Università degli Studi di Padova

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Matematica "Tullio Levi-Civita"

CORSO DI DOTTORATO DI RICERCA IN SCIENZE MATEMATICHE
CURRICULUM MATEMATICA COMPUTAZIONALE
CICLO XXXII

Topics in finite state mean field games and non-Markovian interacting spin systems

Tesi redatta con il contributo finanziario della Fondazione Cassa di Risparmio di Padova e Rovigo

Coordinatore del Corso: Ch.mo Prof. Martino Bardi

Supervisore: Ch.mo Prof. Paolo Dai Pra

Dottorando: Guglielmo Pelino

Riassunto

In questa Tesi si studiano sistemi stocastici con un grande numero di individui microscopici interagenti, sotto determinate ipotesi di simmetria delle interazioni. Gli esempi considerati appartengono a due contesti differenti, a seconda del fatto che il singolo individuo possa controllare la propria dinamica o meno. Nel primo caso, di cui tratta la Parte 1, si è nel contesto dei giochi ad N giocatori e giochi a campo medio, mentre nel secondo, analizzato nella Parte 2 della Tesi, i modelli risultanti vengono detti sistemi di particelle interagenti.

Più precisamente, nella prima parte (Capitoli 1-2) studiamo il problema della convergenza per i giochi a campo medio, il cui fine è di giustificare rigorosamente l'introduzione degli stessi come limite di giochi simmetrici non cooperativi non a somma zero ad N giocatori, quando il numero dei giocatori tende ad infinito. In particolare, l'analisi è incentrata sui cosiddetti giochi a campo medio a stati finiti, in cui lo stato del singolo giocatore appartiene a un insieme discreto finito: trattiamo separatamente il caso in cui si ha unicità di soluzione del gioco a campo medio (Capitolo 1), da quello in cui la formulazione limite ammette più soluzioni (Capitolo 2).

Nella seconda parte invece (Capitoli 3-4), introduciamo alcuni esempi di sistemi di spin in cui la dinamica particellare non interagente è non Markoviana, ottenuti da opportune modifiche di classici modelli ferromagnetici di spin a campo medio. In particolare, rilassiamo l'ipotesi di Markovianità o tramite una procedura di aumento delle variabili che identificano lo stato individuale, o imponendo la presenza di memoria nella dinamica. Sebbene uno degli scopi sia ancora quello di giustificare rigorosamente la formulazione macroscopica dei suddetti modelli, essi presentano alcune caratteristiche di independente interesse, tra cui la presenza di transizioni di fase (Capitoli 3-4), la comparsa di comportamenti periodici auto-sostenuti (Capitolo 3) e la presenza di diversi fenomeni a diverse scale spazio-temporali (Capitolo 4).

Abstract

This Dissertation is devoted to the study of large stochastic systems of small interacting individuals and their macroscopic limit formulations, under symmetric properties of the interactions. The examples we consider belong to two separate contexts, depending on whether the individuals can control their dynamics or not. In the first case, treated in Part 1, we fall into the framework of N-player and mean field games, while in the latter, analyzed in Part 2, the resulting models are examples of interacting particle systems.

More specifically, in the first part (Chapters 1-2) we focus on the convergence problem in mean field games, i.e. on the rigorous justification of mean field games as limits, when the number of players tends to infinity, of Nash equilibria of symmetric non-zero sum non-cooperative N-player games. In particular, we study finite state mean field games, where the state of each player belongs to a discrete finite space, analyzing separately the uniqueness case (Chapter 1) and a scenario with non-uniqueness of solutions to the mean field game (Chapter 2).

In the second part of the Dissertation (Chapters 3-4) we study some examples of interacting spin systems, with non-Markovian individual dynamics, arising as proper modifications of classical ferromagnetic mean field spin systems dynamics. In particular, we focus on two mechanisms for relaxing the Markovianity: a state augmentation procedure, and the insertion of memory effects in the evolution. While one of the goals is still to rigorously justify the passage to a macroscopic description, the models of Part 2 present some features of independent interest, including phase transitions (Chapters 3-4), the emergence of self-sustained oscillations (Chapter 3), and the presence of multiple spatio-temporal scales phenomena (Chapter 4).

Contents

Introduction

Mean field systems of interacting particles were first introduced for applications to physics, and in particular to the kinetic theory of gases for the derivation of the spatially-homogeneous Boltzmann equation ([73]). For the modelization of related physical phenomena the mean field assumption can often seem overly simplicistic. Nevertheless, the interest in mean field models has lasted over the years, due to their analytical tractability and flexibility of application also to other disciplines (such as biology, sociology and economics), in contexts where the mean field assumption appears more natural.

More recently, the integration of mean field interacting particles with control theory has led to the introduction of mean field games ([71, 79]), where the objects of the modelization are large systems of competitively interacting rational agents (rather than particles), which are allowed to control their individual dynamics through some optimization criterion, depending on the other players in a mean field way.

This Dissertation consists of two parts, which we now introduce, where the mean field structure of interactions is a common theme.

Part 1: Finite state mean field games

This part of the Dissertation focuses on the convergence problem for a class of N-player finite state stochastic differential games to the corresponding mean field game limit formulation.

Mean field games were introduced independently by Lasry and Lions [79] and by Huang et al.[71] as limit models for symmetric non-zero-sum non-cooperative N-player dynamic games when the number N of players tends to infinity; see for instance [8, 14, 18], and the recent two-volume comprehensive work [19, 20]. While a wide range of different classes of mean field games has been considered up to now, here we focus on finite time horizon problems with continuous time dynamics under fully symmetric cost structure and complete information, where the position of each agent belongs to a finite state space. In this setting, mean field games were first analyzed in [62] in discrete time, and then in [61] in continuous time.

In the literature, the notion of optimality adopted for the many player games is usually that of a Nash equilibrium. The relation with the limit formulation can then be made rigorous in two opposite directions: either by showing that a solution of the limit model (the mean field game) induces a sequence of approximate Nash equilibria for the N-player games with approximation error tending to zero as N tends to infinity, or by identifying the possible limit points of sequences of N-player Nash equilibria, as solutions, in some sense, of the limit model. While the first problem served as a motivation for the initial development of mean field game theory, and is still relevant for the applications, the latter direction constitutes what we refer to as the *convergence problem in mean field games*.

In the approximation direction results are more common and typically easier to obtain: for the diffusive case without jumps see for instance [9, 18, 21, 71]; the diffusive case with controlled jumps is treated in the recent work [7]. In the finite state space setting, an approximation result is achieved in [4] studying the infinitesimal generator, while in [24] an analogous property is found through a fully probabilistic approach, which allows for less restrictive assumptions on the dynamics and the optimization costs.

On the other hand, results on convergence are fewer and more recent. Important for the convergence problem is the choice of admissible strategies and the resulting definition of Nash equilibrium in the many player games. For Nash equilibria defined in stochastic open-loop strategies, the convergence problem is rather well understood, see [58] and, especially, [77], both in the context of finite horizon games with general diffusive dynamics. In [77], limit points of sequences of N-player Nash equilibria are shown to be concentrated on weak solutions of the corresponding mean field game.

In Part 1 of the Dissertation, we are interested in the convergence problem for Nash equilibria in Markov feedback strategies with full state information. A first result in this direction is given by [61] in our same setting of finite state dynamics. There, convergence of Markovian Nash equilibria to the mean field game limit is proved, but only if the time horizon is small enough. A breakthrough was achieved by Cardaliaguet et. al in [15]. In the setting of games with non-degenerate diffusive dynamics, possibly including common noise, the authors establish convergence to the mean field game limit, in the sense of convergence of value functions as well as propagation of chaos for the optimal state trajectories, for any finite time horizon, provided the so-called master equation associated with the mean field game possesses a unique sufficiently regular solution. The master equation arises as the formal limit of the Hamilton-Jacobi-Bellman systems determining the Markov feedback Nash equilibria. Moreover, the mean field game system can be seen as the characteristics curves for the master equation. If well-posed, the latter yields the optimal value in the mean field game as a function of initial time, state and distribution. It thus also provides the optimal control action, again as a function of time, state, and the measure variable. This allows, in particular, to compare the prelimit Nash equilibria to the solution of the limit model through coupling arguments. Such coupling ultimately allows one to get the desired convergence of the value functions of the N-player game to the solution to the master equation, as well as a propagation of chaos result for the corresponding optimal trajectories, in a similar fashion to the propagation of chaos property for uncontrolled systems (see e.g. [96]).

If the master equation possesses a unique regular solution, which is guaranteed under the Lasry-Lions monotonicity conditions, then the convergence analysis can be considerably refined. In this case, for games with finite state dynamics, in the two independent works [26, 5] the authors obtain a Central Limit Theorem and Large Deviations Principle for the empirical measures associated with Markovian Nash equilibria. In [43, 44], the authors carry out the analysis, enriched by a concentration of measure result, for diffusive dynamics without or with common noise.

We now give a detailed overview of the two chapters of the first part of the Dissertation.

Overview of Chapter 1

In this chapter we analyze the work [26], by A. Cecchin and the author.

Here we focus on the convergence of feedback Nash equilibria for finite state symmetric N-player differential games, where players control their transition rates from state to state. In details, denote the state space as $\Sigma := \{1, \ldots, d\}$. Let $X_i(t)$ be the state of the i-th player at time t, and $\boldsymbol{X}_t^{N,i}$ the state at time t of the other $N-1$ players. The N-player dynamics is given by a system of interacting controlled continuous-time Markov chains, such that, for $h>0$,

$$\mathbb{P}\left[X_i(t+h)=y|X_i(t)=x,\ \boldsymbol{X}_t^{N,i}=\boldsymbol{x}^{N,i}\right]=\alpha_y^i(t,x,\boldsymbol{x}^{N,i})h+o(h),$$

for $x\neq y$ and any $\boldsymbol{x}^{N,i}\in\Sigma^{N-1}$. The control $\alpha_y^i(t,x,\boldsymbol{x}_t^{N,i})$, in Markov feedback form, represents the rate at which player i decides to go from state x to state y, when $\boldsymbol{x}_t^{N,i}$ is the state of the other $N-1$ players at time t.

In our framework, we show that there exists a unique feedback Nash equilibrium for the N-player game (Proposition 1.3). It is provided by the solution to the Hamilton-Jacobi-Bellman (HJB) system of Nd^N coupled ODE's

$$\begin{cases}-\frac{\partial v}{\partial t}^{N,i}(t,\boldsymbol{x})-\sum_{j=1,\ j\neq i}^{N}\alpha^*(x_j,\Delta^j v^{N,j})\cdot\Delta^j v^{N,i}+H(x_i,\Delta^i v^{N,i})=F^{N,i}(\boldsymbol{x}),\\ v^{N,i}(T,\boldsymbol{x})=G^{N,i}(\boldsymbol{x}).\end{cases}\tag{HJB}$$

In the above equation, $F^{N,i}$ and $G^{N,i}$ are respectively the running and terminal interaction costs of the i-th player, H is the Hamiltonian and α^* its unique maximizer, and

$$\Delta^j g(\boldsymbol{x}):=\left(g(x_1,\ldots,y,\ldots,x_N)-g(x_1,\ldots,x_j,\ldots,x_N)\right)_{y=1,\ldots,d}\in\mathbb{R}^d$$

denotes the finite difference of a function $g(\boldsymbol{x})=g(x_1,\ldots,x_N)$ with respect to its j-th entry.

The study of convergence consists in finding a limit for System (HJB) as $N\to+\infty$, under symmetric properties of the game (Proposition 1.5), realized by assuming that the costs $F^{N,i}$ and $G^{N,i}$ satisfy the mean field assumptions, i.e. there exist two functions F and G such that

$$F^{N,i}(\boldsymbol{x})=F(x_i,m_{\boldsymbol{x}}^{N,i}),$$
$$G^{N,i}(\boldsymbol{x})=G(x_i,m_{\boldsymbol{x}}^{N,i}),$$

where $m_{\boldsymbol{x}}^{N,i}:=\frac{1}{N-1}\sum_{j=1,j\neq i}^{N}\delta_{x_j}$ denotes the empirical measure of all the players except for the i-th, which belongs to $P(\Sigma)$, the space of probability measures on Σ. The main result of this chapter is Theorem 1.7, in which we prove the convergence of the value functions $v^{N,i}$'s, as $N\to+\infty$, to the solution to the *master equation*

$$\begin{cases}-\frac{\partial U}{\partial t}+H(x,\Delta^x U)-\int_\Sigma D^m U(t,x,m,y)\cdot\alpha^*(y,\Delta^y U(t,y,m))dm(y)=F(x,m),\\ U(T,x,m)=G(x,m),\quad (x,m)\in\Sigma\times P(\Sigma),\quad t\in[0,T],\end{cases}\tag{M}$$

provided (M) has a unique regular solution. It is a first order PDE in $P(\Sigma)$, the simplex of probability measures in $\mathbb{R}^d$. The corresponding mean field game system, which can be proved to be the system of characteristic curves for (M), is given by the following system of two forward-backward ODEs,

$$\begin{cases}-\frac{d}{dt}u(t,x)+H(x,\Delta^x u(t,x))=F(x,m(t)),\\ \frac{d}{dt}m_x(t)=\sum_y m_y(t)\alpha_x^*(y,\Delta^y u(t,y)),\\ u(T,x)=G(x,m(T)),\\ m_x(t_0)=m_{x,0}\in P(\Sigma),\end{cases}\tag{MFG}$$

where u is the value function of a representative player, and m is the optimal (in the Nash sense) evolution of the distribution of the other infinite players (which coincides with the one of the reference player itself). The convergence result of Theorem 1.7 allows for obtaining the convergence of the optimal trajectories to a collection of i.i.d. limit processes, in terms of a propagation of chaos property, which we state and prove in Theorem 1.8. Moreover, in Section 1.4 we study the fluctuations and large deviations of the optimal empirical measures processes, proving a Central Limit Theorem (Theorem 1.13) and a Large Deviation Principle (Theorem 1.14). Section 1.5 is finally devoted to the study of the well-posedness of the master equation (M) under monotonicity assumptions (in the Lasry-Lions sense) on the costs F and G. We stress, however, that the convergence argument requires only the regularity of a solution to the master equation.

