Title
Stochastic Games: Nash Equilibrium, Pareto Optimality, Price of Anarchy, and Learning

Author
Xu, Renyuan

Publication Date
2019

Peer reviewed|Thesis/dissertation

Stochastic Games: Nash Equilibrium, Pareto Optimality, Price of Anarchy, and Learning

by

Renyuan Xu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Industrial Engineering and Operations Research

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Xin Guo, Chair
Professor Ilan Adler
Professor James Pitman

Summer 2019

Stochastic Games: Nash Equilibrium, Pareto Optimality, Price of Anarchy, and Learning

Abstract

Stochastic Games: Nash Equilibrium, Pareto Optimality, Price of Anarchy, and Learning

by

Renyuan Xu

Doctor of Philosophy in Engineering - Industrial Engineering and Operations Research

University of California, Berkeley

Professor Xin Guo, Chair

Stochastic games with large populations are notoriously difficult to solve due to their intractability and dimensionality. How to analyze game strategies under full information and how to design efficient learning algorithms under partial or no information are among the key questions that need to be answered in order to better understanding such complex stochastic systems.

In this thesis, we provide some attempts to tackle these two questions.

First, we formulate and analyze an N-player stochastic game of the classical fuel follower problem and its mean field game (MFG) counterpart. For the N-player game, we obtain the Nash equilibrium (NE) explicitly by deriving and analyzing a system of Hamilton–Jacobi Bellman (HJB) equations and by establishing the existence of a unique strong solution to the associated Skorokhod problem on an unbounded polyhedron with an oblique reflection. For the MFG, we derive a bang-bang type NE under some mild technical conditions and by the viscosity solution approach. We also show that this solution is an ϵ-NE to the N-player game, with $\epsilon = O\left(\sqrt{\frac{1}{N}}\right)$. The N-player game and the MFG differ in that the NE for the former is state dependent while the NE for the latter is a threshold-type bang-bang policy where the threshold is state independent. Our analysis shows that the NE for a stationary MFG may not be the NE for the corresponding MFG.

Second, we propose a class of stochastic N-player games and discuss its connection to the free boundary problems, where both the associated fully nonlinear partial differential equations (PDEs) and the boundaries separating the action and waiting regions are integral parts of the problems. We show how "moving" boundaries come into play due to the game nature, which distinguishes our results from the existing single-agent literature. We present explicit NE by solving a sequence of Skorokhod problems. For the special case of resource allocation problems, we show how players change their strategies based on different network structures among players and resources, with insights from a sharing economy perspective.

Third, we analyze the Pareto optimality (PO) solution for a class of N-player collaborative games. This is achieved by connecting the collaborative game with an auxiliary central controller problem. The main contributions to solving the auxiliary central controller problem are two-fold. First, we show the regularity $\mathcal{W}^{2,\infty}(\mathbb{R}^N)$ of the central controller's value function, which is the unique solution to a high-dimensional HJB equation with complex gradient constraints. Second we show the optimal strategy is a sequence of Skorokhod problems, where the regularity of the boundary is $\mathcal{W}^{1,\infty}(\mathbb{R}^N)$. With some properties of the PO solution, we then provide an upper bound on the Price of Anarchy (PoA) of this game, which bridges the set of NEs and the PO solution. Some insights are also discussed when $N = 2$, with explicit solutions and exact PoA values.

Fourth, motivated by the advertisement auction problem for online advertisements, we consider the general problem of simultaneous learning and decision-making in a stochastic game setting with a large population. We formulate this type of game with unknown rewards and dynamics as a generalized mean field game (GMFG), incorporating action distributions. We first analyze the existence of the solution to this GMFG and show that naively combining Q-learning with the three-step fixed-point approach in classical MFGs yields unstable algorithms. We then propose an alternating approximating Q-learning algorithm and establish its convergence property and complexity result. The numerical performance of this new algorithm on the repeated Ad auction problem shows superior computational efficiency.

To my parents: Dali Xu and Shumei Ren

Contents

Contents **ii**

List of Figures **iv**

List of Tables **v**

1 Introduction **1**
1.1 N-player Games 1
1.2 Mean field Limit 5
1.3 Computation and Learning on MFGs 10
1.4 Motivation and Organization. 11

2 Stochastic Games for Fuel Followers Problem: N versus MFG **14**
2.1 Introduction 14
2.2 N-Player Fuel Follower Game 17
2.3 MFG for the Fuel Follower Problem 28
2.4 Relation between the N-player game and the MFG 36
2.5 Discussions 40

3 Stochastic Game with Resource Constraints **46**
3.1 Introduction 46
3.2 Problem Setup 49
3.3 NE Game Solution: Verification Theorem and Skorokhod Problem 53
3.4 Nash Equilibrium for Game $\boldsymbol{C_p}$ 59
3.5 Nash Equilibrium for Game $\boldsymbol{C_d}$ 65
3.6 Nash Equilibrium for game $\boldsymbol{C}$ 70
3.7 Comparing Games $\boldsymbol{C_p}$, $\boldsymbol{C_d}$ and $\boldsymbol{C}$ 73

4 Pareto Optimality and Price of Anarchy **77**
4.1 Pareto Optimality (PO) 77
4.2 PO vs NE via Price of Anarchy (PoA) 94

5 Learning Mean Field Game **103**

5.1 Introduction . . . 103
5.2 Framework of General MFG (GMFG) . . . 104
5.3 Solution for GMFGs . . . 107
5.4 RL Algorithms for GMFGs . . . 108
5.5 Experiment: Repeated Auction Game . . . 111
5.6 Conclusion . . . 114

Bibliography **115**

A Chapter 2 **129**
A.1 The Skorokhod Problem (SP) . . . 129
A.2 Well-posedness of Algorithm **1** . . . 133
A.3 Proof of Proposition 15 . . . 134
A.4 Stationary Mean Field Games (SMFGs) . . . 135

B Chapter 3 **136**
B.1 Sketch proof of Theorem 24 . . . 136
B.2 Satisfiability for Assumptions **A1**-**A5** . . . 139
B.3 The unique positive root to (3.4.9) . . . 142

C Chapter 5 **143**
C.1 Distance Metrics and Completeness . . . 143
C.2 Existence and Uniqueness for Stationary NE of GMFGs . . . 144
C.3 Additional Comments on Assumptions . . . 146
C.4 Proof of Theorems 44 and 53 . . . 147
C.5 Proof of Theorem 45 . . . 147
C.6 Naive Algorithm . . . 150
C.7 GMF-V . . . 150
C.8 More Details for the Experiments . . . 151

List of Figures

2.1 Optimal control of the single player problem 18
2.2 Region partition when $N = 3$ 27
2.3 Convergence of v_N with different discount factors 37
2.4 Four NEs when $N = 2$ 42
2.5 Convergence of c_N with different discount factors 43
2.6 N=2 with different α values 45

3.1 Example of adjacent matrix $\boldsymbol{A}$, relationship between the players and resources when $N = 4$ and $M = 6$. 51
3.2 Case $\boldsymbol{C_p}$: MNEP when $N = 2$. 65
3.3 Case $\boldsymbol{C_d}$: MNEP when $N = 2$. 70
3.4 Comparison of projected evolving boundaries for $\boldsymbol{C_p}$, $\boldsymbol{C_d}$, $\boldsymbol{C}$ when $N = 3$. 75

4.1 Worst NE versus PO in PoA. 100
4.2 Worst NE versus PO in PoA (in game values). 101
4.3 Worst NE versus PO in PoA (in game values). 102

5.2 Convergence with different number of inner iterations. 112
5.3 Convergence with different number of states. 112
5.1 Q-tables: GMF-Q vs. GMF-V. 112
5.4 Fluctuations of Naive Algorithm (30 sample paths). 113
5.5 Learning accuracy based on $C(\boldsymbol{\pi})$. 113

A.1 Sequential jumps at time 0 134

List of Tables

5.1 Q-table with $T_k^{\text{GMF-V}} = 5000$. 111

Acknowledgments

I am profoundly indebted to my Ph.D. advisor, Professor Xin Guo. I am grateful that Xin believed in my potential and accepted me into her group four years ago. Joining Xin's research group was the start of every remarkable moment of my graduate school life. It opened the door to prestigious research opportunities and allowed me to pursue my career goals in academia. Xin is a brilliant scholar with deep and broad knowledge, an acute sense of intuition, and an infinite enthusiasm for inquisition and creation. Xin taught me a lot in applied probability, control theory, and game theory. I will always remember the countless hours we spent in her office discussing research problems. The first two years of my graduate life were gloomy because my English was not sufficiently fluent and I made little progress in research. I am grateful that Xin did not give up on me. I would not be able to achieve what I have today without Xin.

I thank my thesis committee members, Professor Ilan Adler and Professor James Pitman. Professor Adler provided me with many great suggestions throughout the years, and his knowledge of game theory from an optimization perspective inspired many of my ideas. Professor Pitman taught me two courses in probability theory. I enjoyed every moment of his classes. Some of his teaching materials turned out to be extremely helpful for my research projects.

It has been a great honor to work with all my collaborators: Anran Hu, Wenpin Tang, Junzi Zhang, Zhengyuan Zhou, Dr. Robert Almgren, Dr. Charles-Albert Lehalle, Professor Jose Blanchet, Professor Thaleia Zariphopoulou, and my advisor Professor Xin Guo. Each one of them is an inspiration to me.

My gratitude extends to all my fellow lab-mates in our research group for their enthusiasm and willingness to provide both technical expertise and research advice, and to all my friends at Berkeley for many hours of mutual support and invaluable discussions. I especially want to thank Haoyang Cao for the fruitful research discussions and emotional support.

Last, but certainly not least, I thank my parents, Dali Xu and Shumei Ren, for their unconditional love and immeasurable emotional encouragement. I also thank my loving boyfriend, Wei, who provided unending support during my Ph.D. studies.

Chapter 1

Introduction

Game Theory is a branch of applied mathematics originally related to economic and political problems. At the begining, John von Neumann and Oskar Morgenstern [181] studied human behavior while making strategic decisions, with the assumption that these decisions were based in rationality. Over the years, Game Theory has been studied and applied to other areas such as ecology, biology, finance, traffic routing, sports, energy system, and social networks.

Besides the wide applications, there is a growing literature on the theoretical side of Game Theory. Examples include Nash equilibrium (NE) for competitive games versus Pareto optimality (PO) for cooperative games; static games versus dynamic games; computational methods with full information versus learning algorithms with partial information; games with indistinguishable players versus games with major-minor players; games with symmetric information versus games with asymmetric information; two-player zero-sum games versus multi-agent nonzero-sum games.

In short, Game Theory is an important field of study that enables us to better understand individual interactions and decision making. In this thesis, we study the connections and the differences between a general class of N-player stochastic games and its mean field counterpart when $N \to \infty$.

1.1 N-player Games

Let us consider the following general stochastic N-player game. Denote $\boldsymbol{X}_t = (X_t^1, \cdots, X_t^N)$ as the joint dynamics and $\boldsymbol{\alpha}_t = (\alpha_t^1, \cdots, \alpha_t^N)$ as the joint controls from N-players. Assume that the dynamics $\boldsymbol{X}_t$ are governed by the following N-dimensional diffusion process:

$$dX_t^i = b^i(t, \boldsymbol{X}_t, \boldsymbol{\alpha}_t)dt + \sigma^i(t, \boldsymbol{X}_t, \boldsymbol{\alpha}_t)dB_t^i + \sigma^0 B_t^0, \quad X_0^i = x^i, \quad (i = 1, \cdots, N), \tag{1.1.1}$$

where $\boldsymbol{B} := (B^0, B^1, \cdots, B^N)$ is a standard $(N+1)$-dimensional Brownian motion on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, with drift $\boldsymbol{b} := (b^1, \cdots, b^N)$ and volatility $\boldsymbol{\sigma} := (\sigma^0, \sigma^1, \cdots, \sigma^N)$ satisfying appropriate regularity conditions. Here (b^i, σ^i) are deterministic

functions: $[0,T]\times\mathbb{R}^N\times A^N \hookrightarrow \mathbb{R}\times\mathbb{R}$. B^0 is the *common noise* that all players are exposed to. This common noise can model the noise correlation among all players. Each player i's control, α_t^i, is in the control set A with some well-defined conditions, for example, $\mathbb{E}[\int_0^T \alpha_t^i dt] < \infty$.

In the game, each player i tries to minimize the following pay-off function J^i over her control $\alpha^i \in A$:

$$J^i(\boldsymbol{x};\boldsymbol{\alpha}) = \mathbb{E}\left[\int_0^T h^i(t,\boldsymbol{X}_t,\boldsymbol{\alpha}_t)dt + g^i(\boldsymbol{X}_T)\right] \tag{1.1.2}$$

subject to (1.1.1). Here $h^i : [0,T]\times\mathbb{R}^N\times\mathbb{A}^N \hookrightarrow \mathbb{R}$ is the running cost function for player i. Note that h^i depends on the current states $\boldsymbol{X}_t$ and actions $\boldsymbol{\alpha}_t$ of all players. $g^i : [0,T]\times\mathbb{R}^N \hookrightarrow \mathbb{R}$ is the terminal cost function for player i.

Open-loop, Closed-loop, or Feedback Strategies

Depending on the information structure available to the players, there are different types of control strategies players can take. Examples include open-loop strategies ($\boldsymbol{B}_{[0,t]}$), closed-loop strategies ($\boldsymbol{X}_{[0,t]}$), and closed-loop strategies in feedback forms ($\boldsymbol{X}_t$). It is important to distinguish the open-loop and the closed-loop strategies because these two types of strategies lead to very different outcomes for both single-agent problems and stochastic games (Sun, Li, and Yong [174] and Carmona, Fouque, and Sun [42]).

Denote $\mathcal{H}_t^{\boldsymbol{B}} := \sigma(\{B_s^0, B_s^1, \cdots, B_s^N\}_{s\le t})$ as the filtration generated by the noises from the system (1.1.1). Similarly, denote $\mathcal{F}_t^{\boldsymbol{X}} := \sigma(\{X_s^1, \cdots, X_s^N\}_{s\le t})$ as the filtration generated by the state processes from the system (1.1.1). The open-loop control is adapted to the filtration generated by $\mathcal{H}^{\boldsymbol{B}}$, and is allowed to depend upon the initial position $\boldsymbol{X}_{0-} = \boldsymbol{x}$; that is, we have $\alpha_t^i \in \{\mathcal{H}_t^{\boldsymbol{B}} \cup \{\boldsymbol{x}\}\}$. Similarly, for closed-loop controls, $\alpha_t^i \in \mathcal{H}_t^{\boldsymbol{X}}$. This means that controls are made based on the historical information of the state evolution. Among all the closed-loop controls, the closed-loop control in feedback forms is the most popular one. That is, α_t^i only depends on the current state $\boldsymbol{X_t}$ rather than the full history. In practice, this strategy is easier to implement since it does not require machine memory to keep the history information. This closed-loop control in feedback forms is often referred to as *Markov controls*. See Carmona [43] for more discussion on these concepts.

Technically speaking, the existence of NE for open-loop control is equivalent to the existence of a coupled *Forward-Backward Stochastic Differential Equation* (FBSDE) system. The existence of NE for closed-loop control is equivalent to the existence of a *Hamilton–Jacobi–Bellman* (HJB) equation system (Sun, Li and Yong [174]).

Here we mention several references, among many, on the topic of stochastic games with open-loop controls and closed-loop controls. For open-loop controls, Lacker and Zariphopoulou [130] study an N-player game and a mean field game (MFG) for optimal investment problem under relative performance criteria; Chiarolla, Ferrari, and Riedel [56] analyze an N-firm stochastic irreversible investment problem under limited sources; Steg [171] invests

an irreversible investment problem in oligopoly; Ferrari, Riedel, and Steg [80] solve the public good contribution problem under uncertainty. For closed-loop controls, Huang, Malhame, and Caines [104] derive the MFG approximation for the NE of N-player game; Bensoussan and Frehse [22] study an N-player game with risk sensitive payoffs; Bardi and Priuli [11] study an linear-quadratic N-player game and the corresponding MFG with ergodic pay-offs, Bensoussan, Sung, Yam, and Yung [24] develop the analysis for linear-quadratic MFG on finite time-horizon.

NE versus PO and the Price of Anarchy

There are various criteria used to measure the performance of strategies in stochastic games. For instance, NE and PO provide two distinct views, with NE focusing on stability and PO on efficiency. The Price of Anarchy (PoA) provides a bridge between NEs and PO and quantifies how far NEs are from being efficient. For convenience, here we introduce NE, PO, and PoA with Markov controls.

NE. Intuitively, NE is a set of strategies that no player will benefit from deviating from this set of strategies. Therefore, it represents a stationary status of the game. Now, let us introduce the formal definition of NE.

Definition 1 (NE)**.** *A tuple of admissible controls $\boldsymbol{\alpha}^* = (\alpha^{1*}, \dots, \alpha^{N*}) \in A^N$ is a NE of the stochastic game (1.1.2), if for any $i = 1, \dots, N$, $\boldsymbol{X}_0 = \boldsymbol{x}$, and any $\alpha \in A^N$, the following inequality holds,*

$$J^i(\boldsymbol{x}; \boldsymbol{\alpha}^*) \leq J^i\left(\boldsymbol{x}; (\boldsymbol{\alpha}^{-i*}, \alpha^i)\right). \tag{1.1.3}$$

Here strategies α^{i*} and α^i are deterministic functions of time t and $\boldsymbol{X}_t = (X_t^1, \dots, X_t^N)$, with the notation $(\boldsymbol{x}^{-i}, y^i) := (x^1, \cdots, x^{i-1}, y^i, x^{i+1}, \cdots, x^N)$ for any $\boldsymbol{x} \in \mathbb{R}^N$. $J^i(\boldsymbol{x}; \boldsymbol{\alpha}^*)$ is called the NE value associated with $\boldsymbol{\alpha}^*$.

(1.1.3) implies that $\boldsymbol{\alpha}^*$ is a NE if no player has the incentive to deviate from this tuple of strategies. That is, conditioning on $\boldsymbol{\alpha}^{-i*}$, the optimal strategy for player i is to follow strategy α^{i*}.

For open-loop strategies with finite-variational controls, NE has been studied by applying a powerful first-order condition, which is a substitute in non-Markovian frameworks for the HJB equation. We refer to Chiarolla, Ferrari, and Riedel [56] for the social planner problem in a market with N firms and limited resources; to Steg [171] for a capital accumulation game; and to Ferrari, Riedel, and Steg [80] for a public good contribution under uncertainty game.

For closed-loop strategies with finite-variational controls, there are a number of papers on non-zero-sum two-player games with singular controls. By treating one as a controller and the other as a stopper, where the controller minimizes the finite variation process and the stopper decides the optimal time to terminate the game, Karatzas and Li [119] prove the

existence of an NE for the game via a BSDE approach. Hernandez-Hernandez, Simon, and Zervos [99] provide an in-depth analysis of the smoothness of the value function and show that the optimal strategy may not be unique when the controller enjoys a first-move advantage. Kwon and Zhang [129] investigate a game of irreversible investment with singular controls and strategic exit. They characterize a class of market perfect equilibria and identify a set of conditions under which the outcome of the game may be unique despite the multiplicity of the equilibria. De Angelis and Ferrari [68] establish the connection between singular controls and optimal stopping times for a non-zero-sum two-player game. Bensoussan and Frehse [21] consider an N-player game with regular controls and obtain the NE via the maximum principle approach. The closest to our problem setting are those of Mannucci [145] and Hamadene and Mu [96]. They consider the fuel follower problem in a finite-time horizon with a bounded velocity and establish the existence of an NE of a two-player game. The former analyzes a strongly coupled parabolic system and the latter uses the BSDE technique.

PO. PO is a game criterion to measure the *efficiency* of the system when players collaborate to reach the social (global) optimality. This type of collaboration game can be found in social welfare maximization (Bartor [15], Coleman [65], Stiglitz [172]), network resource allocation (Teich, Wallenius, Wallenius, and Zionts [178], Lan, Kao, Chiang, and Sabharwal [135]); and recommendation systems (Ribeiro, Lacerda, Veloso, and Ziviani [162]; and Ortega, Sánchez, Bobadilla, and Gutierréz [156]).

Definition 2 (PO)**.** *$\boldsymbol{\alpha}^* \in A^N$ with pay-off functions $\left(J^1, \cdots, J^N\right)$ is a PO if and only if there does not exist $\boldsymbol{\alpha} \in A^N$ such that*

$$J^i\left(\boldsymbol{x}; \boldsymbol{\alpha}\right) \leq J^i\left(\boldsymbol{x}; \boldsymbol{\alpha}^*\right) \text{ for all } i = 1, \ldots, N,$$

and

$$J^j\left(\boldsymbol{x}; \boldsymbol{\alpha}\right) < J^j\left(\boldsymbol{x}; \boldsymbol{\alpha}^*\right),$$

for some $j \in \{1, \ldots, N\}$. The strategies ξ^{i*} and ξ^i are deterministic functions of time t and $\boldsymbol{X}_t$ for all $i = 1, 2, \cdots, N$.

PO can be solved by considering an auxiliary central controller with cost function $\frac{1}{N}\sum_{i=1}^N J^i$. The central controller can *coordinate* the controls from N players to reach her optimal solution. This coordination forms a PO to the original N-player game. Normally, the central controller problem corresponds to a coupled high-dimensional HJB system for which the explicit solution is difficult to derive.

Compared to NE, PO is a less studied subject for N-player stochastic games with finite-variational controls. For PO with two players, Aïd, Basei, and Pham [3] consider a game between a firm and a consumer in an energy market, and Federico and Pham [78] solve the irreversible investment problem where a social planner aims to control its capacity production in order to fit optimally the random demand of a good. For PO in a general N-player game, Ferrari, Riedel, and Steg [80] studies the public good contribution game among N agents.

PoA. PoA was originally introduced to quantify the inefficiency of selfish behavior in routing games (Roughgarden and Tardos [164], Christodoulou and Koutsoupias [59, 58], and Roughgarden [163]). The game-theoretical methods have found many applications involving resource allocation (Altman and Basar [5], Zhu and Pavel [191], and Altman, Boulogne, El-Azouzi, Jiménez, and Wynter [6]) and ranking competition (Bayraktar and Zhang [17]).

Denote $\mathcal{N} := \{\boldsymbol{\alpha}^* \mid \boldsymbol{\alpha}^* \text{ is an NE strategy of game (1.1.2)}\}$ as the set of all NE strategies. The inefficiency of an NE, compared to socially optimal behavior, is quantified by the so-called *Price of Anarchy* (PoA).

Definition 3 (PoA)**.**

$$PoA(\boldsymbol{\alpha}) = \frac{\sup_{\boldsymbol{\alpha^*}\in\mathcal{N}} \left(\sum_{i=1}^{N} J^i(\boldsymbol{x};\boldsymbol{\alpha}^*)\right)}{\sum_{i=1}^{N} J^i\left(\boldsymbol{x};\widetilde{\boldsymbol{\alpha}}\right)},$$

where $\widetilde{\boldsymbol{\alpha}}$ is the social optimality control policy from the central controller.

For static games, this has been studied in Delarue, Lacker, and Ramanan [133]. For the class of continuous-time stochastic differential games, the PoA for MFG with unique NE has been studied in Carmona, Graves, and Tan [49], Achdou and Lauriere [2], Graber [88] and Cardaliaguet and Rainer [40].

Game under Resource Constraints

In practice, players make decisions with respect to various constraints. The resource limit is one of the practical constraints to be considered. For example, resource stands for budget in the investment problems (Björk, Davis, and Landén [26]), inventory in retailing markets (Olivares and Cachon [155]), production level in energy systems (Dong, Huang, Cai, and Liu [71]), computational power in cloud computing, active loads on smart power grids, and communication speed in multimedia wireless networks (Gao, Lu, Sharma, Squillante, and Bosman [83], Georgiadis, Neely, and Tassiulas [84], Levy, Nagarajarao, Pacifici, Spreitzer, Tantawi, and Youssef [141], Samadi, Mohsenian-Rad, Schober, and Wong [165] and Xiao, Song, and Chen [186]).

Game under resource constraints is an important subject. It is the common interest for players to better understand their opponents' behaviors and for mechanism designers to build good platforms, especially when the resource constraints are changing over time.

1.2 Mean field Limit

For each player i ($i = 1, 2, \cdots, N$), if the game dependency on other players only comes from the empirical distribution $\mu_t^{N,i} := \frac{\sum_{j\neq i}\delta_{X_t^j}}{N-1}$, and each player is controlling an independent Brownian motion, game (1.1.1)-(1.1.2) can be rewritten as ($i = 1, \cdots, N$)

$$J^i(\boldsymbol{x};\boldsymbol{\alpha}) = \mathbb{E}\left[\int_0^T h^i(X_t^i, \alpha_t^i, \mu_t^{N,i})dt + g^i(\mu_T^{N,i})\right] \tag{1.2.1}$$

subject to

$$dX_t^i = b^i(X_t^i, \alpha_t^i, \mu_t^{N,i})dt + \sigma^i(X_t^i, \alpha_t^i, \mu_t^{N,i})dB_t^i, \quad X_0^i = x^i. \tag{1.2.2}$$

Assume all players are identically distributed ($b^i = b, \sigma^i = \sigma, h^i = h$ and $g^i = g$). Let $N \to \infty$. If $\mu_t^{N,i}$ converges, then game (1.2.1)-(1.2.2) becomes the following for *a representative* in the MFG:

$$J(\eta;\alpha) = \mathbb{E}\left[\int_0^T h(X_t^{\mu,\alpha}, \alpha_t, \mu_t)dt + g(X_T^{\mu,\alpha})\right] \tag{1.2.3}$$

subject to

$$dX_t^{\mu,\alpha} = b(X_t^{\mu,\alpha}, \alpha_t, \mu_t)dt + \sigma(X_t^{\mu,\alpha}, \alpha_t, \mu_t)dB_t, \quad X_0 \sim \eta. \tag{1.2.4}$$

Here η is the initial distribution of the population, $\alpha \in A$ is the control, and X_t is the dynamics.

Definition 4 (NE for MFG)**.** *A pair consisting of a control policy and a population distribution* $(\alpha^*, \mu^*) := (\{\alpha^*\}_{0\le t\le T}, \{\mu^*\}_{0\le t\le T})$ *is an NE for MFG* (1.2.3)-(1.2.4) *if the following conditions hold:*

- *(Single-player side): Fix* μ^*, α^* *is optimal for the control problem:* $\alpha^* = \arg\max_{\alpha\in A} J(\eta, \alpha|\mu^*)$
- *(Population side):* $\mu_t^* = Law(X_t^{\mu^*,\alpha^*})$*. That is,* μ_t^* *is the law of* $X_t^{\mu^*,\alpha^*}$ *where* $X_t^{\mu^*,\alpha^*}$ *is under the control* α^**.*

The single-player side condition captures the optimality of α^* when the population side μ^* is fixed. The population side condition ensures the "consistency" of the solution: it guarantees that the state distribution flow of the single player matches the population state distribution flow . A more intuitive way to understand the MFG is via a three-step fixed-point perspective.

- **Step 1 Single-agent optimization**: Let the population distribution $\mu := \{\mu_t\}_{0\le t\le T}$ be fixed. For a representative, MFG (1.2.3) becomes a single-agent control problem, denoted as $P(\mu)$. Next, denote the optimal control and the controlled dynamics of problem $P(\mu)$ as α' and $X_t^{\mu,\alpha'}$, respectively. This procedure leads to a mapping $\Gamma_1 : \mathcal{P}(\mathbb{R}) \hookrightarrow A$ such that $\alpha' = \Gamma_1(\mu)$.

- **Step 2 Consistency from population update**: Since all players are identical in the MFG, they will follow the optimal control α', which leads to the update of the population distribution

$$\mu'_t = Law(X_t^{\mu,\alpha'}). \tag{1.2.5}$$

  This leads to the second mapping $\Gamma_2 : \mathcal{A} \hookrightarrow \mathcal{P}(\mathbb{R})$ such that $\mu' = \Gamma_1(\alpha')$.

- **Step 3 Fixed point**: Denote $\Gamma := \Gamma^2 \circ \Gamma^1 : \mathcal{P}(\mathbb{R}) \hookrightarrow \mathcal{P}(\mathbb{R})$. Then a fixed point population distribution μ^* to the mapping Γ is an NE for MFG.

The theory of MFGs has enjoyed tremendous growth since the pioneering works of Huang, Malhamé, and Caines [105] and Lasry and Lions [138]. The MFG provides a tractable approach to the otherwise challenging N-player stochastic games.

Given the MFG formulation in (1.2.3)-(1.2.4), the natural questions that proceed are how to solve the MFG, and when is the solution unique?

Existence.

In terms of existence of the solution, there are mainly three approaches: the PDE approach, the probability approach, and the relaxed control approach.

PDE Approach. In the PDE approach, under some mild technical conditions, the MFG solution can be described by a coupled system with a *backward HJB equation* describing the *conditional optimality* of value function and a *forward Kolmogorov equation* describing the *evolution flow of population distribution*.

First, the value function $v(t,x)$ of problem $P(\mu)$ (defined in Step 1) under fixed population distribution μ follows the following backward HJB equation:

$$\begin{cases} -\partial_t v(t,x) - \sup_a[\mathcal{L}_{a,\mu_t} v(t,x) + h(x,\mu,a)] = 0, \quad \text{on} \quad (0,T)\times\mathbb{R}^d, \\ v(T,x) = g(x,\mu_T), \end{cases} \tag{1.2.6}$$

where the generator $\mathcal{L}_{a,\mu_t}$ is defined as

$$\mathcal{L}_{a,\nu}\phi(x,t) = b(x,\nu,a)\partial_x\phi(x,t) + \frac{1}{2}\sigma(x,\nu,a)^2\phi^2(t,x),$$

for all $\phi(x,t) \in \mathcal{C}^{2,1}(\mathbb{R}\times\mathbb{R}^+)$.

Second, if the optimal control is on the feedback form of $\alpha_t^* = \hat{\alpha}(t, X_t^\mu, \alpha^*)$, the population distribution $\mu_t = \text{Law}(X_t^{\mu,\alpha^*})$ under the MFG control α^*, which is a fixed-point to the three-step approach, satisfies the following Kolmogorov forward equation:

$$\begin{cases} \partial_t\mu_t(x) = -\partial_x(b(x,\nu,\alpha(\hat{t},x))\mu_t(x)) + \frac{1}{2}\partial_x^2(\sigma(x,\nu,\alpha(\hat{t},x))^2\mu_t(x)) \\ \mu_0 = \eta. \end{cases} \tag{1.2.7}$$

There is an extensive study on the existence of the solution to the coupled system (1.2.6)-(1.2.7) under various conditions (Guéant [90], Guéant, Lasry, and Lions [89] and Bardi [12]).

Probability Approach. For the probability approach, which is also referred to as the *FBSDE approach*, a stochastic (Pontryagin) maximum principle is applied, and the MFG is reduced to a forward-backward SDE system of McKean-Vlasov type. Denote $\hat{\alpha}(x, y, \nu) = \arg\max_{a\in\mathcal{A}} H(x, y, \nu, a)$ where the Hamiltonian is defined as $H(x, y, \nu, a) := b(x, \nu, a)y + h(x, \nu, a)$.

$$\begin{cases} dX_t = b(X_t, Y_t, \hat{\alpha}(X_t, Y_t, \mu_t))dt + \sigma B_t, \\ dY_t = -\partial_x H(X_t, Y_t, \mu_t, \hat{\alpha}(X_t, Y_t, \mu_t))dt + Z_t dB_t, \\ X_0 \sim \eta,\ Y_T = \partial_x g(X_T, \mu_T). \end{cases} \tag{1.2.8}$$

The existence of optimal control $\hat{\alpha}(X_t, Y_t, \mu_t)$ can be proved under some standard differentiability and convexity assumptions, where (X_t, Y_t) follows the dynamics in (1.2.8).

This approach is explored by Carmona and Delarue [45, 46], Carmona, Delarue, and Lachapelle [47] and Carmona and Lacker [50], among many others.

Relaxed Control Approach. In both PDE and probability approaches, a key difficulty comes from the forward-backward nature of the problems. In these situations, a more functional-analytic framework is employed in the relaxed control approach. With this approach, there is no need for precise analysis of the optimal feedback control, and the assumptions can be more relaxed. This method is first explored in the regular control (Lacker [132]), and later generalized to the singular control (Fu and Horst [81]).

Uniqueness

There are mainly two sets of conditions that guarantee the uniqueness of the MFG solutions: the monotonicity condition (Lasry and Lions [136], Guéant, Lasry, and Lions [89], and Cardaliaguet, Delarue, Lasry, and Lions [41]) and a small product of certain Lipchitz constants (Huang, Malhamé, and Caines [105], and Huang, Caines, and Malhamé [103]). The monotonicity condition assumes that it is disadvantageous for players' states to be close to one another. A small product of certain Lipchitz constants, on the other hand, implies small variations of the system, which guarantees the contraction mapping of the three-step procedure in MFG derivation.

Several interesting questions have been raised recently on the *reachability* of the MFGs when the uniqueness condition is *violated*: Which MFG is a limit of a sequence of N-player games and which is not? What is the common property for the MFGs to be a limit of some N-player games? For example, see Lacker [131], Nutz, Martin and Tan [152], Cecchin [55], Delarue and Tchuendom [69].

Comparison between N-player Game and MFGs

However, except for the general result that the NE of an MFG is an ϵ-NE to the N-player game (see, for instance Huang, Malhamé, and Caines [105] and Cardaliaguet, Delarue, Lasry,

and Lions [41] for regular controls and Guo and Joon [93] for singular controls), there are very limited results on comparing the NE of N-player games and MFGs. The exceptions are Carmona, Fouque, and Sun [48] for systemic risks, Nutz and Zhang [153] for competition, Lacker and Zariphopoulou [134] for portfolio management, and Bardi [10] for a linear-quadratic problem. All these results, however, are with regular controls. For MFGs with singular controls, notions of relaxed stochastic maximal principle or relaxed admissible controls have been introduced to establish the existence of optimal controls; see, for instance, Fu and Horst [81], Hu, Øksendal, and Sulem [102], and Zhang [190].

Approximations and Convergence

To justify the MFG system, one can use its solution to construct approximate equilibria for the n-player games. There are two types of convergence.

The first type of convergence is in terms of ϵ_N-NE. Namely, a given mean field equilibrium induces an *approximated* NE for a N-player game with an error term ϵ_N for a large N. For instance, see Huang, Malhamé, and Caines [105] and Cardaliaguet, Delarue, Lasry, and Lions [41] for regular controls, and Guo and Joon [93] for singular controls. Under different sets of technical conditions, different orders can be shown for different types of mean field models. For example, $\epsilon_N = O\left(N^{-1/(d+4)}\right)$ in Carmona and Delarue [46] and $\epsilon_N = O\left(N^{-1/2}\right)$ in Huang, Caines, and Malhamé [103].

Another convergence is on the N-player NEs to the mean field limit. This is a more delicate subject. It has been shown that when MFG is not unique, not all MFG solutions are a limit of N-player NE (Laker [131] and Nutz, Martin, and Tan [152]). Therefore, it is important to study when MFG is meaningful and what are the sufficient conditions to apply.

Extension: Common Noise

MFG with common noise describes the scenario when each player faces not only her *private noise* in the dynamics, but also the *common noise* that all players are exposed to. Mathematically speaking, with the presence of common noise, the dynamics of each individual in the N-player game can be written as

$$dX_t^i = b^i(X_t^i, \alpha_t^i, \mu_t^{N,i})dt + \sigma^i(X_t^i, \alpha_t^i, \mu_t^{N,i})dB_t^i + \sigma^0(X_t^i, \alpha_t^i, \mu_t^{N,i})dB_t^0, \quad X_0^i = x^i. \quad (1.2.9)$$

Here, B_t^i is the private noise that drives the dynamics of player i. B_t^0, independent from B_t^i, is the common noise faced by all players ($i = 1, 2, \cdots, N$). Let $N \to \infty$, each player controls the following dynamics in the MFG regime:

$$dX_t^{\mu,\alpha} = b(X_t^{\mu,\alpha}, \alpha_t, \mu_t)dt + \sigma(X_t^{\mu,\alpha}, \alpha_t, \mu_t)dB_t + \sigma^0(X_t^{\mu,\alpha}, \alpha_t, \mu_t)dB_t^0, \quad X_0 \sim \eta. \quad (1.2.10)$$

Here B_t and B_t^0 are independent, with B_t as the private noise and B_t^0 as the common noise.

On an intuitive level, the solution of MFG with common noise can be derived by dealing with the conditional Law $\mu' = \text{Law}(X^{\mu,\alpha}|B^0)$ instead of the regular Law in (1.2.5).

Carmona, Delarue, and Lacker [54] provide the first analysis of MFG problems with common noise and demonstrate existence and uniqueness. Huang, Jaimungal, and Nourian [106] apply MFG with common noise to an optimal execution problem under competition.

Extension: MFG with Multiple Populations

When there are multiple populations in the system, for example a financial network with large banks and small banks, the participants should not be treated equally. This leads to the study of MFG with major-minor players (Carmona and Zhu [52], Nguyen, Luu, and Huang [151] and Huang, Jaimungal, and Nourian [106]) or MFG with multiple populations (Cirant [63] and Bauso, Pesenti, and Tolotti [16]). Another case worth noticing is MFG with a mixture of competition and collaboration: the intra-population interaction is collaboration and inter-population interaction is competition. This hierarchical game structure has been studied in Bensoussan, Huang, and Laurière [23] and Miller and Pham [149].

1.3 Computation and Learning on MFGs

In practice, sometimes it is difficult for players to achieve the goal of reaching NE. There are two main reasons. The first is due to the lack of computational resources when facing complex systems, even when full information is available. The other major reason is when limited information is available to each player. For example, in game (1.1.1), players may have limited knowledge about the parameters b and σ.

To tackle the first challenge with full information, recent developments from the computational front can equip players with an efficient computational method. For the second challenge with partial information, the key is to design efficient reinforcement learning algorithms to help players make decisions while interacting with the unknown system and competing with other players. We will discuss several relevant studies in detail.

Computational Methods. Cardaliaguet and Hadikhanloo [39] introduce a learning procedure (similar to the Fictitious Play) for these games and show its convergence when the MFG is potential and the model is fully observable. Carmona and Mathieu [51] provide a deep-learning-based approach to the MFG solution.

Reinforcement Learning. There are many real-world problems involving a large number of players and unknown systems. Examples include online auction bidding (Gummadi, Key, and Proutiere [91]), massive multi-player online role-playing games (Jeong, Kang, and Kim [114]), high frequency trading (Lehalle and Mouzouni [140]), and the sharing economy (Hamari, Sjöklint, and Ukkonen [97]). Under such circumstances with partial or unknown information, there are some attempts to design a reinforcement learning algorithm to simultaneously learn the system and make decisions.

On learning large population games with mean field approximations, Yang, Ye, Trivedi, Xu, and Zha [187] focus on inverse reinforcement learning for MFGs without decision-making, Yang, Luo, Li, Zhou, Zhang, and Wang [188] study an multi-agent reinforcement learning (MARL) problem with a first-order mean field approximation term modeling the interaction between one player and all the other finite players, and Kizilkale and Caines [125] and Yin, Mehta, Meyn, and Shanbhag [189] consider model-based adaptive learning for MFGs in specific models (*e.g.*, linear-quadratic and oscillator games). More recently, Subramanian and Mahajan [173] consider reinforcement learning in the classical MFG setting, propose a policy-gradient based algorithm and analyze the so-called local NE. For learning large population games without mean field approximation, see Kapoor [117] and Hernandez-Leal, Kartal, and Taylor [100] and the references therein.

1.4 Motivation and Organization.

As introduced previously, N-player non-zero-sum stochastic games are notoriously difficult to solve. The existence or solvability of the game solution can be translated into the existence or solvability of an HJB system for closed-loop controls or FBSDE system for open-loop controls. Normally, the existence of the high-dimensional highly coupled (stochastic) system is hard to analyze, let alone the analytical solutions.

Recently there has been a surge of interest in MFGs, pioneered by the original developments around 2006 (Huang and Malhamé [104], Lasry and Lions [137, 136] and [139]). With an ingenious aggregation approach, MFGs nicely reduce the complexity of N-player games by focusing on $N \to \infty$. Subsequent research, however, has focused largely on theoretical questions of the existence and uniqueness of solutions for the equations governing the particular stochastic differential mean field games. Moreover, there are undesirable consequences of the MFG aggregation approach, and a growing number of studies (Carmona, Fouque, and Sun [42], Guo and Xu [94], Lacker and Zariphopoulou [130]) point to the risk of using MFGs for analyzing N-player games. For instance, NEs of MFGs tend to collapse to that of a single-player game, offering no or limited insight into the general solution structure of N-player games.

Goal. This thesis takes one step back from the study of the existence and uniqueness of different variations of MFGs. Indeed, the goal is to understand some fundamental questions in game theory via the following four pairs of relationships:

1. **2-player games versus N-player games**: What is the missing piece in the literature that obstructs the solvability of the general N-player game compared with the solvable 2-player game?

2. **N-player game versus MFG**: When is MFG a good approximation to the N-player game and when is it not? In what sense is MFG a good approximation? Under what conditions is MFG not a good approximation to the N-player game?

3. **NE versus PO**: NE is a concept of stable strategies under competition, whereas PO is a notion of efficiency under collaboration. What is the relationship between NE and PO? When an NE solution is not unique, how can we distinguish the NEs and what is the proper criterion?

4. **Computation versus Learning**: In practice, players rarely follow the NE, either because they simply do not know how to calculate it or they do not have full information about the system. When players do not have full information, it is important to design efficient algorithms for players to learn how to make decisions while inferring the system and interacting with other players.

Organization. The rest of this thesis is organized as follows:

In Chapter 2, we formulate and analyze an N-player stochastic game of the classical fuel follower problem and its MFG counterpart. For the N-player game, we obtain the NE explicitly by deriving and analyzing a system of HJB equations and by establishing the existence of a unique strong solution to the associated Skorokhod problem on an unbounded polyhedron with an oblique reflection. For the MFG, we derive a bang-bang type NE under some mild technical conditions and by the viscosity solution approach. We also show that this solution is an ϵ-NE to the N-player game, with $\epsilon = O\left(\sqrt{\frac{1}{N}}\right)$. The N-player game and the MFG differ in that the NE for the former is state dependent while the NE for the latter is threshold-type bang-bang policy where the threshold is state independent. Our analysis shows that the NE for a stationary MFG may not be the NE for the corresponding MFG. This is based on work with Professor Xin Guo (UC Berkeley).

In Chapter 3, we propose and analyze a class of stochastic N-player games with some resource constraints. This class of games includes finite fuel stochastic games as a special case. We first derive sufficient conditions for NE in the form of a verification theorem, which reveals an essential game component regarding the interactions among players. It is an analytical representation of the conditional optimality for NEs, largely missing in the existing literature on stochastic games. The derivation of NEs involves first solving a multi-dimensional free boundary problem and then a Skorokhod problem, where the boundary is "moving" in the sense that it depends on both the changes of the system and the interaction among players in the game. This is based on work with Professor Xin Guo (UC Berkeley) and Dr. Wenpin Tang (UC Berkeley).

In Chapter 4, we analyze the PO solution for a class of N-player stochastic games. This is achieved by connecting this collaborative game with an auxiliary central controller problem. The main difficulties are two-fold. The first difficulty is showing the regularity $\mathcal{W}^{2,\infty}(\mathbb{R}^N)$ of the central controller's value function, which is the unique solution to a high-dimensional HJB equation with complex gradient constraints. The second difficulty is showing the existence of the unique optimal solution where the boundary of the reflection region is of $\mathcal{W}^{1,\infty}(\mathbb{R}^N)$. With some properties of the PO solution, we provide an upper bound of the Price of Anarchy, which bridges the set of NEs and the PO solution. Some insights are also discussed when

$N = 2$, with explicit solutions and exact PoA values. This is based on work with Professor Xin Guo (UC Berkeley).

In Chapter 5, we present a general mean field game (GMFG) framework for simultaneous learning and decision-making in stochastic games with a large population. It first establishes the existence of a unique NE to this GMFG, and explains that naively combining Q-learning with the fixed-point approach in classical MFGs yields unstable algorithms. It then proposes a Q-learning algorithm with Boltzmann policy (GMF-Q), with analysis of convergence property and computational complexity. The experiments on repeated Ad auction problems demonstrate that this GMF-Q algorithm is efficient and robust in terms of convergence and learning accuracy. Moreover, its performance is superior in convergence, stability, and learning ability when compared with existing algorithms for multi-agent reinforcement learning. This is based on work with Professor Xin Guo (UC Berkeley), Anran Hu (UC Berkeley), and Junzi Zhang (Stanford University).

Chapter 2

Stochastic Games for Fuel Followers Problem: N versus MFG

2.1 Introduction

The classic fuel follower problem concerns controlling a single moving object on a real line whose movement is modeled by a standard Brownian motion. The controller controls the position of her object in a possibly non-continuous way, i.e., with singular controls. Her objective is to minimize over an infinite-time horizon, the total amount of control and the total L^2 distance of the object to the origin, with a discount factor. The optimal control derived by Beneš, Shepp, and Witsenhausen [20] is shown to be of a "bang-bang" type. That is, there exists a threshold c such that when the object is within $[-c, c]$, it will be idling; and when it is outside $[-c, c]$, the controller will apply the minimal push needed to bring it back within $[-c, c]$. The controlled dynamics is thus a reflected Brownian motion, with local times at c and $-c$ as a result of the minimal push. This problem has a number of generalizations; see, for example, Karatzas [118], Karatzas and Shreve [121], and Shreve and Soner [168]. In particular, Karatzas [118] derives a similar bang-bang type optimal control when the L^2 distance is relaxed to a class of convex and symmetric functions; see Figure 2.1. Due to its simplicity, the fuel follower problem has many applications and has inspired a number of research topics, including reflected stochastic differential equations and semimartingales, Skorokhod problems, and regularities of fully nonlinear PDEs with gradient constraints. See, for instance, Harrison and Williams [98], Soner and Shreve [169], Varadhan and Williams [179], Williams [185], Dai and Williams [66], Kruk [127], Atar and Budhiraja [8], Budhiraja and Ross [30], Evans [76], and Hynd [107].

Our work. In this paper we formulate and analyze an N-player stochastic game of the fuel follower problem and its Mean Field Game (MFG) counterpart. In the N-player game, there are N controllers and N objects with each controller controlling one object. Each controller minimizes her total amount of control and the total distance of her object to *the center*

of the N objects. The interaction among the N controllers in the game is to ensure that their own objects closely follow each other's movement. We derive the Nash Equilibrium (NE) explicitly (Theorem 9). This result is established in two main steps. The first step is to derive and analyze a system of Hamilton–Jacobi–Bellman (HJB) equations for the value functions and to establish a verification theorem (Theorem 7) for the game. After finding the solution to the HJB system, the second step is to construct a feedback control via proving the existence of a (unique strong) solution to an associated Skorokhod problem on an unbounded polyhedron with an oblique reflection (Theorem 8). For the special case of $N = 2$, we exploit the symmetric structure to obtain multiple NEs; see Figure 2.4.

We then consider the corresponding MFG with $N \to \infty$, where each controller minimizes her total amount of control and the total distance of her object to *the mean position* of all objects. Our approach to analyze this MFG is to study directly the two coupled PDEs, the backward parabolic type HJB equation and the forward Kolmogorov equation. By further exploiting the problem structure, we derive an NE which is of a bang-bang type (Theorem 11). The threshold of this bang-bang type NE is state-independent as in the classical fuel follower problem. We finally discuss the relation between the N-player game and the MFG, and show that this NE to the MFG game is an ϵ-NE to the N-player game (Theorem 18).

Our contribution. In general, there are essential technical difficulties in analyzing N-player stochastic games. The underlying HJB system is high dimensional, the existence of its solution is usually hard to analyze, and deriving explicit solutions is even more challenging. Therefore it is in general infeasible to characterize the equilibrium. In the case of the singular control, the HJB equation is even more complex, with additional gradient constraints coming from possible jumps in the control. For MFGs with singular controls, the Hamiltonian for the underlying stochastic control problem diverges and the classical stochastic maximal principle fails. Moreover, due to the possible non-stationarity of the mean information process, the associated HJB equation is parabolic despite the infinite-time horizon setting, making it even more difficult to analyze the regularity of the value functions or to derive explicit solutions.

To the best of our knowledge, our work is the first to provide a complete characterization of the NEs for both the N-player stochastic game and the MFG in a singular control setting. Our explicit solutions are derived for a class of convex and symmetric functions, without the usual linear-quadratic structure for MFGs with regular controls in Bardi [10], Bardi and Priuli [11], Bensoussan, Sung, Yam, and Yung [24].

Moreover, explicit solutions derived in this paper make it possible to directly compare the structural differences between the MFG and the N-player game. It provides useful insights not only for analyzing general N-player games but also for proper formulations of MFGs. Indeed, MFGs may be very different in nature from N-player games: in the fuel follower problem, the MFG degenerates to a single-player game in the sense that its NE is threshold-type bang-bang policy where the threshold is state independent (Proposition 15 and Proposition 16), while the NEs for the N-player game are state dependent (Theorem 9). The collapse of the MFG to the single player problem (Proposition 15) is a side effect by

the *aggregation* in the MFG formulation: players become more *anticipative* when they are assumed to be identical. Our analysis also shows that the NE for a stationary MFG may not be the NE for the corresponding MFG (Remark 14.1).

There are also some noteworthy economic insights from our analysis. For instance, in the N-player game, we show that when the number of players increases, it is more costly for each player to keep track of other players before making decisions, as players will intervene more frequently due to the increasing complexity of the game. Moreover, the bigger the discount factor α, the less frequent players will intervene. (See Remarks 19.1 and 16.1).

Related work on stochastic games. There are a number of papers on non-zero-sum two-player games with singular controls. By treating one as a controller and the other as a stopper, where the controller minimizes the finite variation process and the stopper decides the optimal time to terminate the game, Karatzas and Li [119] prove the existence of an NE for the game via a BSDE approach. Hernandez-Hernandez, Simon, and Zervos [99] provide an in-depth analysis of the smoothness of the value function and show that the optimal strategy may not be unique when the controller enjoys a first-move advantage. Kwon and Zhang [129] investigate a game of irreversible investment with singular controls and strategic exit. They characterize a class of market perfect equilibria and identify a set of conditions under which the outcome of the game may be unique despite the multiplicity of the equilibria. De Angelis and Ferrari [68] establish the connection between singular controls and optimal stopping times for a non-zero-sum two-player game. Bensoussan and Frehse [21] consider an N-player game with regular controls and obtain the NE via the maximum principle approach. The closest to our problem setting are those of Mannucci [145] and Hamadene and Mu [96]. They consider the fuel follower problem in a finite-time horizon with a bounded velocity, and establish the existence of an NE of a two-player game. The former analyzes a strongly coupled parabolic system and the latter uses the BSDE technique.

Related work on MFGs. The theory of MFGs has enjoyed tremendous growth since the pioneering works of Huang, Malhamé, and Caines [105] and Lasry and Lions [138]. The MFG provides a tractable approach to the otherwise challenging N-player stochastic games. However, except for the general result that the NE of an MFG is an ϵ-Nash equilibrium to the N-player game (see, for instance [105] and Cardaliaguet, Delarue, Lasry, and Lions [41] for regular controls and Guo and Joon [93] for singular controls), there are very limited results on comparing the NE of N-player games and MFGs. The exceptions are Carmona, Fouque, and Sun [48] for systemic risks, Nutz and Zhang [153] for competition, Lacker and Zariphopoulou [134] for portfolio management, and [10]. All these results, however, are with regular controls. For MFGs with singular controls, notions of relaxed stochastic maximal principle or relaxed admissible controls have been introduced to establish the existence of optimal controls; see, for instance, Fu and Horst [81], Hu, Øksendal, and Sulem [102], and Zhang [190].

2.2 N-Player Fuel Follower Game

Preliminary: Single Player

The classic fuel follower problem is as follows. Consider a probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ with a standard Brownian motion $\{B_t\}_{t\geq 0}$. The position of the object X_t is assumed to be

$$X_t = x + B_t + \xi_t^+ - \xi_t^-, \quad X_{0-} = x, \tag{2.2.1}$$

where the pair of control (ξ^+, ξ^-) is a non-decreasing, càdlàg process. The goal of the controller is to solve for the value function $v(x)$ of the following optimization problem,

$$v(x) = \inf_{(\xi^+,\xi^-)\in\mathcal{U}} \mathbb{E}\int_0^\infty e^{-\alpha t}\left[h(X_t)dt + d\check{\xi}_t\right], \tag{2.2.2}$$

where the admissible control set $\mathcal{U}$ is

$$\begin{aligned}\mathcal{U} \quad := \Big\{(\xi_t^+, \xi_t^-) \,|\, &\xi_t^+ \text{ and } \xi_t^- \text{ are } \mathcal{F}^{X_{t-}}\text{-progressively measurable, càdlàg, non-decreasing,} \\ &\text{with } \mathbb{E}\left[\int_0^\infty e^{-\alpha t} d\xi_t^+\right] < \infty, \mathbb{E}\left[\int_0^\infty e^{-\alpha t} d\xi_t^-\right] < \infty, \text{and } \xi_{0-}^+ = \xi_{0-}^- = 0\Big\}.\end{aligned}$$

Here $\alpha > 0$ is a discount factor, $\{\mathcal{F}^{X_t}\}_{t\geq 0}$ is the natural filtration of $\{X_t\}_{t\geq 0}$, and $\check{\xi}_t = \xi_t^+ + \xi_t^-$ is the total accumulative amount of controls up to time t, called "fuel usage", hence the term *fuel follower problem.* In addition, under the assumption

A1: The function $h : \mathbb{R} \to \mathbb{R}$ is assumed to be convex, symmetric, twice differentiable, with $h(0) \geq 0$, $h''(x)$ decreasing on $\mathbb{R}^+$, and $0 < k < h''(x) \leq K$ for some constants $K > k > 0$,

Problem (2.2.2) is solved (see [20] and [118]) by analyzing the associated HJB equation

$$\min\left\{\frac{1}{2}v_{xx}(x) + h(x) - \alpha v(x), 1 - v_x(x), 1 + v_x(x)\right\} = 0, \tag{2.2.3}$$

where v_x and v_{xx} are the first and second order derivatives of v with respect to x, respectively. The optimal control $\{\xi_t^{*+}, \xi_t^{*-}\}_{t\geq 0}$ is shown to be of a bang-bang type given by

$$\begin{aligned}\xi_t^{*+} &= \max\left\{0, \max_{0\leq u\leq t}\left\{-x - B_u + \xi_u^{*-} - c\right\}\right\}, \\ \xi_t^{*-} &= \max\left\{0, \max_{0\leq u\leq t}\left\{x + B_u + \xi_u^{*+} - c\right\}\right\},\end{aligned}$$

where the threshold $c > 0$ is the unique positive solution to

$$\frac{1}{\sqrt{2\alpha}}\tanh\left(c\sqrt{2\alpha}\right) = \frac{p_1'(c) - 1}{p_1''(c)}, \tag{2.2.4}$$

with

$$\begin{aligned} p_1(x) &= \mathbb{E}\left[\int_0^\infty e^{-\alpha t} h(x+B_t)dt\right] \\ &= \frac{1}{\sqrt{2\alpha}}\left(e^{-x\sqrt{2\alpha}}\int_{-\infty}^{x} h(z)e^{z\sqrt{2\alpha}}dz + e^{x\sqrt{2\alpha}}\int_x^\infty h(z)e^{-z\sqrt{2\alpha}}dz\right). \end{aligned}$$

The corresponding value function $v(x) \in \mathcal{C}^2(\mathbb{R})$ is given by

$$v(x) = \begin{cases} -\frac{p_1''(c)\cosh\left(x\sqrt{2\alpha}\right)}{2\alpha\cosh\left(c\sqrt{2\alpha}\right)} + p_1(x), & 0 \le x \le c, \\ v(c) + (x-c), & x \ge c, \\ v(-x), & x < 0. \end{cases} \tag{2.2.5}$$

In other words, it is optimal for the controller to apply a "minimal" push to keep the object within $[-c, c]$. Mathematically, the controlled process is a Brownian motion reflected at the boundaries c and $-c$. The minimal push corresponds to the local time of the Brownian motion at c and $-c$. See Figure 2.1.

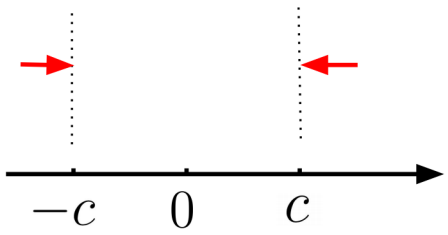


Figure 2.1: Optimal control of the single player problem

N-Player Fuel Follower Game

Now suppose there are N controllers, with each controller controlling one object. For simplicity, let us call such a pair of controller and object a "player". The game is for each player to stay as close as possible to other players.

This N-player game can be formulated as follows. Let $\left(X_t^1, \ldots, X_t^N\right) \in \mathbb{R}^N$ be the positions of players such that for $i = 1, \ldots, N$,

$$X_t^i = x^i + B_t^i + \xi_t^{i,+} - \xi_t^{i,-}, \tag{2.2.6}$$

with $(X_{0-}^1, \cdots, X_{0-}^N) = (x^1, \cdots, x^N) =: \boldsymbol{x}$, where $(B_t^1, \ldots, B_t^N)$ is an N-dimensional standard Brownian motion on $\mathbb{R}^N$. Let $m_t^{(N)} = \frac{\sum_{i=1}^N X_t^i}{N}$ be the center of these N players at time t, with $m_{0-}^{(N)} = \frac{\sum_{i=1}^N x^i}{N}$. Let $h(X_t^i - m_t^{(N)})$ be the distance between player i and the center $m_t^{(N)}$ at time

t. The goal of each player i is to minimize, over all admissible controls $\left(\xi^1, \ldots, \xi^N\right) \in \mathcal{S}_N$, the following payoff function

$$J^i(x^1, \ldots, x^N; \xi^1, \ldots, \xi^N) = \mathbb{E} \int_0^\infty e^{-\alpha_i t} \left[h\left(X_t^i - \rho m_t^{(N)} \right) dt + d\check{\xi}_t^i \right], \qquad \textbf{(N-player)}$$

where $\check{\xi}^i = \xi^{i,+} + \xi^{i,-}$. Here the admissible control set $\mathcal{S}_N$ is defined as

$$\begin{aligned} \mathcal{S}_N := \big\{ (\xi^1, \ldots, \xi^N) \;\big|\; & \xi^j = (\xi^{j,+}, \xi^{j,-}) \in \mathcal{U}_N^j, \mathbb{P}\left(d\xi_t^j(\boldsymbol{x}) d\xi_t^i(\boldsymbol{x}) > 0 \right) = 0, \\ & \text{for any } t > 0, \boldsymbol{x} \in \mathbb{R}^N, i, j \in \{1, \ldots, N\} \text{ and } i \neq j \big\}, \end{aligned} \tag{2.2.7}$$

with

$$\begin{aligned} \mathcal{U}_N^j = \Big\{ (\xi_t^{j,+}, \xi_t^{j,-}) \;\Big|\; & \xi_t^{j,+} \text{ and } \xi_t^{j,-} \text{ are } \mathcal{F}^{(X_{t-}^1, \ldots, X_{t-}^N)}\text{-progressively measurable, càdlàg, non-decreasing,} \\ & \text{with } \mathbb{E}\left[\int_0^\infty e^{-\alpha_j t} d\xi_t^{j,+} \right] < \infty, \;\; \mathbb{E}\left[\int_0^\infty e^{-\alpha_j t} d\xi_t^{j,-} \right] < \infty, \;\; \xi_{0-}^{j,+} = 0, \;\; \xi_{0-}^{j,-} = 0 \Big\}, \end{aligned}$$

where $\alpha_j > 0$ is the discount factor for player j and $\{\mathcal{F}^{(X_t^1, \ldots, X_t^N)}\}_{t \geq 0}$ is the natural filtration of $\{(X_t^1, \ldots, X_t^N)\}_{t \geq 0}$. The condition in Eqn. (4.2.1)

$$\mathbb{P}\left(d\xi_t^i(\boldsymbol{x}) d\xi_t^j(\boldsymbol{x}) > 0 \right) = 0, \qquad \text{for any } \boldsymbol{x} \in \mathbb{R}^N, t \geq 0, i \neq j \tag{2.2.8}$$

is to facilitate designing feasible control policies when controls involve jumps.

Remark 4.1. *Mathematically, one may replace the running cost function $h(X_t^i - m_t^N)$ by $h(X_t^i - \rho m_t^N + \eta)$, with $\rho \geq 0$ indicating the strength of interactions among players as in [103] and [105]. We choose to fix $\rho = 1$ and $\eta = 0$ for clearer model interpretations for the fuel follower problem. Indeed, adding a scaling factor ρ and a constant η will not change the derivation of solutions except for minor notational changes. In fact, as will be shown in Section 2.2 and Appendix A.1, the construction of NEs will be simpler when $\rho \neq 1$.*

Throughout the paper, unless otherwise specified, we will for simplicity and without loss of generality $\alpha_1 = \cdots = \alpha_N = \alpha$. (See Section 3.7 for further sensitivity analysis with respect to α.)

Solution to the N-Player Game

There are various criteria to measure the performance of strategies in stochastic games. For instance, Pareto Optimality (PO) and Nash Equilibrium (NE) provide two distinct views, with NE focusing on stability and PO on efficiency. An NE framework can be further defined depending on the admissible strategies, resulting in open-loop NEs, closed-loop NEs, and the Markovian NEs. See Carmona [43] for more discussions on these concepts.

In this paper, we will focus on the Markovian NE, also known as the closed-loop NE with a feedback form, specified below.

Definition 5. *A tuple of admissible controls $\boldsymbol{\xi}^* = (\xi^{1*}, \dots, \xi^{N*}) \in \mathcal{S}_N$ is a Markovian NE of the stochastic game (**N-player**), if for any $i = 1, \dots, N$, $\boldsymbol{X}_{0-} = \boldsymbol{x}$, and any $(\boldsymbol{\xi}^{-i*}, \xi^i) \in \mathcal{S}_N$, the following inequality holds,*

$$J^i(\boldsymbol{x}; \boldsymbol{\xi}^*) \leq J^i\left(\boldsymbol{x}; (\boldsymbol{\xi}^{-i*}, \xi^i)\right).$$

Here strategies ξ^{i*} and ξ^i are deterministic functions of time t and $\boldsymbol{X}_t = (X_t^1, \dots, X_t^N)$, with the notation $(\boldsymbol{x}^{-i}, y^i) := (x^1, \cdots, x^{i-1}, y^i, x^{i+1}, \cdots, x^N)$ for any $\boldsymbol{x} \in \mathbb{R}^N$. $J^i(\boldsymbol{x}; \boldsymbol{\xi}^*)$ is called the NE value associated with $\boldsymbol{\xi}^*$.

NE Solutions

The NE solution will be derived in two steps. The first is to derive and analyze the associated HJB system. A verification theorem which provides sufficient conditions for the NE values will be presented, along with a solution to the HJB system. The second step is to construct the corresponding NEs, by solving an associated Skorokhod problem.

NE and the HJB System

First,

Definition 6 (*Action* and *waiting* regions)**.** *Player i's* action region $\mathcal{A}_i$ *is defined as*

$$\mathcal{A}_i := \left\{ \boldsymbol{x} \in \mathbb{R}^N \middle| \, d\xi^i(\boldsymbol{x}) \neq 0 \right\},$$

and her waiting region *is $\mathcal{W}_i = \mathbb{R}^N \setminus \mathcal{A}_i$. Denote $\mathcal{A}^{-i} = \cup_{j \neq i} \mathcal{A}_j$ and $\mathcal{W}_{-i} = \cap_{j \neq i} \mathcal{W}_j$.*

Next, a simple heuristic conditional argument via the Dynamic Programming Principle leads to the following HJB system.

Given $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$, for any $i \neq j$,

$$\text{(HJB-N)} \begin{cases} \min\left\{ -\alpha w^i + h\left(\frac{N-1}{N}\left(x^i - \frac{\sum_{j\neq i} x^j}{N-1}\right)\right) + \frac{1}{2}\left(\sum_{j=1}^N w^i_{x^j x^j}\right), 1 - w^i_{x^i}, 1 + w^i_{x^i} \right\} = 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for any } \boldsymbol{x} \in \mathcal{W}_{-i}, \\ w^i_{x^j} = 0, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for any } \boldsymbol{x} \in \mathcal{A}_j, \text{ for any } j \neq i. \end{cases}$$

The derivation of (HJB-N) can be illustrated with the case of $N = 2$. In this case, if $(x^1, x^2) \in \mathcal{A}_2$, $\Delta\xi^{2*} \neq 0$. By the definition of NE, player one is not expected to suffer a loss as otherwise she will have incentives to take actions. Therefore, $w^1(x^1, x^2) = w^1(x^1, x^2 + \Delta\xi^{2*,+} - \Delta\xi^{2*,-})$, letting $\Delta\xi^{2*,\pm} \to 0$, we have $w^1_{x^2} = 0$ in $\mathcal{A}_2$. If $(x^1, x^2) \in \mathcal{W}_2$, $\Delta\xi^{2*} = 0$, then the control problem for player one becomes a classical single player control problem. Therefore, $w^1(x^1, x^2)$ satisfies

$$\min\left\{ -\alpha w^1 + h\left(\frac{x^1 - x^2}{2}\right) + \frac{1}{2}\left(w^1_{x^1 x^1} + w^1_{x^2 x^2}\right), 1 - w^1_{x^1}, 1 + w^1_{x^1} \right\} = 0 \text{ in } \mathcal{W}_2.$$

Here $-\alpha w^1 + h\left(\frac{x^1-x^2}{2}\right) + \frac{1}{2}\left(w^1_{x^1x^1} + w^1_{x^2x^2}\right) = 0$ corresponds to $\Delta\xi^{1*} = 0$, $1 - w^1_{x^1} = 0$ corresponds to $\Delta\xi^{1*,+} > 0$, and $1 + w^1_{x^1} = 0$ corresponds to $\Delta\xi^{1*,-} > 0$. Finally, $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$ ensures Eqn. (2.2.8).

Based on the above HJB system, the following sufficient conditions for an NE can be established.

Theorem 7 (Verification theorem)**.** *For any $i = 1, \dots, N$, suppose $\xi^{i*} \in \mathcal{U}^i_N$ and the corresponding $w^i(.) = J^i(.; \boldsymbol{\xi}^*)$ satisfies the following*

(i) $\boldsymbol{\xi}^* := (\xi^{1*}, \dots, \xi^{N*}) \in \mathcal{S}_N$,

(ii)

$$\min\left\{-\alpha w^i + h\left(\frac{N-1}{N}\left(x^i - \frac{\sum_{j\neq i} x^j}{N-1}\right)\right) + \frac{1}{2}\sum_{j=1}^{N} w^i_{x^jx^j}, 1 - w^i_{x^i}, 1 + w^i_{x^i}\right\} = 0 \quad (2.2.9)$$

for any $\boldsymbol{x} \in \overline{\mathcal{W}_{-i}}$, and

$$w^i_{x^j}(\boldsymbol{x}) = 0,$$

for any $\boldsymbol{x} \in \mathcal{A}_j$.

(iii) (Transversality Condition.) $\limsup_{T\to\infty} \mathbb{E}[e^{-\alpha T} w^i(\boldsymbol{X}_T)] = 0$,

(iv) $w^i(\boldsymbol{x}) \in \mathcal{C}^2(\overline{\mathcal{W}_{-i}})$,

(v) $w^i_{x^j}(\boldsymbol{x})$ *is bounded in* $\overline{\mathcal{W}_{-i}}$*, for any* $j = 1, 2, \cdots, N$,

(vi) there exists a convex function $u^i(\boldsymbol{x}) \in \mathcal{C}^2(\mathbb{R}^N)$ *such that* $u^i(\boldsymbol{x}) = w^i(\boldsymbol{x})$ *on* $\overline{\mathcal{W}_{-i}}$,

(vii) for any $\xi^i \in \mathcal{U}_N$ such that $(\boldsymbol{\xi}^{-i*}, \xi^i) \in \mathcal{S}_N$, the controlled dynamic $(\boldsymbol{X}^{-i*}_t, X^i_t)$ is in $\mathcal{W}_{-i}$ $\mathbb{P}$-a.s. at any time t.

Then $\boldsymbol{\xi}^*$ is an NE with value w^i.

