Topics in the mean-field type approach to pedestrian crowd modeling and conventions

ALEXANDER AURELL

Doctoral Thesis
KTH Royal Institute of Technology
School of Engineering Sciences
Department of Mathematics
Division of Mathematical Statistics
Stockholm, Sweden 2019

Department of Mathematics
KTH Royal Institute of Technology
100 44 Stockholm, Sweden

TRITA-SCI-FOU 2019;47
ISBN 978-91-7873-350-7

Akademisk avhandling som med tillstånd av Kungliga Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen, måndagen den 16 december 2019 klockan 10.00 i Kollegiesalen, Brinellvägen 8, Kungliga Tekniska högskolan, Stockholm.

Print: Universitetsservice US AB, Stockholm, 2019

Till Orkestermaskinen

Abstract

This thesis consists of five appended papers, primarily addressing topics in pedestrian crowd modeling and the formation of conventions.

The first paper generalizes a pedestrian crowd model for competing subcrowds to include nonlocal interactions and an arbitrary (but finite) number of subcrowds. Each pedestrian is granted a 'personal space' and is effected by the presence of other pedestrians within it. The interaction strength may depend on subcrowd affinity. The paper investigates the mean-field type game between subcrowds and derives conditions for the reduction of the game to an optimization problem.

The second paper suggest a model for pedestrians with a predetermined target they have to reach. The fixed and non-negotiable final target leads us to formulate a model with backward stochastic differential equations of mean-field type. Equilibrium in the game between the tagged pedestrians and a surrounding crowd is characterized with the stochastic maximum principle. The model is illustrated by a number of numerical examples.

The third paper introduces sticky reflected stochastic differential equations with boundary diffusion as a means to include walls and obstacles in the mean-field approach to pedestrian crowd modeling. The proposed dynamics allow the pedestrians to move and interact while spending time on the boundary. The model only admits a weak solution, leading to the formulation of a weak optimal control problem.

The fourth paper treats two-player finite-horizon mean-field type games between players whose state trajectories are given by backward stochastic differential equations of mean-field type. The paper validates the stochastic maximum principle for such games. Numerical experiments illustrate equilibrium behavior and the price of anarchy.

The fifth paper treats the formation of conventions in a large population of agents that repeatedly play a finite two-player game. The players access a history of previously used action profiles and form beliefs on how the opposing player will act. A dynamical model where more recent interactions are considered to be more important in the belief-forming process is proposed. Convergence of the history to a collection of minimal CURB blocks and, for a certain class of games, to Nash equilibria is proven.

Keywords: pedestrian crowds, stochastic differential equations, mean field, stochastic control, games, backward dynamics, sticky boundary, stochastic maximum principle, social conventions

Sammanfattning

Den här avhandlingen består av fem artiklar som behandlar några utvalda problem inom matematisk modellering av folkmassors rörelse och uppkomsten av konventioner.

Den första artikeln generaliserar en modell för växelverkan mellan grupper av fotgängare. Varje fotgängare (agent) ges ett 'personligt utrymme' och påverkas av andra agenter som befinner sig i dess utrymme. I artikeln analyseras situationen som ett matematiskt spel av medelfältstyp och villkor för när spelet kan reduceras till ett optimeringsproblem härleds.

I den andra artikeln modelleras fotgängare med ett mål som de är tvungna att nå efter en bestämd (ändlig) tid. Detta ej förhandlingsbara mål leder oss till stokastiska differentialekvationer med ändvillkor. Med den stokastiska maximumprincipen härleds nödvändiga villkor för jämvikt i ett matematisk spel där fotgängarna och en omgivande folkmassa växelverkar i tävlan om den bästa färdvägen. Modellen illustreras med flera numeriska exempel.

I den tredje artikeln introducerar vi reflekterande stokastiska differentialekvationer med limaktiga randvillkor och randdiffusion som ett verktyg för att modellera hur fotgängaren påverkas av väggar och andra fasta hinder. Den föreslagna dynamiska modellen tillåter fotgängarna att spendera tid vid väggar och då också växelverka med omgivningen. Ekvationerna kan endast lösas i en svag mening och därför formuleras modellen som ett styrproblem för fotgängarnas statistiska fördelning.

Artikel fyra behandlar ett spel av medelfältstyp med två spelare vars tillstånd beskrivs av ett system av stokastiska differentialekvationer med ändvillkor. Med den stokastiska maximumprincipen härleds nödvändiga villkor för spelets jämvikt och en numerisk simulering visar på skillnaden i utfall mellan konkurrens och samarbete, alltså mellan spelet och en relaterad styrmodell.

Den femte artikeln handlar om uppkomsten av konventioner i en stor population av agenter som upprepade gånger spelar ett ändligt spel med två roller. När agenterna ska välja strategi har de en historik av tidigare spelade strategier till hjälp. Artikeln introducerar en speldynamik där den senare historiken antas vara viktigare än den tidigare. Vi bevisar konvergens av historiken till strategier i minimala CURB block och, för en specifik klass av spel, till Nashjämvikter.

Nyckelord: folkmassor, stokastiska differentialekvationer, medelfält, stokastisk styrning, limaktiga randvillkor, stokastiska maximumprincipen, dynamik med ändvillkor, spel, konventioner

Acknowledgements

I would like to extend my deepest gratitude to my supervisor Professor Boualem Djehiche. Boualem, you have taken me on an epic mathematical journey. When I now look back and summarize, we have ventured further than I ever thought was possible at the start. You have always tried too steer me towards goals within my reach and you have without exception given me helpful advice at times when things were not going well. Thank you for all the generous support and much needed encouragement.

I would like to thank all the members of my two homes at the department, the division of mathematical statistics and the division of optimization and systems theory, for sharing your knowledge and experience with me in discussions, seminars, and courses. I would also like to thank Robert Thorén at Algorithmica Research AB for supporting me in my teaching duties and for being my window to the world of finance outside academia.

Many thanks to all my fellow PhD students. There is a friendly and open atmosphere among us which I have enjoyed. I hope that it will forever be so that stories of both the horror and the joy of the PhD life are shared and discussed in the corridors. Johan, Felix, and Björn, I will always regard you both as my mentors and my friends. Salah, my travel companion, we have both reached the other end now. I wish you only the best in your next adventure! Göran, you are really passionate about what you do and I hope that we will one day do interesting work together again.

I wish to thank Professor Hamidou Tembine and Mathieu Laurière for inspiring me to work hard and for teaching me about mean-field type games, control theory, and much, much more. Thank you, Gustav Karreskog, for enriching my days with your knowledge about game theory and economics. Thank you, Associate Professor Fredrik Viklund, for giving me trust and responsibility.

Jag vill tacka min familj för att ni så länge jag kan minnas gett mig ert ovillkorliga stöd. Jag vill också tacka alla mina fina vänner för stunderna vi delat under de senaste fem åren. Ni är en mycket viktig del av mitt liv. Tack Helena, för att du ger mig det största stödet. Tack för din fasta tro på jag skulle lyckas slutföra det här arbetet. Du har varit mitt ankare och mina vingar.

Alexander Aurell
Stockholm, December 2019

Table of Contents

Mądry Polak po szkodzie

– Polish proverb

Part I: Introduction

1

Introductory overview

This thesis treats subjects related to the mathematical modeling of crowds and the individual's strategic choice of action. Specifically, it deals with a number of questions and issues linked with the mean-field approach to pedestrian crowd modeling, and also some questions regarding the stability of conventions in normal form games.

In the mean-field approach to pedestrian crowd modeling, a crowd of N pedestrians is approximated with its statistical limit as $N \to \infty$. The phrase *mean field* indicates the affinity with the mean-field theory in physics, which treats systems of large numbers of particles, each with a negligible impact upon the system as a whole. The approximation drastically reduces the dimension of the system of equations that characterizes equilibrium, or optimal, behavior. But there is a trade-off, the mean-field equations are generically nonlinear.

Depending on whether the participants in the crowd are competing or cooperating, the approach leads to two different, but closely related, mathematical problems. On the one hand, mean-field games approximate a very large crowd of competing, interacting pedestrians, each with negligible impact on the behavior of the crowd as a whole. On the other hand, a mean-field type control problem approximates a cooperating crowd whose participants follow the decision of a central planner. The central planner can significantly impact the crowd as a whole. Mean-field type game theory is the multi-central planner generalization of the single central planner mean-field type control.

Ultimately, a model for the pedestrian crowd must be consistent with empirically established facts about pedestrian motion and be able to replicate experimental results. While the interaction between pedestrians is a significant component in pedestrian crowd dynamics, it cannot by itself explain all observed features of a crowd. To name a few other components, individual pedestrians have physical speed constraints, preferred speeds and directions, they conserve energy but are also impatient. In a crowd, the pedestrian motion is constrained by walls, obstacles, and by the crowd density itself.

In the mean-field approach to crowd modeling it is common practice to work under the assumption that aggregate interactions will lead to a crowd which is in Nash equilibrium. The question of whether rational agents will settle in some form of equilibrium does not have a straight-forward answer. Some insight can be gained by analyzing repeated play of games where, turn by turn, each agent updates its belief about the other participants in the game and acts accordingly. After a sufficiently large number of turns, the agents may settle to use only strategies from some subset (possibly a Nash equilibrium) of their admissible strategies and a convention is formed.

The thesis consists of two parts. This part continuous with a chapter that contains background material followed by a chapter that summarizes the appended research papers and also clarifies my contribution to each of them. The scientific contribution is found in the second part, which contains five appended papers. The first three papers deal with questions in the mean-field approach to pedestrian crowd modeling, the fourth paper with the computation of equilibria in the mean-field approach, and the fifth paper treats the formation of conventions in finite normal-form games.

2

Background

This section aims to give both an overview of the modeling approaches to the motion of pedestrian crowds, as well as presenting a background to the mathematical models used in the five papers that comprise the second part of this thesis.

First, normal-form games are defined and the concept of Nash equilibrium is introduced. Then, the theory of stochastic optimal control is briefly summarized. The focus is the type of control problems that are used to model pedestrian crowds in the appended research papers. The idea of a game and a control problem is then, loosely speaking, combined and stochastic differential games are introduced. From there, the step is small to mean-field type control problems and mean-field type games. Finally, this section treats experiments on and contemporary models for pedestrian crowd motion. A special focus is given to the mean-field approach.

The mathematical problems and models presented below have plenty of variations, each one being a field of research on its own. The particular variants introduced here are those of highest relevance for the appended research papers.

2.1 Games and optimal control

Static games

The theory of games dates back to the first half of the 20^{th} century, when it was developed to be a mean for the modeling of conflicting interests and interaction. Economic processes, multinational environmental efforts, and ecological equilibria are examples thereof. Pedestrians competing for the shortest path in a crowded environment, e.g. a train station at rush hours, can also be modeled as a game. The early research culminated with the classical book of von Neumann and Morgenstern [71], which set the foundations of game theory.

The basic entity of a game theoretic model is the **player**, representing an individual, a group, or any other decision making unit. The basic assumptions that underlie the theory are that players pursue well-defined objectives (they are **rational**) and take into account their knowledge or expectations of other players' behavior (they **reason strategically**). Normal-form games, or strategic games, are games in which each player makes a decision once and for all, and the players make their decisions simultaneously and independently. Most of the content presented in the rest of this section can be found in [74].

Definition 1. *A* ***normal-form game*** *consists of*

- *a finite set* $\mathcal{N} := \{1, \ldots, N\}$ *(the set of* ***players****);*
- *for each player* $i \in \mathcal{N}$ *a nonempty set* U_i *(the set of* ***actions (strategies, controls)*** *available to player* i*);*
- *for each player* $i \in \mathcal{N}$ *a preference relation on* $\mathbb{U} := \times_{j\in\mathcal{N}} U_j$ *(the* ***preference relation*** *of player* i*).*

If U_i is a finite set for all $i \in \mathcal{N}$ we call the game **finite**. Generally, a preference relation is a complete reflexive transitive binary relation. If the game is finite, the preference structure of player i can then be replaced by a function $J_i : \mathbb{U} \to \mathbb{R}$ such that player i prefers $u \in \mathbb{U}$ over $v \in \mathbb{U}$ if $J_i(u) < J_i(v)$ and is indifferent if $J_i(u) = J_i(v)$. This is the only kind of preference relation considered in this thesis. The convention will always be that the players' preference is towards lower values of J_i, i.e. they minimize a cost. A normal-form game with preference structure of the type described here is denoted by $\langle \mathcal{N}, (U_i)_{i\in\mathcal{N}}, (J_i)_{i\in\mathcal{N}} \rangle$.

There are plenty of concepts that can serve as solutions to normal-form games. Which one that makes most sense is highly game and context dependent. The most commonly used notion is that of a Nash equilibrium, formalized by Nash [69].

Definition 2. *An N-tuple $\hat{u} := (\hat{u}_1, \dots \hat{u}_N) \in \mathbb{U}$ is called a* ***Nash equilibrium*** *of the strategic game $\langle \mathcal{N}, (U_i)_{i\in\mathcal{N}}, (J_i)_{i\in\mathcal{N}} \rangle$ if*

$$J_i(\hat{u}) \leq J_i(\hat{u}_1, \dots, \hat{u}_{i-1}, u, \hat{u}_{i+1}, \dots, \hat{u}_N), \quad \forall u \in U_i, \forall i \in \mathcal{N}.$$

In words, Definition 2 says that no player can gain from a unilateral deviation to a different strategy. The equilibrium of Definition 2 is often referred to as a **pure** Nash equilibrium for the game $\langle \mathcal{N}, (U_i)_{i\in\mathcal{N}}, (J_i)_{i\in\mathcal{N}} \rangle$. In a generic game, a pure Nash equilibrium does not necessarily exist, and when it does, it might not be unique. Two pure Nash equilibria can yield different values of J_i, a famous example of this situation is the so-called coordination game. Also, playing Nash equilibrium strategies can lead to the worst societal outcome (see the route choice example in [29, Ch. 8]). Solution concepts other than the Nash equilibrium include proper equilibrium, perfect equilibrium, and many more. See for example [74] for an exposition.

Example 3 (Prisoner's dilemma)**.** *To the classical 'Prisoner's dilemma' game there is a unique Nash equilibrium. The preference structure for a generic Prisoner's dilemma is given by payoff matrix below with $b > a > d > c$.*

		Player Y Silent	Player Y Betray
Player X	Silent	(a, a)	(c, b)
Player X	Betray	(b, c)	(d, d)

Table 1: Payoff matrix for Prisoner's dilemma.

The unique pure Nash equilibrium is $(Betray, Betray)$, which individually yields a lower payoff than $(Silent, Silent)$.

Example 4 (Matching pennies)**.** *The 'Matching pennies' game, with payoff matrix defined below, does not have a pure Nash equilibrium in the sense of Definition 2.*

The fact that not all normal-form games have a Nash equilibrium strategy (cf. Example 4) can be resolved by allowing the players to use mixed strategies. A **mixed strategy** is a probability distribution over the action

		Disagreeing player	
		Heads	Tails
Agreeing player	Heads	$(1,-1)$	$(-1,1)$
	Tails	$(-1,1)$	$(1,1)$

Table 2: Payoff matrix for Matching pennies

set. Denote by $\mathcal{P}(U_i)$ the set of probability measures over U_i. Since the players' choices of action are independent, the cost function extends to take a mixed strategy profile $\mu = \otimes_{i\in\mathcal{N}}(\mu_i)$ where $\mu_i \in \mathcal{P}(U_i)$,

$$\mathcal{J}_i(\mu) := \int_{\mathbb{U}} J_i(u)d\mu(u).$$

Definition 5. *The **mixed-strategy extension** of $\langle\mathcal{N}, (U_i), (J_i)\rangle$ is the game $\langle\mathcal{N}, (\mathcal{P}(U_i)), (\mathcal{J}_i)\rangle$. A **mixed Nash equilibrium in a normal-form game** is a Nash equilibrium in its mixed-strategy extension.*

Nash famously proved that every finite normal-form game has a mixed Nash equilibrium [69, 70]. In his Ph.D. thesis [68], John Nash provided two informal interpretations of what his equilibrium concept represents. In the rationalistic interpretation the players have never interacted before and will never interact again, but share common knowledge of everyones rationality and of the game. How will they play? One possibility is a Nash equilibrium strategy, which is compatible with this setup. However, it is now known that there are more strategies than the Nash equilibria which are compatible with common knowledge of rationality and of the game, see [75] for an example. Hence, common knowledge of rationality and the game does not imply the Nash equilibrium play in general. Nonetheless, it does in certain subclasses of normal-form games, such as dominance solvable games (of which the Prisoner's dilemma is an example). Nash called his second interpretation the mass action interpretation. For each role in the game, he considered a population of individuals without complete knowledge, from which a player is selected at random to play the game. The game is played over and over again, and over time, if the players *"accumulate empirical information on the relative advantages of the various pure strategies at their disposal, [...] this 'mass action' interpretation led to the conclusion that the mixed strategies representing the average behavior in each of the populations form an* *equilibrium point"* [68]. Stability and convergence in a certain model of such repeated plays is the topic of the fifth paper of this thesis.

Example 6 (Matching pennies)**.** *The unique mixed Nash equilibrium of the Matching pennies game of Example 4 is when both players uniformly randomize their two strategies.*

A mixed-strategy extension of a finite game is not finite since the set of mixed actions is a subset of the Euclidean space. Games can also be infinite in the sense that the number of players is infinite; this is discussed in Section 2.2, below. As a final remark, note that the Nash equilibrium in a game with only one player is the optimizer of that player's preference, i.e. the game reduces to an optimization problem.

Stochastic optimal control theory

Optimal control theory essentially studies a single player's state trajectory, which evolves over time according to a dynamic constraints involving a control process. The control process is the player's action. This is no longer a game, but a (functional) minimization problem. The theory tries to answer two questions:

- Existence of a minimum/maximum of a performance functional;
- Explicit computation of such a minimum/maximum.

Results on the existence of deterministic optimal controls can be traced back to the work [78] and makes use of the nowadays so-called Roxin condition and a measurable selection theorem. This path was perused for stochastic optimal control in [56, 33], to name a few. Other paths to existence results include approaches based on dynamic programming together with a verification theorem [23], the Girsanov transformation [28], and the martingale representation theorem [54]. Relaxed controls, a concept similar to that of mixed equilibria in games, was introduced in [87] as a compactification device and has been applied to prove existence of optimal stochastic controls [55, 30].

The computation of an optimal control will be addressed later in this section. First, the strong and the weak formulation of the stochastic optimal control problem will be introduced, as well as stochastic differential games.

The strong formulation of stochastic optimal control

Fix a time horizon $T \in (0, \infty)$ and a separable metric space U. Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a filtered probability space satisfying the usual conditions (cf. [86, Def. 2.6]), carrying an m-dimensional standard Brownian motion W and an $\mathbb{R}^n$-valued random variable x, independent of W. Let

$$\mathcal{U} := \{u : [0,T] \times \Omega \to U \mid u \text{ is } \mathbb{F}\text{-adapted}\}.$$

The dynamics of a controlled stochastic system is given by the following stochastic differential equation (SDE)

$$\begin{cases} dX_t = b(t, X_t, u_t)dt + \sigma(t, X_t, u_t)dW_t \\ X_0 = x, \end{cases} \tag{1}$$

where $b : [0,T] \times \mathbb{R}^n \times U \to \mathbb{R}^n$, $\sigma : [0,T] \times \mathbb{R}^n \times U \to \mathbb{R}^{n \times m}$. In the above, X is called the **state trajectory**, x is called the **initial position**, and $u \in \mathcal{U}$ is called the **control** or the action. Consider now a player with state trajectory (1) that takes an action from $\mathcal{U}$ which minimizes the following **cost** functional

$$J(u) := \mathbb{E}\left[\int_0^T f(t, X_t, u_t)dt + g(X_T)\right], \tag{2}$$

where $f : [0,T] \times \mathbb{R}^n \times U \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}$ and the expected value is taken under $\mathbb{P}$. The first and second terms on the right hand side of (2) are referred to as the **running cost** and the **terminal cost**, respectively.

Depending on the choice of coefficient functions it might be so that (1) does not admit a solution (in the strong sense) for all $u \in \mathcal{U}$. Here, one faces a trade-off between assuming regularity on b, σ and restricting the set of controls $\mathcal{U}$. To keep the presentation general we let $\mathcal{U}_{\text{ad}}$ be a subset of $\mathcal{U}$, not necessarily strict, which we call the set of **admissible controls**, i.e. for which

(i) b and σ are such that there exists a unique strong solution of the stochastic system (1) (cf. [86, Def. 6.15]);

(ii) $f(\cdot, X_\cdot, u_\cdot) \in L^1_{\mathbb{F}}([0,T]; \mathbb{R})$ and $g(X_T) \in L^1_{\mathcal{F}_T}(\Omega; \mathbb{R})$, The second assumption makes the cost J well-defined. We may now formulate the **strong stochastic control problem** as follows:

$$(OC)\quad \begin{cases} \textit{Minimize the cost } J \textit{ over the set of admissible controls } \mathcal{U}_{\text{ad}}, \\ \qquad\qquad\qquad\qquad \textit{subject to } (1) \textit{ in the strong sense.} \end{cases}$$

Example 7 (Linear quadratic)**.** *Let ξ be $\mathcal{F}_0$-measurable, real-valued, square-integrable, and independent of W (the Brownian motion introduced above). The linear quadratic version of (OC) has the following structure. The single agent has a state trajectory that satisfies the linear SDE*

$$dX_t = (AX_t + Bu_t)\,dt + \sigma dW_t, \quad X_0 = \xi,$$

and the agent minimizes the quadratic cost

$$J(u) = \mathbb{E}\left[\int_0^T \left(QX_t^2 + Ru_t^2 + 2NX_tu_t\right)dt + FX_T^2\right].$$

Above, Q, R, N, A, B, σ are deterministic constants. If U is unbounded, the natural restriction of $\mathcal{U}$ to admissible controls is

$$\mathcal{U}_{ad} = \left\{u \in \mathcal{U} \mid \mathbb{E}\Big[\int_0^T |u_t|^2 dt\Big] < \infty\right\},$$

since then $\mathbb{E}[|X_t|^2] < \infty$ for all $t \in [0, T]$ and the cost is well defined. If U is bounded, no restriction is needed. If the cost was not quadratic, but of power $\alpha < 2$, the admissibility could be relaxed accordingly. For a power $\alpha > 2$, the problem would be ill-defined unless the assumptions on ξ were strengthened to $\xi \in L^{\alpha}_{\mathcal{F}_0}(\Omega; \mathbb{R})$ (cf. [86, Thm. 6.16]).

Remark 8. *To mention a few of the variations of (OC), there is the random duration control problem, the risk-sensitive problem, partially observable systems, and singular and impulse controls. Further variations include time varying control- and state-constraints and an infinite-time horizon. Variations will in general change the set of admissible controls.*

The weak formulation of stochastic optimal control

The weak formulation of the stochastic optimal control problem consist of varying the probability space and consider it as part of the player's action.

This is reasonable since the player's preference relation (J) depends solely on the distribution of the state trajectory. It might also be so that the prescribed stochastic system fails to have a strong solution, but allows for a weak solution. The probability space then naturally varies with the control, since each control will give rise to a weak solution of the stochastic system.

The approach can be summarized as follows. Take $\Omega := C([0,T];\mathbb{R}^n)$ to be the space of continuous function from $[0,T]$ to $\mathbb{R}^n$. Given $t \in [0,T]$ and $\omega \in \Omega$ let X be the coordinate process $X_t(\omega) := \omega(t)$. Denote by $\mathbb{F}$ the filtration generated by X completed with the $\mathbb{P}$-null sets of Ω. The set of admissible control $\mathcal{U}_{\mathrm{ad}}$ is such that that for each $u \in \mathcal{U}_{\mathrm{ad}}$, there exists a unique weak solution to (1). Note that the set of admissible controls in the weak formulation is not necessarily equal to its counterpart in the strong formulation. Solving (1) weakly for a fixed $u \in \mathcal{U}_{\mathrm{ad}}$ means finding a probability measure $\mathbb{P}^u$ under which the coordinate process is a solution to the stochastic system (1). Hence, this approach induces a family of probability measures $(\mathbb{P}^u, u \in \mathcal{U})$ which is controlled by the player.

When the control does not appear in the diffusion coefficient of (1), the family can be constructed with the Girsanov transformation. In the case of controlled diffusion the control problem is in fact a problem about robustness. All probability measures in the induced family are singular with $\mathbb{P}$ and with each other (for different controls). The so-called second order backward SDE framework has been developed in [79] for the case of controlled diffusion.

Denoting by $\mathbb{E}^u$ expectation with respect to $\mathbb{P}^u$, the cost in the weak formulation is

$$J_W(u) = \mathbb{E}^u\left[\int_0^T f(t, X_t, u_t)dt + g(X_T)\right],$$

and the **weak stochastic optimal control problem** can be stated as follows:

$$(WOC) \quad \begin{cases} \textit{Minimize the cost } J_W \textit{ over the set of admissible controls } \mathcal{U}_{\mathrm{ad}}, \\ \hfill \textit{subject to } (1) \textit{ in the weak sense.} \end{cases}$$

Example 9 (Linear quadratic)**.** *To solve the linear quadratic problem of the previous section in the weak sense, consider the probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *described in this section,* $\mathbb{P}$ *being the probability measure on* $C([0,T];\mathbb{R})$ *under which the coordinate process solves*

$$dX_t = AX_tdt + \sigma dW_t,$$

where W is a $\mathbb{P}$-Wiener process and $\sigma > 0$. Consider the likelihood process

$$dL_t^u = L_t^u B u_t \sigma^{-1} dW_t, \quad L_0^u = 1.$$

The Girsanov transformation $d\mathbb{P}^u = L_T^u d\mathbb{P}$ defines a new probability measure $\mathbb{P}^u$ if L^u is a true martingale, more on this below. Under $\mathbb{P}^u$, the coordinate process satisfies

$$dX_t = (AX_t + Bu_t)dt + \sigma dW_t^u,$$

where W^u is a Wiener process under $\mathbb{P}^u$. An agent with the coordinate process as state trajectory minimizes the cost

$$J_W(u) = \mathbb{E}^u \left[\int_0^T \left(QX_t^2 + Ru_t^2 + 2NX_t u_t \right) dt + FX_T^2 \right]$$

where $\mathbb{E}^u$ denotes expected value under $\mathbb{P}^u$. A key observation is that J_W can be rewritten as an expected value under $\mathbb{E}$; since L^u is a martingale,

$$J_W(u) = \mathbb{E} \left[\int_0^T L_t^u \left(QX_t^2 + Ru_t^2 + 2NX_t u_t \right) dt + FL_T^u X_T^2 \right].$$

In this case, (WOC) can be stated as an optimal control problem in the strong sense on the original probability space,

$$\begin{cases} \min\limits_{u \in \mathcal{U}_{ad}} \ \mathbb{E} \left[\int_0^T L_t^u \left(QX_t^2 + Ru_t^2 + 2NX_t u_t \right) dt + FL_T^u X_T^2 \right], \\ s.t. \qquad dL_t^u = L_t^u B u_t \sigma^{-1} dW_t, \quad L_0 = 1, \\ \qquad\qquad dX_t = AX_t dt + \sigma dW_t, \quad X_0 = \xi. \end{cases}$$

The controlled process is not the state trajectory, but the likelihood process, and the state trajectory does not in any way effect the set of admissible controls. Instead, the set $\mathcal{U}_{ad}$ should be such that L^u is a true martingale (for example satisfy Novikov's condition) and the cost above is well-defined.

Stochastic differential games

As put by Lewin in [63], *"the theory of differential games is a blending of the notions of control theory with the decision structures and solution concepts of classical game theory. A differential game model cannot hence be considered as a double control problem. In general we can reduce a differential game* *model to a control model if we assume that only one player is active and the other is not."*

The last assertion is based on the fact that a the game reduces to a minimization problem whenever there is only one active player. The optimal control problem can therefore be thought of as a single-player game.

Motivated by pursuit and evasion problems, Isaacs laid the foundations for differential game theory in the 1950s [47]. The original work and further early research were incorporated into the book [48]. Around the same time, Pontryagin *et al.* treated optimal control problems in a new way [77] and their seminal work was promptly extended to two-player games in [53].

We will define a stochastic differential game (SDG) on a probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, defined as in Section 2.1 except that now it carries $N \in \mathbb{N}$ standard Wiener processes $(B^i)_i$ and initial conditions $(x_i)_i$, all independent. To work in accordance with Definition 1, which defines a normal-form game, we need to precise a set of players, their sets of actions, and their performance relations in order to construct a game. Consider N players, let U_i be a separable metric space for $i \in \mathcal{N} = \{1, \ldots, N\}$, and let the set of actions for player i be

$$\mathcal{U}_i := \{u \ : \ [0,T] \times \Omega \to U_i \mid u \text{ is } \mathbb{F}\text{-adapted}\}.$$

We might have to restrict $\mathcal{U}_i$ to a set of admissible actions $\mathcal{U}_{i,\mathrm{ad}}$ as in the previous two sections. Given an action vector $\mathbf{u} = (u^1, \ldots, u^N) \in \times_{i=1}^N \mathcal{U}_{i,\mathrm{ad}}$, player i's state is given by the stochastic system

$$\begin{cases} dX_t^i = b_i(t, \mathbf{X}_t, \mathbf{u}_t)dt + \sigma_i(t, \mathbf{X}_t, \mathbf{u}_t)dB_t^i, \\ X_0^i = x_i, \end{cases}$$

where $\mathbf{X} := (X^1, \ldots, X^N)$. Let player i's cost function (performance relation) be

$$J_i(\mathbf{u}) = \mathbb{E}\left[\int_0^T f_i(t, \mathbf{X}_t, \mathbf{u}_t)dt + g_i(\mathbf{X}_T)\right].$$

This setup is a N**-player nonzero-sum non-cooperative** SDG. The same questions that are asked in optimal control theory can be asked here:

- Existence of a Nash equilibrium;
- Explicit computation of Nash equilibria.

The problem presented above is one of many popular variants of a differential game. The theory is rich with adaptations (information structures, risk-sensitivity, etc., cf. Remark 8), each designed to suit one of the many applications.

Computation of optimal stochastic controls and differential game equilibria

Stochastic optimal control problems and differential games are approached in very similar ways when it comes to actually computing optimal controls and Nash equilibria, respectively. There are two predominant tools for characterizing optimal controls in general optimal control theory. Variational methods, sometimes called the probabilistic approach, lead to the so-called **Pontryagin maximum principle**. The dynamic programming method, based on Bellman's principle and sometimes called the partial differential equation (PDE) based approach, leads to the **Hamilton-Jacobi-Bellman** (HJB) equation. Other paths include the direct method (cf. [27] and references therein) and Wiener chaos expansion [46, 82].

The stochastic maximum principle

Pontryagin's maximum principle yields necessary conditions for an optimal control. It is in a way the infinite-dimensional analog of the zero-derivative condition or the KKT condition in finite dimensional function optimization. For the stochastic optimal control problem (OC) it states that if $\hat{u}$ is the optimal control and $\hat{X}$ is the corresponding state trajectory, then there are pairs of processes (p, q) and (P, Q) satisfying the **adjoint equations**, a system of linear backward SDEs (BSDE), such that $\hat{u}$ maximizes the (possibly extended) **Hamiltonian** H pointwisely: for almost every $t \in [0, T]$ and $\mathbb{P}$-a.s.,

$$H(t, \hat{X}_t, \hat{u}_t, p_t, q_t, P_t, Q_t) = \max_{u \in U} H(t, \hat{X}_t, u, p_t, q_t, P_t, Q_t). \tag{3}$$

The infinite-dimensional minimization of J has been replaced by the pointwise minimization of H. The trade-off lies in solving the adjoint equations, that are coupled with the optimal state trajectory. This coupled system has mixed temporal boundary conditions and is a so-called forward-backward SDE (FBSDE). FBSDE theory is an active area of research in itself [3, 42, 85]. The change to pointwise optimization allows for explicit solutions for certain classes of optimal control problems and games such as the linear-quadratic case.

An optimality condition is derived by perturbing an optimal control on a small measurable subset of $[0, T]$ with a spike variation. The variational inequality (3) is then obtained after performing a first order Taylor expansion (or second order if the diffusion is controlled, see [76]) with respect to the perturbation and by sending the perturbation to zero. The adjoint equations appear naturally when integrating the perturbed terminal cost by parts, yielding the so-called **duality relation**.

Example 10 (Linear quadratic)**.** *Consider the example of a strong linear-quadratic optimal control problem. Since only the drift is controlled, we do not have to consider the second order adjoint process (P, Q). The first order adjoint process (p, q) should be constrained at the time horizon T. To clearly see what we should set it to, we have to first introduce the spike variation and look at the variation of the cost. Assume that $\hat{u}$ is an optimal control and consider a spike-variation of $\hat{u}$, i.e. a switch of $\hat{u}$ to an arbitrary $u \in \mathcal{U}_{ad}$ on a set $E_\epsilon \subset [0, T]$ of size $\int_{E_\epsilon} dt = \epsilon$,*

$$u_t^\epsilon = \begin{cases} \hat{u}_t, & t \notin E_\epsilon, \\ u_t, & t \in E_\epsilon. \end{cases}$$

Denote by X^ϵ and $\hat{X}$ the state trajectory when using u^ϵ and $\hat{u}$, respectively, and let $\xi^\epsilon := X^\epsilon - \hat{X}$. Then

$$d\xi_t^\epsilon = (A\xi_t^\epsilon + Bu_t 1_{E_\epsilon}(t))\, dt, \quad \xi_0^\epsilon = 0.$$

For any sufficiently nice function f, estimates of the following form can be derived

$$\mathbb{E}[f(X_t^\epsilon) - f(\hat{X}_t)] = \mathbb{E}[f'(\hat{X}_t)\xi_t^\epsilon] + o(\epsilon).$$

At this point, the optimal cost and the spike-perturbed cost can be compared,

$$\begin{aligned} &J(u^\epsilon) - J(\hat{u}) \\ &= \mathbb{E}\left[\int_0^T \Big(Q((X_t^\epsilon)^2 - \hat{X}_t^2) + Ru_t^2 1_{E_\epsilon}(t) + 2N\xi_t^\epsilon u_t 1_{E_\epsilon}(t)\Big)\, dt\right] \\ &\qquad + \mathbb{E}\left[F\left((X_T^\epsilon)^2 - \hat{X}_T^2\right)\right] \\ &= \mathbb{E}\left[\int_0^T \Big(2Q\hat{X}_t\xi_t^\epsilon + \big(Ru_t^2 + 2N\xi_t^\epsilon u_t\big)\, 1_{E_\epsilon}(t)\Big)\, dt + 2F\hat{X}_T\xi_T^\epsilon\right] + o(\epsilon). \end{aligned}$$

Setting $p_T = -\partial_x g(\hat{X}_T) = -2F\hat{X}_T$, *the terminal cost can be rewritten*

$$
\begin{aligned}
&\mathbb{E}\left[F\left((X_T^\epsilon)^2 - \hat{X}_T^2\right)\right] = -\mathbb{E}\left[p_T\xi_T^\epsilon\right] + o(\epsilon) \\
&= -\mathbb{E}\left[\int_0^T p_t d\xi_t^\epsilon + \int_0^T \xi_t^\epsilon dp_t + \int_0^T d\langle p, \xi^\epsilon\rangle_t\right] + o(\epsilon) \\
&= -\mathbb{E}\left[\int_0^T \left(p_t(A\xi^\epsilon + Bu_t 1_{E_\epsilon}(t)) + \xi_t^\epsilon \mu_p(t)\right) dt + \int_0^T \xi_t^\epsilon \sigma_p(t) dW_t\right] + o(\epsilon) \\
&= -\mathbb{E}\left[\int_0^T \left((Ap_t + \mu_P(t))\xi^\epsilon + Bp_t u_t 1_{E_\epsilon}(t)\right) dt\right] + o(\epsilon),
\end{aligned}
$$

where μ_p *and* σ_p *are the drift and diffusion cofficients of* p, *respectively. This is the so-called duality relation. Gathering terms, we see that*

$$
\begin{aligned}
&J(u^\epsilon) - J(\hat{u}) \\
&\quad = \mathbb{E}\left[\int_0^T \left(2Q\hat{X} - Ap_t - \mu_p(t)\right)\xi_t^\epsilon dt\right] \\
&\qquad + \mathbb{E}\left[\int_0^T \left(Ru_t^2 + 2N\xi_t^\epsilon u_t - Bp_t u_t\right) 1_{E_\epsilon}(t) dt\right] + o(\epsilon) \\
&\quad = -\mathbb{E}\left[\int_0^T \left(H(t, u_t^\epsilon, \hat{X}_t, p_t, q_t) - H(t, \hat{u}_t, \hat{X}_t, p_t, q_t) 1_{E_\epsilon}(t)\right) dt\right] + o(\epsilon),
\end{aligned}
$$

where the Hamiltonian H *is defined as*

$$H(t, u, x, p, q) := -Qx^2 - Ru^2 - 2Nxu + p(Ax + Bu),$$

and μ_p *and* σ_p *have been chosen so that* (p, q) *is the solution to the BSDE*

$$
\begin{cases}
dp_t = -\partial_x H(t, \hat{u}_t, \hat{X}_t, p_t, q_t)dt + q_t dW_t = (Ap_t - 2Q\hat{X}_t)dt + q_t dW_t, \\
p_T = -\partial_x g(\hat{X}_T) = -2\hat{X}_T.
\end{cases}
$$

The BSDE above is the first order adjoint equation. The maximum principle has thus given us necessary conditions for optimality: if $\hat{u}$ *is optimal then it maximizes the quadratic Hamiltonian pointwise,*

$$\hat{u}_t = \arg\max_{u \in U} H(t, u, \hat{X}_t, p_t, q_t), \quad \mathbb{P}\text{-a.s. for a.e. } t, \tag{4}$$

where $\hat{X}$ *and* (p, q) *solves the linear FBSDE system*

$$
\begin{cases}
d\hat{X}_t = \left(A\hat{X}_t + B\hat{u}_t\right) dt + \sigma dW_t, & \hat{X}_0 = \xi, \\
dp_t = \left(Ap_t - 2Q\hat{X}_t\right) dt + q_t dW_t, & p_T = -2\hat{X}_T.
\end{cases} \tag{5}
$$

In the linear-quadratic case with $Q > 0$ the first order optimality condition (4) *is satisfied by $\hat{u}_t = \frac{B}{R} p_t - 2\frac{N}{R}\hat{X}_t$, $\mathbb{P}$-a.s. for almost every t. Inserting this $\hat{u}_t$ into* (5) *we see that $\hat{u}$ is characterized by the following fully coupled linear FBSDE*

$$\begin{cases} d\hat{X}_t = \left((A - 2\frac{NB}{R})\hat{X}_t + \frac{B^2}{R} p_t\right) dt + \sigma dW_t, & \hat{X}_0 = \xi, \\ dp_t = \left(Ap_t - 2Q\hat{X}_t\right) dt + q_t dW_t, & p_T = -2\hat{X}_T. \end{cases}$$

The Hamilton-Jacobi-Bellman equation

In the context of the stochastic optimal control problem (OC), the value function is

$$V(t, x') = \inf_{u \in \mathcal{U}_{\mathrm{ad}}} \mathbb{E}\left[\int_t^T f(t, X_t, u_t) ds + g(X_T) \mid X_t = x'\right],$$

where X is the solution to (1). Dynamic programming tells us that the value function (formally) satisfies the HJB equation

$$\begin{cases} \partial_t V(t,x) + \inf_{u \in U} L\left(t, x, \partial_x V(t,x), \partial^2_{xx} V(t,x), u\right) = 0, \\ V(T,x) = g(x), \end{cases} \tag{6}$$

where

$$L(t,x,y,z,u) := f(t,x,u) + yb(t,x,a) + \frac{1}{2}\mathrm{Tr}\left[z\sigma(t,x,u)\sigma^*(t,x,u)\right].$$

Note that as in the probabilistic approach, the PDE approach converts infinite-dimensional minimization of J to pointwise minimization of L. The trade-off is that we have introduced the HJB equation, which is coupled with the state trajectory and constrained at the terminal time T. An optimal control can be identified in terms of the derivatives of V by solving the pointwise minimization in (6). The approach extends to two-person zero-sum games and other nonzero-sum games, where the HJB equation is replaced by the Hamilton-Jacobi-Isaacs (HJI) equation and system of coupled PDEs (usually also referred to as the HJB system), respectively. Since the dynamic programming approach relies on the Bellman principle it only applies to time-consistent problems. Time-inconsistent problems, such as the minimization of $J(u) = \mathbb{E}[\phi(X^u_T, E[X^u_T])]$ for a nonlinear ϕ, cannot be treated via the dynamic programming as it is based on the tower property of the conditional expectation. Time consistency is not an issue for Pontryagin's maximum principle. The HJB equation and its generalization are highly nonlinear and seldom allow for neither smooth nor unique solutions. The rigorous derivation of the PDEs has engaged many researchers, some of that work includes [32, 80, 21, 49].

Example 11 (Linear quadratic)**.** *For the linear-quadratic problem,*

$$L(t,x,y,z,u) = Qx^2 + Ru^2 + 2Nxu + y(Ax+Bu) + \frac{\sigma^2}{2}z.$$

The minimizer of L is (for positive R) $\hat{u} = -2\frac{N}{R}x - \frac{B}{R}y$ and

$$L(t,x,y,z,\hat{u}) = Qx^2 + Axy + \frac{\sigma^2}{2}z.$$

The density m of the state trajectory when using the control u satisfies the Fokker-Planck equation

$$\begin{cases} \partial_t m + \partial_x \left((Ax+Bu)m\right) - \dfrac{\sigma^2}{2}\partial_{xx} m = 0, \quad (t,x) \in (0,T] \times \mathbb{R}, \\ m(0,x) = m_0(x), \quad x \in \mathbb{R}, \end{cases}$$

where m_0 denotes the density of X_0. The dynamic programming approach hence gives the following necessary condition on an optimal control: if $\hat{u}$ is optimal then $\hat{u}_t = -2\frac{N}{R}x - \frac{B}{R}\partial_x V$ where V is part of the solution to the coupled PDE system

$$\begin{cases} \partial_t V + Qx^2 + Ax\partial_x V + \frac{\sigma^2}{2}\partial_{xx} V = 0, & (t,x) \in [0,T] \times \mathbb{R}, \\ V(T,x) = g(x), & x \in \mathbb{R}, \\ \partial_t m + \partial_x \Big(\big((A - 2\frac{NB}{R})x - \frac{B^2}{R}\partial_x V\big) m \Big) - \frac{\sigma^2}{2}\partial_{xx} m = 0, & (t,x) \in [0,T] \times \mathbb{R}, \\ m(0,x) = m_0(x), & x \in \mathbb{R}. \end{cases}$$

Remark 12. *Comparing Example 10 and 11, the negative gradient of the value function $-\partial_x V$ seems to resemble the adjoint variable p from the maximum principle approach. This is indeed often the case; many examples are given in the preface of [86]. We also note that the PDE system above is not fully coupled, in contrast with the FBSDE in Example 10, but that will no longer be the case in the mean-field regime, where the corresponding PDE system is fully coupled.*

2.2 The mean-field approximation

Are games with many players harder or easier to understand than those with few players? On the one hand, the answer is that they are harder; a strategic game, as we have defined it, grows exponentially with the number of players N. But for many natural classes of games, we might hope that the answer is that they are easier. Perhaps the influence of each player on the outcome is small, which in turn makes the equilibrium behavior easy to characterize. For example, in a large human crowd the action of a single pedestrian has only negligible effect on the evolution of the crowd density. The individual agent is strategically insignificant. The ideal situation of strategic insignificance can only be obtained in models featuring a continuum of agents, as argued by Aumann in [4]. We will in the following sections see the continuum of agents appearing as the probability distribution of the states and controls of the agents. The distributions will replace the explicit agent-to-agent interaction in games with finitely many agents. This replacement is a **mean-field approximation**.

Large population limits of particle systems

Consider the particle system

$$\begin{cases} dX_t^i = \frac{1}{N}\sum_{j=1}^N b(X_t^i, X_t^j)dt + dB_t^i, \\ X_0^i = x_i, \quad i = 1, \dots, N, \end{cases} \tag{7}$$

where (B^i) are independent Wiener processes, x_i are independent and identically distributed random variables also independent of the (B^i). The mapping b is assumed to be Lipschitz continuous. Under these assumptions the system (7) has a unique strong solution and furthermore, the X^i are exchangeable. The law of large numbers implies that for large N,

$$\frac{1}{N}\sum_{j=1}^N b(X_t^i, X_t^j) \approx \int_{\mathbb{R}^d} b\left(X_t^i, x\right) d\mu_t(x),$$

where $\mu_t = \mathbb{P}\circ\left(X_t^i\right)^{-1}$ is the distribution of X_t. In fact, the mean-square limit of the interacting particle system (7) is the **McKean-Vlasov equation**

$$\begin{cases} dX_t^i = \int_{\mathbb{R}^d} b(X_t^i, y) d\mu_t(y) dt + dB_t^i, \\ X_0^i = x_i. \end{cases}$$

The McKean-Vlasov equation also goes under the name **nonlinear diffusion** and **SDE of mean-field type**. The study of the McKean-Vlasov equation, interacting diffusions, and Markov processes originates from Kac's approach to kinetic theory [52] and McKean's work on nonlinear parabolic equations [67]. The monograph by Sznitmann [81] is a classical references for the mean-field limits of particle systems, although in that model the coefficients are linear in the law of the process. The nonlinear case is treated in [50], which also provides a general result on the existence of strong solutions to nonlinear SDEs. Further important contributions include Oelschläger's weak law of large numbers [72, 73] and the derivation of BSDEs of mean-field type [10] (which are vital for stochastic dynamic programming and the SMP).

Example 13 (Linear quadratic)**.** *Consider N interacting agents, each with control u^i and state trajectory X^i. The X^i satisfy the SDE system*

$$\begin{cases} dX_t^i = \Big(AX_t^i + Bu_t^i + \frac{C}{N}\sum_{j=1}^N (X_t^i - X_t^j)\Big)\,dt + \sigma dW_t^i, \\ X_0^i = x_i, \quad i = 1, \dots, N. \end{cases} \tag{8}$$

The coupling in the interacting system (8) *is the average state $\frac{1}{N}\sum_j X_t^j$. If the x_i are exchangeable and identically distributed, the W^i are independent Wiener processes, and the x_i and the W^i are independent, the mean-field approximation of* (8) *is*

$$\begin{cases} d\bar{X}_t^i = \big(A\bar{X}_t^i + Bu_t^i + C(\bar{X}_t^i - \mathbb{E}[\bar{X}_t^i])\big)\,dt + \sigma dW_t^i, \\ \bar{X}_0^i = x_i. \end{cases}$$

Estimates of the following type can be derived

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X_t^i - \bar{X}_t^i|^2\right] \le \frac{C}{N}$$

where C does not depend on N, see for example [50].

Large population limits of symmetric games

Consider a game in normal form with a large number of players who all choose their strategies from the same compact metric space U.

Definition 14. *A normal-form game is called* ***symmetric*** *if $U_i = U$ for all $i \in \mathcal{N}$ and $J_i(u_1, \dots, u_N) = J_{\theta(i)}(u_{\theta(i)}, \dots, u_{\theta(N)})$ for any permutation θ on $\mathcal{N}$.*

Nash proved in [70] that if a finite normal-form game is symmetric, then there exists a symmetric mixed Nash equilibrium $(\hat{\mu}, \dots, \hat{\mu})$, where $\hat{\mu} \in \mathcal{P}(U)$. In a symmetric game the payoff function of player i is by definition a symmetric function. Hence, under suitable bounds and continuity assumptions, we can approximate J_i for large N,

$$J_i(u_1, \dots, u_N) \approx J\left(u_i, \frac{1}{N-1}\sum_{j\neq i}\delta_{u_j}\right), \tag{9}$$

where $J : U \times \mathcal{P}(U) \to \mathbb{R}$. See [16] for a detailed exposition of the arguments and the existence of the function J. The following result from [64] gives sufficient regularity assumption for the mean-field approximation to work in symmetric games:

Assume that J is Lipschitz continuous (with respect to the Wasserstein distance in the measure-valued argument) and that $(\hat{\mu}^{(N)}, \dots, \hat{\mu}^{(N)})$ is a symmetric Nash equilibrium for $\langle \mathcal{N}, (\mathcal{P}(U))_{i=1}^N, (\mathcal{J}_i)_{i=1}^N \rangle$ with $N \in \mathbb{N}$, where

$$\mathcal{J}_i(\mu) := \int_{U^N} J\left(u_i, \frac{1}{N-1}\sum_{j\neq i}\delta_{u_j}\right)\mu(du).$$

Then, up to a subsequence, $\hat{\mu}^{(N)}$ converges to a measure $\hat{\mu}$ as $N \to \infty$ and $\hat{\mu}$ satisfies the mean-field equation

$$\int_U J(v,\bar{\mu})\bar{\mu}(dv) = \inf_{\mu\in\mathcal{P}(U)} \int_U J(v,\bar{\mu})\mu(dv).$$

Compare this with the cooperative scenario, where all agents jointly minimize the average agent cost. Under similar assumptions as stated above, the mean-field equation for $\bar{\mu}$ then becomes

$$\int_U J(v,\bar{\mu})\bar{\mu}(dv) = \inf_{\mu\in\mathcal{P}(U)} \int_U J(v,\mu)\mu(dv).$$

Formally, this can be justified by taking the average player cost, then applying the approximation (9) and finally taking the limit $N \to \infty$.

In both cases, we see that the mean-field approximation replaces agent-to-agent interaction with a dependence on the measure $\bar{\mu}$ which represents an equilibrium (or a socially optimal) configuration of a continuum of agents.

Mean-field type control

Optimal control of an SDE of mean-field type is a nonstandard control problem. One common version of the **mean-field type optimal control problem** (MFTC) reads as follows: minimize

$$J(u) = \mathbb{E}\left[\int_0^T f\left(t, X_t, \mathbb{P}\circ X_t^{-1}, u_t\right) dt + g\left(X_T, \mathbb{P}\circ X_T^{-1}\right)\right],$$

over $u \in \mathcal{U}_{\text{ad}}$ given that X solves the SDE of mean-field type

$$\begin{cases} dX_t = b\left(t, X_t, \mathbb{P}\circ X_t^{-1}, u_t\right) dt + \sigma\left(t, X_t, \mathbb{P}\circ X_t^{-1}, u_t\right) dB_t, \\ X_0 = x. \end{cases}$$

Imagine a cooperating, perfectly informed, and interacting crowd of pedestrians which we would like to model with a single decision maker, the so-called representative agent. The agents cooperate, hence they can influence the group characteristics. Because of the large number of agents, an aggregation effect takes place and the impact of the community can be modeled with the mean-field dependence $\mathbb{P}\circ X_t^{-1}$ (corresponding to μ in the preceding section).

The classical stochastic maximum principle (SMP) was extended to optimal control with mean-field couplings of the form $\mathbb{E}[\phi(X_t)]$ in [2, 11]. Hosking [41] extended the SMP to mean-field couplings of the form $\mathbb{E}[\phi(X_t, u_t)]$. Couplings of the form $\phi(t, X_t, \mathbb{P}\circ X_t^{-1}, u_t)$ are considered in [18, 12]. Stochastic dynamic programming for mean-field type control is considered in [7, 62] and the book [6] outlines solution techniques. Lacker proves in [58] the convergence of a particular particle system (as $N \to \infty$) to the solution of a MFTC using relaxed controls, extending the works [31, 72]. These convergence results motivate the use of MFTC to model cooperating crowds.

Example 15 (Linear quadratic)**.** *Consider the following linear-quadratic control problem of mean-field type,*

$$\begin{cases} \min\limits_{u\in\mathcal{U}_{ad}} \mathbb{E}\left[\int_0^T f(X_t, \mathbb{E}[X_t], u_t)dt + g(X_T, \mathbb{E}[X_T])\right], \\ s.t.\ dX_t = b(X_t, \mathbb{E}[X_t], u_t)dt + \sigma dW_t,\ X_0 = x_0, \end{cases}$$

where

$$\begin{aligned} f(x,y,u) &:= \frac{1}{2}\left(Qx^2 + Q'y^2 + Ru^2\right) + Nxu + N'yu + N''xy, \\ g(x,y) &:= \frac{1}{2}\left(Fx^2 + F'y^2\right) + Gxy, \\ b(x,y,u) &:= Ax + A'y + Bu. \end{aligned}$$

The maximum principle of [2] says that any optimal control $\hat{u}$ maximizes the Hamiltonian pointwise,

$$\hat{u}_t = \arg\max_{u\in U} H(t, \hat{X}_t, \mathbb{E}[\hat{X}_t], u, p_t, q_t), \quad \mathbb{P}\text{-}a.s.,\ a.e.\text{-}t,$$

where $H(t,x,y,u,p,q) := pb(x,y,u) - f(x,y,u)$ and (p,q) solves the adjoint equation

$$\begin{cases} dp_t = -\partial_x H(t, \hat{X}_t, \mathbb{E}[\hat{X}_t], \hat{u}_t, p_t, q_t)dt + q_t dW_t \\ \qquad\qquad - \mathbb{E}\left[\partial_y H(t, \hat{X}_t, \mathbb{E}[\hat{X}_t], \hat{u}_t, p_t, q_t)\right] dt, \\ p_T = -\partial_x g(\hat{X}_T, \mathbb{E}[\hat{X}_T]) - \mathbb{E}\left[\partial_y g(\hat{X}_T, \mathbb{E}[\hat{X}_T])\right]. \end{cases} \tag{10}$$

If the problem does not depend explicitly on the marginal distribution of the state trajectory, it reduces to the one analyzed in Example 10. The first order optimality condition yields

$$\hat{u}_t = \frac{B}{R}p_t - \frac{N}{R}\hat{X}_t - \frac{N'}{R}\mathbb{E}[\hat{X}_t].$$

Explicitly calculating the derivatives in (10) *we get*

$$\begin{aligned} \partial_x H &= Ap - (Qx + Nu + N''y), \\ \partial_y H &= A'p - (Q'y + N'u + N''x), \\ \partial_x g &= Fx + Gy, \\ \partial_y g &= F'y + Gx. \end{aligned}$$

The FBSDE system of mean-field type characterizing the optimal control is

$$
\begin{cases}
d\hat{X}_t = \left\{ \dfrac{B^2}{R} p_t + \left(A - \dfrac{N}{R} \right) \hat{X}_t + \left(A' - \dfrac{N'}{R} \right) \mathbb{E}[\hat{X}_t] \right\} dt + \sigma dW_t, \\
X_0 = x_0 \\
dp_t = -\left\{ \left(A - \dfrac{NB}{R} \right) p_t + \left(A' - \dfrac{N'B}{R} \right) \mathbb{E}[p_t] - \left(Q - \dfrac{N^2}{R} \right) \hat{X}_t \right. \\
\qquad \left. - \left(Q' + 2N'' - \dfrac{2NN' + (N')^2}{R} \right) \mathbb{E}[\hat{X}_t] \right\} dt + q_t dW_t, \\
p_T = -F\hat{X}_T - (F' + 2G)\mathbb{E}[X_T].
\end{cases}
\tag{11}
$$

System (11) *is explicitly solvable up to a system of ODEs, a fact that has been used in papers B and D, below.*

Mean field games

A **mean field game** (MFG) is a system of equations that approximates the asymptotic limit of a non-cooperative differential game in equilibrium with identical and indistinguishable players when the number of players is large. In many cases the MFG provides an approximate Nash equilibrium for a game with a finite number of players. In this setting, individual interactions become less and less relevant as the size of the population grows, and in the limit only the distribution of player states matters. Although there were earlier efforts to deal with large population games, such as [51], MFGs were introduced independently by Huang, Malhamé, and Caines [43, 44], and by Lasry and Lions [59, 60, 61] in the 2000s. The solution to a MFG can be viewed as a fixed point to a two-step scheme: the MFG equilibrium control $\hat{u}$ and the corresponding state trajectory distribution $\hat{\mu}$ are given by:

1. Fix $\mu = (\mu_t)_{t\in[0,T]} \in \mathcal{P}(C([0,T];\mathbb{R}^n))$ and solve the standard optimal control problem

$$
\begin{cases}
\min\limits_{u\in\mathcal{U}_{\mathrm{ad}}} \mathbb{E}\left[\displaystyle\int_0^T f(t, X_t, \mu_t, u_t)dt + g(X_T, \mu_T) \right], \\
\text{s.t. } dX_t = b(t, X_t, \mu_t, u_t)dt + \sigma(t, X_t, \mu_t, u_t)dB_t, \\
\qquad X_0 = x.
\end{cases}
$$

   Call the solution u^μ and the corresponding state trajectory X^μ.

2. Find $\hat{\mu}$ such that $\mathbb{P} \circ \left(X_t^{\hat{\mu}}\right)^{-1} = \hat{\mu}_t$.

A necessary condition for any output of the matching procedure can be stated in terms of an FBSDE system, with the forward part being the state dynamics and the backward part being either the HJB or the adjoint equation (depending on which technique that is used). MFG theory is presented in detail in for example [16, 6, 19], to mention only a few references.

Example 16 (Linear quadratic)**.** *Consider the same setting as in Example 15. Let H be the Hamiltonian*

$$H(t,x,y,u,p) = f(t,x,y,u) - pb(t,x,y,u).$$

The stochastic maximum principle for mean field games [6, Ch. 3.2] says that any control $\hat{u}$ that simultaneously satisfies 1. *and* 2. *above minimizes the Hamiltonian,*

$$\hat{u}_t = \arg\min_{u \in U} H(t, \hat{X}_t, \mathbb{E}[\hat{X}_t], u, p_t), \quad a.e.\ t \in [0,T],\ \mathbb{P}\text{-}a.s.,$$

where the state trajectory corresponding to $\hat{u}$, $\hat{X}$, and the adjoint, p, are given by the FBSDE system

$$\begin{cases} d\hat{X}_t = b(\hat{X}_t, \mathbb{E}[\hat{X}_t], \hat{u}_t)dt + \sigma dW_t, \\ X_0 = x_0 \\ dp_t = -\partial_x H(t, \hat{X}_t, \mathbb{E}[\hat{X}_t], \hat{u}_t, p_t, q_t)dt + qdW_t \\ p_t = -\partial_x g(\hat{X}_T, \mathbb{E}[\hat{X}_T]). \end{cases}$$

Notice how the y-derivatives of Example 15 are not present. The agent no longer has the power to influence its distribution, which is in line with the interpretation of games with a continuum of agents.

Mean-field type games

As optimal control can be extended to differential games, mean-field type optimal control can be extended to **mean-field type games** (MFTG). The element that distinguishes MFTG from SDG (as presented in Section 2.1) is the presence of the player state and control distribution in the dynamics and the cost. An MFTG is a game in which the payoffs and the state dynamics coefficient functions involve not only the state and actions profiles but also the distributions of state-action process (or its marginal distributions). In contrast to the MFG, in which a single player cannot influence the mean-field terms, a single player in an MFTG can have a substantial impact on the mean-field terms.

Consider the SDG from Section 2.1, but let its state trajectories and cost functionals be of mean-field type: for $i = 1, \ldots, N$,

$$\begin{cases} dX_t^i = b_i\left(t, \mathbf{X}_t, \mathbb{P} \circ \mathbf{X}_t^{-1}, \mathbf{u}_t\right) dt + \sigma_i\left(t, \mathbf{X}_t, \mathbb{P} \circ \mathbf{X}_t^{-1}, \mathbf{u}_t\right) dB_t^i, \\ \;X_0^i = x_i, \end{cases} \tag{12}$$

and

$$J_i(\mathbf{u}) = \mathbb{E}\left[\int_0^T f_i\left(t, \mathbf{X}_t, \mathbb{P} \circ \mathbf{X}_t^{-1}, \mathbf{u}_t\right) dt + g\left(\mathbf{X}_T, \mathbb{P} \circ \mathbf{X}_T^{-1}\right)\right]. \tag{13}$$

The conflict that arises when each player minimizes her cost (which is coupled to the other players' strategies) is a nonzero-sum differential N-player MFTG and we may ask when this game is in Nash equilibrium.

The paper [83] reviews the literature on the MFTG (12)–(13) and many of its variants and extensions. The paper also derives optimality conditions with dynamic programming, with an SMP, and using the Wiener chaos expansion. The direct method is applied to solve linear-quadratic MFTGs in [27]. An overview of applications of MFTG theory in the engineering sciences is found in [25].

2.3 Pedestrian crowd dynamics

Large crowds is a routine feature of today's large cities. They can be found not only at mass gatherings, but in train stations, malls, airports, and theaters. Pedestrian security is of real concern since a flawed design of the pedestrian environment can be deadly. Within the last 100 years crowd stampedes have caused over 4000 fatalities and ten times that many serious injuries [38]. To address crowding in the design stage of a project, mathematical models for crowd movement have been developed in the past decades with the aim to replicate and predict human crowd behavior.

Experimental studies of pedestrian behavior

Research on pedestrian behavior started in the 1960s with studies of the pedestrian flows in urban areas. A summary of the history of empirical studies of pedestrians with references is found in [34].

The real-world behavior of individual pedestrian has been studied in both controlled and uncontrolled environments. The experiments deal with the relation between walking speed and crowd density, the factors influencing walking speeds, and pedestrian preferences. Exhaustive overviews of empirical studies of pedestrian motion (and summaries of observed pedestrian characteristics) are found in [84, 37, 9]. Ideally, mathematical models should accommodate all the important pedestrian characteristics. This is not a simple requirement since pedestrian behavior can be highly complex, as the following two examples try to highlight.

- Flows in opposing directions (bidirectional flows) tend to separate, but bidirectional flows of moderate density through bottlenecks most often do not separate, instead the flow direction oscillates. While this effect may be understood as friendly behavior ("you go first, please"), oscillatory flows occur in simulations even in the absence of communication [34].

- Have you ever seen a pillar in front of an exit or an escalator? At very small crowd densities such obstacles hinder the outflow of pedestrians, but at high densities their role may turn out to be essential for lowering the evacuation time in a panic situation. This phenomenon is often explained with Braess' paradox [8], which says that if one or more roads are added to a road network the travel time may increase for all travelers on the network (the paradox can be framed as a Prisoner's Dilemma, in which players would be individually better off if the option to *Betray* was removed).

Mathematical modeling approaches

Modeling a single individual will not lead to a mathematical description of emerging collective behavior, which requires interaction. The interaction is what makes pedestrian crowd modeling interesting from a mathematical point of view. The idea that large-scale collective behavior emerges from local interactions among individuals has become a key concept in the understanding of crowd dynamics.

The most prevalent feature distinguishing between different modeling approaches is the scale:

- In a **microscopic description**, the state of the system is a collection of individual pedestrian states. Mathematical models are generally stated in terms of ODEs/SDEs. This approach includes the "social force" model [35], optimal control [40], and cellular automata [15];
- A **macroscopic description** is used when the state is a gross quantity like density, linear momentum, and kinetic energy. The pedestrians are assimilated to this gross quantity, and the mathematical models are generally stated in terms of conservation laws (PDEs). This approach includes fluid-dynamic models [39, 45], optimal transport [65], and mean-field games [26, 57];
- **Kinetic models** identify the individual states, however their representation is delivered by a suitable probability distribution over the (microscopic) states. Mathematical models describe the evolution of that distribution by means of integro-differential equations. This approach includes swarming [20].

An in-depth survey of modeling approaches is found in [5]. Also, the more recent doctoral thesis [1] includes a survey on state-of-the-art pedestrian crowd modeling.

The mean-field approach

A pedestrian crowd consists of a large number of interacting individuals. Crowd models with the microscopic pedestrian-to-pedestrian interaction replaced by interaction with a mean-field have become popular in this decade. In [40] an MFG-based model is suggested, where the mean-field term approximates the average number of pedestrian interactions. Following Lasry-Lions' seminal papers [59, 60, 61] plenty of MFG-based pedestrian crowd models appeared [26, 57, 24]. Simultaneously, models applying the MFTC theory appeared, see e.g. [13, 14, 24]. Let us first go through some of the strengths of the mean-field approach, then some of its limitations.

The mean-field approach is able to replicate congestion- and aversion-features of pedestrian crowds. For example, the simulation results of [57] suggest that obstacles help the crowd coordinate in the mean-field approach (an example of Braess' paradox). The term congestion covers a broad range of mechanisms in a crowd. Some authors use it to describe the situation when the cost of motion of an individual depends on the crowd density in an increasing manner. This can be incorporated as a cost: with u being the (controlled) velocity, X the pedestrian position, and m the crowd density, consider the following congestion penalty

$$E\left[\int_0^T (c+|u_t|^\alpha)\, m^\beta(t,X_t)dt\right], \tag{14}$$

with $c \geq 0, \alpha > 0$ and $\beta > 1$. As the density increases, movement (i.e. nonzero velocity u) becomes more costly. The presence of the crowd density in (14) makes it a mean-field type congestion cost. The effect described above is the so-called soft congestion. Hard congestion on the other hand refers to physical bounds on the density. This can be included in the mean-field type approach via inequality constraints on the crowd density [66, 17].

The mean-field approach grants the modeled pedestrians the ability to perfectly anticipate future crowd movement. Crisitiani *et al.* explore other information structures in [22]. They define six levels of pedestrian rationality. The study comprehensibly categorizes models after their level of rationality and not surprisingly MFTG/MFG-based models enjoy the highest levels of rationality, implying that their area of application are situations in which pedestrians feel that they can predict crowd movement and act accordingly. A numerical case study is carried out in [14] that showcases the difference between MFTC-based models and the Hughes model. In the latter, the pedestrians only have access to the current state of the crowd. The lowest levels of rationality is observed in panicking crowds, where pedestrians tend move without any coordination, see [36] and references therein.

3

Summary of papers

Here follows short summaries of the five appended papers in Part II that make up the main part of this thesis. Moreover, the contributions of the authors are clearly stated.

Paper A: Mean-field type modeling of nonlocal crowd aversion in pedestrian crowd dynamics.

Paper A of this thesis is an edited version of the paper

- A. Aurell and B. Djehiche. Mean-field type modeling of nonlocal crowd aversion in pedestrian crowd dynamics. *SIAM Journal on Control and Optimization, 56(1):434–455, 2018.*

Summary Pedestrians sense their surroundings and may react to events or obstacles far away from their current position. Paper A addresses nonlocal interactions and competition between many subcrowds in the mean-field approach of [57]. In the paper we model the 'personal space' of each pedestrian. The personal space is a neighbourhood where the pedestrian is concerned with congestion. More specifically, pedestrian i with state trajectory $(X_t^i; t \in [0,T])$ reacts aversively to crowding with the $N-1$ other pedestrians through a nonlocal cost. A smooth kernel ϕ_r defines a personal space of radius r and the crowding cost payed over the time horizon $[0,T]$ is

$$\mathbb{E}\left[\int_0^T \frac{1}{N-1}\sum_{j\neq i}\phi_r(X_t^i - X_t^j)\ dt\right],$$

The crowding cost is combined with energy conservation and a terminal cost realized at $t = T$ to design a nonlocal risk measure J_r. We assume that the pedestrians are cooperating, for example by following signs or a guide, and that they are indistinguishable in terms of their joint state trajectory distribution. These assumptions justify an approximation of the behavior of the crowd with a mean-field type control problem. Optimal behavior is characterized with a Pontryagin's type maximum principle. Furthermore a scenario of multiple non-cooperating crowds, constituting a mean-field type game, is studied. The risk measure is sensitive to crowd affiliation, that is, the level of aversion may vary from one crowd to another. The paper then derives conditions under which the game between crowds is equivalent to an optimal control problem.

Paper B: Modeling tagged pedestrian motion: A mean-field type game approach

Paper B of this thesis is an edited version of the paper

- A. Aurell and B. Djehiche. Modeling tagged pedestrian motion: A mean-field type game approach. *Transportation Research Part B: Methodological, 121:168–183, 2019.*

Summary The paper suggests a model for the so-called tagged pedestrians. Tagged pedestrians plan their motion backwards from a specified target location to an unspecified initial position. In this paper, the target location is to be reached at a certain time. The model is based on the mean-field type game approach and aims to be a decision making tool for the positioning of fire fighters, medical personnel, etc, during mass gatherings. The fixed and non-negotiable final target leads us to model the tagged's state trajectory with backward SDEs (BSDE) of mean-field type. Let $(Y_t)_{t\in[0,T]}$ and $(X_t)_{t\in[0,T]}$ be the state trajectories of agents representative for a tagged group and the surrounding crowd, respectively. The two groups interact and their state trajectories satisfy the forward-backward SDE system,

$$\begin{cases} dY_t = b_y(t, \Theta_y(t), Z_t, \Theta_x(t))dt + Z_t[dB_t^x \ dB_t^y]^*, \\ dX_t = b_x(t, \Theta_y(t), Z_t, \Theta_x(t))dt + \sigma_x(t, \theta_y(t), Z_t, \theta_x(t)dB_t^x, \\ Y_T = y_T, \quad X_0 = x_0, \end{cases}$$

where $\Theta_y(t) := (Y_t, \mathbb{P} \circ Y_t^{-1}, u_t^y)$, $\theta_y(t) := (Y_t, \mathbb{P} \circ Y_t^{-1})$, and Θ_x, θ_x defined correspondingly. The agents Y and X control their movement through the control processes u^y and u^x, respectively. We consider the game between the tagged pedestrians and a surrounding crowd and derive necessary conditions for its Nash equilibrium with an SMP. Scenarios including the bidirectional flow, where tagged pedestrians move towards a target and the crowd moves away from it, is studied numerically.

Paper C: Behavior near walls in the mean field approach to crowd motion

Paper C of this thesis is an edited version of the paper

- A. Aurell and B. Djehiche. Behavior near walls in the mean field approach to crowd motion. ArXiv preprint at arXiv:1907.07407.

The results of the paper were presented at the minisymposium *Mean Field Games: New Trends and Applications – Part 2* at the International Congress on Industrial and Applied Mathematics (ICIAM) in 2019.