Overview of Chapter 2

In this chapter we discuss the results in [25], by A. Cecchin, P. Dai Pra, M. Fischer and the author.

We consider N-player and mean field games in continuous time over a finite horizon, where the position of each agent belongs to a spin-valued state space $\Sigma := \{-1, 1\}$. If there is uniqueness of mean field game solutions, e.g. under monotonicity assumptions, then the convergence results of Chapter 1 apply. In Chapter 2 we instead study an example with anti-monotonic costs, where one expects to find multiple solutions to the mean field game. We identify an element $m \in P(\Sigma)$ with its mean $m_1 - m_{-1}$, which we still denote by $m \in [-1, 1]$. Moreover, denote $z(t) := u(t, -1) - u(t, 1)$, where u is the value function of the reference player in the mean field game limit. In this context each player controls, in feedback Markov form, the *switching rates* from one state to the other. With respect to the setting of the previous chapter, here we set the interaction running cost $F \equiv 0$, we consider an anti-monotonic terminal cost $G(x, m) := -mx$, which favors alignment with the majority, and a simple quadratic Lagrangian $L(x, a) := \frac{a^2}{2}$, penalizing large values of the switching rates.

In this framework, the mean field game system takes the simple form

$$\begin{cases} \dot{z} = \frac{z|z|}{2}, \\ \dot{m} = -m|z| + z, \\ z(T) = 2m(T), \\ m(0) = m_0, \end{cases} \tag{1}$$

which we are able to solve explicitly. In particular, in Proposition 2.1 we prove that System (1) possesses exactly three solutions when the final time horizon of the game $T > 0$ is sufficiently large. Under the same notation, we can associate the corresponding master equation of the limit problem

$$\begin{cases} -\frac{\partial U}{\partial t}(t,x,m) + \frac{1}{2}\left[(\Delta^x U(t,x,m))^-\right]^2 - D^m U(t,x,m,1)\,(\Delta^x U(t,1,m))^-\left(\frac{1+m}{2}\right) \\ \qquad\qquad -D^m U(t,x,m,-1)\,(\Delta^x U(t,-1,m))^-\left(\frac{1-m}{2}\right) = F(x,m), \\ U(T,x,m) = G(x,m), \quad (x,m) \in \{-1,1\} \times [-1,1]. \end{cases} \tag{2}$$

Equation (2) admits an equivalent formulation as a scalar *conservation law* (see Equation (2.13)). Because of the multiple solutions to the mean field game (1), Equation (2) does not admit a regular solution. While multiple weak solutions exist, the associated conservation law has a unique entropy solution (Theorem 2.3), with a discontinuity point, which turns out to be of particular importance for the N-player game. At the prelimit level, the unique N-player Nash equilibrium can again be described in terms of a system of Hamilton-Jacobi-Bellman equations,

$$\begin{cases} -\frac{d}{dt}V^N(t,\mu) + H(V^N(t,1-\mu) - V^N(t,\mu)) \\ \quad = N\mu\left[V^N(t,1-\mu) - V^N(t,\mu)\right]^- \left[V^N\left(t,\mu-\frac{1}{N}\right) - V^N(t,\mu)\right] \\ \qquad +N(1-\mu)\left[V^N\left(t,\mu+\frac{1}{N}\right) - V^N\left(t,1-\mu-\frac{1}{N}\right)\right]^- \left[V^N\left(t,\mu+\frac{1}{N}\right) - V^N(t,\mu)\right], \\ V^N(T,\mu) = -(2\mu-1), \end{cases} \tag{3}$$

written in the alternative variable μ, i.e. the portion of players in state 1, for exploiting the symmetries of the game. System (3) can be shown to reflect the degeneracy of the limit master equation: indeed, for $N \gg 0$, a singularity develops in the symmetric point $\mu = \frac{1}{2}$ (where half of the players is in state 1); see Figure 2.1 for a simulation.

The main result of the chapter (Theorem 2.7) consists in proving that the Nash equilibrium of the N-player game selects, when $N \to +\infty$, the unique entropy solution of the conservation law associated to the master equation (2). For the proof of this fact, we exploit a characterization of the Nash equilibrium (Theorem 2.6), which allows to deduce that the dynamics does not cross the discontinuity point $\mu = \frac{1}{2}$. As for the previous chapter, the convergence theorem allows for obtaining the propagation of chaos property for the N-player optimal trajectories and empirical measures, when they play the Nash equilibrium, but only when we start the dynamics outside the discontinuity point $\mu = \frac{1}{2}$ (Theorem 2.10). When starting the dynamics precisely in the degeneracy point, we expect the limit of the N-player empirical measures to be random, given by a symmetrically weighted sum of two Dirac's deltas over the two non-zero solutions to mean field game (Conjecture 2.1). Our expectations are supported by numerical simulations (see Figure 2.2), and by an analogous result obtained in [42] for the diffusive case.

Finally, in Section 2.2.7 we give another characterization of the multiple solutions to the mean field game system. Indeed, we show that the latter can be viewed as the necessary conditions for optimality, given by the Pontryagin maximum principle, of a *deterministic* optimal control problem in $\mathbb{R}^2$ (Lemma 2.14). We show that the N-player game, in the limit $N \to +\infty$, selects exactly the global minimizer of this problem when it is unique, i.e. when the initial mean of the players m_0 is different from zero (Theorem 2.16).

Part 2: Non-Markovian interacting spin systems

The theory of interacting particle systems, originally motivated by statistical mechanics and dated back to 1960's (see e.g. [82] for a classical textbook), offers popular and powerful tools for the modeling of several complex phenomena in life sciences such as ecology ([99]) and neuroscience ([55]), but also in social sciences and economics ([27, 84, 95, 100]). In particular, interacting particle systems with mean field interactions have been proved to be extremely appealing and successful, due to their mathematical tractability, since the initial pioneering works by McKean ([85, 86]) on Vlasov equations, exploring the connections with nonlinear PDEs.

Typically, the stochastic modelization consists of three main steps: identify variables of interest for each individual, superimpose some form of interaction, and define a Markovian evolution for the state of the whole population. Under the above framework, in the absence of interaction, the individual dynamics is Markovian with respect to the chosen variables of interest.

A question which can arise naturally is about the formulation of interacting particle systems where the non-interacting individual dynamics is itself not Markovian (see e.g. [50, 51]). Assuming a Markovian individual evolution might indeed be restrictive in some contexts: for example, with the initial choice on the quantities of interest, many additional variables get inevitably neglected in the modelization of the individual dynamics, with potential macroscopic effects unaccounted for in the interactive model; moreover, one might be interested in introducing memory effects, which are not captured under the Markovianity assumption (according to which the future depends only on the immediate past, and not on the previous history of the evolution).

In Part 2 of the Dissertation we propose some toy examples of non-Markovian interacting spin systems (i.e. systems with individual states taking values in the binary set $\{-1, 1\}$), which can make the above-mentioned restrictions less dramatic, in some cases revealing a large-scale behavior different from that of the original Markovian version of the model. Although they all arise as proper modifications of classical ferromagnetic mean field spin systems dynamics, the way we relax the Markovianity assumption can differ among the models. In particular, as we motivated above, we realize the non-Markovianity either by an augmentation of state procedure, or by the insertion of memory effects in the individual dynamics.

While having as a purpose to obtain macroscopic descriptions, and study the corresponding limit models of the examples considered, two topics, of independent interest, turn out to be of particular importance in the tractation: the emergence of self-sustained periodic behavior, and the presence of multiscale spatio-temporal phenomena in hierarchical mean field models, respectively analyzed in Chapters 3 and 4.

Specifically, with the term self-sustained periodic behavior we refer to systems where each individual particle has no natural tendency to behave periodically, but the oscillations are rather an effect of self-organization, visible in the macroscopic limit when the number of particles tends to infinity. Among the mechanisms that can lead to or enhance the emergence of this behavior, we cite noise ([36], [92], [98]), dissipation in the interaction potential ([1], [29], [30], [35]), delay in the transmission of information and/or frustration in the interaction network ([31], [50], [97]). In Chapter 3, we shall see that the non-Markovianity of the individual dynamics can as well foster macroscopic oscillations.

Concerning the second topic, hierarchical models were often employed in the literature for applications in population dynamics and genetics, where individuals naturally dispose in groups with a hierarchical structure (families, clans, villages, colonies, populations and so on). A series of papers from the ’90s - ’00s (initiated with [39] and [40] among others), nicely reviewed in [70], deals with different types of hierarchical mean field linearly interacting diffusions (the prototype being linear Wright-Fisher diffusions), where in most cases the macroscopic limits are retrieved at every spatio-temporal scale, and a renormalization map can be defined, allowing one to pass from one hierarchical level to the other. The motivation for focusing on diffusive dynamics as building blocks for the hierarchical models stems from the fact that, with their choices, each individual non-interacting dynamics can itself be obtained as a continuum limit of a corresponding finite state space model of interacting particles: for example, the discrete prelimit counterpart of the Wright-Fisher diffusion is the *voter model* (see e.g. [33]). In Chapter 4, we define hierarchical dynamics of spin-flip type with a ferromagnetic mean field interaction, coupled with a system of linearly interacting diffusions of Ornstein-Uhlenbeck type.

While for a more detailed introduction on the above matters we refer to the beginning of each corresponding chapter, we now proceed with an overview of the results obtained in Part 2 of the Dissertation.

Overview of Chapter 3

The results of this chapter belong to an ongoing work of the author with P. Dai Pra and M. Formentin.

After some preliminaries on spin systems and dynamics, given in Section 3.1, in this chapter we analyze two examples of non-Markovian mean field interacting spin systems. In both cases we consider spin-flip dynamics obtained as modifications of the Curie–Weiss model. In the first example, analyzed in Section 3.2, the individual evolution is obtained by replacing the underlying Poisson process, modeling the jump times in the Markovian case, with a more general renewal process with memory. Let $(\sigma(t))_{t\geq 0}$ denote the resulting spin-valued process, that is an example of two-state semi-Markov process. We can associate a Markovian description to the latter: define $y(t)$ as the time elapsed since the last spin-flip occurred up to time t. Suppose that the waiting times τ of the underlying renewal process satisfy

$$\mathbb{P}(\tau > t) = \varphi(t), \tag{4}$$

for some smooth function $\varphi : [0, +\infty) \to \mathbb{R}$. Then, the pair $(\sigma(t), y(t))_{t\geq 0}$ is Markovian with infinitesimal generator

$$\mathcal{L}f(\sigma, y) = \frac{\partial f}{\partial y}(\sigma, y) + F(y)[f(-\sigma, 0) - f(\sigma, y)], \tag{5}$$

for $f : \{-1, 1\} \times \mathbb{R}^+ \to \mathbb{R}$, with

$$F(y) := -\frac{\varphi'(y)}{\varphi(y)}. \tag{6}$$

We study a corresponding non-Markovian interacting N-particle system for the spins $(\sigma_i(t))_{i=1,\dots,N}$. A mean field type interaction is introduced as a time scaling on the waiting times between two successive particle's jumps, depending on the overall magnetization of the system

$$m^N(t) := \frac{1}{N}\sum_{i=1}^N \sigma_i(t).$$

As above, we associate to each spin $\sigma_i(t)$ the process $y_i(t) \in \mathbb{R}^+$ of the elapsed time since the last jump. Denoting $\boldsymbol{\sigma} := (\sigma_1, \dots, \sigma_N) \in \{-1, 1\}^N$, $\boldsymbol{y} := (y_1, \dots, y_N) \in (\mathbb{R}^+)^N$, the corresponding Markovian N-particle dynamics is defined via the following infinitesimal generator

$$\mathcal{L}^N f(\boldsymbol{\sigma}, \boldsymbol{y}) = \sum_{i=1}^N \frac{\partial f}{\partial y_i}(\boldsymbol{\sigma}, \boldsymbol{y}) + \sum_{i=1}^N F\left(y_i e^{-\beta\sigma_i m^N}\right) e^{-\beta\sigma_i m^N}\left[f(\boldsymbol{\sigma}^i, \boldsymbol{y}^i) - f(\boldsymbol{\sigma}, \boldsymbol{y})\right], \tag{7}$$

where $\boldsymbol{\sigma}^i$ is obtained from $\boldsymbol{\sigma}$ by flipping the i-th spin, while $\boldsymbol{y}^i$ by setting to zero the i-th coordinate. Note that, for $F \equiv 1$, we retrieve the classical Curie–Weiss dynamics for the spins. The macroscopic limit and the propagation of chaos property for this model, as $N \to +\infty$, are studied in Appendix B, when $F(y) := y^\gamma$, $\gamma \in \mathbb{N}$. Under the latter choice, we can associate to the McKean-Vlasov limit process $(\sigma(t), y(t))_{t\geq 0}$ the following Fokker-Planck equation, satisfied by the corresponding density function $f(t, \sigma, y)$:

$$\begin{cases} \frac{\partial}{\partial t} f(t, \sigma, y) + \frac{\partial}{\partial y} f(t, \sigma, y) + y^\gamma e^{-(\gamma+1)\beta\sigma m(t)} f(t, \sigma, y) = 0, \\ f(t, \sigma, 0) = \int_0^{+\infty} y^\gamma e^{(\gamma+1)\beta\sigma m(t)} f(t, -\sigma, y) dy, \\ m(t) = \int_0^\infty [f(t, 1, y) - f(t, -1, y)] dy, \\ 1 = \int_0^\infty [f(t, 1, y) + f(t, -1, y)] dy, \\ f(0, \sigma, y) = f_0(\sigma, y), \text{ for } \sigma \in \{-1, 1\}, \ y \in \mathbb{R}^+. \end{cases} \tag{8}$$

We then find both theoretical and numerical evidence of emerging periodic behavior for the above equation in the cases $\gamma = 1$ and $\gamma = 2$, in terms of a phase transition with respect to the inverse temperature parameter β, via the following approach: in Section 3.2.2, we find a neutral stationary solution of interest to (8) (Proposition 3.2), we linearize formally the dynamics around that equilibrium and we compute the discrete spectrum of the associated linearized operator, which we show to be given by the zeros of an explicit holomorphic function $H_{\beta,\gamma}(\lambda)$ (Propositions 3.5 and 3.6). In Subsection 3.2.2.4 we then study numerically the character of the eigenvalues when the interaction parameter β varies: for both $\gamma = 1, 2$, we find that for all $\beta < \beta_c(\gamma)$ all eigenvalues have negative real part; at $\beta_c(\gamma)$ two eigenvalues are conjugate and purely imaginary, suggesting the possible presence of a Hopf bifurcation in the limit dynamics. These critical values of β are then compared to the ones obtained by simulating the finite particle system in Section 3.2.3, finding a very good accordance.