Proof. Given any $\xi^i \in \mathcal{U}^i_N$ such that $(\boldsymbol{\xi}^{-i*}, \xi^i) \in \mathcal{S}_N$, fixing the control $(\xi^{i,+}_t, \xi^{i,-}_t)$ such that

$$\begin{aligned} X^i_t &= x^i + B^i_t + \xi^{i,+}_t - \xi^{i,-}_t, \\ X^{j*}_t &= x^j + B^j_t + \xi^{j*,+}_t - \xi^{j*,-}_t, \quad j \neq i. \end{aligned}$$

Applying the Itô-Tanaka-Meyers formula (Theorem 14.3.2 in [64]) to $e^{-\alpha t}u^i(\boldsymbol{X}_t^{-i*}, X_t^i)$ yields

$$\begin{aligned}
& \mathbb{E}\left[e^{-\alpha T}u^i(\boldsymbol{X}_T^{-i*}, X_T^i)\right] - u^i(x^1, x^2, \ldots, x^N) \\
= \quad & \mathbb{E}\left[\int_0^T e^{-\alpha t}\left(\frac{1}{2}\sum_{j=1}^N u^i_{x^jx^j}(\boldsymbol{X}_t^{-i*}, X_t^i) - \alpha u^i(\boldsymbol{X}_t^{-i*}, X_t^i)\right)dt\right] \\
+ \quad & \mathbb{E}\left[\int_{[0,T)} e^{-\alpha t}\left((u^i_{x^i}(\boldsymbol{X}_t^{-i*}, X_t^i)d\xi_t^{i,+} - u^i_{x^i}(\boldsymbol{X}_t^{-i*}, X_t^i)d\xi_t^{i,-})\right)\right] \\
+ \quad & \mathbb{E}\left[\sum_{0\le t<T} e^{-\alpha t}\left(\Delta u^i\left(\boldsymbol{X}_t^{-i*}, X_t^i\right) - \nabla u^i\left(\boldsymbol{X}_t^{-i*}, X_t^i\right)\cdot\Delta(\boldsymbol{X}_t^{-i*}, X_t^i)\right)\right] \\
+ \quad & \mathbb{E}\int_0^T e^{-\alpha t}\left(\sum_{j=1}^N u^i_{x^j}(\boldsymbol{X}_t^{-i*}, X_t^i)dB_t^j\right).
\end{aligned}$$

Note that (vii) implies that with control $(\boldsymbol{\xi}^{-i*}, \xi^i) \in \mathcal{S}_N$, $(\boldsymbol{X}_t^{-i*}, X_t^i) \in \mathcal{W}_{-i}$, $\mathbb{P}$-a.s.. By conditions (v) and (vi), $u^i_{x^j}$ is bounded on $\overline{\mathcal{W}_{-i}}$ for any $1 \le j \le N$, therefore $\int_0^T e^{-\alpha t}\left(\sum_{j=1}^N u^i_{x^j}(\boldsymbol{X}_t^{-i*}, X_t^i)dB_t^j\right)$ is square integrable, hence a uniformly integrable martingale. Now conditions (ii), (iv), (v), and (vi) suggest

$$e^{-\alpha T}\mathbb{E}[w^i(\boldsymbol{X}_T^{-i*}, X_T^i)] + \mathbb{E}\int_0^T e^{-\alpha t}\left[h\left(\frac{N-1}{N}\left(X_t^i - \frac{\sum_{j\neq i}X_t^{j*}}{N-1}\right)\right)dt + d\check{\xi}_t^i\right] \ge w^i(x^1, \ldots, x^N).$$

Taking $T \to \infty$, the transversality condition (iii) implies

$$w^i\left(x^1, \ldots, x^N\right) \le J^i\left(x^1, \ldots, x^N; \boldsymbol{\xi}_t^{-i*}, \xi_t^i\right), \tag{2.2.10}$$

for any ξ^i such that $\left(\boldsymbol{\xi}_t^{-i*}, \xi_t^i\right) \in \mathcal{S}_N$. □

The next step is to solve the HJB system, with a focus on a threshold-type solution. That is, there exists a constant $c_N > 0$ (to be determined) such that the action region $\mathcal{A}_i$ and the waiting $\mathcal{W}_i$ of player i can be decomposed into

$$\mathcal{A}_i = \left\{E_i^- \cup E_i^+\right\} \cap Q_i, \quad \mathcal{W}_i = \mathbb{R}^N/\mathcal{A}_i, \tag{2.2.11}$$

where

$$\begin{aligned}
E_i^- &= \left\{\left(x^1, \cdots, x^N\right) \in \mathbb{R}^N \,\middle|\, x^i - \frac{\sum_{j\neq i}x^j}{N-1} \le -c_N\right\}, \\
E_i^+ &= \left\{\left(x^1, \cdots, x^N\right) \in \mathbb{R}^N \,\middle|\, x^i - \frac{\sum_{j\neq i}x^j}{N-1} \ge c_N\right\},
\end{aligned} \tag{2.2.12}$$

with the partition

$$Q_i = \left\{ \boldsymbol{x} \in \mathbb{R}^N \middle| \left| x^i - \frac{\sum_{j\neq i} x^j}{N-1} \right| \geq \left| x^k - \frac{\sum_{j\neq k} x^j}{N-1} \right|, \text{ for any } k < i; \right.$$
$$\left. \left| x^i - \frac{\sum_{j\neq i} x^j}{N-1} \right| > \left| x^k - \frac{\sum_{j\neq k} x^j}{N-1} \right|, \text{ for any } k > i \right\}.$$

Note the modification of the action region $\mathcal{A}_i$ by Q_i is to avoid simultaneous jumps by multiple players. By definition of Q_i, in the event of multiple players in the "action region", the player who is the farthest away from the center intervenes first; in the event that multiple players have the same largest distance to the center, the player with the biggest index intervenes.

Now it is easy to check that

- $\cup_{i=1}^N Q_i = \mathbb{R}^N$, Q_i is a convex cone for any $i = 1, \ldots, N$,
- $\mathcal{W}_i \neq \emptyset$, for any $i = 1, \ldots, N$,
- $\mathcal{A}_i \cap \mathcal{A}_j = 0$, for all $i \neq j$.

Now, a candidate function $w^i(\boldsymbol{x}) \in \mathcal{C}^2(\overline{\mathcal{W}_{-i}})$ should satisfy the following three properties: First, $w^i(\boldsymbol{x})$ is symmetric on $x^i = \frac{\sum_{j\neq i} x^j}{N-1}$ such that

$$w^i_{x^i}\left(\boldsymbol{x}^{-i}, \frac{\sum_{j\neq i} x^j}{N-1}\right) = 0. \tag{2.2.13}$$

Second, if $0 \leq x^i - \frac{\sum_{j\neq i} x^j}{N-1} < c_N$, then $w^i(\boldsymbol{x})$ solves

$$\alpha w^i(\boldsymbol{x}) = h\left(\frac{N-1}{N}\left(x^i - \frac{\sum_{j\neq i} x^j}{N-1}\right)\right) + \frac{1}{2}\sum_{j=1}^N w^i_{x^j,x^j}(\boldsymbol{x}). \tag{2.2.14}$$

Third, if $x^i - \frac{\sum_{j\neq i} x^j}{N-1} \geq c_N$, then player i jumps by a distance of $x^i - \frac{\sum_{j\neq i} x^j}{N-1} - c_N$. Combined,

$$w^i(\boldsymbol{x}) = x^i - \frac{\sum_{j\neq i} x^j}{N-1} - c_N + w^i\left(\boldsymbol{x}^{-i}, \frac{\sum_{j\neq i} x^j}{N-1} + c_N\right). \tag{2.2.15}$$

The general solution satisfying both (2.2.14) and (2.2.13) is given by

$$w^i(\boldsymbol{x}) = B \cdot \cosh\left(\sqrt{\frac{2(N-1)\alpha}{N}}\left(x^i - \frac{\sum_{j\neq i} x^j}{N-1}\right)\right) + p_N\left(x^i - \frac{\sum_{j\neq i} x^j}{N-1}\right),$$

with

$$p_N(x) = \mathbb{E}\left[\int_0^\infty e^{-\alpha t} h\left(\frac{N-1}{N}\left(x + \sqrt{\frac{N}{N-1}}B_t\right)\right) dt\right]. \tag{2.2.16}$$

Here $p_N(x)$ is a particular solution to (2.2.14) and derived from the cost of "doing nothing", and B is constant yet to be determined.

Now matching the values of $w_{x^i}(\boldsymbol{x})$ and $w_{x^i,x^i}(\boldsymbol{x})$ along $x^i = \frac{\sum_{j\neq i} x^j}{N-1} + c_N$ determines c_N and B: c_N is the unique positive solution to

$$\frac{1}{\sqrt{\frac{2(N-1)\alpha}{N}}}\tanh\left(c\sqrt{\frac{2(N-1)\alpha}{N}}\right) = \frac{p_N'(c)-1}{p_N''(c)}, \tag{2.2.17}$$

and

$$B = -\frac{p_N''(c_N)}{\frac{2(N-1)\alpha}{N}\cosh\left(c_N\sqrt{\frac{2(N-1)\alpha}{N}}\right)}.$$

Finally, define

$$u^i(x^1,\dots,x^N) = \begin{cases} u^i\left(x^1,\dots,\frac{\sum_{j\neq i}x^j}{N-1} - c_N,\dots,x^N\right) - c_N - x^i + \frac{\sum_{j\neq i}x^j}{N-1}, & \boldsymbol{x}\in E_i^-, \\ -\frac{p_N''(c_N)\cosh\left(\sqrt{\frac{2(N-1)\alpha}{N}}\left(x^i - \frac{\sum_{j\neq i}x^j}{N-1}\right)\right)}{\frac{2(N-1)\alpha}{N}\cosh\left(c_N\sqrt{\frac{2(N-1)\alpha}{N}}\right)} + p_N\left(x^i - \frac{\sum_{j\neq i}x^j}{N-1}\right), & \\ \qquad\qquad \boldsymbol{x}\in\{E_i^+\cup E_i^-\}^c, & \\ x^i - \frac{\sum_{j\neq i}x^j}{N-1} - c_N + u^i\left(x^1,\dots,\frac{\sum_{j\neq i}x^j}{N-1} + c_N,\dots,x^N\right), & \boldsymbol{x}\in E_i^+. \end{cases}$$

Then it is easy to check that $u^i \in \mathcal{C}^2(\mathbb{R}^N)$ and the candidate solution w^i satisfies (HJB-N) and Theorem 7.

NE and the Skorokhod Problem (SP)

Given the NE solution to the N-player game, the corresponding NE can be constructed by finding a solution to an associated SP on an unbounded polyhedron and with a constant oblique reflection on each face.

First, define $\mathcal{CW}$ the common waiting regions of all players as

$$\begin{aligned} \mathcal{CW} &:= \left\{\boldsymbol{x}\in\mathbb{R}^N \;\middle|\; \left|x^i - \frac{\sum_{j\neq i}x^j}{N-1}\right| < c_N, \text{ for any } i = 1,\dots,N\right\} \\ &= \left\{\boldsymbol{x}\in\mathbb{R}^N \;\middle|\; \boldsymbol{n}_j\cdot\boldsymbol{x} > -c_N\sqrt{\frac{N}{N-1}}, \text{ for } j = 1,\dots,2N\right\} \\ &= \cap_{i=1}^N (E_i^-\cup E_i^+)^c, \end{aligned} \tag{2.2.18}$$

with the normal direction of each face given by

$$\begin{aligned}\boldsymbol{n}_i &= \frac{\sqrt{N-1}}{\sqrt{N}}\left(-\frac{1}{N-1},\cdots,-\frac{1}{N-1},1,-\frac{1}{N-1},\cdots,-\frac{1}{N-1}\right),\\ \boldsymbol{n}_{i+N} &= -\boldsymbol{n}_i.\end{aligned} \tag{2.2.19}$$

where 1 is in the i^{th} position of $\sqrt{\frac{N}{N-1}}\boldsymbol{n}_i$. Note that $\mathcal{CW}$ is an unbounded polyhedron with all of its $2N$ boundaries parallel to the direction $(1,1,\cdots,1)$.

For $j=1,\cdots,2N$, define the $2N$ faces of $\mathcal{CW}$

$$F_j = \{\boldsymbol{x}\in\partial\mathcal{CW} \mid \boldsymbol{n}_j\cdot\boldsymbol{x} = -c_N\}, \tag{2.2.20}$$

and

$$\boldsymbol{d}_i = (0,\cdots,1,\cdots,0),\quad \boldsymbol{d}_{i+N} = -\boldsymbol{d}_i,\quad i=1,\ldots,N, \tag{2.2.21}$$

such that $\boldsymbol{d}_j\cdot\boldsymbol{n}_j = \frac{\sqrt{N-1}}{\sqrt{N}}$, where 1 is in the i^{th} position of $\boldsymbol{d}_i$.

Now, the NE of (**N-player**) can be fully characterized by the solution to the SP with the data $(\boldsymbol{x},\mathcal{CW},(\boldsymbol{d}_1,\cdots,\boldsymbol{d}_{2N}),\{\boldsymbol{B}_t\}_{t\geq 0})$. (See Appendix A.1 for more background materials.)

Theorem 8. *There exits a unique strong solution to SP with the data $(\boldsymbol{x},\mathcal{CW},(\boldsymbol{d}_1,\cdots,\boldsymbol{d}_{2N}),\{\boldsymbol{B}_t\}_{t\geq 0})$ defined in (4.2.2) and (4.2.6). More precisely, the reflected process $\boldsymbol{X}_t^*$ with $\boldsymbol{X}_0^*=\boldsymbol{x}\in\overline{\mathcal{CW}}$ is defined as*

$$X_t^{i*} = x^i + B_t^i + \int_0^t 1_{\{\boldsymbol{X}_s^*\in F_i\}}d\eta^i(s) - \int_0^t 1_{\{\boldsymbol{X}_s^*\in F_{i+N}\}}d\eta^i(s),\quad i=1,2,\cdots,N,$$

where $\eta^j(t)$ is a non-decreasing process with $\eta^j(0)=0$. Moreover, if $\boldsymbol{x}\notin F_k\cap F_j$ for any $k\neq j, k,j=1,2,\cdots,2N$,

$$\mathbb{P}(\boldsymbol{X}_t^*\notin F_k\cap F_j \text{ for any } k\neq j, t\geq 0) = 1. \tag{2.2.22}$$

The idea to prove Theorem 8 is to show first the existence of a weak solution to the SP and next the uniqueness of the strong solution to the SP. Then according to Corollary 3.23 in Karatzas and Shreve [120] and Proposition 1 in Engelbert [75], there exists a unique strong solution to the SP. The existence of a weak solution to the SP is straightforward, following [66]. The uniqueness of a strong solution is established by extending the result of Dupuis and Ishii [73] on a bounded polyhedron to an unbounded one, via the localization technique. Moreover, the reflection vectors $(\boldsymbol{d}_1,\cdots,\boldsymbol{d}_{2N})$ satisfy the *skew symmetry* condition for the polyhedron $\mathcal{CW}$ according to [185], hence an additional localization argument shows that (2.2.22) holds. The detailed proof is provided in Appendix A.1.

Extended Mapping to $\mathbb{R}^N \setminus \overline{\mathcal{CW}}$

Up to now the NE is derived when $\boldsymbol{x} \in \overline{\mathcal{CW}}$. When $\boldsymbol{x} \in \mathbb{R}^N \setminus \overline{\mathcal{CW}}$, the NE would be to jump sequentially to some point $\hat{\boldsymbol{x}} \in \partial\overline{\mathcal{CW}}$, and afterwards continues according to the SP with data $(\hat{\boldsymbol{x}}, \mathcal{CW}, (\boldsymbol{d}_1, \cdots, \boldsymbol{d}_{2N}), \{\boldsymbol{B}_t\}_{t\geq 0})$ where $\hat{\boldsymbol{x}} \in \overline{\mathcal{CW}}$.

Algorithm **1** describes how players sequentially jump to $\overline{\mathcal{CW}}$. In order to show that this algorithm is well defined, one needs to make sure that such jumps stop in finite steps or converge to a limit point on $\hat{\boldsymbol{x}} \in \partial\overline{\mathcal{CW}}$, and that the total distance of such sequential jumps is bounded. The detailed argument is given in Appendix A.2, with the illustration of Figure A.1.

Algorithm 1 Policy: Sequential jumps when $\boldsymbol{x} \notin \overline{\mathcal{CW}}$.

procedure SEQUENTIAL($\boldsymbol{x}$)

Define mapping,

$$\begin{aligned} i = \pi(\boldsymbol{y}) \quad &\text{when} \quad \boldsymbol{y} \in \mathcal{A}_i, \\ \emptyset = \pi(\boldsymbol{y}) \quad &\text{when} \quad \boldsymbol{y} \in \overline{\mathcal{CW}}. \end{aligned} \tag{2.2.23}$$

$\hat{\boldsymbol{x}} \leftarrow \boldsymbol{x}$, $k \leftarrow 0$

while $\pi(\hat{\boldsymbol{x}}) \neq \emptyset$ **do**

$\lambda^* \leftarrow \arg\min \left\{ \lambda > 0 \;\middle|\; \hat{\boldsymbol{x}} + \lambda \boldsymbol{e}_{\pi(\hat{\boldsymbol{x}})} \in \partial E^-_{\pi(\hat{\boldsymbol{x}})} \text{ or } \hat{\boldsymbol{x}} - \lambda \boldsymbol{e}_{\pi(\hat{\boldsymbol{x}})} \in \partial E^+_{\pi(\hat{\boldsymbol{x}})} \right\}$ ▷ e_j is a unit vector in $\mathbb{R}^N$ with jth component to be 1

if $\hat{\boldsymbol{x}} + \lambda^* \boldsymbol{e}_{\pi(\hat{\boldsymbol{x}})} \in \partial E^-_{\pi(\hat{\boldsymbol{x}})}$ **then**

$\boldsymbol{\nu}_0 \leftarrow \boldsymbol{e}_{\pi(\hat{\boldsymbol{x}})}$

else

$\boldsymbol{\nu}_0 \leftarrow -\boldsymbol{e}_{\pi(\hat{\boldsymbol{x}})}$

$\hat{\boldsymbol{x}} \leftarrow \hat{\boldsymbol{x}} + \lambda^* \boldsymbol{\nu_0}$ ▷ Control of player $\pi(\hat{\boldsymbol{x}})$

$\boldsymbol{x}_k \leftarrow \hat{\boldsymbol{x}}$

$k \leftarrow k + 1$

return $\hat{\boldsymbol{x}}, \{\boldsymbol{x}_k\}$ ▷ $\hat{\boldsymbol{x}} \in \partial\mathcal{CW}$

Note that this algorithm gives an ϵ-NE in finite steps. In the case that the starting point is in the intersection of faces, a small perturbation in the algorithm and in the NE value will recover the case of $\boldsymbol{x} \in \mathcal{CW}$. In summary,

Theorem 9 (NE for the N-player game)**.** *Under Assumption* ***A1****, a Markovian NE for game (**N-player**) is given by*

$$\begin{aligned} \xi_t^{i*,+} &= \Delta_0^{i*,+} + \int_0^t \boldsymbol{1}_{\{\boldsymbol{X}_s^* \in F_i\}} d\eta^i(s), \\ \xi_t^{i*,-} &= \Delta_0^{i*,-} + \int_0^t \boldsymbol{1}_{\{\boldsymbol{X}_s^* \in F_{i+N}\}} d\eta^{i+N}(s), \end{aligned} \tag{2.2.24}$$

where $\mathcal{CW}$ *is given in (4.2.2),* $\boldsymbol{X}_t^*$ *is the controlled dynamic with* $\boldsymbol{X}_0^* = \hat{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{\Delta_0}^{*,+} - \boldsymbol{\Delta_0}^{*,-} \in \overline{\mathcal{CW}}$*, with* $\eta^j(t) = \int_0^t \mathbf{1}_{\{\boldsymbol{X}_s^* \in F_j\}} d\eta^j(s)$ *and* $\eta^j(0) = 0$ $(j = 1, 2, \cdots, 2N)$*, the jumps at time* 0 *are*

$$\begin{aligned} \Delta_0^{i*,+} &= \sum_k \mathbf{1}_{\{\boldsymbol{x}_k \in \mathcal{A}_i\}} \left(x_{k+1}^i - x_k^i\right)_+, \\ \Delta_0^{i*,-} &= \sum_k \mathbf{1}_{\{\boldsymbol{x}_k \in \mathcal{A}_i\}} \left(x_k^i - x_{k+1}^i\right)_+, \end{aligned} \tag{2.2.25}$$

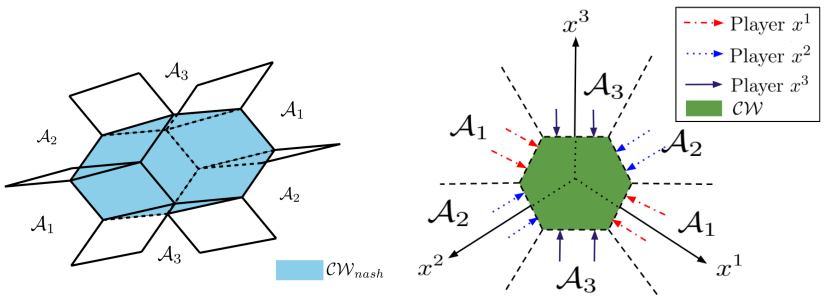
with $\{\boldsymbol{x}_k\}$ *the sequence of jumps prescribed by Algorithm* **1**.

The corresponding NE value $v^i(x^1, \ldots, x^N) := J^i(x^1, \ldots, x^N; \boldsymbol{\xi}^*)$ *is given by*

$$v^i(x^1, \ldots, x^N) = \begin{cases} v^i\left(x^1, \ldots, x^{j-1}, \frac{\sum_{k\neq j} x^k}{N-1} - c_N, x^{j+1}, \ldots, x^N\right), \\ \qquad\qquad \boldsymbol{x} \in E_j^- \cap \mathcal{A}_j, \text{ for any } j \neq i, \\ v^i\left(x^1, \ldots, \frac{\sum_{j\neq i} x^j}{N-1} - c_N, \ldots, x^N\right) - c_N - x^i + \frac{\sum_{j\neq i} x^j}{N-1}, \\ \qquad\qquad \boldsymbol{x} \in E_i^- \cap \overline{\mathcal{W}}_{-i}, \\ -\frac{p_N''(c_N)\cosh\left(\sqrt{\frac{2(N-1)\alpha}{N}}\left(x^i - \frac{\sum_{j\neq i} x^j}{N-1}\right)\right)}{\frac{2(N-1)\alpha}{N}\cosh\left(c_N\sqrt{\frac{2(N-1)\alpha}{N}}\right)} + p_N\left(x^i - \frac{\sum_{j\neq i} x^j}{N-1}\right), \\ \qquad\qquad \boldsymbol{x} \in (E_i^- \cup E_i^+)^c \cap \overline{\mathcal{W}}_{-i}, \\ x^i - \frac{\sum_{j\neq i} x^j}{N-1} - c_N + v^i\left(x^1, \ldots, \frac{\sum_{j\neq i} x^j}{N-1} + c_N, \ldots, x^N\right), \\ \qquad\qquad \boldsymbol{x} \in E_i^+ \cap \overline{\mathcal{W}}_{-i}, \\ v^i\left(x^1, \ldots, x^{j-1}, \frac{\sum_{k\neq j} x^k}{N-1} + c_N, x^{j+1}, \ldots, x^N\right), \\ \qquad\qquad \boldsymbol{x} \in E_j^+ \cap \mathcal{A}_j, \text{ for any } j \neq i. \end{cases} \tag{2.2.26}$$

Here E_i^+*,* E_i^- *are given in (2.2.12), and* $\mathcal{A}_i$ *and* $\mathcal{W}_i$ *defined in (2.2.11).*

Figure (2.2a) shows the region partition when $N = 3$. $\mathcal{CW}$, the unbounded polytope, is surrounded by the action regions $\mathcal{A}_i$, $i = 1, 2, 3$. Figure (2.2b) shows the action region $\mathcal{A}_1$ of player one and the common waiting region $\mathcal{CW}$ of all players.



(a) $\mathcal{CW}$ and $\mathcal{A}_i$ $(i = 1, 2, 3)$

(b) $\mathcal{A}_1$ and $\mathcal{W}_1$, a bird's-eye view from $(1, 1, 1)$ to $(0, 0, 0)$

Figure 2.2: Region partition when $N = 3$

2.3 MFG for the Fuel Follower Problem

Take N identical, rational, and interchangeable players, whose initial positions are random in $\mathbb{R}^N$. Let $N \to \infty$, the MFG for the fuel follower problem is to find a closed-loop control in feedback form of

$$\begin{aligned} v(x) &= \inf_{(\xi^+,\xi^-)\in\mathcal{U}_\infty} J_{(\infty)}(x;\xi_t^+,\xi_t^-) \\ &= \inf_{(\xi^+,\xi^-)\in\mathcal{U}_\infty} \mathbb{E}\int_0^\infty e^{-\alpha t}\left[h(X_t-m_t)dt + d\check{\xi}_t | X_{0-}=x\right], \\ \text{such that} \quad & dX_t = dB_t + d\xi_t^+ - d\xi_t^-, \\ & X_{0-}\sim\mu_{0-},\ m_{0-} = \int x\mu_{0-}(dx), \end{aligned} \tag{2.3.1}$$

where $\mu_t = \lim_{N\to\infty}\frac{\sum_{i=1}^N \mathbf{1}_{\{X_t^i\}}}{N}$ is the distribution of X_t and $m_t = \lim_{N\to\infty}\frac{\sum_{i=1}^N X_t^i}{N} = \int x\mu_t(dx)$ is the mean position of the population at time t, with μ_{0-} symmetric around m_{0-}.

Note that one could write an alternative MFG formulation with

$$\tilde{v}(\mu_{0-}) := \inf_{(\xi^+,\xi^-)\in\mathcal{U}_\infty} \mathbb{E}\int_0^\infty e^{-\alpha t}\left[h(X_t-m_t)dt + d\check{\xi}_t\right].$$

$v(x)$ defined in (5.2.1) can be viewed as $\tilde{v}(\mu_{0-}|X_{0-}=x)$ with $X_{0-}=x$ as some sample drawn from μ_{0-}. Clearly $\tilde{v}(\mu_{0-})$ can be solved by analyzing $v(x)$ as $\tilde{v}(\mu_{0-}) = \mathbb{E}_{\mu_{0-}}[v(X_{0-})]$. This connection is also explored in Section 2.2.2 of [134].

The admissible control set for MFG is

$$\begin{aligned} \mathcal{U}_\infty = \Big\{ &(\xi_t^+,\xi_t^-) \mid \xi_t^+ \text{ and } \xi_t^- \text{ are } \mathcal{F}_t^{(X_{t-},m_{t-})}\text{-progressively measurable, càdlàg, non-decreasing,} \\ &\text{with } \mathbb{E}\left[\int_0^\infty e^{-\alpha t}d\xi_t^+\right] < \infty,\ \mathbb{E}\left[\int_0^\infty e^{-\alpha t}d\xi_t^-\right] < \infty,\ \xi_{0-}^+ = 0,\ \xi_{0-}^- = 0 \Big\}. \end{aligned}$$

NE Solution to the MFG

Definition 10 (NE to MFG (5.2.1))**.** *An NE to the MFG (5.2.1) is a pair of Markovian control $(\xi_t^{*,+},\xi_t^{*,-})_{t\geq 0}$ and a mean function $\{m_t^*\}_{t\geq 0}$ such that*

- $v^*(x) = J_{(\infty)}\left(x;\xi^{*,+},\xi^{*,-}|\{\mu_t^*\}_{t\geq 0}\right) = \min_{\xi\in\mathcal{U}_\infty} J_{(\infty)}\left(x;\xi^+,\xi^-|\{\mu_t^*\}_{t\geq 0}\right)$,
- $P_{X_t^*} = \mu_t^*$, *and* $m_t^* = \int xP_{X_t^*}(dx)$ *is the mean function of* X_t^* *where* X_t^* *is the controlled dynamic under* $(\xi_t^{*,+},\xi_t^{*,-})_{t\geq 0}$.

$v^*(x)$ is called the NE value of the MFG associated with ξ^*.

Theorem 11 (NE to MFG (5.2.1)). *There exists an NE to the MFG (5.2.1),*

$$\begin{aligned}\xi_t^{*,+} = &\max\left\{0, \max_{0\le u\le t}\{m_{0-} - x - B_u + \xi_u^{*,-} - c\}\right\},\\ \xi_t^{*,-} = &\max\left\{0, \max_{0\le u\le t}\{x - m_{0-} + B_u + \xi_u^{*,+} - c\}\right\},\end{aligned} \tag{2.3.2}$$

and the corresponding NE value is

$$v^*(x) = \begin{cases} -\frac{p_1''(c\sqrt{2\alpha})\cosh\left(x\sqrt{2\alpha}\right)}{2\alpha\cosh\left(c\sqrt{2\alpha}\right)} + p_1(x - m_{0-}), & m_{0-} \le x \le m_{0-} + c,\\ v(m_{0-} + c) + (x - m_{0-} - c), & x \ge m_{0-} + c,\\ v(m_{0-} - x), & x < m_{0-},\end{cases} \tag{2.3.3}$$

where c is the solution to (4.2.7).

The proof consists of three steps.

Step 1: Stochastic control problem.

Take the M_1 topology for the Skorokhod space $\mathcal{D}([0,\infty))$ with a Wasserstein distance W_1([**skorokhod1956**, 81]). Fix a mean field measure $\{\mu_t\}_{t\ge 0} \in \mathcal{P}_1(\mathcal{D}([0,\infty)))$, with $m_t = \int x\mu_t(dx)$ and $\mathcal{P}_1$ the class of all probability measures with finite moment of first order. Then (5.2.1) becomes the following time-dependent and state-dependent singular control problem,

$$\begin{aligned}\hat{v}(s,x) = &\inf_{\xi\in\mathcal{U}_\infty} \mathbb{E}\int_s^\infty e^{-\alpha(t-s)}\left[h(X_t - m_t)dt + d\xi_t^+ + d\xi_t^-\right]\\ \text{such that} \quad &dX_t = dB_t + d\xi_t^+ - d\xi_t^-,\ X_{s-} = x,\ m_{s-} = m.\end{aligned} \tag{2.3.4}$$

The corresponding HJB equation for $\hat{v}(s,x)$ is

$$\max\left\{\alpha\hat{v}(s,x) - \hat{v}_t(s,x) - \frac{1}{2}\hat{v}_{xx}(s,x) - h(x-m), -1 + \hat{v}_x(s,x), -1 - \hat{v}_x(s,x)\right\} = 0 \tag{2.3.5}$$

Note that (2.3.5) is a parabolic equation because of μ_t despite the infinite horizon. This is different from the elliptic equation (2.2.3).

We will show that $\hat{v}(s,x)$ in (2.3.4) is a viscosity solution to HJB equation (2.3.5).

First, under a fixed $\{\mu_t\}_{t\ge 0}$, the following dynamic programming principle holds.

Dynamic programming principle (DPP). For all $(s,x) \in \mathbb{R}^+ \times \mathbb{R}$,

$$\hat{v}(s,x) = \inf_{\xi\in\mathcal{U}_\infty} \mathbb{E}\left[\int_s^\theta e^{-\alpha(t-s)}\left(h(X_t - m_t)dt + d\check{\xi}_t\right) + e^{-\alpha(\theta-s)}v(\theta, X_\theta)\right] \tag{2.3.6}$$

for any $\theta \in \mathcal{T}$ and $\theta \ge s$, with $\mathcal{T}$ the set of all $\{\mathcal{F}^{(X_t,m_t)}\}_{t\ge 0}$-stopping times. Here, we adopt the convention that $e^{-\alpha\theta(\omega)} = 0$ when $\theta(\omega) = \infty$. The proof of DPP (2.3.6) follows Guo and Pham [92] by extending the state space from $\mathbb{R}$ to $\mathbb{R}^+ \times \mathbb{R}$.

Definition 12 (Viscosity solution)**.** *$\hat{v}(t,x)$ is a* continuous viscosity solution *to (2.3.5) on $[0,\infty)\times\mathbb{R}$ if*

- *Viscosity super-solution: for any $(t_0,x_0)\in[0,\infty)\times\mathbb{R}$ and for any function $\phi(t_0,x_0)$ such that (t_0,x_0) is a local minimum of $(\hat{v}-\phi)(t,x)$ with $\hat{v}(t_0,x_0)=\phi(t_0,x_0)$,*

$$\max\left\{\alpha\phi(t_0,x_0)-\phi_t(t_0,x_0)-\frac{1}{2}\phi_{x,x}(t_0,x_0)-h(x_0-m),-1+\phi_x(t_0,x_0),-1-\phi_x(t_0,x_0)\right\}\geq 0.$$

- *Viscosity sub-solution: for any $(t_0,x_0)\in[0,\infty)\times\mathbb{R}$ and for any function $\phi(t_0,x_0)$ such that (t_0,x_0) is a local maximum of $(\hat{v}-\phi)(t,x)$ with $\hat{v}(t_0,x_0)=\phi(t_0,x_0)$,*

$$\max\left\{\alpha\phi(t_0,x_0)-\phi_t(t_0,x_0)-\frac{1}{2}\phi_{x,x}(t_0,x_0)-h(x_0-m),-1+\phi_x(t_0,x_0),-1-\phi_x(t_0,x_0)\right\}\leq 0.$$

Proposition 13. *Assume that the value function $\hat{v}(t,x)$ of (2.3.4) is continuous with respect to t. Then $\hat{v}(t,x)$ is a continuous viscosity solution of the HJB equation (2.3.5) on $[s,\infty)\times\mathbb{R}$. Moreover, $\hat{v}(t,x)$ is convex and differentiable in x, and for any $x,y\in\mathbb{R}$,*

$$\hat{v}(s,x)\leq\hat{v}(s,y)+|x-y|. \tag{2.3.7}$$

Proof. Since h is convex and the pay-off function $\mathbb{E}\left[\int_s^\infty e^{-\alpha(t-s)}h(X_t-m_t)dt+d\xi_t^+ +d\xi_t^-\right]$ in problem (2.3.4) is linear in control (ξ^+,ξ^-), the value function $\hat{v}(s,x)$ is convex in x. Since $\hat{v}(s,x)$ is finite and convex on $(-\infty,\infty)$, it is continuous in x. Moreover, consider a special control,

$$\xi_t^+-\xi_t^-=\begin{cases}0, & t=s,\\ y-x, & t\geq s,\end{cases} \tag{2.3.8}$$

clearly $\hat{v}(s,x)\leq\hat{v}(s,y)+|y-x|$.

We now prove that the value function is a viscosity solution of (2.3.5).

- Step A: Viscosity sub-solution.
  For some $(t_0,x_0)\in\mathbb{R}^+\times\mathbb{R}$ and $\phi\in\mathcal{C}^{1,2}(\mathbb{R}^+\times\mathbb{R})$ such that $\hat{v}(t_0,x_0)=\phi(t_0,x_0)$ and $\phi(t_0,x_0)\geq\hat{v}(t_0,x_0)$ for $(t,x)\in B_\epsilon(t_0,x_0)$. That is, $\hat{v}-\phi$ has local maximum at (t_0,x_0). Consider the following admissible control

$$\xi_t^+=\begin{cases}0, & t=t_0,\\ \eta_1, & t\geq t_0,\end{cases} \tag{2.3.9}$$

$$\xi_t^-=\begin{cases}0, & t=t_0,\\ \eta_2, & t\geq t_0,\end{cases} \tag{2.3.10}$$

where $0 \leq \eta_1, \eta_2 \leq \epsilon$. Define the exit time

$$\tau_\epsilon = \inf\left\{t \geq t_0, X_t \notin \bar{B}_\epsilon(t_0, x_0)\right\}. \tag{2.3.11}$$

Notice that X has at most one jump at $t = t_0$ and is continuous on $[t_0, t_0 + \tau_\epsilon)$. By the DPP,

$$\begin{aligned}\phi(t_0, x_0) = \hat{v}(t_0, x_0) \leq \ & \mathbb{E}\int_{t_0}^{t_0+\tau_\epsilon\wedge\delta} e^{-\alpha(t-t_0)}\left[h(X_t - m_t)dt + d\xi_t^+ + d\xi_t^-\right] \\ & + \ \mathbb{E}\left[e^{-\alpha(\tau_\epsilon\wedge\delta)}\phi(t_0 + \tau_\epsilon \wedge \delta, X_{t_0+\tau_\epsilon\wedge\delta})\right].\end{aligned} \tag{2.3.12}$$

By Itô's lemma,

$$\begin{aligned}&\mathbb{E}[e^{-\alpha(\tau_\epsilon\wedge\delta)}\phi(t_0 + \tau_\epsilon \wedge \delta, X_{t_0+\tau_\epsilon\wedge\delta})] \\ = \phi(t_0, x_0) &+ \mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon\wedge\delta} e^{-\alpha(t-t_0)}(-\alpha\phi + \phi_t + \frac{1}{2}\phi_{x,x})(t, X_t)dt\right] \\ &+ \ \mathbb{E}\left[\sum_{t_0\leq t\leq\tau_\epsilon\wedge\delta} e^{-\alpha t}(\phi(t, X_t) - \phi(t, X_{t-}))\right].\end{aligned} \tag{2.3.13}$$

Combining (2.3.12) and (2.3.13),

$$\begin{aligned}&\mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon\wedge\delta} e^{-\alpha(t-t_0)}(\alpha\phi - \phi_t - \frac{1}{2}\phi_{x,x} - h)(t, X_t)dt\right] \\ &-\mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon\wedge\delta} e^{-\alpha(t-t_0)}(d\xi_t^+ + d\xi_t^-)\right] \\ &-\mathbb{E}\left[\sum_{t_0\leq t\leq\tau_\epsilon\wedge\delta} e^{-\alpha t}(\phi(t, X_t) - \phi(t, X_{t-}))\right] \leq 0.\end{aligned} \tag{2.3.14}$$

Now, setting $\eta_1 = \eta_2 = 0$ and letting $\delta \to 0$ leads to $\alpha\phi - \phi_t - \frac{1}{2}\phi_{x,x} - h \leq 0$.

Next, let $\eta_2 = 0$, and note that ξ_t^+ and X_t only jump at time t_0 with a size η_1, therefore

$$\mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon\wedge\delta} e^{-\alpha(t-t_0)}(\alpha\phi - \phi_t - \frac{1}{2}\phi_{x,x} - h)(t, X_t)dt\right] - \eta_1 - \phi(t_0, x_0 + \eta_1) + \phi(t_0, x_0) \leq 0.$$

Now, taking $\delta \to 0$, dividing by η_1, and letting $\eta_1 \to 0$ yields $-1 - \phi_x \leq 0$. Similarly, $-1 + \phi_x \leq 0$. That is, ϕ is the sub-solution to (2.3.5), so that

$$\max\left\{\alpha\phi(t_0, x_0) - \phi_t(t_0, x_0) - \frac{1}{2}\phi_{x,x}(t_0, x_0) - h(x_0 - m), -1 - \phi_x(t_0, x_0), -1 + \phi_x(t_0, x_0)\right\} \leq 0.$$

- Step B: Viscosity Super-solution.
  This is established by a contradiction argument. Suppose otherwise, then there exists (t_0, x_0), $\epsilon, \delta > 0$ $\phi \in C^{1,2}(\mathbb{R}^+ \times \mathbb{R})$ such that for any $(t, x) \in \bar{B}_\epsilon(t_0, x_0)$,

$$\begin{cases} \alpha\phi - \frac{1}{2}\phi_{x,x} - h(x-m) - \phi_t \leq -\delta, \\ -1 + \delta \leq \phi_x \leq 1 - \delta. \end{cases} \tag{2.3.15}$$

  Given any admissible control $(\xi^+, \xi^-) \in \mathcal{U}_\infty$, consider an exit time $\tau_\epsilon = \inf\{t \geq 0, X_{t+t_0} \notin \bar{B}_\epsilon(t_0, x_0)\}$, and apply Itô's lemma to $e^{-\alpha t}\phi(t, X_t)$,

$$\begin{aligned} \mathbb{E}\left[e^{-\alpha\tau_\epsilon}\phi(t_0+\tau_\epsilon, X_{t_0+\tau_\epsilon})\right] = \phi(t_0,x_0) &+ \mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon} e^{-\alpha(t-t_0)}(-\alpha\phi + \phi_t + \frac{1}{2}\phi_{x,x})(t, X_t)dt\right] \\ &+ \mathbb{E}\left[\sum_{t_0 \leq t \leq \tau_\epsilon} e^{-\alpha t}(\phi(t, X_t) - \phi(t, X_{t-}))\right] \\ &+ \mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon} e^{-\alpha t}\phi'(t, X_t)\left(\left(d\xi_t^+\right)^c + \left(d\xi_t^-\right)^c\right)\right]. \end{aligned}$$

  Notice that for any $t_0 \leq t \leq t_0 + \tau_\epsilon$, $(t, X_t) \in \bar{B}_\epsilon(t_0, x_0)$. By the Taylor expansion and $\Delta X_t = \Delta\xi_t^+ - \Delta\xi_t^-$, clearly for any $0 \leq t < \tau_\epsilon$:

$$\begin{aligned} \phi(t, X_t) - \phi(t, X_{t-}) &= \Delta X_t \int_0^1 \phi_x(t, X_t + z\Delta X_t)dz \\ &\geq (-1+\delta)(\Delta\xi_t^+ + \Delta\xi_t^-). \end{aligned} \tag{2.3.16}$$

  Thus,

$$\begin{aligned} &\mathbb{E}[e^{-\alpha\tau_\epsilon}\phi(t_0+\tau_\epsilon, X_{t_0+\tau_\epsilon-})] \\ \geq\ &\phi(t_0,x_0) + \mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon} e^{-\alpha(t-t_0)}(-h+\delta)(t, X_t)dt\right] \\ &+ (\delta-1)\mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon-} e^{-\alpha(t-t_0)}(d\xi_t^+ + d\xi_t^-)\right] \\ =\ &\phi(t_0,x_0) + \mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon} e^{-\alpha(t-t_0)}\left(-h(X_t - m_t)dt - d\xi_t^+ - d\xi_t^-\right)\right] \\ &+ \mathbb{E}\left[e^{-\alpha\tau_\epsilon}(\Delta\xi_{t_0+\tau_\epsilon}^+ + \Delta\xi_{t_0+\tau_\epsilon}^-)\right] + \delta\mathbb{E}[\int_{t_0}^{t_0+\tau_\epsilon} e^{-\alpha t}dt\Big] \\ &+ \delta\mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon-} e^{-\alpha(t-t_0)}(d\xi_t^+ + d\xi_t^-)\right]. \end{aligned} \tag{2.3.17}$$

  By definition of τ_ϵ, $(t_0 + \tau_\epsilon -, X_{t_0+\tau_\epsilon-}) \in \bar{B}_\epsilon(t_0, x_0)$ and $(t_0 + \tau_\epsilon, X_{t_0+\tau_\epsilon})$ is either on the boundary $\partial B_\epsilon(t_0, x_0)$ or out of $\bar{B}_\epsilon(t_0, x_0)$. However, there exists some random variable

$\alpha \in [0,1]$ such that,

$$\begin{aligned} x_\alpha &= X_{t_0+\tau_\epsilon-} + \alpha \Delta X_{t_0+\tau_\epsilon} \\ &= X_{t_0+\tau_\epsilon-} + \alpha(\Delta\xi^+_{t_0+\tau_\epsilon} - \Delta\xi^-_{t_0+\tau_\epsilon}) \in \partial B_\epsilon(t_0,x_0). \end{aligned}$$

Similar as in (2.3.16), we have

$$\phi(t_0+\tau_\epsilon, x_\alpha) - \phi(t_0+\tau_\epsilon, X_{t_0+\tau_\epsilon-}) \geq \alpha(-1+\delta)(\Delta\xi^+_{t_0+\tau_\epsilon} + \Delta\xi^-_{t_0+\tau_\epsilon}). \tag{2.3.18}$$

Notice that $X_{t_0+\tau_\epsilon} = x_\alpha + (1-\alpha)(\Delta\xi^+_{t_0+\tau_\epsilon} - \Delta\xi^-_{t_0+\tau_\epsilon})$, and from (2.3.7),

$$\hat{v}(t_0+\tau_\epsilon, x_\alpha) \leq \hat{v}(t_0+\tau_\epsilon, X_{t_0+\tau_\epsilon}) + (1-\alpha)(\Delta\xi^+_{t_0+\tau_\epsilon} + \Delta\xi^-_{t_0+\tau_\epsilon}). \tag{2.3.19}$$

Recalling $\phi(t_0+\tau_\epsilon, x_\alpha) \leq \hat{v}(t_0+\tau_\epsilon, x_\alpha)$, inequalities (2.3.18) and (2.3.19) imply

$$\phi(t_0+\tau_\epsilon, X_{t_0+\tau_\epsilon-}) \leq \hat{v}(t_0+\tau_\epsilon, X_{t_0+\tau_\epsilon}) + (1-\alpha\delta)(\Delta\xi^+_{t_0+\tau_\epsilon} + \Delta\xi^-_{t_0+\tau_\epsilon}).$$

Plugging the above inequality into (2.3.17), by $\phi(t_0,x_0) = \hat{v}(t_0,x_0)$,

$$\begin{aligned} &\mathbb{E}e^{-\alpha\tau_\epsilon}\left[\int_{t_0}^{t_0+\tau_\epsilon} \left(h(X_t-m_t)dt + d\xi^+_t + d\xi^-_t\right) + \hat{v}(t_0+\tau_\epsilon, X_{t_0+\tau_\epsilon})\right] \\ &\geq \hat{v}(t_0,x_0) + \alpha\delta\mathbb{E}\left[e^{-\alpha\tau_\epsilon}(\Delta\xi^+_{t_0+\tau_\epsilon} + \Delta\xi^-_{t_0+\tau_\epsilon})\right] \\ &\quad + \delta\mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon} e^{-\alpha t}dt\right] + \delta\mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon-} e^{-\alpha(t-t_0)}(d\xi^+_t + d\xi^-_t)\right]. \end{aligned} \tag{2.3.20}$$

There exists a constant $g_0 > 0$ such that for any $(\xi^+,\xi^-) \in \mathcal{U}_\infty$,

$$\alpha\mathbb{E}\left[e^{-\alpha\tau_\epsilon}(\Delta\xi^+_{t_0+\tau_\epsilon} + \Delta\xi^-_{t_0+\tau_\epsilon})\right] + \mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon} e^{-\alpha t}dt\right] + \mathbb{E}\left[\int_{t_0}^{t_0+\tau_\epsilon-} e^{-\alpha(t-t_0)}(d\xi^+_t + d\xi^-_t)\right] \geq g_0.$$

Finally, taking the infimum over all admissible controls $(\xi^+,\xi^-) \in \mathcal{U}_\infty$ in (2.3.20) suggests

$$\hat{v}(t_0,x_0) \geq \hat{v}(t_0,x_0) + \delta g_0, \tag{2.3.21}$$

which is a contradiction.

The differentiability with respect to x can be proved using the convexity of the value function $\hat{v}(s,x)$ to (2.3.5). Since $\hat{v}(s,x)$ is convex, the left and right derivatives with respect to x, $\hat{v}_{x-}(t,x)$ and $\hat{v}_{x+}(t,x)$ exist for any $t \geq s$ and $x \in \mathbb{R}$. Also, $\hat{v}_{x-}(t,x) \leq \hat{v}_{x+}(t,x)$ by convexity. We argue by contradiction and suppose there exists $x_0 \in \mathbb{R}$ and $t_0 \geq 0$ such that $\hat{v}_{x-}(t_0,x_0) < \hat{v}_{x+}(t_0,x_0)$. Fix some q in $(\hat{v}_{x-}(t_0,x_0), \hat{v}_{x+}(t_0,x_0))$ and consider the test function

$$\phi_\epsilon(t,x) = \hat{v}(t_0,x_0) + q(x-x_0) - \frac{1}{2\epsilon}(x-x_0)^2 - \frac{1}{2\epsilon}(t-t_0)^2,$$

with $\epsilon > 0$. Then (t_0, x_0) is a local minimum of $(\hat{v} - \phi_\epsilon)(t, x)$ since $\hat{v}_{x-}(t_0, x_0) < q = \phi_x(t_0, x_0) < \hat{v}_{x+}(t_0, x_0)$ and $\phi_t(t_0, x_0) = 0$. Hence ϕ is a viscosity super-solution by definition. That is,

$$\max\left\{\alpha\phi - \phi_t - \frac{1}{2}\phi_{x,x} - h(x_0 - m), -1 - \phi_x, -1 + \phi_x\right\} \geq 0,$$

which leads to $-\frac{1}{2\epsilon} + h(x_0 - m) - \alpha\phi(t_0, x_0) \geq 0$. Taking $\epsilon > 0$ sufficiently small leads to a contradiction. □

Proposition 14 (Optimal Control)**.** *Assume **A1** and assume that $\hat{v}_t(t, x)$ is continuous with respect to t, the optimal control to (2.3.4) under a fixed $\{\mu_t\}_{t\geq 0} \in \mathcal{P}_1(\mathcal{D}([0, \infty)))$ is of the form*

$$d\hat{\xi}_t = \begin{cases} m_t + c_t - x, & \hat{v}_x(t, x) = 1, \\ 0, & |\hat{v}_x(t, x)| < 1, \\ m_t - c_t - x, & \hat{v}_x(t, x) = -1, \end{cases} \tag{2.3.22}$$

where $t \geq 0$, $m_t = \int x\mu_t(dx)$, and $c_t = \inf\{x \mid \hat{v}_x(t, x) = 1\} - m_t = -\sup\{x \mid \hat{v}_x(t, x) = -1\} + m_t$.

Proof. By Proposition 13, $\hat{v}(t, x)$ is convex and differentiable in x, hence for any fixed $t \in [0, \infty)$, $c_t^1 := \inf\{x \mid \hat{v}_x(t, x) = 1\} - m_t$ and $c_t^2 := -\sup\{x \mid \hat{v}_x(t, x) = -1\} + m_t$ exist. By the symmetry of Problem (5.2.1) under a fixed $\{m_t\}_{t\geq 0}$, $\hat{v}(t, m_t + \delta) = \hat{v}(t, m_t - \delta)$ and $\hat{v}_x(t, m_t + \delta) = -\hat{v}_x(t, m_t - \delta)$ for any fixed t and any $\delta > 0$, hence $c_t^1 = c_t^2$, denoted as c_t.

Because $\hat{v}(t, x)$ is convex in x and continuously differentiable in x and t, one can apply the generalized Itô's formula to $\hat{v}(t, x)$ with (2.3.22) and use a similar argument as the verification theorem in [118] to obtain the optimality of (2.3.22). □

Given the optimal control (2.3.22), define a mapping $\Gamma_1 : \mathcal{P}_1(\mathcal{D}([0, \infty))) \to \mathcal{D}([0, \infty))$ such that

$$\Gamma_1\left(\{\mu_t\}_{t\geq 0}\right) = \{\hat{\xi} \mid \{\mu_t\}_{t\geq 0}\}_{t\geq 0}.$$

Step 2: Consistency.

Given Proposition 14 and a fixed flow $\{\mu_t\}_{t\geq 0}$, the optimal control $(\hat{\xi}_t^+, \hat{\xi}_t^-)$ to (2.3.5) is a bang-bang type and the controlled process $\hat{X}_t$ is a reflected Brownian motion with two time-dependent reflected boundaries $m_t + c_t$ and $m_t - c_t$. $m_t + c_t, m_t - c_t \in \mathcal{C}([0, \infty])$ since $\hat{v}(t, x)$ is continuous and differentiable. By Theorem 2.6 in Burdzy, Kang, and Ramanan [31], there exists a unique solution, $\hat{X}_t$, to the SP with time varying domain $\{(t, x) \mid m_t - c_t \leq x \leq c_t + m_t\}$ such that $\hat{X}_t$ is a càdàg process. Furthermore, by Theorem 2.9 in Burdy, Chen, and Sylvester [32], the Kolmogorov forward equation for $\hat{\mu}_t$ can be described as

$$\begin{cases} p_t(t,x) - \frac{1}{2}p_{x,x}(t,x) = 0, & \text{when } |x - m_t| < c_t, \\ p_x(t,x) + 2(\frac{\partial m_t}{\partial t} + \frac{\partial c_t}{\partial t})p(t,x) = 0, & \text{when } x = m_t + c_t, \\ p_x(t,x) - 2(\frac{\partial m_t}{\partial t} - \frac{\partial c_t}{\partial t})p(t,x) = 0, & \text{when } x = m_t - c_t, \end{cases} \tag{2.3.23}$$

with the initial distribution $p(0,x) = \hat{\mu}_0 \in \mathcal{P}_1(\mathbb{R})$, where

$$\hat{\mu}_0(x) = \begin{cases} 0, & x < m_{0-} - c_0 \text{ or } x > m_{0-} + c_0, \\ \mu_{0-}(x), & |x - m_{0-}| < c_0, \\ \mu_{0-}(x) + \int_{-\infty}^{m_{0-}-c_{0-}} \mu_{0-}(dx), & x = m_{0-} - c_0, \\ \mu_{0-}(x) + \int_{m_{0-}+c_{0+}}^{\infty} \mu_{0-}(dx), & x = m_{0-} + c_0. \end{cases} \tag{2.3.24}$$

By Theorem 2.9 in [32], given $m_t + c_t, m_t - c_t \in \mathcal{C}([0,\infty))$, the Kolmogorov forward equation (2.3.23) with the initial distribution $p(0,x) := \hat{\mu}_0(x)$ has a solution.

Step 3: Fixed point analysis. Denote $\hat{\mu}_t$ as the distribution of $\hat{X}_t$, obviously $\hat{\mu}_t \in \mathcal{P}_1(\mathcal{D}([0,\infty))$. Consequently, define $\Gamma_2 : \mathcal{D}([0,\infty)) \to \mathcal{P}_1(\mathcal{D}([0,\infty))$ such that

$$\Gamma_2\left(\hat{\xi}(t,x|\{\mu_t\}_{t\geq 0})\right) = \{\hat{\mu}_t\}_{t\geq 0}.$$

Now, define a mapping $\Gamma : \mathcal{P}_1(\mathcal{D}([0,\infty)) \to \mathcal{P}_1(\mathcal{D}([0,\infty))$ such that

$$\Gamma(\{\mu_t\}_{t\geq 0}) = \Gamma_2 \circ \Gamma_1(\{\mu_t\}_{t\geq 0}) = \{\hat{\mu}\}_{t\geq 0}.$$

One can then update m'_t, and have

$$\begin{aligned} dm'_t &= d\left(\int x p(t,dx)\right) && (2.3.25) \\ &= \left[\frac{1}{2}\int x p_{x,x}(t,dx)\right]dt && (2.3.26) \\ &= \frac{1}{2}[x p_x(t,x)|_{x=m_t+c_t} - x p_x(t,x)|_{x=m_t-c_t} - p(t,x)|_{x=m_t+c_t} + p(t,x)|_{x=m_t-c_t}]dt && (2.3.27) \\ &= \frac{1}{2}\left[\left(-2\left(\frac{dm_t}{dt} + \frac{dc_t}{dt}\right)x - 1\right)p(t,x)\Big|_{x=m_t+c_t}\right. \\ &\quad \left. - \left(2\left(\frac{dm_t}{dt} - \frac{dc_t}{dt}\right)x - 1\right)p(t,x)\Big|_{x=m_t-c_t}\right]dt && (2.3.28) \end{aligned}$$

(2.3.26) comes from (2.3.23), (2.3.27) is from integration by part, and (2.3.28) follows from the boundary conditions. Since μ_{0-} is symmetric around m_{0-} and the optimal control (2.3.22) is an odd function around m_t for any $t \geq 0$, the distribution $p(t,x)$ is symmetric around m_{0-} for any $t \geq 0$.

$$(2.3.28) = -2\left(\frac{dm_t}{dt}m_t + \frac{dc_t}{dt}c_t\right)p(t, m_t + c_t)dt. \tag{2.3.29}$$

Clearly $m_t = m_{0-}$ is one solution to the fixed point equation (2.3.29). This fixed point to Γ is an NE to the MFG (5.2.1) and the associated NE value is smooth in both x, t.

Remark 14.1. *Note that solution m_t $(= m_{0-})$ is time independent and distribution independent. Consequently $v(t, x)$ is time independent and $\frac{dc_t}{dt} = 0$. In fact, this time independent property of the value function $v(t, x)$ reduces the HJB equation (2.3.5) from a parabolic form to an elliptic one. However, there might be time-dependent NE solution(s) with non-constant mean position $\{m_t\}_{t\geq 0}$ for Eqn. (2.3.29). We are unable to verify the existence/nonexistence of such solutions.*

On a related note, if instead a stationary MFG (SMFG) is specified by replacing $h(X_t - m_t)$ with $h(X_t - \lim_{t\to\infty} m_t)$, the associated HJB equation (2.3.5) will also be elliptic. (See Appendix A.4 for more precise definition of the SMFG formulation.) In this case, one can use the same approach to derive infinitely many NEs of the bang-bang type, with the controlled dynamics reflected at $m - c$ and $m + c$ for any constant m. Note however, the NE for the SMFG when $m \neq m_{0-}$ is not an NE for the MFG (5.2.1).

2.4 Relation between the N-player game and the MFG

Convergence of Game Values

First, from Theorem 9, one can see, with the detailed proof given in Appendix A.3,

Proposition 15. *Given c_N the unique solution to (4.2.4) and $c > 0$ the unique solution to (4.2.7),*

$$\lim_{N\to\infty} c_N = c.$$

When $h(x) = x^2$, c_N is a decreasing function of N.

Remark 15.1. *It is no surprise from our earlier analysis that MFGs are different in nature from N-player games. For instance, the MFG degenerates to a single-player game in the sense that its NE is threshold-type bang-bang policy where the threshold is state independent while the NEs for the N-player game are state dependent. Nevertheless, it is still somewhat unexpected to see the total collapse of the MFG to the single player problem from the above proposition. This could be a result of over aggregation in the MFG formulation: players become more anticipative when they are assumed to be identical.*

Next, denote $v^i_{(N)}$ as the NE value of player i in the N-player game. By (2.2.26), when $x_1 = \cdots = x_N = x$,

$$v^i_{(N)}(x, x, \cdots, x) = \frac{-p_N''(c_N)}{\frac{2(N-1)\alpha}{N}\cosh\left(c_N\sqrt{\frac{2(N-1)\alpha}{N}}\right)} + p_N(0). \tag{2.4.1}$$

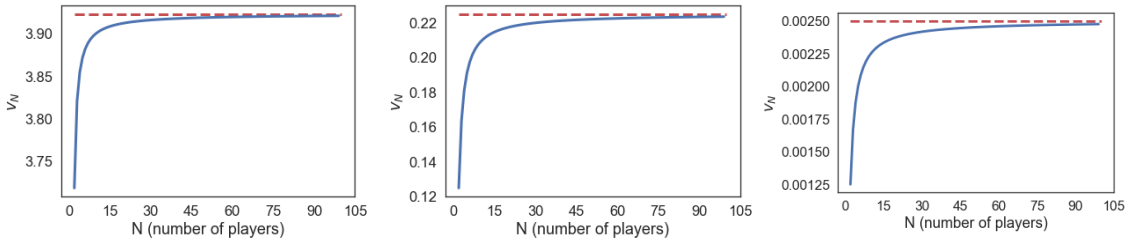
In particular, $v^i_{(N)}(x, x, \cdots, x)$ is independent of x. Moreover, from Proposition 15 and the smoothness of $P_N(x)$, it is easy to verify that

$$\begin{cases} \lim_{N\to\infty} p_N''(c_N) = p_N''(c), \\ \lim_{N\to\infty} \frac{1}{\frac{2(N-1)\alpha}{N}\cosh\left(c_N\sqrt{\frac{2(N-1)\alpha}{N}}\right)} = \frac{1}{2\alpha\cosh\left(c\sqrt{2\alpha}\right)}, \\ \lim_{N\to\infty} p_N(0) = p_1(0). \end{cases}$$

That is,

Proposition 16. *For any $x \in \mathbb{R}$, $\lim_{N\to\infty} v^i_{(N)}(x, x, \cdots, x) = v^*(x)$, where v^* is the NE value of player i in MFG (5.2.1) with $\mu_{0-} = \delta(x)$.*

Figure 2.3 shows the convergence of $v^i_{(N)}(x, x, \cdots, x)$ with $h = x^2$ and with different choices of α. The MFG is illustrated by the dashed red horizontal line.



(a) $\alpha = 0.2$ (b) $\alpha = 2$ (c) $\alpha = 20$

Figure 2.3: Convergence of v_N with different discount factors

Remark 16.1. *Figure 2.3 indicates that v_N is an increasing function of N given any fixed decay parameter α. This implies that when the number of players increases, it is more costly for players to keep track of other players before making decisions. Meanwhile, $v^*(x)$ being a decreasing function of α indicates that the bigger the α, the less frequent players will intervene.*

Approximating the N-player Game by the MFG

One can further show that the NE of MFG given in (2.3.2) is an ϵ-NE for the game in (**N-player**).

Definition 17 (ϵ-NE)**.** *For the game (**N-player**) with an initial distribution μ_{0-}, a control vector $\boldsymbol{\xi} = (\xi^1, \ldots, \xi^N)$ is called its ϵ-NE, if for any $i = 1, \ldots, N$ and any control $\xi^{i'}$ such that $\left(\boldsymbol{\xi}^{-i}, \xi^{i'}\right) = \left(\xi^1, \ldots, \xi^{i-1}, \xi^{i'}, \xi^{i+1}, \ldots, \xi^N\right) \in \mathcal{S}_N$,*

$$\mathbb{E}\left[J^i_{(N)}\left(\boldsymbol{X}_{0-}; \boldsymbol{\xi}\right)\right] \leq \mathbb{E}\left[J^i_{(N)}\left(\boldsymbol{X}_{0-}; \left(\boldsymbol{\xi}^{-i}, \xi^{i'}\right)\right)\right] + \epsilon. \tag{2.4.2}$$

Here X^i_{0-} $(i = 1, 2, \cdots, N)$ are independent samples from distribution μ_{0-}, and $\mathcal{S}_N$ is defined in (4.2.1).

Theorem 18 (ϵ-NE of the N-player game)**.** *Let ξ^* be the NE of MFG given in (2.3.2), then it is an ϵ-NE of the game (**N-player**), with $\epsilon = O\left(\frac{1}{\sqrt{N}}\right)$.*

Proof. Given the game (**N-player**) with $m_{0-} = \int x\mu_{0-}(dx) \in \mathbb{R}$, assume that each player i in the N-player game takes the control $(\xi^{i*,+}_t, \xi^{i*,-}_t)$ according to the NE of the MFG such that

$$\begin{aligned} \xi^{i*,+}_t &= \max\left\{0, \max_{0\leq u\leq t}\left\{m_{0-} - X^i_{0-} - B^i_u + \xi^{i*,-}_u - c\right\}\right\}, \\ \xi^{i*,-}_t &= \max\left\{0, \max_{0\leq u\leq t}\left\{X^i_{0-} - m_{0-} + B^i_u + \xi^{i*,+}_u - c\right\}\right\}. \end{aligned} \tag{2.4.3}$$

To see that $\boldsymbol{\xi}^* = \left(\xi^{1*}, \ldots, \xi^{N*}\right) \in \mathcal{S}_N$, define

$$\begin{aligned} \mathcal{CW}_{mfg} &= \left\{\boldsymbol{x} \in \mathbb{R}^N \middle|\ |x_i - m_{0-}| < c \text{ for } i = 1, 2, \cdots, N\right\}, \\ E^-_{mfg,i} &= \left\{\boldsymbol{x} \in \mathbb{R}^N \middle|\ x^i - m_{0-} \leq -c\right\}, \\ E^+_{mfg,i} &= \left\{\boldsymbol{x} \in \mathbb{R}^N \middle|\ x^i - m_{0-} \geq c\right\}, \end{aligned}$$

with the partition

$$\begin{aligned} Q_{mfg,i} = \Big\{\boldsymbol{x} \in \mathbb{R}^N \Big| \quad & \left|x^i - m_{0-}\right| \geq \left|x^k - m_{0-}\right|, \text{ for any } k < i; \\ & \left|x^i - m_{0-}\right| > \left|x^k - m_{0-}\right|, \text{ for any } k > i\Big\}. \end{aligned}$$

Then the control in (2.4.3) corresponds to the action region $\mathcal{A}_{mfg,i} = \left\{E^-_{mfg,i} \cup E^+_{mfg,i}\right\} \cap Q_{mfg,i}$. The independence of $\{B^1_t, \ldots, B^N_t\}$ and the continuity of $\{X^{1*}_t, \ldots, X^{N*}_t\}_{t>0}$ imply that for any $t \geq 0$ $\mathbb{P}\left(\Pi_{i=1,\ldots,N} d\xi^{i*}_t = 0\right) = 1$.

Suppose that only one player, and without loss of generality, player one, deviates her control $\eta_t = (\eta^+_t, \eta^-_t)$ from all the other players such that $(\boldsymbol{\xi}^{-i*}, \eta) \in \mathcal{S}_N$. Let $\hat{X}^1_t$ be the new position of player one under control (η^+_t, η^-_t) with initial value X^1_{0-}. Then

$$
\begin{aligned}
& h\left(\hat{X}_t^1 - \frac{1}{N}\left(\hat{X}_t^1 + \sum_{j=2,\dots,N} X_t^j\right)\right) \\
= \;& h\left(\hat{X}_t^1 - \frac{1}{N}\hat{X}_t^1 - \frac{N-1}{N}m_{0-}\right) + h'\left(\hat{X}_t^1 - \frac{1}{N}\hat{X}_t^1 - \frac{N-1}{N}m_{0-}\right)\left(\frac{\sum_{j=2,\dots,N} X_t^j}{N} - \frac{N-1}{N}m_{0-}\right) \\
& + \frac{h''(U_t)}{2}\left(\frac{\sum_{j=2,\dots,N} X_t^j}{N} - \frac{N-1}{N}m_{0-}\right)^2,
\end{aligned}
$$

where U_t is a process between $\left(\hat{X}_t^1 - \frac{1}{N}\hat{X}_t^1 - \frac{N-1}{N}m_{0-}\right)$ and $\left(\hat{X}_t^1 - \frac{1}{N}\left(\hat{X}_t^1 + \sum_{j=2}^N X_t^j\right)\right)$.
By Assumption **A1**,

$$
\begin{aligned}
& h\left(\hat{X}_t^1 - \frac{1}{N}\left(\hat{X}_t^1 + \sum_{j=2,\dots,N} X_t^j\right)\right) \\
\leq \;& h\left(\hat{X}_t^1 - \frac{1}{N}\hat{X}_t^1 - \frac{N-1}{N}m_{0-}\right) + h'\left(\hat{X}_t^1 - \frac{1}{N}\hat{X}_t^1 - \frac{N-1}{N}m_{0-}\right)\left(\frac{\sum_{j=2}^N X_t^j}{N} - \frac{N-1}{N}m_{0-}\right) \\
& + \frac{K}{2}\left(\frac{\sum_{j=2}^N X_t^j}{N} - \frac{N-1}{N}m_{0-}\right)^2 \\
\leq \;& h\left(\hat{X}_t^1 - \frac{1}{N}\hat{X}_t^1 - \frac{N-1}{N}m_{0-}\right) + K\left|\hat{X}_t^1 - \frac{1}{N}\hat{X}_t^1 - \frac{N-1}{N}m_{0-}\right| \cdot \left|\frac{\sum_{j=2}^N X_t^j}{N} - \frac{N-1}{N}m_{0-}\right| \\
& + \frac{K}{2}\left(\frac{\sum_{j=2}^N X_t^j}{N} - \frac{N-1}{N}m_{0-}\right)^2.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& h\left(\hat{X}_t^1 - \frac{1}{N}\left(\hat{X}_t^1 + \sum_{j=2,\dots,N} X_t^j\right)\right) \\
\geq \;& h\left(\hat{X}_t^1 - \frac{1}{N}\hat{X}_t^1 - \frac{N-1}{N}m_{0-}\right) - K\left|\hat{X}_t^1 - \frac{1}{N}\hat{X}_t^1 - \frac{N-1}{N}m_{0-}\right| \cdot \left|\frac{\sum_{j=2}^N X_t^j}{N} - \frac{N-1}{N}m_{0-}\right| \\
& - \frac{K}{2}\left(\frac{\sum_{j=2}^N X_t^j}{N} - \frac{N-1}{N}m_{0-}\right)^2.
\end{aligned}
$$

Moreover, under the control (2.4.3), X_t^j $(j = 2, 3, \cdots, N)$ are independent and identically distributed and $|X_t^j - m_{0-}| \leq c$ a.s.. Therefore,

$$
\mathbb{E}\left(\frac{\sum_{j=2}^N X_t^j}{N} - \frac{N-1}{N}m_{0-}\right)^2 = \frac{\sum_{j=1}^N Var(X_t^j)}{N^2} \leq \frac{c^2}{N} = O(\frac{1}{N}),
$$

and

$$\mathbb{E}\left|\frac{\sum_{j=2}^N X_t^j}{N}-\frac{N-1}{N}m_{0-}\right|\le\left(\mathbb{E}\left(\frac{\sum_{j=2}^N X_t^j}{N}-\frac{N-1}{N}m_{0-}\right)^2\right)^{1/2}=O(\frac{1}{\sqrt{N}}).$$

Therefore by the boundedness of X_t^j $(j=2,3,\cdots,N)$ and by the Fubini Theorem,

$$\mathbb{E}\int_0^\infty e^{-\alpha t}\left(\frac{\sum_{j=2}^N X_t^j}{N}-\frac{N-1}{N}m_{0-}\right)^2dt=\int_0^\infty e^{-\alpha t}\mathbb{E}\left(\frac{\sum_{j=2}^N X_t^j}{N}-\frac{N-1}{N}m_{0-}\right)^2dt=O\left(\frac{1}{N}\right).$$

Similarly, when $\hat{X}_t^1$ is under the threshold-type control,

$$\mathbb{E}\int_0^\infty e^{-\alpha t}\left|\hat{X}_t^1-\frac{1}{N}\hat{X}_t^1-\frac{N-1}{N}m_{0-}\right|\cdot\left|\frac{\sum_{j=2}^N X_t^j}{N}-\frac{N-1}{N}m_{0-}\right|dt=O(\frac{1}{\sqrt{N}}). \quad (2.4.4)$$

Now, to minimize the following payoff function

$$\begin{aligned}&\mathbb{E}\int_s^\infty e^{-\alpha t}\left[h\left(\hat{X}_t^1-\frac{1}{N}\hat{X}_t^1-\frac{1}{N}\sum_{j=2}^N X_t^j\right)dt+d\eta_t^+ +d\eta_t^-\right]\\ =\;&\mathbb{E}\int_s^\infty e^{-\alpha t}\left[h\left(\frac{N-1}{N}\hat{X}_t^1-\frac{N-1}{N}m_{0-}\right)dt+d\eta_t^+ +d\eta_t^-\right]+O\left(\frac{1}{\sqrt{N}}\right). \quad (2.4.5)\end{aligned}$$

is equivalent to solving the original fuel follower problem (2.2.2) with a modified running cost $h(\frac{N-1}{N}(\cdot-m_{0-}))$. Since the value function for (2.2.2) is of a linear growth,

$$\begin{aligned}(2.4.5)\;&\geq\;\mathbb{E}\int_s^\infty e^{-\alpha t}\left[h\left(\frac{N-1}{N}(\hat{X}_t^1-m_{0-})\right)dt+d\eta_t^{1*,+}+d\eta_t^{1*,-}\right]+O\left(\frac{1}{\sqrt{N}}\right) \quad (2.4.6)\\ &=\;\mathbb{E}\left[v^*(X_{0-}^1)\right]+O\left(\frac{1}{N}\right)+O\left(\frac{1}{\sqrt{N}}\right) \quad (2.4.7)\end{aligned}$$

where $v^*(x)$ is defined in (2.3.3) and the expectation in (2.4.7) is with respect to the initial distribution μ_{0-}. The above analysis holds for any $(\eta_t^+,\eta_t^-)\in\mathcal{U}_N^i$ such that $(\boldsymbol{\xi}^{-i*},\eta)\in\mathcal{S}_N$. Hence the conclusion. □

2.5 Discussions

Multiple Explicit NEs for $N=2$

When $N=2$, h is symmetric with $h(X_t^1-m_t^{(2)})=h(X_t^2-m_t^{(2)})=h\left(\frac{X_t^1-X_t^2}{2}\right)$. This symmetry simplifies significantly the solution structure and allows for the construction of

multiple NEs. Indeed, given the partition Q_i in (2.2.13) for $N = 2$, $Q_1 = 0$, $Q_2 = \mathbb{R}^2$, one can write the NE and their corresponding values explicitly.

$$\begin{aligned}\xi_t^{2*} &= (\xi_t^{2*,+}, \xi_t^{2*,-}) \\ &= \left(\max\left\{0, \max_{0\le u\le t}\{-x^2 + x^1 - B_u^2 + B_u^1 + \xi_u^{2*,-} - c_2\}\right\},\right. \\ &\quad \left.\max\left\{0, \max_{0\le u\le t}\{x^2 - x^1 + B_u^2 - B_u^1 + \xi_u^{2*,+} - c_2\}\right\}\right), \\ \xi_t^{1*} &= \left(\xi_t^{1*,+}, \xi_t^{1*,-}\right) = (0,0),\end{aligned} \tag{2.5.1}$$

where $c_2 > 0$ is the unique positive solution of

$$\frac{1}{\sqrt{\alpha}}\tanh\left(\sqrt{\alpha}x\right) = \frac{p_2'(x) - 1}{p_2''(x)}, \tag{2.5.2}$$

with

$$p_2(x) = \mathbb{E}\left[\int_0^\infty e^{-\alpha t} h\left(\frac{x}{2} + \frac{\sqrt{2}B_t}{2}\right) dt\right].$$

And the NE values are

$$v^2(x^1,x^2) = \begin{cases} v^2(x^1, x^1 - c_2) - c_2 - x^2 + x^1, & x^2 - x^1 \le -c_2, \\ -\frac{p_2''(c_2)\cosh\left(\sqrt{\alpha}(x^2 - x^1)\right)}{\alpha\cosh\left(c_2\sqrt{\alpha}\right)} + p_2(x^2 - x^1), & |x^2 - x^1| < c_2, \\ x^2 - x^1 - c_2 + v^2(x^1, x^1 + c_2), & x^2 - x^1 \ge c_2, \end{cases} \tag{2.5.3}$$

and

$$v^1(x^1,x^2) = \begin{cases} v^1(x^1, x^1 + c_2), & x^1 - x^2 \le -c_2, \\ -\frac{p_2''(c_2)\cosh\left(\sqrt{\alpha}(x^1 - x^2)\right)}{\alpha\cosh\left(c_2\sqrt{\alpha}\right)} + p_2(x^1 - x^2), & |x^2 - x^1| < c_2, \\ v^1(x^1, x^1 - c_2), & x^1 - x^2 \ge c_2. \end{cases} \tag{2.5.4}$$

There is in fact more than one NE. For instance, in addition to the above constructed NE, labeled as **Case 1**, there are more NEs, including

Case 2: $\mathcal{A}_1 = \{(x^1,x^2) \mid x^1 - x^2 > c_2 \text{ or } x^1 - x^2 < -c_2\}$ and $\mathcal{A}_2 = \emptyset$,

Case 3: $\mathcal{A}_1 = \{(x^1,x^2) \mid x^1 - x^2 < -c_2\}$ and $\mathcal{A}_2 = \{(x^1,x^2) \mid x^1 - x^2 > c_2\}$,

Case 4: $\mathcal{A}_1 = \{(x^1,x^2) \mid x^1 - x^2 > c_2\}$ and $\mathcal{A}_2 = \{(x^1,x^2) \mid x^1 - x^2 < -c_2\}$.