Summary Walls and other obstacles have a twofold effect on pedestrian motion. Pedestrians may see the wall and react to it in advance. If the pedestrian fails to avoid collision she ends up at the wall where her movement is physically constrained. Boundary conditions at walls in the mean-field approach is the topic of Paper C, in which the sticky reflected SDE of mean-field type with boundary diffusion is proposed as an alternative to the popular no-flux boundary condition. The proposed state equation only admits a weak solution. Therefore, we formulate the model as a weak optimal control problem, i.e. the law of the state trajectory $\mathbb{P}^u$ is controlled through a control process u. On a bounded and simply connected domain $\mathcal{D}$ with C^2-boundary $\partial\mathcal{D}$ the state equation of a pedestrian using control u reads

$$\begin{cases} dX_t = 1_{\mathcal{D}}(X_t)\left(\beta(t, X_\cdot, \mathbb{P}^u \circ X_t^{-1}, u_t)dt + dB_t^u\right) \\ \qquad\quad + 1_{\partial\mathcal{D}}(X_t)\left(\pi(X_t)\beta(t, X_\cdot, \mathbb{P}^u \circ X_t^{-1}, u_t) - \dfrac{n(X_t)}{2\gamma}\right)dt \\ \qquad\quad + 1_{\partial\mathcal{D}}(X_t)dB_t^{\partial\mathcal{D},u}, \\ X_0 \in \bar{\mathcal{D}}, \end{cases}$$

where B^u is a $\mathbb{P}^u$-Brownian motion, $B^{\partial\mathcal{D},u}$ is the so-called boundary diffusion under $\mathbb{P}^u$, π is the projection on the tangent space of $\partial\mathcal{D}$ and n is the outward normal of $\partial\mathcal{D}$. The proposed state equation allows the pedestrians move and interact while spending time on the boundary. The paper begins with a proof of existence and uniqueness of weak solutions under assumptions on Lipschitz continuity and linear growth of the involved coefficients. Then, we show how to explicitly solve some linear-quadratic problems and illustrate pedestrian behavior in the model with a numerical experiment where speed profiles from experimental data are replicated. The simulations showcase the flexible boundary behavior in the suggested model.

Contribution The co-author of paper A, B, and C suggested the topic and helped with the formulation of the solved problems. The candidate solved the problems, performed the proofs of all the results, did all the coding and numerical examples, and wrote the manuscripts.

Paper D: Mean-Field Type Games between Two Players Driven by Backward Stochastic Differential Equations

Paper D of this thesis is an edited version of the paper

- A. Aurell. Mean-Field Type Games between Two Players Driven by Backward Stochastic Differential Equations. *Games, 9(4), 88, 2018.*

Summary The paper treats a two-player finite-horizon mean-field type game where the players' state trajectories are terminally constrained. At the same time, no initial condition is specified. To be specific, if the control pair (u^1, u^2) is used, the players' state trajectories Y^1 and Y^2 are given by the system

$$\begin{cases} dY_t^1 = b^1\left(t, \Theta_t^1, \Theta_t^2, Z_t^1, Z_t^2\right) dt + Z_t^1 dW_t, \\ dY_t^2 = b^2\left(t, \Theta_t^2, \Theta_t^1, Z_t^1, Z_t^2\right) dt + Z_t^2 dW_t, \\ Y_T^1 = y^1, \quad Y_T^2 = y^2, \end{cases} \tag{1}$$

where $\Theta_t^i := \left(Y_t^i, \mathbb{P} \circ (Y_t^i)^{-1}, u_t^i\right)$ for $i = 1, 2$. The players' state distributions are present in the drift terms in (1) through Θ^1 and Θ^2, making (1) a system of BSDEs of mean-field type. The class of equations has been well studied in the uncontrolled case [10, 19]. Here, the states are controlled and player i aims to minimize the cost

$$\mathbb{E}\left[\int_0^T f^i\left(t, \Theta_t^i, \Theta_t^{-i}\right) dt + h^i\left(Y_0^i, \mathbb{P} \circ (Y_0^i)^{-1}, Y_0^{-i}, \mathbb{P} \circ (Y_0^{-i})^{-1}\right)\right].$$

The competition between the two players constitutes a mean-field type game of BSDEs. In the paper, we use the stochastic maximum principle to derive necessary conditions for Nash equilibria of the game. The game in equilibrium is compared to an optimally controlled cooperating scenario. In the cooperating scenario the average player cost, the so-called social cost, is minimized.

Paper E: Stochastic Stability and Mixed Equilibria

Paper E of this thesis is an edited version of the paper

- A. Aurell, L. Dinetan, and G. Karreskog. Stochastic Stability and Mixed Equilibria

The results of the paper were presented at the session *Dynamics* at the 15th European Meeting on Game Theory (SING15) in 2019.

Summary This paper is concerned with finite two-player games that are recurrently played in a large population. The players have access to a history of previously used action profiles which they use to form a belief on how the opposing player will act. The paper proposes alternative game dynamics with a recency bias: more recent interactions are considered to be more important in the belief-forming process. Existence and uniqueness of an invariant distribution of historical action profiles is obtained. The paper also studies the convergence of the history of action profiles to a collection of minimal CURB blocks. The paper goes on to prove that in a certain class of two-player games, mainly characterized by that the minimal CURB blocks contain exactly one mixed Nash equilibrium each (e.g. the matching pennies game), the distribution of historical action profiles concentrates on the mixed Nash equilibria.

Contribution All the authors have contributed to all the parts of the work. The idea to use recency bias in the context of conventions is due to G. Karreskog. The idea to use a Krein-Rutman type theorem to analyze convergence of the Markov chain to its stationary distribution is due to L. Dinetan. All the proofs are due to myself and G. Karreskog, except the Lipschitz continuity (Section E.4) and Lemma 12 (Section E.5) which are due to L. Dinetan. G. Karreskog has implemented the code for the numerical experiments. Most of the writing has been done by myself and G. Karreskog.

References

[1] F. Al Reda. *Crowd motion modelisation under some constraints*. PhD thesis, 2017.

[2] D. Andersson and B. Djehiche. A maximum principle for SDEs of mean-field type. *Applied Mathematics & Optimization*, 63(3):341–356, 2011.

[3] F. Antonelli. Backward-forward stochastic differential equations. *The Annals of Applied Probability*, 3(3):777–793, 1993.

[4] R. J. Aumann. Markets with a Continuum of Traders. *Econometrica: Journal of the Econometric Society*, pages 39–50, 1964.

[5] N. Bellomo and C. Dogbe. On the modeling of traffic and crowds: A survey of models, speculations, and perspectives. *SIAM review*, 53(3): 409–463, 2011.

[6] A. Bensoussan, J. Frehse, and P. Yam. *Mean field games and mean field type control theory*, volume 101. Springer, 2013.

[7] A. Bensoussan, J. Frehse, and P. Yam. The master equation in mean field theory. *Journal de Mathématiques Pures et Appliquées*, 103(6): 1441–1474, 2015.

[8] D. Braess. Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung*, 12(1):258–268, 1968.

[9] S. Buchmueller and U. Weidmann. Parameters of pedestrians, pedestrian traffic and walking facilities. *IVT Schriftenreihe*, 132, 2006.

[10] R. Buckdahn, B. Djehiche, J. Li, S. Peng, et al. Mean-field backward stochastic differential equations: a limit approach. *The Annals of Probability*, 37(4):1524–1565, 2009.

[11] R. Buckdahn, B. Djehiche, and J. Li. A general stochastic maximum principle for SDEs of mean-field type. *Applied Mathematics & Optimization*, 64(2):197–216, 2011.

[12] R. Buckdahn, J. Li, and J. Ma. A stochastic maximum principle for general mean-field systems. *Applied Mathematics & Optimization*, 74 (3):507–534, 2016.

[13] M. Burger, M. Di Francesco, P. A. Markowich, and M.-T. Wolfram. On a mean field game optimal control approach modeling fast exit scenarios in human crowds. In *52nd IEEE Conference on Decision and Control*, pages 3128–3133. IEEE, 2013.

[14] M. Burger, M. Di Francesco, P. A. Markowich, and M.-T. Wolfram. Mean field games with nonlinear mobilities in pedestrian dynamics. *Discrete and Continuous Dynamical Systems-Series B*, 19(5):1311–1333, 2014.

[15] C. Burstedde, K. Klauck, A. Schadschneider, and J. Zittartz. Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A: Statistical Mechanics and its Applications*, 295(3-4):507–525, 2001.

[16] P. Cardaliaguet. Notes on mean field games. Technical report, 2010.

[17] P. Cardaliaguet, A. R. Mészáros, and F. Santambrogio. First order mean field games with density constraints: pressure equals price. *SIAM Journal on Control and Optimization*, 54(5):2672–2709, 2016.

[18] R. Carmona and F. Delarue. Forward–backward stochastic differential equations and controlled McKean–Vlasov dynamics. *The Annals of Probability*, 43(5):2647–2700, 2015.

[19] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I-II.* Springer, 2018.

[20] J. A. Carrillo, M. Fornasier, G. Toscani, and F. Vecil. Particle, kinetic, and hydrodynamic models of swarming. In *Mathematical modeling of collective behavior in socio-economic and life sciences*, pages 297–336. Springer, 2010.

[21] M. G. Crandall and P.-L. Lions. Viscosity solutions of Hamilton-Jacobi equations. *Transactions of the American mathematical society*, 277(1): 1–42, 1983.

[22] E. Cristiani, F. S. Priuli, and A. Tosin. Modeling rationality to control self-organization of crowds: an environmental approach. *SIAM Journal on Applied Mathematics*, 75(2):605–629, 2015.

[23] M. Davis. On the existence of optimal policies in stochastic control. *SIAM Journal on Control*, 11(4):587–594, 1973.

[24] B. Djehiche, A. Tcheukam, and H. Tembine. A Mean-Field Game of Evacuation in Multilevel Building. *IEEE Transactions on Automatic Control*, 62(10):5154–5169, 2017.

[25] B. Djehiche, A. Tcheukam, and H. Tembine. Mean-Field-Type Games in Engineering. *AIMS Electronics and Electrical Engineering*, 1 (ElectronEng-01-00018):18, 2017. ISSN 2578-1588.

[26] C. Dogbé. Modeling crowd dynamics by the mean-field limit approach. *Mathematical and Computer Modelling*, 52(9-10):1506–1520, 2010.

[27] T. Duncan and H. Tembine. Linear–Quadratic Mean-Field-Type Games: A Direct Method. *Games*, 9(1):7, 2018.

[28] T. Duncan and P. Varaiya. On the solutions of a stochastic control system. *SIAM Journal on Control*, 9(3):354–371, 1971.

[29] D. Easley and J. Kleinberg. *Networks, crowds, and markets*, volume 8. Cambridge University Press, Cambridge, 2010.

[30] N. El Karoui, N. Du Huu, and M. Jeanblanc-Picqué. Compactification methods in the control of degenerate diffusions: existence of an optimal control. *Stochastics*, 20(3):169–219, 1987.

[31] N. El Karoui, D. H. Nguyen, and M. Jeanblanc-Picqué. Existence of an optimal markovian filter for the control under partial observations. *SIAM journal on control and optimization*, 26(5):1025–1061, 1988.

[32] W. H. Fleming. The Cauchy problem for degenerate parabolic equations. *Journal of Mathematics and Mechanics*, pages 987–1008, 1964.

[33] W. H. Fleming and M. Nisio. On the existence of optimal stochastic controls. *Journal of Mathematics and Mechanics*, 15(5):777–794, 1966.

[34] D. Helbing and A. Johansson. Quantitative agent-based modeling of human interactions in space and time. In *Proceedings of The Fourth Conference of the European Social Simulation Association (ESSA2007)*, pages 623–637. Citeseer, 2007.

[35] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995.

[36] D. Helbing, I. Farkas, and T. Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(6803):487, 2000.

[37] D. Helbing, I. J. Farkas, P. Molnar, and T. Vicsek. Simulation of pedestrian crowds in normal and evacuation situations. *Pedestrian and evacuation dynamics*, 21(2):21–58, 2002.

[38] D. Helbing, L. Buzna, A. Johansson, and T. Werner. Self-organized pedestrian crowd dynamics: Experiments, simulations, and design solutions. *Transportation science*, 39(1):1–24, 2005.

[39] L. Henderson. The statistics of crowd fluids. *Nature*, 229(5284):381, 1971.

[40] S. Hoogendoorn and P. HL Bovy. Simulation of pedestrian flows by optimal control and differential games. *Optimal Control Applications and Methods*, 24(3):153–172, 2003.

[41] J. J. A. Hosking. A stochastic maximum principle for a stochastic differential game of a mean-field type. *Applied Mathematics & Optimization*, 66(3):415–454, 2012.

[42] Y. Hu and S. Peng. Solution of forward-backward stochastic differential equations. *Probability Theory and Related Fields*, 103(2):273–283, 1995.

[43] 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. *Communications in Information & Systems*, 6(3):221–252, 2006.

[44] M. Huang, P. E. Caines, and R. P. Malhamé. Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ε-Nash equilibria. *IEEE transactions on automatic control*, 52(9):1560–1571, 2007.

[45] R. L. Hughes. The flow of human crowds. *Annual review of fluid mechanics*, 35(1):169–182, 2003.

[46] T. Huschto and S. Sager. Stochastic optimal control in the perspective of the Wiener chaos. In *2013 European Control Conference (ECC)*, pages 3059–3064. IEEE, 2013.

[47] R. Isaacs. Differential Games I, II, III, IV. *Research Memoranda RM-1391, RM-1399, RM-1411, RM-1486*, 1954.

[48] R. Isaacs. *Differential games.* John Wiley and Sons, New York, 1965.

[49] R. Jensen. The maximum principle for viscosity solutions of fully non-linear second order partial differential equations. *Archive for Rational Mechanics and Analysis*, 101(1):1–27, 1988.

[50] B. Jourdain, S. Méléard, and W. A. Woyczynski. Nonlinear SDEs driven by Lévy processes and related PDEs. *Alea*, 4:1–29, 2008.

[51] B. Jovanovic and R. W. Rosenthal. Anonymous sequential games. *Journal of Mathematical Economics*, 17(1):77–87, 1988.

[52] M. Kac. Foundations of kinetic theory. In *Proceedings of The third Berkeley symposium on mathematical statistics and probability*, volume 3, pages 171–197. University of California Press Berkeley and Los Angeles, California, 1956.

[53] D. Kelendžeridze. On the theory of optimal pursuit. In *Doklady Akademii Nauk*, volume 138, pages 529–532. Russian Academy of Sciences, 1961.

[54] H. Kushner. Existence results for optimal stochastic controls. *Journal of Optimization Theory and Applications*, 15(4):347–359, 1975.

[55] H. Kushner. Generalized solutions in optimal stochastic control. In *Differential Games and Control Theory III*, pages 147–165. Marel Dekker, 1977.

[56] H. J. Kushner. On the existence of optimal stochastic controls. *Journal of the Society for Industrial and Applied Mathematics, Series A: Control*, 3(3):463–474, 1965.

[57] A. Lachapelle and M.-T. Wolfram. On a mean field game approach modeling congestion and aversion in pedestrian crowds. *Transportation research part B: methodological*, 45(10):1572–1589, 2011.

[58] D. Lacker. Limit theory for controlled McKean–Vlasov dynamics. *SIAM Journal on Control and Optimization*, 55(3):1641–1672, 2017.

[59] J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. I–Le cas stationnaire. *Comptes Rendus Mathématique*, 343(9):619–625, 2006.

[60] J.-M. Lasry and P.-L. Lions. Jeux à champ moyen. II–Horizon fini et contrôle optimal. *Comptes Rendus Mathématique*, 343(10):679–684, 2006.

[61] J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.

[62] M. Laurière and O. Pironneau. Dynamic programming for mean-field type control. *Journal of Optimization Theory and Applications*, 169(3): 902–924, 2016.

[63] J. Lewin. *Differential games: theory and methods for solving game problems with singular surfaces.* Springer Science & Business Media, 2012.

[64] P.-L. Lions. Cours au Collège de France. *Available at www.college-de-france.fr*, 2007.

[65] B. Maury, A. Roudneff-Chupin, and F. Santambrogio. A macroscopic crowd motion model of gradient flow type. *Mathematical Models and Methods in Applied Sciences*, 20(10):1787–1821, 2010.

[66] B. Maury, A. Roudneff-Chupin, F. Santambrogio, and J. Venel. Handling congestion in crowd motion modeling. *Networds & Heterogeneous Media*, 6(3):485–519, 2011.

[67] H. McKean. A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences*, 56(6):1907–1911, 1966.

[68] J. Nash. *Non-cooperative games.* PhD thesis, Princeton University, 1950.

[69] J. F. Nash. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

[70] J. F. Nash. Non-cooperative games. *Annals of mathematics*, pages 286–295, 1951.

[71] J. Neumann Von and O. Morgenstern. Theory of games and economic behavior. *Princeton University Prress, Princeton*, 1944.

[72] K. Oelschläger. A martingale approach to the law of large numbers for weakly interacting stochastic processes. *The Annals of Probability*, 12 (2):458–479, 1984.

[73] K. Oelschläger. A law of large numbers for moderately interacting diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 69(2):279–322, 1985.

[74] M. J. Osborne and A. Rubinstein. *A course in game theory*. MIT press, 1994.

[75] D. G. Pearce. Rationalizable strategic behavior and the problem of perfection. *Econometrica: Journal of the Econometric Society*, pages 1029–1050, 1984.

[76] S. Peng. A general stochastic maximum principle for optimal control problems. *SIAM Journal on control and optimization*, 28(4):966–979, 1990.

[77] L. S. Pontryagin, E. Mishchenko, V. Boltyanskii, and R. Gamkrelidze. *The mathematical theory of optimal processes*. Wiley, 1962.

[78] E. Roxin. The existence of optimal controls. *The Michigan Mathematical Journal*, 9(2):109–119, 1962.

[79] H. M. Soner, N. Touzi, and J. Zhang. Wellposedness of second order backward SDEs. *Probability Theory and Related Fields*, 153(1-2):149–190, 2012.

[80] A. I. Subbotin. A generalization of the basic equation of the theory of differential games. In *Doklady Akademii Nauk*, volume 254, pages 293–297. Russian Academy of Sciences, 1980.

[81] A.-S. Sznitman. Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX—1989*, pages 165–251. Springer, 1991.

[82] H. Tembine. Uncertainty quantification in mean-field-type teams and games. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 4418–4423. IEEE, 2015.

[83] H. Tembine. Mean-field-type games. *AIMS Math*, 2(4):706–735, 2017.

[84] U. Weidmann. Transporttechnik der fußgänger: transporttechnische eigenschaften des fußgängerverkehrs, literaturauswertung. *IVT Schriftenreihe*, 90, 1993.

[85] J. Yong. Forward-backward stochastic differential equations with mixed initial-terminal conditions. *Transactions of the American Mathematical Society*, 362(2):1047–1096, 2010.

[86] J. Yong and X. Y. Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*, volume 43. Springer Science & Business Media, 1999.

[87] L. Young. *Calculus of Variations and Control Theory.* W.B. Saunders, 1969.

Part II: Scientific Papers

Paper A

Mean-field type modeling of nonlocal crowd aversion in pedestrian crowd dynamics

Mean-field type modeling of nonlocal crowd aversion in pedestrian crowd dynamics

by

Alexander Aurell and Boualem Djehiche

Published in
SIAM Journal on Control and Optimization
Volume 56, Number 1 (2018), 434-455.

Abstract

We extend the class of pedestrian crowd models introduced by Lachapelle and Wolfram (2011) to allow for nonlocal crowd aversion and arbitrarily but finitely many interacting crowds. The new crowd aversion feature grants pedestrians a 'personal space' where crowding is undesirable. We derive the model from a particle picture and treat it as a mean-field type game. Solutions to the mean-field type game are characterized via a Pontryagin-type maximum principle. The behavior of pedestrians acting under nonlocal crowd aversion is illustrated by a numerical simulation.

Keywords: crowd dynamics, crowd aversion, mean-field approximation, interacting populations, optimal control, mean-field type game

A.1 Introduction

When a pedestrian is walking through a crowd she chooses her path based not only on her desired final destination but also takes the movement of other surrounding pedestrians into account. The bullet points below are stated in [18] as typical traits of pedestrian behavior.

- Will to reach specific targets. Pedestrians experience a strong interaction with the environment.
- Repulsion from other individuals. Pedestrians may agree to deviate from their preferred path, looking for free surrounding room.
- Deterministic if the crowd is sparse, partially random if the crowd is dense.

These properties appear in classical particle models. Other authors advocate smart particle models with decision-based dynamics. In [18] some fundamental differences between classical and smart particle models are outlined. We list a few of them in Table 1.

CLASSICAL	
Robust	interaction only through collisions
Blindness	dynamics ruled by inertia
Local	interaction is pointwise
SMART	
Fragile	avoidance of collisions and obstacles
Vision	dynamics ruled at least partially by decision
Nonlocal	interaction at a distance

Table 1

A smart particle model lets the pedestrians decide where to walk, with what speed, etc. The choice is based on some rule that takes the available information into account, such as the positioning and movement of other pedestrians. Although more realistic, this approach has some complications. If pedestrian i moves, all pedestrians accessing information on i's state might have to adapt their movements. The large number of connections where information is exchanged within a crowd makes such models difficult to solve in practice, due to their high computational complexity.

The mean-field approach to modeling crowd aversion and congestion for pedestrians was introduced in [15]. The pedestrians are treated as particles following decision-based dynamics and optimize their path by avoiding densely crowded areas. Crowd aversion describes the avoidance of high density areas whereas congestion describes motion hindered by high density. The theory of mean-field games originates from the independent works of Lasry-Lions [16] and Huang-Caines-Malhamé [10]. The cost considered in this early work is not of congestion type, i.e. the energy penalization is independent of the density. The framework was extended to several populations on the torus in [9] and to several populations on a bounded domain with reflecting boundaries in [8], with further studies in [4, 6]. Mean field games with a cost of congestion type was introduced by P-L. Lions in a lecture series 2011 [17]. Congestion has also been studied in the mean-field type setting. In [1] the finite horizon case is considered. In [3, 2] the authors prove existence and uniqueness of weak solutions characterized by an optimization approach based on duality, and propose a numerical method for mean-field type control based on this result for the case of local congestion.

Turning to the crowd aversion model of this paper, a pedestrian with position $X^{i,N}$ in a crowd of N pedestrians controls her velocity such that her risk measure, $J^{i,N}$, is minimized over a finite time horizon $[0, T]$. The risk measure penalizes proximity to others, energy waste and failure to reach a target area. In this paper we advocate for the use of the following nonlocal contribution to the risk measure, reflecting a crowd averse behavior,

$$\mathbb{E}_N \left[\int_0^T \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} \phi_r \left(X_t^{i,N} - X_t^{j,N} \right) dt \right]. \tag{1}$$

The ‘personal space’ of a typical pedestrian is modeled by the function ϕ_r and $X_t^{i,N} - X_t^{j,N}$ is the distance between two pedestrians at time t. The personal space has support within a ball of radius r so for positive r, (1) is a weighted average of the crowding within the personal space and the pedestrian is not effected by crowding outside it. Connecting to the terminology in Table 1, the case of positive r will be referred to as *nonlocal crowd aversion.* In the limit $r \to 0$ the personal space shrinks to a point and only pointwise crowding, that is collisions, will effect the pedestrian. This will be referred to as *local crowd aversion.*

In emergency situations it is often in the interest of all pedestrians to get to a certain place, such an exit. In evacuation planning and crowd management at mass gatherings, it is in the interest of the planner to control the crowd along paths and towards certain areas. Common to such situations is the conflict between attraction to said locations and repulsive interactions in the crowd. Pedestrians acting under nonlocal crowd aversion will order themselves more densely in such places compared to pedestrians acting under local crowd aversion. This effect is caused by the larger personal space, the nonlocal crowd aversion term (1) is an average over a bigger set hence allowing for higher densities in attractive areas. Higher densities will in turn allow for more effective emergency planning when designing for example escape routes. The numerical simulation in the end of this paper confirms this effect. The pedestrians are allowed to move freely, but the observed effect will become even more beneficial for a planner when introducing an environment for the pedestrians to interact with. In reality, crowd management is often done by the strategic placement of obstacles such as pillars and walls. In this paper, we observe that the pedestrians acting under nonlocal crowd aversion travel at an overall lower risk than their local counterpart. This suggests that a crowd with nonlocal crowd averse behavior could potentially move at a higher velocity than its local counterpart which allows for faster and more successful evacuations.

In [15] the mean-field optimal control is characterized through a matching argument. This control is an approximate Nash equilibrium for the crowd. It is, for each pedestrian, the best response to the movement of the rest of the crowd. Furthermore, two crowds are considered where each pedestrian has crowd-specific preferences such as the target location and crowd aversion preference. The authors set up a mean-field game and show that it is equivalent to an optimal control problem. In this paper, we look at the crowd from the bird's-eye view of an evacuation planner. We seek a 'simultaneous' optimal strategy for all the pedestrians involved in the crowd through a mean-field type control approach for the single-crowd case and a mean-field type game approach for the multi-crowd case.

The contributions of this paper are the following. We identify a particle model that is approximated by the mean-field model for crowd aversion proposed in [15]. This gives us insights into how the interaction between pedestrians in the crowd effects the mean-field model and reveals that the crowd of [15] has a locally crowd averse behavior. Our second contribution is a relaxation of the locality of the pedestrian model by allowing for interaction between pedestrians at a distance. Instead of only interacting with other pedestrians through collisions, each pedestrian is here given a personal space where she dislikes crowding. This conceptual change is realistic since pedestrians do not need to be in physical contact to interact. As discussed above, the suggested nonlocal crowd aversion model allows for the following desirable features:

- Higher densities in target areas such as exits or escape routes where the pedestrians have to choose between more crowding and not reaching the target.
- Lower risk, which implies a potential increase in pedestrian velocity allowing for faster exits and a larger flow of people, a very useful feature in the design of evacuation strategies.

Finally, we generalize the model to allow for an arbitrary number of interacting crowds. This multi-crowd scenario is treated as a mean-field type game and is linked to an optimal control problem, for which we prove a sufficient maximum principle.

The paper is organized as follows. After a short section of preliminaries, we consider the single-crowd case in Section A.3. In Section A.4, the multi-crowd case is studied. The results derived in Section A.3 generalize to an arbitrary finite number of interacting crowds and we derive sufficient conditions for a solution with the maximum principle. An example that highlights the difference between local and nonlocal crowd aversion is solved numerically in Section A.5. For the sake of clarity, all technical proofs are moved to an appendix.

A.2 Preliminaries

Given a general Polish space $\mathcal{S}$, let $\mathcal{P}(\mathcal{S})$ denote the space of probability measures on $\mathcal{B}(\mathcal{S})$. For an element $s \in \mathcal{S}$, the Dirac measure on s is an element of $\mathcal{P}(\mathcal{S})$ and will be denoted by δ_s. Let $\mathcal{P}(\mathcal{S})$ be equipped with the topology of weak convergence of probability measures. A metric that induces this topology is the bounded 1-Lipschitz metric,

$$d_{\mathcal{P}(\mathcal{S})}(\mu, \nu) := \|\mu - \nu\|_1 = \sup_{f \in L_1} \langle \mu, f \rangle - \langle \nu, f \rangle ,$$

where L_1 is the set of real-valued functions on $\mathcal{S}$ bounded by 1 and with Lipschitz coefficient 1. With his metric, $\mathcal{P}(\mathcal{S})$ is a Polish space. The space of probability measures on $\mathcal{B}(\mathcal{S})$ with finite second moments will be denoted by $\mathcal{P}_2(\mathcal{S})$,

$$\mathcal{P}_2(\mathcal{S}) := \left\{ \nu \in \mathcal{P}(\mathcal{S}) \ : \ \exists\, s_0 \in \mathcal{S} \text{ that satisfies } \int_{\mathcal{S}} d_{\mathcal{S}}(s, s_0)^2 \nu(ds) < \infty \right\}.$$

Equipped with the topology of weak convergence of measures and convergence of second moments, $\mathcal{P}_2(\mathcal{S})$ is a Polish space. A compatible complete metric is the square Wasserstein metric $d_{\mathcal{P}_2(\mathcal{S})}$, for which the following inequalities will be useful. For all $s_i, \widetilde{s}_i \in \mathcal{S}$ and for all $N \in \mathbb{N}$,

$$d^2_{\mathcal{P}_2(\mathcal{S})}\left(\frac{1}{N} \sum_{i=1}^{N} \delta_{s_i}, \frac{1}{N} \sum_{i=1}^{N} \delta_{\widetilde{s}_i} \right) \leq \frac{1}{N} \sum_{i=1}^{N} d_{\mathcal{S}}(s_i, \widetilde{s}_i)^2. \tag{2}$$

For two random variables X and $\widetilde{X}$ with distribution ν and $\widetilde{\nu}$, respectively,

$$d^2_{\mathcal{P}_2(\mathcal{S})}(\nu, \widetilde{\nu}) \leq \mathbb{E}\left[|X - \widetilde{X}|^2 \right].$$

Let $T > 0$ be a finite time horizon and let $\mathbb{R}^d$, $d \in \mathbb{N}$, be equipped with the Euclidean norm. Let $\mathcal{M}$ and $\mathcal{M}_2$ be the spaces of continuous functions on $[0, T]$ with values in $\mathcal{P}(\mathbb{R}^d)$ and $\mathcal{P}_2(\mathbb{R}^d)$ respectively,

$$\mathcal{M} := C([0, T]; \mathcal{P}(\mathbb{R}^d)), \qquad \mathcal{M}_2 := C([0, T]; \mathcal{P}_2(\mathbb{R}^d)).$$

Equipped with the uniform metrics $d_{\mathcal{M}}$ and $d_{\mathcal{M}}$,

$$\begin{aligned} d_{\mathcal{M}}(m, m') &:= \sup_{t \in [0,T]} d_{\mathcal{P}(\mathbb{R}^d)}(m_t, m'_t), \\ d_{\mathcal{M}_2}(m, m') &:= \sup_{t \in [0,T]} d_{\mathcal{P}_2(\mathbb{R}^d)}(m_t, m'_t), \end{aligned}$$

$\mathcal{M}$ and $\mathcal{M}_2$ are Polish spaces. The mathematical results stated above can be found in [20, Chapter 2] and [11, Chapter 14].

Let A be a compact subset of $\mathbb{R}^d$. Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, denote by $\mathcal{A}$ the set of A-valued $\mathbb{F}$-adapted processes such that

$$\mathbb{E}\left[\int_0^T |a_t|^2 dt \right] < \infty.$$

An element of $\mathcal{A}$ will be called an *admissible control.* From the context, it will be clear which stochastic basis the notation $\mathcal{A}$ is referring to.

Given a vector $x = (x^1, \ldots, x^N)$ in the product space $\mathcal{S}^N$ and an element $y \in \mathcal{S}$, we let

$$\begin{aligned} x^{-i} &:= (x^1, \ldots, x^{i-1}, x^{i+1}, \ldots, x^N), \\ (y, x^{-i}) &:= (x^1, \ldots, x^{i-1}, y, x^{i+1}, \ldots, x^N). \end{aligned}$$

Furthermore, the law of any random quantity X will be denoted by $\mathcal{L}(X)$ and any index set of the form $\{1, \ldots, N\}$ will be denoted by $[\![N]\!]$.

A.3 Single-crowd model for crowd aversion

A.3.1 The particle picture

Let $(\Omega_N, \mathcal{F}^N, \mathbb{F}^N, \mathbb{P}_N)$ be a complete filtered probability space for each $N \in \mathbb{N}$. The filtration $\mathbb{F}^N$ is right-continuous and augmented with $\mathbb{P}_N$-null sets. It carries the independent d-dimensional $\mathbb{F}^N$-Wiener processes $W^{1,N}, \ldots, W^{N,N}$. Let, for each $i \in [\![N]\!]$, the $\mathcal{F}_0^N$-measurable $\mathbb{R}^d$-valued random variable $\xi^{i,N}$ be square-integrable and independent of $(W^{1,N}, \ldots, W^{N,N})$. Given a vector of admissible controls, $\bar{a}^N = (a^{1,N}, \ldots, a^{N,N}) \in \mathcal{A}^N$, consider the system

$$dX_t^{i,N} = b(t, X_t^{i,N}, a_t^{i,N})dt + \sigma(t, X_t^{i,N})dW_t^{i,N}, \quad X_0^{i,N} = \xi^{i,N}, i \in [\![N]\!]. \tag{3}$$

Proposition 1. *Assume that*

(A1) $b : [0,T] \times \mathbb{R}^d \times A \to \mathbb{R}^d$ *and* $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d \times d}$ *are continuous in all arguments.*

(A2) For all $x_1, x_2 \in \mathbb{R}^d$ *and* $a_1, a_2 \in A$*, there exists a constant* $K > 0$ *independent of* (t, x_1, x_2, a_1, a_2) *such that*

$$\begin{aligned} |b(t, x_1, a_1) - b(t, x_2, a_2)| &\leq K(|x_1 - x_2| + |a_1 - a_2|), \\ |\sigma(t, x_1) - \sigma(t, x_2)| &\leq K|x_1 - x_2|, \\ |b(t, x_1, a_1)| + |\sigma(t, x_1)| &\leq K(1 + |x_1| + |a_1|). \end{aligned}$$

Under these assumptions, (3) *has a unique strong solution in the sense that*

$$
\begin{aligned}
&X_0^{i,N} = \xi^{i,N}, \\
&\int_0^t \left|b(s, X_s^{i,N}, a_s^{i,N})\right| + \left|\sigma(s, X_s^{i,N})\right|^2 ds < \infty, \quad t \in [0,T],\ \mathbb{P}\text{-}a.s. \\
&X_t^{i,N} = \xi^{i,N} + \int_0^t b(s, X_s^{i,N}, a_s^{i,N})ds + \int_0^t \sigma(s, X_s^{i,N})dW_s^{i,N}, \quad t \in [0,T].
\end{aligned}
$$

Furthermore, the strong solution $X^{i,N}$ *satisfies the estimate*

$$
\mathbb{E}_N\left[\sup_{s\in[0,t]} |X_s^{i,N}|^2\right] \leq K_t\left(1 + \mathbb{E}_N\left[|\xi^{i,N}|^2\right]\right)
$$

for all $t \in [0,T]$, *for all* $i \in [\![N]\!]$ *and for some positive constant* K_t *depending only on* t.

Proof. A proof can be found in [24, Chapter 1, Theorem 6.16]. Note that K_t is independent of $a^{i,N}$ by compactness of A. □

The process $X^{i,N}$ models the motion of an individual in a crowd of N pedestrians, from now on called an N-crowd, who partially controls her velocity through the control $a^{i,N}$. Since her control is adapted to the full filtration $\mathbb{F}^N$, the model allows for the pedestrian to take every movement in the crowd into account. Her motion is also influenced by external forces, such as the random disturbance driven by $W^{i,N}$. The motion of the pedestrian may be modeled more generally than above by introducing an explicit weak interaction in the drift [10], such as

$$
dX_t^{i,N} = \frac{1}{N}\sum_{j=1}^N \widetilde{b}(t, X_t^{i,N}, a_t^{i,N}, X_t^{j,N})dt + \sigma(t, X_t^{i,N})dW_t^{i,N}.
$$

It is also possible to let a common disturbance effect all pedestrians [13], to model for example evacuations during an earthquake, a fire, a tsunami etc.

Individual i evaluates the state of the N-crowd, given by the control vector $\bar{a}^N = (a^{1,N}, \ldots, a^{N,N})$, according to her measure of risk

$$
\begin{aligned}
&J_r^{i,N}(\bar{a}^N) \\
&\quad := \mathbb{E}_N\left[\int_0^T \left(\frac{1}{2}|a_t^{i,N}|^2 + \int_{\mathbb{R}^d} \phi_r(X_t^{i,N} - y)\mu_t^{-i,N}(dy)\right)dt + \Psi(X_T^{i,N})\right],
\end{aligned}
$$

where $X^{1,N}, \dots, X^{N,N}$ solves (3) given $\bar{a}^N$ and $\mu_t^{-i,N}$ is the empirical measure of $X_t^{-i,N}$. The region where crowding has an influence on the pedestrian's choice of control, her 'personal space', is ideally modeled by a normalized indicator function,

$$\mathbb{I}_r(x) := \begin{cases} \mathrm{Vol}(B_r)^{-1}, & x \in B_r, \\ 0, & x \notin B_r, \end{cases}$$

where $B_r \subset \mathbb{R}^d$ is the ball with radius $r > 0$ centered at the origin and $\mathrm{Vol}(B_r)$ is its volume. The term

$$\int_{\mathbb{R}^d} \mathbb{I}_r(X_t^{i,N} - y)\mu_t^{-i,N}(dy)$$

then represents the number of pedestrians around $X_t^{i,N}$ within a distance less than r at time t [22]. To simplify the calculations we will use a smoothed version of $\mathbb{I}_r$. Let γ_δ be a mollifier, i.e. $\gamma_\delta(x) := \gamma(x/\delta)/\delta$ where γ is a smooth symmetric probability density with compact support. For a fixed $\delta > 0$, we set

$$\phi_r(x) := (\gamma_\delta * \mathbb{I}_r)(x). \tag{4}$$

For convergence estimates later in this section, we assume that the final cost Ψ satisfies the following condition.

(A3) For all $x_1, x_2 \in \mathbb{R}^d$ there exists a constant $K > 0$ independent of (x_1, x_2) such that

$$|\Psi(x_1) - \Psi(x_2)| \le K|x_1 - x_2|.$$

The interpretation of the risk measure is the following. The first term penalizes energy usage whereas the second term penalizes paths through densely crowded areas. The final cost penalizes deviations from specific target regions. Typically the final cost takes large values everywhere except in areas where the pedestrians want to end up, places like meeting points, evacuation doors, etc.

A.3.2 The mean-field type control problem

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be a complete filtered probability space such that the filtration is right-continuous and augmented with $\mathbb{P}$-null sets. Let $\mathbb{F}$ carry a Wiener process W and let ξ be an $\mathcal{F}_0$-measurable and square-integrable $\mathbb{R}^d$-valued random variable independent of W. Given a control $a \in \mathcal{A}$, the mean-field type dynamics is

$$dX_t = b(t, X_t, a_t)dt + \sigma(t, X_t)dW_t, \quad X_0 = \xi. \tag{5}$$

By Proposition 1 there exists a unique strong solution to (5). The mean-field type risk measure is given by

$$J_r(a) = \mathbb{E}\left[\int_0^T \frac{1}{2}|a_t|^2 + \int_{\mathbb{R}^d} \phi_r(X_t - y)\mu_{X_t}(dy)dt + \Psi(X_T)\right]. \tag{6}$$

where μ_{X_t} is the distribution of X_t.

Remark 2. *The difference between a mean-field type control problem and a mean-field game is that in general mean-field games can be reduced to a standard control problem and an equilibrium while a mean-field type control problem is a nonstandard control problem [5, 7]. The matching procedure to find the fixed point (equilibrium) for a mean-field game is pedagogically described as follows [10, 16]:*

(i) Fix a deterministic function $\mu_t : [0, T] \to \mathcal{P}_2(\mathbb{R}^d)$.

(ii) Solve the stochastic control problem

$$\hat{a} = \arg\min_{a \in \mathcal{A}} \mathbb{E}\left[\int_0^T \frac{1}{2}|a_t|^2 + \int_{\mathbb{R}^d} \phi_r(X_t - y)\mu_t(dy)dt + \Psi(X_T)\right],$$

where X is the dynamics corresponding to a.

(iii) Determine the function $\hat{\mu}_t : [0, T] \to \mathcal{P}_2(\mathbb{R}^d)$ such that $\hat{\mu}_t = \mathcal{L}(\hat{X}_t)$ for all $t \in [0, T]$ where $\hat{X}$ is the dynamics corresponding to the optimal control $\hat{a}$.

In the mean-field type control setting the measure-valued process $(\mu_{X_t})_{t \in [0,T]}$ is not considered to be a separate variable but determined by the input control process.

A.3.3 Convergence of the state process

Let the initial data $\xi^{1,N}, \dots, \xi^{N,N}$ satisfy the following assumptions:

(B1) $\sup_{N\in\mathbb{N}} \mathbb{E}_N \left[\frac{1}{N}\sum_{i=1}^N |\xi^{i,N}|^2\right] < \infty$ for all $i \in [\![N]\!]$;

(B2) $(\xi^{1,N}, \dots, \xi^{N,N})$ is exchangeable for all $N \in \mathbb{N}$;

(B3) $\lim_{N\to\infty} \mathcal{L}\left(\frac{1}{N}\sum_{i=1}^N \delta_{\xi^{i,N}}\right) = \delta_{\mu_0}$ in $\mathcal{P}(\mathcal{P}_2(\mathbb{R}^d))$.

Under (B1)-(B3) the sequence $(\xi^{i,N})_{N\in\mathbb{N}}$ is tight and a subsequence can be extracted that converges in distribution to a μ_0-distributed random variable, from now on denoted by ξ. We make the following assumption about the controls:

(B4) The controls are of feedback form, $a_t^{i,N}(\omega) = a^N(t, X_t^{i,N}(\omega))$, where each a^N is an A-valued deterministic function and a^N converge uniformly to a as $N \to \infty$. Furthermore,

$$\sup_{N\in\mathbb{N}} \mathbb{E}_N \left[\int_0^T |a^N(t, X_t^{i,N})|^2\right] < \infty, \quad \forall\, i \in [\![N]\!].$$

Remark 3. *Assumption (B4) implies that, while the paths of the pedestrians in the N-crowd may differ, they are outcomes from a symmetric joint probability distribution. By exchangeability of $(\xi^{i,N}, W^{i,N})_{i=1}^N$,*

$$(a^N(t, X_t^{i,N}))_{i=1}^N \stackrel{d}{=} (a^N(t, X_t^{\pi(i),N}))_{i=1}^N \tag{7}$$

for all permutations π of $[\![N]\!]$. The interpretation is that we cannot distinguish between pedestrians in the crowd, the pedestrians are anonymous.

Proposition 4. *If μ^N is the empirical measure of $X^{1,N}, \dots, X^{N,N}$, the solution of* (3) *given a^N, then $\{\mathcal{L}(\mu^N), N \in \mathbb{N}\}$ is tight in $\mathcal{P}(\mathcal{M}_2)$.*

Proof. The empirical measures are elements of $\mathcal{M}_2$ by Proposition 1 together with (B1) and (B2). The proof of tightness in the case of uncontrolled diffusions is found in [19]. The introduction of a control does not change the situation. $\square$

Recall that a sequence $\{X_n\}$ of random variables converges weakly to X in a Polish space if and only if $\{X_n\}$ is tight and every convergent subsequence of $\{X_n\}$ converges to X. The tightness of the empirical measures implies that along a converging subsequence, μ^N converges weakly to the measure-valued process μ that for all $f \in C_b^2(\mathbb{R}^d)$ satisfies

$$\langle \mu_t, f\rangle - \langle \mu_0, f\rangle = \int_0^t \left\langle \mu_s, b(s,\cdot,a(s,\cdot)) \cdot \nabla f + \frac{1}{2}\mathrm{Tr}\left[\sigma\sigma^T(s,\cdot)\Delta f\right] \right\rangle ds. \quad (8)$$

Since the strong solution of (5) is unique, the weak solution is also unique [23] which is equivalent to uniqueness of solutions to (8) [12]. We have the following result.

Theorem 5. *Let X^i, $i \in \mathbb{N}$, be independent copies of the strong solution of* (5)*. Under assumptions (A1)-(B4), $X^{i,N}$ converges weakly to X^i as $N \to \infty$.*

Proof. Applying Sznitman's propagation of chaos theorem [21], the result follows by the weak convergence of μ^N to the deterministic measure μ. □

A.3.4 Convergence of the risk measure

From the previous section we know that $X^{i,N}$, the strong solution of (3), converges weakly to X, the strong solution of (5), and we know that μ_t^N converges weakly to μ_{X_t}. Applying (2), we have that $d_{\mathcal{P}_2(\mathbb{R}^d)}(\mu_t^{-i,N}, \mu_t^N) \leq 2/N$, so $\mu_t^{-i,N}$ converges weakly to μ_{X_t} as well. By Skorokhod's representation theorem [11, Theorem 3.30] we can represent (up to distribution) all the random variables mentioned above in a common probability space $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ where they converge $\widetilde{\mathbb{P}}$-almost surely. This allows us to write

$$\begin{aligned}
&|J_r^{i,N}(a^N) - J_r(a)| \\
&\leq \mathbb{E}^{\widetilde{\mathbb{P}}}\Bigg[\int_0^T \Bigg\{ \left|\frac{1}{2}|a^N(t, X_t^{i,N})|^2 - \frac{1}{2}|a(t,X_t)|^2\right| \\
&\qquad + \left|\int_{\mathbb{R}^d} \phi_r(X_t^{i,N} - y)\mu_t^{-i,N}(dy) - \int_{\mathbb{R}^d} \phi_r(X_t - y)\mu_{X_t}(dy)\right|\Bigg\} dt \\
&\qquad\qquad + \left|\Psi(X_T^{i,N}) - \Psi(X_T)\right|\Bigg].
\end{aligned}$$

By compactness of A, the continuous mapping theorem, (B4), and the dominated convergence theorem,

$$\lim_{N\to\infty} \mathbb{E}^{\widetilde{\mathbb{P}}}\left[\int_0^T \left|\frac{1}{2}|a^N(t,X_t^{i,N})|^2 - \frac{1}{2}|a(t,X_t)|^2\right|\right] = 0.$$

By (A3), Proposition 1, and the dominated convergence theorem we have that $\mathbb{E}^{\widetilde{\mathbb{P}}}[|\Psi(X_T^{i,N}) - \Psi(X_T)|] = 0$ as $N \to \infty$. Note that

$$\begin{aligned}
&\mathbb{E}^{\widetilde{\mathbb{P}}}\left[\int_0^T \left|\int_{\mathbb{R}^d} \phi_r(X_t^{i,N} - y)\mu_t^{-i,N}(dy) - \int_{\mathbb{R}^d} \phi_r(X_t - y)\mu_{X_t}(dy)\right| dt\right] \\
&\quad\le \mathbb{E}^{\widetilde{\mathbb{P}}}\left[\int_0^T \left|\int_{\mathbb{R}^d} \phi_r(X_t^{i,N} - y)\mu_t^{-i,N}(dy) - \int_{\mathbb{R}^d} \phi_r(X_t^{i,N} - y)\mu_{X_t}(dy)\right| dt\right] \\
&\qquad + \mathbb{E}^{\widetilde{\mathbb{P}}}\left[\int_0^T \left|\int_{\mathbb{R}^d} \phi_r(X_t^{i,N} - y)\mu_{X_t}(dy) - \int_{\mathbb{R}^d} \phi_r(X_t - y)\mu_{X_t}(dy)\right| dt\right].
\end{aligned}$$

As $N \to \infty$, the first term on the right hand side tends to zero by the definition of weak convergence while the second tends to zero by the Continuous Mapping Theorem and Dominated Convergence. We have proved the following result.

Theorem 6. *Let $a \in \mathcal{A}$ and $a^N = (a,\ldots,a) \in \mathcal{A}^N$, then $J_r^{i,N}(a^N) = J_r(a) + \varepsilon_N$ where $\lim_{N\to\infty} \varepsilon_N = 0$.*

A.3.5 Solutions to the N-crowd model and the MFT control problem

The notion of solutions of the the N-crowd model (N-1) and the mean-field type control problem (MFT-1) for crowd aversion will now be defined.

Definition 7 (Solution to N-1)**.** *Let $\hat{a}^N = (\hat{a},\ldots,\hat{a}) \in \mathcal{A}^N$ for some fixed $\hat{a} \in \mathcal{A}$ and let $a^N = (a,\ldots,a) \in \mathcal{A}^N$ for an arbitrary strategy $a \in \mathcal{A}$. Then $\hat{a}^N$ is a solution to N-1 if*

$$J_r^{i,N}(\hat{a}^N) \le J_r^{i,N}(a^N), \quad \forall a \in \mathcal{A},\ \forall i \in [\![N]\!].$$

If, for a given $\varepsilon > 0$, $\hat{a}$ satisfies

$$J_r^{i,N}(\hat{a}^N) \le J_r^{i,N}(a^N) + \varepsilon, \quad \forall a \in \mathcal{A},\ \forall i \in [\![N]\!],$$

then $\hat{a}^N$ is an ε-solution to N-1.

Definition 8 (Solution to MFT-1)**.** *If $\hat{a} \in \mathcal{A}$ satisfies*

$$J_r(\hat{a}) \leq J_r(a), \quad \forall a \in \mathcal{A},$$

then $\hat{a}$ is a solution to MFT-1.

The following result motivates the use of MFT-1 as an approximation to N-1. It confirms that we can construct an approximate solution to N-1 using a solution to MFT-1.

Theorem 9. *If $\hat{a}$ solves MFT-1, then $\hat{a}^N = (\hat{a}, \dots, \hat{a})$ is a ε_N-solution, where $\varepsilon_N \to 0$ as $N \to \infty$, to N-1 among feedback strategies.*

Proof. The proof follows straight away by Theorem 6. □

Remark 10. *It is known that the solution of a mean-field game corresponds to an approximate Nash equilibrium for N-1, see e.g. [10, 16]. To the best of our knowledge, this has not been shown to be true for solutions to mean-field type control problems. Theorem 9 has the following interpretation: a mean-field type optimal control induces an approximate solution for the N-crowd if the crowd consists homogeneous pedestrians and thus a representative pedestrian determines the control of all. This was in fact visible already in Theorem 6.*

A.3.6 Deterministic version of MFT-1

We want to present results in a setting similar to [15] to highlight the differences between the models. To do this, we make the assumption that μ_{X_t} has a density $m_X(t, \cdot)$ for all $t \in [0, T]$. An example of sufficient conditions for the existence is bounded drift and diffusion [19]. Under this assumption, we may rewrite (5)-(6) into a deterministic problem for m_X. Furthermore, an admissible control can not be stochastic in the deterministic problem formulation. The full stochastic problem will be analyzed in future work. We have a new definition of an admissible control.

Definition 11 ($\mathcal{A}_d$)**.** *A square-integrable deterministic function $a : [0, T] \times \mathbb{R}^d \to A$ will be called an admissible control for the deterministic problem and the set of such functions is denoted by $\mathcal{A}_d$.*

By (8) the density m_X satisfies

$$\int_{\mathbb{R}^d} f(x)m_X(t,x)dx - \int_{\mathbb{R}^d} f(x)m_X(0,x)dx =$$
$$\int_0^t \int_{\mathbb{R}^d} \Big(b(s,x,a(s,x)) \cdot \nabla f(x) + \frac{1}{2}\mathrm{Tr}\left[\sigma\sigma^T(s,x)\nabla^2 f(x)\right]\Big) m_X(s,x)dsdx,$$

for all $f \in C_b^2(\mathbb{R}^d)$ and for all $t \in [0,T]$, hence it is a weak solution to

$$\begin{cases} \dfrac{\partial m_X}{\partial t}(t,x) = \dfrac{1}{2}\mathrm{Tr}\left[\nabla^2\sigma\sigma^T m_X(t,x)\right] - \nabla \cdot (b(t,x,a(t,x))m_X(t,x)), \\ m_X(0,x) = \text{ density of } \mu_0. \end{cases}$$

We arrive to a deterministic version of MFT-1 (dMFT-1),

$$\begin{cases} \min\limits_{a \in \mathcal{A}_d} & J_r^{\det}(a) \\ \text{s.t.} & \dfrac{\partial m_X}{\partial t}(t,x) = \dfrac{1}{2}\mathrm{Tr}\left[\nabla^2\sigma\sigma^T m_X(t,x)\right] - \nabla \cdot (b(t,x,a(t,x))m_X(t,x)), \\ & m_X(0,x) = \text{ density of } \mu_0. \end{cases}$$

where

$$J_r^{\det}(a) := \int_{\mathbb{R}^d} \int_0^T \Bigg\{ \frac{1}{2}|a(t,x)|^2 m_X(t,x) +$$
$$\left(\int_{\mathbb{R}^d} \phi_r(x-y)m_X(t,y)dy\right) m_X(t,x)\Bigg\} dtdx + \int_{\mathbb{R}^d} \Psi(x)m(T,x)\Bigg] dx.$$

Remark 12. *Note that ϕ_r converges weakly to δ_0 as $r \to 0$. In this limit, the risk measure tends to*

$$J_0^{det}(a) = \int_{\mathbb{R}^d} \int_0^T \frac{1}{2}|a(t,x)|^2 m_X(t,x) + m_X(t,x)^2 dt + \Psi(x)m(T,x)dx,$$

which is exactly the risk analyzed in the pedestrian crowd model of [15]. Clearly this case corresponds to a situation where the pedestrian will only react to how likely it is for her to 'bump' into other pedestrians. In the case of positive r, the pedestrian is effected by crowding within a personal space of nonzero range and reacts to the level of the density within this range. This is the distinction between locally and nonlocally crowd averse behavior.

A.4 Multi-crowd model for crowd aversion

A.4.1 The particle picture

In this section, crowd averse behavior between several crowds is introduced. The crowds are allowed to differ in their opinions on target areas and/or the level of crowd aversion. This inhomogeneity is introduced in the risk measure. Let the setup be as in Section A.3, except now $\mathbb{F}^N$ carries NM independent $\mathbb{F}^N$-Wiener processes $W^{i,j,N}$, $i \in [\![N]\!]$, $j \in [\![M]\!]$ and there is for all $i \in [\![N]\!]$, $j \in [\![M]\!]$ a square-integrable $\mathcal{F}_0^N$-measurable $\mathbb{R}^d$-valued random variable $\xi^{i,j,N}$ independent of all the Wiener processes. Given NM admissible controls $a^{i,j,N}$, consider the system

$$\begin{cases} dX_t^{i,j,N} = b(t, X_t^{i,j,N}, a_t^{i,j,N})dt + \sigma(t, X_t^{i,j,N})dW_t^{i,j,N}, \\ X_0^{i,j,N} = \xi^{i,j,N}, \qquad i \in [\![N]\!],\ j \in [\![M]\!]. \end{cases} \tag{9}$$

In view of Proposition 1 there exists a unique strong solution to (9). Pedestrian i in crowd j evaluates $\mathbf{a} := (a^{i,j,N})_{i,j}$ according to its individual risk measure

$$\begin{aligned} &J_{r,\Lambda}^{i,j,N}(\mathbf{a}) \\ &\quad := \mathbb{E}_N \left[\int_0^T \frac{1}{2}|a^{i,j,N}|^2 + \int_{\mathbb{R}^d} \phi_r(X_t^{i,j,N} - y)\widetilde{\nu}_{t,\Lambda}^{j,N}(dy)dt + \Psi_j(X_T^{i,j,N}) \right], \end{aligned}$$

where

$$\widetilde{\nu}_{t,\Lambda}^{j,N} := \sum_{k=1}^M \lambda_{jk} \frac{1}{N} \sum_{l=1}^N \delta_{X_t^{l,k,N}},$$

λ_{jk} are bounded and non-negative real numbers, and $\Lambda = (\lambda_{jk})_{jk}$. The weights λ_{jk} quantify the crowd aversion preferences in the model. If λ_{jk} is high, pedestrians in crowd j pay a high price for being close to pedestrians in crowd k. If λ_{jk} is zero, pedestrians in crowd j are indifferent to the positioning of pedestrians in crowd k. Note that if $\lambda_{jk} = 1$ for $j = k$ and 0 otherwise, the crowds are disconnected in the sense that there is no interaction between pedestrians from different crowds.

A.4.2 The mean-field type model

Again, let the setup be as in Section A.3 except that $\mathbb{F}$ now carries M independent $\mathbb{F}$-Wiener processes W^j, $j \in [\![M]\!]$, and there are M square-integrable $\mathcal{F}_0$-measurable $\mathbb{R}^d$-valued random variables ξ^j, $j \in [\![M]\!]$, independent of all the Wiener processes. Given a vector of admissible controls $\bar{a}^M = (a^1, \ldots, a^M)$, the mean-field type dynamics are

$$dX_t^j = b(t, X_t^j, a_t^j)dt + \sigma(t, X_t^j)dW_t^j, \quad X_0^j = \xi^j, \quad j \in [\![M]\!]. \tag{10}$$

There exists a unique strong solution to (10) by Proposition 1. The mean-field type risk measure for crowd $j \in [\![M]\!]$ is given by

$$J_{r,\Lambda}^j(\bar{a}^M) := \mathbb{E}\left[\int_0^T \frac{1}{2}|a^j|^2 + \int_{\mathbb{R}^d} \phi_r(X_t^j - y)\nu_{t,\Lambda}^j(dy)dt + \Psi_j(X_T^j)\right],$$

where $\nu_{t,\Lambda}^j := \sum_{k=1}^M \lambda_{jk}\mu_{X_t^k}$.

A.4.3 Solutions of N-M and MFT-M

The convergence results for the single-crowd case generalize to multiple crowds under the following assumptions:

(C1) $\sup_{N\in\mathbb{N}} \mathbb{E}_N\left[\frac{1}{N}\sum_{i=1}^N |\xi^{i,j,N}|^2\right] < \infty$ for all $j \in [\![M]\!]$;

(C2) $(\xi^{1,j,N}, \ldots, \xi^{N,j,N})$ is exchangeable for all $j \in [\![M]\!]$;

(C3) $\lim_{N\to\infty} \mathcal{L}\left(\frac{1}{N}\sum_{i=1}^N \xi^{i,j,N}\right) = \delta_{\mu_0^j}$ in $\mathcal{P}(\mathcal{P}_2(\mathbb{R}^d))$ for all $j \in [\![M]\!]$;

(C4) The controls are of feedback form, $a_t^{i,j,N}(\omega) = a^{j,N}(t, X_t^{i,j,N}(\omega))$ where each $a^{j,N}$ is a deterministic A-valued function and $a^{j,N}$ converge uniformly to a^j as $N \to \infty$. Furthermore,

$$\sup_{N\in\mathbb{N}} \mathbb{E}_N\left[\int_0^T |a^{j,N}(t, X_t^{i,j,N})|^2\right] < \infty, \quad \forall\, i \in [\![N]\!],\ \forall\, j \in [\![M]\!].$$

Under (A1)–(A3) for all final costs Ψ_j and (C1)–(C4) the results from Section A.3.3 and Section A.3.4 immediately generalize to multiple crowds. Next, solutions to the N-crowd model (N-M) and the mean-field type model (MFT-M) for the multi-crowd case are defined.

Definition 13 (Solution to N-M)**.** *For any $a^j \in \mathcal{A}$, let $(a^j)^N = (a^j, \ldots, a^j) \in \mathcal{A}^N$. The control vector $((\hat{a}^1)^N, \ldots, (\hat{a}^M)^N)$ is a solution to N-M if*

$$J_{r,\Lambda}^{i,j,N}((\hat{a}^1)^N, \ldots, (\hat{a}^M)^N) \le J_{r,\Lambda}^{i,j,N}((a^j)^N, (\hat{a}^{-j})^N), \quad \forall\, a^j \in \mathcal{A},\ \forall\, j \in [\![M]\!].$$

If

$$J^{i,j,N}_{r,\Lambda}((\hat{a}^1)^N, \dots, (\hat{a}^M)^N) \leq J^{i,j,N}_{r,\Lambda}((a^j)^N, (\hat{a}^{-j})^N) + \varepsilon, \ \forall\, a^j \in \mathcal{A}, \ \forall\, j \in [\![M]\!]$$

for $\varepsilon > 0$, $((\hat{a}^1)^N, \dots, (\hat{a}^M)^N)$ *is an* ε*-solution to MFT-M.*

Definition 14 (Solution to MFT-M)**.** *The vector* $\hat{a}^M = (\hat{a}^{1,M}, \dots, \hat{a}^{M,M}) \in \mathcal{A}^M$ *is a solution to MFT-M if*

$$J^j_{r,\Lambda}(\hat{a}^M) \leq J^j_{r,\Lambda}(a, \hat{a}^{-j,M}), \quad \forall\, a \in \mathcal{A}, \ \forall\, j \in [\![M]\!].$$

Remark 15. *There is a fundamental difference between the definition of solutions in the single-crowd case and in the multi-crowd case. The latter is a Nash equilibrium while the former is an optimal control. So, what has changed? We still have anonymity between pedestrians within a crowd but the vector of all controls used in the multi-crowd case,* $((a^{j,N}(t, X^{i,j,N}_t))^N_{i=1})^M_{j=1}$ *for N-M and* $(a^j(t, X^j_t))^M_{j=1}$ *for MFT-M, is not exchangeable (cf.* (7)*). From our point of view, we may distinguish between two pedestrians from different crowds and hence the pedestrians are not anonymous anymore. Thus, it makes sense to look at a game problem between the crowds.*

The approximation result Theorem 9 generalizes to the multi-crowd case.

Theorem 16. *Assume that* $\hat{a}^M$ *is a solution to MFT-M. Then the vector* $((\hat{a}^{1,M})^N, \dots, (\hat{a}^{M,M})^N)$ *is an* ε_N*-solution to N-M.*

Proof. The proof follows exactly the same steps as the proof of Theorem 9. □

Finally, under the assumption that $\mu_{X^j_t}$ admits a density $m_{X^j}(t, \cdot)$, we rewrite MFT-M into a deterministic problem (dMFT-M).

Definition 17 (Solution to dMFT-M)**.** *A control vector* $\hat{a} = (\hat{a}^1, \dots, \hat{a}^M) \in \mathcal{A}^M_d$ *solves dMFT-M if*

$$J^{j,\det}_{r,\Lambda}(\hat{a}) \leq J^{j,\det}_{r,\Lambda}(a, \hat{a}^{-j}), \quad \forall\, a \in \mathcal{A}_d, \ \forall\, j \in [\![M]\!],$$

where

$$J^{j,\det}_{r,\Lambda}(\hat{a}) := \int_{\mathbb{R}^d} \Bigg[\int_0^T \Bigg\{ \frac{1}{2} |\hat{a}^j(t,x)|^2 m_j(t,x) \\ + \sum_{k=1}^M \lambda_{jk} \int_{\mathbb{R}^d} \phi_r(x-y) m_k(t,y) dy\, m_j(t,x) \Bigg\} dt + \Psi_j(x) m_j(T,x) \Bigg] dx$$

and m_j solves

$$\begin{cases} \dfrac{\partial m_j}{\partial t}(t,x) = \dfrac{1}{2}\mathrm{Tr}\left[\nabla^2(\sigma\sigma^T m_j)(t,x)\right] - \nabla\cdot(b(t,x,\hat{a}^j(t,x))m_j(t,x)), \\ m_j(0,t) = \text{ the density of } \mu_0^j. \end{cases}$$

Remark 18. *In the limit $r \to 0$ the risk measure is*

$$\begin{aligned} J_{0,\Lambda}^{j,det}(a) = \int_{\mathbb{R}^d} \Bigg[\int_0^T \Bigg\{ & \frac{1}{2}|a^j(t,x)|^2 m_j(t,x) \\ & + \sum_{k=1}^M \lambda_{j,k} m_k(t,x) m_j(t,x) \Bigg\} dt + \Psi_j(x) m_j(T,x) \Bigg] dx. \end{aligned} \tag{11}$$

The interpretation is the same as in the single-crowd model, when $r \to 0$ the personal space of the pedestrians shrink to a singleton and only collisions have an impact on the choice of control. Note that (11) *with parameters $M = 2$, $\lambda_{11} = \lambda_{22} = 1$, and $\lambda_{12} = \lambda_{21} = \lambda$ is exactly the cost used in [15].*

A.4.4 An optimal control problem equivalent to dMFT-M

In this section an optimal control problem is introduced. It is shown to have the same solution as dMFT-M, so instead of solving the game, an optimal control is characterized by a Pontryagin-type maximum principle. To ease notation, let $\varphi = (\varphi_1, \dots, \varphi_M)$ for $\varphi \in \{\Psi(x), m(t,x), |a(t,x)|^2\}$. Consider the following optimization problem,

$$\begin{cases} \min\limits_{a\in\mathcal{A}_d^M} \; J_{r,\bar{\Lambda}}(a) \\ \quad \text{s.t. } \dfrac{\partial m_j}{\partial t}(t,x) = \dfrac{1}{2}\mathrm{Tr}\left[\nabla^2(\sigma\sigma^T m_j)(t,x)\right] \\ \qquad\qquad\qquad\qquad - \nabla\cdot(b(t,x,a^j(t,x))m_j(t,x)), \\ \qquad\quad m_j(0,x) = \text{density of } \mu_0^j, \quad j \in [\![M]\!], \end{cases} \tag{OC}$$

where

$$J_{r,\bar{\Lambda}}(a) := \int_{\mathbb{R}^d}\left[\int_0^T \left\{\frac{1}{2}|a(t,x)|^2 \cdot m(t,x) + G_{\phi_r}[m]^T(t,x)\bar{\Lambda}m(t,x)\right\}dt \right. \\ \left. + \Psi(x)\cdot m(T,x)\right]dx, \quad \bar{\Lambda}\in\mathbb{R}^{M\times M},$$

$$G_{\phi_r}[m](t,x) \\ := \left(\int_{\mathbb{R}^d}\phi_r(x-y)m_1(t,y)dy, \dots, \int_{\mathbb{R}^d}\phi_r(x-y)m_M(t,y)dy\right).$$

The following proposition is the first link between dMFT-M and (OC).

Proposition 19. *If $\hat{a}$ solves* (OC) *and $\Lambda = \bar{\Lambda} + \bar{\Lambda}^T - \operatorname{diag}(\bar{\Lambda})$, then $\hat{a}$ is a solution to dMFT-M.*

Proof. The proof is found in Appendix A.6.1. □

The condition $\Lambda = \bar{\Lambda} + \bar{\Lambda}^T - \operatorname{diag}(\bar{\Lambda})$ forces Λ to be symmetric and the interpretation is that the aversion between crowds must be symmetric, i.e. if one crowd is averse to another crowd, the latter crowd must be equally averse towards former. One can of course consider other situations, but then it is not possible to rewrite the game into an optimization problem on the form of (OC). Therefore from now Λ is assumed to satisfy the condition of Proposition 19. Note that $\bar{\Lambda}$ does not necessarily have to be symmetric. Towards a characterization of the optimal control, let

$$f(t,x,a,m) := \frac{1}{2}|a(t,x)|^2 \cdot m(t,x) + G_{\phi_r}[m](t,x)^T\bar{\Lambda}m(t,x), \\ g(x,m) := \Psi(x)\cdot m(T,x),$$

and let, with some abuse of notation,

$$\operatorname{Tr}\left[\sigma\sigma^T\nabla^2 p(t,x)\right] \\ := \left(\operatorname{Tr}\left[\sigma\sigma^T\nabla^2 p_1(t,x)\right], \dots, \operatorname{Tr}\left[\sigma\sigma^T\nabla^2 p_M(t,x)\right]\right), \\ \operatorname{Tr}\left[\nabla^2(\sigma\sigma^T m)(t,x)\right] \\ := \left(\operatorname{Tr}\left[\nabla^2(\sigma\sigma^T m_1)(t,x)\right], \dots, \operatorname{Tr}\left[\nabla^2(\sigma\sigma^T m_M)(t,x)\right]\right).$$

Theorem 20 (Sufficient maximum principle for (OC))**.** *Let $\hat{a} \in \mathcal{A}_d^M$, let*

$$H(t,x,a,m,p) := f(t,x,a,m) + \sum_{j=1}^{M} b(t,x,a^j(t,x))m_j(t,x)\cdot\nabla p_j(t,x),$$

and let p solve the adjoint equation

$$\begin{cases} \dfrac{\partial p}{\partial t}(t,x) = -\Bigg(\dfrac{1}{2}|\hat{a}(t,x)|^2 + G_{\phi_r}[\hat{m}]^T(t,x)(\bar{\Lambda}+\bar{\Lambda}^T) \\ \qquad + (b(t,x,\hat{a}^1(t,x))\cdot\nabla p_1(t,x),\dots,b(t,x,\hat{a}^M(t,x))\cdot\nabla p_M(t,x)) \\ \qquad\qquad + \dfrac{1}{2}\mathrm{Tr}\left[\sigma\sigma^T\nabla^2 p(t,x)\right]\Bigg), \\ p(T,x) = \Psi(x). \end{cases} \tag{12}$$

Assume that H is differentiable with respect to a and that

$$(a,m) \mapsto \int_{\mathbb{R}^d} H(t,x,a,m,p)dx \tag{13}$$

is convex for all $t \in [0,T]$. Then $\hat{a}$ solves (OC) *if for all $w^j \in \mathcal{A}_d$ and $j \in [\![M]\!]$ it holds that*

$$\int_{\mathbb{R}^d}\int_0^T D_{a^j}H(t,x,\hat{a}(t,x),\hat{m}(t,x),p)\cdot w^j(t,x)dtdx = 0. \tag{14}$$

Proof. Let $a,\hat{a} \in \mathcal{A}_d^M$ and let $a_\epsilon := \epsilon a + (1-\epsilon)\hat{a}$, $\epsilon \in (0,1)$. Let m^ϵ and $\hat{m}$ satisfy the constraints of (OC) with a_ϵ and $\hat{a}$ respectively, then $\eta := m^\epsilon - \hat{m}$ solves

$$\begin{cases} \dfrac{\partial \eta_j}{\partial t}(t,x) = \dfrac{1}{2}\mathrm{Tr}\left[\sigma^T\sigma(t,x)\nabla^2\eta_j(t,x)\right] \\ \qquad\qquad - \nabla\cdot(b(t,x,\hat{a}^j(t,x))\eta_j(t,x) + \kappa_\epsilon^j(t,x)), \\ \eta_j(0,x) = 0, \quad j \in [\![M]\!], \end{cases}$$

where $\kappa_\epsilon^j := D_a b(t,x,\hat{a}^j(t,x))\epsilon a^j m_j^\epsilon + o(\epsilon a^j)$ is a remainder that will cancel out in the end. Let $\varphi^\epsilon(t,x,p) := \varphi(t,x,a_\epsilon,m^\epsilon,p)$ for $\varphi \in \{f,g,H\}$ and define

$\hat{\varphi}$ in the same way using $\hat{a}$. Note that

$$f^\epsilon(t,x) - \hat{f}(t,x) = H^\epsilon(t,x,p) - \hat{H}(t,x,p) \\ - \sum_{j=1}^{M} \big(b(t,x,\hat{a}^j(t,x))\eta_j(t,x) + \kappa_\epsilon^j(t,x)\big) \cdot \nabla p_j(t,x)$$

and by symmetry of ϕ_r,

$$\int_{\mathbb{R}^d} G_{\phi_r}[\hat{m}](t,x)\bar{\Lambda}\eta(t,x)dx = \int_{\mathbb{R}^d} G_{\phi_r}[\eta](t,x)\bar{\Lambda}^T\hat{m}(t,x)dx. \tag{15}$$

By the convexity assumption on H,

$$\begin{aligned}
&J_{r,\bar{\Lambda}}(a_\epsilon) - J_{r,\bar{\Lambda}}(\hat{a}) \\
&= \int_{\mathbb{R}^d}\int_0^T f^\epsilon(t,x) - \hat{f}(t,x)dtdx + \int_{\mathbb{R}^d} g^\epsilon(x) - \hat{g}(x)dx \\
&\geq \int_{\mathbb{R}^d}\int_0^T \Bigg\{ D_m\hat{H}[\eta](t,x,p) + \sum_{j=1}^{M} D_{a^j}\hat{H}(t,x,p)\cdot(a_\epsilon^j(t,x) - \hat{a}^j(t,x)) \\
&\quad - \sum_{j=1}^{M}\big(b(t,x,\hat{a}^j(t,x))\eta_j(t,x) + \kappa_\epsilon^j(t,x)\big)\cdot\nabla p_j(t,x)\Bigg\}dtdx \\
&\quad + \int_{\mathbb{R}^d}\Psi(x)\cdot\eta(T,x)dx.
\end{aligned}$$

By a variation argument, the m-derivative of $\hat{H}$ is found to be

$$\begin{aligned}
D_m\hat{H}[\eta](t,x,p) &= \frac{1}{2}|\hat{a}(t,x)|^2\cdot\eta(t,x) + G_{\phi_r}[\hat{m}]^T(t,x)\bar{\Lambda}\eta(t,x) \\
&\quad + G_{\phi_r}[\eta]^T(t,x)\bar{\Lambda}\hat{m}(t,x) + \sum_{j=1}^{M} b(t,x,\hat{a}^j(t,x))\eta_j(t,x)\cdot\nabla p_j(t,x).
\end{aligned}$$

The a-derivatives of $\hat{H}$ vanish by the optimality condition (14). Hence, using (15),

$$\begin{aligned}
&J_{r,\bar{\Lambda}}(a_\epsilon) - J_{r,\bar{\Lambda}}(\hat{a}) \\
&\geq \int_0^T\int_{\mathbb{R}^d}\Bigg\{\frac{1}{2}|\hat{a}(t,x)|^2 + G_{\phi_r}[\hat{m}]^T(t,x)(\bar{\Lambda}+\bar{\Lambda}^T) + \frac{1}{2}\mathrm{Tr}\left[\sigma\sigma^T\nabla^2 p(t,x)\right] \\
&\quad + (\hat{a}^1(t,x)\cdot\nabla p_1(t,x),\ldots,\hat{a}^M(t,x)\cdot\nabla p_M(t,x)) + \frac{\partial p}{\partial t}(t,x)\Bigg\}\cdot\eta(t,x)dxdt.
\end{aligned}$$

Applying the adjoint equation (12) now gives $J_{r,\bar\Lambda}(a_\epsilon) - J_{r,\bar\Lambda}(\hat a) \geq 0$ for all convex perturbations a_ϵ of $\hat a$. In the case of a control sets A which is not convex the proof can be carried out in similar fashion by replacing the convex perturbation a_ϵ by a spike variation. □

Note that if

$$\hat a^j(t,x) = -(D_a b(t,x,a(t,x))|_{a=\hat a^j})\nabla p_j(t,x) \tag{16}$$

the optimality condition (14) is satisfied. In the case of linear dynamics, (16) is the well-known solution $\hat a^j(t,x) = -\nabla p_j(t,x)$. No property of $\bar\Lambda$ except boundedness in norm was used in the proof of the maximum principle. The following proposition identifies all matrices $\bar\Lambda$ such that the convexity assumption (13) holds.