In the second model, studied in Section 3.3, the non-Markovianity follows by an augmentation of state procedure, where we double the state space assigning to each microscopic spin another spin-valued variable which produces frustration in the system. Specifically, the state of the i-th particle in the system is identified by a pair of spin-valued variables $(x_i, y_i) \in \{-1, 1\}^2$. The dynamics is given in terms of a continuous time spin-flip type Markov chain on the augmented state space $\{-1, 1\}^{2N}$, where each particle flips one component of its state independently conditioned on the current state of the population, with rates

$$\begin{cases} x_i \to -x_i & \text{with rate} \quad (1 - \varepsilon x_i y_i) e^{-\beta x_i m_x^N}, \\ y_i \to -y_i & \text{with rate} \quad e^{\gamma y_i m_x^N}, \end{cases} \tag{9}$$

where $\gamma, \beta \geq 0$, $0 \leq \varepsilon \leq 1$, and $m_x^N := \frac{1}{N}\sum_{i=1}^N x_i$ is the magnetization of the spins x_i's. Note that, when $\varepsilon = 0$, the restriction of the dynamics to the x_i's is of Curie–Weiss spin-flip type. In addition to the empirical magnetization m_x^N of the x_i's, we also define the analogous quantity m_y^N for the y_i's and

$$m_{xy}^N := \frac{1}{N}\sum_{i=1}^N x_i y_i.$$

It turns out that $\Big((m_x^N(t), m_y^N(t), m_{xy}^N(t))\Big)_{t\geq 0}$ is an order parameter for the above model, in the sense that its dynamics, induced by (9), is Markovian. In the limit

$$\Big((m_x^N(t), m_y^N(t), m_{xy}^N(t))\Big)_{t\geq 0} \to \Big((x(t), y(t), w(t))\Big)_{t\geq 0}$$

for $N \to +\infty$, the macroscopic variables $\Big((x(t), y(t), w(t))\Big)_{t\geq 0}$ satisfy

$$\begin{cases} \dot{x}(t) = -2x(t)\cosh(\beta x(t)) + 2\sinh(\beta x(t)) + 2\varepsilon y(t)\cosh(\beta x(t)) - 2\varepsilon w(t)\sinh(\beta x(t)), \\ \dot{y}(t) = -2y(t)\cosh(\gamma x(t)) - 2\sinh(\gamma x(t)), \\ \dot{w}(t) = -2w(t)\cosh(\gamma x(t)) - 2x(t)\sinh(\gamma x(t)) - 2w(t)\cosh(\beta x(t)) + 2y(t)\sinh(\beta x(t)) \\ \qquad\qquad + 2\varepsilon\cosh(\beta x(t)) - 2\varepsilon x(t)\sinh(\beta x(t)), \\ x(0) = x_0, \quad y(0) = y_0, \quad w(0) = w_0. \end{cases} \tag{10}$$

The limit evolution is thus finite-dimensional, allowing for a deeper analysis of the phase-space diagram with respect to the previous model. After proving the well-posedeness of

System (10) in Section 3.3.1 (Proposition 3.7), in Section 3.3.2 we perform a linear analysis around the disordered equilibrium $(0, 0, \varepsilon/2)$, studying the local phase-diagram when the interaction parameters vary, proving the existence of a supercritical Hopf bifurcation for certain critical values of the parameters (Propositions 3.9 and 3.10); in Section 3.3.3 we find numerically all the equilibria of the dynamics; in Section 3.3.4 we study numerically the local character of the previously found equilibria; finally, in Section 3.3.5 we give detailed illustrations of the dynamics and of the global phase-diagram, via numerical simulations of the macroscopic equations and resorting to the previous analyses.

The results of Chapter 3 strongly suggest that the above models belong to the same universality class: they both feature the presence of a unique stable neutral phase for values of the parameters corresponding to high temperatures, the emergence of periodic orbits in an intermediate range of the parameter values, and a subsequent ferromagnetic ordered phase for increasingly lower temperatures. In particular, both dynamics can generate self-sustained oscillations.

One of the goals of the related literature is to understand which types of microscopic interactions and mechanisms can lead to or enhance the emergence of self-sustained rhythms, in systems where each individual particle has no natural tendency to behave periodically. Although not proved in general, a strong belief in the literature is that, at least for Markovian dynamics, self-sustained oscillations cannot take place if one does not introduce some time-irreversible phenomenon in the dynamics ([10, 60]). While the finite-dimensional model treated in Section 3.3 falls within the above literature (due to the presence of frustration, which is an irreversible phenomenon), the model of Section 3.2, in which we observe that the limit dynamics is still reversible with respect to the stationary distribution around which cycles emerge (see Remark 3.3), suggests that this paradigm could be false for the non-Markovian case.

Overview of Chapter 4

The results analyzed in this chapter are collected from an ongoing work of the author with P. Dai Pra and M. Formentin.

Chapter 4 is devoted to the study of a model of interacting spins with a hierarchical mean field structure, thus serving as an attempt to relax the mean field assumption as well as the Markovianity of the spins through a state augmentation procedure. We refer to the beginning of the chapter for an introduction on hierarchical mean field models and the related literature.

Let V be a set, indexing individuals in a population. Each individual $r \in V$ is identified with a pair of variables (σ_r, x_r): a spin variable $\sigma_r \in \{-1, 1\}$, and a continuous one $x_r \in \mathbb{R}$, representing some aggregated statistics of the remaining characteristics of the individual (and thus being naturally normally distributed for e.g. by a central limit theorem), which would otherwise not be accounted for in the modelization by a spin system. The interaction between each pair of spin variables $\sigma_r, \sigma_s \in V$ is encoded in a variable $J_{rs} \in \mathbb{R}$. Analogously, x_r and x_s interact with a strength proportional to some variables $J'_{rs} \in \mathbb{R}$. The particles $(\sigma_r, x_r)_{r \in V}$ follow stochastic dynamics given by

$$\begin{cases} \sigma_r \mapsto -\sigma_r, \qquad \text{with rate} \quad 1 + \tanh\left[-\sigma_r \sum_{s \in V} J_{rs}(\sigma_s + x_s)\right], \\ dx_r = -\sum_{s \in V} J'_{rs}(x_r - x_s)dt + \sigma dW_r(t), \end{cases} \tag{11}$$

where $W_r(t)$'s are $|V|$ independent Brownian motions, and $\sigma > 0$ is the diffusion coefficient.

As we describe below, the peculiarity of the model consists in the fact that the interaction among different particles scales with what in the related literature (see e.g.

[70]) is referred to as *hierarchical distance*. The main goal of our study is to obtain a limit description of dynamics (11), as $N \to +\infty$, at different spatio-temporal scales, analyzing the possible presence of phase transitions in the system. In this framework, we mainly focus on two choices for V and (deterministic) interaction parameters J_{rs} and J'_{rs}:

- *Ferromagnetic mean field case:*

$$\begin{aligned} V &:= \{1, \dots, N\}, \\ J_{rs} &= \frac{\beta}{N}, \\ J'_{rs} &= \frac{\alpha}{N}, \end{aligned} \tag{12}$$

  with $\alpha, \beta \geq 0$.

- *Ferromagnetic two-level hierarchical case:*

$$V := \{1, \dots, N\} \times \{1, \dots, N\}, \\ \begin{cases} J_{rs} = \frac{\beta_1}{N}, \quad J'_{rs} = \frac{\alpha_1}{N}, \qquad \text{if } |r-s| \leq 1, \\ J_{rs} = \frac{\beta_2}{N^2}, \quad J'_{rs} = \frac{\alpha_2}{N^3}, \qquad \text{if } |r-s| = 2, \end{cases} \tag{13}$$

  with $\alpha_1, \alpha_2, \beta_1, \beta_2 \geq 0$, where the distance $|\cdot|$ between $r := (r_1, r_2)$ and $s := (s_1, s_2)$ is defined by

$$|r-s| := \begin{cases} 0, & \text{if } r_1 = s_1,\ r_2 = s_2 \\ 1, & \text{if } r_1 \neq s_1,\ r_2 = s_2 \\ 2, & \text{otherwise.} \end{cases}$$

In particular, with the term two-level mean field hierarchy we mean a system of N interacting mean field systems of particles, where each mean field is comprised of N particles itself, and the strength of the interaction among different particles scales with their hierarchical distance $|\cdot|$, which can either be 1, if the particles belong to the same mean field, or 2, when they belong to different mean fields.

The hierarchical construction can be reiterated a finite number of times to define a k-level hierarchical model, where $V := \{1, \dots, N\}^k$, $J_{rs} \propto \frac{1}{N^l}$, $J'_{rs} \propto \frac{1}{N^{2l-1}}$ for $|r-s| = l$, with $l = 1, \dots, k$. See Section 4.3.5 for details on the notion of hierarchical distance and for the generalization of our results to each hierarchical level in the subcritical regime, stated in Conjecture 4.1.

The mean field case

The model (11) under the mean field assumptions (12) is analyzed in Section 4.2.

In this case, denote by $(\boldsymbol{\sigma}, \boldsymbol{x}) = (\sigma_j, x_j)_{j=1,\dots,N} \in \mathbb{R}^N \times \{-1, 1\}^N$ a configuration of the entire population. Let $m^N(t) := \frac{1}{N}\sum_{i=1}^N \sigma_i(t)$, and $x^N(t) := \frac{1}{N}\sum_{i=1}^N x_i(t)$ be the empirical averages of the spins and diffusions respectively, which evolve as

$$\begin{cases} m^N \mapsto\ m^N \pm \frac{2}{N}, \qquad \text{with rate} \quad N\frac{1 \mp m^N(t)}{2}\left[1 \pm \tanh(\beta(x^N(t) + m^N(t)))\right], \\ dx^N(t) = \frac{\sigma}{\sqrt{N}} dW(t), \end{cases} \tag{14}$$

where W is a Brownian motion. From the above dynamics, we see that the diffusion $(x^N(t))_{t\geq 0}$ evolves at a timescale of order N, while at times of order 1 it converges to its initial datum for $N \to +\infty$. Assuming $x^N(0) \to x \in \mathbb{R}$ for $N \to +\infty$, it is easy to prove that $(x^N(t), m^N(t)) \to (x, m(t))$ for $N \to +\infty$, where the limit satisfies the ODE (compare with (4.9) - which is stated in alternative variables), parametrized by x,

$$\begin{cases} \dot{m}(t)(x) = 2\tanh(\beta(x + m(t)(x))) - 2m(t)(x), \\ m(0)(x) = m_0(x). \end{cases} \tag{15}$$

Equation (15) features the existence of a phase transition in $\beta = 1$: it possesses a unique stable equilibrium configuration for $\beta < 1$ (*subcritical regime*), whereas for $\beta > 1$ (*supercritical regime*) a region with multiple equilibria, with different stability properties, appears. When we speed up time at order N, the fluctuations of the diffusion $x^N(Nt)$ are not negligible anymore. The main contribution of this section is on the study of the sequence of the accelerated processes $(x^N(Nt), m^N(Nt))_{t\geq 0}$ when $N \to +\infty$. First, we prove that in the subcritical regime (Propositions 4.4 and 4.5), the limit of $(m^N(Nt))_{t\geq 0}$ is a regular diffusion on the unique long-time equilibrium configuration of (15), driven by the limit sped-up diffusion $(x^N(Nt))_{t\geq 0}$. We then address the main result of the section (Theorem 4.7), where we prove that in the supercritical regime such limiting motion turns into a regular diffusion taking place on the two stable branches of equilibria, with *jumps* from one branch to the other when the process reaches the borders of the stable configurations (see Figure 4.3 for a comparison between the two regimes).

