$

We denote by "$\lim_{\|\xi\|_\infty \to 0} v_\xi(p) = v$" for the existence of the sup-asymptotic value v. It is clear that $\lim_{\|\xi\|_\infty \to 0} v_\xi(p) = v$ implies that $\lim_{n\to\infty} v_n(p) = \lim_{\lambda\to\infty} v_\lambda(p) = v$. Even stronger, the existence of sup-asymptotic value implies the existence of TV-asymptotic value, which is defined for the convergence of $v_\xi(p)$ as $TV(\xi) = \sum_{t\geq 1} |\xi_t - \xi_{t+1}|$ vanishes (cf. Sorin [59], Renault [46]).

The following notions ask for uniform properties on approximately optimal strategies.

Definition 5.2.2. *$\underline{v}(p)$ is the **Maxmin** of Γ_∞ if the both hold:*
– *player 1 **can guarantee** $\underline{v}(p)$:*

$$\forall \varepsilon > 0, \exists \sigma_\varepsilon \in \Sigma, \exists N \in \mathbb{N}, \text{ such that } \gamma_n^p(\sigma_\varepsilon, \tau) \geq \underline{v}(p) - \varepsilon, \ \forall \tau \in \mathcal{T}, \forall n \geq N;$$

– *player 2 **can defend** $\underline{v}(p)$:*

$$\forall \varepsilon > 0, \forall \sigma \in \Sigma, \exists \tau_\varepsilon \in \mathcal{T}, \exists N \in \mathbb{N}, \text{ such that } \gamma_n^p(\sigma, \tau_\varepsilon) \leq \underline{v}(p) + \varepsilon, \ \forall n \geq N.$$

The ***Minmax*** $\bar{v}(p)$ of Γ_∞ is defined in a dual way. Γ_∞ has a uniform value if $\bar{v}(p) = \underline{v}(p)$.

We study the asymptotic value as well as the $Maxmin$ and $Minmax$. The main results are:

Theorem 5.2.3. *Sup-asymptotic value exists in Γ_∞.*

Theorem 5.2.4. *$\underline{v}(p)$ exists in Γ_∞, and moreover it is equal to the sup-asymptotic value.*

Theorem 5.2.5. *$\bar{v}(p)$ exists in Γ_∞.*

5.3 Asymptotic analysis

This section is devoted to the proof of Theorem 5.2.3. We first discuss reduced optimal strategies for players in Γ_ξ, which motivates us to introduce a sequence of auxiliary games.

Shapley equation and reduced optimal strategies Below is the Shapley equation defined on $\Delta(K)\times\Omega$ for stochastic games with one-sided incomplete information (cf. Sorin [60]):

$$v_\xi(p,\omega) = \mathrm{Val}_{(x,y)}\Big\{\xi_1 \sum_k p^k g^k(\omega, x^k, y) + \Big(1-\xi_1\Big) \sum_{i,j,k,\omega'} p^k x^k(i) y(j) q^k(\omega'|\omega,i,j) v_{\xi^+}\big(\bar{p}_x[i,j,\omega'], \omega'\big)\Big\}, \tag{5.3.1}$$

where (x,y) takes value in $\big(\Delta(I)\big)^K \times \Delta(J)$, $g^k(\omega, x^k, y)$ is the corresponding linear extensions, $\xi^+ = (\xi_t^+)_{t\geq 1} \in \Delta(\mathbb{N}^*)$ is defined as: $\xi_t^+ = \frac{\xi_{t+1}}{1-\xi_1}, \forall t \geq 1$, and $\bar{p}_x[i,j,\omega'] \in \Delta(K)$ is player 2's *posterior* belief over K conditional on (i,j,ω'). From (5.3.1), player 1 has an optimal strategy that is Markovian in player 2's *posterior* beliefs.

In our model of GBM with one-sided incomplete information, for player 1 to compute $\bar{p}_x[i,j,\omega']$, he needs to know j only when his realized action is $i = Top$. On the other hand, by restricting player 1 to take only a finite number of times the action Top, the induced loss is small if this number is large.

We are lead to consider the following auxiliary game Ξ_M^L where players use reduced strategies. We show that the sup-asymptotic value of Γ_∞ exists and is asymptotically

equal to the value of Ξ_M^L as L and M tends to infinity. Here, $1/L$ can be understood as the mesh of the uniform discretization of a "limit game" on $[0,1]$, and M is a total number of playing the action Top (to induce an absorption with a probability close to one).

Auxiliary game $\Xi_M^L(p)$ For any $L, M \in \mathbb{N}$, $\Xi_M^L(p)$ is played as follows. $k \in K$ is chosen according to p and is communicated to player 1 only; the stochastic states are $\bar{\Omega} = \Omega \times \{0, ..., M\}$, and $(\omega_1, m_1) = (\omega_0, 0)$; at each stage $\ell = 1, ..., L$, players take actions $(i_\ell, j_\ell) \in \{T, B\} \times J$: the induced stage payoff is $\bar{g}^k(\omega_\ell, m_\ell, i_\ell, j_\ell) = g^k(\omega_\ell, i_\ell, j_\ell)$, the transition probability for $(\omega_{\ell+1}, m_{\ell+1})$ is $\bar{q}^k(\cdot|\omega_\ell, m_\ell, i_\ell, j_\ell) \in \Delta(\bar{\Omega})$, and moreover, i_ℓ is public while j_ℓ is public only if $t_\ell = T$.

Denote by $\bar{q}_\omega^k$ (*resp.* $\bar{q}_m^k$) the marginal of $\bar{q}^k$ on Ω (*resp.* on $\{0, ..., M\}$), satisfying: $\forall (i,j) \in I \times J$,

- for $\omega \in \Omega^*$ or $m \leq M-2$:
  - $\bar{q}_\omega^k(\cdot|\omega, m, i, j) = q^k(\cdot|\omega, i, j)$
  - $\bar{q}_m^k(\cdot|\omega, m, i, j) = \mathbb{1}_{\{i=T\}}\delta_{m+1} + \mathbb{1}_{\{i=B\}}\delta_m$
- for $(\omega, m) = (\omega^o, M-1)$: $\bar{q}^k(\cdot|\omega_0, M-1, i, j) = \mathbb{1}_{\{i=T\}}\delta_{(\omega^{j^*}, M)} + \mathbb{1}_{\{i=B\}}\delta_{(\omega^o, M-1)}$.

Let t_m be the random stage of playing the m-th Top, and we write $t_j^m := (t_1, j_{t_1} ..., t_m, j_{t_m})$ for some $m \leq M$ (set $t_0 = 0$ and $t_j^0 = \emptyset$). A behavior strategy for player 1 is written as $\mu = (\mu^k)_k \in \mathcal{Q}_M^K[L]$ with $\mu^k = (\mu_1^k, ..., \mu_M^k)$, where for each k and m, $\mu_{m+1}^k(\cdot|t_j^m)$ is a probability measure over the set $\{t_m + 1, ..., L\}$ for any t_j^m with $t_m < L$. Similarly, a behavior strategy for player 2 is written as $f = (f^1, ..., f^M) \in F_M[L]$ where for each m, $f^{m+1}(\cdot|t_j^m) \in (\Delta(J))^{L-t_m}$ for any t_j^m with $t_m < L$.

The payoff function associated with any pair $(\mu, f) \in \mathcal{Q}_M^K[L] \times F_M[L]$ is:

$$\mathcal{L}^p(\mu, f) = \frac{1}{L}\sum_{\ell=1}^{L} \mathcal{L}_\ell^p(\mu, f) := \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{E}_{\mu,f}^p\Big[\bar{g}_\ell^k(\omega_\ell, m_\ell, i_\ell, j_\ell)\Big] = \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{E}_{\mu,f}^p\Big[g_\ell^k(\omega_\ell, i_\ell, j_\ell)\Big],$$

where $\mathbb{E}_{\mu,f}^p[\cdot]$ denotes for the expectation operator of the unique probability distribution over $K \times (\bar{\Omega} \times I \times J)^L$ induced by $(\mu, f, p, (\bar{q}^k))$. Since I, J, M, L are all finite, $\Xi_M^L(p)$ has a value, which we denote by $w_M^L(p)$.

The proof of Theorem 5.2.3 will be obtained from the dual results Proposition 5.3.1 and Proposition 5.3.4. Indeed, they together imply that

$$\lim_{\|\xi\|_\infty \to 0} v_\xi(p) = \Lambda(p) := \lim_{L,M\to\infty} w_M^L(p).$$

Proof of Theorem 5.2.3 From Proposition 5.3.1 and Proposition 5.3.4,

$$\forall L, M \in \mathbb{N}, \xi \in \Delta(\mathbb{N}^*),\ \sup_{t\geq 1} \xi_t \leq \frac{1}{L^2} :\ \Big|v_\xi(p) - w_M^L(p)\Big| \leq 2C\Big[\frac{1}{L} + \frac{2M}{L} + (1-\chi^-)^M\Big].$$

For each $\varepsilon > 0$, we take $M = \Big\lfloor \frac{\ln \varepsilon}{\ln(1-\chi^-)} \Big\rfloor + 1$ and $2M/L \leq \varepsilon$, then $(1-\chi^-)^M \leq \varepsilon$ thus

$$\Big|v_\xi(p) - w_M^L(p)\Big| \leq 6C\varepsilon. \tag{5.3.2}$$

By taking M and L tend to infinity, we obtain that $\lim_{\|\xi\|_\infty \to 0} v_\xi(p) = \Lambda(p)$. □

The idea for the proof of Proposition 5.3.1 and Proposition 5.3.4 is as follows. For each player, we mimic in Γ_ξ some optimal strategy in Ξ_M^L so as to link the average payoff on a certain block $N(\ell)$ in Γ_ξ to the stage-ℓ payoff in Ξ_M^L. We choose the length of each $N(\ell)$ such that its total weight under ξ is approximately $1/L$, then asymptotically, both players guarantee $w_M^L(p)$.

Proposition 5.3.1. *For any $L, M \in \mathbb{N}$ and $\xi \in \Delta(\mathbb{N}^*)$ with $\sup_{t\geq 1} \xi_t \leq 1/L^2$, we have*

$$v_\xi(p) \geq w_M^L(p) - 2C\Big[(1-\chi^-)^M + \frac{2M}{L} + \frac{1}{L}\Big].$$

Proof. Fix L, M and ξ with $\sup_{t\geq 1} \xi_t \leq 1/L^2$. We introduce the consecutive blocks $N(1), ..., N(L)$ in $\mathbb{N}^*$ satisfying:

$$1/L^2 \leq \xi[\ell] := \sum_{t\in N(\ell)} \xi_t < 1/L + 1/L^2,\ \ell = 1, ..., L-1 \text{ and } N(L) = \mathbb{N}^* / \cup_{\ell=1}^{L-1} N(\ell). \tag{5.3.3}$$

For any $m \in \mathbb{N}$, T_m is the random stage on which player 1 plays the m-th Top in Γ_ξ. We denote $T^m := (T_1, ..., T_m)$ and $(T_j^m) := (T_1, j_{T_1} ..., T_m, j_{T_m})$ (set by convention $T_0 = 0$ and $T_j^0 = \emptyset$). For any sequence T_j^m, we use $t_j^m := (t_1, j_{T_1} ..., t_m, j_{T_1})$ to denote the index of the blocks $N(t^m) := \big(N(t_1), ..., N(t_m)\big)$ with $T_{m'} \in N(t_{m'}), 1 \leq m' \leq m$ and the corresponding moves of player 2. Each t_j^m is identified with a history $(t_1, j_{t_1} ..., t_m, j_{t_m})$ in Ξ_M^L with $j_{t_{m'}} = j_{T_{m'}}, \forall m'$.

Take $\mu = (\mu_1, ..., \mu_M) \in \mathcal{Q}_M^K[L]$ an optimal strategy for player 1 in Ξ_M^L. We define $\sigma := \sigma[\mu; \xi] \in \Sigma$ a behavior strategy in Γ_ξ such that: for all $m = 0, ..., M-1$ and $k \in K$,

- $Prob_\sigma^k(T_{m+1} \in N(\ell) | T_j^m) = \mu^k(\ell | t_j^m),\ \forall t_j^m,\ \forall k,\ \forall \ell = t_m + 1, ..., L;$
- $Prob_\sigma^k(T_{m+1} = s | T_{m+1} \in N(t_{m+1})) = \frac{\xi_s}{\xi[\ell]},\ \forall t_{m+1},\ s \in N(t_{m+1}).$

That is, after each history T_j^m, player 1 first looks at the corresponding history t_j^m in Ξ_M^L, and use $\mu^k(\cdot | t_j^m)$ in game k to choose the (index of) next block of playing Top; within each block $N[\ell]$, the conditional distribution of Top on a stage s is $\frac{\xi_s}{\xi[\ell]}$.

Consider now $\tau \in \mathcal{T}$ a behavior strategy of player 2 to play against σ. Since σ is defined to depend on player 2's past moves that appeared together with a Top, we can assume that τ depends on histories in the same way as σ.

We represent τ by a family of mappings $\tau(\cdot | T_j^m) : \{T_m + 1, ..., L\} \to \Delta(J), \forall T_j^m$. Define now $f = (f^1, ..., f^M) := f[\tau; \xi] \in F_M[L]$ as: for any $t_j^m = (t_1, j_{t_1}, ..., t_m, j_{t_m})$,

$$f^{m+1}(\ell | t_j^m) = \sum_{s\in N(\ell)} \frac{\xi_s}{\xi[\ell]} \mathbb{E}_{\sigma,\tau}^k \Big[j_s | \mathcal{H}(t_j^m)\Big],\ \forall \ell = t_m + 1, ..., L,$$

where $\mathcal{H}(t_j^m)$ is the event of histories in Γ_ξ "$T^m \in N(t^m),\ j_{T_{m'}} = j_{t_{m'}},\ 1 \leq m' \leq m$".

Here, we use the construction of σ to write

$$\begin{aligned}\mathbb{E}_{\sigma,\tau}^k \Big[j_s | \mathcal{H}(t_j^m)\Big] &= \sum_{s^m \in N(t^m)} Prob\Big(T^m = s^m | \mathcal{H}(t_j^m)\Big) \tau\Big(s | s^m, j_{s_{m'}} = j_{t_{m'}}, \forall 1 \leq m' \leq m\Big) \\ &= \sum_{s^m \in N(t^m)} \prod_{1\leq m' \leq m} \frac{\xi(s_{m'})}{\xi[t_{m'}]} \tau\Big(s | s^m, j_{s_{m'}} = j_{t_{m'}}, \forall 1 \leq m' \leq m\Big),\end{aligned}$$

which is independent of k and actually depends on τ and ξ only.

Fix now some $t_j^M = (t_1, j_{t_1}, ..., t_M, j_{t_M})$, and denote for each $k \in K$:

$$\gamma^k_{N(\ell)}(\sigma,\tau|t_j^M) := \mathbb{E}^k_{\sigma,\tau}\Big[\sum_{s\in N(\ell)} \frac{\xi_s}{\xi[\ell]} g^k(\omega_s, i_s, j_s)\Big|\mathcal{H}(t_j^M)\Big]$$

and

$$\mathcal{L}^k_\ell(t_j^M, f) := \mathbb{E}^k_{t^M,f}\Big[\bar{g}^k(\omega_\ell, m_\ell, i_\ell, j_\ell)|\bar{\mathcal{H}}(t_j^M)\Big],$$

where $\bar{\mathcal{H}}(t_j^M)$ is the event in Ξ^L_M of histories "$(t_1, j_{t_1}, ..., t_M, j_{t_M})$".

We compute $\gamma^k_{N(\ell)}(\sigma,\tau|t_j^M)$ and link it to $\mathcal{L}^k_\ell(t_j^M, f)$ by the following:

Claim 5.3.2. *For $\ell \notin \{t_1, ..., t_M\}$, $|\gamma^k_{N(\ell)}(\sigma,\tau|t_j^M) - \mathcal{L}^k_\ell(t_j^M, f)| \le 2C(1-\chi^-)^M$.*

Proof for Claim 5.3.2 Let $m \in \{0, ..., M\}$ with $t_m < \ell < t_{m+1}$ (set $t_{M+1} = L+1$). We take expectation w.r.t. the probability of the absorption on block $N(m'), m' = 1, ..., m$, to obtain:

$$\begin{aligned}
\gamma^k_{N(\ell)}(\sigma,\tau|t_j^M) &= \sum_{1\le m'\le m} Prob^k_{\sigma,\tau}\Big(\theta \in N(t_{m'})|\mathcal{H}(t_j^M)\Big)\overline{\Big\langle a^{k*}, \mathbb{E}_{\sigma,\tau}[j_{T_{m'}}|\mathcal{H}(t_j^M)]\Big\rangle}_{\chi^k} \\
&\quad + Prob^k_{\sigma,\tau}\Big(\theta > T_m|\mathcal{H}(t_j^M)\Big)\Big\langle a^k, \sum_{s\in N(\ell)} \frac{\xi_s}{\xi[\ell]}\mathbb{E}_{\sigma,\tau}[j_s|\mathcal{H}(t_j^M)]\Big\rangle \\
&= \sum_{1\le m'\le m}\prod_{1\le r\le m'}\Big(1-\chi^k(j_{t_r})\Big)\chi^k(j_{t_{m'}})\cdot a^{k*}(j_{t_{m'}}) \\
&\quad + \prod_{1\le r\le m}\Big(1-\chi^k(j_{t_r})\Big)\Big\langle a^k, f^{m+1}(\ell|t_j^m)\Big\rangle,
\end{aligned} \tag{5.3.4}$$

where we have denoted $\overline{\langle a^{k*}, y\rangle}_{\chi^k} = \frac{\sum_j y(j)\chi^k(j)a^{k*}(j)}{\sum_j y(j)\chi^k(j)}$ for $y \in \Delta(J)$.

On the other hand, we have by definition,

$$\begin{aligned}
\mathcal{L}^k_\ell(t_j^M, f) &= \sum_{1\le m'\le m} Prob^k_{\mu,f}\Big(\bar{\theta} = t_{m'}|\bar{\mathcal{H}}(t_j^M)\Big)\cdot a^{k*}(j_{t_{m'}}) \\
&\quad + Prob^k_{\mu,f}\Big(\bar{\theta} > t_M|\bar{\mathcal{H}}(t_j^M)\Big)\Big\langle a^k, f^{m+1}(\ell|t_j^m)\Big\rangle,
\end{aligned} \tag{5.3.5}$$

where we denote $\bar{\theta} = \inf\{\ell \ge 1|\omega_\ell \in \Omega^*\}$ for the stopping time in Ξ^L_M.

Finally, we use the probabilities in (5.3.5) defined by $(\bar{q}^k)$ to approximate the probabilities in (5.3.4) defined by (q^k) (denoting below $Prob^k_{t_j^M,f}(\cdot) := Prob^k_{\mu,f}(\cdot|\bar{\mathcal{H}}(t_j^M))$):

– for $m' < M$:

$$Prob^k_{t_j^M,f}(\bar{\theta} = t_{m'}) = \prod_{1\le r<m'}\Big(1-\chi^k(j_{t_r})\Big)\chi^k(j_{t_{m'}}) \text{ and } Prob^k_{t_j^M,f}(\bar{\theta} > t_m) = \prod_{1\le r\le m}\Big(1-\chi^k(j_{t_r})\Big)$$

– for $m' = M$ (using in Ξ^L_M: $Prob^k_{t_j^M,f}(\bar{\theta} = t_M|\bar{\theta} > t_{M-1}) = 1 - \bar{q}_\omega(\omega^o|\omega^o, M-1, T, j_{T_M}) = 1$),

$$Prob^k_{t_j^M,f}(\bar{\theta} = t_M) = \prod_{1\le r<M}\Big(1-\chi^k(j_{t_r})\Big) \le \prod_{1\le r<M}\Big(1-\chi^k(j_{t_r})\Big)\chi^k(j_{t_M}) + (1-\chi^-)^M$$

and

$$Prob^k_{t_j^M,f}(\bar{\theta} > T_M) = 0 \le \prod_{1\le r\le M}\Big(1-\chi^k(j_{t_r})\Big) \le (1-\chi^-)^M.$$

The above approximations are substituted back into (5.3.4) and (5.3.5), to yield

$$\left|\gamma^k_{N(\ell)}(\sigma,\tau|t^M_j) - \mathcal{L}^k_\ell(t^M_j, f)\right| \le 2C(1-\chi^-)^M,$$

which proves the claim. □

Next we show that the probability distributions of the random sequence $(t_1, j_{t_1}, ..., t_M, j_{t_M})$ are the same in both games. For any $t^M_j = (t(1), j(1), ..., t(M), j(M))$,

Claim 5.3.3. $Prob^k_{\sigma[\mu;\xi],\tau}(\mathcal{H}(t^M_j)) = Prob^k_{\mu,f[\tau;\xi]}(\bar{\mathcal{H}}(t^M_j)),\ \forall k \in K.$

Proof for Claim 5.3.3: We write

$$Prob^k_{\sigma[\mu;\xi],\tau}(\mathcal{H}(t^M_j)) = \prod_{m=0}^{M-1} Prob^k_{\sigma[\mu;\xi],\tau}\Big(T_{m+1} \in N(t(m+1)), j_{T_{m+1}} = j(t_{m+1})|\mathcal{H}(t^m_j)\Big)$$

and

$$Prob^k_{\mu,f[\tau;\xi]}(\bar{\mathcal{H}}(t^M_j)) = \prod_{m=0}^{M-1} Prob^k_{\mu,f[\tau;\xi]}\Big(t_{m+1} = t(m+1), j_{t_{m+1}} = j(m+1)|\bar{\mathcal{H}}(t^m_j)\Big),$$

then it is sufficient for us to show that for any $m = 0, ..., M-1$:

$$\begin{aligned} &Prob^k_{\sigma,\tau}\Big(T_{m+1} \in N(t(m+1)), j_{T_{m+1}} = j(m+1)|\mathcal{H}(t^m_j)\Big) \\ &= Prob^k_{\mu,f}\Big(t_{m+1} = t(m+1), j_{t_{m+1}} = j(m+1)|\bar{\mathcal{H}}(t^m_j)\Big). \end{aligned}$$

Indeed, following the definition of σ and $f^{m+1}(s|t^m_j) = \mathbb{E}_{\sigma,\tau}[j_s|\mathcal{H}(t^m_j)]$, we obtain:

$$\begin{aligned} &Prob^k_{\sigma[\mu;\xi],\tau}\Big(T_{m+1} \in N(t(m+1)), j_{T_{M+1}} = j(m+1)\Big) \\ &= \mu^k_{m+1}\Big(t_{m+1} = t(m+1)|t^m_j\Big) \cdot \sum_{s\in N(t_{m+1})} \frac{\xi_s}{\xi[t_{m+1}]} \cdot Prob^k_{\sigma,\tau}[j_s = j(m+1)|\mathcal{H}(t^m_j)] \\ &= \mu^k_{m+1}\Big(t(m+1)|t^m_j\Big) f^{m+1}(s|t^m_j)[j(m+1)] \\ &= Prob^k_{\mu,f[\tau;\mu]}\Big(t_{m+1} = t(m+1), j_{t_{m+1}} = j(m+1)|\bar{\mathcal{H}}(t^m_j)\Big), \end{aligned}$$

thus the equality is obtained for any m. The proof for the claim is achieved by taking the product of the conditional probabilities on both sides. □

Under each history, there are at most M "exceptional" blocks that contain a Top, so their total weight under ξ is at most $M \cdot (1/L + 1/L^2)$. For other random blocks, we apply Claim 5.3.2 and Claim 5.3.3 together: by taking expectation over all histories and summing over ℓ with the weight $\xi[\ell]$, we obtain

$$\sum_{\ell=1}^{L} \xi[\ell]\left|\gamma^k_{N(\ell)}(\sigma[\mu;N],\tau) - \mathcal{L}^k_\ell(\mu, f[\tau;\xi])\right| \le 2C(1-\chi^-)^M + 2CM(1/L + 1/L^2),\ \forall k$$

This implies that by taking expectation w.r.t. $p \in \Delta(K)$, we have

$$\begin{aligned} &\left|\gamma^p_\xi(\sigma[\mu;\xi],\tau) - \mathcal{L}^p(\mu, f[\tau;\xi])\right| \\ &\le \sum_{\ell=1}^{L} \xi[\ell] \cdot \left|\gamma^p_{N(\ell)}(\sigma,\tau) - \mathcal{L}^p_\ell(\mu, f)\right| + 2C\sum_{\ell=1}^{L}\left|\xi[\ell] - 1/L\right| \\ &\le 2C(1-\chi^-)^M + 2CM(1/L + 1/L^2) + 2C(1/L^2)(L-1) + 2C/L \\ &\le 2C\big[(1-\chi^-)^M + 2M/L^2 + 2/L\big], \end{aligned} \tag{5.3.6}$$

where we have used in the second inequality the fact that $|\xi[\ell]-1/L|\le 1/L^2$, $\forall \ell<L$ and $|\xi[L]-1/L|\le 1/L$. Finally, the optimality of μ in Ξ_M^L implies that

$$v_\xi(p)\ge w_M^L(p)-2C\big[(1-\chi^-)^M+2M/L^2+2/L\big].$$

The proof of the proposition is then complete. □

Proposition 5.3.4. *For any $L,M\in\mathbb{N}$ and $\xi\in\Delta(\mathbb{N}^*)$ with $\sup_{t\ge1}\xi_t\le 1/L^2$, we have*

$$v_\xi(p)\le w_M^L(p)+2C\Big[(1-\chi^-)^M+\frac{M}{L}+\frac{2}{L}\Big].$$

Proof. Fix L,M and ξ with $\sup_{t\ge1}\xi_t\le 1/L^2$. Take some $f=(f^1,...,f^M)\in F_M[L]$ which is optimal for player 2 in Ξ_M^L. We define a behavior strategy $\tau:=\tau[f;\xi]\in\mathcal{T}$ together with a consecutive random blocks $\hat N(1),...,\hat N(L)$ by induction on $m=0,...,M-1$ as follows. For any $T_j^m=(T_1,j_{T_1}...,T_m,j_{T_m})$, denote $t_j^m=(t_1,j_{T_1}...,t_m,j_{T_m})$ with $T_{m'}\in N(t_{m'})$, $1\le m'\le m$.

– For any $\ell=t_m+1,...$ (until t_{m+1}, to be specified):

$$\tau(s|T_j^m)=f^{m+1}(\ell|t_j^m)\text{ for all }s\in\hat N(\ell):=N(\ell)\cap\{T_m+1,...,T_{m+1}\},$$

where the block $N(\ell)$ is defined to start after $\hat N(\ell-1)$ with a length satisfying

$$\xi[\ell]:=\sum_{s\in N(\ell)}\xi_s\in\big[1/L,1/L+1/L^2\big).$$

Set $t_{m+1}>t_m$ with $T_{m+1}\in\hat N(t_{m+1})$ (by convention $t_{m+1}=L+1$ if $T_{m+1}=\infty$)

– after T_M, τ is defined to play any fixed action in $\Delta(J)$ and for the consecutive blocks after $\hat N(t_M)$: $\forall\ell=t_M+1,...,L-1$, $\hat N(\ell)=N(\ell)$ is set with a length satisfying:

$$\xi[\ell]\in\big[1/L,1/L+1/L^2\big).$$

– finally, the rest stages are put in the block

$$\hat N(L):=\mathbb{N}^*/\cup_{\ell=1}^{L-1}\hat N(\ell).$$

Consider now a behavior strategy $\sigma=(\sigma^k)\in\Sigma$ for player 1. It is sufficient for us to consider each σ^k that is pure and depends on histories only through T_j^m. Thus, given any $(j(1),...,j(M))\in J^M$, the sequence $T^M=(T_1,..,T_M)$, thus the blocks $\hat N(\ell),\ell=1,...,L$, is uniquely determined by σ^k in game k for $j_{T_m}=j(m),1\le m\le M$.

We fix now $(j(1),...,j(M))$, thus in game k, T_j^M, $(\hat N(\ell))$ and t_j^M are determined. We identify t_j^M with a history in Ξ_M^L. $\gamma_{N[\ell]}^k(\sigma,\tau|t_j^M)$ and $\mathcal{L}_\ell^k(t_j^M,f)$ are defined as in Proposition 5.3.1. We obtain then analog results.

Claim 5.3.5. *For $\ell\notin\{t_1,...,t_M\}$, $|\gamma_{N[\ell]}^k(\sigma,\tau|t_j^M)-\mathcal{L}_\ell^k(t_j^M,f)|\le 2C(1-\chi^-)^M$, $\forall k\in K$.*

Proof for Claim 5.3.5: Consider first $\ell<t_M$. Let $m\in\{0,...,M-1\}$ with $t_m<\ell<t_{m+1}$. Using the definition of τ, we have the following computation analog to that in Claim

5.3.2:

$$\begin{aligned}
\gamma^k_{N(\ell)}(\sigma,\tau|t^M_j) &= \sum_{1\le m'\le m} Prob^k_{\sigma,\tau}(\theta = T_{m'}|\mathcal{H}(t^M_j))\overline{\Big\langle a^{k*}, \mathbb{E}_{\sigma,\tau}\big[j_{T_{m'}}|\mathcal{H}(t^M_j)]\big)\Big\rangle}_{\chi^k} \\
&\quad + Prob^k_{\sigma,\tau}(\theta > T_m)\Big\langle a^k, \sum_{s\in N(\ell)} \frac{\xi_s}{\xi[\ell]}\mathbb{E}^k_{\sigma,\tau}\big[j_s|\mathcal{H}(t^M_j)]\Big\rangle \\
&= \sum_{1\le m'\le m}\prod_{1\le r<m'}\Big(1-\chi^k(j(r))\Big)\chi^k(j(m'))a^{k*}(j(m')) \\
&\quad + \prod_{1\le r\le m}\Big(1-\chi^k(j(r))\Big)\Big\langle a^k, f^{m+1}(t_{m+1}|t^m_j)\Big\rangle \\
&= \sum_{1\le m'\le m} Prob^k_{t^M_j,f}(\theta = t_{m'})a^{k*}(j(m')) + Prob^k_{t^M_j,f}(\theta > t_m)\Big\langle a^k, f^{m+1}(\ell|t^m)\Big\rangle \\
&= \mathcal{L}^k_\ell(t^M_j, f).
\end{aligned}$$

Next, for any $\ell > t_M$. The computation of $\gamma^k_{N(\ell)}(\sigma,\tau|t^M_j)$ is the same as above except that one has to take into consideration of the absorption after T^M. This event happens with a probability of at most $(1-\chi^-)^M$, hence by the approximation of $\bar{q}^k_\omega(\cdot)$ to $q^k(\cdot)$, we obtain

$$\gamma^k_{N(\ell)}(\sigma,\tau|t^M_j) \le \mathcal{L}^k_\ell(t^M_j, f) + 2C(1-\chi^-)^M.$$

This finishes the proof for the claim. □

We define now some $\mu = (\mu^k) := \mu[\sigma;\xi] \in \mathcal{Q}^K_M[L]$ as follows. By construction of the random blocks, once we have $t^m = (t(1),...,t(m))$ the index of blocks with "$T_{m'} \in \hat{N}(t_{m'}), 1 \le m' \le m$", $(\hat{N}(\ell))^{t_m}_{\ell=1}$ thus the random stages T^m are fixed. For each $t^m_j = (t(1), j(1),...,t(m), j(m))$ identified as a history in Ξ^L_M, let for each $k \in K$:

$\mu^k_{m+1}(\cdot|t^m_j) = \delta_{t_{m+1}}$, where $t_{m+1} > t(m)$ is by definition the random stage $T_{m+1} \in \hat{N}(t_{m+1})$.

Since σ^k is pure, t_{m+1} is deterministic thus μ^k is also pure. We then obtain that for any $t^M_j = (t(1), j(1),...,t(M), j(M))$,

Claim 5.3.6. $Prob^k_{\sigma,\tau[f;\xi]}(\mathcal{H}(t^M_j)) = Prob^k_{\mu[\sigma;\xi],f}(\bar{\mathcal{H}}(t^M_j))$.

Proof for Claim 5.3.6: It is sufficient to prove that for any $m = 0,...,M-1$:

$$\begin{aligned}
&Prob^k_{\sigma,\tau}\Big(T_{m+1} \in \hat{N}(t(m+1)), j_{T_{m+1}} = \ j(m+1)|\mathcal{H}(t^m_j)\Big) \\
&= Prob^k_{\mu,f}\Big(t_{m+1} = t(m+1), j_{t_{m+1}} = j(m+1)|\bar{\mathcal{H}}(t^m_j)\Big).
\end{aligned}$$

Indeed, by the construction "$\tau(s|T^m_j) = f^{m+1}(\ell|t^m_j), \forall s \in \hat{N}(\ell)$" and the definition of μ^k, we have

$$\begin{aligned}
&Prob^k_{\sigma,\tau}\Big(T_{m+1} \in \hat{N}(t(m+1)), j_{T_{m+1}} = j(m+1)|\mathcal{H}(t^m_j)\Big) \\
&= Prob^k_\sigma(t_{m+1} = t(m+1)|T^m_j)\cdot Prob^k_\sigma\Big(j_{T_{m+1}} = j(m+1)|\mathcal{H}(t^m_j), t_{m+1} = t(m+1)\Big) \\
&= \mu^k(t(m+1)|t^m_j)\cdot f^{m+1}(t(m+1)|t^m_j)\,[j(m+1)] \\
&= Prob^k_{\mu,f}\Big(t_{m+1} = t(m+1), j_{t_{m+1}} = j(m+1)|\bar{\mathcal{H}}(t^m_j)\Big).
\end{aligned}$$

The proof for the claim is achieved by taking product probabilities over histories onboth sides. □

For any block $\ell < L$ with $\hat{N}(\ell) = N(\ell)$, we have $\xi[\ell] \geq 1/L$. Under any history, there are at most M "exceptional" blocks for which $\hat{N}(\ell) \neq N(\ell)$ (containing the first M Top's), thus their total weight under μ is bounded by:

$$\sum_{\ell:\ \hat{N}(\ell)\neq N(\ell)} \xi[\ell] + \xi[L] \leq 1 - (L-M)/L = M/L.$$

On the other hand, for blocks $N(\ell)$ with $\hat{N}(\ell) = N(\ell)$, we use together Claim 5.3.5 and Claim 5.3.6. Taking expectation over all histories and summing $\gamma^k_\xi(\sigma, \tau[f;\xi])$ over $\ell = 1, ..., L$ (weighted by $\xi[\ell]$), we obtain

$$\begin{aligned}\gamma^k_\xi(\sigma, \tau[f;\xi]) &\leq \sum_{\ell=1}^{L} \xi[\ell] \cdot \mathcal{L}^k_\ell(\mu[\sigma;\xi], f) + 2CM/L + 2C(1-\chi^-)^M \\ &\leq \mathcal{L}^k(\mu[\sigma;\xi], f) + 4C/L + 2CM/L + 2C(1-\chi^-)^M,\end{aligned}$$

where the error term $4C/L$ is due to the approximation $\sum_{\ell=1}^{L} |\xi[\ell] - 1/L| \leq 2/L$. Next, consider now σ being mixed, we obtain a mixed strategy $\mu[\sigma;\xi] \in \mathcal{Q}^L_M[p]$ following the same probability distribution of σ over the pure ones. We take expectation w.r.t. $p \in \Delta(K)$, to obtain

$$\gamma^p_\xi(\sigma, \tau[f;\xi]) \leq \mathcal{L}^p(\mu[\sigma;\xi], f) + 4C/L + 2CM/L + 2C(1-\chi^-)^M.$$

Finally, since $f \in F_M[L]$ is optimal, we obtain that for any $\xi \in \Delta(\mathbb{N}^*)$ with $\sup_{t\geq 1} \xi_t \leq 1/L^2$:

$$v_\xi(p) \leq \mathcal{L}^p(\mu[\sigma;\xi], f) + 2C\Big[2/L + M/L + (1-\chi^-)^M\Big] \leq w^L_M(p) + 2C\Big[2/L + M/L + (1-\chi^-)^M\Big].$$

The proof for the proposition is complete. □

5.4 Uniform analysis: $Maxmin = \Lambda(p)$

We study $Maxmin$ of Γ_∞ in this section, and prove Theorem 5.2.4 by showing that $\underline{v}(p) = \Lambda(p)$.

In order for us to prove the result, we first give another characterization for $\Lambda(p)$ by games played on $[0,1]$ in Subsection 5.4.1 (cf. Proposition 5.4.4). For this characterization, we prove in Subsection 5.4.2 that player 2 defends it (cf. Proposition 5.4.5), and in Subsection 5.4.1 that player 1 guarantees it (cf. Proposition 5.4.9).

5.4.1 Preparation for the proof

The construction of our proof will be through the induction on the number of Top's being played. We introduce here a generic game whose properties are used at each step of iteration.

$(\hat{\Xi}(p), A^*_\cdot)$ is played on $[0,1]$ where the auxiliary absorbing payoffs (for the auxiliary "absorbing states") are defined by $A^*_\cdot = \{A^*_\cdot(j) | j \in J\}$ a class of functions on $\Delta(K)$ (which can be seen as the space of player 2' *posterior* beliefs). Formally,

- player 1 takes an action in $\mathcal{Q}^K[0] = \{\text{Borel probability measures on } [0,1]\}^K$ and player 2 takes an action in $F[0] = \{\text{measurable functions from } [0,1] \text{ to } \Delta(J)\}$;

– the payoff is defined for each profile $(\mu, f) = ((\mu^k)_{k\in K}, f) \in \mathcal{Q}^K[0]\times F[0]$: $\varphi^p(\mu, f) = \int_0^1 \varphi_t^p(\mu, f)dt$, where for any $t \in [0, 1]$:

$$\begin{aligned}\varphi_t^p(\mu, f) &= \sum_{k\in K} p^k \left(\int_0^t \mu^k(ds)\Big\langle A^*_{\bar{p}_\mu(s)}, f(s)\Big\rangle + (1-\underline{\mu}^k(t))\Big\langle a^k, f(t)\Big\rangle\right) \\ &= \int_0^t \mu(ds)\Big\langle A^*_{\bar{p}_\mu(s)}, f(s)\Big\rangle + (1-\underline{\mu}(t))\Big\langle a_{\tilde{p}_\mu(t)}, f(t)\Big\rangle.\end{aligned} \tag{5.4.1}$$

Here, $\bar{p}_\mu(s)$ and $\tilde{p}_\mu(s)$ are player 2's posteriors over K conditional on respectively "*Top* being played on $[s, s + dt)$" and "*Top* not being played yert until s". Formally, for each $k \in K$,

$$s \mapsto \bar{p}^k_\mu(s) = \frac{p^k\mu^k(ds)}{\sum_k p^k\mu^k(ds)}$$

is the Radon-Nikodyn derivative of $p^k\mu^k(ds)$ w.r.t. $\mu(ds) := \sum_k p^k\mu^k(ds)$, and

$$s \mapsto \tilde{p}^k_\mu(s) = \frac{1-\underline{\mu}^k(s)}{\sum_k p^k(1-\underline{\mu}^k(s))} = \frac{1-\underline{\mu}^k(s)}{1-\underline{\mu}(s)},$$

where we have denoted $\underline{\mu}^k(s) = \mu^k([0, s))$ and $\underline{\mu}(s) = \sum_k p^k\mu^k([0, s))$.

The following two propositions generalize results in Sorin [55], which corresponds to the case $A^*_p = a^*_p = \sum_k p^k a^{k*}$ being affine in $p \in \Delta(K)$.

Proposition 5.4.1. *Assume that for each $j \in J$, the function $A^*_\cdot(j)$ defined on $\Delta(K)$ is concave and C-Lip. Then $(\hat{\Xi}(p), A^*_\cdot)$ has a value,*

$$\hat{w}(p) := \max_{\mu\in\mathcal{Q}^K[0]} \min_{f\in F[0]} \varphi^p(\mu, f) = \min_{f\in F[0]} \max_{\mu\in\mathcal{Q}^K[0]} \varphi^p(\mu, f).$$

Proof. To show that $(\hat{\Xi}(p), A^*_\cdot)$ has a value by Sion's minmax theorem, our proof relies on the following result established in Forges [19] (cf. Mertens *et al.* [35], Ex.4 in p.144).

Lemma 5.4.2 (Forges 1988). [1] *Let K be a finite set, U a separable metric space, and $h : U \times \Delta(K) \to \bar{\mathbb{R}}$ be upper semi-continuous, and concave on $\Delta(K)$ for each $u \in U$. Let for any $P \in \Delta(K \times U)$:*

$$\phi(P) = \int_U h(u, [P(k|u)]_{k\in K}))P(du).$$

Then ϕ is concave and upper semi-continuous and $\{(P, \phi(P))|P$ has finite support$\}$ is dense in the graph of ϕ.