In **Case 4**, clearly

$$\begin{aligned}\xi_t^{1*} &= -\max\left\{0, \max_{0\le u\le t}\left\{0, x^1 - x^2 + B_u^1 - B_u^2 - \xi_u^{2*} - c_2\right\}\right\},\\ \xi_t^{2*} &= -\max\left\{0, \max_{0\le u\le t}\left\{0, x^2 - x^1 + B_u^2 - B_u^1 - \xi_u^{1*} - c_2\right\}\right\},\end{aligned}$$

and the associated NE values are

$$v^1(x^1,x^2) = \begin{cases} v^1(x^1, x^1 + c_2), & x^1 - x^2 \le -c_2,\\ -\frac{p_2''(c_2)\cosh\left(\sqrt{\alpha}(x^1-x^2)\right)}{\alpha\cosh\left(c_2\sqrt{\alpha}\right)} + p_2(x^1 - x^2), & |x^1 - x^2| < c_2,\\ x^1 - x^2 - c_2 + v^1(x^2 + c_2, x^2), & x^1 - x^2 \ge c_2,\end{cases}$$

and

$$v^2(x^1,x^2) = \begin{cases} v^2(x^2 + c_2, x^2), & x^2 - x^1 \le -c_2,\\ -\frac{p_2''(c_2)\cosh\left(\sqrt{\alpha}(x^2-x^1)\right)}{\alpha\cosh\left(c_2\sqrt{\alpha}\right)} + p_2(x^2 - x^1), & |x^2 - x^1| < c_2,\\ x^2 - x^1 - c_2 + v^2(x^1, x^1 + c_2), & x^2 - x^1 \ge c_2.\end{cases}$$

Figure 2.4 illustrates all four NEs.

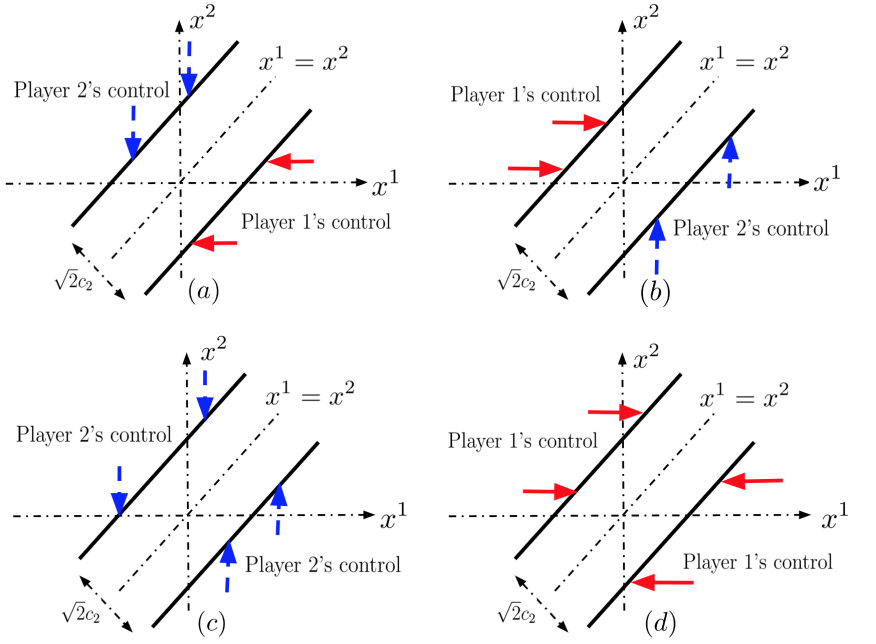


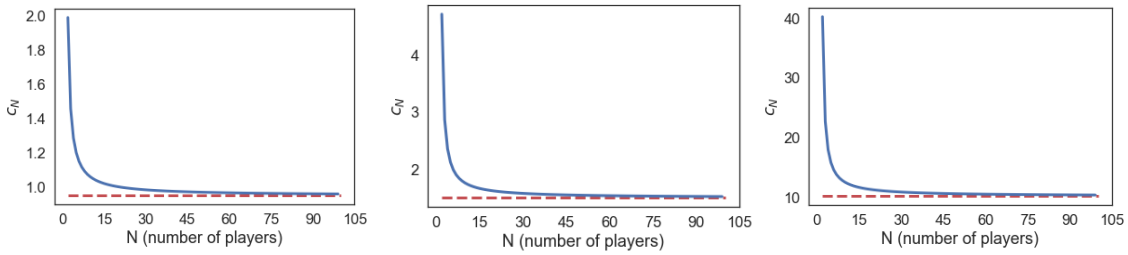
Figure 2.4: Four NEs when $N = 2$

With Varying α

Proposition 19. *When $h(x) = x^2$ and $\alpha \geq 2^{-\frac{1}{3}} \frac{N-1}{N}$, c_N increases with respect to α.*

The proposition follows from simple calculations. Take $h(x) = x^2$, $\frac{p_N'(x)-1}{p_N''(x)} = x - \frac{\alpha k_N^2}{2}$ with $k_N = \frac{N}{N-1} > 1$. Rewrite f_N as $f_N(x,\alpha) = \frac{\sqrt{k_N}}{\sqrt{2\alpha}} \tanh\left(\frac{\sqrt{2\alpha}}{\sqrt{k_N}}x\right) - x + \frac{k_N^2 \alpha}{2}$. Then $\frac{\partial f_N}{\partial x} = -\tanh^2\left(\frac{\sqrt{2\alpha}}{\sqrt{k_N}}x\right)$ and $\frac{\partial f_N}{\partial \alpha} = \frac{x}{2\alpha}\left(1 - \tanh^2\left(\frac{\sqrt{2\alpha}}{\sqrt{k_N}}x\right)\right) - \frac{\sqrt{k_N}}{2\alpha\sqrt{2\alpha}} \tanh\left(\frac{\sqrt{2\alpha}}{\sqrt{k_N}}x\right) + \frac{k_N^2}{2}$. One can verify that $\frac{\partial f_N}{\partial x} < 0$ for any α and $\frac{\partial f_N}{\partial \alpha} > 0$ when $\alpha > 2^{-\frac{1}{3}} k_N^{-1}$. Hence $\frac{\partial c_N}{\partial \alpha} > 0$ when $\alpha > 2^{-\frac{1}{3}} k_N^{-1}$ follows from the chain rule and from $f(c_N(\alpha), \alpha) = 0$ for any N.

Figure 2.5, illustrates the convergence of c_N with different discount factor α. The value of c is shown in the red dash line.



(a) $\alpha = 0.2$ (b) $\alpha = 2$ (c) $\alpha = 20$

Figure 2.5: Convergence of c_N with different discount factors

Remark 19.1. *Figure 2.5 indicates that c_N is a decreasing function of N for any given discount factor α. This implies that players will intervene more frequently with more players in the game. Meanwhile, c being a decreasing function of α indicates that the bigger the α, the less frequent players will intervene. These are consistent with Figure 2.3.*

It is worth noting that the analysis for $\alpha_1 = \cdots = \alpha_N = \alpha$ can be easily extended to the cases when α_i's are different. The exact forms of the NEs, however, may be more complicated, as illustrated in the case of $N = 2$ below.

When $N = 2$, denote α_i as the discount parameter for player i $(i = 1, 2)$. Denote $c_2^{(i)} > 0$ as the unique solution of

$$\frac{1}{\sqrt{\alpha_i}} \tanh\left(\sqrt{\alpha_i} x\right) = \frac{p_2'(x, \alpha_i) - 1}{p_2''(x, \alpha_i)}, \tag{2.5.5}$$

with

$$p_2(x, \alpha_i) = \mathbb{E}\left[\int_0^\infty e^{-\alpha_i t} h\left(\frac{x}{2} + \frac{\sqrt{2} B_t}{2}\right) dt\right].$$

Corollary 19.1 ($N = 2$ with $\alpha_1 \neq \alpha_2$)**.** *Assume* ***A1*** *for game (**N-player**). If $\alpha_2 > \alpha_1 > 2^{-\frac{4}{3}}$, then $c_2^{(2)} > c_2^{(1)}$. The following controls*

$$\begin{aligned}\xi_t^{2*,+} &= \mathbf{1}_{\left\{x^2-x^1<-c_2^{(2)}\right\}}\left(-c_2^{(2)}-x^2+x^1\right),\\ \xi_t^{2*,-} &= \mathbf{1}_{\left\{x^2-x^1>c_2^{(2)}\right\}}\left(c_2^{(2)}-x^2+x^1\right),\end{aligned}$$

and

$$\begin{aligned}\xi_t^{1*,+} &= \max\left\{0, \max_{0\le u\le t}\{x^2+B_u^2+\xi_0^{2*,+}-\xi_0^{2*,-}-x^1-B_u^1+\xi_u^{1*,-}-c_2^{(1)}\}\right\},\\ \xi_t^{1*,-} &= \max\left\{0, \max_{0\le u\le t}\{-x^2-B_u^2-\xi_0^{2*,+}+\xi_0^{2*,-}+x^1+B_u^1+\xi_u^{1*,+}-c_2^{(1)}\}\right\}.\end{aligned}$$

give a Markovian NE. The corresponding NE values are

$$v^1(x^1,x^2)=\begin{cases} v^1(x^1,x^1+c_2^{(2)}) & x^1-x^2\le -c_2^{(2)},\\ v^1(x^2-c_2^{(1)},x^2)+x^2-x^1-c_2^{(1)}, & -c_2^{(2)}\le x^1-x^2\le -c_2^{(1)},\\ -\frac{p_2''(c_2^{(1)})\cosh\left(\sqrt{\alpha}(x^1-x^2)\right)}{\alpha\cosh\left(c_2^{(1)}\sqrt{\alpha}\right)}+p_2(x^1-x^2), & |x^1-x^2|\le c_2^{(1)},\\ x^1-x^2-c_2^{(1)}+v^1(x^2+c_2^{(1)},x^2), & c_2^{(1)}\le x^1-x^2\le c_2^{(2)},\\ v^1(x^1,x^1-c_2^{(2)}) & x^1-x^2\ge c_2^{(2)},\end{cases} \tag{2.5.6}$$

and

$$v^2(x^1,x^2)=\begin{cases} v^2(x^1,x^1-c_2^{(2)})+x^1-x^2-c_2^{(2)}, & x^2-x^1\le -c_2^{(2)},\\ v^2(x^2+c_2^{(1)},x^2) & -c_2^{(2)}\le x^2-x^1\le -c_2^{(1)},\\ -\frac{p_2''(c_2^{(1)})\cosh\left(\sqrt{\alpha}(x^2-x^1)\right)}{\alpha\cosh\left(c_2^{(1)}\sqrt{\alpha}\right)}+p_2(x^2-x^1), & |x^2-x^1|\le c_2^{(1)},\\ v^2(x^2-c_2^{(1)},x^2) & c_2^{(1)}\le x^2-x^1\le c_2^{(2)},\\ x^2-x^1-c_2^{(2)}+v^2(x^1,x^1+c_2^{(2)}), & x^2-x^1\ge c_2^{(2)}.\end{cases} \tag{2.5.7}$$

Figure 2.6 shows the NE defined in Corollary 19.1.

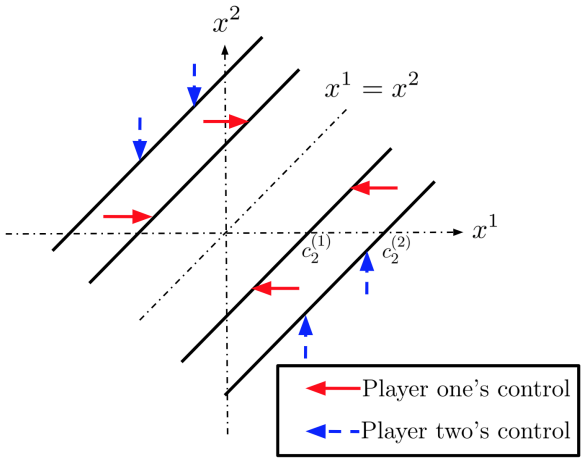


Figure 2.6: N=2 with different α values

Chapter 3

Stochastic Game with Resource Constraints

3.1 Introduction

N-player non-zero-sum stochastic games are notoriously hard. Recently there has been a surge of interest on Mean Field Games (MFGs), pioneered by [104, 137, 136, 139]. With an ingenious aggregation approach, MFGs nicely reduce the complexity of N-player games by focusing on $N \to \infty$. However, there are undesirable consequences of the MFG aggregation approach and a growing number of studies [42, 94, 130] point to the risk of using MFGs for analyzing N-player games. For instance, Nash equilibria (NEs) of MFGs tend to collapse to that of a single-player game, offering no or limited insight into the general solution structure of N-player games.

Motivated by the need for a more in-depth study of N-player stochastic games, in this work we formulate and analyze a classical of stochastic N-player games that originated from the classic finite fuel problem. There are many reasons to consider this type of games. Firstly, the finite fuel problem is one of the landmarks in stochastic control theory and a game formulation is natural [14, 19, 28, 74, 113, 122, 123]. Secondly, its simple yet insightful solution structures have had a wide range of applications including economics and finance [7, 56, 67, 143], operations research [60, 61, 95, 127], and queuing theory [126], in addition to the theory of stochastic controls [8, 29, 30, 57, 62, 76, 98, 168, 169, 66, 179, 184]. Thirdly, there is no prior work analyzing its stochastic game counterpart except for the special case of $N = 2$ and without the fuel constraint [68, 94, 96, 99, 119, 129, 145]. We hope that analyzing this game can shed more light on the fundamental differences between control problems and stochastic games and thus provide useful insights into the intrinsic difficulty of the latter.

The stochastic game presented in this paper goes as follows. There are N players whose dynamics $\boldsymbol{X}_t = (X_t^1, \cdots, X_t^N)$ are governed by the following N-dimensional diffusion process:

$$dX_t^i = b_i(\boldsymbol{X}_{t-})dt + \boldsymbol{\sigma_i}(\boldsymbol{X}_{t-})d\boldsymbol{B}(t) + d\xi_t^{i+} - d\xi_t^{i-}, \quad X_{0-}^i = x^i, \quad (i = 1, \cdots, N), \qquad (3.1.1)$$

where $\boldsymbol{B} := (B^1, \cdots, B^N)$ is a standard N-dimensional Brownian motion in a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, with drift $\boldsymbol{b} := (b_1, \cdots, b_N)$ and covariance matrix $\boldsymbol{\sigma} := (\boldsymbol{\sigma_1}, \cdots, \boldsymbol{\sigma_N})$ satisfying appropriate regularity conditions. Player i's control, $\xi^i = (\xi^{i+}, \xi^{i-})$, is a pair of non-decreasing and càdlàg processes, and of finite variation. Each player has access to some or all of M types of resources. Players interact through their objective functions $h^i(X_t^1, \cdots, X_t^N)$, as well as their shared resources that are the "fuels" of their control. The accessibility of these resources to players and how these resources are consumed by their respective players are governed by a matrix $\boldsymbol{A} := (a_{ij})_{i,j} \in \mathbb{R}^{N\times M}$. The goal of the game is for player i to minimize

$$\mathbb{E}\int_0^\infty e^{-\alpha t} h^i(X_t^1, \cdots, X_t^N)dt.$$

over appropriate admissible game strategies, which are specified in Section 3.2. When $M = 1$ and $\boldsymbol{A} = [1, 1, \cdots, 1]^T \in \mathbb{R}^{N\times 1}$, this is a pooling game $\boldsymbol{C_p}$ corresponding to the N-player finite fuel game where the N players share a fixed amount of the same resource. When $M = N$ and $\boldsymbol{A} = \boldsymbol{I_N}$, this is an N-player game $\boldsymbol{C_d}$ where each player has her individual fixed amount of resource. In general, this matrix $\boldsymbol{A}$ describes the network structure of the N-player game. Note that this N-player game cannot be simply analyzed with an MFG approach as the network structure would collapse if an aggregation approach was applied.

We will analyze the NEs of this stochastic game. We first derive sufficient conditions for the NE policy in the form of a verification theorem (Theorem 22), which reveals an essential game element regarding the interactions among players. This is the Hamilton–Jacobi–Bellman (HJB) representation of the conditional optimality for NE in a stochastic game. To understand the structural properties of the NEs, we proceed further to analyze this stochastic game in terms of the game values, the NE strategies, and the controlled dynamics. Mathematically, the analysis involves first solving a multi-dimensional free boundary problem and then a Skorokhod problem with a *moving* boundary. The boundary is "moving" in that it moves in response to both the changes of the system and the control strategies of other players. The analytical solution is derived by first exploring the two special games $\boldsymbol{C_p}$ and $\boldsymbol{C_d}$. Analyzing these two types of games provides key insights into the solution structure of the general game. Finally, we reformulate the NE strategies in the form of controlled rank-dependent stochastic differential equations (SDEs), and compare game values with games $\boldsymbol{C_p}$ and $\boldsymbol{C_d}$.

Main contributions. (i) In the verification theorem for N-player games, we obtain the form of the HJB equations for general stochastic games with singular controls. Unlike all previous analysis that focused on two-player games, we show that in addition to the standard HJBs that correspond to stochastic control problems, there is an essential term that is unique to stochastic games. This term represents the interactions among players, especially the ones who are active and those who are waiting. This critical term was missing in two-player stochastic games and was simply (mis)understood as a regularity condition (Remark 21.1).

(*ii*) The structural difference between games and control problems is further revealed in the explicit solution to the NEs for N-player games. In a Markovian control problem, a free boundary depends on the state of the system; in stochastic games, however, the "face" of the boundary moves based on the action of herself and interaction among players in the game (Figure 3.4). Note that this free boundary for stochastic games with an infinite time horizon *moves* in a different sense from the one in [57] for finite time control problems where the boundary is time dependent. Rather it moves due to changes of the system and the competition in the game.
(*iii*) This difference is further highlighted in the framework of controlled rank-dependent SDEs. To the best of our knowledges, this is the first time a stochastic game is explicitly connected with rank-dependent SDEs in a more general form. This new form of rank-dependent SDEs presents a fresh class of yet-to-be studied SDEs (Section 3.7).
(*iv*) Finally, stochastic games considered in this paper are resource allocation games. Resource allocation problems have a wide range of applications including cloud computing, smart power grid control, and multimedia wireless networks [83, 84, 141, 165, 186]. However, the existing literature has been unsuccessful in analyzing the resource allocation problem in the setting of stochastic games. Besides the technical contributions, our analysis provides a useful economic insight: in a stochastic game of resource allocations, sharing has lower cost than dividing and pooling yields the lowest cost for each player.

Related work. There are several papers on non-zero-sum games with singular controls [68, 94, 96, 99, 119, 129, 145]. All of these works are games without the fuel constraint and thus are built on one-dimensional stochastic control problems. Furthermore, except for [94], all of these papers are restricted to the case of $N = 2$. Most importantly, because of the restricted problem setting, none of these works managed to discover the critical structural difference between stochastic games and controls. We believe our work is the first to complete the mathematical analysis on an N-player stochastic game based on an original two-dimensional control problem.

There has been some works on reflected SDEs in time-dependent or state-dependent domains. Reflected Brownian motion in smooth time-dependent domains with normal reflection was considered by [34, 33] via the heat equation. The one-dimensional case was also studied by [35] through the Skorokhod problem. Later, [144, 154] give the construction of reflected SDEs in non-smooth time-dependent domains with oblique reflection. There is some work, i.e. [36, 170, 183], on Brownian motion reflected on another Brownian motion, motivated by the study of the Brownian web. However, none of these works involve controls. [25] considers reflected SDEs in the orthant $\mathbb{R}^d_+$ and focuses on the viscosity solution analysis.

In our work the controlled dynamics and the "moving" free boundary are recast in the framework of *controlled* rank-dependent SDEs. The rank-dependent SDEs without controls arise in the "Up the River" problem [4] and in stochastic portfolio theory [79], including the well-studied *Atlas model* for the ergodicity and sample path properties [9, 108, 109, 110, 157, 166, 167] and for the hydrodynamic limit and fluctuations of the Atlas model [37, 70,

177]. Compared to the well-known rank-dependent SDEs, rank-dependent SDEs with an additional control component has not been studied before. We establish the existence of the solution by directly constructing a reflected diffusion process. (See Section 3.7 for further discussions.)

Notations and organization. Throughout the paper, we denote vectors/matrices by bold case letters, e.g., $\boldsymbol{x}$ and $\boldsymbol{X}$. The transpose of a real vector $\boldsymbol{x}$ is denoted as $\boldsymbol{x}^T$. For a vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|$ denotes its l_2 norm. For a matrix $\boldsymbol{X}$, $\|\boldsymbol{X}\|$ denotes its spectral norm.

The paper is organized as follows. Section 3.2 presents the mathematical formulation of the N-player game. Section 3.3 provides verification theorem for sufficient conditions of the NE of the game and the existence of Skorokhod problem for NE strategies. Section 3.4 studies game $\boldsymbol{C_p}$ and Section 3.5 studies game $\boldsymbol{C_d}$. With the insight from these two games, Section 3.6 analyzes the general N-player game $\boldsymbol{C}$. Section 3.7 compares games $\boldsymbol{C_p}$, $\boldsymbol{C_d}$ and $\boldsymbol{C}$, discusses the game values and their economic implications, and unifies their corresponding controlled dynamics in the framework of the controlled rank-dependent SDEs.

3.2 Problem Setup

Now we present the mathematical formulation for the stochastic N-player game.

Controlled dynamics. Let $(X_t^i)_{t\geq 0}$ be the position of player $i, 1 \leq i \leq N$. In the absence of controls, $\boldsymbol{X}_t = (X_t^1, \cdots, X_t^N)$ is governed by the stochastic differential equation (SDE):

$$d\boldsymbol{X}_t = \boldsymbol{b}(\boldsymbol{X}_t)dt + \boldsymbol{\sigma}(\boldsymbol{X}_t)d\boldsymbol{B}(t), \quad \boldsymbol{X}_{0-} = (x^1, \cdots, x^N), \tag{3.2.1}$$

where $\boldsymbol{B} := (B^1, \cdots, B^N)$ is a standard N-dimensional Brownian motion in a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, with the drift $\boldsymbol{b}(\cdot) := (b_1(\cdot), \cdots, b_N(\cdot))$ and the covariance matrix $\boldsymbol{\sigma}(\cdot) := (\sigma_{ij}(\cdot))_{1\leq i,j\leq N}$. To ensure the existence and uniqueness of the SDE, $\boldsymbol{b}(\cdot)$ and $\boldsymbol{\sigma}(\cdot)$ are assumed to satisfy the usual *global Lipschitz condition* and *linear growth condition*:

H1. There exists a constant $L_1 > 0$ and $L_2 > 0$ such that

$$\|\boldsymbol{b}(\boldsymbol{x}) - \boldsymbol{b}(\boldsymbol{y})\| + \|\boldsymbol{\sigma}(\boldsymbol{x}) - \boldsymbol{\sigma}(\boldsymbol{y})\| \leq L_1\|\boldsymbol{x} - \boldsymbol{y}\|,$$
$$\|\boldsymbol{b}(\boldsymbol{x})\| + \|\boldsymbol{\sigma}(\boldsymbol{x})\| \leq L_2\left(1 + \|\boldsymbol{x}\|\right),$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$.

Assumption **H1** ensures the existence of a *strong solution* to (4.1.1) and the solution is square-integrable [124, Theorem 2.9 in Chapter 5]. Here and throughout the rest of the paper, the infinitesimal generator $\mathcal{L}$ is

$$\mathcal{L} := \sum_i b_i(\boldsymbol{x})\frac{\partial}{\partial x^i} + \frac{1}{2}\sum_{i,j}(\boldsymbol{\sigma}(\boldsymbol{x})\boldsymbol{\sigma}(\boldsymbol{x})^T)_{i,j}\frac{\partial^2}{\partial x^i \partial x^j}, \tag{3.2.2}$$

where $\boldsymbol{\sigma}(\boldsymbol{x})\boldsymbol{\sigma}(\boldsymbol{x})^T$ is assumed to be positive-definite for every $\boldsymbol{x} \in \mathbb{R}^N$. See [124, Chapter 5], or [111, Chapter IV] for background on SDEs.

If a control is applied to X_t^i, then X_t^i evolves as

$$dX_t^i = b_i(\boldsymbol{X}_{t-})dt + \boldsymbol{\sigma}_i(\boldsymbol{X}_{t-})d\boldsymbol{B}(t) + d\xi_t^{i+} - d\xi_t^{i-}, \quad X_{0-}^i = x^i, \tag{3.2.3}$$

where $\boldsymbol{\sigma}_i$ is the i^{th} row of the covariance matrix $\boldsymbol{\sigma}$. Here the control (ξ^{i+}, ξ^{i-}) is a pair of non-decreasing and càdlàg processes, and of finite variation. In other words, (ξ^{i+}, ξ^{i-}) is the minimum decomposition of the finite variation process ξ^i such that $\xi^i := \xi^{i+} - \xi^{i-}$.

Game objective. The game is for player i to minimize, for all (ξ^{i+}, ξ^{i-}) in an appropriate admissible control set, over an infinite time horizon, the following objective function,

$$\mathbb{E}\int_0^\infty e^{-\alpha t} h^i(X_t^1, \cdots, X_t^N)dt. \tag{3.2.4}$$

Here $\alpha > 0$ is a constant discount factor. In this game, players interact through their respective objective functions $h^i(\boldsymbol{x}) : \mathbb{R}^N \to \mathbb{R}^+$, which are assumed to be

H2. twice differentiable, with $k \leq ||\nabla^2 h^i(\boldsymbol{x})|| \leq K$ for some $K > k > 0$.

For example, $h^i(\boldsymbol{x}) = h(x^i - \frac{\sum_{j=1}^N x^j}{N})$ is a distance function between the position of player i and the center of all players.

Note that in the objective function (3.2.4), there is no cost of control. With this formulation, the explicit solution structure of the NE for game (3.2.4) is clean. It is entirely possible to consider an N-player game with additional cost of control. For instance, one might study the game formulation of [122] with a proportional cost of control. We conjecture that the solution structure would be similar although the analysis will be more involved. This will be an interesting problem for future analysis.

Admissible control policies. The admissible control set $\mathcal{S}_N(\boldsymbol{x}, \boldsymbol{y})$ for this N-player game is given by

$$\mathcal{S}_N(\boldsymbol{x}, \boldsymbol{y}) := \Bigg\{ \boldsymbol{\xi} : \xi^i \in \mathcal{U}_N^i \text{ for } 1 \leq i \leq N,\ \sum_{i=1}^N \int_0^\infty \frac{a_{ij} Y_{t-}^j}{\sum_{k=1}^M a_{ik} Y_{t-}^k} d\check{\xi}_t^i \leq y^j, 1 \leq j \leq M, \\ \mathbb{P}\left(\Delta\xi_t^i\left(\boldsymbol{X}_{t-}, \boldsymbol{Y}_{t-}\right)\Delta\xi_t^k\left(\boldsymbol{X}_{t-}, \boldsymbol{Y}_{t-}\right) \neq 0\right) = 0 \text{ for all } t \geq 0 \text{ and } i \neq k \Bigg\}, \tag{3.2.5}$$

where

$$\mathcal{U}_N^i := \Big\{ (\xi^+, \xi^-) : \xi^+ \text{ and } \xi^- \text{ are } \mathcal{F}^{\boldsymbol{X_{t-}}, \boldsymbol{Y_{t-}}}\text{-progressively measurable,} \\ \text{càdlàg, non-decreasing, with } \xi_{0-}^+ = \xi_{0-}^- = 0 \Big\},$$

with $\mathcal{F}^{\boldsymbol{X_{t-}},\boldsymbol{Y_{t-}}} := \sigma\left(\cup_{s<t}\, \mathcal{F}^{\boldsymbol{X_s},\boldsymbol{Y_s}}\right)$ the filtrations of $(\boldsymbol{X},\boldsymbol{Y})$ up to time $t-$, and

$$Y_t^j = y^j - \sum_{i=1}^{N} \int_0^t \frac{a_{ij} Y_{s-}^j}{\sum_{k=1}^M a_{ik} Y_{s-}^k} d\check{\xi}_s^i \in \mathbb{R}_+ \quad \text{and} \quad Y_{0-}^j = y^j, \tag{3.2.6}$$

with $a_{ij} = 0$ or 1 for $1 \le i \le N$ and $1 \le j \le M$, $\sum_{j=1}^M a_{ij} > 0$ for all $i = 1, \cdots, N$, and $\sum_{i=1}^N a_{ij} > 0$ for all $j = 1, \cdots, M$. Moreover,

$$\check{\xi}_t^i := \xi_t^{i+} + \xi_t^{i-}, \tag{3.2.7}$$

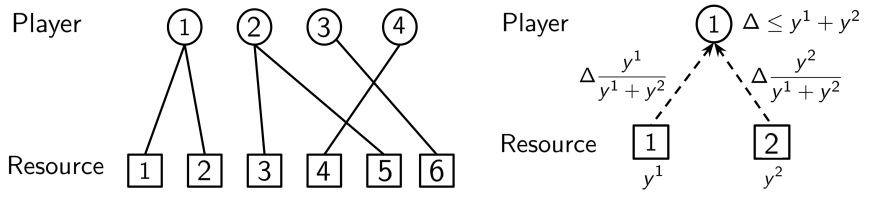
is the accumulative amount of controls/resources consumed by player i up to time t.

The non-decreasing and càdlàg processes $(\xi^{i+}, \xi^{i-}) \in \mathcal{U}_N^i$ can be decomposed in the differential form,

$$d\xi_t^{i\pm} = d(\xi_t^{i\pm})^c + \Delta\xi_t^{i\pm}, \tag{3.2.8}$$

where $d(\xi_t^{i\pm})^c$ is the continuous part and $\Delta\xi_t^{i\pm} := \xi_t^{i\pm} - \xi_{t-}^{i\pm}$ is the jump part of $d\xi_t^{i\pm}$.

Here is the intuition for $\mathcal{S}_N(\boldsymbol{x},\boldsymbol{y})$. In this game, each player i will make a decision based on the current positions of all players and the available resources. In addition to this adaptedness constraint, the admissible control set $\mathcal{S}_N(\boldsymbol{x},\boldsymbol{y})$ specifies the resource allocation policy for each player. For M different types of resources, define $\boldsymbol{A} := (a_{ij})_{i,j} \in \mathbb{R}^{N\times M}$ to be the *adjacent matrix* with $a_{ij} = 0$ or 1. Then $\boldsymbol{A}$ describes the relationship between the players and the types of available resources, with $a_{ij} = 1$ meaning that resource of type j is available to player i, and $a_{ij} = 0$ meaning that resource of type j is inaccessible to player i. The condition $\sum_{j=1}^M a_{ij} > 0$ for all $i = 1, \cdots, N$ implies that each player i has access to at least one resource, and the condition $\sum_{i=1}^N a_{ij} > 0$ for all $j = 1, \cdots, M$ indicates that each resource j is available to at least one player. Moreover, when player i would like to exercise control, she will consume resources proportionally to all the resources available to her. She will stop consuming once all the available resources hit level zero. This results in the form of the integrand in the expression of (3.2.6). Note that the denominator is always no smaller than the numerator hence the integrand is well-defined with the convention $\frac{0}{0} = 0$. See Figure 3.1 for illustration.



(a) Relationship. (b) Resource allocation policy.

Figure 3.1: Example of adjacent matrix $\boldsymbol{A}$, relationship between the players and resources when $N = 4$ and $M = 6$.

Take an example of $N = 4$, $M = 6$, with the matrix $\boldsymbol{A}$ defined as

$$\boldsymbol{A} = \begin{bmatrix} 1,1,0,0,0,0 \\ 0,0,1,0,1,0 \\ 0,0,0,0,0,1 \\ 0,0,0,1,0,0 \end{bmatrix},$$

(Figure 3.1a). The resource allocation policy is illustrated in Figure 3.1b, with the amount of available resource y^1 and y^2 of type one and two respectively. When player one wishes to apply controls of amount Δ, say $\Delta \leq y^1 + y^2$, she will consume resources randomly from type one and two. So player one will take $\Delta \frac{y^1}{y^1+y^2}$ from resource one and $\Delta \frac{y^2}{y^1+y^2}$ from resource two. Finally, the condition $\mathbb{P}(\Delta\xi_t^i \Delta\xi_t^k \neq 0) = 0$ for all $t \geq 0$ and $i \neq k$ excludes the possibility of simultaneous jumps of any two out of N players, which facilitates designing feasible control policies when controls involve jumps. This condition is not a restriction, and instead should be interpreted as a *regularization*. See also [13, 94, 129]. Indeed, when there are multiple players who would like to jump at the same time, one can simply design a proper *order*, for instance by indexing the players and their jump orders, so that they will move *sequentially*.

Game formulation. Let $\boldsymbol{\xi} := (\xi^1, \cdots, \xi^N)$ be the controls from the players. Let $\boldsymbol{x} := (x^1, \cdots, x^N)$ and $\boldsymbol{y} := (y^1, \cdots, y^M)$. Then the stochastic game is for each player i to minimize

$$J^i(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\xi}) := \mathbb{E} \int_0^\infty e^{-\alpha t} h^i(\boldsymbol{X}_t) dt, \tag{3.2.9}$$

subject to the dynamics in (3.2.3) and (3.2.6) with the constraint in (3.2.5). There are two special games of particular interest. One is a game where all players pool their resources such that

$$\sum_{i=1}^N \int_0^\infty d\check{\xi}_s^i < y < \infty. \tag{3.2.10}$$

When $N = 1$, this is a single player game corresponding to the finite fuel control problem which is well studied in [19, 122]. We call this game a pooling game $\boldsymbol{C_p}$. Clearly in terms of the adjacent matrix $\boldsymbol{A}$, this corresponds to $M = 1$, and $\boldsymbol{A} = [1, 1, \cdots, 1]^T \in \mathbb{R}^{N \times 1}$. Another is a game where players divide the resource up front such that

$$\int_0^\infty d\check{\xi}_s^i < y_i, \tag{3.2.11}$$

where y_i is the total amount of controls that player i can exercise. This game is called $\boldsymbol{C_d}$, with $M = N$, and $\boldsymbol{A} = \boldsymbol{I_N}$. Finally, we refer the game with a general matrix $\boldsymbol{A}$ as game $\boldsymbol{C}$.

3.3 NE Game Solution: Verification Theorem and Skorokhod Problem

Verification Theorem

We will analyze the N-player game under the criterion of Markovian NE. See [44] for various concepts of NE of differential games. Recall the definition of a Markovian NE of N-player games.

Definition 20. *A tuple of admissible controls $\boldsymbol{\xi}^* := (\xi^{1*}, \cdots \xi^{N*})$ is a Markovian NE of the N-player game (3.2.9), if for each ξ^i such that $(\boldsymbol{\xi}^{-i*}, \xi^i) \in \mathcal{S}_N(\boldsymbol{x}, \boldsymbol{y})$,*

$$J^i(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\xi}^*) \leq J^i\left(\boldsymbol{x}, \boldsymbol{y}; \left(\boldsymbol{\xi}^{-i*}, \xi^i\right)\right),$$

where $\boldsymbol{\xi}^{-i*} = (\xi^{1*}, \cdots, \xi^{i-1*}, \xi^{i+1*}, \cdots, \xi^{N*})$ and $(\boldsymbol{\xi}^{-i*}, \xi^i) = (\xi^{1*}, \cdots, \xi^{i-1*}, \xi^i, \xi^{i+1*}, \cdots, \xi^{N*})$. Here the strategies ξ^{i*} and ξ^i are functions of time t, $\boldsymbol{X}_t = (X_t^1, \cdots, X_t^N)$, and $\boldsymbol{Y}_t = (Y_t^1, \cdots, Y_t^M)$, with $\boldsymbol{X}_{0-} = \boldsymbol{x}$ and $\boldsymbol{Y}_{0-} = \boldsymbol{y}$. Controls that give Markovian NEs are called the Markovian Nash Equilibrium Points (MNEPs). The associated value function $J^i(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\xi}^*)$ $(i = 1, 2, \cdots, N)$ is called the game value.

We first derive heuristically the associated HJB equations for the game (3.2.9). To this end, we start with some notations of region partitions for each player.

Definition 21 (Action and waiting regions)**.** *The i^{th} player's action region is*

$$\mathcal{A}_i := \{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^N \times \mathbb{R}_+^M : d\xi^i(\boldsymbol{x}, \boldsymbol{y}) \neq 0\},$$

and its waiting region is $\mathcal{W}_i := (\mathbb{R}^N \times \mathbb{R}_+^M) \setminus \mathcal{A}_i$. Let $\mathcal{A}^{-i} := \cup_{j \neq i} \mathcal{A}_j$, and $\mathcal{W}_{-i} := \cap_{j \neq i} \mathcal{W}_j$.

Now the HJB is heuristically derived as follows. When $\mathcal{A}_j \cap \mathcal{A}_i = \emptyset$ for all $i \neq j$ and $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{W}_{-i}$, $\Delta\xi^{j*} = 0$ for $j \neq i$. Thus the game for player i becomes a classical control problem with three choices: $\Delta\xi^{i*} = 0$, $\Delta\xi^{i*,+} > 0$, and $\Delta\xi^{i*,-} > 0$. The case $\Delta\xi^{i*} = 0$ implies, by simple stochastic calculus, $-\alpha v^i + h^i(\boldsymbol{x}) + \mathcal{L}v^i = 0$, the case $\Delta\xi^{i*,+} > 0$ corresponds to $-\sum_{j=1}^M \frac{a_{ij}y^j}{\sum_{k=1}^M a_{ik}y^k} v^i_{y^j} + v^i_{x^i} = 0$, and the case $\Delta\xi^{i*,-} > 0$ corresponds to $-\sum_{j=1}^M \frac{a_{ij}y^j}{\sum_{k=1}^M a_{ik}y^k} v^i_{y^j} - v^i_{x^i} = 0$. [1] One of the three choices will be optimal. In short, we have for $(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{W}_{-i}$,

$$\min\left\{-\alpha v^i + h^i(\boldsymbol{x}) + \mathcal{L}v^i, -\sum_{j=1}^M \frac{a_{ij}y^j}{\sum_{k=1}^M a_{ik}y^k} v^i_{y^j} + v^i_{x^i}, -\sum_{j=1}^M \frac{a_{ij}y^j}{\sum_{k=1}^M a_{ik}y^k} v^i_{y^j} - v^i_{x^i}\right\} = 0, \tag{3.3.1}$$

Since each player i can only control x^i and the resources that are available to her, the above equation is minimizing over $(x^i, \boldsymbol{y})$.

[1] We adopt the convention $\frac{0}{0} = 0$.

When $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{A}_j$, player j will control. Denote the amount of control by player j as $(\Delta\xi^{j*,+}, \Delta\xi^{j*,-})$. When $\mathcal{A}_j \cap \mathcal{A}_i = \emptyset$ for all $i \neq j$, we should have,

$$v^i(\boldsymbol{x},\boldsymbol{y}) = v^i\left(\boldsymbol{x}^{-j}, x^j + \Delta\xi^{j*,+} - \Delta\xi^{j*,-}, \boldsymbol{y} - \left(\frac{a_{j1}y^1}{\sum_{k=1}^M a_{jk}y^k}, \cdots, \frac{a_{jN}y^N}{\sum_{k=1}^M a_{jk}y^k}\right)(\Delta\xi^{j*,+} + \Delta\xi^{j*,-})\right).$$

This leads to

$$\min\left\{-\sum_{k=1}^M \frac{a_{jk}y^k}{\sum_{s=1}^M a_{js}y^s} v^i_{y^k} + v^i_{x^j}, -\sum_{k=1}^M \frac{a_{jk}y^k}{\sum_{s=1}^M a_{js}y^s} v^i_{y^k} - v^i_{x^j}\right\} = 0. \tag{3.3.2}$$

By letting $\Delta\xi^{j*,\pm} \to 0$, (3.3.1) describe the behavior in $\overline{\mathcal{W}}_i$ and near boundary $\partial\mathcal{W}_i$. Moreover, we can show that (3.3.1) is consistent with the jump behaviors in $\mathcal{A}_i$: $-\sum_{j=1}^M \frac{a_{ij}y^j}{\sum_{k=1}^M a_{ik}y^k} v^i_{y^j} \pm v^i_{x^i} = 0$ has a linear solution $v^i(\boldsymbol{x}) = a\left(\pm x_i + \sum_{j=1}^M a_{ij}y^j\right) + b$ for some $a, b \in \mathbb{R}$. And it is easy to check that $\forall \sum_{k=1}^M a_{ik}y^k \geq \Delta > 0$,

$$\frac{a_{ij}y^j - \frac{a_{ij}y^j}{\sum_{k=1}^M a_{ik}y^k}\Delta}{\sum_{k=1}^M a_{ik}y^k - \Delta} = \frac{a_{ij}y^j}{\sum_{k=1}^M a_{ik}y^k},$$

which means the allocation policy (jump direction) outside the waiting region is linear. Hence the dynamics in (3.2.6) satisfies the HJB equation in $\mathcal{A}_i$. The consistency property also holds for (3.3.2).

Remark 21.1. *Note that when $N = 2$, the above equation corresponds to the continuity condition of game values. For general N-player games, it is a mathematical description of interactions between the player in control and those who are not. It guarantees that all players control optimally so that they sequentially push the underlying dynamics until reaching the common waiting region. This is consistent with the intuition that NE is conditionally optimal for each player.*

Remark 21.2. *Under the 'no simultaneous jump' assumption in* (3.2.5)*, there are only two gradient terms in* (3.3.1) *corresponding to the actions from player i. If one removes this 'no simultaneous jump' assumption, there will be $3^N - 1$ terms for gradient constraints, making the problem intractable. Similar analysis holds for* (3.3.2)*.*

Next we present a verification theorem which gives sufficient conditions of an MNEP.

Theorem 22 (Verification theorem)**.** *Assume **H1**-**H2**. Further assume that $\mathcal{A}_j \cap \mathcal{A}_i = \emptyset$ for all $i \neq j$. For each $i = 1, \cdots, N$, suppose that the i^{th} player's strategy $\xi^{i*} \in \mathcal{U}_N^i$ satisfies the following conditions*

(i) $\boldsymbol{\xi}^* := (\xi^{1*}, \cdots, \xi^{N*}) \in \mathcal{S}_N(\boldsymbol{x},\boldsymbol{y})$,

(ii) $v^i(\cdot) = J^i(\cdot;\boldsymbol{\xi}^*)$ *satisfies the HJB equation (3.3.1) for* $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{W}_{-i}$,

(iii) $v^i(\boldsymbol{x},\boldsymbol{y})$ *satisfies the transversality condition*

$$\limsup_{T\to\infty} e^{-\alpha T}\mathbb{E}v^i\left(\boldsymbol{X}_T,\boldsymbol{Y}_T\right) = 0, \tag{3.3.3}$$

for any $(\boldsymbol{X}_t,\boldsymbol{Y}_t)$ *under admissible controls.*

(iv) $v^i(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{C}^2(\overline{\mathcal{W}_{-i}})$ *and* v^i *is convex for all* $(\boldsymbol{x},\boldsymbol{y}) \in \overline{\mathcal{W}_{-i}}$,

(v) $v^i_{x^j}$ *is bounded in* $\overline{\mathcal{W}_{-i}}$ *for each* $j = 1, 2, \cdots, N$,

(vi) for any $\xi^i \in \mathcal{U}_N^i$ *such that* $(\boldsymbol{\xi}^{-i*},\xi^i) \in \mathcal{S}_N(\boldsymbol{x},\boldsymbol{y})$,

$$\mathbb{P}((\boldsymbol{X}_t^{-i*}, X_t^i, \boldsymbol{Y}_t) \in \overline{\mathcal{W}_{-i}}) = 1 \quad \textit{for all } t \geq 0,$$

where $(\boldsymbol{X}_t^{-i*}, X_t^i, \boldsymbol{Y}_t)$ *is under* $(\boldsymbol{\xi}^{-i*},\xi^i)$.

(vii) $v^i(\cdot)$ *satisfies the equation* (3.3.2) *when* $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{A}_j$,

then $\boldsymbol{\xi}^*$ *is an MNEP with value function* v^i.

Proof of Theorem 22. It suffices to prove that for each $i = 1, \cdots, N$,

$$J^i(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\xi}^*) \leq J^i(\boldsymbol{x},\boldsymbol{y};(\boldsymbol{\xi}^{-i*},\xi^i)),$$

for all $(\boldsymbol{\xi}^{-i*},\xi^i) \in \mathcal{S}_N(\boldsymbol{x},\boldsymbol{y})$.

Recall (4.1.1) and (3.2.6). From condition (vi), under control $(\boldsymbol{\xi}^{-i*},\xi^i) \in \mathcal{S}_N(\boldsymbol{x},\boldsymbol{y})$, $(\boldsymbol{X}_t^{-i*}, X_t^i, \boldsymbol{Y}_t) \in \overline{\mathcal{W}_{-i}}$ a.s.. Applying Itô-Meyer's formula [148, Theorem 21] to $e^{-\alpha t}v^i(\boldsymbol{X}_t^{-i*}, X_t^i, \boldsymbol{Y}_t)$ yields

$$\begin{aligned}
&\mathbb{E}[e^{-\alpha T}v^i(\boldsymbol{X}_T^{-i*}, X_T^i, \boldsymbol{Y}_T)] - v^i(\boldsymbol{x},\boldsymbol{y}) \\
=\ & \mathbb{E}\int_0^T e^{-\alpha t}\left(\mathcal{L}v^i - \alpha v^i\right)dt + \mathbb{E}\int_0^T e^{-\alpha t}\sum_{j=1}^N v^i_{x^j}dB_t^j \\
&+ \sum_{j=1,j\neq i}^N \mathbb{E}\int_{[0,T)} e^{-\alpha t}(v^i_{x^j}d\xi_t^{j*,+} - v^i_{x^j}d\xi_t^{j*,-}) - \sum_{j=1,j\neq i}^N \mathbb{E}\int_{[0,T)} e^{-\alpha t}\sum_{k=1}^M \frac{a_{jk}Y_{t-}^k}{\sum_{s=1}^M a_{js}Y_{t-}^s}\left(v^i_{y^k}d\xi_t^{j*,+} + v^i_{y^k}d\xi_t^{j*,-}\right) \\
&+ \mathbb{E}\int_{[0,T)} e^{-\alpha t}(v^i_{x^i}d\xi_t^{i,+} - v^i_{x^i}d\xi_t^{i,-}) - \mathbb{E}\int_{[0,T)} e^{-\alpha t}\sum_{k=1}^M \frac{a_{ij}Y_{t-}^k}{\sum_{s=1}^M a_{is}Y_{t-}^s}\left(v^i_{y^k}d\xi_t^{i,+} + v^i_{y^k}d\xi_t^{i,-}\right) \\
&+ \mathbb{E}\sum_{0\leq t<T} e^{-\alpha t}\left(\Delta v^i - \sum_{j=1}^M v^i_{x^j}\Delta X_t^j - \sum_{k=1}^M v^i_{y^k}\Delta Y_t^k\right).
\end{aligned}$$

Note that condition (v) implies that $\int_0^T e^{-\alpha t}\sum_{j=1}^N v^i_{x^j}dB_t^j$ is a uniformly integrable martingale.

The convexity condition in (iv) implies $\mathbb{E}\sum_{0\le t<T} e^{-\alpha t}(\Delta v^i - v^i_{x^i}\Delta X^i_t - \sum_{j=1}^M v^i_{y^j}\Delta Y^j_t) \ge 0$. Given condition (ii), we see that $v^i(\boldsymbol{x})$ satisfies the HJB equation (3.3.1) on $\mathcal{A}_i$. Therefore

$$\begin{aligned}
&\mathbb{E}\int_{[0,T)} e^{-\alpha t}(v^i_{x^i}d\xi^{i,+}_t - v^i_{x^i}d\xi^{i,-}_t) - \mathbb{E}\int_{[0,T)} e^{-\alpha t}\sum_{k=1}^M \frac{a_{ij}Y^k_{t-}}{\sum_{s=1}^M a_{is}Y^s_{t-}}\left(v^i_{y^k}d\xi^{i,+}_t + v^i_{y^k}d\xi^{i,-}_t\right)\\
=\ &\mathbb{E}\int_{[0,T)} e^{-\alpha t}\left[v^i_{x^i} - \sum_{k=1}^M \frac{a_{ij}Y^k_{t-}}{\sum_{s=1}^M a_{is}Y^s_{t-}}v^i_{y^k}\right]d\xi^{i,+}_t + \mathbb{E}\int_{[0,T)} e^{-\alpha t}\left[-v^i_{x^i} - \sum_{k=1}^M \frac{a_{ij}Y^k_{t-}}{\sum_{s=1}^M a_{is}Y^s_{t-}}v^i_{y^k}\right]d\xi^{i,-}_t \ge 0.
\end{aligned}$$

For each $j \ne i$, almost surely, we have $d\xi^{j*}_t \ne 0$ only when $(\boldsymbol{X}_t, \boldsymbol{Y}_t) \in \partial\mathcal{W}_{-i} \cap \partial\mathcal{A}_j$. Along with the condition (vii),

$$\begin{aligned}
&\mathbb{E}\int_{[0,T)} e^{-\alpha t}(v^i_{x^j}d\xi^{j,+}_t - v^i_{x^j}d\xi^{j,-}_t) - \mathbb{E}\int_{[0,T)} e^{-\alpha t}\sum_{k=1}^M \frac{a_{jk}Y^k_{t-}}{\sum_{s=1}^M a_{js}Y^s_{t-}}\left(v^i_{y^k}d\xi^{j,+}_t + v^i_{y^k}d\xi^{j,-}_t\right)\\
=\ &\mathbb{E}\int_{[0,T)} e^{-\alpha t}\left[v^i_{x^j} - \sum_{k=1}^M \frac{a_{jk}Y^k_{t-}}{\sum_{s=1}^M a_{js}Y^s_{t-}}v^i_{y^k}\right]d\xi^{j*,+}_t + \left[-v^i_{x^j} - \sum_{k=1}^M \frac{a_{jk}Y^k_{t-}}{\sum_{s=1}^M a_{js}Y^s_{t-}}v^i_{y^k}\right]d\xi^{j*,-}_t = 0.
\end{aligned}$$

Condition (ii) also implies $\mathcal{L}v^i - \alpha v^i \ge h$. Combining all of the above,

$$e^{-\alpha T}\mathbb{E}v^i(\boldsymbol{X}^{-i*}_T, X^i_T, \boldsymbol{Y}_T) + \mathbb{E}\int_0^T e^{-\alpha t}h\left(\boldsymbol{X}^{-i*}_t, X^i_t\right)dt \ge v^i(\boldsymbol{x}, \boldsymbol{y}). \tag{3.3.4}$$

By letting $T \to \infty$, the inequality (3.3.4) and condition (iii) lead to the desirable inequality. Along with condition (vii), the equality holds with value $v^i(\boldsymbol{x}, \boldsymbol{y})$. □

Remark 22.1. *Note that, unlike the usual stochastic control problem which requires $\mathcal{C}^2$ regularity in the whole space $\mathbb{R}^N$, in the N-player game (3.2.9), the minimum regularity needed is $\mathcal{C}^2$ in $\mathcal{W}_{-i}$. This is due to the game nature and interactions among players.*

Suppose the game value v^i $(i = 1, 2, \cdots, N)$ that satisfies the verification theorem (Theorem 22) are given, the next step is to construct the corresponding NE strategies. This is by solving a Skorokhod problem, introduced in the next subsection.

Skorokhod Problem

Let $G = \cap_{i\in\mathcal{I}} G_i$ be a nonempty domain in $\mathbb{R}^{n+m}$, where $\mathcal{I}$ is a nonempty finite index set and for each $i \in \mathcal{I}$, G_i is a nonempty domain in $\mathbb{R}^{n+m}$. For simplicity, we assume that $\mathcal{I} = \{1, 2, \cdots, I\}$, with $|\mathcal{I}| = I$. For each $i \in \mathcal{I}$, let $\boldsymbol{n}_i : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ be the unit normal vector field on ∂G_i that points into G_i. And denote $\boldsymbol{r}_i(\cdot) : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ as the reflection direction on ∂G_i. Fix $\boldsymbol{b} \in \mathbb{R}^n$ and $\boldsymbol{\sigma} \in \mathbb{R}^{n\times n}$ as the drift and covariance of the diffusion process without reflection. Let ν denote a probability measure on $(\overline{G}, \mathcal{B}(\overline{G}))$, where $\mathcal{B}(\overline{G})$ is the Borel σ-algebra on $\overline{G}$.

A Skorokhod problem is to find a reflected diffusion process in $\overline{G}$ such that the initial distribution follows ν, the diffusion parameters are $(\boldsymbol{b}, \boldsymbol{\sigma})$, and the reflection direction is $\boldsymbol{r}_i$ on face ∂G_i. For each reflection direction $\boldsymbol{r}_i$ $(i \in \mathcal{I})$, denote $\boldsymbol{r}_i^+ := (r_{i,1}, \cdots, r_{i,n})$ as the vector of the first n components of $\boldsymbol{r}_i$ and denote $\boldsymbol{r}_i^- := (r_{i,n+1}, \cdots, r_{i,n+m})$ as the vector of the next m components of $\boldsymbol{r}_i$. Note that $r_{i,k}^- = r_{i,k+n}$ by the usual index rule $(k = 1, \cdots, m)$.

Definition 23 (Constrained semimartingale reflecting Brownian motion)**.** *A constrained semimartingale reflecting Brownian motion (SRBM) associated with the data* $(G, \boldsymbol{b}, \boldsymbol{\sigma}, \{\boldsymbol{r}_i\}_{i=1}^I, \nu)$ *is an* $\{\mathcal{F}_t\}$*-adapted, n-dimensional process* $\boldsymbol{X}$ *defined on some filtered probability space* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ *such that:*

(i) $\mathbb{P}$*-a.s.,* $\boldsymbol{X}_t = \boldsymbol{W}_t + \sum_{i\in\mathcal{I}} \int_{[0,t)} \boldsymbol{r}_i^+(\boldsymbol{X}_s, \boldsymbol{Y}_s) d\eta_s^i$ *for all* $t \geq 0$*,*

(ii) *under* $\mathbb{P}$*,* $\boldsymbol{W}_t$ *is an n-dimensional* $\mathcal{F}_t$*-Brownian motion with drift vector* $\boldsymbol{b}$*, covariance matrix* $\boldsymbol{\sigma}$ *and initial distribution* ν*,*

(iii) $dY_t^j = \sum_{i\in\mathcal{I}} \int_{[0,t)} \boldsymbol{r}_{i,j}^-(\boldsymbol{X}_t, \boldsymbol{Y}_t) d\eta_t^i$ *and* $Y_t^j \geq 0$ *for* $j = 1, 2, \cdots, m$*,*

(iv) *for each* $i \in \mathcal{I}$*,* η^i *is a one-dimensional process such that* $\mathbb{P}$*-a.s.,*

(a) $\eta_0^i = 0$*,*

(b) η^i *is continuous and nondecreasing,*

(c) $\eta_t^i = \int_{(0,t]} 1_{\{W_s \in \partial G_i \cap \partial G\}} d\eta_s^i$ *for all* $t \geq 0$*,*

(v) $\mathbb{P}$*-a.s.,* $(\boldsymbol{X}_t, \boldsymbol{Y}_t)$ *has continuous paths and* $(\boldsymbol{X}_t, \boldsymbol{Y}_t) \in \overline{G}$ *for all* $t \geq 0$*,*

Remark 23.1. *Specific to the stochastic game in this paper,* $\boldsymbol{X}_t$ *is the controlled diffusion process and* $\boldsymbol{Y}_t$ *is the resource levels. The domain G restricts the dynamics of both* $\boldsymbol{X}_t$ *and* $\boldsymbol{Y}_t$*. Note that the constrained SRBM is slightly different from the standard SRBM (see Kang and Williams [116]) in the sense that the reflection domain depends on both the diffusion process* $\boldsymbol{X}_t$ *and the resource process* $\boldsymbol{Y}_t$*.*

For each $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{n+m}$, let $\mathcal{I}(\boldsymbol{x}, \boldsymbol{y}) = \{i \in \mathcal{I} : (\boldsymbol{x}, \boldsymbol{y}) \in \partial G_i\}$. Let $U_\epsilon(S)$ denote the closed set $\{(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{n+m} : dist((\boldsymbol{x}, \boldsymbol{y}), S) \leq \epsilon\}$ for any $\epsilon > 0$ and $S \subset \mathbb{R}^{n+m}$. If $S = \emptyset$, set $U_\epsilon(S) = \emptyset$ for any $\epsilon > 0$. We propose the following assumptions on domain G and reflection directions $\{\boldsymbol{r}_i, i \in \mathcal{I}\}$:

A1. G is the nonempty domain in $\mathbb{R}^{n+m}$ such that

$$G = \cap_{i\in\mathcal{I}} G_i, \tag{3.3.5}$$

where for each $i \in \mathcal{I}$, G_i is a nonempty domain in $\mathbb{R}^{n+m}$, $G_i \neq \mathbb{R}^{m+n}$ and the boundary ∂G_i is $\mathcal{C}^1$.

A2. For each $\epsilon \in (0,1)$ there exists $R(\epsilon) > 0$ such that for each $i \in \mathcal{I}$, $(\boldsymbol{x},\boldsymbol{y}) \in \partial G_i \cap \partial G$ and $(\boldsymbol{x}',\boldsymbol{y}') \in \overline{G}$ satisfying $\|(\boldsymbol{x},\boldsymbol{y}) - (\boldsymbol{x}',\boldsymbol{y}')\| < R(\epsilon)$, we have

$$\langle \boldsymbol{n}_i(\boldsymbol{x},\boldsymbol{y}), (\boldsymbol{x}',\boldsymbol{y}') - (\boldsymbol{x},\boldsymbol{y})\rangle \geq -\epsilon \|(\boldsymbol{x},\boldsymbol{y}) - (\boldsymbol{x}',\boldsymbol{y}')\|.$$

A3. The function $D : [0,\infty) \to [0,\infty]$ is such that $D(0) = 0$ and

$$D(\epsilon) = \sup_{\mathcal{I}_0 \in \mathcal{I}, \mathcal{I}_0 \neq \emptyset} \sup \left\{ \operatorname{dist}\left((\boldsymbol{x},\boldsymbol{y}), \cap_{i\in\mathcal{I}_0}(\partial G_i \cap \partial G)\right) : (\boldsymbol{x},\boldsymbol{y}) \in \cap_{i\in\mathcal{I}_0} U_\epsilon(\partial G_i \cap \partial G) \right\},$$

for $\epsilon > 0$ satisfies $D(\epsilon) \to 0$ as $\epsilon \to 0$.

A4. There is a constant $L > 0$ such that for each $i \in \mathcal{I}$, $\boldsymbol{r}_i(\cdot)$ is a uniformly Lipschitz continuous function from $\mathbb{R}^{n+m}$ into $\mathbb{R}^{n+m}$ with Lipschitz constant L and $\|\boldsymbol{r}_i(\boldsymbol{x},\boldsymbol{y})\| = 1$ for each $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^{n+m}$.

A5. There is a constant $a \in (0,1)$, and vector valued function $\boldsymbol{c}(\cdot) = (c_1(\cdot), \cdots, c_I(\cdot))$ and $\boldsymbol{d}(\cdot) = (d_1(\cdot), \cdots, d_I(\cdot))$ from ∂G into $\mathbb{R}^I_+$ such that for each $(\boldsymbol{x},\boldsymbol{y}) \in \partial G$,

(i) $\sum_{i\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})} c_i(\boldsymbol{x},\boldsymbol{y}) = 1$,

$$\min_{k\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})} \left\langle \sum_{i\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})} c_i(\boldsymbol{x},\boldsymbol{y})\boldsymbol{n}_i(\boldsymbol{x},\boldsymbol{y}), \boldsymbol{r}_k(\boldsymbol{x},\boldsymbol{y}) \right\rangle \geq a,$$

(ii) $\sum_{i\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})} d_i(\boldsymbol{x},\boldsymbol{y}) = 1$,

$$\min_{k\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})} \left\langle \sum_{i\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})} d_i(\boldsymbol{x},\boldsymbol{y})\boldsymbol{r}_i(\boldsymbol{x},\boldsymbol{y}), \boldsymbol{n}_k(\boldsymbol{x},\boldsymbol{y}) \right\rangle \geq a.$$

Theorem 24. *Given assumptions **A1**-**A5**. Then there exists a constrained SRBM associated with the data* $(G, \boldsymbol{b}, \boldsymbol{\sigma}, \{\boldsymbol{r}_i, i \in \mathcal{I}\}, \nu)$.

The proof of Theorem 24 is adapted from Kang and Williams [116, Theorem 5.1] and combined with [116, Theorem 4.3]. More precisely, we construct a sequence of approximation (random walks) to the constrained SRBM and use the *invariance principle* to establish the weak convergence. The main difference is that the constrained SRMB problem in this paper depends not only on the diffusion process $\boldsymbol{X}_t$ but also on a degenerate process $\boldsymbol{Y}_t$ indicating the remaining resource levels. The detailed proof of Theorem 24 is provided in Appendix B.1.

Now, denote $(\boldsymbol{X}^*_t, \boldsymbol{Y}^*_t)$ as a solution to the Skorokhod problem $(G, \boldsymbol{b}, \boldsymbol{\sigma}, \{\boldsymbol{r}_i, i \in \mathcal{I}\}, \nu)$. If the initial position is in the interior of G, it is not hard to show that $\mathbb{P}((\boldsymbol{X}^*_t, \boldsymbol{Y}^*_t) \in \partial G_i \cap \partial G_j \text{ for } i \neq j, t \geq 0) = 0$. The proof follows the same line as in Williams [184].

In the next three sections, we solve explicitly the game $\boldsymbol{C}$, based on sufficient conditions in the above verification theorem. We will first analyze games $\boldsymbol{C_p}$ and $\boldsymbol{C_d}$ to gain insight into the solution structure. For general $\boldsymbol{b}$ and $\boldsymbol{\sigma}$, explicit solution is almost impossible. Therefore we consider the following $\boldsymbol{b}$, $\boldsymbol{\sigma}$, with a general $\boldsymbol{h}$ for the rest of this paper. That is, we assume

H1′. $b_i = 0, \qquad i = 1, 2, \cdots, N, \qquad \text{and} \qquad \boldsymbol{\sigma} = \boldsymbol{I}_N.$

Moreover, we assume that $h^i(\boldsymbol{x}) := h\left(x^i - \frac{\sum_{j=1}^N x^j}{N}\right)$, such that

H2′. h is convex, symmetric, $h(0) \geq 0$, h'' is non-increasing and $k \leq h'' \leq K$ for some $0 < k < K$.

3.4 Nash Equilibrium for Game C_p

This section analyzes the Markovian NE of game $\boldsymbol{C_p}$. Section 3.4 derives the solution to the HJB equations. Section 3.4 constructs the controlled process from the HJB solution. Section 3.4 derives the NE for the game $\boldsymbol{C_p}$ and specifies the NE for the two-player game with $h(x) = x^2$. Recall that in game $\boldsymbol{C_p}$, $\boldsymbol{A} = [1, 1, \cdots, 1]^T \in \mathbb{R}^{N\times 1}$, and

$$Y_t = y - \sum_{i=1}^N \check{\xi}_t^i \quad \text{and} \quad Y_{0-} = y. \tag{3.4.1}$$

Solving HJB equations

Define

$$\widetilde{x}^i := x^i - \frac{\sum_{j\neq i} x^j}{N-1} \quad \text{for } 1 \leq i \leq N, \tag{3.4.2}$$

to be the relative position from x^i to the center of $(x^j)_{j\neq i}$. For game $\boldsymbol{C_p}$, if $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$, the HJB system simplifies to

$$\textit{(HJB-C}_p\textit{)} \begin{cases} \min\left\{-\alpha v^i + h\left(\frac{N-1}{N}\widetilde{x}^i\right) + \frac{1}{2}\sum_{j=1}^N v^i_{x^jx^j}, -v^i_y + v^i_{x^i}, -v^i_y - v^i_{x^i}\right\} = 0, \\ \qquad\qquad \text{for } (\boldsymbol{x}, y) \in \mathcal{W}_{-i}, \\ \min\left\{-v^i_y + v^i_{x^j}, -v^i_y - v^i_{x^j}\right\} = 0, \qquad \text{for } (\boldsymbol{x}, y) \in \mathcal{A}_j, j \neq i. \end{cases}$$

Now we look for a threshold function $f_N : \mathbb{R} \to \mathbb{R}$ with $f_N(-x) = f_N(x)$ such that the action region $\mathcal{A}_i$ and the waiting region $\mathcal{W}_i$ of the i^{th} player are defined by

$$\mathcal{A}_i := (E_i^+ \cup E_i^-) \cap Q_i \quad \text{and} \quad \mathcal{W}_i := (\mathbb{R}^N \times \mathbb{R}_+) \setminus \mathcal{A}_i, \tag{3.4.3}$$

where

$$E_i^+ := \left\{(\boldsymbol{x}, y) \in \mathbb{R}^N \times \mathbb{R}_+ : \widetilde{x}^i \geq f_N^{-1}(y)\right\} \text{ and } E_i^- := \left\{(\boldsymbol{x}, y) \in \mathbb{R}^N \times \mathbb{R}_+ : \widetilde{x}^i \leq -f_N^{-1}(y)\right\}, \tag{3.4.4}$$

and

$$Q_i := \{(\boldsymbol{x}, y) \in \mathbb{R}^N \times \mathbb{R}_+ : |\widetilde{x}^i| \geq |\widetilde{x}^k| \text{ for } k < i, \ |\widetilde{x}^i| > |\widetilde{x}^k| \text{ for } k > i\}.$$

Note here the partition $\{Q_i\}_{1\le i\le N}$ is introduced to avoid simultaneous jumps by multiple players so that $\mathcal{A}_i\cap\mathcal{A}_j=\emptyset$. The key idea of designing the partition is that if several players are in $E_i^+\cup E_i^-$, the player who is the farthest away from the center controls. If ties occur, the player with the largest index controls. It is easy to see that $\mathcal{W}_i\neq\emptyset$ for $1\le i\le N$, and $\mathcal{A}_i\cap\mathcal{A}_j=\emptyset$ for $i\neq j$.

We seek a solution $v^i(\boldsymbol{x},y)\in\mathcal{C}^2(\overline{\mathcal{W}_{-i}})$ such that if $|\widetilde{x}^i|<f_N^{-1}(y)$, it is of the form,

$$v^i(\boldsymbol{x},y)=p_N(\widetilde{x}^i)+A_N(y)\cosh\left(\widetilde{x}^i\sqrt{\frac{2(N-1)\alpha}{N}}\right), \tag{3.4.5}$$

where

$$p_N(x):=\mathbb{E}\int_0^\infty e^{-\alpha t}h\left(\frac{N-1}{N}x+\sqrt{\frac{N-1}{N}}B_t\right)dt, \tag{3.4.6}$$

with B_t being a one-dimensional Brownian motion. Note that $p_N(\widetilde{x}^i)$ is a solution to $-\alpha v^i+h(\frac{N-1}{N}\widetilde{x}^i)+\frac{1}{2}\sum_{j=1}^N v^i_{x^jx^j}=0$, which corresponds to the waiting region, and $\cosh(\sqrt{\frac{2(N-1)\alpha}{N}}\widetilde{x}^i)$ is a solution to $-\alpha v^i+\frac{1}{2}\sum_{j=1}^N v^i_{x^jx^j}=0$. If there is no resource, then $v^i(\boldsymbol{x},y)=p_N(\widetilde{x}^i)$, so $A_N(0)=0$. The *smooth-fit principle* states that, along the boundary $y=f_N(\tilde{x}^i)$ between the continuation set $\mathcal{W}$ and the action set $\mathcal{A}_i$, v^i has certain regularity properties across the hyperplane. Now applying the smooth-fit principle, we get $v^i_{x^ix^i}=v^i_{yy}=-v^i_{x^iy}$ at the boundary $y=f_N(\widetilde{x}^i)$ with $\widetilde{x}^i>0$. This follows from $v_{x^i}+v_y=0$ and we expect $v^i\in\mathcal{C}^2(\mathcal{W}_{-i})$.

$$\begin{cases} A_N'(f_N)=-p_N'\cosh\left(x\sqrt{\dfrac{2(N-1)\alpha}{N}}\right)+p_N''\sqrt{\dfrac{N}{2(N-1)\alpha}}\sinh\left(x\sqrt{\dfrac{2(N-1)\alpha}{N}}\right),\\ A_N(f_N)=p_N'\sqrt{\dfrac{N}{2(N-1)\alpha}}\sinh\left(x\sqrt{\dfrac{2(N-1)\alpha}{N}}\right)-p_N''\dfrac{N}{2(N-1)\alpha}\cosh\left(x\sqrt{\dfrac{2(N-1)\alpha}{N}}\right). \end{cases}$$

As a consequence,

$$f_N'(x)=\frac{p_N'-\frac{N}{2(N-1)\alpha}p_N'''}{p_N''\sqrt{\frac{N}{2(N-1)\alpha}}\tanh\left(x\sqrt{\frac{2(N-1)\alpha}{N}}\right)-p_N'}, \tag{3.4.7}$$

and

$$A_N(y)=p_N'\sqrt{\frac{N}{2(N-1)\alpha}}\sinh\left(x\sqrt{\frac{2(N-1)\alpha}{N}}\right)-p_N''\frac{N}{2(N-1)\alpha}\cosh\left(x\sqrt{\frac{2(N-1)\alpha}{N}}\right)\Bigg|_{x=f_N^{-1}(y)}. \tag{3.4.8}$$

Moreover, the curve $y=f_N(x)$ intersects $\{x>0\}$ at x_0 such that $A_N(f_N(x_0))=0$. That is, under Assumptions **H1′-H2′**, x_0 is the unique positive root of

$$\sqrt{\frac{2(N-1)\alpha}{N}}\tanh\left(z\sqrt{\frac{2(N-1)\alpha}{N}}\right)=\frac{p_N''(z)}{p_N'(z)}. \tag{3.4.9}$$

The proof of the unique positive root of (3.4.9) is provided in Appendix B.3.