The two-level hierarchical case

The two-level hierarchical case, i.e. dynamics (11) with the choices (13), is studied in Section 4.3.

For any $i, j = 1, \ldots, N$, we identify the i-th individual of the j-th population with the pair of state variables (σ_{ij}, x_{ij}). Define the first-level empirical magnetization of the j-th population

$$m_j^N(t) := \frac{1}{N}\sum_{i=1}^N \sigma_{ij}(t),$$

and the analogous quantity for $x_j^N(t)$. Moreover, denote the two-level empirical magnetization as

$$M^N(t) := \frac{1}{N^2}\sum_{i,j=1}^N \sigma_{ij}(t) = \frac{1}{N}\sum_{j=1}^N m_j^N(t),$$

and the same for $X^N(t) := \frac{1}{N^2}\sum_{ij} x_{ij}(t) = \frac{1}{N}\sum_{j=1}^N x_j^N(t)$. With this notation, the first-level averages $(x_j^N(t), m_j^N(t))_{j=1,\ldots,N}$ form a system of N interacting particles, with dynamics

$$\begin{cases} m_j^N \mapsto m_j^N \pm \frac{2}{N}, \text{ rate } N\frac{1\mp m_j^N(t)}{2}\Big[1 \pm \tanh(\beta_1(x_j^N(t) + m_j^N(t)) + \beta_2(X^N(t) + M^N(t)))\Big] \\ dx_j^N(t) = -\frac{\alpha_2}{N}\Big[x_j^N(t) - X^N(t)\Big]dt + \frac{\sigma}{\sqrt{N}}dW_j(t), \end{cases} \tag{16}$$

with W_j's N independent Brownian motions. In particular, we see that the diffusions x_j^N's move at times of order N, while, by summing over j the second equation in (16), we deduce that X^N moves at times of order N^2. Assuming that $x_j^N(0) = x_j \overset{iid}{\sim} \mu_0(dx)$ (with $\mathbb{E}[x_j] = 0$ for simplicity), at times of order 1 we prove (Theorem 4.13) a propagation of chaos for the system of magnetizations (the corresponding property for the diffusions is trivial): namely, $m_j^N(t) \to \tilde{m}_j(t) := m(t)(x_j)$ for $N \to +\infty$, where $m(t)(x)$ solves

$$\begin{cases} \dot{m}(t)(x) = 2\tanh(\beta_1(x + m(t)(x)) + \beta_2 M(t)) - 2m(t)(x), \\ M(t) = \int_{\mathbb{R}} m(t)(x)\mu_0(dx). \end{cases} \tag{17}$$

We observe that the limit i.i.d. processes $\tilde{m}_j(t)$ are of McKean-Vlasov type, because of the integral condition on $M(t)$ in Equation (17). Due to the presence of the latter, the long-time behavior of (17) is harder to study than for its corresponding mean field version (15). For a rigorous analysis on the further timescales, we thus restrict to the subcritical regime $\beta_1 + \beta_2 < 1$, where Equation (17) possesses a unique stable equilibrium configuration (Proposition 4.16) given by the solution to

$$\begin{cases} \overline{m}(x) = \tanh(\beta_1(x + \overline{m}(x)) + \beta_2 \overline{M}), \\ \overline{M} = \int_{\mathbb{R}} \overline{m}(x) \mu_0(dx) = 0, \end{cases} \tag{18}$$

where we assumed as above, for simplicity, $\mathbb{E}[x_j] = 0$. In Theorem 4.18 we control the uniform ergodicity of the N-particle system behavior when we rescale time up to a certain $C(N)$, which is allowed to grow with N.

When we accelerate time further, at a timescale of order N (Section 4.3.3), we prove that the N-particle system $(x_j^N(Nt), m_j^N(Nt))_{j=1,\dots,N}$ still propagates chaos, where the magnetizations are converging to i.i.d. copies of a McKean-Vlasov process living onto the equilibrium configuration curve, driven by the i.i.d. limit non-trivial dynamics of the accelerated diffusions $(x_j^N(Nt))_{t\geq 0}$'s (Theorem 4.20). In this case, the integral McKean-Vlasov condition on $M(t)$ is with respect to the law at time t of the Ornstein-Uhlenbeck diffusion $(x(t))_{t\geq 0}$, the limit of $x_j^N(Nt) \to x(t)$ as $N \to +\infty$, which solves

$$dx(t) = -\alpha_2 x(t)dt + \sigma dW(t).$$

Finally, at a timescale of order N^2 (Section 4.3.4), the diffusions $(x_j^N(N^2 t))_{t\geq 0}$ fastly reach their stationary distribution; we thus characterize the limit of the sequence of second-level empirical magnetizations $(M^N(N^2 t))_{t\geq 0}$, proving that it converges, with respect to all its finite time dimensional distributions, to a limit stochastic process $(M(t))_{t\geq 0}$ (note that at the previous timescales the convergence was to a deterministic object instead), which is the average of the equilibrium curve with respect to a mixture of the stationary distributions of the first-level diffusions x_j^N's (Theorem 4.29), where the motion is driven by the non-trivial limit second-level diffusion $(X(t))_{t\geq 0}$, which solves

$$dX(t) = \sigma dW(t),$$

with W a Brownian motion. After generalizing the above results to the k-level hierarchical case in the subcritical regime (Section 4.3.5), we develop heuristic arguments, reinforced by numerical simulations, for studying the two-level supercritical regime in the limit case of null temperature $\beta_1 = \beta_2 = +\infty$ (Section 4.3.6).

Notation and preliminaries

The notation we adopt in the first part of the Dissertation differs in general from that of the second part. As they are based on works belonging to two different - but related - mathematical literatures (mean field games and mean field interacting spin systems respectively), we try to adhere to the classical notation employed by each corresponding scientific community. While we postpone the specific notation to the beginning of each part, in this section we recall some preliminary classical facts on the common mathematical tools employed throughout the Dissertation.

Let $\mathbb{R}^+$ denote the positive real numbers, including 0. Let E be a Polish space. Every model we consider features E-valued processes with sample paths taking values in (or being embedded in) the Skorohod space of càdlàg functions, which we denote by $\mathcal{D}(\mathbb{R}^+; E)$. When not specified otherwise, we refer to the weak convergence of stochastic processes in the above Skorohod space: namely, we say that a sequence $(X_n)_{n\geq 1}$ of stochastic processes converges to a limiting process X if, for any $T > 0$, X_n converges in distribution on the path space $\mathcal{D}([0,T]; E)$ to X, as $n \to +\infty$. In what follows, we consider different choices for E depending on the context: $E := \mathbb{R}^d, E := \mathbb{R}^+ \times \{-1, 1\}$, $E := [-1, 1]$ and $E := \mathbb{R} \times [-1, 1]$ should cover all the possible options.

A general theme is about obtaining macroscopic descriptions for stochastic systems with a large numbers of individuals (either controlled, in which case we restrict to Nash equilibria configurations, or uncontrolled). In this regard, an important role is played by the so-called *propagation of chaos* property. We recall its definition (see e.g. [67])

Definition (Propagation of chaos). *Let Q be a probability measure on E and Q^N a probability measure on E^N. The sequence $(Q^N)_{N\geq 1}$ is Q-chaotic if for any fixed integer k and any continuous bounded functions $f_1, \ldots, f_k$ on E,*

$$\lim_{N\to\infty} \langle Q^N, f_1 \otimes \ldots \otimes f_k \otimes 1^{N-k} \rangle = \prod_{i=1}^{k} \langle Q, f_i \rangle.$$

The above definition means that, asymptotically in N, any k coordinates become independent, all with the same distribution Q. In the context of N interacting stochastic processes, we apply the above definition by considering Q^N to be the joint law of the N processes on the product path space, and Q the law of some limiting process. All the processes which we consider in the Dissertation enjoy additional symmetries, due to the mean field type interactions involved: in particular, the joint law is always invariant by permutation of the individual components (this property is referred to as *exchangeability*). For such systems, proving the above propagation of chaos is equivalent to proving a Law of Large Numbers for the associated empirical measure processes (see e.g. [67, Prop. 4.2]), which we interchangeably refer to as the propagation of chaos property.

When needed, we endow the space of probability measures on $(E, |\cdot|)$ with finite first moment, $\mathcal{P}_1(E)$, with the 1-Wasserstein distance, which we denote by $\boldsymbol{d_1}$,

$$\boldsymbol{d_1}(\mu,\nu) := \inf_{\gamma\in\Gamma(\mu,\nu)} \int_{E\times E} |x-y| d\gamma(x,y),$$

with Γ the set of measures on $E \times E$ having first and second marginals equal to μ and ν respectively. We recall an inequality which we use repeatedly. It follows easily by the Kantorovich-Rubinstein duality theorem (see e.g. [91, Ch. 5]): fix $(x_1, \ldots, x_N)$ and $(y_1, \ldots, y_N) \in E^N$, and let $\mu^N := \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$, $\nu^N := \frac{1}{N}\sum_{i=1}^N \delta_{y_i}$ be two empirical measures. Then,

$$\boldsymbol{d_1}(\mu^N, \nu^N) \le \frac{1}{N}\sum_{i=1}^N |x_i - y_i|. \tag{19}$$

Part I

Finite state mean field games

CHAPTER 1

The uniqueness case: convergence, fluctuations and large deviations via the master equation

In this chapter we discuss the convergence problem, analyzed in [26], for finite state symmetric N-player games under uniqueness assumptions on the limit. The limit dynamics is given by a finite state mean field game system made of two coupled forward-backward ODEs. We exploit the master equation approach (introduced in [15] for the diffusive case), which in this finite-dimensional framework is a first order PDE in the simplex of probability measures, obtaining the convergence of the feedback Nash equilibria, the value functions and the optimal trajectories. The convergence argument requires only the regularity of a solution to the master equation. Moreover, we show that the convergence results imply the propagation of chaos as well as refined asymptotics for the N-player empirical measures, in terms of a Central Limit Theorem and a Large Deviation Principle. The key point for proving such results is to compare the prelimit optimal trajectories with the ones in which each player chooses the control induced by the master equation. The fluctuations are then found by analyzing the associated infinitesimal generator, while the Large Deviation properties are derived using a result in [54]. Finally, we study the well-posedness and regularity of solution to the master equation under monotonicity assumptions.

Finite state mean field games have been studied by several authors in the last years, starting from [62] in discrete time, and then in the continuous time setting by [61] and [68], in the latter with applications to graphs. For a probabilistic approach to finite state mean field games we refer to [24]. On the convergence problem, a first result was given in [62], but only for a small enough time horizon. An equation similar to the master equation of this chapter, but holding in the whole space $\mathbb{R}^d$, was analyzed in [83], proving the well-posedness and regularity under stronger assumptions. The works [57] and [61] deal also with the problem of convergence, as T tends to infinity, to the stationary mean field game. The master equation was formally discussed in [64], [65] as well as in [61], in the first two with a particular focus on the two state problem, a context which is related to the model we study in Chapter 2. On numerical methods, we acknowledge the work [63] in the finite state case under monotonicity. A class of mean field games with major and minor agents was analyzed in [22], showing the relation with the N-player game in the approximation direction. Finally, we mention [5], which appeared online together with [26], in which the authors independently obtain the same convergence results we prove here, by using again the master equation approach of [15], but considering a probabilistic representation of the dynamics different from ours. Moreover, they also obtain a Central

Limit Theorem for the fluctuations of the empirical measure processes. However, they prove it in a different way, that is, via a martingale Central Limit Theorem.

Let us mention that a Central Limit Theorem and a Large Deviation Principle for mean field games, enriched by a concentration of measure result, were then established also in the diffusive case, via the master equation approach, in the two separate works [43, 44].

1.1 Introducing the model

In this section we introduce the equations in play at a formal level. Let $X_i(t)$ be the state of the i-th player at time t. The dynamics of the N players is given by the system of controlled SDEs:

$$X_i(t) = Z_i + \int_0^t \int_\Xi f(X_i(s^-), \xi, \alpha^i(s, \boldsymbol{X}_{s^-}))\mathcal{N}_i(ds, d\xi), \tag{1.1}$$

for $i = 1, \dots, N$, where each $X_i(t)$ is a process taking values in the finite space $\Sigma = \{1, \dots, d\}$ and we denote by $\boldsymbol{X}_t := (X_1(t), \dots, X_N(t))$ the vector of the N processes; $\mathcal{N}_i$ are N i.i.d. Poisson measures on $[0, T] \times \Xi$, with $\Xi \subset \mathbb{R}^d$, and the controls $\alpha^i \in A \subset \mathbb{R}^d$ are only in feedback form. The function f is crucial for the definition of the dynamics (1.1): it models the possible jumps of the Markov chain, while the Poisson measures prescribe their random occurrences. Following an idea of [66] which we repeatedly use throughout the Dissertation, we define the function f so that the control $\alpha^i_y(t, x, \boldsymbol{x}^{N,i}_t)$ represents the rate at which player i decides to go from state x to state y, when $x \neq y$, $\boldsymbol{x}^{N,i}_t$ being the states of the other $N-1$ players at time t; c.f. (1.2) and (1.18) below. Let us remark that, while Cardaliaguet et al. ([15]) study the convergence problem also in the presence of a noise (Brownian motion) common to all the players, which makes things even more difficult, we do not consider here any common noise. In the discrete setting, this would result in considering dynamics with simultaneous jumps, which can be realized by adding another Poisson measure in (1.1), common to all the players.