We first prove that for any f in $F'[0]$, the payoff function $\varphi^p(\cdot, f)$ is *u.s.c.* and *concave* in $\mu \in \mathcal{Q}^K[0]$ by Lemma 5.4.2. To do this, we rewrite the integral payoff $\varphi^p(\mu, f)$ to make it the sum of two parts such that the second part is affine in μ and the first takes the form of $\phi(\cdot)$ in Lemma 5.4.2.

We apply Fubini's theorem for $\varphi^p(\mu, f)$, to have

$$\varphi^p(\mu, f) = \int_0^1 \left(\Big\langle A^*_{\bar{p}_\mu(t_1)}, f(t_1)\Big\rangle(1-t_1)\right)\mu(dt_1) + \int_0^1 \left(\int_0^{t_1} \Big\langle a_{\bar{p}_\mu(t)}, f(t)\Big\rangle dt\right)\mu(dt_1). \tag{5.4.2}$$

1. Fabien Gensbittel is acknowledged for pointing out this result to the author.

The second part (non-absorbing payoff) in (5.4.2) is

$$\int_0^1 \Big(\int_0^{t_1} \Big\langle a_{\bar{p}_\mu(t)}, f(t) \Big\rangle dt \Big) \mu(dt_1) = \sum_k p^k \int_0^1 \Big(\int_0^{t_1} \Big\langle a^k, f(t) \Big\rangle dt \Big) \mu^k(dt_1),$$

which is affine in μ. As for the first part (absorbing payoff), we set as in Lemma 5.4.2 the following:

- $U = [0,1]$ and $u = t_1$;
- $P = p \otimes \mu \in \Delta(K \times [0,1])$;
- $[P(k|t_1)]_{k \in K} = \bar{p}_\mu(t_1)$;
- $[0,1] \times \Delta(K) \ni (t_1, \bar{p}) \mapsto h(t_1, \bar{p}) = \langle A^*_{\bar{p}}, f(t_1) \rangle \cdot (1 - t_1)$.

The above notations enable us to write

$$\phi(p \otimes \mu) = \int_U h(u, [P(k|u)]_{k \in K})) P(du) = \int_0^1 \Big(\Big\langle A^*_{\bar{p}_\mu(t_1)}, f(t_1) \Big\rangle \cdot (1 - t_1) \Big) \mu(dt_1).$$

Next we check that the conditions in Lemma 5.4.2 are satisfied, i.e. the function

$$(t_1, \bar{p}) \mapsto h(t_1, \bar{p}) = \langle A^*_{\bar{p}}, f(t_1) \rangle \cdot (1 - t_1)$$

defined from $[0,1] \times \Delta(K)$ to $\mathbb{R}$ is *u.s.c* in $(t_1, \bar{p})$, and is *concave* in $\bar{p}$ for each $t_1 \in [0,1]$.

Indeed, by assumption $\bar{p} \mapsto A^*_{\bar{p}}(j)$ is concave and Liptchitz continuous for each $j \in J$, and so is its linear extension to $\langle A^*_{\bar{p}}, f(t_1) \rangle$; moreover, since $f \in F'[0]$, the function $(t_1, \bar{p}) \mapsto h(t_1, \bar{p}) = \langle A^*_{\bar{p}}, f(t_1) \rangle \cdot (1 - t_1)$ is joint continuous. Lemma 5.4.2 applies for us to obtain that $\phi(p \otimes \mu)$ is concave and *u.s.c.* in $p \otimes \mu$, which is thus in particular concave and *u.s.c.* in μ. To sum up the two parts, we see that $\varphi^p(\mu, f)$ is concave and *u.s.c.* in μ.

Consider now the restricted game $\hat{\Xi}'(p)$ of $\hat{\Xi}(p)$ where player 2's action set is reduced to $F'[0]$. $\mathcal{Q}^K[0]$ is convex, weakly-$*$ compact and $F'[0]$ is convex; the payoff $\varphi^p(\cdot, \cdot)$ is affine in $f \in F'[0]$, and is concave and *u.s.c.* (*w.r.t.* the weak-$*$ topology) in $\mu \in \mathcal{Q}^K[0]$. Sion's minmax theorem applies to have the existence of its value $\hat{w}'(p)$.

Next we have

$$\underline{\hat{w}}(p) := \max_{\mu \in \mathcal{Q}^K[0]} \min_{f \in F[0]} \int_0^1 \varphi^p_t(\mu, f) dt \geq \max_{\mu \in \mathcal{Q}^K[0]} \min_{f \in F'[0]} \int_0^1 \varphi^p_t(\mu, f) dt = \hat{w}'(p).$$

Indeed, for each $\mu \in \mathcal{Q}^K[0]$ and $\varepsilon > 0$, let $f \in F[0]$ with $\int_0^1 \varphi^p_t(\mu, f) dt \leq \underline{\hat{w}}(p) + \varepsilon$. There exists, by Lusin's Theorem, some $f' \in F'[0]$ such that

$$\Big| \int_0^1 \varphi^p_t(\mu, f) dt - \int_0^1 \varphi^p_t(\mu, f') dt \Big| \leq \varepsilon.$$

This implies that $\hat{w}'(p) \leq \underline{\hat{w}}(p)$, thus

$$\max_{\mu \in \mathcal{Q}^K[0]} \min_{f \in F[0]} \int_0^1 \varphi^p_t(\mu, f) dt \geq \min_{f \in F'[0]} \max_{\mu \in \mathcal{Q}^K[0]} \int_0^1 \varphi^p_t(\mu, f) dt \geq \min_{f \in F[0]} \max_{\mu \in \mathcal{Q}^K[0]} \int_0^1 \varphi^p_t(\mu, f) dt.$$

This proves the existence of value $\hat{w}(p)$ in $\hat{\Xi}(p)$. □

Below we define a family of games $(\hat{\Xi}_m(p), A^{*,m}_{\cdot})$ with $A^{*,m}_{\cdot}$ being set in a recursive way:

– for $m = 1$: $A_p^{*,1}(j) = a_p^*(j)$, which is affine in p.
$\hat{\Xi}_1(p)$ has a value, which is denoted as $\hat{w}_1(p)$. Moreover, $\hat{w}_1(p)$ is *concave* and C-*Lip.* in p, as it is the value of an incomplete information game (cf. Sorin [59]).
– for $m \geq 2$: $A_p^{*,m}(j) := \chi^p(j)a^*_{\hat{p}(j)}(j) + (1-\chi^p(j))\hat{w}_{m-1}(\bar{p}(j))$, $\forall p \in \Delta(K)$, where $\hat{p}(j)$ and $\bar{p}(j)$ are *posteriors* of p given by: $\forall k \in K$,

$$\hat{p}^k(j) := \frac{p^k\chi^k(j)}{\sum_k p^k\chi^k(j)} \text{ and } \bar{p}^k(j) := \frac{p^k(1-\chi^k(j))}{\sum_k p^k(1-\chi^k(j))}.$$

By inductive assumption, $\hat{w}_{m-1}(\cdot)$ exists and is *concave* and C-*Lip.* Below we use Lemma 5.4.2 to show that $A_p^{*,m}(j)$ is *concave* and C-*Lip.* in p. This implies that Proposition 5.4.1 is applicable for $\hat{\Xi}_m$ to have a value $\hat{w}_m(\cdot)$, which is moreover *concave* and C-*Lip.* Thus our inductive definition of the recursive family $(\hat{\Xi}_m(p), A_\cdot^{*,m}(j))$ is complete.

Lemma 5.4.3. *For any $m = 1, ..., M$ and any $j \in J$, the mapping defined on $\Delta(K)$*

$$p \mapsto A_p^{*,m}(j) = \chi^p(j)a^*_{\hat{p}(j)}(j) + (1-\chi^p(j))\hat{w}_{m-1}(\bar{p}(j))$$

is concave and C-Lip.

Proof. We fix m, j and set as in Lemma 5.4.2 the following:
– $U = \{u^*, u^o\}$ endowed with the discrete topology, where u^* refers to "absorbing" and u^o refers to "non-absorbing";
– $P = p \otimes \chi(j) \in \Delta(K \times \{u^*, u^o\})$ defined as $\Big($denote $\chi(j) := (\chi^k(j))_{k\in K}\Big)$:

$$P(k, u^*) = p^k\chi^k(j) \text{ and } P(k, u^o) = p^k(1-\chi^k(j)) \text{ for all } k.$$

– $[P(k|u^*)]_{k\in K} = \hat{p}(j)$ and $[P(k|u^o)]_{k\in K} = \bar{p}(j)$.
– $h(\cdot,\cdot)$ on $U \times \Delta(K)$ is defined as: $h(u^*, q) = a_q^*(j)$ and $h(u^o, q) = \hat{w}_{m-1}(q)$ for all $q \in \Delta(K)$.

The above notations enable us to write

$$\phi(p \otimes \chi(j)) = \int_U h(u, [P(k|u)]_{k\in K})p \otimes \chi(j)(du) = \chi^p a^*_{\hat{p}(j)}(j) + (1-\chi^p)\hat{w}_{m-1}(\bar{p}(j)).$$

Now it is easy to verify that the conditions in Lemma 5.4.2 are satisfied: $h(\cdot,\cdot)$ is continuous in $(u, q) \in U \times \Delta(K)$ since $\hat{w}_{m-1}(q)$ is C-Lip. in q; for fixed $u \in U$, $h(u, q)$ is concave in q: either $u = u^*$, it is $a_q^*(j)$ thus linear in q, or $u = u^o$ it is $\hat{w}_{m-1}(q)$ thus concave in q. □

The following result implies that $\Lambda(p) = \lim_{M\to\infty} \hat{w}_M(p)$, thus a second characterization of the asymptotic value.

Proposition 5.4.4. *For any $M \in \mathbb{N}$, $\hat{w}_M(p) = \lim_{L\to\infty} w_M^L(p)$.*

We first introduce the uniform discretization of $(\hat{\Xi}(p), A_\cdot^*)$. For each $L \in \mathbb{N}$, let $(\hat{\Xi}^L(p), A_\cdot^*)$ be the auxiliary game defined by:
– players take actions $(\mu, f) \in \mathcal{Q}^K[L] \times F[L] = (\Delta(\{1, ..., L\}))^K \times (\Delta(J))^L$;
– the payoff function is $\hat{\mathcal{L}}^p(\mu, f) = \frac{1}{L}\sum_{\ell=1}^L \hat{\mathcal{L}}_\ell^p(\mu, f)$, where

$$\hat{\mathcal{L}}_\ell^p(\mu, f) = \sum_{1\leq\ell'\leq\ell} \mu(\ell')\langle A^*_{\bar{p}_\mu(\ell')}, f(\ell')\rangle + (1-\underline{\mu}(\ell))\langle a_{\tilde{p}_\mu(\ell)}, f(\ell)\rangle.$$

By exchanging the order of sum, we can also write:

$$\hat{\mathcal{L}}^p(\mu, f) = \frac{1}{L}\sum_{\ell=1}^L \mu(\ell)\Big[(L-\ell)\langle A^*_{\bar{p}_\mu(\ell)}, f(\ell)\rangle + \sum_{\ell'=1}^{\ell} \langle a_{\bar{p}_\mu(\ell)}, f(\ell')\rangle\Big] \tag{5.4.3}$$

Suppose that for each $j \in J$, $A^*(j)$. is concave and C-Lip. in $\bar{p} \in \Delta(K)$, then the same argument as in Proposition 5.4.1 applies to have the existence of its value, which we denote by $\hat{w}^L(p)$. Consider now the recursive family of discretization, and denote $\hat{w}^L_m(p)$ for the value of $(\hat{\Xi}^L_m, A^{*,m}_{\cdot})$, $m = 1, ..., M$.

Our proof for Proposition 5.4.4 is divided into the following two parts (denoting $\hat{\nu}_m(p) := \lim_{L\to\infty} \hat{w}^L_m(p)$ whenever it exists):

- **Part A**. $\hat{w}_m(p) = \hat{\nu}_m(p)$, $\forall p \in \Delta(K)$, $\forall m = 1, ..., M$. Eventually, we show that the optimal strategies in continuous time game give asymptotic optimal strategies in discretized game as the mesh of the discretization vanishes (cf. Sorin [55] or Mertens *et al.* [35] for the model of BM). We prove in the **Appendix** the generic result "$\hat{w}(p) = \hat{\nu}(p) := \lim_{L\to\infty} \hat{\nu}^L(p)$", which implies the result of **Part A**.
- **Part B**. $\lim_{L\to\infty} w^L_m(p) = \hat{\nu}_m(p)$, $\forall p \in \Delta(K)$, $\forall m = 1, ..., M$. We compare the recursive equations for $w^L_m(p)$ and $\hat{w}^L_m(p)$, and deduce their asymptotic convergence to the same limit as L tends to infinity. The proof is again put in the **Appendix**.

5.4.2 Player 2 defends $\Lambda(p)$

This whole subsection is devoted to the proof for

Proposition 5.4.5. *Player 2 defends $\Lambda(p)$ in Γ_∞.*

For fixed $\varepsilon > 0$, we take $M = \left\lfloor \frac{\ln \varepsilon}{\ln(1-\chi^-)} \right\rfloor + 1$ and and set L large enough satisfying

$$\hat{w}^L_m(\bar{p}) \leq \hat{w}_m(\bar{p}) + \varepsilon, \ \forall m \in \{1, ..., M\}, \forall \bar{p} \in \Delta(K).$$

For each $m = M, ..., 1$, we fix f^m a family of strategies in $F[L]$ such that $f^m[\bar{p}]$ is ε-optimal in $(\hat{\Xi}^L_m, A^{*,m}_{\cdot}(j))$ for each $\bar{p} \in \Delta(K)$. Denote $f := (f^1, ..., f^M)$. Consider any $\sigma \in \Sigma$ a behavior strategy in Γ_∞. We are going to construct a behavior strategy $\tau := \tau[f; \sigma] \in \mathcal{T}$ such that:

$$\gamma_n(\sigma, \tau[f; \sigma]) \leq \hat{w}^L_M(p) + \varepsilon$$

for all n sufficiently large.

An overview of the proof of Proposition 5.4.5

To define τ, we first present in **Step I** a generic construction; in **Step II**, we iterate this construction for M times; in **Step III** we compute the expected payoff to conclude.

We shortly explain here the idea of the generic construction, which is analog to Sorin [55]. Consider the game $(\hat{\Xi}(p), A^*_{\cdot})$ with an auxiliary absorbing payoff equal to the expected payoff that can be defended by player 2 after a Top. Let L large with $\hat{w}^L(p) \leq \hat{w}(p) + \varepsilon$ and consider $h \in F[L]$ an ε-optimal strategy in the discretized game $(\hat{\Xi}^L(p), A^*_{\cdot})$.

1. First, we use σ and h to compute the distribution of the stopping times for playing Top on the path of Γ_∞, if player 2 is "following" h. By "following" h, we mean that player 2 takes the action $h(\ell)$ *i.i.d.* until the probability of playing Top is almost exhausted, and then to take $h(\ell + 1)$ *i.i.d* until the probability of playing Top is almost exhausted ... etc. This defines in each state k a sequence of probabilities to hit Top on the consecutive L blocks, thus the measure $\mu = (\mu^k) \in \Delta(\{1, ..., L\})^K$;
2. Second, suppose that player 2 follows h until some ℓ and then to take the action there $h(\ell)$ *i.i.d.* forever. This induces in Γ_∞ an expected average payoff around $\varphi^p_l(\mu, h)$ in any long game.

3. Finally, as h is ε-optimal, there exists some $\ell^* \in \{1, ..., L\}$ satisfying $\varphi^p_{\ell^*}(\mu, h) \leq \hat{w}^L(p) + \varepsilon$. Take the behavior strategy $\tau_{\ell^*}[h; \sigma]$ to follow h until ℓ^*, then it defends $\hat{w}^L(p) \leq \hat{w}(p) + \varepsilon$ against σ.

To iterate, we define $\tau(f; \sigma)$ in **Step II** to play $\tau_{\ell^*_1}[f^1; \sigma]$ until stage T_1,...,to play $\tau_{\ell^*_m}[f^m; \sigma]$ until stage T_m,...,where for each m, the strategy $\tau_{\ell^*_m}[f^m; \sigma]$ is constructed as in **Step I** for the game $\langle \hat{\Xi}^L_{M-m+1}(\bar{p}_{m-1}), A^{*,m}_{\cdot} \rangle$ with $\bar{p}_{m-1}$ the *posterior* belief after T_{m-1}. Finally in **Step III**, our computation is based on induction.

Below we formalize the above analysis of three steps.

Step I. Defining a sequence of generic behavior strategies $\tau = (\tau_\ell)_{\ell=1}^L := \tau[h; \sigma]$ and the measures $\mu = (\mu^k)_{k \in K} := \mu[h; \sigma]$ associated with any $h \in F[L]$ and $\sigma \in \Sigma$.

Consider the generic auxiliary game $(\hat{\Xi}^L(p), A^*_{\cdot}(j))$. We fix $h \in \Delta(J)^L = F[L]$ a strategy for player 2 with $h(\ell)$ its component ℓ. For each $\ell = 1, ..., L$, τ_ℓ represents the behavior strategy that follows h until ℓ. They are used, together with σ, to compute in Γ_∞ the distribution of the stopping time $\tilde{T}$ for playing (the first) *Top*.

We define $\tau_\ell := \tau_\ell[h; \sigma]$ by induction on $\ell = 1, ..., L$ as follows.

– τ_1 is to: play the action $h(1)$ *i.i.d.*
  Let $N_1 \geq 1$ satisfy:

$$\mu^k(1) := Prob^k_{\sigma,\tau_1}(\tilde{T} \leq N_1) \geq Prob^k_{\sigma,\tau_1}(\tilde{T} < \infty) - \varepsilon, \ \forall k \in K.$$

– Similarly, for $\ell = 2, ..., L$, τ_ℓ is to: follow $\tau_{\ell-1}$ until stage $N_{\ell-1}$, and then to play the action $h(\ell)$ *i.i.d.*
  Let $N_\ell \geq N_{\ell-1} + 1$ satisfy:

$$\mu^k(\ell) := Prob^k_{\sigma,\tau_1}(N_{\ell-1} < \tilde{T} \leq N_\ell) \geq Prob^k_{\sigma,\tau_1}(N_{\ell-1} < \tilde{T} < \infty) - \varepsilon, \ \forall k \in K.$$

Denote by $B_\ell := \{N_{\ell-1} + 1, ..., N_\ell\}, \ell = 1, ..., L$ the consecutive random blocks.

Take $N_0 = \max\{N_\ell | 1 \leq \ell \leq L\}$, which is uniformly bounded in all histories. Fixing now any $\ell^* \in \{1, ..., L\}$ and $n \geq N_0$, let us compute the expected n-stage payoff induced by (σ, τ_{ℓ^*}), in the following recursive form:

Lemma 5.4.6. *For all $k \in K$,*

$$\begin{aligned}\mathbb{E}^k_{\sigma,\tau_{\ell^*}}[g_n] \leq \sum_{1 \leq \ell \leq \ell^*} \mu^k(\ell) \Big[\chi^k(h(\ell)) \overline{\langle a^{k*}, h(\ell) \rangle}_{\chi^k} + \big(1 - \chi^k(h(\ell))\big) \mathbb{E}^k_{\sigma,\tau_{\ell^*}} \big[g_n | \tilde{T} \in B_\ell, \theta > \tilde{T} \big] \Big] \\ + (1 - \underline{\mu}^k(\ell^*)) \langle a^k, h(\ell^*) \rangle + (2C + 1)\varepsilon.\end{aligned}$$

Proof. Fix any $n \geq N_0 \geq N_{\ell^*}$ for some ℓ^*. We write:

$$\begin{aligned}&\mathbb{E}^k_{\sigma,\tau_{\ell^*}}[g_n] \\ &= \sum_{1 \leq \ell \leq \ell^*} Prob^k_{\sigma,\tau_{\ell^*}}(\tilde{T} \in B_\ell) \Big[\chi^k(h(\ell)) \overline{\langle a^{k*}, h(\ell) \rangle}_{\chi^k} + \big(1 - \chi^k(h(\ell))\big) \mathbb{E}^k_{\sigma,\tau_{\ell^*}} \big[g_n | \tilde{T} \in B_\ell, \theta > \tilde{T} \big] \Big] \\ &+ Prob^k_{\sigma,\tau_{\ell^*}}(\tilde{T} \geq n) \cdot \big\langle a^k, h(\ell^*) \big\rangle + Prob^k_{\sigma,\tau_{\ell^*}}(N_{\ell^*} < \tilde{T} < n) \cdot \mathbb{E}^k_{\sigma,\tau_{\ell^*}} \big[g_n | N_{\ell^*} < \tilde{T} < n \big].\end{aligned}$$

By construction of the block B_ℓ and definition of the measure $\rho^k(\ell)$, we have

$$Prob^k_{\sigma,\tau_{\ell^*}}(\tilde{T} \in B_\ell) = \mu^k(\ell),\ \forall \ell = 1, ..., \ell^*.$$

Moreover, by definition of the stage N_{ℓ^*},

$$Prob^k_{\sigma,\tau_{\ell^*}}(N_{\ell^*} < \tilde{T} < n) \leq Prob^k_{\sigma,\tau_{\ell^*}}(N_{\ell^*} < \tilde{T} < \infty) \leq \varepsilon.$$

This implies that, to approximate $Prob^k_{\sigma,\tau_{\ell^*}}(\tilde{T} \geq n)$ by $Prob^k_{\sigma,\tau_{\ell^*}}(\tilde{T} > N_{\ell^*}) = (1 - \underline{\mu}^k(\ell^*))$ in the expression of $\mathbb{E}^k_{\sigma,\tau_{\ell^*}}[g_n]$, there is an error term $2C\varepsilon$. We obtain then the claim of the lemma. □

Step II. Defining the uniform best reply strategy τ^* as an iteration of $\tau_{\ell^*_1}$, ..., $\tau_{\ell^*_M}$.

For any $\bar{p} \in \Delta(K)$, $\ell^\sharp_m \in \{1, ..., L\}$ and $\bar{\sigma} \in \Sigma$, let $\tau_{\ell^\sharp_m} := \tau_{\ell^\sharp_m}[f^m; \bar{\sigma}] \in \mathcal{T}$ be the behavior strategy to follow $f^m[\bar{p}]$ until $\ell^\sharp_m$, as was defined in **Step I**. The corresponding blocks are

$$B^m_\ell = \{N^m_{\ell-1} + 1, ..., N^m_\ell\},\ \text{for } \ell = 1, ..., \ell^\sharp_m.$$

On each B^m_ℓ, $\tau_{\ell^\sharp_m}$ is playing $f^m[\bar{p}](\ell)$ *i.i.d*, and $N^m_\ell > N^m_{\ell-1}$ is set with (denote $N^m_1 = T_{m-1}$):

$$\rho^k_m(\ell) := Prob^k_{\bar{\sigma},\tau_{\ell^\sharp_m}}(N^m_{\ell-1} < T_m \leq N^m_\ell) \geq Prob^k_{\bar{\sigma},\tau_{\ell^\sharp_m}}(N^m_{\ell-1} < T_m < \infty) - \varepsilon,\ \forall k \in K.$$

We write $\rho_m = (\rho^k_m)_{k \in K}$ and then $\rho = (\rho_1, ..., \rho_M)$.

For any generic vector $\ell^\sharp = (\ell^\sharp_1, ..., \ell^\sharp_M) \in \{1, ..., L\}^M$, the behavior strategy $\tau_{\ell^\sharp} := \tau_{\ell^\sharp}[f; \sigma]$ is defined to iterate $\tau_{\ell_1}, ..., \tau_{\ell_M}$ as follows:

- for $m = 1$ (before T_1): $\tau_{\ell^\sharp}$ is to follow $\tau_{\ell^\sharp_1} = \tau_{\ell^\sharp_1}[f^1; \sigma]$ until the random stage T_1.

  Let us index by $\ell_1 \in \{1, ..., \ell^\sharp_1\}$ the block $B^1_{\ell_1}$ on which appears T_1, and denote by $\bar{p}_\rho(\ell^1_{\hat{\jmath}})$ player 2's posterior belief over K conditional on the event

  $$''T_1 \in B^1_{\ell_1}, j_{T_1} = \hat{\jmath}_1\ \&\ \theta > T_1''.$$

- for $m = 2, ..., M$ (after T^{m-1} and before T_m): τ_{ℓ^*} is to follow $\tau_{\ell^*_{m-1}}$ until the random stage T_{m-1}, and then to follow $\tau_{\ell^*_m} = \tau_{\ell^*_m}[f^m, \sigma(h_{T_{m-1}})]$ until the random stage T_m, where
  - $\sigma(h_{T_{m-1}})$ is the continuation of σ after $h_{T_{m-1}}$;
  - $\bar{p}_\rho(\ell^{m-1}_{\hat{\jmath}})$ is player 2's *posterior* belief over K conditional on the event

  $$''\ T_{m'} \in B^{m'}_{\ell_{m'}}, j_{T_{m'}} = \hat{\jmath}_{m'}, \forall\ m' \in \{1, ..., m-1\},\ \&\ \theta > T_{m-1}\ ''.$$

  We index by $\ell_m \in \{1, ..., \ell^\sharp_m\}$ the block $B^m_{\ell_m}$ on which appears T_m, and denote by $\bar{p}_\rho(\ell^m_{\hat{\jmath}})$ player 2's *posterior* belief over K conditional on the event

  $$''\ T_{m'} \in B^{m'}_{\ell_{m'}}, j_{T_{m'}} = \hat{\jmath}_{m'}, \forall\ m' \in \{1, ..., m\},\ \&\ \theta > T_{m-1}\ ''.$$

Next, we fix a random vector $\ell^* = (\ell^*_1, ..., \ell^*_M) \in \{1, ..., L\}^M$ along the play.

Notation 5.4.7. *For each $m = 0, ..., M-1$, we write below $\mathbf{q}_m := \bar{p}_\rho(\ell^m_{\hat{\jmath}})$ and $\mathcal{L}^{\mathbf{q}_m, M-m}$ for the payoff function in $\hat{\Xi}^L_{M-m}(\mathbf{q}_m)$.*

Each $f^{m+1}[\mathbf{q}_m]$ is ε-optimal, thus:

$$\mathcal{L}^{\mathbf{q}_m,M-m}(\rho_{m+1},f^{m+1})=\frac{1}{L}\sum_{\ell=1}^{L}\mathcal{L}_{\ell}^{\mathbf{q}_m,M-m}(\rho_{m+1},f^{m+1})\leq\hat{w}_{M-m}^{L}(\mathbf{q}_m)+\varepsilon\leq\hat{w}_{M-m}(\mathbf{q}_m)+2\varepsilon.$$

This implies that there exists some $\ell^*_{m+1}\in\{1,...,L\}$ with

$$\mathcal{L}_{\ell^*_{m+1}}^{\mathbf{q}_m,M-m}(\rho_{m+1},f^{m+1})\leq\hat{w}_{M-m}(\mathbf{q}_m)+2\varepsilon.$$

Finally, we define the behavior strategy $\tau_{\ell^*}=\tau_{\ell^*}[f;\sigma]$ associated with $\ell^*=\{\ell_1^*,...,\ell_M^*\}$.

Step III. Conclusion of the proof

Let $\bar{N}=\max\{N_\ell^m|1\leq\ell\leq L,1\leq m\leq M\}$. We prove the following

Proposition 5.4.8. *For any $n\geq\bar{N}$,*

$$\mathbb{E}^p_{\sigma,\tau_{\ell^*}}[g_n]\leq\Lambda(p)+(4C+3)\varepsilon+(4C+2)\frac{\varepsilon\ln\varepsilon}{\ln(1-\chi^-)}.$$

Proof. The expectation $\mathbb{E}^p_{\sigma,\tau_{\ell^*}}[g_n]$ is firstly taken conditional on the event "$T_1\in B^1_{\ell_1},\ \ell_1=1,...,\ell_1^*$", so Lemma 5.4.6 applies for $\tilde{T}=T_1$ under $(\sigma,\tau_{\ell_1^*})$, yielding:

$$\begin{aligned}
&\mathbb{E}^p_{\sigma,\tau_{\ell^*}}[g_n]=\sum_k p^k\mathbb{E}^k_{\sigma,\tau_{\ell^*}}[g_n]\\
\leq&\sum_k p^k\Bigg\{\sum_{\ell_1=1}^{\ell_1^*}\rho_1^k(\ell_1)\sum_{\hat{\jmath}_1}f^1(\ell_1)[\hat{\jmath}]\Big[\chi^k(\hat{\jmath}_1)a^{k*}(\hat{\jmath}_1)+\Big(1-\chi^k(\hat{\jmath}_1)\Big)\mathbb{E}^k_{\sigma,\tau_{\ell^*}}[g_n|\mathcal{A}_1]\Big]\\
&\qquad+(1-\underline{\rho}_1^k(\ell_1^*))\Big\langle a^k,f^1(\ell_1^*)\Big\rangle\Bigg\}+2C\varepsilon \qquad (5.4.4)\\
=&\sum_{\ell_1=1}^{\ell_1^*}\rho_1(\ell_1)\sum_{\hat{\jmath}_1}f^1(\ell_1)[\hat{\jmath}_1]\Big[\chi^{\bar{p}_\rho(\ell_1)}(\hat{\jmath}_1)a^*_{\hat{p}_\rho(\ell^1_{\hat{\jmath}})}(\hat{\jmath}_1)+\Big(1-\chi^{\bar{p}_\rho(\ell_1)}(\hat{\jmath}_1)\Big)\mathbb{E}^{\mathbf{q}_1}_{\sigma,\tau_{\ell^*}}[g_n|\mathcal{A}_1]\Big]\\
&\qquad+(1-\underline{\rho}_1(\ell_1^*))\Big\langle a_{\tilde{p}_\rho(\ell_1^*)},f^1(\ell_1^*)\Big\rangle+2C\varepsilon,
\end{aligned}$$

where $\hat{p}_\rho(\ell^1_{\hat{\jmath}})$ is player 2's *posterior* belief conditional on "$T_1\in B^1_{\ell_1},j_{T_1}=\hat{\jmath}_1,\theta=T_1$", $\tilde{p}_\rho(\ell_1^*)$ is player 2's *posterior* belief conditional on "$T_1>N^1_{\ell_1^*}$", and $\mathcal{A}_1$ denotes the event

$$''T_1\in B^1_{\ell_1},j_{T_1}=\hat{\jmath}_1\ \&\ \theta>T_1''.$$

One compares the payoff on the right hand side of (5.4.4) to $\hat{\mathcal{L}}^{p,M}_{\ell_1^*}(\rho_1,f^1)$, to obtain:

$$\begin{aligned}
&\mathbb{E}^p_{\sigma,\tau_{\ell^*}}[g_n]-\hat{\mathcal{L}}^{p,M}_{\ell_1^*}(\rho_1,f^1)\\
=&\sum_{\ell_1=1}^{\ell_1^*}\sum_{\hat{\jmath}_1\in J}\rho_1(\ell_1)f^1(\ell_1)[\hat{\jmath}_1]\Big(1-\chi^{\bar{p}_\rho(\ell_1)}(\hat{\jmath}_1)\Big)\Bigg[\mathbb{E}^{\mathbf{q}_1}_{\sigma,\tau_{\ell^*}}[g_n|\mathcal{A}_1]-\hat{w}_{M-1}(\mathbf{q}_1)\Bigg]+2C\varepsilon\\
\leq&(1-\chi^-)\sum_{\ell_1=1}^{\ell_1^*}\sum_{\hat{\jmath}_1\in J}\rho_1(\ell_1)f^1(\ell_1)[\hat{\jmath}_1]\Bigg[\mathbb{E}^{\mathbf{q}_1}_{\sigma,\tau_{\ell^*}}[g_n|\mathcal{A}_1]-\hat{w}_{M-1}(\mathbf{q}_1)\Bigg]+2C\varepsilon
\end{aligned}$$

Moreover, as we have chosen ℓ_1^* in **Step II** satisfying $\hat{L}^{p,M}_{\ell_1^*}(\rho_1, f^1) \leq \hat{w}_M(p)+2\varepsilon$, we obtain thus the following recursive inequality

$$\begin{aligned}&\mathbb{E}^p_{\sigma,\tau_{\ell^*}}[g_n] - \hat{w}_M(p)\\ \leq\ &(1-\chi^-)\sum_{\ell_1=1}^{\ell_1^*}\sum_{\hat{\jmath}_1\in J}\rho_1(\ell_1)f^1(\ell_1)[\hat{\jmath}_1]\left[\mathbb{E}^{\mathbf{q}_1}_{\sigma,\tau_{\ell^*}}[g_n|\mathcal{A}_1] - \hat{w}_{M-1}(\mathbf{q}_1)\right] + 2(C+1)\varepsilon.\end{aligned}$$

Similarly, we use Lemma 5.4.6 consequentially for $m = 2, ..., M$ the random times $\tilde{T} = T_m$ under $(\sigma(h_{T_{m-1}}), \tau_{\ell^*_m})$ to obtain:

$$\begin{aligned}&\mathbb{E}^{\mathbf{q}_{\mathrm{m}-1}}_{\sigma,\tau_{\ell^*}}[g_n|\mathcal{A}_{m-1}] - \hat{w}_{M-m+1}(\mathbf{q}_{\mathrm{m}-1})\\ \leq\ &(1-\chi^-)\sum_{\ell_m=1}^{\ell_m^*}\sum_{\hat{\jmath}_m\in J}\rho_m(\ell_m)f^m(\ell_m|\mathbf{q}_{\mathrm{m}-1})[\hat{\jmath}_m]\left[\mathbb{E}^{\mathbf{q}_{\mathrm{m}}}_{\sigma,\tau_{\ell^*}}[g_n|\mathcal{A}_m] - \hat{w}_{M-m}(\mathbf{q}_{\mathrm{m}})\right] + 2(C+1)\varepsilon,\end{aligned}$$

where $\hat{w}_0(\mathbf{q}_{\mathrm{M}}) = a^*_{\mathbf{q}_{\mathrm{M}}}$ and $\mathcal{A}_m$ is the event "$T_{m'} \in B^{m'}_{\ell_{m'}}, j_{T_{m'}} = \hat{\jmath}_{m'}, \forall m' \in \{1, ..., m\}$ & $\theta > T_m$".

Finally, we take expectation iteratively over histories to yield:

$$\begin{aligned}&\mathbb{E}^p_{\sigma,\tau_{\ell^*}}[g_n] - \hat{w}_M(p)\\ \leq\ &(1-\chi^-)^M\sum_{\ell^M\leq\ell^*}\sum_{\hat{\jmath}^M\in J^M}\rho(\ell^M)\prod_{m=1}^{M}f^m(\ell_m|\mathbf{q}_{\mathrm{m}-1})[\hat{\jmath}_m]\left(\mathbb{E}^{\mathbf{q}_{\mathrm{M}}}_{\sigma,\tau_{\ell^*}}[g_n|\mathcal{A}_M] - a^*_{\mathbf{q}_{\mathrm{M}}}(\hat{\jmath}_M)\right) + 2M(C+1)\varepsilon.\end{aligned}$$

where we have denoted $''\ell^M \leq \ell^{*''}$ for $''1 \leq \ell_m \leq \ell^*_m,\ m = 1, ..., M''$, $\hat{\jmath}^M = (\hat{\jmath}_1, ..., \hat{\jmath}_M)$ and $\rho(\ell^M) = \prod_{1\leq m\leq M}\rho_m(\ell_m)$.

By definition of $M = \left\lfloor\frac{\ln\varepsilon}{\ln(1-\chi^-)}\right\rfloor + 1$, $(1-\chi^-)^M \leq \varepsilon$, thus

$$\mathbb{E}^p_{\sigma,\tau_{\ell^*}}[g_n] \leq \hat{w}_M(p) + (4C+2)M\varepsilon \leq \hat{w}_M(p) + (4C+2)\Big[\frac{\ln\varepsilon}{\ln(1-\chi^-)} + 1\Big]\varepsilon.$$

Moreover, from the proof of Theorem 5.2.3 (cf. (5.3.2)), we have for such M:

$$|\hat{w}_M(p) - \Lambda(p)| \leq \varepsilon.$$

Thus we obtain that for any $n \geq \bar{N}$:

$$\mathbb{E}^p_{\sigma,\tau_{\ell^*}}[g_n] \leq \Lambda(p) + (4C+3)\varepsilon + (4C+2)\frac{\varepsilon\ln\varepsilon}{\ln(1-\chi^-)}.$$

This completes the proof for the proposition. □

Proof for Proposition 5.4.5: Take now $\varepsilon' = (4C+3)\varepsilon + (4C+2)\frac{\varepsilon\ln\varepsilon}{\ln(1-\chi^-)}$ and consider any $n \geq N_0 := \bar{N}/\varepsilon'$, we obtain $\gamma^p_n(\sigma, \tau_{\ell^*}) \leq \Lambda(p) + 2\varepsilon'$.. Since $\varepsilon > 0$ is arbitrary and $\frac{\varepsilon\ln\varepsilon}{\ln(1-\chi^-)}$ vanishes as ε tends to zero, this proves that player 2 defends $\Lambda(p)$. □

5.4.3 Player 1 guarantees $\Lambda(p)$

This part is devoted to the proof for

Proposition 5.4.9. *Player 1 guarantees $\Lambda(p)$ in Γ_∞.*

We present our proof in three parts.
A) First, we establish and study the "equalizing" property (cf. Prop. 5.4.10) of the limit game $(\hat{\Xi}(p), A^*_\cdot)$. This generalizes the result in Sorin [55] (cf. Prop. VIII.4.6, Mertens *et al.* [35]).
B) Next, we construct the ε-optimal strategies for player 1 in Γ_∞. In Sorin [55], the ε-optimal strategies are constructed upon a strategy pair having the "equalizing" property. Here, we take an iteration of such construction for M times, where at each time, the "equalizing" property in the limit game $\hat{\Xi}_m$ is used.
C) Finally we conclude by computing the expected averaging payoff.

A) Preliminaries: the "equalizing" property and its discrete approximation

The following proposition states that in the limit game, there is a strategy pair with its associated payoff being constant on $[0, 1]$.

Proposition 5.4.10. *Assume that for each $j \in J$, the function $A^*_\cdot(j)$ defined on $\Delta(K)$ is concave and C-Lip. For any $\mu \in \mathcal{Q}^K[0]$ that is optimal for player 1 in $(\hat{\Xi}(p), A^*_\cdot)$, there exists some $f \in F'[0]$ such that*

$$\varphi^p_t(\mu, f) = \hat{w}(p), \ \forall t \in [0, 1).$$

The proof for this proposition is close to the case of BM with one-sided incomplete information (cf. Prop.VIII.4.6. in Mertens *et al.* [35]). For the sake of completeness, we give it here and put in the **Appendix**.

Fix now a strategy pair (μ, f) with the "equalizing" property as in Proposition 5.4.10. For the aim of obtaining a uniform bound on the error term, we introduce below the discrete approximation. See also Merten *et al.* [35] (VIII.4, p.463) and Sorin [55] (Lemma 28) for results in BM with one-sided incomplete information.

Lemma 5.4.11. *For any $\varepsilon > 0$, we fix an $\omega_\sharp < 1$ with $\mu([\omega_\sharp, 1)) \leq \varepsilon$. There exists a partition $\{\omega_\ell\}$ of $[0, 1]$ such that the following conditions are satisfied:*

1. *$0 = \omega_0 < \cdots < \omega_L < \omega_{L+1} = 1$ with $\omega_L = \omega_\sharp$.*
2. *each interval has small length: $\omega_\ell - \omega_{\ell-1} \leq \varepsilon$ for $\ell = 1, ..., L + 1$.*
3. *each interval has small weight: $\mu((\omega_{\ell-1}, \omega_\ell)) \leq (1 - \omega_\sharp)\varepsilon$ for $\ell = 1, ..., L$.*
4. *$|f(t) - f(t')| \leq \varepsilon(1 - \omega_\sharp)$ for all $t, t' \in (\omega_{\ell-1}, \omega_\ell]$ with $\ell = 1, ..., L$.*
5. *atomic point is put into the set of partition points: $\omega \in \{\omega_r\}$ whenever $\mu(\{\omega\}) > \varepsilon$.*

Proof. $f \in F'[0]$ is continuous, so for any $\varepsilon > 0$, there is some δ in $(0, \varepsilon]$ such that $|f(t) - f(t')| \leq (1 - \omega_\sharp)\varepsilon$ for all $|t - t'| \leq \delta$. We take then a sufficiently fine partition $\{\omega_\ell\}_{\ell=0}^L$ of $[0, \omega_\sharp]$ such that: $\omega_\ell - \omega_{\ell-1} \leq \delta$ and $\mu((\omega_{\ell-1}, \omega_\ell)) \leq (1 - \omega_\sharp)\varepsilon$; whenever there is an atomic point, i.e. $\mu(\{\omega\}) > \varepsilon$, let $\omega \in \{\omega_\ell\}$. The partition $\{\omega_\ell\}_{\ell=0}^{L+1}$ satisfies then all these conditions. □

We define now a pair $(\hat{\mu}, \hat{f}) \in \mathcal{Q}^K[0] \times F[0]$ as a discrete approximation of (μ, f) w.r.t. the partition $\{\omega_\ell\}$:

- $\forall k \in K$, $\hat{\mu}^k(\{\omega_1\}) = \hat{\mu}^k([0, \omega_1])$ and $\hat{\mu}^k(\{\omega_\ell\}) = \mu^k((\omega_{\ell-1}, \omega_\ell])$ for $\ell = 2, ..., L + 1$;
- $\hat{f}(\cdot)$ is piece-wise constant: $\hat{f}(t) = f(\omega_\ell)$ for $t \in (\omega_{\ell-1}, \omega_\ell]$ with $\ell = 1, ..., L$ and $\hat{f}(t) = f(\omega_\sharp)$ for $t \in (\omega_L, 1]$.