Proposition 21. *Condition* (13) *holds if and only if*

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \phi_r(x-y)(m(t,y)-m'(t,y))^T\bar\Lambda(m(t,x)-m'(t,x))dydx \geq 0,$$

for all densities m and m' and $t \in [0,T]$

Proof. The convexity of H in a is trivial. H is convex in m if

$$\begin{aligned}&\int_{\mathbb{R}^d} H(t,x,a,\alpha m + (1-\alpha)m', p)dx\\ &\qquad \leq \alpha\int_{\mathbb{R}^d} H(t,x,a,m,p)dx + (1-\alpha)\int_{\mathbb{R}^d} H(t,x,a,m',p)dx.\end{aligned}$$

The inequality above can be rearranged into

$$\begin{aligned}0 &\geq (\alpha^2-\alpha)\int_{\mathbb{R}^d} G_{\phi_r}[\widetilde m](t,x)\bar\Lambda\widetilde m(t,x)dx\\ &= (\alpha^2-\alpha)\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\phi_r(x-y)\widetilde m(t,x)^T\bar\Lambda\widetilde m(t,x)dydx,\end{aligned}$$

where $\widetilde m := m - m'$. The fact that $(\alpha^2-\alpha) < 0$ concludes the proof. □

The opposite direction of Proposition 19 can now be proven.

Proposition 22. *If $\hat a$ solves dMFT-M, $\hat m$ satisfies the constraints of* (OC) *with control $\hat a$ and p satisfies the adjoint equation* (12)*, then $\hat a$ solves* (OC).

Proof. The proof is found in Appendix A.6.2. □

The local risk measure, introduced in Remark 18, will naturally yield a different Hamiltonian and adjoint equation than above. Anyhow, results analogous to Proposition 19, Theorem 20, and Proposition 22 hold for the local case, and their proofs are carried out following the same steps as in the nonlocal case. The most notable structural change is that in the local case, H is convex if and only if $\bar{\Lambda}$ is positive semidefinite.

A.5 Numerical example

With the following numerical example we want to illustrate the difference between local and nonlocal crowd aversion. We consider the following simple crowd model on the one-dimensional torus $\mathbb{T}$,

$$\begin{cases} \min\limits_{a \in \mathcal{A}_d} & \displaystyle\int_{\mathbb{T}} \int_0^T \left\{ \frac{a^2(t,x)}{2} + C \int_{\mathbb{T}} \phi_r(x-y) m(t,y) dy \right\} m(t,x) dt dx \\ & \displaystyle\qquad\qquad\qquad\qquad + \int_{\mathbb{T}} \Psi(x) m(T,x) dx \\ \text{s.t.} & \displaystyle\frac{\partial m}{\partial t}(t,x) = \frac{1}{2} \frac{\partial^2 m}{\partial x^2}(t,x) - \frac{\partial}{\partial x}(a(t,x) m(t,x)), \\ & m(0,x) = m_0(x). \end{cases} \tag{17}$$

To make the comparison we also consider the corresponding local crowd aversion problem

$$\begin{cases} \min\limits_{a \in \mathcal{A}_d} & \displaystyle\int_{\mathbb{T}} \int_0^T \left\{ \frac{a^2(t,x)}{2} + C m(t,x) \right\} m(t,x) dt + \Psi(x) m(T,x) dx \\ \text{s.t.} & \displaystyle\frac{\partial m}{\partial t}(t,x) = \frac{1}{2} \frac{\partial^2 m}{\partial x^2}(t,x) - \frac{\partial}{\partial x}(a(t,x) m(t,x)), \\ & m(0,x) = m_0(x). \end{cases} \tag{18}$$

The constraint in (17) and (18) corresponds to the dynamics of a pedestrian that controls her velocity but is disturbed by Brownian noise,

$$dX_t = a(t, X_t) dt + dW_t.$$

The constant C has been introduced to reweigh the contribution of crowd aversion. By up-weighting this term, emphasis is given to the impact of the preference, local or nonlocal, and the difference between the two crowds will be more clear. To solve (17) and (18) the gradient decent method (GDM) of [15] is used.

A.5.1 Simulations and discussions

We let $T = 1$ and $C = 500$, and m_0 and ϕ_r are set to the functions presented in Figure 1. Most pedestrians are initially gathered around $x = 0$ and they have an incentive to end up around $x = 0.5$ at time $t = 1$. The personal space of a pedestrian is modeled as

$$\hat{\phi}_{0.2}(x) := 5\mathbb{I}_{[0,.2]}(x).$$

In the calculations, $\hat{\phi}_{0.2}$ is smoothed with a mollifier (cf. (4)). Note that

$$\int_{\mathbb{T}} \hat{\phi}_{0.2}(x-y)m(t,y)dy = 5\mathbb{P}\left(x - X_t \in [0, 0.2]\right),$$

The use of an indicator to model the personal space thus has the following interpretation: the pedestrian acting under nonlocal crowd aversion is affected by the probability of other pedestrians being closer than 0.2 to her.

The optimal controls for (17) and (18) are found by the GDM-scheme of [15]. The convergence of the risk is presented in Figure 2. In Figure 3, a comparison between the solutions of (17) and (18) is displayed. The crowds behave similarly until time begins to approach $t = 1$. The crowd acting under nonlocal crowd aversion then gathers more densely in the low cost area. Since the crowding experienced by a pedestrian in the nonlocal model is an average over a larger neighborhood, she cares less about pointwise high densities and the benefits of reaching the low cost area around $x = 0.5$ has a stronger impact in the nonlocal model, resulting in a more concentrated density. This is visualized in Figure 4, where on the left the difference between crowd aversion penalties,

$$\int_{\mathbb{T}} \hat{\phi}_{0.2}(x-y)m_{\text{non-local}}(t,y)dy - m_{\text{local}}(t,x),$$

is plotted. On the right plot, we display

$$m_{\text{non-local}}(t,x) - m_{\text{local}}(t,x).$$

Note that even though the densities differ at $t = 1$, the two crowds experience approximately the same amount of crowding at that time $t = 1$.

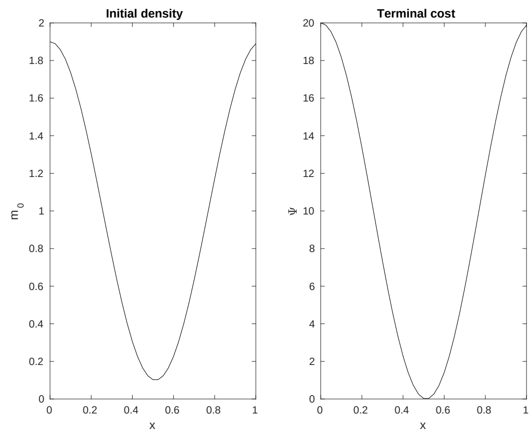


Figure 1: The initial density and terminal cost used in the simulations. Initially the pedestrians are crowded around $x = 0$ but they will quickly flatten the density to heed their crowd aversion preferences. The low cost around $x = 0.5$ will give the pedestrians an incentive to end up around this part of the domain at $t = 1$.

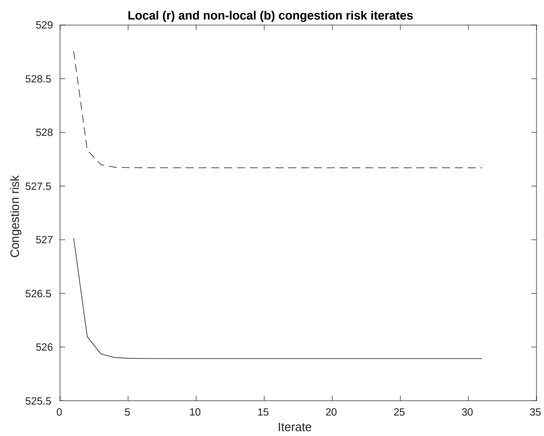


Figure 2: In each iteration of the GDM the control function a is updated. The method is run until the risk measure, under local (dashed line) and nonlocal (solid line) crowd aversion, has converged to a minimum.

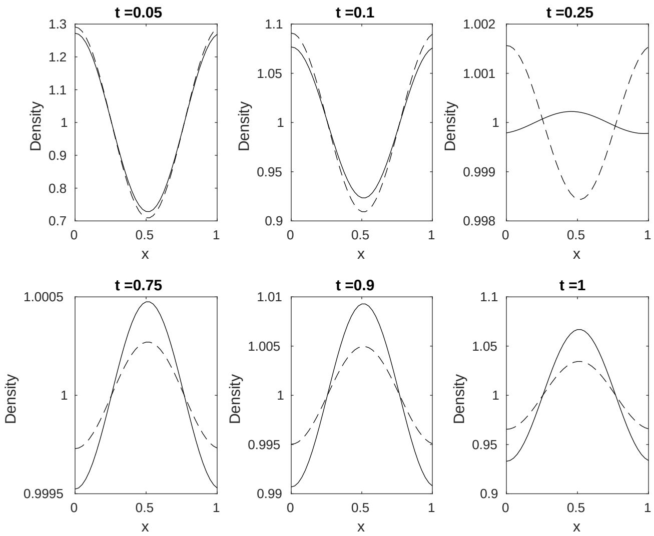


Figure 3: The optimally controlled density under local (dashed) and nonlocal (solid) crowd aversion at six instants.

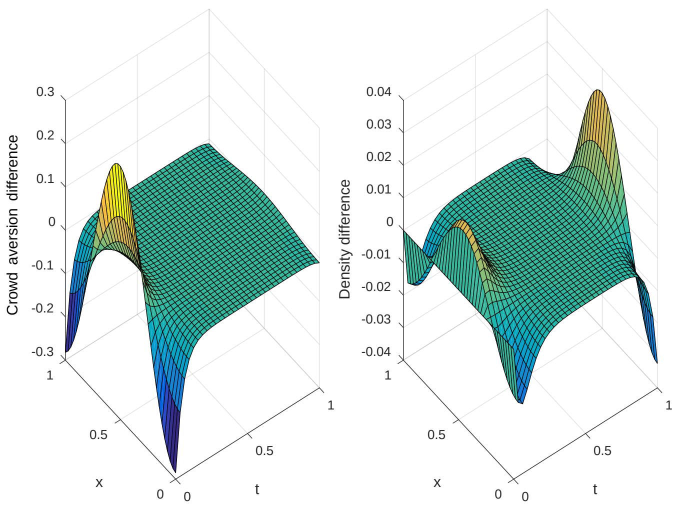


Figure 4: Differences between the two crowd models. *Left:* Nonlocal minus local crowd aversion penalty. *Right:* Nonlocal minus local density.

A.6 Appendix

A.6.1 Proof of Proposition 19

The proof extends the results in [15] to an arbitrary finite number of crowds and to nonlocal crowd aversion.

Proof. Let the entries in $\bar{\Lambda}$ be denoted by $\bar{\lambda}_{jk}$. For each $j \in [\![M]\!]$,

$$\begin{aligned} &J_{r,\bar{\Lambda}}(a) - J^{j,\det}_{r,\Lambda}(a^j, a^{-j}) \\ &= \sum_{k \neq j} \left(\int_{\mathbb{R}^d} \int_0^T \frac{1}{2} |a^k(t,x)|^2 m_k(t,x) dt dx + \int_{\mathbb{R}^d} \Psi_k(x) m_k(T,x) dx \right) \\ &\qquad + \sum_{k,l=1}^{M} \left(\int_{\mathbb{R}^d} \int_0^T G_{\phi_r}[m_k]^T \bar{\lambda}_{kl} m_l dt dx \right) \\ &\qquad\qquad - \sum_{k=1}^{M} \left(\int_{\mathbb{R}^d} \int_0^T G_{\phi_r}[m_k]^T \lambda_{kj} m_j dt dx \right). \end{aligned} \tag{19}$$

Note that by symmetry of ϕ, the indices of $G_\phi[m_k]$ and m_l may be swapped under the integral sign and the last two lines of (19) can be rewritten as

$$\begin{aligned} &\sum_{\substack{k,l=1 \\ l,k \neq j}}^{M} \left(\int_{\mathbb{R}^d} \int_0^T G_{\phi_r}[m_k]^T \bar{\lambda}_{kl} m_l dt dx \right) \\ &\qquad + \int_{\mathbb{R}^d} \int_0^T \sum_{\substack{k=1 \\ k \neq j}}^{M} \Big(G_{\phi_r}[m_k]^T (\bar{\lambda}_{kj} + \lambda_{jk} - \lambda_{kj}) m_j \Big) \\ &\qquad\qquad + G_{\phi_r}[m_j]^T (\bar{\lambda}_{jj} - \lambda_{jj}) m_j dt dx. \end{aligned}$$

The last two lines of the last equation vanish since $\Lambda = \bar{\Lambda} + \bar{\Lambda}^T - \operatorname{diag}(\bar{\Lambda})$ and $J_{r,\bar{\Lambda}}(a) - J^{j,\det}_{r,\Lambda}(a^j, a^{-j})$ is independent of (a^j, m_j). Therefore the optimality of $\hat{a}$ implies that

$$J^{j,\det}_{r,\Lambda}(\hat{a}) \leq J^{j,\det}_{r,\Lambda}(a^j, \hat{a}^{-j}), \quad \forall\, a_j \in \mathcal{A}_d,\ j \in [\![M]\!]. \tag{20}$$

Since (20) holds for all $j \in [\![M]\!]$, $\hat{a}$ is a solution to dMFT-M. □

A.6.2 Proof of Proposition 22

This proof is a variation of [14, Proposition 4.2.1] and extends it to an arbitrary finite number of crowds and to nonlocal crowd aversion.

Proof. Let, for a given $\epsilon > 0$, a^j_ϵ be the first order perturbation of $\hat{a}^j$ for some arbitrary w^j such that

$$a^j_\epsilon(t,x) := \hat{a}^j(t,x) + \epsilon w^j(t,x) \in \mathcal{A}_d.$$

Let m^ϵ_j satisfy the constraints in (OC) with a^j_ϵ and let

$$m^\epsilon_j(t,x) := \hat{m}_j(t,x) + \epsilon h^\epsilon_j(t,x) + \mathcal{O}({h^\epsilon_j}^2).$$

Then h^ϵ_j satisfies the equation

$$\begin{cases} \dfrac{\partial h^\epsilon_j}{\partial t}(t,x) = \dfrac{1}{2}\mathrm{Tr}\left[\nabla^2(\sigma\sigma^T h^\epsilon_j)(t,x)\right] - \nabla\cdot\big(b(t,x,\hat{a}^j(t,x))h^\epsilon_j(t,x)\big) \\ \qquad\qquad - \nabla\cdot\left(\dfrac{b(t,x,a^j_\epsilon(t,x)) - b(t,x,\hat{a}^j(t,x)}{\epsilon}m^\epsilon_j(t,x))\right), \\ h^\epsilon_j(0,x) = 0. \end{cases} \tag{21}$$

Let $\mathcal{J}^j : \epsilon \to J^{j,\det}_{r,\Lambda}(a^j_\epsilon, \hat{a}^{-j})$. Since the functional is convex, $\hat{a}$ solves dMFT-M if and only if

$$\frac{\partial \mathcal{J}^j_{r,\Lambda}}{\partial \epsilon}(0) = 0, \quad \forall w^j \text{ such that } \hat{a}^j + \epsilon w^j \in \mathcal{A}_d,\ \forall\ j \in [\![M]\!]. \tag{22}$$

Condition (22) is equivalent to

$$\begin{aligned} 0 = \int_{\mathbb{R}^d}\Bigg[\int_0^T \Bigg\{ & \hat{a}^j(t,x)\hat{m}_j(t,x)\cdot w^j(t,x) + \frac{1}{2}|\hat{a}^j(t,x)|^2 h^0_j(t,x) \\ & + 2\lambda_{jj}\left(\int_{\mathbb{R}^d}\phi_r(x-y)\hat{m}_j(t,y)dy\right)h^0_j(t,x) \\ + \sum_{\substack{k=1\\k\neq j}}^M \lambda_{jk}\left(\int_{\mathbb{R}^d}\phi_r(x-y)\hat{m}_k(t,y)dy\right) & h^0_k(t,x)\Bigg\}dt + \Psi_j(x)h^0_j(T,x)\Bigg]dx \end{aligned} \tag{23}$$

where h_j^0 is the solution of (21) in the limit $\epsilon \to 0$. Since p satisfies the adjoint equation, $\Psi_j(x) = p_j(T,x)$ and

$$\begin{aligned}
&\int_{\mathbb{R}^d} p_j(T,x) h_j^0(T,x)dx \\
&= \int_{\mathbb{R}^d} \int_0^T \Bigg\{ -\frac{1}{2}|a^j(t,x)|^2 - 2\lambda_{jj} \left(\int_{\mathbb{R}^d} \phi_r(x-y)\hat{m}_j(t,y)dy \right) \\
&- \sum_{k \neq j}^{M} \lambda_{jk} \left(\int_{\mathbb{R}^d} \phi_r(x-y)\hat{m}_k(t,y)dy \right) - b(t,x,\hat{a}^j(t,x)) \cdot \nabla p_j(t,x) \\
&- \frac{1}{2}\mathrm{Tr}\left[\sigma\sigma^T \nabla^2 p_j(t,x)\right] \Bigg\} h_j^0(t,x) \\
&+ \Bigg\{ \frac{1}{2}\mathrm{Tr}\left[\nabla^2(\sigma\sigma^T h_j^0)(t,x)\right] - \nabla \cdot \left(b(t,x,\hat{a}^j(t,x)) h_j^0(t,x) \right) \\
&- \nabla \cdot \left(D_{a^j} b(t,x,\hat{a}^j(t,x)) w^j(t,x) \hat{m}_j(t,x) \right) \Bigg\} p_j(t,x) dt dx.
\end{aligned} \tag{24}$$

Inserting (24) into (23) yields

$$\int_{\mathbb{R}^d} \int_0^T \left(\hat{a}^j(t,x) + D_{a^j} b(t,x,\hat{a}^j(t,x))^T \nabla p_j(t,x) \right) \cdot w_j(t,x)\hat{m}_j(t,x) dx dt = 0.$$

The last equation holds for all $j \in [\![M]\!]$ so $\hat{a}$ satisfies the optimality condition in Theorem 20 and therefore $\hat{a}$ is a solution to (OC) by Theorem 20. $\square$

References

[1] Y. Achdou and M. Laurière. On the system of partial differential equations arising in mean field type control. *Discrete & Continuous Dynamical Systems-A*, 35(9):3879–3900, 2015.

[2] Y. Achdou and M. Laurière. Mean Field Type Control with Congestion (II): An Augmented Lagrangian Method. *Applied Mathematics & Optimization*, 74(3):535–578, 2016.

[3] Y. Achdou and M. Laurière. Mean field type control with congestion. *Applied Mathematics & Optimization*, 73(3):393–418, 2016.

[4] Y. Achdou, M. Bardi, and M. Cirant. Mean field games models of segregation. *Mathematical Models and Methods in Applied Sciences*, 27 (01):75–113, 2017.

[5] D. Andersson and B. Djehiche. A maximum principle for SDEs of mean-field type. *Applied Mathematics & Optimization*, 63(3):341–356, 2011.

[6] M. Bardi and M. Cirant. Uniqueness of solutions in mean field games with several populations and Neumann conditions. In *PDE Models for Multi-Agent Phenomena*, pages 1–20. Springer, 2018.

[7] A. Bensoussan, J. Frehse, P. Yam, et al. *Mean field games and mean field type control theory*, volume 101. Springer, 2013.

[8] M. Cirant. Multi-population mean field games systems with Neumann boundary conditions. *Journal de Mathématiques Pures et Appliquées*, 103(5):1294–1315, 2015.

[9] E. Feleqi. The derivation of ergodic mean field game equations for several populations of players. *Dynamic Games and Applications*, 3(4): 523–536, 2013.

[10] M. Huang, R. P. Malhamé, P. E. Caines, et al. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information & Systems*, 6(3):221–252, 2006.

[11] O. Kallenberg. *Foundations of modern probability.* Springer Science & Business Media, 2006.

[12] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science & Business Media, 2012.

[13] V. Kolokoltsov and M. Troeva. On the mean field games with common noise and the McKean-Vlasov SPDEs. *arXiv preprint arXiv:1506.04594*, 2015.

[14] A. Lachapelle. *Quelques problèmes de transport et de contrôle en économie: aspects théoriques et numériques.* PhD thesis, Université Paris Dauphine-Paris IX, 2010.

[15] A. Lachapelle and M.-T. Wolfram. On a mean field game approach modeling congestion and aversion in pedestrian crowds. *Transportation research part B: methodological*, 45(10):1572–1589, 2011.

[16] J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese Journal of Mathematics*, 2(1):229–260, 2007.

[17] P.-L. Lions. Cours du Collège de France. URL `https://www.college-de-france.fr/site/en-pierre-louis-lions/course-2011-2012.htm`. Accessed: 2017-09-29.

[18] G. Naldi, L. Pareschi, and G. Toscani. *Mathematical modeling of collective behavior in socio-economic and life sciences*. Springer Science & Business Media, 2010.

[19] K. Oelschläger. A law of large numbers for moderately interacting diffusion processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 69(2):279–322, 1985.

[20] K. R. Parthasarathy. *Probability measures on metric spaces*, volume 352. American Mathematical Soc., 1967.

[21] A.-S. Sznitman. Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX—1989*, pages 165–251. Springer, 1991.

[22] A. Tcheukam, B. Djehiche, and H. Tembine. Evacuation of multi-level building: Design, control and strategic flow. In *Control Conference (CCC), 2016 35th Chinese*, pages 9218–9223. IEEE, 2016.

[23] T. Yamada, S. Watanabe, et al. On the uniqueness of solutions of stochastic differential equations. *Journal of Mathematics of Kyoto University*, 11(1):155–167, 1971.

[24] J. Yong and X. Y. Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*, volume 43. Springer Science & Business Media, 1999.

Paper B

Modeling tagged pedestrian motion: a mean-field type game approach

Modeling tagged pedestrian motion: a mean-field type game approach

by

Alexander Aurell and Boualem Djehiche

Published in
Transportation Research Part B: Methodological
Volume 121 (2019), 168-183.

Abstract

This paper suggests a model for the motion of *tagged pedestrians*: pedestrians moving towards a specified targeted destination, which they are forced to reach. It aims to be a decision-making tool for the positioning of fire fighters, security personnel and other services in a pedestrian environment. Taking interaction with the surrounding crowd into account leads to a differential nonzero-sum game model where the tagged pedestrians compete with the surrounding crowd. When deciding how to act, pedestrians consider crowd distribution-dependent effects, like congestion and crowd aversion. Including such effects in the parameters of the game makes it a *mean-field type game*. The equilibrium control is characterized, and special cases are discussed. Behavior in the model is studied by numerical simulations.

Keywords: pedestrian dynamics, backward-forward stochastic differential equations, mean-field type games, congestion, crowd aversion, evacuation planning

B.1 Introduction

Tagged pedestrians are individuals that plan their motion from an *unspecified initial position* in order to reach a *specified target location* in a certain time. The model for tagged pedestrian motion proposed in this paper is based on mean-field type game theory, and is a decision making tool for the positioning of fire fighters, medical personnel, etc, during mass gatherings. The tagged's prime objective is the pre-set final destination which is considered essential to reach; ending up in proximity of the final destination is not acceptable, it has to be reached. This is in sharp contrast to the standard finite-horizon models cited below, where pedestrians are penalized if their final position deviates from a target position, such penalization is a 'soft' constraint and can be broken at a cost. The tagged's initial position is chosen rationally. Therefore, we are inclined to think of the tagged as external entities to be deployed in the crowd. Where they (rationally) ought to be deployed is subject to an offline calculation made by a coordinator: a *central planner*. Besides tagged pedestrian motion, possible applications of the model include cancer cell dynamics and smart medicine in the human body, and malware propagation in a network, among other.

The central planner's decision making is based on knowledge of the pedestrian distribution. As noted in [43], the pedestrian behavior in a dense crowd is empirically random to some extent, likely due to the large number of external inputs. In a noisy environment, the central planner anticipates the behavior of the crowd and predicts the tagged's path to the target. As is standard in the mean-field approach, pedestrian interaction is modeled as a reactions to the state distribution of the crowd. This leads us to formulate a mean-field type game based model, which in certain scenarios reduces to an optimal control based model.

B.1.1 Related work

B.1.1.1 Optimal control and games of mean-field type

Rational pedestrian behavior is, in this paper, characterized by either a game equilibrium, or an optimal control. The tool used to find the equilibrium/optimal behavior is Pontryagin's stochastic maximum principle (SMP). For stochastic control problems, the SMP yields, when available, necessary conditions that must be satisfied by any optimal solution. The necessary conditions become sufficient under additional convexity conditions. Early results show that an optimal control along with the corresponding optimal state trajectory must solve the so-called Hamiltonian system, which is a two-point (forward-backward) boundary value problem, together with a maximum condition of the so-called Hamiltonian function (see [49] for a detailed account). In stochastic differential games, both zero-sum and nonzero-sum, the SMP is one of the main tools for obtaining conditions for an equilibrium, and was essentially inherited from the theory of stochastic optimal control. For recent examples of the use of the SMP in stochastic differential game theory, see [41, 4].

In stochastic systems, the backward equation is fundamentally different from the forward equation, if the solution is required to be adapted. An adapted solution to a backward stochastic differential equation (BSDE) is a pair of adapted stochastic processes $(Y_\cdot, Z_\cdot)$, where $Z_\cdot$ corrects any "non-adaptedness" caused by the terminal condition of $Y_\cdot$. As pointed out in [36], the first component $Y_\cdot$ corresponds to the mean evolution of the dynamics, and $Z_\cdot$ to the risk between current time and terminal time. Linear BSDEs extend to non-linear BSDEs [44] and backward-forward SDEs (BFSDE) [3, 29]. BSDEs with distribution-dependent coefficients, mean-field BSDEs, are by now well-understood objects [11, 12]. Mean-field BFSDEs arise naturally in the probabilistic analysis of mean-field games (MFG), mean-field type games (MFTG) and optimal control of mean-field type equations.

The theory of optimal control of mean-field SDEs, initiated in [2], treats stochastic control problems with coefficients dependent on the marginal state-distribution. This theory is by now well developed for forward stochastic dynamics, i.e., with initial conditions on state [10, 23, 14, 17]. With SMPs for optimal control problems of mean-field type at hand, MFTG theory can inherit these techniques, just like stochastic differential game theory does in the mean-field free case. See [45] for a review of solution approaches to MFTGs.

Optimal control of mean-field BSDEs has recently gained attention. In [40] the mean-field LQ BSDE control problem with deterministic coefficients is studied. Assuming the control space is linear, linear perturbation is used to derive a stationarity condition which together with a mean-field BFSDE system characterizes the optimal control. Other recent work on the control of mean-field BSDEs makes use of the SMP approach of [48] to control of BFSDEs.

Optimal control problems of mean-field type can be interpreted as large population limits of cooperative games, where the players collaborate to optimize a joint objective [38]. A close relative to mean-field type control is MFG, a class of non-cooperative stochastic differential games, initiated independently by [32] and [39]. Both mean-field type control problems and MFGs approximate games between a large number of indistinguishable (anonymous) players, interacting weakly through a mean-field coupling term. Weak player-to-player interaction through the mean-field coupling restricts the influence one player has on any other player to be inversely proportional to the number of players, hence the level of influence of any specific player is very small. In contrast to the MFG, players in a MFTG can be influential, and distinguishable (non-anonymous). That is, state dynamics and/or cost need not be of the same form over the whole player population, and a single player can have a major influence on the other players' dynamics and/or cost.

B.1.1.2 Pedestrian crowd modeling

There is a variety of mathematical approaches to the modeling of pedestrian crowd motion. *Microscopic force-based* models [27, 18], and in particular the social force model, represent pedestrian behavior as a reaction to forces and potentials, applied not only by the surrounding environment but also by the pedestrian's internal motivation and desire. A *cellular automata* approach to the microscopic modeling of pedestrian crowds can be found in [15, 35, 34], to mention only a few sources. *Macroscopic* models view the crowd as a continuum, described by averaged quantities such as density and pressure. The Hughes model [33, 31, 46] couples a conservation law, representing the physics of the crowd, with an Eikonal equation modeling a common task of the pedestrians. Its variations are manifold. *Kinetic* and other *multi-scale* models [6, 20, 7] constitute an intermediate step between the micro- and the macro scales. *Microscopic game* and *optimal control* models for pedestrian crowd dynamics, with their relevant continuum limits in the form of *mean-field games* and *mean-field type optimal control*, have gained interest in the last decade. In [28], a simplified MFG was used to model the rational behavior of a pedestrian in a crowd. Following Lasry and Lion's paper on MFG, [25] applied MFG to pedestrian crowd motion and [37] used MFG to model local congestion effects, i.e. the relationship between energy needed to walk/run and local crowd density. MFG based models have also been used to simulate evacuation of pedestrians [24].

The mean-field approach rests on an exchangeability assumption; pedestrians are anonymous, they may have different paths but one individual cannot be distinguished from another. While interesting when modeling circumstances where pedestrians can be considered indistinguishable (e.g. a train station during rush hour or fast exits of an area in case of an alarm) there are situations where an anonymous crowd model is not satisfactory. The mean-field type game is a tool to model distinguishable sub-crowds and influential individuals, and has been applied in e.g. [5]. Other ways to break the anonymity within the mean-field approach are *multi-population MFGs* [26, 19, 1] and *major agent models* [30, and references therein].

Another important characteristic of standard MFGs and MFTGs is the assumption of anticipative players. Each player is assumed to predict how the whole population will act in the future, and then pick a strategy accordingly. In a pedestrian crowd setting, this would correspond to pedestrians knowing, most likely by experience, how the surrounding crowd will behave. This is a very high 'level of rationality', and certainly not appropriate in all scenarios. We refer to [21] for a precise discussion on the level of rationality of pedestrian crowd models, including mean-field games and optimal control of mean-field type based models.

B.1.2 Paper contribution and outline

This paper investigates a new modeling approach to the motion of pedestrians whose primary objective is to reach a specific target, at a specific time. The model can represent a large group that is steered by a central planner, or a single individual. Even though these pedestrians have certain non-standard goals, they are constrained by the same physical limitations as ordinary pedestrians, and the central planner takes this into consideration. Moreover, the central planner has access to complete information of the surrounding environment, and utilizes it in the decision making process.

The contribution of this paper is a mean-field type game based model for the motion of tagged pedestrians in a surrounding crowd of 'ordinary pedestrians'. The players in the game are *crowds*, and act rationally under general distribution-dependent dynamics and cost. The tagged pedestrians have a 'hard' terminal condition, while the ordinary pedestrians have a 'hard' initial condition, and this results in a state equation in the form of a mean-field BFSDE, representing i) the predicted motion of the tagged towards the target, coupled with ii) the evolution of the surrounding crowd from a known initial configuration.

Rational behavior in the model is characterized with a version of the SMP, tailored for our mean-field BFSDE with mean-field type costs. The central section of this paper is the solved examples, where we illustrate pedestrian behavior in the model. Further directions of research are also outlined.

The tagged pedestrian model is presented in Section B.2, which begins in a deterministic setting, to which we gradually add components until the full model is reached. The SMP that gives necessary and sufficient conditions for equilibrium controls in the mean-field type game is presented in Theorem 2. Examples of tagged motion are studied in Section B.3. All technical proofs and background theory have been moved to the appendices.

B.2 The tagged pedestrian model

In this section the tagged pedestrian model is introduced. Velocity fields are the driving components in the model and include both small-scale pedestrian interactions and path planning components. The latter is implemented by pedestrians in a rational way: a cost functional summarizing pedestrian preferences is minimized. The former describes involuntary movement over which the pedestrian has no control.

List of symbols

$T \in (0, \infty)$	the time horizon
$(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$	the underlying filtered probability space
$\mathcal{P}(\mathbb{R}^d)$	the space of probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$
$\mathcal{P}_2(\mathbb{R}^d)$	all square-integrable measures in $\mathcal{P}(\mathbb{R}^d)$
$\mathcal{L}(X) \in \mathcal{P}(\mathbb{R}^k)$	the distribution of the random variable $X \in \mathbb{R}^k$ under $\mathbb{P}$
$X.$	the stochastic process $\{X_t;\, t \geq 0\}$
$L^2_{\mathcal{F}}(\mathbb{R}^d)$	the space of $\mathcal{F}$-measurable square-integrable $\mathbb{R}^d$-valued random variables

B.2.1 Pedestrian dynamics

In a deterministic setting the pedestrian state dynamics are described by ordinary differential equations (ODE). An *ordinary pedestrian* is initiated at some location $x_0 \in \mathbb{R}^d$, and moves according to an ODE with an initial condition,

$$\begin{cases} \frac{d}{dt}X_t = b_t^x, & t \in [0,T], \\ X_0 = x_0. \end{cases} \tag{1}$$

The pedestrian influences her velocity through a *control* function, $u_\cdot^x$. The control is assumed to take values in the set $U^x \subset \mathbb{R}^{d_x}$, $d_x \geq 1$. Alongside the control function, b_t^x may depend on the interaction with other pedestrians, for example collisions. In the literature, the velocity is often split into a *behavioral velocity* (the control) and an *interaction velocity*,

$$b_t^x = u_t^x + b_t^{x,\mathrm{int}}.$$

So, on top of any interaction velocity, the tagged influences her movement through a control, and this grants her some smartness. As was discussed in the introduction, the pedestrian may foresee crowd movement and act in advance to avoid congested areas and other obstacles. Pedestrian models that consider the behavioral velocity to be an internal choice of the pedestrian leaves the framework of classical particle models, they are decision-based smart models. A summary of the difference between the model classes is found in [43]. Alongside ordinary pedestrians, *tagged pedestrians* are deployed. They represent a person on a mission, who has to reach a target location $y_T \in \mathbb{R}^d$ at time $t = T$. In the deterministic setting, the tagged moves according to an ODE with terminal condition,

$$\begin{cases} \frac{d}{dt}Y_t = u_t^y + b_t^{y,\mathrm{int}}, & t \in [0,T],\ u_t^y \in U^y \subset \mathbb{R}^{d_y},\ d_y \geq 1, \\ Y_T = y_T. \end{cases} \tag{2}$$

Just like $u_\cdot^x$ influences the terminal position of an ordinary pedestrian, so does $u_\cdot^y$ influence the initial position of a tagged pedestrian. This should be interpreted in the following way: the initial position of the tagged pedestrian is not pre-determined, but depends on the pedestrian's choice of behavioral velocity. The choice is subject to the terminal condition, and at the same time it reflects other preferences. For example, if there is a high risk of injury at a certain location y_T (doors, stairs, etc.), where is the best spot for a medic to be positioned, so that she can reach y_T in time T? Certainly not at y_T, since it is a high risk area. The medic's initial location is preferably a safe spot, from which it is easy for her to access y_T, taking surrounding pedestrians and environment into account. This is what should be reflected in the choice of $u_\cdot^y$.

Pedestrian motion can be considered deterministic if the crowd is sparse, but partially random if the crowd is dense. To capture this, (1)-(2) is extended to its stochastic counterpart. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space, endowed with the filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$, satisfying the usual conditions. Let the space carry $B^x_\cdot$ and $B^y_\cdot$, independent w_x- and w_y-dimensional $\mathbb{F}$-Wiener processes, and a random variable $y_T \in L^2_{\mathcal{F}_T}(\mathbb{R}^d)$ independent of $B_\cdot := (B^x_\cdot, B^y_\cdot)^*$. The Brownian motion $B_\cdot$ is split into $B^x_\cdot$ and $B^y_\cdot$ to emphasize modeling features: $B^x_\cdot$ is the noise that explicitly effects the ordinary pedestrian's diffusion, while $B^y_\cdot$ may be used to model any noise that in addition to $B^x_\cdot$ effects the tagged. All information in the model up to time t is contained in $\mathcal{F}_t$, and a process that depends only on past and current information is called an $\mathcal{F}_t$-*adapted process*. It is natural to require pedestrian motion to be adapted, since pedestrians react causally to the environment. In this random environment, we consider a control to be feasible if it is open-loop adapted and square integrable, i.e. belongs to the sets $\mathcal{U}^x$ and $\mathcal{U}^y$ for ordinary and tagged pedestrians respectively,

$$\begin{aligned} \mathcal{U}^x &:= \left\{ u : \Omega \times [0,T] \to U^x \;\middle|\; u_\cdot \text{ is } \mathbb{F}\text{-adapted, } \mathbb{E}\left[\int_0^T |u_s|^2 ds\right] < \infty \right\}, \\ \mathcal{U}^y &:= \left\{ u : \Omega \times [0,T] \to U^y \;\middle|\; u_\cdot \text{ is } \mathbb{F}\text{-adapted, } \mathbb{E}\left[\int_0^T |u_s|^2 ds\right] < \infty \right\}. \end{aligned} \tag{3}$$

As a demonstration of how randomness effects the tagged model, consider the tagged dynamics with Brownian small-scale interactions and a random terminal condition $y_T \in L^2_{\mathcal{F}_T}(\mathbb{R}^d)$,

$$\begin{cases} dY_t = (u^y_t + B^y_t)dt, \quad \forall\ t \in [0,T],\ u^y_\cdot \in \mathcal{U}^y, \\ Y_T = y_T. \end{cases}$$

The naive solution $Y_t = y_T - \int_t^T (u^y_s + B^y_s) ds$ is not $\mathcal{F}_t$-adapted, it depends on $(B^y_s; t \leq s \leq T)$! On the other hand,

$$Y_t = \mathbb{E}\left[y_T - \int_t^T u^y_s + B^y_s ds \mid \mathcal{F}_t \right] \tag{4}$$

is $\mathcal{F}_t$-adapted. If Y_t from (4) is square-integrable for all $t \in [0,T]$, which it is in the current setup, the martingale representation theorem grants existence of a unique square-integrable $\mathbb{R}^{d \times (w_x + w_y)}$-valued and $\mathcal{F}_t$-adapted process Z.

such that
$$\begin{cases} Y_t - \int_0^t (u_s^y + B_s^y)ds = \int_0^t Z_s dB_s, \quad \forall t \in [0,T], \\ Y_T = y_T. \end{cases} \tag{5}$$

Equation (5) constitutes a BSDE. The conditional expectation (4) can be interpreted as the L^2-projection of the tagged's future path onto currently available information. Therefore, a practical interpretation of $Z_\cdot$ is that it is a supplementary control used by the tagged, to make her path to y_T the 'best prediction' at any instant in time. From a modeling point of view, the tagged pedestrian thus uses two control processes:

- $u_\cdot^y$ - to heed preferences on initial position, speed, congestion and more. It is the tagged's subjective best response to the environment.
- $Z_\cdot$ - to predict the best path to y_T given $u_\cdot^y$. It is a square-integrable process, implicitly given by the martingale representation theorem.

Interaction between pedestrians at time t is introduced via the mean-field of ordinary pedestrians, $\mathcal{L}(X_t) := \mathbb{P} \circ (X_t)^{-1}$, and tagged pedestrians, $\mathcal{L}(Y_t)$. The distributions approximate the over-all behavior of the crowds in the large population limit, under the assumption that within each of the two groups individuals are indistinguishable (anonymous). This assumption is in place throughout the paper. An example of a mean-field dependent preference is the following: to safely accomplish its mission, a security team prefers that no individual deviates too far away from the mean position of the team. Also, effects like congestion, the extra effort needed when moving in a high density area, and aversion, repulsion from other pedestrians, can be captured with distribution dependent coefficients.

In a random environment, with mean-field interactions, we formulate the dynamics of representative group members (ordinary and tagged) in the model as
$$\begin{cases} dY_t = b^y(t, \Theta_t^y, Z_t, \Theta_t^x)dt + Z_t dB_t, \\ dX_t = b^x(t, \Theta_t^y, Z_t, \Theta_t^x)dt + \sigma^x(t, \theta_t^y, Z_t, \theta_t^x)dB_t^x, \\ Y_T = y_T, \qquad X_0 = x_0, \end{cases} \tag{6}$$

where $\Theta_t^y := (Y_t, \mathcal{L}(Y_t), u_t^y)$, $\theta_t^y := (Y_t, \mathcal{L}(Y_t))$ and
$$b^x, b^y : \Omega \times [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times U^y \times \mathbb{R}^{d\times(w_x+w_y)} \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times U^x \to \mathbb{R},$$
$$\sigma^x : \Omega \times [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^{d\times(w_x+w_y)} \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}.$$

Θ_t^x and θ_t^x are defined correspondingly. Distribution-dependence makes (6) a so-called *controlled mean-field BFSDE*. Appendix B.5 summarizes results on existence of unique solutions to mean-field BFSDEs and states assumptions strong enough so that there exists a unique solution to (6) given any feasible control pair $(u_\cdot^x, u_\cdot^y) \in \mathcal{U}^x \times \mathcal{U}^y$. The assumptions are in force throughout this paper.

Remark 1. *Fundamental diagrams, that describe the marginal relations between speed, density and flow in a crowd, are not necessary in the construction of b^x and b^y, as functions of crowd density. It is pointed out in [21] that the use of fundamental diagrams in pedestrian crowd models is an artifact from road traffic models, without proper justification in the case of two-dimensional flows. Instead, the more natural (for the purpose of modeling pedestrian crowd motion) multidimensional velocity fields b^x and b^y are used here.*

B.2.2 Pedestrian preferences

Modeling pedestrian preference is a delicate task. Not only is the gathering of and the calibration to empirical data difficult for many reasons, but different setups lead to vastly different mathematical formulations of rationality. The focus in this paper is on setups where pedestrian groups are controlled by a central planner. Other possible setups are discussed in Section B.4.

The ordinary and tagged pedestrians pay the cost f^x and f^y per time unit, respectively,

$$f^x, f^y : \Omega \times [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times U^y \times \mathbb{R}^{d\times(w_x+w_y)} \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times U^x \to \mathbb{R}.$$

On top of this, ordinary pedestrians pay a terminal cost h^x at time T, and tagged pedestrians pay an initial cost h^y at time 0,

$$h^x, h^y : \Omega \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}.$$

Given $(u_\cdot^x, u_\cdot^y)$, a pair of feasible controls, the total cost is $J^x : \mathcal{U}^x \to \mathbb{R}$ for the ordinary pedestrian, and $J^y : \mathcal{U}^y \to \mathbb{R}$ for the tagged,

$$\begin{aligned} J^x(u_\cdot^x; u_\cdot^y) &:= \mathbb{E}\left[\int_0^T f^x(t, \Theta^y, Z_t, \Theta_t^x)dt + h^x(\theta_T^y, \theta_T^x)\right], \\ J^y(u_\cdot^y; u_\cdot^x) &:= \mathbb{E}\left[\int_0^T f^y(t, \Theta^y, Z_t, \Theta_t^x)dt + h^y(\theta_0^y, \theta_0^x)\right]. \end{aligned}$$

B.2.2.1 The mean-field type game

Consider the following situation. Within each crowd, *pedestrians cooperate*, but on a group-level, the *crowds compete*. This constitutes a so-called *mean-field type game* between the crowds. A Nash equilibrium in the game is a pair of feasible controls, $(\hat{u}^x_\cdot, \hat{u}^y_\cdot) \in \mathcal{U}^x \times \mathcal{U}^y$, satisfying the inequalities

$$\begin{cases} J^x(u_\cdot; \hat{u}^y_\cdot) \geq J^x(\hat{u}^x_\cdot; \hat{u}^y_\cdot), & \forall\, u_\cdot \in \mathcal{U}^x, \\ J^y(u_\cdot; \hat{u}^x_\cdot) \geq J^y(\hat{u}^y_\cdot; \hat{u}^x_\cdot), & \forall\, u_\cdot \in \mathcal{U}^y. \end{cases} \tag{7}$$

The next result is a Pontryagin's type stochastic maximum principle, and yields necessary and sufficient conditions for any control pair satisfying (7). A proof is provided in Appendix B.7.

Theorem 2. *Suppose that $(\hat{u}^x_\cdot, \hat{u}^y_\cdot)$ is a Nash equilibrium, i.e. satisfies* (7)*, and let the regularity assumptions of Lemma 5 be in force. Denote the corresponding state processes, given by* (6)*, by $\hat{X}_\cdot$ and $(\hat{Y}_\cdot, \hat{Z}_\cdot)$ respectively. Let $(p^{xx}_\cdot, q^{xx}_\cdot, q^{xy}_\cdot)$, $p^{xy}_\cdot$, $p^{yy}_\cdot$ and $(p^{yx}_\cdot, q^{yx}_\cdot, q^{yy}_\cdot)$ solve the adjoint equations*

$$\begin{cases} dp^{xx}_t = -\left\{\partial_x \hat{H}^x_t + \mathbb{E}\left[{}^*(\partial_{\mu^x} \hat{H}^x_t)\right]\right\} dt + q^{xx}_t dB^x_t + q^{xy}_t dB^y_t, \\ dp^{xy}_t = -\left\{\partial_y \hat{H}^x_t + \mathbb{E}\left[{}^*(\partial_{\mu^y} \hat{H}^x_t\right]\right\} dt - \partial_z \hat{H}^x_t dB_t, \\ dp^{yy}_t = -\left\{\partial_y \hat{H}^y_t + \mathbb{E}\left[{}^*(\partial_{\mu^y} \hat{H}^y_t)\right]\right\} dt - \partial_z \hat{H}^y_t dB_t, \\ dp^{yx}_t = -\left\{\partial_x \hat{H}^y_t + \mathbb{E}\left[{}^*(\partial_{\mu^x} \hat{H}^y_t)\right]\right\} dt + q^{yx}_t dB^x_t + q^{yy}_t dB^y_t, \\ \quad p^{xx}_T = -\left\{\partial_x \hat{h}^x + \mathbb{E}\left[{}^*(\partial_{\mu^x} \hat{h}^x)\right]\right\}, \quad p^{xy}_0 = 0, \\ \quad p^{yy}_0 = \partial_y \hat{h}^y + \mathbb{E}\left[{}^*(\partial_{\mu^y} \hat{h}^y)\right], \quad p^{yx}_T = 0, \end{cases} \tag{8}$$

where H^i, $i \in \{x, y\}$, is the Hamiltonian, defined by

$$\begin{aligned} &H^i(\omega, t, y, \mu^y, v, z, x, \mu^x, u, p^{ix}, p^{iy}, q^{ix}) \\ &\quad := \sum_{j \in \{x,y\}} b^j(\omega, t, y, \mu^y, v, z, x, \mu^x, u) p^{ij} \\ &\qquad\quad + \sigma^x(\omega, t, y, \mu^y, x, \mu^x) q^{ix} - f^i(\omega, t, y, \mu^y, v, z, x, \mu^x, u). \end{aligned} \tag{9}$$

Then

$$\begin{aligned} \hat{u}^x_t &= \underset{v \in U^x}{\arg\max}\ H^x(t, \hat{\Theta}^y_t, \hat{Z}_t, \hat{\theta}^x_t, v, p^{xx}_t, p^{xy}_t, q^{xx}_t), \quad a.e.\text{-}t,\ \mathbb{P}\text{-}a.s. \\ \hat{u}^y_t &= \underset{v \in U^y}{\arg\max}\ H^y(t, \hat{\theta}^y_t, v, \hat{Z}_t, \hat{\Theta}^x_t, p^{yx}_t, p^{yy}_t, q^{yx}_t), \quad a.e.\text{-}t,\ \mathbb{P}\text{-}a.s. \end{aligned} \tag{10}$$

If furthermore H^i is concave in $(y, \mu^y, v, z, x, \mu^x, u)$, and h^i is convex in (y, μ^y, x, μ^x, z), $i \in \{x, y\}$, then any feasible control pair satisfying (10) *is a Nash equilibrium in the mean-field type game.*

Remark 3. *Note that in the mean-field type game, the tagged can be thought of as a major player, influencing the ordinary crowd. Furthermore, the central planner has access to a model of the ordinary crowd and this is necessary for the determination of the equilibrium control.*

B.2.2.2 Optimal control of mean-field type

If the central planner does not have access to a model of the crowd of ordinary pedestrians, any interaction with the surrounding environment will then enter the tagged model as a random signal. This is covered by the (ω, t)-dependence of b^y and f^y. The mean-field type game then reduces to a so-called *optimal control problem of mean-field type*,

$$\begin{cases} \min\limits_{u^y_\cdot \in \mathcal{U}^y} \mathbb{E}\left[\int_0^T f^y(t, \Theta^y_t, Z_t)dt + h^y(\theta^y_0, Z_0)\right], \\ \quad \text{s.t.} \;\; dY_t = b^y(t, \Theta^y_t, Z_t)dt + Z_t dB_t, \\ \qquad\quad\; Y_T = y_T. \end{cases} \tag{11}$$

Problem (11) is a special case of (7) and necessary and sufficient conditions follow as a corollary to Theorem 2.

Corollary 4. *Suppose that $\hat{u}^y_\cdot$ solves* (11) *and denote the corresponding tagged state $(\hat{Y}_\cdot, \hat{Z}_\cdot)$. Let $p_\cdot$ solve the adjoint equation*

$$\begin{cases} dp_t = -\left\{\partial_y \mathcal{H}(t, \hat{\Theta}^y_t, \hat{Z}_t, p_t) + \mathbb{E}\left[{}^*\left(\partial_\mu \mathcal{H}(t, \hat{\Theta}^y_t, \hat{Z}_t, p_t)\right)\right]\right\} dt \\ \qquad\qquad - \partial_z \mathcal{H}(t, \hat{\Theta}^y_t, \hat{Z}_t, p_t) dB_t, \\ p_0 = \partial_y h^y(\hat{\theta}^y_0) + \mathbb{E}\left[{}^*\left(\partial_\mu h^y(\hat{\theta}^y_0)\right)\right], \end{cases}$$

where $\mathcal{H}$ is the Hamiltonian

$$\mathcal{H}(\omega, t, y, \mu, v, z, p) := b^y(\omega, t, y, \mu, v, z)p - f^y(t, y, \mu, v, z).$$

Then

$$\hat{u}^y_t = \arg\max_{v \in U^y} \mathcal{H}(t, \hat{\theta}^y_t, v, \hat{Z}_t, p_t), \quad \textit{a.e.-}t,\ \mathbb{P}\textit{-a.s.} \tag{12}$$

If moreover $\mathcal{H}$ is concave in (y, μ, v, z) and that h is convex in (y, μ), then any feasible $\hat{u}^y_\cdot$ that satisfies (12)*, almost surely for a.e. t, is an optimal control to* (11)*.*

B.3 Simulations

This section is devoted to numerical simulations of the tagged model under various external influences and internal preferences. First, two scenarios in the optimal control version of the model are considered, including preferences on velocity, avoidance and interaction via the mean position of the group. Secondly, asymmetric bidirectional flow is simulated with the full mean-field type game version of the model. None of the parameters used in the simulations stem from real world measurements, but the velocity profiles and the asymmetric bidirectional flow are compared qualitatively with the experimental studies [42] and [50].

B.3.1 Optimal control: Keeping a tagged group together

In this scenario, the common goal of the tagged pedestrians is to stay close to the mean position of the group, while conserving energy and initiating in the proximity of $y_0 \in \mathbb{R}^2$. Distance to the group mean is one of the simplest mean-field effects that can be considered. Nonetheless it is a distribution-dependent quantity, and the control problem characterizing tagged pedestrian behavior in this scenario is the nonstandard optimization problem (13). The components of (13) are presented in Table 2. There is no surrounding crowd present, only tagged pedestrians occupy the space.

$$\begin{cases} \min\limits_{u_\cdot \in \mathcal{U}} \dfrac{1}{2}\mathbb{E}\Big[\displaystyle\int_0^T \Big(\lambda_{\text{cont}}\,(u^y_t)^2 + \lambda_{\text{attr}}\,(Y_t - \mathbb{E}[Y_t])^2\Big)\,dt \\ \qquad\qquad\qquad\qquad\qquad + \lambda_{\text{init}}(Y_0 - y_0)^2\Big], \\ \text{s.t.}\;\; dY_t = (u^y_t + \lambda_{\text{noise}} B_t)dt + Z_t dB_t, \\ \qquad Y_T = y_T. \end{cases} \tag{13}$$

The scenario is simulated for two sets of parameters, see Table 3. The mean-field BFSDE system of equations characterizing optimal behavior, given by

Velocity component	**Form**
Internal velocity (control)	$u^y_\cdot \in \mathcal{U}^y$
Acceleration noise	$\lambda_{\text{noise}} B^y_\cdot$
Preference (penalty)	**Form**
Energy usage per unit time	$\lambda_{\text{cont}} \left(u^y_\cdot\right)^2$
Distance from group mean per unit time	$\lambda_{\text{attr}} \left(Y_\cdot - \mathbb{E}[Y_\cdot]\right)^2$
Distance from $y_0 \in \mathbb{R}^2$ at $t = 0$	$\lambda_{\text{init}}(Y_0 - y_0)^2$

Table 2: Keeping a tagged group together: Components.

Corollary 4, is solved with the least-square Monte Carlo method of [9]. The result is presented in Figure 1. The group walks approximately on the straight line from the starting area to the target point. Remember that the initial position of the tagged is chosen rationally by solving (13). There is a trade-off between starting close to y_0 and walking with high speed, and the groups rationally initiates not at y_0, but somewhere between y_0 and y_T. The group that prefers proximity to other group members does indeed move in a more compact formation. The difference is clearly seen when looking at the mean distance to the center of the tagged group, see Figure 2.

	λ_{noise}	λ_{cont}	λ_{attr}	λ_{init}	$\mathbb{E}[y_0]$	y_T	T
Set 1	1	50	50	10	[0.1,0.1]	[2,2]	1
Set 2	1	50	0	10	[0.1,0.1]	[2,2]	1

Table 3: Keeping a tagged group together: Parameter values. Set 1 are the parameters use in the scenario with distance-to-mean penalty, Set 2 without distance-to-mean penalty. The preferred initial position is normally distributed around $[0.1, 0.1]$.

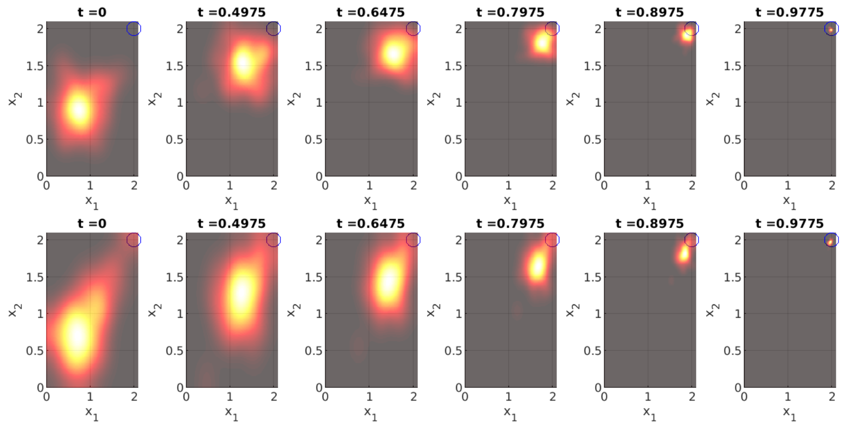


Figure 1: Top row: Crowd density evolution when $\lambda_2 = 50$. Bottom row: Crowd density evolution when $\lambda_2 = 0$.

B.3.2 Optimal control: Desired velocity

Linear-quadratic scenarios have accessible closed form solutions. We want to mention the case where the tagged's goal is to move at her desired velocity v_{des}, similar to what was originally introduced as *the desired speed and direction* in [27]. This scenario is an important special case, since desired velocity is measurable in live experiments. See for example [42] for the speed profile of pedestrian walking in a straight corridor, starting from standing still. The scenario is formulated as standard optimal control problem, (14), and the setting is summarized in Table 4.

$$\begin{cases} \min_{u \in \mathcal{U}[0,T]} \frac{1}{2}\mathbb{E}\Big[\int_0^T \Big(\lambda_{\text{cont}}\,(u_t^y)^2 \\ \qquad\qquad + \lambda_{\text{des}}\,(u_t^y - v_{\text{des}}(t))^2 + \lambda_{\text{rep}}(Y_t - Q)^2\Big)dt\Big], \\ \text{s.t.}\;\; dY_t = (u_t^y + \lambda_{\text{noise}} B_t^y)dt + Z_t dB_t, \\ \qquad Y_T = y_T. \end{cases} \tag{14}$$

In view of Corollary 4, the optimal control is

$$\hat{u}_t^y = \frac{p_t + \lambda_{\text{des}} v_{\text{des}}(t)}{\lambda_{\text{cont}} + \lambda_{\text{rep}}},$$

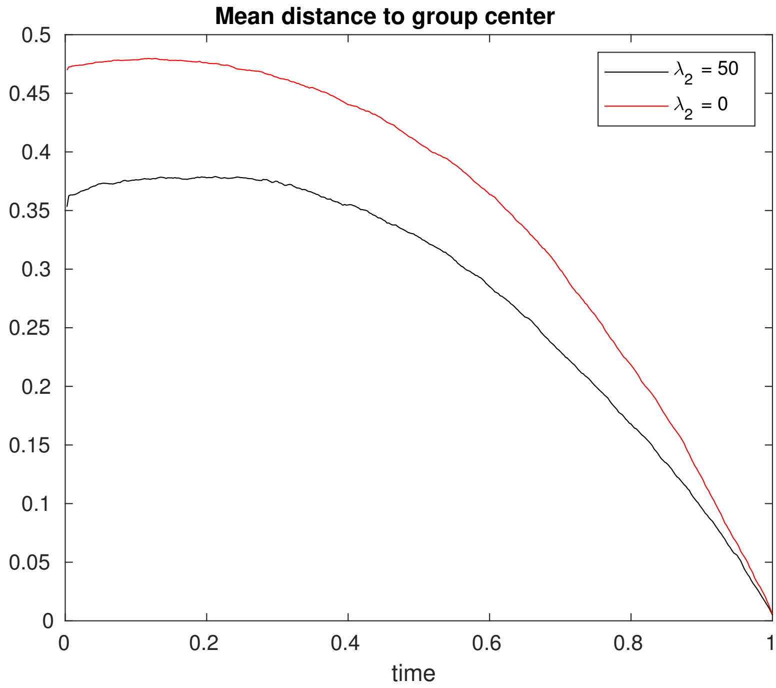


Figure 2: A tagged group acting under preferences with a penalty on the distance to group center indeed moves in a more compact formation than a group without the penalty.

With the ansatz $\hat{Y}_t = \gamma(t)p_t + \eta(t)B_t^y + \theta_t$, $\gamma(T) = \eta(T) = 0$ and $\theta(T) = y_T$, a matching argument gives the optimally controlled dynamics up to a system of ODEs:

$$\begin{cases} \frac{d}{dt}\gamma(t) = -\lambda_{\text{rep}}\gamma(t)^2 + \frac{1}{\lambda_{\text{cont}}+\lambda_{\text{des}}}, & \gamma(T) = 0, \\ \frac{d}{dt}\eta(t) = -\lambda_{\text{rep}}\gamma(t)\eta(t) + \lambda_{\text{noise}}, & \eta(T) = 0, \\ \frac{d}{dt}\theta(t) = -\lambda_{\text{rep}}\gamma(t)(\theta(t) - Q) + \frac{\lambda_{\text{des}}v_{\text{ des}}(t)}{\lambda_{\text{cont}}+\lambda_{\text{des}}}, & \theta(T) = y_T, \\ \hat{Z}_t = \eta(t). \end{cases}$$

In Figure 3, the simulated tagged crowd density is presented for two values of Q. The desired velocity is set to be negative in both directions for $t \in [0, T/2]$, and positive for $t \in (T/2, T]$, which corresponds to a preference to first move south-west during the first half of the time period and then turn

Velocity component	**Form**
Internal velocity (control)	$u^y_\cdot \in \mathcal{U}^y$
Acceleration noise	$\lambda_{\text{noise}} B^y_\cdot$
Preference	**Form**
Energy usage per unit time	$\lambda_{\text{cont}} \left(u^y_\cdot\right)^2$
Uncomfortable velocity per unit time	$\lambda_{\text{des}} \left(u^y_\cdot - v_{\text{des}}(t)\right)^2$
Distance from $Q \in \mathbb{R}^2$ per unit time	$\lambda_{\text{rep}}(Y_\cdot - Q)^2$

Table 4: Desired velocity: Components

around and move north-east. The parameter λ_{rep} is set to a negative value, hence the tagged prefers to avoid Q. The parameters used in the simulation are summarized in Table 5. The trade-off between walking in the desired velocity and walking close to the diamond Q before reaching the target circle is clearly visible in the plot. Recall that the initial position is determined by the optimization procedure. In this scenario, there is no preference on initial position and the tagged group compensates for the location of Q by changing its initial position!

In [42] the average time-dependent velocity of a pedestrian initially standing still is measured experimentally in the absence of interactions. The result is a relationship between speed and time, than can be used as data for v_{des}. Approximating the graph presented in [42] with

$$v_{\text{des}}(t) = \max\{0.1, \arctan(\pi t - 1.6)\}, \tag{15}$$

the scenario is simulated with the two presented in Table 6. The result is presented in Figure 4. The parameter set with a higher penalty on from deviation from desired velocity naturally results in a velocity profile closer to $v_{\text{des}}(t)$.

	λ_{noise}	λ_{cont}	λ_{des}	λ_{rep}
Set 1	0.5	0.5	1	-2
Set 2	0.5	0.5	1	-2
	Q	$v_{\text{des}}(t)$	y_T	T
Set 1	$-[0.5, 0.5]$	$\text{sign}\left(t - \frac{T}{2}\right)[3, 3]$	$[0.1, 0.1]$	1
Set 2	$[1.5, 1.5]$	$\text{sign}\left(t - \frac{T}{2}\right)[3, 3]$	$[0.1, 0.1]$	1

Table 5: Desired velocity: Parameter values.

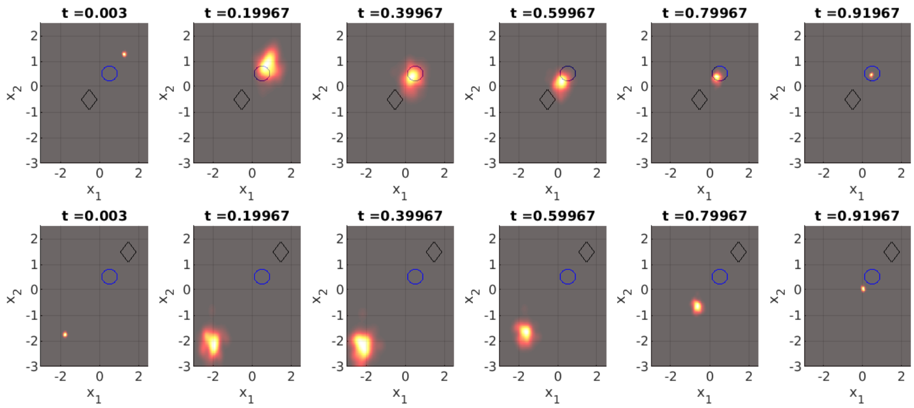
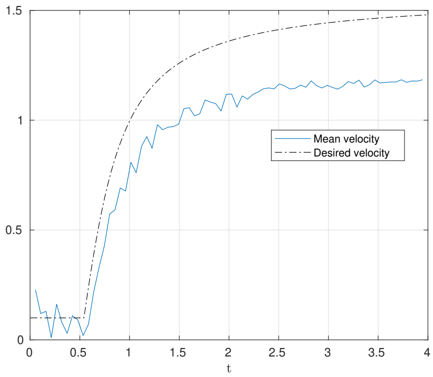


Figure 3: *Top row:* Desired velocity with Set 1 parameters from Table 5. *Bottom row:* Desired velocity with Set 2 parameters from Table 5. The tagged crowd moves first south-west and then north-east, following their desired velocity, while avoiding the diamond on their way to the circle.

	λ_{noise}	λ_{cont}	λ_{des}	λ_{rep}	Q	$v_{\text{des}}(t)$	y_T	T
Set 3	0.1	0.5	2	0	$[0, 0]$	Eq. (15)	$[0, 0]$	4
Set 4	0.1	0.5	10	0	$[0, 0]$	Eq. (15)	$[0, 0]$	4

Table 6: Desired velocity: Parameter values.



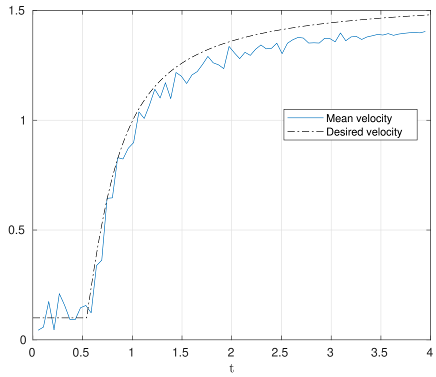


Figure 4: *Left:* Desired velocity with Set 3 parameters from Table 6. Right: Desired velocity with Set 4 from Table 6. The mean velocity measured in the scenario (14) compared to the desired velocity (15).

B.3.3 Mean-field type game: Asymmetric bidirectional flow

Consider now a scenario where ordinary pedestrians initiate at $x_0 \in \mathbb{R}^2$, close $y_T \in \mathbb{R}^2$, the location of an incident. They begin to walk towards the safe spot $x_T \in \mathbb{R}^2$. A tagged pedestrian is to end up at the location of the incident y_T at time $t = T$. The tagged pedestrian is repelled by the mean of the ordinary pedestrian crowd, while the ordinary pedestrian crowd is repelled by the tagged pedestrian. This scenario is implemented as the MFTG (16), summarized in Table 7.

$$\begin{cases} J^x(u^x_\cdot; u^y_\cdot) = \dfrac{1}{2}\mathbb{E}\Bigg[\displaystyle\int_0^T \Big(\lambda^x_{\text{cont}} \, (u^x_t)^2 + \lambda^x_{\text{rep}}(X_t - Y_t)^2\Big)\, dt \\ \qquad\qquad\qquad\qquad\qquad\qquad + \lambda^x_{\text{term}}(X_T - x_T)^2\Bigg], \\ dX_t = u^x_t dt + \sigma dB^x_t, \quad \sigma \in \mathbb{R}, \quad X_0 = x_0, \\ J^y(u^y_\cdot; u^x_\cdot) = \dfrac{1}{2}\mathbb{E}\Bigg[\displaystyle\int_0^T \Big(\lambda^y_{\text{cont}} \, (u^y_t)^2 + \lambda^y_{\text{red}}(Y_t - \mathbb{E}[X_t])^2\Big)\, dt \\ \qquad\qquad\qquad\qquad\qquad\qquad + \lambda^y_{\text{init}}(Y_0 - y_0)^2\Bigg], \\ dY_t = (u^y_t + \lambda^y_{\text{noise}} dB^y_t) + Z_t dB_t. \end{cases} \tag{16}$$

ORDINARY PEDESTRIAN	
Velocity component	**Form**
Internal velocity (control)	$u^x_\cdot \in \mathcal{U}^x$
Diffusion noise level	σ
Preference	**Form**
Energy usage per unit time	$\lambda^x_{\text{cont}}\,(u^x_\cdot)^2$
Repulsion per unit time	$\lambda^x_{\text{rep}}(X_\cdot - Y_\cdot)^2$
Proximity to x_T at $t = T$	$\lambda^x_{\text{term}}(X_T - x_T)^2$
TAGGED PEDESTRIAN	
Velocity component	**Form**
Internal velocity (control)	$u^y_\cdot \in \mathcal{U}^y$
Acceleration noise	$\lambda^y_{\text{noise}} B^y_\cdot$
Preference	**Form**
Energy usage per unit time	$\lambda^y_{\text{cont}}\,(u^y_\cdot)^2$
Repulsion per unit time	$\lambda^y_{\text{rep}}(Y_\cdot - \mathbb{E}[X_\cdot])^2$
Proximity to y_0 at $t = 0$	$\lambda^y_{\text{init}}(Y_0 - y_0)^2$

Table 7: Asymmetric bidirectional flow: Components

In Figure 5 and Figure 6 the scenario is simulated for the parameter sets presented in Table 8. In Figure 5, the ordinary and the tagged do not have to cross paths to go from their initial to their terminal positions. The simulated paths (top plot of Figure 5) are similar in shape to both the outcome of the corridor experiment under 'condition 3' (no obstacle) of [42] and the BFR-SSL experiment of [50]. These experimental studies were conducted in a controlled environment that is outside the tagged pedestrian model of this paper. Anyhow, the tagged model replicates the separation of lanes in a bidirectional pedestrian flow and the randomness that pedestrian motion exhibits. The density snapshots (bottom row of Figure 5) reveal that in the simulated scenario, the tagged moves at almost constant velocity towards y_T, while the ordinary group lingers a while at x_0 before it starts to move towards x_T. In Figure 6, the tagged's and the ordinary's straight path from initial position to target cross each other. In this scenario, the ordinary pedestrians resolve this by taking walking in a half-circle around the tagged, before moving towards their preferred terminal position x_T.