Specializing to the case $h(x) = x^2$, we get

$$p_N^{sq}(x) = \left(\frac{N-1}{N}\right)^2 \frac{x^2}{\alpha} + \frac{N-1}{N\alpha^2}, \tag{3.4.10}$$

$$f_N^{sq}(x) = \int_{c\sqrt{\frac{N}{2(N-1)\alpha}}}^{|x|\wedge c\sqrt{\frac{N}{2(N-1)\alpha}}} \left(\frac{1}{z}\sqrt{\frac{N}{2(N-1)\alpha}}\tanh\left(z\sqrt{\frac{2(N-1)\alpha}{N}}\right) - 1\right)^{-1} dz, \tag{3.4.11}$$

where c is the unique positive root of $z\tanh z = 1$, and

$$A_N^{sq}(y) = -\frac{N}{N-1}\alpha^2(\cosh z - z\sinh(z))\Big|_{z=f_N^{-1}(y)\sqrt{\frac{2(N-1)\alpha}{N}}}. \tag{3.4.12}$$

Controlled dynamics

Given the candidate solution to (HJB-C_p), we derive the corresponding NEP by showing the existence of a weak solution $(\boldsymbol{X}_t, Y_t)$ to a Skorokhod problem with an unbounded domain, where the boundary of the domain depends on both the diffusion term $\boldsymbol{X}_t$ and the degenerate term $\boldsymbol{Y}_t$.

To start, let

$$\begin{aligned}\mathcal{W}_{NE} :&= \{(\boldsymbol{x}, y) \in \mathbb{R}^{N+1} : |\widetilde{x}^i| < f_N^{-1}(y) \text{ for } 1 \le i \le N\} \\ &= \left\{(\boldsymbol{x}, y) \in \mathbb{R}^{N+1} : \boldsymbol{n}_i \cdot \boldsymbol{x} > -\sqrt{\frac{N-1}{N}} f_N^{-1}(y) \text{ for } 1 \le i \le 2N\right\} \\ &= \cap_{i=1}^N \left(E_i^- \cup E_i^+\right)^c.\end{aligned} \tag{3.4.13}$$

The normal direction of each face is given by $(i = 1, 2, \cdots, N)$

$$\begin{aligned}\boldsymbol{n}_i &= c_i\left(-\frac{1}{N-1}, \cdots, -\frac{1}{N-1}, 1, -\frac{1}{N-1}, \cdots, -\frac{1}{N-1}, (f_N^{-1})'(y)\right), \\ \boldsymbol{n}_{i+N} &= c_{i+N}\left(\frac{1}{N-1}, \cdots, \frac{1}{N-1}, -1, \frac{1}{N-1}, \cdots, \frac{1}{N-1}, (f_N^{-1})'(y)\right),\end{aligned}$$

with the i^{th} component to be ± 1. c_i and c_{N+i} are normalizing constants such that $\|\boldsymbol{n}_i\| = \|\boldsymbol{n}_{N+i}\| = 1$.

Note that $\mathcal{W}_{NE}$ is an unbounded domain in $\mathbb{R}^{N+1}$ with $2N$ boundaries. For $i = 1, 2, \cdots, N$, define the $2N$ faces of $\mathcal{W}_{NE}$

$$\begin{aligned}F_i &= \{(\boldsymbol{x}, y) \in \partial\mathcal{W}_{NE} \mid (\boldsymbol{x}, y) \in \partial E_i^+\}, \\ F_{i+N} &= \{(\boldsymbol{x}, y) \in \partial\mathcal{W}_{NE} \mid (\boldsymbol{x}, y) \in \partial E_i^-\}.\end{aligned}$$

Denote the reflection direction on each face as

$$\begin{aligned}\boldsymbol{r}_i &= c_i' \left(0 \cdots, -1, \cdots 0, -1\right), \\ \boldsymbol{r}_{N+i} &= c_{N+i}' \left(0 \cdots, 1, \cdots 0, -1\right),\end{aligned}$$

with the i^{th} component to be ± 1. c_i' and c_{N+i}' are normalizing constants such that $\|\boldsymbol{r}_i\| = \|\boldsymbol{r}_{N+i}\| = 1$. NE strategy is defined as follows.

Case 1: $(\boldsymbol{X}_{0-}, Y_{0-}) = (\boldsymbol{x}, y) \in \overline{\mathcal{W}_{NE}}$. One can check that $\mathcal{W}_{NE}$ defined in (3.4.13) and $\{\boldsymbol{r}_i\}_{i=1}^{2N}$ defined above satisfies assumptions **A1**-**A5**. According to Theorem 24, there exists a weak solution to the Skorokhod problem with data $\left(\mathcal{W}_{NE}, \{\boldsymbol{r}_i\}_{i=1}^{2N}, \boldsymbol{b}, \boldsymbol{\sigma}, \boldsymbol{x} \in \overline{\mathcal{W}_{NE}}\right)$. (See Appendix B.2 for the satisfiability of **A1**-**A5**.)

Case 2: $(\boldsymbol{X}_{0-}, Y_{0-}) = (\boldsymbol{x}, y) \notin \overline{\mathcal{W}_{NE}}$, that is, there exists $i \in \{1, \cdots, N\}$ such that $(\boldsymbol{X}_{0-}, Y_{0-}) \in \mathcal{A}_i$. We show that the controlled process $(\boldsymbol{X}, Y)$ jumps sequentially to a point $(\widehat{\boldsymbol{x}}, \widehat{y}) \in \overline{\mathcal{W}_{NE}}$ for some $0 \leq \widehat{y} < y$, and then follows the solution to the Skorokhod problem starting at $(\widehat{\boldsymbol{x}}, \widehat{y}) \in \overline{\mathcal{W}_{NE}}$. In this case, the jumps will either stop in finite steps, or converge to a limit point $(\widehat{\boldsymbol{x}}, \widehat{y}) \in \overline{\mathcal{W}_{NE}}$ for $0 \leq \widehat{y} < y$.

For each $k \geq 1$, let $\boldsymbol{x}_k = (x_k^1, \cdots, x_k^N)$ be the positions, and y_k be the remaining resource after the k^{th} jump. If $(\boldsymbol{x}_k, y_k) \in \mathcal{A}_i$, then the i^{th} player will jump until $\boldsymbol{X}$ hits $\partial E_i^+ \cup \partial E_i^-$. Suppose that the jumps do not stop in finite steps. At the k^{th} step, let $x_k^{(1)} \leq \cdots \leq x_k^{(N)}$ be the order statistics of $\boldsymbol{x}_k$. Note that only the player with position $x_k^{(1)}$ or $x_k^{(N)}$ intervenes. Then $(x_k^{(1)})_{k\geq 0}$ is non-decreasing and bounded from above by $x_0^{(N)}$, therefore $(x_k^{(1)})_{k\geq 0}$ converges, and so does $(x_k^{(N)})_{k\geq 0}$. Hence $(\boldsymbol{x}_k)_{k\geq 0}$ converges. Since $(y_k)_{k\geq 0}$ is decreasing and bounded below by 0, it converges to some point $\hat{y}$. Now suppose that $(\boldsymbol{x}_k, y_k) \to (\widehat{\boldsymbol{x}}, \widehat{y}) \notin \partial \mathcal{W}_{NE}$. Let $i_* \in \{1, \cdots, N\}$ such that $\widehat{\boldsymbol{x}} \in \mathcal{A}_{i_*}$. For k sufficiently large, we have $|\boldsymbol{x}_k - \widehat{\boldsymbol{x}}| < \varepsilon$ and by the triangle inequality,

$$\left| x_k^{i_*} - \frac{\sum_{j\neq i_*} x_k^j}{N-1} \right| \geq \max_{1\leq i\leq N} \left\{ \left| \widehat{x}_k^i - \frac{\sum_{j\neq i} \widehat{x}_k^j}{N-1} \right| - f_N^{-1}(\widehat{y}) \right\} - 2\varepsilon.$$

Thus the i_*^{th} player should jump at least $\left(\max_{1\leq i\leq N} \left\{ \left| \widehat{x}_k^i - \frac{\sum_{j\neq i} \widehat{x}_k^j}{N-1} \right| - f_N^{-1}(\widehat{y}) \right\} - 2\varepsilon \right) \wedge \widehat{y}$ in the $(k+1)^{th}$ step. It suffices to take ε sufficiently small to get a contradiction.

In summary, the controlled process inherits a rich structure from the candidate solution.

- If starting at a point in the common waiting region of all N players, then the controlled process is a reflected Brownian motion with an evolving free boundary.
- If staring at a point outside the common waiting region, then the controlled process follows rank-dependent dynamics with a moving origin.

NE for the N-player game

Combining the results in Sections 3.4 and 3.4, and based on the verification theorem developed in Section 3.3, we have the following theorem of the NE for the N-player game (3.2.9) with constraint (3.4.1).

Theorem 25 (NE for the N-player game $\boldsymbol{C_p}$)**.** *Assume* ***H1'-H2'****. Let $v^i : \mathbb{R}^N \times \mathbb{R}_+ \to \mathbb{R}$ be defined by*

$$v^i(\boldsymbol{x}, y) = \begin{cases} p_N(\widetilde{x}^i) + A_N(y)\cosh\left(\widetilde{x}^i\sqrt{\frac{2(N-1)\alpha}{N}}\right) & \text{if } (\boldsymbol{x}, y) \in \mathcal{W}_{-i} \cap \mathcal{W}_i, \\ v^i\left(\boldsymbol{x}^{-i}, x^i_+ + \frac{\sum_{k\neq i} x^k}{N-1}, f_N(x^i_+)\right) & \text{if } (\boldsymbol{x}, y) \in \mathcal{W}_{-i} \cap E^+_i, \\ v^i\left(\boldsymbol{x}^{-i}, \frac{\sum_{k\neq i} x^k}{N-1} - x^i_-, f_N(x^i_-)\right) & \text{if } (\boldsymbol{x}, y) \in \mathcal{W}_{-i} \cap E^-_i, \\ v^i\left(\boldsymbol{x}^{-j}, x^j_+ + \frac{\sum_{k\neq j} x^k}{N-1}, f_N(x^j_+)\right) & \text{if } (\boldsymbol{x}, y) \in \mathcal{A}_j \cap E^+_j \text{ for } j \neq i, \\ v^i\left(\boldsymbol{x}^{-j}, \frac{\sum_{k\neq j} x^k}{N-1} - x^j_-, f_N(x^j_-)\right) & \text{if } (\boldsymbol{x}, y) \in \mathcal{A}_j \cap E^-_j \text{ for } j \neq i, \end{cases} \tag{3.4.14}$$

where

- *$\mathcal{A}_i$ and $\mathcal{W}_i$ are given in* (3.4.3)*, and $E^\pm_i$ is given in* (3.4.4) *with $f_N(\cdot)$ defined by* (3.4.7)-(3.4.9)*,*
- *$\widetilde{x}^i$ is defined by* (3.4.2)*, and $A_N(\cdot)$ is defined by* (3.4.8)*,*
- *x^i_+ is the unique positive root of $z - f_N(z) = \widetilde{x}^i - y$, and x^i_- is the unique negative root of $z + f_N(z) = \widetilde{x}^i + y$.*

Then v^i is the game value associated with an MNEP $\boldsymbol{\xi}^* = (\xi^{1*}, \cdots, \xi^{N*})$. That is,

$$v^i(\boldsymbol{x}, y) = J^i_{C_p}(\boldsymbol{x}, y; \boldsymbol{\xi}^*).$$

Moreover, the controlled process $(\boldsymbol{X}^*, Y^*)$ under $\boldsymbol{\xi}^*$ is given in Section 3.4.

Proof. Now we check that conditions (i)-(vii) in Theorem 22 are satisfied.

(i) Based on the analysis in Section 3.4, when $(\boldsymbol{x}, y) \in \overline{\mathcal{W}_{NE}}$, the NE strategy is a solution to the Skorokhod problem specified in Case 2, which is a continuous process. When $(\boldsymbol{x}, y) \notin \mathcal{W}_{NE}$, the sequential push specified in Case 1 satisfies the "no simultaneous jump" condition.

(ii) Solution (3.4.14) satisfies the derivation in Section 3.4 and hence satisfies the HJB in $\mathcal{W}_{-i}$.

(iii) Since $\|\nabla^2 v^i\| \leq K$, and the control $\boldsymbol{\xi} \in \mathcal{S}_N(\boldsymbol{x}, y)$ has finite variations, the transversality condition (iii) is satisfied.

(iv) Solution (3.4.14) satisfies the smooth-fit principle in Section 3.4, therefore, $v^i \in \mathcal{C}^2(\mathcal{W}_{-i})$. v^i is convex in $\overline{\mathcal{W}_{-i}}$ since $p_N(\widetilde{x}^i) + A_N(y)\cosh\left(\widetilde{x}^i\sqrt{\frac{2(N-1)\alpha}{N}}\right)$ is convex.

(v) Since f_N^{-1} is non-increasing, in $\mathcal{W}_{-i}$, $\tilde{x}^i \le f_N^{-1}(y) \le f_N^{-1}(0) < \infty$. This implies that $\tilde{x}^i$ is bounded in $\overline{\mathcal{W}}_{-i}$. By the definition of $A_N(y)$ in (3.4.5), $A_N(y)$ is bounded in $\mathcal{W}_{-i}$. Hence $v^i_{x_j}$ is bounded in $\mathcal{W}_{-i}$ by definition (3.4.8).

(vi) By the construction of Case 1 and Case 2, when $(\boldsymbol{x}, y) \notin \overline{\mathcal{W}_{-i}}$, there is a sequential push at time 0 to move the joint position to some point $(\hat{\boldsymbol{x}}, \hat{y}) \in \partial\mathcal{W}_{-i}$. when $(\boldsymbol{x}, y) \in \overline{\mathcal{W}_{-i}}$, $(\boldsymbol{\xi}^{-i*}, , \xi^i)$ forms a solution to the Skorokhod problem in $\cap_{j\neq i}(E_j^- \cup E_j^+)^c$. It is easy to verify that $\cap_{j\neq i}(E_j^- \cup E_j^+)^c \subset \mathcal{W}_{-i}$ and the Skorokhod problem with $\cap_{j\neq i}(E_j^- \cup E_j^+)^c$ has a weak solution. Therefore condition (vi) is satisfied.

(vii) Since v^i has the same value before and after player j's control, equation (3.3.2) is trivially satisfied.

□

To illustrate, we specialize Theorem 25 to the case $N = 2$ and $h(x) = x^2$. In this case, we can also construct the strong solution of NE strategies.

Corollary 25.1 (NE for the two-player game $\boldsymbol{C_p}$)**.** *Assume* ***H1′-H2′****. The following controls*

$$\begin{cases} \xi_t^{1*,+} &= 0, \\ \xi_t^{1*,-} &= 0, \\ \xi_t^{2*,+} &= \max\left\{0, \max_{0\le s\le t}\{0, x^1 - x^2 + B_s^1 - B_s^2 - \xi_s^{2*,+} + \xi_s^{2*,-} - (f_2^{sq})^{-1}(y - \xi_s^{2*,+} - \xi_s^{2*,-})\}\right\}, \\ \xi_t^{2*,-} &= \max\left\{0, \max_{0\le u\le t}\{0, x^2 - x^1 + B_s^2 - B_s^1 + \xi_s^{2*,+} - \xi_s^{2*,-} - (f_2^{sq})^{-1}(y - \xi_s^{2*,+} - \xi_s^{2*,-})\}\right\}, \end{cases}$$

give an MNEP for the two-player game (3.2.9) *with* (3.4.1) *and* $h(x) = x^2$*, where* $(f_2^{sq})^{-1}$ *is defined in (3.4.11). Moreover, let* v^1 *and* v^2 *be the associated values of the above MNEP* (ξ^{1*}, ξ^{2*})*, then*

$$v^1(x^1, x^2, y) = \begin{cases} \frac{(x^1-x^2)^2}{4\alpha} + \frac{1}{2\alpha^2} + A(y)\cosh\left((x^1 - x^2)\sqrt{\alpha}\right) & \text{if } |x^1 - x^2| \le (f_2^{sq})^{-1}(y), \\ v^1(x^1, x^1 + x_+^2, f_2(x_+^2)) & \text{if } x^1 \le x^2 - (f_2^{sq})^{-1}(y), \\ v^1(x^1, x^1 - x_-^2, f_2(x_-^2)) & \text{if } x^1 \ge x^2 + (f_2^{sq})^{-1}(y), \end{cases} \tag{3.4.15}$$

and

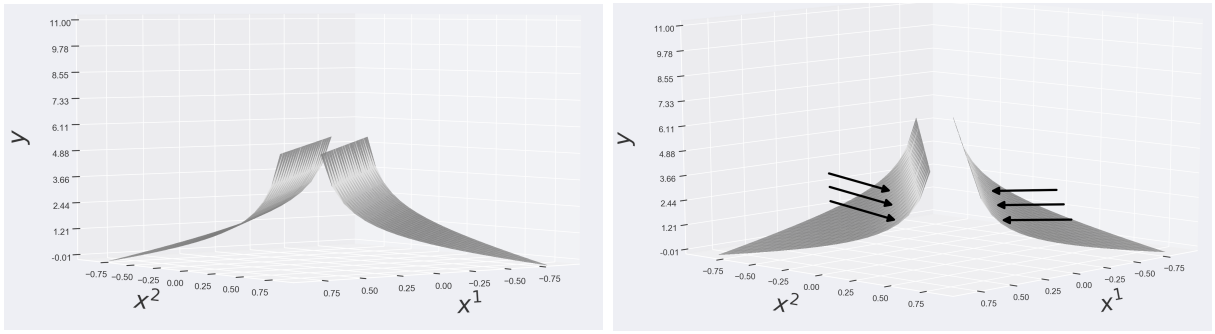
$$v^2(x^1, x^2, y) = \begin{cases} \frac{(x^2-x^1)^2}{4\alpha} + \frac{1}{2\alpha^2} + A(y)\cosh\left((x^2 - x^1)\sqrt{\alpha}\right) & \text{if } |x^2 - x^1| \le (f_2^{sq})^{-1}(y), \\ v^2(x^1, x^1 - x_-^2, f_2(x_-^2)) & \text{if } x^2 \le x^1 - (f_2^{sq})^{-1}(y), \\ v^2(x^1, x^1 + x_+^2, f_2(x_+^2))) & \text{if } x^2 \ge x^1 + (f_2^{sq})^{-1}(y), \end{cases} \tag{3.4.16}$$

where

$$A(y) = -2\alpha^2(\cosh(z) - z\sinh(z))|_{z=\sqrt{\alpha}(f_2^{sq})^{-1}(y)}, \tag{3.4.17}$$

and x_+^2 *is the unique root of* $z - f_2^{sq}(z) = x^1 - y$, *and* x_-^2 *is the unique root of* $z + f_2^{sq}(z) = x^1 + y$.

Note that under partition $\{Q_i\}_{i=1,2}$, we have $\mathcal{A}_1 = \emptyset$, hence $(\xi^{1*,+}, \xi^{1*,-}) = (0, 0)$.



(a) No control from player one.

(b) Control from player two.

Figure 3.2: Case $\boldsymbol{C_p}$: MNEP when $N = 2$.

3.5 Nash Equilibrium for Game C_d

In this section, we study the MNEP of the N-player game $\boldsymbol{C_d}$. That is $A = \boldsymbol{I_N} \in \mathbb{R}^{N\times N}$, and

$$Y_t^i = y^i - \check{\xi}_t^i \quad \text{with} \quad Y_{0-}^i = y^i. \tag{3.5.1}$$

Recall that the major difference between game $\boldsymbol{C_p}$ and game $\boldsymbol{C_d}$ is that, in the former all N players share a fixed amount of the same resource, while in the latter each player has her own individual fixed resource constraint. This difference is reflected in $(HJB - C_p)$ and $(HJB - C_d)$ in terms of their dimensionality, and in each player's control based on the remaining resources. In particular, $(HJB - C_p)$ and the state space $(\boldsymbol{x}, y)$ of $\boldsymbol{C_p}$ are of dimension $N+1$, whereas $(HJB - C_d)$ and the state space $(\boldsymbol{x}, \boldsymbol{y})$ of $\boldsymbol{C_d}$ are of dimension $2N$. Moreover, in game $\boldsymbol{C_p}$, the gradient constraint is $-v_y^i \pm v_{x^i}^i$ for player i. In contrast, in game $\boldsymbol{C_d}$, each player controls her own resource level, the gradient constraint becomes $-v_{y^i}^i \pm v_{x^i}^i$ for player i. So if $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$, the HJB equation for $v^i(\boldsymbol{x}, \boldsymbol{y})$ in game $\boldsymbol{C}_d$ is as follows.

$$\textit{(HJB-C}_d\textit{)} \begin{cases} \min\left\{-\alpha v^i + h\left(\dfrac{N-1}{N}\widetilde{x}^i\right) + \dfrac{1}{2}\displaystyle\sum_{j=1}^{N} v_{x^j x^j}^i, -v_{y^i}^i + v_{x^i}^i, -v_{y^i}^i - v_{x^i}^i\right\} = 0, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{for } (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{W}_{-i}, \\ \min\left\{-v_{y^j}^i + v_{x^j}^i, -v_{y^j}^i - v_{x^j}^i\right\} = 0, \qquad\qquad \text{for } (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{A}_j, j \neq i. \end{cases}$$

Note that the control policy of the i^{th} player only depends on $(\boldsymbol{x}, y^i)$ in $\mathcal{W}_{-i}$. As seen in Section 3.4, for the controlled process of type $\boldsymbol{C_p}$, upon hitting the boundary of the

polyhedron, the polyhedron will expand in all directions. While for the controlled process of type $\boldsymbol{C_d}$, only one direction of the the polyhedron will move once hit.

To proceed, similar to Section 3.4, define the action region $\mathcal{A}_i \in \mathbb{R}^N \times \mathbb{R}^N_+$ and the waiting region $\mathcal{W}_i$ of the i^{th} player by

$$\mathcal{A}_i := (E_i^+ \cup E_i^-) \cap Q_i \quad \text{and} \quad \mathcal{W}_i := \mathbb{R}^N \times \mathbb{R}^N_+ \setminus \mathcal{A}_i, \tag{3.5.2}$$

where

$$\begin{aligned} Q_i := \big\{(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^N \times \mathbb{R}^N_+ : |\widetilde{x}^i| - f_N^{-1}(y^i) \geq |\widetilde{x}^k| - f_N^{-1}(y^k) \text{ for } k < i, \\ |\widetilde{x}^i| - f_N^{-1}(y^i) > |\widetilde{x}^k| - f_N^{-1}(y^k) \text{ for } k > i\big\}, \end{aligned}$$

and

$$E_i^+ := \big\{(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^N \times \mathbb{R}^N_+ : \widetilde{x}^i \geq f_N^{-1}(y^i)\big\} \text{ and } E_i^- := \big\{(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^N \times \mathbb{R}^N_+ : \widetilde{x}^i \leq -f_N^{-1}(y^i)\big\}. \tag{3.5.3}$$

Recall the definition of the threshold function $f_N(\cdot)$ from (3.4.7)-(3.4.9), we now investigate control of player i which only depends on $(\boldsymbol{x}, y^i)$ in $\mathcal{W}_{-i}$. That is, for $|\widetilde{x}^i| < f_N^{-1}(y^i)$,

$$v^i(\boldsymbol{x},\boldsymbol{y}) = p_N(\widetilde{x}^i) + A_N(y^i)\cosh\left(\widetilde{x}^i\sqrt{\frac{2(N-1)\alpha}{N}}\right), \tag{3.5.4}$$

is a solution to (HJB-C_d), where $p_N(\cdot)$ is defined by (3.4.6), and $A_N(\cdot)$ defined by (3.4.8).

The next step is to construct the controlled process $(\boldsymbol{X},\boldsymbol{Y})$ corresponding to the HJB solution (3.5.4). Let

$$\begin{aligned} \mathcal{W}_{NE} :&= \{(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^N \times \mathbb{R}^N_+ : |\widetilde{x}^i| < f_N^{-1}(y^i) \text{ for } 1 \leq i \leq N\} \\ &= \cap_{i=1}^N \left(E_i^- \cup E_i^+\right)^c. \end{aligned} \tag{3.5.5}$$

The normal direction on each face is given by

$$\begin{aligned} \boldsymbol{n}_i &= c_i\left(\frac{1}{N-1},\cdots,\frac{1}{N-1}-1,\frac{1}{N-1}\cdots,\frac{1}{N-1};0,\cdots,0,(f_N^{-1})'\left(y^i\right),0,\cdots,0\right), \\ \boldsymbol{n}_{N+i} &= c_{N+i}\left(-\frac{1}{N-1},\cdots,-\frac{1}{N-1},1,-\frac{1}{N-1},\cdots,-\frac{1}{N-1};0,\cdots,0,(f_N^{-1})'\left(y^i\right),0,\cdots,0\right), \end{aligned}$$

with the i^{th} component to be ± 1 and the $(N+i)^{th}$ component to be $(f_N^{-1})'(y^i)$. c_i and c_{N+i} are normalizing constants such that $\|\boldsymbol{n}_i\| = \|\boldsymbol{n}_{N+i}\| = 1$.

Note that $\mathcal{W}_{NE}$ is an unbounded domain in $\mathbb{R}^{2N}$ with $2N$ boundaries. For $i = 1, 2, \cdots, N$, define the $2N$ faces of $\mathcal{W}_{NE}$

$$\begin{aligned} F_i &= \{(\boldsymbol{x},\boldsymbol{y}) \in \partial\mathcal{W}_{NE} \mid (\boldsymbol{x},\boldsymbol{y}) \in \partial E_i^+\}, \\ F_{i+N} &= \{(\boldsymbol{x},\boldsymbol{y}) \in \partial\mathcal{W}_{NE} \mid (\boldsymbol{x},\boldsymbol{y}) \in \partial E_i^-\}. \end{aligned}$$

Denote the reflection direction on each face as

$$\begin{aligned} \boldsymbol{r}_i &= c_i' \left(0 \cdots, 0, -1, 0, \cdots 0; 0, \cdots, 0, -1, 0, \cdots, 0\right), \\ \boldsymbol{r}_{N+i} &= c_{N+i}' \left(0 \cdots, 0, 1, 0, \cdots 0; 0, \cdots, 0, -1, 0, \cdots, 0\right), \end{aligned}$$

with the i^{th} component to be ± 1 and the $(N+i)^{th}$ component to be 1. c_i' and c_{N+i}' are normalizing constants such that $\|\boldsymbol{r}_i\| = \|\boldsymbol{r}_{N+i}\| = 1$.

The NE strategy is defined as follows.

Case 1: $(\boldsymbol{X}_{0-}, \boldsymbol{Y}_{0-}) = (\boldsymbol{x}, \boldsymbol{y}) \in \overline{\mathcal{W}_{NE}}$. One can check that $\mathcal{W}_{NE}$ defined in (3.5.5) and $\{\boldsymbol{r}_i\}_{i=1}^{2N}$ defined above satisfies assumptions **A1**-**A5**. Therefore, there exists a weak solution to the Skorokhod problem with data $\left(\mathcal{W}_{NE}, \{\boldsymbol{r}_i\}_{i=1}^{2N}, \boldsymbol{b}, \boldsymbol{\sigma}, \boldsymbol{x} \in \overline{\mathcal{W}_{NE}}\right)$. (See Appendix B.2 for the satisfiability of **A1**-**A5**.)

Case 2: $(\boldsymbol{X}_{0-}, \boldsymbol{Y}_{0-}) = (\boldsymbol{x}, \boldsymbol{y}) \notin \overline{\mathcal{W}_{NE}}$. There exists $i \in \{1, \cdots, N\}$ such that $(\boldsymbol{X}_{0-}, \boldsymbol{Y}_{0-}) \in \mathcal{A}_i$. For each $k \geq 1$, let $\boldsymbol{x}_k = (x_k^1, \cdots, x_k^N)$ be the positions, and $\boldsymbol{y}_k = (y_k^1, \cdots, y_k^N)$ be the resource remaining after the k^{th} control. If $(\boldsymbol{x}_k, \boldsymbol{y}_k) \in \mathcal{A}_i$, then the i^{th} player will control until $\boldsymbol{X}$ hits $\partial E_i^+ \cup \partial E_i^-$. The argument in Section 3.4 shows that the controlled process $\boldsymbol{X}$ controls sequentially to a point $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) \in \overline{\mathcal{W}_{NE}}$ for $\boldsymbol{0} \leq \widehat{\boldsymbol{y}} \leq \boldsymbol{y}$. Then $(\boldsymbol{X}, \boldsymbol{Y})$ follows the solution to the Skorokhod problem starting at $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}})$.

In summary, the NE for the N-player game (3.2.9) with constraint $\boldsymbol{C_d}$ is stated as follows.

Theorem 26 (NE for the N-player game $\boldsymbol{C_d}$)**.** *Assume* ***H1'*****-*****H2'****. Let $v^i : \mathbb{R}^N \times \mathbb{R}_+^N \to \mathbb{R}$ be defined by*

$$v^i(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} p_N(\widetilde{x}^i) + A_N(y^i) \cosh\left(\widetilde{x}^i \sqrt{\frac{2(N-1)\alpha}{N}}\right) & \text{if } (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{W}_{-i} \cap \mathcal{W}_i, \\ v^i\left(\boldsymbol{x}^{-i}, x_+^i + \frac{\sum_{k \neq i} x^k}{N-1}, f_N(x_+^i)\right) & \text{if } (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{W}_{-i} \cap E_i^+, \\ v^i\left(\boldsymbol{x}^{-i}, \frac{\sum_{k \neq i} x^k}{N-1} - x_-^i, f_N(x_-^i)\right) & \text{if } (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{W}_{-i} \cap E_i^-, \\ v^i\left(\boldsymbol{x}^{-j}, x_+^j + \frac{\sum_{k \neq j} x^k}{N-1}, y^i\right) & \text{if } (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{A}_j \cap E_j^+ \text{ for } j \neq i, \\ v^i\left(\boldsymbol{x}^{-j}, \frac{\sum_{k \neq j} x^k}{N-1} - x_-^j, y^i\right) & \text{if } (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{A}_j \cap E_j^- \text{ for } j \neq i, \end{cases} \tag{3.5.6}$$

where

- *$\mathcal{A}_i$ and $\mathcal{W}_i$ are given in* (3.5.2)*, and $E_i^\pm$ is given in* (3.5.3) *with $f_N(\cdot)$ defined by* (3.4.7)-(3.4.9)*,*
- *$\widetilde{x}^i$ is defined by* (3.4.2)*, and $A_N(\cdot)$ is defined by* (3.4.8)*,*
- *x_+^i is the unique positive root of $z - f_N(z) = \widetilde{x}^i - y$, and x_-^i is the unique negative root of $z + f_N(z) = \widetilde{x}^i + y$.*

Then v^i is the game value associated with an MNEP $\boldsymbol{\xi}^* = (\xi^{1*}, \cdots, \xi^{N*})$. That is,

$$v^i(\boldsymbol{x}, \boldsymbol{y}) = J^i_{C_d}(\boldsymbol{x}, \boldsymbol{y}; \boldsymbol{\xi}^*).$$

Moreover, the controlled process $(\boldsymbol{X}^*, \boldsymbol{Y}^*)$ under $\boldsymbol{\xi}^*$ is given in this section:

Case 1 *if $(\boldsymbol{x}, \boldsymbol{y}) \in \overline{\mathcal{W}_{NE}}$, and*

Case 2 *if $(\boldsymbol{x}, \boldsymbol{y}) \notin \overline{\mathcal{W}_{NE}}$.*

Theorem 26 can be verified in a similar way as Theorem 25. Specializing to the two-player game with $h(x) = x^2$, we have the following result.

Corollary 26.1 (NE for $N = 2$ for game $\boldsymbol{C_d}$)**.** *Assume **H1′**-**H2′**. The following controls*

$$\begin{cases} \xi_t^{1*,+} := \Delta\xi_0^{1*,+} + \displaystyle\int_0^{t\wedge\tau_1} \boldsymbol{1}_{\{\boldsymbol{X}_s^* \in F_1(Y_s^{1*})\}} \boldsymbol{1}_{\{Y_s^{1*} > Y_s^{2*}\}} d\eta_s^1, \\ \xi_t^{1*,-} := \Delta\xi_0^{1*,-} + \displaystyle\int_0^{t\wedge\tau_1} \boldsymbol{1}_{\{\boldsymbol{X}_s^* \in F_3(Y_s^{1*})\}} \boldsymbol{1}_{\{Y_s^{1*} > Y_s^{2*}\}} d\eta_s^3, \\ Y_t^{1*} := y^1 - \hat{\xi}_t^{1*}, \quad \tau_1 := \inf\{t \geq 0 : Y_t^{1*} = 0\}, \\ \xi_t^{2*,+} := \Delta\xi_0^{2*,+} + \displaystyle\int_0^{t\wedge\tau_2} \boldsymbol{1}_{\{\boldsymbol{X}_s^* \in F_2(Y_s^{2*})\}} \boldsymbol{1}_{\{Y_s^{2*} \geq Y_s^{1*}\}} d\eta_s^2, \\ \xi_t^{2*,-} := \Delta\xi_0^{2*,-} + \displaystyle\int_0^{t\wedge\tau_2} \boldsymbol{1}_{\{\boldsymbol{X}_t^* \in F_4(Y_s^{2*})\}} \boldsymbol{1}_{\{Y_s^{2*} \geq Y_s^{1*}\}} d\eta_s^4, \\ Y_t^{2*} := y^2 - \hat{\xi}_t^{2*}, \quad \tau_2 := \inf\{t \geq 0 : Y_t^{2*} = 0\}, \end{cases} \tag{3.5.7}$$

give an MNEP for the two-player game $\boldsymbol{C_d}$ with $h(x) = x^2$, where

- $F_1(y) = F_4(y) = \Big\{(x^1, x^2) : x^1 - x^2 = -(f_2^{sq})^{-1}(y)\Big\}$,
- $F_2(y) = F_3(y) = \Big\{(x^1, x^2) : x^1 - x^2 = (f_2^{sq})^{-1}(y)\Big\}$,
- η_t^{i*} *are non-decreasing processes with* $\eta_{0-}^{i*} = 0$ $(i = 1, 2, 3, 4)$,
- $$\Delta\xi_0^{2*,+} = \begin{cases} x_-^2, \ \textit{if } y^2 \geq y^1 \ \textit{ and } \ x^2 \leq x^1 - (f_2^{sq})^{-1}(y^2), \\ x_-^2, \ \textit{if } y^2 < y^1 \ \textit{and } \ x^2 \leq x_+^1 - (f_2^{sq})^{-1}(y^2), \end{cases}$$

  $$\Delta\xi_0^{2*,-} = \begin{cases} x_+^2, \ \textit{if } y^2 \geq y^1 \ \textit{ and } \ x^2 \geq x^1 - (f_2^{sq})^{-1}(y^2), \\ x_+^2, \ \textit{if } y^2 < y^1 \ \textit{and } \ x^2 \geq x_-^1 - (f_2^{sq})^{-1}(y^2), \end{cases}$$

  $$\Delta\xi_0^{1*,+} = \begin{cases} x_-^1, \ \textit{if } y^1 > y^2 \ \textit{ and } \ x^1 \leq x^2 - (f_2^{sq})^{-1}(y^1), \\ x_-^1, \ \textit{if } y^1 < y^2 \ \textit{and } \ x^1 \leq x_+^2 - (f_2^{sq})^{-1}(y^1), \end{cases}$$

$$\Delta\xi_0^{1*,-} = \begin{cases} x_+^1, \text{ if } y^1 > y^2 \text{ and } x^1 \geq x^2 - (f_2^{sq})^{-1}(y^1), \\ x_+^1, \text{ if } y^1 < y^2 \text{ and } x^1 \geq x_-^2 - (f_2^{sq})^{-1}(y^1), \end{cases}$$

- x_+^i *is the unique root of* $z - f_2^{sq}(z) = x^j - y$, x_-^i *is the unique root of* $z + f_2^{sq}(z) = x^j + y$, *with* $f_2^{sq}(\cdot)$ *is given by* (3.4.11). *(*$i, j = 1, 2$ *and* $i \neq j$*).*

Moreover, let v^1 *and* v^2 *be the corresponding values of the above MNEP* (ξ^{1*}, ξ^{2*}). *Then if* $y^1 > y^2$,

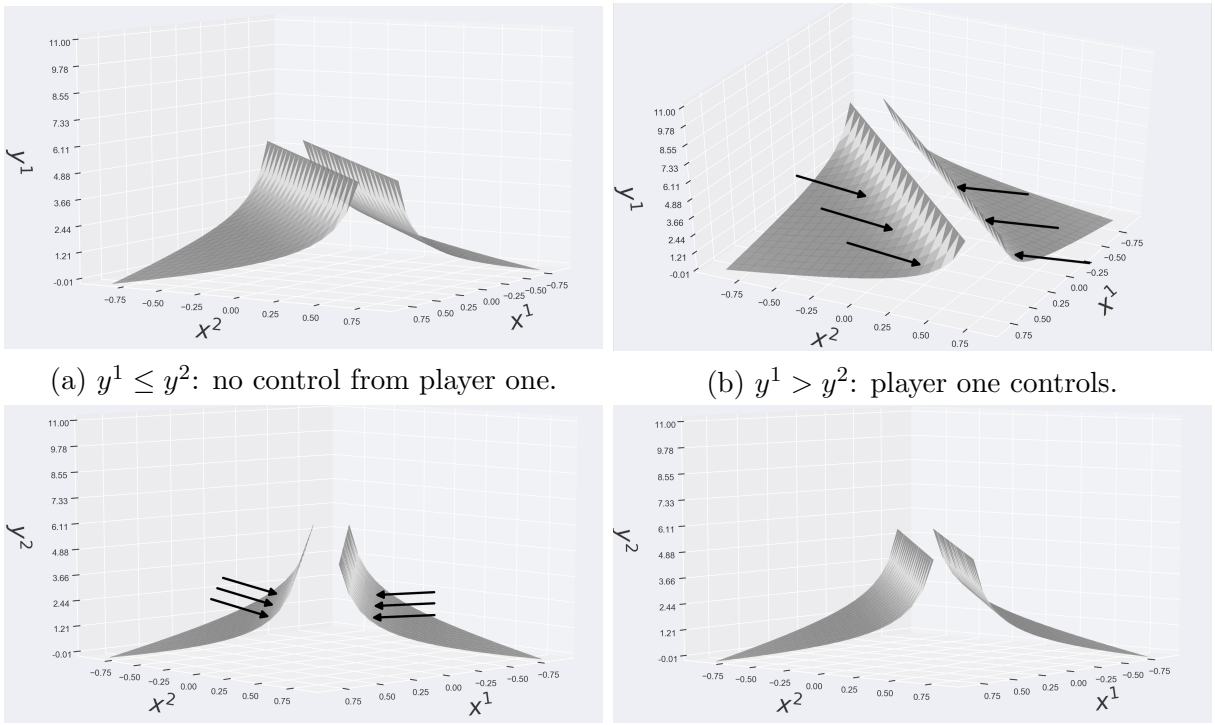
$$\begin{cases} v^1(x^1, x^2, y^1) = \begin{cases} \frac{(x^1-x^2)^2}{4\alpha} + \frac{1}{2\alpha^2} + A(y^1)\cosh\left((x^1 - x^2)\sqrt{\alpha}\right) & \text{if } |x^1 - x^2| \leq (f_2^{sq})^{-1}(y^1), \\ v^1(x_-^1, x^2 - x_-^1, f_2^{sq}(x_-^1)) & \text{if } x^1 \leq x^2 - (f_2^{sq})^{-1}(y^1), \\ v^1(x_+^1, x^2 + x_+^1, f_2^{sq}(x_+^1)) & \text{if } x^1 \geq x^2 + (f_2^{sq})^{-1}(y^1), \end{cases} \\ v^2(x^1, x^2, y^2) = \begin{cases} \frac{(x^2-x^1)^2}{4\alpha} + \frac{1}{2\alpha^2} + A(y^2)\cosh\left((x^2 - x^1)\sqrt{\alpha}\right) & \text{if } |x^2 - x^1| \leq (f_2^{sq})^{-1}(y^2), \\ v^2(x_+^1, x^2, y^2) & \text{if } x^2 \leq x^1 - (f_2^{sq})^{-1}(y^2), \\ v^2(x_-^1, x^2, y^2) & \text{if } x^2 \geq x^1 + (f_2^{sq})^{-1}(y^2); \end{cases} \end{cases} \tag{3.5.8}$$

and if $y^1 \leq y^2$,

$$\begin{cases} v^1(x^1, x^2, y^1) = \begin{cases} \frac{(x^1-x^2)^2}{4\alpha} + \frac{1}{2\alpha^2} + A(y^1)\cosh\left((x^1 - x^2)\sqrt{\alpha}\right) & \text{if } |x^1 - x^2| \leq (f_2^{sq})^{-1}(y^1), \\ v^1(x^1, x_+^2, y^1) & \text{if } x^1 \leq x^2 - (f_2^{sq})^{-1}(y^1), \\ v^1(x^1, x_-^2, y^1) & \text{if } x^1 \geq x^2 + (f_2^{sq})^{-1}(y^1), \end{cases} \\ v^2(x^1, x^2, y^2) = \begin{cases} \frac{(x^2-x^1)^2}{4\alpha} + \frac{1}{2\alpha^2} + A(y^2)\cosh\left((x^2 - x^1)\sqrt{\alpha}\right) & \text{if } |x^2 - x^1| \leq (f_2^{sq})^{-1}(y^2), \\ v^2(x^1, x^1 + x_+^2, f_2^{sq}(x_+^2)) & \text{if } x^2 \leq x^1 - (f_2^{sq})^{-1}(y^2), \\ v^2(x^1, x^1 - x_-^2, f_2^{sq}(x_-^2)) & \text{if } x^2 \geq x^1 + (f_2^{sq})^{-1}(y^2), \end{cases} \end{cases} \tag{3.5.9}$$

where $A(\cdot)$ *is given by* (3.4.17).

Comparison of Corollary 25.1 and Corollary 26.1. Consider $N = 2$ and $h(x) = x^2$. In game $\boldsymbol{C_p}$, only player two controls the two separating hyperplanes whereas player one does nothing, see Figure 3.2. In game $\boldsymbol{C_p}$, player one controls the two separating hyperplanes when $y^1 > y^2$ and she does nothing when $y^2 \geq y^1$. See Figure 3.3.



(a) $y^1 \leq y^2$: no control from player one.

(b) $y^1 > y^2$: player one controls.

(c) $y^1 \leq y^2$: player two controls.

(d) $y^1 > y^2$: no control from player two.

Figure 3.3: Case $\boldsymbol{C_d}$: MNEP when $N = 2$.

3.6 Nash Equilibrium for game $\boldsymbol{C}$

In the previous two sections, we have dealt with two special games $\boldsymbol{C_p}$ and $\boldsymbol{C_d}$. Analysis of these two games provides important insight into the solution structure of the general game $\boldsymbol{C}$. Namely, the NE strategy depends on the positions of players and their remaining resource levels. With these two special cases in mind, now recall that in game $\boldsymbol{C}$,

$$dY_t^j = -\sum_{i=1}^{N} \frac{a_{ij} Y_{t-}^j}{\sum_{k=1}^{M} a_{ik} Y_{t-}^k} d\check{\xi}_t^i \quad \text{and} \quad Y_{0-}^j = y^j \geq 0. \tag{3.6.1}$$

For the HJB equation $(HJB - C)$, the gradient constraint is more complicated than the two special cases $\boldsymbol{C_p}$ and $\boldsymbol{C_d}$. When $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$,

$$\textit{(HJB-C)} \begin{cases} \min\left\{ -\alpha v^i + h + \frac{1}{2}\sum_{j=1}^{N} v^i_{x^j x^j}, -\sum_{j=1}^{M} \frac{a_{ij} y^j}{\sum_{k=1}^{M} a_{ik} y^k} v^i_{y^j} + v^i_{x^i}, -\sum_{j=1}^{M} \frac{a_{ij} y^j}{\sum_{k=1}^{M} a_{ik} y^k} v^i_{y^j} - v^i_{x^i} \right\} = 0, \\ \qquad\qquad \text{for } (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{W}_{-i}, \\ \min\left\{ -\sum_{k=1}^{M} \frac{a_{jk} y^k}{\sum_{s=1}^{M} a_{js} y^s} v^i_{y^k} + v^i_{x^j}, -\sum_{k=1}^{M} \frac{a_{jk} y^k}{\sum_{s=1}^{M} a_{js} y^s} v^i_{y^k} - v^i_{x^j} \right\} = 0, \\ \qquad\qquad \text{for } (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{A}_j, j \neq i. \end{cases}$$

In particular, if $\boldsymbol{A} = [1, 1, \cdots, 1]^T \in \mathbb{R}^{N\times 1}$, then $(HJB - C)$ becomes $(HJB - C_p)$; and if $\boldsymbol{A} = \boldsymbol{I_N}$, then it is $(HJB - C_d)$.

Similar to Section 3.4, define the action region $\mathcal{A}_i \in \mathbb{R}^N \times \mathbb{R}^M_+$ and the waiting region $\mathcal{W}_i$ of the i^{th} player by

$$\mathcal{A}_i := (E_i^+ \cup E_i^-) \cap Q_i \quad \text{and} \quad \mathcal{W}_i := \mathbb{R}^N \times \mathbb{R}^N_+ \setminus \mathcal{A}_i, \tag{3.6.2}$$

where

$$Q_i := \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^N \times \mathbb{R}^M_+ : |\widetilde{x}^i| - f_N^{-1}\left(\sum_{j=1}^M a_{ij}y^j\right) \geq |\widetilde{x}^k| - f_N^{-1}\left(\sum_{j=1}^M a_{kj}y^j\right) \text{ for } k < i, \right.$$
$$\left. |\widetilde{x}^i| - f_N^{-1}\left(\sum_{j=1}^M a_{ij}y^j\right) > |\widetilde{x}^k| - f_N^{-1}\left(\sum_{j=1}^M a_{kj}y^j\right) \text{ for } k > i \right\},$$

and

$$E_i^+ := \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^N \times \mathbb{R}^M_+ : \widetilde{x}^i \geq f_N^{-1}\left(\sum_{j=1}^M a_{ij}y^j\right) \right\} \text{ and } E_i^- := \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^N \times \mathbb{R}^M_+ : \widetilde{x}^i \leq -f_N^{-1}\left(\sum_{j=1}^M a_{ij}y^j\right) \right\}. \tag{3.6.3}$$

From the analysis in Sections 3.4 and 3.5, and the "guess" that the control policy of player i only depends on $(\boldsymbol{x}, \sum_{j=1}^M a_{ij}y^j)$ when in $\mathcal{W}_{-i}$, we get for $|\widetilde{x}^i| < f_N^{-1}(\sum_{j=1}^M a_{ij}y^j)$,

$$v^i(\boldsymbol{x}, \boldsymbol{y}) = p_N(\widetilde{x}^i) + A_N\left(\sum_{j=1}^M a_{ij}y^j\right) \cosh\left(\widetilde{x}^i\sqrt{\frac{2(N-1)\alpha}{N}}\right), \tag{3.6.4}$$

is a solution to (HJB-C), where $p_N(\cdot)$ is defined by (3.4.6), and $A_N(\cdot)$ defined by (3.4.8).

The next step is to construct the controlled process $(\boldsymbol{X}, \boldsymbol{Y})$ corresponding to the HJB solution (3.6.4).

$$\begin{aligned} \mathcal{W}_{NE} :&= \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^N \times \mathbb{R}^M_+ : |\widetilde{x}^i| < f_N^{-1}\left(\sum_{j=1}^M a_{ij}y^j\right) \text{ for } 1 \leq i \leq N \right\} \\ &= \cap_{i=1}^N \left(E_i^- \cup E_i^+\right)^c. \end{aligned} \tag{3.6.5}$$

The normal direction on each face is given by

$$\begin{aligned} \boldsymbol{n}_i &= c_i \left(\frac{1}{N-1}, \cdots, -1, \cdots, \frac{1}{N-1}; (f_N^{-1})'\left(\sum_{j=1}^M a_{ij}y^j\right) a_{i1}, \cdots, (f_N^{-1})'\left(\sum_{j=1}^M a_{ij}y^j\right) a_{iM} \right), \\ \boldsymbol{n}_{N+i} &= c_{N+i} \left(-\frac{1}{N-1}, \cdots, 1, \cdots, -\frac{1}{N-1}; (f_N^{-1})'\left(\sum_{j=1}^M a_{ij}y^j\right) a_{i1}, \cdots, (f_N^{-1})'\left(\sum_{j=1}^M a_{ij}y^j\right) a_{iM} \right), \end{aligned}$$

with the i^{th} component being ± 1, and c_i and c_{N+i} the normalizing constants such that $\|\boldsymbol{n}_i\| = \|\boldsymbol{n}_{N+i}\| = 1$.

Note that $\mathcal{W}_{NE}$ is an unbounded domain in $\mathbb{R}^{2N}$ with $2N$ boundaries. For $i = 1, 2, \cdots, N$, define the $2N$ faces of $\mathcal{W}_{NE}$

$$\begin{aligned} F_i &= \{(\boldsymbol{x},\boldsymbol{y}) \in \partial\mathcal{W}_{NE} \mid (\boldsymbol{x},\boldsymbol{y}) \in \partial E_i^+\}, \\ F_{i+N} &= \{(\boldsymbol{x},\boldsymbol{y}) \in \partial\mathcal{W}_{NE} \mid (\boldsymbol{x},\boldsymbol{y}) \in \partial E_i^-\}. \end{aligned}$$

Denote the reflection direction on each face as

$$\begin{aligned} \boldsymbol{r}_i &= c_i' \left(0 \cdots, -1, \cdots 0; -\frac{a_{i1}y^1}{\sum_{j=1}^M a_{ij}y^j}, \cdots, -\frac{a_{iM}y^M}{\sum_{j=1}^M a_{ij}y^j}\right), \\ \boldsymbol{r}_{N+i} &= c_{N+i}' \left(0 \cdots, 1, \cdots 0; -\frac{a_{i1}y^1}{\sum_{j=1}^M a_{ij}y^j}, \cdots, -\frac{a_{iM}y^M}{\sum_{j=1}^M a_{ij}y^j}\right), \end{aligned}$$

with the i^{th} component to be ± 1. c_i' and c_{N+i}' are normalizing constants such that $\|\boldsymbol{r}_i\| = \|\boldsymbol{r}_{N+i}\| = 1$.

NE strategy is defined as follows.

Case 1: $(\boldsymbol{X}_{0-}, \boldsymbol{Y}_{0-}) = (\boldsymbol{x},\boldsymbol{y}) \in \overline{\mathcal{W}_{NE}}$. One can check that $\mathcal{W}_{NE}$ defined in (3.6.5) and $\{\boldsymbol{r}_i\}_{i=1}^{2N}$ defined above satisfies assumptions **A1**-**A5**. Therefore, there exists a weak solution to the Skorokhod problem with data $\left(\mathcal{W}_{NE}, \{\boldsymbol{r}_i\}_{i=1}^{2N}, \boldsymbol{b}, \boldsymbol{\sigma}, \boldsymbol{x} \in \overline{\mathcal{W}_{NE}}\right)$. (See Appendix B.2 for the satisfiability of **A1**-**A5**.)

Case 2: $(\boldsymbol{X}_{0-}, \boldsymbol{Y}_{0-}) = (\boldsymbol{x},\boldsymbol{y}) \in \overline{\mathcal{W}_{NE}}$. There exists $i \in \{1, \cdots, N\}$ such that $(\boldsymbol{X}_{0-}, \boldsymbol{Y}_{0-}) \in \mathcal{A}_i$. For each $k \geq 1$, let $\boldsymbol{x}_k = (x_k^1, \cdots, x_k^N)$ be the positions, and $\boldsymbol{y}_k = (y_k^1, \cdots, y_k^M)$ be the remaining resource level after the k^{th} jump. If $(\boldsymbol{x}_k, \boldsymbol{y}_k) \in \mathcal{A}_i$, then the i^{th} player will jump until $\boldsymbol{X}$ hits $\partial E_i^+ \cup \partial E_i^-$. The argument in Section 3.4 shows that the controlled process $(\boldsymbol{X}, \boldsymbol{Y})$ jumps sequentially to a point $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) \in \overline{\mathcal{W}_{NE}}$ for $\boldsymbol{0} \leq \widehat{\boldsymbol{y}} \leq \boldsymbol{y}$. Then $(\boldsymbol{X}, \boldsymbol{Y})$ follows the solution to the Skorokhod problem starting at $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}})$.

The NE for the N-player game (3.2.9) with constraint $\boldsymbol{C}$ is stated as follows.

Theorem 27 (NE for the N-player game $\boldsymbol{C}$)**.** *Assume* ***H1'***-***H2'***. *Let* $v^i : \mathbb{R}^N \times \mathbb{R}_+^M \to \mathbb{R}$ *be defined*

by

$$v^i(\boldsymbol{x},\boldsymbol{y}) = \begin{cases} p_N(\widetilde{x}^i) + A_N(\sum_{j=1}^M a_{ij}y^j)\cosh\left(\widetilde{x}^i\sqrt{\frac{2(N-1)\alpha}{N}}\right) & \text{if } (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{W}_{-i}\cap\mathcal{W}_i, \\ v^i\left(\boldsymbol{x}^{-i}, x^i_+ + \frac{\sum_{k\neq i}x^k}{N-1}, f_N(x^i_+)\right) & \text{if } (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{W}_{-i}\cap E^+_i, \\ v^i\left(\boldsymbol{x}^{-i}, \frac{\sum_{k\neq i}x^k}{N-1} - x^i_-, f_N(x^i_-)\right) & \text{if } (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{W}_{-i}\cap E^-_i, \\ v^i\left(\boldsymbol{x}^{-j}, x^j_+ + \frac{\sum_{k\neq j}x^k}{N-1}, y^i\right) & \text{if } (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{A}_j\cap E^+_j \text{ for } j\neq i, \\ v^i\left(\boldsymbol{x}^{-j}, \frac{\sum_{k\neq j}x^k}{N-1} - x^j_-, y^i\right) & \text{if } (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{A}_j\cap E^-_j \text{ for } j\neq i, \end{cases} \tag{3.6.6}$$

where

- $\mathcal{A}_i$ *and* $\mathcal{W}_i$ *are given in* (3.6.2)*, and* $E^\pm_i$ *is given in* (3.6.3) *with* $f_N(\cdot)$ *defined by* (3.4.7)-(3.4.9)*,*
- $\widetilde{x}^i$ *is defined by* (3.4.2)*, and* $A_N(\cdot)$ *defined by* (3.4.8)*,*
- x^i_+ *is the unique positive root of* $z - f_N(z) = \widetilde{x}^i - \sum_{j=1}^M a_{ij}y^j$*, and* x^i_- *is the unique negative root of* $z + f_N(z) = \widetilde{x}^i + \sum_{j=1}^M a_{ij}y^j$*.*

Then v^i *is the value associated with a MNEP* $\boldsymbol{\xi}^* = (\xi^{1*},\cdots,\xi^{N*})$*. That is,*

$$v^i(\boldsymbol{x},\boldsymbol{y}) = J^i_C(\boldsymbol{x},\boldsymbol{y};\boldsymbol{\xi}^*).$$

Moreover, the controlled process $(\boldsymbol{X}^*,\boldsymbol{Y}^*)$ *under* $\boldsymbol{\xi}^*$ *is a solution to a Skorokhod problem as described in* ***Case 1*** *if* $(\boldsymbol{x},\boldsymbol{y}) \in \overline{\mathcal{W}_{NE}}$*, and described as* ***Case 2*** *if* $(\boldsymbol{x},\boldsymbol{y}) \notin \overline{\mathcal{W}_{NE}}$*.*

Remark 27.1. *Since each player makes decisions based on the total available resource and is indifferent to the resource identity, we assume the boundary in the smooth-fit principle satisfies* $A_N(y_1,\cdots,y_M) = A_N(\sum_{j=1}^M a_{ij}y^j)$ *for player* i*. Note that the value function depends on* $\boldsymbol{y}$ *only through* $\sum_j a_{ij}y^j$*. Therefore if we denote* $\tilde{v}^i(\boldsymbol{x},z) := v^i(\boldsymbol{x},\boldsymbol{y})$ *and* $z = \sum_{j=1}^N a_{ij}y^j$*, it is easy to verify that* $\sum_{j=1}^M \frac{a_{ij}y^j}{\sum_{k=1}^M a_{ik}y^k} v^i_{y^j} = \sum_{j=1}^M \frac{a_{ij}y^j}{\sum_{k=1}^M a_{ik}y^k} a_{ij}\tilde{v}^i_z = \tilde{v}^i_z$*. Hence the calculation is reduced to that for Theorem 25.*

3.7 Comparing Games C_p, C_d and C

In this section, we compare the games $\boldsymbol{C_p}$, $\boldsymbol{C_d}$ and $\boldsymbol{C}$. We will first compare their game values and discuss their economic implications. We will then discuss their difference in terms of the NEP. Finally, we discuss their perspective NEs in the framework of controlled rank-dependent SDEs.

To make the games comparable, let us assume $y = \sum_{j=1}^N y^j$. Let us also consider a special sharing game $\boldsymbol{C_s}$ which can be connected with both $\boldsymbol{C_d}$ and $\boldsymbol{C_p}$:

$\boldsymbol{C_s}$**:** $M = N$ and $a_{ii} = 1$ for $i = 1, 2, \cdots, N$.

Pooling, Dividing, and Sharing

Denote the game value and waiting region for each player i as $v^i_{C_p}$ and $\mathcal{W}_i^{C_p}$ respectively for game $\boldsymbol{C_p}$. Similar notations are defined for $\boldsymbol{C_d}$ and $\boldsymbol{C_s}$.

Comparing game values.

Proposition 28 (Game values comparison)**.** *Assume* ***H1′-H2′****. For each* $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^N \times \mathbb{R}^N_+$*, if* $(\boldsymbol{x}, y) \in \mathcal{W}_i^{C_p}$*, and* $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{W}_i^{C_d} \cap \mathcal{W}_i^{C_s}$*, then,*

$$v^i_{C_p}(\boldsymbol{x}, y) \le v^i_{C_s}(\boldsymbol{x}, y) \le v^i_{C_d}(\boldsymbol{x},\boldsymbol{y}),\ i = 1, 2, \cdots, N.$$

Proof. The comparison is by direct computation. Indeed, recall that in case $\boldsymbol{C_p}$, when $(\boldsymbol{x}, y) \in \mathcal{W}_i^{C_p}$,

$$v^i_{C_p}(\boldsymbol{x}, y) = p_N(\widetilde{x}^i) + A_N(y)\cosh\left(\widetilde{x}^i\sqrt{\frac{2(N-1)\alpha}{N}}\right),$$

for $i = 1, 2, \cdots, N$, where $\widetilde{x}^i$ is defined in (3.4.2) and A_N is defined in (3.4.8).

Similarly, in case $\boldsymbol{C_d}$, when $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{W}_i^{C_d}$,

$$v^i_{C_d}(\boldsymbol{x},\boldsymbol{y}) = p_N(\widetilde{x}^i) + A_N(y^i)\cosh\left(\widetilde{x}^i\sqrt{\frac{2(N-1)\alpha}{N}}\right),$$

for each $i = 1, 2, \cdots, N$. And, in case $\boldsymbol{C_s}$, when $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{W}_i^{C_s}$,

$$v^i_{C_s}(\boldsymbol{x},\boldsymbol{y}) = p_N(\widetilde{x}^i) + A_N\left(\sum_{j=1}^N a_{ij}y^j\right)\cosh\left(\widetilde{x}^i\sqrt{\frac{2(N-1)\alpha}{N}}\right),$$

for each $i = 1, 2, \cdots, N$. By elementary calculations,

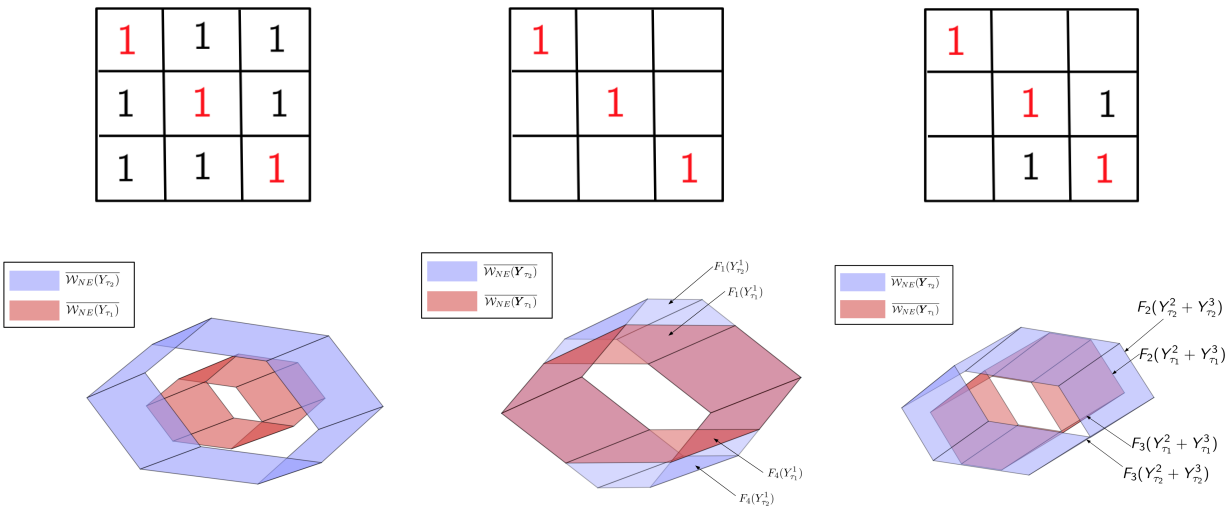
$$A'_N(y) < 0.$$

Therefore, when $y = \sum_{j=1}^N y^j$, $(\boldsymbol{x}, y) \in \mathcal{W}_i^{C_p}$, and $(\boldsymbol{x},\boldsymbol{y}) \in \mathcal{W}_i^{C_d} \cap \mathcal{W}_i^{C_s}$,

$$v^i_{C_p}(\boldsymbol{x}, y) \le v^i_{C_s}(\boldsymbol{x}, y) \le v^i_{C_d}(\boldsymbol{x},\boldsymbol{y}).$$

The first inequality holds because $y = \sum_{i=1}^N y^i \ge \sum_{i=1}^N a_{ij}y^j$ and the equality holds if and only if $a_{ij} = 1$ for each $j = 1, 2, \cdots, N$. The second inequality holds because $a_{ii} = 1$ and the equality holds if and only if $a_{ij} = 0$ for each $j \ne i$. □

This result has a clear economic interpretation. In a stochastic game where players have the options to share resources, versus the possibility to divide resources in advance, sharing will have lower cost than dividing. Pooling yields the lowest cost for each player.



(a) $\boldsymbol{C_p}$ (b) $\boldsymbol{C_d}$ (c) $\boldsymbol{C}$

Figure 3.4: Comparison of projected evolving boundaries for $\boldsymbol{C_p}$, $\boldsymbol{C_d}$, $\boldsymbol{C}$ when $N = 3$.

Define the projected common waiting region

$$\mathcal{W}_{NE}(\boldsymbol{y}) := \left\{ \boldsymbol{x} \in \mathbb{R}^N : |\widetilde{x}^i| < f_N^{-1}\left(\sum_{j=1}^{M} a_{ij} y^j\right) \text{ for } 1 \le i \le N \right\},$$

for any fixed resource level $\boldsymbol{y}$. Then $\mathcal{W}_{NE}(\boldsymbol{y})$ is a polyhedron with $2N$ boundary faces. Figure 3.4a shows a pooling game $\boldsymbol{C_p}$. After one player exercises controls, all the faces of the boundary move. Figure 3.4b corresponds to a dividing game $\boldsymbol{C_d}$. After player i exercises controls, her faces of F_i and F_{i+N} move. Here $i = 1, N = 3$. For a sharing game $\boldsymbol{C}$, shown in Figure 3.4c, after one player exercises her controls, the faces of the players who are connected with her will move, while the faces for other players remain unchanged. Here $i = 2$ and player 2 and 3 are connected.

NEs for the games and controlled rank-dependent SDEs

In the previous sections, the controlled dynamics is constructed directly via the reflected Brownian motion. This class of SDEs can also be cast in the framework of rank-dependent SDEs. Indeed, the controlled dynamics of NE in the action regions of the N-player can be written as a *controlled rank-dependent SDEs*:

$$\begin{aligned} dX_t^i &= \sum_{j=1}^{N} 1_{F^i(\boldsymbol{X}_t,\boldsymbol{Y}_t)=F^{(j)}(\boldsymbol{X}_t,\boldsymbol{Y}_t)} \left(\delta_j dt + \sigma_j dB_t^j + d\xi_t^{j,+} - d\xi_t^{j,-}\right), \\ dY_t^j &= -\sum_{i=1}^{N} \frac{a_{ij} Y_{s-}^j}{\sum_{j=1}^{M} a_{ij} Y_{s-}^j} d\check{\xi}_s^i, \end{aligned}$$

with $(\xi^{i,+}, \xi^{i,-})$ the controls, $F^i : \mathbb{R}^N \times \mathbb{R}_+^M \to \mathbb{R}$ a rank function depending on both $\boldsymbol{X}$ and $\boldsymbol{Y}$, $F^{(1)} \leq \cdots \leq F^{(N)}$ the order statistics of $(F^i)_{1\leq i\leq N}$, and $\delta_i \in \mathbb{R}$, $\sigma_i \geq 0$.

In game $\boldsymbol{C_p}$, the controlled dynamics in the action regions satisfies the SDEs with $F^i_{C_p}(\boldsymbol{x}, \boldsymbol{y}) = |x^i - \frac{\sum_{j\neq i} x^j}{N-1}|$, $\delta_i = 0$ and $\sigma_i = 0$ for each $i = 1, \cdots N$, and

$$\xi^{i,\pm} = 0 \quad \text{for each } i = 1, \cdots, N-1 \quad \text{and} \quad \xi^{N,\pm} \neq 0.$$

In game $\boldsymbol{C_d}$,

$$F^i_{C_d}(\boldsymbol{x}, \boldsymbol{y}) = \left| x^i - \frac{\sum_{j\neq i} x^j}{N-1} - f_N^{-1}(y^i) \right|.$$

For the general game $\boldsymbol{C}$, the controlled process in the action regions is governed by the rank-dependent dynamics with $F^i_C(\boldsymbol{x}, \boldsymbol{y}) = |x^i - \frac{\sum_{j\neq i} x^j}{N-1} - f_N^{-1}(\sum_{j=1}^M a_{ij} y^j)|$ where f_N is a threshold function defined in (3.4.7)-(3.4.9), and δ_i, σ_i and $\xi^{i,\pm}$ satisfy the same condition as before.

Note that the special case without controls, i.e., $F^i(\boldsymbol{x}, \boldsymbol{y}) = x^i$ and $\xi^{i,\pm} = 0$, corresponds to the *rank-dependent SDEs*. In particular, the rank-dependent SDEs with $\delta_1 = 1$, $\delta_2 = \cdots \delta_N = 0$ is known as the *Atlas model*. To the best of our knowledge, rank-dependent SDEs with additional controls or a general rank function F^i has not been studied before. There are various aspects including uniqueness and sample path properties that await further investigation and we leave them to interested readers.

Chapter 4

Pareto Optimality and Price of Anarchy

4.1 Pareto Optimality (PO)

In this section, we introduce a class of N-player game, the definition of Pareto optimality and its connection to a auxiliary central controller problem.

Mathematical Formulation

Let us first define the N-player game.

Controlled dynamics. Let $(X_t^i)_{t\geq 0} \in \mathbb{R}$ denote the location of player $i, 1 \leq i \leq N$. In the absence of controls, $\boldsymbol{X}_t = (X_t^1, \ldots, X_t^N) \in \mathbb{R}^N$ follows a stochastic differential equation (SDE):

$$d\boldsymbol{X}_t = \boldsymbol{\mu} dt + \boldsymbol{\sigma} d\boldsymbol{B}_t, \quad \boldsymbol{X}_0 = (x^1, \ldots, x^N), \tag{4.1.1}$$

where $\boldsymbol{B} := (B^1, \ldots, B^N) \in \mathbb{R}^N$ is a standard N-dimensional Brownian motion in a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$, with a drift $\boldsymbol{\mu} := (\mu_1, \ldots, \mu_N)$ and a covariance matrix $\boldsymbol{\sigma} := (\sigma_{ij})_{1\leq i,j\leq N}$. Here μ_i, σ_{ij} are constants.

If player i applies controls (of a finite variation type) ξ_t^i to X_t^i, then X_t^i evolves as

$$dX_t^i = \mu^i dt + \boldsymbol{\sigma}^i \cdot d\boldsymbol{B}_t + d\xi_t^i, \quad X_{0-}^i = x^i, \quad i = 1, \ldots, N,$$

where $\boldsymbol{\sigma}_i$ is the i^{th} row of the covariance matrix $\boldsymbol{\sigma}$. Assume the diffusion matrix $\boldsymbol{\sigma}\boldsymbol{\sigma}^T$ is positive definite and there exists $a > 0$ such that $\boldsymbol{\sigma}\boldsymbol{\sigma}^T > aI$, where $I \in \mathbb{R}^{N\times N}$ is the identity matrix.

Denoting the pair of non-decreasing and càdlàg processes (ξ^{i+}, ξ^{i-}) as the minimum decomposition of the finite variational process $\xi^i := (\xi_t^i)_{t\geq 0}$ such that $\xi^i := \xi^{i+} - \xi^{i-}$, then the above controlled dynamics can be written as

$$dX_t^i = \mu^i dt + \boldsymbol{\sigma}^i \cdot d\boldsymbol{B}_t + d\xi_t^{i,+} - d\xi_t^{i,-}, \quad X_{0-}^i = x^i, \tag{4.1.2}$$

Note that the non-decreasing and càdlàg processes ξ^{i+} and ξ^{i-} can be further decomposed in a differential form,

$$d\xi_t^{i\pm} = d(\xi_t^{i\pm})^c + \Delta\xi_t^{i\pm},$$

with $d(\xi_t^{i\pm})^c$ the continuous and $\Delta\xi_t^{i\pm} := \xi_t^{i\pm} - \xi_{t-}^{i\pm}$ the jump part of $d\xi_t^{i\pm}$.

Game objective. The game is for player i to minimize, among all (ξ^{i+}, ξ^{i-}) from an appropriate admissible control set $\mathcal{U}_N^i$ (to be specified below), over an infinite time horizon, the following objective function,

$$J^i(\boldsymbol{x};\boldsymbol{\xi}) = \mathbb{E}\int_0^\infty e^{-\alpha t}\left[h^i(\boldsymbol{X}_t)dt + K_i^+ d\xi_t^{i,+} + K_i^- d\xi_t^{i,-}\right], \qquad \textbf{(N-player)}$$

Here $\alpha > 0$ is a constant discount factor.