We see that the "equalizing" property of (μ, f) is approximately preserved by $(\hat{\mu}, \hat{f})$.

Lemma 5.4.12. *For any $t \in (0, \omega_\sharp]$,*

$$\left|\varphi_t^p(\hat{\mu}, \hat{f}) - \varphi_t^p(\mu, f)\right| \leq 3C(1-\omega_\sharp)\varepsilon.$$

This implies in particular that

$$\left|\varphi^p(\hat{\mu}, \hat{f}) - \hat{w}(p)\right| \leq 5C(1-\omega_\sharp)\varepsilon.$$

Further, there exists some $\hat{\omega}_\sharp \in (\omega_\sharp, 1)$ such that for all $t \in (\hat{\omega}_\sharp, \omega_\sharp)$:

$$\left|\varphi_t^p(\hat{\mu}, \hat{f}) - \varphi_t^p(\mu, f)\right| \leq 3C\varepsilon.$$

Proof. Consider first any $t \in (0, \omega_\sharp]$ and let $t \in (\omega_{\ell-1}, \omega_\ell]$ for some $\ell = 1, ..., L$. We look at $|\varphi_t^p(\hat{\mu}, \hat{f}) - \varphi_t^p(\mu, f)|$. Indeed, there are two differences: either the absorption appears in $(\omega_{\ell-1}, \omega_\ell)$ under $\mu(\cdot)$, which has probability at most $(1-\omega_\sharp)\varepsilon$ according to *Point. 3* in Lemma 5.4.12; otherwise, the difference is bounded by $\sup_{1\leq \ell' \leq \ell} \sup_{s\in(\omega_{\ell'-1}, \omega_{\ell'})} \|_1 f(s) - f(\omega_{\ell'})\|_1 C$, which is at most by $C(1-\omega_\sharp)\varepsilon$ following *Point. 4* in Lemma 5.4.12. We obtain thus

$$\left|\varphi_t^p(\hat{\mu}, \hat{f}) - \varphi_t^p(\mu, f)\right| \leq (1-\omega_\sharp)\varepsilon 2C + C(1-\omega_\sharp)\varepsilon = 3C(1-\omega_\sharp)\varepsilon.$$

Now we consider the points close to $\omega_\sharp$ from the right. Take $\hat{\omega}_\sharp \in (\omega_\sharp, 1)$ with $\|f(t) - f(\omega_\sharp)\|_1 \leq \varepsilon$ for all $t \in (\omega_\sharp, \hat{\omega}_\sharp)$. By definition, $\hat{f}(t) = f(\omega_\sharp)$ for all $t \in (\omega_\sharp, 1]$ and $\mu(\omega_\sharp, 1) \leq \varepsilon$, thus we apply the same argument as above (for $t \leq \omega_\sharp$) to any $t \in (\omega_\sharp, \hat{\omega}_\sharp)$ to obtain [2]

$$\left|\varphi_t^p(\hat{\mu}, \hat{f}) - \varphi_t^p(\mu, f)\right| \leq \varepsilon 2C + C\varepsilon = 3C\varepsilon.$$

□

Notation: For each $k \in K$ and $\ell = 1, ..., L+1$, we write $\hat{\mu}^k(\ell) := \hat{\mu}^k(\{\omega_\ell\})$, $\bar{p}_{\hat{\mu}}(\ell) := \bar{p}_{\hat{\mu}}(\omega_\ell)$ and $\tilde{p}_{\hat{\mu}}(\ell) := \tilde{p}_{\hat{\mu}}(\omega_\ell)$, $\hat{f}(\ell) := \hat{f}(\omega_\ell)$.

Lemma 5.4.13. *Suppose there is some $\hat{\mathfrak{y}} = \{\hat{y}_\ell\}_{\ell=1}^L$ in $\Delta(J)$ such that: for all $\ell \in \{1, ..., L\}$, if*

$$\sum_k p^k \left(1 - \underline{\hat{\mu}}^k(\ell')\right) \langle a^k, \hat{y}_{\ell'}\rangle \leq \sum_k p^k \left(1 - \underline{\hat{\mu}}^k(\ell')\right) \langle a^k, \hat{f}(\ell')\rangle + C(1-\omega_\sharp)\varepsilon \tag{5.4.5}$$

for all $\ell' \in \{1, ..., \ell\}$, then it implies that

$$\sum_{r=1}^{\ell} \hat{\mu}(r)\langle A^*_{\bar{p}_{\hat{\mu}}(r)}, \hat{y}_r\rangle \geq \sum_{r=1}^{\ell} \hat{\mu}(r)\langle A^*_{\bar{p}_{\hat{\mu}}(r)}, \hat{f}(r)\rangle - 5C\varepsilon. \tag{5.4.6}$$

Proof. Suppose the claim is not true, and let $\hat{\mathfrak{y}} = \{\hat{y}_\ell\}_{\ell=1}^L$ such that for some $\ell \in \{1, ..., L\}$, (5.4.5) is satisfied for every $\ell' \in \{1, ..., \ell\}$, while (5.4.6) is not true. So it is possible for us to take some subset $\{\ell_1, ..., \ell_m\}$ of $\{1, ..., \ell\}$ such that

$$\hat{\mu}(\ell_q)\langle A^*_{\bar{p}_{\hat{\mu}}(\ell_q)}, \hat{y}_{\ell_q} - \hat{f}(\ell_q)\rangle < 0 \text{ for each } \ell_q, q = 1, ..., m,$$

2. Note that the error term is now controlled by ε instead of $(1-\omega_\sharp)\varepsilon$. This is due to the fact that $\omega_\sharp$ is chosen after ε. Moreover, the control is only valid for points close to $\omega_\sharp$ from right but not for the whole interval $[\omega_\sharp, 1]$.

and

$$\sum_{q=1}^{m} \hat{\mu}(\ell_q)\langle A^*_{\bar{p}_{\hat{\mu}}(\ell_q)}, \hat{y}_{\ell_q}\rangle < \sum_{q=1}^{m} \hat{\mu}(\ell_q)\langle A^*_{\bar{p}_{\hat{\mu}}(\ell_q)}, \hat{f}(\omega_{\ell_q})\rangle - 5C\varepsilon. \tag{5.4.7}$$

Next, let us define g by modifying the value of f on the points $\{\ell_1, ..., \ell_m\}$ to be $\hat{y}_{\ell_q}$ for each ω_{ℓ_q}. This changes the payoff $\varphi^p(\mu, \cdot)$ (defined from $\hat{f}$ to g) by:

$$\begin{aligned}\varphi^p(\hat{\mu}, g) - \varphi^p(\hat{\mu}, \hat{f}) \leq & \sum_{q=1}^{m} (\omega_{\ell_q} - \omega_{\ell_q - 1}) \sum_k p^k \left(1 - \underline{\hat{\mu}}^k(\ell_q)\right) \left\langle a^k, \hat{y}_{\ell_q} - \hat{f}(\ell_q)\right\rangle \\ & + \sum_{q=1}^{m} (1 - \omega_{\ell_q})\hat{\mu}(\ell_q) \left\langle A^*_{\bar{p}_{\mu}(\ell_q)}, \hat{y}_{\ell_q} - \hat{f}(\ell_q)\right\rangle.\end{aligned} \tag{5.4.8}$$

Let us bound the two partial sums on the right-hand-side of inequality (5.4.8) as follows

– (5.4.5) is used to bound the first part, which yields:

$$\sum_{q=1}^{m} (\omega_{\ell_q} - \omega_{\ell_q - 1}) \sum_k p^k \left(1 - \underline{\hat{\mu}}^k(\ell_q)\right) \left\langle a^k, \hat{y}_{\ell_q} - \hat{f}(\ell_q)\right\rangle \leq C(1 - \omega_\sharp)\varepsilon.$$

– Since $\hat{\mu}(\ell_q) \left\langle A^*_{\bar{p}_{\hat{\mu}}(\ell_q)}, \hat{y}_{\ell_q} - \hat{f}(\ell_q)\right\rangle < 0$ for each ℓ_q, thus we obtain for the first part:

$$\begin{aligned}\sum_{q=1}^{m} (1 - \omega_{\ell_q})\hat{\mu}(\ell_q) \left\langle A^*_{\bar{p}_{\mu}(\ell_q)}, \hat{y}_{\ell_q} - \hat{f}(\ell_q)\right\rangle &\leq (1 - \omega_\sharp) \sum_{q=1}^{m} \mu(\ell_q) \left\langle A^*_{\bar{p}_{\mu}(\ell_q)}, \hat{y}_{\ell_q} - f(\ell_q)\right\rangle \\ (\text{ using } ''(5.4.7)'') &< (1 - \omega_\sharp)(-5C\varepsilon).\end{aligned}$$

We sum the above two parts to obtain in (5.4.8):

$$\varphi^p(\hat{\mu}, g) - \varphi^p(\hat{\mu}, \hat{f}) \leq C(1 - \omega_\sharp)\varepsilon - 5C(1 - \omega_\sharp)\varepsilon = -4C(1 - \omega_\sharp)\varepsilon,$$

thus

$$\varphi^p(\hat{\mu}, g) \leq \varphi^p(\hat{\mu}, \hat{f}) - 4C(1 - \omega_\sharp)\varepsilon \leq \hat{w}(p) - C(1 - \omega_\sharp)\varepsilon,$$

according to Lemma 5.4.12.

Finally, observing that $g(\cdot)$ is piece-wise constant on $(\omega_{\ell-1}, \omega_\ell)$, we obtain

$$\varphi^p(\mu, g) = \varphi^p(\hat{\mu}, g) \leq \hat{w}(p) - C(1 - \omega_\sharp)\varepsilon,$$

which leads to a contradiction to the optimality of μ in $(\hat{\Xi}(p), A^*_{\cdot}(j))$. □

Lemma 5.4.14. *For any $\hat{y} \in \Delta(J)$,*

$$\sum_k p^k \left(1 - \underline{\mu}^k(L+1)\right) \langle a^k, \hat{y}\rangle \geq \sum_k p^k \left(1 - \underline{\mu}^k(L+1)\right) \langle a^k, \hat{f}(L+1)\rangle - 4C\varepsilon.$$

Proof. As otherwise, we define g to be $\hat{f}$ on intervals $(\omega_{\ell-1}, \omega_\ell]$ for $\ell = 1, ..., L$ and to be $\hat{y}$ on $(\omega_\sharp, 1]$. The induced difference in payoff is then

$$\varphi^p(\hat{\mu}, g) - \varphi^p(\hat{\mu}, \hat{f}) = (1 - \omega_\sharp) \sum_k p^k \left(1 - \underline{\mu}^k(L+1)\right) \langle a^k, \hat{y} - \hat{f}(L+1)\rangle < -4C(1 - \omega_\sharp)\varepsilon.$$

Considering $|\varphi^p(\hat{\mu}, \hat{f}) - \hat{w}(p)| \leq 3C(1 - \omega_\sharp)\varepsilon$ and $\varphi^p(\hat{\mu}, g) = \varphi^p(\mu, g)$ as $g(\cdot)$ is piece-wise constant, we obtain a contradiction to the optimality of μ in $(\hat{\Xi}(p), A^*_{\cdot}(j))$. □

B) ε-optimal strategies

B.1) BM associated with the payoff ψ and the level z

We first recall here the properties of some ε-optimal strategy α in BM corresponding to a (non-absorbing) payoff $\psi \in \mathbb{R}^J$ and a level $z \in [-C, +C]$.

Notation $\bar{j}_n := \frac{\sum_{1\leq s\leq n} \delta_{j_s}}{n}$ denotes for player 2's average empirical frequency until stage n.

Proposition 5.4.15. *Consider the complete information BM with non-absorbing payoff vector ψ and the level z. For any $\varepsilon, \eta > 0$, there exits some $N_0 \in \mathbb{N}$ and some behavior strategy α for player 1 such that for all $n \geq N_0$ and for any of player 2's strategy τ:*

$$\langle \psi, \bar{j}_n \rangle \leq z - C\varepsilon \Longrightarrow \mathbb{P}_{\alpha,\tau}(\tilde{T} \leq n) \geq 1 - \varepsilon \quad \& \quad \left(\mathbb{E}_{\alpha,\tau}\left[\langle \psi, j_{\tilde{T}} \rangle | \tilde{T} \leq n\right] - z\right) \mathbb{P}_{\alpha,\tau}(\tilde{T} \leq n) \leq \eta\varepsilon,$$

where $\tilde{T}$ denotes the random time of playing Top.

For a reference, see Mertens *et al.* [35] (Equation (8)-(9), p.381) or Sorin [55] (Prop. 29).

B.2) The generic behavior strategy $\bar{\sigma}[\hat{\mu}; \hat{f}]$ associated with an "equalizing" pair

Take $(\bar{\mu}, \bar{f})$ an "equalizing" pair for the limit game $\hat{\Xi}(p)$ satisfying Proposition 5.4.10. We fix a partition $\{\omega_\ell\}_{\ell=1}^L$ satisfying Lemma 5.4.12 and $(\hat{\mu}, \hat{f})$ the corresponding discrete approximation.

We are going to introduce a generic behavior strategy $\bar{\sigma} := \bar{\sigma}[\hat{\mu}; \hat{f}]$ associated with $(\hat{\mu}, \hat{f})$. Indeed, when Γ_∞ is the model of BM with one-sided incomplete information (i.e., $M = 1$), $\bar{\sigma}$ guarantees $\Lambda(p)$ (cf. Mertens *et al.* [35], sec.VIII.4). In our model, the construction of the ε-optimal strategies will be an iteration of $\bar{\sigma}$ for M times.

For any $\ell = 1..., L+1$, let

$$\bar{\psi}_\ell := \left(1 - \underline{\hat{\mu}}(\ell)\right) a_{\tilde{p}_{\hat{\mu}}(\ell)} \in \mathbb{R}^J \quad \text{and} \quad \bar{z}_\ell := \langle \bar{\psi}_\ell, f(\ell) \rangle = \left(1 - \underline{\hat{\mu}}(\ell)\right) \langle a_{\tilde{p}_{\hat{\mu}}(\ell)}, \hat{f}(\ell) \rangle \in [-C, +C]$$

For $\ell = 1, ..., L$, we consider the complete information BM game associated with the non-absorbing payoff vector $\bar{\psi}_\ell$ and the level $\bar{z}_\ell$, and let $\bar{\alpha}_\ell := \bar{\alpha}_\ell[\bar{\psi}_\ell, \bar{z}_\ell]$ be an ε-optimal strategy (with $\bar{N}_\ell \in \mathbb{N}$ the corresponding uniform stage bound) satisfying Proposition 5.4.15.

For each k, $\bar{\sigma}(k, \cdot)$ is to: first select $\ell^* \in \{1, ..., L+1\}$ according to the measure $\hat{\mu}^k(\cdot)$, and then play $Bottom$ until the random stage $\bar{\theta}_{\ell^*}$, on which to play Top with probability 1, where the stopping times $\{\bar{\theta}_\ell\}$ are defined by the following inductive procedure:

- for $\ell = 1, ..., L$, $(\bar{\theta}_\ell - \bar{\theta}_{\ell-1})$ follows the law of $\tilde{T}$ induced by $\bar{\alpha}_\ell$ (set by convention $\bar{\theta}_0 = 0$).
- For $\ell = L+1$, set $\bar{\theta}_{L+1} = \infty$, that is, $\bar{\sigma}(k, \cdot)$ is to play $Bottom$ *i.i.d.* forever if $\ell^* = L+1$.

B.3) ε-optimal strategy σ^* taking the generic form $\bar{\sigma}$ in sequentially M times

Let $M \in \mathbb{N}$. We consider the recursive family $\{\hat{\Xi}_m | 1 \leq m \leq M\}$, and for each $\hat{\Xi}_{M-m+1}(\bar{p})$, let $(\mu_m(\bar{p}), f^m(\bar{p}))$ be an "equalizing" pair satisfying Proposition 5.4.10. Moreover, we fix $(\hat{\mu}_m(\bar{p}), \hat{f}^m(\bar{p}))$ its discrete approximation w.r.t. the partition $\{\omega_\ell^m\}_{\ell=1}^{L_m}$ which satisfies the conditions in Lemma 5.4.12.

For $m = 1, ..., M$, define $\sigma(m)$ to be: for any $h_{T_{m-1}} = (\omega_1, i_1, j_1, ..., \omega_{T_{m-1}}, i_{T_{m-1}}, j_{T_{m-1}})$,

$$\sigma(m; h_{T_{m-1}})[k, \cdot] = \bar{\sigma}\left[\hat{\mu}_m(\mathbf{q}_{\mathrm{m}-1}); \hat{f}^m(\mathbf{q}_{\mathrm{m}-1})\right][k, \cdot], \ \forall k \in K$$

where $\mathbf{q}_{\mathrm{m}-1}$ is the *posterior* of p conditional on the history "$h_{T_{m-1}}$ & $\theta > T_{m-1}$".

Finally, we define $\sigma^* := \sigma[\hat{\mu}; \hat{f}] = \sigma(1) \odot_{T_1} \sigma(2) \cdots \odot_{T_{M-1}} \sigma(M)$ as the concatenations of $\sigma(1), ..., \sigma(M)$ at the random times $T_1, ..., T_{M-1}$. To precise, we introduce the following notation. For any two histories $h_{n'} = (\omega_1, i_1, j_1, ..., \omega_{n'}, i_{n'}, j_{n'})$ and $h_n = (h_{n'}, \omega_{n'+1}, i_{n'+1}, j_{n'+1}, ..., \omega_n, i_n, j_n)$, we write $h_n/h_{n'} = (\omega_{n'+1}, i_{n'+1}, j_{n'+1}, ..., \omega_n, i_n, j_n)$. Then for all $(k, h_n) \in K \times (\Omega \times I \times J)^n$:

$$\sigma(1) \odot_{T_1} \sigma(2)[k, h_n] = \begin{cases} \sigma(1)[k, h_n] \text{ if } n \leq T_1; \\ \sigma(2; h_{T_1})[k, h_n/h_{T_1}] \text{ if } n > T_1. \end{cases}$$

Below we detail a bit more the behavior strategy σ^* and introduce some more notations for further use.

Notation 5.4.16. *For each $m \in \{1, ..., M\}$ and $\ell \in \{1, ..., L_m\}$, define $\tilde{\mathbf{q}}_{\mathrm{m}-1}(\ell)$ as:*

$$\tilde{\mathbf{q}}^{\mathrm{k}}_{\mathrm{m}-1}(\ell) = \frac{\mathbf{q}^{\mathrm{k}}_{\mathrm{m}-1}(1 - \underline{\hat{\mu}}^k_m(\mathbf{q}_{\mathrm{m}-1})(\ell)}{\sum_k \mathbf{q}^{\mathrm{k}}_{\mathrm{m}-1}(1 - \underline{\hat{\mu}}^k_m(\mathbf{q}_{\mathrm{m}-1})(\ell)}, \ k \in K.$$

Let α^m_ℓ be the ε-optimal strategy in the complete information BM associated with the following non-absorbing payoff vector and level (N^m_r is the corresponding uniform stage number):

$$\psi^m_\ell = (1 - \underline{\hat{\mu}}_m(\mathbf{q}_{\mathrm{m}-1})(\ell)) a_{\tilde{\mathbf{q}}_{\mathrm{m}-1}(\ell)},$$

$$z^m_\ell = \left\langle \psi^m_\ell, \hat{f}^m(\mathbf{q}_{\mathrm{m}-1})(\ell) \right\rangle = (1 - \underline{\hat{\mu}}_m(\mathbf{q}_{\mathrm{m}-1})(\ell)) \left\langle a_{\tilde{\mathbf{q}}_{\mathrm{m}-1}(\ell)}, \hat{f}^m(\mathbf{q}_{\mathrm{m}-1})(\ell) \right\rangle.$$

Let $\mathcal{P}^m = \{T_{m-1} + 1, ..., T_m\}$ (set $T_0 = 0$) be the m-th *phrase* on which σ^* follows $\sigma(m)$, that is, $(\alpha^m_\ell)^{L_m}_{\ell=1}$ is sequentially played until the random stage T_{m+1}. We denote (set $\theta^m_0 = T_{m-1}$)

- θ^m_ℓ for the stopping time of the block $B^m_\ell = \{\theta^m_{\ell-1} + 1, ..., \theta^m_\ell\}$, that is, $\theta^m_\ell - \theta^m_{\ell-1}$ follows the law of $\tilde{T}$ under α^m_ℓ;
- $\ell_m \in \{1, ..., L_m\}$ for the index (following $\hat{\mu}^k_m(\mathbf{q}_{\mathrm{m}-1})$ in game k) of the block such that $T_m = \theta^m_{\ell_m} \in B^m_{\ell_m}$.

We see that $\tilde{\mathbf{q}}_{\mathrm{m}-1}(\ell)$ is the *posterior* of $\mathbf{q}_{\mathrm{m}-1}$ conditional on "$T_m > \theta^m_\ell$". Notice that the *posterior* beliefs depend on a history h_{T_m} only through $(\omega_{T_1}, T_1, j_{T_1}, ..., \omega_{T_m}, T_m, j_{T_m})$.

C) Conclusion of the proof: computing the expected average payoff

For each $n \in \mathbb{N}$, $m \in \{1, ..., M\}$ and $\ell \in \{1, ..., L_m\}$: let $\hat{\theta}^m_\ell = \min\{\theta^m_\ell, n\}$, $\hat{B}^m_\ell = \{\hat{\theta}^m_{\ell-1} + 1, ..., \hat{\theta}^m_\ell\} = B^m_\ell \cap \{1, ..., n\}$. Put $L_m(n) = \max\{1 \leq \ell < L_m | \theta^m_\ell \leq n\}$, then $n \in \hat{B}^m_{L_m(n)+1}$.

Notation 5.4.17. *For any $\ell^m_{\hat{\jmath}} = (\ell_1, ..., \ell_m, \hat{\jmath}_1, ..., \hat{\jmath}_m) \in \times_{m'=1}^{m}\{1, ..., L_{m'}\} \times J^m$, let (set $\mathcal{A}_0 = \emptyset$)*

$$\mathcal{A}_m := \mathcal{A}(\ell^m_{\hat{\jmath}}) \text{ be the event } '' \ T_{m'} \in B^{m'}_{\ell_{m'}},\ j_{T_{m'}} = \hat{\jmath}_{m'}, \forall m' \in \{1, ..., m\} \ \& \ \theta > T_m \ ''.$$

$\mathbb{E}^p_{\sigma^*,\tau}[j_{T_m}|\mathcal{A}_{m-1}]$ is the conditional expectation of player 2's mixed move at the random T_m and we use $j(B) := \sum_{t\in B} \delta_{j_t}/(\sharp B)$ to denote for player 2's empirical frequency on a block B.

Fix now any $\varepsilon > 0$ and we take as in Proposition 5.4.15 the parameter $\eta = C(1-\omega^m_\sharp)\varepsilon$ for each $\bar{\alpha}^m_\ell$. Let $\bar{N}^{m+1} = \max_{\ell\in\{1,...,L_m\}} N^{m+1}_\ell < \infty$.

During each phrase $\hat{\mathcal{P}}^m$, player 2's action (empirical frequency) is "blocked" by $\sigma(m)$, which is summarized in the following preliminary proposition (cf. Mertens *et al.* [35], sec. VIII.4 for the model of BM with incomplete information).

Notation 5.4.18. *Let $\mathbf{q}^-_{\mathrm{m}}$ be the posterior of p conditional on the event $''\mathcal{A}^-(\ell^m_{\hat{\jmath}}) := \mathcal{A}(\ell^{m-1}_{\hat{\jmath}})\ \&\ T_m \in B^m_{\ell_m}\ ''$ for given $\ell^{m-1}_{\hat{\jmath}} = (\ell_1, ..., \ell_{m-1}, \hat{\jmath}_1, ..., \hat{\jmath}_{m-1}) \in \times_{m'=1}^{m-1}\{1, ..., L_{m'}\} \times J^{m-1}$ and $\ell_m \in \{1, ..., L_m\}$. We obtain*

$$\mathbf{q}^{-\mathrm{k}}_{\mathrm{m}} = \frac{\mathbf{q}^{\mathrm{k}}_{\mathrm{m-1}}\hat{\mu}^k_m(\mathbf{q}^{\mathrm{k}}_{\mathrm{m-1}})(\ell_m)}{\sum_k \mathbf{q}^{\mathrm{k}}_{\mathrm{m-1}}\hat{\mu}^k_m(\mathbf{q}^{\mathrm{k}}_{\mathrm{m-1}})(\ell_m)},\ \forall k \in K.$$

We denote throughout the rest of the proof:

$$A^*_{\mathbf{q}^-_{\mathrm{m}}}(j) := A^{*,M-m+1}_{\mathbf{q}^-_{\mathrm{m}}}(j),\ \forall j \in J \quad \text{and} \quad \hat{\mu}_{\mathbf{m}} := \hat{\mu}_m(\mathbf{q}_{\mathrm{m-1}}),\ \hat{f}^{\mathbf{m}} := \hat{f}^m(\mathbf{q}_{\mathrm{m-1}}).$$

Proposition 5.4.19. *For any $n \in \mathbb{N}, m \in \{1, .., M\}$ and $\ell^{m-1}_{\hat{\jmath}} \in \times_{m'=1}^{m-1}\{1, ..., L_{m'}\}\times J^{m-1}$,*

$$\begin{aligned}&\mathbb{E}^p_{\sigma^*,\tau}\left[\sum_{\ell=0}^{L_m(n)} (\sharp\hat{B}^m_{\ell+1})\left(\sum_{\ell_m=1}^{\ell} \hat{\mu}_{\mathbf{m}}(\ell_m)\langle A^*_{\mathbf{q}^-_{\mathrm{m}}}(j), j_{T_m}\rangle + \left(1 - \underline{\hat{\mu}}_{\mathbf{m}}(\ell+1)\right)\left\langle a_{\tilde{\mathbf{q}}_{\mathrm{m-1}}(\ell+1)}, j(\hat{B}^m_{\ell+1})\right\rangle\right)\middle|\mathcal{A}(\ell^{m-1}_{\hat{\jmath}})\right]\\ \geq&(n - T_{m-1})\Big[\hat{w}_{M-m+1}(\mathbf{q}_{\mathrm{m-1}}) - 12C\varepsilon\Big] - 2C\bar{N}^m L_m.\end{aligned} \tag{5.4.9}$$

Proof. We consider any history with $n > T_{m-1}$ as otherwise $\sharp\hat{\mathcal{P}}^m = 0$ and the conditional expectation on it can be defined arbitrarily. We fix any $\ell^{m-1}_{\hat{\jmath}}$ and write $\mathcal{A}_{m-1}$ for $\mathcal{A}(\ell^{m-1}_{\hat{\jmath}})$.

As σ^* is defined to depend on any $h_{T_{m-1}} \in \mathcal{A}_{m-1}$ only through $\ell^{m-1}_{\hat{\jmath}}$, and we consider any τ to be against σ^*, it is thus with no confusion and w.o.l.g. to write directly $\mathbb{E}^p_{\sigma^*,\tau}[\cdot|\mathcal{A}_{m-1}]$ instead of writing $\mathbb{E}^p_{\sigma^*,\tau}[\cdot|h_{T_{m-1}}]$ and then taking the expectation. We keep this throughout the rest.

i). the expected non-absorbing part defined through $j(\hat{B}^m_{\ell+1})$.

On each block $\hat{B}^m_{\ell+1}$ for $1 \leq \ell + 1 \leq L_m$, the strategy $\alpha^m_{\ell+1} = \alpha^m_{\ell+1}[\psi^m_{\ell+1}, z^m_{\ell+1}]$ is used. Consider any ℓ with $\sharp\hat{B}^m_{\ell+1} \geq \bar{N}^m$: Proposition 5.4.15 applies so we have that with probability at least $1 - \varepsilon$, player 2's empirical frequency $j(\hat{B}^m_{\ell+1})$ satisfies:

$$\left\langle \psi^m_{\ell+1}, j(\hat{B}^m_{\ell+1})\right\rangle \geq z^m_{\ell+1} - C\varepsilon,$$

thus in expectation, we obtain

$$\mathbb{E}^p_{\sigma^*,\tau}\Big[\Big\langle \psi^m_{\ell+1}, j(\hat{B}^m_{\ell+1})\Big\rangle \Big| \mathcal{A}_{m-1}\Big] \geq \; z^m_{\ell+1} - 3C\varepsilon. \tag{5.4.10}$$

For $\ell = L-1$, *Bottom* is played *i.i.d.* on $\hat{B}^m_L$. According to Lemma 5.4.14,

$$\Big\langle \psi^m_L, j(\hat{B}^m_L)\Big\rangle \geq z^m_L - 4C\varepsilon. \tag{5.4.11}$$

ii). <u>the expected absorbing part defined through $\mathbb{E}^p_{\sigma^*,\tau}[j_{T_m}|\mathcal{A}(\ell^{m-1}_{\hat{j}})]$</u>

To calculate the auxiliary "absorbing" payoff on block $\hat{B}^m_{\ell+1}$, we look at the possible auxiliary "absorption" at any previous block (i.e., the action T_m being played on $\hat{B}^m_{\ell_m}$ for $\ell_m = 1, ..., \ell$). For this aim, let us prove that: for each $\ell = 1, ..., L_m(n)$,

$$\mathbb{E}^p_{\sigma^*,\tau}\Big[\sum_{\ell_m=1}^{\ell} \hat{\mu}_{\mathbf{m}}(\ell_m)\Big\langle A^*_{\mathbf{q}^-_{\mathrm{m}}}, j_{T_m}\Big\rangle \Big| \mathcal{A}_{m-1}\Big] \geq \sum_{\ell_m=1}^{\ell} \hat{\mu}_{\mathbf{m}}(\ell_m)\Big\langle A^*_{\mathbf{q}^-_{\mathrm{m}}}, \hat{f}^{\mathbf{m}}(\ell_m)\Big\rangle - 5C\varepsilon \tag{5.4.12}$$

We use Lemma 5.4.13 to derive our result. In fact, let us set in the lemma

$$\hat{\mathfrak{y}} = \{\hat{y}_\ell\}_{\ell=1}^L \text{ with } \hat{y}_\ell := \mathbb{E}^p_{\sigma^*,\tau}[j_{\theta^m_\ell}|T_m \leq n \;\&\; \mathcal{A}_{m-1}].$$

Notice that $\mathbb{E}^p_{\sigma^*,\tau}[j_{\theta^m_\ell}|T_m \leq n \;\&\; \mathcal{A}_{m-1}] = \mathbb{E}^p_{\sigma^*,\tau}[j_{\hat{\theta}^m_\ell}|\mathcal{A}_{m-1}]$ for $\ell \leq L_m(n)$.

Take any $\ell' \in \{1, ..., \ell\}$. By the construction of σ^*, $\alpha^m_{\ell'}$ is played on the block $\hat{B}_{\ell'}$. Since $Prob^p_{\sigma^*,\tau}(\theta^m_{\ell'} \leq n|\mathcal{A}_{m-1}) = 1$ for $\ell' \leq \ell \leq L_m(n)$, (i.e., the block $\hat{B}^m_{\ell'}$ occurs with probability 1 within the first n stages), according to Proposition 5.4.15, player 2's expected frequency at the stopping time $\theta^m_{\ell'}$ satisfies:

$$\mathbb{E}^p_{\sigma^*,\tau}\Big[\langle \psi^m_{\ell'}, j_{\hat{\theta}^m_{\ell'}}\rangle \Big| \mathcal{A}_{m-1}\Big] \leq \bar{z}_{\ell'} + C(1-\omega^m_\sharp)\varepsilon,$$

which is written as

$$(1-\underline{\hat{\mu}}_{\mathbf{m}}(\ell'))\langle a_{\tilde{\mathbf{q}}_{\mathrm{m}-1}(\ell')}, \mathbb{E}^p_{\sigma^*,\tau}[j_{\hat{\theta}^m_{\ell'}}|\mathcal{A}_{m-1}] - \hat{f}^{\mathbf{m}}(\ell')\rangle \leq C(1-\omega^m_\sharp)\varepsilon.$$

Condition in Lemma 5.4.13 is satisfied, so by taking $\ell_m = \ell'$ and $T_m = \hat{\theta}^m_{\ell_m}$, we obtain:

$$\sum_{\ell_m=1}^{\ell} \hat{\mu}_{\mathbf{m}}(\ell_m)\Big\langle A^*_{\mathbf{q}^-_{\mathrm{m}}}, \mathbb{E}^p_{\sigma^*,\tau}[j_{T_m}|\mathcal{A}_{m-1}] - \hat{f}^{\mathbf{m}}(\ell_m)\Big\rangle \geq -5C\varepsilon,$$

which proves (5.4.12).

iii). <u>the conclusion</u>

We put together the expected non-absorbing payoff (computed as in (5.4.10) or (5.4.11)) and the expected auxiliary "absorbing" payoff (computed as in (5.4.12)) to obtain:

$$\begin{aligned}
&\mathbb{E}^p_{\sigma^*,\tau}\Bigg[\sum_{\ell=0}^{L_m(n)} (\sharp\hat{B}^m_{\ell+1})\Big(\sum_{\ell_m=1}^{\ell} \hat{\mu}_{\mathbf{m}}(\ell_m)\Big\langle A^*_{\mathbf{q}^-_{\mathrm{m}}}, j_{T_m}\Big\rangle + \Big(1-\underline{\hat{\mu}}_{\mathbf{m}}(\ell+1)\Big)\Big\langle a_{\tilde{\mathbf{q}}_{\mathrm{m}-1}(\ell+1)}, j(\hat{B}^m_{\ell+1})\Big\rangle\Big)\Big|\mathcal{A}_{m-1}\Bigg] \\
&\geq \mathbb{E}^p_{\sigma^*,\tau}\Bigg[\sum_{\ell=0}^{L_m(n)} (\sharp\hat{B}^m_{\ell+1})\Big(\sum_{\ell_m=1}^{\ell} \hat{\mu}_{\mathbf{m}}(\ell_m)\Big\langle A^*_{\mathbf{q}^-_{\mathrm{m}}}, \hat{f}^{\mathbf{m}}(\ell_m)\Big\rangle \\
&\qquad + \Big(1-\underline{\hat{\mu}}_{\mathbf{m}}(\ell+1)\Big)\Big\langle a_{\tilde{\mathbf{q}}_{\mathrm{m}-1}(\ell+1)}, \hat{f}^{\mathbf{m}}(\ell+1))\Big\rangle - 9C\varepsilon\Big)\Big|\mathcal{A}_{m-1}\Bigg] - 2C\bar{N}^m L_m
\end{aligned} \tag{5.4.13}$$

where the error term $\bar{N}^m L_m 2C$ is due to the payoff on blocks of small length (no larger than $\bar{N}^m$), and there are at most L_m such blocks in total.

Next, we look at the following term on the right hand side of (5.4.13), each of them corresponding to $\sharp\hat{B}^m_{\ell+1}$ for some $\ell+1 \in \{1, ..., L+1\}$:

$$\phi(\ell+1) := \sum_{\ell_m=1}^{\ell} \hat{\mu}_{\mathbf{m}}(\ell_m) \left\langle A^*_{\mathbf{q}^-_{\mathrm{m}}}, \hat{f}^{\mathbf{m}}(\ell_m) \right\rangle + \left(1 - \underline{\hat{\mu}}_{\mathbf{m}}(\ell+1)\right) \left\langle a_{\tilde{\mathbf{q}}_{\mathrm{m}-1}(\ell+1)}, \hat{f}^{\mathbf{m}}(\ell+1)) \right\rangle.$$

Below $\varphi^{\mathbf{q}_{\mathrm{m}-1}}$ denotes the payoff function in $\hat{\Xi}_{M-m+1}(\mathbf{q}_{\mathrm{m}-1})$. Then the above term is:

$$\phi(\ell+1) = \varphi^{\mathbf{q}_{\mathrm{m}-1}}_{\omega^m_{\ell+1}}\left(\hat{\mu}_{\mathbf{m}}, \hat{f}^{\mathbf{m}}\right) \text{ for } \ell+1 = 1, ..., L, \text{ and } \phi(L+1) = \varphi^{\mathbf{q}_{\mathrm{m}-1}}_{\omega^m_\sharp}\left(\hat{\mu}_{\mathbf{m}}, \hat{f}^{\mathbf{m}}\right).$$

The pair $(\hat{\mu}_{\mathbf{m}}, \hat{f}^{\mathbf{m}})$ approximately preserves the "equalizing" property of $(\mu_{\mathbf{m}}, f^{\mathbf{m}}) := (\hat{\mu}^m(\mathbf{q}_{\mathrm{m}-1}), \hat{f}^m(\mathbf{q}_{\mathrm{m}-1}))$. According to Lemma 5.4.12 and Proposition 5.4.10, we obtain: for $\ell+1 = 1, ..., L$ or $\sharp$,

$$\varphi^{\mathbf{q}_{\mathrm{m}-1}}_{\omega^m_{\ell+1}}(\hat{\mu}_{\mathbf{m}}, \hat{f}^{\mathbf{m}}) \leq \varphi^{\mathbf{q}_{\mathrm{m}-1}}_{\omega^m_{\ell+1}}(\mu_{\mathbf{m}}, f^{\mathbf{m}}) - 3C(1-\omega^m_\sharp)\varepsilon = \hat{w}_{M-m+1}(\mathbf{q}_{\mathrm{m}-1}) - 3C(1-\omega^m_\sharp)\varepsilon.$$

The above inequality is put back to (5.4.13) to obtain:

$$\begin{aligned}
&\mathbb{E}^p_{\sigma^*,\tau}\left[\sum_{\ell=0}^{L_m(n)} (\sharp\hat{B}^m_{\ell+1})\left(\sum_{\ell_m=1}^{\ell} \hat{\mu}_{\mathbf{m}}(\ell_m)\left\langle A^*_{\mathbf{q}^-_{\mathrm{m}}}, j_{T_m}\right\rangle + (1-\underline{\hat{\mu}}_{\mathbf{m}}(\ell+1))\left\langle a_{\tilde{\mathbf{q}}_{\mathrm{m}}(\ell+1)}, j(\hat{B}^m_{\ell+1})\right\rangle\right)\middle|\mathcal{A}_{m-1}\right] \\
&\geq \mathbb{E}^p_{\sigma^*,\tau}\left[\sum_{\ell=0}^{L_m(n)} (\sharp\hat{B}^m_{\ell+1})\middle|\mathcal{A}_{m-1}\right]\left(\hat{w}_{M-m+1}(\mathbf{q}_{\mathrm{m}-1}) - 12C\varepsilon\right) - 2C\bar{N}^m L_m \\
&= (n - T_{m-1})\left[\hat{w}_{M-m+1}(\mathbf{q}_{\mathrm{m}-1}) - 12C\varepsilon\right] - 2C\bar{N}^m L_m
\end{aligned}$$

This completes the proof for the proposition. □

Let $\hat{\mathcal{P}}^m = \cup_{\ell=1}^{\ell_m+1} \hat{B}^m_\ell = \{T_{m-1}+1, ..., T_m\} \cap \{1, ..., n\}$. We use the previous result to compute the expected average payoff on each phrase so as to obtain

Proposition 5.4.20.

$$n\gamma^p_n(\sigma^*, \tau) \geq n\hat{w}_M(p) - 2C\sum_{m=1}^{M}(\bar{N}^m L_m) - n12C\varepsilon M - n2C(1-\chi^-)^M. \tag{5.4.14}$$

Proof. We can write

$$n\gamma^p_n(\sigma^*, \tau) = \mathbb{E}_{\sigma^*,\tau}\left[\sum_{m=1}^{M} (\sharp\hat{\mathcal{P}}^m)\gamma^p_{\hat{\mathcal{P}}^m}(\sigma^*, \tau)\right],$$

where $\gamma^p_{\hat{\mathcal{P}}^m}(\sigma^*, \tau)$ is the expected average payoff on phrase $\hat{\mathcal{P}}^m$. We shall calculate the average payoff by phrase from backward. Indeed, let us prove the following recursive result, which includes the claim (5.4.14) of the proposition for $m = 0$:

Lemma 5.4.21. *For any* $\ell_{\hat{\jmath}}^m = (\ell_1, ..., \ell_m, \hat{\jmath}_1, ..., \hat{\jmath}_m) \in \times_{m'=1}^m \{1, ..., L_{m'}\} \times J^m, m \in \{0, ..., M-1\}$*:*

$$\mathbb{E}^p_{\sigma^*,\tau}\Big[\sum_{\hbar=m+1}^{M} (\sharp\hat{\mathcal{P}}^{\hbar})\gamma^p_{\hat{\mathcal{P}}^{\hbar}}(\bar{\sigma},\tau)|\mathcal{A}(\ell_{\hat{\jmath}}^m)\Big]$$
$$\geq (n-T_m)\Big[\hat{w}_{M-m}(\mathbf{q}_{\mathrm{m}}) - 2C\frac{\sum_{\hbar=m+1}^{M}\bar{N}^{\hbar}L_{\hbar}}{(n-T_m)} - 12C\varepsilon(M-m) - 2C\varepsilon(1-\chi^-)^{M-m}\Big]. \tag{5.4.15}$$

Proof for Lemma 5.4.21:

i). <u>To initialize our inductive proof, consider $m = M-1$.</u>

Once Top being played, there is a probability of at least χ^- the state is absorbed. On this sub-event ("$\theta = T_M$"), we use Proposition 5.4.19 to bound the conditional expected payoff:

$$\mathbb{E}^p_{\sigma^*,\tau}\Big[(\sharp\hat{\mathcal{P}}^M)\gamma^{\mathbf{q}_{\mathrm{M}}}_{\hat{\mathcal{P}}^M}(\sigma^*,\tau)\Big|\mathcal{A}_{M-1}\ \&\ \theta = T_M\Big] \geq (\sharp\hat{\mathcal{P}}^M)\Big[\hat{w}_1(\mathbf{q}_{\mathrm{M}-1}) - 2C\bar{N}^M L_M/(\sharp\hat{\mathcal{P}}^M) - 12C\varepsilon\Big].$$

The conditional probability for its complementary event is at most $(1-\chi^-)$, thus in expectation:

$$\mathbb{E}^p_{\sigma^*,\tau}\Big[(\sharp\hat{\mathcal{P}}^M)\gamma^{\mathbf{q}_{\mathrm{M}}}_{\hat{\mathcal{P}}^M}(\sigma^*,\tau)\Big|\mathcal{A}_{M-1}\Big] \geq (\sharp\hat{\mathcal{P}}^M)\Big[\hat{w}_1(\mathbf{q}_{\mathrm{M}-1}) - 2C\bar{N}^M L_M/(\sharp\hat{\mathcal{P}}^M) - 12C\varepsilon - 2C(1-\chi^-)\Big].$$

which proves (5.4.15) for $\sharp\hat{\mathcal{P}}^M = n - T_{M-1}$ conditional on $\mathcal{A}_{m-1}$ and $n \geq T_{M-1}$. This starts our inductive proof on m from backward.

ii) <u>The inductive proof for any $m = M-2, ..., 0$.</u>

For fixed $(\ell_{\hat{\jmath}}^{m-1}, \ell_m)$, we write for short $\mathcal{A}_m^- := \mathcal{A}^-(\ell_{\hat{\jmath}}^{m-1}, \ell_m)$, which is by definition the event

$$''\ T_{m'} \in B^{m'}_{\ell_{m'}},\ j_{T_{m'}} = \hat{\jmath}_{m'}, \forall m' \in \{1, ..., m-1\}\ ,\ \theta > T_{m-1}\ \&\ T_m \in B^m_{\ell_m}\ ''.$$

Suppose that (5.4.15) is proved for some $m \in \{M-1, ..., 1\}$, and we prove the result for $m-1$. We write

$$\begin{aligned}&\mathbb{E}^p_{\sigma^*,\tau}\Big[\sum_{\hbar=m}^{M}(\sharp\hat{\mathcal{P}}^{\hbar})\gamma^{\mathbf{q}_{\mathrm{m}-1}}_{\hat{\mathcal{P}}^{\hbar}}(\sigma^*,\tau)\Big|\mathcal{A}_{m-1}\Big]\\ &= \sum_{\ell_m=1}^{L_m(n)+1}\hat{\mu}_{\mathbf{m}}(\ell_m)\mathbb{E}^p_{\sigma^*,\tau}\Big\{(\sharp\hat{\mathcal{P}}^m)\gamma^{\mathbf{q}_{\mathrm{m}}^-}_{\hat{\mathcal{P}}^m}(\sigma^*,\tau) + \Big[\sum_{\hbar=m+1}^{M}(\sharp\hat{\mathcal{P}}^{\hbar})\gamma^{\mathbf{q}_{\mathrm{m}}^-}_{\hat{\mathcal{P}}^{\hbar}}(\sigma^*,\tau)\Big]\Big|\mathcal{A}_m^-\Big\}.\end{aligned} \tag{5.4.16}$$

Let us compute the above sum in two parts separately: the payoffs on $\hat{\mathcal{P}}^m$ and after that.