In our framework, we show that there exists a unique feedback Nash equilibrium for the N-player game. It is provided by the Hamilton-Jacobi-Bellman (HJB) system of Nd^N coupled ODE's

$$\begin{cases} -\frac{\partial v}{\partial t}^{N,i}(t, \boldsymbol{x}) - \sum_{j=1,\, j\neq i}^N \alpha^*(x_j, \Delta^j v^{N,j}) \cdot \Delta^j v^{N,i} + H(x_i, \Delta^i v^{N,i}) = F^{N,i}(\boldsymbol{x}), \\ v^{N,i}(T, \boldsymbol{x}) = G^{N,i}(\boldsymbol{x}). \end{cases} \tag{HJB}$$

In the above equation, $F^{N,i}$ and $G^{N,i}$ are respectively the running and terminal costs, H is the Hamiltonian and α^* its unique maximizer, and

$$\Delta^j g(\boldsymbol{x}) := (g(x_1, \dots, y, \dots, x_N) - g(x_1, \dots, x_j, \dots, x_N))_{y=1,\dots,d} \in \mathbb{R}^d$$

denotes the finite difference of a function $g(\boldsymbol{x}) = g(x_1, \dots, x_N)$ with respect to its j-th entry.

The study of convergence consists in finding a limit for System (HJB) as N tends to infinity. To this end, we assume symmetric properties of the game. Namely, the costs $F^{N,i}$ and $G^{N,i}$ satisfy the mean field assumptions, i.e. there exist two functions F and G such that $F^{N,i}(\boldsymbol{x}) = F(x_i, m^{N,i}_{\boldsymbol{x}})$ and $G^{N,i}(\boldsymbol{x}) = G(x_i, m^{N,i}_{\boldsymbol{x}})$, where $m^{N,i}_{\boldsymbol{x}}$ denotes the empirical measure of all the players except for the i-th, which belongs to $P(\Sigma)$, the space of probability measures on Σ. Thanks to these mean field assumptions, we shall say that the solution $v^{N,i}$ of System (HJB) can be found in the form $v^{N,i}(t,\boldsymbol{x}) = V^N(t, x_i, m_{\boldsymbol{x}}^{N,i})$, for a suitable function V^N of time, space and measure; this makes the convergence problem more tractable. At a formal level, we can introduce the limit equation assuming the existence of a function U such that $V^N(t, x_i, m_{\boldsymbol{x}}^{N,i}) \sim U(t, x_i, m_{\boldsymbol{x}}^{N,i})$ for large N. Then, let us analyze the different components of System (HJB) and which should be their corresponding limits in terms of U. First, the i-th difference of $v^{N,i}$ should converge to

$$\begin{aligned}\Delta^i v^{N,i}(t,\boldsymbol{x}) &= \left(v^{N,i}\left(t, y, m_{\boldsymbol{x}}^{N,i}\right) - v^{N,i}\left(t, x_i, m_{\boldsymbol{x}}^{N,i}\right)\right)_{y=1,\dots,d} \\ &\to \left(U(t,y,m) - U(t,x_i,m)\right)_{y=1,\dots,d} = \Delta^x U(t,x_i,m).\end{aligned}$$

For $j \neq i$ we should instead get

$$\begin{aligned}&\Delta^j v^{N,i}(t,\boldsymbol{x}) = \\ &\quad\left(v^{N,i}\left(t, x_i, \frac{1}{N-1}\sum_{k\neq j,i}\delta_{x_k} + \frac{1}{N-1}\delta_y\right) - v^{N,i}\left(t, x_i, \frac{1}{N-1}\sum_{k\neq i}\delta_{x_k}\right)\right)_{y=1,\dots,d} \\ &\quad\sim \frac{1}{N-1} D^m U(t, x_i, m_{\boldsymbol{x}}^{N,i}, x_j),\end{aligned}$$

modulo terms of order $O(1/N^2)$, where a precise definition of $D^m U$, the derivative with respect to a probability measure, will be given in the next section. Then, $H(x_i, \Delta^i v^{N,i}) \to H(x_i, \Delta^x U)$, and we should obtain

$$\begin{aligned}&\sum_{j=1, j\neq i}^{N} \alpha^*(x_j, \Delta^j v^{N,j}) \cdot \Delta^j v^{N,i} \\ &\qquad\sim \frac{1}{N-1}\sum_{j=1,j\neq i}^{N} \alpha^*(x_j, \Delta^x U(t, x_j, m_{\boldsymbol{x}}^{N,i})) \cdot D_m U(t, x_i, m_{\boldsymbol{x}}^{N,i}, x_j) \\ &\qquad\sim \int_\Sigma \alpha^*(y, \Delta^y U(t, y, m_{\boldsymbol{x}}^{N,i})) \cdot D^m U(t, x_i, m_{\boldsymbol{x}}^{N,i}, y) dm_{\boldsymbol{x}}^{N,i}(y) \\ &\qquad\to \int_\Sigma D^m U(t,x,m,y) \cdot \alpha^*(y, \Delta^y U(t,y,m)) dm(y).\end{aligned}$$

Thus, we are able to introduce the master equation, that is the equation to which we would like to prove convergence

$$\begin{cases} -\frac{\partial U}{\partial t} + H(x, \Delta^x U) - \int_\Sigma D^m U(t,x,m,y)\cdot \alpha^*(y, \Delta^y U(t,y,m)) dm(y) = F(x,m), \\ U(T,x,m) = G(x,m), \quad (x,m) \in \Sigma \times P(\Sigma), \quad t \in [0,T]. \end{cases} \tag{M}$$

It is a first order PDE in $P(\Sigma)$, the simplex of probability measures in $\mathbb{R}^d$. We solve it using the strategy developed in [15], which relies on the method of characteristics. Indeed, as remarked above, the classical mean field game system can be seen as the characteristic curves of (M). In our finite state setting, the mean field game system consists of two coupled ODEs: a Hamilton-Jacobi-Bellman equation giving the value function of the limit control problem and a Kolmogorov-Fokker-Planck describing the evolution of the limit deterministic flow of probability measures. We solve the mean field game system for any initial time and initial distribution: this defines a candidate solution to (M) and, in order to prove that it is differentiable with respect to the initial condition, we introduce and analyze a linearized mean field game system. To prove the well posedness of (M) for any time horizon, the sufficient hypotheses we make are the monotonicity assumptions of Lasry and Lions. However, we stress again that these assumptions play no role in the convergence argument, as it requires only the existence of a regular solution to (M).

The rest of the chapter is organized as follows. In Section 1.2, we start with the notation and the definition of derivatives in the simplex. So we present the two sets of assumptions we make use of: one for the convergence, the fluctuations and the large deviation results, while the other, stronger, for the well posedness of the master equation; we also show an example in which the assumptions are satisfied. Then we give a detailed description of both the N-player game and the limit model. Section 1.3 contains the convergence results and their proofs, while in Section 1.4 we employ the convergence argument to derive refined asymptotics for the empirical measure process, that is, a Central Limit Theorem and a Large Deviation Principle. Section 1.5 analyzes the well-posedness and regularity of the solution to the master equation. We conclude with Section 1.6 by summarizing all the main results.

1.2 Model and assumptions

1.2.1 Notation

Here we briefly clarify the notation used throughout the chapter. Part of the notation is employed also in Chapter 2. First of all, we are considering $\Sigma = \{1, \dots, d\}$ to be the finite state space of any player. Let T be the finite time horizon and $A := [\kappa, M]^d$, for $\kappa, M > 0$, be the compact space of control values. Denote by

$$P(\Sigma) := \Big\{ m \in \mathbb{R}^d \ : \ m_j \geq 0, \quad m_1 + \cdots + m_d = 1 \Big\}$$

the space of probability measures on Σ. Besides the euclidean distance in $\mathbb{R}^d$, denoted with $|\cdot|$, we may interchangeably use the Wasserstein metric $\boldsymbol{d}_1$ on $P(\Sigma)$ since all metrics are equivalent. We observe that the simplex $P(\Sigma)$ is a compact and convex subset of $\mathbb{R}^d$.

Let $\Xi := [0, M]^d$. In the dynamics given by (1.1), the function $f : \Sigma \times \Xi \times A \to \{-d, \dots, d\}$ modeling the jumps has to be a measurable function such that $f(x, \xi, a) \in \{1 - x, \dots, d - x\}$. Specifically, throughout the chapter we set, for $x \in \Sigma$, $\xi = (\xi_y)_{y \in \Sigma}$ and $a = (a_y)_{y \in \Sigma}$,

$$f(x, \xi, a) := \sum_{y \in \Sigma} (y - x) \mathbb{1}_{]0, a_y[}(\xi_y). \tag{1.2}$$

The measures $\mathcal{N}_i$ appearing in (1.1) are N i.i.d. stationary Poisson random measures on $[0, T] \times \Xi$, with intensity measure ν on Ξ given by

$$\nu(E) := \sum_{j=1}^{d} \ell(E \cap \Xi_j), \tag{1.3}$$

for any E in the Borel σ-algebra $\mathcal{B}(\Xi)$ of Ξ, where $\Xi_j := \{u \in \Xi \ : \ u_i = 0 \ \ \forall \, i \neq j\}$ is viewed as a subset of $\mathbb{R}$, and ℓ is the Lebesgue measure on $\mathbb{R}$. We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and denote by $\mathbb{F} = (\mathcal{F}_t)_{t \in [0,T]}$ the filtration generated by the Poisson measures. These definitions of f and ν ensure that the control is exactly the transition rate of the Markov chain; see (1.18) below.

The initial datum of the N-player game is represented by N i.i.d. random variables $Z_1, \dots, Z_N$ with values in Σ and distributed as $m_0 \in P(\Sigma)$. The vector $\boldsymbol{Z} = (Z_1, \dots, Z_N)$ is in particular *exchangeable*, in the sense that the joint distribution is invariant under permutations, and is assumed to be $\mathcal{F}_0$-measurable, i.e. independent of the noise.

The state of player i at time t is denoted by $X_i(t)$, with $\boldsymbol{X}_t := (X_1(t), \ldots, X_N(t))$. The trajectories of each X_i are in $\mathcal{D}([0,T];\Sigma)$, the space of càdlàg functions from $[0,T]$ to Σ endowed with the Skorokhod metric. For $\boldsymbol{x} = (x_1, \ldots, x_N) \in \Sigma^N$, denote the empirical measures

$$m_{\boldsymbol{x}}^N := \frac{1}{N}\sum_{j=1}^N \delta_{x_j} \qquad m_{\boldsymbol{x}}^{N,i} := \frac{1}{N-1}\sum_{j=1, j\neq i}^N \delta_{x_j}.$$

Thus, $m_{\boldsymbol{X}}^N(t) := m_{\boldsymbol{X}_t}^N$ is the empirical measure of the N players and $m_{\boldsymbol{X}}^{N,i}(t) := m_{\boldsymbol{X}_t}^{N,i}$ is the empirical measure of all the players except the i-th. Clearly, they are $P(\Sigma)$-valued stochastic processes. In the limit dynamics, the empirical measure is replaced by a deterministic flow of probability measures $m : [0,T] \to P(\Sigma)$.

In choosing his/her strategy, each player minimizes the sum of three costs: a Lagrangian $L : \Sigma \times A \longrightarrow \mathbb{R}$, a running cost $F : \Sigma \times P(\Sigma) \longrightarrow \mathbb{R}$ and a final cost $G : \Sigma \times P(\Sigma) \longrightarrow \mathbb{R}$ (see next section for the precise definition of the N-player game). The Hamiltonian H is defined as the Legendre transform of L:

$$H(x,p) := \sup_{\alpha \in A} \left\{-\alpha \cdot p - L(x,\alpha)\right\}, \tag{1.4}$$

for $x \in \Sigma$ and $p \in \mathbb{R}^d$.

Given a function $g : \Sigma \to \mathbb{R}$ we denote its first finite difference $\Delta g(x) \in \mathbb{R}^d$ by

$$\Delta g(x) := \begin{pmatrix} g(1) - g(x) \\ \cdot \\ \cdot \\ \cdot \\ g(d) - g(x) \end{pmatrix}.$$

When we have a function $g : \Sigma^N \to \mathbb{R}$, we denote with $\Delta^j g(\boldsymbol{x}) \in \mathbb{R}^d$ the first finite difference with respect to the j-th coordinate, namely

$$\Delta^j g(\boldsymbol{x}) := \begin{pmatrix} g(x_1, \ldots, x_{j-1}, 1, x_{j+1}, \ldots, x_N) - g(\boldsymbol{x}) \\ \cdot \\ \cdot \\ \cdot \\ g(x_1, \ldots, x_{j-1}, d, x_{j+1}, \ldots, x_N) - g(\boldsymbol{x}) \end{pmatrix}.$$

For future use, let us observe that, for $g : \Sigma \to \mathbb{R}$,

$$|\Delta g(x)| \le \max_y [\Delta g(x)]_y \le 2 \max_x |g(x)| \le C|g|. \tag{1.5}$$

For a function $u : [t_0, T] \times \Sigma \longrightarrow \mathbb{R}$, we denote

$$||u|| := \sup_{t \in [t_0,T]} \max_{x \in \Sigma} |u(t,x)|. \tag{1.6}$$

We also use the notation $u(t) := (u_1(t), \ldots, u_d(t)) = (u(t,1), \ldots, u(t,d))$. When considering a function u with values in $\mathbb{R}^d$, its norm is defined as in (1.6), but where $|\cdot|$ denotes the euclidean norm in $\mathbb{R}^d$.