In this game, players interact through their respective objective functions $h^i(\boldsymbol{x}) : \mathbb{R}^N \to \mathbb{R}^+$. For example, $h^i(\boldsymbol{x}) = h\left(x^i - \rho\frac{\sum_{j=1}^N x^j}{N}\right)$, with h a general distance function and $\rho \in [0,1]$, is a game where players aim to stay as close as possible to each other during the game.

Admissible control $\mathcal{U}_N^i$. The admissible control set for player i is of a Markovian type and defined as

$$\begin{aligned}\mathcal{U}_N^i = \Big\{(\xi_t^{i,+}, \xi_t^{i,-}) \mid \xi_t^{i,+} \text{ and } \xi_t^{i,-} \text{ are } \mathcal{F}_{t-}^{(X^1,\ldots,X^N)}\text{-progressively measurable, càdlàg non-decreasing,} \\ \text{with } \mathbb{E}\left[\int_0^\infty e^{-\alpha t}d\xi_t^{i,+}\right] < \infty, \mathbb{E}\left[\int_0^\infty e^{-\alpha t}d\xi_t^{i,-}\right] < \infty, \xi_{0-}^{i,+} = 0, \xi_{0-}^{i,-} = 0\Big\},\end{aligned} \tag{4.1.3}$$

with $\mathcal{F}^{\boldsymbol{X}_{t-}} := \sigma\left(\cup_{s<t}\mathcal{F}^{\boldsymbol{X}_s}\right)$ the filtration generated by $\boldsymbol{X}$ up to time $t-$.

PO and the Auxiliary Central Controller Problem

In this Section, we will analyze the game (**N-player**) in the sense of PO. Recall that

Definition 29 (PO)**.** *Given game* (**N-player**)*, $\boldsymbol{\xi}^* \in \mathcal{U}_N := (\mathcal{U}_N^1, \cdots, \mathcal{U}_N^N)$ with pay-off functions $\left(J^1(\boldsymbol{x};\boldsymbol{\xi}^*), \ldots, J^N(\boldsymbol{x};\boldsymbol{\xi}^*)\right)$ is a PO if and only if there does not exist $\boldsymbol{\xi} \in \mathcal{U}_N$ such that*

$$J^i(\boldsymbol{x};\boldsymbol{\xi}) \le J^i(\boldsymbol{x};\boldsymbol{\xi}^*) \text{ for all } i = 1, \ldots, N,$$

and

$$J^j(\boldsymbol{x};\boldsymbol{\xi}) < J^j(\boldsymbol{x};\boldsymbol{\xi}^*),$$

for some $j \in \{1, \ldots, N\}$. Here the strategies ξ^{i*} and ξ^i are deterministic functions of time t and $\boldsymbol{X}_t$ (with $\boldsymbol{X}_{0-} = \boldsymbol{x}$) for all $i = 1, 2, \cdots, N$.

We will derive PO by analyzing an associated N-dimensional stochastic control problem, called the central controller problem.

Mathematically, the central controller problem is the following minimization problem

$$v(\boldsymbol{x}) = \min_{\boldsymbol{\xi}\in\mathcal{U}_N} J(\boldsymbol{x};\boldsymbol{\xi}), \tag{4.1.4}$$

with $\boldsymbol{\xi} \in \mathcal{U}_N$, subject to the dynamics (4.1.2) with the pay-off function $J(\boldsymbol{x};\boldsymbol{\xi})$ defined as the weighted average pay-off function of all players such that

$$\begin{aligned} J(\boldsymbol{x};\boldsymbol{\xi}) &= \sum_{i=1}^{N} \alpha^i J^i(\boldsymbol{x},\boldsymbol{\xi}) \\ &= \mathbb{E}\int_0^\infty e^{-\alpha t}\left[H(\boldsymbol{X}_t)dt + \sum_{i=1}^{N}\alpha^i K_i^+ d\xi_t^{i,+} + \sum_{i=1}^{N}\alpha^i K_i^- d\xi_t^{i,-}\right]. \end{aligned}$$

Here $H(\boldsymbol{x}) = \sum_{i=1}^N \alpha_i h^i(\boldsymbol{x})$ is the weighted running cost with $\alpha_i > 0$ and $\sum_{i=1}^N \alpha_i = 1$. Note that when $\alpha_i = \frac{1}{N}$ $(i = 1, 2, \cdots, N)$, the central controller treats all players equally, often adopted in the social welfare optimization problem.

This type of control problem (4.1.4) has been analyzed in [147] where the optimal control is shown to be the limit of a sequence of control problems of bounded velocity with an increasing upper bound on the velocity. In this paper, we will study the regularity of the value function and the property of optimal control. This property of optimal control is critical fr analyzing the property of PO and its comparison with NEs in the context of Price of Anarchy (Section 4.2). The regularity of a similar yet simpler N-dimensional control problem has been studied in [128]. The gradient constraint in their problem is $\nabla v(\boldsymbol{x})$, which is easier to analyze compared to our case (defined in (4.1.7)).

Firstly, we have

Theorem 30. *The optimal control of problem (4.1.4) is a PO to the game (**N-player**).*

Proof. Given the payoff function J^i defined as in (**N-player**), $v(\boldsymbol{x})$ the value function of the "central controller" which is defined as in (4.1.4), and $\boldsymbol{\xi}^* := (\xi^{1*}, \ldots, \xi^{N*})$ the optimal control to problem (4.1.4). Then for any $\boldsymbol{\xi} := (\xi^1, \ldots, \xi^N) \in \mathcal{U}_N$,

$$\sum_{i=1}^{N} \alpha^i J^i(\boldsymbol{x};\boldsymbol{\xi}) \geq v(\boldsymbol{x}), \tag{4.1.5}$$

where value $v(\boldsymbol{x})$ is reached when player i takes the control ξ_t^{i*} $(i = 1, 2, \ldots, N)$.

If there is another $\boldsymbol{\xi'} := (\xi^{1'}, \ldots, \xi^{N'}) \in \mathcal{U}_N$ and $k \in \{1, \ldots, N\}$ such that

$$J^k(\boldsymbol{x}; \xi^{1'}, \ldots, \xi^{N'}) < J^k(\boldsymbol{x}; \xi^{1*}, \ldots, \xi^{N*}),$$

then given $\alpha_i > 0$ for all i, there must exists $j \in \{1, \ldots, N\}$ such that

$$J^j(\boldsymbol{x}; \xi^{1'}, \ldots, \xi^{N'}) > J^j(\boldsymbol{x}; \xi^{1*}, \ldots, \xi^{N*}).$$

Hence the control $\boldsymbol{\xi}^*$ is a PO by definition.

□

Next, we establish the existence of PO, with some technical assumptions which ensure the well-definedness of the game (**N-player**).

Assumptions. There exist $C > c > 0$ such that $H(\boldsymbol{x})$ and $h^i(\boldsymbol{x})$ $(i = 1, 2, \cdots, N)$ satisfy the following conditions.

A1. $\forall \boldsymbol{x} \in \mathbb{R}^N$, $0 \leq H(\boldsymbol{x}) \leq C(1 + \|\boldsymbol{x}\|^2)$.

A2. $\forall \boldsymbol{x}, \boldsymbol{x'} \in \mathbb{R}^N$, $|H(\boldsymbol{x}) - H(\boldsymbol{x'})| \leq C(1 + \|\boldsymbol{x}\| + \|\boldsymbol{x'}\|)\|\boldsymbol{x} - \boldsymbol{x'}\|$.

A3. $H(\boldsymbol{x}) \in \mathcal{C}^{2,1}(\mathbb{R}^N)$, H is convex, with $0 < c \leq \frac{\partial^2 H(\boldsymbol{x})}{\partial z^2} \leq C$ for all unit direction $z \in \mathbb{R}^N$.

Besides **A1**-**A3**, We need another two assumptions to insure the existence of a unique PO solution. Assumption **A4** (specified later) describes the regularity of central controller's value function. Assumption **A5** (specified later) describes the existence of a Lipchitz mapping for initial jumps. Now,

Theorem 31 (PO Solution)**.** *Under Assumptions **A1**-**A5**, there exists a unique solution to the N-player game (**N-player**) and the dynamics under optimal control is a solution to a Skorokhod problem. Moreover, fix any weight $\boldsymbol{\alpha} \in \{\boldsymbol{b} \mid \boldsymbol{b} \in \mathbb{R}^N_{++}$ and $\sum_{i=1}^N b_i = 1\}$, the optimal control to the N-player game (**N-player**) will provide a PO solution. The set of POs forms a Pareto frontier parameterized by $\boldsymbol{\alpha}$.*

Now we provide the proof of Theorem 31.

Derivation of PO via Analyzing the Central Controller Problem.

To prove Theorem 31, it suffices to establish the existence and uniqueness of solution to the central control problem (4.1.4), along with some characterizations of the optimal policy. We will first establish the regularity properties of the value function $v(\boldsymbol{x})$ in Section 4.1. We will then establish the optimal control associated $v(\boldsymbol{x})$ in Sections 4.1 and 4.1.

Regularities analysis

First, by the Dynamic Programming Principle (DPP), the Hamilton-Jacobi-Bellman (HJB) equation associated with (4.1.4) is

$$\max\{\alpha u - \mathcal{L}u - H(\boldsymbol{x}), \beta(\nabla u) - 1\} = 0, \tag{4.1.6}$$

with the operator

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^N \boldsymbol{\sigma}^i \cdot \boldsymbol{\sigma}^j \frac{\partial^2}{\partial x^i \partial x^j} + \sum_{i=1}^N \mu^i \frac{\partial}{\partial x^i},$$

and

$$\beta(\boldsymbol{q}) = \max_{1 \leq i \leq N} \left[\left(\frac{q^i}{K_i^+}\right)^+ \vee \left(\frac{q^i}{K_i^-}\right)^- \right], \tag{4.1.7}$$

where $\boldsymbol{q} := (q^1, \cdots, q^N)$, $(a)^+ = \max\{0, a\}$ and $(a)^- = \max\{0, -a\}$ for any $a \in \mathbb{R}$. Note that operator β is well defined as $K_i^{\pm} > 0$ for all $i = 1, 2, \cdots, N$.

Next, define the waiting region $\mathcal{C}$ as

$$\mathcal{C} = \{\boldsymbol{x} \mid \beta(\nabla v(\boldsymbol{x})) < 1\}. \tag{4.1.8}$$

Clearly, under Assumptions **A3**, $\mathcal{C}$ is bounded. Moreover,

Proposition 32. *Under Assumptions **A1**-**A3**, $v(\boldsymbol{x}) \in C^{4,\alpha}(\mathcal{C})$ and $v(\boldsymbol{x})$ is strictly convex in $\mathcal{C}$.*

Proof. Let B be any open ball such that $\bar{B} \in \mathcal{C}$. By Theorem 6.13 in [86], the Dirichlet problem in B,

$$\begin{cases} \alpha\tilde{v} - \mathcal{L}\tilde{v} = H(\boldsymbol{x}), & \forall x \in B, \\ \tilde{v} = v, & \forall x \in \partial B, \end{cases} \tag{4.1.9}$$

has a solution $\tilde{v} \in \mathcal{C}^0(\bar{B}) \cap \mathcal{C}^{2,\alpha}(B)$. In particular, $\tilde{v} - v \in \mathcal{W}^{2,\infty}(B)$, therefore by (4.1.9), $\tilde{v} - v \in \mathcal{W}_0^{1,2}(B)$. By Theorem 8.9 of [86], $v = \tilde{v}$ in B, thus $v \in \mathcal{C}^{2,\alpha}(B)$. By Theorem 6.17 of [86], $v \in \mathcal{C}^{4,\alpha}(B)$ thus $v \in \mathcal{C}^{4,\alpha}(\mathcal{C})$ for all $\alpha \in (0, 1)$. □

Theorem 33 (Regularity of $v(\boldsymbol{x})$)**.** *Under Assumptions **A1**- **A3**, the value function $v(\boldsymbol{x})$ to the control problem is the unique $\mathcal{W}_{loc}^{2,\infty}(\mathbb{R}^N)$ solution to (4.1.6). Moreover, there exists $K > 0$ such that*

(i) $0 \le v(\boldsymbol{x}) \le K(1 + \|\boldsymbol{x}\|^2)$, $\forall \boldsymbol{x} \in \mathbb{R}^N$,

(ii) $|v(\boldsymbol{x}) - v(\boldsymbol{x}')| \le K(1 + \|\boldsymbol{x}\| + \|\boldsymbol{x}'\|)\|\boldsymbol{x} - \boldsymbol{x}'\|$, $\forall \boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^N$,

(iii) $0 \le \frac{\partial^2}{\partial z^2} v(\boldsymbol{x}) \le K$ *for any second order directional derivative* $\frac{\partial^2}{\partial z^2}$.

Remark 33.1. *Note that this regularity property $v(x) \in \mathcal{W}_{loc}^{2,\infty}(\mathbb{R}^N)$ is essential for the existence and uniqueness of the optimal control, indeed, it is needed for the Skorokhod solution.*

Proof. We will prove Theorem 33 in five steps. For simplicity, K and k will be used for generic positive constants which may represent different constants for different estimates.

Step (i). First, $v(\boldsymbol{x}) \ge 0$ is clear by the non-negativity of $H(\boldsymbol{x})$. Moreover, by the property that $\boldsymbol{\sigma}\boldsymbol{\sigma}^T \ge a\ \mathbf{I}$, it follows from a known estimate and martingale argument [146, (2.15)] that the solution $\{\tilde{\boldsymbol{X}}_t\}_{t\ge 0} := \{\boldsymbol{x} + \boldsymbol{\mu}t + \boldsymbol{\sigma}\boldsymbol{B}_t\}_{t\ge 0}$ with $\boldsymbol{\xi} = \mathbf{0}$ satisfying

$$\mathbb{E}\int_0^\infty e^{-\alpha t}\|\tilde{\boldsymbol{X}}_t\|^2 dt \le K(1 + \|\boldsymbol{x}\|^2), \quad \forall \boldsymbol{x} \in \mathbb{R}^N$$

for some constant $K > 0$. By Assumption **A2**, there exists some constant $K > 0$ such that

$$v(\boldsymbol{x}) \le J(\boldsymbol{x}, \mathbf{0}) \le K(1 + \|\boldsymbol{x}\|^2), \forall \boldsymbol{x} \in \mathbb{R}^N.$$

Thus (i) is established.

Step (ii). For each fixed $\boldsymbol{x} \in \mathbb{R}^N$, let

$$\mathcal{U}_{\boldsymbol{x}} = \{\boldsymbol{\xi} \in \mathcal{U} : J(\boldsymbol{x}, \boldsymbol{\xi}) \leq J(\boldsymbol{x}; \mathbf{0})\}. \tag{4.1.10}$$

By Assumption **A2**,

$$\mathbb{E} \int_0^\infty e^{-\alpha t} \|\boldsymbol{X}_t\|^2 dt \leq K(1 + \|\boldsymbol{x}\|^2), \qquad \forall \boldsymbol{x} \in \mathbb{R}^N, \boldsymbol{\xi} \in \mathcal{U}_{\boldsymbol{x}}. \tag{4.1.11}$$

For $\boldsymbol{\xi} \in \mathcal{U}_{\boldsymbol{x}}$, it is easy to verify that

$$\mathbb{E} \int_0^\infty e^{-\alpha t} \|\boldsymbol{\xi}_t\|^2 dt \leq K(1 + \|\boldsymbol{x}\|^2), \tag{4.1.12}$$

and

$$|v(\boldsymbol{x}) - v(\boldsymbol{x}')| \leq \sup \left\{ |J(\boldsymbol{x}; \boldsymbol{\xi}) - J(\boldsymbol{x}'; \boldsymbol{\xi})| : \boldsymbol{\xi} \in \mathcal{U}_{\boldsymbol{x}} \cup \mathcal{U}_{\boldsymbol{x}'} \right\}, \forall \boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^N.$$

Meanwhile,

$$|J(\boldsymbol{x}; \boldsymbol{\xi}) - J(\boldsymbol{x}'; \boldsymbol{\xi})| \leq \mathbb{E} \int_0^\infty e^{-\alpha t} |H(\boldsymbol{X}_t^{\boldsymbol{x}}) - H(\boldsymbol{X}_t^{\boldsymbol{x}'})| dt.$$

Statement (ii) for v follows from this by using Assumption **A3**, the facts that $\boldsymbol{X}_t^{\boldsymbol{x}} - \boldsymbol{X}_t^{\boldsymbol{x}'} = \boldsymbol{x} - \boldsymbol{x}'$, and that for any $\boldsymbol{\xi} \in \mathcal{U}_{\boldsymbol{x}} \cup \mathcal{U}_{\boldsymbol{x}'}$,

$$\begin{aligned} \mathbb{E} \int_0^\infty e^{-\alpha t} \|\boldsymbol{X}_t^{\boldsymbol{x}}\| dt &\leq K(1 + \|\boldsymbol{x}\| + \|\boldsymbol{x}'\|), \\ \mathbb{E} \int_0^\infty e^{-\alpha t} \|\boldsymbol{X}_t^{\boldsymbol{x}'}\| dt &\leq K(1 + \|\boldsymbol{x}\| + \|\boldsymbol{x}'\|). \end{aligned} \tag{4.1.13}$$

In fact, if $\boldsymbol{\xi} \in \mathcal{U}_{\boldsymbol{x}}$, (4.1.13) follows immediately from (4.1.12) by the Hölder inequality. Meanwhile, if $\boldsymbol{\xi} \in \mathcal{U}_{\boldsymbol{x}'}$, (4.1.13) holds beacause

$$\|\boldsymbol{X}_t^{\boldsymbol{x}}\| \leq \|\boldsymbol{X}_t^{\boldsymbol{x}'}\| + \|\boldsymbol{x} - \boldsymbol{x}'\| \leq \|\boldsymbol{X}_t^{\boldsymbol{x}'}\| + \|\boldsymbol{x}\| + \|\boldsymbol{x}'\|.$$

Step (iii). For $i = 1, 2, \cdots, N$, let $\Delta_i \boldsymbol{x} := (0, \cdots, 0, \Delta x^i, 0, \cdots, 0)$ be the N-dimensional row vector with the i-th entry being Δx^i. For any function $F : \mathbb{R}^N \to \mathbb{R}$, define the second difference of F in the x^i direction by

$$\delta_i^2 F(\boldsymbol{x}) = F(\boldsymbol{x} + \Delta_i \boldsymbol{x}) + F(\boldsymbol{x} - \Delta_i \boldsymbol{x}) - 2F(\boldsymbol{x}). \tag{4.1.14}$$

It is easy to check that

$$\delta_i^2 v(\boldsymbol{x}) \leq \sup\{\delta_i^2 J(\boldsymbol{x}; \boldsymbol{\xi}) : \boldsymbol{\xi} \in \mathcal{U}_{\boldsymbol{x}}\}. \tag{4.1.15}$$

Since $H \in \mathcal{C}^2(\mathbb{R}^N)$, for $\boldsymbol{x} \in \mathbb{R}^N$,

$$\delta_i^2 H(\boldsymbol{x}) = (\Delta x^i)^2 \int_0^1 \int_{-\lambda}^{\lambda} \frac{\partial^2 H}{\partial (x^i)^2}(x^1, \ldots, x^i + \mu \Delta x^i, \ldots, x^N) d\mu d\lambda. \tag{4.1.16}$$

By Assumption **A3**,

$$\delta_i^2 H(\boldsymbol{x}) \le K(\Delta x^i)^2 \int_0^1 \int_{-\lambda}^{\lambda} d\mu d\lambda = (\Delta x^i)^2 K. \tag{4.1.17}$$

Hence

$$0 \le \delta_i^2 v(\boldsymbol{x}) \le K(\Delta x^i)^2,\ \boldsymbol{x} \in \mathbb{R}^N, |\Delta x^i| \le 1. \tag{4.1.18}$$

To prove the lower bound of (4.1.18), it suffices to prove the convexity of v, which follows from the joint convexity of $J(\boldsymbol{x};\boldsymbol{\xi})$ in the following sense:

$$J(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{x}') \le \theta J(\boldsymbol{x};\boldsymbol{\xi}) + (1-\theta)J(\boldsymbol{x}';\boldsymbol{\xi}'), \tag{4.1.19}$$

for any $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^N$ and any $\boldsymbol{\xi}, \boldsymbol{\xi}' \in \mathcal{U}$. The convexity of J in $(\boldsymbol{x};\boldsymbol{\xi})$ is then obvious since $\boldsymbol{X}_t^{\boldsymbol{x}}$ depends linearly on $(\boldsymbol{x},\boldsymbol{\xi})$ and the set $\mathcal{U}$ and the function H are both convex.

Step (iv). To prove $v \in \mathcal{W}_{loc}^{2,\infty}$, let B be any open ball and let $\psi \in C_0^\infty(\mathbb{R}^N)$ be any test function with a support contained in B. Since $(\Delta x^i)^{-2}\delta_i^2 v(\boldsymbol{x})$ is bounded on B for $|\Delta x^i| \le 1$, there is a sequence $\eta_k \to 0+$ as $k \to \infty$ such that, denoting by g_k the result of replacing Δx^i by η_k in $(\Delta x^i)^{-2}\delta_i^2 v(\boldsymbol{x})$, we have $g_k \to Q$ weakly in $L^p(B)$ for some p with $1 < p < \infty$. It is then easy to see that

$$\int_{\mathbb{R}^N} \psi(\boldsymbol{x}) Q(\boldsymbol{x}) d\boldsymbol{x} = \int_{\mathbb{R}^N} \frac{\partial^2 \psi}{\partial x^i x^i} v(\boldsymbol{x}) d\boldsymbol{x}, \quad \forall \psi \in \mathcal{C}_0^\infty(B). \tag{4.1.20}$$

Here $Q = \frac{\partial^2 v}{\partial x^i x^i}$ is the generalized derivative. The existence and local boundedness of mixed second order generalized derivatives are now immediate: for $k = 1, 2, \ldots, N$, let $\boldsymbol{e}_k$ denote the unit vector in the direction of the positive x_k axis. For any fixed $i \ne j$ with $1 \le i, j \le N$, let $\boldsymbol{y}$ be a new coordinate whose axis points in the $\frac{\boldsymbol{e}_i + \boldsymbol{e}_j}{\sqrt{2}}$ direction, then $\frac{\partial^2 v}{\partial x^i \partial x^j} = \frac{\partial^2 v}{\partial \boldsymbol{y}^2} - \frac{1}{2}(\frac{\partial^2 v}{\partial x^i x^i} + \frac{\partial^2 v}{\partial x^j x^j})$.

Step (v). To show that v is the unique solution to HJB, we proceed by a contradiction argument. Suppose v_1 and v_2 are two non-negative solutions. Let y_0 be the point where v_2 attains its minimum value. Given $\delta > 0$, define

$$\phi_\delta(\boldsymbol{x}) := v_1(\boldsymbol{x}) - v_2(\boldsymbol{x}) - \delta\|\boldsymbol{x} - \boldsymbol{y}_0\|^2, \qquad \forall \boldsymbol{x} \in \mathbb{R}^N.$$

The function ϕ_δ attains its maximum at some $x_\delta \in \mathbb{R}^N$ and

$$0 = \nabla\phi_\delta(\boldsymbol{x}_\delta) = \nabla v_1(\boldsymbol{x}) - \nabla v_2(\boldsymbol{x}) - 2\delta(\boldsymbol{x}_\delta - \boldsymbol{y}_0). \tag{4.1.21}$$

This leads to

$$\nabla v_1(\boldsymbol{x}_\delta) = \nabla v_2(\boldsymbol{x}_\delta) + 2\delta(\boldsymbol{x}_\delta - \boldsymbol{y}_0).$$

Consequently,

$$1 \ge \beta(\nabla v_1(\boldsymbol{x}_\delta)) = \beta(\nabla v_2(\boldsymbol{x}_\delta) + 2\delta(\boldsymbol{x}_\delta - \boldsymbol{y}_0)).$$

Since y_0 is the minimal point of v_2, we have

$$\nabla v_2(\boldsymbol{x}_\delta) \cdot (\boldsymbol{x}_\delta - \boldsymbol{y}_0) \geq 0.$$

This means that either $\beta(\Delta v_2(\boldsymbol{x}_\delta)) < 1$, or for any $i \in \arg\max \beta(\nabla v_2(\boldsymbol{x}_\delta))$, $(\boldsymbol{x}_\delta - \boldsymbol{y}_0)^i = 0$. Suppose the latter, then by (4.1.21), we have

$$0 = D_i v_1(\boldsymbol{x}_\delta) - D_i v_2(\boldsymbol{x}_\delta) - 2\delta(\boldsymbol{x}_\delta - \boldsymbol{y}_0)^i.$$

Hence

$$D_i v_1(\boldsymbol{x}_\delta) = D_i v_2(\boldsymbol{x}_\delta) = 0,$$

for $i \in \arg\max \beta(\nabla v_2(\boldsymbol{x}_\delta))$. This implies $\beta(\Delta v_2(\boldsymbol{x}_\delta)) < 1$. Meanwhile from (4.1.6), we know

$$\Delta v_2(\boldsymbol{x}_\delta) = v_2(\boldsymbol{x}_\delta) - H(\boldsymbol{x}_\delta).$$

By Bony's maximum principle ([142]),

$$\begin{aligned} 0 &\geq \liminf \operatorname{ess}_{\boldsymbol{x}\to\boldsymbol{x}_\delta} \Delta\phi_\delta(\boldsymbol{x}) \\ &= \liminf \operatorname{ess}_{\boldsymbol{x}\to\boldsymbol{x}_\delta} \Delta v_1(\boldsymbol{x}) - \Delta v_2(\boldsymbol{x}) - 4\delta \\ &\geq v_1(\boldsymbol{x}_\delta) - v_2(\boldsymbol{x}_\delta) - 4\delta. \end{aligned}$$

It follows that for any $\boldsymbol{x} \in \mathbb{R}^N$,

$$v_1(\boldsymbol{x}) - v_2(\boldsymbol{x}) = \phi_\delta(\boldsymbol{x}) + \delta\|\boldsymbol{x} - \boldsymbol{y}_0\|^2 \leq \phi_\delta(\boldsymbol{x}_\delta) + \delta\|\boldsymbol{x} - \boldsymbol{y}_0\|^2 \leq \delta(4 + \|\boldsymbol{x} - \boldsymbol{y}_0\|^2).$$

Letting $\delta \to 0$, we have $v_1(\boldsymbol{x}) \leq v_2(\boldsymbol{x})$. Similarly, we have $v_2(\boldsymbol{x}) \leq v_1(\boldsymbol{x})$.

Finally given the regularity of the value function, the existence of PO strategy to (4.1.4) is straightforward according to [147]. □

PO strategy when $\boldsymbol{x} \in \overline{\mathcal{C}}$

In the floowing, we will present a complete characterization of the PO strategy. We first derive the PO strategies when the initial position $\boldsymbol{x}$ is in the closure of non-action region $\overline{\mathcal{C}}$. We then discuss the PO strategies when $\boldsymbol{x}$ is not in $\overline{\mathcal{C}}$.

When the initial position $\boldsymbol{x} \in \mathcal{C}$, the optimal control can be constructed as the limit of a sequence of ϵ-optimal policies via an appropriate Skorokhod problem with piece-wise $\mathcal{C}^1$ boundaries and the corresponding controlled dynamics are N-dimensional reflected diffusion processes on a bounded region. Recall,

Definition 34 (Skorokhod Problem)**.** *Let G be an open domain in $\mathbb{R}^N$ with $S = \partial G$. Let $\boldsymbol{x} \in \bar{G}$ and let $\boldsymbol{r}$ be a unit vector field defined on S. That is, for each $\boldsymbol{x} \in S$, $|\boldsymbol{r}(\boldsymbol{x})| = 1$, pointing inside G (in particular, nontangential to S), we say that a continuous process*

$$\boldsymbol{\xi}_t = \int_0^t \boldsymbol{N}_s d\eta_s, \tag{4.1.22}$$

with $\eta_t = \bigvee_{[0,t]} \boldsymbol{\xi}$, is a solution to a Skorokhod problem with data $(\boldsymbol{B}_t, G, \boldsymbol{r}, \boldsymbol{x})$ if

(a) *$|\boldsymbol{N}_t| = 1$, η_t is continuous and nondecreasing;*

(b) *the process $\boldsymbol{X}_t = \boldsymbol{x} + \boldsymbol{\mu}t + \boldsymbol{\sigma}\boldsymbol{\sigma}^T\boldsymbol{B}_t + \int_0^t \boldsymbol{N}_s d\eta_s$ satisfies $\boldsymbol{X}_t \in \bar{G}$, $0 \le t < \infty$, a.s;*

(c) *for every $0 \le t < \infty$,*

$$\eta_t = \int_0^t \boldsymbol{1}_{(\boldsymbol{X}_s \in \partial G, \boldsymbol{N}_s = \boldsymbol{r}(\boldsymbol{X}_s))} d\eta_s.$$

To ensure the existence of a unique PO, we assume the value function v has the following property on $\mathcal{C}$.

A4. The Hessian matrix of v, $\nabla^2 v$, is diagonal-dominated in $\mathcal{C}$:

$$v_{x^i x^i}(\boldsymbol{x}) > \left| \sum_{j \neq i} v_{x^i x^j}(\boldsymbol{x}) \right|, \forall i, = 1, 2, \cdots, N \quad \text{and} \quad \boldsymbol{x} \in \mathcal{C}. \tag{4.1.23}$$

Assumption **A4** implies that the diagonal dominates in the row/column of Hessian matrix $\nabla^2 v$. A similar assumption is used in [87, Assumption 3] to ensure the existence of a unique NE.

Theorem 35 (Optimal Policy)**.** *Assume that $\boldsymbol{x} \in \overline{\mathcal{C}}$. Under Assumptions **A1**- **A4**, the unique optimal control to problem (4.1.4) exits. The optimal control $\boldsymbol{\xi}^*$ is a solution of a Skorokhod problem such that $\boldsymbol{X}_t^* \in \overline{\mathcal{C}}$.*

Proof of Theorem 35 consists of several steps. The first step is to construct the ϵ-optimal policies by considering a piece-wise smooth $(\boldsymbol{N}^\epsilon, \boldsymbol{\xi}^\epsilon)$ to the Skorokhod problem in a regions with piece-wise $\mathcal{C}^1$ boundaries. The second step is to show the convergence of $(\boldsymbol{N}^\epsilon, \boldsymbol{\xi}^\epsilon)_{\epsilon>0}$ to the desired optimal control.

Step 1: Skorokhod problem with piece-wise smooth boundary. We first construct an approximation $\mathcal{C}_\epsilon$ of $\mathcal{C}$ that has piecewise $\mathcal{C}^1$ boundaries. Clearly, if $\partial\mathcal{C}$ itself is C^2, the $\mathcal{C}_\epsilon = \mathcal{C}$.

Let $\phi^\epsilon(x) \in C^\infty(\mathbb{R}^N, \mathbb{R}_+)$ be such that $\phi^\epsilon(\boldsymbol{x}) = 0$ for $|\boldsymbol{x}| \ge \epsilon$ and

$$\int_{\mathbb{R}^N} \phi^\epsilon(\boldsymbol{x}) d\boldsymbol{x} = 1. \tag{4.1.24}$$

Since $v(\boldsymbol{x}) \in \mathcal{W}^{2,\infty}_{loc}(\mathbb{R}^N)$, consider a regularization of $v(\boldsymbol{x})$ via ϕ^ϵ, such that

$$v^\epsilon(\boldsymbol{x}) = \phi^\epsilon * v(\boldsymbol{x}). \tag{4.1.25}$$

Because v, ∇v, $D^2 v$ are bounded on $B_R(0)$, with $\bar{\mathcal{C}} \subset B_{R-1}(0)$, thus H^ϵ, v^ϵ are bounded uniformly on $\bar{\mathcal{C}}$ for $\epsilon < 1$, and

$$v^\epsilon \to v, \quad \nabla v^\epsilon \to \nabla v, \quad H^\epsilon \to H \quad \text{uniformly in } \bar{\mathcal{C}}.$$

Therefore, for any $\epsilon_k > 0$, there exists $\delta_k > 0$ such that $\delta_k \leq \epsilon_k$ and for all $\delta \in [0, \delta_k]$, $\|\nabla v^\epsilon - \nabla v\|_{L_1} < \epsilon_k$. Take a sequence $\{\epsilon_k\}_k$ such that $\epsilon_k > 0$ and non-increasing with $\lim_{k\to\infty} \epsilon_k = 0$. Denote $w^{\delta_k}(\boldsymbol{x}) = \beta(\nabla v^{\delta_k}(\boldsymbol{x}))$ and $\mathcal{C}_{\epsilon_k} := \{\boldsymbol{x} \mid w^{\delta_k}(\boldsymbol{x}) < 1 - 2\epsilon_k\} = \cap_{j=1}^{2N} G_j^{\epsilon_k}$, where $i = 1, 2, \cdots, N$,

$$\begin{aligned} G_i^{\epsilon_k} &= \{\boldsymbol{x} \mid v_{x^i}^{\delta_k}(\boldsymbol{x}) < 1 - 2\epsilon_k\}, \\ G_{i+N}^{\epsilon_k} &= \{\boldsymbol{x} \mid v_{x^i}^{\delta_k}(\boldsymbol{x}) > -1 + 2\epsilon_k\}. \end{aligned}$$

Since $\|\nabla v^{\delta_k} - \nabla v\|_{L_1} < \epsilon_k$, we have $\mathcal{C}_{\epsilon_k} \subset \mathcal{C}$. Also notice that $\partial G_j^{\epsilon_k} \cap \overline{\mathcal{C}}_{\epsilon_k} \in \mathcal{C}^2$ because v^{δ_k} is smooth. Now, let us take any ϵ from the sequence $\{\epsilon_k\}_k$.

Define the vector field γ_j on each face G_j^ϵ as

$$\begin{aligned} \gamma_i &= -\boldsymbol{e}_i, \\ \gamma_{i+N} &= \boldsymbol{e}_i, \end{aligned}$$

for $i = 1, 2, \cdots, N$, where $\boldsymbol{e}_i = (0, \cdots, 0, 1, 0, \cdots, 0)$ with the i^{th} component to be 1. Then define the directions of reflection by

$$r_\epsilon(\boldsymbol{x}) = \left\{ \sum_{j \in I_\epsilon(\boldsymbol{x})} \alpha_j \gamma_j(\boldsymbol{x}) \;\middle|\; \alpha_i \geq 0 \;\; and \left| \sum_{j \in I_\epsilon(\boldsymbol{x})} \alpha_j \gamma_j(\boldsymbol{x}) \right| = 1 \right\}. \tag{4.1.26}$$

When $\epsilon = 0$, denote $I(\boldsymbol{x}) := I_0(\boldsymbol{x})$ and $r(\boldsymbol{x}) := r_0(\boldsymbol{x})$ for the index set and reflection cone of region $\mathcal{C}$, respectively.

Define the normal direction on face G_j^ϵ as n_j $(j = 1, 2, \cdots, 2N)$:

$$\begin{aligned} n_i &= -\frac{\nabla v_{x^i}}{\|\nabla v_{x^i}\|_2}, \\ n_{i+N} &= \frac{\nabla v_{x^i}}{\|\nabla v_{x^i}\|_2}, \qquad i = 1, 2, \cdots, N. \end{aligned}$$

n_j is well-defined under Assumption **A3**. Under Assumption **A4**, $n_i \cdot \gamma_i = \frac{v_{x^i x^i}}{\|\nabla v_{x^i}\|_2} > 0$ and $n_{i+N} \cdot \gamma_{i+N} = \frac{v_{x^i x^i}}{\|\nabla v_{x^i}\|_2} > 0$. Moreover, at each point $\boldsymbol{x} \in S_\epsilon$, there exists $\gamma \in r_\epsilon(\boldsymbol{x})$ pointing into $\mathcal{C}_\epsilon$. This is because

- There is no $\boldsymbol{x} \in \partial\mathcal{C}_\epsilon$ such that $i, i+N \in I_\epsilon(\boldsymbol{x})$ for all $i = 1, 2, \cdots, N$. This implies $|I_\epsilon(\boldsymbol{x})| \leq N$ for all $\boldsymbol{x} \in \partial\mathcal{C}_\epsilon$.
- For any $\boldsymbol{x} \in \partial\mathcal{C}_\epsilon$, there exists $\alpha_j \geq 0$ for $j \in I_\epsilon(\boldsymbol{x})$,

$$\left\langle \sum_{j \in I_\epsilon(\boldsymbol{x})} \alpha_j \gamma_j(x), n_k(x) \right\rangle > 0 \tag{4.1.27}$$

  for $k \in I_\epsilon(\boldsymbol{x})$.
- Taking $\alpha_j = 1$ for $j \in I_\epsilon(\boldsymbol{x})$, **A4** implies (4.1.27) holds.

Along with Assumption **A4**, the following condition ((3.8) in Dupuis and Ishii [72]) holds: the existence of scalars $b_j \geq 0$ $j \in I_\epsilon(\boldsymbol{x})$, such that

$$b_j \langle \gamma_j(\boldsymbol{x}), n_j(\boldsymbol{x}) \rangle > \sum_{k \in I_{\epsilon(\boldsymbol{x}) \setminus \{i\}}} b_k \left| \langle \gamma_k(\boldsymbol{x}), n_k(\boldsymbol{x}) \rangle \right|$$

Therefore, by Theorem 4.8 in [72], there exists a solution to the SP with $(\boldsymbol{X}_t, \mathcal{C}_\epsilon, r_\epsilon(\cdot), \boldsymbol{x})$.

Step 2. ϵ-optimal ploicy. Now we need to show that the solution to the Skorokhod problem with data$(\boldsymbol{x} + \boldsymbol{\mu} t + \boldsymbol{\sigma} d\boldsymbol{B}_t, \mathcal{C}_\epsilon, \boldsymbol{r}_\epsilon, \boldsymbol{x})$ is an ϵ-optimal policy of the control problem (4.1.4) with

$$\boldsymbol{\xi}_t^\epsilon = \int_0^t \boldsymbol{N}_s^\epsilon \cdot d\eta_s^\epsilon, \tag{4.1.28}$$

and $\boldsymbol{N}^\epsilon(\boldsymbol{x}) = \boldsymbol{r}^\epsilon(\boldsymbol{x})$ on $\mathcal{S}_\epsilon$. To see this, denote $\boldsymbol{X}_t^\epsilon = \boldsymbol{x} + \int_0^t \boldsymbol{\mu}(\boldsymbol{X}_s) ds + \int_0^t \boldsymbol{\sigma}(\boldsymbol{X}_s) d\boldsymbol{B}_s + \boldsymbol{\xi}_t^\epsilon$, where $\boldsymbol{\xi}_t^\epsilon$ is defined in (4.1.28). Then,

$$\begin{aligned}
v(\boldsymbol{x}) &= \mathbb{E}_x \int_0^\infty e^{-\alpha t} \left[H(\boldsymbol{X}_t^\epsilon) dt + \nabla v(\boldsymbol{X}_t^\epsilon) \cdot \boldsymbol{N}_t^\epsilon d\boldsymbol{\eta}_t^\epsilon \right] \\
&= \mathbb{E}_{\boldsymbol{x}} \int_0^\infty e^{-\alpha t} \left[H(\boldsymbol{X}_t^\epsilon) dt + (1 - 2\epsilon) \left[(\boldsymbol{N}_t^\epsilon)^+ \cdot \boldsymbol{K}^+ + (\boldsymbol{N}_t^\epsilon)^- \cdot \boldsymbol{K}^- \right] d\eta_t^\epsilon \right] \\
&= \mathbb{E}_{\boldsymbol{x}} \int_0^\infty e^{-\alpha t} \left[H(\boldsymbol{X}_t^\epsilon) dt + \left[(\boldsymbol{N}_t^\epsilon)^+ \cdot \boldsymbol{K}^+ + (\boldsymbol{N}_t^\epsilon)^- \cdot \boldsymbol{K}^- \right] d\eta_t^\epsilon \right] \\
&\quad -2\epsilon \mathbb{E}_{\boldsymbol{x}} \int_0^\infty e^{-\alpha t} \left[(\boldsymbol{N}_t^\epsilon)^+ \cdot \boldsymbol{K}^+ + (\boldsymbol{N}_t^\epsilon)^- \cdot \boldsymbol{K}^- \right] d\eta_t^\epsilon \\
&\geq \mathbb{E}_{\boldsymbol{x}} \int_0^\infty e^{-\alpha t} \left[H(\boldsymbol{X}_t^\epsilon) dt + \left[(\boldsymbol{N}_t^\epsilon)^+ \cdot \boldsymbol{K}^+ + (\boldsymbol{N}_t^\epsilon)^- \cdot \boldsymbol{K}^- \right] d\eta_t^\epsilon \right] \\
&\quad -2\epsilon K_{\max} \mathbb{E}_{\boldsymbol{x}} \int_0^\infty e^{-\alpha t} d\eta_t^\epsilon
\end{aligned}$$

where $\boldsymbol{N}^\epsilon(\boldsymbol{x}) = \boldsymbol{\gamma}^\epsilon(\boldsymbol{x})$ on $\mathcal{S}_\epsilon$ and $K_{\max} = \max_{1 \leq i \leq N} \{K_i^+, K_i^-\}$.
Moreover, there exists constant $C > 0$ such that $\mathbb{E}_{\boldsymbol{x}} \left[\int_0^\infty e^{-\alpha t} d\eta_t^\epsilon \right] \leq C$ for all $\epsilon < \frac{1}{2}$. Hence

$$v(\boldsymbol{x}) \geq J(\boldsymbol{x}; \boldsymbol{\xi}_t^\epsilon) - 2\epsilon C K_{\max}.$$

When $\epsilon_k \to 0$, we have $J(\boldsymbol{x}; \boldsymbol{\xi}_t^\epsilon) \to v(\boldsymbol{x})$.

Step 3: Existence and uniqueness of optimal control. Now we show that If $J(\boldsymbol{x}; \boldsymbol{\xi}^\epsilon) \to v(\boldsymbol{x})$ as $\epsilon \to 0$, then $\xi_t^\epsilon(\omega)$ converges in measure m_T. Here m_T is a measure on $([0,T] \times \Omega, \mathcal{B}[0,T] \times \mathcal{F})$. Furthermore, there exists an unique optimal policy $\boldsymbol{\xi}^*$ which is the limit of a subsequence of $\{\boldsymbol{\xi}^\epsilon\}_\epsilon$.

The existence follows with an appropriate modification of Theorem 4.5 and Corollary 4.11 in [147], as below. From [147], if $(\boldsymbol{N}^{\epsilon_k}, \boldsymbol{\xi}^{\epsilon_k})$ is a sequence of ϵ_k-optimal policies for $\boldsymbol{x}$ and $\lim_{k \to \infty} \epsilon_k \to 0$, then one can extract a subsequence $\epsilon_{k'}$ such that

$$\boldsymbol{\xi}_t^{\epsilon_{k'}} = \int_0^t \boldsymbol{N}_s^{\epsilon_{k'}} d\eta_s^{\epsilon_{k'}} \to \boldsymbol{\xi}_t^* \tag{4.1.29}$$

for Leb $\times \mathbb{P}$ almost all (t, ω), where Leb is the Lebesgue measure on $[0, \infty)$.

From the analysis in Step 1 and Step 2, we know there exits a sequence of ϵ_k-optimal policy and $\epsilon_k \to 0$ when $k \to \infty$. Therefore, the optimal control exists.

Let

$$A = \left\{\omega \mid \boldsymbol{X}_t^{\epsilon_{k'}}(\omega) \in \overline{\mathcal{C}}_{\epsilon_{k'}} \text{ for all } 0 \le t < \infty \text{ and all } k' \ge 0\right\}.$$

By definition (4.1.28), $P(A) = 1$. Also define

$$B = \left\{\omega \mid \boldsymbol{X}_t^{\epsilon_{k'}} \to \boldsymbol{X}_t \text{ a.e. } L_{eb} \text{ on } [0, \infty)\right\}.$$

Then by (4.1.29), $P(B) = 1$. For all $\omega \in A \cap B$, since $\overline{\mathcal{C}}$ is closed,

$$\boldsymbol{X}_t(\omega) \in \overline{\mathcal{C}} \text{ Leb a.e. on } [0, \infty)$$

It remains to show the uniqueness of optimal control. This can be proved by a contradiction argument. Suppose there are two optimal controls $\{\boldsymbol{\xi}^*\}_{t\ge 0}$ and $\{\boldsymbol{\xi}^{**}\}_{t\ge 0}$ such that $\boldsymbol{\xi}^* \neq \boldsymbol{\xi}^{**}$ almost surely. Let $\{\boldsymbol{X}_t^*\}_{t\ge 0}$ and $\{\boldsymbol{X}_t^{**}\}_{t\ge 0}$ be the corresponding trajectories. Let $\boldsymbol{\xi}_t = \frac{\boldsymbol{\xi}_t^* + \boldsymbol{\xi}_t^{**}}{2}$ and $\boldsymbol{X}_t = \frac{\boldsymbol{X}_t^* + \boldsymbol{X}_t^{**}}{2}$. Then

$$\begin{aligned} v(\boldsymbol{x}) - J(\boldsymbol{x}; \boldsymbol{\xi}_t) &= \frac{(J(\boldsymbol{x}; \boldsymbol{\xi}^*) + J(\boldsymbol{x}; \boldsymbol{\xi}^{**}))}{2} - J(x; \boldsymbol{\xi}) \\ &\ge \mathbb{E}\int_0^\infty e^{-\alpha t} \frac{H(\boldsymbol{X}_t^* + H(\boldsymbol{X}_t^{**}))}{2} - H\left(\frac{\boldsymbol{X}_t^* + \boldsymbol{X}_t^{**}}{2}\right) dt > 0. \end{aligned}$$

It is easy to check that $\frac{H(\boldsymbol{x}+H(\boldsymbol{y}))}{2} - H(\frac{\boldsymbol{x}+\boldsymbol{y}}{2}) > 0$ if $\boldsymbol{x} \neq \boldsymbol{y}$ by Assumption **A1**. Therefore we have $v(\boldsymbol{x}) > J(\boldsymbol{x}; \boldsymbol{\xi})$, which contradicts the optimality of $\{\boldsymbol{\xi}^*\}_{t\ge 0}$ and $\{\boldsymbol{\xi}^{**}\}_{t\ge 0}$. Hence the optimal control is unique.

Theorem 36. *When $\boldsymbol{x} \in \bar{\mathcal{C}}$, under conditions **A1** - **A4**, the optimal policy defined in (4.1.29) acts only on $\partial\mathcal{C}$, and its push direction $\vartheta(\boldsymbol{x})$ is in $r(\boldsymbol{x})$.*

Proof of Theorem 36. Recall the definition of smooth function ϕ^ϵ in (4.1.24) and the smooth version of value function v^ϵ in (4.1.25). Let $H^\epsilon(\boldsymbol{x}) = \phi^\epsilon * H(\boldsymbol{x})$. From the HJB Equation (4.1.6),

$$\alpha v - \mathcal{L}v \le H, \quad \gamma(\nabla v) \le 1 \text{ in } \mathbb{R}^N,$$

and

$$\alpha v^\epsilon - \mathcal{L}v^\epsilon \le H^\epsilon, \quad \gamma(\nabla v^\epsilon) \le 1 \text{ in } \mathbb{R}^N. \tag{4.1.30}$$

Let $T > 0$ and apply the Meyer's version of Itô's formula (Theorem 14.3.2 in [64]) to $e^{-\alpha t} v^\epsilon(\boldsymbol{x})$,

$$\begin{aligned} \mathbb{E}_{\boldsymbol{x}}\left[e^{-\alpha T} v^\epsilon(\boldsymbol{X}_T)\right] = \quad & v^\epsilon(\boldsymbol{x}) + \mathbb{E}_{\boldsymbol{x}}\int_0^T e^{-\alpha T}\left(\mathcal{L}v^\epsilon - \alpha v^\epsilon\right)(\boldsymbol{X}_t) dt \\ + \quad & \mathbb{E}_{\boldsymbol{x}}\int_0^T e^{-\alpha T} \nabla v^\epsilon(\boldsymbol{X}_t) \cdot \boldsymbol{N}_t d\eta_t \\ + \quad & \mathbb{E}_{\boldsymbol{x}}\int_0^T \sum_{0\le t<T} e^{-\alpha t}(v^\epsilon(\boldsymbol{X}_t) - v^\epsilon(\boldsymbol{X}_{t-}) - \nabla v^\epsilon(\boldsymbol{X}_t)(\eta_t - \eta_{t-})), \end{aligned}$$

where the last term comes from the jumps of $\boldsymbol{X}_t$. By (4.1.30),

$$\begin{aligned}\mathbb{E}_{\boldsymbol{x}}\left[e^{-\alpha T}v^\epsilon(\boldsymbol{X}_T)\right]+\mathbb{E}_{\boldsymbol{x}}\int_0^T e^{-\alpha T}H^\epsilon(\boldsymbol{X}_t)dt-\mathbb{E}_{\boldsymbol{x}}\int_0^T e^{-\alpha T}\nabla v^\epsilon(\boldsymbol{X}_t)\cdot\boldsymbol{N}_t d\eta_t\\ +\mathbb{E}_{\boldsymbol{x}}\int_0^T\sum_{0\le t<T}e^{-\alpha t}(v^\epsilon(\boldsymbol{X}_t)-v^\epsilon(\boldsymbol{X}_{t-})-\nabla v^\epsilon(\boldsymbol{X}_t)(\eta_t-\eta_{t-}))\ge v^\epsilon(\boldsymbol{x}),\end{aligned}\tag{4.1.31}$$

as $\boldsymbol{X}_t\in\bar{\mathcal{C}}$ for all $t\ge 0$ a.s. and $\mathcal{C}$ is bounded. Because v, ∇v, D^2v are bounded on $B_R(0)$, with $\bar{\mathcal{C}}\subset B_{R-1}(0)$, thus H^ϵ, v^ϵ are bounded uniformly on $\bar{\mathcal{C}}$ for $\epsilon<1$, and

$$v^\epsilon\to v,\quad \nabla v^\epsilon\to\nabla v,\quad H^\epsilon\to H\quad \text{uniformly in } \bar{\mathcal{C}}.$$

On the other hand for $\forall\boldsymbol{x}\in\mathcal{C}$,

$$v(\boldsymbol{x})=\mathbb{E}_{\boldsymbol{x}}\int_0^\infty e^{-\alpha t}\left[H(\boldsymbol{X}_t)dt+\left[(\boldsymbol{N}_t^*)^+\cdot\boldsymbol{K}^+-(\boldsymbol{N}_t^*)^-\cdot\boldsymbol{K}^-\right]d\eta_t^*\right],\tag{4.1.32}$$

where $\boldsymbol{X}_t^*=\boldsymbol{x}+\boldsymbol{B}_t+\xi_t^*$ with $\xi_t^*:=\int_0^t\boldsymbol{N}_s^*d\eta_s^*$ the optimal control. In particular,

$$\mathbb{E}_x\int_0^\infty e^{-\alpha t}d\eta_t^*<\infty,\tag{4.1.33}$$

which leads to

$$\mathbb{E}_{\boldsymbol{x}}\int_0^T e^{-\alpha t}\left[(\boldsymbol{N}_t^*)^+\cdot\boldsymbol{K}^+-(\boldsymbol{N}_t^*)^-\cdot\boldsymbol{K}^-\right]d\eta_t^*<\infty.$$

Thus, by the bounded convergence theorem, we get from (4.1.31)

$$\begin{aligned}\mathbb{E}_{\boldsymbol{x}}\left[e^{-\alpha T}v(\boldsymbol{X}_T)\right]+\mathbb{E}_x\int_0^T e^{-\alpha T}H(X_t)dt-\mathbb{E}_{\boldsymbol{x}}\int_0^T e^{-\alpha T}\nabla v(\boldsymbol{X}_t)\cdot\boldsymbol{N}_t d\eta_t\\ +\mathbb{E}_{\boldsymbol{x}}\int_0^T\sum_{0\le t<T}e^{-\alpha t}(v(\boldsymbol{X}_t)-v^\epsilon(\boldsymbol{X}_{t-})-\nabla v(\boldsymbol{X}_t)(\eta_t-\eta_{t-}))\ge v(\boldsymbol{x}).\end{aligned}\tag{4.1.34}$$

The last term on the left-hand side is nonpositive because of convexity of v, hence

$$\mathbb{E}_{\boldsymbol{x}}\left[e^{-\alpha T}v(\boldsymbol{X}_T)\right]+\mathbb{E}_{\boldsymbol{x}}\int_0^T e^{-\alpha T}H(\boldsymbol{X}_t)dt-\mathbb{E}_{\boldsymbol{x}}\int_0^T e^{-\alpha T}\nabla v(\boldsymbol{X}_t)\cdot\boldsymbol{N}_t d\eta_t\ge v(\boldsymbol{x}).$$

Letting $T\to\infty$, by the boundedness of $\boldsymbol{X}_t^*$, $\gamma(\nabla v)\le 1$, $|\boldsymbol{N}_t^*|=1$, (4.1.39), and (4.1.32), we have

$$0\ge\mathbb{E}_{\boldsymbol{x}}\int_0^\infty e^{-\alpha t}\left[\left[\nabla v(\boldsymbol{X}_t^*)+\boldsymbol{K}^+\cdot(\boldsymbol{N}_t^*)^+\right]d\eta_t+\left[\nabla v(\boldsymbol{X}_t^*)-\boldsymbol{K}^-\cdot(\boldsymbol{N}_t^*)^-\right]d\eta_t\right].$$

Given $\gamma(\nabla v)\le 1$, we have

$$-K_i^-\le v_{x^i}(\boldsymbol{x})\le K_i^+,\quad \forall x\in\mathbb{R}^N \text{ and } i=1,2,\cdots,N.$$

Hence

$$0\ge\mathbb{E}_{\boldsymbol{x}}\int_0^\infty e^{-\alpha t}\left[\left[\nabla v(\boldsymbol{X}_t^*)+\boldsymbol{K}^+\cdot(\boldsymbol{N}_t^*)^+\right]d\eta_t+\left[\nabla v(\boldsymbol{X}_t^*)-\boldsymbol{K}^-\cdot(\boldsymbol{N}_t^*)^-\right]d\eta_t\right]\ge 0.$$

This implies $d\eta_t^*=0$ when $\gamma(\nabla v(\boldsymbol{X}_t^*))<1$ a.e. t. Also, when $d\eta_t^*\ne 0$, $\boldsymbol{N}_t^*(\boldsymbol{x})\in r(\boldsymbol{x})$ for $\boldsymbol{x}\in\mathcal{S}$ a.e. $t\in[0,\infty)$, where the reflection cone $r(\boldsymbol{x})$ is defined in (4.1.26).

□

PO when $\boldsymbol{x} \notin \overline{\mathcal{C}}$

When $\boldsymbol{x} \notin \overline{\mathcal{C}}$, the optimal policy is to jump immediately to some point $\hat{x} \in \mathcal{C}$ and then follows the optimal policy in $\overline{\mathcal{C}}$. In order to define directions of optimal jumps, we start by assuming the existence of a certain map that projects points in $\mathbb{R}^N$ onto $\mathcal{C}$ in a way that is compatible with the directions of reflection $r(\cdot)$.

We will need the following assumption so that the refection field of the Skorokhod problem is extendable to the $\mathbb{R}^N$ plane (Dupuis and Ishii [72]). Note that **A5** follows from conditions **A1**-**A3** when $N = 2$.

A5. There is a map $\pi : \mathbb{R}^N \to \mathcal{C}$ satisfying $\pi(\boldsymbol{x}) = \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathcal{C}$ and $\pi_\epsilon(\boldsymbol{x}) - \boldsymbol{x} \in r(\pi(\boldsymbol{x}))$.

Theorem 37. *Under conditions **A1**-**A3**, and **A5**, for any $\boldsymbol{x} \notin \overline{\mathcal{C}}$, there exists an optimal policy π such that $\pi(\boldsymbol{x}) \in \partial\mathcal{C}$ at time 0 and*

$$v(\boldsymbol{x}) = v(\pi(\boldsymbol{x})) + \|\boldsymbol{x} - \pi(\boldsymbol{x})\|.$$

Proof. Define $l(\boldsymbol{y}) = \sum_i l_i(y_i)$ where

$$l_i(y_i) = \begin{cases} K_i^- y_i \text{ if } y_i \geq 0, \\ -K_i^+ y_i \text{ if } y_i < 0. \end{cases} \tag{4.1.35}$$

Notice that $l(\boldsymbol{y})$ is a convex function and

$$l_i(y_i) = \max_{-K_i^+ \leq k \leq K_i^-} \{k y_i\} = \max\{-K_i^+ y_i, K_i^- y_i\} \text{ for } y_i \in \mathbb{R}.$$

Define $u_\epsilon(\boldsymbol{x}) = v(\hat{\boldsymbol{x}}_\epsilon) + l(\boldsymbol{x} - \hat{\boldsymbol{x}}_\epsilon)$, $u_\epsilon(\boldsymbol{x})$ correspond to the policy that push $\boldsymbol{x}$ to $\hat{\boldsymbol{x}}_\epsilon \in \mathcal{S}_\epsilon$ when $\boldsymbol{x} \in \mathbb{R}^N \setminus \bar{\mathcal{C}}$ and keep to play optimal policy in $\bar{\mathcal{C}}$ afterwards.

Here we define two linear approximations which are the both lower bound and upper bound of the value function $v(\boldsymbol{x})$, respectively.

For $\boldsymbol{x} \notin \overline{\mathcal{C}}$ define

$$\begin{aligned} u_1(\boldsymbol{x}) &= v(\pi(\boldsymbol{x})) + \nabla v(\pi(\boldsymbol{x}))(\boldsymbol{x} - \pi(\boldsymbol{x})), \\ u_2(\boldsymbol{x}) &= v(\pi(\boldsymbol{x})) + l(\boldsymbol{x} - \pi(\boldsymbol{x})). \end{aligned} \tag{4.1.36}$$

Notice that $u_2(\boldsymbol{x}) \geq v(\boldsymbol{x})$ because it corresponds to a sub-optimal strategy. Meanwhile, $u_1(\boldsymbol{x}) \leq v(\boldsymbol{x})$ because of convexity. Thus,

$$u_1(\boldsymbol{x}) \leq v(\boldsymbol{x}) \leq u_2(\boldsymbol{x}). \tag{4.1.37}$$

Our next step is to show $u_1(\boldsymbol{x}) = u_2(\boldsymbol{x})$.

By Assumption **A5**, we can rewrite u_1 and u_2 in (4.1.36) by the following expressions,

$$\begin{aligned} u_1(\boldsymbol{x}) &= v(\pi(\boldsymbol{x})) + \nabla v(\pi(\boldsymbol{x})) \cdot d(\pi(\boldsymbol{x}))\|\boldsymbol{x} - \pi(\boldsymbol{x})\|, \\ u_2(\boldsymbol{x}) &= v(\pi(\boldsymbol{x})) + \boldsymbol{K}(\pi(\boldsymbol{x})) \cdot d(\pi(\boldsymbol{x}))\|\boldsymbol{x} - \pi(\boldsymbol{x})\|. \end{aligned}$$

where $d(\pi(\boldsymbol{x})) \in r(\pi(\boldsymbol{x}))$, $\boldsymbol{K}(\boldsymbol{x}) = (K_1, \cdots, K_N)(\boldsymbol{x})$ and

$$K_i(\boldsymbol{x}) = K_i^+ \mathbf{1}(\nabla v(\boldsymbol{x}) > 0) + K_i^- \mathbf{1}(\nabla v(\boldsymbol{x}) < 0).$$

Therefore $u_1(\boldsymbol{x}) = u_2(\boldsymbol{x})$.

□

McKean-Vlasov Approximation for PO

In this section, we connect the PO of a class of N-player games with an appropriate McKean-Vlasov (MKV) control problem. We show that the PO can be approximated by the solution of this MKV problem with an error $\epsilon = \left(\frac{1}{\sqrt{N}}\right)$.

Here we restrict the N-player game to a symmetric case as *law of large number (LLN)* works for a large population with homogeneous individuals and with weak interactions, as in the case for mean-field games.

To start, define

$$J^i(\boldsymbol{x};\boldsymbol{\xi}) = \mathbb{E}\int_0^\infty e^{-\alpha t}\left[h(X_t^i, \bar{\boldsymbol{X}}_t^{-i})dt + K^+ d\xi_t^{i,+} + K^- d\xi_t^{i,-}\right], \qquad \textbf{(N-player}')$$

with

$$dX_t^i = \mu dt + \sigma dB_t^i + d\xi_t^{i,+} - d\xi_t^{i,-}, \text{ and } X_{0-}^i = x^i.$$

Here $\bar{x}^{-i} = \sum_{j\neq i}\beta^{i,j}x^j$ is the weighted average of all players' positions other than player i, with $\sum_{j\neq i}\beta^{i,j} = 1$ and $\beta^{i,j} \geq 0$ $(i,j = 1,2,\cdots,N)$. In the objective function (**N-player′**), player i interacts with other players only through the mean term $\bar{\boldsymbol{X}}_t^{-i}$. And all players share the same model specification: μ, σ, h, K^+ and K^-. For simplicity, assume $\beta^{i,j} = \frac{1}{N-1}$ for all $1 \leq i \neq j \leq N$.

By the symmetry of problem (**N-player′**), it is easy to check that $\mathbb{P}_{X_t^{i*}} = \mathbb{P}_{X_t^{j*}}$ for any $1 \leq i \neq j \leq N$ and $t \geq 0$ where X_t^{i*} and X_t^{j*} are dynamics of player i and j under PO.

Denote $\nabla_i h(x^1,x^2) = \partial_{x^i}h_y(x^1,x^2)$ and $\nabla^2_{ij}h(x^1,x^2) = \partial^2_{x^ix^j}h(x^1,x^2)$ for $i,j = 1,2$. Further assume that there exists a constant $\tilde{K} > 0$ such that

$$|\nabla_2 h(x^1,x^2)| \leq \tilde{K}(|x^1| + |x^2|).$$

Letting $N \to \infty$, the corresponding MKV problem becomes

$$\begin{aligned} u(\mathbb{P}_Z) &:= \inf_{(\xi^+,\xi^-)\in\mathcal{U}} J_{MV}(\xi_t^+,\xi_t^-) \\ &:= \inf_{(\xi^+,\xi^-)\in\mathcal{U}} \mathbb{E}\int_0^\infty e^{-\alpha t}\left[h\left(X_t, \mathbb{E}[X_t]\right)dt + K^+ d\xi_t^+ + K^- d\xi_t^-\right], \qquad (4.1.38) \\ \text{such that} \quad & dX_t = \mu dt + \sigma dB_t + d\xi_t^+ - d\xi_t^-, \;\; \mathbb{P}_{X_{0-}} = \mathbb{P}_Z \in \mathcal{P}(\mathbb{R}), \end{aligned}$$

for any $Z \in L^2(\mathbb{R})$. Here $\mathcal{P}(\mathbb{R})$ is the probability measure on $\mathbb{R}$. The admissible control set $\mathcal{U}$ is defined as

$$\mathcal{U} = \left\{(\xi_t^+,\xi_t^-) \;\middle|\; \xi_t^+ \text{ and } \xi_t^- \text{ are } \mathcal{F}_t^{(X_{t-},\mathbb{P}_{X_{t-}})}\text{-progressively measurable, càdlàg, non-decreasing,} \right.$$
$$\left. \text{with } \mathbb{E}\left[\int_0^\infty e^{-\alpha t}d\xi_t^+\right] < \infty, \;\; \mathbb{E}\left[\int_0^\infty e^{-\alpha t}d\xi_t^-\right] < \infty, \;\; \xi_{0-}^+ = 0, \;\; \xi_{0-}^- = 0\right\}.$$

Moreover $h : \mathbb{R} \to \mathbb{R}^+$ is assumed to be convex, symmetric, non-negative, and there exist $C_0 > c_0 > 0$ such that $c_0 \leq \|\nabla h\|_2 \leq C_0$.

The difference between (**N-player′**) and (5.2.1) is the running cost term. In (**N-player′**), the second component of the running cost is $\bar{X}_t^{-i}$. In contrast, the second component of the running cost is $\mathbb{E}[X_t]$ in (5.2.1).

Theorem 38 (MKV Approximation for PO)**.** *Denote $\boldsymbol{\xi}^* := (\xi^{1*}, \cdots, \xi^{N*})$ as the PO solution to (**N-player**). In addition, denote $\bar{\boldsymbol{\xi}} := (\bar{\xi}^1, \ldots, \bar{\xi}^N)$ as the solution to the MKV control problem (5.2.1). Assume **A1-A4**. Assume further there exists a constant c_N such that the PO solution $\{\boldsymbol{X}_t^*\}_{t\geq 0}$ satisfies*

$$\int_0^\infty e^{-\alpha t}\mathbb{E}\left(X_t^{i*} - \mathbb{E}[X_t^{i*}]\right)\left(X_t^{j*} - \mathbb{E}[X_t^{j*}]\right)dt \leq c_N^2, \tag{4.1.39}$$

for any $i \neq j$, and there exists a constant C_N such that a.e.

$$\left|X_t^{i*}\right| \leq C_N. \tag{4.1.40}$$

Then

$$\left|\frac{1}{N}\sum_{i=1}^N \mathbb{E}_{\boldsymbol{X}_{0-}} J^i\left(\boldsymbol{X}_{0-}; \boldsymbol{\xi}^*\right) - \frac{1}{N}\sum_{i=1}^N \mathbb{E}_{\boldsymbol{X}_{0-}} J^i\left(\boldsymbol{X}_{0-}; \bar{\boldsymbol{\xi}}\right)\right| = O\left(c_N + \frac{C_N}{N} + C_N\sqrt{\frac{C_N}{N} + c_N}\right).$$

Proof. Assume $\boldsymbol{X}_t$ is the dynamics under controls $\boldsymbol{\xi}_t$ such that $\mathbb{P}_{X_t^i} = \mathbb{P}_{X_t^j}$ ($i \neq j$ and $t \geq 0$) and (4.1.39)-(4.1.40) are satisfied.

It is easy to check that under Assumptions (4.1.39) and (4.1.40), PO solution satisfies above conditions.

By the Taylor's expansion,

$$\begin{aligned} & h\left(X_t^1, \overline{X}_t^{-i}\right) \\ = \;& h\left(X_t^1, \mathbb{E}[X_t^1]\right) + \nabla_2 h\left(X_t^1, \mathbb{E}[X_t^1]\right)\left(\frac{\sum_{j=1}^N \left(X_t^j - \mathbb{E}[X_t^j]\right)}{N}\right) \\ & + \nabla_{22}^2 \frac{h(X_t^1, U_t)}{2}\left(\frac{\sum_{j=1}^N \left(X_t^j - \mathbb{E}[X_t^j]\right)}{N}\right)^2, \end{aligned}$$

where $U_t \in [\min\{\frac{1}{N}\sum_{j=1}^N X_t^j, \mathbb{E}[X_t^1]\}, \max\{\frac{1}{N}\sum_{j=1}^N X_t^j, \mathbb{E}[X_t^1]\}]$. That is, U_t is a process between $\frac{1}{N}\sum_{j=1}^N X_t^j$ and $\mathbb{E}[X_t^1]$.

Moreover,

$$\begin{aligned} & h\left(X_t^1, \overline{X}_t^{-i}\right) \\ = \; & h\left(X_t^1, \mathbb{E}[X_t^1]\right) + \nabla_2 h\left(X_t^1, \mathbb{E}[X_t^1]\right)\left(\frac{\sum_{j=1}^N \left(X_t^j - \mathbb{E}[X_t^j]\right)}{N}\right) \\ & + \nabla_{22}^2 \frac{h(X_t^1, U_t)}{2}\left(\frac{\sum_{j=1}^N \left(X_t^j - \mathbb{E}[X_t^j]\right)}{N}\right)^2 \\ \leq \; & h\left(X_t^1, \mathbb{E}[X_t^1]\right) + 2\tilde{K} C_N \left|\frac{\sum_{j=1}^N \left(X_t^j - \mathbb{E}[X_t^j]\right)}{N}\right| \\ & + \frac{K}{2}\left(\frac{\sum_{j=1}^N \left(X_t^j - \mathbb{E}[X_t^j]\right)}{N}\right)^2 \end{aligned}$$

Similarly,

$$\begin{aligned} & h\left(X_t^1, \overline{X}_t^{-i}\right) \\ \geq \; & h\left(X_t^1, \mathbb{E}[X_t^1]\right) - 2\tilde{K} C_N \left|\frac{\sum_{j=1}^N \left(X_t^j - \mathbb{E}[X_t^j]\right)}{N}\right| \\ & - \frac{K}{2}\left(\frac{\sum_{j=1}^N \left(X_t^j - \mathbb{E}[X_t^j]\right)}{N}\right)^2. \end{aligned}$$

Since

$$\mathbb{E}\left(\frac{\sum_{j=1}^N \left(X_t^j - \mathbb{E}[X_t^j]\right)}{N}\right)^2 = \frac{\mathbb{E}\left[X_t^1 - \mathbb{E}[X_t^1]\right]^2}{N} + \frac{N-1}{N}\mathbb{E}\left[\left(X_t^1 - \mathbb{E}[X_t^1]\right)\left(X_t^2 - \mathbb{E}[X_t^2]\right)\right]$$

and

$$\mathbb{E}\left|\frac{\sum_{j=1}^N \left(X_t^j - \mathbb{E}[X_t^j]\right)}{N}\right| \leq \left[\mathbb{E}\left(\frac{\sum_{j=1}^N \left(X_t^j - \mathbb{E}[X_t^j]\right)}{N}\right)^2\right]^{1/2}$$

Now, by assumptions (4.1.39)-(4.1.40), solving the central controller problem (4.1.4) is equivalent to solving the following problem,

$$\frac{1}{N}\sum_{i=1}^N \left(\mathbb{E}\int_s^\infty e^{-\alpha t}\left[h\left(X_t^i - \rho\mathbb{E}[X_t^i]\right)dt + K d\xi_t^{i,+} + K d\xi_t^{i,-}\right]\right) + O\left(c_N + \frac{C_N}{N} + C_N\sqrt{\frac{C_N}{N} + c_N}\right),$$

which is a decentralized problem of MKV-type for each individual player. □

Remark 38.1. (4.1.39) *implies that the long-run discounted correlation between X_t^{j*} and X_t^{i*} are bounded by c_N, where $\boldsymbol{X}_t^*$ is the dynamics under PO control in game* (**N-player**). *If $c_N \to 0$ as $N \to \infty$, there is a weak correlation among players when the number of players is large. This assumption is common for for MKV problems, see [53, 47] for example.*

Remark 38.2. *Denote $\mathcal{C}_N$ as the non-action region of PO solution for (* ***N-player****). When support(Z) $\subset \mathcal{C}_N$ for all N, $c_N = \frac{1}{\sqrt{N}}$.*

4.2 PO vs NE via Price of Anarchy (PoA)

In this section, we compare PO and NE under the notion of Price of Anarchy (PoA). For sake of comparison, we specify $K_i^+ = K_i^- = 1$, $h^i(\boldsymbol{x}) = h\left(x^i - \rho\frac{\sum_{j=1}^N x^i}{N}\right)$, $0 < \rho < 1$, $\mu^i = 0$ and $\boldsymbol{\sigma}^i = (0, \cdots, 0, 1, 0, \cdots, 0) \in \mathbb{R}^N$ with the i^{th} component to be 1. The distance function $h \geq 0$ is assumed to be symmetric, convex and there exists $0 < c < C$ such that $c \leq h'' \leq C$ and $\frac{c}{C} \geq \frac{\rho^2}{N-1}$. We denote this game specification as (**N-player-a**).