Conditional on $\mathcal{A}_m^-$ for each ℓ_m, the payoff on a block $\hat{B}^m_\ell, \ell \leq \ell_m$ is non-absorbing, thus the expected payoff sum on $\hat{\mathcal{P}}^m$ writes as

$$\sum_{\ell_m=1}^{L_m(n)+1}\hat{\mu}_{\mathbf{m}}(\ell_m)\mathbb{E}^p_{\sigma^*,\tau}\Big[(\sharp\hat{\mathcal{P}}^m)\gamma^{\mathbf{q}_{\mathrm{m}}^-}_{\hat{\mathcal{P}}^m}(\sigma^*,\tau)|\mathcal{A}_m^-\Big] = \sum_{\ell_m=1}^{L_m(n)+1}\hat{\mu}_{\mathbf{m}}(\ell_m)\mathbb{E}^p_{\sigma^*,\tau}\Big[\sum_{\ell=0}^{\ell_m-1}(\sharp\hat{B}^m_{\ell+1})\langle a_{\mathbf{q}_{\mathrm{m}}^-}, j(\hat{B}^m_{\ell+1})\rangle|\mathcal{A}_m^-\Big]. \tag{5.4.17}$$

Now we compute the expected payoff sum after $\hat{\mathcal{P}}^m$. We fix any $\ell_m \in \{\ell_{m-1}+1, ...L_m(n)+1\}$, and consider the expectation conditional on $\mathcal{A}_m^-$. For any realization of j_{T_m}, we first take the expectation conditional on the sub-event "$\theta = T_m$" or "$\theta > T_m$", to yield:

$$\begin{aligned}
&\mathbb{E}^p_{\sigma^*,\tau}\left\{\sum_{\hbar=m+1}^{M}(\sharp\hat{\mathcal{P}}^{\hbar})\gamma^{\mathbf{q}_{\mathrm{m}}^-}_{\hat{\mathcal{P}}^{\hbar}}(\sigma^*,\tau)\Big|\mathcal{A}_m^-\right\}\\
&= \mathbb{E}^p_{\sigma^*,\tau}\left\{(n-T_m)\chi^{\mathbf{q}_{\mathrm{m}}^-}(j_{T_m})\,a^*_{\hat{\mathbf{q}}_{\mathrm{m}}}(j_{T_m}) + \Big(1-\chi^{\mathbf{q}_{\mathrm{m}}^-}(j_{T_m})\Big)\mathbb{E}^p_{\bar{\sigma},\tau}\Big[\sum_{\hbar=m+1}^{M}(\sharp\hat{\mathcal{P}}^{\hbar})\gamma^{\mathbf{q}_{\mathrm{m}}}_{\hat{\mathcal{P}}^{\hbar}}(\sigma^*,\tau)|\mathcal{A}_m\Big]\Big|\mathcal{A}_m^-\right\},
\end{aligned} \tag{5.4.18}$$

where $\hat{\mathbf{q}}_{\mathrm{m}}$ denotes the *posterior* of p conditional on "$\mathcal{A}(\ell_{\hat{\jmath}}^{m-1}), j_{T_m}, T_m \in \hat{B}^m_{\ell_m}$ & $\theta = T_m$".

The inductive assumption for m applies for the part in Equation (5.4.18) conditional on $\mathcal{A}_m$:

$$\mathbb{E}^p_{\bar{\sigma},\tau}\Big[\sum_{\hbar=m+1}^{M}(\sharp\hat{\mathcal{P}}^{\hbar})\gamma^{\mathbf{q}_{\mathrm{m}}}_{\hat{\mathcal{P}}^{\hbar}}(\sigma^*,\tau)|\mathcal{A}_m\Big] \geq (n-T_m)\big[\hat{w}_{M-m}(\mathbf{q}_{\mathrm{m}}) - \mathbf{e}_{\mathbf{m}}\big]$$

where the error term is denoted as

$$\mathbf{e}_{\mathbf{m}} := 2C\frac{\sum_{\hbar=m+1}^{M}\bar{N}^{\hbar}L_{\hbar}}{n-T_m} + 12C\varepsilon(M-m) + 2C(1-\chi^-)^{M-m}.$$

This implies that in Equation (5.4.18)

$$\begin{aligned}
&\mathbb{E}^p_{\sigma^*,\tau}\Big[\sum_{\hbar=m+1}^{M}(\sharp\hat{\mathcal{P}}^{\hbar})\gamma^{\mathbf{q}_{\mathrm{m}}^-}_{\hat{\mathcal{P}}^{\hbar}}(\bar{\sigma},\tau)|\mathcal{A}_m^-\Big]\\
=&\mathbb{E}^p_{\sigma^*,\tau}\Big[(n-T_m)\chi^{\mathbf{q}_{\mathrm{m}}^-}(j_{T_m})a^*_{\hat{\mathbf{q}}_{\mathrm{m}}}(j_{T_m}) + (n-T_m)\Big(1-\chi^{\mathbf{q}_{\mathrm{m}}^-}(j_{T_m})\Big)\big[\hat{w}_{M-m}(\mathbf{q}_{\mathrm{m}}) - \mathbf{e}_{\mathbf{m}}\big]\Big|\mathcal{A}_m^-\Big]\\
\geq&\ \mathbb{E}^p_{\sigma^*,\tau}\Big[(n-T_m)\big[A^*_{\mathbf{q}_{\mathrm{m}}^-}(j_{T_m}) - (1-\chi^-)\mathbf{e}_{\mathbf{m}}\big]\Big|\mathcal{A}_m^-\Big],
\end{aligned} \tag{5.4.19}$$

where we have used in the last inequality the definition

$$A^*_{\mathbf{q}_{\mathrm{m}}^-}(j_{T_m}) = \chi^{\mathbf{q}_{\mathrm{m}}^-}(j_{T_m})\cdot a^*_{\hat{\mathbf{q}}_{\mathrm{m}}}(j_{T_m}) + \Big(1-\chi^{\mathbf{q}_{\mathrm{m}}^-}(j_{T_m})\Big)\cdot\hat{w}_{M-m}(\mathbf{q}_{\mathrm{m}}).$$

Finally, (5.4.17) and (5.4.19) are substituted back into (5.4.16) to obtain:

$$\begin{aligned}
&\mathbb{E}^p_{\sigma^*,\tau}\left[\sum_{\hbar=m}^{M}(\sharp\hat{\mathcal{P}}^{\hbar})\gamma^{\mathbf{q}_{\mathrm{m}-1}}_{\hat{\mathcal{P}}^{\hbar}}(\sigma^*,\tau)\Big|\mathcal{A}_{m-1}\right]\\
\geq&\sum_{\ell_m=1}^{L_m(n)+1}\hat{\mu}_{\mathbf{m}}(\ell_m)\mathbb{E}^p_{\sigma^*,\tau}\Big[\sum_{\ell=0}^{\ell_m+1}(\sharp\hat{B}^m_{\ell+1})\Big\langle a_{\mathbf{q}_{\mathrm{m}}^-}, j(B^m_{\ell+1})\Big\rangle + (n-T_m)\big[A^*_{\mathbf{q}_{\mathrm{m}}^-}(j_{T_m}) - (1-\chi^-)\mathbf{e}_{\mathbf{m}}\big]|\mathcal{A}_m^-\Big]
\end{aligned} \tag{5.4.20}$$

Let us look the right hand side of (5.4.20) aside the error term $(n-T_m)(1-\chi^-)\mathbf{e}_{\mathbf{m}}$. By construction, the block $\hat{B}^m_\ell = \{\hat{\theta}^m_{\ell-1}+1, ..., \hat{\theta}^m_\ell\}$ is computed as if $T_m = \infty$ thus

$\sum_{\ell=\ell_m}^{L_m(n)}(\sharp\hat{B}^m_{\ell+1}) = n - T_m$ under every history. This enables us to write

$$\begin{aligned}
&\sum_{\ell_m=1}^{L_m(n)+1}\hat{\mu}_{\mathbf{m}}(\ell_m)\mathbb{E}^p_{\sigma^*,\tau}\Bigg\{\sum_{\ell=0}^{\ell_m-1}(\sharp\hat{B}^m_{\ell+1})\Big\langle a_{\mathbf{q}^-_{\mathrm{m}}}, j(\hat{B}^m_{\ell+1})\Big\rangle + (n-T_m)A^*_{\mathbf{q}^-_{\mathrm{m}}}(j_{T_m})\Big|\mathcal{A}^-_m\Bigg\}\\
=&\sum_{\ell_m=1}^{L_m(n)+1}\hat{\mu}_{\mathbf{m}}(\ell_m)\mathbb{E}^p_{\sigma^*,\tau}\Bigg\{\sum_{\ell=0}^{\ell_m-1}(\sharp\hat{B}^m_{r+1})\Big\langle a_{\mathbf{q}^-_{\mathrm{m}}}, j(\hat{B}^m_{\ell+1})\Big\rangle + \sum_{\ell=\ell_m}^{L_m(n)}(\sharp\hat{B}^m_{\ell+1})A^*_{\mathbf{q}^-_{\mathrm{m}}}(j_{T_m})\Big|\mathcal{A}^-_m\Bigg\}\\
=&\mathbb{E}^p_{\sigma^*,\tau}\Bigg[\sum_{\ell=0}^{L_m(n)}(\sharp\hat{B}^m_{\ell+1})\Bigg(\sum_{\ell_m=1}^{\ell}\hat{\mu}_{\mathbf{m}}(\ell_m)A^*_{\mathbf{q}^-_{\mathrm{m}}}(j_{T_m}) + \Big(1-\underline{\hat{\mu}}_{\mathbf{m}}(\ell+1)\Big)\Big\langle a_{\tilde{\mathbf{q}}_{\mathrm{m}}(\ell+1)}, j(\hat{B}^m_{\ell+1})\Big\rangle\Bigg)\Big|\mathcal{A}_{m-1}\Bigg],
\end{aligned}$$

where to obtain the last equality, we made the sum of the expected payoff by blocks, conditional on T_m before or after each block.

We can apply now Proposition 5.4.19 to the above expression to obtain:

$$\begin{aligned}
&\sum_{\ell_m=1}^{L_m(n)+1}\hat{\mu}_{\mathbf{m}}(\ell_m)\mathbb{E}^p_{\sigma^*,\tau}\Big[\sum_{\ell=0}^{\ell_m-1}(\sharp\hat{B}^m_{r+1})\Big\langle a_{\mathbf{q}^-_{\mathrm{m}}}, j(\hat{B}^m_{r+1})\Big\rangle + (n-T_m)A^*_{\mathbf{q}^-_{\mathrm{m}}}(j_{T_m})\Big|\mathcal{A}_{m-1}\Big]\\
&\geq (n-T_{m-1})\Big[\hat{w}_{M-m+1}(\mathbf{q}_{\mathrm{m}-1}) - 12C\varepsilon\Big] - 2C\bar{N}^m L_m,
\end{aligned}$$

which is substituted back into (5.4.20), adding the error term $(n-T_m)(1-\chi^-)\mathbf{e}_{\mathrm{m}}$, to yield:

$$\begin{aligned}
&\mathbb{E}^p_{\sigma^*,\tau}\Bigg\{\sum_{\hbar=m}^{M}(\sharp\hat{\mathcal{P}}^{\hbar})\gamma^{\mathbf{q}_{\mathrm{m}-1}}_{\hat{\mathcal{P}}^{\hbar}}(\sigma^*,\tau)\Big|\mathcal{A}_{m-1}\Bigg\}\\
\geq&\ (n-T_{m-1})\Big[\hat{w}_{M-m+1}(\mathbf{q}_{\mathrm{m}-1}) - 12C\varepsilon\Big] - 2C\bar{N}^m L_m\\
&-(n-T_m)\Big[2C\frac{\sum_{\hbar=m+1}^{M}\bar{N}^{\hbar}L_{\hbar}}{n-T_m} + 12C\varepsilon(M-m) + 2C(1-\chi^-)^{M-m+1}\Big]\\
\geq&\ (n-T_{m-1})\Big[\hat{w}_{M-m+1}(\mathbf{q}_{\mathrm{m}-1}) - \Big(2C\frac{\sum_{\hbar=m}^{M}\bar{N}^{\hbar}L_{\hbar}}{n-T_{m-1}} + 12C\varepsilon(M-m+1) + 2C(1-\chi^-)^{M-m+1}\Big)\Big]\\
=&(n-T_{m-1})\Big[\hat{w}_{M-m+1}(\mathbf{q}_{\mathrm{m}-1}) - \mathbf{e_{m-1}}\Big].
\end{aligned} \tag{5.4.21}$$

This proves (5.4.15) for $m-1$, thus the induction procedure for Lemma 5.4.21 is finished. □

In particular, our proof for Proposition 5.4.20 is achieved for $m = 0$ in Lemma 5.4.21. □

Proof for Proposition 5.4.9: Proposition 5.4.20 implies that for any $\varepsilon > 0$, there is σ^* and N_0 such that for any τ and $n \geq N_0 := \frac{2C\sum_{m=1}^{M}(\bar{N}^m L_m)}{\varepsilon}$, $M > 0$

$$\gamma^p_n(\sigma^*,\tau) \geq \hat{w}_M(p) - \varepsilon - 13C\varepsilon M - 2C(1-\chi^-)^M.$$

Take $M = \left\lfloor \frac{\ln\varepsilon}{\ln(1-\chi^-)}\right\rfloor + 1$, then $(1-\chi^-)^M \leq \varepsilon$ and $\hat{w}_M(p) \geq \Lambda(p) \geq -\varepsilon$, following the proof for Theorem 5.2.3. Moreover, $\varepsilon M \leq \frac{\varepsilon\ln\varepsilon}{\ln(1-\chi^-)}$, which vanishes as ε tends to zero. As $\varepsilon > 0$ being arbitrary, this completes the proof that player 1 guarantees $\Lambda(p)$ in Γ_∞. □

5.5 Uniform analysis: *Minmax*

This section is devoted for the proof of Theorem 5.2.5. To do this, we first introduce an auxiliary game $\Theta_M(p)$, and then prove that for M large, its value is guaranteed by player 2 and is defended by player 1.

Auxiliary game $\Theta_M(p)$ For any $M > 0$ and$p \in \Delta(K)$, the game $\Theta_M(p)$ is defined as:
- player 1 takes an action $x = \big(M(k)\big)_{k\in K} \in \{0, 1, ..., M\}^K$.
- player 2 takes an action $y = (j_1, ..., j_{M+1}) \in J^{M+1}$.
- the payoff function is $\phi^p(x, y) = \sum_{k\in K} p^k \phi^k(M(k), y)$, where

$$\phi^k(M(k), y) = \sum_{m'=0}^{M(k)} \prod_{s=0}^{m'-1} (1 - \chi^k(j_s))\chi^k(j_{m'})a^{k*}(j_{m'}) + \prod_{s=0}^{M(k)} (1 - \chi^k(j_s))a^k(j_{M(k)+1}).$$

This is a finite game, thus it has a value, which we denote by $u_M(p)$. We show that $\bar{v}(p) = \lim_{M\to\infty} u_M(p)$. As usual, the proof is divided into two parts: player 2 guarantees it and player 1 defends it.

Part I: player 2 guarantees $\liminf_{M\to\infty} u_M(p)$

We fix an $\varepsilon > 0$, and take $M \in \mathbb{N}$ large satisfying

$$(1 - \chi^-)^M \leq \varepsilon \quad \text{and} \quad u_M(p) \leq \liminf_{m\to\infty} u_m(p) + \varepsilon.$$

For any $y = (y_1, ..., y_{M+1}) \in \big(\Delta(J)\big)^{M+1}$ a mixed strategy for player 2 in $\Theta_M(p)$, we define $\tau := \tau[y] \in \mathcal{T}$ a behavior strategy in Γ_∞ to:
- start playing y_1 *i.i.d.* until T_1 the stage of the first *Top*;
- for $m = 1, ..., M - 1$: after $T^m = (T_1, ...T_m)$ the stages of the first m *Top*'s, play y_{m+1} *i.i.d.* until T_{m+1};
- after T_M: play y_{M+1} *i.i.d.* forever.

Consider any $\sigma \in \Sigma$ to play against τ. We are going to construct some $\mu = (\mu^k) := \mu[\sigma; y] \in (\Delta\{0, ..., M\})^K$ in $\Theta_M(p)$ such that for all n sufficiently large,

$$\gamma_n^p(\sigma, \tau[y]) \leq \phi^p(\mu[\sigma; y], y) + 6C\varepsilon.$$

We then choose y to be optimal, and this implies that player 2 guarantee $\liminf_{M\to\infty} u_M(p)$.

Fix now a $\sigma \in \Sigma$. First of all, by definition of M, it is with a loss of $2C\varepsilon$ to assume that *Top* appears at most M times under σ. Then, under this assumption, there is some $\bar{N} \in \mathbb{N}$ such that

$$Prob^k_{\sigma,\tau}(i_n = Top) \leq \varepsilon, \ \forall n > \bar{N}.$$

From now on, we work on the following event $(\mathcal{M})$:

"no Top after $\bar{N}$ and at most M times the action Top within the first $\bar{N}$ stages".

For any play $h = (\omega_1, i_1, j_1, ..., \omega_s, i_s, j_s, ...,) \in H_\infty$, let $T^m[h] = (T_1, ..., T_m)$ be the associated sequence of stages appearing *Top*'s, where $m := m[h] = \sharp\{s \in \mathbb{N} | i_s(h) = Top\}$. We have that: for any $n > \bar{N}$ and $k \in K$,

$$\mathbb{E}^k_{\tau[y]}\Big[g_n | T^m[h]\Big] = \sum_{m'=0}^{m} \prod_{s=0}^{m'-1} (1 - \chi^k(y_s))\chi^k(y_{m'})\overline{\langle a^{k*}, y_{m'}\rangle}_{\chi^k} + \prod_{m'=0}^{m} (1 - \chi^k(y_{m'}))a^k(y_{m+1}),$$

which is equal to $\phi^k(m[h], y)$.

We take expectation over all histories, to obtain:

$$\mathbb{E}^k_{\sigma,\tau[y]}[g_n] = \int_{H_\infty} \mathbb{E}^k_{\tau[y]}[g_n|T^m[h]]\mathrm{d}Prob^k_{\sigma,\tau[y]}(h) = \int_{H_\infty} \phi^k(m[h], y)\mathrm{d}Prob^k_{\sigma,\tau[y]}(h). \quad (5.5.1)$$

Define now a class of probability measures $\mu = (\mu^k) := \mu[\sigma; y]$ with $\mu^k \in \Delta(\{0, ..., M\}), \forall k \in K$:

$$\mu^k(m') = \int_{H_\infty} \mathbb{1}_{\{m[h]=m'\}}\mathrm{d}Prob^k_{\sigma,\tau[y]}(h), \ \forall m' \in \{0, .., M\}.$$

Using the deinifion of $\mu^k(\cdot)$, we re-write Equation (5.5.1) as:

$$\mathbb{E}^k_{\sigma,\tau[y]}[g_n] = \sum_{m'\in\{0,...,M\}} \mu^k(m')\phi^k(m', y) = \phi^k(\mu^k, y)$$

Now we take expectation w.r.t. $p \in \Delta(K)$, to obtain

$$\mathbb{E}^p_{\sigma,\tau[y]}[g_n] = \sum_k p^k\phi^k(\mu^k, y) = \phi^p(\mu[\sigma; y], y), \ \forall n \geq \bar{N}.$$

Finally, we take y to be optimal, and consider any $n \geq \bar{N}/\varepsilon$. This gives us:

$$\gamma^p_n(\sigma, \tau[y]) \leq \phi^p(\mu[\sigma; y], y) + 2C\varepsilon + 4C\varepsilon \leq u_M(p) + 6C\varepsilon \leq \liminf_{m\to\infty} u_m(p) + \varepsilon(1 + 6C),$$

where the error term $4C\varepsilon$ is to bound the payoff outside the event $(\mathcal{M})$. As ε is arbitrary, this proves that player 2 guarantees $\liminf_{M\to\infty} u_M(p)$.

Part II: player 1 defends $\limsup_{M\to\infty} u_M(p)$

Fix an $\varepsilon > 0$ and we take $M > 0$ sufficiently large such that

$$(1 - \chi^-)^M \leq \varepsilon \ \text{ and } \ u_M(p) \geq \limsup_{m\to\infty} u_m(p) - \varepsilon.$$

Given $\tau \in \mathcal{T}$ player 2's behavior strategy in Γ_∞ and $x \in \Delta(\{0, ..., M\}^K)$, we are going to construct some mixed strategy $\sigma := \sigma[x; \tau]$ in Γ_∞ and $y := y[\tau; x] \in (\Delta(J))^{M+1}$ such that: for all n sufficiently large,

$$\gamma^p_n(\sigma, \tau) \geq \phi^p(x, y) - \varepsilon.$$

By choosing x optimal, the result will be obtained.

Fix now a $\tau \in \mathcal{T}$ and $x \in \Delta(\{0, ..., M\}^K)$. We write $x = \sum_{r=1}^R \lambda^r\delta_{x^r}$ for some $(\lambda^r, x^r)_{r=1}^R$ where $x^r = (M^r(k)) \in \{0, ..., M\}^K$ is a pure action for any r and $(\lambda^r)_{r\in R} \in \Delta(\{1, ..., R\})$ is the random devise.

We define, at the same time, a sequence of random times $(T(m))_{m=0}^{M+1}$ and a sequence of pure behavior strategies $(\sigma(m))_{m=0}^{M}$ along the play as follows:

- $T(0) = 0$ and $\sigma(0)$ is to play *Bottom* forever;
- for any $m = 0, ..., M$:
  - $T(m + 1)$ is equal to $T(m) + 1$ if $\{(r, k)|M^r(k) = m\} = \emptyset$ and is otherwise an $\frac{\varepsilon}{M+1}$-optimal solution to for following optimization problem:

$$\min_{n>T(m)} \sum_{(r,k)}^{M^r(k)=m} p^k\lambda^r\Big\langle a^k, \mathbb{E}_{\sigma(m),\tau}[j_n|\mathcal{H}_{T(m)}]\Big\rangle; \quad (5.5.2)$$

– $\sigma(m+1)$ is to follow $\sigma(m)$ until the stage $T(m+1)-1$, to play *Top* at $T(m+1)$ and then to play *Bottom*.

Note that the sequence $(T(m))$ can be taken uniformly bounded, thus we fix $\bar{N}>0$ with $T_{M+1}\leq \bar{N}$ *a.s.* for all histories. For fixed $m\in\{0,...,M\}$ and $n\geq\bar{N}$, we obtain: $\forall k\in K$,

$$\begin{aligned}
&\mathbb{E}^k_{\sigma(m),\tau}[g_n]\\
&=\sum_{0\leq s\leq m} Prob^k_{\sigma(m),\tau}(\theta=T(s))\overline{\langle a^{k*},\mathbb{E}_{\sigma(m),\tau}[j_{T(s)}]\rangle}_{\chi^k}+Prob^k_{\sigma(m),\tau}(\theta>T(m))\langle a^k,\mathbb{E}_{\sigma(m),\tau}[j_n]\rangle\\
&=\sum_{1\leq m'\leq m}\prod_{0\leq s<m'}(1-\chi^k(y_s))\chi^k(y_{m'})\overline{\langle a^{k*},y_{m'}\rangle}_{\chi^k}+\prod_{0\leq s\leq m}(1-\chi^k(y_s))\langle a^k,y^{(n)}_m\rangle,
\end{aligned}\tag{5.5.3}$$

(we write throughout the rest $y_0=\emptyset$ and $\chi^k(\emptyset)=0$)

where we have denoted $y_m:=\mathbb{E}_{\sigma(M),\tau}[j_{T(m)}]$ for each $m=1,...,M+1$ and $y^{(n)}_m:=\mathbb{E}_{\sigma(m),\tau}[j_n]$ for each $m=1,...,M$. As $\sigma(M)$ is a sequence of pure moves, the expectation $\mathbb{E}_{\sigma(M),\tau}[j_{T(m)}]$ is independent of k.

We set $y=(y_1,...,y_{M+1}):=y[\tau;x]\in(\Delta(J))^{M+1}$. Define now the ε-uniform reply strategy $\sigma:=\sigma[x;\tau]\in\Sigma$ to be: with probability λ^r to play the pure moves $\sigma(M^r(k))$ in state k, that is, to play *Top* only on those stages $T(1),...,T(M^r(k))$. We prove the following

Claim 5.5.1. $\mathbb{E}^p_{\sigma[x;\tau],\tau}[g_n]\geq\phi^p(x,y[\tau;x])-\varepsilon$, $\forall n\geq\bar{N}$.

Proof for Claim 5.5.1: We take expectation w.r.t. $(\lambda^r)\in\Delta(R)$ and $p\in\Delta(K)$, and for each pair (r,k), we use the expression in Equation (5.5.3) for $m=M^r(k)$, to obtain:

$$\begin{aligned}
&\mathbb{E}^p_{\sigma[x;\tau],\tau}[g_n]=\sum_{r,k}p^k\lambda^r\mathbb{E}^k_{\sigma(M^r(k)),\tau}[g_n]\\
&=\sum_{r,k}p^k\lambda^r\left\{\sum_{m'=1}^{M^r(k)}\prod_{0\leq s<m'}(1-\chi^k(y_s))\chi^k(y_{m'})\overline{\langle a^{k*},y_{m'}\rangle}_{\chi^k}\right\}+\sum_{r,k}p^k\lambda^r\left\{\prod_{s=0}^{M^r(k)}(1-\chi^k(y_s))\langle a^k,y^{(n)}_{M^r(k)}\rangle\right\}.
\end{aligned}\tag{5.5.4}$$

Let us look at the second part (non-absorbing) of the right hand side of Equation (5.5.4). We re-arrange the terms indexed by (r,k) according to $M^r(k)=m$, for $m=0,...,M$, to obtain:

$$\sum_{r,k}p^k\lambda^r\left\{\prod_{s=0}^{M^r(k)}(1-\chi^k(y_s))\langle a^k,y^{(n)}_{M^r(k)}\rangle\right\}=\sum_{0\leq m\leq M}\sum_{(r,k)}^{M^r(k)=m}p^k\lambda^r\prod_{0\leq s\leq m}(1-\chi^k(y_s))\langle a^k,y^{(n)}_m\rangle.\tag{5.5.5}$$

Next for each $m\in\{0,...,M\}$ with $\sum_{(r,k):\ M^r(k)=m}p^k\lambda^r>0$, we use the definition of $T(m+1)$ in (5.5.2) to obtain: for any $h_{T(m)}\in\mathcal{H}_{T(m)}$,

$$\begin{aligned}
&\sum_{(r,k)}^{M^r(k)=m}p^k\lambda^r\prod_{0\leq s\leq m}(1-\chi^k(y_s))\langle a^k,\mathbb{E}_{\sigma(m),\tau}[j_n|h_{T(m)}]\rangle\\
\geq&\sum_{(r,k)}^{M^r(k)=m}p^k\lambda^r\prod_{0\leq s\leq m}(1-\chi^k(y_s))\langle a^k,\mathbb{E}_{\sigma(m),\tau}[j_{T(m+1)}|h_{T(m)}]\rangle-\frac{\varepsilon}{M+1},
\end{aligned}$$

thus in expectation over all histories, we have

$$\begin{aligned}
&\sum_{(r,k)}^{M^r(k)=m} p^k\lambda^r \prod_{0\le s\le m}(1-\chi^k(y_s))\langle a^k, \mathbb{E}_{\sigma(m),\tau}[j_n]\rangle \\
\ge &\sum_{(r,k)}^{M^r(k)=m} p^k\lambda^r \prod_{0\le s\le m}(1-\chi^k(y_s))\langle a^k, \mathbb{E}_{\sigma(m),\tau}[j_{T(m+1)}]\rangle - \frac{\varepsilon}{M+1} \qquad (5.5.6)\\
= &\sum_{(r,k)}^{M^r(k)=m} p^k\lambda^r \prod_{0\le s\le m}(1-\chi^k(y_s))\langle a^k, y_{m+1}\rangle - \frac{\varepsilon}{M+1}.
\end{aligned}$$

We substitute (5.5.6) back into the right hand side of (5.5.5), by summing over $m = 0, ..., M$, to obtain:

$$\begin{aligned}
&\sum_{r,k} p^k\lambda^r\left\{\prod_{s=0}^{M^r(k)}(1-\chi^k(y_s))\langle a^k, y^{(n)}_{M^r(k)}\rangle\right\} \\
\ge &\sum_{0\le m\le M}\left\{\sum_{(r,k)}^{M^r(k)=m} p^k\lambda^r \prod_{0\le s\le m}(1-\chi^k(y_s))\langle a^k, y_{m+1}\rangle\right\} - \varepsilon \qquad (5.5.7)\\
= &\sum_{r,k} p^k\lambda^r \prod_{s=0}^{M^r(k)}(1-\chi^k(y_s))\langle a^k, y_{M^r(k)+1}\rangle - \varepsilon.
\end{aligned}$$

Finally, we put (5.5.7) into (5.5.4) and use the definition of $\phi^k(\cdot,\cdot)$ to obtain:

$$\begin{aligned}
&\mathbb{E}^p_{\sigma[x;\tau],\tau}[g_n] + \varepsilon \\
\ge &\sum_{r,k} p^k\lambda^r\left\{\sum_{m'=1}^{M^r(k)}\prod_{0\le s<m'}(1-\chi^k(y_s))\chi^k(y_{m'})\overline{\langle a^{k*}, y_{m'}\rangle}_{\chi^k} + \prod_{s=0}^{M^r(k)}(1-\chi^k(y_s))\langle a^k, y_{M^r(k)+1}\rangle\right\} \\
= &\sum_{r,k} p^k\lambda^r\Big\{\phi^k(M^r(k), y)\Big\} = \sum_r \lambda^r\Big\{\sum_k p^k\phi^k(M^r(k), y)\Big\} \\
= &\sum_r \lambda^r\phi^p(x^r, y) = \phi^p(x, y).
\end{aligned}$$

The proof of the claim is then complete. □

Take now $x = (x^r, \lambda^r)$ to be optimal, we obtain from Claim 5.5.1 that: $\forall n \ge \bar{N}/\varepsilon$,

$$\gamma_n(\sigma[x;\tau],\tau) \ge u_M(p) - (2C+1)\varepsilon \ge \limsup_{m\to\infty} u_m(p) - (2C+1)\varepsilon.$$

By our construction, σ is a mixed strategy. According to Kukn's theorem, there exists some behavior strategy which guarantees player 1 the same payoff. This proves that player 1 defends $\limsup_{M\to\infty} u_M(p)$.

To conclude from the results in both **Part I** and **Part II**, we obtain that $\bar{v}(p) = \lim_{M\to\infty} u_M(p)$, thus the proof for Theorem 5.2.5 is achieved.

5.6 Appendix

5.6.1 Proof for Proposition 5.4.4

The proof for **Part A of Proposition 5.4.4** is divided into two parts, Lemma 5.6.1 and Lemma 5.6.2.

Lemma 5.6.1. $\liminf_{L\to\infty} \hat{w}^L(p) \geq \hat{w}(p)$.

Proof for Lemma 5.6.1: Let $\mu = (\mu^k) \in \mathcal{Q}^K[0]$ be optimal to guarantee $\hat{w}(p)$ in $(\hat{\Xi}, A^*_{\cdot}(j))$. For each $L \in \mathbb{N}$, define $\hat{\mu} := \hat{\mu}[\mu; L] = (\hat{\mu}^k) \in \mathcal{Q}^K[L]$ as an atomic approximation of μ w.r.t. $\{\ell/L\}$, i.e. $\hat{\mu}^k(1) = \mu^k([0, 1/L])$ and $\hat{\mu}^k(\ell) = \mu^k\Big(((\ell-1)/L, \ell/L]\Big)$ for all $k \in K$ and $\ell = 2, ..., L$. Consider now any of player 2's strategy $\hat{f} \in F[L]$. We identified $\hat{f}$ with a function $f := f[\hat{f}; L] \in F[0]$ that is piece-wise constant on $((\ell-1)/L, \ell/L]$, i.e. $f(t) = \hat{f}(\ell), \forall t \in ((\ell-1)/L, \ell/L]$.