We now introduce the concept of variation with respect to a probability measure m of a function $U : P(\Sigma) \to \mathbb{R}$. Let us remark that the usual notion of gradient cannot be defined for such a function: since the domain is $P(\Sigma)$ we are not allowed to define e.g. the directional derivative $\frac{\partial}{\partial m_1}$, as we would have to extend the definition of U outside the simplex.

Definition 1.1. *We say that a function $U : P(\Sigma) \to \mathbb{R}$ is differentiable if there exists a function $D^m U : P(\Sigma) \times \Sigma \to \mathbb{R}^d$ given by*

$$[D^m U(m,y)]_z := \lim_{s \to 0^+} \frac{U(m + s(\delta_z - \delta_y)) - U(m)}{s}. \tag{1.7}$$

for $z = 1, \dots, d$. Moreover, we say that U is C^1 if the function $D^m U$ is continuous in m.

Morally, we can think of $[D^m U(m,y)]_z$ as the (right) directional derivative of U with respect to m along the direction $\delta_z - \delta_y$. We also observe that $m + s(\delta_z - \delta_y)$ might be outside the probability simplex (e.g. when we are at the boundary), in which case we consider the limit only across admissible directions. However, note that, for our purposes, this is not really a problem: since in the limit $m(t)$ will be the distribution of the reference player, the bound from below for the control ensures that the boundary of the simplex will never be touched.

Together with the definition, we state an identity which will come useful in the following sections:

$$[D^m U(m,y)]_z = [D^m U(m,x)]_z + [D^m U(m,y)]_x, \tag{1.8}$$

for any $x, y, z \in \Sigma$. Its derivation is an immediate consequence of the linearity of the directional derivative.

We can easily extend the above definition to the case of derivative with respect to a direction $\mu \in P_0(\Sigma)$, with

$$P_0(\Sigma) := \left\{ \mu \in \mathbb{R}^d \ : \ \mu_1 + \dots + \mu_d = 0 \right\}.$$

Indeed, an element $\mu = (\mu_1, \dots, \mu_d) = \sum_{z \in \Sigma} \mu_z \in P_0(\Sigma)$ can be rewritten as a linear combination of $\delta_z - \delta_y$ as follows

$$\mu = \sum_{z \neq y} \mu_z (\delta_z - \delta_y),$$

for each $y \in \Sigma$, since $\sum_{z \neq y} \mu_z (\delta_z - \delta_y) = \sum_{z \neq y} \mu_z \delta_z - \left(\sum_{z \neq y} \mu_z \right) \delta_y$, and $\sum_{z \neq y} \mu_z = -\mu_y$.

This remark allows us to define the derivative of $U(m)$ along the direction $\mu \in P_0(\Sigma)$ as a map $\frac{\partial}{\partial \mu} U : P(\Sigma) \times \Sigma \to \mathbb{R}$, defined for each $y \in \Sigma$ by

$$\frac{\partial}{\partial \mu} U(m,y) := \sum_{z \neq y}^{d} \mu_z \left[D^m U(m,y) \right]_z = \mu \cdot D^m U(m,y), \tag{1.9}$$

where the last equality comes from the fact that $[D^m U(m,y)]_y = 0$.

We also note that the definition of $\frac{\partial}{\partial \mu} U(m,y)$ does not actually depend on y, i.e.

$$\frac{\partial}{\partial \mu} U(m,y) = \frac{\partial}{\partial \mu} U(m,1) \tag{1.10}$$

for every $y \in \Sigma$ and for this reason we will fix $y = 1$ when needed in the equations. Indeed, by means of identity (1.8) and the fact that $\mu \in P_0(\Sigma)$, for each $y \in \Sigma$

$$\begin{aligned} \frac{\partial}{\partial \mu} U(m,1) &= \sum_{z=1}^{d} \mu_z \left[D^m U(m,1) \right]_z = [\text{identity (1.8)}] \\ &= \sum_{z=1}^{d} \left([D^m U(m,y)]_z + [D^m U(m,1)]_y \right) \mu_z \end{aligned}$$

$$= \sum_{z=1}^{d} [D^m U(m,y)]_z \mu_z + [D^m U(m,1)]_y \sum_{z=1}^{d} \mu_z$$
$$= \sum_{z=1}^{d} [D^m U(m,y)]_z \mu_z = \frac{\partial}{\partial \mu} U(m,y).$$

For a function $U : \Sigma \times P(\Sigma) \to \mathbb{R}$ we denote the variation with respect to the first coordinate in a point $(x,m) \in \Sigma \times P(\Sigma)$ by $\Delta^x U(x,m)$. Also, denote by $\Gamma^\dagger$ the transpose of a matrix Γ.

1.2.2 Assumptions

We now summarize the assumptions we make, which can vary according to the different results.

Because of the compactness of A, the continuity of L with respect to its second argument is sufficient for guaranteeing the existence and finiteness of the supremum in (1.4) for each (x,p). Moreover, we assume that there exists a unique maximizer $\alpha^*(x,p)$ in the definition of H for every (x,p):

$$\alpha^*(x,p) := \arg\min_{\alpha \in A} \{L(x,\alpha) + \alpha \cdot p\} = \arg\max_{\alpha \in A} \{-L(x,\alpha) - \alpha \cdot p\}. \tag{1.11}$$

With our choices for f in (1.2) and the intensity measure ν in (1.3), a sufficient condition for the above assertion is given by the strict convexity of L in α (see Lemma 3 in [24]). If L is uniformly convex, such optimum α^* is globally Lipschitz in p, and whenever H is differentiable it can be explicitly expressed as $\alpha^*(x,p) = -D_p H(x,p)$; see Proposition 1 in [61] for the proof.

We will work with two sets of assumptions on H. We first observe that it is enough to give hypotheses for $H(x,\cdot)$ on a sufficiently big compact subset of $\mathbb{R}^d$, i.e. for $|p| \le K$, because of the uniform boundedness of $\Delta^i v^{N,i}$: see next section for details (Remark 1.4). In what follows, the constant K is fixed:

(H1) If $|p| \le K$ then H and α^* are Lipschitz continuous in p.

We stress the fact that the above assumptions, together with the existence of a regular solution to (M), are alone sufficient for proving the convergence of the N-player game to the limit mean-field game dynamics.

In order to establish the well-posedness and the needed regularity for the master equation we make use of the following additional assumptions:

(RegH) If $|p| \le K$, H is C^2 with respect to p; H, $D_p H$ and $D^2_{pp} H$ are Lipschitz in p and the second derivative is bounded away from 0, i.e. there exists a constant C such that

$$D^2_{pp} H(x,p) \ge C^{-1}; \tag{1.12}$$

(Mon) The cost functions F and G are monotone in m in the Lasry-Lions sense, i.e., for every $m, m' \in P(\Sigma)$,

$$\sum_{x \in \Sigma} (F(x,m) - F(x,m'))(m(x) - m'(x)) \ge 0, \tag{1.13}$$

and the same holds for G;

(RegFG) The cost functions F and G are C^1 with respect to m, with D^mF and D^mG bounded and Lipschitz continuous. In this case (1.13) is equivalent to say that

$$\sum_x \mu_x[D^m F(x,m,1)\cdot\mu] \geq 0 \tag{1.14}$$

for any $m \in P(\Sigma)$ and $\mu \in P_0(\Sigma)$.

Observe that the assumptions on H allow for quadratic Hamiltonian. As we will see, the above assumptions imply both the boundedness and Lipschitz continuity of $\Delta^x U$ and $D^m U$ with respect to m. We conclude the section with an example for which all the assumptions are satisfied.

Example 1.1. The easiest example for the costs F and G is $F(x,m) = G(x,m) = m(x)$. Slightly more in general, one can consider $F(x,m) = \nabla\phi(m)(x)$, ϕ being a real convex function on $\mathbb{R}^d$.

For the choice of the Lagrangian L, a bit of work is needed in order to recover the regularity for H, since the maximization in the definition (1.4) of H is performed only on the compact subset $A = [\kappa, M]^d$ of $\mathbb{R}^d$.

Consider the Lagrangian, not depending on x, defined by

$$L(\alpha) := b|\alpha - a|^2, \tag{1.15}$$

with $a := \left(\frac{\kappa+M}{2}\right)(1,\dots,1)^\dagger$ and b a large enough constant to be chosen later. The computation of $H := \sup_{\alpha\in[\kappa,M]}\{-p\cdot\alpha - L(\alpha)\}$ for such choice of L gives

$$H(p) = \frac{p^2}{4b} - a\cdot p, \tag{1.16}$$

for $|p| \leq b(M-\kappa)$, while H is linear outside this interval. It is trivial to verify that H is in $C^1(\mathbb{R}^d)$, and thus **(H1)** is satisfied, while H is not in $C^2(\mathbb{R}^d)$ because of the linear components. Nevertheless, (1.12) is satisfied whenever $|p| \leq K$, with the choice $b := \frac{K}{M-\kappa}$. Moreover, the Lipschitz continuity of D_pH and $D^2_{pp}H$ is trivially holding because of expression (1.16) for $|p| \leq K$ and the linearity outside, and **(RegH)** follows. Note that p represents the gradient of the value functions and thus it belongs to a compact $[-K, K]$, where K is independent of b; c.f. Remark 1 below.

1.2.3 N-player game

In this section we describe the N-player game in a general setting. Namely, we suppose that each individual has complete information on the states of all the other players and we do not require the players to be symmetric. Then, we show the relation between System (HJB) and the concept of Nash equilibria for the game through a classical Verification Theorem. We conclude the section by introducing the mean field assumptions and stating a consequence on the symmetry of the solution to (HJB). We remark that most of the results of this section were found also in [61], but in a slightly different framework. Namely, there the authors assumed a priori that the value functions depend on the empirical measure, assuming hence symmetry. Moreover, they studied the infinitesimal generator of the processes, while here we employ our probabilistic representation.

In the prelimit the dynamics is given by the system of N controlled SDEs

$$X_i(t) = Z_i + \int_0^t \int_\Xi f(X_i(s^-), \xi, \alpha^i(s, \boldsymbol{X}_{s^-}))\mathcal{N}_i(ds, d\xi), \tag{1.17}$$

for $i = 1, \dots, N$, where f is given by (1.2) and $\boldsymbol{X}_t = (X_1(t), \dots, X_N(t))$. Each player is allowed to choose his/her control α^i having complete information on the state of the other players. We consider only controls $\boldsymbol{\alpha}^N := (\alpha^1, \dots, \alpha^N)$ in feedback form, i.e. the controls are deterministic functions of time and space $\alpha^i : [0,T] \times \Sigma^N \longrightarrow A$, $\alpha^i = \alpha^i(t, \boldsymbol{x})$. We say that $\alpha^i \in \mathcal{A}$, for each i, if it is a measurable function of time. We denote by $\mathcal{A}^N$ the set of feedback strategy vectors $\boldsymbol{\alpha}^N = (\alpha^1, \dots, \alpha^N)$, each α^i belonging to $\mathcal{A}$.

We remark that the dynamics (1.17) is always well-posed, for any admissible choice of the control, since the state space is finite and the coefficients are then trivially Lipschitz continuous. Namely, for any $\boldsymbol{\alpha}^N \in \mathcal{A}^N$ there exists a unique strong solution to (1.17), in the sense that $(\boldsymbol{X}_t)_{t\in[0,T]}$ is adapted to the filtration $\mathbb{F}$ generated by the Poisson random measures.

With the definition of f in (1.2) and the intensity measure ν in (1.3), the dynamics of any player remains in Σ for any time and the feedback controls are exactly the transition rates of the jump processes $(X_i(t))_{i=1,\dots,N}$. Indeed, one can prove - see [24] - that, for $x \neq y$ and $\boldsymbol{x}^{N,i} \in \Sigma^{N-1}$,

$$\mathbb{P}\left[X_i(t+h) = y | X_i(t) = x,\ \boldsymbol{X}_t^{N,i} = \boldsymbol{x}^{N,i}\right] = \alpha_y^i(t, x, \boldsymbol{x}^{N,i})h + o(h). \tag{1.18}$$

In more rigorous terms, with the above choices, for any $\boldsymbol{\alpha}^N \in \mathcal{A}^N$ the state evolution of the N players $\boldsymbol{X}_t := (X_i(t))_{i=1}^N$ is a Markov process, whose law is uniquely determined as the solution to the martingale problem for the time-dependent generator

$$\mathcal{L}_t f(\boldsymbol{x}) = \sum_{i=1}^N \sum_{y\in\Sigma} \alpha_y^i(t, \boldsymbol{x}) \left[f([\boldsymbol{x}^i, y]) - f(\boldsymbol{x})\right],$$

where

$$[\boldsymbol{x}^i, y]_j = \begin{cases} x_j & \text{for } j \neq i \\ y & \text{for } j = i. \end{cases}$$

Since α is the vector of the transition rates of the Markov chain, we set $\alpha_x^i(x) = -\sum_{y\neq x} \alpha_y^i(x)$. We remark that the boundedness from below of the controls ($\alpha^i \in [\kappa, M]^d$, $\kappa > 0$) guarantees that $P(X_i(t) = x) > 0$ for every x in Σ and $t > 0$, for any player i.