To start, let us recall the notion of NE, a stable solution under competition.

NE

Definition 39 (Markovian NE)**.** *A tuple of admissible controls $\boldsymbol{\xi}^* = (\xi^{1*}, \cdots, \xi^{N*})$ is a Markovian NE of the stochastic game* ***N-player*** *if for any $i = 1, 2, \cdots, N$, $\boldsymbol{X}_{0-} = \boldsymbol{x}$, and any $(\boldsymbol{\xi}^{-i*}, \xi^i) \in \mathcal{S}_N$, the following inequality holds,*

$$J^i(\boldsymbol{x}; \boldsymbol{\xi}^*) \leq J^i(\boldsymbol{x}; (\boldsymbol{\xi}^{-i*}, \xi^i)).$$

Here strategies ξ^{i*} and ξ^i are functions of time t and state $\boldsymbol{X}_t = (X_t^1 \cdots, X_t^N)$, with the notation $(\boldsymbol{x}^{-i}, y^i) := (x^1, \cdots, x^{i-1}, y^i, x^{i+1}, \cdots, x^N)$ for any $\boldsymbol{x} \in \mathbb{R}^N$. $J^i(\boldsymbol{x}; \boldsymbol{\xi}^*)$ is called the NE value associated with $\boldsymbol{\xi}^*$.

NE solution. Here the admissible control set $\mathcal{S}_N$ for the NE solution is defined as

$$\begin{aligned}\mathcal{S}_N := \Big\{(\xi^1, \ldots, \xi^N) \;\Big|\; \xi^i = (\xi^{i,+}, \xi^{i,-}) \in \mathcal{U}_N^i, \mathbb{P}\left(\Delta\xi_t^i(\boldsymbol{x})\Delta\xi_t^j(\boldsymbol{x}) > 0\right) = 0,\\ \text{for any } t > 0, \boldsymbol{x} \in \mathbb{R}^N, i, j \in \{1, \ldots, N\} \text{ and } i \neq j\Big\},\end{aligned} \tag{4.2.1}$$

with

$$\begin{aligned}\mathcal{U}_N^i = \Big\{(\xi_t^{i,+}, \xi_t^{i,-}) \;\Big|\; &\xi_t^{i,+} \text{ and } \xi_t^{i,-} \text{ are } \mathcal{F}^{(X_{t-}^1, \ldots, X_{t-}^N)}\text{-progressively measurable, càdlàg, non-decreasing,}\\ &\text{with } \mathbb{E}\left[\int_0^\infty e^{-\alpha t} d\xi_t^{i,+}\right] < \infty,\ \mathbb{E}\left[\int_0^\infty e^{-\alpha t} d\xi_t^{i,-}\right] < \infty,\ \xi_{0-}^{i,+} = 0,\ \xi_{0-}^{i,-} = 0\Big\},\end{aligned}$$

where $\alpha > 0$ is the discount factor for player j and $\{\mathcal{F}^{(X_t^1, \ldots, X_t^N)}\}_{t \geq 0}$ is the natural filtration of $\{(X_t^1, \ldots, X_t^N)\}_{t \geq 0}$.

Following the approach in ([94]) for the case of $\rho = 1$, we can show that deriving NEs is reduced to solving a Skorokhod problem and

Theorem 40 (NE.). *When the starting position $\boldsymbol{x} \in \mathcal{CW}$, NE of game (**N-player**) is a solution to the Skorokhod problem with data $(\boldsymbol{B}_t, \mathcal{CW}, \{\boldsymbol{d}_j\}_{j=1}^{2N}, \boldsymbol{x})$, where*

$$\begin{aligned} \mathcal{CW} &:= \left\{ \boldsymbol{x} \in \mathbb{R}^N \;\middle|\; \left| x^i - \rho \frac{\sum_{j\neq i} x^j}{N-1} \right| < c_N, \text{ for any } i = 1, \dots, N \right\} \\ &= \left\{ \boldsymbol{x} \in \mathbb{R}^N \;\middle|\; \boldsymbol{n}_j \cdot \boldsymbol{x} > -c_N \sqrt{\frac{N}{N-1}}, \text{ for } j = 1, \dots, 2N \right\} \\ &= \cap_{i=1}^N (E_i^- \cup E_i^+)^c. \end{aligned} \tag{4.2.2}$$

The normal direction of each face is given by

$$\begin{aligned} \boldsymbol{n}_i &= \frac{\sqrt{N-1}}{\sqrt{N}} \left(-\rho \frac{1}{N-1}, \cdots, -\rho \frac{1}{N-1}, 1, -\rho \frac{1}{N-1}, \cdots, -\rho \frac{1}{N-1} \right), \\ \boldsymbol{n}_{i+N} &= -\boldsymbol{n}_i, \end{aligned} \tag{4.2.3}$$

where 1 is in the i^{th} position of $\sqrt{\frac{N}{N-1}}\boldsymbol{n}_i$. Here the threshold c_N is the unique positive solution to

$$\frac{1}{\sqrt{\frac{2(N-1)\alpha}{N}}} \tanh\left(c \sqrt{\frac{2(N-1)\alpha}{N}} \right) = \frac{p_N'(c) - 1}{p_N''(c)}, \tag{4.2.4}$$

with

$$p_N(x) = \mathbb{E}\left[\int_0^\infty e^{-\alpha t} \left(\frac{N-\rho}{N} x + \sqrt{\frac{N-1}{N} + \frac{(1-\rho)^2}{N}} B_t \right)^2 dt \right]. \tag{4.2.5}$$

Finally, the reflection directions are given by

$$\boldsymbol{d}_i = (0, \cdots, 1, \cdots, 0), \quad \boldsymbol{d}_{i+N} = -\boldsymbol{d}_i, \quad i = 1, \dots, N, \tag{4.2.6}$$

Property of $\mathcal{CW}$. Note that since matrix

$$N = \begin{bmatrix} 1 & -\frac{\rho}{N-1} & -\frac{\rho}{N-1} & \cdots & -\frac{\rho}{N-1} \\ -\frac{\rho}{N-1} & 1 & -\frac{\rho}{N-1} & \cdots & -\frac{\rho}{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{\rho}{N-1} & -\frac{\rho}{N-1} & -\frac{\rho}{N-1} & \cdots & 1 \end{bmatrix}$$

has full-rank when $\rho \in [0, 1)$, it is easy to check that $\mathcal{CW}$ is a bounded polyhedron with $2N$ faces. This property is useful for analyzing the PoA.

To avoid simultaneous jumps, one could impose the players to jump *sequentially* if π (defined in Assumption **A5**) involves jumps from multiple players, in both NE and PO policies. For more technical details, one is refered to [94].

PoA

Definition 41 (PoA)**.**

$$PoA(\boldsymbol{x}) = \frac{\sup_{\boldsymbol{\xi^*}\in\mathcal{N}}\left(\sum_{i=1}^N J^i(\boldsymbol{x};\boldsymbol{\xi}^*)\right)}{\sum_{i=1}^N J^i\left(\boldsymbol{x};\widetilde{\boldsymbol{\xi}}\right)},$$

where $\widetilde{\boldsymbol{\xi}}$ *is the solution to the central controller's problem* (4.1.4)*, and*

$$\mathcal{N} := \left\{\boldsymbol{\xi}^* \mid \boldsymbol{\xi}^* \textit{is NE strategy of game } (J^i, {X_t}^i, \xi^i)_{i=1}^N\right\}$$

is the set of all NE strategies.

Note that by definition PoA($\boldsymbol{x}$) ≥ 1. Furthermore, the smaller the value of the PoA, the more efficient the strategy.

For the N-player game (**N-player-a**) with $0 \leq \rho < 1$, we have

Theorem 42 (Upper bound on PoA)**.** *For game* ***N-player-a****,*

$$PoA(\boldsymbol{x}) \leq \frac{N(c_N + c)^2 + \alpha\sum_{i=1}^N u(x^i)}{N\tilde{c}_N^2},$$

for $\boldsymbol{x} \in \mathcal{CW}$*, where* $\mathcal{CW}$ *is defined in* (4.2.2)*,* $\tilde{c}_N = Diam(\mathcal{C})$*, the threshold* $c_N > 0$ *is the unique positive solution to* (4.2.4) *with* $p_N(x)$ *defined in* (4.2.5)*. The threshold* $c > 0$ *is the unique positive solution to*

$$\frac{1}{\sqrt{2\alpha}}\tanh\left(c\sqrt{2\alpha}\right) = \frac{p_1^{'}(c) - 1}{p_1^{''}(c)}, \tag{4.2.7}$$

with

$$\begin{aligned} p_1(x) &= \mathbb{E}\left[\int_0^\infty e^{-\alpha t}h(x + B_t)dt\right] \\ &= \frac{1}{\sqrt{2\alpha}}\left(e^{-x\sqrt{2\alpha}}\int_{-\infty}^x h(z)e^{z\sqrt{2\alpha}}dz + e^{x\sqrt{2\alpha}}\int_x^\infty h(z)e^{-z\sqrt{2\alpha}}dz\right). \end{aligned}$$

Finally

$$u(x) = \begin{cases} -\frac{p_1^{''}(c)\cosh\left(x\sqrt{2\alpha}\right)}{2\alpha\cosh\left(c\sqrt{2\alpha}\right)} + p_1(x), & 0 \leq x \leq c, \\ v(c) + (x - c), & x \geq c, \\ v(-x), & x < 0. \end{cases} \tag{4.2.8}$$

Proof. In order to bound the PoA, we first start with two common properties of the NE strategies for game (**N-player-a**). First, given Theorem 40, we have $\left|X_t^{i*} - \rho\overline{\boldsymbol{X}}_t^{-i*}\right| \leq c_N$ almost surely for every $t \geq 0$.

Next, denote $\xi^{i\sharp}$ as the following decentrlized strategy:

$$\xi_t^{i\sharp,+} = \max\left\{0, \max_{0\le s\le t}\left\{-x^i - B_s^i + \xi_t^{i\sharp,-} - c\right\}\right\}, \tag{4.2.9}$$

$$\xi_t^{i\sharp,-} = \max\left\{0, \max_{0\le s\le t}\left\{x^i + B_s^i + \xi_t^{i\sharp,+} - c\right\}\right\}, \tag{4.2.10}$$

which is the optimal control of the single agent problem with constant c is defined in (4.2.7). Define a matrix

$$\tilde{N} = \begin{bmatrix} 1 & -\frac{\rho}{N-1} & -\frac{\rho}{N-1} & \cdots & -\frac{\rho}{N-1} \\ -\frac{\rho}{N-1} & 1 & -\frac{\rho}{N-1} & \cdots & -\frac{\rho}{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 1 & \cdots & 0 \\ -\frac{\rho}{N-1} & -\frac{\rho}{N-1} & -\frac{\rho}{N-1} & \cdots & 1 \end{bmatrix},$$

where the i-th row are all zeros except for the (i,i) component to be 1. Since matrix N has full-rank when $\rho \in [0,1)$, we know $(\boldsymbol{X}_t^{-i*}, X_t^{i\sharp})_{t\ge 0}$ is on a bounded region for all $i = 1, 2, \cdots, N$. Moreover, we have $\left|X_t^{i\sharp} - \rho\overline{\boldsymbol{X}}_t^{-i*}\right| \le c_N + c$ almost surely.

Therefore,

$$\begin{aligned} J^i(\boldsymbol{x}, \xi^{i*}|\boldsymbol{\xi^{-i*}}) &\le J^i(\boldsymbol{x}, \xi^{i\sharp}|\boldsymbol{\xi^{-i*}}) \\ &= \mathbb{E}\int_0^\infty e^{-\alpha t}\left[\left(X_t^{i\sharp} - \rho\frac{\sum_{j\ne i} X_t^{j*}}{N-1}\right)^2 dt + d\xi_t^{i\sharp,+} + d\xi_t^{i\sharp,-}\right] \\ &\le \mathbb{E}\int_0^\infty e^{-\alpha t}\left[\left[(c_N+c)^2 + \left(X_t^{i\sharp}\right)^2\right] dt + d\xi_t^{i\sharp,+} + d\xi_t^{i\sharp,-}\right] \\ &= \frac{1}{\alpha}(c_N+c)^2 + u(x^i), \end{aligned}$$

with u defined in (4.2.8).

Therefore,

$$\sup_{\boldsymbol{\xi^*}\in\mathcal{N}}\left(\sum_{i=1}^N J^i(\boldsymbol{x};\boldsymbol{\xi}^*)\right) \le N(c_N+c)^2 + \sum_{i=1}^N u(x^i)$$

Meanwhile, denote $\tilde{\boldsymbol{\xi}}$ as the PO described in Theorem 35 and $\tilde{\boldsymbol{X}}_t$ as the controlled dynamics under $\overline{\boldsymbol{\xi}}$. From Theorem 35, $\tilde{\boldsymbol{X}}_t \in \mathcal{C}$ almost surely. Therefore there exists a constant $\tilde{c}_N$ such that $\left|\tilde{X}_t^i - \rho\overline{\tilde{\boldsymbol{X}}}_t^{-i}\right| \le \tilde{c}_N$ almost surely.

Hence we have $PoA \le \frac{N\frac{1}{\alpha}(c_N+c)^2 + \sum_{i=1}^N u(x^i)}{N\frac{1}{\alpha}\tilde{c}_N^2}$, where $\tilde{c}_N = \mathrm{D}iam(\mathcal{C})$.

□

Corollary 42.1. *In the degenerate case of $N = 2$ with $\rho = 1$. That is, $h_1(\boldsymbol{x}) = h_2(\boldsymbol{x}) = h\left(\frac{x^1 - x^2}{2}\right)$. Assume $K_1^\pm = k_1 > 0$ and $K_2^\pm = k_2 > 0$, and WLOG $k_2 > k_1 > 0$. Then*

$$PoA(x^1,x^2)=\frac{v^1(x^1,x^2)+v^2(x^1,x^2)}{2\tilde{v}(x^1,x^2)},$$

where v^1, v^2 *are the NE values such that*

$$v^1(x^1,x^2)=\begin{cases} v^1(x^1,x^1+\tilde{c}_2^{(2)}) & x^1-x^2\le -\tilde{c}_2^{(2)},\\ v^1(x^2-\tilde{c}_2^{(1)},x^2)+k_1\left(x^2-x^1-\tilde{c}_2^{(1)}\right), & -\tilde{c}_2^{(2)}\le x^1-x^2\le -\tilde{c}_2^{(1)},\\ -\frac{p_1''(c_2^{(1)})\cosh\left(\sqrt{\alpha}(x^1-x^2)\right)}{\alpha\cosh\left(\tilde{c}_2^{(1)}\sqrt{\alpha}\right)}+p_1(x^1-x^2), & |x^1-x^2|\le c_2^{(1)},\\ v^1(x^2+\tilde{c}_2^{(1)},x^2)+k_1\left(x^1-x^2-\tilde{c}_2^{(1)}\right), & \tilde{c}_2^{(1)}\le x^1-x^2\le \tilde{c}_2^{(2)},\\ v^1(x^1,x^1-\tilde{c}_2^{(2)}), & x^1-x^2\ge \tilde{c}_2^{(2)}, \end{cases} \tag{4.2.11}$$

$$v^2(x^1,x^2)=\begin{cases} v^2(x^1,x^1-\tilde{c}_2^{(2)})+k_2\left(x^1-x^2-\tilde{c}_2^{(2)}\right), & x^2-x^1\le -\tilde{c}_2^{(2)},\\ v^2(x^2+\tilde{c}_2^{(1)},x^2) & -\tilde{c}_2^{(2)}\le x^2-x^1\le -\tilde{c}_2^{(1)},\\ -\frac{p_1''(\tilde{c}_2^{(1)})\cosh\left(\sqrt{\alpha}(x^2-x^1)\right)}{\alpha\cosh\left(c_2^{(1)}\sqrt{\alpha}\right)}+p_1(x^2-x^1), & |x^2-x^1|\le \tilde{c}_2^{(1)},\\ v^2(x^2-\tilde{c}_2^{(1)},x^2), & \tilde{c}_2^{(1)}\le x^2-x^1\le \tilde{c}_2^{(2)},\\ v^2(x^1,x^1+\tilde{c}_2^{(2)})+k_2\left(x^2-x^1-\tilde{c}_2^{(2)}\right), & x^2-x^1\ge \tilde{c}_2^{(2)}, \end{cases} \tag{4.2.12}$$

and $\tilde{v}$ *is the value function of the central controller for game (**N-player-a**) such that*

$$\tilde{v}(x^1,x^2)=\begin{cases} -\frac{p_1''(\tilde{c}_1)\cosh\left(x\sqrt{\alpha}\right)}{\alpha\cosh\left(\tilde{c}_1\sqrt{\alpha}\right)}+p_1(x_1-x_2), & 0\le\ x_1-x_2\le\tilde{c}_1,\\ v(\tilde{c}_1)+\frac{k_1}{2}(x-\tilde{c}_1), & x_1-x-2\ge\tilde{c}_1,\\ v(-x_1,-x_2), & x_1-x_2<0. \end{cases} \tag{4.2.13}$$

Here $\tilde{c}_2^{(2)}>\tilde{c}_2^{(1)}>\tilde{c}_1$, *with* $\tilde{c}_2^{(2)}>\tilde{c}_2^{(1)}$, $\tilde{c}_2^{(i)}>0$ *the unique solutions of*

$$\frac{1}{\sqrt{\alpha}}\tanh\left(\sqrt{\alpha}x\right)=\frac{p_1'(x)-k_i}{p_1''(x)}, \tag{4.2.14}$$

and $\tilde{c}_1$ *the unique solution to*

$$\frac{1}{\sqrt{\alpha}}tanh\left(\sqrt{\alpha}x\right)=\frac{p_1^{'}(x)-\frac{k_1}{2}}{p_1^{''}(x)}. \tag{4.2.15}$$

Proof. Simple analysis confirms that the thresholds $\tilde{c}_2^{(1)}$ for player one and $\tilde{c}_2^{(2)}$ for player two defined in (4.2.15) are unique under bang-bang type of controls. Given the unique thresholds $\tilde{c}_2^{(1)}$ and $\tilde{c}_2^{(2)}$, there are multiple NEs. Player one and player two can coordinate the jumps in region $\left\{(x^1,x^2)\mid |x^1-x^2|>\tilde{c}_2^{(2)}\right\}$. See for example,

$$\xi_t^{2*,+} = \mathbf{1}_{\{x^2-x^1<-c\}}\left(-c-x^2+x^1\right), \tag{4.2.16}$$

$$\xi_t^{2*,-} = \mathbf{1}_{\{x^2-x^1>c\}}\left(c-x^2+x^1\right), \tag{4.2.17}$$

and

$$\begin{aligned}\xi_t^{1*,+} &= \max\left\{0, \max_{0\le u\le t}\{x^2 + B_u^2 + \xi_0^{2*,+} - \xi_0^{2*,-} - x^1 - B_u^1 + \xi_u^{1*,-} - c\}\right\},\\ \xi_t^{1*,-} &= \max\left\{0, \max_{0\le u\le t}\{-x^2 - B_u^2 - \xi_0^{2*,+} + \xi_0^{2*,-} + x^1 + B_u^1 + \xi_u^{1*,+} - c\}\right\},\end{aligned}$$

is an NE solution for any $c \ge \tilde{c}_2^{(2)}$. When player two is in charge of region $\left\{(x^1,x^2) \mid |x^1-x^2| > \tilde{c}_2^{(2)}\right\}$ ($c = \tilde{c}_2^{(2)}$), the cost J^1+J^2 is the highest since it is more expensive for player two to control. Therefore, the "worst" NE given by

$$\xi_t^{2*,+} = \mathbf{1}_{\left\{x^2-x^1<-\tilde{c}_2^{(2)}\right\}}\left(-\tilde{c}_2^{(2)} - x^2 + x^1\right), \tag{4.2.18}$$

$$\xi_t^{2*,-} = \mathbf{1}_{\left\{x^2-x^1>\tilde{c}_2^{(2)}\right\}}\left(\tilde{c}_2^{(2)} - x^2 + x^1\right), \tag{4.2.19}$$

and

$$\begin{aligned}\xi_t^{1*,+} &= \max\left\{0, \max_{0\le u\le t}\{x^2 + B_u^2 + \xi_0^{2*,+} - \xi_0^{2*,-} - x^1 - B_u^1 + \xi_u^{1*,-} - \tilde{c}_2^{(1)}\}\right\},\\ \xi_t^{1*,-} &= \max\left\{0, \max_{0\le u\le t}\{-x^2 - B_u^2 - \xi_0^{2*,+} + \xi_0^{2*,-} + x^1 + B_u^1 + \xi_u^{1*,+} - \tilde{c}_2^{(1)}\}\right\}.\end{aligned}$$

The associate value functions v^1 and v^2 for player one and player two are defined in (4.2.11)-(4.2.12), respectively. In summary,

$$v^1(x^1,x^2) + v^2(x^1,x^2) = \sup_{(\phi^1,\phi^2)\in\mathcal{N}} J^1(x^1,x^2,(\phi^1,\phi^2)) + J^2(x^1,x^2,(\phi^1,\phi^2)),$$

where $\mathcal{N}$ is the set of all admissible NE policies. Meanwhile, the central controller is facing the following optimization problem:

$$\begin{aligned}V(x^1,x^2) &= \min_{(\xi^1,\xi^2)\in\mathcal{U}_2} J^1(x^1,x^2,\xi^1,\xi^2) + J^2(x^1,x^2,\xi^1,\xi^2)\\ &= \min_{(\xi^1,\xi^2)\in\mathcal{U}_2} \mathbb{E}\left[\int_0^\infty e^{-\alpha t}\left(h\left(\frac{X_t^1 - X_t^2}{2}\right)dt + d\xi_t^{1,+} + d\xi_t^{1,-} + d\xi_t^{2,+} + d\xi_t^{2,-}\right)\right]\end{aligned} \tag{4.2.20}$$

subject to

$$\begin{aligned}dX_t^1 &= dB_t^1 + d\xi_t^{1,+} - d\xi_t^{1,-}\\ dX_t^2 &= dB_t^2 + d\xi_t^{2,+} - d\xi_t^{2,-}\end{aligned} \tag{4.2.21}$$

Problem (4.2.20)-(4.2.21) is equivalent to the following problem (4.2.22)-(4.2.23) since $\xi_t^{1,+}$ and $\xi_t^{2,-}$ ($\xi_t^{1,-}$ and $\xi_t^{2,+}$) are equivalent controls to the central controller; and $B_t^1 + B_t^2 \sim \sqrt{2}B_t$, where B_t is a standard Brownian motion.

$$V(x^1,x^2) = \min_{\xi\in\mathcal{U}_1} \mathbb{E}\left[\int_0^\infty e^{-\alpha t}\left(h\left(\frac{X_t}{2}\right)dt + d\xi_t^+ + d\xi_t^-\right)\right] \tag{4.2.22}$$

subject to

$$dX_t = \sqrt{2}dB_t + d\xi_t^+ - d\xi_t^-. \tag{4.2.23}$$

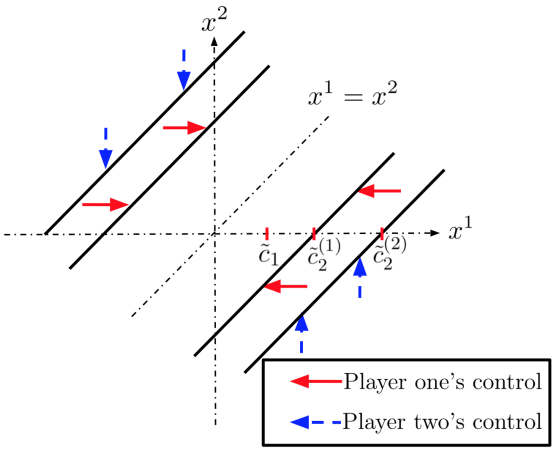
By solving the one-dimensional control problem (4.2.22)-(4.2.23), one can see the following control yields the PO for the game:

$$\begin{aligned} \boldsymbol{\xi}_t^{1*} &= (\xi_t^{1*,+}, \xi_t^{1*,-}), \\ \boldsymbol{\xi}_t^{2*} &= (\xi_t^{2*,+}, \xi_t^{2*,-}) = (0,0), \end{aligned} \tag{4.2.24}$$

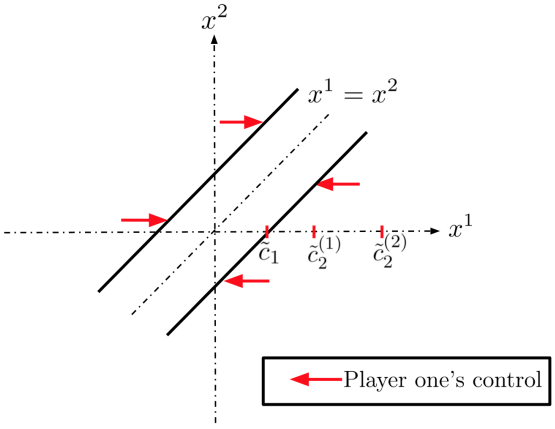
where

$$\begin{aligned} \xi_t^{1*,+} &= \max\left\{0, \max_{0\le u\le t}\left\{-(x^2-x^1) - B_u^2 + B_u^1 + \xi_u^{1*,-} - \tilde{c}_1\right\}\right\}, \\ \xi_t^{1*,-} &= \max\left\{0, \max_{0\le u\le t}\left\{x^2 - x^1 + B_u^2 - B_u^1 + \xi_u^{1*,+} - \tilde{c}_1\right\}\right\}, \end{aligned}$$

□



(a) Worst NE policies: five regions corresponding to five different actions: two action regions from player two, two action regions from player one, and a common non-action region in the middle.



(b) PO policies

Figure 4.1: Worst NE versus PO in PoA.

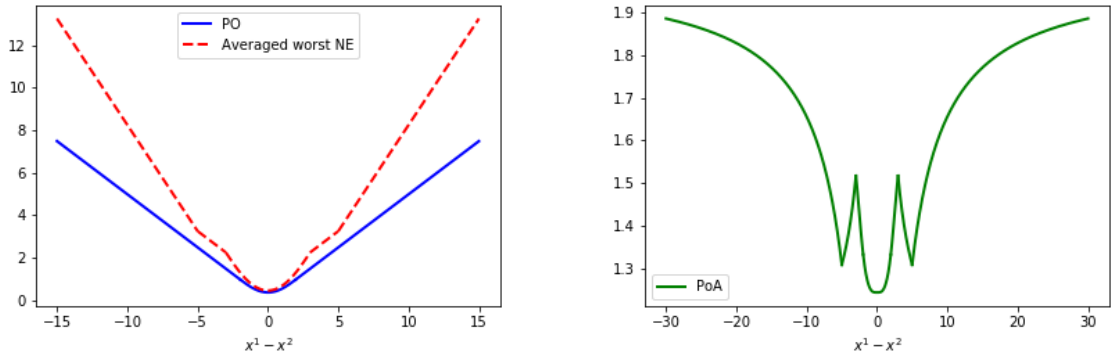
There are some insights from the analysis of PoA.

$\tilde{c}_2^{(1)}$ versus $\tilde{c}_2^{(2)}$. When the cost of controls is different for player one and player two ($k_1 \neq k_2$), the threshold is also different $\left(\tilde{c}_2^{(1)} \neq \tilde{c}_2^{(2)}\right)$. One can show that a bigger cost of control implies a bigger threshold. See Figure 4.1a. Intuition: when control is expensive, player is more reluctant to control, resulting in a larger non-action region.

PO vs NE via PoA. Due to the particular form of the cost function h, PoA only depends on the relative distance $|x^1 - x^2|$. One can show that when $k_1 \neq k_2$, PoA$\to \frac{\max\{k_1,k_2\}}{\min\{k_1,k_2\}}$ when $|x^1 - x^2| \to \infty$. This also suggests that the worst NE is *asymptotically efficient* when two players are further away from each other. Figure 4.2 shows the asymptotic limit of PoA being 2 with $\alpha = 1, k_1 = 1$ and $k_2 = 2$.

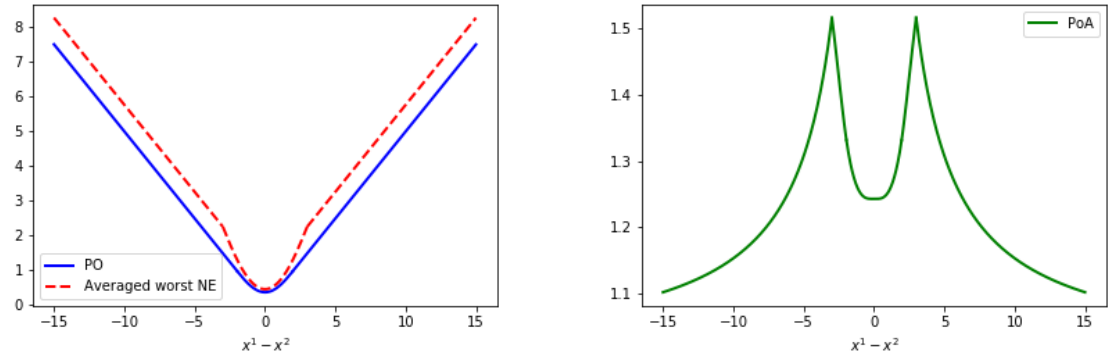
Since the central controller for PO has the freedom to coordinate player one and player two. She will *always* choose the player with the cheaper cost (player one) to control and let the other player (player two) do nothing. This is societally optimal, hence PO is efficient via the centralized coordination.

See Figure 4.3 for a special case when two players are symmetric ($k_1 = k_2$ and $\alpha = 1$). In this case, $\tilde{c}_2^{(1)} \neq \tilde{c}_2^{(2)}$. It is equally optimal for the central controller to pick player 1 or 2. Moreover, at the free boundary between the action and the non-action regions of the worst NE, PoA reaches the largest value (two peaks in Figure 4.3b). Reason: by the NE policy player one frequently exercises controls where by the PO policy there is no control. This leads to the least efficiency of the worst NE.



(a) PO versus Worst NE in game values (b) PoA

Figure 4.2: Worst NE versus PO in PoA (in game values).



(a) PO versus Worst NE in game values (b) PoA

Figure 4.3: Worst NE versus PO in PoA (in game values).

Chapter 5

Learning Mean Field Game

5.1 Introduction

Motivating example. This paper is motivated by the following Ad auction problem for an advertiser. An Ad auction is a stochastic game on an Ad exchange platform among a large number of players, the advertisers. In between the time a web user requests a page and the time the page is displayed, usually within a millisecond, a Vickrey-type of second-best-price auction is run to incentivize interested advertisers to bid for an Ad slot to display advertisement. Each advertiser has limited information before each bid: first, her own *valuation* for a slot depends on an unknown conversion of clicks for the item; secondly, she, should she win the bid, only knows the reward *after* the user's activities on the website are finished. In addition, she has a budget constraint in this repeated auction.

The question is, how should she bid in this online sequential repeated game when there is a *large* population of bidders competing on the Ad platform, with *unknown* distributions of the conversion of clicks and rewards?

Besides the Ad auction, there are many real-world problems involving a large number of players and unknown systems. Examples include massive multi-player online role-playing games [114], high frequency tradings [140], and the sharing economy [97].

Our work. Motivated by these problems, we consider a general framework of simultaneous learning and decision-making in stochastic games with a large population. We formulate a general mean-field-game (GMFG) with incorporation of action distributions, and with unknown rewards and dynamics. This general framework can also be viewed as a generalized version of MFGs of McKean-Vlasov type [1], which is a different paradigm from the classical MFG. It is also beyond the scope of the existing Q-learning framework for Markov decision problem (MDP) with unknown distributions, as MDP is technically equivalent to a single player stochastic game.

On the theory front, this general framework differs from all existing MFGs. We establish under appropriate technical conditions, the existence and uniqueness of the Nash equilibrium (NE) to this GMFG. On the computational front, we show that naively combining Q-learning with the three-step fixed-point approach in classical MFGs yields unstable algorithms. We then propose a Q-learning algorithm with Boltzmann policy (GMF-Q), establish its convergence property and analyze its

computational complexity. Finally, we apply this GMF-Q algorithm to the Ad auction problem, where this GMF-Q algorithm demonstrates its efficiency and robustness in terms of convergence and learning. Moreover, its performance is superior, when compared with existing algorithms for multi-agent reinforcement learning for convergence, stability, and learning accuracy.

Related works. On learning large population games with mean-field approximations, [187] focuses on inverse reinforcement learning for MFGs without decision making, [188] studies an MARL problem with a first-order mean-field approximation term modeling the interaction between one player and all the other finite players, and [125] and [189] consider model-based adaptive learning for MFGs in specific models (*e.g.*, linear-quadratic and oscillator games). More recently, [173] considers reinforcement learning in the classical MFG setting, and proposes a policy-gradient based algorithm and analyzes the so-called local NE. For learning large population games without mean-field approximation, see [117, 100] and the references therein.

In the specific topic of learning auctions with a large number of advertisers, [38] and [115] explore reinforcement learning techniques to search for social optimal solutions with real-word data, and [112] uses MFGs to model the auction system with unknown conversion of clicks within a Bayesian framework.

However, none of these works consider the problem of simultaneous learning and decision-making in a general MFG framework. Neither do they establish the existence and uniqueness of the NE, nor do they present model-free learning algorithms with complexity analysis and convergence to the NE.

5.2 Framework of General MFG (GMFG)

Background: Classical N-player Markovian Game and MFG

Let us first recall the classical N-player game. There are N players in a game. At each step t, the state of player i $(= 1, 2, \cdots, N)$ is $s_t^i \in \mathcal{S} \subseteq \mathbb{R}^d$ and she takes an action $a_t^i \in \mathcal{A} \subseteq \mathbb{R}^p$. Here d, p are positive integers, and $\mathcal{S}$ and $\mathcal{A}$ are compact (for example, finite) state space and action space, respectively. Given the current state profile of N-players $\mathbf{s}_t = (s_t^1, \ldots, s_t^N) \in \mathcal{S}^N$ and the action a_t^i, player i will receive a reward $r^i(\mathbf{s}_t, a_t^i)$ and her state will change to s_{t+1}^i according to a transition probability function $P^i(\mathbf{s}_t, a_t^i)$.

A Markovian game further restricts the admissible policy/control for player i to be of the form $a_t^i = \pi_t^i(\mathbf{s}^t)$. That is, $\pi_t^i : \mathcal{S}^N \to \mathcal{P}(\mathcal{A})$ maps each state profile $\mathbf{s} \in \mathcal{S}^N$ to a randomized action, with $\mathcal{P}(\mathcal{X})$ the space of probability measures on space $\mathcal{X}$. The accumulated reward (a.k.a. the value function) for player i, given the initial state profile $\mathbf{s}$ and the policy profile sequence $\boldsymbol{\pi} := \{\boldsymbol{\pi}_t\}_{t=0}^{\infty}$ with $\boldsymbol{\pi}_t = (\pi_t^1, \ldots, \pi_t^N)$, is then defined as

$$V^i(\mathbf{s}, \boldsymbol{\pi}) := \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r^i(\mathbf{s}_t, a_t^i) \Big| \mathbf{s}_0 = \mathbf{s}\right], \tag{5.2.1}$$

where $\gamma \in (0, 1)$ is the discount factor, $a_t^i \sim \pi_t^i(\mathbf{s}^t)$, and $s_{t+1}^i \sim P^i(\mathbf{s}_t, a_t^i)$. The goal of each player is to maximize her value function over all admissible policy sequences.

In general, this type of stochastic N-player game is notoriously hard to analyze, especially when N is large. Mean field game (MFG), pioneered by [105] and [138], provides an ingenious and tractable aggregation approach to approximate the otherwise challenging N-player stochastic games. The basic idea for an MFG goes as follows. Assume all players are identical, indistinguishable and interchangeable, when $N \to \infty$, one can view the limit of other players' states $\mathbf{s}_t^{-i} = (s_t^1, \ldots, s_t^{i-1}, s_t^{i+1}, \ldots, s_t^N)$ as a population state distribution $\mu_t := \lim_{N\to\infty} \frac{\sum_{j=1, j\neq i}^{N} \mathbf{1}(s_t^j)}{N}$. Due to the homogeneity of the players, one can then focus on a single (representative) player. That is, in an MFG, one may consider instead the following optimization problem,

$$\begin{aligned} \text{maximize}_{\boldsymbol{\pi}} \quad & V(s, \boldsymbol{\pi}, \boldsymbol{\mu}) := \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t, \mu_t) | s_0 = s\right] \\ \text{subject to} \quad & s_{t+1} \sim P(s_t, a_t, \mu_t), \quad a_t \sim \pi_t(s_t, \mu_t), \end{aligned}$$

where $\boldsymbol{\pi} := \{\pi_t\}_{t=0}^{\infty}$ denotes the policy sequence and $\boldsymbol{\mu} := \{\mu_t\}_{t=0}^{\infty}$ the distribution flow. In this MFG setting, at time t, after the representative player chooses her action a_t according to some policy π_t, she will receive reward $r(s_t, a_t, \mu_t)$ and her state will evolve under a *controlled stochastic dynamics* of a mean-field type $P(\cdot|s_t, a_t, \mu_t)$. Here the policy π_t depends on both the current state s_t and the current population state distribution μ_t such that $\pi : \mathcal{S} \times \mathcal{P}(\mathcal{S}) \to \mathcal{P}(\mathcal{A})$.

General MFG (GMFG)

In the classical MFG setting, the reward and the dynamic for each player are known. They depend only on s_t the state of the player, a_t the action of this particular player, and μ_t the population state distribution. In contrast, in the motivating auction example, the reward and the dynamic are unknown; they rely on the actions of *all* players, as well as on s_t and μ_t.

We therefore define the following general MFG (GMFG) framework. At time t, after the representative player chooses her action a_t according to some policy $\pi : \mathcal{S} \times \mathcal{P}(\mathcal{S}) \to \mathcal{P}(\mathcal{A})$, she will receive a reward $r(s_t, a_t, \mathcal{L}_t)$ and her state will evolve according to $P(\cdot|s_t, a_t, \mathcal{L}_t)$, where r and P are possibly unknown. The objective of the player is to solve the following control problem:

$$\begin{aligned} \text{maximize}_{\boldsymbol{\pi}} \quad & V(s, \boldsymbol{\pi}, \boldsymbol{\mathcal{L}}) := \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t, \mathcal{L}_t) | s_0 = s\right] \\ \text{subject to} \quad & s_{t+1} \sim P(s_t, a_t, \mathcal{L}_t), \quad a_t \sim \pi_t(s_t, \mu_t). \end{aligned} \tag{GMFG}$$

Here, $\boldsymbol{\mathcal{L}} := \{\mathcal{L}_t\}_{t=0}^{\infty}$, with $\mathcal{L}_t = \mathbb{P}_{s_t, a_t} \in \mathcal{P}(\mathcal{S} \times \mathcal{A})$ the joint distribution of the state and the action (*i.e.*, the population state-action pair). $\mathcal{L}_t$ has marginal distributions α_t for the population action and μ_t for the population state.

In this framework, we adopt the well-known Nash Equilibrium (NE) for analyzing stochastic games.

Definition 43 (NE for GMFGs)**.** *In* (GMFG)*, a player-population profile* $(\boldsymbol{\pi}^\star, \boldsymbol{\mathcal{L}}^\star) := (\{\pi_t^\star\}_{t=0}^{\infty}, \{\mathcal{L}_t^\star\}_{t=0}^{\infty})$ *is called an NE if*

1. *(Single player side) Fix* $\boldsymbol{\mathcal{L}}^\star$*, for any policy sequence* $\boldsymbol{\pi} := \{\pi_t\}_{t=0}^{\infty}$ *and any initial state* $s \in \mathcal{S}$*,*

$$V\left(s, \boldsymbol{\pi}^\star, \boldsymbol{\mathcal{L}}^\star\right) \geq V\left(s, \boldsymbol{\pi}, \boldsymbol{\mathcal{L}}^\star\right). \tag{5.2.2}$$

2. *(Population side)* $\mathbb{P}_{s_t,a_t} = \mathcal{L}_t^\star$ *for all* $t \geq 0$*, where* $\{s_t, a_t\}_{t=0}^\infty$ *is the dynamics under the policy sequence* $\boldsymbol{\pi}^\star$ *starting from* $s_0 \sim \mu_0^\star$*, with* $a_t \sim \pi_t^\star(s_t, \mu_t^\star)$*,* $s_{t+1} \sim P(\cdot|s_t, a_t, \mathcal{L}_t^\star)$*, and* $\mu_t^\star$ *being the population state marginal of* $\mathcal{L}_t^\star$*.*

The single player side condition captures the optimality of $\boldsymbol{\pi}^\star$, when the population side is fixed. The population side condition ensures the "consistency" of the solution: it guarantees that the state and action distribution flow of the single player does match the population state and action sequence $\boldsymbol{\mathcal{L}}^\star$.

Example: GMFG for the Repeated Auction

Now, consider the repeated Vickrey auction with a budget constraint in Section 5.1. Take a representative advertiser in the auction. Denote $s_t \in \{0, 1, 2, \cdots, s_{\max}\}$ as the budget of this player at time t, where $s_{\max} \in \mathbb{N}^+$ is the maximum budget allowed on the Ad exchange with a unit bidding price. Denote a_t as the bid price submitted by this player and α_t as the bidding/(action) distribution of the population. The reward for this advertiser with bid a_t and budget s_t is

$$r_t = \mathbf{I}_{w_t^M=1}\left[(v_t - a_t^M) - (1+\rho)\mathbf{I}_{s_t<a_t^M}(a_t^M - s_t)\right]. \tag{5.2.3}$$

Here w_t^M takes values 1 and 0, with $w_t^M = 1$ meaning this player winning the bid and 0 otherwise. The probability of winning the bid would depend on M, the index for the game intensity, and α_t. (See discussion on M in Appendix C.8.) The conversion of clicks at time t is v_t and follows an unknown distribution. a_t^M is the value of the second largest bid at time t, taking values from 0 to $s_{\max}$, and depends on both M and $\mathcal{L}_t$. Should the player win the bid, the reward r_t consists of two parts, corresponding to the two terms in (5.2.3). The first term is the profit of wining the auction, as the winner only needs to pay for the second best bid a^M in a Vickrey auction. The second term is the penalty of overshooting if the payment exceeds her budget, with a penalty rate ρ. At each time t, the budget dynamics s_t follows,

$$s_{t+1} = \begin{cases} s_t, & w_t^M \neq 1, \\ s_t - a_t^M, & w_t^M = 1 \text{ and } a_t^M \leq s_t, \\ 0, & w_t^M = 1 \text{ and } a_t^M > s_t. \end{cases}$$

That is, if this player does not win the bid, the budget will remain the same. If she wins and has enough money to pay, her budget will decrease from s_t to $s_t - a_t^M$. However, if she wins but does not have enough money, her budget will be 0 after the payment and there will be a penalty in the reward function. Note that in this game, both the rewards r_t and the dynamics s_t are unknown *a priori*.

In practice, one often modifies the dynamics of s_{t+1} with a non-negative random budget fulfillment $\Delta(s_{t+1})$ after the auction clearing [91], such that

$$\hat{s}_{t+1} = s_{t+1} + \Delta(s_{t+1}). \tag{5.2.4}$$

One may see some particular choices of $\Delta(s_{t+1})$ in the experiment section (Section 5.5).

5.3 Solution for GMFGs

We now establish the existence and uniqueness of the NE to (GMFG), by generalizing the classical fixed-point approach for MFGs to this GMFG setting. (See [105] and [138] for the classical case). It consists of three steps.

Step A. Fix $\boldsymbol{\mathcal{L}} := \{\mathcal{L}_t\}_{t=0}^{\infty}$, (GMFG) becomes the classical optimization problem. Indeed, with $\boldsymbol{\mathcal{L}}$ fixed, the population state distribution sequence $\boldsymbol{\mu} := \{\mu_t\}_{t=0}^{\infty}$ is also fixed, hence the space of admissible policies is reduced to the single-player case. Solving (GMFG) is now reduced to finding a policy sequence $\pi^{\star}_{t,\boldsymbol{\mathcal{L}}} \in \Pi := \{\pi \,|\, \pi : \mathcal{S} \to \mathcal{P}(\mathcal{A})\}$ over all admissible $\boldsymbol{\pi}_{\boldsymbol{\mathcal{L}}} = \{\pi_{t,\boldsymbol{\mathcal{L}}}\}_{t=0}^{\infty}$, to maximize

$$\begin{aligned} V(s, \boldsymbol{\pi}_{\boldsymbol{\mathcal{L}}}, \boldsymbol{\mathcal{L}}) := \quad & \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t, \mathcal{L}_t) | s_0 = s\right], \\ \text{subject to} \quad & s_{t+1} \sim P(s_t, a_t, \mathcal{L}_t), \quad a_t \sim \pi_{t,\boldsymbol{\mathcal{L}}}(s_t). \end{aligned}$$

Notice that with $\boldsymbol{\mathcal{L}}$ fixed, one can safely suppress the dependency on μ_t in the admissible policies. Moreover, given this fixed $\boldsymbol{\mathcal{L}}$ sequence and the solution $\boldsymbol{\pi}^{\star}_{\boldsymbol{\mathcal{L}}} := \{\pi^{\star}_{t,\boldsymbol{\mathcal{L}}}\}_{t=0}^{\infty}$, one can define a mapping from the fixed population distribution sequence $\boldsymbol{\mathcal{L}}$ to an arbitrarily chosen optimal randomized policy sequence. That is,

$$\Gamma_1 : \{\mathcal{P}(\mathcal{S} \times \mathcal{A})\}_{t=0}^{\infty} \to \{\Pi\}_{t=0}^{\infty},$$

such that $\boldsymbol{\pi}^{\star}_{\boldsymbol{\mathcal{L}}} = \Gamma_1(\boldsymbol{\mathcal{L}})$. Note that this $\boldsymbol{\pi}^{\star}_{\boldsymbol{\mathcal{L}}}$ sequence satisfies the single player side condition in Definition 43 for the population state-action pair sequence $\boldsymbol{\mathcal{L}}$. That is, $V\left(s, \boldsymbol{\pi}^{\star}_{\boldsymbol{\mathcal{L}}}, \boldsymbol{\mathcal{L}}\right) \geq V\left(s, \boldsymbol{\pi}, \boldsymbol{\mathcal{L}}\right)$, for any policy sequence $\boldsymbol{\pi} = \{\pi_t\}_{t=0}^{\infty}$ and any initial state $s \in \mathcal{S}$.

As in the classical MFG literature [105], a feedback regularity (FR) condition is needed for the analysis of Step A.

Assumption 1. *There exists a constant $d_1 \geq 0$, such that for any $\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{L}}' \in \{\mathcal{P}(\mathcal{S} \times \mathcal{A})\}_{t=0}^{\infty}$,*

$$D(\Gamma_1(\boldsymbol{\mathcal{L}}), \Gamma_1(\boldsymbol{\mathcal{L}}')) \leq d_1 \mathcal{W}_1(\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{L}}'), \tag{5.3.1}$$

where

$$\begin{aligned} D(\boldsymbol{\pi}, \boldsymbol{\pi}') &:= \sup_{s \in \mathcal{S}} \mathcal{W}_1(\boldsymbol{\pi}(s), \boldsymbol{\pi}'(s)) = \sup_{s \in \mathcal{S}} \sup_{t \in \mathbb{N}} W_1(\pi_t(s), \pi'_t(s)), \\ \mathcal{W}_1(\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{L}}') &:= \sup_{t \in \mathbb{N}} W_1(\mathcal{L}_t, \mathcal{L}'_t), \end{aligned} \tag{5.3.2}$$

and W_1 is the ℓ_1-Wasserstein distance between probability measures [85, 180, 160].

Step B. Based on the analysis in Step A and $\boldsymbol{\pi}^{\star}_{\boldsymbol{\mathcal{L}}} = \{\pi^{\star}_{t,\boldsymbol{\mathcal{L}}}\}_{t=0}^{\infty}$, update the initial sequence $\boldsymbol{\mathcal{L}}$ to $\boldsymbol{\mathcal{L}}'$ following the controlled dynamics $P(\cdot|s_t, a_t, \mathcal{L}_t)$.

Accordingly, for any admissible policy sequence $\boldsymbol{\pi} \in \{\Pi\}_{t=0}^{\infty}$ and a joint population state-action pair sequence $\boldsymbol{\mathcal{L}} \in \{\mathcal{P}(\mathcal{S} \times \mathcal{A})\}_{t=0}^{\infty}$, define a mapping $\Gamma_2 : \{\Pi\}_{t=0}^{\infty} \times \{\mathcal{P}(\mathcal{S} \times \mathcal{A})\}_{t=0}^{\infty} \to \{\mathcal{P}(\mathcal{S} \times \mathcal{A})\}_{t=0}^{\infty}$ as follows:

$$\Gamma_2(\boldsymbol{\pi}, \boldsymbol{\mathcal{L}}) := \hat{\boldsymbol{\mathcal{L}}} = \{\mathbb{P}_{s_t, a_t}\}_{t=0}^{\infty}, \tag{5.3.3}$$

where $s_{t+1} \sim \mu_t P(\cdot|\cdot, a_t, \mathcal{L}_t)$, $a_t \sim \pi_t(s_t)$, $s_0 \sim \mu_0$, and μ_t is the population state marginal of $\mathcal{L}_t$.

One needs a standard assumption in this step.

Assumption 2. *There exist constants $d_2,\ d_3 \geq 0$, such that for any admissible policy sequences $\boldsymbol{\pi}, \boldsymbol{\pi}^1, \boldsymbol{\pi}^2$ and joint distribution sequences $\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{L}}^1, \boldsymbol{\mathcal{L}}^2$,*

$$\mathcal{W}_1(\Gamma_2(\boldsymbol{\pi}^1, \boldsymbol{\mathcal{L}}), \Gamma_2(\boldsymbol{\pi}^2, \boldsymbol{\mathcal{L}})) \leq d_2 D(\boldsymbol{\pi}^1, \boldsymbol{\pi}^2), \tag{5.3.4}$$

$$\mathcal{W}_1(\Gamma_2(\boldsymbol{\pi}, \boldsymbol{\mathcal{L}}^1), \Gamma_2(\boldsymbol{\pi}, \boldsymbol{\mathcal{L}}^2)) \leq d_3 \mathcal{W}_1(\boldsymbol{\mathcal{L}}^1, \boldsymbol{\mathcal{L}}^2). \tag{5.3.5}$$

Assumption 2 can be reduced to Lipschitz continuity and boundedness of the transition dynamics P. (See the Appendix for more details.)

Step C. Repeat Step A and Step B until $\boldsymbol{\mathcal{L}}'$ matches $\boldsymbol{\mathcal{L}}$.

This step is to take care of the population side condition. To ensure the convergence of the combined step A and step B, it suffices if $\Gamma : \{\mathcal{P}(\mathcal{S} \times \mathcal{A})\}_{t=0}^{\infty} \to \{\mathcal{P}(\mathcal{S} \times \mathcal{A})\}_{t=0}^{\infty}$ is a contractive mapping under the $\mathcal{W}_1$ distance, with $\Gamma(\boldsymbol{\mathcal{L}}) := \Gamma_2(\Gamma_1(\boldsymbol{\mathcal{L}}), \boldsymbol{\mathcal{L}})$. Then by the Banach fixed point theorem and the completeness of the related metric spaces, there exists a unique NE to the GMFG.

In summary, we have

Theorem 44 (Existence and Uniqueness of GMFG solution)**.** *Given Assumptions 1 and 2, and assume $d_1 d_2 + d_3 < 1$. Then there exists a unique NE to* (GMFG)*.*

5.4 RL Algorithms for GMFGs

In this section, we design the computational algorithm for the GMFG. Since the reward and transition distributions are unknown, this is simultaneously learning the system and finding the NE of the game. We will focus on the case with finite state and action spaces, *i.e.*, $|\mathcal{S}|, |\mathcal{A}| < \infty$. We will look for stationary (time independent) NEs. Accordingly, we abbreviate $\boldsymbol{\pi} := \{\pi\}_{t=0}^{\infty}$ and $\boldsymbol{\mathcal{L}} : \{\mathcal{L}\}_{t=0}^{\infty}$ as π and $\mathcal{L}$, respectively. This stationarity property enables developing appropriate time-independent Q-learning algorithm, suitable for an infinite time horizon game. Modification from the GMFG framework to this special stationary setting is straightforward, and is left in the Appendix.

The algorithm consists of two steps, parallel to Step A and Step B in Section 5.3.

Step 1: Q-learning with stability for fixed $\mathcal{L}$. With $\mathcal{L}$ fixed, it becomes a standard learning problem for an infinite horizon MDP. We will focus on the Q-learning algorithm [175, 161].

The Q-learning algorithm approximates the value iteration by stochastic approximation. At each step with the state s and an action a, the system reaches state s' according to the controlled dynamics and the Q-function is updated according to

$$Q_{\mathcal{L}}(s,a) \leftarrow\ (1 - \beta_t(s,a)) Q_{\mathcal{L}}(s,a) + \beta_t(s,a) \left[r(s,a,\mathcal{L}) + \gamma \max_{\tilde{a}} Q_{\mathcal{L}}(s', \tilde{a}) \right], \tag{5.4.1}$$

where the step size $\beta_t(s,a)$ can be chosen as ([77])

$$\beta_t(s,a) = \begin{cases} |\#(s,a,t) + 1|^{-h}, & (s,a) = (s_t, a_t), \\ 0, & \text{otherwise.} \end{cases}$$

with $h \in (1/2, 1)$. Here $\#(s, a, t)$ is the number of times up to time t that one visits the pair (s, a). The algorithm then proceeds to choose action a' based on $Q_{\mathcal{L}}$ with appropriate exploration strategies, including the ϵ-greedy strategy.

After obtaining the approximate $\hat{Q}^{\star}_{\mathcal{L}}$, in order to retrieve an approximately optimal policy, it would be natural to define an **argmax-e** operator so that actions with equal maximum Q-values would have equal probabilities to be selected. Unfortunately, the discontinuity and sensitivity of **argmax-e** could lead to an unstable algorithm (see Figure 5.4 for the corresponding naive Algorithm 3 in Appendix). [1]

Instead, we consider a Boltzmann policy based on the operator $\mathbf{softmax}_c : \mathbb{R}^n \to \mathbb{R}^n$, defined as

$$\mathbf{softmax}_c(x)_i = \frac{\exp(cx_i)}{\sum_{j=1}^{n} \exp(cx_j)}. \tag{5.4.2}$$

This operator is smooth and close to the **argmax-e** (see Lemma 56 in the Appendix). Moreover even though Boltzmann policies are not optimal, the difference between the Boltzmann and the optimal one can always be controlled by choosing the hyper-parameter c appropriately in the **softmax** operator (5.4.2).

Step 2: error control in updating $\mathcal{L}$. Given the sub-optimality of the Boltzmann policy, one needs to characterize the difference between the optimal policy and the non-optimal ones. In particular, one can define the action gap between the best action and the second best action in terms of the Q-value as $\delta^s(\mathcal{L}) := \max_{a' \in \mathcal{A}} Q^{\star}_{\mathcal{L}}(s, a') - \max_{a \notin \operatorname{argmax}_{a \in \mathcal{A}} Q^{\star}_{\mathcal{L}}(s,a)} Q^{\star}_{\mathcal{L}}(s, a) > 0$. Action gap is important for approximation algorithms [18], and are closely related to the problem-dependent bounds for regret analysis in reinforcement learning and multi-armed bandits, and advantage learning algorithms including A2C [150].

The problem is: in order for the learning algorithm to converge in terms of $\mathcal{L}$ (Theorem 45), one needs to ensure a definite differentiation between the optimal policy and the sub-optimal ones. This is problematic as the infimum of $\delta^s(\mathcal{L})$ over an infinite number of $\mathcal{L}$ can be 0. To address this, the population distribution at step k, say $\mathcal{L}_k$, needs to be projected to a finite grid, called ϵ-net. The relation between the ϵ-net and action gaps is as follows:

For any $\epsilon > 0$, there exist a positive function $\phi(\epsilon)$ and an ϵ-net $S_\epsilon := \{\mathcal{L}^{(1)}, \ldots, \mathcal{L}^{(N_\epsilon)}\} \subseteq \mathcal{P}(\mathcal{S} \times \mathcal{A})$, with the properties that $\min_{i=1,\ldots,N_\epsilon} d_{TV}(\mathcal{L}, \mathcal{L}^{(i)}) \leq \epsilon$ for any $\mathcal{L} \in \mathcal{P}(\mathcal{S} \times \mathcal{A})$, and that $\max_{a' \in \mathcal{A}} Q^{\star}_{\mathcal{L}^{(i)}}(s, a') - Q^{\star}_{\mathcal{L}^{(i)}}(s, a) \geq \phi(\epsilon)$ for any $i = 1, \ldots, N_\epsilon$, $s \in \mathcal{S}$, and any $a \notin argmax_{a \in \mathcal{A}} Q^{\star}_{\mathcal{L}^{(i)}}(s, a)$.

Here the existence of ϵ-nets is trivial due to the compactness of the probability simplex $\mathcal{P}(\mathcal{S} \times \mathcal{A})$, and the existence of $\phi(\epsilon)$ comes from the finiteness of the action set $\mathcal{A}$.

In practice, $\phi(\epsilon)$ often takes the form of $D\epsilon^{\alpha}$ with $D > 0$ and the exponent $\alpha > 0$ characterizing the decay rate of the action gaps, and S_ϵ takes a uniform grid with appropriate grid sizes.

Finally, to enable Q-learning, it is assumed that one has access to a population simulator (See [159, 182]). That is, for any policy $\pi \in \Pi$, given the current state $s \in \mathcal{S}$, for any population

[1] **argmax-e** is not continuous: Let $x = (1, 1)$, then **argmax-e**$(x) = (1/2, 1/2)$. For any $\epsilon > 0$, let $y = (1, 1 - \epsilon)$, then **argmax-e**$(y) = (1, 0)$.

distribution $\mathcal{L}$, one can obtain the next state $s' \sim P(\cdot|s, \pi(s,\mu), \mathcal{L})$, a reward $r = r(s, \pi(s,\mu), \mathcal{L})$, and the next population distribution $\mathcal{L}' = \mathbb{P}_{s',\pi(s',\mu)}$. For brevity, we denote the simulator as $(s', r, \mathcal{L}') = \mathcal{G}(s, \pi, \mathcal{L})$. Here μ is the state marginal distribution of $\mathcal{L}$.

In summary, we propose the following Algorithm 2.

Algorithm 2 Q-learning for GMFGs (GMF-Q)

1: **Input**: Initial $\mathcal{L}_0$, tolerance $\epsilon > 0$.
2: **for** $k = 0, 1, \cdots$ **do**
3: Perform Q-learning for T_k iterations to find the approximate Q-function $\hat{Q}^{\star}_k(s,a) = \hat{Q}^{\star}_{\mathcal{L}_k}(s,a)$ of an MDP with dynamics $P_{\mathcal{L}_k}(s'|s,a)$ and rewards $r_{\mathcal{L}_k}(s,a)$.
4: Compute $\pi_k \in \Pi$ with $\pi_k(s) = \mathbf{softmax}_c(\hat{Q}^{\star}_k(s,\cdot))$.
5: Sample $s \sim \mu_k$, where μ_k is the population state marginal of $\mathcal{L}_k$, and obtain $\tilde{\mathcal{L}}_{k+1}$ from $\mathcal{G}(s, \pi_k, \mathcal{L}_k)$.
6: Find $\mathcal{L}_{k+1} = \mathbf{Proj}_{S_\epsilon}(\tilde{\mathcal{L}}_{k+1})$

Here $\mathbf{Proj}_{S_\epsilon}(\mathcal{L}) = \text{argmin}_{\mathcal{L}^{(1)},\ldots,\mathcal{L}^{(N_\epsilon)}} d_{TV}(\mathcal{L}^{(i)}, \mathcal{L})$. See Lemma 57 and Theorem 45 for details about the choices of hyper-parameters c and T_k.

In the special case when the rewards $r_{\mathcal{L}}$ and transition dynamics $P(\cdot|s,a,\mathcal{L})$ are known, one can replace the Q-learning step in the above Algorithm 2 by a value iteration, resulting in the GMF-V Algorithm 4 in the Appendix.

We next show the convergence of this GMF-Q algorithm (Algorithm 2) to an ϵ-Nash of (GMFG), with complexity analysis.

Theorem 45 (Convergence and complexity of GMF-Q)**.** *Assume the same conditions in Theorem 44 and Lemma 57 in the Appendix. For any tolerances $\epsilon,\ \delta > 0$, set $\delta_k = \delta/K_{\epsilon,\eta}$, $\epsilon_k = (k+1)^{-(1+\eta)}$ for some $\eta \in (0,1]$ $(k = 0, \ldots, K_{\epsilon,\eta}-1)$, $T_k = T^{\mathcal{M}_{\mathcal{L}_k}}(\delta_k, \epsilon_k)$ (defined in Lemma 57 in the Appendix) and $c = \frac{\log(1/\epsilon)}{\phi(\epsilon)}$. Then with probability at least $1-2\delta$, $W_1(\mathcal{L}_{K_{\epsilon,\eta}}, \mathcal{L}^\star) \le C\epsilon$.*
Moreover, the total number of iterations $T = \sum_{k=0}^{K_{\epsilon,\eta}-1} T^{\mathcal{M}_{\mathcal{L}_k}}(\delta_k, \epsilon_k)$ is bounded by

$$T = O\left(K_{\epsilon,\eta}^{1+\frac{4}{h}} \left(\log(K_{\epsilon,\eta}/\delta)\right)^{\frac{2}{1-h}+\frac{2}{h}+3}\right). \tag{5.4.3}$$

Here $K_{\epsilon,\eta} := \left\lceil 2\max\left\{(\eta\epsilon)^{-1/\eta}, \log_d(\epsilon/\max\{diam(\mathcal{S})diam(\mathcal{A}), 1\}) + 1\right)\right\}\right\rceil$ is the number of outer iterations, h is the step-size exponent in Q-learning (defined in Lemma 57 in the Appendix), and the constant C is independent of δ, ϵ and η.

The proof of Theorem 45 in the Appendix depends on the Lipschitz continuity of the **softmax** operator [82], the closeness between **softmax** and the **argmax-e** (Lemma 56 in the Appendix), and the complexity of Q-learning for the MDP (Lemma 57 in the Appendix). Lemma 57 also provides guidance on how to choose the number of inner iterations T_k in Algorithm 2.

Table 5.1: Q-table with $T_k^{\text{GMF-V}} = 5000$.

$T_k^{\text{GMF-Q}}$	1000	3000	5000	10000
ΔQ	0.21263	0.1294	0.10258	0.0989

5.5 Experiment: Repeated Auction Game

In this section, we report the performance of the proposed GMF-Q Algorithm. The objectives of the experiments include 1) testing the convergence, stability, and learning ability of GMF-Q in the GMFG setting, and 2) comparing GMF-Q with existing multi-agent reinforcement learning algorithms, including IL algorithm and MF-Q algorithm.

We take the GMFG framework for the repeated auction game from Section 5.2. Here each advertiser learns to bid in the auction with a budget constraint.

Parameters. The model parameters are set as: $|\mathcal{S}| = |\mathcal{A}| = 10$, the overbidding penalty $\rho = 0.2$, the distributions of the conversion rate $v \sim$ uniform[4], and the competition intensity index $M = 5$. The random fulfillment is chosen as: if $s < s_{\max}$, $\Delta(s) = 1$ with probability $\frac{1}{2}$ and $\Delta(s) = 0$ with probability $\frac{1}{2}$; if $s = s_{\max}$, $\Delta(s) = 0$.

The algorithm parameters are (unless otherwise specified): the temperature parameter $c = 4.0$, the discount factor $\gamma = 0.8$, the parameter h from Lemma 57 in the Appendix being $h = 0.87$, and the baseline inner iteration being 2000. Recall that for GMF-Q, both v and the dynamics of P for s are unknown *a priori*. The 90%-confidence intervals are calculated with 20 sample paths.

Performance evaluation in the GMFG setting. Our experiment shows that the GMF-Q Algorithm is efficient and robust, and learns well.

Convergence and stability of GMF-Q. GMF-Q is efficient and robust. First, GMF-Q converges after about 10 outer iterations; secondly, as the number of inner iterations increases, the error decreases (Figure 5.2); and finally, the convergence is robust with respect to both the change of number of states and the initial population distribution (Figure 5.3).

In contrast, the Naive algorithm does not converge even with 10000 inner iterations, and the joint distribution $\mathcal{L}_t$ keeps fluctuating (Figure 5.4).

Learning accuracy of GMF-Q. GMF-Q learns well. Its learning accuracy is tested against its special form GMF-V (Appendix C.7), with the latter assuming a known distribution of conversation rate v and the dynamics P for the budget s. The relative L_2 distance between the Q-tables of these two algorithms is $\Delta Q := \frac{\|Q_{\text{GMF-V}} - Q_{\text{GMF-Q}}\|_2}{\|Q_{\text{GMF-V}}\|_2} = 0.098879$. This implies that GMF-Q learns the true GMFG solution with 90-percent accuracy with 10000 inner iterations.

The heatmap in Figure 5.1a is the Q-table for GMF-Q Algorithm after 20 outer iterations. Within each outer iteration, there are $T_k^{\text{GMF-Q}} = 10000$ inner iterations. The heatmap in Figure 5.1b is the Q-table for GMF-Q Algorithm after 20 outer iterations. Within each outer iteration, there are $T_k^{\text{GMF-V}} = 5000$ inner iterations.

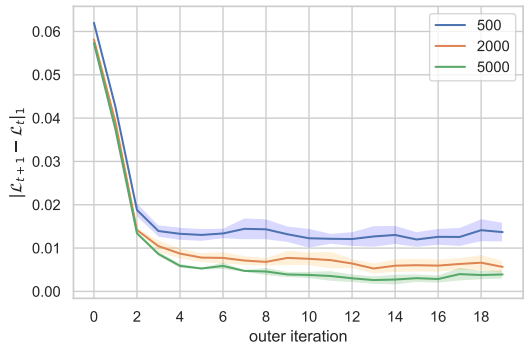


Figure 5.2: Convergence with different number of inner iterations.

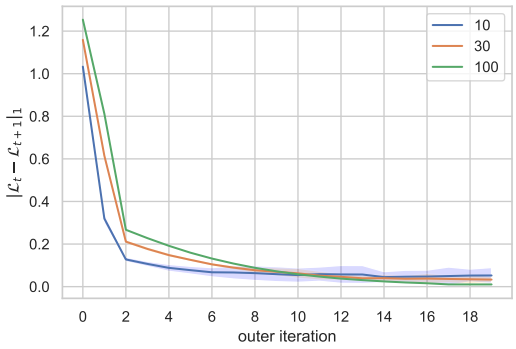
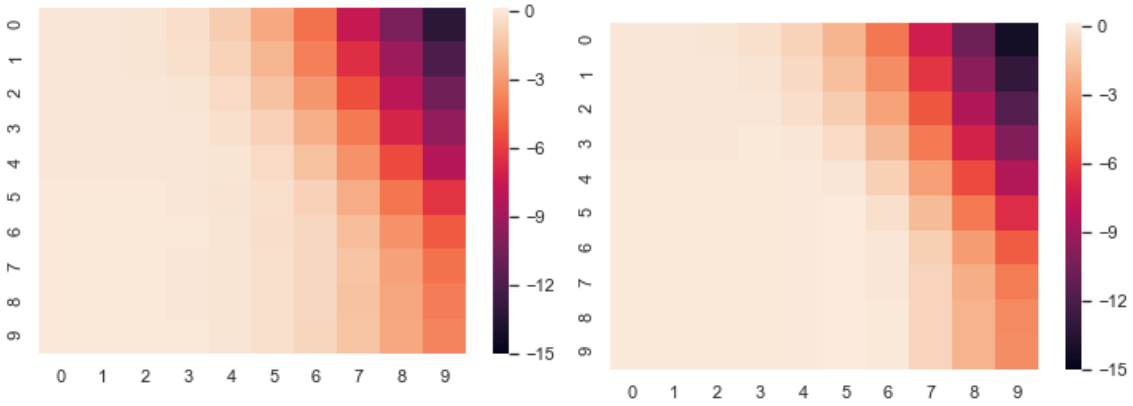


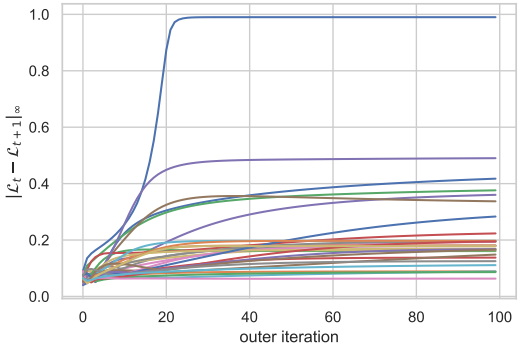
Figure 5.3: Convergence with different number of states.



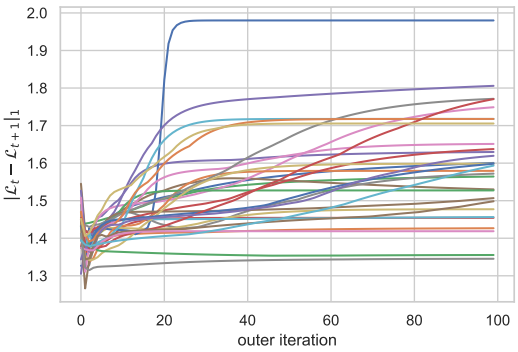
(a) GMF-Q.

(b) GMF-V.

Figure 5.1: Q-tables: GMF-Q vs. GMF-V.

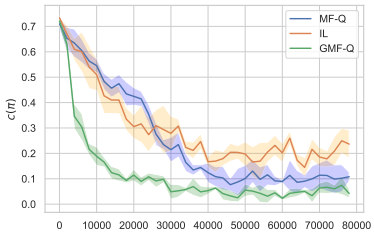


(a) fluctuation in l_∞.

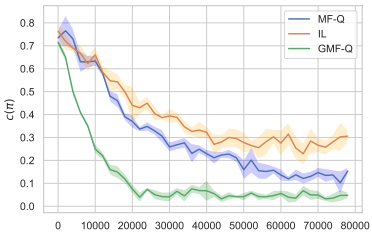


(b) fluctuation in l_1.

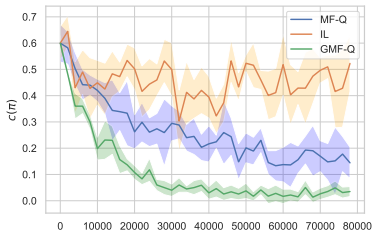
Figure 5.4: Fluctuations of Naive Algorithm (30 sample paths).



(a) $|\mathcal{S}| = |\mathcal{A}| = 10, N = 20$.