TAGGED PED.	λ^y_{noise}	λ^y_{cont}	λ^y_{rep}	λ^y_{init}	y_0	$\mathbb{E}[y_T]$	T
Bidirectional flow	0.7	1	-2	3	[0,1]	[10,0]	1
Twist	0.7	1	-1	3	[0,-3]	[10,-1]	1
ORDINARY PED.	σ	λ^x_{cont}	λ^x_{rep}	λ^x_{end}	$\mathbb{E}[x_0]$	x_T	T
Bidirectional flow	0.7	1	-1	10	[10,-1]	[0,0]	1
Twist	0.7	1	-1.7	10	[10,-2]	[0,0]	1

Table 8: Asymmetric bidirectional flow: Parameters values. The initial and terminal constraints are normally distribution with mean tabled above, and standard deviation 0.3 and 0.1 for the tagged and ordinary pedestrian, respectively.

B.4 Concluding remarks and research perspectives

A mean-field type game model for so-called tagged pedestrian motion has been presented and the reliability of the model has been studied through simulations. To perform simulations, necessary and sufficient conditions for a Nash equilibrium are provided in Theorem 2. The theorem is proven under quite restrictive conditions on involved coefficient functions. However, necessary conditions for a Nash equilibrium in similar games are available

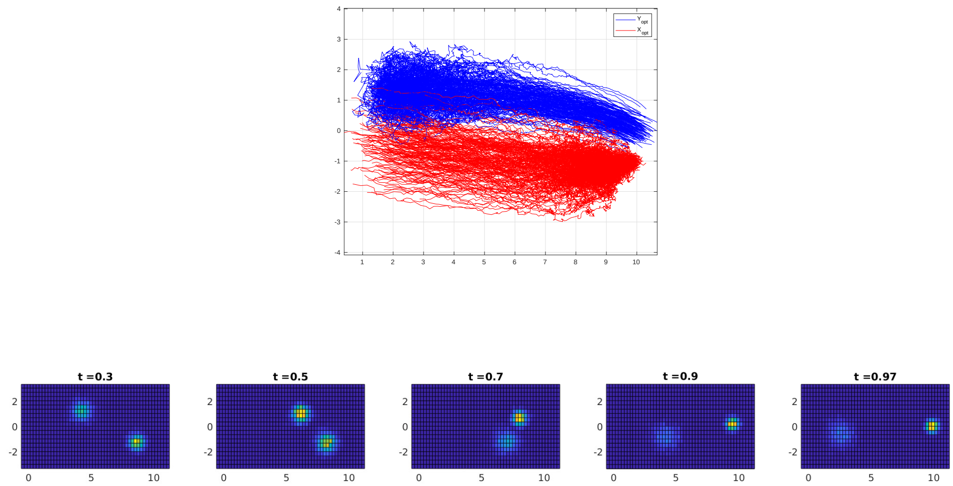


Figure 5: With parameter set 'Bidirectional flow', Table 8. *Top row:* Pedestrian paths. *Bottom row:* Pedestrian density.

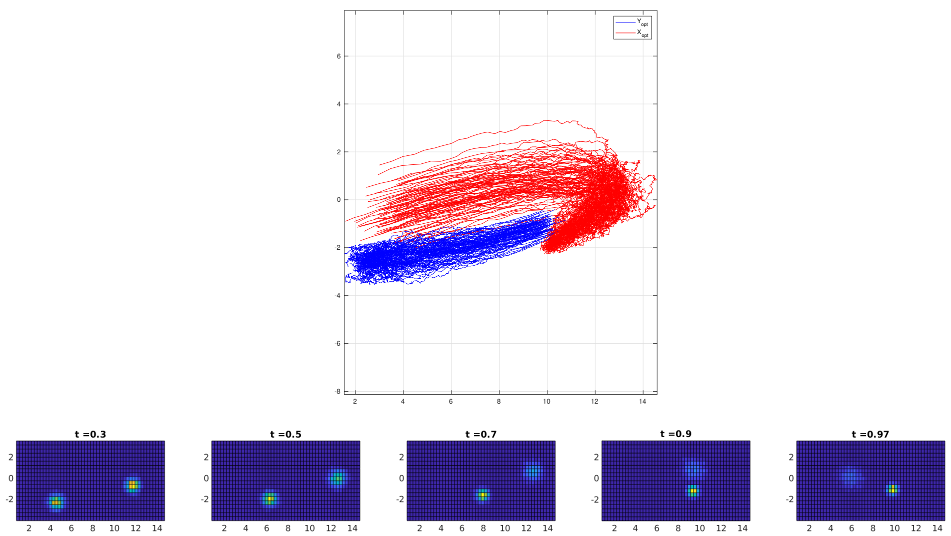


Figure 6: With parameter set 'Twist', Table 8. *Top row:* Pedestrian paths. *Bottom row:* Pedestrian density.

under less restrictive conditions and since our proof follows a standard path, the conditions can certainly be relaxed. The model captures both game-like and minor agent-type scenarios. In the latter, the tagged cannot effect crowd movement while in the former, the tagged and the surrounding crowd have conflicting interests, interact, and compete. The rational pedestrian behavior in the competitive game-like scenario is to use an equilibrium strategy.

There are many variations to the mean-field type game approach. The scenarios that have been considered in this paper fall into two rather extreme categories; our pedestrians have acted under either *basic* or *optimal rationality*. The terms are defined in [20]. When pedestrians neither have access to information about their surroundings nor the ability to anticipate crowd motion, the best they can do is to implement a control policy based on their own position and their target position. This is a basic level of rationality. If the full model is available and pedestrians cooperate, they can implement a control policy of optimal rationality. If the tagged can observe crowd densities at each instant in time *but not anticipate future crowd movement*, dynamic pedestrian preference may be modeled as a set of control problems: for each $\tau \in [0,T]$,

$$\begin{cases} \min\limits_{u^y_\cdot(\tau)\in\mathcal{U}^y_\tau} \mathbb{E}\left[\int_\tau^T f^y(t,Y_t,\mathcal{L}(Y_\tau),u^y_t(\tau))dt + \mathbb{I}\{\tau=0\}h^y(Y_0,\mathcal{L}(Y_0))\right], \\ \quad\text{s.t. } dY_t = b^y(t,Y_t,\mathcal{L}(Y_\tau),Z_t,u^y_t(\tau))dt + Z_t dB_t, \\ \quad\phantom{\text{s.t. }} Y_T = y_T, \end{cases}$$

where $\mathcal{U}^y_\tau$ is defined in the same way as $\mathcal{U}^y$, but with the interval $[0,T]$ replaced by $[\tau,T]$, cf. (3). This is an intermediate level between basic and optimal rationality. Pedestrian decision making can also be modeled as a decentralized mechanism, i.e. instead of cooperating within the crowd, the pedestrians compete in a game-like manner. Decentralized crowd formation can be modeled as a MFG:

(i) Fix a deterministic function $\mu_\cdot : [0,T] \to \mathcal{P}_2(\mathbb{R}^d)$.

(ii) Solve the stochastic control problem

$$\begin{cases} \min\limits_{u^y_\cdot\in\mathcal{U}^y} \mathbb{E}\left[\int_0^T f^y(t,Y_t,\mu_t,u^y_t)dt + h^y(Y_0,\mu_0)\right], \\ \quad\text{s.t. } dY_t = b(t,Y_t,\mu_t,Z_t,u^y_t)dt + Z_t dB_t, \\ \quad\phantom{\text{s.t. }} Y_T = y_T. \end{cases}$$

(iii) Determine the function $\hat{\mu}_\cdot : [0,T] \to \mathcal{P}_2(\mathbb{R}^d)$ such that $\hat{\mu}_t$ is the marginal law of the optimally controlled state from Step (ii).

Furthermore, minimal exit time (evacuation) problems can be posed at all levels of rationality.

The numerical simulation of mean-field BFSDE systems has been done either with the least-square Monte Carlo method of [8], or by a reduction to a system of ODEs by the method of matching. The downside with the least-square Monte Carlo method is that it is not clear which basis functions to use. The matching method is feasible only for linear-quadratic problems. Other simulation approaches include deep learning [47] and fixed-point schemes [22]. Fast and stable numerical solvers for mean-field BFSDEs beyond the linear-quadratic case is an area of research that would benefit many applied fields. In pedestrian crowd modeling, improved solvers would facilitate simulation whenever effects like congestion, crowd aversion, and anisotropic preferences are present.

B.5 Appendix 1: Mean-field BFSDE

Given a control pair $(u^x_\cdot, u^y_\cdot) \in \mathcal{U}^x \times \mathcal{U}^y$, systems of the form (6) have been studied in the context of optimal control of mean-field type, where they naturally arise as necessary optimality conditions. This appendix summarizes some of the results on existence and uniqueness of solutions to MF-BFSDEs. Let

$$\mathbb{H}^{2,d} := \left\{ V_\cdot\ \mathbb{R}^d\text{-valued and prog. meas.} : \mathbb{E}\left[\int_0^T |V_s|^2 ds\right] < \infty \right\},$$
$$\mathbb{S}^{2,d} := \left\{ V_\cdot\ \mathbb{R}^d\text{-valued and prog. meas.} : \mathbb{E}\left[\sup_{s\in[0,T]} |V_s|^2\right] < \infty \right\},$$

and recall that, in the case of a fixed control pair, b^x and b^y are functions of

$$(\omega, t, y, \mu^y, z, x, \mu^x) \in \Omega \times [0,T] \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^{d\times(w_x+w_y)} \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d),$$

and σ^x is a function of

$$(\omega, y, \mu^y, z, x, \mu^x) \in \Omega \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^{d\times(w_x+w_y)} \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d).$$

B.5.1 Quadratic-type constraints

For $t \in [0,T]$, $\mu^y, \mu^x \in \mathcal{P}_2(\mathbb{R}^d)$, $x, \bar{x}, y, \bar{y} \in \mathbb{R}^d$ and $z, \bar{z} \in \mathbb{R}^{d\times(w_x+w_y)}$, let

$$\begin{aligned}
&\mathcal{A}(t, y, \bar{y}, \mu^y, v, z, \bar{z}, x, \bar{x}, \mu^x) \\
&= (b^y(t, y, \mu^y, v, z, x, \mu^x) - b^y(t, \bar{y}, \mu^y, v, \bar{z}, \bar{x}, \mu^x)) \cdot (y - \bar{y}) \\
&\quad + (b^x(t, y, \mu^y, v, z, x, \mu^x) - b^x(t, \bar{y}, \mu^y, v, \bar{z}, \bar{x}, \mu^x)) \cdot (x - \bar{x}) \\
&\quad + (\sigma^x(t, y, \mu^y, z, x, \mu^x) - \sigma^x(t, \bar{y}, \mu^y, \bar{z}, \bar{x}, \mu^x)) \cdot (z - \bar{z}).
\end{aligned}$$

In [22], the authors provide conditions on $\mathcal{A}$ under which, alongside standard assumptions, (6) has a unique solution $(X., Y., Z.)$ in $\mathbb{H}^{2,d} \times \mathbb{H}^{2,d} \times \mathbb{H}^{2,d\times(w_x+w_y)}$ for all T.

B.5.2 Small time constraint

Under standard Lipschitz- and linear growth-conditions and for a nondegenerate diffusion σ^x, (6) has a unique solution $(X., Y., Z.) \in \mathbb{S}^{2,d} \times \mathbb{S}^{2,d} \times \mathbb{H}^{2,d\times(w_x+w_y)}$ for small enough T by [17]. The bound on T depends on the Lipschitz coefficients. In [17], the authors provide an example where uniqueness fails.

B.6 Appendix 2: Differentiation of $f : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$

The differentiation of functions taking probability measures is handled with the lifting technique, introduced by P.-L. Lions and outlined in for example [16, 13, 17]. Assume that the underlying probability space is rich enough, so that for every $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ there is a random variable $Y \in L^2_{\mathcal{F}}(\Omega; \mathbb{R}^d)$ such that $\mathcal{L}(Y) = \mu$. A probability space with this property is $([0,1], \mathcal{B}([0,1]), dx)$. Under this assumption, any function $f : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ induces a function $F : L^2(\mathbb{R}^d) \to \mathbb{R}$ so that $F(Y) := f(\mathcal{L}(Y))$. The Fréchet derivative of F at Y, whenever it exists, is the continuous linear functional $DF[Y]$ that satisfies

$$F(Y') - F(Y) = \mathbb{E}\left[DF[Y] \cdot (Y' - Y)\right] + o(\|Y' - Y\|_2)$$

for all $Y' \in L^2_{\mathcal{F}}(\Omega; \mathbb{R}^d)$. Riesz' Representation Theorem yields uniqueness of $DF[Y]$. Furthermore, there exists a Borel function $\varphi[\mu] : \mathbb{R}^d \to \mathbb{R}^d$, independent of the version of Y, such that $DF[Y] = \varphi[\mathcal{L}(Y)](Y)$, see [16]. Therefore

$$f(\mathcal{L}(Y')) - f(\mathcal{L}(Y)) = \mathbb{E}\left[\varphi[Y](Y)(Y' - Y)\right] + o(\|Y' - Y\|) \tag{17}$$

for all $Y' \in L^2_{\mathcal{F}}(\Omega;\mathbb{R}^d)$. Denote $\partial_\mu f(\mu;x) := \varphi[\mu](x)$ for $x \in \mathbb{R}^d$ and let $\partial_\mu f(\mathcal{L}(Y)) := \partial_\mu f(\mathcal{L}(Y);Y)$. The following identity characterizes derivatives with respect to elements in $\mathcal{P}_2(\mathbb{R}^d)$,

$$DF[Y] = \varphi[\mathcal{L}(Y)](Y) = \partial_\mu f(\mathcal{L}(Y)).$$

Equation (17) is the Taylor approximation of a function over $\mathcal{P}_2(\mathbb{R}^d)$. Consider now an f that besides the measure takes another argument, ξ. Then

$$f(\xi,\mathcal{L}(Y')) - f(\xi,\mathcal{L}(Y)) = \mathbb{E}\left[\partial_\mu f(\widetilde{\xi},\mathcal{L}(Y);Y)(Y'-Y)\right] + o(\|Y-Y'\|_2),$$

where the expectation is taken over *non-tilded random variables.* This is abbreviated as

$$\mathbb{E}\left[\partial_\mu f(\widetilde{\xi},\mathcal{L}(Y);Y)(Y'-Y)\right] =: \mathbb{E}\left[(\partial_\mu f(\xi,\mathcal{L}(Y)))^*(Y'-Y)\right]. \tag{18}$$

Note that $\mathcal{L}(Y)$ is deterministic and the expectation is only taken over the 'directional argument' of $\partial_\mu f$, that is Y. Also, the expected value (18) is stochastic, since it is *not taken over* ξ. Taking another expectation and changing the order of integration yields

$$\mathbb{E}\left[\widetilde{\mathbb{E}}\left[\partial_\mu f(\widetilde{\xi},\mathcal{L}(Y);Y)\right](Y'-Y)\right],$$

where the tilded expectation is taken *over tilded random variables.* This is abbreviated as

$$\widetilde{E}\left[\partial_\mu f(\widetilde{\xi},\mathcal{L}(Y);Y)\right] =: \mathbb{E}\left[{}^*(\partial_\mu f(\xi,\mathcal{L}(Y)))\right]$$

Example from Section B.3.1 The following measure derivative appears in Section 3.1,

$$\mathbb{E}\left[{}^*\left(\partial_\mu\left(Y_t - \int_{\mathbb{R}^2} y\mathcal{L}(Y_t)(dy)\right)^2\right)\right].$$

Note that $\left(Y_t - \int_{\mathbb{R}^2} y\mathcal{L}(Y_t)(dy)\right)^2 = (Y_t - \mathbb{E}[M])^2 =: F(M)$ where M is a random variable with law $\mathcal{L}(Y_t)$. By the Taylor expansion, $DF[M] = 2(Y_t - \mathbb{E}[M])$ and therefore

$$\mathbb{E}\left[{}^*\left(\partial_\mu\left(Y_t - \int_{\mathbb{R}^2} y\mathcal{L}(Y_t)(dy)\right)^2\right)\right] = \mathbb{E}\left[2(Y_t - \mathbb{E}[M])\right] = 0.$$

B.7 Appendix 3: Proof of Theorem 2

Let $\bar{u}^{x,\epsilon}_{\cdot}$ be a spike variation of $\hat{u}^x_{\cdot}$,

$$\bar{u}^{x,\epsilon}_t := \begin{cases} \hat{u}^x_t, & t \in [0,T]\backslash E_\epsilon, \\ u_t, & t \in E_\epsilon, \end{cases}$$

where $u_{\cdot} \in \mathcal{U}^x$ and E_ϵ is a subset of $[0,T]$ of measure ϵ. Given the control pair $(\bar{u}^{x,\epsilon}_{\cdot}, \hat{u}^y_{\cdot})$, denote the corresponding solution to the state equation (6) by $\bar{X}^\epsilon_{\cdot}$ and $(\bar{Y}^\epsilon_{\cdot}, \bar{Z}^\epsilon_{\cdot})$. To ease notation, let for $\vartheta \in \{b^x, b^y, \sigma^x, f^x, f^y, h^x, h^y\}$,

$$\begin{aligned} &\bar{\vartheta}^\epsilon_t := \vartheta(t, \bar{\theta}^{y,\epsilon}_t, \hat{u}^y_t, \bar{Z}^\epsilon_t, \bar{\Theta}^{x,\epsilon}_t), \quad \hat{\vartheta}^\epsilon_t := \vartheta(t, \hat{\Theta}^y_t, \hat{Z}_t, \hat{\Theta}^x_t), \\ &\delta_x \vartheta(t) := \vartheta(t, \hat{\Theta}^y_t, \hat{Z}_t, \hat{\theta}^x_t, \bar{u}^{x,\epsilon}_t) - \hat{\vartheta}^\epsilon_t. \end{aligned} \tag{19}$$

Consider the ordinary pedestrian's potential loss, would she switch from the equilibrium control $\hat{u}^x_{\cdot}$ to the perturbed $\bar{u}^{x,\epsilon}_{\cdot}$,

$$J^x(\bar{u}^{x,\epsilon}_{\cdot}; \hat{u}^y_{\cdot}) - J(\hat{u}^x_{\cdot}; \hat{u}^y_{\cdot}) = \mathbb{E}\left[\int_0^T \left(\bar{f}^{x,\epsilon}_t - \hat{f}^x_t\right) dt + \bar{h}^{x,\epsilon}_T - \hat{h}^x_T\right].$$

A Taylor expansion of the terminal cost difference yields

$$\mathbb{E}\left[\bar{h}^{x,\epsilon}_T - \hat{h}^x_T\right] = \mathbb{E}\left[\left(\partial_y \hat{h}^x_t + \mathbb{E}[^*(\partial_\mu \hat{h}^x_T)]\right)\left(\bar{X}^\epsilon_t - \hat{X}_t\right)\right] + o\left(\|\bar{X}^\epsilon_T - \hat{X}_T\|_2\right),$$

Let $\widetilde{X}^x_{\cdot}$ and $(\widetilde{Y}^x_{\cdot}, \widetilde{Z}^x_{\cdot})$ be the first order variation processes, solving the linear BFSDE system

$$\begin{cases} d\widetilde{X}^x_t = \Big\{ \left(\partial_y \hat{b}^x_t + \mathbb{E}[^*(\partial_{\mu^y} \hat{b}^x_t)]\right) \widetilde{Y}^x_t + \partial_z \hat{b}^x_t \widetilde{Z}^x_t + \left(\partial_x \hat{b}^x_t + \mathbb{E}[^*(\partial_{\mu^x} \hat{b}^x_t)]\right) \widetilde{X}^x_t \\ \qquad\qquad + \delta_x b^x(t) 1_{E_\epsilon}(t) \Big\} dt + \\ \quad \Big\{ (\partial_y \hat{\sigma}^x_t + \mathbb{E}[^*(\partial_{\mu^y} \hat{\sigma}^x_t)]) \widetilde{Y}^x_t + \partial_z \hat{\sigma}^x_t \widetilde{Z}^x_t + (\partial_x \hat{\sigma}^x_t + \mathbb{E}[^*(\partial_{\mu^x} \hat{\sigma}^x_t)]) \widetilde{X}^x_t \Big\} dB^x_t, \\ d\widetilde{Y}^x_t = \Big\{ \left(\partial_y \hat{b}^y_t + \mathbb{E}[^*(\partial_{\mu^y} \hat{b}^y_t)]\right) \widetilde{Y}^x_t + \partial_z \hat{b}^y_t \widetilde{Z}^x_t + \left(\partial_x \hat{b}^y_t + \mathbb{E}[^*(\partial_{\mu^x} \hat{b}^y_t)]\right) \widetilde{X}^x_t \\ \qquad\qquad + \delta_x b^y(t) 1_{E_\epsilon}(t) \Big\} dt + \widetilde{Z}^x_t dB_t, \\ \widetilde{X}^x_0 = 0, \quad \widetilde{Y}^x_T = 0. \end{cases}$$

Lemma 5. *Assume that b^x are b^y are Lipschitz in the controls, that b^x, b^y, f^x, f^y, σ^x are differentiable at the equilibrium point almost surely for all t, that their derivatives are bounded almost surely for all t, and that*

$$\partial_x \hat{h}^x + \mathbb{E}\left[{}^*(\partial_{\mu^x}\hat{h}^x)\right] \in L^2_{\mathcal{F}_T}(\mathbb{R}^d), \quad \partial_y \hat{h}^y + \mathbb{E}\left[{}^*(\partial_{\mu^y}\hat{h}^y)\right] \in L^2_{\mathcal{F}_0}(\mathbb{R}^d).$$

Then for some positive constant C,

$$\sup_{t\in[0,T]} \mathbb{E}\left[|\widetilde{X}^x_t|^2 + |\widetilde{Y}^x_t|^2 + \int_0^t \|\widetilde{Z}^x_s\|_F \, ds\right] \leq C\epsilon^2,$$

$$\sup_{t\in[0,T]} \mathbb{E}\left[|\bar{X}^\epsilon_t - \hat{X}_t - \widetilde{X}^x_t|^2 + |\bar{Y}^\epsilon_t - \hat{Y}_t - \widetilde{Y}^x_t|^2\right] + \sup_{t\in[0,T]} \mathbb{E}\left[\int_0^t \|\bar{Z}^\epsilon_s - \hat{Z}_s - \widetilde{Z}^x_s\|_F \, ds\right] \leq C\epsilon^2.$$

Proof. The proof is a combination of standard estimates, see [49] for the SDE terms and [4] for the BSDE terms. □

The adjoint processes and the Hamiltonian, defined in (8) and (9) respectively, yield together with Lemma 5 and integration by parts that

$$\begin{aligned}
J^x(\bar{u}^{x,\epsilon}_\cdot; \hat{u}^y_\cdot) - J(\hat{u}^x_\cdot; \hat{u}^y_\cdot) &= \mathbb{E}\Bigg[\int_0^T \Bigg\{ \left(\bar{b}^{x,\epsilon}_t - \hat{b}^x_t\right) p^{xx}_t + \left(\bar{b}^{y,\epsilon}_t - \hat{b}^y_t\right) p^{xy}_t \\
&\qquad + \left(\bar{\sigma}^{x,\epsilon}_t - \hat{\sigma}^x_t\right) q^{xx}_t - \left(\bar{H}^{x,\epsilon}_t - \hat{H}^x_t\right) \Bigg\} dt - p^{xx}_T \widetilde{X}^x_T \Bigg] + o(\epsilon) \\
&= \mathbb{E}\Bigg[\int_0^T \Bigg\{ \left(\bar{b}^{x,\epsilon}_t - \hat{b}^x_t\right) p^{xx}_t + \left(\bar{b}^{y,\epsilon}_t - \hat{b}^y_t\right) p^{xy}_t \\
&\qquad + \left(\bar{\sigma}^{x,\epsilon}_t - \hat{\sigma}^x_t\right) q^{xx}_t - \left(\bar{H}^{x,\epsilon}_t - \hat{H}^x_t\right) \Bigg\} dt \\
&\qquad - \int_0^T \widetilde{X}^x_t dp^{xx}_t - \int_0^T p^{xx}_t d\widetilde{X}^x_t - \int_0^T d\langle p^{xx}_\cdot, \widetilde{X}^x_\cdot \rangle_t \\
&\qquad - \int_0^T \widetilde{Y}^x_t dp^{xy}_t - \int_0^T p^{xy}_t d\widetilde{Y}^x_t - \int_0^T d\langle p^{xy}_\cdot, \widetilde{Y}^x_\cdot \rangle_t \Bigg] + o(\epsilon) \\
&= -\mathbb{E}\left[\int_0^T \delta_x H^x(t) 1_{E_\epsilon}(t) dt\right] + o(\epsilon),
\end{aligned} \tag{20}$$

where $\bar{H}_t^{x,\epsilon}$, $\hat{H}_t^x$ and $\delta_x H^x(t)$ are defined in line with (19). The final equality is retrieved by expanding all differences on the third row of (20), canceling all but $\delta_x H^x(t) 1_{E_\epsilon}(t)$ with the forth and fifth row, while making use of the estimates from Lemma 5.

Consider now a spike variation of the tagged's control,

$$\check{u}_t^{y,\epsilon} := \begin{cases} \hat{u}_t^y, & t \in [0,T] \backslash E_\epsilon, \\ u_t, & t \in E_\epsilon, \end{cases}$$

where $u_\cdot \in \mathcal{U}^y$. Following the same lines of calculations as above, one finds that if $p_\cdot^{yx}$, $p_\cdot^{yy}$ are given by the adjoint equations (8) and the Hamiltonian H^y by (9), then

$$J^y(\check{u}_\cdot^{y,\epsilon}; \hat{u}_\cdot^x) - J^y(\hat{u}_\cdot^y; \hat{u}_\cdot^x) = -\mathbb{E}\left[\int_0^T \delta_y H^y(t) 1_{E_\epsilon} dt\right] + o(\epsilon),$$

where $\delta_y H^y(t)$ is defined in line with (19), for the spike perturbation $\check{u}_t^{y,\epsilon}$. The rest of the proof is standard, and can be found in for example [49].

References

[1] Y. Achdou, M. Bardi, and M. Cirant. Mean field games models of segregation. *Mathematical Models and Methods in Applied Sciences*, 27 (01):75–113, 2017.

[2] D. Andersson and B. Djehiche. A maximum principle for SDEs of mean-field type. *Applied Mathematics & Optimization*, 63(3):341–356, 2011.

[3] F. Antonelli. Backward-forward stochastic differential equations. *The Annals of Applied Probability*, pages 777–793, 1993.

[4] A. Aurell. Mean-field type games between two players driven by backward stochastic differential equations. *Games*, 9(4):88, 2018.

[5] A. Aurell and B. Djehiche. Mean-field type modeling of nonlocal crowd aversion in pedestrian crowd dynamics. *SIAM Journal on Control and Optimization*, 56(1):434–455, 2018.

[6] N. Bellomo, A. Bellouquid, and D. Knopoff. From the microscale to collective crowd dynamics. *Multiscale Modeling & Simulation*, 11(3): 943–963, 2013.

[7] N. Bellomo, D. Clarke, L. Gibelli, P. Townsend, and B. Vreugdenhil. Human behaviours in evacuation crowd dynamics: from modelling to "big data" toward crisis management. *Physics of life reviews*, 18:1–21, 2016.

[8] C. Bender and J. Steiner. A posteriori estimates for backward SDEs. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1):139–163, 2013.

[9] C. Bender, J. Zhang, et al. Time discretization and markovian iteration for coupled FBSDEs. *The Annals of Applied Probability*, 18(1):143–177, 2008.

[10] A. Bensoussan, J. Frehse, P. Yam, et al. *Mean field games and mean field type control theory*, volume 101. Springer, 2013.

[11] R. Buckdahn, B. Djehiche, J. Li, S. Peng, et al. Mean-field backward stochastic differential equations: a limit approach. *The Annals of Probability*, 37(4):1524–1565, 2009.

[12] R. Buckdahn, J. Li, and S. Peng. Mean-field backward stochastic differential equations and related partial differential equations. *Stochastic Processes and their Applications*, 119(10):3133–3154, 2009.

[13] R. Buckdahn, B. Djehiche, and J. Li. A general stochastic maximum principle for SDEs of mean-field type. *Applied Mathematics & Optimization*, 64(2):197–216, 2011.

[14] R. Buckdahn, J. Li, and J. Ma. A stochastic maximum principle for general mean-field systems. *Applied Mathematics & Optimization*, 74 (3):507–534, 2016.

[15] C. Burstedde, K. Klauck, A. Schadschneider, and J. Zittartz. Simulation of pedestrian dynamics using a two-dimensional cellular automaton. *Physica A: Statistical Mechanics and its Applications*, 295(3):507–525, 2001.

[16] P. Cardaliaguet. Notes on mean field games. Technical report, 2010.

[17] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I-II.* Springer, 2018.

[18] M. Chraibi, A. U. K. Wagoum, A. Schadschneider, and A. Seyfried. Force-based models of pedestrian dynamics. *NHM*, 6(3):425–442, 2011.

[19] M. Cirant. Multi-population mean field games systems with Neumann boundary conditions. *Journal de Mathématiques Pures et Appliquées*, 103(5):1294–1315, 2015.

[20] E. Cristiani, B. Piccoli, and A. Tosin. Multiscale modeling of granular flows with application to crowd dynamics. *Multiscale Modeling & Simulation*, 9(1):155–182, 2011.

[21] E. Cristiani, F. S. Priuli, and A. Tosin. Modeling rationality to control self-organization of crowds: an environmental approach. *SIAM Journal on Applied Mathematics*, 75(2):605–629, 2015.

[22] B. Djehiche and S. Hamadéne. Mean-field backward-forward stochastic differential equations and nonzero-sum differential games. Preprint, 2018.

[23] B. Djehiche, H. Tembine, and R. Tempone. A stochastic maximum principle for risk-sensitive mean-field type control. *IEEE Transactions on Automatic Control*, 60(10):2640–2649, 2015.

[24] B. Djehiche, A. Tcheukam, and H. Tembine. A mean-field game of evacuation in multi-level building. *IEEE Transactions on Automatic Control*, 62(10):5154–5169, 2017.

[25] C. Dogbé. Modeling crowd dynamics by the mean-field limit approach. *Mathematical and Computer Modelling*, 52(9):1506–1520, 2010.

[26] E. Feleqi. The derivation of ergodic mean field game equations for several populations of players. *Dynamic Games and Applications*, 3(4): 523–536, 2013.

[27] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995.

[28] S. P. Hoogendoorn and P. H. Bovy. Pedestrian route-choice and activity scheduling theory and models. *Transportation Research Part B: Methodological*, 38(2):169–190, 2004.

[29] Y. Hu and S. Peng. Solution of forward-backward stochastic differential equations. *Probability Theory and Related Fields*, 103(2):273–283, 1995.

[30] J. Huang, S. Wang, and Z. Wu. Backward-forward linear-quadratic mean-field games with major and minor agents. *Probability, Uncertainty and Quantitative Risk*, 1(1):8, 2016.

[31] L. Huang, S. Wong, M. Zhang, C.-W. Shu, and W. H. Lam. Revisiting Hughes' dynamic continuum model for pedestrian flow and the development of an efficient solution algorithm. *Transportation Research Part B: Methodological*, 43(1):127–141, 2009.

[32] M. Huang, R. P. Malhamé, P. E. Caines, et al. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information & Systems*, 6(3):221–252, 2006.

[33] R. L. Hughes. A continuum theory for the flow of pedestrians. *Transportation Research Part B: Methodological*, 36(6):507–535, 2002.

[34] N. Ibrahim, F. H. Hassan, R. Abdullah, and A. T. Khader. Features of microscopic horizontal transition of cellular automaton based pedestrian movement in normal and panic situation. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 9(2-12):163–169, 2017.

[35] C.-J. Jin, R. Jiang, J.-L. Yin, L.-Y. Dong, and D. Li. Simulating bidirectional pedestrian flow in a cellular automaton model considering the body-turning behavior. *Physica A: Statistical Mechanics and its Applications*, 2017.

[36] M. Kohlmann and X. Y. Zhou. Relationship between backward stochastic differential equations and stochastic controls: a linear-quadratic approach. *SIAM Journal on Control and Optimization*, 38(5):1392–1407, 2000.

[37] A. Lachapelle and M.-T. Wolfram. On a mean field game approach modeling congestion and aversion in pedestrian crowds. *Transportation research part B: methodological*, 45(10):1572–1589, 2011.

[38] D. Lacker. Limit Theory for Controlled McKean–Vlasov Dynamics. *SIAM Journal on Control and Optimization*, 55(3):1641–1672, 2017.

[39] J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.

[40] X. Li, J. Sun, and J. Xiong. Linear quadratic optimal control problems for mean-field backward stochastic differential equations. *Applied Mathematics & Optimization*, pages 1–28, 2016.

[41] J. Moon, T. E. Duncan, and T. Basar. Risk-sensitive zero-sum differential games. *IEEE Transactions on Automatic Control*, 2018.

[42] M. Moussaïd, D. Helbing, S. Garnier, A. Johansson, M. Combe, and G. Theraulaz. Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proceedings of the Royal Society of London B: Biological Sciences*, pages rspb–2009, 2009.

[43] G. Naldi, L. Pareschi, and G. Toscani. *Mathematical modeling of collective behavior in socio-economic and life sciences*. Springer Science & Business Media, 2010.

[44] E. Pardoux and S. Peng. Adapted solution of a backward stochastic differential equation. *Systems & Control Letters*, 14(1):55–61, 1990.

[45] H. Tembine. Mean-field-type games. *AIMS Math*, 2:706–735, 2017.

[46] M. Twarogowska, P. Goatin, and R. Duvigneau. Macroscopic modeling and simulations of room evacuation. *Applied Mathematical Modelling*, 38(24):5781–5795, 2014.

[47] E. Weinan, J. Han, and A. Jentzen. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Communications in Mathematics and Statistics*, 5(4):349–380, 2017.

[48] J. Yong. Forward-backward stochastic differential equations with mixed initial-terminal conditions. *Transactions of the American Mathematical Society*, 362(2):1047–1096, 2010.

[49] J. Yong and X. Y. Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*, volume 43. Springer Science & Business Media, 1999.

[50] J. Zhang, W. Klingsch, A. Schadschneider, and A. Seyfried. Ordering in bidirectional pedestrian flows and its influence on the fundamental diagram. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(02):P02002, 2012.

Paper C

Behavior near walls in the mean-field approach to crowd dynamics

Behavior near walls in the mean-field approach to crowd dynamics

by

Alexander Aurell and Boualem Djehiche

Manuscript under review for
SIAM Journal on Applied Mathematics

Abstract

This paper introduces a system of stochastic differential equations (SDE) of mean-field type that models pedestrian motion. The system lets the pedestrians spend time at, and move along, walls, by means of sticky boundaries and boundary diffusion. As an alternative to Neumann-type boundary conditions, sticky boundaries and boundary diffusion has a 'smoothing' effect on pedestrian motion. When these effects are active, the pedestrian paths are semimartingales with first-variation part absolutely continuous with respect to the Lebesgue measure dt, rather than an increasing processes (which in general induces a measure singular with respect to dt) as is the case under Neumann boundary conditions. We show that the proposed mean-field model for pedestrian motion admits a unique weak solution and that it is possible to control the system in the weak sense, using a Pontryagin-type maximum principle. We also relate the mean-field type control problem to the social cost minimization in an interacting particle system. We study the novel model features numerically and we confirm empirical findings on pedestrian crowd motion in congested corridors.

Keywords: pedestrian crowd modeling, mean-field type control, sticky boundary conditions, boundary diffusion

C.1 Introduction

Models for pedestrian motion in confined domains must consider interaction with solid obstacles such as pillars and walls. The pedestrian response to a restriction of movement has been included into crowd models either as boundary conditions or repulsive forces. Up until today, the Neumann condition and its variants (e.g. no-flux) have been especially popular among the boundary conditions. The Neumann condition suffers from a drawback related to its microscopic (pathwise) interpretation. A Neumann condition on the crowd density corresponds to pedestrian paths reflecting in the boundary. In reality, pedestrians do not bounce off walls in the manner of classical Newtonian particles, but their movement is slowed down by the impact and a positive amount of time is needed to choose a new direction of motion. It is natural to think that whenever a pedestrian is forced (or decides) to make contact with a wall, she stays there for some time. During this time, she can move and interact with other pedestrians, before re-entering the interior of the domain.

C.1.1 Mathematical modeling of pedestrian-wall interaction

Today there is more than one conventional approach to the mathematical modeling of pedestrian motion. This section aims to summarize how they incorporate the interaction between pedestrians and walls.

Microscopic force-based models, among which the social force model has gained the most attention, describes pedestrians as Newton-like particles. From the initial work [30] and onward, the influence a wall has on the pedestrian is modeled as a repulsive force. The shape of the corresponding potential has been studied experimentally, for example in [40]. The cellular automata is another widely used microscopic approach to pedestrian crowd modeling. Walls are modeled as cells to which pedestrians cannot transition, already the original work [37] considers this viewpoint. In the continuum limits of cellular automata, as for example in [10, 14], boundary conditions are often set to no-flux conditions of the same type as (1) below.

The focus of macroscopic models is the global pedestrian density, either in a stationary or a dynamic regime. Inspired by fluid dynamics [33] treats the crowd as a 'thinking fluid' that moves at maximum speed towards a target location while taking environmental factors into account, such as the congestion of the crowd. In this category of models, boundary conditions at impenetrable walls are most often implemented as Neumann conditions for the pedestrian density. The pathwise interpretation of a Neumann boundary condition is instantaneous reflection. A nonlocal projection of pedestrian velocity in the normal and the tangential direction of the boundary respectively is suggested in [6] and implemented in [7], allowing for nonlocal interaction with boundaries.

Mean-field games and mean-field type control/games are macroscopic models of rational pedestrians with the ability to anticipate crowd movement, and adapt accordingly. These models can capture competition between individuals as well as crowd/sub-crowd cooperation. In the mean-field approach to pedestrian crowd modeling pedestrian-to-pedestrian interaction is assumed to be symmetric and weak, thus plausibly replaced by an interaction with a mean field (typically a functional of the pedestrian density). One of the most attractive features of the mean-field approach is that it connects the macroscopic (pedestrian density) and the microscopic (pedestrian path) point-of-view, typically through results on the near-optimality/equilibrium of mean-field optimal controls/equilibria. The connection permits us to infer individual pedestrian behavior from crowd density simulations, and vice versa. In what follows, the crowd density is denoted by m. In [38], the density is subjected to $n(x) \cdot \nabla m(t,x) = 0$ at walls, where $n(x)$ is the outward normal at x. Under this constraint, the normal velocity of the pedestrian is zero at any wall. Taking conservation of probability mass into account, [12] derives the following boundary condition

$$-n(x) \cdot (\nabla m(t,x) - G(m)v(t,x)) = 0, \tag{1}$$

where $G(m)v$ is a general form of the pedestrian velocity. The constraint (1) represents reflection at the boundary since in the corresponding microscopic interpretation pedestrians make a classical Newtonian bounce whenever they hit the boundary. The same type of constraint is used in [2]. The case of several interacting populations in a bounded domain with reflecting boundaries has been studied in the stationary and dynamic case [17, 1, 5]. In these papers, the crowd density at walls is constrained by

$$n(x) \cdot (\nabla m(t,x) + m(t,x)\partial_p H(x, \nabla u)) = 0.$$

The constraint is a reflection and the term $-\partial_p H(x, \nabla u)$ is the velocity of pedestrians that use the mean-field equilibrium strategy.

C.1.2 Sticky reflected stochastic differential equations

The sticky reflected Brownian motion was discovered by Feller [23, 24, 25]. He studied the infinitesimal generator of strong Markov processes on $[0, \infty)$ that behave like Brownian motion in $(0, \infty)$, and showed that it is possible for the process to be 'sticky' on the boundary, i.e. to sojourn at 0. So 'sticky reflection' was appended to the list of boundary conditions for diffusions, which already included instantaneous reflection, absorption, and the elastic Robin condition. Wentzell [44] extended the result to more general domains.

Itô and McKean [34] constructed sample paths to the one-dimensional sticky reflected Brownian motion

$$dX_t = 2\mu 1_{\{X_t=0\}}dt + 1_{\{X_t>0\}}dW_t, \quad \mu > 0, \tag{2}$$

whose infinitesimal generator is the one studied by Feller. Skorokhod conjectured that the sticky reflected Brownian motion has no strong solution. A proof that (2) has a unique weak solution can be found in for example [46, IV.7].

Chitashvili published the technical report [15] in 1989 claiming a proof of Skorokhod's conjecture. Around that time, the process was studied by several authors, e.g. [29, 26, 3, 47], to name a few. Warren [45] provided a proof of Skorokhod's conjecture in 1997 and in 2014 Engelbert and Peskir [22] published a proof useful for further generalizations. The fact that the system has no strong solution has consequences for how optimal control of the system can be approached, as we will see in this paper.

Building on [22], interacting particle systems of sticky reflected Brownian motions are considered in [27]. Interaction is introduced via a Girsanov transformation. See [27, Sect. 3.2] for the construction. Under assumptions on the 'shape' of the interaction and integrability of the Girsanov kernel, the interacting system is well-defined. Since the process no longer behaves like a Brownian motion in the interior of the domain, it is now referred to as a sticky reflected SDE. The boundary behavior is shown to be sticky in the sense that the process spends a (dt-)positive time on the boundary.

Sticky reflected SDEs with boundary diffusion are considered in [28]. The paths defined by such a system are allowed to move on the (sufficiently smooth) boundary $\partial\mathcal{D}$ of some bounded domain $\mathcal{D} \subset \mathbb{R}^d$. Under smoothness conditions on $\partial\mathcal{D}$, the authors show that this type of SDE has a unique weak solution. Furthermore, an interacting system is studied, where interaction is introduced via a Girsanov transformation.

C.1.3 Synopsis

In this paper, the sticky reflected SDE with boundary diffusion of [28] is proposed as a model for pedestrian crowd motion in confined domains. We begin by considering a (non-transformed) sticky reflected SDE with boundary diffusion on $\mathcal{D}$, a non-empty bounded subset of $\mathbb{R}^n$ with C^2-smooth boundary $\Gamma := \partial\mathcal{D}$ (see Section C.2.2, below) and outward normal n,

$$dX_t = \left(1_{\mathcal{D}}(X_t) + 1_{\Gamma}(X_t)\pi(X_t)\right) dB_t - 1_{\Gamma}(X_t)\frac{1}{2}\Big(\frac{1}{\gamma} + \kappa(X_t)\Big)n(X_t)dt, \quad (3)$$

where $\pi(X_t)$ is the projection onto the tangent space of Γ at X_t, $\kappa(X_t)$ is the mean curvature of Γ at X_t, and γ is a positive constant representing the stickiness of Γ, cf. Remark 5 in Section E.2 below. All relevant technical details can be found in Section C.2. Equation (3) admits a unique weak solution $\mathbb{P}$, but no strong solution. To control an equation that admits only a weak solution is to control a probability measure on $(\Omega, \mathcal{F})$, under which the state process $X. := \{X_t\}_{t\in[0,T]}$ is interpreted as the coordinate process $X_t(\omega) = \omega(t)$. If all the admissible distributions of $X.$ are absolutely continuous with respect the reference measure $\mathbb{P}$, then Girsanov's theorem can be used to implement the control. This corresponds to the case when the drift of (3) is controlled. In the controlled diffusion case, admissible measures are all singular with $\mathbb{P}$ and with one another (for different controls), and the control problem is in fact a robustness problem over all admissible measures, which led to the so-called second order backward SDE framework [42]. In this paper we treat the case with controlled drift, the controlled diffusion case will be treated elsewhere. A mean-field dependent drift β is introduced into the coordinate process through the Girsanov transformation

$$\frac{d\mathbb{P}^u}{d\mathbb{P}}\Big|_{\mathcal{F}_t} = L_t^u := \mathcal{E}_t\left(\int_0^{\cdot} \beta\left(t, X., \mathbb{P}^u(t), u_t\right)^* dB_t\right),$$

where $\mathbb{P}^u(t) := \mathbb{P}^u \circ X_t^{-1}$ is the marginal distribution of X_t under $\mathbb{P}^u$, β^* denotes the transpose of β, and $\mathcal{E}$ is the Doléans-Dade exponential defined for a continuous local martingale M as

$$\mathcal{E}_t(M) := \exp\left(M_t - \frac{1}{2}\langle M\rangle_t\right). \quad (4)$$

The path of a typical pedestrian in the interacting crowd is then (under $\mathbb{P}^u$) described by

$$\begin{cases} dX_t = 1_{\mathcal{D}}(X_t)\Big(\beta\left(t, X_\cdot, \mathbb{P}^u(t), u_t\right) dt + dB_t^u\Big) \\ \qquad\quad + 1_\Gamma(X_t)\left(\pi(X_t)\beta\left(t, X_\cdot, \mathbb{P}^u(t), u_t\right) - \dfrac{n(X_t)}{2\gamma}\right) dt \\ \qquad\quad + 1_\Gamma(X_t) dB_t^{\Gamma,u}, \\ dB_t^{\Gamma,u} = \pi(X_t) dB_t^u - \dfrac{1}{2}\kappa(X_t) n(X_t) dt, \end{cases} \tag{5}$$

where B^u is a $\mathbb{P}^u$-Brownian motion. We provide a proof of the existence and uniqueness of the controlled probability measure $\mathbb{P}^u$ based on a fixed-point argument involving the total variation distance (cf. [20]).

Pedestrians are assumed to be cooperating and controlled by a rational central planner. The central planner represents an authority that gives directions to the crowd through signs, mobile devices, or security personnel, and the crowd follows the instructions. This setup has been used to study evacuation in for example [11, 13, 21]. For a discussion on the goals, the degrees of cooperation, and the information structure in a pedestrian crowd, see [18]. The central planner's goal is to minimize the finite-horizon cost functional

$$J(u) := E^u\left[\int_0^T f\left(t, X_\cdot, \mathbb{P}^u(t), u_t\right) dt + g\left(X_T, \mathbb{P}^u(T)\right)\right], \tag{6}$$

where f is the instantaneous cost and g is the terminal cost (see Section C.4 for conditions on the functions f and g). The minimization of (6) subject to (5) is equivalent to the following mean-field type control problem, stated in the strong sense in the original probability space with measure $\mathbb{P}$,

$$\begin{cases} \inf\limits_{u\in\mathcal{U}} \; E\left[\displaystyle\int_0^T L_t^u f\left(t, X_\cdot, \mathbb{P}^u(t), u_t\right) dt + L_T^u g\left(X_T, \mathbb{P}^u(T)\right)\right], \\ \text{s.t.} \;\; dL_t^u = L_t^u \beta\left(t, X_\cdot, \mathbb{P}^u(t), u_t\right)^* dB_t, \; L_0^u = 1. \end{cases} \tag{7}$$

The validity of (7) is justified in Section C.4 below. Problem (7) is nowadays a standard mean-field type control problem and a stochastic maximum principle yielding necessary conditions for an optimal control can be found in [9]. Solving the general problem (7) with a Pontryagin-type maximum principle poses some practical difficulties, the main one being the necessity of a second order adjoint process. However, most difficulties can be tackled by imposing assumptions plausible for models of pedestrian crowd motion. With the aim to replicate the pedestrian behavior observed in the empirical studies [49] and [50], we consider a special case of (7) in Section C.5 where u_t takes values in a convex set and $\mathbb{P}^u(t)$ is replaced by $E^u[r(X_t)]$, where the function $r : \mathbb{R}^d \to \mathbb{R}^d$ can be different for each of the coefficients involved.

C.1.4 Paper contribution and outline

The main contribution of this paper is a new approach to boundary conditions in pedestrian crowd modeling. The sticky reflected SDE of mean-field type with boundary diffusion is proposed as an alternative to reflected SDEs of mean-field type to model pedestrian paths in optimal-control based models. Sticky boundaries and boundary diffusion allows the pedestrian to spend time and move along the boundary (walls, pillars, etc.), in contrast to models based on reflected SDEs where pedestrians are immediately reflected. Existence and uniqueness of the mean-field type version of the sticky reflected SDE with boundary diffusion is treated. The model can be optimally controlled (in the weak sense) and a Pontryagin-type stochastic maximum principle is applied to derive necessary optimality conditions. Furthermore, the mean-field type control problem has a microscopic interpretation in the form of a system of interacting sticky reflected SDEs with boundary diffusion. The new features of sticky boundaries and boundary diffusion yield more flexibility when modeling pedestrian behavior at boundaries. A scenario of unidirectional pedestrian flow in a long narrow corridor is studied numerically to highlight these novel characteristics and to replicate experimental findings as a first step in model validation.

The rest of the paper is organized as follows. Section C.2 defines notation and summarizes relevant background theory. Section C.3 introduces sticky reflected SDEs of mean-field type with boundary diffusion. Conditions under which the equation has a unique weak solution are presented. In Section C.4 the finite horizon optimal control of the state equation introduced in Section C.3 is considered. In the uncontrolled case, the convergence on an interacting (non-mean-field) particle system to the sticky reflected SDE of mean-field type is proved. Finally, Section C.5 presents analytic examples and numerical results based on the particle system approximation concerning unidirectional flow in a long narrow corridor.

C.2 Preliminaries

The domain $\mathcal{D}$ is a non-empty bounded subset of $\mathbb{R}^d$ with C^2-smooth boundary $\Gamma := \partial\mathcal{D}$. The closure of $\mathcal{D}$ is denoted $\bar{\mathcal{D}}$. The Euclidean norm is denoted $|\cdot|$. A finite time horizon $T > 0$ is fixed throughout the paper. The path of a stochastic process is denoted $X_\cdot := \{X_t\}_{t\in[0,T]}$, and C is a generic positive constant.

C.2.1 The coordinate process and probability metrics

Let $(\mathcal{X}, d)$ be a metric space. The set of Borel probability measures on $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. By $\mathcal{P}_p(\mathcal{X}) \subset \mathcal{P}(\mathcal{X})$ we denote the set of all $\mu \in \mathcal{P}(\mathcal{X})$ such that $(\|\mu\|_p)^p := \int d(y_0, y)^p \mu(dy) < \infty$ for an arbitrary $y_0 \in \mathcal{X}$.

Let $\Omega := C([0,T];\mathbb{R}^d)$ be endowed with the metric $|\omega|_T := \sup_{t\in[0,T]} |\omega(t)|$ for $\omega \in \Omega$. Denote by $\mathcal{F}$ the Borel σ-field over Ω. Given $t \in [0,T]$ and $\omega \in \Omega$, put $X_t(\omega) = \omega(t)$ and denote by $\mathcal{F}_t^0 := \sigma(X_s; s \le t)$ the filtration generated by $X_\cdot$. $X_\cdot$ is the so-called *coordinate process*. For any $P \in \mathcal{P}(\Omega)$ (the set of Borel probability measures on Ω) we denote by $\mathbb{F}^P := (\mathcal{F}_t^P; t \in [0,T])$ the completion of $\mathbb{F}^0 := (\mathcal{F}_t^0; t \in [0,T])$ with the P-null sets of Ω.

Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and let $\mathcal{B}(\mathbb{R}^d)$ be the Borel σ-algebra on $\mathbb{R}^d$. The total variation metric on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is

$$d_{TV}(\mu, \nu) := 2 \sup_{A\in\mathcal{B}(\mathbb{R}^d)} |\mu(A) - \nu(A)| .$$

On the filtration $\mathbb{F}^P$, where $P \in \mathcal{P}(\Omega)$, the total variation metric between $m, m' \in \mathcal{P}(\Omega)$ is

$$D_t(m, m') := 2 \sup_{A\in\mathcal{F}_t^P} \left|m(A) - m'(A)\right|, \quad 0 \le t \le T, ,$$

and satisfies $D_s(m, m') \le D_t(m, m')$ for $0 \le s \le t$. Consider the coordinate process $X_\cdot$, then for $m, m' \in \mathcal{P}(\Omega)$,

$$d_{TV}\left(m \circ X_t^{-1}, m' \circ X_t^{-1}\right) \le D_t(m, m'), \quad 0 \le t \le T.$$

Endowed with the metric D_T, $\mathcal{P}(\Omega)$ is a complete metric space. The total variation metric is connected to the Kullback-Leibler divergence through the Csiszár-Kullback-Pinsker inequality,

$$D_t^2(m, m') \le 2E^m \left[\log\left(dm/dm'\right)\right], \tag{8}$$

where E^m denotes expectation with respect to m.

C.2.2 Boundary diffusion

In this subsection we introduce the boundary diffusion B^Γ and review the necessary parts of the background theory presented in [28, Sect. 2].

Definition 1. *Γ is Lipschitz continuous (resp. C^k-smooth) if for every $x \in \Gamma$ there exists a neighborhood $V \subset \mathbb{R}^d$ of x such that $\Gamma \cap V$ is the graph of a Lipschitz continuous (resp. C^k-smooth) function and $\mathcal{D} \cap V$ is located on one side of the graph, i.e., there exists new orthogonal coordinates $(y_1, \ldots, y_d)$ given by an orthogonal map T, a reference point $z \in \mathbb{R}^{d-1}$, real numbers $r, h > 0$, and a Lipschitz continuous (resp. C^k-smooth) function $\varphi : \mathbb{R}^{d-1} \to \mathbb{R}$ such that*

(i) $V = \{y \in \mathbb{R}^d : |y_{-d} - z| < r, |y_d - \varphi(y_{-d})| < h\}$

(ii) $\mathcal{D} \cap V = \{y \in V : -h < y_d - \varphi(y_{-d}) < 0\}$

iii) $\Gamma \cap V = \{y \in V : y_d = \varphi(y_{-d})\}$

Definition 2. *For $y \in V$, let*

$$\widetilde{n}(y) := \frac{(-\nabla\varphi(y_{-d}), 1)}{\sqrt{|\nabla\varphi(y_{-d})|^2 + 1}}.$$

Let $x \in \Gamma$ and $T \in \mathbb{R}^{d\times d}$ be the orthogonal transformation from Definition 1. Then the outward normal vector at x is defined by $n(x) := T^{-1}\widetilde{n}(Tx)$.

Definition 3. *Let $x \in \Gamma$ and $\pi(x) := E - n(x)n(x)^* \in \mathbb{R}^{d\times d}$, where E is the identity matrix. $\pi(x)$ is the orthogonal projection on the tangent space at x.*

Note that for $z \in \mathbb{R}^d$, $\pi(x)z = z - (n(x), z)\, n(x)$.

Definition 4. *Let $f \in C^1(\bar{\mathcal{D}})$ and $x \in \Gamma$. Whenever Γ is sufficiently smooth at x, $\nabla_\Gamma f(x) := \pi(x)\nabla f(x)$ and if $f \in C^2(\bar{\mathcal{D}})$, $\Delta_\Gamma f(x) := Tr(\nabla^2_\Gamma f(x))$. If n is differentiable at x the mean curvature of Γ at x is*

$$\kappa(x) := div_\Gamma n(x) = (\pi(x)\nabla) \cdot n(x).$$

In [28] it is noted that whenever Γ is C^2-smooth,

$$(\pi\nabla)^*\, \pi = -\kappa n.$$

A Brownian motion $B^\Gamma_\cdot$ on a smooth boundary Γ is a Γ-valued stochastic process generated by $\frac{1}{2}\Delta_\Gamma$. This is in analogy with the standard Brownian motion on $\mathbb{R}^d$, in the sense that $B^\Gamma_\cdot$ solves the martingale problem for $(\frac{1}{2}\Delta_\Gamma, C^\infty(\Gamma))$. A solution to the Stratonovich SDE

$$dB^\Gamma_t = \pi(B^\Gamma_t) \circ dB_t,$$

where $B_\cdot$ is a standard Brownian motion on $\mathbb{R}^d$, is a Brownian motion on Γ [32, Chap. 3, Sect. 2]. By the Itô-Stratonovich transformation rule, the Brownian motion on Γ solves

$$dB^\Gamma_t = -\frac{1}{2}\kappa(B^\Gamma_t)n(B^\Gamma_t)dt + \pi(B^\Gamma_t)dB_t.$$

C.3 Sticky reflected SDEs of mean-field type with boundary diffusion

In this section we provide conditions for the existence and uniqueness of a weak solution to the sticky reflected SDE of mean-field type with boundary diffusion. Consider the reflected sticky SDE with boundary diffusion,

$$\begin{cases} dX_t = -1_\Gamma(X_t)\frac{1}{2}\left(\frac{1}{\gamma} + \kappa(X_t)\right) n(X_t)dt \\ \qquad\qquad + \left(1_\mathcal{D}(X_t) + 1_\Gamma(X_t)\pi(X_t)\right) dB_t, \\ X_0 = x_0 \in \bar{\mathcal{D}}, \end{cases} \tag{9}$$

which from now on will be written in short-hand notation as

$$dX_t = a(X_t)dt + \sigma(X_t)dB_t, \tag{10}$$

where $a : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are bounded functions over $[0,T] \times \bar{\mathcal{D}}$, defined as

$$a(x) := -1_\Gamma(x)\frac{1}{2}\left(\frac{1}{\gamma} + \kappa(x)\right) n(x), \quad \sigma(x) := 1_\mathcal{D}(x) + 1_\Gamma(x)\pi(x).$$

By [28, Thm 3.9 & 3.17], (9) has a unique weak solution, i.e. there is a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ that solves the corresponding martingale problem (cf. [35, Thm 18.7]), and the solution $X_\cdot$ is $C([0,T]; \bar{\mathcal{D}})$-valued $\mathbb{P}$-a.s. The result [28, Thm 3.9] relies on some conditions, lets verify them for the sake of completeness. The weight functions α and β, introduced on [28, pp. 6], are in (9) set to be everywhere constant and positive such that $\alpha/\beta = 1/\gamma$ (cf. Remark 5, below). Condition 3.12 of [28] therefore holds: $\partial\mathcal{D}$ is C^2 and the constant positive weight functions have the required regularity. This justifies the use of [28, Thm 3.9], and no further conditions are required for [28, Thm 3.17]. To simplify notation, from now on through out the rest of this paper let $\mathbb{F}$ denote the completion of $\mathbb{F}^0$ with the $\mathbb{P}$-null sets of Ω, i.e. $\mathbb{F} = (\mathcal{F}_t; t \geq 0) := \mathbb{F}^{\mathbb{P}}$.

Remark 5. *The coordinate process is composed of three essential parts:*

- *Interior diffusion* $1_{\mathcal{D}}(X_t)dB_t$*;*
- *Boundary diffusion* $1_\Gamma(X_t)(\pi(X_t)dB_t - \frac{1}{2}(\kappa n)(X_t)dt) = 1_\Gamma(X_t)dB_t^\Gamma$*;*
- *Normal sticky reflection* $-1_\Gamma(X_t)\frac{1}{2\gamma}n(X_t)dt$*.*

The constant γ *is connected to the level of stickiness of the boundary* Γ*. It also influences the invariant distribution of* X_t*. Let* λ *and* s *denote the Lebesgue measure on* $\mathbb{R}^d$ *and the surface measure on* Γ*, respectively. Consider the measure* $\rho := 1_{\mathcal{D}}\alpha\lambda + 1_\Gamma\alpha' s$*,* $\alpha, \alpha' \in \mathbb{R}$*. By choosing* $\alpha = \bar{\alpha}/\lambda(\mathcal{D})$ *and* $\alpha' = (1-\bar{\alpha})/s(\Gamma)$*,* $\bar{\alpha} \in [0,1]$*,* ρ *becomes a probability measure on* $\mathbb{R}^d$ *with support in* $\bar{\mathcal{D}}$ *and* ρ *is in fact the invariant distribution of* (9) *whenever*

$$\frac{1}{\gamma} = \frac{\bar{\alpha}}{(1-\bar{\alpha})}\frac{s(\Gamma)}{\lambda(\mathcal{D})}.$$

Hence $\bar{\alpha} \to 1$ *as* $\gamma \to 0$ *and the invariant distribution of* (9) *concentrates on the interior* $\mathcal{D}$*. But as* $\gamma \to \infty$*, it concentrates on the boundary* Γ*. We say that the more probability mass that* ρ *locates on* Γ*, the stickier* Γ *is.*

Next, we introduce mean-field interactions and a control process in (9) through a Girsanov transformation.

Definition 6. *Let the set of control values* U *be a subset of* $\mathbb{R}^d$*. The set of admissible controls is*

$$\mathcal{U} := \{u : [0,T] \times \Omega \to U \mid u \ \mathbb{F}\text{-prog. measurable}\}.$$

Let $\mathbb{Q}(t) := \mathbb{Q} \circ X_t^{-1}$ denote the t-marginal distribution of the coordinate process under $\mathbb{Q} \in \mathcal{P}(\Omega)$. Let β be a measurable function from $[0,T] \times \Omega \times \mathcal{P}(\mathbb{R}^d) \times U$ into $\mathbb{R}^d$ such that:

Assumption 1. *For every* $\mathbb{Q} \in \mathcal{P}(\Omega)$ *and* $u \in \mathcal{U}$, $(\beta(t, X_\cdot, \mathbb{Q}(t), u_t))_{t\in[0,T]}$ *is progressively measurable with respect to* $\mathbb{F}$*, the completion of the filtration generated by the coordinate process with the* $\mathbb{P}$*-null sets of* Ω.

Assumption 2. *For every* $t \in [0, T]$, $\omega \in \Omega$, $u \in U$, *and* $\mu \in \mathcal{P}(\mathbb{R}^d)$,

$$|\beta(t, \omega, \mu, u)| \leq C \left(1 + |\omega|_T + \int_{\mathbb{R}^d} |y|\mu(dy)\right).$$

Assumption 3. *For every* $t \in [0, T]$, $\omega \in \Omega$, $u \in U$, *and* $\mu, \mu' \in \mathcal{P}(\mathbb{R}^d)$,

$$\left|\beta\left(t, \omega, \mu, u\right) - \beta\left(t, \omega, \mu', u\right)\right| \leq C d_{TV}(\mu, \mu').$$

Given $\mathbb{Q} \in \mathcal{P}(\Omega)$ and $u \in \mathcal{U}$, let

$$L_t^{u,\mathbb{Q}} := \mathcal{E}_t\left(\int_0^\cdot \beta\left(s, X_\cdot, \mathbb{Q}(s), u_s\right) dB_s\right), \tag{11}$$

where $\mathcal{E}$ is the Doléans-Dade exponential (cf. (4)).

Lemma 7. *The positive measure* $\mathbb{P}^{u,\mathbb{Q}}$ *defined by* $d\mathbb{P}^{u,\mathbb{Q}} = L_t^{u,\mathbb{Q}} d\mathbb{P}$ *on* $\mathcal{F}_t$ *for all* $t \in [0, T]$*, is well-defined and is a probability measure on* Ω*. Moreover,* $\mathbb{P}^{u,\mathbb{Q}} \in \mathcal{P}_p(\Omega)$ *for all* $p \in [1, \infty)$ *and under* $\mathbb{P}^{u,\mathbb{Q}}$ *the coordinate process satisfies*

$$X_t = x_0 + \int_0^t \Big(\sigma(X_s)\beta\left(s, X_\cdot, \mathbb{Q}(s), u_s\right) + a(X_s)\Big) ds + \int_0^t \sigma(X_s) dB_s^{\mathbb{Q}}, \tag{12}$$

where $B^{\mathbb{Q}}$ *is a standard* $\mathbb{P}^{u,\mathbb{Q}}$*-Brownian motion.*

Proof. Assume that $\varphi_\cdot$ is a process such that $\mathbb{P}^\varphi$, defined by $d\mathbb{P}^\varphi = L_t^\varphi d\mathbb{P}$ on $\mathcal{F}_t$ where $L_t^\varphi := \mathcal{E}_t(\int_0^\cdot \varphi_s dB_s)$, is a probability measure on Ω. By Girsanov's theorem, the coordinate process under $\mathbb{P}^\varphi$ satisfies

$$dX_t = \left(\sigma(X_t)\varphi_t + a(X_t)\right) dt + \sigma(X_t) dB_t^\varphi,$$

where $B_\cdot^\varphi$ is a $\mathbb{P}^\varphi$-Brownian motion. The C^2-smoothness of the boundary Γ grants a bounded orthogonal projection onto the tangent space of Γ and a bounded mean curvature of Γ. By the Burkholder-Davis-Gundy inequality we have for $1 \leq p < \infty$

$$\begin{aligned} E^{\varphi}\left[|X|_T^p\right] \leq E^{\varphi}\Bigg[C\Bigg(|X_0|^p + \int_0^T |\sigma(X_s)\varphi_s|^p ds + \int_0^T |a(X_s)|^p ds \\ + \left|\int_0^{\cdot} \sigma(X_s) dB_s^{\varphi}\right|_T^p\Bigg)\Bigg] \\ \leq C\left(1 + \int_0^T E^{\varphi}[|\varphi_s|^p] ds\right), \end{aligned}$$

where E^{φ} denotes expectation taken under $\mathbb{P}^{\varphi}$. By Assumption 3 it holds for every $t \in [0,T]$, $\omega \in \Omega$, $\mu \in \mathcal{P}(\mathbb{R}^d)$, and $u \in U$ that

$$|\beta(t,\omega,\mu,u)| \leq C\Big(d_{TV}(\mu,\mathbb{P}(t)) + |\beta(t,\omega,\mathbb{P}(t),u)|\Big). \tag{13}$$

In view of (13), Assumption 2 and 3, and the fact that the total variation between two probability measures is uniformly bounded, we have for all $t \in [0,T]$,

$$\begin{aligned} |\beta(t,X.,\mathbb{Q}(t),u_t)| &\leq C\left(d_{TV}(\mathbb{Q}(t),\mathbb{P}(t)) + |\beta(t,X.,\mathbb{P}(t),u_t)|\right) \\ &\leq C\left(1 + |X|_T + \int_{\mathbb{R}^d} |y|\mathbb{P}(t)(dy)\right) \\ &\leq C\left(\sup\{|y| : y \in \bar{\mathcal{D}}\}\right) =: \bar{C} < \infty, \quad \mathbb{P}\text{-a.s.} \end{aligned} \tag{14}$$

The third inequality of (14) holds $\mathbb{P}$-a.s. since under $\mathbb{P}$, $X. \in C\left([0,T];\bar{\mathcal{D}}\right)$ almost surely. We note that (14) implies that Novikov's condition is satisfied,

$$E\left[\exp\left(\frac{1}{2}\int_0^T \sup_{s\in[0,t]} |\beta(s,X.,\mathbb{Q}(s),u_s)|^2 dt\right)\right] \leq E\left[\exp\left(\frac{T\bar{C}^2}{2}\right)\right] < \infty,$$

where E denotes expectation with respect to $\mathbb{P}$. Hence the Doléans-Dade exponential defined in (11) is an $(\mathcal{F}_t,\mathbb{P})$-martingale and $\mathbb{P}^{u,\mathbb{Q}}$ is indeed a probability measure, i.e. $\mathbb{P}^{u,\mathbb{Q}} \in \mathcal{P}(\Omega)$. To show that $\mathbb{P}^{u,\mathbb{Q}} \in \mathcal{P}_p(\Omega)$ for any

$p \in [1, \infty)$, we simply note that

$$\begin{aligned}
& E^{u,\mathbb{Q}} \left[|X|_T^p \right] \\
&= E^{u,\mathbb{Q}} \left[|X|_T^p \left(1_{\{X. \in C([0,T];\bar{\mathcal{D}})\}} + 1_{\{X. \notin C([0,T];\bar{\mathcal{D}})\}} \right) \right] \\
&= E \left[L_T^u |X|_T^p \left(1_{\{X. \in C([0,T];\bar{\mathcal{D}})\}} + 1_{\{X. \notin C([0,T];\bar{\mathcal{D}})\}} \right) \right] \\
&\leq \sup\{|y|^p : y \in \bar{\mathcal{D}}\} E \left[L_T^u 1_{\{X. \in C([0,T];\bar{\mathcal{D}})\}} \right] \\
&= \sup\{|y|^p : y \in \bar{\mathcal{D}}\}.
\end{aligned} \tag{15}$$

Finally, by Girsanov's theorem the coordinate process under $\mathbb{P}^{u,\mathbb{Q}}$ satisfies (12). □

For a given $u \in \mathcal{U}$, consider the map

$$\Phi_u : \mathcal{P}(\Omega) \ni \mathbb{Q} \mapsto \mathbb{P}^{u,\mathbb{Q}} \in \mathcal{P}(\Omega),$$

such that $d\mathbb{P}^{u,\mathbb{Q}} = L_t^u d\mathbb{P}$ on $\mathcal{F}_t$, where L^u is given by (11).

Proposition 8. *The map Φ_u is well-defined and admits a unique fixed point for all $u \in \mathcal{U}$. Moreover, for every $p \in [1, \infty)$ the fixed point, denoted $\mathbb{P}^u$, belongs to $\mathcal{P}_p(\Omega)$. In particular,*

$$E^u \left[|X|_T^p \right] \leq \sup_{y \in \bar{\mathcal{D}}} |y|^p,$$

where E^u denotes expectation with respect to $\mathbb{P}^u$.