Notation: Denote ${}^+((\ell-1)/L, \ell/L]$ for $((\ell-1)/L, \ell/L]$ when $\ell = 2, ..., L$ and for $[0, 1/L]$ when $\ell = 1$. Let $\mu(\cdot|[\ell])$ be the conditional probability distribution of μ given "$\theta \in {}^+((\ell-1)/L, \frac{\ell}{L}]$" for $\ell = 1, ..., L$, that is,

$$\forall B \in \mathcal{B}\Big({}^+(\frac{\ell-1}{L}, \frac{\ell}{L}]\Big), \quad \mu(B|[\ell]) = \frac{\mu(B)}{\mu\Big({}^+(\frac{\ell-1}{L}, \ell/L]\Big)} = \frac{\mu(B)}{\hat{\mu}(\ell)}, \quad \text{for } \ell = 1, ..., L.$$

We write

$$\begin{aligned}
\varphi^p(\mu, f) &= \int_0^1 \mu(dt)\Big[\Big\langle A^*_{\bar{p}_\mu(t)}, f(t)\Big\rangle(1-t) + \int_0^t \Big\langle a_{\bar{p}_\mu(t)}, f(s)\Big\rangle ds\Big] \\
&= \sum_{\ell=1}^L \hat{\mu}(\ell) \int_{+\left(\frac{\ell-1}{L}, \frac{\ell}{L}\right]} \Big[\Big\langle A^*_{\bar{p}_\mu(t)}, f(t)\Big\rangle(1-t) + \int_0^t \Big\langle a_{\bar{p}_\mu(t)}, f(s)\Big\rangle ds\Big]\mu\Big(dt|[\ell]\Big) \\
&= \sum_{\ell=1}^L \hat{\mu}(\ell) \int_{+\left(\frac{\ell-1}{L}, \frac{\ell}{L}\right]} \Big[\Big\langle A^*_{\bar{p}_\mu(t)}, \hat{f}(\ell)\Big\rangle(1-t) + \int_0^t \Big\langle a_{\bar{p}_\mu(t)}, f(s)\Big\rangle ds\Big]\mu\Big(dt|[\ell]\Big),
\end{aligned} \tag{5.6.1}$$

and compare it with

$$\hat{\mathcal{L}}^p(\hat{\mu}, \hat{f}) = \sum_{\ell=1}^L \hat{\mu}(\ell)\Big[\langle A^*_{\bar{p}_{\hat{\mu}}(\ell)}, \hat{f}(\ell)\rangle(1 - \frac{\ell}{L}) + \frac{1}{L}\sum_{\ell'=1}^{\ell} \langle a_{\bar{p}_{\hat{\mu}}(\ell)}, \hat{f}(\ell')\rangle\Big].$$

Conditional on "$\theta \in {}^+((\ell-1)/L, \ell/L]$" for each $\ell = 1, ..., L$, we have:
– for the non-absorbing part:

$$\begin{aligned}
&\int_{+\left(\frac{\ell-1}{L}, \frac{\ell}{L}\right]} \Big[\int_0^t \Big\langle a_{\bar{p}_\mu(t)}, f(s)\Big\rangle ds\Big]\mu\Big(dt|[\ell]\Big) \\
\leq &\int_{+\left(\frac{\ell-1}{L}, \frac{\ell}{L}\right]} \Big[\int_0^{\frac{\ell-1}{L}} \Big\langle a_{\bar{p}_\mu(t)}, f(s)\Big\rangle ds\Big]\mu\Big(dt|[\ell]\Big) + C/L \\
= &\int_{+\left(\frac{\ell-1}{L}, \frac{\ell}{L}\right]} \Big[\sum_{\ell'=1}^{\ell-1} \Big\langle a_{\bar{p}_\mu(t)}, \hat{f}(\ell')\Big\rangle\Big]\mu\Big(dt|[\ell]\Big) + C/L \\
= &\sum_{\ell'=1}^{\ell-1} \Big[\int_{+\left(\frac{\ell-1}{L}, \frac{\ell}{L}\right]} \Big\langle a_{\bar{p}_\mu(t)}, \hat{f}(\ell')\Big\rangle \mu\Big(dt|[\ell]\Big)\Big] + C/L \\
= &\sum_{\ell'=1}^{\ell-1} \Big\langle \int_{+\left(\frac{\ell-1}{L}, \frac{\ell}{L}\right]} a_{\bar{p}_\mu(t)}\mu\Big(dt|[\ell]\Big), \hat{f}(\ell')\Big\rangle + C/L = \frac{1}{L}\sum_{\ell'=1}^{\ell-1} \langle a_{\bar{p}_{\hat{\mu}}(\ell)}, \hat{f}(\ell')\rangle + C/L,
\end{aligned} \tag{5.6.2}$$

where the second equality uses "Fubini's theorem" and the last equality relies on the fact that $a_{\bar{p}_\mu(t)}$ is linear in $\bar{p}_\mu(t)$ and $t \mapsto \bar{p}_\mu(t)$ defines a *posterior* of $\bar{p}_{\hat{\mu}}(\ell)$ conditional on $t \in {}^+(\frac{\ell-1}{L}, \frac{\ell}{L}]$, i.e.,

$$\forall k \in K, \quad \int_{+\left(\frac{\ell-1}{L},\frac{\ell}{L}\right]} \bar{p}^k_\mu(t)\mu\Big(dt|[\ell]\Big) = \int_{+\left(\frac{\ell-1}{L},\frac{\ell}{L}\right]} \frac{p^k\mu^k(dt)}{\mu(dt)} \cdot \frac{\mu(dt)}{\hat{\mu}(\ell)} = \bar{p}_{\hat{\mu}}(\ell);$$

– for the absorbing part:

$$\int_{+\left(\frac{\ell-1}{L},\frac{\ell}{L}\right]} A^*_{\bar{p}_\mu(t)}\mu\Big(dt|[\ell]\Big) \leq A^*_{\bar{p}_{\hat{\mu}}(\ell)}$$

since $A^*_{\bar{p}_\mu(t)}$ is concave in $\bar{p}_\mu(t)$ and $\int_{+\left(\frac{\ell-1}{L},\frac{\ell}{L}\right]} \bar{p}_\mu(t)\mu\Big(dt|[\ell]\Big) = \bar{p}_{\hat{\mu}}(\ell)$. This implies that

$$\begin{aligned}\int_{+\left(\frac{\ell-1}{L},\frac{\ell}{L}\right]} \Big\langle A^*_{\bar{p}_\mu(t)}, \hat{f}(\ell)\Big\rangle(1-t)\mu\Big(dt|[\ell]\Big) &\leq (1-\frac{\ell}{L})\Big\langle \int_{+\left(\frac{\ell-1}{L},\frac{\ell}{L}\right]} A^*_{\bar{p}_\mu(t)}\mu\Big(dt|[\ell]\Big), \hat{f}(\ell)\Big\rangle + C/L \\ &\leq (1-\frac{\ell}{L})\langle A^*_{\bar{p}_{\hat{\mu}}(\ell)}, \hat{f}(\ell)\rangle + C/L\end{aligned} \tag{5.6.3}$$

We substitute (5.6.2) and (5.6.3) back into (5.6.1) to obtain:

$$\begin{aligned}\varphi^p(\mu, f) &\leq \sum_{\ell=1}^{L} \hat{\mu}(\ell)\Big[\langle A^*_{\bar{p}_{\hat{\mu}}(\ell)}, \hat{f}(\ell)\rangle(1-\frac{\ell}{L}) + \frac{1}{L}\sum_{\ell'=1}^{\ell}\langle a_{\bar{p}_{\hat{\mu}}(\ell)}, \hat{f}(\ell')\rangle\Big] + 3C/L \\ &= \hat{\mathcal{L}}^p(\hat{\mu}, \hat{f}) + 3C/L.\end{aligned}$$

Take now $L \in \mathbb{N}$ large with $L \geq 3C/\varepsilon$ and $\hat{w}^L(p) \leq \liminf_{L\to\infty} \hat{w}^L(p) + \varepsilon$, we obtain that against any $\hat{f} \in F[L]$:

$$\hat{\mathcal{L}}^p(\hat{\mu}, \hat{f}) \geq \hat{w}(p) - \varepsilon,$$

thus

$$\liminf_{L\to\infty} \hat{w}^L(p) \geq \hat{w}^L(p) - \varepsilon \geq \hat{w}(p) - 2\varepsilon.$$

The proof for the lemma is then complete. □

Lemma 5.6.2. $\limsup_{L\to\infty} \hat{w}^L(p) \leq \hat{w}(p)$.

Proof for Lemma 5.6.2 For fixed $\varepsilon > 0$, let $f \in F'[0]$ be an ε-optimal strategy in $(\hat{\Xi}(p), A^*_\cdot(j))$. By the (uniform) continuity of f on $[0,1]$, we take $L \in \mathbb{N}$ large with

$$|f(t) - f(t')| \leq \varepsilon, \quad \text{for all } t, t' \in [0,1], \quad |t - t'| \leq 1/L.$$

Define $\hat{f} := \hat{f}[f; L] \in F[L]$ with $\hat{f}(\ell) = f(\frac{\ell}{L})$ for all $\ell = 1, ..., L$. For any $\hat{\mu} \in \mathcal{Q}^K[L]$, let us identify it with an atomic measure $\mu := \mu[\hat{\mu}; L] \in \mathcal{Q}^K[0]$ satisfying $\mu^k(\{\frac{\ell}{L}\}) = \hat{\mu}^k(\ell), \forall \ell = 1, ..., L, \forall k \in K$.

We obtain that for any $\ell = 1, ..., L$ and $t \in (\frac{\ell-1}{L}, \frac{\ell}{L}]$,

$$\begin{aligned}\varphi^p_t(\mu, f) &= \sum_{1\leq\ell'\leq\ell-1} \mu\Big(\{\frac{\ell'}{L}\}\Big)\Big\langle A^*_{\bar{p}_\mu(\frac{\ell'}{L})}, f(\frac{\ell'}{L})\Big\rangle + \Big(1 - \underline{\mu}(\frac{\ell-1}{L})\Big)\Big\langle a_{\bar{p}_\mu(\frac{\ell}{L})}, f(t)\Big\rangle \\ &= \sum_{1\leq\ell'\leq\ell-1} \hat{\mu}(\ell')\Big\langle A^*_{\bar{p}_{\hat{\mu}}(\ell')}, \hat{f}(\ell')\Big\rangle + \Big(1 - \underline{\hat{\mu}}(\ell-1)\Big)\Big\langle a_{\bar{p}_{\hat{\mu}}(\ell)}, f(t)\Big\rangle,\end{aligned}$$

which implies that

$$\varphi^p_{\frac{\ell}{L}}(\mu,f)=\hat{\mathcal{L}}^p_\ell(\hat{\mu},\hat{f})\quad\text{and}\quad\left|\varphi^p_{\frac{\ell}{L}}(\mu,f)-\varphi^p_t(\mu,f)\right|\le C\|f(t)-f(\ell/L)\|_1\le C\varepsilon.$$

Integrating over $t\in(\frac{\ell-1}{L},\frac{\ell}{L}]$ and summing over $\ell=1,...,L$, one obtains

$$\left|\varphi^p(\mu,f)-\hat{\mathcal{L}}^p(\hat{\mu},\hat{f})\right|\le\sum_{\ell=1}^L\Big|\int_{(\frac{\ell-1}{L},\frac{\ell}{L}]}\varphi^p_t(\mu,f)dt-\frac{1}{L}\hat{\mathcal{L}}^p_\ell(\hat{\mu},\hat{f})\Big|\le C\varepsilon.$$

Thus, by the ε-optimality of f,

$$\hat{\mathcal{L}}^p(\hat{\mu},\hat{f})\le\varphi^p(\mu,f)+C\varepsilon\le\hat{w}(p)+(C+1)\varepsilon.$$

Take further L large with $\hat{w}^L(p)\ge\limsup_{\ell\to\infty}\hat{w}^\ell(p)-\varepsilon$. As $\hat{\mu}\in\mathcal{Q}^K[L]$ is arbitrary, we obtain:

$$\limsup_{\ell\to\infty}\hat{w}^\ell(p)\le\hat{w}^L(p)+\varepsilon\le\hat{w}(p)+(C+2)\varepsilon.$$

The proof for the lemma is then complete. □

Proof for PART B of Proposition 5.4.4: Let us prove by induction on $m=1,...,M$. First, the claim is true for $m=1$ since we have $\hat{w}^L_1(p)=w^L_1(p)$ for any $L\in\mathbb{N}$. Suppose now that for some m and for all $\bar{p}\in\Delta(K)$, the limit $\lim_{L\to\infty}w^L_m(\bar{p})$ exists and is equal to $\hat{\nu}_m(\bar{p})$. We prove below that for any $p\in\Delta(K)$, $\lim_{L\to\infty}w^L_{m+1}(p)=\hat{\nu}_{m+1}(p)$.

For any fixed $\varepsilon>0$, we take $L_0\in\mathbb{N}$ such that

$$|w^\ell_m(\bar{p})-\hat{\nu}_m(\bar{p})|\le\varepsilon,\ \forall\ell\ge L_0,\ \forall\bar{p}\in\Delta(K).$$

Notice that here the existence of such L_0 uniformly in $\Delta(K)$ is due to a compactness argument: $\hat{\nu}_m(p)$ is C-Lip. on the compact simplex $\Delta(K)$.

Consider now $m+1$ and $p\in\Delta(K)$. Denote by $\bar{w}_{m+1}(p)$ any accumulation point of the sequence $(w^L_{m+1}(p))$. We take $L\ge L_0/\varepsilon$ realizing both limits as follows:

$$\left|w^L_{m+1}(p)-\bar{w}_{m+1}(p)\right|\le\varepsilon\quad\text{and}\quad\left|\hat{w}^L_{m+1}(p)-\hat{\nu}_{m+1}(p)\right|\le\varepsilon\tag{5.6.4}$$

Below (μ,f) takes value in $\mathcal{Q}^K[L]\times F[L]$. Let $T\in\{1,...,L\}$ is the random stage of playing *Top*. For any $(\ell,j)\in\{1,...,L\}\times J$, $\bar{p}_\mu(\ell)$ is the *posterior* belief conditional on "$T=\ell$"; $\bar{p}_\mu(\ell,j)$ is the *posterior* belief conditional on "$T=\ell,j_T=j,\theta=T$"; $\bar{p}_\mu(\ell,j)$ is the *posterior* belief conditional on "$T=\ell,j_T=j,\theta>T$".

As was proved in **Part A**, we can replace $\hat{w}_m(\cdot)$ by $\hat{\nu}_m(\cdot)$ in the definition of $\hat{w}^L_{m+1}(p)$, which yields (cf. (5.4.3)):

$$\begin{aligned}\hat{w}^L_{m+1}(p)=&\mathrm{Val}_{(\mu,f)}\ \frac{1}{L}\sum_{\ell=1}^L\mu(\ell)\Bigg\{\sum_{\ell'=1}^\ell\langle a_{\bar{p}_\mu(\ell)},f(\ell')\rangle\\&+(L-\ell)\sum_j f(\ell)[j]\Big[\chi^{\bar{p}_\mu(\ell)}(j)a^*_{\hat{p}_\mu(\ell,j)}(j)+(1-\chi^{\bar{p}_\mu(\ell)}(j))\hat{\nu}_m(\bar{p}_\mu(\ell,j))\Big]\Bigg\}.\end{aligned}\tag{5.6.5}$$

Further, the following analog recursive equation is satisfied for $w^L_m(\cdot)$:

$$\begin{aligned}w^L_{m+1}(p)=&\mathrm{Val}_{(\mu,f)}\ \frac{1}{L}\sum_{\ell=1}^L\mu(\ell)\Bigg\{\sum_{\ell'=1}^\ell\langle a_{\bar{p}_\mu(\ell)},f(\ell')\rangle\\&+(L-\ell)\sum_j f(\ell)[j]\Big[\chi^{\bar{p}_\mu(\ell)}(j)a^*_{\hat{p}_\mu(\ell,j)}(j)+(1-\chi^{\bar{p}_\mu(\ell)}(j))w^{L-\ell}_m(\bar{p}_\mu(\ell,j))\Big]\Bigg\}.\end{aligned}\tag{5.6.6}$$

Denote $L_\varepsilon := \lfloor (1-\varepsilon)L \rfloor$ and consider the reduced game where $(\hat{\mu}, f) \in \mathcal{Q}^K[L_\varepsilon] \times F[L]$:

$$w^{L_\varepsilon}_{m+1}(p) = \mathrm{Val}_{(\hat{\mu},f)} \frac{1}{L}\sum_{\ell=1}^{L_\varepsilon} \hat{\mu}(\ell)\Big\{ \sum_{\ell'=1}^{\ell} \langle a_{\bar{p}_{\hat{\mu}}(\ell)}, f(\ell')\rangle + (L-\ell)\sum_j f(\ell)[j]\Big[\chi^{\bar{p}_{\hat{\mu}}}(j) a^*_{\bar{p}_{\hat{\mu}}(\ell,j)}(j) + (1-\chi^{\bar{p}_{\hat{\mu}}}(j)) w^{L-\ell}_m(\bar{p}_{\hat{\mu}}(\ell,j))\Big]\Big\}. \tag{5.6.7}$$

We can exchange the order of sum to re-write the payoffs in Equation (5.6.6) and Equation (5.6.7) in the form of a average payoff of L stages. This implies that

$$w^L_{m+1}(p) \geq w^{L_\varepsilon}_{m+1}(p) \geq w^L_{m+1}(p) - 2C\varepsilon,$$

where the error term $2C\varepsilon$ is due to the average loss of not playing *Top* during the last ε duration of the game.

Moreover, for any $\ell \leq L_\varepsilon$: one has $L - \ell \geq L_0$, thus

$$\big|w^{L-\ell}_m(\bar{p}_\mu(\ell,j)) - \hat{\nu}_m(\bar{p}_\mu(\ell,j))\big| \leq \varepsilon, \ \forall j \in J.$$

We replacing in Equation (5.6.7) the variable $w^{L-\ell}_m(\cdot)$ by $\hat{\nu}_m(\cdot)$, we obtain from Equation (5.6.5):

$$\big|w^{L_\varepsilon}_{m+1}(p) - \hat{w}^L_{m+1}(p)\big| \leq 2C\varepsilon, \quad \text{thus} \quad \big|w^L_{m+1}(p) - \hat{w}^L_{m+1}(p)\big| \leq 4C\varepsilon.$$

Finally, we combine inequalities in (5.6.4) to obtain

$$\big|\bar{w}_{m+1}(p) - \hat{\nu}_{m+1}(p)\big| \leq (4C+2)\varepsilon.$$

This finishes the inductive proof for **PART B:** $\lim_{L\to\infty} w^L_m(p) = \hat{\nu}_m(p), \forall m, \ p \in \Delta(K)$. □

5.6.2 Proof for Proposition 5.4.10

Proposition 5.4.10 Assume that for each $j \in J$, the function $A^*_\cdot(j)$ defined on $\Delta(K)$ is concave and C-Lip. Then for any $\mu \in \mathcal{Q}^K[0]$ that is optimal for player 1 in $(\hat{\Xi}(p), A^*_\cdot(j))$, there exists some $f \in F'[0]$ such that

$$\varphi^p_t(\mu, f) = \hat{w}(p), \ \forall t \in [0,1).$$

Proof for Proposition 5.4.10: Let $\mu_\varepsilon \in \mathcal{Q}^K[0]$ be a non-atomic ε-optimal strategy for player 1 in $\hat{\Xi}(p)$. We consider an auxiliary game $\mathcal{G}(\mu_\varepsilon)$ where player 1 chooses at random a point t in $[0,1]$ and player chooses a function $f \in F'[0]$. The corresponding payoff is $\varphi^p_t(\mu_\varepsilon, f)$. This game has a value ν_ε. Indeed, the strategy set of player 1 (*resp.* player 2) is convex, compact (*resp.* convex). Moreover, the mapping $f \mapsto \varphi_t(\mu_\varepsilon, f)$ is affine and the mapping $t \mapsto \varphi_t(\mu_\varepsilon, f)$ is continuous. Obviously one has $\nu_\varepsilon(p) \geq \hat{w}(p) - \varepsilon$, since player 1 can use the $\ell(\cdot)$ the Lebesgue measure to choose t and then the payoff is precisely $\varphi^p(\mu_\varepsilon, f)$.

Below let us prove that $\nu_\varepsilon \leq \hat{w}$. In fact, let $b(\cdot)$ be an optimal (compactness) strategy of player 1 so that $\int_0^1 \varphi^p_t(\mu_\varepsilon, f) b(dt) \geq \nu_\varepsilon$ for all $f \in F'[0]$. Replacing ν_ε by $\nu_\varepsilon - \delta$, we can assume that $\underline{b}(t) = b([0,t])$ is a strictly increasing continuous function from $[0,1]$ to itself with $\underline{b}(0) = 0$ and $\underline{b}(1) = 1$. We now use $\underline{b}$ to re-scale time, namely, we define $\tilde{\mu}_\varepsilon$ in $\mathcal{Q}^K[0]$ and $\tilde{f}$ in F' by $\underline{\tilde{\mu}_\varepsilon}(\underline{b}(t)) = \underline{\mu_\varepsilon}(t)$ and $\tilde{f}(\underline{b}(t) = f(t)$. That is, $\underline{\tilde{\mu}_\varepsilon} = \underline{\mu_\varepsilon} \circ \underline{b}^{-1}$ and $\tilde{f} = f \circ b^{-1}$. We prove that

Claim 5.6.3.

$$\int_0^1 \varphi_t^p(\mu_\varepsilon, f) b(dt) = \int_0^1 \varphi_t^p(\tilde{\mu}_\varepsilon, \tilde{f}) \ell(dt).$$

Proof for Claim 5.6.3: By definition, μ_ε is the image measure $\underline{b}^{-1}\sharp\tilde{\mu}_\varepsilon$ as $b(\cdot)$ is bijective. It is sufficient for us to show that for all $t \in [0,1]$,

$$\varphi_t^p(\tilde{\mu}_\varepsilon, \tilde{f}) = \varphi^p_{\underline{b}^{-1}(t)}(\mu_\varepsilon, f), \tag{5.6.8}$$

and then use a change of variable "$t = \underline{b}^{-1}(\omega)$" in the integration $\int_0^1 \varphi_t^p(\mu_\varepsilon, f) b(dt)$ to obtain the claim. The expression $\varphi_t^p(\tilde{\mu}_\varepsilon, \tilde{f})$ defined by $A^*_\cdot$ is :

$$\varphi_t^p(\tilde{\mu}_\varepsilon, \tilde{f}) = \int_0^t \langle A^*_{\bar{p}_{\tilde{\mu}_\varepsilon}(t')}, \tilde{f}(t')\rangle \tilde{\mu}_\varepsilon(dt') + \left(1 - \underline{\tilde{\mu}_\varepsilon}(t)\right) \langle a_{\tilde{p}_{\tilde{\mu}_\varepsilon}(t)}, \tilde{f}(t)\rangle. \tag{5.6.9}$$

By definition of $\bar{p}_\mu(\cdot)$ and $\tilde{p}_\mu(\cdot)$, we obtain by the change of variable "$t' = \underline{b}(\omega')$" for all $t' \in [0,t]$ and $\omega' \in [0,\omega]$:

$$\bar{p}^k_{\tilde{\mu}_\varepsilon}(t') = \frac{p^k \tilde{\mu}^k_\varepsilon(dt')}{\sum_{k\in K} p^k \tilde{\mu}^k_\varepsilon(dt')} = \frac{p^k \mu^k_\varepsilon(d\omega')}{\sum_{k\in K} p^k \mu^k_\varepsilon(d\omega')} = \bar{p}^k_{\mu_\varepsilon}(\omega')$$

and

$$\tilde{p}^k_{\tilde{\mu}_\varepsilon}(t) = \frac{p^k \left(1 - \underline{\tilde{\mu}^k_\varepsilon}(\underline{b}(\omega))\right)}{\sum_{k\in K} p^k \left(1 - \underline{\tilde{\mu}^k_\varepsilon}(\underline{b}(\omega))\right)} = \frac{p^k \left(1 - \underline{\mu^k_\varepsilon}(\omega)\right)}{\sum_{k\in K} p^k \left(1 - \underline{\mu^k_\varepsilon}(\omega)\right)} = \bar{p}^k_{\mu_\varepsilon}(\omega)$$

for all $k \in K$. Finally, we use the change of variable "$t' = \underline{b}(\omega')$" for all $t' \in [0,t]$ and $\omega' \in [0,\omega]$ in the integration (5.6.9) to obtain:

$$\begin{aligned}
\varphi_t^p(\tilde{\mu}_\varepsilon, \tilde{f}) &= \int_0^t \langle A^*_{\bar{p}_{\tilde{\mu}_\varepsilon}(t')}, \tilde{f}(t')\rangle \tilde{\mu}_\varepsilon(dt') + \left(1 - \underline{\tilde{\mu}_\varepsilon}(t)\right) \langle a_{\tilde{p}_{\tilde{\mu}_\varepsilon}(t)}, \tilde{f}(t)\rangle \\
&= \int_0^\omega \langle A^*_{\bar{p}_{\mu_\varepsilon}(\omega')}, f(\omega')\rangle \mu_\varepsilon(d\omega') + \left(1 - \underline{\mu_\varepsilon}(\omega)\right) \langle a_{\tilde{p}_{\mu_\varepsilon}(\omega)}, f(\omega)\rangle \\
&= \varphi^p_\omega(\mu_\varepsilon, f) \\
&= \varphi^p_{\underline{b}^{-1}(t)}(\mu_\varepsilon, f),
\end{aligned}$$

which proves (5.6.8). □

Since $b(\cdot)$ defines a one-to-one mapping on $F'[0]$, following Claim 5.6.3, we have: $\varphi(\tilde{\mu}_\varepsilon, \tilde{f}) \geq \nu_\varepsilon(p) - \delta$ for all $f \in F'[0]$. This implies that the measure $\tilde{\mu}_\varepsilon = \tilde{\mu}_\varepsilon \circ \underline{b}^{-1}$ guarantees $\nu_\varepsilon(p) - \delta$ in $\hat{\Xi}(p)$, thus $\nu_\varepsilon(p) - \delta \leq \hat{w}(p)$. Since $\delta > 0$ is arbitrary, the inequality $\nu_\varepsilon(p) \leq \hat{w}(p)$ is obtained and we have $\hat{w}(p) - \varepsilon \leq \nu_\varepsilon(p) = \hat{w}(p)$ for all $\varepsilon > 0$.

Consider now for each $\varepsilon = 1/n$, denote $\mu_n = \mu_\varepsilon$, and let (μ_n) converges weakly to μ: $\varphi_t^p(\mu_n, f') \xrightarrow{n\to\infty} \varphi_t^p(\mu, f')$ for any $f' \in F'[0]$ and $t \in [0,1]$. For each $n \geq 1$, let now $f_n \in F'[0]$ be the $\frac{1}{n}$-optimal strategy of player 2 in the auxiliary game $\mathcal{G}(\mu_n)$: in particular against any of player 1's strategy playing the Dirac mass, $\varphi_t^p(\mu_n, f_n) \leq \nu_{1/n} + 1/n \leq \hat{w} + 1/n$ for all $t \in [0,1)$. Finally, let $f \in F[0]$ be an accumulation point of (f_n) satisfying $\varphi_t(\mu, f_n) \xrightarrow{n\to\infty} \varphi_t(\mu, f)$ for all $t \in [0,1)$. This implies that for all $t \in [0,1)$, on can take n large such that:

$$\varphi_t(\mu, f) \leq \varphi_t(\mu, f_n) + 1/n \leq \varphi_t(\mu_n, f_n) + 2/n \leq \hat{w}(p) + 2/n$$

thus $\varphi_t(\mu, f) \leq \hat{w}(p),\ \forall t \in [0,1]$. Since μ is optimal in $\hat{\Xi}(p)$, this implies that $\varphi_t(\mu, f) = \hat{w}(p)$ for all $t \in [0,1]$. $\square$

ACKNOWLEDGMENT This paper is written during the course of my PhD thesis, I wish to thank my supervisor Sylvain Sorin for fruitful comments. I thank also Fabien Gensbittel and Guillaume Vigeral for helpful discussions. Part of this work was done during the author's stay at *Hausdorff Research Institute for Mathematics* on the workshop "Stochastic Dynamics in Economics and Finance", May-July 2013 in Bonn, Germany.

Chapter 6

Jeux récursifs: valeur uniforme, théorème Taubérian et la conjecture de Mertens

Résumé Nous étudions les jeux récursifs à somme nulle avec un espace d'état dénombrable. Lorsque la famille des fonctions valeur à n étapes $\{v_n, n \geq 1\}$ est totalement bornée pour la norme uniforme, nous prouvons l'existence de la valeur uniforme. En particulier, la convergence uniforme de (v_n) implique la convergence uniforme de la suite (v_λ) des fonctions valeur escompté. À l'aide d'un résultat dans Rosenberg et Vieille [52], nous obtenons un théorème taubérien uniforme pour les jeux récursifs: (v_n) converge uniformément si et seulement si (v_λ) converge uniformément.

Nous appliquons notre résultat principal à une sous-classe du modèle général des jeux répétés avec un espace d'état fini, des ensembles d'actions finis et des ensembles de signaux finis. Ce sont des jeux récursifs avec signaux (où les joueurs n'observent que les signaux sur l'état et sur les actions jouées) et le maximiseur est toujours plus informé que le minimiseur. Nous prouvons pour cette classe de jeux répétés la conjecture de Mertens: "$Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$". Enfin, on en déduit l'existence de la valeur uniforme dans les jeux récursifs finis avec signaux symétriques.

Mots-clés Jeux stochastiques, jeux récursifs, valeur asymptotique, valeur uniforme, théorème taubérien, Maxmin

---

Ce chapitre est issu de l'article *Recursive games: uniform value, Tauberian theorem and Merten's conjecture*, en collaboration avec Xavier Venel, et il est accepté pour publication dans la revue *International Journal of Game Theory (spacial issue in honor of Abraham Neyman).*

Recursive games: Uniform value, Tauberian theorem and Merten's conjecture

joint with Xavier Venel (Paris-Sorbonne)

To appear in *International Journal of Game Theory*

Abstract We study two-player zero-sum recursive games with a countable state space and finite action space at each state. When the family of n-stage values $\{v_n, n \geq 1\}$ is totally bounded for the uniform norm, we prove the existence of the uniform value. Together with a result in Rosenberg and Vieille [52], we obtain a uniform Tauberian theorem for recursive game: (v_n) converges uniformly if and only if (v_λ) converges uniformly.

We apply our main result to finite recursive games with signals (where players observe only signals on the state and on past actions). When the maximizer is more informed than the minimizer, we prove the Mertens conjecture $Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$. Finally, we deduce the existence of the uniform value in finite recursive game with symmetric information.

Keywords Stochastic games, recursive games, asymptotic value, uniform value, Tauberian theorem, Maxmin

6.1 Introduction

Stochastic games were introduced by Shapley [53] to model a multiplayer dynamic interaction, where players' collective decisions influence the current payoff and also the future state. In this article, we focus on two-player zero-sum recursive games introduced by Everett [18]. The specificity of a recursive game is that the state space is divided into two sets: absorbing states and active states. On absorbing states, the process is absorbed and the payoff is fixed. On active (non-absorbing) states, the payoff is always equal to 0.

There are several ways to evaluate the payoff stream in a zero-sum stochastic game. Given a positive integer n, the n-stage payoff is the expected average payoff during the first n stages. Given $\lambda \in (0, 1]$, the λ-discounted payoff is the Abel mean of the infinite stage payoffs with a weight $\lambda(1-\lambda)^{t-1}$ for stage t. We will focus on the concept of uniform value. A stochastic game admits a uniform value if both players can approximately guarantee the same payoff level in all sufficiently long n-stage games without knowing *a priori* the length of the game.

Mertens and Neyman [34] proved that a stochastic game with a finite state space and finite set of actions where the players observe the current state and the stage payoffs

admits a uniform value. Their proof uses the fact that the function $\lambda \mapsto v_\lambda$ has bounded variation, where v_λ is the λ-discounted value (Bewley and Kohlberg [10]). For stochastic games with an infinite state space, this argument in general does not apply.

Markovian decision processes (henceforth MDP) are stochastic games with only one player. Lehrer and Sorin [28] showed that in a MDP, the uniform convergence of (v_λ) (w.r.t. the initial state) as λ tends to zero is equivalent to the uniform convergence of the n-stage values (v_n) as n tends to infinity. Nevertheless, uniform convergence of (v_n) or (v_λ) is not sufficient for the existence of the uniform value (cf. Monderer and Sorin [38] or Lehrer and Monderer [27]).

For recursive games, the situation seems to be different. There are two results giving sufficient conditions for a recursive game with countable state space to have a uniform value. The first one can be derived from Rosenberg and Vieille [52]: if (v_λ) converges uniformly to some function v, then the recursive game has a uniform value, which is equal to v. The second one is due to Solan and Vieille [54]: if, except on a finite subset, the *limsup value*[1] is above a strictly positive constant on the non-absorbing states, then the recursive game has a uniform value, which is equal to the limsup value.

The main result of this paper is that the uniform convergence of the n-stage values is a sufficient condition for the existence of the uniform value. In fact we prove a stronger result: for any recursive game with countable state space, if the family $\{v_n, n \geq 1\}$ is totally bounded for the uniform norm, then the uniform value exists. Our proof follows the same idea as Solan and Vieille [54] and we will use several of their results.

Our result together with the result of Rosenberg and Vieille [52] provides a uniform Tauberian theorem for recursive games: (v_n) converges uniformly if and only if (v_λ) converges uniformly, and in case of convergence, both limits are the same. For general stochastic games, Ziliotto [70] provided recently a direct proof of this result.

Finally, we apply our main result to finite recursive games with signals. In a recursive game with signals, players do not perfectly observe the state and actions at every stage anymore, rather they receive a private signal. Mertens [33] conjectured that in a general model of zero-sum repeated games, if player 1 (the maximizer) is always more informed than player 2 (the minimizer) during the play (in the sense that player 2's private signal can be deduced from player 1's private signal) then $Maxmin = \lim_{n\to\infty} v_n = \lim_{\lambda\to 0} v_\lambda$, *i.e.*, the uniform maxmin and the asymptotic value both exist and are equal.

Ziliotto [68] showed that the result is false in general. Nevertheless, several positive results have been obtained for subclasses of games including Sorin [55] and Sorin [57] for Big match with one-sided incomplete information, Rosenberg *et al.* [50], Renault [45] and Gensbittel *et al.* [21] for a more informed controller, and Rosenberg and Vieille [52] for recursive games with one-sided incomplete information.

We prove the Mertens conjecture in finite recursive games with signals, where player 1 is always more informed than player 2 during the play. The proof uses several results from Gensbittel *et al.* [21], concerning the n-stage value functions in a repeated game where player 1 is more informed than player 2. Our result generalizes Rosenberg and Vieille [52], which deals with the model where player 1 is informed of a private signal on the state at the beginning of the game. Moreover, we deduce the existence of the uniform value in finite recursive games with symmetric information.

The organization of the article is as follows: in Section 6.2 we introduce the model of recursive games; in Section 6.3 we present the main result and several corollaries; Section

1. The limsup value is the value of the game in which the global payoff to player 1 is the limsup of the stage payoff stream.

6.4 is dedicated to the proofs; finally in Section 6.5 we apply the result to finite recursive games with signals.

6.2 Preliminaries: model and notations

Notation Given any metric space S, endowed with the Borelian σ-algebra, we denote by $\Delta(S)$ the set of probabilities on S and we denote by $\Delta_f(S)$ the set of probabilities with finite support.

6.2.1 The model

A two-player zero sum stochastic game $\Gamma = \langle X, A, B, g, q\rangle$ is given by
- a state space X.
- player 1's action set A, and for any $x \in X$, $A(x)$ is a finite subset of A.
- player 2's action set B, and for any $x \in X$, $B(x)$ is a finite subset of B.
- a payoff function: $g : X \times A \times B \to [-1, +1]$.
- a transition probability function: q: $X \times A \times B \to \Delta_f(X)$.

Play of the game The stochastic game with initial state $x_1 \in X$ is denoted by $\Gamma(x_1)$, and is played as follows: at each stage $t \geq 1$, after observing $(x_1, a_1, b_1, \dots \dots, a_{t-1}, b_{t-1}, x_t)$, player 1 and player 2 choose simultaneously actions $a_t \in A(x_t)$ and $b_t \in B(x_t)$. The stage payoff is $g(x_t, a_t, b_t)$ and a new state x_{t+1} is drawn according to the probability distribution $q(x_t, a_t, b_t)$. Both players observe the action pair (a_t, b_t) and the state x_{t+1}. The game then proceeds to stage $t + 1$.

Note that we did not make any measurability assumption on the model. As the transition probability distribution is supposed to be finitely supported, given an initial state, the set of actions and states that might appear in the infinite game are in fact countable. Therefore probability distributions are well defined.

Recursive game Γ is a recursive game if there exist a set of active states denoted by X^0 and a set of absorbing states denoted by X^* with $X^0 \cup X^* = X$ and $X^0 \cap X^* = \emptyset$, such that:
- the stage payoff is 0 on active states: $\forall x \in X^0$, $g(x, a, b) = 0, \forall (a, b) \in A(x) \times B(x)$;
- states in X^* are absorbing: $\forall x \in X^*$, $q(x, a, b)(x) = 1$, $\forall (a, b) \in A(x) \times B(x)$, and $g(x, a, b)$ depends only on x.

6.2.2 Definition of strategies and evaluations

History At stage t, the space of finite histories is $H_t = (X \times A \times B)^{t-1} \times X$. Set $H_\infty = (X \times A \times B)^\infty$ to be the space of infinite *plays*. We consider the discrete topology on X, A and B. For every $t \geq 1$, we identify any $h_t \in H_t$ with a cylinder set in H_∞ and denote by $\mathcal{H}_t$ the σ-field of H_t induced on H_∞. The product σ-field on H_∞ is $\mathcal{H}_\infty = \sigma(\mathcal{H}_t, t \geq 1)$.

Strategy A *(behavior) strategy* for player 1 is a sequence of functions $\sigma = (\sigma_t)_{t\geq 1}$ with each $t \geq 1$, $\sigma_t : (H_t, \mathcal{H}_t) \to \Delta(A)$ such that for every $h_t \in H_t$, $\sigma_t(h_t)(A(x_t)) = 1$. If for every $t \geq 1$ and $h_t \in H_t$, there exists $a \in A(x_t)$ such that $\sigma_t(h_t)[a] = 1$, then the strategy is *pure*. We define similarly a behavior strategy τ for player 2. Denote by Σ and $\mathcal{T}$ respectively player 1's and player 2's sets of behavior strategies. Denote by $\widehat{\Sigma}$ and $\widehat{\mathcal{T}}$

respectively player 1's and player 2's subsets of strategies that depend on the histories only through the states but not on the actions.

Evaluations Let us describe several ways to evaluate the payoff in Γ. By Kolmogorov's extension theorem, any triple $(x_1, \sigma, \tau) \in X \times \Sigma \times \mathcal{T}$ induces a unique probability distribution over $(H_\infty, \mathcal{H}_\infty)$ denoted by $\mathbb{P}_{x_1,\sigma,\tau}$. Let $\mathbb{E}_{x_1,\sigma,\tau}$ be the corresponding expectation.

n-stage average For each positive $n \geq 1$, the expected average payoff up to stage n, induced by the couple of strategies (σ, τ) and the initial state x_1 is given by

$$\gamma_n(x_1, \sigma, \tau) = \mathbb{E}_{x_1,\sigma,\tau}\left(\frac{1}{n}\sum_{t=1}^{n} g(x_t, a_t, b_t)\right).$$

The game with expected n-stage average payoff and initial state x_1 is denoted as $\Gamma_n(x_1)$.

λ-discounted average For each $\lambda \in (0, 1]$, the expected λ-discounted average payoff, induced by the couple of strategies (σ, τ) and the initial state x_1 is given by

$$\gamma_\lambda(x_1, \sigma, \tau) = \mathbb{E}_{x_1,\sigma,\tau}\left(\lambda\sum_{t=1}^{\infty}(1-\lambda)^{(t-1)} g(x_t, a_t, b_t)\right).$$

The game with expected λ-discounted average payoff and initial state x_1 is denoted as $\Gamma_\lambda(x_1)$.

In either $\Gamma_n(x_1)$ or $\Gamma_\lambda(x_1)$, player 1 maximizes the expected average payoff and player 2 minimizes it. For a fixed x_1 the game $\Gamma_n(x_1)$ is finite, so there exists a value $v_n(x_1)$ by minmax theorem. The existence of the discounted value $v_\lambda(x_1)$ is also standard, and we refer to Mertens *et al.* [35] (Section VII.1.) for a general presentation.

6.2.3 Stopping time and concatenation of strategies

A function $\theta : (H_\infty, \mathcal{H}_\infty) \to \mathbb{N}$ is called a *stopping time* if the set $\{h \in H_\infty | \theta(h) = t\}$ is $\mathcal{H}_t$-measurable for all $t \geq 1$. Explicitly for any $h, h' \in H_\infty$ and $n \geq 1$: if h and h' coincide until stage n and $\theta(h) = n$ then $\theta(h') = n$. Let θ and θ' be two stopping times, we write $\theta \leq \theta'$ if for every $h \in H_\infty$, $\theta(h) \leq \theta'(h)$.

Given a sequence of strategies $(\sigma^{[\ell]})_{\ell\geq 1}$ and a sequence of increasing stopping time $(\theta_\ell)_{\ell\geq 1}$, we define $\sigma^* := \sigma^{[1]}\theta_1\sigma^{[2]}\theta_2 \cdots$ as the *concatenation* of $(\sigma^{[\ell]})_{\ell\geq 1}$ along $(\theta_\ell)_{\ell\geq 1}$. Given $n \geq t \geq 1$ and $h \in H_\infty$, let h_n be the projection of h on H_n and h_n^t be the history of h between stage t and n. The strategy σ^* is defined by $\sigma_n^*(h_n) = \sigma_n^{[1]}(h_n)$ if $n < \theta_1(h)$; $\sigma_n^*(h_n) = \sigma_{n-\theta_{m-1}}^{[m]}(h_n^{\theta_{m-1}})$ if $\theta_{m-1} \leq n < \theta_m$. Informally, for every $\ell \geq 1$ at stage θ_ℓ, the player forgets the past and starts to play $\sigma_{\ell+1}$ at the current state.

6.2.4 Uniform value

Uniformly guarantee Player 1 *uniformly guarantees* w if for every $\varepsilon > 0$, there exists σ_ε in Σ and $N_0 \geq 1$ such that for every $x_1 \in X^0$,

$$\gamma_n(x_1, \sigma_\varepsilon, \tau) \geq w(x_1) - \varepsilon, \quad \forall n \geq N_0, \ \forall \tau \in \mathcal{T}.$$

We say that the strategy σ_ε uniformly guarantees $w - \varepsilon$. Similarly, player 2 uniformly guarantees w if for every $\varepsilon > 0$, there exists τ_ε in $\mathcal{T}$ and $N_0 \geq 1$ such that for every

$x_1 \in X^0$,

$$\gamma_n(x_1, \sigma, \tau_\varepsilon) \leq w(x_1) + \varepsilon, \quad \forall n \geq N_0, \ \forall \sigma \in \Sigma.$$

Uniform value $v_\infty : X \to \mathbb{R}$ is the uniform value of the game Γ if both players uniformly guarantee v_∞. A strategy for player 1 (*resp.* player 2) that uniformly guarantees $v_\infty - \varepsilon$ (*resp.* $v_\infty + \varepsilon$) is called *uniform ε-optimal.* If both players can uniformly guarantee v_∞ with pure strategies, Γ has *a uniform value in pure strategies.*

Remark 6.2.1. *In defining the uniform value, we ask N_0 to be independent of the initial state x_1. One direct consequence of the existence of the uniform value v_∞ is the uniform convergence of $(v_n)_{n\geq 1}$ to v_∞. This is stronger than the definition where the existence of the uniform value is considered state by state (see for example Solan and Vieille [54], Definitions 3-4)*

6.3 Main results

In this section, we present the main result of the paper, namely Theorem 6.3.1, as well as several corollaries. We also provide an example that does not satisfy the condition of Theorem 6.3.1 and does not have a uniform value.