Next, we define the object of the minimization. Let $\boldsymbol{\alpha}^N = (\alpha^1, \dots, \alpha^N) \in \mathcal{A}^N$ be a strategy vector and $\boldsymbol{X} = (X_1, \dots, X_N)$ the corresponding solution to (1.17). For $i = 1, \dots, N$ and given functions $F^{N,i}, G^{N,i} : \Sigma^N \longrightarrow \mathbb{R}$, we associate to the i-th player the cost functional

$$J_i^N(\boldsymbol{\alpha}^N) := \mathbb{E}\left[\int_0^T \left[L(X_i(t), \alpha^i(t, \boldsymbol{X}_t)) + F^{N,i}(\boldsymbol{X}_t)\right] dt + G^{N,i}(\boldsymbol{X}_T)\right]. \tag{1.19}$$

The optimality condition for the N-player game is given by the usual concept of Nash equilibria. For a strategy vector $\boldsymbol{\alpha}^N = (\alpha^1, \dots, \alpha^N) \in \mathcal{A}^N$ and $\beta \in \mathcal{A}$, denote by $[\boldsymbol{\alpha}^{N,-i}; \beta]$ the perturbed strategy vector given by

$$[\boldsymbol{\alpha}^{N,-i}; \beta]^j := \begin{cases} \alpha^j, & j \neq i \\ \beta, & j = i. \end{cases}$$

Then, we can introduce the following

Definition 1.2. *A strategy vector $\boldsymbol{\alpha}^N$ is said to be a Nash equilibrium for the N-player game if for each $i = 1, \dots, N$*

$$J_i^N(\boldsymbol{\alpha}^N) = \inf_{\beta\in\mathcal{A}} J_i^N([\boldsymbol{\alpha}^{N,-i}; \beta]).$$

Let us now introduce the functional

$$J_i^N(t,\boldsymbol{x},\boldsymbol{\alpha}^N) := \mathbb{E}\left[\int_t^T [L(X_i^{t,\boldsymbol{x}}(s),\alpha^i(s,\boldsymbol{X}_s^{t,\boldsymbol{x}})) + F^{N,i}(\boldsymbol{X}_s^{t,\boldsymbol{x}})]ds + G^{N,i}(\boldsymbol{X}_T^{t,\boldsymbol{x}})\right], \quad (1.20)$$

where

$$X_i^{t,\boldsymbol{x}}(s) = x_i + \int_t^s\int_\Xi f(X_i^{t,\boldsymbol{x}}(r^-),\xi,\alpha^i(r,\boldsymbol{X}_{r^-}^{t,\boldsymbol{x}}))\mathcal{N}_i(dr,d\xi) \qquad s\in[t,T].$$

We work under hypotheses that guarantee the existence of a unique maximizer $\alpha^*(x,p)$ defined in (1.11). With this notation, the Hamilton-Jacobi-Bellman system associated to the above differential game is given by System (HJB) of Section 1.1:

$$\begin{cases} -\frac{\partial v}{\partial t}^{N,i}(t,\boldsymbol{x}) - \sum_{j=1,\ j\neq i}^N \alpha^*(x_j,\Delta^j v^{N,j})\cdot\Delta^j v^{N,i} + H(x_i,\Delta^i v^{N,i}) = F^{N,i}(\boldsymbol{x}), \\ v^{N,i}(T,\boldsymbol{x}) = G^{N,i}(\boldsymbol{x}). \end{cases}$$

This is a system of Nd^N coupled ODE's, whose well-posedness for all $T>0$ can be proved through standard ODEs techniques, because of the Lipschitz continuity of the vector fields involved in the equations.

We are now able to relate System (HJB) to the Nash equilibria for the N-player game through the following

Proposition 1.3 (Verification Theorem). *Let $v^{N,i}$, $i=1,\dots,N$ be a classical solution to System* (HJB)*. Then the feedback strategy vector $\boldsymbol{\alpha}^{N*} = (\alpha^{1,*},\dots,\alpha^{N,*})$ defined by*

$$\alpha^{i,*}(t,\boldsymbol{x}) := \alpha^*(x_i,\Delta^i v^{N,i}(t,\boldsymbol{x})) \qquad i=1,\dots,N, \quad (1.21)$$

is the unique Nash equilibrium for the N-player game and the $v^{N,i}$'s are the value functions *of the game, i.e.*

$$v^{N,i}(t,\boldsymbol{x}) = J_i^N(t,\boldsymbol{x},\boldsymbol{\alpha}^{N*}) = \inf_{\beta\in\mathcal{A}} J_i^N(t,\boldsymbol{x},[\boldsymbol{\alpha}^{N*,-i};\beta]). \quad (1.22)$$

Proof. Let $\beta\in\mathcal{A}$ be any feedback and $\boldsymbol{X}^{t,\boldsymbol{x}}$ the corresponding solution to (1.17), given the strategy vector $[\boldsymbol{\alpha}^{N*-i};\beta]$; denote for simplicity $\boldsymbol{X}=\boldsymbol{X}^{t,\boldsymbol{x}}$. Fixing $i\in\{1,\dots,N\}$, because of the uniqueness of the maximizer in (1.11), we have

$$\begin{aligned} \frac{\partial v}{\partial t}^{N,i} &+ \sum_{j\neq i}\sum_{y=1}^d \alpha_y^*(t,x_j,\Delta^j v^{N,j})[\Delta^j v^{N,i}(t,\boldsymbol{x})]_y \\ &+ \beta(t,\boldsymbol{x})\cdot\Delta^i v^{N,i}(t,\boldsymbol{x}) + L(x_i,\beta(t,\boldsymbol{x})) + F^{N,i}(\boldsymbol{x}) \geq 0, \end{aligned}$$

for any $t,\boldsymbol{x}$. Applying first Itô formula (Theorem II.5.1 in [72], p. 66) and then Lemma 3 in [24] and the above inequality, we obtain

$$\begin{aligned} v^{N,i}(t,\boldsymbol{x}) &= \mathbb{E}\left[v^{N,i}(T,\boldsymbol{X}_T) - \int_t^T \frac{\partial v}{\partial t}^{N,i}(s,\boldsymbol{X}_s)ds\right] \\ &- \sum_{j=1}^N \mathbb{E}\Bigg[\int_t^T\int_\Xi \Big[v^{N,i}(X_1(s),\dots,X_j(s)+f(X_j(s),\xi,[\boldsymbol{\alpha}^{N*,-i};\beta](s,\boldsymbol{X}_s)),\dots,X_N(s)) \\ &\qquad -v^{N,i}(\boldsymbol{X}_s)\Big]\nu(d\xi)ds\Bigg] \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\Bigg[v^{N,i}(T,\boldsymbol{X}_T) - \int_t^T \Bigg(\frac{\partial v}{\partial t}^{N,i}(s,\boldsymbol{X}_s) \\
&+ \sum_{j\neq i} \alpha^{j,*}(s,\boldsymbol{X}_s)\cdot\Delta^j v^{N,i}(s,\boldsymbol{X}_s) + \beta(t,\boldsymbol{X}_s)\cdot\Delta^i v^{N,i}(t,\boldsymbol{X}_s)\Bigg) ds\Bigg] \\
&\leq \mathbb{E}\left[G^{N,i}(T,\boldsymbol{X}_T) + \int_t^T \Big(L(X_i(s),\beta(s,\boldsymbol{X}_s)) + F^{N,i}(\boldsymbol{X}_s)\Big) ds\right] \\
&=: J_i^N(t,\boldsymbol{x},[\boldsymbol{\alpha}^{N*,-i};\beta]).
\end{aligned}
$$

Replacing β by $\alpha^{i,*}$ the inequalities become equalities. □

Remark 1.4. *It is important to observe that the solution $v^{N,i}$ to* (HJB) *is uniformly bounded with respect to N. Namely, there exists a constant $K > 0$ such that*

$$
\sup_{\boldsymbol{x}\in\Sigma^N} |v^{N,i}(t,\boldsymbol{x})| \leq K,
$$

where the constant K is independent of N, i and t. This and (1.5) *immediately imply an analogous bound for $|\Delta^i v^{N,i}(t,\boldsymbol{x})|$: it is for this reason that the only local regularity (assumptions **(H1)** and **(RegH)**) for $H(x,p)$ with respect to p is enough for getting the convergence and the well-posedness results.*

We are interested in studying the limit of System (HJB) as $N \to +\infty$ under symmetric properties for the N-player game. Namely, we assume that the players are all identical and indistinguishable. In practice, this symmetry is expressed through the following mean-field assumptions on the costs:

$$
\begin{aligned}
F^{N,i}(\boldsymbol{x}) &= F(x_i, m_{\boldsymbol{x}}^{N,i}), \\
G^{N,i}(\boldsymbol{x}) &= G(x_i, m_{\boldsymbol{x}}^{N,i}),
\end{aligned}
\tag{M-F}
$$

for some F and $G : \Sigma \times P(\Sigma) \to \mathbb{R}$. An easy but crucial consequence of assumptions (M-F) and the uniqueness of solution to System (HJB) is that the solution $v^{N,i}$ of such system enjoys symmetric properties:

Proposition 1.5. *Under the mean-field assumptions* (M-F)*, there exists $v^N : [0,T] \times \Sigma^N \to \mathbb{R}^d$ such that the solutions $v^{N,i}$ to System* (HJB) *satisfy, for $i = 1,\ldots,N$,*

$$
v^{N,i}(t,\boldsymbol{x}) = v^N(t, x_i, (x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_N)),
\tag{1.23}
$$

for any $(t,x) \in [0,T]\times\Sigma$, and the function

$$
\Sigma^{N-1} \ni (y_1,\ldots,y_{N-1}) \to v^N(t,x,(y_1,\ldots,y_{N-1}))
$$

is invariant under permutations of $(y_1,\ldots,y_{N-1})$.

Proof. Let $\tilde{\boldsymbol{x}}$ be defined from $\boldsymbol{x}$ after exchanging x_k with x_j, for $j \neq k \neq i$. Because of (M-F), we have that $F^{N,i}(\boldsymbol{x}) = F^{N,i}(\tilde{\boldsymbol{x}})$ and $G^{N,i}(\boldsymbol{x}) = G^{N,i}(\tilde{\boldsymbol{x}})$ and thus, by the uniqueness of solution to (HJB) we conclude $v^{N,i}(t,\boldsymbol{x}) = v^{N,i}(t,\tilde{\boldsymbol{x}})$. □

The above proposition motivates the study of a possible convergence of System (HJB) to a limiting system, by analyzing directly the limit of the functions v^N.

1.2.4 Mean field game and master equation

The mean field game describes the limit for $N \to +\infty$ of the N-players dynamics when they play the Nash equilibrium strategy. Here we illustrate it heuristically, assuming the empirical measure of the process corresponding to the Nash equilibrium obeys a Law of Large Numbers, i.e. it converges to a deterministic flow of probability measures $m : [0,T] \to P(\Sigma)$.

The resulting dynamics for $N \to +\infty$ is characterized by a continuum of i.i.d. players in which the *representative agent* (also referred to as *reference player*) evolves according to

$$X(t) = Z + \int_0^t \int_\Xi f(X(s^-), \xi, \alpha(s, X(s^-)))\mathcal{N}(ds, d\xi), \quad t \in [0,T], \tag{1.24}$$

where the law of the initial condition Z is m_0 and $\mathcal{N}$ is a Poisson random measure with intensity measure ν defined in (1.3). The controls are in feedback form, i.e. they belong to the space of measurable functions $\alpha : [0,T] \times \Sigma \longrightarrow A$. The associated cost is

$$J(\alpha, m) := \mathbb{E}\left[\int_0^T [L(X(t), \alpha(t, X(t))) + F(X(t), m(t))]\, dt + G(X(T), m(T))\right]. \tag{1.25}$$

The reference player thus faces the following problem:

(i) the player controls its jump intensities $\alpha_y : [0,T] \times \Sigma \to [0, +\infty)$, $y \in \Sigma$, via feedback controls depending on time and on his/her own state;

(ii) for a given deterministic flow of probability measures $m : [0,T] \to P(\Sigma)$, the player aims at minimizing the cost (1.25);

(iii) denote by $\alpha^{*,m}$ the optimal control for the above problem, and let $(X^{*,m}(t))_{t\in[0,T]}$ be the corresponding optimal process. The above-mentioned Law of Large Number predicts that the flow $(m(t))_{t\in[0,T]}$ should be chosen so that the following consistency relation, known as mean field equilibrium condition, holds:

$$m(t) = \text{Law}(X^{*,m}(t)) \tag{1.26}$$

for every $t \in [0,T]$.

In literature, such limit dynamics is described by the celebrated mean field game system, whose unknowns are two functions (u, m). The equation in u describes the dynamics of the value function of the reference player, which optimizes his/her payoff under the influence of the collective behaviour of the others, while the equation in m describes the evolution of the distribution of the players. In our discrete setting the mean field game system takes the following form of a strongly coupled system of ODEs:

$$\begin{cases} -\frac{d}{dt}u(t,x) + H(x, \Delta^x u(t,x)) = F(x, m(t)), \\ \frac{d}{dt}m_x(t) = \sum_y m_y(t)\alpha_x^*(y, \Delta^y u(t,y)), \\ u(T,x) = G(x, m(T)), \\ m_x(t_0) = m_{x,0}, \end{cases} \tag{MFG}$$

with $\alpha^*(x,p)$ defined in (1.11) and $u, m : [0,T] \times \Sigma \longrightarrow \mathbb{R}$. A solution (u,m) to (MFG) can be seen as a fixed point of the following procedure, which indeed mimics the problem faced by the representative agent described above: starting with a flow m, solve the first equation - the backward Hamilton-Jacobi-Bellman equation for u - which yields a unique optimal feedback control $\alpha^{*,m}$ for the given m; then, impose that the distribution of the player's corresponding dynamics (1.24) is exactly m, giving the second equation - the forward Kolmogorov-Fokker-Planck (KFP). As a consequence, for a solution (u, m) to (MFG), we have

$$J(\alpha, m) \leq J(\beta, m)$$

for any admissible feedback β, where $\alpha(t,x) = \alpha^*(x, \Delta^x u(t,x))$, and the mean field equilibrium condition (1.26) holds.