Proof. By Lemma 7, the mapping is well defined. We first show the contraction property of the map Φ_u in the complete metric space $\mathcal{P}(\Omega)$, endowed with the total variation distance D_T. The proof is an adaptation of the proof of [16, Thm. 8]. For each $t \in [0, T]$, let $\beta_t^{\mathbb{Q}} := \beta(t, X., \mathbb{Q}(t), u_t)$. Given $\mathbb{Q}, \widetilde{\mathbb{Q}} \in \mathcal{P}(\Omega)$, the Csiszár-Kullback-Pinsker inequality (8) and the fact that

$\int_0^\cdot (dB_s - \beta_s^{\mathbb{Q}} ds)$ is a martingale under $\Phi_u(\mathbb{Q}) = \mathbb{P}^{u,\mathbb{Q}}$ yields

$$\begin{aligned}
&D_T^2\left(\Phi_u(\mathbb{Q}), \Phi_u(\widetilde{\mathbb{Q}})\right) \le 2E^{u,\mathbb{Q}}\left[\log\left(L_T^{\mathbb{Q}}/L_T^{\widetilde{\mathbb{Q}}}\right)\right]\\
&= 2E^{u,\mathbb{Q}}\left[\int_0^T \left(\beta_s^{\mathbb{Q}} - \beta_s^{\widetilde{\mathbb{Q}}}\right) dB_s - \frac{1}{2}\int_0^T \left(\beta_s^{\mathbb{Q}}\right)^2 - \left(\beta_s^{\widetilde{\mathbb{Q}}}\right)^2 ds\right]\\
&= 2E^{u,\mathbb{Q}}\left[\int_0^T \left(\beta_s^{\mathbb{Q}} - \beta_s^{\widetilde{\mathbb{Q}}}\right) \beta_s^{\mathbb{Q}} - \frac{1}{2}\left(\beta_s^{\mathbb{Q}}\right)^2 + \frac{1}{2}\left(\beta_s^{\widetilde{\mathbb{Q}}}\right)^2 ds\right]\\
&= \int_0^T \mathbb{E}^{u,\mathbb{Q}}\left[\left(\beta_s^{\mathbb{Q}} - \beta_s^{\widetilde{\mathbb{Q}}}\right)^2\right] ds\\
&\le C\int_0^T d_{TV}^2\left(\mathbb{Q}(s), \widetilde{\mathbb{Q}}(s)\right) ds \le C\int_0^T D_s^2\left(\mathbb{Q}, \widetilde{\mathbb{Q}}\right) ds.
\end{aligned}$$

Iterating the inequality, we obtain for every $N \in \mathbb{N}$,

$$D_T^2\left(\Phi_u^N(\mathbb{Q}), \Phi_u^N(\widetilde{\mathbb{Q}})\right) \le \frac{C^N T^N}{N!} D_T^2\left(\mathbb{Q}, \widetilde{\mathbb{Q}}\right),$$

where Φ_u^N denotes the N-fold composition of Φ_u. Hence Φ_u^N is a contraction for N large enough, thus admitting a unique fixed point, which is also the unique fixed point for Φ_u. Under $\mathbb{P}^u$, the fixed point of Φ_u, the coordinate process satisfies

$$dX_t = (\sigma(X_t)\beta\,(t, X_\cdot, \mathbb{P}^u(t), u_t) + a(X_t))\,dt + \sigma(X_t)dB_t^u,$$

where B^u is a $\mathbb{P}^u$-Brownian motion. Following the calculations from Lemma 7 that lead to (15), we get the estimate

$$(\|\mathbb{P}^u\|_p)^p = E^u\left[|X|_T^p\right] \le \sup_{y\in\bar{\mathcal{D}}} |y|^p,$$

where $p \in [1, \infty)$. $\square$

From now on, we will denote the Brownian motion corresponding to $\mathbb{P}^u$ by B^u. To summarize this section, we have proved the following result under Assumption 1-3.

Theorem 9. *Given $u \in \mathcal{U}$, there exists a unique weak solution to the sticky reflected SDE of mean-field type with boundary diffusion*

$$dX_t = (\sigma(X_t)\beta\,(t, X_\cdot, \mathbb{P}^u(t), u_t) + a(X_t))\,dt + \sigma(X_t)dB_t^u. \tag{16}$$

Under $\mathbb{P}^u$ the t-marginal distribution of $X_\cdot$ is $\mathbb{P}^u(t)$ for $t \in [0, T]$ and $X_\cdot$ is almost surely $C([0, T]; \bar{\mathcal{D}})$-valued. Furthermore, $\mathbb{P}^u \in \mathcal{P}_p(\Omega)$.

Proof. We are left to show that $\mathbb{P}^u\left(X. \in C([0,T];\bar{\mathcal{D}})\right) = 1$, all other statements of the theorem have been proved. Since $\mathbb{P}(X. \notin C([0,T];\bar{\mathcal{D}}) = 0$,

$$\mathbb{P}^u\left(X. \notin C([0,T];\bar{\mathcal{D}})\right) = \mathbb{E}\left[L_T^u 1_{\{X.\notin C([0,T];\bar{\mathcal{D}})\}}\right] = 0,$$

which proves that $X.$ is $\mathbb{P}^u$-almost surely $C([0,T];\bar{\mathcal{D}})$-valued. □

Remark 10. *The drift component β is projected in the tangential direction of the boundary by σ whenever the process is at the boundary (cf.* (9)*). The drift component a is not effected by the transformation. From a modeling perspective, the interpretation is that the pedestrian's tangential movement is partially controllable but also influenced by other pedestrians through the mean field. The normal direction is an uncontrolled delayed reflection.*

C.4 Mean-field type optimal control

Let E^u denote expectation taken under $\mathbb{P}^u$. To apply the stochastic maximum principle of [8], we make the assumption that the mean-field type Girsanov kernel β depends linearly on $\mathbb{P}^u$.

Assumption 4. *Let $\widetilde{\beta} : [0,T]\times\Omega\times\mathbb{R}^d\times U \to \mathbb{R}^d$ and let $r_\beta : \mathbb{R}^d \to \mathbb{R}^d$, and assume that*

$$\beta\left(t, X., \mathbb{P}^u(t), u_t\right) = \widetilde{\beta}\left(t, X., E^u\left[r_\beta(X_t)\right], u_t\right).$$

With some abuse of notation, we will continue to denote the Girsanov kernel by β, although from now this refers to $\widetilde{\beta}$. Let $f : [0,T]\times\Omega\times\mathbb{R}^d\times U \to \mathbb{R}$, $g : \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}$, $r_f : \mathbb{R}^d \to \mathbb{R}^d$, and $r_g : \mathbb{R}^d \to \mathbb{R}^d$.

Assumption 5. *For every $u \in \mathcal{U}$, the process $\left(f(t, X., E^u[r_f(X_t)], u_t)\right)_t$ is progressively measurable with respect to $\mathbb{F}$ and $(x,y) \mapsto g(x,y)$ is Borel measurable.*

Consider the finite horizon mean-field type cost functional $J : \mathcal{U} \to \mathbb{R}$,

$$J(u) := E^u\left[\int_0^T f\left(t, X., E^u\left[r_f(X_t)\right], u_t\right)dt + g\left(X_T, E^u\left[r_g(X_T)\right]\right)\right]. \quad (17)$$

The control problem considered in this section is the minimization of J with respect to $u \in \mathcal{U}$ under the constraint that the coordinate process for any given u satisfies (16). The integration in (17) is with respect to a measure absolutely continuous with respect to $\mathbb{P}$. Changing measure, we get

$$\begin{aligned} J(u) = E\bigg[\int_0^T & L_t^u f\left(t, X_\cdot, E[L_t^u r_f(X_t)], u_t\right) dt \\ & + L_T^u g\left(X_T, E[L_T^u r_g(X_T)]\right) \bigg], \end{aligned} \tag{18}$$

where E is the expectation taken under the original probability measure $\mathbb{P}$ and L^u the controlled likelihood process, given by the SDE of mean-field type

$$dL_t^u = L_t^u \beta\left(t, X_\cdot, E\left[L_t^u r_\beta(X_t)\right], u_t\right)^* dB_t, \quad L_0^u = 1. \tag{19}$$

C.4.1 Necessary optimality conditions

After making one final assumption about the regularity of β, f, and g (Assumption 6 below), the stochastic maximum principle yields necessary conditions on an optimal control for the minimization of (18) subject to (19). Assumption 4 and 6 are stated in their current form for the sake of technical, not conceptual, simplicity and may be relaxed.

Assumption 6. *The functions* $(t, x, y, u) \mapsto (f, \beta)(t, x, y, u)$ *and* $(x, y) \mapsto g(x, y)$ *are twice continuously differentiable with respect to* y*. Moreover,* β, f *and* g *and all their derivatives up to second order with respect to* y *are continuous in* (y, u)*, and bounded.*

The next result is a slight generalization of [8, Thm 2.1]. The paper [8] treats an optimal control problem of mean-field type with deterministic coefficients. The approach of [8], which goes back to [41], extends without any further conditions to include random coefficients, as shown in [31]. Moreover, in our case the coefficients are not bounded functions, they are linear in the likelihood. This could seem to violate the conditions of [8, Thm 2.1], but an application of Grönwall's lemma yields $E[(L_t^u)^p] \leq \exp(C(p)t)$ for all $t \in [0, T]$ and $p \geq 2$, where $C(p)$ is a bounded constant, and the estimates of [8] can be recovered after an application of Hölder's inequality.

Theorem 11. *Assume that* $(\hat{u}, L^{\hat{u}})$ *solves the optimal control problem* (18)-(19)*. Then there are two pairs of* $\mathbb{F}$*-adapted processes,* (p, q) *and* (P, Q)*, that* *satisfy the first and second order adjoint equations*

$$\begin{cases} dp_t = -\Big(q_t\beta_t^{\hat{u}} + E\left[q_t L_t^{\hat{u}}\nabla_y\beta_t^{\hat{u}}\right] r_\beta(X_t) \\ \qquad\qquad - f_t^{\hat{u}} - E\left[L_t^{\hat{u}}\nabla_y f_t^{\hat{u}}\right] r_f(X_t)\Big)dt + q_t dB_t, \\ p_T = -g_T^{\hat{u}} - E\left[L_T^{\hat{u}}\nabla_y g_T^{\hat{u}}\right] r_g(X_T), \end{cases}$$

$$\begin{cases} dP_t = -\Big(\left|\beta_t^{\hat{u}} + E\left[L_t^{\hat{u}}\nabla_y\beta_t^{\hat{u}}\right] r_\beta(X_t)\right|^2 P_t \\ \qquad\qquad + 2Q_t\left(\beta_t^{\hat{u}} + E\left[L_t^{\hat{u}}\nabla_y\beta_t^{\hat{u}}\right] r_\beta(X_t)\right)\Big)dt + Q_t dB_t, \\ P_T = 0, \end{cases}$$

where ∇_y *denotes differentiation with respect to the* $\mathbb{R}^d$*-valued argument. Furthermore,* (p,q) *and* (P,Q) *satisfy*

$$E\left[\sup_{t\in[0,T]}|p_t|^2 + \int_0^T |q_t|^2 dt\right] < \infty, \quad E\left[\sup_{t\in[0,T]}|P_t|^2 + \int_0^T |Q_t|^2 dt\right] < \infty,$$

and for every $u \in U$ *and a.e.* $t \in [0,T]$*, it holds* $\mathbb{P}$*-a.s. that*

$$\mathcal{H}\left(L_t^{\hat{u}}, u, p_t, q_t\right) - \mathcal{H}\left(L_t^{\hat{u}}, \hat{u}_t, p_t, q_t\right) + \frac{1}{2}\left[\delta\left(L\beta\right)(t)\right]^T P_t \left[\delta\left(L\beta\right)(t)\right] \le 0, \quad (20)$$

where $\mathcal{H}(L_t^u, u_t, p_t, q_t) := L_t^u\beta_t^u q_t - L_t^u f_t^u$ *and*

$$\delta(L\beta)(t) := L_t^{\hat{u}}\left(\beta\left(t, X_\cdot, E[L_t^{\hat{u}} r_\beta(X_t)], u\right) - \beta_t^{\hat{u}}\right).$$

The following local form of the optimality condition (20) can be found in e.g. [48, pp. 120], and will be useful for the computations in Section C.5. If U is a convex set and $\mathcal{H}$ is differentiable with respect to u, then (20) implies

$$(u - \hat{u}_t)^* \nabla_u \mathcal{H}\left(L_t^{\hat{u}}, \hat{u}_t, p_t, q_t\right) \le 0, \quad \forall\, u \in U, \text{ a.e. } t \in [0,T], \ \mathbb{P}\text{-a.s.} \quad (21)$$

Remark 12. *Sufficient conditions for weak optimal controls will seldom be satisfied since they typically require the Hamiltonian to be convex (or concave) in at least state* (L_t^u) *and control* (u_t)*. This is false even for the simplest version of our problem. Assume that* $\beta(t,\omega,y,u) = u$ *and* $f = 0$*, then* $(\ell, u) \mapsto \mathcal{H}(\ell, u, p, q) = \ell u q$*, which is neither convex nor concave. However, necessary optimality conditions can be useful as we will see in Section C.5.*

C.4.2 Microscopic interpretation of the mean-field type control problem

In this section, we give a microscopic interpretation of the mean-field type control problem (7) in the form of an interacting particle system (collaboratively) minimizing the social cost. Our means will be the propagation of chaos result [39, Thm. 2.6]. We will work under all the assumptions stated so far, but we will use the notation from Section C.3 for β, f, and g.

We will fix a closed-loop control and we will assume that all the interacting particles are using this control. This assumption is made in order to extract the approximating property of any solution to the mean-field optimal control problem that is on closed-loop form. In Section C.5, we will see examples of such controls.

We introduce an interacting system of sticky reflected SDEs with boundary diffusion. Each equation has an initial value with distribution λ, where λ is a nonatomic measure and $\lambda(\bar{\mathcal{D}}) = 1$. See Remark 10 in [39] for the necessity of the random initial condition.

Consider the measure $\mathbb{P}^{\otimes N}$ on $(\Omega^N, \mathcal{B}(\Omega^N))$, the weak solution to a system of $N \in \mathbb{N}$ i.i.d. sticky reflected Brownian motions with boundary diffusion

$$dX_t^{N,i} = a(X_t^{N,i})dt + \sigma(X_t^{N,i})dB_t^i, \quad X_0^{N,i} = \xi^{N,i}, \quad i = 1, \dots, N,$$

where $\xi_1, \dots, \xi_N$ are random variables with law λ and $B^1, \dots, B^N$ are independent $\mathbb{F}$-Wiener processes, all independent of each other. The functions a and σ are defined as in (10). Given controls $u^i \in \mathcal{U}$ (now $\mathbb{F}$-progressively measurable), $i = 1, 2, \dots$, define the likelihood process $L_{\mathbf{u},t}^{N,i}$ as the solution to

$$dL_{\mathbf{u},t}^{N,i} = L_{\mathbf{u},t}^{N,i} \beta\left(t, X_\cdot^{N,i}, \mu_t^N, u_t^i\right)^* dB_t^i, \quad L_{\mathbf{u},0}^{N,i} = 1, \quad i = 1, \dots, N,$$

where μ^N is the empirical measure of the coordinate processes,

$$\mu^N := \frac{1}{N} \sum_{i=1}^N \delta_{X_\cdot^i} \in \mathcal{P}(\Omega).$$

Then $L_{\mathbf{u},t}^N := \prod_{i=1}^N L_{\mathbf{u},t}^{N,i}$ is the Radon-Nikodym derivative for the Girsanov-type change of measure from $\mathbb{P}^{\otimes N}$ to $\mathbb{P}^{N,\mathbf{u}}$, under which the coordinate processes satisfy

$$\begin{cases} dX_t^{N,i} = \Big(a(X_t^{N,i}) + \sigma(X_t^{N,i})\beta\left(t, X_\cdot^{N,i}, \mu_t^N, u_t^i\right)\Big)\,dt + \sigma(X_t^{N,i})d\widetilde{B}_t^i, \\ X_0^{N,i} = \xi^{N,i}, \quad i = 1, \dots, N, \end{cases} \tag{22}$$

where $\widetilde{B}^1, \dots$ are $\mathbb{P}^{N,\mathbf{u}}$-Brownian motions and $\mathbf{u} := (u^1, \dots, u^N)$. We note that $\mathbb{P}^{N,\mathbf{u}}$ is the law of an interacting system of diffusion processes. The social cost of the system (22) is defined as

$$\frac{1}{N}\sum_{i=1}^N J^i(\mathbf{u}) := \frac{1}{N}\sum_{i=1}^N E^{N,\mathbf{u}}\left[\int_0^T f(t, X_\cdot^{N,i}, \mu_t^N, u_t^i)dt + g(X_T^i, \mu_T^N)\right].$$

The following theorem is an adaptation of [39, Thm. 2.6] where the drift $b := a + \sigma\beta$ and the Girsanov kernel $\sigma^{-1}b := \beta$.

Theorem 13. *Let $u \in \mathcal{U}$ be a closed-loop control, i.e. $u_t(\omega) = \varphi(\omega_{\cdot\wedge t})$ for some measurable function $\varphi : (\Omega, \mathcal{F}) \to (U, \mathcal{B}(U))$. Given the control u and a random variable ξ with law λ (nonatomic with support only on $\bar{\mathcal{D}}$), the sticky reflected SDE of mean-field type with boundary diffusion*

$$\begin{cases} dX_t = (a(X_t) + \sigma(X_t)\beta(t, X_\cdot, \mathbb{P}^u(t), \varphi(X_{\cdot\wedge t})))\,dt + \sigma(X_t)dB_t, \\ X_0 = \xi, \end{cases}$$

can be approximated by the interacting particle system (22) with all components using the fixed closed-loop control u. Furthermore, the value of the mean-field cost functional J at u is the asymptotic social cost of the interacting particle system as $N \to \infty$ when all the $X^{N,i}$s are using the fixed control u. More specifically,

$$\lim_{N\to\infty} D_T\left(\mathbb{P}^{N,\mathbf{u}} \circ (X_\cdot^{N,1}, \dots, X_\cdot^{N,k})^{-1}, (\mathbb{P}^u \circ X_\cdot^{-1})^{\otimes k}\right) = 0, \tag{23}$$

with $\mathbf{u} = (u, \dots, u)$, and

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N J^i(u, \dots, u) \to J(u).$$

Proof. We denote by $\mathcal{E}(\mathcal{P}(\Omega))$ the smallest σ-field on $\mathcal{P}(\Omega)$ such that the map $\mu \mapsto \int_\Omega \phi d\mu$ is measurable for all bounded and measurable $\phi : \Omega \to \mathbb{R}$.

As pointed out in [39], $\mathcal{E}(\mathcal{P}(\Omega))$ coincides with the Borel σ-field on $\mathcal{P}(\Omega)$ generated by the topology of weak convergence.

To verify the assumptions of [39, Thm. 2.6], we note that β is progressively measurable with respect to $\mathbb{F}$ and that β is Lipschitz continuous in the measure-valued argument with respect to d_{TV}. This implies condition $(\mathcal{E})$ in [39], the $\mathcal{E}(\mathcal{P}(\Omega))$-measurability of the function

$$\begin{aligned}
&F_{s,t} : \mathcal{P}(\Omega) \to \mathbb{R},\\
&\qquad F_{s,t}(\nu) = \int_\Omega \int_s^t |\beta(u,\omega,\nu_t) - \beta(u,\omega,\mathbb{P}^u(t))|^2 \, du \; \nu(d\omega),
\end{aligned}$$

the $\tau(\Omega)$-continuity of $F_{s,t}$, and the inequality (2.3) from [39, Thm. 2.6]. Furthermore, β is bounded, implying condition (A) in [39]. So the propagation of chaos (23) holds.

By [43, Prop. 2.2], the propagation of chaos implies that $\mathcal{P}(\mathcal{P}(\Omega)) \ni M^N := \mathbb{P}^{N,\mathbf{u}} \circ (\mu^N)^{-1} \to \delta_{\mathbb{P}^u \circ X_\cdot^{-1}}$ in the weak topology. By assumption, f and g are bounded and continuous in the y-argument. Hence,

$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N J^i(u,\dots,u)\\
&= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N E^{N,\mathbf{u}} \left[\int_0^T f\left(t, X^{N,i}, \mu_t^N, \varphi(X^{N,i}_{\cdot\wedge t})\right) dt + g(X_T^{N,i}, \mu_T^N) \right]\\
&= \lim_{N\to\infty} E^{N,\mathbf{u}} \Bigg[\int_0^T \int_\Omega f\left(t, \omega', \mu_t^N, \varphi(\omega'_{\cdot\wedge t})\right) \mu^N(d\omega') dt\\
&\qquad\qquad\qquad + \int_\Omega g(\omega'(T), \mu_T^N) \mu^N(d\omega') \Bigg]\\
&= \lim_{N\to\infty} \int_0^T \int_{\mathcal{P}(\Omega)} \left\{ \int_\Omega f\left(t, \omega', \int_\Omega r_f(\omega''(t)) m(d\omega''), \varphi(\omega'_{\cdot\wedge t})\right) m(d\omega') \right\}\\
&\qquad\qquad\qquad\qquad\qquad\qquad M^N(dm) dt\\
&+ \lim_{N\to\infty} \int_{\mathcal{P}(\Omega)} \int_\Omega g\left(\omega'(T), \int_\Omega r_g(\omega''(T)) m(d\omega'')\right) m(d\omega') M^N(dm).\\
&= E^u \left[\int_0^T f\left(t, X_\cdot, \mathbb{P}^u(t)\right) dt + g\left(X_T, \mathbb{P}^u(T)\right) \right] = J(u).
\end{aligned}$$

□

C.5 Examples

As a first step in model validation, experimental results on pedestrian speed profiles in a long narrow corridor are replicated in this section. The application of the proposed approach also displays the new features it offers regarding behavior near walls. From the necessary optimality conditions we derive an expression for the optimal control valid in the following two toy examples and the corridor scenario. The numerical simulations are based on the particle system approximation derived in Section C.4.2.

Throughout the rest of this section it is assumed that the set U is convex and sufficiently large so that all the controls in the following analytical expressions are admissible. Furthermore, it is assumed that r_g is differentiable and that $(\hat{u}, L^{\hat{u}})$ is optimal for the mean-field type control problem (18)-(19). We recall the first order adjoint equation,

$$\begin{cases} dp_t = -\Big(q_t\beta_t^{\hat{u}} + E\left[q_t L_t^{\hat{u}}\nabla_y\beta_t^{\hat{u}}\right] r_\beta(X_t) \\ \qquad\qquad\qquad - f_t^{\hat{u}} - E\left[L_t^{\hat{u}}\nabla_y f_t^{\hat{u}}\right] r_f(X_t)\Big)dt + q_t dB_t, \\ p_T = -g_T^{\hat{u}} - E\left[L_T^{\hat{u}}\nabla_y g_T^{\hat{u}}\right] r_g(X_T). \end{cases} \tag{24}$$

Rewriting $E[L_t^{\hat{u}}Y_t] = E^{\hat{u}}[Y_t]$ and changing measure to $\mathbb{P}^{\hat{u}}$, (24) becomes

$$\begin{cases} dp_t = -A_t dt + q_t dB_t^{\hat{u}}, \\ p_T = -g_T^{\hat{u}} - E^{\hat{u}}\left[\nabla_y g_T^{\hat{u}}\right] r_g(X_T), \end{cases} \tag{25}$$

where $A_t := E^{\hat{u}}\left[q_t\nabla_y\beta_t^{\hat{u}}\right] r_\beta(X_t) - f_t^{\hat{u}} - E^{\hat{u}}\left[\nabla_y f_t^{\hat{u}}\right] r_f(X_t)$. By the martingale representation theorem (see e.g. [36, pp. 182]) p can be written as the conditional expectation

$$p_t = -E^{\hat{u}}\left[g_T^{\hat{u}} + E^{\hat{u}}[\nabla_y g_T^{\hat{u}}] r_g(X_T) \mid \mathcal{F}_t\right] + E^{\hat{u}}\left[\int_t^T A_s ds \mid \mathcal{F}_t\right]. \tag{26}$$

The theorem applies to our problem since g and its y-derivative are assumed to be bounded. Let

$$\phi(t, X_t) := g\left(X_t, E^{\hat{u}}[r_g(X_t)]\right) + E^{\hat{u}}[\nabla_y g_t^{\hat{u}}] r_g(X_t).$$

By Dynkin's formula,

$$E^{\hat{u}}[\phi(T, X_T) \mid \mathcal{F}_t] = \phi(t, X_t) + \int_t^T E^{\hat{u}}\left[(\mathcal{G} + \partial_s)\phi(s, X_s) \mid \mathcal{F}_t\right] ds,$$

where $\mathcal{G}$ is the generator of the coordinate process and ∂_s denotes differentiation with respect to time. Hence, by applying Itô's formula on p in (26), where only $X.$ contributes to the diffusion part, and matching the diffusion parts of that and p from (25), we get

$$q_s = -\nabla_x \phi(s, X_s)\sigma(X_s). \tag{27}$$

The local optimality condition in the case of a convex U and coefficients differentiable in u, given in (21) right below Theorem 11, can be used to write $\hat{u}$ in terms of the other processes. To use it, we make the following assumption.

Assumption 7. *The functions $(t, x, y, u) \mapsto (f, \beta)(t, x, y, u)$ are differentiable with respect to u.*

With Assumption 7 in force, an optimal control $\hat{u}$ satisfies the the local optimality condition. The local optimality condition is satisfied by any $\hat{u}$ such that $\nabla_u \mathcal{H}(L_t^{\hat{u}}, \hat{u}_t, p_t, q_t) = 0$ for almost every $t \in [0, T]$, $\mathbb{P}$-a.s., i.e.

$$q_t \nabla_u \beta_t^{\hat{u}} = \nabla_u f_t^{\hat{u}}, \quad \text{a.e. } t \in [0, T], \ \mathbb{P}\text{-a.s.}. \tag{28}$$

Since $\mathbb{P}^{\hat{u}}$ is absolutely continuous with respect to $\mathbb{P}$, the equality above also holds for almost every $t \in [0, T]$ $\mathbb{P}^{\hat{u}}$-a.s. We have now at hand an expression for the optimal control whenever we can solve (27)-(28) for $\hat{u}$.

C.5.1 Linear-quadratic problems with convex U

C.5.1.1 A non-mean-field example

Let $\mathcal{D} \subset \mathbb{R}^d$ be an admissible domain and $\mathbb{P}$ the probability measure on the space of continuous paths under which the coordinate process solves (9). Consider the following linear-quadratic optimal control problem on $\mathcal{D}$,

$$\begin{cases} \min\limits_{u \in \mathcal{U}} \dfrac{1}{2} E\left[\displaystyle\int_0^T L_t^u |u_t|^2 dt + L_T^u |X_T - x_T|^2\right], \\ \text{s.t. } dL_t^u = L_t^u u_t^* dB_t, \quad L_0^u = 1, \end{cases}$$

where B is a $\mathbb{P}$-Brownian motion. The necessary optimality condition (28) yields

$$\hat{u}_t = q_t^*, \quad \mathbb{P}\text{-a.s., a.e. } t \in [0, T].$$

Matching the diffusion coefficients gives us the optimal control,

$$\hat{u}_t = -\sigma(X_t)\left(X_t - x_T\right), \quad \mathbb{P}\text{-a.s., a.e. } t \in [0,T].$$

The corresponding likelihood process solves

$$dL_t^{\hat{u}} = -L_t^{\hat{u}}\left(X_t - x_T\right)^* \sigma(X_t) dB_t, \quad L_0^{\hat{u}} = 1,$$

and under $\mathbb{P}^{\hat{u}}$, the optimally controlled path distribution, the coordinate process solves

$$\begin{aligned} dX_t &= a(X_t)dt + \sigma(X_t)dB_t \\ &= a(X_t)dt + \sigma(X_t)\left(-\sigma(X_t)\left(X_t - x_T\right)dt + dB_t^{\hat{u}}\right) \\ &= \left(a(X_t) - \sigma(X_t)\left(X_t - x_T\right)\right)dt + \sigma(X_t)dB_t^{\hat{u}}. \end{aligned}$$

We have used the fact that $\pi^2 = \pi = \pi^*$, which holds since π is an orthogonal projection.

C.5.1.2 A mean-field example

Consider now on some admissible domain $\mathcal{D} \subset \mathbb{R}^d$ the mean-field type optimal control problem

$$\begin{cases} \min\limits_{u \in \mathcal{U}} \dfrac{1}{2} E\left[\displaystyle\int_0^T L_t^u |u_t|^2 dt + L_T^u \left|X_T - E\left[L_T^u X_T\right]\right|^2\right], \\ \text{s.t. } dL_t^u = L_t^u u_t^* dB_t, \quad L_0^u = 1. \end{cases}$$

As before, B is a $\mathbb{P}$-Brownian motion, where $\mathbb{P}$ is a probability measure on the path space under which the coordinate process solves (9). Then $E^{\hat{u}}[\nabla_y g_t^{\hat{u}}] = 0$, so (since $r_g(x) = x$ here)

$$\nabla_x \phi\left(t, X_t\right) = \left(X_t - E^{\hat{u}}[X_t]\right)^*,$$

and (28) yields $\hat{u}_t = -\sigma(X_t)(X_t - E^{\hat{u}}[X_t])$ $\mathbb{P}$-a.s. for almost every $t \in [0,T]$. Under $\mathbb{P}^{\hat{u}}$ the coordinate process solves

$$dX_t = \left(a(X_t) - \sigma(X_t)\left(X_t - E^{\hat{u}}\left[X_t\right]\right)\right)dt + \sigma(X_t)dB_t^{\hat{u}}.$$

C.5.2 Unidirectional pedestrian motion in a corridor

Experimental studies have been conducted on the impact of proximity to walls on pedestrian speed. Pedestrian speed profiles heavily depend on circumstances like location, weather, and congestion. In this section, we will replicate two scenarios of unidirectional motion in a confined domain with the proposed mean-field type optimal control model. Especially, we are interested in how the proposed model behaves on the boundary and if boundary movement characteristics can be influenced through the running cost f. Sticky boundaries and boundary diffusion grants our pedestrians controlled movement at the boundary. By altering the internal parameters of these effect, we are able to shape the mean speed profile at the boundary.

Zanlungo *et al.* [49] observe that in a tunnel connecting a shopping center with a railway station in Osaka, Japan, pedestrians tend to lower their walking speed when walking close to the walls. The authors obtain a concave cross-section average speed profile from their experiment, with its maximum approximately at the center of the corridor. The average speed at the center of the corridor is about 10% higher than that of near-wall walkers.

Daamen and Hoogendoorn [19] on the other hand observe (in a controlled environment) pedestrian speeds that are higher at the boundary than in the interior of the domain. In their experiment, a unidirectional stream of pedestrians walk in a wide corridor that at a certain point, at a *bottleneck*, shrinks into a tight corridor. Upstream from the bottleneck, pedestrians close to the corridor walls move more freely due to less congestion, compared to those at the center of the corridor. The experiment results in a cross-section speed profile with more than twice as high average pedestrian speed in the low-density regions along corridor walls compared to the center of the corridor.

By modeling congestion with simple mean-dependent effects, we can replicate the overall shape of the average speed profiles of both [49] and [19] (not the density profile, to achieve this one needs a more sophisticated mean-field model). Our reason for implementing only mean-dependent effects instead of nonlocal distribution-dependent effects (like those considered in for example [4]) is solely to simplify the analysis.

Consider a long narrow corridor with walls parallel to the x-axis at $y = -0.1$ and $y = 0.1$. Our analysis requires $\mathcal{D}$ to be C^2-smooth, so the effective corridor (the corridor perceived by the pedestrians) has rounded corners. However, the corners will not have any substantial effect on the simulation results since the crowd is initiated so far away from the target that under the chosen coefficient values, the pedestrians will not reach it ahead of the time horizon $T = 1$. On this domain, crowd behavior is modeled with the following optimal control problem

$$\begin{cases} \min_{u.\in\mathcal{U}} \frac{1}{2} E\left[\int_0^1 L_t^u f\left(t, X., E\left[L_t^u r_f(X_t)\right], u_t\right) dt + L_T^u \left|X_T - x_T\right|^2\right], \\ \text{s.t. } dL_t^u = L_t^u u_t dB_t, \quad L_0^u = 1, \end{cases}$$

where B is a Brownian motion under $\mathbb{P}$, the probability measure under which $X.$ solves (9) with $\gamma = 0.5$, and x_T is the location of an exit at the end of the corridor. The choice of γ is made so that the plots below are visually comparable. The running cost f is of congestion-type,

$$f\left(t, X., E\left[L_t^u r_f(X_t)\right], u_t\right) = \mathcal{C}(X_t)\Big(c_f + h\left(t, X., E^u\left[r_f(X_t)\right]\right)\Big)u_t^2,$$

where $c_f > 0$ is a positive constant, $c_f u^2$ is the cost of moving in free space, and hu^2 the additional cost to move in congested areas. The coefficient $\mathcal{C}(X_t) := c_\Gamma 1_\Gamma(X_t) + 1_\mathcal{D}(X_t)$, $c_\Gamma > 0$, is used to monitor f (though it is not our control process) on the boundary Γ. The cost of moving on the boundary is increasing with c_Γ, so for high values of c_Γ we expect a lower speed on the boundary. We know from (28)-(27) that

$$q_t^* = \mathcal{C}(X_t)\Big(c_f + h\left(t, X., E^{\hat{u}}\left[r_f(X_t)\right]\right)\Big)\hat{u}_t, \quad q_t = -(X_t - x_T)^* \sigma(X_t). \tag{29}$$

Matching the expressions in (29) yields the optimal control

$$\hat{u}_t = \frac{\sigma(X_t)\left(X_t - x_T\right)}{\mathcal{C}(X_t)\Big(c_f + h\left(t, X., E^{\hat{u}}\left[r_f(X_t)\right]\right)\Big)}.$$

It implements the following strategy: move towards the target location x_T, but scale the speed according to the local congestion. Consider the two congestion penalties

$$h_1 := \left|X_2(t) - E^{\hat{u}}\left[X_2(t)\right]\right|, \quad h_2 := \frac{1}{\left|X_2(t) - E^{\hat{u}}\left[X_2(t)\right]\right|}, \tag{30}$$

where $X_2(t)$ is the second (the y-)component of the coordinate process, i.e. the component in the direction perpendicular to the corridor walls. Stickiness is set to $\gamma = 0.5$. The choice of h in (30) means that we have set $r_f(X_t) = X_2(t)$.

The corridor is split into 9 segments parallel with the corridor walls. The mean speed is estimated in each segment for four different values of c_Γ and the results corresponding to congestion penalty h_1 and h_2 are presented in Figure 1 and 2, respectively. The profiles plotted in Figure 1 attains the concave shape observed by [49], mimicking the fast track in the middle of the lane. In Figure 2 the profiles follow the convex shape observed by [19], taking into account that movement in the crowded center (mean of the group) is costly. When c_Γ is small, the pedestrians can travel further on the boundary for the same cost. Heuristically, the higher γ is the longer it takes for the pedestrian to re-enter $\mathcal{D}$ and therefore a high γ combined with a small c_Γ yields the highest boundary speed. This effect is evident in the figures, where smaller values of c_Γ results in higher mean speed at the boundary. We note that we are able to shape the mean speed at the boundary by our choice of model parameters.

C.6 Conclusion and discussion

In this paper, we propose a to the best of our knowledge new variation of the mean-field approach to crowd modeling based on sticky reflected SDEs. The proposed model accounts for pedestrians that spend some time at the boundary and have the possibility to choose a new direction of motion while being on the boundary.

We provide conditions for the proposed model equations to admit a unique weak solution, which is the best we can hope for (cf. [22]). Then, we consider mean-field type optimal control of the proposed dynamic model and give necessary conditions for optimality with a Pontryagin-type stochastic maximum principle. There is a microscopic interpretation of the model even on the boundary of the domain and thus it can be used to approximate optimal (or equilibrium) behavior of a pedestrian crowd on a microscopic, i.e. individual, level.

Pedestrians do often see and react to walls at a distance. This has been studied empirically, some experiments are mentioned in the introduction.

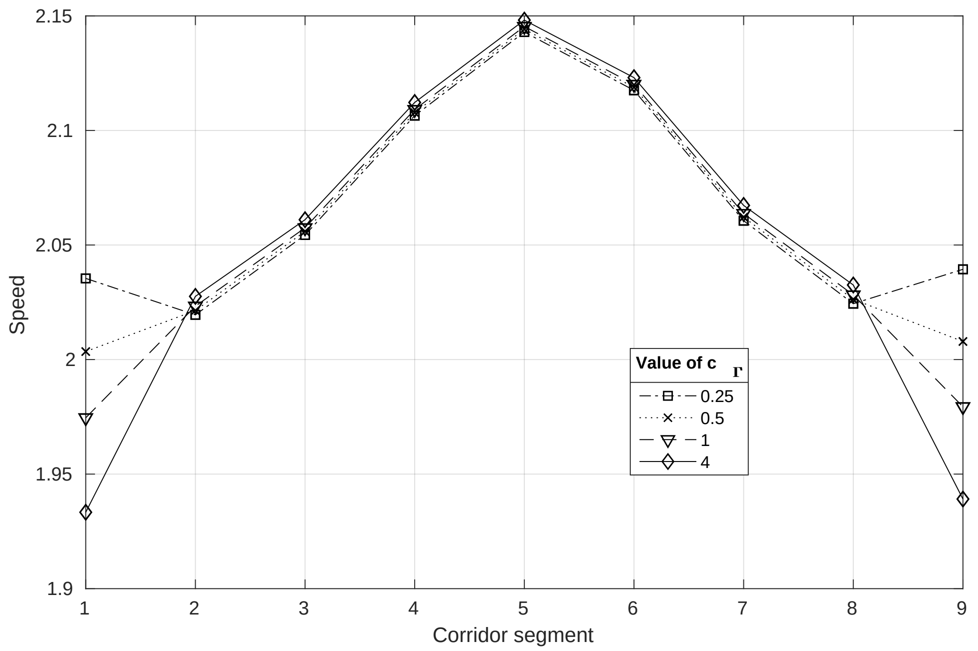


Figure 1: Mean speed in 9 segments of the corridor when $h = h_1$ and $c_f = 1$, estimated from 4000 realizations of the controlled coordinate process.

Force-based models can implement repulsing potential forces spiking to infinity at boundaries to keep the pedestrians away from the walls and inside the domain, effectively making it impossible for any pedestrian to reach a wall. A ranged, *nonlocal*, interaction with walls will have a smoothing effect on pedestrian density, just like nonlocal pedestrian-to-pedestrian interaction has, as is noted in [4]. Nonlocal interaction is an important aspect of pedestrian crowd modeling, but cannot give an answer to what will happen whenever a pedestrian actually reaches a wall. Interaction with walls at a distance can be included in our proposed model either in the drift, as is the case in force-based models, or through the cost functional, as in agent-based models.

An extension of the proposed framework would be to let the pedestrian control her stickiness, i.e. her motion in the normal direction of the boundary at the boundary. Stickiness is not necessarily a physical feature of the

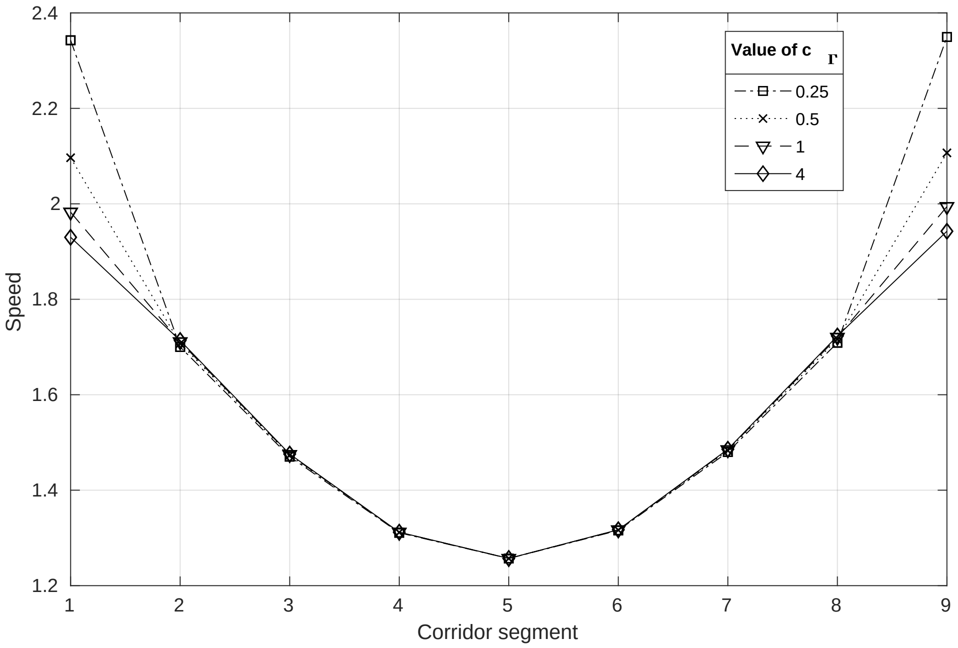


Figure 2: Mean speed in 9 segments of the corridor when $h = h_2$ and $c_f = 0.1$, estimated from 4000 realizations of the controlled coordinate process.

domain, but the time spent on the boundary may be subject to the pedestrian's preference. This aspect cannot be described by the proposed model, since the Girsanov change of measure does not effect stickiness (cf. Remark 10). Another extension would be to consider the controlled diffusion case mentioned in the introduction.

References

[1] Y. Achdou, M. Bardi, and M. Cirant. Mean field games models of segregation. *Mathematical Models and Methods in Applied Sciences*, 27 (01):75–113, 2017.

[2] G. Albi, Y.-P. Choi, M. Fornasier, and D. Kalise. Mean field control hierarchy. *Applied Mathematics & Optimization*, 76(1):93–135, 2017.

[3] M. Amir. Sticky Brownian motion as the strong limit of a sequence of random walks. *Stochastic processes and their applications*, 39(2):221–237, 1991.

[4] A. Aurell and B. Djehiche. Mean-field type modeling of nonlocal crowd aversion in pedestrian crowd dynamics. *SIAM Journal on Control and Optimization*, 56(1):434–455, 2018.

[5] M. Bardi and M. Cirant. Uniqueness of solutions in mean field games with several populations and Neumann conditions. In *PDE Models for Multi-Agent Phenomena*, pages 1–20. Springer, 2018.

[6] N. Bellomo and L. Gibelli. Toward a mathematical theory of behavioral-social dynamics for pedestrian crowds. *Mathematical Models and Methods in Applied Sciences*, 25(13):2417–2437, 2015.

[7] N. Bellomo and L. Gibelli. Behavioral crowds: Modeling and Monte Carlo simulations toward validation. *Computers & Fluids*, 141:13–21, 2016.

[8] R. Buckdahn, B. Djehiche, and J. Li. A general stochastic maximum principle for SDEs of mean-field type. *Applied Mathematics & Optimization*, 64(2):197–216, 2011.

[9] R. Buckdahn, J. Li, and J. Ma. A stochastic maximum principle for general mean-field systems. *Applied Mathematics & Optimization*, 74 (3):507–534, 2016.

[10] M. Burger, P. Markowich, and J.-F. Pietschmann. Continuous limit of a crowd motion and herding model: analysis and numerical simulations. *Kinet. Relat. Models*, 4(4):1025–1047, 2011.

[11] M. Burger, M. Di Francesco, P. A. Markowich, and M.-T. Wolfram. On a mean field game optimal control approach modeling fast exit scenarios in human crowds. In *52nd IEEE Conference on Decision and Control*, pages 3128–3133. IEEE, 2013.

[12] M. Burger, M. Di Francesco, P. Markowich, and M.-T. Wolfram. Mean field games with nonlinear mobilities in pedestrian dynamics. *Discrete & Continuous Dynamical Systems-B*, 19(5):1311–1333, 2014.

[13] M. Burger, M. Di Francesco, P. A. Markowich, and M.-T. Wolfram. Mean field games with nonlinear mobilities in pedestrian dynamics. *Discrete and Continuous Dynamical Systems-Series B*, 19(5):1311–1333, 2014.

[14] M. Burger, S. Hittmeir, H. Ranetbauer, and M.-T. Wolfram. Lane formation by side-stepping. *SIAM Journal on Mathematical Analysis*, 48(2):981–1005, 2016.

[15] R. Chitashvili. On the nonexistence of a strong solution in the boundary problem for a sticky brownian motion. *Department of Operations Research and System Theory [BS]*, (R 8901), 1989.

[16] S. E. Choutri, B. Djehiche, and H. Tembine. Optimal control and zero-sum games for Markov chains of mean-field type. *Mathematical Control and Related Fields*, 2018.

[17] M. Cirant. Multi-population mean field games systems with neumann boundary conditions. *Journal de Mathématiques Pures et Appliquées*, 103(5):1294–1315, 2015.

[18] E. Cristiani, F. S. Priuli, and A. Tosin. Modeling rationality to control self-organization of crowds: an environmental approach. *SIAM Journal on Applied Mathematics*, 75(2):605–629, 2015.

[19] W. Daamen and S. P. Hoogendoorn. Flow-density relations for pedestrian traffic. In *Traffic and granular flow'05*, pages 315–322. Springer, 2007.

[20] B. Djehiche and S. Hamadène. Optimal control and zero-sum stochastic differential game problems of mean-field type. *Applied Mathematics & Optimization*, pages 1–28, 2018.

[21] B. Djehiche, A. Tcheukam, and H. Tembine. A Mean-Field Game of Evacuation in Multilevel Building. *IEEE Transactions on Automatic Control*, 62(10):5154–5169, 2017.

[22] H.-J. Engelbert and G. Peskir. Stochastic differential equations for sticky Brownian motion. *Stochastics An International Journal of Probability and Stochastic Processes*, 86(6):993–1021, 2014.

[23] W. Feller. The parabolic differential equations and the associated semigroups of transformations. *Annals of Mathematics*, pages 468–519, 1952.

[24] W. Feller. Diffusion processes in one dimension. *Transactions of the American Mathematical Society*, 77(1):1–31, 1954.

[25] W. Feller et al. Generalized second order differential operators and their lateral conditions. *Illinois journal of mathematics*, 1(4):459–504, 1957.

[26] C. Graham. The martingale problem with sticky reflection conditions, and a system of particles interacting at the boundary. *Annales de l'IHP Probabilités et statistiques*, 24(1):45–72, 1988.

[27] M. Grothaus and R. Voßhall. Strong Feller property of sticky reflected distorted Brownian motion. *Journal of Theoretical Probability*, 31(2): 827–852, 2018.

[28] M. Grothaus, R. Voßhall, et al. Stochastic differential equations with sticky reflection and boundary diffusion. *Electronic Journal of Probability*, 22, 2017.

[29] J. M. Harrison and A. J. Lemoine. Sticky Brownian motion as the limit of storage processes. *Journal of Applied Probability*, 18(1):216–226, 1981.

[30] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995.

[31] J. J. A. Hosking. A stochastic maximum principle for a stochastic differential game of a mean-field type. *Applied Mathematics & Optimization*, 66(3):415–454, 2012.

[32] E. P. Hsu. *Stochastic analysis on manifolds*, volume 38. American Mathematical Soc., 2002.

[33] R. Hughes. The flow of large crowds of pedestrians. *Mathematics and Computers in Simulation*, 53(4-6):367–370, 2000.

[34] K. Itô and H. P. McKean. Brownian motions on a half line. *Illinois journal of mathematics*, 7(2):181–231, 1963.

[35] O. Kallenberg. *Foundations of modern probability.* Springer, New York, 1997.

[36] I. Karatzas and S. E. Shreve. *Brownian Motion and Stochastic Calculus.* Springer, 1988.

[37] A. Kirchner and A. Schadschneider. Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics. *Physica A: statistical mechanics and its applications*, 312(1-2): 260–276, 2002.

[38] A. Lachapelle and M.-T. Wolfram. On a mean field game approach modeling congestion and aversion in pedestrian crowds. *Transportation research part B: methodological*, 45(10):1572–1589, 2011.

[39] D. Lacker et al. On a strong form of propagation of chaos for McKean-Vlasov equations. *Electronic Communications in Probability*, 23, 2018.

[40] J. Ma, W.-G. Song, Z.-M. Fang, S.-M. Lo, and G.-X. Liao. Experimental study on microscopic moving characteristics of pedestrians in built corridor based on digital image processing. *Building and Environment*, 45(10):2160–2169, 2010.

[41] S. Peng. A general stochastic maximum principle for optimal control problems. *SIAM Journal on control and optimization*, 28(4):966–979, 1990.

[42] H. M. Soner, N. Touzi, and J. Zhang. Wellposedness of second order backward SDEs. *Probability Theory and Related Fields*, 153(1-2):149–190, 2012.

[43] A.-S. Sznitman. Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX—1989*, pages 165–251. Springer, 1991.

[44] A. D. Venttsel'. On boundary conditions for multidimensional diffusion processes. *Theory of Probability & Its Applications*, 4(2):164–177, 1959.

[45] J. Warren. Branching processes, the Ray-Knight theorem, and sticky Brownian motion. In *Séminaire de Probabilités XXXI*, pages 1–15. Springer, 1997.

[46] S. Watanabe and N. Ikeda. *Stochastic differential equations and diffusion processes*. Elsevier Science, 1981.

[47] K. Yamada. Reflecting or sticky Markov processes with Levy generators as the limit of storage processes. *Stochastic Processes and their Applications*, 52(1):135–164, 1994.

[48] J. Yong and X. Y. Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*, volume 43. Springer Science & Business Media, 1999.

[49] F. Zanlungo, T. Ikeda, and T. Kanda. A microscopic "social norm" model to obtain realistic macroscopic velocity and density pedestrian distributions. *PloS one*, 7(12):e50720, 2012.

[50] J. Zhang. *Pedestrian fundamental diagrams: Comparative analysis of experiments in different geometries*, volume 14. Forschungszentrum Jülich, 2012.

Paper D

Mean-field type games between two players driven by backward stochastic differential equations

Mean-field type games between two players driven by backward stochastic differential equations

by

Alexander Aurell

Published in
Games Volume 9, Number 4 (2018), 88

Abstract

In this paper, mean-field type games between two players with backward stochastic dynamics are defined and studied. They make up a class of non-zero-sum, non-cooperating, differential games where the players' state dynamics solve backward stochastic differential equations (BSDE) that depend on the marginal distributions of player states. Players try to minimize their individual cost functionals, also depending on the marginal state distributions. Under some regularity conditions, we derive necessary and sufficient conditions for existence of Nash equilibria. Player behavior is illustrated by numerical examples, and is compared to a centrally planned solution where the social cost, the sum of player costs, is minimized. The inefficiency of a Nash equilibrium, compared to socially optimal behavior, is quantified by the so-called price of anarchy. Numerical simulations of the price of anarchy indicate how the improvement in social cost achievable by a central planner depends on problem parameters.

Keywords: mean-field type game, non-zero-sum differential game, cooperative game, backward stochastic differential equations, linear-quadratic stochastic control, social cost, price of anarchy

D.1 Introduction

Mean-field type games (MFTG) is a class of games in which payoffs and dynamics depend not only on the state and control profiles of the players, but also on the distribution of the state and control processes. MFTGs has by now a plethora of applications in the engineering sciences, see [16] and the references therein. This paper studies MFTGs between two players, with state-distribution dependent cost functionals $J^i : \mathcal{U}^i \to \mathbb{R}$, $i = 1, 2$, and mean-field BSDE state dynamics. The Nash solution $(\hat{u}^1_\cdot, \hat{u}^2_\cdot) \in \mathcal{U}^1 \times \mathcal{U}^2$ is dictated by the pair of inequalities

$$\begin{cases} J^1(\hat{u}^1_\cdot; \hat{u}^2_\cdot) \leq J^1(u^1_\cdot; \hat{u}^2_\cdot), & \forall u^1_\cdot \in \mathcal{U}^1, \\ J^2(\hat{u}^2_\cdot; \hat{u}^1_\cdot) \leq J^2(u^2_\cdot; \hat{u}^1_\cdot), & \forall u^2_\cdot \in \mathcal{U}^2. \end{cases} \tag{1}$$

Following the path laid-out in [1], we establish a Pontryagin type maximum principle, yielding necessary and sufficient conditions for any pair of controls satisfying (1). Behavior in the equilibrium (1) is compared to the socially optimal solution, that minimizes the social cost $J := J^1 + J^2$.

D.1.1 Related Work

Pontryagin's maximum principle is the tool, alongside dynamic programming, to characterize optimal controls in both deterministic and stochastic settings. It can treat not only standard stochastic systems, but generalizes to optimal stopping, singular controls, risk-sensitive controls and partially observed models. Pontryagin's maximum principle yields necessary conditions that must be satisfied by any solution. The necessary conditions become sufficient under additional convexity conditions. Early results showed that an optimal control along with the corresponding optimal state trajectory must solve the so-called Hamiltonian system, which is a two-point (forward-backward) boundary value problem, together with a maximum condition on the so-called Hamiltonian function. A very useful aspect of this result is that minimization of the cost functional (over a set of control functions) may reduce to pointwise maximization of the Hamiltonian, at each point in time (over the set of control values). Pontryagin's technique for deterministic systems and stochastic systems with uncontrolled diffusion can be summarized as follows: assume that there exists an optimal control, make a spike-variation of it and then consider the first order term of the Taylor expansion with respect to the perturbation. This leads to variational inequality and the result follows from duality. If the diffusion is controlled, second order terms in the Taylor expansion have to be considered. In this case, one ends up with two forward-backward SDEs and a more involved maximum condition for the Hamiltonian. See [49] for a detailed account. For stochastic systems, the backward equation is fundamentally different from the forward equation, whenever one is looking for adapted solutions. An adapted solution to a BSDE is a pair of adapted stochastic processes $(Y_\cdot, Z_\cdot)$, where $Z_\cdot$ corrects "non-adaptiveness" caused by the terminal condition of $Y_\cdot$. As pointed out by [24], the first component $Y_\cdot$ corresponds to the mean evolution of the dynamics, and $Z_\cdot$ to the risk between current time and terminal time. The linear BSDE extends to non-linear BSDEs [39], with applications not only within stochastic optimal control but also in stochastic analysis [40] and finance [19, 21], and to forward-backward SDEs (FBSDE). A BSDE with distribution-dependent coefficients, the mean-field BSDE, is derived in [7] as the limit of a particle system. Existence and uniqueness results and a comparison theorem for mean-field BSDEs are provided in [8].

In stochastic differential games, both zero-sum and nonzero-sum, Pontryagin's stochastic maximum principle (SMP) and dynamic programming are the main tools for obtaining conditions for an equilibrium. These tools were essentially inherited from the theory of stochastic optimal control. As in the optimal control setting, the latter deals with solving systems of second-order parabolic partial differential equations, while the former is related to analyzing FBSDEs where, in the case of initial state constraints, the adjoint process is a BSDE. For a recent example of the use of the SMP in stochastic differential game theory, see [36].

The theory of mean-field type control (MFTC), initiated in [1], treats stochastic control problems with coefficients dependent on the marginal state-distribution. This theory is by now well developed for forward stochastic dynamics, i.e., with initial conditions on state [6, 15, 9, 12]. With SMPs for MFTC problems at hand, MFTG theory can inherit these techniques like stochastic differential game theory does in the mean-field free case. See [44] for a review of solution approaches to MFTGs. A MFTC problem can be interpreted as a large population limit of a cooperative game, where the players share a joint goal to optimize some objective [27]. A close relative to MFTC is the mean field game (MFG). MFG is a class of non-cooperative stochastic differential games where a large number of indistinguishable (anonymous) players interact weakly through a mean-field coupling term, initiated by

[22, 28] independently, and followed up by, among many others, [30, 45, 35]. Weak player-to-player interaction through a mean-field coupling restricts the influence one player has on any other player to be inversely proportional to the number of players, hence the level of influence of any specific player in a large game is very small. The coupling of player state dynamics leads to conflicting objectives, making the large game hard to analyze. The MFG equilibrium then often provides an approximate solution of the mass behavior in the pre-limit (finite population) game. In contrast to the MFG, players in a MFTG can be influential, and distinguishable (non-anonymous). That is, the state dynamics and the cost need not be of the same form over the whole player population, and a single player can have a major influence on other players' dynamics and cost.

Already in [41], an SMP in local form was derived for a controlled nonlinear BSDE. By first finding a global estimate for the variation of the second component of the BSDE solution, an SMP in global form was derived in [17]. A reinterpretation of BSDEs as forward stochastic optimal control problems [24] opened up for a new solution approach in the field of control of BSDEs. Inspired by the reinterpretation, optimal control of linear-quadric (LQ) BSDEs was solved in [32] by constructing a sequence of forward control problems with an additional state constraint, whose limit solution is the solution to the original LQ BSDE control problem. This approach was later used by [47] to solve a general FBSDE control problem, where the authors overcome the difficulty of controlling the diffusion in the forward process. Instead of writing down a second-order adjoint equation for the full system, the technique of [32] is used. Previous to that, [48] studied optimal control problem for general coupled forward–backward stochastic differential equations (FBSDEs) with controlled diffusions. A maximum principle of Pontryagin's type for the optimal control is derived, by means of spike variation techniques.

Optimal control of mean-field BSDEs has recently gained attention. In [31] the mean-field LQ BSDE control problem with deterministic coefficients is studied. Assuming the control space is linear, linear perturbation is used to derive a stationarity condition which together with a mean-field FBSDE system characterizes the optimal control. Existence of optimal controls is also proven under convexity assumptions. Other recent work on the control of BSDEs includes [43, 29], both using the FBSDE approach of [48].

D.1.2 Potential Applications of MFTG with Mean-Field BSDE Dynamics

In [3], a model is proposed for groups of pedestrians moving towards targets they are forced to reach, such as deliveries and emergency personnel. The strict terminal condition leads to the formulation of a dynamic model for crowd motion where the state dynamics is a mean-field BSDE. Mean-field effects appear in pedestrian crowd models as approximations of aggregate human interaction, so the game would in fact be a MFTG [2]. A game between such groups is of interest since it can be a tool for decentralized decision making under conflicting interests. Other areas of application include strategies for financial investments, where often future conditions are specified [14, 21] and lead to dynamic models including BSDEs. The already mentioned study [16] presents a lengthy list of applications of forward MFTGs in engineering sciences.

D.1.3 Paper Contribution and Outline

In this paper, control of mean-field BSDEs is extended to games between players whose state dynamics are mean-field BSDEs. Such games are in fact MFTGs, since the distribution of each player is effected by both players' choice of strategy. Our MFTG could be viewed as a game between mean-field FBSDEs, where the backward equation is the state equation, and the forward equation is pure noise. A Pontryagin's type SMP is derived, resulting in a verification theorem and conditions for existence of a Nash equilibrium. This solution approach is similar to that of [41, 17, 31]. The use of spike-perturbation requires minimal assumptions on the set of admissible controls, and differentiation of functions over the space of probability measure makes it possible to go beyond linear-quadratic mean-field cost and dynamics. The state BSDE is not converted to a forward optimization problem in the spirit of [32]. As a consequence, *the adjoint equation in our SMP is a forward SDE*. For the sake of comparison, optimality conditions for the cooperative situation are derived. In this setting, the players work together to optimize social cost, which is the the sum of player costs. The approach used is a straight-forward adaptation of the techniques used in control of SDEs of mean-field type; again, we do not need to take the route via some equivalent forward optimization problem to solve the backward MFTC problem. This cooperative game is a MFTC problem, and our result here is basically a special case of the FBSDE results in [48] or [47] mentioned above, although mean-field terms are present. Numerical simulations are done in the linear-quadratic case, which is explicitly solvable up to a system of ODEs. The examples pinpoint differences between player behavior in the game versus the centrally planned solution. The fraction between the social cost in the game equilibrium and the social cost optimum quantifies the game efficiency and was first studied in [25] for traffic coordination on networks under the name *coordination ratio*. This fraction was later renamed in [37] to *the price of anarchy*. We notice that paying a high price for using large control values, or deviating from a preferred initial position makes the problem stiffer, in the sense that the improvement by team optimality is decreasing, while paying a high price for mean-field related costs makes the problem less stiff.

The rest of this paper is organized as follows. The problem formulation is given in Section D.2. Sections D.3 and D.4 deal with necessary and sufficient conditions for any Nash equilibrium and social optimum; maximum principles for the MFTG and the MFTC are derived. An LQ problem is solved explicitly in Section D.5, and numerical results are presented. The paper concludes with some remarks on possible extensions in Section D.6, followed by an appendix containing proofs.

D.2 Problem formulation

List of Symbols

$T \in (0, \infty)$	the time horizon
$(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$	the underlying filtered probability space
$\mathcal{L}(X)$	the distribution of a random variable X under $\mathbb{P}$
$L^2_{\mathcal{F}_t}(\Omega; \mathbb{R}^d)$	the set of $\mathbb{R}^d$-valued $\mathcal{F}_t$-measurable random variables X such that $\mathbb{E}[\lvert X\rvert^2] < \infty$
$\mathcal{G}$	the progressive σ-algebra
$X.$	a stochastic process $\{X_t\}_{t \geq 0}$
$\mathbb{S}^{2,k}$	the set of $\mathbb{R}^k$-valued, continuous $\mathcal{G}$-measurable processes $X.$ such that $\mathbb{E}[\sup_{t\in[0,T]} \lvert X_t\rvert^2] < \infty$
$\mathbb{H}^{2,k}$	the set of $\mathbb{R}^k$-valued $\mathcal{G}$-measurable processes $X.$ such that $\mathbb{E}[\int_0^T \lvert X_s\rvert^2 ds] < \infty$

$\mathcal{U}^i$	the set of admissible controls for player i
$\mathcal{P}(\mathcal{X})$	the set of probability measures on $\mathcal{X}$
$\mathcal{P}_2(\mathcal{X})$	the elements of $\mathcal{P}(\mathcal{X})$ with finite second moment
Θ_t^i	the t-marginal of the state-, law- and control-tuple of player i
$\Vert Z\Vert_F$	the trace (Frobenius) norm of the matrix Z
$\partial_{y^i} f(y^i)$	derivative of the function f with domain $\mathbb{R}^d$
$\partial_{\mu^i} f(\mu^i)$	derivative of the function f with domain $\mathcal{P}_2(\mathbb{R}^d)$, see Appendix D.7 for details

Let $T > 0$ be a finite real number representing the time horizon of the game. Consider a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ on which two independent standard Brownian motions $W^1_\cdot$, $W^2_\cdot$ are defined, d_1- and d_2-dimensional respectively. Additionally, y^1_T, $y^2_T \in L^2_{\mathcal{F}_T}(\Omega;\mathbb{R}^d)$ and ξ, $\mathcal{F}_0$-measurable, are defined on the space. We assume that these five random objects are independent and that they generate the filtration $\mathbb{F} := \{\mathcal{F}_t\}_{t\geq 0}$. Notice that ξ makes $\mathcal{F}_0$ non-trivial. Let $\mathcal{G}$ be the σ-algebra on $[0,T]\times\Omega$ of $\mathcal{F}_t$-progressively measurable sets. For $k \geq 1$, let $\mathbb{S}^{2,k}$ be the set of $\mathbb{R}^k$-valued and continuous $\mathcal{G}$-measurable processes $X_\cdot := \{X_t : t \in [0,T]\}$ such that $\mathbb{E}[\sup_{t\in[0,T]}|X_t|^2] < \infty$, and let $\mathbb{H}^{2,k}$ be the set of $\mathbb{R}^k$-valued $\mathcal{G}$-measurable processes $X_\cdot$ such that $\mathbb{E}[\int_0^T |X_s|^2 ds] < \infty$.

Let (U^i, d_{U^i}) be a separable metric space, $i = 1, 2$. Player i picks her control $u^i_\cdot$ from the set

$$\mathcal{U}^i := \left\{ u : [0,T]\times\Omega \to U^i \mid u_\cdot\ \mathbb{F}\text{-adapted},\ \mathbb{E}\left[\int_0^T d_{U^i}(u_s)^2 ds\right] < \infty \right\}.$$

The distribution of any random variable $\xi \in \mathcal{X}$ will be denoted by $\mathcal{L}(\xi) \in \mathcal{P}(\mathcal{X})$, and $-i$ will denote the index $\{1,2\}\backslash i$. Given a pair of controls $(u^1_\cdot, u^2_\cdot) \in \mathcal{U}^1 \times \mathcal{U}^2$, consider the system of controlled BSDEs

$$dY^i_t = b^i(t, \Theta^i_t, \Theta^{-i}_t, Z_t)dt + Z^{i,1}_t dW^1_t + Z^{i,2}_t dW^2_t, \quad Y^i_T = y^i_T,\ i = 1, 2, \quad (2)$$

where $\Theta^i_t := (Y^i_t, \mathcal{L}(Y^i_t), u^i_t)$ and $Z_t := [Z^{1,1}_t\ Z^{1,2}_t\ Z^{2,1}_t\ Z^{2,2}_t]$. Furthermore,

$$b^i : \Omega \times [0,T] \times S \times U^i \times S \times U^{-i} \times \mathbb{R}^{d\times(2d_1+2d_2)} \to \mathbb{R}^d,$$

where $S := \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ is equipped with the norm $\Vert(y,\mu)\Vert_S := |y| + d_2(\mu)$, d_2 being the 2-Wasserstein metric on $\mathcal{P}(\mathbb{R}^d)$. $\mathbb{R}^{d\times(2d_1+2d_2)}$ is equipped with the trace norm $\|Z\|_F = \mathrm{tr}(ZZ^*)^{1/2}$. Note that if X is a square integrable random variable in $\mathbb{R}^d$, then $d_2(\mathcal{L}(X)) < \infty$ and $\mathcal{L}(X) \in \mathcal{P}_2(\mathbb{R}^d)$, the space of measures with finite d_2-norm.

Given $(u^1_\cdot, u^2_\cdot) \in \mathcal{U}^1 \times \mathcal{U}^2$, a pair of $\mathbb{R}^d \times \mathbb{R}^{d\times(d_1+d_2)}$-valued $\mathcal{G}$-measurable processes, $(Y^i_\cdot, [Z^{i,1}_\cdot Z^{i,2}_\cdot])$, $i = 1, 2$, is a solution to (2) if for all $t \in [0, T]$,

$$Y^i_t = y^i_T - \int_t^T b^i(s, \Theta^i_s, \Theta^{-i}_s, Z_s)ds - \sum_{j=1}^{2} \int_t^T Z^{i,j}_s dW^j_s, \quad \mathbb{P}\text{-a.s.},$$

and $(Y^i_\cdot, [Z^{i,1}_\cdot Z^{i,2}_\cdot]) \in \mathbb{S}^{2,d} \times \mathbb{H}^{2,d\times(d_1+d_2)}$.

Remark 1. *Any terminal condition $y^i_T \in L^2_{\mathcal{F}_T}(\Omega; \mathbb{R}^d)$ naturally induces an $\mathbb{F}$-martingale $Y^i_t := \mathbb{E}[y^i_T \mid \mathcal{F}_t]$. The martingale representation theorem then gives existence of a unique process $[Z^{i,1}_\cdot, Z^{i,2}_\cdot] \in \mathbb{H}^{2,d\times(d_1+d_2)}$ such that $Y^i_t = y^i_T + \int_t^T Z^{i,1}_s dW^1_s + \int_t^T Z^{i,2}_s dW^2_s$, i.e., $[Z^{i,1}_\cdot, Z^{i,2}_\cdot]$ plays the role of the projection and without it, $Y^i_\cdot$ would not be $\mathcal{G}$-measurable. Hence the noise $(W^1_\cdot, W^2_\cdot)$ generating the filtration is common to both players, and $[Z^{i,1}_\cdot, Z^{i,2}_\cdot]$, $i = 1, 2$ is their respective reaction to it. Player i may actually be effected by all the noise in the filtration even if only some components of $(W^1_\cdot, W^2_\cdot)$ appear in b^i. An interpretation of $[Z^{i,1}_\cdot, Z^{i,2}_\cdot]$ is that it is a second control of player i: first she plays $u^i_\cdot$ to heed preferences on energy use, initial position etc., then she picks $[Z^{i,1}_\cdot, Z^{i,2}_\cdot]$ so that her path to y^i_T is the optimal prediction based on available information in the filtration at any given time.*

Existence and uniqueness of (2) is given by a slight variation of the results of [8], where the one-dimensional case is treated. For the d-dimensional mean-field free case, see [38].

Assumption 1. *The process $b^i(\omega, \cdot, 0, \dots, 0)$, $i = 1, 2$, belongs to $\mathbb{H}^{2,d}$ and for any $v^i = (y^i, \mu^i, u^i, y^{-i}, \mu^{-i}, u^{-i}, z) \in S \times U^1 \times S \times U^2 \times \mathbb{R}^{d\times(2d_1+2d_2)}$, $b^i(\omega, \cdot, v^i)$, $i = 1, 2$, is $\mathcal{G}$-measurable.*

Assumption 2. *Given a pair of control values $(u^1, u^2) \in U^1 \times U^2$, there exists a constant $L > 0$ such that for all $t \in [0, T]$ and tuples $(y^1, \mu^1, y^2, \mu^2, z)$, $(\bar{y}^1, \bar{\mu}^1, \bar{y}^2, \bar{\mu}^2, \bar{z}) \in S \times S \times \mathbb{R}^{d(2d_1+2d_2)}$,*

$$\begin{aligned}&|b^i(t, y^i, \mu^i, u^i, y^{-i}, \mu^{-i}, u^{-i}, z) - b^i(t, \bar{y}^i, \bar{\mu}^i, u^i, \bar{y}^{-i}, \bar{\mu}^{-i}, u^{-i}, \bar{z})| \\ &\qquad \le L\left(\sum_{j=1}^{2} \|(y^j, \mu^j) - (\bar{y}^j, \bar{\mu}^j)\|_S + \|z - \bar{z}\|_F\right), \quad \mathbb{P}\text{-}a.s.,\ i = 1, 2.\end{aligned}$$

Theorem 2. *Let Assumptions 1 and 2 hold. Then, for any terminal conditions* $y_T^1, y_T^2 \in L^2(\Omega, \mathcal{F}_T, \mathbb{P}; \mathbb{R}^d)$ *and* $(u^1_\cdot, u^2_\cdot) \in \mathcal{U}^1 \times \mathcal{U}^2$*, the system of mean-field BSDEs* (2) *has a unique solution* $(Y^i_\cdot, [Z^{i,1}_\cdot, Z^{i,2}_\cdot]) \in \mathbb{S}^{2,d} \times \mathbb{H}^{2,d\times(d_1+d_2)}$, $i = 1, 2$.

Next, we introduce the best reply of player i as follows:

$$J^i(u^i_\cdot; u^{-i}_\cdot) := \mathbb{E}\left[\int_0^T f^i(t, \Theta^i_t, \Theta^{-i}_t)dt + h^i(Y^i_0, \mathcal{L}(Y^i_0), Y^{-i}_0, \mathcal{L}(Y^{-i}_0))\right]$$

for given maps $f^i : [0,T] \times S \times U^i \times S \times U^{-i} \to \mathbb{R}$ and $h^i : \Omega \times S \times S \to \mathbb{R}$.

Assumption 3. *For any pair of controls* $(u^1_\cdot, u^2_\cdot) \in \mathcal{U}^1 \times \mathcal{U}^2$*,* $f^i(\cdot, \Theta^i_\cdot, \Theta^{-i}_\cdot) \in L^1_{\mathcal{F}}(0,T;\mathbb{R})$ *and* $h(Y^i_0, \mathcal{L}(Y^i_0), Y^{-i}_0, \mathcal{L}(Y^{-i}_0)) \in L^1_{\mathcal{F}_0}(\Omega;\mathbb{R})$.