6.3.1 Sufficient condition for the existence of the uniform value

Denote by $\mathbf{B}(X)$ the set of functions from X to $[-1, 1]$ with the uniform norm $\|\cdot\|_\infty$. Recall that a set of functions F in $(\mathbf{B}(X), \|.\|_\infty)$ is *totally bounded* if for every $\varepsilon > 0$, there exists a finite subset $F_R = \{f_r : 1 \leq r \leq R\} \subseteq F$ such that for any $f \in F$, there is $f_r \in F_R$ with $||f - f_r||_\infty \leq \varepsilon$.

Theorem 6.3.1. *Suppose that the space $\{v_n, n \geq 1\}$ is totally bounded for the uniform norm, then the recursive game Γ has a uniform value v_∞. Moreover both players can uniformly guarantee v_∞ with strategies that depend only on the history of states and not on past actions.*

We deduce from the previous result a uniform Tauberian theorem in recursive games.

Corollary 6.3.2. *The sequence of n-stage values $(v_n)_{n\geq 1}$ converges uniformly as n tends to infinity if and only if the sequence of λ-discounted values $(v_\lambda)_{\lambda\in(0,1]}$ converges uniformly as λ tends to zero. In case of convergence, both limits are the same.*

On one hand, if (v_n) converges uniformly, the family is totally bounded, thus the uniform value exists, and this implies the uniform convergence of (v_λ) (Sorin [59], Lemma 3.1). On the other hand, the converse result is established in Rosenberg and Vieille [52] (see Remark 6, Theorem 1 and Theorem 3).

Remark 6.3.3. *The equivalence of the uniform convergences of $(v_n)_{n\geq 1}$ and $(v_\lambda)_{\lambda\in(0,1]}$ has been proven in MDP by Lehrer and Sorin [28]. Ziliotto [70] recently showed that it is also true for stochastic games whenever the Shapley operator is well defined.*

If, in addition, for every $n \geq 1$ the n-stage value $v_n(x)$ exists in pure strategies, then Γ has a uniform value in pure strategies.

Corollary 6.3.4. *Suppose that for every $n \geq 1$, both players have pure optimal strategies in the n-stage game, and $\{v_n, n \geq 1\}$ is totally bounded for the uniform norm. Then Γ has a uniform value v_∞ in pure strategies. Moreover, both players can uniformly guarantee v_∞ with strategies that depend only on the history of states and not on past actions.*

Remark 6.3.5. *The result in Corollary 6.3.4 extends to games with general action sets $A(x)$ and $B(x)$ provided that for any $n \geq 1$, the n-stage game has a value and both players have pure optimal strategies.*

The proof of Corollary 6.3.4 is similar to that of Theorem 6.3.1. The key difference involves a technical lemma (Lemma 6.4.18) for the existence of a (pure) stopping time which is used in the definition of players' optimal strategies (see the proof of Proposition 6.4.3). We discuss this point and present the proof in Subsection 6.4.3.

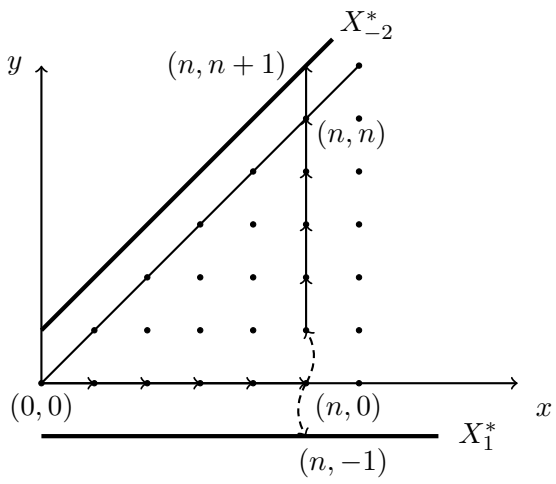
6.3.2 A recursive game without uniform value

We present here an example of a recursive game with countable state space where $\{v_n,\ n \geq 1\}$ is not totally bounded and there is no uniform value (See **Figure 3** below for illustration). This is an adaptation to our framework of an example in Lehrer and Sorin [28].

The state space is a subset of $\mathbb{Z} \times \mathbb{Z}$. The set of active states is $X^0 = \{(x, y) \in \mathbb{N} \times \mathbb{N} \,|0 \leq y \leq x\}$ and the set of absorbing states is $X^* = X_1^* \bigcup X_{-2}^*$ (two types), where $X_1^* = \mathbb{N} \times \{-1\}$ and $X_{-2}^* = \{(x, x+1)|x \geq 0\}$. The payoff is 1 on X_1^* and is -2 on X_{-2}^*. There is only one player (maximizer), whose action set is $\{R(ight), J(ump)\}$. The transition rule is given by:

- at $(x, 0) \in X^0$: $q((x, 0), R)(x+1, 0) = 1$, and $q((x, 0), J)(x, -1) = q((x, 0), J)(x, 1) = \frac{1}{2}$;
- at $(x, y) \in X^0$ with $0 < y \leq x$: $q((x, y), a)(x, y+1) = 1$, $\forall a \in \{R, J\}$.

Starting at $(0, 0)$, one optimal strategy for an n-stage game is to go *Right* for half of the game, and then to *Jump*. This gives an expected average payoff around $\frac{1}{4}$, thus $\lim_{n\to\infty} v_n(0, 0) = \frac{1}{4}$. In a λ-discounted game, the optimal stage to *Jump* is approximately $\frac{\ln(\frac{2-\lambda}{4})}{\ln(1-\lambda)}$. It follows that $v_\lambda(0, 0) \approx \frac{2-\lambda}{16}$ and thus $\lim_{\lambda\to 0} v_\lambda(0, 0) = \frac{1}{8}$. This implies that there is no uniform value. On the other hand, $\{v_n, n \geq 1\}$ is not totally bounded for the uniform norm. Indeed, the convergence of (v_n) is not uniform: for any $x \geq 1$, $\lim_{n\to\infty} v_n(x, 1) = -2$ while $v_x(x, 1) = 0$.



The figure on the left illustrates a play $(R, ..., R, J)$ jumping after n steps: with probability 1/2 the state is absorbed at $(n, n+1) \in X_{-2}^*$; with probability 1/2 the state is absorbed at $(n, -1) \in X_1^*$.

$\longrightarrow$: a deterministic transition;
$- \dashrightarrow$: a probabilistic transition.

Figure 3

6.4 Proofs

In the first subsection, we introduce and establish preliminary results for a subclass of recursive game, which will be called *positive-valued recursive games.* In the second subsection, we prove Theorem 6.3.1 by a reduction of any recursive game to a positive-valued recursive game. The proof for Corollary 6.3.4 is given in the third subsection.

6.4.1 The case of positive-valued recursive game

Definition 6.4.1. *A recursive game is* positive-valued *if there exist $M > 0$ and $n_0 \geq 1$ such that for every non-absorbing state $x \in X^0$, there exists $n(x) \leq n_0$ such that $v_{n(x)}(x) \geq M$.*

In order to state the next proposition, we first introduce the notion of uniformly terminating strategy.

Definition 6.4.2. *Denote by ρ the stopping time of absorption in X^*: $\rho = \inf\{n \geq 1, x_n \in X^*\}$. The strategy σ is said to be* uniformly terminating *if for any $\varepsilon > 0$, there exists $N \geq 1$ such that for every $x_1 \in X^0$ and for every $\tau \in \mathcal{T}$, $\mathbb{P}_{x_1,\sigma,\tau}(\rho \leq N) \geq 1 - \varepsilon$.*

Proposition 6.4.3. *Let Γ be a positive-valued recursive game. We fix the numbers $M > 0, n_0 \geq 1$ and the mapping $n(\cdot) : X^0 \longrightarrow \{1, ..., n_0\}$ such that $v_{n(x)}(x) \geq M, \forall x \in X^0$. Then player 1 uniformly guarantees $v_{n(\cdot)}(\cdot)$ with uniformly terminating strategies that depends only on states: for all $\varepsilon > 0$, there exists σ^* in $\widehat{\Sigma}$ and $N_0 \geq 1$ such that for every $x_1 \in X^0$ and every τ in $\mathcal{T}$,*

$$(i)\ \mathbb{P}_{x_1,\sigma^*,\tau}(\rho \leq N_0) \geq 1 - \varepsilon \text{ and } (ii)\ \gamma_n(x_1, \sigma^*, \tau) \geq v_{n(x_1)}(x_1) - \varepsilon,\ \forall n \geq N_0.$$

Proof. Let $\hat{\sigma}$ be a profile of strategies such that for every $x \in X^0$, $\hat{\sigma}(x)$ is optimal in the $n(x)$-stage game $\Gamma_{n(x)}(x)$. Let $\tilde{k} := \tilde{k}(x)$ be a random stage uniformly chosen in $\{1, ..., n(x)\}$. For any $\tau \in \mathcal{T}$ and $x \in X^0$, $(x, \hat{\sigma}, \tau)$ and $\tilde{k}$ induce a probability distribution over $H_\infty \times \{1, ...n(x)\}$, which we denote by $\widetilde{\mathbb{P}}_{x,\hat{\sigma},\tau}$. Let $\widetilde{\mathbb{E}}_{x,\hat{\sigma},\tau}$ be the corresponding expectation. We obtain:

$$\widetilde{\mathbb{E}}_{x,\hat{\sigma},\tau}[g(x_{\tilde{k}})] = \mathbb{E}_{x,\hat{\sigma},\tau}\Big[\frac{1}{n(x)}\sum_{l=1}^{n(x)} g(x_t)\Big] \geq \inf_{\tau'} \mathbb{E}_{x,\hat{\sigma},\tau'}\Big[\frac{1}{n(x)}\sum_{l=1}^{n(x)} g(x_t)\Big] \geq v_{n(x)}(x) \geq M.$$

It follows that

$$\widetilde{\mathbb{E}}_{x,\hat{\sigma},\tau}\left[g(x_{\tilde{k}})\mathbb{1}_{\rho\leq\tilde{k}} + g(x_{\tilde{k}})\mathbb{1}_{\rho>\tilde{k}}\right] \geq M.$$

On the event $\{\rho > \tilde{k}\}$, $g(x_{\tilde{k}}) = 0$, whereas on the event $\{\rho \leq \tilde{k}\}$, we have $g(x_{\tilde{k}}) = g(x_\rho)$. This implies that

$$\widetilde{\mathbb{P}}_{x,\hat{\sigma},\tau}(\rho \leq \tilde{k})\widetilde{\mathbb{E}}_{x,\hat{\sigma},\tau}\left[g(x_\rho) \mid \rho \leq \tilde{k}\right] = \widetilde{\mathbb{E}}_{x,\hat{\sigma},\tau}\left[g\left(x_{\tilde{k}}\right)\right] \geq v_{n(x)}(x) \geq M. \tag{6.4.1}$$

Using the fact that the payoff function g has maximal norm 1, we deduce from (6.4.1):

$$\widetilde{\mathbb{P}}_{x,\hat{\sigma},\tau}(\rho \leq \tilde{k}) \geq M. \tag{6.4.2}$$

Define the strategy[2] σ^* as concatenations of $(\hat{\sigma}(x_{u_l}))_{l\geq 0}$ at the random stages $(u_\ell)_{\ell\geq 0}$, where u_ℓ is defined inductively along the play by $u_0 = 1$ and $u_{\ell+1} - u_\ell = \tilde{k}(x_{u_\ell})$ follows the uniform distribution over $\{1, ..., n(x_\ell)\}$. Let $\widetilde{\mathbb{P}}_{x,\sigma^*,\tau}$ be the (product) probability

2. The strategy σ^* is a generalized mixed strategy, which is equivalent to a behavior strategy by Kuhn's theorem.

distribution over $H_\infty \times \{1, ..., n_0\}^{\mathbb{N}}$ induced by (x, σ^*, τ), and $\widetilde{\mathbb{E}}_{x,\sigma^*,\tau}$ the corresponding expectation. Let $\varepsilon > 0$.
(i) We show that σ^* is uniformly terminating. By (6.4.2), the conditional probability of absorbing on each block $\{u_{l-1}, ..., u_l - 1\}$ is no smaller than M. Thus for any τ and $x_1 \in X^0$,

$$\widetilde{\mathbb{P}}_{x_1,\sigma^*,\tau}(\rho \geq u_l) \leq (1-M)^l, \ \forall l \geq 1.$$

The length of each block is uniformly bounded by n_0, thus if we put $l^* \geq \frac{\ln(\varepsilon)}{\ln(1-M)}$:

$$\widetilde{\mathbb{P}}_{x_1,\sigma^*,\tau}(\rho \leq n_0 l^*) \geq \widetilde{\mathbb{P}}_{x_1,\sigma^*,\tau}(\rho \leq u_{l^*}) \geq 1 - (1-M)^{l^*} \geq 1 - \varepsilon. \tag{6.4.3}$$

(ii) We now argue that σ^* uniformly guarantees $v_{n(x_1)}(x_1) - 3\varepsilon$. Let $N_0 = n_0 l^*/\varepsilon$. For any $\tau \in \mathcal{T}$, $x_1 \in X^0$ and $n \geq n_0 l^*$, we have

$$\begin{aligned}
\mathbb{E}_{x_1,\sigma^*,\tau}[g(x_n)] &= \widetilde{\mathbb{E}}_{x_1,\sigma^*,\tau}\left[\sum_{l=0}^{\ell^*-1} g(x_n)\mathbb{1}_{u_l \leq \rho < u_{l+1}} + g(x_n)\mathbb{1}_{u_{\ell^*} \leq \rho}\right] \\
&= \sum_{l=0}^{\ell^*-1} \widetilde{\mathbb{P}}_{x_1,\sigma^*,\tau}(u_l \leq \rho < u_{l+1})\widetilde{\mathbb{E}}_{x_1,\sigma^*,\tau}[g(x_\rho)|u_l \leq \rho < u_{l+1}] \\
&+ \widetilde{\mathbb{P}}_{x_1,\sigma^*,\tau}(u_{l^*} \leq \rho)\widetilde{\mathbb{E}}_{x_1,\sigma^*,\tau}[g(x_n)|\rho \geq u_{l^*}].
\end{aligned}$$

According to (6.4.3), $\mathbb{P}_{x_1,\sigma^*,\tau}(\rho \geq u_{l^*}) \leq \varepsilon$, thus we focus on an absorption before u_{ℓ^*}:

$$\mathbb{E}_{x_1,\sigma^*,\tau}[g(x_n)] \geq \sum_{l=0}^{l^*-1} \widetilde{\mathbb{P}}_{x_1,\sigma^*,\tau}(u_l \leq \rho < u_{l+1})\widetilde{\mathbb{E}}_{x_1,\sigma^*,\tau}[g(x_\rho)|u_l \leq \rho < u_{l+1}] - \varepsilon \tag{6.4.4}$$

For each $l \geq 0$, σ^* is following $\hat{\sigma}(x_{u_l})$ for $u_{l+1} - u_l = \tilde{k}(x_{u_\ell})$ stages. Thus (6.4.1) applies, and we obtain: for $\ell \geq 1$,

$$\widetilde{\mathbb{P}}_{x_1,\sigma^*,\tau}(u_l \leq \rho < u_{l+1})\widetilde{\mathbb{E}}_{x_1,\sigma^*,\tau}[g(x_\rho)|u_l \leq \rho < u_{l+1}] \geq \widetilde{\mathbb{P}}_{x_1,\sigma^*,\tau}(\rho > u_\ell)M > 0,$$

and for $l = 0$,

$$\widetilde{\mathbb{P}}_{x_1,\sigma^*,\tau}(1 \leq \rho < u_1)\widetilde{\mathbb{E}}_{x_1,\sigma^*,\tau}[g(x_\rho)|1 \leq \rho < u_1] \geq v_{n(x_1)}(x_1).$$

By substituting the two previous inequalities into (6.4.4), we obtain that

$$\forall n \geq n_0 l^*, \forall x_1 \in X^0, \ \mathbb{E}_{x_1,\sigma^*,\tau}[g(x_n)] \geq v_{n(x_1)}(x_1) - \varepsilon. \tag{6.4.5}$$

Now for $n \geq N_0$, we deduce that $\gamma_n(x_1, \sigma^*, \tau) \geq v_{n(x_1)}(x_1) - 3\varepsilon$. □

One can deduce from Proposition 6.4.3 a first result on recursive games with the condition that the sequence of n-stage values converges uniformly to a function bounded away from 0.

Corollary 6.4.4. *Assume that in a recursive game Γ, the sequence of n-stage values $(v_n)_{n\geq 1}$ converges uniformly to a function v satisfying for every $x \in X^0$, $v(x) \geq M' > 0$ for some M'. Then Γ is positive-valued and player 1 uniformly guarantees v with uniformly terminating strategies.*

6.4.2 Existence of the uniform value (proof of Theorem 6.3.1)

This subsection is devoted to the proof of Theorem 6.3.1: the total boundedness of $\{v_n, n \geq 1\}$ implies the existence of the uniform value v_∞. We prove that player 1 guarantees the point-wise limit superior value $x \mapsto v(x) := \limsup_n v_n(x)$. By symmetry, player 2 guarantees $\liminf_n v_n(x)$, and the result follows.

The uniform ε-optimal strategy will use alternatively two different types of strategies. This approach is classical for recursive games and has been used for example in Rosenberg and Vieille [52] and in Solan and Vieille [54]. Our construction is close to Solan and Vieille [54] in which some similar "positive-valued recursive game" is introduced to make a reduction for the general case.

The proof is decomposed into three parts. In the first one, we introduce a family of auxiliary positive-valued recursive games and define the first type of strategies. In the second part, we define the second type of strategies. Finally, we construct the strategy σ^* and prove that it is uniform ε-optimal.

Before proceeding to the proof, let us first prove a preliminary result, which shows that due to the total boundedness of $\{v_n\}$, the point-wise limit superior of (v_n) can be realized along uniform convergent subsequences. We fix a recursive game Γ for the rest of this section.

Proposition 6.4.5. *For every $x \in X$, we have*

$$v(x) = \limsup_n v_n(x) = \max_{f \in F} f(x),$$

where F is the set of limit points of the sequence $(v_n)_{n\geq 1}$ in $(\mathbf{B}(X), \|.\|_\infty)$.

Proof. $(\mathbf{B}(X), \|.\|_\infty)$ is a complete metric space and $(\{v_n\}, \|\cdot\|_\infty)$ is totally bounded, therefore F is compact and non-empty. For every $x \in X$, we denote $w(x) := \max_{f\in F} f(x)$. Fix $x \in X$. Since $v(x)$ is the largest limit point of $(v_n(x))_{n\geq 1}$, we have $w(x) \leq v(x)$. By definition of the limit superior, there exists a subsequence $(v_{n_k}(x))_{k\geq 1}$ which converges to $\limsup v_n(x)$. There exists a subsequence of $(v_{n_k})_{k\geq 1}$ that converges in $(\mathbf{B}(X), \|.\|_\infty)$ to some $f^* \in F$, therefore

$$\max_{f\in F} f(x) \geq f^*(x) = v(x).$$

□

A) Reduction: auxiliary recursive games

Auxiliary recursive games Let $\theta : X \to \{0, 1\}$. We define the auxiliary recursive game $\Gamma^\theta = \langle A, B, X = X^0_\theta \bigcup X^*_\theta, q_\theta, g_\theta\rangle$ where any active state $x \in X^0$ such that $\theta(x) = 1$ is seen as an absorbing state: the active state space of Γ^θ is $X^0_\theta = \{x \in X^0, \theta(x) = 0\}$ and the absorbing state space is $X^*_\theta = X^* \bigcup \{x \in X^0, \theta(x) = 1\}$. The transition q_θ is equal to q and the payoff g_θ is equal to g on all states except $\{x \in X^0, \theta(x) = 1\}$, on which the state is absorbing and the absorbing payoff is $g_\theta = v$. For every $n \geq 1$, let v^θ_n be the value of the n-stage auxiliary game Γ^θ_n.

Proposition 6.4.6. *Let $\eta > 0$ and $\theta : X \to \{0, 1\}$. There exists $n_0 \geq 1$ such that for every $x_1 \in X^0_\theta$, there exists $n(x_1) \leq n_0$ with $v^\theta_{n(x_1)}(x_1) \geq v(x_1) - 4\eta$.*

Proof. Let $\eta > 0$ be fixed and $F_R = \{f_1, ..., f_R\} \subseteq F$ be a finite cover of size $\frac{\eta}{2}$ of the set F. As $\{v_n, n \geq 1\}$ is totally bounded, there exists some stage $n(\eta) \in \mathbb{N}$, after which any n-stage value v_n is $\frac{\eta}{2}$-close to F its set of accumulation points, hence η-close to F_R:

$$\exists n(\eta) \in \mathbb{N},\ \forall n \geq n(\eta),\ \exists f_r \in \{f_1, ..., f_R\},\ s.t.\ \ ||v_n - f_r||_\infty \leq \eta; \tag{6.4.6}$$

Moreover for every $r \in \{1, .., R\}$, f_r is an accumulation point of $\{v_n, n \geq 1\}$, therefore there exists some $n_r > \frac{n(\eta)}{\eta}$ such that v_{n_r} is η-close to f_r:

$$\forall f_r \in F_R,\ \exists n_r > \frac{n(\eta)}{\eta},\ s.t.\ \ ||v_{n_r} - f_r||_\infty \leq \eta. \tag{6.4.7}$$

Finally we take $n_0 = \max\{n_r : 1 \leq r \leq R\}$. The integers n_r are chosen such that when absorption in X^*_θ occurs in the game of length n_r the remaining number of stages is either a fraction smaller than η of the total length of the game or greater than $n(\eta)$ and Equations (6.4.6) applies.

Let $x_1 \in X^0_\theta$ be any non-absorbing state in the auxiliary game Γ^θ. By compactness of F, there exists $f \in F$ such that $f(x_1) = v(x_1)$ and $f_r \in F_R$ with $||f - f_r||_\infty \leq \frac{\eta}{2}$. In particular at state x_1,

$$f_r(x_1) \geq f(x_1) - \frac{\eta}{2} = v(x_1) - \frac{\eta}{2},$$

which together with (6.4.7) implies that

$$v_{n_r}(x_1) \geq f_r(x_1) - \eta \geq v(x_1) - \frac{3}{2}\eta. \tag{6.4.8}$$

We now prove that

$$v^\theta_{n_r}(x_1) \geq v_{n_r}(x_1) - 2\eta. \tag{6.4.9}$$

Denote by

$$\rho_\theta = \inf_{t \geq 1}\{x_t \in X^*_\theta\} = \inf_{t \geq 1}\{x_t \in X^* \text{ or } \theta(x_t) = 1\}$$

the stopping time associated to absorption in Γ^θ, and set $\rho^{n_r}_\theta = \min(\rho_\theta, n_r)$. An adaptation of standard proof technique of the Shapley equation gives us:

$$v_{n_r}(x_1) = \max_{\sigma \in \Sigma} \min_{\tau \in \mathcal{T}} \mathbb{E}_{x_1,\sigma,\tau}\left(\frac{1}{n_r}\left(\sum_{t=1}^{\rho^{n_r}_\theta - 1} g(x_t)\right) + \frac{n_r - \rho^{n_r}_\theta + 1}{n_r} v_{n_r - \rho^{n_r}_\theta + 1}(x_{\rho^{n_r}_\theta})\right).$$

We separate the histories into two sets depending on whether $n_r - \rho^{n_r}_\theta(h) + 1 > n(\eta)$ in which cases Equation (6.4.6) applies, or $n_r - \rho^{n_r}_\theta(h) + 1 \leq n(\eta)$ in which cases $\frac{n_r - \rho^{n_r}_\theta(h) + 1}{n_r} \leq \eta$ (by definition $n_r \geq \frac{n(\eta)}{\eta}$), and deduce that

$$v_{n_r}(x_1) \leq \max_{\sigma \in \Sigma} \min_{\tau \in \mathcal{T}} \mathbb{E}_{x_1,\sigma,\tau}\left(\frac{1}{n_r}\sum_{t=1}^{\rho^{n_r}_\theta - 1} g(x_t) + \frac{n_r - \rho^{n_r}_\theta + 1}{n_r} f'_h(x_{\rho^{n_r}_\theta})\right) + 2\eta,$$

with $f'_h \in F_R$ depending on the history given by Equation (6.4.6) applied to $v_{n_r - \rho^{n_r}_\theta + 1}$ when $n_r - \rho^{n_r}_\theta + 1 > n(\eta)$, and any function in F_r otherwise. Therefore, by considering v as the supremum of $f \in F$ at each point $x_{\rho^{n_r}_\theta} \in X$, we have $f'_h(x_{\rho^{n_r}_\theta}) \leq v(x_{\rho^{n_r}_\theta})$, thus

$$\begin{aligned} v_{n_r}(x_1) &\leq \max_{\sigma \in \Sigma} \min_{\tau \in \mathcal{T}} \mathbb{E}_{x_1,\sigma,\tau}\left(\frac{1}{n_r}\sum_{t=1}^{\rho^{n_r}_\theta - 1} g(x_t) + \frac{n_r - \rho^{n_r}_\theta + 1}{n_r} v(x_{\rho^{n_r}_\theta})\right) + 2\eta \\ &= v^\theta_{n_r}(x_1) + 2\eta. \end{aligned}$$

This proves inequality (6.4.9). We now use Equation (6.4.8) and Equation (6.4.9) to conclude:

$$v^{\theta}_{n_r}(x_1) \geq v(x_1) - 4\eta.$$

It means that for each $x_1 \in X^0_\theta$, there exists $n(x_1) := n_r \leq n_0 = \max\{n_r : \ 1 \leq r \leq R\}$, such that $v^{\theta}_{n(x_1)}(x_1) \geq v(x_1) - 4\eta$. □

.

Remark 6.4.7. *Proposition 6.4.6 is also true if θ is a deterministic stopping time and not only a function on the state. The auxiliary game would be defined on a larger state space: the set of finite histories of the original game. The proof in itself is similar.*

Fix now any $\varepsilon > 0$ and define $\theta_\varepsilon : X \to \{0,1\}$ such that $\{x \in X, \ \theta_\varepsilon(x) = 1\} = \{x \in X, v(x) < \varepsilon\}$. We denote by $\Gamma^\varepsilon = \langle A, B, X = X^0_\varepsilon \bigcup X^*_\varepsilon, q_\varepsilon, g_\varepsilon\rangle$ the auxiliary game associated to Γ defined by the stopping time θ_ε.

Corollary 6.4.8. *In the game Γ^ε, Player 1 uniformly guarantees v with uniformly terminating strategies that depend only on past states.*

Proof. Let $\eta \in (0, \varepsilon/8]$, by Proposition 6.4.6 there exists $n_0 \geq 1$ such that for every $x_1 \in X^0_\varepsilon$, there exists $n(x_1) \leq n_0$ with

$$v^{\varepsilon}_{n(x_1)}(x_1) \geq v(x_1) - 4\eta \geq \varepsilon/2, \tag{6.4.10}$$

where the second inequality comes from the definition of X^0_ε. Therefore, Γ^ε is a positive-valued recursive game and by Proposition 6.4.3, player 1 uniformly guarantees $v^{\varepsilon}_{n(\cdot)}(\cdot)$ with uniformly terminating strategies in $\widehat{\Sigma}$. By Equation (6.4.10), it follows that for every $\eta > 0$, player 1 uniformly guarantees $v - 4\eta$ with uniformly terminating strategies. □

Fix now a strategy σ^*_ε that is uniformly terminating in Γ^ε, depends only on past states and guarantees $v(x_1) - \varepsilon^2$ in $\Gamma^\varepsilon(x_1)$ for every $x_1 \in X^0_\varepsilon$.

B) One-shot game

One-shot game G^f For each $f : X \to [-1, +1]$ and $x_1 \in X$, we define the one-shot game G^f as follows: player 1's action set is $A(x_1)$, player 2's action set is $B(x_1)$, and the payoff is for each $(s,t) \in \Delta(A) \times \Delta(B)$,

$$\mathbb{E}_{q(x_1,s,t)}[f(x_2)] = \sum_{a\in A, b\in B} s(a)t(b) \left(\sum_{x_2 \in X} q(x_1, a, b)(x_2) f(x_2) \right).$$

Lemma 6.4.9. *For any limit point $f \in F$, the one-shot game G^f has a value equal to f.*

Proof. Let $n \geq 1$, it is known that (cf. Vigeral [63] p.40, Lemma 4.2.2)

$$\|v_n - v_{n+1}\|_\infty \leq \frac{2}{n+1},$$

and by Shapley's formula (see **Appendix**) that

$$\begin{aligned} v_{n+1}(x_1) &= \sup_{s\in\Delta\left(A(x_1)\right)} \inf_{t\in\Delta(B(x_1))} \mathbb{E}_{q(x_1,s,t)} \left[\frac{1}{n+1} g(x_1) + \frac{n}{n+1} v_n(x_2) \right] \\ &= \inf_{t\in\Delta(B(x_1))} \sup_{s\in\Delta\left(A(x_1)\right)} \mathbb{E}_{q(x_1,s,t)} \left[\frac{1}{n+1} g(x_1) + \frac{n}{n+1} v_n(x_2) \right]. \end{aligned}$$

We obtain the result by taking the limit along a subsequence converging uniformly to $f \in F$. □

Following Proposition 6.4.5, one can take for each $x \in X$ some $f^* \in F$ such that $v(x) = f^*(x) \geq f(x), \forall f \in F$. Then the following result is a direct consequence of Lemma 6.4.9 .

Corollary 6.4.10. *For every $x_1 \in X$, there exists $s^*(x_1) \in \Delta(A(x_1))$ such that*

$$\forall b \in B(x_1),\ \mathbb{E}_{q(x_1,s^*(x_1),b)}\left[v(x_2)\right] \geq v(x_1).$$

Fix now $s^* := \left(s^*(x_1)\right)_{x_1 \in X}$ a profile of strategies satisfying the conclusion of Corollary 6.4.10.

C) Optimal strategy

Roughly speaking, we build $\bar{\sigma}$ a uniform ε-optimal strategy for player 1 to play σ^*_ε in Γ^ε on the states with value v above 2ε, and to play s^* in G^v on the states with value v below ε. And for the states with value v between ε and 2ε, $\bar{\sigma}$ will be either of the two depending on the regime.

Construction of $\bar{\sigma}$ Define a sequence of stopping times $(u_l)_{l\geq 1}$ and the concatenated strategy $\overline{\sigma} := s^* u_1 \sigma^*_\varepsilon u_2 s^* u_3 \sigma^*_\varepsilon u_4 \cdots$ in Γ as follows:

– $\overline{\sigma}$ is to play $s^*(x_n)$ at each stage n up to stage (not included)

$$u_1 = \inf\{n \geq 1,\ v(x_n) > 2\varepsilon\};$$

and then to play $\sigma^*_\varepsilon(x_{u_1})$ up to stage (not included)

$$u_2 = \inf\{n \geq u_1,\ v(x_n) < \varepsilon\}.$$

– In general: for each $r \geq 1$, $\overline{\sigma}$ is to play $\sigma^*_\varepsilon(x_{u_{2r-1}})$ from stage u_{2r-1} (*the odd phase*) up to stage (not included)

$$u_{2r} = \inf\{n \geq u_{2r-1},\ v(x_n) < \varepsilon\}.$$

and then to play $s^*(x_n)$ at each stage $n \geq u_{2r}$ (*the even phase*), up to stage (not included)

$$u_{2r+1} = \inf\{n \geq u_{2r},\ v(x_n) > 2\varepsilon\}.$$

Remark 6.4.11. *The idea of alternating between two types of strategies is common in Rosenberg and Vieille [52], Solan and Vieille [54] and this article. The main difference is the definition of the target function v used to define how to switch from one type of strategies to the other. Rosenberg and Vieille [52] use the limit of discounted values and σ^*_ε is an optimal strategy in some λ-discounted game (for λ close to zero). Solan and Vieille [54] use the limsup value and introduce an auxiliary positive-valued game. We adopt a similar approach to Solan and Vieille [54] but with v the largest limit point of (v_n).*

By construction, $\bar{\sigma}$ depends on the histories only through the states and not the actions. Let us show that $\bar{\sigma}$ uniformly guarantees $v - 25\varepsilon$ for player 1, which finishes the proof of Theorem 6.3.1.

Fix from now on any $x_1 \in X$. Recall that ρ denotes the absorption time in the game Γ. The next result shows that the process $(v(x_{\min(\rho,u_l)}))_{l\geq 1}$, which is the value of v at switching times (u_l), is almost a submartingale up to an error of ε^2.

Proposition 6.4.12. *For every $l \geq 1$ and every $\tau \in \mathcal{T}$:*

$$\mathbb{E}_{x_1,\overline{\sigma},\tau}[v(x_{\min(\rho,u_{l+1})})|\mathcal{H}_{\min(\rho,u_l)}] \geq v(x_{\min(\rho,u_l)}) - \varepsilon^2 \mathbb{1}_{\rho>u_l},$$

on the event $\min(\rho, u_l) < +\infty$.

Proof. Take any τ in $\mathcal{T}$. The result is true if $\rho \leq u_l$. Suppose that l is even and $\rho > u_l$: by construction the strategy $(s^*(x_n))$ is used during the phrase $n \in \{u_l, ..., u_{l+1}-1\}$, thus:

$$\mathbb{E}_{x_1,\overline{\sigma},\tau}[v(x_{n+1})|\mathcal{H}_n] \geq v(x_n), \text{ for all } u_l \leq n < \min(\rho, u_{l+1}).$$

Therefore $(v(x_n))$ is a bounded submartingale and by Doob's stopping theorem,

$$\mathbb{E}_{x_1,\overline{\sigma},\tau}[v(x_{\min(\rho,u_{l+1})})|\mathcal{H}_{\min(\rho,u_l)}] \geq v(x_{\min(\rho,u_l)}).$$

Suppose that l is odd and $\rho > u_l$. By construction, player 1 is using $\sigma^*_\varepsilon(x_{u_l})$, which uniformly guarantees $v(x_{u_l}) - \varepsilon^2$ in the auxiliary game $\Gamma^\varepsilon(x_{u_l})$:

$$\exists N_0 \geq 1, \quad \mathbb{E}_{x_1,\bar{\sigma},\tau}\left[\frac{1}{n}\sum_{t=u_l+1}^{u_l+n} g_\varepsilon(x_t)|\mathcal{H}_{u_l}\right] \geq v(x_{u_l}) - \varepsilon^2 \text{ for all } n \geq N_0. \tag{6.4.11}$$

Denote by $\rho^\varepsilon = \min\{m \geq u_l + 1 : x_m \in X^*_\varepsilon\}$ the absorption time in $\Gamma^\varepsilon(x_{u_l})$. Since in recursive games the payoff is zero before absorption, we have

$$\mathbb{E}_{x_1,\bar{\sigma},\tau}\left[g_\varepsilon(x_{\rho_\varepsilon})|\mathcal{H}_{u_l}\right] = \mathbb{E}_{x_1,\bar{\sigma},\tau}\left[\lim_{n\to\infty}\frac{1}{n}\sum_{t=u_l+1}^{u_l+n} g_\varepsilon(x_t)|\mathcal{H}_{u_l}\right]. \tag{6.4.12}$$

By the dominated convergence theorem,

$$\mathbb{E}_{x_1,\bar{\sigma},\tau}\left[\lim_{n\to\infty}\frac{1}{n}\sum_{t=u_l+1}^{u_l+n} g_\varepsilon(x_t)|\mathcal{H}_{u_l}\right] = \lim_{n\to\infty}\mathbb{E}_{x_1,\bar{\sigma},\tau}\left[\frac{1}{n}\sum_{t=u_l+1}^{u_l+n} g_\varepsilon(x_t)|\mathcal{H}_{u_l}\right]. \tag{6.4.13}$$

We deduce from (6.4.11)-(6.4.13) that

$$\mathbb{E}_{x_1,\bar{\sigma},\tau}\left[g_\varepsilon(x_{\rho^\varepsilon})|\mathcal{H}_{u_l}\right] \geq v(x_{u_l}) - \varepsilon^2.$$

Moreover, $g_\varepsilon(x_{\rho^\varepsilon}) = v(x_{\rho^\varepsilon})$ and conditionally on $\rho > u_l$, $\rho^\varepsilon = \min(u_{l+1}, \rho)$. It follows that

$$\mathbb{E}_{x_1,\overline{\sigma},\tau}[v(x_{\min(\rho,u_{l+1})})|\mathcal{H}_{u_l}] \geq v(x_{u_l}) - \varepsilon^2.$$

□

Due to the possible error term ε^2, the sequence $(v(x_{\min(\rho,u_l)}))_{l\geq 1}$ is not a submartingale. Nevertheless, one can prove a lemma similar to the usual upcrossing lemma for submartingale. Indeed, the value is a martingale excepts if it crosses upwards the interval $[\varepsilon, 2\varepsilon]$. When this happens, the value may decreases of at most ε^2. With the submartingale property established in Proposition 6.4.12, an easy adaptation of the standard result on upcrossing number of submartingale implies the following result, as was shown in Proposition 3 of Rosenberg and Vieille [52]:

Lemma 6.4.13. *Let $N = \sup\{p \geq 1 : u_{2p-1} < +\infty\}$ be the number of times the process $(v(x_{u_l}))$ crosses upward the interval $[\varepsilon, 2\varepsilon]$.For every $\tau \in \mathcal{T}$,*

$$\mathbb{E}_{x_1,\overline{\sigma},\tau}[N] \leq \frac{1}{\varepsilon - \varepsilon^2}.$$

By construction, σ^*_ε is *uniformly terminating* within the auxiliary absorbing states X^*_ε. That is to say, any play between stages u_{2p-1} and u_{2p} (on an odd phase) has bounded length with high probability under the strategy $\sigma^*_\varepsilon(x_{u_{2p-1}})$, uniformly over any starting state $x_{u_{2p-1}} \in X^0_\varepsilon$. Since Lemma 6.4.13 implies that the number of odd phases is bounded in expectations, the total frequency of stages on all odd phases is negligible for n large. Let us formalize this fact.

Recall that ρ^ε denotes the absorption time in the auxiliary game Γ^ε. It follows that there exists $N_1 > 0$ such that

$$\forall x \in X^0_\varepsilon \text{ and } \tau \in \mathcal{T} : \mathbb{P}_{x,\sigma^*_\varepsilon(x),\tau}(\rho^\varepsilon > N_1) \leq \varepsilon^3. \tag{6.4.14}$$

For each $n \in \mathbb{N}$, define $A_n = \{u_{2p-1} \leq n < \min(\rho, u_{2p}), u_{2p-1} < \rho, \text{ for some } p\} \subseteq H_\infty$. These are all infinite plays where stage n is in an odd phrase, *i.e.*, the stages between u_{2p-1} and u_{2p} on which $\sigma^*_\varepsilon(x_{u_{2p-1}})$ is used. We fix for the rest of subsection the uniform stage number N_1 satisfying (6.4.14).

Lemma 6.4.14. *For every $\tau \in \mathcal{T}$ and every $n \geq \frac{N_1}{\varepsilon^3}$,*

$$\frac{1}{n}\sum_{k=1}^{n} \mathbb{P}_{x_1,\overline{\sigma}(x_1),\tau}(A_k) \leq 5\varepsilon.$$

The proof for this lemma relies on the upcrossing property established in Lemma 6.4.13, and takes the same form as Lemma 27 in Solan and Vieille [54]. Solan and Veille [54] make some *finiteness* assumption (on the set of non-absorbing states on which the target function is not bounded away from zero) in order to obtain the existence of X^1_ε a subset of X^0_ε and a uniform bound $N_1 \geq 1$ such that

$$\forall x \in X^1_\varepsilon \text{ and } \tau \in \mathcal{T} : \mathbb{P}_{x,\sigma^*_\varepsilon(x),\tau}(\rho^\varepsilon > N_1) \leq \varepsilon^3.$$

Under the assumption that $\{v_n, n \geq 1\}$ is totally bounded, we showed in Section 6.4.2 (cf. the condition defined in (6.4.14)) that we can consider X^1_ε to be the whole set X^0_ε.

The following result is a reformulation of the submartingale property in Lemma 6.4.12.

Lemma 6.4.15. *For any $m_0 \geq 1$, we have*

$$\mathbb{E}_{x_1,\overline{\sigma},\tau}[v(x_{m_0})] \geq v(x_1) - \varepsilon^2 \cdot \mathbb{E}_{x_1,\overline{\sigma},\tau}[N] - 2\mathbb{P}_{x_1,\overline{\sigma},\tau}(A_{m_0}) - \varepsilon.$$

Proof. For a proof, we refer to Proposition 28 in Solan and Vieille [54], where our lemma is stated as Equation (4) in their proof. □

Now we use Lemma 6.4.13, Lemma 6.4.14 and Lemma 6.4.15 to prove the following proposition, which concludes the proof of Theorem 6.3.1.