(b) $|\mathcal{S}| = |\mathcal{A}| = 20, N = 20$.



(c) $|\mathcal{S}| = |\mathcal{A}| = 10, N = 40$.

Figure 5.5: Learning accuracy based on $C(\boldsymbol{\pi})$.

Comparison with existing algorithms for N-player games. To test the effectiveness of GMF-Q for approximating N-player games, we next compare GMF-Q with IL algorithm and MF-Q algorithm. IL algorithm [176] considers N independent players and each player solves a

decentralized reinforcement learning problem ignoring other players in the system. The MF-Q algorithm [188] extends the NASH-Q Learning algorithm for the N-player game introduced in [101], adds the aggregate actions ($\bar{\boldsymbol{a}}_{-i} = \frac{\sum_{j\neq i} a_j}{N-1}$) from the opponents, and works for the class of games where the interactions are only through the average actions of N players.

Performance metric. We adopt the following metric to measure the difference between a given policy π and an NE (here $\epsilon_0 > 0$ is a safeguard, and is taken as 0.1 in the experiments):

$$C(\boldsymbol{\pi}) = \frac{1}{N|\mathcal{S}|^N} \sum\nolimits_{i=1}^{N} \sum\nolimits_{\boldsymbol{s}\in\mathcal{S}^N} \frac{\max_{\pi^i} V_i(\boldsymbol{s}, (\boldsymbol{\pi}^{-i}, \pi^i)) - V_i(\boldsymbol{s}, \boldsymbol{\pi})}{|\max_{\pi^i} V_i(\boldsymbol{s}, (\boldsymbol{\pi}^{-i}, \pi^i))| + \epsilon_0}.$$

Clearly $C(\boldsymbol{\pi}) \geq 0$, and $C(\boldsymbol{\pi}^*) = 0$ if and only if $\boldsymbol{\pi}^*$ is an NE. Policy $\arg\max_{\pi_i} V_i(\boldsymbol{s}, (\boldsymbol{\pi}^{-i}, \pi_i))$ is called the best response to $\boldsymbol{\pi}^{-i}$. A similar metric without normalization has been adopted in [158].

Our experiment (Figure 5.5) shows that GMF-Q is superior in terms of convergence rate, accuracy, and stability for approximating an N-player game: GMF-Q converges faster than IL and MF-Q, with the smallest error, and with the lowest variance, as ϵ-net improves the stability.

For instance, when $N = 20$, IL Algorithm converges with the largest error 0.220. The error from MF-Q is 0.101, smaller than IL but still bigger than the error from GMF-Q. The GMF-Q converges with the lowest error 0.065. Moreover, as N increases, the error of GMF-Q deceases while the errors of both MF-Q and IL increase significantly. As $|\mathcal{S}|$ and $|\mathcal{A}|$ increase, GMF-Q is robust with respect to this increase of dimensionality, while both MF-Q and IL clearly suffer from the increase of the dimensionality with decreased convergence rate and accuracy. Therefore, GMF-Q is more scalable than IL and MF-Q, when the system is complex and the number of players N is large.

5.6 Conclusion

This paper builds a GMFG framework for simultaneous learning and decision-making, establishes the existence and uniqueness of NE, and proposes a Q-learning algorithm GMF-Q with convergence and complexity analysis. Experiments demonstrate superior performance of GMF-Q.

Bibliography

[1] B. Acciaio, J. Backhoff, and R. Carmona. "Extended mean field control problems: stochastic maximum principle and transport perspective". In: *Arxiv Preprint:1802.05754* (2018).

[2] Yves Achdou and Mathieu Lauriere. "On the system of partial differential equations arising in mean field type control". In: *arXiv preprint arXiv:1503.05044* (2015).

[3] René Aïd, Matteo Basei, and Huyên Pham. "The coordination of centralised and distributed generation". In: *arXiv preprint arXiv:1705.01302* (2017).

[4] David Aldous. ""Up the River" game stoSry". In: (2002). Available at http://www.stat.berkeley.edu/dous/Research/OP/river.pdf.

[5] Eitan Altman and Tamer Basar. "Multiuser rate-based flow control". In: *IEEE Transactions on Communications* 46.7 (1998), pp. 940–949.

[6] Eitan Altman et al. "A survey on networking games in telecommunications". In: *Computers & Operations Research* 33.2 (2006), pp. 286–311.

[7] Luis Alvarez and Larry Shepp. "Optimal harvesting of stochastically fluctuating populations". In: *Journal of Mathematical Biology* 37.2 (1998), pp. 155–177.

[8] R. Atar and A. Budhiraja. "Singular control with state constraints on unbounded domain". In: *The Annals of Probability* 34.5 (2006), pp. 1864–1909.

[9] Adrian D. Banner, Robert Fernholz, and Ioannis Karatzas. "Atlas models of equity markets". In: *Annals of Applied Probability* 15.4 (2005), pp. 2296–2330. ISSN: 1050-5164. DOI: 10.1214/105051605000000449. URL: http://dx.doi.org/10.1214/105051605000000449.

[10] M. Bardi. "Explicit solutions of some linear-quadratic mean field games". In: *Networks and Heterogeneous Media* 7.2 (2012), pp. 243–261.

[11] M. Bardi and F.S. Priuli. "Linear-quadratic N-person and mean-field games with ergodic cost". In: *SIAM Journal on Control and Optimization* 52.5 (2014), pp. 3022–3052.

[12] Martino Bardi. "Explicit solutions of some linear-quadratic mean field games". In: (2012).

[13] Matteo Basei, Haoyang Cao, and Xin Guo. "Nonzero-sum stochastic games with impulse controls". In: *arXiv preprint arXiv:1901.08085* (2019).

[14] John Bather and Herman Chernoff. "Sequential decisions in the control of a space-ship (finite fuel)". In: *Journal of Applied Probability* 4.3 (1967), pp. 584–604.

[15] Francis M Bator. "The simple analytics of welfare maximization". In: *The American Economic Review* 47.1 (1957), pp. 22–59.

[16] Dario Bauso, Raffaele Pesenti, and Marco Tolotti. "Opinion dynamics and stubbornness via multi-population mean-field games". In: *Journal of Optimization Theory and Applications* 170.1 (2016), pp. 266–293.

[17] Erhan Bayraktar and Yuchong Zhang. "Terminal Ranking Games". In: *arXiv preprint arXiv:1906.09628* (2019).

[18] M. G. Bellemare et al. "Increasing the Action Gap: new Operators for Reinforcement Learning". In: *AAAI Conference on Artificial Intelligence*. 2016, pp. 1476–1483.

[19] V. Beneš, L. Shepp, and H. Witsenhausen. "Some solvable stochastic control problems". In: *Stochastics: An International Journal of Probability and Stochastic Processes* 4.1 (1980), pp. 39–83.

[20] Václav E Beneš, Lawrence A Shepp, and Hans S Witsenhausen. "Some solvable stochastic control problemst". In: *Stochastics: An International Journal of Probability and Stochastic Processes* 4.1 (1980), pp. 39–83.

[21] A. Bensoussan and J. Frehse. "Stochastic games for N players". In: *Journal of Optimization Theory and Applications* 105.3 (2000), pp. 543–565.

[22] A. Bensoussan and J. Frehse. "Stochastic games with risk sensitive payoffs for N players". In: *Stochastic Analysis and Related Topics VIII*. Vol. 55. 4. Springer, 2003, pp. 29–66.

[23] Alain Bensoussan, Tao Huang, and Mathieu Laurière. "Mean field control and mean field game models with several populations". In: *arXiv preprint arXiv:1810.00783* (2018).

[24] A. Bensoussan et al. "Linear-quadratic mean field games". In: *Journal of Optimization Theory and Applications* 169.2 (2016), pp. 496–529.

[25] Anup Biswas et al. "On viscosity solution of hjb equations with state constraints and reflection control". In: *SIAM Journal on Control and Optimization* 55.1 (2017), pp. 365–396.

[26] Tomas Björk, Mark HA Davis, and Camilla Landén. "Optimal investment under partial information". In: *Mathematical Methods of Operations Research* 71.2 (2010), pp. 371–399.

[27] F. Bolley. "Separability and completeness for the Wasserstein distance". In: *Séminaire de Probabilités XLI* (2008), pp. 371–377.

[28] David Bridge and Steven Shreve. "Multi-dimensional finite-fuel singular stochastic control". In: *Applied Stochastic Analysis*. 1992, pp. 38–58.

[29] Peter Brockwell, Sidney Resnick, and Richard Tweedie. "Storage processes with general release rule and additive inputs". In: *Advances in Applied Probability* 14.2 (1982), pp. 392–433.

[30] A. Budhiraja and K. Ross. "Existence of optimal controls for singular control problems with state constraints". In: *The Annals of Applied Probability* 16.4 (2006), pp. 2235–2255.

[31] K. Burdzy, W. Kang, and K. Ramanan. "The Skorokhod problem in a time-dependent interval". In: *Stochastic Processes and their Applications* 119.2 (2009), pp. 428–452.

[32] Krzysztof Burdzy, Zhen-Qing Chen, and John Sylvester. "The heat equation and reflected Brownian motion in time-dependent domains". In: *The Annals of Probability* 32.1B (2004), pp. 775–804.

[33] Krzysztof Burdzy, Zhen-Qing Chen, and John Sylvester. "The heat equation and reflected Brownian motion in time-dependent domains". In: *Annals of Probability* 32.1B (2004), pp. 775–804.

[34] Krzysztof Burdzy, Zhen-Qing Chen, and John Sylvester. "The heat equation and reflected Brownian motion in time-dependent domains.: II. Singularities of solutions". In: *J. Funct. Anal.* 204.1 (2003), pp. 1–34.

[35] Krzysztof Burdzy, Weining Kang, and Kavita Ramanan. "The Skorokhod problem in a time-dependent interval". In: *Stochastic Process. Appl.* 119.2 (2009), pp. 428–452.

[36] Krzysztof Burdzy and David Nualart. "Brownian motion reflected on Brownian motion". In: *Probab. Theory Related Fields* 122.4 (2002), pp. 471–493.

[37] Manuel Cabezas et al. "Brownian particles of rank-dependent drifts: out of equilibrium behavior". In: (2017). arXiv:1708.01918.

[38] H. Cai et al. "Real-time bidding by reinforcement learning in display advertising". In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM. 2017, pp. 661–670.

[39] Pierre Cardaliaguet and Saeed Hadikhanloo. "Learning in mean field games: the fictitious play". In: *ESAIM: Control, Optimisation and Calculus of Variations* 23.2 (2017), pp. 569–591.

[40] Pierre Cardaliaguet and Catherine Rainer. "On the (in) efficiency of MFG equilibria". In: *arXiv preprint arXiv:1802.06637* (2018).

[41] Pierre Cardaliaguet et al. "The master equation and the convergence problem in mean field games". In: *ArXiv Preprint: 1509.02505* (2015).

[42] R. Carmona, J. P. Fouque, and L. H. Sun. "Mean field games and systemic risk". In: *Communications in Mathematical Sciences* 13.4 (2015), pp. 911–933.

[43] René Carmona. *Lectures on BSDEs, Stochastic Control, and Stochastic Differential Games with Financial Applications*. SIAM, 2016.

[44] René Carmona. *Lectures on BSDEs, stochastic control, and stochastic differential games with financial applications*. Vol. 1. SIAM, 2016.

[45] René Carmona and François Delarue. "Mean field forward-backward stochastic differential equations". In: *Electronic Communications in Probability* 18 (2013).

[46] René Carmona and François Delarue. "Probabilistic analysis of mean-field games". In: *SIAM Journal on Control and Optimization* 51.4 (2013), pp. 2705–2734.

[47] René Carmona, François Delarue, and Aimé Lachapelle. "Control of McKean–Vlasov dynamics versus mean field games". In: *Mathematics and Financial Economics* 7.2 (2013), pp. 131–166.

[48] Rene Carmona, Jean-Pierre Fouque, and Li-Hsien Sun. "Mean field games and systemic risk". In: *Available at SSRN 2307814* (2013).

[49] René Carmona, Christy V Graves, and Zongjun Tan. "Price of anarchy for Mean Field Games". In: *ESAIM: Proceedings and Surveys* 65 (2019), pp. 349–383.

[50] René Carmona and Daniel Lacker. "A probabilistic weak formulation of mean field games and applications". In: *The Annals of Applied Probability* 25.3 (2015), pp. 1189–1231.

[51] René Carmona and Mathieu Laurière. "Convergence Analysis of Machine Learning Algorithms for the Numerical Solution of Mean Field Control and Games: I–The Ergodic Case". In: *arXiv preprint arXiv:1907.05980* (2019).

[52] Rene Carmona and Xiuneng Zhu. "A probabilistic approach to mean field games with major and minor players". In: *The Annals of Applied Probability* 26.3 (2016), pp. 1535–1580.

[53] René Carmona, François Delarue, et al. "Forward–backward stochastic differential equations and controlled McKean–Vlasov dynamics". In: *The Annals of Probability* 43.5 (2015), pp. 2647–2700.

[54] René Carmona, François Delarue, Daniel Lacker, et al. "Mean field games with common noise". In: *The Annals of Probability* 44.6 (2016), pp. 3740–3803.

[55] Alekos Cecchin et al. "On the convergence problem in mean field games: a two state model without uniqueness". In: *SIAM Journal on Control and Optimization* 57.4 (2019), pp. 2443–2466.

[56] Maria B Chiarolla, Giorgio Ferrari, and Frank Riedel. "Generalized Kuhn–Tucker Conditions for N-Firm Stochastic Irreversible Investment under Limited Resources". In: *SIAM Journal on Control and Optimization* 51.5 (2013), pp. 3863–3885.

[57] Pao–Liu Chow, José–Luis Menaldi, and Maurice Robin. "Additive control of stochastic linear systems with finite horizon". In: *SIAM Journal on Control and Optimization* 23.6 (1985), pp. 858–899.

[58] George Christodoulou and Elias Koutsoupias. "On the price of anarchy and stability of correlated equilibria of linear congestion games". In: *European Symposium on Algorithms*. Springer. 2005, pp. 59–70.

[59] George Christodoulou and Elias Koutsoupias. "The price of anarchy of finite congestion games". In: *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*. ACM. 2005, pp. 67–73.

[60] E. Çinlar. "A local time for a storage process". In: *Annals of Probability* (1975), pp. 930–950.

[61] E. Çinlar and M. Pinsky. "A stochastic integral in storage theory". In: *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 17.3 (1971), pp. 227–240.

[62] E. Çinlar and M. Pinsky. "On dams with additive inputs and a general release rule". In: *Journal of Applied Probability* 9.2 (1972), pp. 422–429.

[63] Marco Cirant. "Multi-population mean field games systems with Neumann boundary conditions". In: *Journal de Mathématiques Pures et Appliquées* 103.5 (2015), pp. 1294–1315.

[64] Samuel N Cohen and Robert James Elliott. *Stochastic Calculus and Applications*. Vol. 2. Springer, 2015.

[65] Jules L Coleman. "Efficiency, utility, and wealth maximization". In: *Hofstra L. Rev.* 8 (1979), p. 509.

[66] J.G. Dai and R.J. Williams. "Existence and uniqueness of semimartingale reflecting Brownian motions in convex polyhedrons". In: *Theory of Probability and its Applications* 40.1 (1996), pp. 1–40.

[67] Mark Davis and Andrew Norman. "Portfolio selection with transaction costs". In: *Mathematics of operations research* 15.4 (1990), pp. 676–713.

[68] T. De Angelis and G. Ferrari. "Stochastic non-zero-sum games: a new connection between singular control and optimal stopping". In: *ArXiv Preprint: 1601.05709* (2016).

[69] François Delarue and Rinel Foguen Tchuendom. "Selection of equilibria in a linear quadratic mean-field game". In: *Stochastic Processes and their Applications* (2019).

[70] Amir Dembo and Li-Cheng Tsai. "Equilibrium fluctuation of the Atlas model". In: *Annals of Probability* 45.6b (2017), pp. 4529–4560.

[71] C Dong et al. "Robust planning of energy management systems with environmental and constraint-conservative considerations under multiple uncertainties". In: *Energy Conversion and Management* 65 (2013), pp. 471–486.

[72] Paul Dupuis and Hitoshi Ishii. "On Lipschitz continuity of the solution mapping to the Skorokhod problem, with applications". In: *Stochastics: An International Journal of Probability and Stochastic Processes* 35.1 (1991), pp. 31–62.

[73] Paul Dupuis and Hitoshi Ishii. "SDEs with oblique reflection on nonsmooth domains". In: *The Annals of Probability* 21.1 (1993), pp. 554–580.

[74] Nicole El Karoui and Ioannis Karatzas. "Probabilistic aspects of finite-fuel, reflected follower problems". In: *Acta Applicandae Mathematica* 11.3 (1988), pp. 223–258.

[75] HJ Engelbert. "On the theorem of T. Yamada and S. Watanabe". In: *Stochastics: An International Journal of Probability and Stochastic Processes* 36.3-4 (1991), pp. 205–216.

[76] L.C. Evans. "A second order elliptic equation with gradient constraint". In: *Communications in Partial Differential Equations* 4.5 (1979), pp. 555–572.

[77] E. Even-Dar and Y. Mansour. "Learning rates for Q-learning". In: *Journal of Machine Learning Research* 5(Dec) (2003), pp. 1–25.

[78] Salvatore Federico and Huyen Pham. "Smooth-fit principle for a degenerate two-dimensional singular stochastic control problem arising in irreversible investment". In: *preprint* (2012).

[79] Robert Fernholz. *Stochastic Portfolio Theory*. Vol. 48. Applications of Mathematics (New York). Stochastic Modelling and Applied Probability. Springer-Verlag, New York, 2002, pp. xiv+177. ISBN: 0-387-95405-8. DOI: 10.1007/978-1-4757-3699-1. URL: http://dx.doi.org/10.1007/978-1-4757-3699-1.

[80] Giorgio Ferrari, Frank Riedel, and Jan-Henrik Steg. "Continuous-time public good contribution under uncertainty: a stochastic control approach". In: *Applied Mathematics & Optimization* 75.3 (2017), pp. 429–470.

[81] Guanxing Fu and Ulrich Horst. "Mean field games with singular controls". In: *SIAM Journal on Control and Optimization* 55.6 (2017), pp. 3833–3868.

[82] B. Gao and L. Pavel. "On the Properties of the Softmax Function with Application in Game Theory and Reinforcement Learning". In: *Arxiv Preprint:1704.00805* (2017).

[83] Xuefeng Gao et al. "Bounded-Velocity Stochastic Control for Dynamic Resource Allocation". In: *arXiv preprint arXiv:1801.01221* (2018).

[84] Leonidas Georgiadis, Michael J Neely, and Leandros Tassiulas. "Resource allocation and cross-layer control in wireless networks". In: *Foundations and Trends® in Networking* 1.1 (2006), pp. 1–144.

[85] A. L. Gibbs and F. E. Su. "On choosing and bounding probability metrics". In: *International Statistical Review* 70(3) (2002), pp. 419–435.

[86] David Gilbarg and Neil S Trudinger. *Elliptic Partial Differential Equations of Second Order*. springer, 2015.

[87] Diogo A Gomes, Joana Mohr, and Rafael Rigao Souza. "Discrete time, finite state space mean field games". In: *Journal de mathématiques pures et appliquées* 93.3 (2010), pp. 308–328.

[88] P Jameson Graber. "Linear quadratic mean field type control and mean field games with common noise, with application to production of an exhaustible resource". In: *Applied Mathematics & Optimization* 74.3 (2016), pp. 459–486.

[89] O. Guéant, J.M. Lasry, and P.L. Lions. "Mean field games and applications". In: *Paris-Princeton Lectures on Mathematical Finance 2010* (2011), pp. 205–266.

[90] Olivier Guéant. "A reference case for mean field games models". In: *Journal de mathématiques pures et appliquées* 92.3 (2009), pp. 276–294.

[91] R. Gummadi, P. Key, and A. Proutiere. "Repeated auctions under budget constraints: Optimal bidding strategies and equilibria". In: *the Eighth Ad Auction Workshop*. 2012.

[92] X. Guo and H. Pham. "Optimal partially reversible investment with entry decision and general production function". In: *Stochastic Processes and their Applications* 115.5 (2005), pp. 705–736.

[93] Xin Guo and Joon Seok Lee. "Stochastic Games and Mean Field Games with Singular Controls". In: *Preprint* (2018).

[94] Xin Guo and Renyuan Xu. "Stochastic Games for Fuel Followers Problem: N vs MFG". In: *arXiv preprint arXiv:1803.02925* (2018).

[95] Xin Guo et al. "Optimal spot market inventory strategies in the presence of cost and price risk". In: *Mathematical Methods of Operations Research* 73.1 (2011), pp. 109–137.

[96] S. Hamadène and R. Mu. "Bang–bang-type Nash equilibrium point for Markovian non-zero-sum stochastic differential game". In: *Comptes Rendus Mathematique* 352.9 (2014), pp. 699–706.

[97] J. Hamari, M. Sjöklint, and A. Ukkonen. "The sharing economy: Why people participate in collaborative consumption". In: *Journal of the Association for Information Science and Technology* 67(9) (2016), pp. 2047–2059.

[98] J.M. Harrison and R.J. Williams. "Multidimensional reflected Brownian motions having exponential stationary distributions". In: *The Annals of Probability* 15.1 (1987), pp. 115–137.

[99] D. Hernandez-Hernandez, R.S. Simon, and M. Zervos. "A zero-sum game between a singular stochastic controller and a discretionary stopper". In: *The Annals of Applied Probability* 25.1 (2015), pp. 46–80.

[100] P. Hernandez-Leal, B. Kartal, and M. E. Taylor. "Is multiagent deep reinforcement learning the answer or the question? A brief survey". In: *Arxiv Preprint:1810.05587* (2018).

[101] J. Hu and M. P. Wellman. "Nash Q-learning for general-sum stochastic games". In: *Journal of Machine Learning Research* 4.Nov (2003), pp. 1039–1069.

[102] Y. Hu, B. Øksendal, and A. Sulem. "Singular mean-field control games with applications to optimal harvesting and investment problems". In: *ArXiv Preprint: 1406.1863* (2014).

[103] M. Huang, P.E. Caines, and R.P. Malhamé. "Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized epsilon-Nash equilibria". In: *IEEE transactions on automatic control* 52.9 (2007), pp. 1560–1571.

[104] M. Huang, R. P. Malhamé, and P. E. Caines. "Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle". In: *Communications in Information and Systems* 6.3 (2006), pp. 221–252.

[105] Minyi Huang, Roland P Malhamé, and Peter E Caines. "Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle". In: *Communications in Information and Systems* 6.3 (2006), pp. 221–252.

[106] Xuancheng Huang, Sebastian Jaimungal, and Mojtaba Nourian. "Mean-field game strategies for optimal execution". In: *Applied Mathematical Finance* 26.2 (2019), pp. 153–185.

[107] R.C. Hynd. *Partial Differential Equations with Gradient Constraints Arising in the Optimal Control of Singular Stochastic Processes*. Ph.D. dissertation, University of California, Berkeley, 2010.

[108] Tomoyuki Ichiba and Ioannis Karatzas. "On collisions of Brownian particles". In: *Annals of Applied Probability* 20.3 (2010), pp. 951–977. ISSN: 1050-5164. DOI: 10.1214/09-AAP641. URL: http://dx.doi.org/10.1214/09-AAP641.

[109] Tomoyuki Ichiba, Ioannis Karatzas, and Mykhaylo Shkolnikov. "Strong solutions of stochastic equations with rank-based coefficients". In: *Probab. Theory Related Fields* 156.1-2 (2013), pp. 229–248. ISSN: 0178-8051. DOI: 10.1007/s00440-012-0426-3. URL: http://dx.doi.org/10.1007/s00440-012-0426-3.

[110] Tomoyuki Ichiba et al. "Hybrid atlas models". In: *Annals of Applied Probability* 21.2 (2011), pp. 609–644. ISSN: 1050-5164. DOI: 10.1214/10-AAP706. URL: http://dx.doi.org/10.1214/10-AAP706.

[111] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic Differential Equations and Diffusion Processes*. Vol. 24. Elsevier, 2014.

[112] K. Iyer, R. Johari, and M. Sundararajan. "Mean field equilibria of dynamic auctions with learning". In: *ACM SIGecom Exchanges* 10.3 (2011), pp. 10–14.

[113] S. D. Jacka. "A finite fuel stochastic control problem". In: *Stochastics* 10.2 (1983), pp. 103–113.

[114] S. H. Jeong, A. R. Kang, and H. K. Kim. "Analysis of Game Bot's Behavioral Characteristics in Social Interaction Networks of MMORPG". In: *ACM SIGCOMM Computer Communication Review* 45(4) (2015), pp. 99–100.

[115] J. Jin et al. "Real-Time Bidding with Multi-Agent Reinforcement Learning in Display Advertising". In: *Arxiv Preprint:1802.09756* (2018).

[116] Weining Kang, Ruth J Williams, et al. "An invariance principle for semimartingale reflecting Brownian motions in domains with piecewise smooth boundaries". In: *The Annals of Applied Probability* 17.2 (2007), pp. 741–779.

[117] S. Kapoor. "Multi-Agent Reinforcement Learning: A Report on Challenges and Approaches". In: *Arxiv Preprint:1807.09427* (2018).

[118] I. Karatzas. "A class of singular stochastic control problems". In: *Advances in Applied Probability* 15.2 (1983), pp. 225–254.

[119] I. Karatzas and Q. Li. "BSDE approach to non-zero-sum stochastic differential games of control and stopping". In: *Stochastic Processes, Finance and Control* 1 (2011), pp. 105–153.

[120] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Vol. 113. Springer Science and Business Media, 2012.

[121] I. Karatzas and S.E. Shreve. "Connections between optimal stopping and singular stochastic control II. Reflected follower problems". In: *SIAM Journal on Control and Optimization* 23.3 (1985), pp. 433–451.

[122] Ioannis Karatzas. "A class of singular stochastic control problems". In: *Adv. in Appl. Probab.* 15.2 (1983), pp. 225–254. ISSN: 0001-8678. DOI: 10.2307/1426435. URL: https://doi.org/10.2307/1426435.

[123] Ioannis Karatzas and Steven Shreve. "Equivalent models for finite-fuel stochastic control". In: *Stochastics* 18.3-4 (1986), pp. 245–276.

[124] Ioannis Karatzas and Steven E Shreve. "Brownian motion". In: *Brownian Motion and Stochastic Calculus*. Springer, 1998, pp. 47–127.

[125] A. C Kizilkale and P. E Caines. "Mean field stochastic adaptive control". In: *IEEE Transactions on Automatic Control* 58.4 (2013), pp. 905–920.

[126] E. V. Krichagina and M. I. Taksar. "Diffusion approximation for GI/G/1 controlled queues". In: *Queueing systems* 12.3-4 (1992), pp. 333–367.

[127] L. Kruk. "Optimal policies for N-dimensional singular stochastic control problems part I: The Skorokhod problem". In: *SIAM Journal on Control and Optimization* 38.5 (2000), pp. 1603–1622.

[128] Lukasz Kruk. "Optimal policies for n-dimensional singular stochastic control problems part I: The Skorokhod problem". In: *SIAM Journal on Control and Optimization* 38.5 (2000), pp. 1603–1622.

[129] H.D. Kwon and H. Zhang. "Game of singular stochastic control and strategic exit". In: *Mathematics of Operations Research* 40.4 (2015), pp. 869–887.

[130] D. Lacker and T. Zariphopoulou. "Mean field and N-agent games for optimal investment under relative performance criteria". In: *arXiv:1703.07685* (2017).

[131] Daniel Lacker. "A general characterization of the mean field limit for stochastic differential games". In: *Probability Theory and Related Fields* 165.3-4 (2016), pp. 581–648.

[132] Daniel Lacker. "Mean field games via controlled martingale problems: existence of Markovian equilibria". In: *Stochastic Processes and their Applications* 125.7 (2015), pp. 2856–2894.

[133] Daniel Lacker and Kavita Ramanan. "Rare nash equilibria and the price of anarchy in large static games". In: *Mathematics of Operations Research* 44.2 (2018), pp. 400–422.

[134] Daniel Lacker and Thaleia Zariphopoulou. "Mean field and n-agent games for optimal investment under relative performance criteria". In: *Mathematical Finance* (2017).

[135] Tian Lan et al. *An axiomatic theory of fairness in network resource allocation*. IEEE, 2010.

[136] Jean-Michel Lasry and Pierre-Louis Lions. "Jeux à champ moyen II–horizon fini et contrôle optimal". In: *Comptes Rendus Mathématique* 343.10 (2006), pp. 679–684.

[137] Jean-Michel Lasry and Pierre-Louis Lions. "Jeux à champ moyen I–le cas stationnaire". In: *Comptes Rendus Mathématique* 343.9 (2006), pp. 619–625.

[138] Jean-Michel Lasry and Pierre-Louis Lions. "Mean field games". In: *Japanese journal of mathematics* 2.1 (2007), pp. 229–260.

[139] Jean-Michel Lasry and Pierre-Louis Lions. "Mean field games". In: *Japanese Journal of Mathematics* 2.1 (2007), pp. 229–260.

[140] C-A. Lehalle and C. Mouzouni. "A Mean Field Game of Portfolio Trading and Its Consequences On Perceived Correlations". In: *ArXiv Preprint:1902.09606* (2019).

[141] Ron Levy et al. "Performance management for cluster based web services". In: *Integrated Network Management VIII*. Springer, 2003, pp. 247–261.

[142] P-L Lions. "A remark on Bony maximum principle". In: *Proceedings of the American Mathematical Society* 88.3 (1983), pp. 503–508.

[143] Arne Løkka and Mihail Zervos. "Optimal dividend and issuance of equity policies in the presence of proportional costs". In: *Insurance: Mathematics and Economics* 42.3 (2008), pp. 954–961.

[144] Niklas Lundström and Thomas Önskog. "Stochastic and partial differential equations on non-smooth time-dependent domains". In: (2015). arXiv:1503.05433.

[145] P. Mannucci. "Non-zero-sum stochastic differential games with discontinuous feedback". In: *SIAM Journal on Control and Optimization* 43.4 (2004), pp. 1222–1233.

[146] José-Luis Menaldi and Maurice Robin. "On some cheap control problems for diffusion processes". In: *Transactions of the American Mathematical Society* 278.2 (1983), pp. 771–802.

[147] José-Luis Menaldi and Michael I Taksar. "Optimal correction problem of a multidimensional stochastic system". In: *Automatica* 25.2 (1989), pp. 223–232.

[148] P. A. Meyer. "Martingales locales changement de variables, formules exponentielles". In: *Séminaire de Probabilités X*. Springer, 1976, pp. 291–331.

[149] Enzo Miller and Huyen Pham. "Linear-Quadratic McKean-Vlasov Stochastic Differential Games". In: *Modeling, Stochastic Control, Optimization, and Applications*. Springer, 2019, pp. 451–481.

[150] V. M. Minh et al. "Asynchronous Methods for Deep Reinforcement Learning". In: *International Conference on Machine Learning*. 2016.

[151] Son Luu Nguyen and Minyi Huang. "Mean field LQG games with mass behavior responsive to a major player". In: *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE. 2012, pp. 5792–5797.

[152] Marcel Nutz, Jaime San Martin, and Xiaowei Tan. "Convergence to the mean field game limit: a case study". In: *arXiv preprint arXiv:1806.00817* (2018).

[153] Marcel Nutz and Yuchong Zhang. "A Mean Field Competition". In: *ArXiv Preprint: 1708.01308* (2017).

[154] Kaj Nyström and Thomas Önskog. "The Skorohod oblique reflection problem in time-dependent domains". In: *Annals of Probability* 38.6 (2010), pp. 2170–2223.

[155] Marcelo Olivares and Gérard P Cachon. "Competing retailers and inventory: An empirical investigation of General Motors' dealerships in isolated US markets". In: *Management science* 55.9 (2009), pp. 1586–1604.

[156] Fernando Ortega et al. "Improving collaborative filtering-based recommender systems results using Pareto dominance". In: *Information Sciences* 239 (2013), pp. 50–61.

[157] Soumik Pal and Jim Pitman. "One-dimensional Brownian particle systems with rank-dependent drifts". In: *Annals of Applied Probability* 18.6 (2008), pp. 2179–2207. ISSN: 1050-5164. DOI: 10.1214/08-AAP516. URL: http://dx.doi.org/10.1214/08-AAP516.

[158] J. Pérolat, B. Piot, and O. Pietquin. "Actor-Critic Fictitious Play in Simultaneous Move Multistage Games". In: *International Conference on Artificial Intelligence and Statistics*. 2018.

[159] J. Pérolat et al. "Learning Nash Equilibrium for General-Sum Markov Games from Batch Data". In: *Arxiv Preprint:1606.08718* (2016).

[160] G. Peyré and M. Cuturi. "Computational optimal transport". In: *Foundations and Trends in Machine Learning* 11(5-6) (2019), pp. 355–607.

[161] B. Recht. "A tour of reinforcement learning: The view from continuous control". In: *Annual Review of Control, Robotics, and Autonomous Systems* (2018).

[162] Marco Tulio Ribeiro et al. "Pareto-efficient hybridization for multi-objective recommender systems". In: *Proceedings of the sixth ACM conference on Recommender systems*. ACM. 2012, pp. 19–26.

[163] Tim Roughgarden. "Intrinsic robustness of the price of anarchy". In: *Proceedings of the forty-first annual ACM symposium on Theory of computing*. ACM. 2009, pp. 513–522.

[164] Tim Roughgarden and Éva Tardos. "How bad is selfish routing?" In: *Journal of the ACM (JACM)* 49.2 (2002), pp. 236–259.

[165] Pedram Samadi et al. "Advanced demand side management for the future smart grid using mechanism design". In: *IEEE Transactions on Smart Grid* 3.3 (2012), pp. 1170–1180.

[166] Andrey Sarantsev. "Triple and simultaneous collisions of competing Brownian particles". In: *Electron. J. Probab.* 20 (2015), no. 29, 28. ISSN: 1083-6489. DOI: 10.1214/EJP.v20-3279. URL: http://dx.doi.org/10.1214/EJP.v20-3279.

[167] Mykhaylo Shkolnikov. "Competing particle systems evolving by interacting Lévy processes". In: *Annals of Applied Probability* 21.5 (2011), pp. 1911–1932. ISSN: 1050-5164. DOI: 10.1214/10-AAP743. URL: http://dx.doi.org/10.1214/10-AAP743.

[168] S.E. Shreve and H.M. Soner. "A free boundary problem related to singular stochastic control". In: *Applied Stochastic Analysis (London, 1989)* 16.2 and 3 (1991), pp. 265–301.

[169] H Mete Soner and Shreve E Shreve. "Regularity of the value function for a two-dimensional singular stochastic control problem". In: *SIAM Journal on Control and Optimization* 27.4 (1989), pp. 876–907.

[170] Florin Soucaliuc, Bálint Tóth, and Wendelin Werner. "Reflection and coalescence between independent one-dimensional Brownian paths". In: 36.4 (2000), pp. 509–545.

[171] Jan-Henrik Steg. "Irreversible investment in oligopoly". In: *Finance and Stochastics* 16.2 (2012), pp. 207–224.

[172] Joseph E Stiglitz. "Pareto efficient and optimal taxation and the new new welfare economics". In: *Handbook of public economics*. Vol. 2. Elsevier, 1987, pp. 991–1042.

[173] J. Subramanian and A. Mahajan. "Reinforcement Learning in Stationary Mean-field Games". In: *International Conference on Autonomous Agents and Multiagent Systems* (2019).

[174] Jingrui Sun, Xun Li, and Jiongmin Yong. "Open-loop and closed-loop solvabilities for stochastic linear quadratic optimal control problems". In: *SIAM Journal on Control and Optimization* 54.5 (2016), pp. 2274–2308.

[175] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[176] M. Tan. "Multi-agent reinforcement learning: independent vs. cooperative agents". In: *International Conference on Machine Learning*. 1993, pp. 330–337.

[177] Wenpin Tang and Li-Cheng Tsai. "Optimal surviving strategy for drifted Brownian motions with absorption". In: *Annals of Probability* 46.3 (2018), pp. 1597–1650.

[178] Jeffrey E Teich et al. "Identifying Pareto-optimal settlements for two-party resource allocation negotiations". In: *European Journal of Operational Research* 93.3 (1996), pp. 536–549.

[179] S.R. Varadhan and R.J. Williams. "Brownian motion in a wedge with oblique reflection". In: *Communications on Pure and Applied Mathematics* 38.4 (1985), pp. 405–443.

[180] C. Villani. *Optimal transport: old and new*. Vol. 338. Springer Science & Business Media, 2008.

[181] John Von Neumann, Oskar Morgenstern, and Harold William Kuhn. *Theory of games and economic behavior (commemorative edition)*. Princeton university press, 2007.

[182] H. T. Wai et al. "Multi-Agent Reinforcement Learning via Double Averaging Primal-Dual Optimization". In: *Advances in Neural Information Processing Systems*. 2018, pp. 9672–9683.

[183] Jon Warren. "Dyson's Brownian motions, intertwining and interlacing". In: *Electron. J. Probab.* 12 (2007), pp. 573–590.

[184] R. J. Williams. "Reflected Brownian motion with skew symmetric data in a polyhedral domain". In: *Probab. Theory Related Fields* 75.4 (1987), pp. 459–485. ISSN: 0178-8051. DOI: 10.1007/BF00320328. URL: http://dx.doi.org/10.1007/BF00320328.

[185] R.J. Williams. "Reflected Brownian motion with skew symmetric data in a polyhedral domain". In: *Probability Theory and Related Fields* 75.4 (1987), pp. 459–485.

[186] Zhen Xiao, Weijia Song, and Qi Chen. "Dynamic resource allocation using virtual machines for cloud computing environment." In: *IEEE Trans. Parallel Distrib. Syst.* 24.6 (2013), pp. 1107–1117.

[187] J. Yang et al. "Deep mean field games for learning optimal behavior policy of large populations". In: *Arxiv Preprint:1711.03156* (2017).

[188] Y. Yang et al. "Mean Field Multi-Agent Reinforcement Learning". In: *Arxiv Preprint:1802.05438* (2018).

[189] H. Yin et al. "Learning in mean-field games". In: *IEEE Transactions on Automatic Control* 59.3 (2014), pp. 629–644.

[190] L. Zhang. "The Relaxed Stochastic Maximum Principle in the Mean-field Singular Controls". In: *ArXiv Preprint: 1202.4129* (2012).

[191] Quanyan Zhu and Lacra Pavel. "State-space approach to pricing design in osnr nash game". In: *IFAC Proceedings Volumes* 41.2 (2008), pp. 12001–12006.

Appendix A

Chapter 2

A.1 The Skorokhod Problem (SP)

First, some notation for a general polyhedron G.

Take a fixed integer l ($l \geq 1$), let $\boldsymbol{J} = \{1, 2, \cdots, l\}$. Given an $l-$dimensional vector $\boldsymbol{b} = (b_1, \cdots, b_l)$ and N-dimensional unit vectors $\{\boldsymbol{n}_j, j \in \boldsymbol{J}\}$, a polyhedron G is defined by

$$G = \{\boldsymbol{x} \in \mathbb{R}^N \mid \boldsymbol{n}_j \cdot \boldsymbol{x} > b_j \text{ for any } j \in \boldsymbol{J}\}.$$

Assume the faces $F_j = \{\boldsymbol{x} \in \bar{G} \mid \boldsymbol{n}_j \cdot \boldsymbol{x} = b_j\}$ ($j \in \boldsymbol{J}$) are of dimension $N-1$.

Next, take another set of $N-$dimensional vectors $\{\boldsymbol{d}_j, j \in \boldsymbol{J}\}$, we can define the SP problem on a polyhedron with oblique reflections $(\boldsymbol{d_1}, \cdots, \boldsymbol{d_l})$, in both the strong sense and the weak sense.

Definition 46 (Strong solution to SP)**.** *Given a polyhedron G, a vector field $(\boldsymbol{d_1}, \cdots, \boldsymbol{d_l})$, and $\boldsymbol{x} \in \overline{G}$. Given an N-dimensional Brownian motion $\{\boldsymbol{B}_t\}_{t\geq 0}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P}_{\boldsymbol{x}})$, a strong solution to the SP with the data $(\boldsymbol{x}, G, (\boldsymbol{d_1}, \cdots, \boldsymbol{d_l}), \{B_t\}_{t\geq 0})$ is an $\mathcal{F}_t^B$-adapted process $\boldsymbol{X}_t$ such that*

(a) $\boldsymbol{X}_t = \boldsymbol{x} + \boldsymbol{B}_t + \boldsymbol{\eta}_t D$, *with* $D = \begin{bmatrix} \boldsymbol{d}^1 \\ \cdots \\ \boldsymbol{d}^l \end{bmatrix} \in \mathbb{R}^{l\times N}$,

(b) $\boldsymbol{X}_t$ *has a continuous path in* $\overline{G}$,

(c) $\boldsymbol{X}_t \in \overline{G}$, *for any* $t \geq 0$ *a.s.,*

(d) $\eta_0^j = 0$, η_t^j *is continuous and nondecreasing,* η_t^j *increases only when* $\boldsymbol{X}_t$ *is on the face* F_j. *That is,*

$$\eta_t^j = \int_0^t \boldsymbol{1}_{\{\boldsymbol{X}_s \in \partial F_j\}} d\eta_s^j,$$

(e) the reflection direction $\gamma(\boldsymbol{x}) := \boldsymbol{d}_j$, *if* $\boldsymbol{x} \in F_j$ *for* $j \in \boldsymbol{J}$.

Definition 47 (Weak solution to SP)**.** *Given a polyhedron G, a vector field $(\boldsymbol{d_1}, \cdots, \boldsymbol{d_l})$, and $\boldsymbol{x} \in \bar{G}$. A weak solution to the SP with the data $(\boldsymbol{x}, G, (\boldsymbol{d_1}, \cdots, \boldsymbol{d_l}))$ is an adapted N-dimensional process $\boldsymbol{X}_t$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P}_{\boldsymbol{x}})$ such that*

(a) $\boldsymbol{X}_t = \boldsymbol{W}_t + \boldsymbol{\eta}_t D$, with $D = \begin{bmatrix} \boldsymbol{d}^1 \\ \cdots \\ \boldsymbol{d}^l \end{bmatrix} \in \mathbb{R}^{l \times N}$ and $\boldsymbol{W}$ an N-dimensional Brownian motion under $\mathbb{P}_{\boldsymbol{x}}$, with $\boldsymbol{X}_0 = \boldsymbol{x}$, $\mathbb{P}_{\boldsymbol{x}}$-a.s.,

(b) $\boldsymbol{X}_t$ has a continuous path in $\overline{G}$, $\mathbb{P}_{\boldsymbol{x}}$-a.s.,

(c) $\eta_0^j = 0$, η_t^j is continuous and nondecreasing, $\eta^j(t)$ can increase only when $\boldsymbol{X}_t$ is on the face F_j. That is,

$$\eta_t^j = \int_0^t \boldsymbol{1}_{\{\boldsymbol{X}_s \in \partial F_j\}} d\eta_s^j,$$

(d) the reflection direction $\gamma(\boldsymbol{x}) := \boldsymbol{d}_j$, if $\boldsymbol{x} \in F_j$ for $j \in \boldsymbol{J}$.

Now the proof of Theorem 8 follows from the following two lemmas.

Lemma 48 (Existence of the weak solution to SP)**.** *Fix $\boldsymbol{x} \in \overline{\mathcal{CW}}$. There exists a weak solution to the SP with the data $(\mathcal{CW}, (\boldsymbol{d}_1, \cdots, \boldsymbol{d}_{2N}), \boldsymbol{x})$ with $\boldsymbol{d}_j$ $(j = 1, 2, \cdots, 2N)$ defined in (4.2.6) and $\mathcal{CW}$ defined in* (4.2.2)*. It is a semimartingale reflected Brownian motion (SRBM) starting from $\boldsymbol{x}$.*

In fact, this weak solution is unique in a weak sense, see [66].

Proof of Lemma 48. Following the notation in [66], define the *maximal set* to characterize the points on $\partial\mathcal{CW}$ as follows. Take $\mathbf{J} = \{1, 2, \cdots, 2N\}$ the index set of the $2N$ faces of $\mathcal{CW}$. For each $\varnothing \neq \boldsymbol{K} \subset \mathbf{J}$, define $F_{\mathbf{K}} = \cap_{j \in \mathbf{K}} F_j$. Let $F_{\varnothing} = \mathcal{CW}$. A set $\boldsymbol{K} \subset \boldsymbol{J}$ is *maximal* if $\mathbf{K} \neq \varnothing$, $F_{\mathbf{K}} \neq \varnothing$, and $F_{\mathbf{K}} \neq F_{\tilde{\mathbf{K}}}$ for any $\tilde{\mathbf{K}} \supset \mathbf{K}$ such that $\tilde{\mathbf{K}} \neq \mathbf{K}$. Now, it suffices to show that for each *maximal* $\boldsymbol{K} \subset \boldsymbol{J}$,

(S.a) there is a positive linear combination $\boldsymbol{d} = \sum_{j \in \boldsymbol{K}} a_j \boldsymbol{d}_j$ $(a_j > 0,\ \forall j \in \boldsymbol{K})$ of the $\{\boldsymbol{d}_j,\ j \in \boldsymbol{K}\}$ such that $\boldsymbol{n}_j \cdot \boldsymbol{d} > 0$ for any $j \in \boldsymbol{K}$;

(S.b) there is a positive linear combination $\boldsymbol{n} = \sum_{j \in \boldsymbol{K}} c_j \boldsymbol{n}_j$ $(c_j > 0,\ \forall j \in \boldsymbol{K})$ of the $\{\boldsymbol{n}_j,\ j \in \boldsymbol{K}\}$ such that $\boldsymbol{d}_j \cdot \boldsymbol{n} > 0$ for any $j \in \boldsymbol{K}$.

Let us first show that for any maximal $\boldsymbol{K}$, $|\boldsymbol{K}| \leq N - 1$. To see this claim, denote

$$N_{\text{mat}} := \begin{bmatrix} \boldsymbol{n}_1 \\ \boldsymbol{n}_2 \\ \vdots \\ \boldsymbol{n}_{2N} \end{bmatrix} = \frac{\sqrt{N-1}}{\sqrt{N}} \begin{bmatrix} 1 & -\frac{1}{N-1} & -\frac{1}{N-1} & \cdots & -\frac{1}{N-1} \\ -\frac{1}{N-1} & 1 & -\frac{1}{N-1} & \cdots & -\frac{1}{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{N-1} & -\frac{1}{N-1} & -\frac{1}{N-1} & \cdots & 1 \\ -1 & \frac{1}{N-1} & \frac{1}{N-1} & \cdots & \frac{1}{N-1} \\ \frac{1}{N-1} & -1 & \frac{1}{N-1} & \cdots & \frac{1}{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{N-1} & \frac{1}{N-1} & \frac{1}{N-1} & \cdots & -1 \end{bmatrix} \in \mathbb{R}^{2N \times N}.$$

It follows from some calculations that $\det(N_{\text{mat}}) = N-1$, implying that for any $\boldsymbol{K} \subset \boldsymbol{J}$ with $|\boldsymbol{K}| = N$, $\cap F_{j\in\boldsymbol{K}} = \varnothing$. Moreover, for any maximal $\mathbf{K}$, $|\boldsymbol{K}| \leq N-1$. Now checking the conditions **(S.a)** and **(S.b)** for any maximal $\boldsymbol{K}$ reduces to checking these conditions for the maximal $\boldsymbol{K}$ with $|\boldsymbol{K}| = N-1$.

Note that for any $i = 1, \cdots, N$, F_i and F_{N+i} are parallel faces such that $F_i \cap F_{N+i} = \varnothing$, there is no maximal $\boldsymbol{K}$ for which both $i \in \boldsymbol{K}$ and $N+i \in \boldsymbol{K}$. Thus, take any $\boldsymbol{K} = \{i_1, \cdots, i_{N-1}\}$, where $i_k \in \{k, N+k\}$ for $k = 1, 2, \cdots, N-1$. Denote m as the number of indexes in $\boldsymbol{K}$ which is strictly smaller than N, then $N-1-m$ is the number of indexes in $\boldsymbol{K}$ that are greater than N.

To check **(S.a)**, define $\boldsymbol{n} = \sum_{k=1}^{N-1} \boldsymbol{n}_{i_k}$, then for any $k \in \{1, 2, \cdots, N-1\}$,

$$\boldsymbol{n} \cdot \boldsymbol{d}_{i_k} = \frac{\sqrt{N-1}}{\sqrt{N}} \left[1 + \frac{1}{N-1}\left[\mathbf{1}_{(i_k \leq N)}(N-2m) + \mathbf{1}_{(i_k > N)}(-N+2m+2)\right]\right] > 0.$$

To check **(S.b)**, define $d = \sum_{k=1}^{N-1} d_{i_k}$, then for any $k \in \{1, 2, \cdots, N-1\}$,

$$\boldsymbol{d} \cdot \boldsymbol{n}_{i_k} = \frac{\sqrt{N-1}}{\sqrt{N}} \left[1 + \frac{1}{N-1}\left[\mathbf{1}_{(i_k \leq N)}(N-2m) + \mathbf{1}_{(i_k > N)}(-N+2m+2)\right]\right] > 0.$$

□

Next, the uniqueness of solution in the strong sense is established by the localization technique. That is, construct a sequence of bounded region $\mathcal{W}_k$ ($k \in \mathbb{N}^+$) such that

$$\mathcal{W}_1 \subset \mathcal{W}_2 \subset \cdots \subset \mathcal{CW},$$

where $\mathcal{W}_k$ satisfies the condition in [73]. Then define a sequence of stopping times associated with $\mathcal{W}_k$ ($k \in \mathbb{N}^+$) and extend the strong uniqueness result on bounded regions in [73].

Lemma 49 (Uniqueness of the strong solution to SP)**.** *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, suppose there are two strong solutions $\boldsymbol{X}_t^*$ and $\boldsymbol{X}_t^{*\prime}$ to the SP with the data $(\boldsymbol{x}, \mathcal{CW}, (\boldsymbol{d}_1, \cdots, \boldsymbol{d}_{2N}), \{\boldsymbol{B}_t\}_{t\geq 0})$ with $\boldsymbol{d}_i$ ($i = 1 \cdots, 2N$) defined in (4.2.6). Then*

$$\mathbb{P}_{\boldsymbol{x}}\left(\boldsymbol{X}_t^* = \boldsymbol{X}_t^{*\prime}; \quad 0 \leq t < \infty\right) = 1.$$

Proof of Lemma 49. First, the uniqueness on a bounded region. To this end, define the bounded region $\mathcal{W}_k = \mathcal{CW} \cap \left\{\boldsymbol{x} \;\middle|\; \left|\sum_{i=1}^N x_i\right| < k\right\}$ for $k \in \mathbb{N}_+$. Clearly, $\mathcal{W}_k \subseteq \mathcal{W}_{k+1} \subseteq \mathcal{CW}$ and $\mathcal{CW} = \cup_k \mathcal{W}_k$. Define the boundaries of $\mathcal{W}_k$ as

$$\partial\mathcal{W}_k = \cup_{i=1}^{2N} F_i^{(k)} \cup F_{2N+1}^{(k)} \cup F_{2N+2}^{(k)},$$

where $F_i^{(k)} = F_i \cap \overline{\mathcal{W}}_k$ for $i = 1, \ldots, 2N$, $F_{2N+1}^{(k)} = \overline{\mathcal{W}}_k \cap \{\boldsymbol{x} \mid \sum_{i=1}^N x_i = k\}$, and $F_{2N+2}^{(k)} = \overline{\mathcal{W}}_k \cap \{\boldsymbol{x} \mid \sum_{i=1}^N x_i = -k\}$. Define the reflection direction $\gamma^{(l)}(\cdot)$ on $\partial\mathcal{W}_k$ as

$$\gamma^{(k)}(\boldsymbol{x}) = \begin{cases} \boldsymbol{d}_{2N+1} = \boldsymbol{n}_{2N+1} = \frac{1}{\sqrt{N}}(-1, -1, \ldots, -1), \ \boldsymbol{x} \in F_{2N+1}^{(k)}, \\ \boldsymbol{d}_{2N+2} = \boldsymbol{n}_{2N+2} = \frac{1}{\sqrt{N}}(1, 1, \ldots, 1), \ \boldsymbol{x} \in F_{2N+2}^{(k)}, \\ \boldsymbol{d}_i, \ \boldsymbol{x} \in F_i^{(k)}, \ \text{for } i = 1, 2, \cdots, 2N. \end{cases} \tag{A.1.1}$$

For $\boldsymbol{x} \in \partial\mathcal{W}_k$, define $I_k(\boldsymbol{x}) := \left\{ i : \boldsymbol{x} \in F_i^{(k)} \right\}$ as the *index set* of $\boldsymbol{x}$. Following [73], we will show that, for each $\boldsymbol{x} \in \partial\mathcal{W}_k$, there exists $b_i \geq 0$, $i \in I_k(\boldsymbol{x})$, such that

$$b_i \boldsymbol{d}_i \cdot \boldsymbol{n}_i > \sum_{j \in I_k(\boldsymbol{x}) \setminus \{i\}} b_j |\boldsymbol{d}_j \cdot \boldsymbol{n}_i|. \qquad \textbf{(S.c)}$$

Define $b_i = 1$ for any $i = 1, 2, \cdots, 2N+2$. It is sufficient to verify **(S.c)** for $\boldsymbol{x} \in \partial\mathcal{W}_k$ such that $|I_k(\boldsymbol{x})| = N$. In this case, either $2N+1 \in I_k(\boldsymbol{x})$ or $2N+2 \in I_k(\boldsymbol{x})$. Take any $i_0 \in I_k(\boldsymbol{x}) \setminus \{2N+1, 2N+2\}$,

$$\begin{aligned} |\boldsymbol{n}_{i_0} \cdot \boldsymbol{d}_{i_0}| &= \frac{\sqrt{N-1}}{\sqrt{N}}, \\ |\boldsymbol{n}_{i_0} \cdot \boldsymbol{d}_j| &= \frac{\sqrt{N-1}}{\sqrt{N}} \frac{1}{N-1}, \quad \text{for } j \in I_k(\boldsymbol{x}) \setminus \{2N+1, 2N+2, i_0\}, \\ |\boldsymbol{n}_{i_0} \cdot \boldsymbol{d}_j| &= 0, \quad \text{for } j \in \{2N+1, 2N+2\}. \end{aligned}$$

Hence (**S.c**) holds with $\boldsymbol{d}_{i_0} \cdot \boldsymbol{n}_{i_0} = \frac{\sqrt{N-1}}{\sqrt{N}}$ and $\sum_{j \in I_k(\boldsymbol{x}) \setminus \{i_0\}} |\boldsymbol{d}_j \cdot \boldsymbol{n}_{i_0}| = \frac{\sqrt{N-1}}{\sqrt{N}} \frac{N-2}{N-1}$. By [73], there exists a unique strong solution $(\boldsymbol{X}_t^k, \boldsymbol{\eta}^k)_{t \geq 0}$ to the SP with the data $(\boldsymbol{x}, \mathcal{W}_k, (\boldsymbol{d}_1, \cdots, \boldsymbol{d}_{2N+2}), \{\boldsymbol{B}_t\}_{t \geq 0})$ such that $\boldsymbol{x} \in \overline{\mathcal{W}}_k$.

Now, let $(\boldsymbol{X}_t^{k\prime}, \boldsymbol{\eta}^{k\prime})$ be the strong solution to the SP with the data $(\boldsymbol{x}', \mathcal{W}_k, (\boldsymbol{d}_1, \cdots, \boldsymbol{d}_{2N+2}), \{\boldsymbol{B}_t'\}_{t \geq 0})$. Then by [73], there exists a constant $C_k < \infty$ such that for any $0 \leq t \leq T$,

$$\mathbb{E}\left(\sup_{0 \leq s \leq t} \|\boldsymbol{X}_s^k - \boldsymbol{X}_s^{k\prime}|^2 \right) \leq C_k \left\{ \|\boldsymbol{x} - \boldsymbol{x}'\|^2 + \int_0^t \mathbb{E}\left(\sup_{0 \leq u \leq s} \|\boldsymbol{B}_s - \boldsymbol{B}_s'\|^2 \right) ds \right\}. \tag{A.1.2}$$

To finish the proof, now suppose that there are two strong solutions $(\boldsymbol{X}_t^*, \boldsymbol{\eta}^*)_{t \geq 0}$ and $(\boldsymbol{X}_t^{*\prime}, \boldsymbol{\eta}^{*\prime})_{t \geq 0}$ to the SP with the data $(\boldsymbol{x}, \mathcal{CW}, (\boldsymbol{d}_1, \cdots, \boldsymbol{d}_{2N}), \{\boldsymbol{B}_t\}_{t \geq 0})$, with $\boldsymbol{d}_i$ $(i = 1, 2, \cdots, 2N)$ defined in (4.2.6) and $\boldsymbol{X}_0^* = \boldsymbol{X}_0^{*\prime} = \boldsymbol{x} \in \overline{\mathcal{CW}}$. Suppose there exists $M := M(\boldsymbol{x})$ such that $\boldsymbol{x} \in \overline{\mathcal{W}}_k$ for $k \geq M$. Define $\tau_k = \inf\{t : \boldsymbol{X}_t^* \in F_{2N+1}^{(k)} \cup F_{2N+2}^{(k)}\}$ and $\tau_k' = \inf\{t : \boldsymbol{X}_t^{*\prime} \in F_{2N+1}^{(k)} \cup F_{2N+2}^{(k)}\}$. Then the uniqueness of the strong solution to SP with the data $(\boldsymbol{x}, \mathcal{W}_k, \gamma^{(k)}, \{\boldsymbol{B}_t\}_{t \geq 0})$ implies that for $k \geq M$,

$$\begin{aligned} &\mathbb{P}_{\boldsymbol{x}}\left(\boldsymbol{X}_t^* = \boldsymbol{X}_t^{*\prime},\ t \leq \tau_k \right) = 1, \\ &\mathbb{P}_{\boldsymbol{x}}\left(\tau_k = \tau_k' \right) = 1. \end{aligned} \tag{A.1.3}$$

By the continuity of the probability measure,

$$\mathbb{P}_{\boldsymbol{x}}\left(\boldsymbol{X}_t^* = \boldsymbol{X}_t^{*\prime},\ t \leq \tau_k, k \to \infty \right) = \lim_{k \to \infty} \mathbb{P}_{\boldsymbol{x}}\left(\boldsymbol{X}_t^* = \boldsymbol{X}_t^{*\prime},\ t \leq \tau_k \right) = 1. \tag{A.1.4}$$

Now it remains to show $\lim_{k \to \infty} \tau_k = \infty$ a.s.. Suppose otherwise, then there exists $\tau^* = \tau^*(\omega) < \infty$ such that $\lim_{k \to \infty} \tau_k = \tau^*$ pathwise. Therefore,

$$\mathbb{P}_{\boldsymbol{x}}\left(\left| \sum_{i=1}^N X_{\tau^*}^{i*} \right| = \infty \right) = \mathbb{P}_{\boldsymbol{x}}\left(\left| \sum_{i=1}^N x^i + B_{\tau^*}^i + \eta_{\tau^*}^{i*} - \eta_{\tau^*}^{N+i*} \right| = \infty \right) = 1,$$

which implies, from the bounded variation property of $\{\boldsymbol{\eta}^*\}_{t \geq 0}$, $\mathbb{P}_{\boldsymbol{x}}\left(\left| \sum_{i=1}^N B_{\tau^*}^i \right| = \infty \right) = 1$. This contradicts with the property of Brownian motion, thus $\lim_{k \to \infty} \tau_k = \infty$ a.s.. $\square$

A.2 Well-posedness of Algorithm 1

If $\boldsymbol{x} = (x^1, \cdots, x^N) \notin \overline{\mathcal{CW}}$, then there exists an i such that $\boldsymbol{x} \in \mathcal{A}_i$. For any $k > 1$, denote the point after the k-th jump as $\boldsymbol{x}_k = (x_k^1, \ldots, x_k^N)$. In step $k+1$, if $\boldsymbol{x}_k \in \mathcal{A}_i$, player i will apply a minimal push to reach the boundary $\partial E_i^- \cup \partial E_i^+$.

If the jumps do not stop in finite steps, an argument by contradiction will show that they converge to $\hat{\boldsymbol{x}} \in \partial\mathcal{CW}$. Let us first show that $\{\boldsymbol{x}_k\}_{k\geq 1}$ converges. At each step $k \geq 1$, denote $x_k^{(1)} \leq \cdots \leq x_k^{(N)}$ as the ordered points of $x_k^1, \ldots, x_k^N$. At each step k, only the player with position $x_k^{(1)}$ or $x_k^{(N)}$ will jump. Therefore $\{x_k^{(1)}\}_{k\geq 1}$ is a non-decreasing sequence with an upper bound $\max_{i\leq N} x^i$. Hence the limit exists, denoted as $x_*^{(1)}$. Similarly, the bounded non-increasing sequence $x_k^{(N)}$ has a limit, denoted as $x_*^{(N)}$. Then by the sandwich argument, $\{\boldsymbol{x}_k\}_{k\geq 1}$ converges.

Next, denote the distance $d_k^i = |x_k^i - \frac{\sum_{i\neq j} x_k^j}{N-1}|$. By definition of $\mathcal{A}_i$, the player with the biggest d_k^i will jump in step $k+1$. Suppose $\hat{\boldsymbol{x}} = \lim_{k\to\infty} \boldsymbol{x}_k \notin \partial\mathcal{CW}$, then there exists an $m \in \{1, \ldots, N\}$ such that $\hat{\boldsymbol{x}} \in \mathcal{A}_m$. Denote the distance $\Delta = |\hat{x}^m - \frac{\sum_{j\neq m} \hat{x}^j}{N-1}| - c_N = \max_{i=1,2,\cdots,N}\{|\hat{x}^i - \frac{\sum_{j\neq i} \hat{x}^j}{N-1}|\} - c_N > 0$. Given $\epsilon > 0$ so that $\epsilon < \frac{\Delta}{8N}$ and $\overline{\mathcal{CW}} \cap B_\epsilon(\hat{\boldsymbol{x}}) = \emptyset$, there exists a sufficiently large $K > 0$ such that for any $k' > K$, $\boldsymbol{x}_{k'} \in B_\epsilon(\hat{\boldsymbol{x}})$. That is, $\sum_{i=1}^N |x_{k'}^i - \hat{x}^i|^2 \leq \epsilon^2$. By the triangle inequality,

$$\begin{aligned}
\left|x_{k'}^m - \frac{\sum_{j\neq m} x_{k'}^j}{N-1}\right| - c_N &\geq \left|\hat{x}^m - \frac{\sum_{j\neq m} \hat{x}^j}{N-1}\right| - c_N - \left|\left(x_{k'}^m - \frac{\sum_{j\neq m} x_{k'}^j}{N-1}\right) - \left(\hat{x}^m - \frac{\sum_{j\neq m} \hat{x}^j}{N-1}\right)\right| \\
&\geq \left|\hat{x}^m - \frac{\sum_{j\neq m} \hat{x}^j}{N-1}\right| - c_N - |x_{k'}^m - \hat{x}^m| - \frac{1}{N-1}\sum_{j\neq m}\left|x_{k'}^j - \hat{x}^j\right| \\
&\geq \Delta - 2\epsilon \geq \frac{4N-1}{4N}\Delta > 2\epsilon.
\end{aligned}$$

Thus in step $k'+1$, the player should jump at a minimum distance of $\frac{4N-1}{4N}\Delta$, which is strictly greater than 2ϵ when $N > 1$. Therefore $\boldsymbol{x}_{k'+1} \notin B_\epsilon(\hat{\boldsymbol{x}})$, which is a contradiction. Hence $\hat{\boldsymbol{x}} = \lim_{k\to\infty} \boldsymbol{x}_k \in \partial\overline{\mathcal{CW}}$.

To see that the total distance of sequential jumps is bounded, rewrite d_k^i in the form of $d_k^i = \frac{N-1}{N}\left|x_k^i - \bar{x}_k\right|$, where $\bar{x}_k = \frac{\sum_{j=1}^N x_k^j}{N}$. Clearly, in step $k+1$, either the player with value $x_k^{(1)}$ or the player with value $x_k^{(N)}$ will jump. By the monotonicity property of $\{x_k^{(N)}\}_k$ and $\{x_k^{(1)}\}_k$, the total distance of jumps is bounded pointwise.

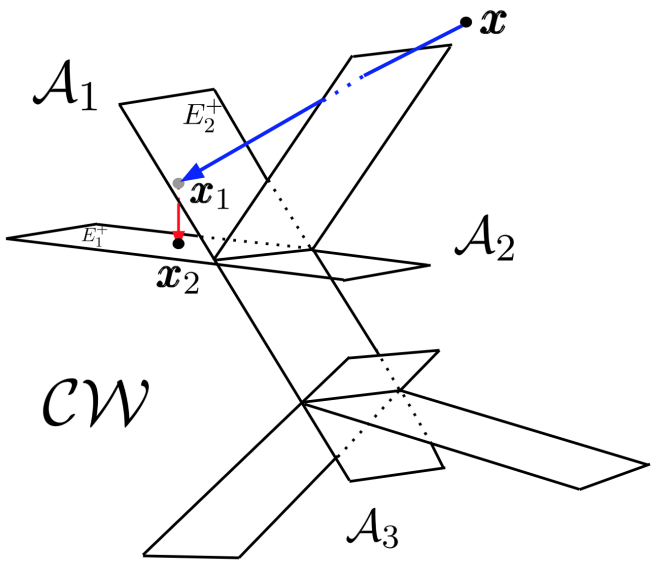


Figure A.1: Sequential jumps at time 0

A.3 Proof of Proposition 15

Proof. First, denote for $N \geq 2$,

$$f_N(x) = \frac{1}{\sqrt{\frac{2(N-1)\alpha}{N}}} \tanh\left(x\sqrt{\frac{2(N-1)\alpha}{N}}\right) - \frac{p_N'(x)-1}{p_N''(x)},$$

and

$$f_1(x) = \frac{1}{\sqrt{2\alpha}} \tanh\left(x\sqrt{2\alpha}\right) - \frac{p_1'(x)-1}{p_1''(x)}.$$

Then there exists a unique $c > 0$ such that $f_1(c) = 0$ and there exists a unique $c_N > 0$ such that $f_N(c_N) = 0$ for $N \geq 2$. Denote $m_1(x) = \frac{p_1'(x)-1}{p_1''(x)}$ and $m_N(x) = \frac{p_N'(x)-1}{p_N''(x)}$. There exists $\tilde{c}_N > 0$ such that $m_N'(x) \geq 1$ on $(\tilde{c}_N, \infty)$ with $0 < \tilde{c}_N < c_N < \infty$ for $N \geq 2$. And there exists $\tilde{c} > 0$ such that $m_1'(x) \geq 1$ on $(\tilde{c}, \infty)$ with $0 < \tilde{c} < c < \infty$. Now $0 < \tanh'(x) = 1 - \tanh^2(x) < 1$ for any $x \in (0, \infty)$, therefore $f_N'(x) < 0$ on (c_N, ∞) for $N \geq 2$ and $f_1'(x) < 0$ on (c, ∞). Since f_N converges to f_1 pointwise, for any $\epsilon > 0$, there exists an N_ϵ such that for any $n \geq N_\epsilon$, $|f_n(c)| = |f_n(c) - f_1(c)| \leq \epsilon$. By the uniqueness of the zeros for each function f_N, $c_N \to c$ as $N \to \infty$.

Secondly, when $h = x^2$, f_N reduces to

$$f_N(x) = \frac{1}{\sqrt{\frac{2(N-1)\alpha}{N}}} \tanh\left(\sqrt{\frac{2(N-1)\alpha}{N}}x\right) - x + \frac{\alpha}{2\left(\frac{N-1}{N}\right)^2},$$

with $f_N(c_N) = 0$. Therefore, $\frac{\partial c_N}{\partial N} = -\frac{\partial f_N}{\partial N} \cdot \frac{1}{\frac{\partial f_N}{\partial c_N}}$ with $\frac{\partial f_N}{\partial c_N} = -\tanh^2\left(\sqrt{\frac{2(N-1)\alpha}{N}}x\right) < 0$, the conclusion follows after simple computations.

□

A.4 Stationary Mean Field Games (SMFGs)

SMFG for (5.2.1). An SMFG version of (5.2.1) can be formulated as the follows.

$$\begin{aligned} v_\infty(x) &= \inf_{(\xi^+,\xi^-)\in\mathcal{U}_\infty} J_{(\infty)}(x;\xi_t^+,\xi_t^-) \\ &= \inf_{(\xi^+,\xi^-)\in\mathcal{U}_\infty} \mathbb{E}\int_0^\infty e^{-\alpha t}\left[h(X_t-m_\infty)dt + d\xi_t^{i,+} + d\xi_t^{i,-}\,\middle|\, X_{0-}=x\right], \\ \text{such that} \quad & X_t = B_t + \xi_t^+ - \xi_t^- + x, \\ & \mu_{0-}=\mu,\ \ X_{0-}\sim\mu,\ \ m_{0-}=m=\int x\mu_{0-}(dx), \end{aligned} \tag{A.4.1}$$

where

- $\mu_t = \lim_{N\to\infty}\frac{\sum_{i=1}^N \delta_{X_t^i}}{N}$ is the distribution of X_t,
- $m_t = \lim_{N\to\infty}\frac{\sum_{i=1}^N X_t^i}{N}$ is the mean position of the population at time t,
- $m_\infty = \lim_{t\to\infty} m_t$ is the limiting mean position if it exists.

The admissible control set for SMFG is $\mathcal{U}_\infty$ as defined in Section 2.3. SMFG is a game with the long-term mean-field aggregation.

Definition 50 (NE to SMFG (A.4.1))**.** *An NE to the SMFG (A.4.1) is a pair of Markovian control* $(\xi_t^{*,+},\xi_t^{*,-})_{t\geq 0}$ *and a limiting mean position* m^* *such that*

- $v^*(x) = J_{(\infty)}\left(x;\xi^{*,+},\xi^{*,-}|\, m^*\right) = \min_{\xi\in\mathcal{U}_\infty} J_{(\infty)}\left(x;\xi^+,\xi^-|m^*\right)$,
- $m^* = \lim_{t\to\infty}\mathbb{E}[X_t^*]$ *where* X_t^* *is the controlled dynamic under* $(\xi_t^{*,+},\xi_t^{*,-})_{t\geq 0}$.

$v^*(x)$ *is called the NE value of the SMFG associated with* ξ^*.