The problems we consider in the following sections are

1. The Mean-field Type Game (MFTG): find the Nash equilibrium controls of
$$\begin{cases} \inf\limits_{u^i_\cdot \in \mathcal{U}^i} J^i(u^i_\cdot; u^{-i}_\cdot), \\ \text{s.t. } dY^i_t = b^i(t, \Theta^i_t, \Theta^{-i}_t, Z_t)dt + Z^{i,1}dW^1_t + Z^{i,2}_t dW^2_t, \\ \qquad Y^i_T = y^i_T, \quad i = 1, 2, \end{cases} \tag{3}$$

2. The Mean-field Type Control Problem (MFTC): find the optimal control pair of
$$\begin{cases} \inf\limits_{(u^1_\cdot, u^2_\cdot) \in \mathcal{U}^1 \times \mathcal{U}^2} J(u^1_\cdot, u^2_\cdot) := J^1(u^1_\cdot; u^2_\cdot) + J^2(u^2_\cdot; u^1_\cdot), \\ \text{s.t.} \qquad dY^i_t = b^i(t, \Theta^i_t, \Theta^{-i}_t, Z_t)dt + Z^{i,1}dW^1_t + Z^{i,2}_t dW^2_t, \\ \qquad\qquad Y^i_T = y^i_T, \qquad i = 1, 2. \end{cases} \tag{4}$$

In the game each player assumes that the other player acts rationally, i.e., minimizes cost, and picks her control as the best response to that. This leads to a set of two inequalities, characterizing any control pair $(u^1_\cdot, u^2_\cdot)$ that constitutes a Nash equilibrium. In this paper, each player is aware of the other player's control set, best response function and state dynamics. Therefore, even though the decision process is decentralized, both players solve the same set of inequalities. When there is not a unique Nash equilibrium, there is an ambiguity around which equilibrium strategy to play if the players do not communicate. In the control problem, a central planner decides what strategies are played by both of the players. The central planner might just be the two players cooperating towards a common goal, or some superior decision maker. The goal is to find the control pair that minimizes the *social cost* J.

The notion of a centrally planned/cooperative solution is related to the concept of *team optimality* in *team problems* [4]. In a team problem, the players share a common objective. A team-optimal solution is then the solution to the joint minimization of the common objective. In our case, the social cost J is a common objective in the MFTC. The Nash solution to the team problem is given by the control pair that satisfies the two inequalities

$$\begin{cases} J(\hat{u}^1_\cdot, \hat{u}^2_\cdot) \leq J(u^1_\cdot, \hat{u}^2_\cdot), & \forall u^1_\cdot \in \mathcal{U}^1, \\ J(\hat{u}^1_\cdot, \hat{u}^2_\cdot) \leq J(\hat{u}^1_\cdot, u^2_\cdot), & \forall u^2_\cdot \in \mathcal{U}^2. \end{cases} \tag{5}$$

In (5), each player is minimizing the social cost with respect to its marginal, under the assumption that the other player is minimizing its marginal. This is the so-called *player-by-player optimality* of a control pair in a team problem. Notice that if we set $J^1(u^1_\cdot; u^2_\cdot) = J^2(u^2_\cdot; u^1_\cdot)$ in (3), it becomes a team problem. The solution to the MFTG (3) will then be the player-by-player optimal solution to the minimization of the social cost.

Logically, we expect the optimal social cost to be lower than the social cost in a Nash equilibrium. The ratio between the worst case social cost in the game and the optimal social cost is called the price of anarchy, and we will highlight it in the numerical simulations in Section D.5 where we also observe behavioral differences between MFTG and MFTC given identical data.

D.3 Problem 1: MFTG

This section is the derivation of necessary and sufficient equilibrium conditions of (3). Given the existence of such a pair of controls, we derive the conditions by the means of a Pontryagin type stochastic maximum principle.

Assume that $(\hat{u}^1_\cdot, \hat{u}^2_\cdot)$ is a Nash equilibrium for the MFTG, i.e., satisfies the following system of inequalities,

$$\begin{cases} J^1(\hat{u}^1_\cdot; \hat{u}^2_\cdot) \leq J^1(u^1_\cdot; \hat{u}^2_\cdot), & u^1_\cdot \in \mathcal{U}^1, \\ J^2(\hat{u}^2_\cdot; \hat{u}^1_\cdot) \leq J^2(u^2_\cdot; \hat{u}^1_\cdot), & u^2_\cdot \in \mathcal{U}^2. \end{cases}$$

Consider the first inequality, with $\bar{u}^{\varepsilon,1}$ chosen as a spike-perturbation of $\hat{u}^1$. That is, for $u_\cdot \in \mathcal{U}^1$,

$$\bar{u}^{\varepsilon,1}_t := \begin{cases} \hat{u}^1_t, & t \in [0,T]\backslash E_\varepsilon, \\ u_t, & t \in E_\varepsilon. \end{cases}$$

Here, E_ε is any subset of $[0,T]$ of Lebesgue measure ε. Clearly, $\bar{u}^{\varepsilon,1}_\cdot \in \mathcal{U}^1$. When player 1 plays the spike-perturbed control $\bar{u}^{\varepsilon,1}_\cdot$ and player 2 plays the equilibrium control $\hat{u}^2_\cdot$, we denote the dynamics by

$$\begin{cases} d\bar{Y}^{\varepsilon,1}_t = b^1(t, \bar{\Theta}^{\varepsilon,1}_t, \bar{Y}^{\varepsilon,2}_t, \mathcal{L}(\bar{Y}^{\varepsilon,2}_t), \hat{u}^2_t, \bar{Z}^\varepsilon_t)dt + \bar{Z}^{\varepsilon,1,1}_t dW^1_t + \bar{Z}^{\varepsilon,1,2}_t dW^2_t, \\ \bar{Y}^1_T = y^1, \\ d\bar{Y}^{\varepsilon,2}_t = b^2(t, \bar{Y}^{\varepsilon,2}_t, \mathcal{L}(\bar{Y}^{\varepsilon,2}_t), \hat{u}^2_t, \bar{\Theta}^{\varepsilon,1}_t, \bar{Z}^\varepsilon_t)dt + \bar{Z}^{\varepsilon,2,1}_t dW^1_t + \bar{Z}^{\varepsilon,2,2}_t dW^2_t, \\ \bar{Y}^2_T = y^2. \end{cases} \tag{6}$$

The performance of the perturbed dynamics (6) will be compared with that of the equilibrium dynamics

$$\begin{cases} d\hat{Y}^1_t = b^1(t, \hat{\Theta}^1_t, \hat{\Theta}^2_t, \hat{Z}_t)dt + \hat{Z}^{1,1}_t dW^1_t + \hat{Z}^{1,2}_t dW^2_t,\ \hat{Y}^1_T = y^1, \\ d\hat{Y}^2_t = b^2(t, \hat{\Theta}^2_t, \hat{\Theta}^1_t, \hat{Z}_t)dt + \hat{Z}^{2,1}_t dW^1_t + \hat{Z}^{2,2}_t dW^2_t,\ \hat{Y}^2_T = y^2. \end{cases}$$

For simplicity, we write for $\varphi \in \{b^1, f^1, h^1\}$, $\psi \in \{b^2, f^2, h^2\}$, and $\vartheta \in \{b^i, f^i, h^i, i = 1, 2\}$,

$$\begin{aligned} \bar{\varphi}^\varepsilon_t &:= \varphi(t, \bar{\Theta}^{\varepsilon,1}_t, \bar{Y}^{\varepsilon,2}_t, \mathcal{L}(\bar{Y}^{\varepsilon,2}_t), \hat{u}^2_t, \bar{Z}^\varepsilon_t), \\ \bar{\psi}^\varepsilon_t &:= \psi(t, \bar{Y}^{\varepsilon,2}_t, \mathcal{L}(\bar{Y}^{\varepsilon,2}_t), \hat{u}^2_t, \bar{\Theta}^{\varepsilon,1}_t, \bar{Z}^\varepsilon_t), \\ \hat{\vartheta}_t &:= \vartheta(t, \hat{\Theta}^i_t, \hat{\Theta}^{-i}_t, \hat{Z}_t). \end{aligned}$$

In this shorthand notation, which will be used from now on, the difference in performance is

$$J^1(\bar{u}^{\varepsilon,1}_\cdot; \hat{u}^2_\cdot) - J^1(\hat{u}^1_\cdot; \hat{u}^2_\cdot) = \mathbb{E}\left[\int_0^T \bar{f}^{\varepsilon,1}_t - \hat{f}^1_t dt + \bar{h}^{\varepsilon,1}_0 - \hat{h}^1_0\right].$$

Any derivative of $f : a \to f(a)$ will be denoted $\partial_a f$, indifferent of the space the function is mapping from/to.

Assumption 4. *The functions*

$$\begin{aligned}(y^1,\mu^1,u^1,y^2,\mu^2,u^2,z) &\mapsto b^i(t,y^i,\mu^i,u^i,y^{-i},\mu^{-i},u^{-i},z)\\ (y^1,\mu^1,u^1,y^2,\mu^2,u^2) &\mapsto f^i(t,y^i,\mu^i,u^i,y^{-i},\mu^{-i},u^{-i})\\ (y^1,\mu^1,y^2,\mu^2) &\mapsto h^i(y^i,\mu^i,y^{-i},\mu^{-i})\end{aligned}$$

are for all $t \in [0,T]$ *a.s. differentiable at* $(\hat{Y}_0^1, \mathcal{L}(\hat{Y}_0^1), \hat{Y}_0^2, \mathcal{L}(\hat{Y}_0^2))$, $(\hat{\Theta}_t^1, \hat{\Theta}_t^2, \hat{Z}_t)$, *and* $(\hat{\Theta}_t^1, \hat{\Theta}_t^2)$ *respectively. Furthermore,*

$$\partial_{y^j}\hat{b}_t^i,\ \partial_{\mu^j}\hat{b}_t^i,\ \partial_{y^j}\hat{f}_t^i,\ \partial_{\mu^j}\hat{f}_t^i, \quad i,j=1,2,$$

are for all t *a.s. uniformly bounded, and*

$$\partial_{y^j}\hat{h}_0^i + \mathbb{E}\left[{}^*(\partial_{\mu^j}\hat{h}_0^i)\right] \in L^2_{\mathcal{F}_0}(\Omega;\mathbb{R}^d).$$

For $i=1,2$,

$$\begin{aligned}\bar{h}_0^{\varepsilon,i} - \hat{h}_0^i = &\sum_{j=1}^2 \left\{ \partial_{y^j}\hat{h}_0^i(\bar{Y}_0^{\varepsilon,j} - \hat{Y}_0^j) + \mathbb{E}\left[(\partial_{\mu^j}\hat{h}_0^i)^*(\bar{Y}_0^{\varepsilon,j} - \hat{Y}_0^j)\right]\right\}\\ &+ \sum_{j=1}^2 \left\{ o\left(|\bar{Y}_0^{\varepsilon,j} - \hat{Y}_0^j|\right) + o\left(\mathbb{E}[|\bar{Y}_0^{\varepsilon,j} - \hat{Y}_0^j|^2]^{1/2}\right)\right\}.\end{aligned} \tag{7}$$

A brief overview on differentiation of functions from $\mathcal{P}_2(\mathbb{R}^d)$ to $\mathbb{R}$ is found in Appendix D.7 which also defines the notation $(\partial_{\mu^j}\hat{h}_0^i)^*$ in (29). Both $\bar{Y}_t^{\varepsilon,1} - \hat{Y}_t^1$ and $\bar{Y}_t^{\varepsilon,2} - \hat{Y}_t^2$ appear in (7), this suggests that we need to introduce two first order variation processes. That is, we want $(\widetilde{Y}_\cdot^i, [\widetilde{Z}_\cdot^{i,1}, \widetilde{Z}_\cdot^{i,2}])$, $i=1,2$, that for some $C>0$ satisfies

$$\begin{aligned}&\sup_{0\le t\le T} \mathbb{E}\left[|\widetilde{Y}_t^i|^2 + \sum_{j=1}^2 \int_0^t \|\widetilde{Z}_s^{i,j}\|_F^2\, ds\right] \le C\varepsilon^2,\\ &\sup_{0\le t\le T} \mathbb{E}\left[|\bar{Y}_t^{\varepsilon,i} - \hat{Y}_t^i - \widetilde{Y}_t^i|^2 + \sum_{j=1}^2 \int_0^t \|\bar{Z}_s^{\varepsilon,i,j} - \hat{Z}_s^{i,j} - \widetilde{Z}_s^{i,j}\|_F^2\, ds\right] \le C\varepsilon^2.\end{aligned} \tag{8}$$

Let δ_i denote variation in $u_\cdot^i$ so that for $\vartheta \in \{f^i, b^i, i=1,2\}$,

$$\delta_i\vartheta(t) := \vartheta(t, \hat{Y}_t^i, \mathcal{L}(\hat{Y}_t^i), \bar{u}_t^{\varepsilon,i}, \hat{\Theta}_t^{-i}, \hat{Z}_t) - \hat{\vartheta}_t.$$

Assumption 5. *For $y^i, \mu^i \in S$, $i = 1, 2$, $z \in \mathbb{R}^{d\times(2d_1+2d_2)}$ and (u^1, u^2), $(v^1, v^2) \in U^1 \times U^2$, there exists a constant $L > 0$ such that*

$$
\begin{aligned}
&|b^i(t, y^i, \mu^i, u^i, y^{-i}, \mu^{-i}, u^{-i}, z) - b^i(t, y^i, \mu^i, v^i, y^{-i}, \mu^{-i}, v^{-i}, z)| \\
&\qquad \leq L \sum_{j=1}^{2} d_{U^j}(u^j, v^j),
\end{aligned}
$$

a.s. for all $t \in [0, T]$.

Lemma 3. *Let Assumption 1, 2, 4, and 5 be in force. The first order variation processes that satisfy* (8) *is given by the following system of BSDEs,*

$$
\begin{cases}
d\widetilde{Y}^i_t = \left(\displaystyle\sum_{j=1}^{2} \left\{ \partial_{y^j} \hat{b}^i_t \widetilde{Y}^j_t + \mathbb{E}\left[(\partial_{\mu^j} \hat{b}^i_t)^* \widetilde{Y}^j_t \right] \right\} \right. \\
\qquad\qquad \left. + \displaystyle\sum_{j,k=1}^{2} \partial_{z^{j,k}} \hat{b}^i_t \widetilde{Z}^{j,k}_t + \delta_1 b^i(t) 1_{E_\varepsilon}(t) \right) dt + \displaystyle\sum_{j=1}^{2} \widetilde{Z}^{i,j}_t dW^j_t, \\
\widetilde{Y}^i_T = 0, \qquad i = 1, 2.
\end{cases}
$$

A proof of the lemma is found in the appendices. By Lemma 3,

$$
\begin{aligned}
\mathbb{E}\left[\bar{h}^{\varepsilon,1}_0 - \hat{h}^1_0 \right] &= \mathbb{E}\left[\sum_{j=1}^{2} \partial_{y^j} \hat{h}^1_0 \widetilde{Y}^j_0 + \mathbb{E}\left[(\partial_{\mu^j} \hat{h}^1_0)^* \widetilde{Y}^j_0 \right] \right] + o(\varepsilon) \\
&= \mathbb{E}\left[\sum_{j=1}^{2} p^{1,j}_0 \widetilde{Y}^j_0 \right] + o(\varepsilon),
\end{aligned}
$$

where the costates $p^{1,j}_\cdot$, $j = 1, 2$, satisfy $p^{1,j}_0 := \partial_{y^j} \hat{h}^1_0 + \mathbb{E}\left[{}^*(\partial_{\mu^j} \hat{h}^1_0) \right]$. The notation ${}^*(\partial_{\mu^j} \hat{h}^1_0)$ is defined in (30) in the appendices. Assumption 4 grants us existence and uniqueness of solutions to (9) below.

Lemma 4 (Duality relation). *Let Assumption 1, 2, and 4 hold. Let $p^{1,j}_\cdot$ be the solution to the SDE*

$$
\begin{cases}
dp^{1,j}_t = -\left\{ \partial_{y^j} \hat{H}^1_t + \mathbb{E}\left[{}^*(\partial_{\mu^j} \hat{H}^1_t) \right] \right\} dt - \displaystyle\sum_{k=1}^{2} \partial_{z^{j,k}} \hat{H}^1_t dW^k_t, \\
p^{1,j}_0 = \partial_{y^j} \hat{h}^1_0 + \mathbb{E}\left[{}^*(\partial_{\mu^j} \hat{h}^1_0) \right],
\end{cases}
\tag{9}
$$

where for $(y^i, \mu^i) \in S$, $i = 1, 2$, *and* $(u^1, u^2, z) \in U^1 \times U^2 \times \mathbb{R}^{d\times(2d_1+2d_2)}$,

$$\begin{aligned} &H^1(\omega, t, y^1, \mu^1, u^1, y^2, \mu^2, u^2, z, p_t^{1,1}, p_t^{1,2}) \\ &:= \sum_{j=1}^{2} b^j(\omega, t, y^j, \mu^j, u^j, y^{-j}, \mu^{-j}, u^{-j}, z) p_t^{1,j} \\ &\qquad - f^1(t, y^1, \mu^1, u^1, y^2, \mu^2, u^2). \end{aligned} \tag{10}$$

Then the following duality relation holds,

$$\mathbb{E}\left[\sum_{j=1}^2 p_0^{1,j}\widetilde{Y}_0^j\right] = -\mathbb{E}\left[\int_0^T \sum_{j=1}^2 p_t^{1,j}\delta_1 b^j(t) 1_{E_\varepsilon}(t) + \widetilde{Y}_t^j \left(\partial_{y^j}\hat{f}_t^1 + \mathbb{E}\left[{}^*(\partial_{\mu^j}\hat{f}_t^1)\right]\right) dt\right].$$

The proof of the duality relation has been passed on to the appendices. We have that

$$\begin{aligned} \bar{f}_t^{\varepsilon,i} - \hat{f}_t^i &= \sum_{j=1}^{2}\left\{\partial_{y^j}\hat{f}_t^i(\bar{Y}_t^{\varepsilon,j} - \hat{Y}_t^j) + \mathbb{E}\left[(\partial_{\mu^j}\hat{f}_t^i)^*(\bar{Y}_t^{\varepsilon,j} - \hat{Y}_t^j)\right]\right\} \\ &+ \delta_1 f^i(t) 1_{E_\varepsilon}(t) + \sum_{j=1}^{2}\left\{o\left(|\bar{Y}_t^{\varepsilon,j} - \hat{Y}_t^j|\right) + o\left(\mathbb{E}[|\bar{Y}_t^{\varepsilon,j} - \hat{Y}_t^j|^2]^{1/2}\right)\right\}. \end{aligned} \tag{11}$$

By the expansion (11) and Lemma 3,

$$\begin{aligned} &\mathbb{E}\left[\int_0^T \bar{f}_t^{\varepsilon,1} - \hat{f}_t^1 dt\right] \\ &= \mathbb{E}\left[\int_0^T \sum_{j=1}^{2} \widetilde{Y}_t^j \left(\partial_{y^j}\hat{f}_t^1 + \mathbb{E}\left[{}^*(\partial_{\mu^j}\hat{f}_t^1)\right]\right) + \delta_1 f^1(t) 1_{E_\varepsilon}(t) dt\right] + o(\varepsilon), \end{aligned}$$

which yields

$$\begin{aligned} &J^1(\bar{u}_\cdot^{\varepsilon,1}; \hat{u}_\cdot^2) - J^1(\hat{u}_\cdot^1; \hat{u}_\cdot^2) \\ &= \mathbb{E}\left[\int_0^T \left\{-p_t^{1,1}\delta_1 b^1(t) - p_t^{1,2}\delta_1 b^2(t) + \delta_1 f^1(t)\right\} 1_{E_\varepsilon}(t) dt\right] + o(\varepsilon). \end{aligned}$$

Therefore

$$J^1(\bar{u}_\cdot^{\varepsilon,1}; \hat{u}_\cdot^2) - J^1(\hat{u}_\cdot^1; \hat{u}_\cdot^2) = -\mathbb{E}\left[\int_0^T \delta_1 H^1(t) 1_{E_\varepsilon}(t) dt\right] + o(\varepsilon). \tag{12}$$

From the last identity, we can derive necessary and sufficient conditions for player 1's best response to $\hat{u}^2_\cdot$.

The same argument can be carried out for players 2's best response to $\hat{u}^1_\cdot$. Naturally, we need to impose the corresponding assumptions on player 2's control. For completeness and later reference, we state now the second player's version of Lemma 4.

Lemma 5. *(Duality relation, player 2) Let Assumptions 1, 2, and 4 hold, and let $p^{2,j}_\cdot$ be the solution to the SDE*

$$\begin{cases} dp^{2,j}_t = -\left\{\partial_{y^j}\hat{H}^2_t + \mathbb{E}\left[{}^*(\partial_{\mu^j}\hat{H}^2_t)\right]\right\}dt - \sum_{k=1}^{2}\partial_{z^{j,k}}\hat{H}^2_t dW^k_t, \\ p^{2,j}_0 = \partial_{y^j}\hat{h}^2_0 + \mathbb{E}\left[{}^*(\partial_{\mu^j}\hat{h}^2_0)\right], \end{cases} \tag{13}$$

where for $(y^i,\mu^i) \in S$, $i=1,2$, and $(u^1,u^2,z) \in U^1 \times U^2 \times \mathbb{R}^{d\times(2d_1+2d_2)}$,

$$\begin{aligned} &H^2(\omega,t,y^2,\mu^2,u^2,y^1,\mu^1,u^1,z,p^{2,1}_t,p^{2,2}_t) \\ &:= \sum_{j=1}^{2} b^j(\omega,t,y^j,\mu^j,u^j,y^{-j},\mu^{-j},u^{-j},z)p^{2,j}_t \\ &\quad - f^2(t,y^2,\mu^2,u^2,y^1,\mu^1,u^1). \end{aligned} \tag{14}$$

Then the following duality relation holds,

$$\begin{aligned} &\mathbb{E}\left[\sum_{j=1}^{2} p^{2,j}_0 \widetilde{Y}^j_0\right] \\ &= -\mathbb{E}\left[\int_0^T \sum_{j=1}^{2} p^{2,j}_t \delta_2 b^j(t) 1_{E_\varepsilon}(t) + \widetilde{Y}^j_t\left(\partial_{y^j}\hat{f}^2_t + \mathbb{E}\left[{}^*(\partial_{y^j}\hat{f}^2_t)\right]\right)dt\right]. \end{aligned}$$

Necessary conditions for an equilibrium can be stated as a system of six equations, the two state BSDEs and the four costate (adjoint) SDEs. Sufficient conditions for a Nash equilibrium can now be stated as convexity conditions on the four functions $H^i, h^i, i = 1,2$. We let Assumption 1–5 be in place.

Theorem 6 (Necessary equilibrium conditions). *Suppose that $(\hat{u}^1_\cdot, \hat{u}^2_\cdot)$ is an equilibrium control for the MFTG and that $p^{i,j}_\cdot$, $i, j = 1, 2$, solve* (9) *and* (13). *Then, for $i = 1, 2$,*

$$\hat{u}^i_t = \arg\max_{\alpha \in U^i} H^i(t, \hat{Y}^i_t, \mathcal{L}(\hat{Y}^i_t), \alpha, \hat{\Theta}^{-i}_t, \hat{Z}_t, p^{i,1}_t, p^{i,2}_t), \quad a.e.\text{-}t,\ \mathbb{P}\text{-}a.s. \tag{15}$$

Proof. Let $E_\varepsilon := [s, s+\varepsilon]$, $u_\cdot \in \mathcal{U}^1$ and $A \in \mathcal{F}_t$ for $t \in E_\epsilon$. Consider the spike-perturbation

$$u^\varepsilon_t := \begin{cases} u_t 1_A + \hat{u}^1_t 1_{A^c}, & t \in E_\varepsilon, \\ \hat{u}^1_t, & t \in [0,T] \backslash E_\varepsilon. \end{cases}$$

Then

$$\begin{aligned} &\hat{H}^1_t - H^1(t, \hat{Y}^1_t, \mathcal{L}(\hat{Y}^1_t), u^\varepsilon_t, \hat{\Theta}^2_t, \hat{Z}_t, p^{1,1}_t, p^{1,2}_t) = \\ &\qquad \left(\hat{H}^1_t - H^1(t, \hat{Y}^1_t, \mathcal{L}(\hat{Y}^1_t), u_t, \hat{\Theta}^2_t, \hat{Z}_t, p^{1,1}_t, p^{1,2}_t)\right) 1_A 1_{E_\varepsilon}(t). \end{aligned}$$

Applying (12), we obtain

$$\frac{1}{\varepsilon}\mathbb{E}\left[\int_s^{s+\varepsilon} \left(\hat{H}^1_t - H^1(t, \hat{Y}^1_t, \mathcal{L}(\hat{Y}^1_t), u_t, \hat{\Theta}^2_t, \hat{Z}_t, p^{1,1}_t, p^{1,2}_t)\right) 1_A dt\right] \geq \frac{1}{\varepsilon} o(\varepsilon).$$

Sending ε to zero yields for a.e. $s \in [0,T]$

$$\mathbb{E}\left[\left(\hat{H}^1_s - H^1(s, \hat{Y}^1_s, \mathcal{L}(\hat{Y}^1_s), u_s, \hat{\Theta}^2_s, \hat{Z}_s, p^{1,1}_s, p^{1,2}_s)\right) 1_A\right] \geq 0.$$

The last inequality holds for all $A \in \mathcal{F}_s$, thus for a.e. $s \in [0,T]$, it holds $\mathbb{P}$-a.s. that

$$\mathbb{E}\left[\left(\hat{H}^1_s - H^1(s, \hat{Y}^1_s, \mathcal{L}(\hat{Y}^1_s), u_s, \hat{\Theta}^1_s, \hat{Z}_s, p^{1,1}_s, p^{1,2}_s)\right) \mid \mathcal{F}_s\right] \geq 0. \tag{16}$$

By measurability of the integrand in (16),

$$\hat{u}^1_t = \arg\max_{\alpha \in U^1} H^1(t, \hat{Y}^1_t, \mathcal{L}(\hat{Y}^1_t), \alpha, \hat{\Theta}^2_t, \hat{Z}_t, p^{1,1}_t, p^{1,2}_t), \quad a.e.\ t \in [0,T],\ \mathbb{P}\text{-a.s.}$$

The same argument yields

$$\hat{u}^2_t = \arg\max_{\alpha \in U^2} H^2(t, \hat{Y}^2_t, \mathcal{L}(\hat{Y}^2_t), \alpha, \hat{\Theta}^1_t, \hat{Z}_t, p^{2,1}_t, p^{2,2}_t), \quad a.e.\ t \in [0,T],\ \mathbb{P}\text{-a.s.}$$

□

Theorem 7 (Sufficient equilibrium conditions)**.** *Suppose that* $\hat{u}^1_\cdot$ *and* $\hat{u}^2_\cdot$ *satisfy* (15)*. Suppose furthermore that for* $(t, p^{i,1}, p^{i,2}, z) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d\times(2d_1+2d_2)}$*,* $i = 1, 2$*,*

$$(y^1, \mu^1, u^1, y^2, \mu^2, u^2) \mapsto H^i(t, y^i, \mu^i, u^i, y^{-i}, \mu^{-i}, u^{-i}, z, p^{i,1}, p^{i,2})$$

is concave $\mathbb{P}$*-a.s. and*

$$(y^1, \mu^1, y^2, \mu^2) \mapsto h^i(y^i, \mu^i, y^{-i}, \mu^{-i})$$

is convex $\mathbb{P}$*-a.s. Then* $\hat{u}^1_\cdot, \hat{u}^2_\cdot$ *constitute an equilibrium control and the pair* $(\hat{Y}^i_\cdot, [\hat{Z}^{i,1}_\cdot, \hat{Z}^{i,2}_\cdot], \hat{u}^i_\cdot)$*,* $i = 1, 2$*, is an equilibrium for the MFTG.*

Proof. By assumption, $\delta_i H^i(t) \leq 0$ for any spike variation, almost surely for a.e. t. Applying the convexity and concavity assumptions in the expansion steps results in the inequality

$$0 \leq -\mathbb{E}\left[\int_0^T \delta_i H^i(t) 1_{E_\varepsilon}(t)dt\right] \leq J^i(u^i_\cdot; \hat{u}^{-i}_\cdot) - J^i(\hat{u}^i_\cdot; \hat{u}^{-i}_\cdot).$$

□

D.4 Problem 2: MFTC

Carrying out a similar argument to that of the previous section, we find necessary optimality conditions for problem (4). Also, we readily get a verification theorem. The pair $(\hat{u}^1_\cdot, \hat{u}^2_\cdot) \in \mathcal{U}^1 \times \mathcal{U}^2$ is optimal if

$$J(\hat{u}^1_\cdot, \hat{u}^2_\cdot) \leq J(u^1_\cdot, u^2_\cdot), \quad (u^1_\cdot, u^2_\cdot) \in \mathcal{U}^1 \times \mathcal{U}^2. \tag{17}$$

Assume from now on that $(\hat{u}^1_\cdot, \hat{u}^2_\cdot)$ is an optimal control. We study the inequality (17) when $(\check{u}^{\varepsilon,1}_\cdot, \check{u}^{\varepsilon,2}_\cdot)$ is a spike-perturbation of $(\hat{u}^1_\cdot, \hat{u}^2_\cdot)$,

$$(\check{u}^{\varepsilon,1}_t, \check{u}^{\varepsilon,2}_t) := \begin{cases} (\hat{u}^1_t, \hat{u}^2_t), & t \in [0,T] \backslash E_\varepsilon, \\ (u^1_t, u^2_t), & t \in E_\varepsilon, \end{cases}$$

where E_ε is any subset of $[0,T]$ of Lebesgue measure ε and $(u^1_\cdot, u^2_\cdot) \in \mathcal{U}^1 \times \mathcal{U}^2$. When the players use the perturbed control, we denote the state dynamics by

$$\begin{cases} d\check{Y}^{\varepsilon,1}_t = b^1(t, \check{\Theta}^{\varepsilon,1}_t, \check{\Theta}^{\varepsilon,2}_t, \check{Z}^\varepsilon_t)dt + \check{Z}^{\varepsilon,1,1}_t dW^1_t + \check{Z}^{\varepsilon,1,2}_t dW^2_t, \ \check{Y}^1_T = y^1, \\ d\check{Y}^{\varepsilon,2}_t = b^2(t, \check{\Theta}^{\varepsilon,2}_t, \check{\Theta}^{\varepsilon,1}_t, \check{Z}^\varepsilon_t)dt + \check{Z}^{\varepsilon,2,1}_t dW^1_t + \check{Z}^{\varepsilon,2,2}_t dW^2_t, \ \check{Y}^2_T = y^2, \end{cases}$$

and we will compare their performance to that of the optimally controlled state dynamics

$$\begin{cases} d\hat{Y}_t^1 = b^1(t, \hat{\Theta}_t^1, \hat{\Theta}_t^2, \hat{Z}_t)dt + \hat{Z}_t^{1,1}dW_t^1 + \hat{Z}_t^{1,2}dW_t^2,\ \hat{Y}_T^1 = y^1, \\ d\hat{Y}_t^2 = b^2(t, \hat{\Theta}_t^2, \hat{\Theta}_t^1, \hat{Z}_t)dt + \hat{Z}_t^{2,1}dW_t^1 + \hat{Z}_t^{2,2}dW_t^2,\ \hat{Y}_T^2 = y^2. \end{cases}$$

To simplify notation we write

$$\begin{aligned} \check{\vartheta}_t^{\varepsilon} &:= \vartheta(t, \check{\Theta}_t^{\varepsilon,i}, \check{\Theta}_t^{\varepsilon,-i}, \check{Z}_t^{\varepsilon}), \\ \hat{\vartheta}_t &:= \vartheta(t, \hat{\Theta}_t^{i}, \hat{\Theta}_t^{-i}, \hat{Z}_t), \end{aligned}$$

for $\vartheta \in \{b^i, f^i, h^i, i = 1, 2\}$ and in this notation

$$\begin{aligned} &J(\check{u}_\cdot^{\varepsilon,1}, \check{u}_\cdot^{\varepsilon,2}) - J(\hat{u}_\cdot^1, \hat{u}_\cdot^2) \\ &= \mathbb{E}\left[\int_0^T \check{f}_t^{\varepsilon,1} + \check{f}_t^{\varepsilon,2} - \hat{f}_t^1 - \hat{f}_t^2 dt + \check{h}_0^{\varepsilon,1} + \check{h}_0^{\varepsilon,2} - \hat{h}_0^1 - \hat{h}_0^2\right] \\ &= \mathbb{E}\left[\int_0^T \check{f}_t^{\varepsilon} - \hat{f}_t dt + \check{h}_0^{\varepsilon} - \hat{h}_0\right] \end{aligned}$$

where $f_t := f_t^1 + f_t^2$ and $h_t := h_t^1 + h_t^2$. Again, we want to find first order variation processes $(\widetilde{Y}_\cdot^i, [\widetilde{Z}_\cdot^{i,1}, \widetilde{Z}_\cdot^{i,2}])$, $i = 1, 2$, that satisfy (8) with $(\bar{Y}_\cdot^{\varepsilon,i}, [\bar{Z}_\cdot^{\varepsilon,i,1}, \bar{Z}_\cdot^{\varepsilon,i,2}])$ replaced by its 'checked' counterpart $(\check{Y}_\cdot^{\varepsilon,i}, [\check{Z}_\cdot^{\varepsilon,i,1}, \check{Z}_\cdot^{\varepsilon,i,2}])$.

Assumption 6. *The functions*

$$\begin{aligned} (y^1, \mu^1, u^1, y^2, \mu^2, u^2, z) &\mapsto b^i(t, y^i, \mu^i, u^i, y^{-i}, \mu^{-i}, u^{-i}, z) \\ (y^1, \mu^1, u^1, y^2, \mu^2, u^2) &\mapsto f^i(t, y^i, \mu^i, u^i, y^{-i}, \mu^{-i}, u^{-i}) \\ (y^1, \mu^1, y^2, \mu^2) &\mapsto h^i(y^i, \mu^i, y^{-i}, \mu^{-i}) \end{aligned}$$

are for all t a.s. differentiable at $(\hat{Y}_0^1, \mathcal{L}(\hat{Y}_0^1), \hat{Y}_0^2, \mathcal{L}(\hat{Y}_0^2))$, $(\hat{\Theta}_t^1, \hat{\Theta}_t^2, \hat{Z}_t)$, and $(\hat{\Theta}_t^1, \hat{\Theta}_t^2)$ respectively. Furthermore,

$$\partial_{y^j}\hat{b}_t^i,\ \partial_{\mu^j}\hat{b}_t^i,\ \partial_{y^j}\hat{f}_t^i,\ \partial_{\mu^j}\hat{f}_t^i, \quad i, j = 1, 2,$$

are for all t a.s. uniformly bounded and $\partial_{y^j}\hat{h}_0^i + \mathbb{E}\left[{}^*(\partial_\mu^j \hat{h}_0^i)\right] \in L^2_{\mathcal{F}_0}(\Omega; \mathbb{R}^d)$.

Notice that the point of differentiability is generally not the same in Assumption 4 and 6. Above, $(\hat{u}_\cdot^1, \hat{u}_\cdot^2)$ is an optimal control while in Assumption 4, it is an equilibrium control. Let δ denote simultaneous variation in controls, for $\vartheta \in \{f^i, b^i, i = 1, 2\}$,

$$\delta\vartheta(t) := \delta_1\vartheta(t) + \delta_2\vartheta(t).$$

Lemma 8. *Let Assumption 1, 2, 5, and 6 be in force. The first order variation processes that satisfy the 'checked' version of* (8) *are given by the following system of BSDEs,*

$$\begin{cases} d\widetilde{Y}_t^i = \left(\sum_{j=1}^2 \left\{ \partial_{y^j} \hat{b}^i \widetilde{Y}_t^j + \mathbb{E}\left[(\partial_{\mu^j} \hat{b}^i)^* \widetilde{Y}_t^j \right] \right\} \right. \\ \qquad\qquad \left. + \delta b^i(t) 1_{E_\varepsilon}(t) + \sum_{j,k=1}^2 \partial_{z^{j,k}} \hat{b}^i \widetilde{Z}_t^{j,k} \right) dt + \sum_{j=1}^2 \widetilde{Z}_t^{i,j} dW_t^j, \\ \widetilde{Y}_T^i = 0, \quad i = 1, 2. \end{cases}$$

The proof follows the same steps as the proof of Lemma 3. By Lemma 8,

$$\mathbb{E}\left[\check{h}_0^\varepsilon - \hat{h}_0 \right] = \mathbb{E}\left[\sum_{j=1}^2 p_0^j \widetilde{Y}_0^j \right] + o(\varepsilon),$$

where $p_0^j := \partial_{y^j} \hat{h}_0 + \mathbb{E}\left[{}^*(\partial_{\mu^j} \hat{h}_0) \right]$.

Lemma 9 (Duality relation)**.** *Let Assumption 1, 2, and 6 hold. Let $p_\cdot^j$ be the solution to the SDE*

$$\begin{cases} dp_t^j = -\left\{ \partial_{y^j} \hat{H}_t + \mathbb{E}\left[{}^*(\partial_{\mu^j} \hat{H}_t) \right] \right\} dt - \sum_{k=1}^2 \partial_{z^{j,k}} \hat{H}_t dW_t^k, \\ p_0^j = \partial_{y^j} \hat{h}_0 + \mathbb{E}\left[{}^*(\partial_{\mu^j} \hat{h}_0) \right], \end{cases} \tag{18}$$

where for $(y^i, \mu^i) \in S$, $i = 1, 2$, and $(u^1, u^2, z) \in U^1 \times U^2 \times \mathbb{R}^{d \times (2d_1 + 2d_2)}$,

$$\begin{aligned} &H(\omega, t, y^1, \mu^1, u^1, y^2, \mu^2, u^2, z, p_t^1, p_t^2) \\ &:= \sum_{j=1}^2 b^j(\omega, t, y^j, \mu^j, u^j, y^{-j}, \mu^{-j}, u^{-j}, z) p_t^j \\ &\qquad - \sum_{j=1}^2 f^j(t, y^j, \mu^j, u^j, y^{-j}, \mu^{-j}, u^{-j}). \end{aligned}$$

Then the following duality relation holds,

$$\begin{aligned}&\mathbb{E}\left[\sum_{j=1}^{2} p_0^j \widetilde{Y}_0^j\right]\\&= -\mathbb{E}\left[\int_0^T \sum_{j=1}^{2} p_t^j \delta b^j(t) 1_{E_\varepsilon}(t) + \widetilde{Y}_t^j \left(\partial_{y^j} \hat{f}_t + \mathbb{E}\left[{}^*(\partial_{\mu^j} \hat{f}_t)\right]\right) dt\right].\end{aligned}$$

The proof of Lemma 9 and Lemma 4 are almost identical. By Lemma 8,

$$\begin{aligned}&J(\check{u}^{\varepsilon,1}_\cdot, \check{u}^{\varepsilon,2}_\cdot) - J(\hat{u}^1_\cdot, \hat{u}^2_\cdot)\\&\quad= \mathbb{E}\left[\int_0^T \left\{-p_t^1 \delta b^1(t) - p_t^2 \delta b^2(t) + \delta f(t)\right\} 1_{E_\varepsilon}(t) dt\right] + o(\varepsilon).\end{aligned}$$

Thus

$$J(\check{u}^{\varepsilon,1}_\cdot, \check{u}^{\varepsilon,2}_\cdot) - J(\hat{u}^1_\cdot, \hat{u}^2_\cdot) = -\mathbb{E}\left[\int_0^T \delta H(t) 1_{E_\varepsilon}(t) dt\right] + o(\varepsilon).$$

In the following two theorems, Assumptions 1–3 and 5–6 are in force.

Theorem 10 (Necessary optimality conditions)**.** *Suppose that $(\hat{u}^1_\cdot, \hat{u}^2_\cdot)$ is an optimal control for the MFTC and that $p^i_\cdot$, $i = 1, 2$, solves* (18)*. Then, for $i = 1, 2$,*

$$(\hat{u}^1_t, \hat{u}^2_t) = \underset{(v,w)\in U^1\times U^2}{\arg\max}\, H(t, \hat{Y}^1_t, \mathcal{L}(\hat{Y}^1_t), v, \hat{Y}^2_t, \mathcal{L}(\hat{Y}^2_t), w, \hat{Z}_t, p^1_t, p^2_t) \tag{19}$$

for almost every t, $\mathbb{P}$-a.s.

Theorem 11 (Sufficient optimality conditions)**.** *Suppose $(\hat{u}^1_\cdot, \hat{u}^2_\cdot)$ satisfy* (19) *and that, for $(t, p^1, p^2, z) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d\times(2d_1+2d_2)}$,*

$$(y^1, \mu^1, u^1, y^2, \mu^2, u^2) \mapsto H(t, y^1, \mu^1, u^1, y^2, \mu^2 s, u^2, z, p^1, p^2)$$

is concave $\mathbb{P}$-a.s. and

$$(y^1, \mu^1, y^2, \mu^2) \mapsto h(y^1, \mu^1, y^2, \mu^2)$$

is convex $\mathbb{P}$-a.s. Then $(\hat{u}^1_\cdot, \hat{u}^2_\cdot)$ is an optimal control and $(\hat{Y}^i_\cdot, [\hat{Z}^{i,1}_\cdot, \hat{Z}^{i,2}_\cdot], \hat{u}^i_\cdot)$, $i = 1, 2$ solves the MFTC.

D.5 Example: the linear-quadratic case

In this section we consider a linear-quadratic version of (3) and (4), in the one-dimensional case. Let $a_i, c_{i,j}, q_{i,j}, \bar{q}_{i,j}, \widetilde{q}_{i,j}, \bar{s}_{i,j}, s_i, \bar{s}_i^E, r_i : [0,T] \mapsto \mathbb{R}$, $i,j = 1,2$ be deterministic coefficient functions, uniformly bounded over $[0,T]$. Additionally, $r_i(t) \geq \epsilon > 0$ for $i = 1,2$. Define

$$
\begin{aligned}
b^i(t, \Theta_t^i, \Theta_t^{-i}, Z_t) &= a_i(t)u_t^i + \sum_{j=1}^{2} c_{i,j}(t) W_t^j, \\
f^i(t, \Theta_t^i, \Theta_t^{-i}) &= \sum_{j=1}^{2} \Bigg\{ \frac{1}{2} q_{i,j}(t)(Y_t^j)^2 + \frac{1}{2}\bar{q}_{i,j}(t)\mathbb{E}[Y_t^j]^2 \\
&\quad + \widetilde{q}_{i,j}(t) Y_t^j \mathbb{E}[Y_t^j] + \bar{s}_{i,j}(t)\mathbb{E}[Y_t^j] Y_t^{-j} \Bigg\} \\
&\quad + s_i(t) Y_t^i Y_t^{-i} + \bar{s}_i^E(t)\mathbb{E}[Y_t^i]\mathbb{E}[Y_t^{-i}] + \frac{1}{2} r_i(t)(u_t^i)^2.
\end{aligned} \tag{20}
$$

The uniform boundedness of the coefficients implies Assumption 1–6, given initial costs h^1, h^2, satisfying Assumption 4 and 6. Assumption 3 (integrability of f^i) follows by classical BSDE estimates [50]. The Hamiltonian functions of the two player game, defined in (10) and (14), are in this example given by

$$
\begin{aligned}
&H^i(t, \Theta_t^i, \Theta_t^{-i}, Z_t) \\
&= \left(a_1(t)u_t^1 + \sum_{j=1}^{2} c_{1,j} W_t^j \right) p^{i,1} + \left(a_2(t)u_t^2 + \sum_{j=1}^{2} c_{2,j} W_t^j \right) p^{i,2} \\
&\quad - \sum_{j=1}^{2} \Bigg\{ \frac{1}{2} q_{i,j}(t)(Y_t^j)^2 + \frac{1}{2}\bar{q}_{i,j}(t)\mathbb{E}[Y_t^j]^2 \\
&\quad + \widetilde{q}_{i,j}(t) Y_t^j \mathbb{E}[Y_t^j] + \bar{s}_{i,j}(t)\mathbb{E}[Y_t^j] Y_t^{-j} \Bigg\} \\
&\quad - s_i(t) Y_t^i Y_t^{-i} - \bar{s}_i^E(t)\mathbb{E}[Y_t^i]\mathbb{E}[Y_t^{-i}] - \frac{1}{2} r_i(t)(u_t^i)^2.
\end{aligned}
$$

The Hessian of $(y^1,\dots,u^2)\mapsto H^1(t,y^1,\dots,u^2,z,p^{1,1},p^{1,2})$ is

$$\mathcal{H}^1(t) := -\begin{bmatrix} q_{1,1}(t) & \widetilde{q}_{1,1}(t) & 0 & s_1(t) & \bar{s}_{1,2}(t) & 0 \\ \widetilde{q}_{1,1}(t) & \bar{q}_{1,1}(t) & 0 & \bar{s}_{1,1}(t) & \bar{s}_1^E(t) & 0 \\ 0 & 0 & r_1(t) & 0 & 0 & 0 \\ s_1(t) & \bar{s}_{1,1}(t) & 0 & q_{1,2}(t) & \widetilde{q}_{1,2}(t) & 0 \\ \bar{s}_{1,2}(t) & \bar{s}_1^E(t) & 0 & \widetilde{q}_{1,2}(t) & \bar{q}_{1,2}(t) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and the Hessian of $(y^1,\dots,u^2)\mapsto H^2(t,y^1,\dots,u^2,z,p^{2,1},p^{2,2})$ is

$$\mathcal{H}^2(t) := -\begin{bmatrix} q_{2,1}(t) & \widetilde{q}_{2,1}(t) & 0 & s_2(t) & \bar{s}_{2,2}(t) & 0 \\ \widetilde{q}_{2,1}(t) & \bar{q}_{2,1}(t) & 0 & \bar{s}_{2,1}(t) & \bar{s}_2^E(t) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ s_2(t) & \bar{s}_{2,1}(t) & 0 & q_{2,2}(t) & \widetilde{q}_{2,2}(t) & 0 \\ \bar{s}_{2,2}(t) & \bar{s}_2^E(t) & 0 & \widetilde{q}_{2,2}(t) & \bar{q}_{2,2}(t) & 0 \\ 0 & 0 & 0 & 0 & 0 & r_2(t) \end{bmatrix}.$$

The coefficients are further assumed to be such that $\mathcal{H}^1(t)$ and $\mathcal{H}^2(t)$ are negative semi-definite for all $t\in[0,T]$. Also, we assume that $(y^1,\dots,\mu^2)\mapsto h^i(y^1,\dots,\mu^2)$, yet unspecified, is convex. Theorem 7 yields

$$\hat{u}^i_t = a_i(t)r_i^{-1}(t)p^{i,i}_t,$$

where $p^{i,i}_\cdot$ solves (9) if $i=1$ and (13) if $i=2$. In fact, the equilibrium is unique in this case, since $\hat{u}^i_\cdot$ is the unique pointwise solution to (15) and $p^{i,i}_\cdot$ is unique, see (35)-(36). By Theorem 11,

$$\hat{u}^i_t = a_i(t)r_i^{-1}(t)p^i_t,$$

where $p^i_\cdot$ solves (18), is an optimal control for the linear-quadratic MFTC and it is unique.

D.5.1 MFTG

The equilibrium dynamics are

$$d\hat{Y}^i_t = \left(a_i^2(t)r_i^{-1}(t)p^{i,i}_t + \sum_{j=1}^2 c_{i,j}W^j_t\right)dt + \hat{Z}^{i,1}_t dW^1_t + \hat{Z}^{i,2}_t dW^2_t,\ \hat{Y}^i_T = y^i_T.$$

We see that only two costate processes, $p_\cdot^{1,1}$ and $p_\cdot^{2,2}$, are relevant here. This is a consequence of the lack of explicit dependence on u^{-i} in the b^i and f^i specified in (20). Nevertheless, the running cost f^i depends implicitly on u^{-i} through player $-i$'s state and expected state. We make the following ansatz: there exists deterministic functions $\alpha_i, \bar{\alpha}_i, \beta_i, \bar{\beta}_i, \theta_i : [0,T] \to \mathbb{R}$, $i = 1,2$ and $\gamma_{i,j} : [0,T] \to \mathbb{R}$, $i,j = 1,2$, such that

$$\begin{aligned}\hat{Y}_t^i &= \alpha_i(t) p_t^{i,i} + \bar{\alpha}_i(t)\mathbb{E}[p_t^{i,i}] + \beta_i(t) p_t^{-i,-i} + \bar{\beta}_i(t)\mathbb{E}[p_t^{-i,-i}] \\ &+ \quad \gamma_{i,1}(t) W_t^1 + \gamma_{i,2}(t) W_t^2 + \theta_i(t).\end{aligned} \tag{21}$$

Clearly, we need to impose the terminal conditions

$$\alpha_i(T) = 0, \quad \bar{\alpha}_i(T) = 0, \quad \beta_i(T) = 0, \quad \bar{\beta}_i(T) = 0, \quad \gamma_{i,j}(T) = 0, \quad \theta_i(T) = y_T^i.$$

Calculations presented in the appendix identifies coefficients and yields the following system of ODEs determining $\alpha_i(\cdot), \ldots, \theta_i(\cdot)$,

$$\begin{cases} \dot{\alpha}_i(t) + \alpha_i(t) P^i(t) + \beta_i(t) R^{-i}(t) = a_i^2(t) r_i^{-1}(t), \\ \dot{\bar{\alpha}}_i(t) + \alpha_i(t)\bar{P}_i(t) + \bar{\alpha}_i(t)(P^i(t) + \bar{P}^i(t)) + \\ \qquad \beta_i(t)\bar{R}^{-i}(t) + \bar{\beta}_i(t)(R^{-i}(t) + \bar{R}^{-i}(t)) = 0, \\ \dot{\beta}_i(t) + \alpha_i(t) R^i(t) + \beta_i(t) P^{-i}(t) = 0, \\ \dot{\bar{\beta}}_i(t) + \alpha_i(t)\bar{R}^i(t) + \bar{\alpha}_i(t)(R^i(t) + \bar{R}^i(t)) + \\ \qquad \beta_i(t)\bar{P}^{-i}(t) + \bar{\beta}_i(t)(P^{-i}(t) + \bar{P}^{-i}(t)) = 0, \\ \dot{\gamma}_{i,1}(t) + \alpha_i(t)\Phi^i(t) + \beta_i(t)\Phi^{-i}(t) = c_{i,1}(t), \\ \dot{\gamma}_{i,2}(t) + \alpha_i(t)\Psi^i(t) + \beta_i(t)\Psi^{-i}(t) = c_{i,2}(t), \\ \dot{\theta}_i(t) + \theta_i(t)\Big((\alpha_i(t) + \bar{\alpha}_i(t))(Q_i(t) + \bar{Q}_i(t)) + \\ \qquad (\beta_i(t) + \bar{\beta}_i(t))(S_{-i}(t) + \bar{S}_{-i}(t))\Big) + \\ \qquad \theta_{-i}(t)\Big((\alpha_i(t) + \bar{\alpha}_i(t))(S_i(t) + \bar{S}_i(t)) + \\ \qquad (\beta_i(t) + \bar{\beta}_i(t))(Q_{-i}(t) + \bar{Q}_{-i}(t))\Big) = 0, \end{cases} \tag{22}$$

where

$$
\begin{aligned}
P^i(t) &:= Q_i(t)\alpha_i(t) + S_i(t)\beta_{-i}(t), \;\; R^i(t) := Q_i(t)\beta_i(t) + S_i(t)\alpha_{-i}(t),\\
\bar{P}^i(t) &:= Q_i(t)\bar{\alpha}_i(t) + \bar{Q}_i(t)(\alpha_i(t) + \bar{\alpha}_i(t))\\
&\quad + S_i(t)\bar{\beta}_{-i}(t) + \bar{S}_i(t)(\beta_{-i}(t) + \bar{\beta}_{-i}(t)),\\
\bar{R}^i(t) &:= Q_i(t)\bar{\beta}_i(t) + \bar{Q}_i(t)(\beta_i(t) + \bar{\beta}_i(t)) + S_i(t)\bar{\alpha}_{-i}(t)\\
&\quad + \bar{S}_i(t)(\alpha_{-i}(t) + \bar{\alpha}_{-i}(t)),\\
\Phi^i(t) &:= (Q_i(t)\gamma_{i,1}(t) + S_i(t)\gamma_{-i,1}(t)),\\
\Psi^i(t) &:= (Q_i(t)\gamma_{i,2}(t) + S_i(t)\gamma_{-i,2}(t)),\\
Q_i(t) &:= q_{i,i}(t) + \widetilde{q}_{i,i}(t), \qquad \bar{Q}_i(t) := \widetilde{q}_{i,i}(t) + \bar{q}_{i,i}(t),\\
S_i(t) &:= s_i(t) + \bar{s}_{i,i}(t), \qquad \bar{S}_i(t) := \bar{s}_{i,-i}(t) + \bar{s}_i^E(t).
\end{aligned}
\tag{23}
$$

Now (21)–(23) gives us the equilibrium dynamics. In this fashion, it is possible to solve LQ problems more general than (20).

D.5.2 MFTC

The optimally controlled dynamics are

$$
d\hat{Y}_t^i = \left(a_i^2(t) r_i^{-1}(t) p_t^i + \sum_{j=1}^{2} c_{i,j} W_t^j \right) dt + \hat{Z}_t^{i,1} dW_t^1 + \hat{Z}_t^{i,2} dW_t^2, \; \hat{Y}_T^i = y_T^i.
$$

We make almost the same ansatz as before. Assume that there exists deterministic functions $\alpha_i, \bar{\alpha}_i, \beta_i, \bar{\beta}_i, \theta_i : [0,T] \to \mathbb{R}$, $i = 1,2$ and $\gamma_{i,j} : [0,T] \to \mathbb{R}$, $i,j = 1,2$, with terminal conditions

$$
\alpha_i(T) = 0, \; \bar{\alpha}_i(T) = 0, \; \beta_i(T) = 0, \; \bar{\beta}_i(T) = 0, \; \gamma_{i,j}(T) = 0, \; \theta_i(T) = y_T^i \tag{24}
$$

such that

$$
\begin{aligned}
\hat{Y}_t^i = \alpha_i(t) p_t^i + \bar{\alpha}_i(t)\mathbb{E}[p_t^i] + \beta_i(t) p_t^{-i} + \bar{\beta}_i(t)\mathbb{E}[p_t^{-i}]\\
+ \gamma_{i,1}(t) W_t^1 + \gamma_{i,2}(t) W_t^2 + \theta_i(t).
\end{aligned}
\tag{25}
$$

By redefining $Q_i, \bar{Q}_i, S_i, \bar{S}_i$ in (23),

$$
\begin{aligned}
Q_i(t) &:= q_{1,i}(t) + q_{2,i}(t) + \widetilde{q}_{1,i}(t) + \widetilde{q}_{2,i}(t),\\
\bar{Q}_i(t) &:= \widetilde{q}_{1,i}(t) + \widetilde{q}_{2,i}(t) + \bar{q}_{1,i}(t) + \bar{q}_{2,i}(t),\\
S_i(t) &:= s_1(t) + s_2(t) + \bar{s}_{1,i}(t) + \bar{s}_{2,i}(t),\\
\bar{S}_i(t) &:= \bar{s}_{1,-i}(t) + \bar{s}_{2,-i}(t) + \bar{s}_1^E(t) + \bar{s}_2^E(t),
\end{aligned}
$$

(22), (23) and (24), (25) gives us the optimally controlled state dynamics.

D.5.3 Simulation and the Price of Anarchy

Let $T := 1$, $\xi := (y_0^1, y_0^2) \in L^2_{\mathcal{F}_0}(\Omega; \mathbb{R}^d \times \mathbb{R}^d)$ be preferred initial positions for player 1 and 2 respectively, and

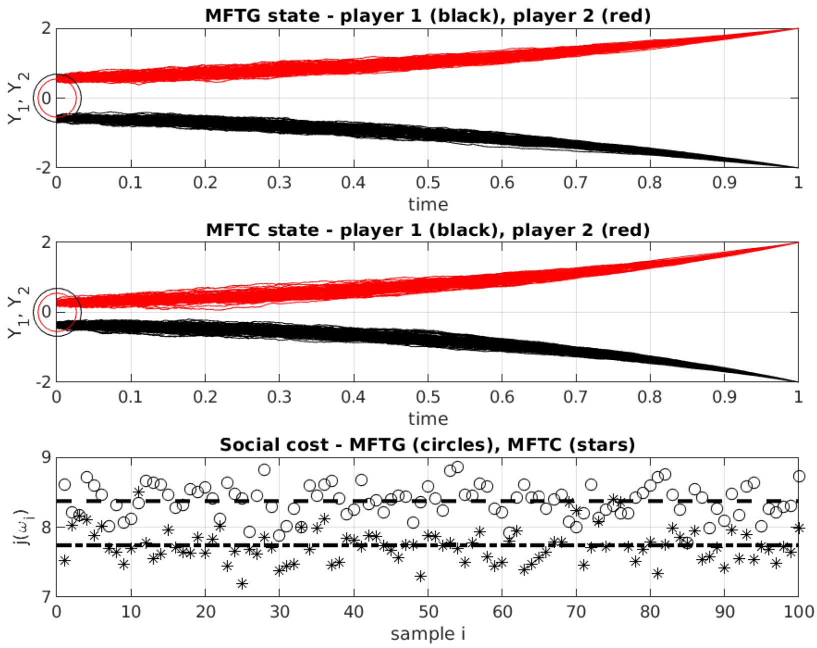
$$f_t^i := \frac{1}{2}\Big(r_i(u_t^i)^2 + \rho_i(Y_t^i - \mathbb{E}[Y_t^{-i}])^2\Big), \quad h_t^i := \frac{\nu_i}{2}(Y_0^i - y_0^i)^2.$$

In this setup, $\mathcal{H}^1$ and $\mathcal{H}^2$ are negative semi-definite if $r_i, \rho_i > 0$, h^i is convex if $\nu_i > 0$. In Figure 1 numerical simulations of MFTG and MFTC are presented. In (a), the two players have identical preferences, but different terminal conditions. The situation is symmetric in the sense that we expect the realized paths of player 1 reflected through the line $y = 0$ to be approximately paths of player 2. In (c), preferences are asymmetric and as a consequence, the realized paths are not each other's mirrored images.

The central planner in a MFTC uses more information than a single player does. In fact, in our example, $\gamma^{i,j}(t) = 0$ when $i \neq j$ in the MFTG. The interpretation is that in the game, player i does not care about player $-i$'s noise, only her mean state. For the central planner however, $\gamma^{i,j}$ is not identically zero for $i \neq j$. This can be observed in (b), where the central planner makes the player states evolve under some common noise.

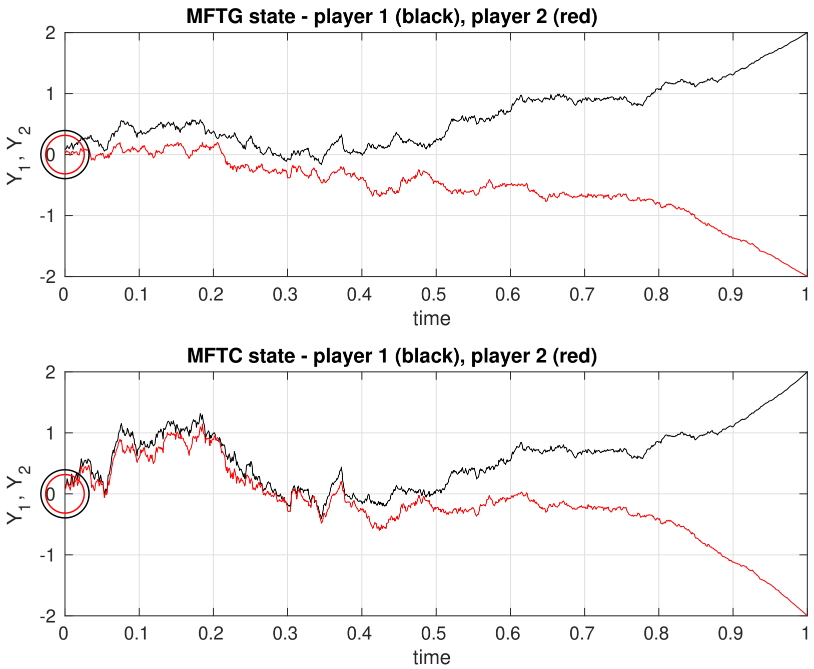
In (c) we see an interesting contrast between the MFTG and the MFTC. Player 1 (black) feels no attraction to player 2 ($\rho_1 = 0$) while player 2 is attracted to the mean position of player 1 ($\rho_2 > 0$). In the game, player 1 travels on the straight line from $(t, y) \approx (0, -1)$ to its terminal position $(t, y) = (1, -2)$. Player 2, on the other hand, deviates far from its preferred initial position at time $t = 0$, only to be in the proximity of player 1. In the MFTC, the central planner makes player 1 linger around $y = 0$ for some time, before turning south towards the terminal position. The result is less movement movement by player 2. Even though player 1 pays a higher individual cost, the social cost is reduced by approximately 33%. The *social cost* J is approximated by

$$J(u^1, u^2) \approx \frac{1}{N}\sum_{i=1}^{N} j(\omega_i),$$



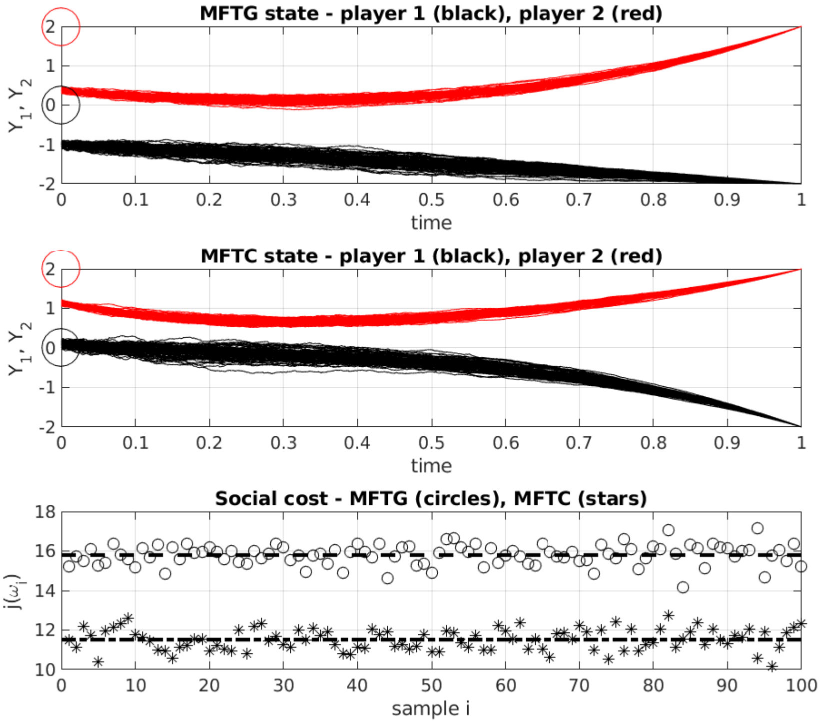
(a)

y_T^1	a_1	c_{11}	c_{12}	r_1	ρ_1	ν_1	y_0^1
-2	1	0.3	0	1	1	1	$\mathcal{N}(0, 0.1)$
y_T^2	a_2	c_{21}	c_{22}	r_2	ρ_2	ν_2	y_0^2
2	1	0	0.3	1	1	1	$\mathcal{N}(0, 0.1)$



(**b**)

$\boldsymbol{y_T^1}$	$\boldsymbol{a_1}$	$\boldsymbol{c_{11}}$	$\boldsymbol{c_{12}}$	$\boldsymbol{r_1}$	$\boldsymbol{\rho_1}$	$\boldsymbol{\nu_1}$	$\boldsymbol{y_0^1}$
-2	1	3	0	1	10	1	$\mathcal{N}(0, 0.1)$
$\boldsymbol{y_T^2}$	$\boldsymbol{a_2}$	$\boldsymbol{c_{21}}$	$\boldsymbol{c_{22}}$	$\boldsymbol{r_2}$	$\boldsymbol{\rho_2}$	$\boldsymbol{\nu_2}$	$\boldsymbol{y_0^2}$
2	1	0	3	1	10	1	$\mathcal{N}(0, 0.1)$



(**c**)

y_T^1	a_1	c_{11}	c_{12}	r_1	ρ_1	ν_1	y_0^1
-2	1	0.3	0	1	4	1	$\mathcal{N}(0, 0.1)$
y_T^2	a_2	c_{21}	c_{22}	r_2	ρ_2	ν_2	y_0^2
2	1	0	0.3	1	0	1	$\mathcal{N}(2, 0.1)$

Figure 1: Numerical examples: (**a**) symmetric preference, (**b**) single path sample, (**c**) asymmetric attraction and initial position. Circles indicate the preferred initial positions.

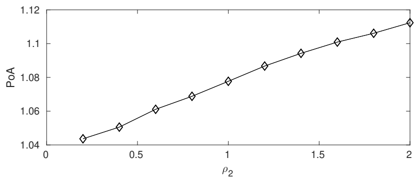
where $j(\omega_i) := \sum_{j=1}^{2} \int_0^T f_t^j(\omega_i)dt + h^j(\omega_i)$. In (a) and (c), the outcomes of j (circles for equilibrium control, stars for optimal control) are presented along with the approximation of J (dashed lines) for $N = 100$. The optimal control yields the lower social cost in both cases. This is expected, the general inefficiency of a Nash equilibrium in nonzero-sum games is well known [18]. The price of anarchy quantifies the inefficiency due to non-cooperation, see for static games [25, 26], for differential games [5] and for linear-quadratic mean-field type games [20]. The price of anarchy in mean-field games has been studied recently in [13, 11]. It is defined as the largest ratio between social cost for an equilibrium (MFTG) to the optimal social cost (MFTC),

$$PoA := \sup_{\substack{(\hat{u}^1_\cdot, \hat{u}^2_\cdot) \text{ MFTG} \\ \text{equilibrium}}} \frac{J(\hat{u}^1_\cdot, \hat{u}^2_\cdot)}{\min\limits_{u^i_\cdot \in \mathcal{U}^i, i=1,2} J(u^1_\cdot, u^2_\cdot)}.$$

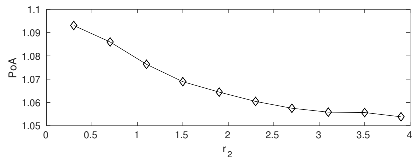
Taking the parameter set of (a) as a point of reference, see Table 2, we vary one parameter at the time and study PoA. The result is presented in Figure 2. In the intervals studied, PoA is increasing in ρ_i and T and decreasing in ν_i and r_i. The reason is that the players become less flexible when ν_i and/or r_i are increased, and the improvement a central planner can do decreases. On the other hand, an increased time horizon gives the central planner more time to improve the social cost. Also, an increased preference on attraction rewards the unegoistic behavior in the MFTC model.

Table 2: Parameter values in the symmetric case (a).

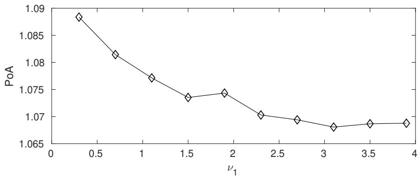
y_T^1	a_1	c_{11}	c_{12}	r_1	ρ_1	ν_1	y_0^1
-2	1	0.3	0	1	1	1	$\mathcal{N}(0, 0.1)$
y_T^2	a_2	c_{21}	c_{22}	r_2	ρ_2	ν_2	y_0^2
2	1	0	0.3	1	1	1	$\mathcal{N}(0, 0.1)$



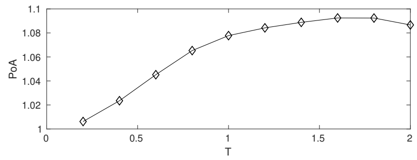
(**a**) Variation of ρ_2 in $[0.2, 2]$.



(**b**) Variation of r_2 in $[0.2, 4]$.



(**c**) Variation of ν_1 in $[0.2, 4]$.



(**d**) Variation of T in $[0.2, 2]$.

Figure 2: Numerical approximations ($N = 5000$) of the price of anarchy PoA.

D.6 Conclusions and discussion

Mean-field type games with backward stochastic dynamics, where the coefficients are allowed to depend on the marginal distributions of the player states, have been defined in this paper. Under regularity assumptions necessary conditions for a Nash equilibrium have been derived in the form of a stochastic maximum principle. Additional convexity assumptions yield sufficient conditions. In linear-quadric examples, player behavior in the MFTG is compared to the centrally planned solution in the MFTC. The efficiency of the MFTG Nash equilibrium, quantified by the price of anarchy, and its dependence on problem parameters is studied.

The framework presented in this paper has many possible extensions, towards both theory and applications. The theory for martingale-driven BSDEs is now standard, and one could exchange $W^1_\cdot, W^2_\cdot$ throughout this paper for two martingales $M^1_\cdot, M^2_\cdot$, possibly jump processes, and approach the game with the theory of forward-backward SDEs. Indeed, the topic of games between mean-field FBSDEs seems yet unexplored. These kind of problems would have immediate applications in finance.

With our definition of $\mathcal{U}^i$, we have restricted ourselves to open loop adapted controls in this paper. Other information structures, such as perfect/partial state- and/or law feedback controls, lagged or noise-perturbed controls are possible. Furthermore, both players have perfect information about each other in this paper. Taking inspiration from for example [34, 33], the access to information could be restricted, so that the players have only partial information on states/laws. These types of extensions are interesting both from the theoretical and applied point of view. Depending on the application, the information structure of the problem will naturally change.

Exploring conditions for the MFTG to be a potential game, or an S-modular game, can open a door for applications in for example interference management and resource allocation [42, 46, 23] to make use of this framework.