Proposition 6.4.16. *For any $x_1 \in X^0$ and for any τ,*

$$\mathbb{E}_{x_1,\bar{\sigma},\tau}\left[\frac{1}{n}\sum_{m=1}^{n} g(x_m)\right] \geq v(x_1) - 25\varepsilon, \ \forall n \geq \frac{N_1}{\varepsilon^3}.$$

Proof. Take $x_1 \in X^0$ and fix any τ. In this proof h will denote a pure play. We use the fact that $g(x_m) \geq v(x_m) - 2\varepsilon$ if $h \notin A_m$: indeed, either the play has absorbed so $g(x_m) = v(x_m)$, or we have $v(x_m) < 2\varepsilon$ and $g(x_m) = 0$. Moreover, if $h \in A_m$, we use $g(x_m) \geq -1$. This gives us:

$$\begin{aligned}\mathbb{E}_{x_1,\overline{\sigma},\tau}\left[\frac{1}{n}\sum_{m=1}^{n} g(x_m)\right] &\geq \frac{1}{n}\mathbb{E}_{x_1,\overline{\sigma},\tau}\left[\sum_{m=1}^{n} \mathbb{1}_{h\notin A_m}(v(x_m)-2\varepsilon) + \sum_{m=1}^{n}\mathbb{1}_{h\in A_m}(-1)\right] \\ &\geq \frac{1}{n}\mathbb{E}_{x_1,\overline{\sigma},\tau}\left[\sum_{m=1}^{n} v(x_m)\right] + \frac{1}{n}\mathbb{E}_{x_1,\overline{\sigma},\tau}\left[\sum_{m=1}^{n}\mathbb{1}_{h\in A_m}\big(-1-(v(x_m)-2\varepsilon)\big)\right] - 2\varepsilon.\end{aligned} \tag{6.4.15}$$

Lemma 6.4.15 (taking average sum on $m_0 = 1, ..., n$) implies that

$$\frac{1}{n}\mathbb{E}_{x_1,\overline{\sigma},\tau}\left[\sum_{m=1}^{n} v(x_m)\right] \geq v(x_1) - \varepsilon^2 \cdot \mathbb{E}_{x_1,\overline{\sigma},\tau}[N] - \frac{2}{n}\sum_{m=1}^{n}\mathbb{P}_{x_1,\overline{\sigma},\tau}(A_m) - \varepsilon. \tag{6.4.16}$$

Moreover, the bound $v(x_m) \leq 1$ gives

$$\begin{aligned}\frac{1}{n}\mathbb{E}_{x_1,\overline{\sigma},\tau}\left[\sum_{m=1}^{n}\mathbb{1}_{h\in A_m}\big(-1-v(x_m)+2\varepsilon\big)\right] &\geq \frac{1}{n}\mathbb{E}_{x_1,\overline{\sigma},\tau}\left[\sum_{m=1}^{n}\mathbb{1}_{h\in A_m}(-2+2\varepsilon)\right] \\ &= (-2+2\varepsilon)\frac{1}{n}\sum_{m=1}^{n}\mathbb{P}_{x_1,\overline{\sigma},\tau}(A_m).\end{aligned} \tag{6.4.17}$$

We substitute (6.4.16) and (6.4.17) back into (6.4.15) to obtain

$$\mathbb{E}_{x_1,\overline{\sigma},\tau}\left[\frac{1}{n}\sum_{m=1}^{n} g(x_m)\right] \geq v(x_1) - \varepsilon^2 \cdot \mathbb{E}_{x_1,\overline{\sigma},\tau}[N] - 3\varepsilon + (-4+2\varepsilon)\cdot\left(\frac{1}{n}\sum_{m=1}^{n}\mathbb{P}_{x_1,\overline{\sigma},\tau}(A_m)\right).$$

Finally, we use Lemma 6.4.13 and Lemma 6.4.14 in the equality to have that: $\forall n \geq \frac{N_1}{\varepsilon^3}$ and $\forall \varepsilon \leq \frac{1}{2}$,

$$\mathbb{E}_{x_1,\overline{\sigma},\tau}\left[\frac{1}{n}\sum_{m=1}^{n} g(x_m)\right] \geq v(x_1) - \frac{\varepsilon^2}{\varepsilon-\varepsilon^2} - 3\varepsilon - 20\varepsilon \geq v(x_1) - 25\varepsilon.$$

Note that N_1 does not depend on the particular choice of x_1 in X^0, so the strategy $\overline{\sigma}$ uniformly guarantees $v - 25\varepsilon$ in the infinite game Γ. □

6.4.3 Pure optimal strategy (proof of Corollary 6.3.4)

To prove the result, it is sufficient to show that both the strategy s^* and the strategy σ^*_ε defined in the proof of Theorem 6.3.1 can be chosen pure and depending only on the history of states.

By assumption, the n-stage game $\Gamma_n(x)$ has a value in pure strategies. It follows that Shapley's equation for any v_n is satisfied with pure strategies, and so is Lemma 6.4.9. We deduce that there exists a pure action s^* that satisfies the conclusion of Corollary 6.4.10.

The construction of the strategy σ^*_ε appeared in the proof of Proposition 6.4.2, where it was defined as the concatenation of a sequence of strategies $(\hat{\sigma}(x_{u_\ell}))_{\ell\geq 1}$ at the random stages $(u_\ell)_{\ell\geq 1}$. As each $\hat{\sigma}(x)$ is optimal in the $n(x)$-stage game $\Gamma_{n(x)}(x)$, $\hat{\sigma}(x_{u_\ell})$ can be taken pure. The definition of the random stages u_ℓ involved a randomized stopping time $\tilde{k} \in \{1, ..., n(x)\}$ satisfying:

$$\forall \tau \in \mathcal{T},\ \widetilde{\mathbb{E}}_{x,\hat{\sigma},\tau}\left[g(x_{\tilde{k}})\right] \geq \min_{\tau'\in\mathcal{T}} \mathbb{E}_{x,\hat{\sigma},\tau'}\Big[\frac{1}{n(x)}\sum_{t=1}^{n(x)} g(x_t)\Big].$$

To obtain a pure strategy σ^*_ε, we show that the random stopping time $\tilde{k}$ can be replaced by a stopping time (pure one), which depends only on the history of states and not on the actions. In order to build this stopping time, we restrict ourselves to strategies in $\widehat{\Sigma}$, *i.e.*, strategies which depend only on past states. Note that each $\hat{\sigma}(x_{u_\ell})$, as an optimal strategy in $\Gamma_{n(x_{u_\ell})}(x_{u_\ell})$, can be taken in $\widehat{\Sigma}$.

Lemma 6.4.17. *Fix any $\hat{\sigma} \in \widehat{\Sigma}$ and x_1. For any $\tau \in \mathcal{T}$, there exists some $\hat{\tau} \in \widehat{\mathcal{T}}$ such that $\mathbb{P}_{x_1,\hat{\sigma},\hat{\tau}}(x_1, ..., x_t) = \mathbb{P}_{x_1,\hat{\sigma},\tau}(x_1, ..., x_t)$ for any $(x_1, ..., x_t) \in X^t, t \geq 1$.*

Proof. For all $t \geq 1$, we denote by $s_t := (x_1, ..., x_t)$ the t first states. For any $\tau \in \mathcal{T}$, define the reduced strategy $\hat{\tau} \in \widehat{\mathcal{T}}$ as:

$$\hat{\tau}_t(s_t) = \sum_{h_t \in H_t(s_t)} \mathbb{P}_{x_1,\hat{\sigma},\tau}(h_t|s_t)\tau_t(h_t), \ \forall s_t, \ \forall t \geq 1.$$

where $H_t(s_t)$ denotes the histories in H_t containing s_t. Then we obtain by definition:

$$\mathbb{P}_{x_1,\hat{\sigma},\hat{\tau}}(s_{t+1}) = \sum_{h_t \in H_t(s_t)} \mathbb{P}_{x_1,\hat{\sigma},\tau}(h_t|s_t)\mathbb{P}_{x_1,\hat{\sigma},\tau}(s_{t+1}|h_t) = \mathbb{P}_{x_1,\hat{\sigma},\tau}(s_{t+1}).$$

□

Lemma 6.4.18. *Fix any $x_1 \in X^0$ and $\sigma \in \widehat{\Sigma}$. For any $n \geq 1$, there exists a stopping time $\theta : \bigcup_{1 \leq t \leq n} X^t \to \{1, ..., n\}$ such that for every strategy τ of player 2:*

$$\mathbb{E}_{x_1,\sigma,\tau}\left[g(x_\theta)\right] \geq \min_{\tau'} \mathbb{E}_{x_1,\sigma,\tau'}\left[\frac{1}{n}\sum_{t=1}^{n} g(x_t)\right].$$

Proof. By Lemma 6.4.17, we can assume that $\tau \in \widehat{\mathcal{T}}$. Let us prove the result by induction. For every $x_1 \in X^0$, the result is true for $n = 1$. Suppose that the claim is true for $n - 1$. Let $x_1 \in X^0$. By applying the inductive assumption to the different states possible at stage 2, we obtain that there is some stopping time $\theta^+ : \bigcup_{t=1}^{n-1} X^t \to \{2, ..., n\}$ such that

$$\mathbb{E}_{x_1,\sigma,\tau}\big[g(x_{\theta^+})|x_2\big] \geq \min_{\tau'} \mathbb{E}_{x_1,\sigma,\tau'}\Big[\frac{1}{n-1}\sum_{t=2}^{n} g(x_t)|x_2\Big] := w_{n-1}(\sigma, x_1, x_2). \tag{6.4.18}$$

Denote $w_{n-1}(\sigma, x_1) = \inf_{y \in \Delta(J)} \mathbb{E}_{x_1,\sigma,y}\big[w_{n-1}(\sigma, x_1, x_2)\big]$. We define the stopping time $\theta : \bigcup_{t=1}^{n} X^t \to \{1, ..., n\}$ by

$$\forall (x_1, .., x_t) \in X^t, \quad \theta(x_1, ..., x_t) = \begin{cases} 1 & \text{if } 0 \geq w_{n-1}(\sigma, x_1), \\ \theta^+(x_2, ..., x_t) & \text{otherwise.} \end{cases}$$

According to the definition of θ and the inductive assumption (6.4.18) for θ^+:

$$\begin{aligned} \mathbb{E}_{x_1,\sigma,\tau}[g(x_\theta)] &= g(x_1)\mathbb{1}_{0 \geq w_{n-1}(\sigma,x_1)} + \mathbb{E}_{x_1,\sigma,\tau}\Big[\mathbb{E}_{x_1,\sigma,\tau}\big[g(x_{\theta^+})|x_2\big]\Big]\mathbb{1}_{0 < w_{n-1}(\sigma,x_1)} \\ &\geq \max\left\{0, w_{n-1}(\sigma, x_1)\right\} \\ &\geq \frac{n-1}{n} w_{n-1}(\sigma, x_1). \end{aligned}$$

Finally $g(x_1) = 0$, therefore

$$\frac{n-1}{n} w_{n-1}(\sigma, x_1) = \left(\frac{n-1}{n}\right) \inf_{\tau'} \mathbb{E}_{x_1,\sigma,\tau'}\left[\frac{1}{n-1}\sum_{t=2}^{n-1} g(x_t)\right] = \min_{\tau'} \mathbb{E}_{x_1,\sigma,\tau'}\left[\frac{1}{n}\sum_{t=1}^{n} g(x_t)\right].$$

This concludes the inductive proof. □

Remark 6.4.19. *Let Γ be a stochastic game where the payoff function depends only on the state but not the actions, the proof for the above result follows the same way.*

6.5 Application to recursive games with signals

In this last section, we apply our result to the model of finite recursive games with signals where one player is more informed than the other player. Introducing an auxiliary stochastic game similar to the one defined in Gensbittel *et al.* [21], we show that the study of such a recursive game can be reduced to the study of recursive game with a countable state space satisfying the assumption of Corollary 6.3.4.

6.5.1 Model

The following mode of general repeated games is introduced in Mertens *et al.* [35]. A repeated game $\Gamma = (K, I, J, C, D, g, q)$ is given by

- a finite state space: K.
- two finite action spaces I and J.
- two finite signal spaces C and D.
- a payoff function: $g : K \times I \times J \to [-1, +1]$.
- a transition probability function (on states and signals): q from $K \times I \times J$ to $\Delta(K \times C \times D)$.

Denote by $\Gamma(\pi)$ the game with an initial probability distribution $\pi \in \Delta(K \times C \times D)$, which is played as follows. Initially, the triple (k_1, c_1, d_1) is drawn according to π. At stage 1: player 1 learns c_1 and player 2 learns d_1. Then simultaneously player 1 chooses an action $i_1 \in I$ and player 2 chooses an action $j_1 \in J$. The stage payoff is $g(k_1, i_1, j_1)$, and the new triple (k_2, c_2, d_2) is drawn according to $q(k_1, i_1, j_1)$. The game then proceeds to stage 2: player 1 observes c_2, and player 2 observes d_2 etc...

We assume that each player's signal contains his own action. Formally, there exists $\hat{\imath} : C \to I$ and $\hat{\jmath} : D \to J$ such that

$$\forall k \in K, \sum_{k',c,d} q\big(k, \hat{\imath}(c), \hat{\jmath}(d)\big)(k', c, d) = 1.$$

We will focus on repeated games with the following two features: *recursive* and *one player is more informed than the other.*

Definition 6.5.1. *The repeated game Γ is* recursive *if there exist K^0 and K^*, a partition of K such that:*

- *the stage payoff is 0 on active states: $\forall (k, i, j) \in K^0 \times I \times J$, $g(k, i, j) = 0$.*
- *states in K^* are absorbing: $\forall k \in K^*$, $\sum_{c \in C, d \in D} q(k, i, j)(k, c, d) = 1$ for all $(i, j) \in I \times J$ and $g(k, i, j)$ depends only on k.*

In the rest of the paper, a recursive repeated game will be called a recursive games with signals.

Definition 6.5.2. *Player 1 is* more informed *than player 2 in the recursive game Γ if there exists a mapping $\hat{d} : C \to D$ such that, if E denotes $\{(k, c, d) \in K \times C \times D,\ \hat{d}(c) = d\}$, then*

$$q(k, i, j)(E) = 1,\ \forall (k, i, j) \in K \times I \times J.$$

Notation 6.5.3. *We denote by:* $\Delta^1(K \times C \times D) = \{\pi | \pi(E) = 1\}$.

We define similarly that player 2 is more informed than player 1. Whenever player 1 is more informed than player 2 and player 2 is more informed than player 1, Γ is a repeated game with *symmetric signals*. We denote by $\Delta^*(K \times C \times D)$ the set of symmetric initial distributions.

Remark 6.5.4. *By assumption, if player 1 is more informed than player* 2*, he learns especially the action played by player* 2 *since it is included in the signal of player* 2*. Player 2 is in general not informed of the action played by player 1.*

In Gensbittel et al. [21], the authors considered a weaker notion of "a more informed player" but they made a different assumption on the transition function, especially that the less informed player has no influence on the evolution of beliefs of both players. It is not clear if our result still holds under this weaker assumption.

6.5.2 Evaluation

At stage t, the space of past histories of player 1 is $H_t^1 = (C \times I)^{t-1} \times C$ and the space of past histories of player 2 is $H_t^2 = (D \times J)^{t-1} \times D$. Set $H_\infty = (K \times C \times D \times I \times J)^\infty$ to be the space of infinite *plays*. For any play $h = (k_s, c_s, d_s, i_s, j_s)_{s \geq 1}$, we denote by h_t its projection on H_t, by h_t^1 its projection on H_t^1, and by h_t^2 its projection on H_t^2.

A *(behavior) strategy* for player 1 is a sequence $(\sigma_t)_{t \geq 1}$ of functions $\sigma_t : H_t^1 \to \Delta(I)$. A *(behavior) strategy* for player 2 is a sequence $\tau = (\tau_t)_{t \geq 1}$ of functions $\tau_t : H_t^2 \to \Delta(J)$. We denote by Σ and $\mathcal{T}$ player's respective sets of strategies. An initial distribution $\pi \in \Delta(K \times C \times D)$ and a couple of strategies (σ, τ) define a probability over the set of infinite plays, which we denote by $\mathbb{P}^\pi_{\sigma,\tau}$.

For any given $\pi \in \Delta(K \times C \times D)$, let $\gamma_n(\pi, \sigma, \tau)$ (resp. $\gamma_\lambda(\pi, \sigma, \tau)$) be the expected n-stage payoff (resp. λ-discounted payoff) associated with $(\sigma, \tau) \in \Sigma \times \mathcal{T}$. We denoted by $v_n(\pi)$ the n-stage value and by $v_\lambda(\pi)$ the λ-discounted value.

Definition 6.5.5. *Given an initial distribution* $\pi \in \Delta(K \times C \times D)$*, the game* $\Gamma(\pi)$ *has an asymptotic value* $v(\pi)$ *if:*

$$v(\pi) = \lim_{n \to \infty} v_n(\pi) = \lim_{\lambda \to 0} v_\lambda(\pi).$$

Definition 6.5.6. *Given an initial distribution* $\pi \in \Delta(K \times C \times D)$*, the game* $\Gamma(\pi)$ *has a* uniform maxmin $\underline{v}_\infty(\pi)$ *if:*

- *Player* 1 *can* guarantee $\underline{v}_\infty(\pi)$*, i.e. for all* $\varepsilon > 0$ *there exists a strategy* $\sigma^* \in \Sigma$ *of player* 1 *and* $n_0 \geq 1$ *such that*

$$\forall n \geq n_0,\ \forall \tau \in \mathcal{T},\ \gamma_n(\pi, \sigma^*, \tau) \geq \underline{v}_\infty(\pi) - \varepsilon.$$

- *Player* 2 *can* defend $\underline{v}_\infty(\pi)$ *i.e. for all* $\varepsilon > 0$ *and for every strategy* $\sigma \in \Sigma$ *of player* 1*, there exists* $n_0 \geq 1$ *and* $\tau^* \in \mathcal{T}$ *such that*

$$\forall n \geq n_0,\ \gamma_n(\pi, \sigma, \tau^*) \leq \underline{v}_\infty(\pi) + \varepsilon.$$

The game $\Gamma(\pi)$ has a *uniform minmax* $\overline{v}_\infty(\pi)$ is define similarly if player 2 can guarantee it and player 1 can defend it.

Definition 6.5.7. *Given an initial distribution* $\pi \in \Delta(K \times C \times D)$*, we say that* $\Gamma(\pi)$ *has a uniform value if both* $\bar{v}_\infty(\pi)$ *and* $\underline{v}_\infty(\pi)$ *exist and are equal. Whenever the uniform value exists, we denote it by* $v_\infty(\pi)$*.*

6.5.3 Results

Theorem 6.5.8. *Let Γ be a recursive game such that player 1 is more informed than player 2. Then for every distribution $\pi \in \Delta^1(K \times C \times D)$, both the asymptotic value the uniform maxmin exist and are equal:*

$$\underline{v}_\infty(\pi) = \lim v_n(\pi) = \lim v_\lambda(\pi)$$

By symmetry, we deduce a similar result by exchanging the roles of player 1 and player 2. When the information is symmetric, both results are true and we obtain the existence of the uniform value.

Corollary 6.5.9. *Let Γ be a recursive game with symmetric signals. Then for every $\pi \in \Delta^*(K \times C \times D)$, the game $\Gamma(\pi)$ has a uniform value.*

It is known from Ziliotto [68] that stochastic games with symmetric signals may have no uniform value. Therefore recursive games have very particular properties. It is a challenging task to identify the subclass of repeated games with $\underline{v}_\infty(\pi) = \lim v_n(\pi) = \lim v_\lambda(\pi)$.

Remark 6.5.10. *Note that we have assumed that the stage payoff on absorbing states does not depend on the actions played. Under this assumption, players' strategies have only an influence on non-absorbing plays. Therefore, without loss of generality, we assume in the following that players observe whenever an absorption occurs and in which state it is.*

If we consider that the payoff in absorbing states still depends on the actions played, then our proof does not work. Indeed the auxiliary game introduced in Proposition 6.5.16 is not recursive anymore. The result $\underline{v}_\infty(\pi) = \lim v_n(\pi) = \lim v_\lambda(\pi)$ is unknown for this general case.

Remark 6.5.11. *It is not known whether recursive games with any structure of signals have a uniform value. As highlighted in Rosenberg and Vieille [**?**], the equicontinuity of the λ-discounted value functions is sufficient in order to deduce the existence of the uniform value for recursive games (with perfect observations). For a recursive game with any structure of signals, one can introduce the game associated with a universal belief space but we do not know a metric on this space such that the λ-discounted values or the n-stage values are equicontinuous/totally bounded.*

6.5.4 Proof for Theorem 6.5.8

We introduce some notations concerning different belief hierarchies. Denote by $B_1 = \Delta(K)$ the set of beliefs of player 1 on the state variable. Denote by $B_2 = \Delta_f(B_1) = \Delta_f(\Delta(K))$ the set of beliefs of player 2 on the (first-order) beliefs of player 1. Finally, we denote by $\Delta_f(B_2) = \Delta_f(\Delta_f(\Delta(K)))$ the set of probability distibutions over the second-order beliefs of player 2.

An overview of the proof

A) We fix Γ a recursive game with signals such that player 1 is more informed than player 2. The first subsection presents general properties for repeated games with one player more informed than the other. Given any $\pi \in \Delta^1(K \times C \times D)$, we can define from it a distribution of the beliefs of player 2 on the beliefs of player 1 about the state. This defines a function Φ from $\Delta^1(K \times C \times D)$ to $\Delta_f(B_2)$. Applying results in Gensbittel *et al.* [21], we know that $v_n(\pi)$ depends on π only through $\Phi(\pi)$. This enables us to

show that the value function v_n, defined on $\Delta^1(K \times C \times D)$, induces a canonical function $\hat{v}_n$ defined on B_2 such that $v_n(\pi) = \hat{v}_n(\Phi(\pi))$ and the family $\{\hat{v}_n, n \geq 1\}$ is totally bounded.

B) In the second subsection, we introduce an auxiliary recursive game (with perfect observations) $\mathcal{G}$ which is defined on B_2 and is played with pure actions. We prove in Proposition 6.5.17 that its n-stage game $\mathcal{G}_n$ has a value w_n, which coincides with $\hat{v}_n$ on B_2. A direct consequence is that $\{w_n, n \geq 1\}$ is totally bounded, therefore $\mathcal{G}$ satisfies the conditions of Corollary 6.3.4 and it has a uniform value w_∞. We deduce in $\Gamma(\pi)$ the existene of the asymptotic value given by $w_\infty(\Phi(\pi))$ through the equality $v_n(\pi) = \hat{v}_n(\Phi(\pi)) = w_n(\Phi(\pi))$ (samely for $v_\lambda(\pi)$).

C) The third subsection proves that (cf. Proposition 6.5.22) player 1 can uniformly guarantee $w_\infty(\Phi(\pi))$ in $\Gamma(\pi)$ by mimicking uniform ε-optimal strategies in $\mathcal{G}(\Phi(\pi))$.

D) The last subsection proves that (cf. Proposition 6.5.25) that player 2 can uniformly defend $w_\infty(\Phi(\pi))$ by introducing a second auxiliary recursive game $\mathcal{R}$.

A) Canonical value function $\hat{v}_n$

We follow in this subsection Gensbittel *et al.* [21] to introduce the canonical function $\hat{v}_n$. Note that to obtain results in this subsection, the additional assumption that player 1 controls the transition (made later in their paper) is not used in Gensbittel *et al.* [21].

For convenience, we extend the definition of $\Gamma(\pi)$ to a larger family of initial probability distributions. Given any two finite sets C' and D' and $\pi \in \Delta^1(K \times C' \times D')$, $\Gamma(\pi)$ is the game where (k, c', d') is drawn at stage 1 according to π, player 1 observes c', player 2 observes d' (which is contained in c' $\pi - a.s.$) and then from stage 2 on, the game is played as previously described with signals in C and D.

For any random variable ξ defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathcal{F}$ a sub σ-algebra of $\mathcal{A}$, let $\mathcal{L}_{\mathbb{P}}(\xi \mid \mathcal{F})$ denote the conditional distribution of ξ given $\mathcal{F}$, which is seen as a $\mathcal{F}$-measurable random variable [3] and let $\mathcal{L}_{\mathbb{P}}(\xi)$ denote the distribution of ξ.

Notation 6.5.12. *For every strategy profile $(\sigma, \tau) \in \Sigma \times \mathcal{T}$, we denote the first-order belief of player 1 on K at stage n given h_n^1 by $p_n \in B_1$, the second-order belief of player 2, i.e., his belief about the belief of player 1 on K at stage n given h_n^2 by $x_n \in B_2$, and the distribution of x_n by $\eta_n \in \Delta_f(B_2)$, i.e.,*

$$p_n \triangleq \mathcal{L}_{\mathbb{P}^\pi_{\sigma\tau}}(k_n \mid h_n^1), \quad x_n \triangleq \mathcal{L}_{\mathbb{P}^\pi_{\sigma\tau}}(p_n | h_n^2), \quad \textit{and} \quad \eta_n \triangleq \mathcal{L}_{\mathbb{P}^\pi_{\sigma\tau}}(x_n).$$

Notation 6.5.13. *For any $\pi \in \Delta^1(K \times C' \times D')$ where C' and D' are two finite sets, the* image *of π is given by the following function in $\Delta_f(B_2)$:*

$$\begin{aligned}\Phi(\pi) &\triangleq \mathcal{L}_\pi\left(\mathcal{L}_\pi\left(\mathcal{L}_\pi(k_1|c_1)|d_1\right)\right), \\ &= \sum_{d \in D'} \pi(d) \delta_{\left(\sum_{c \in C'} \pi(c|d) \delta_{\pi(.|c,d)}\right)}.\end{aligned}$$

The interpretation of $\Phi(\pi)$ is as follows: with probability $\pi(d)$, player 2 observes the signal d and beieves that: player 1 received the signal c with probability $\pi(c|d)$ and therefore player 1's belief over K is $\pi(.|c, d)$.

3. All random variables appearing here take only finitely many values so that the definition of conditional laws does not require any additional care about measurability.

The assumptions imply that if $\pi \in \Delta^1(K \times C' \times D')$, then π satisfies the following two properties:

P1) $\pi(c)\pi(k,c,d) = \pi(k,c)\pi(c,d),\ \forall (k,c,d) \in K \times C' \times D'$.

P2) There exists a map $f_1 = f_1^\pi : C' \to \Delta(B_2)$ such that $x_1 = f_1(c_1),\ \pi$-almost surely.

Under P1) and P2), Proposition 1 of Gensbittel *et al.* [21] applies and we obtain the following result, which states that the value of any n-stage game depends on any initial distribution π only through its image $\Phi(\pi)$.

Proposition 6.5.14 (Gensbittel et al. 2014)**.** *Let C' and D' be two finite sets. Let $\pi, \pi' \in \Delta^1(K \times C' \times D')$ and $n \geq 1$. If $\Phi(\pi) = \Phi(\pi')$, then $v_n(\pi) = v_n(\pi')$.*

Reciprocally, given $\eta \in \Delta_f(B_2)$, let us construct a *canonical distribution* π satisfying $\Phi(\pi) = \eta$.

The canonical game $\hat{\Gamma}(\eta)$. Given $\eta \in \Delta_f(B_2)$. Define two finite sets $D' := supp(\eta) \subseteq B_2$ and $C' := D' \times \left(\cup_{x \in supp(\eta)} supp(x)\right)$, and a probability distribution $\pi(\eta) \in \Delta(K \times C' \times D')$ by

$$\forall (k,p) \in K \times \Delta(K),\ x, x' \in B_2,\ \pi(k,(p,x),x') := \begin{cases} \eta(x)x(p)p(k) & \text{if } x = x' \\ 0 & \text{if } x \neq x'. \end{cases}$$

By construction, $\pi(\eta)$ can be seen as an element of $\Delta^1(K \times C' \times D')$, and satisfies $\Phi(\pi(\eta)) = \eta$. The *canonical game* of $\Gamma(\pi)$ is denoted as $\hat{\Gamma}(\eta)$. Its value, denoted by $\hat{v}_n(\eta)$, is equal to $v_n(\pi(\eta))$ the value of $\Gamma_n(\pi(\eta))$. If $\eta = \delta_x$ for some $x \in B_2$, we denote $\hat{v}_n(x)$ for $\hat{v}_n(\delta_x)$.

Informally, the game $\hat{\Gamma}(\eta)$ proceeds as follows: η is common knowledge, player 2 is informed about the realization x of a random variable with law η (player 2 learns his beliefs). Then player 1 is informed about x (his opponent's beliefs) and about the realization p of a random variable with law x (his own beliefs). The state variable is finally chosen according to p, but no player observes it.

By the above construction, one obtains that: $v_n(\pi) = \hat{v}_n\big(\Phi(\pi)\big)$ for any $\pi \in \Delta^1(K \times C' \times D')$.

The result below follows from Proposition 2 of Gensbittel *et al.* [21]. The Wasserstein metric $\mathbf{d}$ on $\Delta(\Delta(K))$ is defined by:

$$\forall x, y \in \Delta(\Delta(K)),\ \mathbf{d}(x,y) = \sup_{f \in \mathcal{D}_1} \left| \int_{\Delta(K)} f(p)x(dp) - \int_{\Delta(K)} f(p)y(dp) \right|,$$

where $\mathcal{D}_1$ is the set of 1-Lipschitz function from $(\Delta(K), \|.\|_1)$ to $[-1,1]$.

Proposition 6.5.15 (Gensbittel et al. 2014)**.** *Let $\eta \in \Delta_f(B_2)$, $n \geq 1$ and $x \in B_2$. Then, $\hat{v}_n(\eta)$ is linear on $\Delta_f(B_2)$ and, as a mapping on B_2, $\hat{v}_n(x)$ is 1-Lipschitz for the Wasserstein metric $\mathbf{d}$.*

Since the state space B_2 is totally bounded for the Wasserstein metric, we deduce by Arzela-Ascoli theorem that the set of functions $\{\hat{v}_n, n \geq 1\}$ is totally bounded.

B) Auxiliary recursive game and asymptotic value

In this subsection we define from Γ an auxiliary game $\mathcal{G}$. We will prove that for every $n \geq 1$, $\mathcal{G}_n$ admits a value in pure strategies and that this value is equal to $\hat{v}_n$. Moreover this game is also recursive. Using Corollary 6.3.4, we deduce the existence of the uniform value in $\mathcal{G}$ and the existence of the asymptotic value in Γ.

Let $\mathcal{G} = (X, A, B, G, \ell)$ be the stochastic game played in pure strategies, defined by:
- the state space $X = \Delta_f(\Delta(K))$ (endowed with the Wasserstein metric $\mathbf{d}$)
- the action space $A = \{f : \Delta(K) \to \Delta(I)\}$ and for all $x \in X$, $A(x) = \{\text{supp}(x) \to \Delta(I)\}$ for player 1
- the action space $B = \Delta(J)$ for player 2
- the payoff function $G : X \to [-1, 1]$, defined for any $x \in X$ by $G(x) := \sum_{p \in \Delta(X)} g(p)x(p)$
- the transition function $\ell : X \times A \times B \to \Delta_f(X)$ defined as $\ell(x, a, b) := \Phi(Q(x, a, b))$. Here, $Q(x, a, b) \in \Delta_f(K \times (\Delta(K) \times C) \times D)$ is the joint distribution of $(k_2, (p, c_2), d_2)$ in the canonical game $\widehat{\Gamma}(\delta_x)$ when the players play $(\sigma_1, \tau_1) = (a, b)$ at stage 1. The sets K, C, D and $\text{supp}(x)$ being finite, Q can be seen as an element in $\Delta^1(K \times C' \times D')$ with C' a finite subset of $\Delta(K) \times C$ and $D' = D$

For any $x \in X$, we denote by $\mathcal{G}(x)$ the game starting at x. We extend the definition to $\mathcal{G}(z)$ for any $z \in \Delta_f(X)$ such that the initial state is chosen randomly along z.

Since players observes when and where absorption occurs, players' beliefs (first and second-order) are either supported on K^0 (therefore respectively in $\Delta(K^0)$ and in $\Delta(\Delta(K^0))$) or supported on each single point $k \in K^*$ (to be δ_k and to be δ_{δ_k}).

Proposition 6.5.16. *Let $X_r = \Delta_f(\Delta(K^0)) \cup \{\delta_{\delta_k} : k \in K^*\}$. The set X_r is a subset of X. The game $\mathcal{G}^r = (X_r, A, B, G, \ell)$ with the state space X_r is well defined and is recursive with the absorbing states $\{\delta_{\delta_k} : k \in K^*\}$.*

In the following, we identify each δ_{δ_k} with k itself for any $k \in K^*$, and write $X_r = \Delta_f(\Delta(K^0)) \cup K^*$. By abuse of notation, we write again X for X_r and $\mathcal{G}$ for $\mathcal{G}^r$.

Proposition 6.5.17. *For every $n \geq 1$, the n-stage game $\mathcal{G}_n$ has a value w_n in pure strategies. Moreover, for every $x \in X$, $w_n(x) = \hat{v}_n(x)$.*

Proof. We prove the result by induction on $n \geq 1$. Let $n = 1$. Given $x \in X$, the game $\mathcal{G}_1(x)$ has a value $w_1(x)$ and it is equal to $w_1(x) = G(x) = \sum_p g(p)x(p)$. It is equal to $\hat{v}_1(x)$ by construction. This initializes our induction. Let $n \geq 1$ such that w_n, the value of $\mathcal{G}_n$, exists in pure strategies, and for every $x \in X$, $w_n(x) = \hat{v}_n(x)$. Gensbittel *et al.* [21] showed in the proof of their Proposition 5 that the family $\{\hat{v}_n, n \geq 1\}$ satisfies the Shapley equation: for every $x \in X$ and for every $n \geq 1$,

$$\begin{aligned}\hat{v}_{n+1}(x) &= \sup_{a \in A(x)} \inf_{b \in B} \mathbb{E}_{\ell(x,a,b)} \left[\frac{1}{n+1} g(x, a, b) + \frac{n}{n+1} \hat{v}_n(x') \right] \\ &= \inf_{b \in B} \sup_{a \in A(x)} \mathbb{E}_{\ell(x,a,b)} \left[\frac{1}{n+1} g(x, a, b) + \frac{n}{n+1} \hat{v}_n(x') \right],\end{aligned}$$

where the random variable $x' \in X$ is chosen along the law $\ell(x, a, b)(\cdot)$. By the inductive assumption, we can replace $\hat{v}_n$ by w_n on the right hand side of above equation, to obtain

that:

$$\forall x \in X, \quad \hat{v}_{n+1}(x) = \sup_{a \in A(x)} \inf_{b \in B} \mathbb{E}_{\ell(x,a,b)} \left[\frac{1}{n+1} g(x) + \frac{n}{n+1} w_n(x') \right]$$
$$= \inf_{b \in B} \sup_{a \in A(x)} \mathbb{E}_{\ell(x,a,b)} \left[\frac{1}{n+1} g(x) + \frac{n}{n+1} w_n(x') \right].$$

We now use the above equation to show that both players can guarantee $\hat{v}_{n+1}(x)$ in $\mathcal{G}_{n+1}(x)$ in pure strategies. Let $x \in X$ be fixed and a^* be an action of player 1 such that

$$\inf_{b \in B} \mathbb{E}_{\ell(x,a^*,b)} \left[\frac{1}{n+1} g(x) + \frac{n}{n+1} w_n(x') \right] \geq \hat{v}_{n+1}(x). \tag{6.5.1}$$

Again by inductive assumption, let $\sigma_n^*(x')$ be an optimal pure strategy in $\mathcal{G}_n(x'), \forall x' \in X$. We define the strategy $\sigma_{n+1}^*(x)$ to play a^* at the first stage and then $\sigma_n^*(x')$ where x' is the current state at stage 2. $\sigma_{n+1}^*(x)$ is pure and guarantees player 1 the payoff in $\mathcal{G}_{n+1}(x)$ no smaller than the left hand side of Equation (6.5.1), hence $\hat{v}_{n+1}(x)$. A similar construction for player 2 finishes the inductive proof. □

Therefore, by Proposition 6.5.15 the family of n-stage values $\{w_n\}$ is totally bounded for the uniform norm, and we can apply Corollary 6.3.4 (with non finite sets of actions) for the game $\mathcal{G}$.

Proposition 6.5.18. *For every $z \in \Delta_f(X)$, the game $\mathcal{G}(z)$ has a uniform value denoted by $w_\infty^*(z)$. Moreover both players can uniformly guarantee the value with pure strategies that depend on the history of states but not on the past actions.*

Since $v_n(\pi) = \hat{v}_n(\Phi(\pi)) = w_n(\Phi(\pi))$, and the same construction of the canonical value function $\hat{v}_\lambda$ implies $v_\lambda(\pi) = \hat{v}_\lambda(\Phi(\pi)) = w_\lambda(\Phi(\pi))$, we deduce the existence of the asymptotic value in the game $\Gamma(\pi)$ for every $\pi \in \Delta^1(K \times C \times D)$:

Proposition 6.5.19. $\lim_{n\to\infty} v_n(\pi) = \lim_{\lambda\to 0} v_\lambda(\pi) = w_\infty^*(\Phi(\pi))$.

C) Player 1 uniformly guarantees w_∞^*

We first show that player 1 is able to compute in the original game $(p_t)_{t\geq 1}$ his first-order beliefs and $(x_t)_{t\geq 1}$ the second-order beliefs of player 2 without knowing the strategy of player 2.

Lemma 6.5.20. *Let (σ, τ) be a couple of strategies in $\Gamma(\pi)$. For every $t \geq 1$, $p_t = \mathcal{L}_{\mathbb{P}^\pi_{\sigma,\tau}}(k_t|h_t^1)$ and $x_t = \mathcal{L}_{\mathbb{P}^\pi_{\sigma,\tau}}(p_t|h_t^2)$ are independent of τ for all h_t^1, h_t^2.*

Proof. Let (σ, τ) be a pair of strategies and $\pi \in \Delta(K \times C \times D)$, we write $\mathbb{P} := \mathbb{P}^\pi_{\sigma,\tau}$ for short. Let $h = (k_s, c_s, d_s, i_s, j_s)_{s\geq 1} \in H_\infty$. For any $t \geq 1$, we define

$$\beta(h_t)\pi(k_1, c_1, d_1) = \prod_{\ell=1}^{t-1} q(k_\ell, i_\ell, j_\ell)(k_{\ell+1}, c_{\ell+1}, d_{\ell+1})$$

with the convention $\beta(k_1, c_1, d_1) = \pi(k_1, c_1, d_1)$. These notations help to write

$$\mathbb{P}(h_t) = \beta(h_t) \prod_{\ell=1}^{t-1} \sigma_t(h_\ell^1)[i_\ell]\tau_\ell(h_\ell^2)[j_\ell].$$

The key point is that under $\mathbb{P}(\cdot)$, i_{t-1}, j_{t-1} and d_t are c_t-measurable whereas j_{t-1} is d_t-measurable. It follows that after observing $(c_1, ..., c_t)$, player 1's belief is:

$$\begin{aligned} p_t(k_t) &= \mathbb{P}\left(k_t|c_1, ..., c_t\right) = \mathbb{P}\left(k_t|c_1, d_1, i_1, j_1, ..., c_t, d_t\right) \\ &= \frac{\sum_{k'_1,...,k'_{t-1}} \mathbb{P}(k'_1, c_1, d_1, i_1, j_1..., k_t, c_t, d_t)}{\sum_{k'_1,...,k'_{t-1},k'_t} \mathbb{P}(k'_1, c_1, d_1, i_1, ..., k'_t, c_t, d_t)} = \frac{\sum_{k'_1,...,k'_{t-1}} \beta(k'_1, c_1, d_1, i_1, j_1, ..., k_t, c_t, d_t)}{\sum_{k'_1,...,k'_{t-1},k_t} \beta(k'_1, c_1, d_1, i_1, j_1, ..., k'_t, c_t, d_t)}, \end{aligned}$$

which depends on neither σ nor τ. We now consider $x_t = \mathcal{L}_{\mathbb{P}}(p_t|d_1, ...d_t)$ for a given observed history $h_t^2 = (d_1, ..., d_t)$ of player 2, which is decomposed as:

$$x_t = \mathcal{L}_{\mathbb{P}}\Big(\mathcal{L}_{\mathbb{P}}(k_{t+1}|c_1, ..., c_t)|d_1, .., d_t\Big) = \sum_{c'_1,...,c'_t} \mathbb{P}(c'_1, ..., c'_t|d_1, ..., d_t)\delta_{\mathcal{L}_{\mathbb{P}}(k_{t+1}|c'_1,...,c'_t)}.$$

By the previous result that $\mathcal{L}_{\mathbb{P}}(k_{t+1}|c'_1, ..., c'_t)$ does not depend on τ, it is sufficient to prove that $\mathbb{P}(c'_1, ..., c'_t|d_1, ..., d_t)$ is independent of τ. Let us consider a sequence of signals $(c'_1, ..., c'_t)$ inducing $(d_1, ..., d_t)$ that we complete with $(i'_1, ..., i'_t)$ the sequence of actions it contains. This gives us

$$\begin{aligned} \mathbb{P}(c'_1, ..., c'_t|d_1, ..., d_t) &= \mathbb{P}(c'_1, d_1, i'_1, j_1, ..., c'_t, d_t|d_1, j_1, ..., d_t) = \frac{\mathbb{P}(c'_1, d_1, i'_1, j_1, ..., c'_t, d_t)}{\mathbb{P}(d_1, j_1, ..., d_t)} \\ &= \frac{\sum_{k'_1,...,k'_t} \beta(h'_t) \prod_{\ell=1}^{t-1} \sigma_\ell(h'^1_\ell)[i'_\ell]}{\sum_{k'_1,...,k'_t} \sum_{c'_1,...,c'_t} \beta(h'_t) \prod_{\ell=1}^{t-1} \sigma_\ell(h'^1_\ell)[i'_\ell]}, \end{aligned}$$

where $h'_\ell = (k'_1, c'_1, d_1, i'_1, j_1, ..., k'_\ell, c'_\ell, d_\ell)$ is the history of stage ℓ and $h'^1_\ell = (c'_1, i'_1, ..., c'_\ell)$ is the private history of player 1 of stage ℓ. The right hand side of the above equation does not depend on the strategy of player 2 and the result is obtained. □

Before building the strategy of player 1, we prove that the transition rule $\ell(\cdot) : X \times A \times B \to \Delta_f(X)$ of the auxiliary game is linear with respect to B (the action of player 2).