Appendix B

Chapter 3

B.1 Sketch proof of Theorem 24

Proof. By assumption **A5**(ii), for each $(\boldsymbol{x},\boldsymbol{y}) \in \partial G$, there is $\boldsymbol{d}(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^I_+$ such that

$$\sum_{i\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})} d_i(\boldsymbol{x},\boldsymbol{y}) = 1, \quad \text{and} \quad \min_{j\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})} \left\langle \sum_{i\in\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})} d_i(\boldsymbol{x},\boldsymbol{y})\boldsymbol{r}_i(\boldsymbol{x},\boldsymbol{y}), \boldsymbol{n}_j(\boldsymbol{x},\boldsymbol{y}) \right\rangle \geq a. \tag{B.1.1}$$

By (B.1.1), [116, Lemma 2.1] and the fact that $\boldsymbol{n}_i$ is continuous on ∂G_i for each $i \in \mathcal{I}$, we have that for each $(\boldsymbol{x},\boldsymbol{y}) \in \partial G$, there is $r_{\boldsymbol{x},\boldsymbol{y}} \in (0,\delta)$ such that for each $(\boldsymbol{x}',\boldsymbol{y}') \in B_{r_{\boldsymbol{x},\boldsymbol{y}}}(\boldsymbol{x},\boldsymbol{y}) \cap \partial G$,

$$l(\boldsymbol{x}',\boldsymbol{y}') \subset l(\boldsymbol{x},\boldsymbol{y}), \tag{B.1.2}$$

and

$$\min_{j\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})} \left\langle \sum_{i\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})} d_i(\boldsymbol{x},\boldsymbol{y})\boldsymbol{r}_i(\boldsymbol{x},\boldsymbol{y}), \boldsymbol{n}_j(\boldsymbol{x}',\boldsymbol{y}') \right\rangle \geq \frac{a}{2}. \tag{B.1.3}$$

Since ∂G_i is $\mathcal{C}^1$, for each $(\boldsymbol{x},\boldsymbol{y}) \in \partial G$, there is $m(\boldsymbol{x},\boldsymbol{y}) > 0$ and $r_{\boldsymbol{x},\boldsymbol{y}} \in (0,\delta)$ ($r_{\boldsymbol{x},\boldsymbol{y}}$ can be chosen even smaller if necessary) such that for each $(\boldsymbol{x}',\boldsymbol{y}') \in B_{r_{\boldsymbol{x},\boldsymbol{y}}}(\boldsymbol{x},\boldsymbol{y}) \cap \partial G$, (B.1.2)-(B.1.3) hold and

$$(\boldsymbol{x}',\boldsymbol{y}') + \lambda \sum_{i\in\mathcal{I}(\boldsymbol{x},\boldsymbol{y})} d_i(\boldsymbol{x},\boldsymbol{y})\boldsymbol{r}_i(\boldsymbol{x},\boldsymbol{y}) \in G \quad \text{for all } \lambda \in (0, m(\boldsymbol{x},\boldsymbol{y})). \tag{B.1.4}$$

Let $B^o_{r_{\boldsymbol{x},\boldsymbol{y}}}(\boldsymbol{x},\boldsymbol{y})$ denote the interior of the closed ball $B_{r_{\boldsymbol{x},\boldsymbol{y}}}(\boldsymbol{x},\boldsymbol{y})$. There exists a countable set $\{(\boldsymbol{x}_k,\boldsymbol{y}_k)\}$ such that $\partial G \in \cup_k B_{r_{\boldsymbol{x}_k,\boldsymbol{y}_k}}$ and $\{(\boldsymbol{x}_k,\boldsymbol{y}_k)\} \cap B_N(0)$ is a finite set for each integer $N \geq 1$. We can further choose the set $\{(\boldsymbol{x}_k,\boldsymbol{y}_k)\}$ to be minimal in the sense that for each strict subset C of $\{\boldsymbol{x}_k,\boldsymbol{y}_k\}$, $\{B_{r_{\boldsymbol{x},\boldsymbol{y}}} : (\boldsymbol{x},\boldsymbol{y}) \in C\}$ does not cover ∂G. Let $D_k = \left(B_{r_{\boldsymbol{x}_k,\boldsymbol{y}_k}}\right) \setminus \left(\cup_{i=1}^{k-1} B_{r_{\boldsymbol{x}_i,\boldsymbol{y}_i}}\right) \cap \partial G$ for each k. Then $D_k \neq \emptyset$ for each k, $\{D_k\}$ is a partition of ∂G, and for each $(\boldsymbol{x},\boldsymbol{y}) \in \partial G$ there is a unique index $i(\boldsymbol{x},\boldsymbol{y})$ such that $(\boldsymbol{x},\boldsymbol{y}) \in D_{i(\boldsymbol{x},\boldsymbol{y})}$. For each $i(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^{n+m}$, let

$$(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}}) = \begin{cases} (\boldsymbol{x},\boldsymbol{y}), & \text{if } (\boldsymbol{x},\boldsymbol{y}) \notin \partial G, \\ (\boldsymbol{x}_{i(\boldsymbol{x},\boldsymbol{y})},\boldsymbol{y}_{i(\boldsymbol{x},\boldsymbol{y})}), & \text{if } (\boldsymbol{x},\boldsymbol{y}) \in \partial G. \end{cases}$$

Note that for all $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^{n+m}$,

$$\|(\boldsymbol{x},\boldsymbol{y}) - (\bar{\boldsymbol{x}},\bar{\boldsymbol{y}})\| < \delta. \tag{B.1.5}$$

For each $i \in l$ and $(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}^{n+m}$, let

$$\boldsymbol{r}_i^\delta(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{r}_i(\bar{\boldsymbol{x}},\bar{\boldsymbol{y}}). \tag{B.1.6}$$

We construct $(\boldsymbol{X}^\delta, \boldsymbol{Y}^\delta)$ as follows. Let $\boldsymbol{W}$ be defined on some filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$ be a d-dimensional $\{\mathcal{F}_t\}$-Brownian motion with drift $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\sigma}$ such that $\boldsymbol{W}$ is continuous almost surely and W_0 has distribution ν. Let $\tau_1 := \inf\{t \geq 0 : \boldsymbol{W}_t \in \partial G\}$ and

$$\boldsymbol{X}_t^\delta = \boldsymbol{W}_t, \qquad \boldsymbol{\eta}_t^\delta = 0, \qquad \text{and } \boldsymbol{Y}_t^\delta = 0, \qquad \text{for } \ 0 \leq t < \tau_1.$$

Note that $\boldsymbol{X}^\delta_{\tau_1-}$ exists on $\{\tau_1 < \infty\}$ since $\boldsymbol{W}$ has continuous paths and in the case that $\tau_1 = 0$, $\boldsymbol{X}^\delta_{0-} = \boldsymbol{W}_0$. On $\{\tau_1 < \infty\}$, define

$$\eta^{i,\delta}_{\tau_1} = \begin{cases} 0, & i \notin l\left(\overline{\boldsymbol{X}^\delta_{\tau_1-}}, \overline{\boldsymbol{Y}^\delta_{\tau_1-}}\right), \\ d_i\left(\overline{\boldsymbol{X}^\delta_{\tau_1-}}, \overline{\boldsymbol{Y}^\delta_{\tau_1-}}\right)\left(\frac{m\left(\overline{\boldsymbol{X}^\delta_{\tau_1-}}, \overline{\boldsymbol{Y}^\delta_{\tau_1-}}\right)}{2} \wedge \delta\right), & i \in l\left(\overline{\boldsymbol{X}^\delta_{\tau_1-}}, \overline{\boldsymbol{Y}^\delta_{\tau_1-}}\right), \end{cases}$$

$$\boldsymbol{X}^\delta_{\tau_1} = \boldsymbol{X}_{\tau_1} + \left(\frac{m\left(\overline{\boldsymbol{X}^\delta_{\tau_1-}}, \overline{\boldsymbol{Y}^\delta_{\tau_1-}}\right)}{2} \wedge \delta\right)\left(\sum_{i \in l\left(\overline{\boldsymbol{X}^\delta_{\tau_1-}}, \overline{\boldsymbol{Y}^\delta_{\tau_1-}}\right)} d_i\left(\overline{\boldsymbol{X}^\delta_{\tau_1-}}, \overline{\boldsymbol{Y}^\delta_{\tau_1-}}\right) \boldsymbol{r}_i^{+,\delta}(\overline{\boldsymbol{X}^\delta_{\tau_1-}}, \overline{\boldsymbol{Y}^\delta_{\tau_1-}})\right),$$

and

$$Y^{j,\delta}_{\tau_1} = Y^{j,\delta}_{\tau_1-} + \sum_{i \in l} r_{ij}^{-,\delta}\left(\boldsymbol{X}^\delta_{\tau_1-}, \boldsymbol{Y}^\delta_{\tau_1-}\right)\left(\eta^{i,\delta}_{\tau_1} - \eta^{i,\delta}_{\tau_1-}\right) \quad \text{for } j = 1, 2, \cdots m.$$

So $\boldsymbol{X}^\delta, \boldsymbol{\eta}^\delta$ and $\boldsymbol{Y}^\delta$ have been defined on $[0, \tau_1)$ and at τ_1 on $\{\tau_1 < \infty\}$, such that

(i) $\boldsymbol{X}_t^\delta = \boldsymbol{W}_t + \sum_{i \in l} \boldsymbol{r}_i^{+,\delta}(\boldsymbol{X}^\delta_{0-})\eta_0^{i,\delta} + \sum_{i \in l}\int_{(0,t]} \boldsymbol{r}_i^{+,\delta}(\boldsymbol{X}^\delta_{s-}, \boldsymbol{Y}^\delta_{s-}) d\eta_s^{i,\delta}$ and
$\boldsymbol{Y}_t^\delta = \sum_{i \in l} \boldsymbol{r}_i^{-,\delta}(\boldsymbol{Y}^\delta_{0-})\eta_0^{i,\delta} + \sum_{i \in l}\int_{(0,t]} \boldsymbol{r}_i^{-,\delta}(\boldsymbol{X}^\delta_{s-}, \boldsymbol{Y}^\delta_{s-}) d\eta_s^{i,\delta}$ for all $t \in [0, \tau_1] \cap [0, \infty)$,
where $\boldsymbol{X}^\delta_{0-} = \boldsymbol{W}_0$ and $\boldsymbol{Y}^\delta_{0-} = 0$.

(ii) $(\boldsymbol{X}_t^\delta, \boldsymbol{Y}_t^\delta) \in \overline{G}$

(iii) for $i \in l$,

- **(a)** $\eta^{i,\delta} \geq 0$,
- **(b)** $\eta^{i,\delta}$ is nondecreasing on $[0, \tau_1] \cap [0, \infty)$,
- **(c)** $\eta^{i,\delta} = \eta_0^{i,\delta} + \int_{(0,t]} 1_{\{(\boldsymbol{X}^\delta_s, \boldsymbol{Y}^\delta_s) \in U_{2\delta}(\partial G \cap \partial G_i)\}} d\eta_s^{i,\delta}$ for $t \in [0, \tau_1] \cap [0, \infty)$,

(iv) $\|\Delta\boldsymbol{\eta}_t^\delta\| = \|\boldsymbol{\eta}_t^\delta - \boldsymbol{\eta}_{t-}^\delta\| \leq \delta$ for $t \in [0, \tau_1] \cap [0, \infty)$, where where $\boldsymbol{\eta}^\delta_{0-} = \boldsymbol{0}$.

Proceeding by induction, we assume that for some $n \geq 2$, $\tau_1 \leq \cdots \leq \tau_{n-1}$ have been defined, and $(\boldsymbol{X}^\delta, \boldsymbol{Y}^\delta, \boldsymbol{\eta}^\delta)$ has been defined on $[0, \tau_{t-1})$ and at τ_{t-1} on $\{\tau_{t-1} < \infty\}$, such that $(i) - (iv)$ above hold with τ_{n-1} in place of τ_1. Then we define $\tau_n = \infty$ on $\{\tau_{n-1} = \infty\}$ and on $\{\tau_{n-1} < \infty\}$ we define

$$\tau_n = \inf\{t \geq \tau_{n-1} : \left(\boldsymbol{X}^\delta_{\tau_{n-1}} + \boldsymbol{W}_t - \boldsymbol{W}_{\tau_{n-1}}, \boldsymbol{Y}^\delta_{\tau_{n-1}}\right) \in \partial G\}.$$

Note that between τ_{n-1} and τ_n, the resource level $\boldsymbol{Y}^\delta_t$ remains constant while X^δ_t behaves like a Brownian motion.

For $\tau_{n-1} \leq t < \tau_n$, let

$$\begin{aligned} \boldsymbol{\eta}^\delta_t &= \boldsymbol{\eta}^\delta_{\tau_{n-1}}, \\ \boldsymbol{Y}^\delta_t &= \boldsymbol{Y}^\delta_{\tau_{n-1}}, \\ \boldsymbol{X}^\delta_t &= \boldsymbol{X}^\delta_{\tau_{n-1}} + \boldsymbol{W}_t - \boldsymbol{W}_{\tau_{n-1}}. \end{aligned}$$

On $\{\tau_n < \infty\}$, let

$$\eta^{i,\delta}_{\tau_n} = \begin{cases} \eta^{i,\delta}_{\tau_{n-1}}, & i \notin l\left(\overline{\boldsymbol{X}^\delta_{\tau_n-}}, \overline{\boldsymbol{Y}^\delta_{\tau_n-}}\right), \\ d_i\left(\overline{\boldsymbol{X}^\delta_{\tau_{n-1}-}}, \overline{\boldsymbol{Y}^\delta_{\tau_{n-1}-}}\right)\left(\frac{m\left(\overline{\boldsymbol{X}^\delta_{\tau_{n-1}-}}, \overline{\boldsymbol{Y}^\delta_{\tau_{n-1}-}}\right)}{2} \wedge \delta\right), & i \in l\left(\overline{\boldsymbol{X}^\delta_{\tau_n-}}, \overline{\boldsymbol{Y}^\delta_{\tau_n-}}\right), \end{cases}$$

$$\boldsymbol{X}^\delta_{\tau_n} = \boldsymbol{X}_{\tau_n-} + \left(\frac{m\left(\overline{\boldsymbol{X}^\delta_{\tau_n-}}, \overline{\boldsymbol{Y}^\delta_{\tau_n-}}\right)}{2} \wedge \delta\right)\left(\sum_{i \in l\left(\overline{\boldsymbol{X}^\delta_{\tau_n-}}, \overline{\boldsymbol{Y}^\delta_{\tau_n-}}\right)} d_i\left(\overline{\boldsymbol{X}^\delta_{\tau_n-}}, \overline{\boldsymbol{Y}^\delta_{\tau_n-}}\right) \boldsymbol{r}^{+,\delta}_i(\overline{\boldsymbol{X}^\delta_{\tau_n-}}, \overline{\boldsymbol{Y}^\delta_{\tau_n-}})\right),$$

and

$$Y^{j,\delta}_{\tau_n} = Y^{j,\delta}_{\tau_n-} + \sum_{i \in l} r^{-,\delta}_{ij}\left(\boldsymbol{X}^\delta_{\tau_n-}, \boldsymbol{Y}^\delta_{\tau_n-}\right)\left(\eta^{i,\delta}_{\tau_n} - \eta^{i,\delta}_{\tau_n-}\right) \quad \text{for } j = 1, 2, \cdots m.$$

In this way, $\boldsymbol{X}^\delta$, $\boldsymbol{\eta}^\delta$ and $\boldsymbol{Y}^\delta$ have been defined on $[0, \tau_n)$ and at τ_n on $\{\tau_n < \infty\}$ such that (i)-(iv) hold with τ_n in place of τ_1. By construction $\{\tau_n\}_{n=1}^\infty$ is a nondecreasing sequence of stopping times. Let $\tau = \lim_{n\to\infty} \tau_n$. On $\{\tau = \infty\}$, the construction of $(\boldsymbol{X}^\delta, \boldsymbol{\eta}^\delta, \boldsymbol{Y}^\delta)$ is complete. A similar argument in [116, Theorem 5.1] shows that $\{\tau < \infty\} = \emptyset$.

Consider a sequence of sufficiently small δ's, denoted by $\{\delta^n\}$, such that $\delta^n \downarrow 0$ as $n \to \infty$. For each δ^n, let $(\boldsymbol{X}^{\delta^n}, \boldsymbol{Y}^{\delta^n}, \boldsymbol{\eta}^{\delta^n})$ be the tuple constructed as above for the same diffusion process $\boldsymbol{W}$ with drift $\boldsymbol{b}$ and covariance matrix $\boldsymbol{\sigma}$. Assumption 4.1 in [116] is satisfied with $\alpha^n = 0$, $\beta^n = 0$ and and $2\delta^n$ in place of δ^n. Denote $\boldsymbol{W}^n = \boldsymbol{W} + \sum_{i \in l} \boldsymbol{r}^{+,\delta^n}_i(\boldsymbol{X}^{\delta^n}_{0-}, \boldsymbol{Y}^{\delta^n}_{0-})\eta^{i,\delta^n}_0$. Consequently, $\{\boldsymbol{\mathcal{Z}}^{\delta^n}\}_{n=1}^\infty := \{(\boldsymbol{W}^{\delta^n}, \boldsymbol{X}^{\delta^n}, \boldsymbol{Y}^{\delta^n}, \boldsymbol{\eta}^{\delta^n})\}_{n=1}^\infty$ is C-tight and any weak limit point $\boldsymbol{\mathcal{Z}}$ of this sequence satisfies conditions (i), (iii), (iv) and (v) and in Definition 23 with $\mathcal{F}_t = \sigma(\boldsymbol{Z}_s : 0 \leq s \leq t)$, $t \geq 0$.

It is straightforward that $\boldsymbol{W}^n$ converges to Brownian motion with drift $\boldsymbol{b}$ in D. In addition, $\boldsymbol{M}^{\delta^n} := \{\boldsymbol{W}^{\delta^n}_t - \boldsymbol{W}^{\delta^n}_0 - \boldsymbol{b}t, t \geq 0\} = \{\boldsymbol{W}_t - \boldsymbol{W}_0 - \boldsymbol{b}t, t \geq 0\}$ is a martingale with respect to $\boldsymbol{W}$ which

$(\boldsymbol{X}^{\delta^n}, \boldsymbol{Y}^{\delta^n}, \boldsymbol{\eta}^{\delta^n})$ is adapted to. Therefore $\boldsymbol{M}^{\delta^n}$ is a martingale with respect to $(\boldsymbol{W}^{\delta^n}, \boldsymbol{X}^{\delta^n}, \boldsymbol{Y}^{\delta^n}, \boldsymbol{\eta}^{\delta^n}))$ and it is also uniformly integrable.

Hence by Proposition 4.1 in [116], any weak limit point of $\{\boldsymbol{Z}^{\delta^n}\}_{n=1}^{\infty}$ is an extended constrained SRBM with data $(G, \boldsymbol{b}, \boldsymbol{\sigma}, \{\boldsymbol{r}_i, i \in l\}, \nu)$.

□

B.2 Satisfiability for Assumptions A1-A5

Take $n = N$, $m = M$ and $I = 2N$ in Definition 23. We then check the satisfiability for Assumptions **A1**-**A5** for game $\boldsymbol{C}$. $\boldsymbol{C_p}$ and $\boldsymbol{C_d}$ are two special cases.

A1

Assumption **A1** is trivially satisfied by definition. We write

$$G = \cap_{j=1}^{2N} G_j,$$

where

$$G_i = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{N+M} \middle| \, \tilde{x}_i \leq f_N^{-1}\left(\sum_{j=1}^{M} a_{ij} y^j \right) \right\},$$

$$G_{N+i} = \left\{ (\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{N+M} \middle| \, \tilde{x}_i \geq -f_N^{-1}\left(\sum_{j=1}^{M} a_{ij} y^j \right) \right\},$$

for $i = 1, 2, \cdots, N$. The boundary of G_i is smooth since f_N^{-1} is smooth.

A2

Assumption **A2** is satisfied since f_N^{-1} is smooth and decreasing. It satisfies the uniform exterior cone condition. At any boundary point $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \partial G_j$, we can put a truncated closed right circular cone $V_{(\boldsymbol{x}_0, \boldsymbol{y}_0)}$ satisfying $V_{(\boldsymbol{x}_0, \boldsymbol{y}_0)} \cap \bar{G} = \{(\boldsymbol{x}_0, \boldsymbol{y}_0)\}$.

A3

Assumption **A3** can be shown by contradiction. The proof is inspired from that of [116, Lemma (A.2)] which is for bounded region with tightness argument. We modify the proof via a shifting argument.

Suppose that Assumption **A3** does not hold. Then, since there are only finite many $l_0 \in l$, $l_0 \neq \emptyset$, there is an $\epsilon > 0$, a nonempty set $l_0 \subset l$, a sequence $\{\epsilon_n\} \subset (0, \infty)$ with $\epsilon_n \to 0$ as $n \to \infty$, a sequence $\{(\boldsymbol{x}_n, \boldsymbol{y}_n)\} \subset \mathbb{R}^{N+M}$ such that for each n, $(\boldsymbol{x}_n, \boldsymbol{y}_n) \in \cap_{j \in l_0} U_{\epsilon_n}(\partial G_j \cap \partial G)$ and $\text{dist}((\boldsymbol{x}_n, \boldsymbol{y}_n), \cap_{j \in l_0}(\partial G_j \cap \partial G)) \geq \epsilon$.

By exploiting the special structure of region G, $\text{dist}((\boldsymbol{x}, \boldsymbol{y}), \cap_{j \in l_0}(\partial G_j \cap \partial G)) = \text{dist}((\boldsymbol{x} - a\boldsymbol{1}, \boldsymbol{y}), \cap_{j \in l_0}(\partial G_j \cap \partial G))$ for any $a \in \mathbb{R}$ and $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{N+M}$. Here $\boldsymbol{1} \in \mathbb{R}^N$ is a vector with all

ones. Intuitively, this is because for any fixed $\boldsymbol{y}$, the projection of G onto $\boldsymbol{x}$-space is a polyhedron unbounded along the directions of $\mathbf{1} \in \mathbb{R}^N$. This is consistent with the model where we only look at the relative distance between positions.

Therefore, for each $(\boldsymbol{x}_n, \boldsymbol{y}_n)$, there exists $a_n \in \mathbb{R}$ such that $\|\boldsymbol{x}_n - a_n\mathbf{1}\| \leq 1$. Denote $\tilde{\boldsymbol{x}}_n = \boldsymbol{x}_n - a_n\mathbf{1}$. Hence $(\tilde{\boldsymbol{x}}_n, \boldsymbol{y}_n)$ is a bounded sequence in $\mathbb{R}^{N+M}$ and $\text{dist}((\tilde{\boldsymbol{x}}_n, \boldsymbol{y}_n), \cap_{j\in l_0}(\partial G_j \cap \partial G)) \geq \epsilon$. WLOG, we may assume that $(\tilde{\boldsymbol{x}}_n, \boldsymbol{y}_n) \to (\boldsymbol{x}, \boldsymbol{y})$ as $n \to \infty$ for some $(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{R}^{N+M}$. It follows that $(\boldsymbol{x}, \boldsymbol{y}) \in \cap_{j\in l_0}(\partial G_j \cap \partial G)$, since for each $j \in l_0$,

$$\text{dist}((\boldsymbol{x}, \boldsymbol{y}), \partial G_j \cap \partial G) \leq \|(\tilde{\boldsymbol{x}}_n, \boldsymbol{y}_n) - (\boldsymbol{x}, \boldsymbol{y})\| + \text{dist}((\tilde{\boldsymbol{x}}_n, \boldsymbol{y}_n), \partial G_j \cap \partial G) \leq \|(\tilde{\boldsymbol{x}}_n, \boldsymbol{y}_n) - (\boldsymbol{x}, \boldsymbol{y})\| + \epsilon_n \to 0,$$

as $n \to \infty$. This contradicts with the fact that $(\tilde{\boldsymbol{x}}_n, \boldsymbol{y}_n) \to (\boldsymbol{x}, \boldsymbol{y})$ and $\text{dist}((\tilde{\boldsymbol{x}}_n, \boldsymbol{y}_n), \cap_{j\in l_0}(\partial G_j \cap \partial G)) \geq \epsilon$.

A4

Recall that for $i = 1, 2, \cdots, N$,

$$\begin{aligned}
\boldsymbol{r}_i &= c_i' \left(0 \cdots, -1, \cdots 0; -\frac{a_{i1}y^1}{\sum_{j=1}^M a_{ij}y^j}, \cdots, -\frac{a_{iM}y^M}{\sum_{j=1}^M a_{ij}y^j}\right), \\
\boldsymbol{r}_{N+i} &= c_{N+i}' \left(0 \cdots, 1, \cdots 0; -\frac{a_{i1}y^1}{\sum_{j=1}^M a_{ij}y^j}, \cdots, -\frac{a_{iM}y^M}{\sum_{j=1}^M a_{ij}y^j}\right),
\end{aligned}$$

where c_j' is a normalizing constant such that $\|\boldsymbol{r}_j\| = 1$ $(j = 1, 2, \cdots, 2N)$.

On each face $j = 1, 2, \cdots, 2N$, $\boldsymbol{r}_j$ is a function of $\boldsymbol{y}$, which is bounded. Moreover, $\boldsymbol{r}_j$ is smooth and $D_{\boldsymbol{y}}\boldsymbol{r}_j$ is bounded. Therefore, $\boldsymbol{r}_j(\cdot)$ is uniformly Lipschitz continuous function. Note that when the adjacent matrix $A = \{a_{kj}\}_{1\leq k,j\leq N}$ is an identity matrix or matrix with all ones, $\boldsymbol{r}_i$ is constant on ∂G_i for all $i \in l$.

A5

Denote $g := f_N^{-1}$. First we show that g is a non-negative decreasing function on $[0, y]$ where $y := \sum_{j=1}^M y^j$ is the total resource.

Recall that

$$f_N'(x) = \frac{p_N' - \frac{N}{2(N-1)\alpha}p_N'''}{p_N''\sqrt{\frac{N}{2(N-1)\alpha}}\tanh\left(x\sqrt{\frac{2(N-1)\alpha}{N}}\right) - p_N'}.$$

We claim that $f_N'(x) < 0$ when $x \geq 0$ and $\lim_{x\downarrow 0} f_N'(x) = -\infty$. Since $h'(x) \geq 0$ and $h'''(x) \leq 0$ for $x \geq 0$, we have $p_N' - \frac{N}{2(N-1)\alpha}p_N''' \geq 0$ for $x \geq 0$. Denote $q(x) = p_N''\sqrt{\frac{N}{2(N-1)\alpha}}\tanh\left(x\sqrt{\frac{2(N-1)\alpha}{N}}\right) - p_N'$. It is easy to check that $q(0) = 0$. Moreover, $q'(x) = p_N'''\sqrt{\frac{N}{2(N-1)\alpha}}\tanh\left(x\sqrt{\frac{2(N-1)\alpha}{N}}\right) + p_N''\frac{1}{\cosh^2\left(x\sqrt{\frac{2(N-1)\alpha}{N}}\right)} - p_N'' < 0$ for $x > 0$ and $q'(x) = 0$ for $x = 0$. This is because $h''' \leq 0$

$(x \geq 0)$, $\cosh(x) \geq 1$ $(x \geq 0)$, and $\cosh(x) = 1$ if and only if $x = 0$. Moreover, given the fact that $\lim_{x\downarrow 0} f_N(x) = \infty$, $f'_N(x)$ is not bounded as $x \downarrow 0$, we have $\lim_{x\downarrow 0} f'_N(x) = -\infty$.

Combining all above, $f'_N(x) < 0$ for $x \geq 0$. Therefore, there exists $0 < \tilde{k}(y) < \tilde{K}(y) < \infty$ such that $-\infty < -\tilde{K}(y) < f'_N(z) < -\tilde{k}(y) < 0$ when $z \in [\underline{x}, \overline{x}]$. Here $\underline{x} = g(y) > 0$ and $\overline{x} = g(0)$. Note that $g'(\cdot) = \frac{1}{f'(f^{-1}(\cdot))}$, therefore $-\frac{1}{\tilde{k}(y)} \leq g'(w) \leq -\frac{1}{\tilde{K}(y)}$ when $w \in [0, y]$. Now let $k(y) := \frac{1}{\tilde{K}(y)}$ and $K(y) := \frac{1}{\tilde{k}(y)}$.

Next, Recall that

$$\begin{aligned}
\boldsymbol{n}_i &= c_i \left(\frac{1}{N-1}, \cdots, -1, \cdots, \frac{1}{N-1}; g'\left(\sum_{j=1}^{M} a_{ij} y^j\right) a_{i1}, \cdots, g'\left(\sum_{j=1}^{M} a_{ij} y^j\right) a_{iM} \right), \\
\boldsymbol{r}_i &= c'_i \left(0 \cdots, -1, \cdots 0; -\frac{a_{i1} y^1}{\sum_{j=1}^{M} a_{ij} y^j}, \cdots, -\frac{a_{iM} y^M}{\sum_{j=1}^{M} a_{ij} y^j} \right), \\
\boldsymbol{n}_{N+i} &= c_{N+i} \left(-\frac{1}{N-1}, \cdots, 1, \cdots, -\frac{1}{N-1}; g'\left(\sum_{j=1}^{M} a_{ij} y^j\right) a_{i1}, \cdots, g'\left(\sum_{j=1}^{M} a_{ij} y^j\right) a_{iM} \right), \\
\boldsymbol{r}_{N+i} &= c'_{N+i} \left(0 \cdots, 1, \cdots 0; -\frac{a_{i1} y^1}{\sum_{j=1}^{M} a_{ij} y^j}, \cdots, -\frac{a_{iM} y^M}{\sum_{j=1}^{M} a_{ij} y^j} \right),
\end{aligned}$$

where ± 1 is on the i-th position. Obviously all the latter M components in $\boldsymbol{n}_j$ and $\boldsymbol{r}_j$ are non-positive $(1 \leq j \leq 2N)$.

By simple calculation, we have $\frac{1}{\sqrt{\frac{N}{N-1} + K(y)N}} \leq c_j \leq \frac{1}{\sqrt{\frac{N}{N-1} + \epsilon}}$ and $\sqrt{\frac{N}{N+1}} \leq c'_j \leq \frac{1}{\sqrt{2}}$for all $1 \leq j \leq N$. Similar to the definition of $\boldsymbol{r}_j^+$ and $\boldsymbol{r}_j^-$, denote $\boldsymbol{n}_j^+$ as the first N components in $\boldsymbol{n}^j$ and $\boldsymbol{n}_j^-$ as the latter M components in $\boldsymbol{n}^j$.

Since face i and $N+i$ are parallel to each other $(i = 1, 2, \cdots, N)$, there are at most N faces intersecting with each other. It suffices to consider $(\boldsymbol{x}, \boldsymbol{y})$ such that $|\mathcal{I}((\boldsymbol{x}, \boldsymbol{y}))| = N$. For these points, consider $c_i = \frac{1}{N}$ and $d_i = \frac{1}{N}$ $(i = 1, 2, \cdots, N)$. Therefore, for $i^* \in \{i, N+i\}$ with $i = 1, 2, \cdots, N$,

$$\left\langle \frac{\sum_{i=1}^{N} \boldsymbol{n}_{i*}}{N}, \boldsymbol{r}_{i*} \right\rangle \geq \frac{1}{N} \langle \boldsymbol{n}_{i*}^-, \boldsymbol{r}_{i*}^- \rangle = \frac{1}{N} c'_{i^*} c_{i^*} \langle \boldsymbol{n}_{i*}^-, \boldsymbol{r}_{i*}^- \rangle = -c'_{i^*} c_{i^*} g'\left(\sum_{j=1}^{M} a_{ij} y^j\right) \geq \frac{1}{\sqrt{\frac{N+1}{N-1} + (N+1)K(y)}} k(y).$$

Similarly, for $i^* \in \{i, N+i\}$ with $i = 1, 2, \cdots, N$,

$$\left\langle \frac{\sum_{i=1}^{N} \boldsymbol{r}_{i*}}{N}, n_{i*} \right\rangle \geq \frac{1}{N} \langle \boldsymbol{n}_{i*}^-, \boldsymbol{r}_{i*}^- \rangle = \frac{1}{N} \langle \boldsymbol{n}_{i*}^-, \boldsymbol{r}_{i*}^- \rangle = -c'_{i^*} c_{i^*} g'\left(\sum_{j=1}^{M} a_{ij} y^j\right) \geq \frac{1}{\sqrt{\frac{N+1}{N-1} + (N+1)K(y)}} k(y).$$

B.3 The unique positive root to (3.4.9)

Define $q(z) = \frac{p_N''(z)}{p_N'(z)}$ where $p_N(x)$ is defined in (3.4.6). Note that

$$q(0) = \frac{p_N''(0)}{p_N'(0)} = \frac{\mathbb{E}\int_0^\infty e^{-\alpha t} h''\left(\sqrt{\frac{N-1}{N}} B_t\right) dt}{\mathbb{E}\int_0^\infty e^{-\alpha t} h'\left(\sqrt{\frac{N-1}{N}} B_t\right) dt}.$$

Under Assumption **H2′**, $p_N'(0) = 0$, $\frac{k}{\alpha} < p_N''(0) < \frac{K}{\alpha}$, and

$$q'(z) = \frac{p_N'''(z)p_N'(z) - (p_N''(z))^2}{(p_N'(z))^2}.$$

Moreover, Assumption **H2′** implies that $h'''(z) \le 0$ and $h'(z) \ge 0$ for $z \ge 0$. Therefore, $q(0) = \infty$ and $q'(z) \le 0$. Furthermore, since $k \le h'' \le K$ and $h' \ge kx + c$ for some constant c, we have $\lim_{x\to\infty} q(x) = 0$.

On the other hand, define $f(x) = \sqrt{\frac{2(N-1)\alpha}{N}} \tanh\left(z\sqrt{\frac{2(N-1)\alpha}{N}}\right)$. It is easy to check that $f(0) = 0$, $f'(x) > 0$ for $x \ge 0$, and $\lim_{x\to\infty} f(x) = \sqrt{\frac{2(N-1)\alpha}{N}}$. Therefore, $f(x) = q(x)$ has a unique positive solution.

Appendix C

Chapter 5

C.1 Distance Metrics and Completeness

This section reviews some basic properties of the Wasserstein distance. It then proves that the metrics defined in the main text are indeed distance functions and define complete metric spaces.

ℓ_1-Wasserstein distance and dual representation. The ℓ_1 Wasserstein distance over $\mathcal{P}(\mathcal{X})$ for $\mathcal{X} \subseteq \mathbb{R}^k$ is defined as

$$W_1(\nu, \nu') := \inf_{M \in \mathcal{M}(\nu,\nu')} \int_{\mathcal{X}\times\mathcal{X}} \|x - y\|_2 \mathrm{d}M(x, y). \tag{C.1.1}$$

where $\mathcal{M}(\nu, \nu')$ is the set of all measures (couplings) on $\mathcal{X} \times \mathcal{X}$, with marginals ν and ν' on the two components, respectively.

The Kantorovich duality theorem enables the following equivalent dual representation of W_1:

$$W_1(\nu, \nu') = \sup_{\|f\|_L \leq 1} \left| \int_{\mathcal{X}} f d\nu - \int_{\mathcal{X}} f d\nu' \right|, \tag{C.1.2}$$

where the supremum is taken over all 1-Lipschitz functions f, *i.e.*, f satisfying $|f(x) - f(y)| \leq \|x - y\|_2$ for all $x, y \in \mathcal{X}$.

The Wasserstein distance W_1 can also be related to the total variation distance via the following inequalities [85]:

$$d_{\min}(\mathcal{X}) d_{TV}(\nu, \nu') \leq W_1(\nu, \nu') \leq \mathrm{diam}(\mathcal{X}) d_{TV}(\nu, \nu'), \tag{C.1.3}$$

where $d_{\min}(\mathcal{X}) = \min_{x \neq y \in \mathcal{X}} \|x - y\|_2$, which is guaranteed to be positive when $\mathcal{X}$ is finite.

When $\mathcal{S}$ and $\mathcal{A}$ are compact, for any compact subset $\mathcal{X} \subseteq \mathbb{R}^k$, and for any $\nu, \nu' \in \mathcal{P}(\mathcal{X})$, $W_1(\nu, \nu') \leq \mathrm{diam}(\mathcal{X}) d_{TV}(\nu, \nu') \leq \mathrm{diam}(\mathcal{X}) < \infty$, where $\mathrm{diam}(\mathcal{X}) = \sup_{x,y \in \mathcal{X}} \|x - y\|_2$ and d_{TV} is the total variation distance. Moreover, one can verify

Lemma 51. *Both D and $\mathcal{W}_1$ are distance functions, and they are finite for any input distribution pairs. In addition, both $(\{\Pi\}_{t=0}^{\infty}, D)$ and $(\{\mathcal{P}(\mathcal{S} \times \mathcal{A})\}_{t=0}^{\infty}, \mathcal{W}_1)$ are complete metric spaces.*

These facts enable the usage of Banach fixed-point mapping theorem for the proof of existence and uniqueness (Theorems 44 and 53).

Proof of Lemma 51. It is known that for any compact set $\mathcal{X} \subseteq \mathbb{R}^k$, $(\mathcal{P}(\mathcal{X}), W_1)$ defines a complete metric space [27]. Since $W_1(\nu, \nu') \leq \text{diam}(\mathcal{X})$ is uniformly bounded for any $\nu,\ \nu' \in \mathcal{P}(\mathcal{X})$, we know that $\mathcal{W}_1(\boldsymbol{\mathcal{L}}, \boldsymbol{\mathcal{L}}') \leq \text{diam}(\mathcal{X})$ and $D(\boldsymbol{\pi}, \boldsymbol{\pi}') \leq \text{diam}(\mathcal{X})$ as well, so they are both finite for any input distribution pairs. It is clear that they are distance functions based on the fact that W_1 is a distance function.

Finally, we show the completeness of the two metric spaces $(\{\Pi\}_{t=0}^{\infty}, D)$ and $(\{\mathcal{P}(\mathcal{S}\times\mathcal{A})\}_{t=0}^{\infty}, \mathcal{W}_1)$. Take $(\{\Pi\}_{t=0}^{\infty}, D)$ for example. Suppose that $\boldsymbol{\pi}^k$ is a Cauchy sequence in $(\{\Pi\}_{t=0}^{\infty}, D)$. Then for any $\epsilon > 0$, there exists a positive integer N, such that for any $m,\ n \geq N$,

$$D(\boldsymbol{\pi}^n, \boldsymbol{\pi}^m) \leq \epsilon \Longrightarrow W_1(\pi_t^n(s), \pi_t^m(s)) \leq \epsilon \text{ for any } s \in \mathcal{S},\ t \in \mathbb{N}, \tag{C.1.4}$$

which implies that $\pi_t^k(s)$ forms a Cauchy sequence in $(\mathcal{P}(\mathcal{A}), W_1)$, and hence by the completeness of $(\mathcal{P}(\mathcal{A}), W_1)$, $\pi_t^k(s)$ converges to some $\pi_t(s) \in \mathcal{P}(\mathcal{A})$. As a result, $\boldsymbol{\pi}^n \to \boldsymbol{\pi} \in \{\Pi\}_{t=0}^{\infty}$ under metric D, which shows that $(\{\Pi\}_{t=0}^{\infty}, D)$ is complete.

The completeness of $(\{\mathcal{P}(\mathcal{S} \times \mathcal{A})\}_{t=0}^{\infty}, \mathcal{W}_1)$ can be proved similarly. □

The same argument for Lemma 51 shows that both D and W_1 are distance functions and are finite for any input distribution pairs, with both (Π, D) and $(\mathcal{P}(\mathcal{S} \times \mathcal{A}), W_1)$ again complete metric spaces.

C.2 Existence and Uniqueness for Stationary NE of GMFGs

Definition 52 (Stationary NE for GMFGs)**.** *In* (GMFG)*, a player-population profile $(\pi^\star,\ \mathcal{L}^\star)$ is called a stationary NE if*

1. *(Single player side) For any policy π and any initial state $s \in \mathcal{S}$,*

$$V\left(s, \pi^\star, \mathcal{L}^\star\right) \geq V\left(s, \pi, \mathcal{L}^\star\right). \tag{C.2.1}$$

2. *(Population side) $\mathbb{P}_{s_t, a_t} = \mathcal{L}^\star$ for all $t \geq 0$, where $\{s_t, a_t\}_{t=0}^{\infty}$ is the dynamics under the policy $\pi^\star$ starting from $s_0 \sim \mu^\star$, with $a_t \sim \pi^\star(s_t, \mu^\star)$, $s_{t+1} \sim P(\cdot|s_t, a_t, \mathcal{L}^\star)$, and $\mu^\star$ being the population state marginal of $\mathcal{L}^\star$.*

The existence and uniqueness of the NE to (GMFG) in the stationary setting can be established by modifying appropriately the same fixed-point approach for the GMFG in the main text.

Step 1. Fix $\mathcal{L}$, the GMFG becomes the classical optimization problem. That is, solving (GMFG) is now reduced to finding a policy $\pi_{\mathcal{L}}^\star \in \Pi := \{\pi \,|\, \pi : \mathcal{S} \to \mathcal{P}(\mathcal{A})\}$ to maximize

$$\begin{aligned} V(s, \pi_{\mathcal{L}}, \mathcal{L}) := \quad & \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t, \mathcal{L}) | s_0 = s\right], \\ \text{subject to} \quad & s_{t+1} \sim P(s_t, a_t, \mathcal{L}), \quad a_t \sim \pi_{\mathcal{L}}(s_t). \end{aligned}$$

Now given this fixed $\mathcal{L}$ and the solution $\pi^\star_{\mathcal{L}}$ to the above optimization problem, one can again define

$$\Gamma_1 : \mathcal{P}(\mathcal{S} \times \mathcal{A}) \to \Pi,$$

such that $\pi^\star_{\mathcal{L}} = \Gamma_1(\mathcal{L})$. Note that this $\pi^\star_{\mathcal{L}}$ satisfies the single player side condition for the population state-action pair L,

$$V\left(s, \pi^\star_{\mathcal{L}}, \mathcal{L}\right) \geq V\left(s, \pi, \mathcal{L}\right), \tag{C.2.2}$$

for any policy π and any initial state $s \in \mathcal{S}$.

Accordingly, a similar feedback regularity (FR) condition is needed in this step.

Assumption 3. *There exists a constant $d_1 \geq 0$, such that for any $\mathcal{L}, \mathcal{L}' \in \mathcal{P}(\mathcal{S} \times \mathcal{A})$,*

$$D(\Gamma_1(\mathcal{L}), \Gamma_1(\mathcal{L}')) \leq d_1 W_1(\mathcal{L}, \mathcal{L}'), \tag{C.2.3}$$

where

$$D(\pi, \pi') := \sup_{s \in \mathcal{S}} W_1(\pi(s), \pi'(s)), \tag{C.2.4}$$

and W_1 is the ℓ_1-Wasserstein distance (a.k.a. earth mover distance) between probability measures.

Step 2. Based on the analysis of Step 1 and $\pi^\star_{\mathcal{L}}$, update the initial $\mathcal{L}$ to $\mathcal{L}'$ following the controlled dynamics $P(\cdot|s_t, a_t, \mathcal{L})$.

Accordingly, define a mapping $\Gamma_2 : \Pi \times \mathcal{P}(\mathcal{S} \times \mathcal{A}) \to \mathcal{P}(\mathcal{S} \times \mathcal{A})$ as follows:

$$\Gamma_2(\pi, \mathcal{L}) := \hat{\mathcal{L}} = \mathbb{P}_{s_1, a_1}, \tag{C.2.5}$$

where $a_1 \sim \pi(s_1)$, $s_1 \sim \mu P(\cdot|\cdot, a_0, \mathcal{L})$, $a_0 \sim \pi(s_0)$, $s_0 \sim \mu$, and μ is the population state marginal of $\mathcal{L}$.

One also needs a similar assumption in this step.

Assumption 4. *There exist constants $d_2,\ d_3 \geq 0$, such that for any admissible policies π, π_1, π_2 and joint distributions $\mathcal{L}, \mathcal{L}_1, \mathcal{L}_2$,*

$$W_1(\Gamma_2(\pi_1, \mathcal{L}), \Gamma_2(\pi_2, \mathcal{L})) \leq d_2 D(\pi_1, \pi_2), \tag{C.2.6}$$

$$W_1(\Gamma_2(\pi, \mathcal{L}_1), \Gamma_2(\pi, \mathcal{L}_2)) \leq d_3 W_1(\mathcal{L}_1, \mathcal{L}_2). \tag{C.2.7}$$

Step 3. Repeat until $\mathcal{L}'$ matches $\mathcal{L}$.

This step is to ensure the population side condition. To ensure the convergence of the combined step one and step two, it suffices if $\Gamma : \mathcal{P}(\mathcal{S} \times \mathcal{A}) \to \mathcal{P}(\mathcal{S} \times \mathcal{A})$ with $\Gamma(\mathcal{L}) := \Gamma_2(\Gamma_1(\mathcal{L}), \mathcal{L})$ is a contractive mapping (under the W_1 distance).

Similar to the proof of Theorem 44, again by the Banach fixed point theorem and the completeness of the related metric spaces, there exists a unique stationary NE of the GMFG. That is,

Theorem 53 (Existence and Uniqueness of stationary MFG solution)**.** *Given Assumptions 3 and 4, and assume $d_1 d_2 + d_3 < 1$. Then there exists a unique stationary NE to* (GMFG)*.*

C.3 Additional Comments on Assumptions

As mentioned in the main text, the single player side Assumption 1 and its counterpart Assumption 3 for the stationary version correspond to the feedback regularity (FR) condition in the classical MFG literature. Here we add some comments on the population side Assumption 2 and its stationary version Assumption 4. For simplicity and clarity, let us consider the stationary case with finite state and action spaces. Then we have the following result.

Lemma 54. *Suppose that* $\max_{s,a,\mathcal{L},s'} P(s'|s,a,\mathcal{L}) \le c_1$*, and that* $P(s'|s,a,\cdot)$ *is* c_2*-Lipschitz in* W_1*, i.e.,*

$$|P(s'|s,a,\mathcal{L}_1) - P(s'|s,a,\mathcal{L}_2)| \le c_2 W_1(\mathcal{L}_1,\mathcal{L}_2). \tag{C.3.1}$$

Then in Assumption 4, d_2 *and* d_3 *can be chosen as*

$$d_2 = \frac{2diam(\mathcal{S})diam(\mathcal{A})|\mathcal{S}|c_1}{d_{\min}(\mathcal{A})} \tag{C.3.2}$$

and $d_3 = \frac{diam(\mathcal{S})diam(\mathcal{A})c_2}{2}$*, respectively.*

Lemma 54 provides an explicit characterization of the population side assumptions based only on the boundedness and Lipschitz properties of the transition dynamics P. In particular, c_1 becomes smaller when the transition dynamics becomes more diverse and the state space becomes larger.

Proof. (Lemma 54) We begin by noticing that $\mathcal{L}' = \Gamma_2(\pi,\mathcal{L})$ can be expanded and computed as follows:

$$\mu'(s') = \sum\nolimits_{s\in\mathcal{S},a\in\mathcal{A}} \mu(s)P(s'|s,a,\mathcal{L})\pi(s,a), \quad \mathcal{L}'(s',a') = \mu'(s')\pi(s',a'), \tag{C.3.3}$$

where μ is the state marginal distribution of $\mathcal{L}$.

Now by the inequalities (C.1.3), we have

$$\begin{aligned}
&W_1(\Gamma_2(\pi_1,\mathcal{L}),\Gamma_2(\pi_2,\mathcal{L})) \le \mathrm{diam}(\mathcal{S}\times\mathcal{A}) d_{TV}(\Gamma_2(\pi_1,\mathcal{L}),\Gamma_2(\pi_2,\mathcal{L}))\\
=&\frac{\mathrm{diam}(\mathcal{S}\times\mathcal{A})}{2}\sum_{s'\in\mathcal{S},a'\in\mathcal{A}}\left|\sum_{s\in\mathcal{S},a\in\mathcal{A}}\mu(s)P(s'|s,a,\mathcal{L})\left(\pi_1(s,a)\pi_1(s',a')-\pi_2(s,a)\pi_2(s',a')\right)\right|\\
\le&\frac{\mathrm{diam}(\mathcal{S}\times\mathcal{A})}{2}\max_{s,a,\mathcal{L},s'}P(s'|s,a,\mathcal{L})\sum_{s,a,s',a'}\mu(s)(\pi_1(s,a)+\pi_2(s,a))|\pi_1(s',a')-\pi_2(s',a')|\\
\le&\frac{\mathrm{diam}(\mathcal{S}\times\mathcal{A})}{2}\max_{s,a,\mathcal{L},s'}P(s'|s,a,\mathcal{L})\sum_{s',a'}|\pi_1(s',a')-\pi_2(s',a')|\cdot(1+1)\\
=&2\mathrm{diam}(\mathcal{S}\times\mathcal{A})\max_{s,a,\mathcal{L},s'}P(s'|s,a,\mathcal{L})\sum_{s'}d_{TV}(\pi_1(s'),\pi_2(s'))\\
\le&\frac{2\mathrm{diam}(\mathcal{S}\times\mathcal{A})\max_{s,a,\mathcal{L},s'}P(s'|s,a,\mathcal{L})|\mathcal{S}|}{d_{\min}(\mathcal{A})}D(\pi_1,\pi_2)=\frac{2\mathrm{diam}(\mathcal{S})\mathrm{diam}(\mathcal{A})|\mathcal{S}|c_1}{d_{\min}(\mathcal{A})}D(\pi_1,\pi_2).
\end{aligned} \tag{C.3.4}$$

Similarly, we have

$$\begin{aligned}
&W_1(\Gamma_2(\pi,\mathcal{L}_1),\Gamma_2(\pi,\mathcal{L}_2)) \le \operatorname{diam}(\mathcal{S}\times\mathcal{A}) d_{TV}(\Gamma_2(\pi,\mathcal{L}_1),\Gamma_2(\pi,\mathcal{L}_2))\\
=&\frac{\operatorname{diam}(\mathcal{S}\times\mathcal{A})}{2}\sum_{s'\in\mathcal{S},a'\in\mathcal{A}}\left|\sum_{s\in\mathcal{S},a\in\mathcal{A}}\mu(s)\pi(s,a)\pi(s',a')\left(P(s'|s,a,\mathcal{L}_1)-P(s'|s,a,\mathcal{L}_2)\right)\right|\\
\le&\frac{\operatorname{diam}(\mathcal{S}\times\mathcal{A})}{2}\sum_{s,a,s',a'}\mu(s)\pi(s,a)\pi(s',a')\left|P(s'|s,a,\mathcal{L}_1)-P(s'|s,a,\mathcal{L}_2)\right|\\
\le&\frac{\operatorname{diam}(\mathcal{S})\operatorname{diam}(\mathcal{A})c_2}{2}.
\end{aligned} \tag{C.3.5}$$

This completes the proof. □

C.4 Proof of Theorems 44 and 53

For notational simplicity, we only present the proof for the stationary case (Theorem 53). The proof of Theorems 44 is the same with appropriate notational changes.

First by Definition 52 and the definitions of Γ_i $(i=1,2)$, $(\pi,\mathcal{L})$ is a stationary NE iff $\mathcal{L}=\Gamma(\mathcal{L})=\Gamma_2(\Gamma_1(\mathcal{L}),\mathcal{L})$ and $\pi=\Gamma_1(\mathcal{L})$, where $\Gamma(\mathcal{L})=\Gamma_2(\Gamma_1(\mathcal{L}),\mathcal{L})$. This indicates that for any $\mathcal{L}_1,\mathcal{L}_2\in\mathcal{P}(\mathcal{S}\times\mathcal{A})$,

$$\begin{aligned}
&W_1(\Gamma(\mathcal{L}_1),\Gamma(\mathcal{L}_2)) = W_1(\Gamma_2(\Gamma_1(\mathcal{L}_1),\mathcal{L}_1),\Gamma_2(\Gamma_1(\mathcal{L}_2),\mathcal{L}_2))\\
&\le W_1(\Gamma_2(\Gamma_1(\mathcal{L}_1),\mathcal{L}_1),\Gamma_2(\Gamma_1(\mathcal{L}_2),\mathcal{L}_1)) + W_1(\Gamma_2(\Gamma_1(\mathcal{L}_2),\mathcal{L}_1),\Gamma_2(\Gamma_1(\mathcal{L}_2),\mathcal{L}_2))\\
&\le (d_1d_2+d_3)W_1(\mathcal{L}_1,\mathcal{L}_2).
\end{aligned} \tag{C.4.1}$$

And since $d_1d_2+d_3\in[0,1)$, by the Banach fixed-point theorem, we conclude that there exists a unique fixed-point of Γ, or equivalently, a unique stationary MFG solution to (GMFG).

C.5 Proof of Theorem 45

The proof of Theorem 45 relies on the following lemmas.

Lemma 55 ([82]). *The softmax function is c-Lipschitz, i.e.,* $\|\boldsymbol{softmax}_c(x)-\boldsymbol{softmax}_c(y)\|_2\le c\|x-y\|_2$ *for any* $x,\ y\in\mathbb{R}^n$.

Notice that for a finite set $\mathcal{X}\subseteq\mathbb{R}^k$ and any two (discrete) distributions $\nu,\ \nu'$ over $\mathcal{X}$, we have

$$W_1(\nu,\nu')\le\operatorname{diam}(\mathcal{X})d_{TV}(\nu,\nu')=\frac{\operatorname{diam}(\mathcal{X})}{2}\|\nu-\nu'\|_1\le\frac{\operatorname{diam}(\mathcal{X})}{2}\|\nu-\nu'\|_2, \tag{C.5.1}$$

where in computing the ℓ_1-norm, $\nu,\ \nu'$ are viewed as vectors of length $|\mathcal{X}|$.

Hence Lemma 55 implies that for any $x,\ y\in\mathbb{R}^{|\mathcal{X}|}$, when $\mathbf{softmax}_c(x)$ and $\mathbf{softmax}_c(y)$ are viewed as probability distributions over $\mathcal{X}$, we have

$$W_1(\mathbf{softmax}_c(x),\mathbf{softmax}_c(y))\le\frac{\operatorname{diam}(\mathcal{X})c}{2}\|x-y\|_2\le\frac{\operatorname{diam}(\mathcal{X})\sqrt{|\mathcal{X}|}c}{2}\|x-y\|_\infty.$$

Lemma 56. *The distance between the softmax and the argmax mapping is bounded by*

$$\|\boldsymbol{softmax}_c(x) - \boldsymbol{argmax\text{-}e}(x)\|_2 \le 2n\exp(-c\delta),$$

where $\delta = x_{\max} - \max_{x_j < x_{\max}} x_j$, $x_{\max} = \max_{i=1,\dots,n} x_i$, *and* $\delta := \infty$ *when all* x_j *are equal.*

Similar to Lemma 55, Lemma 56 implies that for any $x \in \mathbb{R}^{|\mathcal{X}|}$, viewing $\mathbf{softmax}_c(x)$ as probability distributions over $\mathcal{X}$ leads to

$$W_1(\mathbf{softmax}_c(x), \mathbf{argmax\text{-}e}(x)) \le \mathrm{diam}(\mathcal{X})|\mathcal{X}|\exp(-c\delta).$$

Proof of Lemma 56. Without loss of generality, assume that $x_1 = x_2 = \cdots = x_m = \max_{i=1,\dots,n} x_i = x^\star > x_j$ for all $m < j \le n$. Then

$$\mathbf{argmax\text{-}e}(x)_i = \begin{cases} \frac{1}{m}, & i \le m, \\ 0, & otherwise. \end{cases}$$

$$\mathbf{softmax}_c(x)_i = \begin{cases} \frac{e^{cx^\star}}{me^{cx^\star} + \sum_{j=m+1}^n e^{cx_j}}, & i \le m, \\ \frac{e^{cx_i}}{me^{cx^\star} + \sum_{j=m+1}^n e^{cx_j}}, & otherwise. \end{cases}$$

Therefore

$$\begin{aligned} &\|\mathbf{softmax}_c(x) - \mathbf{argmax\text{-}e}(x)\|_2 \le \|\mathbf{softmax}_c(x) - \mathbf{argmax\text{-}e}(x)\|_1 \\ =& m\left(\frac{1}{m} - \frac{e^{cx^\star}}{me^{cx^\star} + \sum_{j=m+1}^n e^{cx_j}}\right) + \frac{\sum_{i=m+1}^n e^{cx_i}}{me^{cx^\star} + \sum_{j=m+1}^n e^{cx_j}} \\ =& \frac{2\sum_{i=m+1}^n e^{cx_i}}{me^{cx^\star} + \sum_{i=m+1}^n e^{cx_i}} = \frac{2\sum_{i=m+1}^n e^{-c\delta_i}}{m + \sum_{i=m+1}^n e^{-c\delta_i}} \\ \le& \frac{2}{m}\sum_{i=m+1}^n e^{-c\delta_i} \le \frac{2(n-m)}{m}e^{-c\delta} \le 2ne^{-c\delta}, \end{aligned}$$

with $\delta_i = x_i - x^\star$. □

Lemma 57 ([77])**.** *For an MDP, say* $\mathcal{M}$, *suppose that the Q-learning algorithm takes step-sizes*

$$\beta_t(s,a) = \begin{cases} |\#(s,a,t)+1|^{-h}, & (s,a) = (s_t, a_t), \\ 0, & otherwise. \end{cases}$$

with $h \in (1/2, 1)$. *Here* $\#(s,a,t)$ *is the number of times up to time* t *that one visits the state-action pair* (s,a). *Also suppose that the covering time of the state-action pairs is bounded by* L *with probability at least* $1-p$ *for some* $p \in (0,1)$. *Then* $\|Q_{T^{\mathcal{M}}(\delta,\epsilon)} - Q^\star\|_\infty \le \epsilon$ *with probability at least* $1-2\delta$. *Here* Q_T *is the* T*-th update in Q-learning, and* $Q^\star$ *is the (optimal) Q-function, given that*

$$T^{\mathcal{M}}(\delta,\epsilon) = \Omega\left(\left(\frac{L\log_p(\delta)}{\beta}\log\frac{V_{\max}}{\epsilon}\right)^{\frac{1}{1-h}} + \left(\frac{\left(L\log_p(\delta)\right)^{1+3h} V_{\max}^2 \log\left(\frac{|\mathcal{S}||\mathcal{A}|V_{\max}}{\delta\beta\epsilon}\right)}{\beta^2\epsilon^2}\right)^{\frac{1}{h}}\right),$$

where $\beta = (1-\gamma)/2$, $V_{\max} = R_{\max}/(1-\gamma)$, *and* $R_{\max}$ *is an upper bound on the extreme difference between the expected rewards, i.e.,* $\max_{s,a,\mu} r(s,a,\mu) - \min_{s,a,\mu} r(s,a,\mu) \le R_{\max}$.

Here the covering time L of a state-action pair sequence is defined to be the number of steps needed to visit all state-action pairs starting from any arbitrary state-action pair, and $T^{\mathcal{M}}(\delta, \epsilon)$ is the number of inner iterations T_k set in Algorithm 2. This will guarantee the convergence in Theorem 45. Also notice that the l_∞ norm above is defined in an element-wise sense, *i.e.*, for $M \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{A}|}$, we have $\|M\|_\infty = \max_{s\in\mathcal{S},a\in\mathcal{A}} |M(s,a)|$.

Proof of Theorem 45. Define $\hat{\Gamma}_1^k(\mathcal{L}) := \mathbf{softmax}_c\left(\hat{Q}^\star_{\mathcal{L}_k}\right)$. In the following, $\pi = \mathbf{softmax}_c(Q_{\mathcal{L}})$ is understood as the policy π with $\pi(s) = \mathbf{softmax}_c(Q_{\mathcal{L}}(s,\cdot))$. Let $\mathcal{L}^\star$ be the population state-action pair in a stationary NE of (GMFG). Then $\pi_k = \hat{\Gamma}_1^k(\mathcal{L}_k)$. Denoting $d := d_1 d_2 + d_3$, we see

$$\begin{aligned}
W_1(\tilde{\mathcal{L}}_{k+1}, \mathcal{L}^\star) =& W_1(\Gamma_2(\pi_k, \mathcal{L}_k), \Gamma_2(\Gamma_1(\mathcal{L}^\star), \mathcal{L}^\star)) \\
\leq& W_1(\Gamma_2(\Gamma_1(\mathcal{L}_k), \mathcal{L}_k), \Gamma_2(\Gamma_1(\mathcal{L}^\star), \mathcal{L}^\star)) + W_1(\Gamma_2(\Gamma_1(\mathcal{L}_k), \mathcal{L}_k), \Gamma_2(\hat{\Gamma}_1^k(\mathcal{L}_k), \mathcal{L}_k)) \\
\leq& W_1(\Gamma(\mathcal{L}_k), \Gamma(\mathcal{L}^\star)) + d_2 D(\Gamma_1(\mathcal{L}_k), \hat{\Gamma}_1^k(\mathcal{L}_k)) \\
\leq& (d_1 d_2 + d_3) W_1(\mathcal{L}_k, \mathcal{L}^\star) + d_2 D(\mathbf{argmax\text{-}e}(Q^\star_{\mathcal{L}_k}), \mathbf{softmax}_c(\hat{Q}^\star_{\mathcal{L}_k})) \\
\leq& d W_1(\mathcal{L}_k, \mathcal{L}^\star) + d_2 D(\mathbf{softmax}_c(\hat{Q}^\star_{\mathcal{L}_k}), \mathbf{softmax}_c(Q^\star_{\mathcal{L}_k})) \\
&+ d_2 D(\mathbf{argmax\text{-}e}(Q^\star_{\mathcal{L}_k}), \mathbf{softmax}_c(Q^\star_{\mathcal{L}_k})) \\
\leq& d W_1(\mathcal{L}_k, \mathcal{L}^\star) + \frac{c d_2 \mathrm{diam}(\mathcal{A})\sqrt{|\mathcal{A}|}}{2} \|\hat{Q}^\star_{\mu_k} - Q^\star_{\mu_k}\|_\infty \\
&+ d_2 D(\mathbf{argmax\text{-}e}(Q^\star_{\mathcal{L}_k}), \mathbf{softmax}_c(Q^\star_{\mathcal{L}_k})).
\end{aligned}$$

Then since $\mathcal{L}_k \in S_\epsilon$ by the projection step, Lemma 56, and Lemma 57 with the choice of $T_k = T^{\mathcal{M}_\mu}(\delta_k, \epsilon_k))$, we have, with probability at least $1 - 2\delta_k$,

$$W_1(\tilde{\mathcal{L}}_{k+1}, \mathcal{L}^\star) \leq d W_1(\mathcal{L}_k, \mathcal{L}^\star) + \frac{c d_2 \mathrm{diam}(\mathcal{A})\sqrt{|\mathcal{A}|}}{2}\epsilon_k + d_2 \mathrm{diam}(\mathcal{A})|\mathcal{A}| e^{-c\phi(\epsilon)}. \tag{C.5.2}$$

Finally, it is clear that with probability at least $1 - 2\delta_k$,

$$\begin{aligned}
W_1(\mathcal{L}_{k+1}, \mathcal{L}^\star) &\leq W_1(\tilde{\mathcal{L}}_{k+1}, \mathcal{L}^\star) + W_1(\tilde{\mathcal{L}}_{k+1}, \mathbf{Proj}_{S_\epsilon}(\tilde{\mathcal{L}}_{k+1})) \\
&\leq d W_1(\mathcal{L}_k, \mathcal{L}^\star) + \frac{c d_2 \mathrm{diam}(\mathcal{A})\sqrt{|\mathcal{A}|}}{2}\epsilon_k + d_2 \mathrm{diam}(\mathcal{A})|\mathcal{A}| e^{-c\phi(\epsilon)} + \epsilon.
\end{aligned}$$

By telescoping, this implies that with probability at least $1 - 2\sum_{k=0}^{K-1}\delta_k$,

$$\begin{aligned}
W_1(\mathcal{L}_K, \mathcal{L}^\star) \leq& d^K W_1(\mathcal{L}_0, \mathcal{L}^\star) + \frac{c d_2 \mathrm{diam}(\mathcal{A})\sqrt{|\mathcal{A}|}}{2} \sum_{k=0}^{K-1} d^{K-k}\epsilon_k \\
&+ \frac{(d_2 \mathrm{diam}(\mathcal{A})|\mathcal{A}| e^{-c\phi(\epsilon)} + \epsilon)(1 - d^K)}{1 - d}.
\end{aligned} \tag{C.5.3}$$

Since ϵ_k is summable, hence $\sup_{k\geq 0}\epsilon_k < \infty$), $\sum_{k=0}^{K-1} d^{K-k}\epsilon_k \leq \frac{\sup_{k\geq 0}\epsilon_k}{1-d} d^{\lfloor (K-1)/2 \rfloor} + \sum_{k=\lceil (K-1)/2 \rceil}^{\infty}\epsilon_k$.

Now plugging in $K = K_{\epsilon,\eta}$, with the choice of δ_k and $c = \frac{\log(1/\epsilon)}{\phi(\epsilon)}$, and noticing that $d \in [0, 1)$, it is clear that with probability at least $1 - 2\delta$,

$$\begin{aligned} W_1(\mathcal{L}_{K_{\epsilon,\eta}}, \mathcal{L}^\star) \leq & d^{K_{\epsilon,\eta}} W_1(\mathcal{L}_0, \mathcal{L}^\star) \\ & + \frac{c d_2 \text{diam}(\mathcal{A})\sqrt{|\mathcal{A}|}}{2} \left(\frac{\sup_{k\geq 0} \epsilon_k}{1-d} d^{\lfloor (K_{\epsilon,\eta}-1)/2 \rfloor} + \sum_{k=\lceil (K_{\epsilon,\eta}-1)/2 \rceil}^{\infty} \epsilon_k \right) \\ & + \frac{(d_2 \text{diam}(\mathcal{A})|\mathcal{A}| + 1)\epsilon}{1-d}. \end{aligned} \tag{C.5.4}$$

Setting $\epsilon_k = (k+1)^{-(1+\eta)}$, then when $K_{\epsilon,\eta} \geq 2(\log_d \epsilon + 1)$,

$$\frac{\sup_{k\geq 0} \epsilon_k}{1-d} d^{\lfloor (K_{\epsilon,\eta}-1)/2 \rfloor} \leq \frac{\epsilon}{1-d}.$$

Similarly, when $K_{\epsilon,\eta} \geq 2(\eta\epsilon)^{-1/\eta}$, $\sum_{k=\lceil \frac{K_{\epsilon,\eta}-1}{2} \rceil}^{\infty} \epsilon_k \leq \epsilon$.

Finally, when $K_{\epsilon,\eta} \geq \log_d(\epsilon/(\text{diam}(\mathcal{S})\text{diam}(\mathcal{A})))$, $d^{K_{\epsilon,\eta}} W_1(\mathcal{L}_0, \mathcal{L}^\star) \leq \epsilon$, since $W_1(\mathcal{L}_0, \mathcal{L}^\star) \leq \text{diam}(\mathcal{S} \times \mathcal{A}) = \text{diam}(\mathcal{S})\text{diam}(\mathcal{A})$.

In summary, if $K_{\epsilon,\eta} = \lceil 2 \max\{(\eta\epsilon)^{-1/\eta}, \log_d(\epsilon/\max\{\text{diam}(\mathcal{S})\text{diam}(\mathcal{A}), 1\}) + 1\} \rceil$, then with probability at least $1 - 2\delta$,

$$W_1(\mathcal{L}_{K_{\epsilon,\eta}}, \mathcal{L}^\star) \leq \left(1 + \frac{c d_2 \text{diam}(\mathcal{A})\sqrt{|\mathcal{A}|}(2-d)}{2(1-d)} + \frac{(d_2 \text{diam}(\mathcal{A})|\mathcal{A}| + 1)}{1-d} \right) \epsilon = O(\epsilon).$$

Finally, plugging in ϵ_k and δ_k into $T^{\mathcal{M}_L}(\delta_k, \epsilon_k)$, and noticing that $k \geq K_{\epsilon,\eta}$ and $\sum_{k=0}^{K_{\epsilon,\eta}-1} (k+1)^\alpha \leq \frac{K_{\epsilon,\eta}^{\alpha+1}}{\alpha+1}$, it is immediate that

$$T = O\left((\log(K_{\epsilon,\eta}/\delta))^{\frac{1}{1-h}} K_{\epsilon,\eta} (\log K_{\epsilon,\eta})^{\frac{1}{1-h}} + (\log(K_{\epsilon,\eta}/\delta))^{\frac{1}{h}+3} \frac{K_{\epsilon,\eta}^{1+\frac{2(1+\eta)}{h}}}{1+\frac{2(1+\eta)}{h}} (\log(K_{\epsilon,\eta}/\delta))^{\frac{1}{h}} \right).$$

By further relaxing η to 1 and merging the terms, (5.4.3) follows. □

C.6 Naive Algorithm

The Naive iterative algorithm (Algorithm 3) is to replace Step A in the three-step fixed-point approach of GMFGs with Q-learning iterations. The limitation of this Naive algorithm has been discussed in the main text (Step 1, Section 5.4) and empirically verified in Section 5.5 (Figure 5.4).

C.7 GMF-V

GMF-V, briefly mentioned in Section 5.4, is the value-iteration version of our main algorithm GMF-Q. GMF-V applies to the GMFG setting with fully known transition dynamics P and rewards r.

Algorithm 3 Alternating Q-learning for GMFGs (Naive)

Input: Initial population state-action pair L_0
for $k = 0, 1, \cdots$ **do**
    Perform Q-learning to find the Q-function $Q_k^\star(s,a) = Q_{L_k}^\star(s,a)$ of an MDP with dynamics $P_{L_k}(s'|s,a)$ and rewards $r_{L_k}(s,a)$.
    Solve $\pi_k \in \Pi$ with $\pi_k(s) = \textbf{argmax-e}\,(Q_k^\star(s,\cdot))$.
    Sample $s \sim \mu_k$, where μ_k is the population state marginal of L_k, and obtain L_{k+1} from $\mathcal{G}(s, \pi_k, L_k)$.

Algorithm 4 Value Iteration for GMFGs (GMF-V)

Input: Initial L_0, tolerance $\epsilon > 0$.
for $k = 0, 1, \cdots$ **do**
    Perform value iteration for T_k iterations to find the approximate Q-function Q_{L_k} and value function V_{L_k}:
    for $t = 1, 2, \cdots, T_k$ **do**
        for all $s \in \mathcal{S}$ and $s \in \mathcal{A}$ **do**
            $Q_{L_k}(s,a) \leftarrow \mathbb{E}[r(s,a,L_k)] + \gamma \sum_{s'} P(s'|s,a,L_k) V_{L_k}(s')$
            $V_{L_k}(s) \leftarrow \max_a Q_{L_k}(s,a)$
    Compute a policy $\pi_k \in \Pi$:
      $\pi_k(s) = \textbf{softmax}_c(Q_{L_k}(s,\cdot))$.
    Sample $s \sim \mu_k$, where μ_k is the population state marginal of L_k, and obtain $\tilde{L}_{k+1}$ from $\mathcal{G}(s, \pi_k, L_k)$.
    Find $L_{k+1} = \textbf{Proj}_{S_\epsilon}(\tilde{L}_{k+1})$

C.8 More Details for the Experiments

Competition Intensity Index M.

In the experiment, the competition index M is interpreted and implemented as the number of selected players in each auction competition. That is, in each round, $M-1$ players will be randomly selected from the population to compete with the *representative* advertiser for the auction. Therefore, the population distribution $\mathcal{L}_t$, the winner indicator w_t^M, and second-best price a_t^M all depend on M. This parameter M is also referred to as the *auction thickness* in the auction literature [112].

Adjustment for Algorithm MF-Q.

For MF-Q, [188] assumes all N players have a joint state s. In the auction experiment, we make the following adjustment for MF-Q for computational efficiency and model comparability: each player i makes decision based on her own private state and table Q^i is a functional of s^i, a^i and $\frac{\sum_{j\neq i} a^j}{N-1}$.