The following abbreviations have been used:

BSDE	Backward stochastic differential equation
FBSDE	Forward-backward stochastic differential equation
LQ	Linear-quadratic
MFTC	Mean-field type control problem
MFTG	Mean-field type game
ODE	Ordinary differential equation
PoA	Price of Anarchy
SDE	Stochastic differential equation

D.7 Appendix 1: Differentiation of $f : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$

Derivatives of real-valued functions with domain $\mathcal{P}_2(\mathbb{R}^d)$ will be defined with the lifting technique, outlined for example in [9, 10, 12]. Consider the function $f : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$. We assume that our probability space is rich enough, so that for every $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, there exists a square-integrable random variable X whose distribution is μ, i.e., $\mu = \mathcal{L}(X)$. For example, $([0,1], \mathcal{B}([0,1]), dx)$ has this property. Then we may write $f(\mu) =: F(X)$ and we can differentiate F in Fréchet-sense whenever there exists a continuous linear functional $DF[X] : L^2(\mathcal{F}; \mathbb{R}^d) \to \mathbb{R}$ such that

$$F(X+Y) - F(X) = \mathbb{E}[DF[X]Y] + o(\|Y\|_2) =: D_Y f(\mu) + o(\|Y\|_2), \quad (26)$$

where $\|Y\|_2^2 := \mathbb{E}[Y^2]$. $D_Y f(\mu)$ is the Fréchet derivative of f at μ, in the direction Y and we have that for $Y \in L^2(\mathcal{F};\mathbb{R}^d)$ and $\mu = \mathcal{L}(X)$,

$$D_Y f(\mu) = \mathbb{E}[DF[X]Y] =: \lim_{t\to 0} \frac{\mathbb{E}[F(X+tY) - F(X)]}{t}.$$

By Riesz' Representation Theorem, $DF[X]$ is unique and it is known [9] that there exists a Borel function $\varphi[\mu] : \mathbb{R}^d \to \mathbb{R}^d$, independent of the version of X, such that $DF[X] = \varphi[\mu](X)$. Therefore, with $\mu' = \mathcal{L}(X')$ for some random variable X', (26) can be written as

$$f(\mu') - f(\mu) = E[h[X](X'), X' - X] + o(\|X' - X\|_2), \tag{27}$$

for all $X' \in L^2(\mathcal{F};\mathbb{R}^d)$. We denote $\partial_\mu f(\mu; x) := h[\mu](x)$, where $x \in \mathbb{R}^d$, and $\partial_\mu f(\mathcal{L}(X); X) =: \partial_\mu f(\mathcal{L}(X))$, and have the identity

$$DF[X] = h[\mathcal{L}(X)](X) = \partial_\mu f(\mathcal{L}(X)).$$

Example 12. *If* $f(\mu) = (\int_{\mathbb{R}^d} x d\mu(x))^2$ *then*

$$\lim_{t\to 0} \frac{\mathbb{E}[X+tY]^2 - \mathbb{E}[X]^2}{t} = \mathbb{E}[2\mathbb{E}[X]Y],$$

and $\partial_\mu f(\mu) = 2\int_{\mathbb{R}^d} x d\mu(x)$.

Example 13. *If* $f(\mu) = \int_{\mathbb{R}^d} x d\mu(x)$ *then* $\partial_\mu f(\mu) = 1$.

The Taylor approximation of a measure-valued function is given by (27), and we will write

$$f(\mathcal{L}(X')) - f(\mathcal{L}(X)) = \mathbb{E}\left[\partial_\mu f(\mathcal{L}(X))(X' - X)\right] + o(\|X' - X\|_2).$$

Assume now that f takes another argument, ξ. Then

$$\begin{aligned} &f(\xi, \mathcal{L}(X')) - f(\xi, \mathcal{L}(X)) \\ &= \mathbb{E}\left[\partial_\mu f(\widetilde{\xi}, \mathcal{L}(X); X)(X' - X)\right] + o(\|X' - X\|_2), \end{aligned} \tag{28}$$

where the expectation is *not taken over the tilded variable.* Note that $\mathcal{L}(X)$ is deterministic. In situations where the expected value is taken only over the directional argument of $\partial_\mu f$, we will write

$$\mathbb{E}\left[\partial_\mu f(\widetilde{\xi}, \mathcal{L}(X); X)(X' - X)\right] =: \mathbb{E}\left[(\partial_\mu f(\xi, \mathcal{L}(X)))^*(X' - X)\right]. \tag{29}$$

The expected value in (28) is a random quantity because of $\widetilde{\xi}$. Taking another expected value, and changing the order of integration, leads to

$$\mathbb{E}\left[\widetilde{\mathbb{E}}[\partial_\mu f(\widetilde{\xi}, \mathcal{L}(X); X)](X' - X)\right],$$

where the tilded expectation is taken only over the tilded variable. The notation for this will be

$$\widetilde{\mathbb{E}}[\partial_\mu f(\widetilde{\xi}, \mathcal{L}(X); X)] =: \mathbb{E}\left[{}^*(\partial_\mu f(\xi, \mathcal{L}(X)))\right]. \tag{30}$$

D.8 Appendix 2: Proofs

D.8.1 Proof of Lemma 3

Let

$$\widetilde{b}_t^i := \sum_{j=1}^{2} \left\{ \partial_{y^j} \hat{b}_t^i \widetilde{Y}_t^j + \mathbb{E}\left[(\partial_{\mu^j} \hat{b}_t^i)^* \widetilde{Y}_t^j \right] \right\} + \sum_{j,k=1}^{2} \partial_{z^{j,k}} \hat{b}_t^i \widetilde{Z}_t^{i,j},$$

then $\widetilde{Y}_t^i = -\int_t^T \widetilde{b}_s^i + \delta_1 b^i(s) 1_{E_\varepsilon}(s) ds - \sum_{j=1}^2 \int_t^T \widetilde{Z}_s^{i,j} dW_s$. An application of Ito's formula to $|\widetilde{Y}_t^1|^2 + |\widetilde{Y}_t^2|^2$ yields

$$\begin{aligned}\sum_{i=1}^{2} |\widetilde{Y}_t^i|^2 + \int_t^T \sum_{i,j=1}^{2} \|\widetilde{Z}_s^{i,j}\|_F^2 ds = \int_t^T 2 \sum_{i=1}^{2} \langle \widetilde{Y}_s^i, \widetilde{b}_s^i + \delta_1 b^i(s) 1_{E_\varepsilon}(s) \rangle ds \\ + \sum_{i,j=1}^{2} \int_t^T \langle \widetilde{Y}_s^i, \widetilde{Z}_s^{i,j} dW_s^j \rangle.\end{aligned} \tag{31}$$

Let D denote the largest bound for all the derivatives of b^1 and b^2 present. By Jensen's and Young's inequalities,

$$2 \sum_{i=1}^{2} \langle \widetilde{Y}_s^i, \widetilde{b}_s^i \rangle \leq \sum_{i=1}^{2} \left\{ (6D + 16D^2) |\widetilde{Y}_s^i|^2 + 2D\mathbb{E}[|\widetilde{Y}_s^i|^2] \right\} + \frac{1}{2} \sum_{i,j=1}^{2} \|\widetilde{Z}_s^{i,j}\|_F^2.$$

The stochastic integrals in (31) are local martingales and vanish under an expectation [38]. Therefore, with $K_0 := 8D + 16D^2$,

$$\begin{aligned}&\mathbb{E}\left[\sum_{i=1}^{2}|\widetilde{Y}_s^i|^2 + \frac{1}{2}\sum_{i,j=1}^{2}\int_t^T \|\widetilde{Z}_s^{i,j}\|_F^2\, ds\right]\\ &\leq K_0\int_t^T \mathbb{E}\left[\sum_{i=1}^{2}|\widetilde{Y}_s^i|^2\right] ds + 2\int_t^T \mathbb{E}\left[\sum_{i=1}^{2}\langle \widetilde{Y}_s^i, \delta_1 b^i(s) 1_{E_\varepsilon}(s)\rangle\right] ds.\end{aligned} \tag{32}$$

Let $\tau \in [0,T]$, then

$$\sup_{(T-\tau)\leq t\leq T} K_0\int_t^T \mathbb{E}\left[\sum_{i=1}^{2}|\widetilde{Y}_s^i|^2\right] ds \leq K_0\delta \sup_{(T-\tau)\leq t\leq T}\mathbb{E}\left[\sum_{i=1}^{2}|\widetilde{Y}_s^i|^2\right], \tag{33}$$

and by Hölder's and Young's inequalities,

$$\begin{aligned}&\sup_{(T-\tau)\leq t\leq T}\int_t^T \mathbb{E}\left[\sum_{i=1}^{2}\langle \widetilde{Y}_s^i, \delta_1 b^i(s) 1_{E_\varepsilon}(s)\rangle\right] ds\\ &\leq \sup_{(T-\tau)\leq t\leq T}\int_t^T \sum_{i=1}^{2}\mathbb{E}\left[|\widetilde{Y}_s^i|^2\right]^{1/2}\mathbb{E}\left[|\delta_1 b^i(s) 1_{E_\varepsilon}(s)|^2\right]^{1/2} ds\\ &\leq \sum_{i=1}^{2}\left\{\sup_{(T-\tau)\leq t\leq T}\mathbb{E}\left[|\widetilde{Y}_s^i|^2\right]^{1/2}\right\}\int_{T-\tau}^T \mathbb{E}\left[|\delta_1 b^i(s)|^2\right]^{1/2} 1_{E_\varepsilon}(s) ds\\ &\leq \sum_{i=1}^{2}\frac{\delta}{2}\left\{\sup_{(T-\tau)\leq t\leq T}\mathbb{E}\left[|\widetilde{Y}_s^i|^2\right]\right\}\\ &\quad + \frac{1}{2\delta}\left(\int_{T-\delta}^T \mathbb{E}\left[|\delta_1 b^i(s)|^2\right]^{1/2} 1_{E_\varepsilon}(s) ds\right)^2.\end{aligned} \tag{34}$$

By Assumption 5 and the definition of $\mathcal{U}^1$, we have for some $K_1 > 0$,

$$\frac{1}{2\delta}\left(\int_{T-\delta}^T \mathbb{E}\left[|\delta_1 b^i(s)|^2\right]^{1/2} 1_{E_\varepsilon}(s) ds\right)^2 \leq K_1\varepsilon^2.$$

Plugging (33) and (34) into (32) yields

$$\sup_{(T-\delta)\leq t\leq T}\mathbb{E}\left[(1-(K_0+1)\delta)\sum_{i=1}^{2}|\widetilde{Y}_t^i|^2 + \frac{1}{2}\sum_{i,j=1}^{2}\int_t^T \|\widetilde{Z}_s^{i,j}\|_F^2\, ds\right] \leq K_1\varepsilon^2.$$

For $\delta < (K_0+1)^{-1}$, we conclude that

$$\sup_{(T-\delta)\leq t\leq T} \mathbb{E}\left[\sum_{i=1}^{2} |\widetilde{Y}_t^i|^2 + \sum_{i,j=1}^{2} \int_t^T \|\widetilde{Z}_s^{i,j}\|_F^2 \, ds\right] \leq K_2\varepsilon^2,$$

where $K_2 > 0$ depends on δ, the bound D, the Lipschitz coefficient of b^i and the integration bound in the definition of $\mathcal{U}^1$. The steps above can be repeated for the intervals $[T-2\delta, T-\delta]$, $[T-3\delta, T-2\delta]$, etc. until 0 is reached. After a finite number of iterations, we have

$$\sup_{0\leq t\leq T} \mathbb{E}\left[\sum_{i=1}^{2} |\widetilde{Y}_t^i|^2 + \sum_{i,j=1}^{2} \int_t^T \|\widetilde{Z}_s^{i,j}\|_F^2 \, ds\right] \leq K_3\varepsilon^2,$$

where K_3 depends on K_2 and T. This is the first estimate in (8). The second estimate follows from similar calculations.

D.8.2 Proof of Lemma 4

Integration by parts yields

$$\mathbb{E}\left[\sum_{j=1}^{2} \widetilde{Y}_0^j p_0^{1,j}\right] = -\mathbb{E}\left[\int_0^T \sum_{j=1}^{2} \widetilde{Y}_t^j dp_t^{1,j} + p_t^{1,j} d\widetilde{Y}_t^j + d\langle \widetilde{Y}_t^j, p^{1,j}\rangle_t dt\right].$$

Assume that $dp_t^{1,j} = \beta_t^j dt + \sigma_t^{j,1} dW_t^1 + \sigma_t^{j,2} dW_t^2$, then

$$\begin{aligned}
&\sum_{i=1}^{2} \widetilde{Y}_t^i dp_t^{1,i} + p_t^{1,i} d\widetilde{Y}_t^i + d\langle p^{1,i}, \widetilde{Y}^i\rangle_t = \sum_{i=1}^{2} \Bigg[\widetilde{Y}_t^i \left(\beta_t^i dt + \sigma_t^{i,1} dW_t^1 + \sigma_t^{i,2} dW_t^2\right) \\
&+ p_t^{1,i}\left(\sum_{j=1}^{2} \left\{\partial_{y^j} \hat{b}_t^i \widetilde{Y}_t^j + \mathbb{E}\left[(\partial_{\mu^j} \hat{b}_t^i)^* \widetilde{Y}_t^j\right] + \sum_{k=1}^{2} \partial_{z^{j,k}} \hat{b}_t^i \widetilde{Z}_t^{j,k}\right\} + \delta_1 b^i(t) 1_{E_\varepsilon}(t)\right) \\
&+ \sigma_t^{i,1} \widetilde{Z}_t^{i,1} + \sigma_t^{i,2} \widetilde{Z}_t^{i,2}\Bigg] dt + (\dots) dW_t^1 + (\dots) dW_t^2.
\end{aligned}$$

Hence, the lemma is equivalent to that, under expectations, we have

$$
\begin{aligned}
&-\mathbb{E}\Bigg[\int_0^T \Bigg\{ \widetilde{Y}_t^1 \beta_t^1 + \widetilde{Y}_t^2 \beta_t^2 \\
&\quad + \widetilde{Y}_t^1 \Big\{ p_t^{1,1} \partial_{y^1} \hat{b}_t^1 + p_t^{1,2} \partial_{y^1} \hat{b}_t^2 + \mathbb{E}\left[{}^*(\partial_{\mu^1} \hat{b}_t^1) p_t^{1,1}\right] + \mathbb{E}\left[{}^*(\partial_{\mu^1} \hat{b}_t^2) p_t^{1,2}\right] \Big\} \\
&\quad + \widetilde{Y}_t^2 \Big\{ p_t^{1,1} \partial_{y^2} \hat{b}_t^1 + p_t^{1,2} \partial_{y^2} \hat{b}_t^2 + \mathbb{E}\left[{}^*(\partial_{\mu^2} \hat{b}_t^1) p_t^{1,1}\right] + \mathbb{E}\left[{}^*(\partial_{\mu^2} \hat{b}_t^2) p_t^{1,2}\right] \Big\} \\
&\quad + (p_t^{1,1} \partial_{z^{1,1}} \hat{b}_t^1 + p_t^{1,2} \partial_{z^{1,1}} \hat{b}_t^2 + \sigma_t^{1,1}) \widetilde{Z}_t^{1,1} \\
&\quad + (p_t^{1,1} \partial_{z^{1,2}} \hat{b}_t^1 + p_t^{1,2} \partial_{z^{1,2}} \hat{b}_t^2 + \sigma_t^{1,2}) \widetilde{Z}_t^{1,2} \\
&\quad + (p_t^{1,1} \partial_{z^{2,1}} \hat{b}_t^1 + p_t^{1,2} \partial_{z^{2,1}} \hat{b}_t^2 + \sigma_t^{2,1}) \widetilde{Z}_t^{2,1} \\
&\quad + (p_t^{1,1} \partial_{z^{2,2}} \hat{b}_t^1 + p_t^{1,2} \partial_{z^{2,2}} \hat{b}_t^2 + \sigma_t^{2,2}) \widetilde{Z}_t^{2,2} \\
&\quad + (p_t^{1,1} \delta_1 b^1(t) + p_t^{1,2} \delta_1 b^2(t)) 1_{E_\varepsilon}(t) \Bigg\} dt \Bigg] \\
&= -\mathbb{E}\left[\int_0^T \sum_{i=1}^2 \left(\widetilde{Y}_t^i \left\{ \partial_{y^i} \hat{f}_t^1 + \mathbb{E}\left[{}^*(\partial_{\mu^i} \hat{f}_t^1)\right] \right\} - p_t^{1,i} \delta_1 b^i(t) 1_{E_\varepsilon}(t) \right) dt \right].
\end{aligned}
$$

We match coefficients and get

$$
\begin{aligned}
\beta_t^j &= -\left(p_t^{1,1} \partial_{y^j} \hat{b}_t^1 + p_t^{1,2} \partial_{y^j} \hat{b}_t^2 + \mathbb{E}\left[{}^*(\partial_{\mu^i} \hat{b}_t^1) p_t^{1,1}\right] + \mathbb{E}\left[{}^*(\partial_{\mu^i} \hat{b}_t^2) p_t^{1,2}\right] \right) \\
&\qquad + \partial_{y^j} \hat{b}_t^1 + \mathbb{E}\left[{}^*(\partial_{\mu^j} \hat{b}_t^1)\right] \\
&= -\left\{ \partial_{y^j} \hat{H}_t^1 + \mathbb{E}\left[{}^*(\partial_{\mu^j} \hat{H}_t^1\right] \right\}, \\
\sigma_t^{j,k} &= -\left(p_t^{1,1} \partial_{z^{j,k}} \hat{b}_t^1 + p_t^{1,2} \partial_{z^{j,k}} \hat{b}_t^2 \right).
\end{aligned}
$$

D.8.3 Linear-quadratic MFTG: derivation of the ODE system

Under the ansatz, the adjoint equation is

$$
\begin{aligned}
&dp_t^{i,i} \\
&= \Big\{(q_{i,i}(t) + \widetilde{q}_{i,i}(t))\hat{Y}_t^i + (\widetilde{q}_{i,i}(t) + \bar{q}_{i,i}(t))\mathbb{E}[\hat{Y}_t^i] \\
&+ (s_{i,1}(t) + s_{i,2}(t)\bar{s}_{i,i}(t))\hat{Y}_t^{-i} + (\bar{s}_{i,-i}(t) + \bar{s}_{i,1}^E + \bar{s}_{i,2}^E(t))\mathbb{E}[\hat{Y}_t^{-i}]\Big\}dt \\
&=: \Big\{Q_i(t)\hat{Y}_t^i + \bar{Q}_i(t)\mathbb{E}[\hat{Y}_t^i] + S_i(t)\hat{Y}_t^{-i} + \bar{S}_i(t)\mathbb{E}[\hat{Y}_t^{-i}]\Big\}dt \\
&= \Big\{Q_i(t)\Big(\alpha_i(t)p_t^{i,i} + \bar{\alpha}_i(t)\mathbb{E}[p_t^{i,i}] + \beta_i(t)p_t^{-i,-i} + \bar{\beta}_i(t)\mathbb{E}[p_t^{-i,-i}] \\
&\qquad\qquad + \gamma_{i,1}(t)W_t^1 + \gamma_{i,2}W_t^2 + \theta_i(t)\Big) \\
&\quad + \bar{Q}_i(t)\Big((\alpha_i(t) + \bar{\alpha}_i(t))\mathbb{E}[p_t^{i,i}] + (\beta_i(t) + \bar{\beta}_i(t))\mathbb{E}[p_t^{-i,-i}] + \theta_i(t)\Big) \\
&\quad + S_i(t)\Big(\alpha_{-i}(t)p_t^{-i,-i} + \bar{\alpha}_{-i}(t)\mathbb{E}[p_t^{-i,-i}] + \beta_{-i}(t)p_t^{i,i} \\
&\qquad\qquad + \bar{\beta}_{-i}(t)\mathbb{E}[p_t^{i,i}] + \gamma_{-i,1}W_t^1 + \gamma_{-i,2}W_t^2 + \theta_{-i}(t)\Big) \\
&\quad + \bar{S}_i(t)\Big((\alpha_{-i}(t) + \bar{\alpha}_{-i}(t))\mathbb{E}[p_t^{-i,-i}] \\
&\qquad\qquad + (\beta_{-i}(t) + \bar{\beta}_{-i}(t))\mathbb{E}[p_t^{i,i}] + \theta_{-i}(t)\Big)\Big\}dt \\
&= \Big\{p_t^{i,i}\Big(Q_i(t)\alpha_i(t) + S_i(t)\beta_{-i}(t)\Big) \\
&\quad + \mathbb{E}[p_t^{i,i}]\Big(Q_i(t)\bar{\alpha}_i(t) + \bar{Q}_i(t)(\alpha_i(t) + \bar{\alpha}_i(t)) \\
&\qquad\qquad + S_i(t)\bar{\beta}_{-i}(t) + \bar{S}_i(t)(\beta_{-i}(t) + \bar{\beta}_{-i}(t))\Big) \\
&\quad + p_t^{-i,-i}\Big(Q_i(t)\beta_i(t) + S_i(t)\alpha_{-i}(t)\Big) \\
&\quad + \mathbb{E}[p_t^{-i,-i}]\Big(Q_i(t)\bar{\beta}_i(t) + \bar{Q}_i(t)(\beta_i(t) + \bar{\beta}_i(t)) \\
&\qquad\qquad + S_i(t)\bar{\alpha}_{-i}(t) + \bar{S}_i(t)(\alpha_{-i}(t) + \bar{\alpha}_{-i}(t))\Big) \\
&\quad + W_t^1(Q_i(t)\gamma_{i,1}(t) + S_i(t)\gamma_{i,2}(t)) + W_t^2(Q_i(t)\gamma_{i,2} + S_i(t)\gamma_{-i,2}) \\
&\quad + \theta_i(t)(Q_i(t) + \bar{Q}_i(t)) + \theta_{-i}(t)(S_i(t) + \bar{S}_i(t))\Big\}dt \\
&=: \Big\{p_t^{i,i}P^i(t) + \mathbb{E}[p_t^{i,i}]\bar{P}^i(t) + p_t^{-i,-i}R^i(t) + \mathbb{E}[p_t^{-i,-i}]\bar{R}^i(t) \\
&\qquad + W_t^1\Phi^i(t) + W_t^2\Psi^i(t) \\
&\qquad + \theta_i(t)(Q_i(t) + \bar{Q}_i(t)) + \theta_{-i}(t)(S_i(t) + \bar{S}_i(t))\Big\}dt,
\end{aligned}
\tag{35}
$$

and the expected value of $p_\cdot^{i,i}$ solves

$$\begin{aligned} d(\mathbb{E}[p_t^{i,i}]) = \Big\{ & \mathbb{E}[p_t^{i,i}](P^i(t) + \bar{P}^i(t)) + \mathbb{E}[p_t^{-i,-i}](R^i(t) + \bar{R}^i(t)) \\ & + \theta_i(t)(Q_i(t) + \bar{Q}_i(t)) + \theta_{-i}(t)(S_i(t) + \bar{S}_i(t)) \Big\} dt. \end{aligned} \tag{36}$$

The initial conditions $p_0^{i,i}$, $\mathbb{E}[p_0^{i,i}]$, $p_0^{-i,-i}$, $\mathbb{E}[p_0^{-i,-i}]$ are given by a system of linear equations, which is derived is the same way as (35) and (36). Applying Ito's formula to the ansatz, and using (35) and (36), we get

$$\begin{aligned} d\hat{Y}_t^i = & \Big(\dot{\alpha}_i(t) p_t^{i,i} + \dot{\bar{\alpha}}_i(t)\mathbb{E}[p_t^{i,i}] + \dot{\beta}_i(t) p_t^{-i,-i} + \dot{\bar{\beta}}_i(t)\mathbb{E}[p_t^{-i,-i}] \\ & + \dot{\gamma}_{i,1}(t) W_t^1 + \dot{\gamma}_{i,2}(t) W_t^2 + \dot{\theta}_i(t) \Big) dt \\ & + \alpha_i(t) dp_t^{i,i} + \bar{\alpha}_i(t) d(\mathbb{E}[p_t^{i,i}]) + \beta_i(t) dp_t^{-i,-i} + \bar{\beta}_i(t) d(\mathbb{E}[p_t^{-i,-i}]) \\ & + \gamma_{i,1}(t) dW_t^1 + \gamma_{i,2}(t) dW_t^2 \\ = & \Big\{ p_t^{i,i} \Big(\dot{\alpha}_i(t) + \alpha_i(t) P^i(t) + \beta_i(t) R^i(t) \Big) \\ & + \mathbb{E}[p_t^{i,i}] \Big(\dot{\bar{\alpha}}_i(t) + \alpha_i(t)\bar{P}^i(t) + \bar{\alpha}_i(t)(P^i(t) + \bar{P}^i(t)) \\ & \qquad + \beta_i(t)\bar{R}^{-i}(t) + \bar{\beta}_i(t)(R^{-i}(t) + \bar{R}^{-i}(t)) \Big) \\ & + p_t^{-i,-i} \Big(\dot{\beta}_i(t) + \alpha_i(t) R^i(t) + \beta_i(t) P^{-i}(t) \Big) \\ & + \mathbb{E}[p_t^{-i,-i}] \Big(\dot{\bar{\beta}}_i(t) + \alpha_i(t)\bar{R}^i(t) + \bar{\alpha}_i(t)(R^i(t) + \bar{R}^i(t)) \\ & \qquad + \beta_i(t)\bar{P}^{-i}(t) + \bar{\beta}_i(t)(P^{-i}(t) + \bar{P}^{-i}(t)) \Big) \\ & + W_t^1 \Big(\dot{\gamma}_{i,1} + \alpha_i(t)\Phi^i(t) + \beta_i(t)\Phi^{-i}(t) \Big) \\ & + W_t^2 \Big(\dot{\gamma}_{i,2} + \alpha_i(t)\Psi^i(t) + \beta_i(t)\Psi^{-i}(t) \Big) \\ & + \dot{\theta}_i(t) + \theta_i(t) \Big((\alpha_i(t) + \bar{\alpha}_i(t))(Q_i(t) + \bar{Q}_i(t)) \\ & \qquad + (\beta_i(t) + \bar{\beta}_i(t))(S_{-i}(t) + \bar{S}_{-i}(t)) \Big) \\ & + \theta_{-i} \Big((\alpha_i(t) + \bar{\alpha}_i(t))(S_i(t) + \bar{S}_i(t)) \\ & \quad + (\beta_i(t) + \bar{\beta}_i(t))(Q_i(t) + \bar{Q}_i(t)) \Big) \Big\} dt + \gamma_{i,1}(t) dW_t^1 + \gamma_{i,2} dW_t^2. \end{aligned}$$

We can now match these dynamics with the true state dynamics and we get the system of ODEs (22) and $\gamma_{i,j}(t) = \hat{Z}_t^{i,j}$.

References

[1] D. Andersson and B. Djehiche. A maximum principle for SDEs of mean-field type. *Applied Mathematics & Optimization*, 63(3):341–356, 2011.

[2] A. Aurell and B. Djehiche. Mean-field type modeling of nonlocal crowd aversion in pedestrian crowd dynamics. *SIAM Journal on Control and Optimization*, 56(1):434–455, 2018.

[3] A. Aurell and B. Djehiche. Modeling tagged pedestrian motion: A mean-field type game approach. *Transportation research part B: methodological*, 121:168–183, 2019.

[4] T. Basar and G. J. Olsder. *Dynamic noncooperative game theory*, volume 23. Siam, 1999.

[5] T. Başar and Q. Zhu. Prices of anarchy, information, and cooperation in differential games. *Dynamic Games and Applications*, 1(1):50–73, 2011.

[6] A. Bensoussan, J. Frehse, P. Yam, et al. *Mean field games and mean field type control theory*, volume 101. Springer, 2013.

[7] R. Buckdahn, B. Djehiche, J. Li, S. Peng, et al. Mean-field backward stochastic differential equations: a limit approach. *The Annals of Probability*, 37(4):1524–1565, 2009.

[8] R. Buckdahn, J. Li, and S. Peng. Mean-field backward stochastic differential equations and related partial differential equations. *Stochastic Processes and their Applications*, 119(10):3133–3154, 2009.

[9] R. Buckdahn, J. Li, and J. Ma. A stochastic maximum principle for general mean-field systems. *Applied Mathematics & Optimization*, 74 (3):507–534, 2016.

[10] P. Cardaliaguet. Notes on mean field games. Technical report, 2010.

[11] P. Cardaliaguet and C. Rainer. On the (in) efficiency of MFG equilibria. *SIAM Journal on Control and Optimization*, 57(4):2292–2314, 2019.

[12] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I-II.* Springer, 2018.

[13] R. Carmona, C. V. Graves, and Z. Tan. Price of anarchy for mean field games. *ESAIM: Proceedings and Surveys*, 65:349–383, 2019.

[14] Z. Chen and L. Epstein. Ambiguity, risk, and asset returns in continuous time. *Econometrica*, 70(4):1403–1443, 2002.

[15] B. Djehiche, H. Tembine, and R. Tempone. A stochastic maximum principle for risk-sensitive mean-field type control. *IEEE Transactions on Automatic Control*, 60(10):2640–2649, 2015.

[16] B. Djehiche, A. Tcheukam, and H. Tembine. Mean-field-type games in engineering. *AIMS Electronics and Electrical Engineering*, 1(1):18–73, 2017.

[17] N. Dokuchaev and X. Y. Zhou. Stochastic controls with terminal contingent conditions. *Journal of Mathematical Analysis and Applications*, 238(1):143–165, 1999.

[18] P. Dubey. Inefficiency of nash equilibria. *Mathematics of Operations Research*, 11(1):1–8, 1986.

[19] D. Duffie and L. G. Epstein. Stochastic differential utility. *Econometrica: Journal of the Econometric Society*, pages 353–394, 1992.

[20] T. E. Duncan and H. Tembine. Linear–quadratic mean-field-type games: A direct method. *Games*, 9(1):7, 2018.

[21] N. El Karoui, S. Peng, and M. C. Quenez. Backward stochastic differential equations in finance. *Mathematical finance*, 7(1):1–71, 1997.

[22] M. Huang, R. P. Malhamé, P. E. Caines, et al. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle. *Communications in Information & Systems*, 6(3):221–252, 2006.

[23] G. Katsinis, E. E. Tsiropoulou, and S. Papavassiliou. Multicell interference management in device to device underlay cellular networks. *Future Internet*, 9(3):44, 2017.

[24] M. Kohlmann and X. Y. Zhou. Relationship between backward stochastic differential equations and stochastic controls: a linear-quadratic approach. *SIAM Journal on Control and Optimization*, 38(5):1392–1407, 2000.

[25] E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. In *Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413. Springer, 1999.

[26] E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. *Computer science review*, 3(2):65–69, 2009.

[27] D. Lacker. Limit Theory for Controlled McKean–Vlasov Dynamics. *SIAM Journal on Control and Optimization*, 55(3):1641–1672, 2017.

[28] J.-M. Lasry and P.-L. Lions. Mean field games. *Japanese journal of mathematics*, 2(1):229–260, 2007.

[29] J. Li and H. Min. Controlled mean-field backward stochastic differential equations with jumps involving the value function. *Journal of Systems Science and Complexity*, 29(5):1238–1268, 2016.

[30] T. Li and J.-F. Zhang. Asymptotically optimal decentralized control for large population stochastic multiagent systems. *IEEE Transactions on Automatic Control*, 53(7):1643–1660, 2008.

[31] X. Li, J. Sun, and J. Xiong. Linear quadratic optimal control problems for mean-field backward stochastic differential equations. *Applied Mathematics & Optimization*, pages 1–28, 2016.

[32] A. E. Lim and X. Y. Zhou. Linear-quadratic control of backward stochastic differential equations. *SIAM journal on control and optimization*, 40(2):450–474, 2001.

[33] H. Ma and B. Liu. Linear-quadratic optimal control problem for partially observed forward-backward stochastic differential equations of mean-field type. *Asian Journal of Control*, 18(6):2146–2157, 2016.

[34] H. Ma and B. Liu. Maximum principle for partially observed risk-sensitive optimal control problems of mean-field type. *European Journal of Control*, 32:16–23, 2016.

[35] J. Moon and T. Basar. Linear quadratic risk-sensitive and robust mean field games. *IEEE Trans. Automat. Contr.*, 62(3):1062–1077, 2017.

[36] J. Moon, T. E. Duncan, and T. Basar. Risk-sensitive zero-sum differential games. *IEEE Transactions on Automatic Control*, 2018.

[37] C. Papadimitriou. Algorithms, games, and the internet. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 749–753. ACM, 2001.

[38] É. Pardoux. BSDEs, weak convergence and homogenization of semilinear PDEs. In *Nonlinear analysis, differential equations and control*, pages 503–549. Springer, 1999.

[39] É. Pardoux and S. Peng. Adapted solution of a backward stochastic differential equation. *Systems & Control Letters*, 14(1):55–61, 1990.

[40] S. Peng. Probabilistic interpretation for systems of quasilinear parabolic partial differential equations. *Stochastics and stochastics reports*, 37(1-2):61–74, 1991.

[41] S. Peng. Backward stochastic differential equations and applications to optimal control. *Applied Mathematics and Optimization*, 27(2):125–144, 1993.

[42] P. Semasinghe, S. Maghsudi, and E. Hossain. Game theoretic mechanisms for resource management in massive wireless iot systems. *IEEE Communications Magazine*, 55(2):121–127, 2017.

[43] M. Tang and Q. Meng. Linear-quadratic optimal control problems for mean-field backward stochastic differential equations with jumps. arXiv preprint arXiv:1611.06434, 2016.

[44] H. Tembine. Mean-field-type games. *AIMS Math*, 2:706–735, 2017.

[45] H. Tembine, Q. Zhu, and T. Başar. Risk-sensitive mean-field games. *IEEE Transactions on Automatic Control*, 59(4):835–850, 2014.

[46] E. E. Tsiropoulou, P. Vamvakas, and S. Papavassiliou. Joint customized price and power control for energy-efficient multi-service wireless networks via s-modular theory. *IEEE Transactions on Green Communications and Networking*, 1(1):17–28, 2017.

[47] Z. Wu. A general maximum principle for optimal control of forward–backward stochastic systems. *Automatica*, 49(5):1473–1480, 2013.

[48] J. Yong. Forward-backward stochastic differential equations with mixed initial-terminal conditions. *Transactions of the American Mathematical Society*, 362(2):1047–1096, 2010.

[49] J. Yong and X. Y. Zhou. *Stochastic controls: Hamiltonian systems and HJB equations*, volume 43. Springer Science & Business Media, 1999.

[50] J. Zhang. *Backward Stochastic Differential Equations: From Linear to Fully Nonlinear Theory*, volume 86. Springer, 2017.

Paper E

Stochastic stability of mixed equilibria

Stochastic stability of mixed equilibria

by

Alexander Aurell, Lee Dinetan and Gustav Karreskog

Manuscript

Abstract

It is common for models of learning in games to make assumptions so that either a deterministic model can be studied, or the resulting stochastic dynamic is a finite Markov chain. However, in many settings where a stochastic process is considered the assumptions necessary to get a finite-state space model might lead to undesired modeling artifacts. With simulations we show how the assumptions in [25], a well-studied model for stochastic stability, lead to unexpected behavior in games without strict equilibria, such as Matching Pennies. Furthermore, we argue that this behavior should be considered a modeling artifact. In this paper we propose a continuous-state space model with recency bias that can converge to mixed equilibria. The model is similar in spirit to that of Young. We derive known properties of finite-state space models for adaptive play for our model, such as the convergence to and existence of a unique invariant distribution of the state process, and the concentration of that distribution on minimal CURB blocks. Then, we establish conditions under which the state process converges to mixed equilibria inside minimal CURB blocks. While deriving the results for the specific model considered, we develop a methodology that is relevant for a larger class of continuous-state space learning models.

Keywords: conventions, recency bias, learning mixed equilibria, minimal CURB blocks, stochastic stability

E.1 Introduction

The general setting we consider is the evolution of social conventions as introduced in [25]. We consider large populations, one for each player role, from which players are randomly drawn to play a normal form game. Before deciding which action to take the players get access to a sample of historical interactions. The players use this sample to form beliefs about the other players' behavior, and thereafter responds to the mixed strategy induced by that sample. Once they have played, the history is updated, new players are sampled, and the process continues from the updated history.

$$\begin{tikzcd}
\text{History} \arrow[rr] & & \text{Expectations} \arrow[dl] \\
& \text{Actions} \arrow[ul] &
\end{tikzcd}$$

The idea is that in many real life situations, for example when buying a house, each player might not have played the exact game before, but has knowledge about some, but not all, previous interactions and assumes that the player interacting with her will behave similarly to how players have historically behaved. The goal is to see which actions will be taken in the long run, and therefore which stable conventions, if any, will arise.

To get results for the long run distribution of states of the process, it is helpful if the implementation has two properties. The state should be given by a Markov process, and the resulting process should be ergodic so that a unique invariant distribution exists. In the original formulation of [25] this is achieved by defining the state as finite memory of length m with the last m interactions, and a small mistake probability $\varepsilon > 0$ with which a uniformly random action is taken instead of a best reply. The space of possible length m memories is finite, and $\varepsilon > 0$ ensures that it is possible to go from any state to any other state. The resulting process is an ergodic finite state Markov chain and by standard results has a unique invariant distribution that can be studied.

Most of the work building on the original model contain this basic finite memory and noisy action structure, and it is indeed well suited to study the relative stability of different pure (i.e. strict) Nash equilibria or minimal CURB blocks[1]. However, this state space is ill-suited to answer questions

[1] A subset of strategy profiles C is called Closed Under Rational Behavior if the best

	1	**2**
1	$1,-1$	$-1,1$
2	$-1,1$	$1,-1$

Figure 1: Matching Pennies

about behavior around mixed equilibria, since one then has to keep track of the ordering of the history, and exhibits behavior around even simple mixed-equilibria that is better viewed as a modeling artifact than as a realistic description of behavior. The purpose of this paper is to show that by changing the structure of the state space, i.e. the history, we get a process that converges to some minimal CURB configuration and behaves reasonably also inside minimal CURB and around mixed Nash equilibria.

To better understand the limitations of the standard finite memory process, consider perhaps the simplest normal form game with a unique mixed Nash equilibrium: Matching Pennies.

Consider the case where the length of the histroy is $m = 9$, and both players sample the whole history and play without a mistake. Assume that the history contains four interactions where both players took action **1**, followed by five interactions where both took action **2**. The row player will then take action **2** and the column player action **1**. However, since the interaction that falls out of the history is one where the column player played **1**, the sample to which the row player responds will not change until the **1**:s in the end of the history have all fallen out and the first interaction with a **2** falls out of the history. At that point, the history contains five interactions where the column player played **1**, so now the row player wants to play **1** as well. However, by now all the interactions in the history are such that the row player played **2**. So for the coming five interactions they will both take action **1**.

$$\begin{pmatrix} 111122222 \\ 111122222 \end{pmatrix} \to \begin{pmatrix} 222222222 \\ 222211111 \end{pmatrix} \to \begin{pmatrix} 222211111 \\ 111111111 \end{pmatrix} \to \begin{pmatrix} 111111111 \\ 111122222 \end{pmatrix} \to \cdots$$

The behavior in the next period depends as much on what falls out of the history as what is added, and this generates the cycling behavior. The cycling

replies to any strategy profile with support in C is also in C. It is called a minimal CURB block if it does not contain any strictly smaller CURB block [3].

behavior does not only happen in this special case but is a general feature observed when simulating the process.

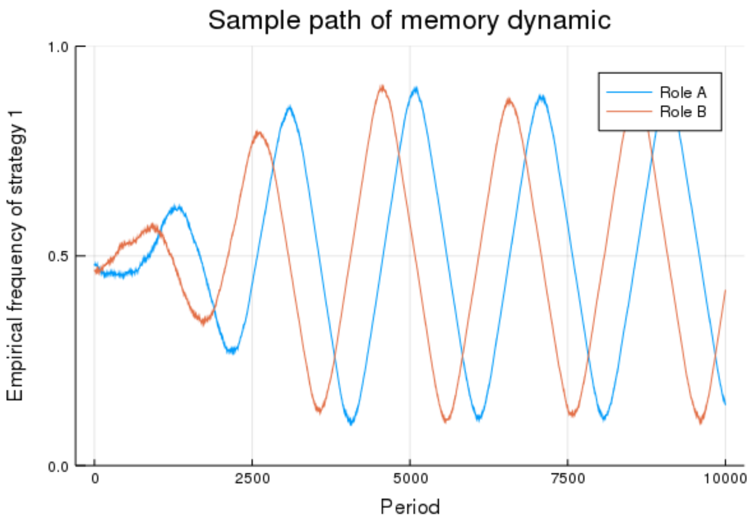


Figure 2: A 10 000 period simulation of the finite memory dynamic on matching pennies with $m = 1000$, $k = 20$, $\varepsilon = 0.05$. Initiated at the 50/50 equilibrium.

To address this problem we introduce a new dynamic that differs from previous works in the structure of the history. In this dynamic, the history is assumed to be infinite, but more recent interactions are more likely to be sampled. The probability of sampling an interaction decreases with a factor β, $0 < \beta < 1$, for each step back in the history. This decrease allows us to reduce the state to simply the sampling probabilities for the different strategies while maintaining the Markovian property of the process (i.e. the history). We can in a meaningful way analyze the process at a finer level inside the minimal CURB blocks (and determine properties of the distribution of play inside a minimal CURB block). A simulation of this new process can be seen in Figure 3. The recency process stays in small neighborhood of the mixed Nash equilibrium $(0.5, 0.5)$.

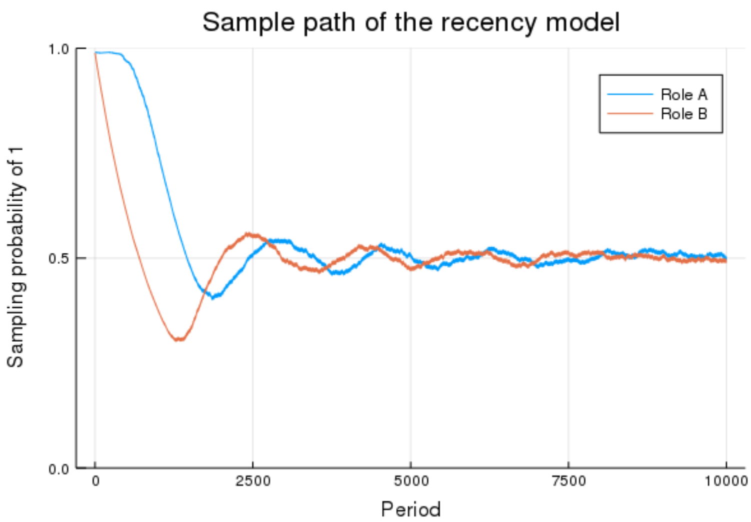


Figure 3: A 10 000 period simulation of the recency dynamic on matching pennies with $\beta = 0.999$, $k = 20$, $\varepsilon = 0.05$.

We analyze the resulting state process with tools from the theory of continuous state space Markov processes. The results and tools used here are applicable to a larger set of processes than the specific example we consider, and showing how to analyze such a population dynamic is in itself a relevant contribution.

E.1.1 Related Literature

Already in his dissertation John Nash gave a second interpretation of the Nash equilibrium, the *mass action* interpretation [19]. He considers a setting where a large population is associated to each player role, and one player per role is selected in each period to play the game, and that the individual players accumulate empirical information on the relative advantage of the different available pure strategies. He then argues, informally, that in such a setting, the stable points correspond to Nash equilibria and those points should eventually be reached by the process.

The mass action interpretation is appealing since its assumptions about bounded rationality and repeated interactions are more credible than those underlying the rationalistic interpretation built on assumptions of perfect rationality and common knowledge[2]. Furthermore, experimental evidence clearly favors some kind of learning and adjustment over the rationalistic motivation. The general result is that in a one-shot interaction, play rarely corresponds to a Nash equilibrium, but if the players have a chance to learn and adjust, play often (but far from always) moves to a Nash equilibrium. See [8, Ch. 6] for an overview of experimental models and results.

Appealing as the motivation might be, the theoretical picture has turned out to be considerably more complicated than indicated by Nash's informal argument and what many researchers might initially have thought. One of the first, and most studied, models formalizing a setting similar in spirit to the mass-action interpretation is that of fictitious play by Brown in [7]. Even though Brown thought fictitious play would in general converge to a Nash equilibrium, it was shown in [23] that even in a game with a unique Nash equilibrium there might only exist a stable cycle and no convergence to the mixed equilibrium. In general, it is the case that if the process has a stationary point, it must be a Nash equilibrium, but the existence of such a stationary point is not guaranteed. See e.g. [14], [24], or [22] for overview of such results. The general results that exist are not about convergence to stable points, which would then normally correspond to Nash equilibria, but convergence to stable sets. [21] show set-convergence results for evolutionary dynamics and [2] for best reply dynamics. Similarly [17] and [26] show set-convergence results for dynamics similar to those studied in this paper.

Smooth fictitious play, first introduced in [12], is a version of fictitious play where players respond with a perturbed best response. In contrast to the standard version of fictitious play, not only the empirical frequency but also actual play can converge to a Nash equilibrium. In [4] and [16], global convergence results are shown for some games with unique Nash equilibria, including interior ESS, two-player zero-sum, supermodular and potential games.

A downside with standard versions of both fictitious play and smooth fictitious play is that the stepsize decreases with time, and the state moves slower and slower. So the initial starting point is crucial, and when behavior is cyclic the cycles take longer and longer time to complete. Introducing some kind of recency, similar to that used in this paper, gives a dynamic with constant stepsize, which for many applications is more natural. Such

[2] Especially since perfect rationality and common knowledge by itself only leads to rationalizability but not all the way to Nash equilibrium.

dynamics are studied in [5], where the time average in unstable games is studied, and [13].

The one class of dynamics for which there exists quite general results for convergence to equilibrium rely on a combination of noisy behavior and satisfaction. A given player randomly explores actions until she is satisfied, e.g. her received payoff is higher than some threshold or close enough to the maximum payoff observed. Then she keeps taking that action as long as she is still satisfied. The exact setting and formulation of results vary between the cited papers, but in general this kind of dynamic seems to able to converge to a Nash equilibrium under quite general circumstances. The unsatisfactory part of this class of dynamics is that players are in a sense maybe too unsophisticated, at least if the game is known, and that the path to equilibrium thus might be very long and somewhat unrealistic [11, 20, 15, 6].

The existing literature building on [25, 26], has not focused on questions about mixed equilibria, but instead focused on questions about speed of convergence, e.g. [18], or improving tools for finding stochastically stable subsets, e.g. [10], but to our knowledge no one has studied mixed equilibria more carefully.

E.1.2 Summary and outline

Since we make considerable changes to the underlying model, and most crucially the state-space, we cannot rely directly on any existing results for the memory (state) process. In Section E.2 the new model is formalized and we introduce the tools we need to analyze the process. We prove weak convergence to a unique invariant distribution for a large class of memory processes within our proposed model and that in the limit as the error-rate $\varepsilon \to 0$, the invariant distribution will concentrate on the minimal CURB blocks of the game. Once we have recovered these crucial results, we turn to the question of behavior inside minimal CURB blocks that are non-singleton, and show that for any generic game where the minimal CURB blocks are at most 2×2 play will eventually concentrate around Nash equilibria or, when the sample size k is small, close to some point on the k-grid spanning the simplices which is also close to the Nash equilibrium. Most proofs have been appended in the end of the paper.

E.2 The Model

We consider a two-player finite game G, iteratively played by two new players drawn from large populations. The game has two asymmetric player roles, 1 and 2. The sets of pure strategies in the game are S_1 and S_2, containing $m_1 \in \mathbb{N}$ and $m_2 \in \mathbb{N}$ pure strategies respectively; the spaces of mixed strategies are thus $\Delta(S_1)$ and $\Delta(S_2)$.

For $\sigma_{-i} \in \Delta(S_{-i})$, we will denote by $BR_i(\sigma_{-i}) \subset S_i$ the set of best replies of player i to the mixed strategy σ_{-i}. We identify $\Delta(S_i)$ with the $(m_i - 1)$-dimensional simplex, so that the space $\square(S) := \Delta(S_1) \times \Delta(S_2)$ endowed with its usual uniform distance $|\cdot|$ is compact.

E.2.1 Sampling strategies

Each play is recorded as a pair (s_1, s_2), with $s_1 \in S_1$ and $s_2 \in S_2$ the strategies played by each player. Denoting $s_1(t)$ and $s_2(t)$ the strategies played at time t, the history is thus a sequence of plays

$$((s_1(t), s_2(t)))_{t \in \mathbb{Z}}.$$

Notice that for $t < 0$, the history is just some infinite history, coding for fictional plays for the purposes of our mechanisms.

At each time t, each player of role $i \in \{1, 2\}$ samples $k \in \mathbb{N}$ plays (with replacement) from the history of the opposing player role $-i$. Each sample is drawn independently and samples are drawn with bias towards more recent plays in a geometric fashion. Namely, players of role i have a bias $\beta \in [0, 1)$ called the recency parameter, such that at time t the probability of selecting the time period $t - \tau$, $\tau \in \{1, \ldots, t\}$, is

$$(1 - \beta)\,\beta^{\tau - 1}.$$

Therefore, a play of the strategy $s \in S_{-i}$ will be sampled by player i with total probability

$$p_{-i,s}(t) = (1 - \beta) \sum_{\tau=1}^{\infty} \beta^{\tau-1} 1_s(s_{-i}(t - \tau)),$$

where 1_s is the indicator function on s.

We will call $p_{-i}(t) := \left(p_{-i,1}(t), \ldots p_{-i,m_{-i}}(t)\right)$ the state process of player role $-i$ at time t. It is a vector of sampling probabilities obtained by player

i from player $-i$'s history and is an element of $\Delta(S_{-i})$. The result of the sampling is a random vector $\left(n_{-i,1}(t), \ldots n_{-i,m_{-i}}(t)\right)$ of integers, following a multinomial distribution with parameters k and $p_{-i}(t)$. For $s \in S_i$, let $\overrightarrow{1_{i,s}} \in \Delta(S_i)$ be the unit vector representing the pure strategy $s \in S_i$, i.e. a vector of of size m_i with 0 everywhere except at position s, where it is 1. From her sample, player i forms an empirical (average) opposing strategy profile

$$D_{-i}(t) := \frac{1}{k} \sum_{s=1}^{m_{-i}} n_{-i,s}(t) \overrightarrow{1_{-i,s}} \in \Delta(S_{-i}). \tag{1}$$

Player i now deems her opponent will play at turn t accordingly to the mixed strategy $D_{-i}(t)$ and tries to play the best response to it. However, player i can make a mistake. Player i's error parameter (or mistake frequency) $\varepsilon \in (0,1]$ indicates the probability she will fail to play a strategy in $BR_i(D_{-i}(t))$, and instead plays, with uniform probability, an arbitrary strategy in S_i. If $BR_i(D_{-i}(t))$ is not a singleton, the realized action is sampled uniformly from all the elements of $BR_i(D_{-i}(t))$. We denote the outcome of the uniform sampling between all best replies to $x \in \Delta S_{-i}$ by $\widehat{BR}_i(x) \in S_i$. The distinction we want to emphasize with this notation is that $BR_i(x)$ is set-valued (the set of all best replies to x) while $\widehat{BR}_i(x)$ is S_i-valued and random (one of the best replies has been randomly selected).

In the end, player i will play $\widehat{BR}_i(D_{-i}(t))$, with $D_{-i}(t)$ obtained as described above, with a probability of $1-\varepsilon$; and additionally, play any strategy $s \in S_i$ with probability ε/m_i. We complete this section by calling

$$\widetilde{BR}_i(p_{-i}) \in S_i$$

the random choice of strategy obtained by a player i through the following process:

1. Looking at a history where plays of strategies by the opposing role get sampled with probabilities given by p_{-i};

2. Sampling k of them to form the empirical expectation D_{-i};

3. Actually playing the best response $\widehat{BR}_i(D_{-i})$, except in a fraction ε of the time when a randomly selected strategy is played.

E.2.2 Run of the plays

At $t = 0$, a initial infinite history $((s_1(u), s_2(u)))_{u \in \mathbb{Z}_-}$ is given. At each time $t \in \mathbb{N}_0$, two new individuals are assigned to the roles. They use same value of the parameters k and ε. After sampling from the history with recency parameter β, they play $s_i(t) = \widetilde{BR}_i(p_{-i}(t))$, $i = 1, 2$, where $p_{-i}(t)$ is exactly the historical distribution of plays with recency bias. The realized plays $(s_1(t), s_2(t))$ are appended to the history, and the step restarts. The exponential nature of sampling leads to the following important property of the state process $(p_1(t), p_2(t)) \in \square(S)$.

Proposition 1. *The state process $p_i(t) \in \Delta(S_i)$ obeys the equation*

$$p_i(t+1) = \beta p_i(t) + (1-\beta)\overrightarrow{1_{i,s_i(t)}}$$

where $s_i(t) = \widetilde{BR}_i(p_{-i}(t))$ is drawn randomly according to the model.

The order of historical plays is not necessary to characterize the model, since the vectors $(p_1(t), p_2(t)) \in \square(S)$ capture all the relevant information. It follows that the history of games, summarized by $(p_1(\cdot), p_2(\cdot))$, is a Markov process over $\square(S)$, and its Markov transition kernel may be expressed as a linear stochastic operator K. The operator K is analyzed further in Section E.2.3 below. In particular, from the position $(p_1(t), p_2(t)) \in \square(S)$, at most $m_1 m_2$ different points $(p_1(t+1), p_2(t+1))$ may be reached. Conditioned on $p(t)$, for any $s_1 \in S_1$ and $s_2 \in S_2$ the point

$$\left(\beta p_1(t) + (1-\beta)\overrightarrow{1_{1,s_1}}, \beta p_2(t) + (1-\beta)\overrightarrow{1_{2,s_2}}\right)$$

will be reached when $s_1(t) = s_1$ and $s_2(t) = s_2$, which happens with probability

$$\prod_{i=1}^{2} \mathbb{P}\left(\widetilde{BR}_i(p_{-i}(t)) = s_i \mid p_{-i}(t)\right),$$

since players sample independently, and

$$\mathbb{P}\left(\widetilde{BR}_i(p_{-i}(t)) = s_i\right) = (1-\varepsilon)\,\mathbb{P}\left(\widehat{BR}_i(D_{-i}(t)) = s_i\right) + \varepsilon/m_i, \quad (2)$$

where we remember that $D_{-i}(t) \in \Delta(S_{-i})$ is a multinomial combination of strategies (with parameters k and $p_{-i}(t)$). Importantly, the probability in (2) is a Lipschitz-continuous function of the state $(p_1(t), p_2(t))$.

Proposition 2. *For $i \in \{1, 2\}$, the functions*

$$p_{-i} \mapsto \mathbb{P}\left(\widetilde{BR}_i(p_{-i}) = s_i\right)$$

are Lipschitz over $\Delta(S_{-i})$ for all for all $s_i \in S_i$, with Lipschitz constants at most $(1-\varepsilon)\, km_{-i}$ with respect to the norm $\|\cdot\|_\infty$ over $\Delta(S_{-i})$.

E.2.3 The kernel K

The kernel K of our Markovian state process p will be defined in this section. The state process takes values in the continuous set $\square(S)$ and therefore its kernel is defined as a functional taking continuous real-valued functions on $\square(S)$. The kernel is the continuous-state space equivalent of the transition rate matrix in models with a discrete state space.

Let $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ be an underlying filtered probability space. The filtration is the natural filtration of $(p_1(t), p_2(t)\ ;\ t \in \mathbb{N})$ and satisfies the usual conditions. Let $E := C(\square(S); \mathbb{R})$ be the space of continuous real-valued functions over $\square(S)$, endowed with the uniform norm $\|\cdot\|_\infty$. Let $\mathcal{P}(\square(S))$ denote the set of Borel probability measures on $\square(S)$.

We define K as the following linear operator from E to E

$$\begin{aligned} K &: E \to E \\ f(p_1, p_2) &\mapsto \sum_{s_1=1}^{m_1} \sum_{s_2=1}^{m_2} \mathbb{P}\left(\widetilde{BR}_1(p_2) = s_1, \widetilde{BR}_2(p_1) = s_2\right) \times \\ &\qquad f\left(\beta p_1 + (1-\beta)\overrightarrow{1_{1,s_1}}, \beta p_2 + (1-\beta)\overrightarrow{1_{2,s_2}}\right). \end{aligned} \tag{3}$$

From this definition comes the elementary properties of K.

Proposition 3. *K is an operator from E to E. It is linear, continuous, and stochastic.*

E.3 Main Results

E.3.1 The unique probability measure invariant with respect to K

The first thing we need to perform the subsequent analysis is to prove that our Markovian state process actually has a unique invariant distribution. To do this we will make use of a newly developed theorem that relies on what can be called *Lipschitz-friendly* operators [9, Thm. 17]. The full justification has been passed on to the appendices.

The theorem itself is more general than what is needed for our goals and is a contribution in itself. The result holds for any Markov process with the same state space and recency-weighted state updating, as long as the stochastic best response function is Lipschitz-continuous in the state, and there is a non-zero lower bound $\varepsilon > 0$ with which any given strategy is played in any state.

Theorem 4. *Assume that the map*

$$\square(S) \ni p \mapsto \mathbb{P}(\widetilde{BR}(p) = s) \in [0,1],$$

is Lipschitz-continuous and uniformly bounded from below by $\epsilon > 0$ for all $s \in S$, and that the recency parameter $\beta \in \left(1 - \frac{1}{\max\{m_1,m_2\}}, 1\right)$. Then, for every $\nu \in \mathcal{P}(\square(S))$, there are $c \in \mathbb{R}_+$, $\lambda \in (0,1)$, and a probability measure $\mu \in \mathcal{P}(\square(S))$ such that

$$\|\nu K^n - \mu\|_W \leq c\lambda^n, \tag{4}$$

where K is the kernel defined in (3) and $\|\cdot\|_W$ is the Wasserstein distance on $\mathcal{P}(\square(S))$. Furthermore, μ is the unique probability measure invariant with respect to K.

In other words, Theorem 4 says that for whichever initial distribution ν that $p(0)$ is drawn from, the distribution of the Markov process will converge "geometrically uniformly" to the probability measure μ which is the unique solution of $\mu K = \mu$.

Examples of other best response functions $\widetilde{BR}$ than the one studied here for which Theorem 4 applies are logit best reply, i.e.

$$\mathbb{P}\left[\widetilde{BR}_i(p) = s_i\right] = \frac{\exp(\eta\pi_i(s_i, p_{-i}))}{\sum_{a\in S_i}\exp(\eta\pi_i(a, p_{-i}))}$$

for some $\eta > 0$, or a model where k itself is a random parameter, or where only robust best responses to the sample are considered.

E.3.2 Convergence to minimal CURB configurations

Before we turn to the question of convergence to a minimal CURB block, one minor technical detail most be resolved. A minimal CURB block is a collection of strategy profiles $C_1 \times C_2 \subset S$ such that the best reply to all mixed strategies in in the sub-simplex spanned by those strategies is always inside the set, i.e. $BR(\sigma) \subset C$ for all $\sigma \in \square(C)$, where $\square(C) := \Delta(C_1) \times \Delta(C_2)$. However, since our agents only reply to samples of size k, it might be the case that the mixed strategy from the simplex that has a best reply outside a non-CURB block simply never is sampled. The game below is a simple illustration of this point.

	1	**2**
1	$2, -100$	$-100, 2$
2	$-100, 2$	$2, -100$
3	$1, 0$	$1, 0$

If $k = 1$ only best replies to pure strategies will ever be considered. If the process initially has support only on the block $\{\mathbf{1}, \mathbf{2}\} \times \{\mathbf{1}, \mathbf{2}\}$, the best reply to any sample will be inside that block, even though **3** is the best reply to most properly mixed strategies. We could call this smaller set of blocks that are closed under best replies to any strategies on the k-lattice spanned, $\square^k(C)$ k-CURB blocks. In most settings, a relatively small k is enough for the k-CURB blocks to coincide with the CURB blocks. In the rest of the paper, we will speak of CURB blocks and by that mean k-CURB blocks. Alternatively, we could have assumed that k is sufficiently large for them to coincide.

In what follows, we first prove that the model concentrates on minimal CURB blocks for general two player games. Then we prove convergence of paths to an approximate mixed Nash equilibrium in the special case $m_1 = m_2 = 2$ and a unique mixed Nash equilibrium. Finally, the results are combined and we show that in general games with minimal CURB blocks of size at most 2×2, each containing a unique mixed Nash equilibrium, the process spends arbitrarily long time, controlled by β, in a neighbourhood of approximate Nash equilibria.

E.3.2.1 Concentration on minimal CURB blocks

The underlying idea of the proof is similar to the one used in results for the original finite memory dynamics. Our case is complicated by the fact that once a strategy has been played, it never truly disappears from memory but always has a positive probability of being sampled. However, that probability decreases over time as long as the strategy is not played again. We will therefore consider a neighbourhood $B_\delta(C)$, $\delta > 0$ around strategy blocks $C := C_1 \times C_2 \subset S$, defined as all pairs (p_1, p_2) in $\square(S)$ such that each of the components puts at least $1 - \delta$ probability on the block.

Definition 5. *The set $B_\delta(C) \subset \square(S)$ is defined for any $\delta > 0$ by*

$$B_\delta(C) := \left\{ p = (p_1, p_2) \in \square(S) \ \mid \ \sum_{s=1}^{m_i} p_{i,s} 1_{C_i}(s) \geq 1 - \delta, \ i = 1, 2 \right\}.$$

We will analyze the time it takes for the state process to enter such neighbourhoods, and the expected time it will spend in there once entered. Let $\mathcal{C}$ denote the union of all minimal CURB blocks in the game. We will show that expected time to go from $B_\delta(\mathcal{C})^c$ to $B_\delta(\mathcal{C})$ is always bounded, but the expected time spent inside $B_\delta(\mathcal{C})$ once entered goes to infinity as ε goes to zero. This in turn will imply that as ε goes to zero, the invariant distribution concentrates on the neighbourhood $\mathcal{C}$, the union of all minimal CURB blocks.

Theorem 6. *Let $\mathcal{C}$ be the union of all the minimal CURB blocks of the game and let $B_\delta(\mathcal{C})$ be the set given by Definition 5. It holds for all $\delta > 0$ that as $\varepsilon \to 0$, the invariant distribution μ^*_ε concentrates on $B_\delta(\mathcal{C})$,*

$$\lim_{\varepsilon \to \infty} \mu^*_\varepsilon \left(B_\delta(\mathcal{C}) \right) = 1.$$

Proof. The proof consists of three steps.

Step 1. Bounding the probability of reaching $B_\delta(\mathcal{C})$ in finite time.
To find a lower bound for the probability to go from an arbitrary point $p(t) \in B_\delta(\mathcal{C})^c$ to $B_\delta(\mathcal{C})$ in finite time we create a particular path of positive probability that does exactly that. Let $p(t) \in \square(S)$ be given and let $s^1 \in S_1 \times S_2$ be the strategy profile played at in period t. Either s^1 is a CURB block, or the best reply set to s^1 contains a strategy not in s^1, $BR(\overrightarrow{1_{s^1}}) \not\subset s^1$. If the former statement is true this step of the proof is complete. That is not always the case, therefore assume that we are in the case of the latter statement, i.e. that the best reply set to s^1 contains a strategy not in s^1. Then, the probability of both players only sampling s^1 at time $t+1$ is bounded from below by $(1-\beta)^{2k}$. Hence the probability of a strategy profile

$s^2 \in BR(\overrightarrow{1_{s^1}}), s^2 \neq s^1$, being played is bounded from below by

$$\mathbb{P}\left(\widetilde{BR}(p(t)) = s^2 \mid p(t)\right) \geq \frac{(1-\beta)^{2k}}{m_1 m_2}(1-\varepsilon)^2.$$

Now let F_2 be the smallest block $F_2 \in S_1 \times S_2$ that contains $\{s^1, s^2\}$. Either F_2 is a CURB block or $BR(\Delta(F_2)) \not\subset F_2$, in which case there is at least one sample D of size k from F_2 such that $BR(D) \not\subset F_2$. The probability of sampling that particular D, and the best replies to D being such that at least one of them is not in F_2, is again bounded away from zero. Until we have sampled a sequence of strategy profiles, each extending the set F_i, such that F_i is a CURB block, there is always some sample with positive sampling probability such that $BR(D) \not\subset F_i$. The probability of playing a strategy s^i which is a best reply to D which is not in F_i, $s^i \in BR(D) \cap (F_i)^c$, is therefore bounded from below by

$$\mathbb{P}\left(\widetilde{BR}(p(t+i-1)) = s^i \mid p(t+i-1)\right) \geq \frac{\left(\beta^{i-1}(1-\beta)\right)^{2k}}{m_1 m_2}(1-\varepsilon)^2.$$

Keep filling $F_i, F_{i+1}, F_{i+2}, \ldots$ with strategies from the CURB block in this fashion, so that F_T spans a CURB block and $T \leq m_1 + m_2$ [17, Lemma 1]. To get a uniform lower bound, assume that $T = m_1 + m_2$ and that once F_i is a CURB block the following $T - i$ strategy profiles are inside the CURB block. The probability of this progression of plays is bounded from below: let $\mathcal{E}$ be the event that $p(t+T)$ puts at most β^{T+1} mass outside the CURB block spanned by F_T, then

$$\mathbb{P}(\mathcal{E}) \geq \frac{\left(\beta^{2k}\right)^{(T-1)!}(1-\beta)^{2Tk}}{m_1^T m_2^T}(1-\varepsilon)^{2T}.$$

Inside the CURB block spanned by F_T, there is a minimal CURB block which we denote by $C = C_1 \times C_2$. The probability of both players sampling from C given the state $p(t+T)$ (as described above) is greater or equal to

$$\mathbb{P}\left((D_1/k, D_2/k) \in \square(C) \mid D \text{ from } p(t+T)\right) \geq \left(\beta^T(1-\beta)\right)^{2k}(1-\varepsilon)^2.$$

Starting from $p(t) \in B_\delta(C)^c$, a sequence of plays that results in $p(t+T+T^*) \in B_\delta(C)$ is to play T strategies to fill F_T followed by T^* strategies from the minimal CURB block C. Conditional on $p(t) \in B_\delta(\mathcal{C})^c$ and the aforementioned event $\mathcal{E}$, the probability that $p(t+T+T^*) \in B_\delta(C) \subset B_\delta(\mathcal{C})$ is bounded from below by

$$\mathbb{P}\Big((D_1, D_2)(t+T+i) \in \square(C), i = 0, \ldots, T^* - 1 \mid p(t+T) \text{ as above}\Big)$$
$$\geq \big(\beta^T(1-\beta)(1-\varepsilon)\big)^{2kT^*} =: \gamma(\varepsilon, T, T^*).$$

Now $p(t+T+T^*)$ gives at most β^{T^*} probability to all strategy profiles outside $\square(C)$. Therefore, we pick $\delta > 0$ and let $T^* \in \mathbb{N}$ be such that $\beta^{T^*} < \delta$ and, summarizing the analysis in this step, we have derived a bound on the probability of moving from any point $p(t) \in B_\delta(\mathcal{C})^c$ to $B_\delta(\mathcal{C})$ in $T+T^*$ steps. We denote this bound by $\underline{K}$ and it is given by

$$K^{T+T^*}(p(t), B_\delta(\mathcal{C}))$$
$$\geq \frac{\left(\beta^{2k}\right)^{(T-1)!}(1-\beta)^{2TK}(1-\varepsilon)^{2T}}{m_1^T m_2^T}\gamma(\varepsilon, T, T^*) =: \underline{K}.$$

Step 2. Expected exit time from $B_\delta(\mathcal{C})$.
Once in $B_\delta(\mathcal{C})$, one of two things must happen for the process to leave. Either one player makes a mistake or one player samples at least one strategy from outside the minimal CURB block C the process is currently centered around. So instead of calculating the time to the first exit, denoted τ_ε, we calculate the expected time until one of these two things happen the first time. Let τ_ε^* denote the time, starting from $t = 0$, until either a strategy is sampled outside C or one player makes an ε-tremble. We denote the expression for the probability that $\tau_\varepsilon^* > t^*$, $t^* \in \mathbb{N}$, with $Q_\varepsilon(t^*)$,

$$Q_\varepsilon(t^*) := \mathbb{P}\left(\tau_\varepsilon^* > t^* \mid p(0) \in B_\delta(C)\right) = \prod_{t=0}^{t^*}(1-\beta^t\delta)^{2k}(1-\varepsilon)^2.$$

For the case $\varepsilon = 0$, we use the fact that $\sum_{t=0}^{\infty}\beta^t\delta$ is convergent to conclude that $\prod_{t=0}^{\infty}(1-\beta^t\delta)^{2k}$ approaches a non-zero limit. Since Q_ε is decreasing and non-negative,

$$\lim_{t^*\to\infty} Q_\varepsilon(t^*) = \begin{cases} Q^* \in (0,1), & \text{if } \varepsilon = 0, \\ 0, & \text{if } \varepsilon > 0. \end{cases} \tag{5}$$

We can now derive a bound for τ_ε, the expected time to exit from $B_\delta(\mathcal{C})$,

$$\begin{aligned}\mathbb{E}[\tau_\varepsilon] &\geq \mathbb{E}[\tau^*_\varepsilon] \\ &\geq \mathbb{E}[\tau^*_\varepsilon \mid \tau^*_\varepsilon \geq t^*, p(0) \in B_\delta(C)] \\ &\qquad \times \mathbb{P}(\tau^*_\varepsilon \geq t^* \mid p(0) \in B_\delta(C))\mathbb{P}(p(0) \in B_\delta(C))) \\ &\geq t^* Q_\varepsilon(t^*)\nu(B_\delta(C)),\end{aligned}$$

where ν is the initial distribution of the state process and $\nu(B_\delta(C))$ is the probability that $p(0) \in B_\delta(C)$. We know that the state process converges weakly to the invariant distribution for all initial distributions and therefore ν is any distribution on $\square(S)$ of our choice. Choosing ν as the distribution of the constructed $p(t + T + T^*)$ from above,

$$\begin{aligned}E[\tau_\varepsilon] &\geq t^* \prod_{t=0}^{t^*}(1 - \beta^t\delta)^{2k}(1 - \varepsilon)^2 \\ &= t^*(1 - \varepsilon)^{2t^*} Q_0(t^*) \\ &\geq t^*(1 - \varepsilon)^{2t^*} Q^*,\end{aligned}$$

where t^* is any positive integer. For a fixed ε, the function $t^* \mapsto t^*(1 - \varepsilon)^{2t^*}$ is maximized by $t^*(\varepsilon) = -(2\ln(1 - \varepsilon))^{-1}$. There is therefore a decreasing sequence of positive numbers $(\varepsilon_j)_{j=1}^{\infty}$, tending to zero as $j \to \infty$, such that $t^*(\varepsilon_j)$ is an integer and

$$\mathbb{E}[\tau_\varepsilon] \geq -\frac{Q^*}{2e\ln(1 - \varepsilon_j)},$$

which diverges to ∞ as $j \to \infty$.

Step 3. Putting it all together.
We know that for any $\varepsilon > 0$ there exists a unique invariant probability measure μ^*_ε. We also have a lower bound for $K(x, B_\delta(\mathcal{C}))$ uniform over $x \in B_\delta(\mathcal{C})^c$, and a lower bound for the expected time the process stays in $B_\delta(\mathcal{C})$ once it has entered.