Lemma 6.5.21. *For any $(x, a) \in X \times A$ and $\bar{b} = \sum_{s\in S} \lambda_s b_s$ a convex combination in B, we have*

$$\ell(x, a, \bar{b}) = \ell\left(x, a, \sum_{s\in S} \lambda_s b_s\right) = \sum_{s\in S} \lambda_s \ell\left(x, a, b_s\right).$$

Proof. Let $(x, a) \in X \times A$ and $b \in B$. Recall that $Q := Q(x, a, b)$ denotes a distribution in $\Delta_f(K \times (\Delta(K) \times C) \times D)$, which can be seen as as an element in $\Delta^1(K \times C' \times D')$ with $C' = supp(x) \times C$ a finite subset of $\Delta(K)$ and $D' = D$ We have by definition of the image mapping $\Phi(\cdot)$: $\ell(x, a, b) = \Phi(Q) = \sum_{d'\in D'} Q(d')\delta_{\mathcal{L}_Q(\mathcal{L}_Q(k|c')|d')}$.

Similarly to the previous lemma, for every $(c', d') = ((p, c), d') \in C' \times D'$, $\mathcal{L}_Q\Big(\mathcal{L}_Q(k|(p, c))|d'\Big)$ does not depend on b. Indeed, the signal (c', d') contains the action $(i_1, j_1) = (\hat{\imath}(c'), \hat{\jmath}(d'))$ and c' contains d' *a.s.* It follows that

$$\mathbb{P}_Q(k|p, c) = \frac{a(p)[\hat{\imath}(c)]b[\hat{\jmath}(c)]q^{K\times C}(p, \hat{\imath}(c), \hat{\jmath}(c))(k, c')}{a(p)[\hat{\imath}(c)]b[\hat{\jmath}(c)]q^{C}(p, \hat{\imath}(c), \hat{\jmath}(c))(c')} = \frac{q^{K\times C}(p, \hat{\imath}(c), \hat{\jmath}(c))(k, c')}{q^{C}(p, \hat{\imath}(c), \hat{\jmath}(c))(c')}$$

and

$$\mathbb{P}_Q(p, c|d') = \frac{x(p)q^{C}(p, a(p), \hat{\jmath}(d'))(c)}{\sum_{p\in supp(x)} x(p)q^{D}(p, a(p), \hat{\jmath}(d'))(d')}.$$

Since these quantities do not depend on b, we will not precise b in the following. The application $Q(x,a,b)$ being linear in b, we can easily deduce the announced result:

$$\begin{aligned}\Phi\left(Q\left(x,a,\bar{b}\right)\right) &= \sum_{d'\in D'} Q\left(x,a,\bar{b}\right)(d')\delta_{\mathcal{L}_{Q(x,a,.)}(\mathcal{L}_{Q(x,a,.)}(k|c')|d')} \\ &= \sum_{d'\in D'}\left(\sum_{s\in S}\lambda_s Q(x,a,b_s)(d')\delta_{\mathcal{L}_{Q(x,a,.)}(\mathcal{L}_{Q(x,a,.)}(k|c')|d')}\right) \\ &= \sum_{s\in S}\lambda_s\Phi(Q(x,a,b_s).\end{aligned}$$

□

We now deduce from the two previous lemmas that player 1 uniformly guarantees $w^*_\infty(\Phi(\pi))$ in the game $\Gamma(\pi)$.

Proposition 6.5.22. *Player 1 uniformly guarantees $w^*_\infty(\Phi(\pi))$ in $\Gamma(\pi)$.*

Proof. Fix any $\varepsilon > 0$. We divide the proof into three steps. First, we define the optimal strategy $\hat{\sigma}$ in $\Gamma(\pi)$. Then we show how to link the distribution over the states in $\mathcal{G}$ to the distribution of second-order beliefs in Γ. Finally, we deduce that the strategy $\hat{\sigma}$ is uniform ε-optimal.

Step I: Defining the strategy.

Consider the auxiliary game $\mathcal{G}(z)$ with $z = \Phi(\pi) \in \Delta_f(X)$. According to Proposition 6.5.18, player 1 has pure uniform ε-optimal strategies which depend on histories only through the states but not the actions. With a slight abuse of notations, there exists $\hat{\sigma}^* : \bigcup_{t=1}^\infty X^t \to A = \{a : \Delta(K) \to \Delta(I)\}$ and $N_0 \geq 1$ such that

$$\hat{\gamma}_n(z,\hat{\sigma}^*,\hat{\tau}) \geq w^*_\infty(z) - \varepsilon \text{ for all } n \geq N_0 \text{ and for all } \hat{\tau} : \bigcup_{t=1}^\infty X^t \to B = \Delta(J)$$

where $\hat{\gamma}_n(z,\hat{\sigma}^*,\hat{\tau})$ is the expected n-stage average payoff in the auxiliary game $\mathcal{G}(z)$ induced by $(z,\hat{\sigma}^*,\hat{\tau})$.

We define the strategy $\sigma^* \in \Sigma$ in the game $\Gamma(\pi)$ such that for any h^1_t,

$$\sigma^*(h^1_t) = \hat{\sigma}^*(x_1,...,x_t)[p_t] \text{ with } p_t = \mathcal{L}_{\mathbb{P}^\pi_{\sigma^*}}(k_t|h^1_t) \text{ and } x_t = \mathcal{L}_{\mathbb{P}^\pi_{\sigma^*}}(p_t|h^2_t).$$

By Lemma 6.5.20, this is a well defined strategy of player 1 since he can compute p_t and x_t at every stage $t \geq 1$. We now check that the strategy σ^* uniformly guarantees $w^*_\infty(z) - \varepsilon$ in $\Gamma(\pi)$.

Step II: Linking the probability law of beliefs

Let $\tau \in \mathcal{T}$ be a strategy in $\Gamma(\pi)$. We define a strategy $\hat{\tau}$ in $\mathcal{G}(\Phi(\pi))$ such that (π,σ^*,τ) and $(\Phi(\pi),\hat{\sigma},\hat{\tau})$ generate the same probability law for $(x_1,...,x_t,...)$. With a slight abuse in notation, we denote by $\hat{\tau}$ the strategy in $\mathcal{G}$ such that for all $(x_1,...,x_t) \in X^t$,

$$\hat{\tau}(x_1,...,x_t) = \sum_{h^2_t\in H^2_t(x_1,...,x_t)} \mathbb{P}^\pi_{\sigma^*,\tau}(h^2_t|x_1,...,x_t)\tau(h^2_t),$$

where $H^2_t(x_1,...,x_t) = \{h^2_t \in H^2_t | \mathcal{L}_{\mathbb{P}^\pi_{\sigma^*,\tau}}(k_l|h^2_\ell) = x_\ell,\ 1 \leq \ell \leq t\}$ denotes the set of player 2's t-stage histories in Γ that induce the beliefs $(x_1,...,x_t)$.

Lemma 6.5.23. *Let σ^* and $\hat{\tau}$ be constructed as above given $\hat{\sigma}^*$ and τ, we have:*

$$\forall t \geq 1,\ \mathcal{L}_{\mathbb{P}^{\pi}_{\sigma^*,\tau}}(x_1, ..., x_t) = \mathcal{L}_{\mathbb{P}^{z}_{\hat{\sigma}^*,\hat{\tau}}}(x_1, ..., x_t).$$

Proof of Lemma 6.5.23: We prove the lemma by induction on $t \geq 1$. For $t = 1$, the law of x_1 is independent of the strategy profile. By definition of the image mapping $\Phi(\cdot)$,

$$\mathcal{L}_{\mathbb{P}^{\pi}_{\sigma^*,\tau}}(x_1) = \mathcal{L}_{\pi}\left(\mathcal{L}_{\pi}(\mathcal{L}_{\pi}(k_1|c_1)|d_1)\right) = \Phi(\pi).$$

As $\Phi(\pi) = z$, the probability law to choose the initial state $x_1 \in X$ in $\mathcal{G}(z)$, $\mathcal{L}_{\mathbb{P}^{z}_{\hat{\sigma}^*,\hat{\tau}}}(x_1) = \Phi(\pi)$.

Suppose now that we have proved that $\mathcal{L}_{\mathbb{P}^{\pi}_{\sigma^*,\tau}}(x_1, ..., x_t) = \mathcal{L}_{\mathbb{P}^{z}_{\hat{\sigma}^*,\hat{\tau}}}(x_1, ..., x_t)$ for some $t \geq 1$. It is then sufficient to prove that conditional on any realization [4] $\tilde{s}_t := (\tilde{x}_1, ..., \tilde{x}_t) \in (B_2)^t$,

$$\mathcal{L}_{\mathbb{P}^{\pi}_{\sigma^*,\tau}}(x_{t+1}|x_1 = \tilde{x}_1, ..., x_t = \tilde{x}_t) = \mathcal{L}_{\mathbb{P}^{z}_{\hat{\sigma}^*,\hat{\tau}}}(x_{t+1}|x_1 = \tilde{x}_1, ..., x_t = \tilde{x}_t).$$

Fix some $\tilde{s}_t = (\tilde{x}_1, ..., \tilde{x}_t) \in X^t$. By definition of $\hat{\tau}$ and the linearity of ℓ showed in Lemma 6.5.21, we know that

$$\begin{aligned}\mathcal{L}_{\mathbb{P}^{z}_{\hat{\sigma}^*,\hat{\tau}}}(x_{t+1}|x_1 = \tilde{x}_1, ..., x_t = \tilde{x}_t) &= \ell(\tilde{x}_t, \hat{\sigma}^*(\tilde{s}_t), \hat{\tau}(\tilde{s}_t)) \\ &= \sum_{h^2_t \in H^2_t(\tilde{s}_t)} \mathbb{P}^{\pi}_{\sigma^*,\tau}(h^2_t|\tilde{s}_t)\ell(\tilde{x}_t, \hat{\sigma}^*(\tilde{s}_t), \tau(h^2_t)). \end{aligned} \tag{6.5.2}$$

By definition of the conditional expectation, we have in Γ,

$$\mathcal{L}_{\mathbb{P}^{\pi}_{\sigma^*,\tau}}(x_{t+1}|x_1 = \tilde{x}_1, ..., x_t = \tilde{x}_t) = \sum_{h^2_t \in H^2_t(\tilde{s}_t)} \mathbb{P}^{\pi}_{\sigma^*,\tau}(h^2_t|\tilde{s}_t)\mathcal{L}_{\mathbb{P}^{\pi}_{\sigma^*,\tau}}(x_{t+1}|h^2_t).$$

Thus, it is sufficient to prove that for every $h^2_t \in H^2_t(\tilde{s}_t)$, $\ell(\tilde{x}_t, \hat{\sigma}^*(\tilde{s}_t), \tau(h^2_t)) = \mathcal{L}_{\mathbb{P}^{\pi}_{\sigma^*,\tau}}(x_{t+1}|h^2_t)$. Let $h^2_t \in H^2_t(\tilde{x}^t)$ and $Q[h^2_t] := Q\left(\tilde{x}_t, \hat{\sigma}^*(\tilde{s}_t), \tau(h^2_t)\right) \in \Delta_f\left(K \times (\Delta(K) \times C) \times D\right)$ the joint distribution of $(k_{t+1}, (p_t, c_{t+1}), d_{t+1})$ in the canonical game $\hat{\Gamma}(\delta_{\tilde{x}_t})$ when $\left(\hat{\sigma}^*(\tilde{s}_t), \tau(h^2_t)\right) \in A \times B$ is played. By definition of the image mapping $\Phi(\cdot)$ and σ^*, we obtain

$$\mathcal{L}_{\mathbb{P}^{\pi}_{\sigma^*,\tau}}(x_{t+1}|h^2_t) = \mathcal{L}_{Q[h^2_t]}\left(\mathcal{L}_{Q[h^2_t]}(\mathcal{L}_{Q[h^2_t]}(k_{t+1}|c_{t+1})|d_{t+1})\right) = \Phi(Q[h^2_t]) = \ell(\tilde{x}_t, \hat{\sigma}^*(\tilde{s}_t), \tau(h^2_t)).$$

□

Step III: Conclusion of the proof

Finally, let us compare the payoffs in both games. If $k^* \in K^*$, we have $G(k^*) = g(k^*) = \mathbb{E}^{\pi}_{\sigma^*,\tau}[g(k_t)|x_t = k^*]$. If $x_t \in \left(\Delta_f\left(\Delta(K^0)\right)\right)$, we have $G(x_t) = 0 = \mathbb{E}^{\pi}_{\sigma^*,\tau}[g(k_t)|x_t]$. It follows that for every $x_t \in X$, we have $G(x_t) = \mathbb{E}^{\pi}_{\sigma^*,\tau}\left[g(k_t)|x_t\right]$. By taking conditional expectation, Lemma 6.5.23 implies that $\mathbb{E}^{z}_{\hat{\sigma}^*,\hat{\tau}}[G(x_t)] = \mathbb{E}^{\pi}_{\sigma^*,\tau}[g(k_t)]$. Since $\hat{\sigma}^*$ is uniform ε-optimal in the auxiliary game $\mathcal{G}\left(\Phi(\pi)\right)$, we obtain

$$\gamma_n(\pi, \sigma^*, \tau) = \hat{\gamma}_n(\Phi(\pi), \hat{\sigma}^*, \hat{\tau}) \geq w^*_{\infty}(\Phi(\pi)) - \varepsilon \text{ for all } n \geq N_0.$$

Therefore, the strategy σ^* uniformly guarantees $w^*_{\infty}(\Phi(\pi)) - \varepsilon$ in $\Gamma(\pi)$. □

4. For this part of the proof it is convenient to differentiate the random variable describing the second order belief (or the state in $\mathcal{G}$) that will be denoted by x_t from its realization denoted by $\tilde{x}_t$.

D) Player 2 uniformly defends w^*_∞

We now prove that player 2 can defend $w^*_\infty(\Phi(\pi)) = \lim v_n(\pi) = \lim v_\lambda(\pi)$. The situation of player 2 is different since he is allowed to know the strategy of player 1. In order to prove this result, we introduce another auxiliary recursive game $\mathcal{R}$.

For any $n \geq 1$, let $H'_n \subseteq H_n$ be the set of n-stage histories such that player 1 can deduce player 2's private signals, and $H^0_n \subseteq H_n$ be the set of n-stage histories containing only non-absorbing states. We consider the following game $\mathcal{R}$ where the set of states is almost the set of distribution over all finite histories. It is defined as follows:

- the state space is $Z = Z_0 \bigcup K^*$ where $Z_0 = \bigcup_{n\geq 1} \Delta(H^0_n \cap H'_n)$,
- the action space of player 1 is $A = \bigcup_{n\geq 1}\{f : H^1_n \to \Delta(I)\}$ and for any $\pi_n \in \Delta(H_n)$, $A(\pi_n) = \{f : H^1_n \to \Delta(I)\}$,
- the action space of player 2 is $B = \bigcup_{n\geq 1}\{f : H^2_n \to \Delta(J)\}$ and for any $\pi_n \in \Delta(H_n)$, $B(\pi_n) = \{f : H^2_n \to \Delta(J)\}$,
- the transition $Q : Z \times A \times B \to \Delta_f(Z)$ is given by:

$$\forall (k^*, a, b) \in K^* \times A \times B,\ Q(k^*, a, b) = \delta_{k^*},$$

and

$$\forall (z, a, b) \in Z_0 \times A \times B,\ Q(z, a, b) = Q^0(z, a, b)\delta_{\pi^0} + \sum_{k\in K^*} Q(z, a, b)(k^*)\delta_{k^*},$$

where $Q(z, a, b)(k^*)$ is the probability of absorption in state k^* at the next stage given by

$$Q(z, a, b)(k^*) = \sum_{h_n,i,j,c,d} z(h_n)a(h^1_n)[i]b(h^2_n)[j]q(k_n, i, j)(k^*, c, d);$$

$Q^0(z, a, b)$ is the probability of no absorption given by

$$Q^0(z, a, b) = \sum_{h_n,i,j,c,d} \sum_{k\in K^0} z(h_n)a(h^1_n)[i]b(h^2_n)[j]q(k_n, i, j)(k, c, d),$$

and $\pi^0 \in Z_0$ is the conditional probability on not having absorbed, *i.e.*,

$$\forall (h_n, k, i, j, c, d) \in H_n \times K \times I \times J \times C \times D,\ \pi^0(h_n, i, j, c, d) = \frac{z(h_n)a(h^1_n)[i]b(h^2_n)[j]q(k_n, i, j)(k, c, d)}{Q^0(z, a, b)}$$

- the stage payoff function $R : Z \times A \times B \to [-1, +1]$ is given by

$$\forall (\pi_n, a, b) \in Z_0 \times A \times B,\ R(\pi_n, a, b) = 0,$$

and

$$\forall (k^*, a, b) \in K^* \times A \times B,\ R(k^*, a, b) = g(k^*).$$

By construction, the game $\mathcal{R}$ is recursive. We denote by $\widetilde{\Sigma}$ (resp. $\widetilde{\mathcal{T}}$) the set of behavior strategy for player 1 (resp. for player 2) in the game $\mathcal{R}$.

Proposition 6.5.24. *For every $\pi \in Z_0$ and every $n \geq 1$, the n-stage game $\mathcal{R}_n(\pi)$ has a value in history independent pure strategies, which is denoted by $\widetilde{v}_n(\pi)$ and*

$$\widetilde{v}_n(\pi) = v_n(\pi).$$

Moreover, if player 2 can uniformly defend some payoff level v in the game $\mathcal{R}(\pi)$ with pure strategies then he can also uniformly defend v in the game $\Gamma(\pi)$.

Proof. First, a strategy σ of player 1 in Γ is a sequence of applications $(\sigma_n)_{n\geq 1}$ such that σ_n is a mapping from H_1^n to $\Delta(I)$. By definition, this is a sequence of actions in the game $\mathcal{R}$, *i.e.*, a history independent pure strategy in $\mathcal{R}$. Similarly, a strategy τ of player 2 in Γ induces a sequence of actions in $\mathcal{R}$. By definition of Q and R, it follows that for every $\pi \in Z_0$, $\sigma \in \Sigma$ and $\tau \in \mathcal{T}$,

$$\gamma_n(\pi, \sigma, \tau) = \widetilde{\gamma}_n(\pi, \sigma, \tau). \tag{6.5.3}$$

Let σ be an optimal strategy of player 1 in the game $\Gamma_n(\pi)$. Consider now a pure strategy $\widetilde{\tau} \in \widetilde{\mathcal{T}}$. The triple $(\pi, \sigma, \widetilde{\tau})$ generates a probability distribution $\mathbb{P}$ on $(Z \times A \times B)^{\mathbb{N}}$ such that there exists at most one play $(\pi_t, a_t, b_t)_{t\geq 1}$ that is non absorbing $\mathbb{P} - a.s.$, *i.e.*, $(\pi_t, a_t, b_t)_{t\geq 1} \in (Z_0 \times A \times B)^{\mathbb{N}}$. Define the strategy $\tau \in \mathcal{T}$ of player 2 in $\Gamma(\pi)$ by setting $\tau_t = b_t$ for all $t \geq 1$. We obtain

$$\widetilde{\gamma}_n(\pi, \sigma, \widetilde{\tau}) = \widetilde{\gamma}_n(\pi, \sigma, \tau) = \gamma_n(\pi, \sigma, \tau) \geq v_n(\pi).$$

Therefore, player 1 guarantees the payoff $v_n(\pi)$ in $\mathcal{R}(\pi)$ with the history independent pure strategy σ. Similarly, player 2 can guarantee $v_n(\pi)$ with a history independent pure strategy and $\widetilde{v}_n(\pi) = v_n(\pi)$.

Finally, let us assume that player 2 can uniformly defend the payoff level v with pure strategies in the game $\mathcal{R}(\pi)$. Let $\varepsilon > 0$ and $\sigma \in \Sigma$. Interpreting σ as an history-independent strategy in $\mathcal{R}$, there exist $N_0 \geq 1$ and a pure strategy $\widetilde{\tau} \in \widetilde{\mathcal{T}}$ such that

$$\forall n \geq N_0,\ \widetilde{\gamma}_n(\pi, \sigma, \widetilde{\tau}) \leq v + \varepsilon. \tag{6.5.4}$$

As in the previous paragraph, we can associate to the triple $(\pi, \sigma, \widetilde{\tau})$ a unique play $(\pi_t, a_t, b_t)_{t\geq 1}$ in $(Z_0 \times A \times B)^{\mathbb{N}}$ and define the strategy $\tau \in \mathcal{T}$ of player 2 in $\Gamma(\pi)$ by setting $\tau_t = b_t$ for all $t \geq 1$. We obtain

$$\forall n \geq N_0,\ \gamma_n(\pi, \sigma, \tau) = \widetilde{\gamma}_n(\pi, \sigma, \tau) = \widetilde{\gamma}_n(\pi, \sigma, \widetilde{\tau}) \leq v + \varepsilon.$$

This proves that player 2 can uniformly defend v in $\Gamma(\pi)$. □

We conclude by showing that the game $\mathcal{R}$ fulfills the conditions of Corollary 6.3.4.

Proposition 6.5.25. *Player 2 uniformly defends* $w_\infty^*(\Phi(\pi))$ *in* $\Gamma(\pi)$.

Proof. We already noticed that the game $\mathcal{R}$ is recursive. Let $\pi \in \Delta(H_n^0 \cap H_n') \subseteq Z_0$ for some $n \geq 1$. Since player 1 is more informed than player 2 (π supported on H_n'), π can be identified as an element in $\Delta^1(K \times C' \times D')$ for some finite C' and D'. By Proposition 6.5.24, we obtain that for any $\pi \in Z_0$,

$$\widetilde{v}_n(\pi) = v_n(\pi) = \hat{v}_n(\Phi(\pi)).$$

According to Corollary 6.5.15, the family $\{\hat{v}_n, n \geq 1\}$ considered as functions on B_2 is totally bounded, and so is the family of their linear extensions to $\Delta_f(B_2)$.

By Corollary 6.3.4, $\mathcal{R}(\pi)$ has a uniform value $w_\infty^*(\Phi(\pi))$ in pure strategies for every $\pi \in \Delta^1(K \times C \times D)$. It follows from Proposition 6.5.24 that player 2 can uniformly defend $w_\infty^*(\Phi(\pi))$ in $\Gamma(\pi)$. □

6.6 Appendix

Lemma 6.6.1. *[Generalized Shapley Equation] Let $n \geq 1$ and $\theta \leq n$ be any stopping time, we have*

$$\begin{aligned} v_n(x_1) &= \max_{\sigma\in\Sigma}\min_{\tau\in\mathcal{T}} \mathbb{E}_{x_1,\sigma,\tau}\left(\frac{1}{n}\sum_{t=1}^{\theta} g(x_t) + \frac{n-\theta}{n} v_{n-\theta}(x_{\theta+1})\right) \\ &= \min_{\tau\in\mathcal{T}}\max_{\sigma\in\Sigma} \mathbb{E}_{x_1,\sigma,\tau}\left(\frac{1}{n}\sum_{t=1}^{\theta} g(x_t) + \frac{n-\theta}{n} v_{n-\theta}(x_{\theta+1})\right). \end{aligned}$$

Proof. We prove that

$$v_n(x_1) \leq \max_{\sigma\in\Sigma}\min_{\tau\in\mathcal{T}} \mathbb{E}_{x_1,\sigma,\tau}\left(\frac{1}{n}\sum_{t=1}^{\theta} g(x_t) + \frac{n-\theta}{n} v_{n-\theta}(x_{\theta+1})\right). \tag{6.6.1}$$

Similarly, one can prove by reversing the role of player 1 and player 2 that the min max is smaller than $v_n(x_1)$. Since the max min is always smaller than the min max, the result follows.

Let σ in Σ be an optimal strategy of player 1 in the game of length n. Fix τ' a best reply of player 2 to σ in the auxiliary game with payoff

$$\mathbb{E}_{x_1,\sigma,\tau}\left(\frac{1}{n}\sum_{t=1}^{\theta} g(x_t) + \frac{n-\theta}{n} v_{n-\theta}(x_{\theta+1})\right).$$

For every history $h \in H_u$ with $u \leq n$, let $\tau^*(h)$ be a best response strategy to $\sigma(h)$ in the remaining game Γ_{n-u} of length $n-u$. For fixed σ in Σ and τ' in $\mathcal{T}$, we shall define the concatenated strategy $\tilde{\tau}$ to follow τ' until the stopping time θ and then to play a best reply $\tau^*(h(\theta))$ to the conditional strategy $\sigma(h(\theta))$. Formally, for any $h \in H_m, m \leq n$, define:

$$\tilde{\tau}(h) = \begin{cases} \tau'(h), & \text{for all } h \in H_m \text{ with } m < \theta(h) \\ \tau^*[h(\theta)](h^\theta), & \text{for all } h \in H_m \text{ with } m \geq \theta(h). \end{cases}$$

Denote by $\tilde{\mathbb{E}}[\cdot]$ the expectation operator $\mathbb{E}_{x_1,\sigma,\tilde{\tau}}[\cdot]$ and by $\mathbb{E}^*[\cdot \mid h(\theta)]$ the conditional expectation operator $\mathbb{E}_{x_1,\sigma,\tau^*}[\cdot \mid h(\theta)]$. We fix any $x_1 \in X$. Since σ is optimal in Γ_n, we have

$$v_n(x_1) \leq \tilde{\mathbb{E}}\left(\frac{1}{n}\sum_{t=1}^{n} g(x_t)\right) \tag{6.6.2}$$

By definition, $\tilde{\tau}(h(\theta)) = \tau^*(h(\theta))$. As the expected n-stage payoff sum can be decomposed conditional on the stopping time θ and history $h(\theta)$, we have

$$\tilde{\mathbb{E}}\left(\sum_{t=1}^{n} g(x_t)\right) = \tilde{\mathbb{E}}\left[\sum_{t=1}^{\theta} g(x_t) + \left(\sum_{t=\theta+1}^{n} g(x_t)\right)\right] = \tilde{\mathbb{E}}\left[\sum_{t=1}^{\theta} g(x_t) + \mathbb{E}^*\left(\sum_{t=\theta+1}^{n} g(x_t) \mid h(\theta)\right)\right]. \tag{6.6.3}$$

Conditional on any $h(\theta)$, $\tau^*[h(\theta)]$ is best response to $\sigma[h(\theta)]$ in the game $\Gamma_{n-\theta}$, thus we have

$$\mathbb{E}^*\left(\sum_{t=\theta+1}^{n} g(x_t) \mid h(\theta)\right) \leq (n-\theta) v_{n-\theta}(x_{\theta+1}),$$

which is substituted into (6.6.3), and together with (6.6.2) to have

$$v_n(x_1) \leq \tilde{\mathbb{E}}\left[\frac{1}{n}\sum_{t=1}^{\theta} g(x_t) + \frac{n-\theta}{n} v_{n-\theta}(x_{\theta+1})\right] \tag{6.6.4}$$

By construction of the strategy $\tilde{\tau}$ the law of history $h(\theta)$ is the same under $(\sigma, \tilde{\tau})$ and under (σ, τ'). Then we have from (6.6.4):

$$v_n(x_1) \leq \tilde{\mathbb{E}}\left[\frac{1}{n}\sum_{t=1}^{\theta} g(x_t) + \frac{n-\theta}{n} v_{n-\theta}(x_{\theta+1})\right] = \mathbb{E}_{x_1,\sigma,\tau'}\left[\frac{1}{n}\sum_{t=1}^{\theta} g(x_t) + \frac{n-\theta}{n} v_{n-\theta}(x_{\theta+1})\right]$$

By definition, τ' is best response to σ in the auxiliary game, thus we have

$$\begin{aligned} v_n(x_1) \leq & \mathbb{E}_{x_1,\sigma,\tau'}\left[\frac{1}{n}\sum_{t=1}^{\theta} g(x_t) + \frac{n-\theta}{n} v_{n-\theta}(x_{\theta+1})\right] \\ \leq & \inf_{\tau \in T} \mathbb{E}_{x_1,\sigma,\tau}\left[\frac{1}{n}\sum_{t=1}^{\theta} g(x_t) + \frac{n-\theta}{n} v_{n-\theta}(x_{\theta+1})\right], \end{aligned}$$

which proves (6.6.1) thus the proof is finished. □

Acknowledgements The authors thank Sylvain Sorin for his careful reading of earlier versions of this paper, whose comments have significantly improved its presentation. The authors also thank an associated editor and an anonymous referee (of *International Journal of Game Theory*) for their numeruous helpful remarks. The authors gratefully acknowledge the support of the Agence Nationale de la Recherche, under grant ANR JEUDY, ANR-10-BLAN 0112.

Bibliography

[1] O. Alvarez and M. Bardi. Ergodicity, stabilization, and singular pertubations for Bellman-Isaacs equations. *Memoirs of the Amer. Math. Soc.*, 204:1–90, 2010.

[2] M. Arisawa. Ergodic problem for the Hamilton-Jacobi-Belmann equations II. *Ann. Inst. Henri Poincaré, Analyse Nonlinéaire*, 15:1–24, 1998.

[3] M. Arisawa and P.L. Lions. On ergodic stochastic control. *Comm. Partial Differential Equations*, 23:2187–2217, 1998.

[4] R. Aumann. Mixed and behavior strategies in infinite extensive games. *Advances in Game Theory, Annals of Mathematics Studies 52, M. Dresher, L. S. Shapley, and A. W. Tucker (eds.)*, 627–650, 1964.

[5] R. Aumann, M. Maschler, and R.E. Stearns. *Repeated Games with Incomplete Information.* The MIT Press, 1995.

[6] M. Bardi and F. Priuli. LQG Mean-Field Games with ergodic cost. *Proc. 52nd IEEE Conference on Decision and Control*, pages 2493–2498, 2013.

[7] R. Bellman. The theory of dynamic programming. *Bulletin of the American Mathematical Society*, 60:503–516, 1954.

[8] R. Bellman. A Markovian decision process. *Journal of Mathematics and Mechanics*, 6:679–684, 1957.

[9] A. Bensoussan. *Perturbation Methods in Optimal Control.* Wiley/Gauthiers-Villas, Chichester, 1988.

[10] T. Bewley and E. Kohlberg. The asymptotic theory of stochastic games. *Mathematics of Operations Research*, 1:197–208, 1976.

[11] D. Blackwell. An analog of the minmax theorem for vector payoffs. *Pacific Journal of Mathematics*, 6:1–8, 1956.

[12] D. Blackwell. Discrete dynamic programming. *Annals of Mathematical Statistics*, 33:719–726, 1962.

[13] D. Blackwell and T. S. Ferguson. The big match. *Annals of Mathematical Statistics*, 39:159–163, 1968.

[14] R Buckdahn, D. Goreac, and M. Quincampoix. Existence of asymptotic values for nonexpansive stochastic control systems. *Applied Mathematics and Optimization*, 70:1–28, 2014.

[15] P. Cardaliaguet. Differential games with asymmetric information. *SIAM Journal on Control and Optimization*, 46:816–838, 2007.

[16] P. Cardaliaguet, R. Laraki, and S. Sorin. A continuous time approach for the asymptotic value in two-person zero-sum repeated games. *SIAM Journal on Control and Optimization*, 50:1573–1596, 2012.

[17] M. Coulomb. Games with a recursive structure. *Stochastic Games and Applications (Chapter 28), A. Neyman and S. Sorin (eds.), NATO Science Series C, Mathematical and Physical Sciences*, pages 427–442, 2003.

[18] H. Everett. Recursive games. *Contributions to the Theory of Games III, M. Dresher, A.W. Tucker and P. Wolfe (eds.), Annales of Mathematical Studies 39, Princeton University Press*, 47–78, 1957.

[19] F. Forges. Communication equilibra in repeated games with incomplete information. *Mathematics of Operations Research*, 13:191–231, 1988.

[20] V. Gaitsgory. On the use of the averaging method in control problems. *(Russian) Differentsialnye Uravneniya*, 22:1876–1886, 1986.

[21] F. Gensbittel, M. Oliu-Barton, and X. Venel. Existence of the uniform value in repeated games with a more informed controller. *Journal of Dynamics and Games*, 1:411–445, 2014.

[22] D. Gillette. Stochastic games with zero stop probabilities. *Contributions to the Theory of Games III, M. Dresher, A.W. Tucker and P. Wolfe (eds.), Annales of Mathematical Studies 39, Princeton University Press*, 178–187, 1957.

[23] D. Goreac. A note on general tauberian-type results for controlled stochastic dynamics. *Preprint, hal:01120513*, 2015.

[24] R.Z. Khasminskii. On the averaging principle for Itô stochastic equations. *Kybernetika*, 4:260–279, 1968.

[25] D. Khlopin. On uniform tauberian theorems for dynamic games. *arXiv:1412.7331*, 2015.

[26] E. Kohlberg. Repeated games with absorbing states. *The Annals of Statistics*, 2:724–738, 1974.

[27] E. Lehrer and D. Monderer. Discounting versus averaging in dynamic programming. *Games and Economic Behavior*, 6:97–113, 1994.

[28] E. Lehrer and S. Sorin. A uniform Tauberian theorem in dynamic programming. *Mathematics of Operations Research*, 17:303–307, 1992.

[29] X. Li. Limit value for compact nonexpansive optimal control with general evaluations. *Preprint*, 2015.

[30] X. Li, M. Quincampoix, and J. Renault. Limit value for optimal control with general means. *arXiv:1503.05238*, 2015.

[31] X. Li and X. Venel. Recursive games: uniform value, Tauberian theorem and Mertens conjecture. *arXiv:1506.00949*, 2015.

[32] S.A. Lippman. Criterion equivalence in discrete dynamic programming. *Operations Research*, 17:920–923, 1968.

[33] J.-F. Mertens. Repeated games. In *Proceedings of the International Congress of Mathematicians, (Berkeley, 1986)*. American Mathematical Society, 1528–1577, 1987.

[34] J.-F. Mertens and A. Neyman. Stochastic games. *International Journal of Game Theory*, 10:53–66, 1981.

[35] J.-F. Mertens, S. Sorin, and S. Zamir. *Repeated Games*. Cambridge University Press, 2015.

[36] J.-F. Mertens and S. Zamir. The value of two-person zero-sum repeated games with lack of information on both sides. *International Journal of Game Theory*, 1:39–64, 1971.

[37] J.-F. Mertens and S. Zamir. On a repeated game without a recursive structure. *International Journal of Game Theory*, 5:173–182, 1976.

[38] D. Monderer and S. Sorin. Asymptotic properties in dynamic programming. *International Journal of Game Theory*, 22:1–11, 1993.

[39] A. Neyman and S. Sorin. Equilibria in repeated games of incomplete information: the general symmetric case. *International Journal of Game Theory*, 27:201–210, 1998.

[40] M. Oliu-Barton and G. Vigeral. A uniform Tauberian theorem in optimal control. *Advances in Dynamic Games, Annals of the International Society of Dynamic Games, P. Cardaliaguet and R. Cressman (eds.)*, 12:199–215, 2013.

[41] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko. *The Mathematical Theory of Optimal Processes.* Wiley, New York, 1962.

[42] M. Quincampoix and J. Renault. On the existence of a limit value in some nonexpansive optimal control problems. *SIAM Journal on Control and Optimization*, 49:2118, 2011.

[43] J. Renault. The value of Markov chain games with lack of information on one side. *Mathematics of Operations Research*, 31:490–512, 2006.

[44] J. Renault. Uniform value in dynamic programming. *Journal of the European Mathematical Society*, 13:309–330, 2011.

[45] J. Renault. The value of repeated games with an informed controller. *Mathematics of Operations Research*, 37:154–179, 2012.

[46] J. Renault. General limit value in dynamic programming. *Journal of Dynamics and Games*, 1:471–484, 2014.

[47] J. Renault and X. Venel. A distance on some probability spaces, with applications to Markov decision processes and repeated games. *hal-00674998*, 2013.

[48] D. Rosenberg. Zero sum absorbing games with incomplete information on one side: asymptotic analysis. *SIAM Journal on Control and Optimization*, 39:208–225, 2000.

[49] D. Rosenberg, E. Solan, and N. Vieille. Blackwell optimality in Markov decision processes with partial observation. *The Annals of Statistics*, 30:1178–1193, 2002.

[50] D. Rosenberg, E. Solan, and N. Vieille. Stochastic games with a single controller and incomplete information. *SIAM Journal of Control and Optimization*, 43:86–110, 2004.

[51] D. Rosenberg and S. Sorin. An operator approach to zero-sum repeated games. *Israel Journal of Mathematics*, 121:221–246, 2001.

[52] D. Rosenberg and N. Vieille. The maxmin of recursive games with incomplete information on one side. *Mathematics of Operations Research*, 25:23–35, 2000.

[53] L.S. Shapley. Stochastic games. *Proc. Nat. Acad. Sci.*, 39:1095–1100, 1953.

[54] E. Solan and N. Vieille. Uniform value in recursive games. *The Annals of Applied Probability*, 12:1185–1201, 2002.

[55] S. Sorin. "Big match" with lack of information on one side (i). *International Journal of Game Theory*, 13:201–255, 1984.

[56] S. Sorin. "Big match" with lack of information on one side (ii). *International Journal of Game Theory*, 14:173–204, 1985.

[57] S. Sorin. On a repeated game with state dependent signalling matrices. *International Journal of Game Theory*, 14:249–272, 1985.

[58] S. Sorin. On repeated games without a recursive structure: existence of lim v(n). *International Journal of Game Theory*, 18:45–55, 1989.

[59] S. Sorin. *A First Course on Zero-sum Repeated Games*. Springer, 2002.

[60] S. Sorin. Stochastic games with incomplete information. *Stochastic Games and Applications (Chapter 25), A. Neyman and S. Sorin (eds.), NATO Science Series C, Mathematical and Physical Sciences*, 375–395, 2003.

[61] S. Sorin and S. Zamir. "Big match" with lack of information on one side (iii). *Stochastic Games and Releated Topics, T.E. S. Raghavan et al.(eds.)*, 101–112, 1991.

[62] X. Venel and B. Ziliotto. Existence of pure epsilon-optimal strategies in gambling houses. *Preprint*, 2015.

[63] G. Vigeral. *Propritétés asymptotiques des jeux répétés à somme nulle*. Ph.D. thesis of Université Paris-6, 2009.

[64] C. Villani. *Optimal Transportation: Old and New*. Springer, 2009.

[65] O.J. Vrieze and F. Thuijsman. On equilibria in repeated games with absorbing states. *International Journal of Games Theory*, 18:293–310, 1989.

[66] D.H. Wagner. Survey of measurable selection theorems. *SIAM Journal on Control and Optimization*, 15:859–903.

[67] C. Waternaux. Solution of a class of repeated games without a recursive structure. *International Journal of Game Theory*, 12:129–160, 1983.

[68] B. Ziliotto. Zero-sum repeated games: counterexamples to the existence of the asymptotic value and the conjecture maxmin= lim v (n). *arXiv:1305.4778*, 2013.

[69] B. Ziliotto. General limit value in stochastic games. *arXiv:1410.5231*, 2014.

[70] B. Ziliotto. A Tauberian theorem for nonexpansive operators and applications to zero-sum stochastic games. *arXiv:1501.06525*, 2015.