The probability given by the invariant distribution to the set $B_\delta(\mathcal{C})$ is at least the sum over n of the probability of: the state process not being in it $(n + 1)(T + T^*)$ steps ago, but in it $n(T + T^*)$ steps ago, and then staying there for at least $n(T+T^*)$ time steps,

$$\begin{aligned}
1 \geq \mu_\varepsilon^*(B_\delta(\mathcal{C})) &\geq \sum_{n=0}^{\infty} \left(\int_{B_\delta(\mathcal{C})^c} K^{T+T^*}\left(x, B_\delta(\mathcal{C})\right) d\mu_\varepsilon^*(x) \right) \mathbb{P}\left(\tau_\varepsilon \geq n(T+T^*)\right) \\
&\geq \mu_\varepsilon^*\left(B_\delta(\mathcal{C})^c\right) \underline{K} \left(\sum_{n=0}^{\infty} \mathbb{P}\left(\frac{\tau_\varepsilon}{T+T^*} \geq n \right) \right) \\
&\geq \mu_\varepsilon^*\left(B_\delta(\mathcal{C})^c\right) \frac{\underline{K}}{T+T^*} \mathbb{E}\left[\tau_\varepsilon^*\right].
\end{aligned}$$

Since $\underline{K} > 0$ increases as $\varepsilon \to 0$, $\mathbb{E}\left[\tau_\varepsilon^*\right] \to \infty$ as $\varepsilon \to 0$, and $T+T^*$ does not depend on ε, we conclude that $\mu_\varepsilon^*\left(B_\delta(\mathcal{C})^c\right) \to 0$ as $\varepsilon \to 0$. □

The following corollary illuminates the usefulness of Theorem 6. To evaluate the asymptotic probability of any strategy block, it suffices to evaluate its intersection with the neighbourhood of each minimal CURB block. Below, the upper bound 1/2 for δ prevents two neighbourhoods to overlap.

Corollary 7. *Let C_j, $j = 1, \ldots, J$, be the minimal CURB blocks of the game. For any measurable set $\sigma \subset \square(S)$ and for any $0 < \delta < 1/2$*

$$\lim_{\varepsilon \to 0} \mu_\varepsilon^*(\sigma) = \lim_{\varepsilon \to 0} \sum_{j=1}^{J} \mu_\varepsilon^*\left(B_\delta(C_j) \cap \sigma\right).$$

E.3.2.2 Concentration around approximate Nash equilibria

This section is devoted to proving that the properties of the state process implies the concentration of its distribution around Nash equilibria in any two player game with only 1×1 and 2×2 minimal CURB blocks, each containing a single Nash equilibrium. The 1×1-case follows by the previous section, so our focus here will be on the 2×2-case.

Assumption 1. *Each minimal CURB block C_j, $j = 1, \ldots, J$, of the game G is a 2×2 block and contains exactly one completely mixed Nash equilibrium N_j^*.*

Denote by $n_j^* = (n_{j,1}^*, n_{j,2}^*) \in \square(C_j)$ the fixed point to the expected best reply

$$\mathbb{E}\left[\widetilde{BR}_i(n_{j,-i}^*)\right] = n_{j,i}^*, \quad i = 1, 2, \; j = 1, \ldots, J, \tag{6}$$

which exists and is (minimal CURB-wise) unique by Lemma 13 in the appendix. Denote the unique fixed point of (6) when the players draw k samples by $n_j^*\{k\}$. Then, as is proven in Lemma 14, for large k it holds when $\varepsilon > 0$ that $\|N_j^* - n_j^*\{k\}\| < 1/\sqrt{k'}$ where $k' \in \mathbb{N}$ is greater than a bound that depends only on ε.

The following theorem is the concentration result for our state process on two-player games satisfying assumption 1.

Theorem 8. *Let the game satisfy assumption 1, let n_j^* be the fixed point to the expected pure best reply system* (6) *in the minimal CURB block C_j, and let $(p(t))_t$ be the state process with recency parameter*

$$\beta \in \left(\max\left\{1 - \frac{1}{\max\{m_1, m_2\}}, \bar{\beta}\right\}, 1\right),$$

where the lower bound $\bar{\beta} \in (0,1)$ depends only on k and n_j^* and is defined in (26)*. Then, for all $\eta > 0$, it holds that*

$$\lim_{\varepsilon \to 0} \lim_{t \to \infty} \mathbb{P}\left(\min_j \|p(t) - n_j^*\| < \eta\right) = 1 - O(1-\beta),$$

Proof. We know from Theorem 6 that $\lim_{\varepsilon \to 0} \mu^*(B_\delta(\mathcal{C})) \to 1$. Take any point ρ in $B_\delta(C_j)$ for some j and let $p(\cdot)$ be a state process with $p(t) = \rho$. Denote by $\tau_\varepsilon^*(t)$ the time since either a strategy outside C_j was sampled or an ε-tremble happened, counting backward from the current time t. This means that for the $\tau_\varepsilon^*(t)$ periods $\{t - \tau_\varepsilon^*(t), \ldots, t\}$ the state process has behaved as a process with $\varepsilon = 0$ on only the subgame given by C_j. Let furthermore n denote the mean process started at $p(t - \tau_\varepsilon^*(t))$,

$$\begin{cases} n_i(t+1) = \beta n_i(t) + (1-\beta)\mathbb{E}\left[\widetilde{BR}_i(n_{-i}(t))\right], \\ n_i(t - \tau_\varepsilon^*(t)) = p_i(t - \tau_\varepsilon^*(t)), \quad i = 1, 2, \quad s > 0. \end{cases}$$

Let Θ be the event $\Theta := \{\|p(t) - n_j^*\| \geq \eta + \|n_j(t) - n_j^*\|\}$. We have

$$\begin{aligned}
\mathbb{P}(\Theta) &= \sum_{r=0}^{R} \mathbb{P}\left(\Theta \mid \tau_\varepsilon^*(t) = r\right) \mathbb{P}\left(\tau_\varepsilon^*(t) = r\right) \\
&\quad + \mathbb{P}\left(\Theta \mid \tau_\varepsilon^*(t) > R\right) \mathbb{P}\left(\tau_\varepsilon^*(t) > R\right) \\
&= \sum_{r=0}^{R} \mathbb{P}\left(\Theta \mid \tau_\varepsilon^*(t) = r\right) \\
&\qquad \times \Big\{ \mathbb{P}\left(\tau_\varepsilon^*(t) = r \mid p(t-r) \in B_\delta(\mathcal{C})\right) \mathbb{P}\left(p(t-r) \in B_\delta(\mathcal{C})\right) \\
&\qquad + \mathbb{P}\left(\tau_\varepsilon^*(t) = r \mid p(t-r) \in B_\delta(\mathcal{C})^c\right) \mathbb{P}\left(p(t-r) \in B_\delta(\mathcal{C})^c\right) \Big\} \\
&\quad + \mathbb{P}\left(\Theta \mid \tau_\varepsilon^*(t) > R\right) \mathbb{P}\left(\tau_\varepsilon^*(t) > R\right).
\end{aligned} \tag{7}$$

Take now the limit $t \to \infty$ of the sum,

$$\begin{aligned}
&\lim_{t\to\infty} \sum_{r=0}^{R} \mathbb{P}\left(\Theta \mid \tau_\varepsilon^*(t) = r\right) \\
&\qquad \times \Big\{ \mathbb{P}\left(\tau_\varepsilon^*(t) = r \mid p(t-r) \in B_\delta(\mathcal{C})\right) \mathbb{P}\left(p(t-r) \in B_\delta(\mathcal{C})\right) \\
&\qquad + \mathbb{P}\left(\tau_\varepsilon^*(t) = r \mid p(t-r) \in B_\delta(\mathcal{C})^c\right) \mathbb{P}\left(p(t-r) \in B_\delta(\mathcal{C})^c\right) \Big\} \\
&= \sum_{r=0}^{R} \mathbb{P}\left(\Theta \mid \tau_\varepsilon^* = r\right) \Big\{ \mathbb{P}\left(\tau_\varepsilon^* = r \mid p(0) \in B_\delta(\mathcal{C})\right) \mu_\varepsilon^*\left(B_\delta(\mathcal{C})\right) \\
&\qquad\qquad + \mathbb{P}\left(\tau_\varepsilon^* = r \mid p(0) \in B_\delta(\mathcal{C})^c\right) \mu_\varepsilon^*\left(B_\delta(\mathcal{C})^c\right) \Big\}.
\end{aligned}$$

Later, we will take $\varepsilon \to 0$ and the term on the last line of the expression above will vanish (by Theorem 6). Therefore, we continue to study only the first term

$$\sum_{r=0}^{R} \mathbb{P}\left(\Theta \mid \tau_\varepsilon^* = r\right) \mathbb{P}\left(\tau_\varepsilon^* = r \mid p(0) \in B_\delta(\mathcal{C})\right). \tag{8}$$

Lemma 15 together with the triangle inequality and Proposition 17 says that for any $\eta > 0$,

$$\begin{aligned}
\mathbb{P}\left(\Theta \mid \tau_\varepsilon^* = r\right) &\leq \mathbb{P}\left(\|p(t) - n_j(t)\| \geq \eta \mid \tau_\varepsilon^*(t) = r\right) \\
&\leq \frac{c\beta\left(\beta^r + (1-\beta)\right)}{\eta^2}.
\end{aligned}$$

The probability that the process has behaved as a 2×2 minimal CURB game for the past r time periods is

$$
\begin{aligned}
&\mathbb{P}\left(\tau_\varepsilon^* = r \mid p(0) \in B_\delta(\mathcal{C})\right) \\
&= \mathbb{P}\left(\tau_\varepsilon^* > r-1 \mid p(0) \in B_\delta(\mathcal{C})\right) - \mathbb{P}\left(\tau_\varepsilon^* > r \mid p(0) \in B_\delta(\mathcal{C})\right) \\
&= Q_\varepsilon(r-1) - Q_\varepsilon(r) \\
&= \Big(1 - (1-\beta^r\delta)^{2k}(1-\varepsilon)^2\Big) \prod_{t=0}^{r-1} (1-\beta^t\delta)^{2k}(1-\varepsilon)^2 \\
&= \Big(1 - (1-\beta^r\delta)^{2k}(1-\varepsilon)^2\Big)\, (1-\varepsilon)^{2(r-1)} Q_o(r-1) \\
&\leq \left(\frac{1-(1-\varepsilon)^2}{(1-\varepsilon)^2} + 2k\beta^r\delta\right)(1-\varepsilon)^{2r}.
\end{aligned}
$$

The last inequality follows from examining the remainder in the first order Taylor expansion of $(1-\beta^r\delta)^{2k}$ around $\beta^r\delta = 0$. The remainder is

$$
R_1(\beta^r\delta) = \frac{2k(2k-1)(1-c)^{2k-2}}{2!}(\beta^r\delta)^2
$$

for some $c \in [0, \beta^r\delta]$. The remainder is non-negative, $R_1(\beta^r\delta) \geq 0$, for all admissible values of β, δ, and c. Combining the estimates above, we can bound (8),

$$
\begin{aligned}
&\sum_{r=0}^{R} \mathbb{P}\left(\Theta \mid \tau_\varepsilon^* = r\right) \mathbb{P}\left(\tau_\varepsilon^* = r \mid p(0) \in B_\delta(\mathcal{C})\right) \\
&\leq \sum_{r=0}^{R} \left(\frac{c\beta\left(\beta^r + 1 - \beta\right)}{\eta^2}\right)(1-\varepsilon)^{2r}\left(\frac{1-(1-\varepsilon)^2}{(1-\varepsilon)^2} + 2k\beta^r\delta\right) \\
&= \frac{c\beta}{\eta^2}\Bigg\{\left(\frac{1-(1-\varepsilon)^2}{(1-\varepsilon)^2}\right)\sum_{r=0}^{R}\left(\beta^r + 1 - \beta\right)(1-\varepsilon)^{2r} \\
&\qquad\qquad + 2k\delta \sum_{r=0}^{R}\left(\beta^r + 1 - \beta\right)\beta^r(1-\varepsilon)^{2r}\Bigg\}.
\end{aligned}
$$

All the series are convergent, letting $R \to \infty$ we get

$$\begin{aligned}
&\lim_{R\to\infty} \sum_{r=0}^{R} \mathbb{P}\left(\Theta \mid \tau_\varepsilon^* = r\right) \mathbb{P}\left(\tau_\varepsilon^* = r \mid p(0) \in B_\delta(\mathcal{C})\right) \\
&\qquad \leq \frac{c\beta}{\eta^2} \left\{ \frac{1-(1-\varepsilon)^2}{(1-\varepsilon)^2(1-\beta(1-\varepsilon)^2)} + \frac{1-\beta}{(1-\varepsilon)^2} \right. \\
&\qquad\qquad \left. + 2k\delta \left(\frac{1}{1-\beta^2(1-\varepsilon)^2} + \frac{(1-\beta)}{1-\beta(1-\varepsilon)^2} \right) \right\}.
\end{aligned}$$

Sending $\varepsilon \to 0$, we get

$$\begin{aligned}
&\lim_{\varepsilon\to 0} \lim_{R\to\infty} \sum_{r=0}^{R} \mathbb{P}\left(\Theta \mid \tau_\varepsilon^* = r\right) \mathbb{P}\left(\tau_\varepsilon^* = r \mid p(0) \in B_\delta(\mathcal{C})\right) \\
&\qquad \leq \frac{c\beta}{\eta^2} \left\{ 1-\beta + 2k\delta \left(\frac{1}{1-\beta^2} + 1 \right) \right\} \\
&\qquad \leq \frac{c\beta}{\eta^2} \left\{ 1-\beta + 2k\delta \left(\frac{1}{1-\beta} + 1 \right) \right\}.
\end{aligned}$$

Finally, choosing $\delta = (1-\beta)^2/k$, we get

$$\begin{aligned}
&\lim_{\varepsilon\to 0} \lim_{R\to\infty} \sum_{r=0}^{R} \mathbb{P}\left(\Theta \mid \tau_\varepsilon^* = r\right) \mathbb{P}\left(\tau_\varepsilon^* = r \mid p(0) \in B_{(1-\beta)^2/k}(\mathcal{C})\right) \\
&\qquad \leq \frac{5c\beta}{\eta^2}(1-\beta).
\end{aligned}$$

The term $\mathbb{P}(\tau_\varepsilon^* > R) = 0$ vanishes as $R \to \infty$ by the same analysis that showed that $Q^* = 0$ when $\varepsilon \neq 0$, cf. (5). Going all the way back to (7) and plugging in our estimates yields

$$\lim_{\varepsilon\to 0} \mu_\varepsilon^*(\Theta) \leq \frac{5c\beta}{\eta^2}(1-\beta),$$

and hence

$$\lim_{\varepsilon\to 0} \mu_\varepsilon^* \left(\{ p \in \square(S) \mid \min_j \|p - n_j^*\| < \eta \} \right) \geq 1 - O(1-\beta)$$

where the ordo O is uniform. $\square$

Note that if the whole game is a 2×2-minimal CURB block, δ cannot be defined. However, in that case Proposition 17 right away yields a bound of order $1 - \beta$.

E.4 Conclusions and outlook

We conclude the paper with some reflections. The model with recency bias proposed in this paper enjoys many of the theoretical properties of the original model for evolution of social conventions, as shown by Theorem 4 and Theorem 6. This paper then goes beyond what is known for the original model when it comes to behavior within stable blocks of strategies, so-called minimal CURB blocks. Theorem 8 provides conditions under which the model with recency bias converges to Nash equilibria inside minimal CURB blocks. To get to these results we choose to work in a continuous-state space framework. We regard the history as a distribution over past plays (with recency bias) instead a vector of the exact past history of plays and this requires us to use other theoretical tools. However, the complication is only mathematical, we have not observed any modeling artifacts in our simulations.

As a final note, we formulate a conjecture: Theorem 10 together with convergence properties of the state process yields concentration of the invariant distribution around Nash equilibria in any two player game. That is, we expect that Assumption 1 can be relaxed.

Conjecture 9. *Assume that the process defined by the restriction of the game G to the subgame C_j, which is a minimal CURB block of any size, has a unique completely mixed Nash equilibrium N_j^*. Let $\mu_\varepsilon^*|_{C_j}$ be the invariant distribution of the state process on the subgame. If it holds for all $\eta > 0$ that*

$$\lim_{\varepsilon \to 0} \mu_\varepsilon^*\Big|_{C_j} \big(p \in \square(C_j) \mid \|p - n_j^*\| < \eta\big) = 1 - O(1-\beta), \tag{9}$$

then

$$\lim_{\varepsilon \to 0} \mu_\varepsilon^* \left(p \in \square(S) \;\middle|\; \min_j \|p - n_j^*\| < \eta\right) = 1 - O(1-\beta).$$

We believe the conjecture to be true and the reason is the following argument. The set of probability measures is a compact space when equipped with the Wasserstein metric, so instead of looking at $\lim_{\varepsilon \to 0} \mu_\varepsilon^*$ we may consider the compact set of accumulation points for μ^*_ε, and take μ^*_0 in that set which must be an invariant distribution (not necessarily unique) for the game with no deficiencies, i.e. with $\varepsilon = 0$. But without deficiencies, once the position is close to a minimal CURB block, the probability of ever getting out is a geometrically decaying series so by Borel-Cantelli almost surely the process is eventually stuck in the minimal CURB block. Then, we apply the hypothesis on the restricted game.

Further directions of research, after proving or disproving Conjecture 9, include proving the estimate (9) for minimal CURB blocks of any size and extending the analysis to games with minimal CURB blocks containing multiple Nash equilibria. Another interesting direction would be to replicate Young's analysis on which minimal CURB block or blocks the state process concentrates on.

E.5 Proofs for $m_1 \times m_2$-games

E.5.1 Exponential history

Let us prove Proposition 1. Starting from the definition, we have

$$p_{-i,s}\,(t+1) = (1-\beta) \sum_{\tau=1}^{\infty} \beta^{\tau-1} 1_s(s_{-i}(t-\tau+1)).$$

After index substitution $v = \tau - 1$, splitting the term $v = 0$ yields

$$p_{-i,s}\,(t+1) = (1-\beta) \left(1_s(s_{-i}(t)) + \sum_{v=1}^{\infty} \beta^{v} 1_s(s_{-i}(t-v)) \right).$$

In other words,

$$p_{-i,s}\,(t+1) = (1-\beta)\,\beta \sum_{v=1}^{\infty} \beta^{v-1} 1_s(s_{-i}(t-v)) + (1-\beta)\, 1_s(s_{-i}(t)).$$

We recognize the first term as $p_{-i,s}\,(t)$, so we are left for every $s \in S_{-i}$ with

$$p_{-i,s}\,(t+1) = \beta p_{-i,s}\,(t) + (1-\beta)\, 1_s(s_{-i}(t)),$$

which is the representation we seek.

E.5.2 Lipschitz continuity

Here we are going to prove the Proposition 2. At the beginning there is a sample with respect to probabilities $p := p_{-i}(t)$, yielding a random vector $N := \left(n_{-i,1}(t), \ldots n_{-i,m_{-i}}(t)\right)$ of integers from the (discrete) probability distribution

$$\mathbb{P}\left(N = \left(n_1, \ldots n_{m_{-i}}\right)\right) = k! \prod_{j=1}^{m_{-i}} \frac{p_s^{n_s}}{n_s!}.$$

Each N will lead to an empirical opposing strategy profile D, that must belong to some finite 'simplex grid'

$$\Delta^{(-i,k)} := \left\{ \frac{1}{k} \sum_{s \in S_{-i}} n_s \overrightarrow{1_{-i,s}} \; ; \; n_s \in \mathbb{N}_0, \sum_{s \in S_{-i}} n_s = k \right\}.$$

Now let us form m_i subsets from $\Delta^{(-i,k)}$ (which is finite), named $\Delta_s^{(-i,k)}$ for $s \in S_i$, where $x \in \Delta_s^{(-i,k)}$ whenever $s \in BR_i(x)$. Note that $(\Delta_s^{(-i,k)})_s$ is not a disjoint cover of $\Delta^{(-i,k)}$ except in the special case when each $x \in \Delta^{(-i,k)}$ has a unique best response. Also, $\cup_s \Delta_s^{(-i,k)} = \Delta^{(-i,k)}$ since the best response set is never empty.

For $a \leq m_i$, the probability that $\widetilde{BR}_i(p) = a$ is going to be played is thus obtained as follows :

- When deficiency occurs, which happens a fraction ε of the time, a is played with a probability $1/m_i$, totalling ε/m_i.
- When deficiency does not occur, which happens a fraction $1 - \varepsilon$ of the time, the player selects its best response, so it will be a with the probability $\mathbb{P}\left(D \in \Delta_a^{(-i,k)}, \widehat{BR}_i(D) = a\right)$.

In short,

$$\mathbb{P}\left(\widetilde{BR}_i(p) = a\right) = \varepsilon/m_i + (1-\varepsilon) \sum_{x \in \Delta_a^{(-i,k)}} \mathbb{P}\left(\widehat{BR}_i(x) = a\right) \mathbb{P}(D = x). \quad (10)$$

However $D = x$ is an event of the shape $N = \left(n_1, \ldots n_{m_{-i}}\right)$, so considering $\mathbb{P}(D = x)$ as a function of $p_1, \ldots p_n$, we get

$$\frac{\partial \mathbb{P}\left(N = \left(n_1, \ldots n_{m_{-i}}\right)\right)}{\partial p_b} = k! \frac{p_b^{n_b - 1}}{(n_b - 1)!} \prod_{j \neq b} \frac{p_j^{n_j}}{n_j!},$$

with the convention $1/(-1)! = 0$ for continuity. So relatively to the norm $\|.\|_\infty$ over $\Delta(S_{-i})$, the Lipschitz constant of the probabilities $\mathbb{P}\left(D \in \Delta_a^{(-i,k)}\right)$ are at most

$$\sum_{b=1}^{m_{-i}} \left| \frac{\partial \mathbb{P}\left(D \in \Delta_a^{(-i,k)}\right)}{\partial p_b} \right| \leq \sum_{b=1}^{m_{-i}} \sum_{x \in \Delta_a^{(-i,k)}} k! \frac{p_b^{n_b - 1}}{(n_b - 1)!} \prod_{\substack{j=1 \\ j \neq b}}^{m_{-i}} \frac{p_j^{n_j}}{n_j!}.$$

However we know that

$$\sum_{x \in \Delta^{(-i,k)}} \frac{p_b^{n_b - 1}}{(n_b - 1)!} \prod_{\substack{j=1 \\ j \neq b}}^{m_{-i}} \frac{p_j^{n_j}}{n_j!} = \frac{1}{(k-1)!},$$

as this is the multinomial formula for $k-1$ draws. Since $\Delta_a^{(-i,k)} \subset \Delta^{(-i,k)}$, the Lipschitz constant of $\mathbb{P}\left(D \in \Delta_a^{(-i,k)}\right)$ is at most

$$\sum_{b=1}^{m_{-i}} k! \frac{1}{(k-1)!} = k m_{-i}.$$

Bounding $\mathbb{P}\left(\widehat{BR}_i(x) = a\right)$ from (10) by 1, the Lipschitz constant for

$$p \mapsto \mathbb{P}\left(\widetilde{BR}_i(p) = a\right)$$

is at most $(1 - \varepsilon) k m_{-i}$.

E.5.3 Elementary properties of K

Now let us prove the Proposition 3. Knowing that by the next time step, the players will have chosen strategies $(s_1, s_2) \in S$, let us denote

$$\Gamma\left((p_1, p_2), (s_1, s_2)\right) := \left(\beta p_1 + (1 - \beta) \overrightarrow{1_{1,s_1}}, \beta p_2 + (1 - \beta) \overrightarrow{1_{2,s_2}}\right)$$

the next state of the history, $(p_1(t+1), p_2(t+1))$, when s is played from $p_i(t) = p_i$. In particular, as a function of (p_1, p_2), Γ is Lipschitz of constant $\beta < 1$. Also, let us call

$$\sigma\left((p_1, p_2), (s_1, s_2)\right) := \mathbb{P}\left(\widetilde{BR}_1(p_2) = s_1, \widetilde{BR}_2(p_1) = s_2\right)$$

the probability of getting to $\Gamma((p_1,p_2),(s_1,s_2))$ from (p_1,p_2). Thanks to Proposition 2 and the Lipschitz constant of $\mathbb{P}\left(\widetilde{BR}_i(\cdot)=s\right)$ being at most $(1-\varepsilon)\,km_{-i}$, the Lipschitz constant of σ seen as a function of (p_1,p_2) with (s_1,s_2) constant is at most

$$(1-\varepsilon)\,k\,(m_1+m_2)\,. \tag{11}$$

For $p=(p_1,p_2)\in A$ and $s=(s_1,s_2)\in S$, we have thus

$$Kf : p \mapsto \sum_{s_1=1}^{m_1}\sum_{s_2=1}^{m_2}\sigma(p,s)\,f(\Gamma(p,s))\,.$$

Since f is continuous by hypothesis, continuity of Kf comes immediately. Therefore K operates from E to E.

Linearity is immediate when looking at the definition of K. Continuity follows from linearity because K is bounded; indeed, when $\|f\|_\infty \le 1$, since probabilities sum to 1 we also get $\|Kf\|_\infty \le 1$. K is positive, in the sense that when $f \ge 0$ everywhere, also $Kf \ge 0$ everywhere on $\square(S)$. Finally, noting by $\overrightarrow{1}$ the unit function, $K\overrightarrow{1}=\overrightarrow{1}$ for the same reason.

E.5.4 Dinetan's theorem for Lipschitz-friendly operators

Let $L \subset E := C(\square(S);\mathbb{R})$ be the subset of Lipschitz functions over $\square(S)$, endowed with the semi-norm

$$\|f\|_L := \sup_{p\neq p'}\left(\frac{|f(p)-f(p')|}{|p-p'|}\right).$$

Notice that L is a Banach space with the norm $\|f\|_Z := \max(\|f\|_\infty, \|f\|_L)$. The following theorem is an adaptation of the result of [9].

Theorem 10. *Let K be a linear, continuous, stochastic, Lipschitz-friendly and infra-ergodic operator from E to E. Provided that K has no witness set for any class, there are $c\in\mathbb{R}_+$, $\lambda\in(0,1)$, and a unique probability measure $\mu\in\mathcal{P}(\square(S))$ invariant with respect to K such that for every $n\in\mathbb{N}$ and $f\in L$,*

$$\|K^n f - \overrightarrow{1}\mu f\|_\infty \le c\lambda^n\|f\|_Z \tag{12}$$

As pointed out in [9], the natural norm on the topological dual of L with the norm $\|\cdot\|_Z$ is the Wasserstein norm $\|\cdot\|_W$. Restricted to probability measures, the Wasserstein norm translates to the topology of weak convergence and this justifies the rewriting of (12) to the more easily interpreted (4) in Theorem 4.

To prove the existence of a unique invariant probability measure, we need to prove the different properties required in Theorem 10. Proofs of the properties are found below. We want to emphasize that we only need some quite general properties of our process to justify the result. So checking the result for other processes with the same state space should be easy. Specifically, we need a lower bound for the probability for any pure strategy to be played, which e.g. a logit best reply function has for a finite game, and that the best reply function has certain Lipschitz continuity properties. So any stochastic process with the same state space and the same updating conditional on the strategies being played has a unique invariant distribution if for all $s_i \in S_i$ and $i \in \{1, 2\}$ it holds that

$$\min_{p_{-i} \in \Delta(S_{-i})} \mathbb{P}\left(\widetilde{BR}_i(p_{-i}) = s_i\right) > \varepsilon$$

for some constant $\varepsilon > 0$ and $\mathbb{P}\left(\widetilde{BR}_i(\cdot) = s_i\right)$ is Lipschitz continuous.

The continuity, stochasticity, and linearity have already been proved. The infra-ergodicity is proved in Section E.5.6 and comes from the fact that every pure strategy has positive probability of being played in every period, so it is possible to get from any state p to any open subset of $\square(S)$ in finite time. The non-existence of witness-sets is true since the state space $\square(S)$ is connected.

Remark 11. *An operator from E to E is compact if it sends the unit ball in E to a compact set. Viewing K as an operator on E, as in* (3)*, it is a non-compact operator. Indeed, K sends the unit ball in E to a squeezed unit ball in E, which is non-compact. This is actually the case for all operators from E to E containing integration with respect to point masses. The non-compactness of K is the reason the standard version of the Krein-Rutman theorem cannot be applied, and why we apply Dinetan's version.*

E.5.5 K is Lipschitz-friendly

We recall that K is Lipschitz-friendly when there are $\lambda^+ > \lambda^- \in \mathbb{R}^+$ such that for every $f \in E^+$ (non-negative continuous functions on $\square(S)$) satisfying

$\|f\|_L \leq \lambda^+$ and $\|f\|_\infty \leq 1$, it holds that $\|Kf\|_L \leq \lambda^-$ and $\|Kf\|_\infty \leq 1$. The norm $\|\cdot\|_L$ is defined in Section E.5.4.

Let $f \in E^+$ hold these properties, let $p, p' \in \square(S)$, and consider the difference

$$Kf\left(p'\right) - Kf\left(p\right) = \sum_{s \in S} \sigma\left(p', s\right) f\left(\Gamma\left(p', s\right)\right) - \sigma\left(p, s\right) f\left(\Gamma\left(p, s\right)\right). \tag{13}$$

Here, let us use the identity

$$ab - cd = \frac{1}{2}\left(a + c\right)\left(b - d\right) + \frac{1}{2}\left(a - c\right)\left(b + d\right)$$

to transform the summand in (13) into

$$\begin{aligned} &\frac{1}{2}\left(\sigma\left(p', s\right) + \sigma\left(p, s\right)\right)\left(f\left(\Gamma\left(p', s\right)\right) - f\left(\Gamma\left(p, s\right)\right)\right) \\ &\qquad + \frac{1}{2}\left(\sigma\left(p', s\right) - \sigma\left(p, s\right)\right)\left(f\left(\Gamma\left(p', s\right)\right) + f\left(\Gamma\left(p, s\right)\right)\right). \end{aligned} \tag{14}$$

On the top row of (14), from Γ Lipschitz of constant β, we have

$$\left|f\left(\Gamma\left(p', s\right)\right) - f\left(\Gamma\left(p, s\right)\right)\right| \leq \left\|p' - p\right\|_\infty \beta \|f\|_L.$$

Therefore we get

$$\begin{aligned} &\left|\sum_{s \in S}\left(\sigma\left(p', s\right) + \sigma\left(p, s\right)\right)\left(f\left(\Gamma\left(p', s\right)\right) - f\left(\Gamma\left(p, s\right)\right)\right)\right| \\ &\leq \sum_{s \in S}\left(\sigma\left(p', s\right) + \sigma\left(p, s\right)\right)\left\|p' - p\right\|_\infty \beta \|f\|_L \\ &= 2\beta \|f\|_L \left\|p' - p\right\|_\infty. \end{aligned}$$

The last equality holds since the sum is taken over a disjoint partition of the space of outcome. From (11), we know that σ is Lipschitz continuous in its first argument. To simplify notation, denote its Lipschitz constant by L_σ. For the bottom row of (14) we then have the estimate

$$\left|\sigma\left(p', s\right) - \sigma\left(p, s\right)\right| \leq L_\sigma \left\|p' - p\right\|_\infty,$$

and we get

$$\begin{aligned}&\left|\sum_{s\in S}\left(\sigma\left(p',s\right)-\sigma\left(p,s\right)\right)\left(f\left(\Gamma\left(p',s\right)\right)+f\left(\Gamma\left(p,s\right)\right)\right)\right|\\&\leq\sum_{s\in S}L_\sigma\left\|p'-p\right\|_\infty 2\left\|f\right\|_\infty\\&\leq 2L_\sigma\left\|f\right\|_\infty m_1m_2\left\|p'-p\right\|_\infty .\end{aligned}$$

Let us denote $c := \max\{1, m_1m_2L_\sigma\}$ so that in the end we have

$$\left|Kf\left(p'\right)-Kf\left(p\right)\right|\leq\left(\beta\|f\|_L+c\left\|f\right\|_\infty\right)\left\|p'-p\right\|_\infty .$$

Therefore, whenever $\|f\|_\infty \leq 1$ and

$$\|f\|_L\leq\lambda^+ := \frac{2c}{1-\beta},$$

the Lipschitz constant of Kf will be at most $\beta\lambda^+ + c =: \lambda^- < \lambda^+$ by choice of λ^+. It follows that f is Lipschitz-friendly with these λ^+ and λ^-.

E.5.6 K is infra-ergodic

Let $f \in E^+ \setminus \{0\}$ and $p \in \square(S)$. We will prove that there is $n \in \mathbb{N}$ such that $(K^n f)(p) > 0$, implying infra-ergodicity of K.

For $i \in \{1, 2\}$, $j \leq m_{-i}$, and $t \in \mathbb{N}$, let $\omega_{-i,j,t} := 1_j(s_{-i}(t))$ be the indicator of a play j by the opponent of player i at time t, so that

$$p_{-i,j}(t) = (1-\beta)\sum_{\tau=1}^{\infty}\beta^{\tau-1}\omega_{-i,j,t-\tau}.$$

We will call $\Sigma^{(-i)} := \{0,1\}^{m_{-i}\times\mathbb{N}}$ the set of binary arrays, indexed by $s \leq m_{-i}$ and $t \in \mathbb{N}$, such that for every t there is exactly one s such that $\Sigma^{(-i)}_{s,t} = 1$. In other words, $\Sigma^{(-i)}$ represents the history for player $-i$, where a 1 at the entry (s,t) indicates that s was played at time t. Likewise, for $n \in \mathbb{N}$, we will call $\Sigma^{(-i,N)} := \{0,1\}^{m_{-i}\times N}$ the set of binary arrays indexed by $s \leq m_{-i}$ and $t \in \{1, \dots N\}$ obeying the same condition, in other words the history up to time N.

E.5.6.1 Approximative history

Let $p_{-i} \in \Delta(S_{-i})$. We are going to exhibit a sequence of plays of finite length N, i.e. a $\omega \in \Sigma^{(-i,N)}$ for some $N \in \mathbb{N}$, such that the partial sums

$$p_{-i,j}^{(N)} := (1-\beta)\sum_{t=1}^{N} \beta^{t-1}\omega_{-i,j,t}$$

fall close to p_{-i}. Namely, we want to prove the following.

Lemma 12. *Let $p_{-i} \in \Delta(S_{-i})$ and $\varepsilon > 0$. We assume that $(1-\beta)\, m_{-i} \leq 1$. There are $N_{-i} \in \mathbb{N}$ and $\omega_{-i} \in \Sigma^{(-i,N)}$ such that for every $j \leq m_{-i}$ and $N \geq N_{-i}$,*

$$p_{-i,j}^{(N)} = (1-\beta)\sum_{t=1}^{N} \beta^{t-1}\omega_{-i,j,t} \in \left[p_{-i,j} - \varepsilon, p_{-i,j}\right].$$

Proof. The following algorithm provides a proof of Lemma 12. Start by setting $p_{-i,j}^{(0)} = 0$ for every j, and ω_{-i} to the empty array of dimensions 0 and m_{-i}. Define N_{-i} as the smallest $N \in \mathbb{N}$ such that $\beta^N < \varepsilon$. For $t \in \{1, \ldots N_{-i}\}$, repeat the following steps :

1. Look for $j \leq m_{-i}$ such that $p_{-i,j} - p_{-i,j}^{(t-1)}$ is maximal, and call any of these pure strategies a.
2. Append at the end of the array ω_{-i} the values $\omega_{-i,a,t} = 1$ and $\omega_{-i,j,t} = 0$ for $j \neq a$.
3. Compute $p_{-i,j}^{(t)}$ accordingly to the updated history ω_{-i}.

Return the final history ω_{-i} and values $p_{-i,j}^{(N)}$. □

We are going to prove recursively that for every $t \in \mathbb{N}$, we always have $p_{-i,j}^{(t)} \leq p_{-i,j}$ and

$$\sum_{j=1}^{m_{-i}} p_{-i,j}^{(t)} = 1 - \beta^t.$$

For $t = 0$ this is because $p_{-i,j}$ is non-negative and $p_{-i,j}^{(0)} = 0$. Assuming these at time t, since the probabilities $p_{-i,j}$ add to 1, the maximum value among

$p_{-i,j} - p^{(t)}_{-i,j}$ must be at least β^t / m_{-i}. By definition, then $\omega_{-i,a,t+1} = 1$ and

$$p^{(t+1)}_{-i,a} = p^{(t)}_{-i,a} + (1-\beta)\,\beta^t \leq p_{-i,j} - \frac{\beta^t}{m_{-i}} + (1-\beta)\,\beta^t.$$

Therefore, since $(1-\beta)\,m_{-i} \leq 1$ as assumed, the right-hand side is also bounded by $p_{-i,j}$. As for other strategies $j \neq a$, since $p^{(t+1)}_{-i,j} = p^{(t)}_{-i,j}$ the inequality $p^{(t)}_{-i,j} \leq p_{-i,j}$ still holds. Now we also know that

$$p^{(t+1)}_{-i,j} - p^{(t)}_{-i,j} = (1-\beta)\,\beta^t \omega_{-i,j,t+1},$$

and since exactly one among the m_{-i} bits $\omega_{-i,j,t+1}$ is 1, we have

$$\sum_{j=1}^{m_{-i}} \left(p^{(t+1)}_{-i,j} - p^{(t)}_{-i,j} \right) = (1-\beta)\,\beta^t.$$

The recursion hypothesis thus leads us to

$$\sum_{j=1}^{m_{-i}} p^{(t+1)}_{-i,j} = 1 - \beta^t + (1-\beta)\,\beta^t = 1 - \beta^{t+1},$$

which proves our properties by recursion. So in particular after time N_{-i}, by choice of N_{-i}, for every $N \geq N_{-i}$ we have

$$\sum_{j=1}^{m_{-i}} p^{(N)}_{-i,j} > 1 - \varepsilon,$$

while $p^{(N)}_{-i,j} \leq p_{-i,j}$ for every j. Since $\sum_j p_{-i,j} = 1$, this is possible only if each $p^{(N)}_{-i,j}$ is at least $p_{-i,j} - \varepsilon$, leading to the result.

E.5.6.2 Proof of ergodicity

Now let us look again at our $f \in E^+ \setminus \{0\}$ and $p \in \square\,(S)$ from the beginning of the proof. Since f is continuous, the set $f^{-1}\left((0,\infty)\right)$ is open and nonempty, therefore contains an open ball B centered on $x \in \square\,(S)$ of radius $\varepsilon > 0$.

Let us write $x = (x_1, x_2)$ for $x_i \in \Delta(S_i)$. We apply Lemma 12 to x_1 and to x_2, yielding play records ω_1 and ω_2, and values $p_{-i,j}^{(N_{-i})}$ such that for every $j \leq m_{-i}$ and $N \geq N_{-i}$,

$$p_{-i,j}^{(N)} \in [p_{-i,j} - \varepsilon, p_{-i,j}] .$$

So let us call $N = \max(N_1, N_2)$ and look for $\left(K^N f\right)(x)$. As is known,

$$(Kf)(x) = \sum_{s_1=1}^{m_1} \sum_{s_2=1}^{m_2} \sigma(x, s)\, f(\Gamma(x, s))$$

is a convex combination of values of f at the points $\Gamma(x, s) \in \square(S)$ accessible from x in the history dynamics, once $s = (s_1, s_2)$ is played. In particular, the probabilities $\sigma(x, s)$ are bounded from below by $\eta = \varepsilon/(m_1 m_2) > 0$.

The play records ω_1 and ω_2 up to time N are now read in reverse time order. At each time step $t \in \{0, \dots N-1\}$, there is a probability at least η^2 that player 1 chooses the strategy $a \leq m_1$ given by $\omega_{1,a,N-t} = 1$, and player 2 chooses the strategy $b \leq m_2$ given by $\omega_{2,b,N-t} = 1$. Therefore the plays up to time N have a probability at least $\eta^{2N} > 0$ of being dictated by ω_1 and ω_2. When this happens, thanks to the Proposition 1, a history having started by $(p_1(0), p_2(0)) = p$ will now be at the position

$$\begin{aligned}(p_1(N), p_2(N)) \quad &= \quad \sum_{t=1}^{N} \left((1-\beta)\, \beta^{t-1} \overrightarrow{1_{1,s_1(N-t)}}, (1-\beta)\, \beta^{t-1} \overrightarrow{1_{2,s_2(N-t)}} \right) \\ &\quad + \left(\beta^N p_1(0), \beta^N p_2(0)\right).\end{aligned}$$

Since $s_i(N-t)$ is then the strategy $a \leq m_i$ given by $\omega_{i,a,t} = 1$, in fact we have

$$p_{-i,j}(N) = \sum_{t=1}^{N} (1-\beta)\, \beta^{t-1} \omega_{-i,j,t} + \beta^N p_{i,j}(0). \tag{17}$$

By Lemma 12, the choice of the records ω_1 and ω_2 makes the sum (17) to take a value between $x_1 - \varepsilon$ and x_1. As we also have $\beta^N < \varepsilon$ and $p_{i,j}(0) \leq 1$, we get $p_{i,j}(N) \in (p_{i,j} \pm \varepsilon)$, thus

$$\|(p_1(N), p_2(N)) - (x_1, x_2)\| < \varepsilon.$$

In other words, it means that this point $y = (p_1(N), p_2(N)) \in B$ is accessible from p in N steps. It follows that in the convex combination of $\left(K^N f\right)(p)$, there is the therm $f(y)$ with a positive coefficient, therefore $\left(K^N f\right)(p) > 0$. This holding for every p and f shows that K infra-ergodic.

E.6 Proofs for 2×2 minimal CURB blocks

Before turning to the proof of Theorem 8 we present three lemmas. Consider the following version of Assumption 1 for a single 2×2 minimal CURB block.

Assumption 2. *$m_1 = m_2 = 2$ and the whole game is a minimal CURB block, containing a unique Nash equilibrium N^* that is completely mixed.*

Generically, all 2×2 games without pure Nash equilibria must have the basic Matching Pennies structure. One player will be 'agreeing' and the other 'disagreeing' in the sense that the best reply of the agreeing player is to play the same strategy (0 or 1) as the disagreeing player. On the other hand, the disagreeing player's best reply is to not play the same strategy as the agreeing player. Any other situation will generically yield at least one minimal CURB block of size 1×1, contradicting the assumption. For the rest of this section, we will refer to the player 1 and 2 as the agreeing and the disagreeing player, respectively.

The following lemma is a consequence of the uniqueness of N^*.

Lemma 13. *Let k (the number of samples) be an integer such that $N_1^* k \notin \mathbb{N}$ and $N_2^* k \notin \mathbb{N}$. Then there exists a unique fixed point $n^* = (n_1^*, n_2^*) \in \square(S)$ to the system*

$$\begin{cases} \mathbb{E}\left[\widetilde{BR}_1(n_2^*)\right] = n_1^*, \\ \mathbb{E}\left[\widetilde{BR}_2(n_1^*)\right] = n_2^*. \end{cases} \tag{18}$$

Proof. The Nash equilibrium $N^* = (N_1^*, N_2^*)$, which is unique by assumption, defines 'cut-off's $M_i := \lfloor N_i^* r_k \rfloor$, $i = 1, 2$, such that if more than M_1 of the agreeing player's k samples from the disagreeing player's history are 1, he plays 1. The disagreeing player will play strategy 1 if more than M_2 of his k samples from the agreeing player's history of plays are 0. Consider the function

$$p_{k,M}(x) := (1-\varepsilon) \sum_{i=M+1}^{k} \binom{k}{i} x^i (1-x)^{k-i} + \varepsilon/2. \tag{19}$$

Given that player history is in state (a, d), the probability that the agreeing and disagreeing player plays strategy 1 is $p_a(d) := p_{k,M_2}(d)$ and $p_d(a) := 1 - p_{k,M_1}(a)$, respectively. We can now rewrite (18) as

$$p_a(n_2^*) = n_1^*, \qquad p_d(n_1^*) = n_2^*.$$

The range of p_a and p_b is $I_\varepsilon := [\varepsilon/2, 1-\varepsilon/2]$. Therefore, by the strict monotonicity and the continuity of p_a and p_d, we may rewrite (18) again, now as

$$(p_a \circ p_d)(n_1^*) = n_1^*, \quad n_1^* \in I_\varepsilon,$$
$$(p_d \circ p_a)(n_2^*) = n_2^*, \quad n_2^* \in I_\varepsilon.$$

Note that since p_a and p_d are strictly increasing and decreasing, respectively, both $p_a \circ p_d$ and $p_d \circ p_a$ are strictly decreasing functions from $[0,1]$ to $[p_d(1-\varepsilon/2), p_d(\varepsilon/2)]$ and $[p_a(\varepsilon/2), p_a(1-\varepsilon/2)]$, respectively. Therefore

$$\min\{p_a \circ p_d(\varepsilon/2), p_d \circ p_a(\varepsilon/2)\} \geq \min\{p_d(1-\varepsilon/2), p_a(\varepsilon/2)\} > \varepsilon/2,$$

and

$$\begin{aligned}\max\{p_a \circ p_d(1-\varepsilon/2), p_d \circ p_a(1-\varepsilon/2)\}\\ \leq \max\{p_d(\varepsilon/2), p_a(1-\varepsilon/2)\} < 1-\varepsilon/2.\end{aligned}$$

Hence, since $p_a \circ p_d$ and $p_d \circ p_a$ are continuous, they intersect the straight line $x = y$ at a (function-wise) unique point in their respective images and these intersection points are n_1^* and n_2^*. □

The k- and ε-dependent fixed point n^* is in general not equal to the Nash equilibrium N^*. However, as the following lemma shows, it tends to N^* as $k \to \infty$ when ε is small enough.

Lemma 14. *Let $n^*\{k\}$ denote the fixed point from Lemma 13 when $k \in \mathbb{N}$ samples are drawn by each of the players. Assume that $\varepsilon < \max\{N_1^*, N_2^*, 1-N_1^*, 1-N_2^*\}$. Then there is a bound $\underline{k}(\varepsilon) \in \mathbb{N}$, depending only on ε, such that for all $k' \geq \underline{k}(\varepsilon)$, $|N_i^* - n_i^*\{k\}| < 1/\sqrt{k'}$ for all $k > k'$.*

Proof. The sample D_i from (1) is binomially distributed in the 2×2-game setting. To be more clear, we denote the sample drawn by player $-i$ from player i's history, when the history is in state $n_i^*\{k\}$, by

$$D_i^k := \frac{1}{k}\sum_{j=1}^{k} \eta_{i,jk}.$$

The draws $\eta_{i,jk}$ are $\mathrm{Ber}(n_i^*\{k\})$-distributed for $j = 1, \ldots, k$ and $k \in \mathbb{N}$, and $\eta_{i,1k}, \ldots, \eta_{i,kk}$ are independent. We denote the mean, variance, and skewness of $\eta_{i,jk}$ by $\mu_{i,k}$, $\sigma_{i,k}$, and $\rho_{i,k}$, respectively,

$$\mu_{i,k} = n_i^*\{k\}, \quad \sigma_{i,k}^2 = n_i^*\{k\}(1-n_i^*\{k\}), \quad \rho_{i,k} = \frac{1-2n_i^*\{k\}}{\sqrt{n_i^*\{k\}(1-n_i^*\{k\})}}.$$

By Lemma 13, $n_i^*\{k\} \in (\varepsilon/2, 1-\varepsilon/2)$ and hence the skewness is finite and the variance is bounded away from zero. For $i = 1, 2$, $j = 1, \ldots, k$, and $k \in \mathbb{N}$ we define the centralized draws $\bar{\eta}_{i,jk} := \eta_{i,jk} - \mu_{i,k}$. Consider $\bar{D}_i^k$, a centralized and scaled sample defined by the relation

$$D_i^k = \frac{\sigma_{i,k}}{\sqrt{k}} \bar{D}_i^k + \mu_{i,k}.$$

Denoting by $\bar{F}_{i,k}$ the distribution function of $\bar{D}_i^k$, the Berry-Essén theorem tells us that

$$|\bar{F}_{i,k}(x) + \mathcal{N}(x)| \leq 3\rho_{i,k}/\sigma_{i,k}^3\sqrt{k}$$

where $\mathcal{N}$ is the standard normal distribution function. With the bounds for the skewness and variance derived above, we have that $\bar{F}_{i,k}(x) = \mathcal{N}(x) + \bar{f}_i(k)$ where $\bar{f}_i \in O(1/(\varepsilon^2\sqrt{k}))$ and we can rewrite the system for $n^*\{k\}$ as

$$\begin{aligned}
n_1^*\{k\} &= \mathbb{E}\left[\widetilde{BR}_1(n_2^*\{k\})\right] \\
&= (1-\varepsilon)\left(\mathbb{P}(D_2^k > N_2^*) + \frac{1}{2}\mathbb{P}(D_2^k = N_2^*)\right) + \varepsilon/2 \\
&= (1-\varepsilon)\Big(1 - \mathbb{P}(\bar{D}_2^k \leq \sqrt{k}(N_2^* - n_2^*\{k\})/\sigma_{i,k}) \\
&\qquad\qquad + C_1\mathbb{P}(\bar{D}_2^k = \sqrt{k}(N_2^* - n_2^*\{k\})/\sigma_{i,k})\Big) + \varepsilon/2 \\
&= (1-\varepsilon)\left(1 - \mathcal{N}(\sqrt{k}(N_2^* - n_2^*\{k\})/\sigma_{2,k}) + f_2(k)\right) + \varepsilon/2, \\
n_2^*\{k\} &= (1-\varepsilon)\left(\mathcal{N}(\sqrt{k}(N_1^* - n_1^*\{k\})/\sigma_{1,k}) + f_1(k)\right) + \varepsilon/2.
\end{aligned}$$

Reordering the equation system above yields

$$\begin{aligned}
\mathcal{N}(\sqrt{k}(N_1^* - n_1^*\{k\})/\sigma_{1,k}) &= \frac{n_2^*\{k\} - \varepsilon/2}{1-\varepsilon} + f_1(k), \\
\mathcal{N}(\sqrt{k}(N_2^* - n_2^*\{k\})/\sigma_{2,k}) &= \frac{1 - \varepsilon/2 - n_1^*\{k\}}{1-\varepsilon} - f_2(k).
\end{aligned} \tag{20}$$

We aim to prove that for all k greater than some bound $\underline{k}$, $\|N^* - n^*\{k\}\| \leq \delta$ where δ is small and in a sense uniform. Ideally, we want $\delta = 1/\sqrt{\underline{k}}$ and that $\underline{k}$ depends only on ε. The result would then be: choosing a ε such that the Nash equilibrium is located in an interior region defined by this ε, the bound will hold for $\underline{k}(\varepsilon)$.

To simplify notation, let $\varepsilon' := (\varepsilon/2)/(1-\varepsilon)$. Apart from the $1/\sqrt{k}$-factor, the Berry-Essén bound can be made uniform over k by taking the lowest possible standard deviation and the largest possible skewness. Call this ε-dependent uniform bound $C(\varepsilon)$. Consider a positive integer $\underline{k}(\varepsilon)$ such that

A) $\sqrt{\underline{k}(\varepsilon)} > 2C(\varepsilon)/\varepsilon'$,

B) $\sqrt{\underline{k}(\varepsilon)} > \max_{x\in\mathcal{N}} \left\{ \dfrac{\mathcal{N}^{-1}(1-\varepsilon'/2)}{(x-\varepsilon)/4} \right\}$, $\mathcal{N} := \{N_1^*, N_2^*, 1-N_1^*, 1-N_2^*\}$.

These assumptions are crucial for the argument that is to follow, and the set of integers satisfying them simultaneously is non-empty for all $\varepsilon > 0$. The argument is necessary since we cannot take the limit $k \to \infty$. The calculations would simplify drastically in the limit, but they would be carried out only in a formal sense. There are plenty of reasons that hinders us from justifying the formal limit, the main one being the degeneracy of the approximating distribution in the limit.

Suppose that there is a $k \in \mathbb{N}$ such that $k > \underline{k}(\varepsilon) \in \mathbb{N}$ and $|N_1^* - n_1^*\{k\}| > \mathcal{N}^{-1}(1-\varepsilon'/2)/\sqrt{\underline{k}(\varepsilon)}$. By assumption B), $\mathcal{N}^{-1}(1-\varepsilon'/2)/\sqrt{\underline{k}(\varepsilon)} < 1/4$ and the lower bound for $|N_1^* - n_1^*\{k\}|$ is not a contradiction to $N_1^* \in (\varepsilon, 1-\varepsilon)$ and $n_1^*\{k\} \in (\varepsilon/2, 1-\varepsilon/2)$ for small enough ε. Under the assumptions stated in this paragraph, (20) says that one of the following two statements holds true

$$
\begin{aligned}
n_2^*\{k\} &< \varepsilon/2 + (1-\epsilon)\left(\mathcal{N}\left(-\sqrt{k\mathcal{N}^{-1}(1-\varepsilon'/2)/\underline{k}(\varepsilon)}\right) - f_1(k)\right),\\
n_2^*\{k\} &> \varepsilon/2 + (1-\epsilon)\left(\mathcal{N}\left(\sqrt{k\mathcal{N}^{-1}(1-\varepsilon'/2)/\underline{k}(\varepsilon)}\right) - f_1(k)\right).
\end{aligned}
$$

If

$$
k > \underline{k}(\varepsilon)\left(\frac{\max\left\{\mathcal{N}^{-1}\left(1-\varepsilon'-f_1(k)\right),\ \mathcal{N}^{-1}\left(1-\varepsilon'+f_1(k)\right)\right\}}{\mathcal{N}^{-1}(1-\varepsilon'/2)}\right)^2, \tag{21}
$$

where $\varepsilon' := (\varepsilon/2)/(1-\varepsilon)$, then $n_2^*\{k\} \in (0,\varepsilon) \cup (1-\varepsilon, 1)$. The inequality (21) holds if

$$
k > \underline{k}(\varepsilon)\left(\frac{\mathcal{N}^{-1}(1-\varepsilon' + C(\varepsilon)/\sqrt{\underline{k}(\varepsilon)})}{\mathcal{N}^{-1}(1-\varepsilon'/2)}\right)^2,
$$

since $|f_1(k)| \le C(\varepsilon)/\sqrt{k} < C(\varepsilon)/\sqrt{\underline{k}(\varepsilon)}$. By assumption B), $\underline{k}(\varepsilon)$ dominates $(2C(\varepsilon)/\varepsilon')^2$ and hence the inequality (21) holds true.

Consider the possibility that $n_2^*\{k\} \in (0, \varepsilon)$. Using assumption B), we get that

$$\mathcal{N}(\sqrt{k}(N_2^* - n_2^*\{k\})/\sigma_{2,k}) > 1 - \varepsilon/4$$

and by (20) and assumption B), $n_1^*\{k\} < \varepsilon$. Let $r_{i,k} := \varepsilon - n_i^*\{k\}$, $i = 1, 2$ be the rest, a sequence of numbers in $(0, \varepsilon)$. Then

$$\begin{aligned}\varepsilon - r_{2,k} &= \mathbb{E}\left[\widetilde{BR}_2(\varepsilon - r_{1,k})\right] \\ &= (1-\varepsilon)\left(\mathcal{N}\left(\sqrt{k}(N_1^* - \varepsilon + r_{1,k})/\sigma_{1,k}\right) + f_1(k)\right) + \varepsilon/2,\end{aligned}$$

hence

$$\varepsilon' - \frac{r_{2,k}}{1-\varepsilon} - f_1(k) = \mathcal{N}(\sqrt{k}(N_1^* - \varepsilon + r_{1,k})/\sigma_{1,k}). \tag{22}$$

By Berry-Essén and assumption A),

$$\varepsilon' - \frac{r_{2,k}}{1-\varepsilon} - f_1(k) \le \varepsilon' + \frac{C(\varepsilon)}{\sqrt{\underline{k}(\varepsilon)}} < 3\varepsilon'/2.$$

which is also a bound for the right hand side of (22),

$$\mathcal{N}(\sqrt{k}(N_1^* - \varepsilon + r_{1,k})/\sigma_{1,k}) < 3\varepsilon'/2.$$

Solving for N_1^* and using the fact that $3\varepsilon'/2 < 1/2$ leads to the contradiction

$$N_1^* < \varepsilon - r_{1,k} + \sigma_{1,k}\mathcal{N}^{-1}(3\varepsilon'/2)/\sqrt{k} < \varepsilon.$$

The other three cases, $n^* \in (0, \varepsilon) \times (1-\varepsilon, 1)$, $n^* \in (1-\varepsilon, 1) \times (0, \varepsilon)$, and $n^* \in (1-\varepsilon)^2$, can be treated in a similar way. Each will lead to a contradiction of the fundamental assumption $\varepsilon < \max\{N_1^*, N_2^*, 1 - N_1^*, 1 - N_2^*\}$.

By contradiction, we have the result: for any bound $\underline{k}(\varepsilon)$ satisfying A) and B) it holds that $|N_1^* - n_1^*\{k\}| < \mathcal{N}^{-1}(1-\varepsilon'/2)/\sqrt{\underline{k}(\varepsilon)}$ for all $k > \underline{k}(\varepsilon)$. $\square$

The factor $N^{-1}(1 - \varepsilon'/2)$ in the final line of the proof will never cause us a problem; if $\underline{k}(\varepsilon)$ satisfies assumption A) and B), then

$$\bar{k}(\varepsilon) := \left(\mathcal{N}^{-1}(1-\varepsilon'/2)\right)^2 \underline{k}(\varepsilon) > \underline{k}(\varepsilon)$$

will also satisfy assumption A) and B) and work as a bound.

Consider the mean process $n(\cdot) = (n_1(\cdot), n_2(\cdot))$ with dynamics

$$\begin{cases} n_i(t+1) = \beta n_i(t) + (1-\beta)\mathbb{E}\left[\widetilde{BR}_i(n_{-i}(t))\right], \\ \quad n_i(0) = p(0), \quad i = 1, 2. \end{cases} \tag{23}$$

The next lemma shows that as time increases, the mean process n tends to the fixed point n^*.

Lemma 15. *For all $\beta \in (\bar{\beta}, 1)$, where $\bar{\beta} > 0$ depends on k and n^* and is defined in* (26) *below, the fixed point n^* is exponentially asymptotically stable for the mean process* (23)*, i.e. there exists positive constants M, λ (possibly dependent on k, ε and n^*) such that*

$$\|n(t) - n^*\| \leq M\|n(0) - n^*\| \exp(-\lambda t).$$

Proof. Let f_i, $i = 1, 2$, be the update of the mean process centralized around the fixed point n^*,

$$n_i(t+1) - n_i^* = \beta n_i(t) + (1-\beta)\mathbb{E}\left[\widetilde{BR}_i(n_{-i}(t))\right] - n^* =: f_i(n_1(t), n_2(t)). \tag{24}$$

Linearization of (24) yields

$$n(t+1) - n^* = F(n^*)(n(t) - n^*) + \rho(n(t) - n^*)\|n(t) - n^*\|_2, \tag{25}$$

where F is the Jacobian of (f_1, f_2),

$$F(n^*) = \begin{bmatrix} \beta & F_1(n^*) \\ F_2(n^*) & \beta. \end{bmatrix},$$

and

$$\begin{aligned} F_1(n^*) &:= (1-\beta)(1-\varepsilon)p_a'(n_2^*), \\ F_2(n^*) &:= -(1-\beta)(1-\varepsilon)p_d'(n_1^*), \end{aligned}$$

and ρ is a continuous function that exists by the continuous differentiability of f_i, which follows from (19). The eigenvalues for F are $\beta \pm i(1-\beta)(1-\varepsilon)\sqrt{p_a'(n_2^*)p_d'(n_1^*)}$. The linear system is uniformly asymptotically stable if all the eigenvalues of F are located within the unit circle. Both p_a and p_d are cumulative distribution functions, so their derivative is positive. The absolute value of each of the eigenvalues is $\beta^2+(1-\beta)^2(1-\varepsilon)^2 p'_a(n_2^*)p'_d(n_1^*)$, which is smaller than 1 whenever

$$\bar{\beta} := \frac{p'_a(n_2^*)p'_d(n_1^*)-1}{p'_a(n_2^*)p'_d(n_1^*)+1} < \beta < 1. \tag{26}$$

Notice that $\bar{\beta}$ does not depend on ε, all the terms containing ε cancel in the expression above. Uniform asymptotical stability of n^* for the linearized system (25) implies exponential asymptotical stability of n^* for the mean dynamics [1, Ch. 5]. □

Consider now the variance process,

$$v(t) = (v_1(t), v_2(t)) := (\mathbb{V}(p_1(t)), \mathbb{V}(p_2(t))) .$$

Lemma 16. *For all $t \in \mathbb{N}$,*

$$v_1(t)+v_2(t) \leq c\beta\left(\beta^t + (1-\beta)\right), \tag{27}$$

where c is a positive constant that depends only on k.

Proof. As in Lemma 13, let M_1 and M_2 be the cut-offs defined by the unique Nash equilibrium N^*, let n^* be the fixed point (18), and let

$$p_{k,M}(x) := (1-\varepsilon)\sum_{i=M+1}^{k}\binom{k}{i}x^i(1-x)^{k-i}+\varepsilon/2.$$

Consider A^*, the 'shifted' state space

$$A^* = (A_1^*, A_2^*) := [-n_1^*, 1-n_1^*]\times[-n_2^*, 1-n_2^*].$$

Extend the function $d \mapsto p_{k,M_1}(n_2^*+d)$ to A_2^*/β and $a \mapsto p_{k,M_2}(n_1^*+a)$ to A_1^*/β by keeping the same expression, and define the functions g_d and g_a over A_1^* and A_2^*, respectively, by

$$g_d(a) := \int_0^a \Big(p_{k,M_2}(n_1^*+z/\beta)-(1-n_2^*)\Big)dz,$$
$$g_a(d) := \int_0^d \Big(p_{k,M_1}(n_2^*+z/\beta)-n_1^*\Big)dz.$$

The functions g_a and g_d are smooth for all M_1, M_2 and all $k<\infty$, and $g_a(0)=g_d(0)=0$. Furthermore, $g'_a(d)=p_{k,M_1}(n_2^*+d/\beta)-n_1^*$ and likewise differentiation of g_d yields the integrand evaluated in the argument. Hence, since p_{k,M_1} and p_{k,M_2} are strictly increasing, g_a and g_d are strictly convex. We will make use the following estimates of g_a and g_d: there exists four (in general different from each other) positive constants c_{a-}, c_{a+}, c_{d-}, and c_{d+} such that

$$\begin{aligned} p_{k,M_1}(n_2^* + y) &\geq c_{a+}y + n_1^*, && y \in [0, 1 - n_2^*], \\ p_{k,M_1}(n_2^* + y) &\leq c_{a-}y + n_1^*, && y \in [-n_2^*, 0], \\ p_{k,M_2}(n_1^* + y) &\geq c_{d+}y + (1 - n_2^*), && y \in [0, 1 - n_1^*], \\ p_{k,M_2}(n_1^* + y) &\leq c_{d-}y + (1 - n_2^*), && y \in [-n_1^*, 0]. \end{aligned}$$

The estimates imply that for all $(a, d) \in A^*$,

$$g_a(d) + g_d(a) \geq \min\{c_{a+}, c_{a-}, c_{d+}, c_{d-}\} \frac{1}{2\beta} \left(a^2 + d^2\right). \tag{28}$$

Now consider the shifted states $A(t) := p_1(t) - n_1^*$ and $B(t) := p_2(t) - n_2^*$. The update of the shifted state is

$$\begin{aligned} A(t+1) &= \beta A(t) + (1 - \beta)\left(\widetilde{BR}_1(n_2^* + B(t)) - n_1^*\right), \quad A(0) = p_1(0) - n_1^*, \\ B(t+1) &= \beta B(t) + (1 - \beta)\left(\widetilde{BR}_2(n_1^* + A(t)) - n_2^*\right), \quad B(0) = p_2(0) - n_2^*. \end{aligned}$$

For $(a, d) \in A^*$, let $G(a, d) := g_d(a) + g_a(d)$. Expanding G with the Taylor formula yields

$$\begin{aligned} &G\left(A\left(t+1\right), B\left(t+1\right)\right) \\ &= G\left(\beta A\left(t\right), \beta B\left(t\right)\right) + g_d'\left(\beta A\left(t\right)\right)\left(1 - \beta\right)\left(\widetilde{BR}_1(n_2^* + B(t)) - n_1^*\right) \\ &\quad + g_a'\left(\beta B\left(t\right)\right)\left(1 - \beta\right)\left(\widetilde{BR}_2(n_1^* + A(t)) - n_2^*\right) + O\left(\left(1 - \beta\right)^2\right) \\ &= G\left(\beta A\left(t\right), \beta B\left(t\right)\right) \\ &\quad + (1 - \beta)\Bigg(\left(p_{k,M_2}(n_1^* + A(t)) - (1 - n_2^*)\right)\left(\widetilde{BR}_1(n_2^* + B(t)) - n_1^*\right) \\ &\quad + \left(p_{k,M_1}(n_2^* + B(t)) - n_1^*\right)\left(\widetilde{BR}_2(n_1^* A(t)) - n_2^*\right)\Bigg) + O\left(\left(1 - \beta\right)^2\right), \end{aligned}$$

where the ordo is uniform since g_a and g_d are C^2 over the compact state space. However

$$\mathbb{E}\left[\widetilde{BR}_1(n_2^*+B(t)) \mid \mathcal{F}_t\right] = p_{k,M_1}(n_2^*+B(t)),$$
$$\mathbb{E}\left[\widetilde{BR}_2(n_1^*+A(t)) \mid \mathcal{F}_t\right] = 1-p_{k,M_2}(n_1^*+A(t)).$$

Therefore the line of order 1 has conditional expectation zero and we are left with

$$\mathbb{E}\left[G\left(A\left(t+1\right),B\left(t+1\right)\right)\ \mid \mathcal{F}_t\right] \leq G\left(\beta A\left(t\right),\beta B\left(t\right)\right)+M\left(1-\beta\right)^2$$

for some uniform constant M. By convexity of g_a and g_d,

$$\mathbb{E}\left[G\left(A\left(t+1\right),B\left(t+1\right)\right)\ \mid \mathcal{F}_t\right] \leq \beta G\left(A\left(t\right),B\left(t\right)\right)+M\left(1-\beta\right)^2$$

By repeated use of the argument above together with the tower property of conditional expectations we get

$$\begin{aligned}&\mathbb{E}\left[G\left(A\left(t+1\right),B\left(t+1\right)\right)\ \mid \mathcal{F}_t\right]\\&\qquad\leq \beta^{\tau+1}G\left(A\left(0\right),B\left(0\right)\right)+\sum_{\tau=0}^{t}\beta^{\tau}M\left(1-\beta\right)^2.\end{aligned}$$

So when $t\to\infty$,

$$\limsup_{t\to\infty}\mathbb{E}\left[G\left(A\left(t\right),B\left(t\right)\right)\right]\leq M\left(1-\beta\right). \tag{29}$$

From (28) and (29) it follows that

$$\limsup_{t\to\infty}\mathbb{E}\left[A^2\left(t\right)+B^2\left(t\right)\right]\leq\frac{2\beta M}{\min\{c_{a+},c_{a-},c_{d+},c_{d-}\}}\left(1-\beta\right).$$

The proof is completed by simply noting that

$$v_1(t)+v_2(t)\leq\mathbb{E}\left[\left(p_1(t)-n_1^*\right)^2+\left(p_2(t)-n_2^*\right)^2\right]=\mathbb{E}[A^2(t)+B^2(t)].$$

$\square$

As $t\to\infty$, the bound (27) is controlled by β: the right-hand side is continuous in β and as $\beta\to 1$, it goes to zero. The following concentration result is a straight away consequence of Lemma 16 and Chebyshev's inequality.

Proposition 17. *For all $\eta > 0$,*

$$\mathbb{P}\left(\|p(t) - n(t)\|_2 \geq \eta\right) \leq \frac{c\beta\left(\beta^t + (1-\beta)\right)}{\eta^2},$$

where c is a positive constant that depends only on k.

References

[1] R. P. Agarwal. *Difference equations and inequalities: Theory, Methods, and Applications.* CRC Press, 2000.

[2] D. Balkenborg, J. Hofbauer, and C. Kuzmics. Refined best reply correspondence and dynamics. *Theoretical Economics*, 8(1):165–192, jan 2013.

[3] K. Basu and J. W. Weibull. Strategy Subsets Closed Under Rational Behavior. *Economics Letters*, 36(2):141–146, 1991.

[4] M. Benaïm and M. W. Hirsch. Mixed Equilibria and Dynamical Systems Arising from Fictitious Play in Perturbed Games. *Games and Economic Behavior*, 29(1-2):36–72, oct 1999.

[5] M. Benaïm, J. Hofbauer, and E. Hopkins. Learning in games with unstable equilibria. *Journal of Economic Theory*, 144(4):1694–1709, 2009.

[6] J. I. Block, D. Fudenberg, and D. K. Levine. Learning dynamics with social comparisons and limited memory 1. *Theoretical Economics*, 14 (1):135–172, 2019.

[7] G. W. Brown. Iterative solution of games by fictitious play. *Activity analysis of production and allocation*, 13(1):374–376, 1951.

[8] C. F. Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton University Press, 2003.

[9] L. Dinetan. On the positivity condition for the Krein-Rutman theorem. Unpublished, 2018.

[10] G. Ellison. Basins of Attraction, Long-Run Stochastic Stability, and the Speed of Step-by-Step Evolution. *Review of Economic Studies*, 67(1): 17–45, 2000.

[11] D. P. Foster and H. P. Young. Learning, hypothesis testing, and nash equilibrium. *Games and Economic Behavior*, 45(1):73–96, 2003.

[12] D. Fudenberg and D. M. Kreps. Learning Mixed Equilibria. *Games and Economic Behavior*, 5:320–367, 1993.

[13] D. Fudenberg and D. K. Levine. Learning with recency bias. *Proceedings of the National Academy of Sciences*, 111:10826–10829, 2014.

[14] D. Fudenberg, F. Drew, D. K. Levine, and D. K. Levine. *The theory of learning in games*, volume 2. MIT press, 1998.

[15] S. Hart and A. Mas-Colell. Stochastic uncoupled dynamics and Nash equilibrium. *Games and Economic Behavior*, 57(2):286–303, 2006.

[16] J. Hofbauer and W. H. Sandholm. On the Global Convergence of Stochastic Fictitious Play. *Econometrica*, 70(6):2265–2294, nov 2002.

[17] S. Hurkens. Learning by Forgetful Players. *Games and Economic Behavior*, 11(2):304–329, 1995.

[18] G. E. Kreindler and H. P. Young. Fast convergence in evolutionary equilibrium selection. *Games and Economic Behavior*, 80:39–67, 2013.

[19] J. Nash. *Non-cooperative games.* PhD thesis, Princeton University, 1950.

[20] H. Peyton Young and D. P. Foster. Regret testing: learning to play Nash equilibrium without knowing you have an opponent. *Theoretical Economics*, 1(3):341–367, 2006.

[21] K. Ritzberger and J. W. Weibull. Evolutionary Selection in Normal-Form Games. *Econometrica*, 63(6):1371, nov 1995.

[22] W. H. Sandholm. *Population games and evolutionary dynamics.* MIT press, 2010.

[23] L. Shapley. Some topics in two-person games. *Advances in game theory*, 52:1–29, 1964.

[24] J. W. Weibull. *Evolutionary game theory.* MIT press, 1997.

[25] H. P. Young. The evolution of conventions. *Econometrica*, 61:57–84, 1993.

[26] H. P. Young. *Individual Strategy and Social Structure An Evolutionary Theory of Institutions.* Princeton University Press, 1998.