As already mentioned, recently in [15] a new technique involving the so-called master equation was introduced to rigorously justify the passage from symmetric N-player differential games to mean field games. Generally speaking, the master equation summarizes all the information needed to find solutions to the mean field game: System (MFG) provides the characteristic curves for (M) (see Section 1.5 below). Indeed, $U(t_0, x, m_0) := u(t_0, x)$ solves (M), (u, m) being the solution to the mean field game system (MFG) starting at time t_0 up to time T, with $m(t_0) = m_0$. Moreover, in the Introduction we already motivated heuristically the convergence result of System (HJB) to the master equation (M). As it will be clear from the convergence argument, all that is needed is the existence of a regular solution to (M).

To be specific on the needed regularity, we conclude this section with the definition of regular solution to (M).

Definition 1.6. *A function $U : [0,T] \times \Sigma \times P(\Sigma) \to \mathbb{R}$ is said to be a* classical solution *to* (M) *if it is continuous in all its arguments, C^1 in t and C^1 in m and, for any $(t,x,m) \in [0,T] \times \Sigma \times P(\Sigma)$, we have*

$$\begin{cases} -\frac{\partial U}{\partial t} + H(x, \Delta^x U) - \int_\Sigma D^m U(t,x,m,y) \cdot \alpha^*(y, \Delta^y U(t,y,m)) dm(y) = F(x,m), \\ U(T,x,m) = G(x,m), \quad (x,m) \in \Sigma \times P(\Sigma). \end{cases}$$

In particular,

$$\Delta^x U(t,x,\cdot) : P(\Sigma) \to \mathbb{R}^d$$

is bounded and Lipschitz continuous, and

$$D^m U(t,x,\cdot) : P(\Sigma) \to \mathbb{R}^{d \times d}$$

is bounded.

Moreover, we say that U is a regular solution *to* (M) *if it is a classical solution and $D^m U(t,x,\cdot)$ is also Lipschitz continuous in m, uniformly in (t,x).*

Let us observe that in the master equation we could replace $D^m U(t,x,m,y)$ by $D^m U(t,x,m,1)$, thanks to property (1.8) of the derivative. Under sufficient conditions, we will prove in Section 1.5 the existence and uniqueness of a regular solution to (M).

1.3 The convergence argument

In this section we take for granted the well-posedness of the master equation (M) and focus on the study of the convergence. We give the precise statement of the convergence in terms of two theorems: the first one describes the convergence in average of the value functions, while the second one is a propagation of chaos for the optimal trajectories.

For any $i \in \{1, \dots, N\}$ and $x \in \Sigma$, set

$$w^{N,i}(t_0, x, m_0) := \sum_{x_1=1}^{d} \cdots \sum_{x_{i-1}=1}^{d} \sum_{x_{i+1}=1}^{d} \cdots \sum_{x_N=1}^{d} v^{N,i}(t_0, \boldsymbol{x}) \prod_{j \neq i} m_0(x_j),$$

where $\boldsymbol{x} = (x_1, \dots, x_N)$, and

$$||w^{N,i}(t_0, \cdot, m_0) - U(t_0, \cdot, m_0)||_{L^1(m_0)} := \sum_{x=1}^{d} |w^{N,i}(t_0, x, m_0) - U(t_0, x, m_0)|m_0(x).$$

The main result is given by the following

Theorem 1.7. *Assume **(H1)** and that (M) admits a unique regular solution U in the sense of Definition 1.6. Fix $N \geq 1$, $(t_0, m_0) \in [0, T] \times P(\Sigma)$, $\boldsymbol{x} \in \Sigma^N$ and let $(v^{N,i})_{i=1,\dots,N}$ be the solution to (HJB). Then*

$$\frac{1}{N}\sum_{i=1}^{N} |v^{N,i}(t_0, \boldsymbol{x}) - U(t_0, x_i, m^N_{\boldsymbol{x}})| \leq \frac{C}{N} \tag{1.27}$$

$$||w^{N,i}(t_0, \cdot, m_0) - U(t_0, \cdot, m_0)||_{L^1(m_0)} \leq \frac{C}{\sqrt{N}}. \tag{1.28}$$

In (1.27) *and* (1.28)*, the constant C does not depend on i, t_0, m_0, $\boldsymbol{x}$ nor N.*

As stated above, the convergence can be studied also in terms of the optimal trajectories. Consider the optimal process $\boldsymbol{Y}_t = (Y_1(t), \dots, Y_N(t))_{t\in[0,T]}$ for the N-player game:

$$Y_i(t) = Z_i + \int_0^t \int_\Xi \sum_{y\in\Sigma} (y - Y_i(s^-))\mathbb{1}_{]0,\alpha^i_y(s,\boldsymbol{Y}_{s^-})[}(\xi_y)\mathcal{N}_i(ds, d\xi), \quad t \in [0, T] \tag{1.29}$$

where $\alpha^i_y(t, \boldsymbol{Y}_t)$ is the optimal feedback, i.e. $\alpha^i_y(t, \boldsymbol{y}) := [\alpha^*(y_i, \Delta^i v^{N,i}(t, \boldsymbol{y}))]_y$. Moreover, let $\tilde{\boldsymbol{X}}_t = (\tilde{X}_1(t), \dots, \tilde{X}_N(t))_{t\in[0,T]}$ be the i.i.d. process solution to

$$\tilde{X}_{i,t} = Z_i + \int_0^t \int_\Xi \sum_{y\in\Sigma} (y - \tilde{X}_i(s^-))\mathbb{1}_{]0,\tilde{\alpha}^i_y(s,\tilde{\boldsymbol{X}}_{s^-})[}(\xi_y)\mathcal{N}_i(ds, d\xi), \quad t \in [0, T] \tag{1.30}$$

with $\tilde{\alpha}^i_y(t, \tilde{\boldsymbol{X}}_t) := [\alpha^*\left(\tilde{X}_i(t), \Delta^x U(t, \tilde{X}_i(t), Law(\tilde{X}_i(t)))\right)]_y$. We remark that

$$Law(\tilde{X}_i(t)) = m(t),$$

with m the solution to the mean field game.

Theorem 1.8. *Under the same assumptions of Theorem 1.7, for any $N \geq 1$ and any $i \in \{1, \dots, N\}$, we have*

$$\mathbb{E}\left[\sup_{t\in[0,T]} \left|Y_i(t) - \tilde{X}_i(t)\right|\right] \leq CN^{-\frac{1}{9}} \tag{1.31}$$

for some constant $C > 0$ independent of m_0 and N. In particular, we obtain the Law of Large Numbers

$$\mathbb{E}\left[\sup_{t\in[0,T]} \left|m^N_{\boldsymbol{Y}}(t) - m(t)\right|\right] \leq CN^{-\frac{1}{9}}. \tag{1.32}$$

Note that the supremum is taken inside the mean, giving the convergence in the space of trajectories. For this reason, we have a slow convergence of order $N^{-1/9}$, coming from a result in [90] about the convergence of the empirical measures of a decoupled system (c.f. Lemma 1 below). Instead, if the supremum is taken outside the mean, the convergence would be of order $N^{-1/2}$, thanks to a result in [59].

1.3.1 Approximating the optimal trajectories

The first step in the proof of these results is to show that the projection of U onto empirical measures

$$u^{N,i}(t,\boldsymbol{x}) := U(t,x_i,m_{\boldsymbol{x}}^{N,i}) \tag{1.33}$$

satisfies the system (HJB) up to a term of order $O(\frac{1}{N})$. The following proposition makes rigorous the intuition we already used in the heuristic derivation of the master equation (M). In what follows, C will denote any constant independent of $i, N, m_0, \boldsymbol{x}$ which is allowed to change from line to line.

Proposition 1.9. *Let U be a regular solution to* (M) *and $u^{N,i}(t,\boldsymbol{x})$ be defined as in (1.33). Then, for $j \neq i$,*

$$\Delta^j u^{N,i}(t,\boldsymbol{x}) = \frac{1}{N-1} D^m U(t,x_i,m_{\boldsymbol{x}}^{N,i},x_j) + \tau^{N,i,j}(t,\boldsymbol{x}), \tag{1.34}$$

where $\tau^{N,i,j} \in C^0([0,T]\times\Sigma^N;\mathbb{R}^d)$, $||\tau^{N,i,j}|| \leq \frac{C}{(N-1)^2}$.

Proof. Observe first that $[\Delta^j u^{N,i}(t,\boldsymbol{x})]_{x_j} = 0 = [D^m U(t,x_i,m_{\boldsymbol{x}}^{N,i},x_j)]_{x_j}$ by definition, so we set $[\tau^{N,i,j}(t,\boldsymbol{x})]_{x_j} = 0$. Consider then $h \neq x_j$: $[\Delta^j u^{N,i}(t,\boldsymbol{x})]_h = U(t,x_i,\frac{1}{N-1}\sum_{k\neq i,j}\delta_{x_k} + \frac{1}{N-1}\delta_h) - U(t,x_i,m_{\boldsymbol{x}}^{N,i})$ by definition. By standard computations we get

$$\begin{aligned}
&U\left(t,x_i,\frac{1}{N-1}\sum_{k\neq i,j}\delta_{x_k} + \frac{1}{N-1}\delta_h\right) - U(t,x_i,m_{\boldsymbol{x}}^{N,i})\\
&= U\left(t,x_i,m_{\boldsymbol{x}}^{N,i} + \frac{1}{N-1}(\delta_h-\delta_{x_j})\right) - U(t,x_i,m_{\boldsymbol{x}}^{N,i})\\
&= \int_0^{\frac{1}{N-1}} \Big[D^m U(m_{\boldsymbol{x}}^{N,i} + s(\delta_h-\delta_{x_j}),x_j)\Big]_h ds\\
&= \int_0^{\frac{1}{N-1}} \Bigg(\Big[D^m U(m_{\boldsymbol{x}}^{N,i} + s(\delta_h-\delta_{x_j}),x_j)\Big]_h + \Big[D^m U(m_{\boldsymbol{x}}^{N,i},x_j)\Big]_h\\
&\qquad\qquad - \Big[D^m U(m_{\boldsymbol{x}}^{N,i},x_j)\Big]_h\Bigg) ds\\
&= \frac{1}{N-1}\Big[D^m U(m_{\boldsymbol{x}}^{N,i},x_j)\Big]_h\\
&\qquad + \int_0^{\frac{1}{N-1}} \left(\Big[D^m U(m_{\boldsymbol{x}}^{N,i} + s(\delta_h-\delta_{x_j}),x_j)\Big]_h - \Big[D^m U(m_{\boldsymbol{x}}^{N,i},x_j)\Big]_h\right) ds\\
&= \frac{1}{N-1}\Big[D^m U(t,x_i,m_{\boldsymbol{x}}^{N,i},x_j)\Big]_h + O\left(\frac{1}{(N-1)^2}\right),
\end{aligned}$$

where the last equality is derived by exploiting the Lipschitz continuity in m of $D^m U$

$$\begin{aligned}
&\left|\int_0^{\frac{1}{N-1}} \left(\Big[D^m U(m_{\boldsymbol{x}}^{N,i} + s(\delta_h-\delta_{x_j}),x_j)\Big]_h - \Big[D^m U(m_{\boldsymbol{x}}^{N,i},x_j)\Big]_h\right) ds\right|\\
&\qquad \leq C\int_0^{\frac{1}{N-1}} \Big|s(\delta_h-\delta_{x_j})\Big| ds = O\left(\frac{1}{(N-1)^2}\right).
\end{aligned}$$

For every component h of $D^m U$ we proved the assertion of the proposition, and thus the same holds for the whole vector. □

In the next proposition we show that the $u^{N,i}$'s almost solve the system (HJB):

Proposition 1.10. *Under the assumptions of Theorem 1.7, the functions* $(u^{N,i})_{i=1,\dots,N}$ *solve*

$$\begin{cases} -\frac{\partial u^{N,i}}{\partial t}(t,\boldsymbol{x}) - \sum_{j=1,\ j\neq i}^{N} \alpha^*(x_j, \Delta^j u^{N,j}) \cdot \Delta^j u^{N,i} + H(x_i, \Delta^i u^{N,i}) = F^{N,i}(\boldsymbol{x}) + r^{N,i}(t,\boldsymbol{x}) \\ u^{N,i}(T,\boldsymbol{x}) = G(x_i, m_{\boldsymbol{x}}^{N,i}), \end{cases} \tag{1.35}$$

with $r^{N,i} \in C^0([0,T]\times\Sigma^N)$, $||r^{N,i}|| \le \frac{C}{N}$.

Proof. We know that U solves

$$-\partial_t U + H(x, \Delta^x U) - \int_\Sigma D^m U(t,x,m,y)\cdot \alpha^*(y, \Delta^y U(t,y,m)) dm(y) = F(x,m),$$

and $U(T,x,m) = G(x,m)$. Computing the equation in $(t, x_i, m_{\boldsymbol{x}}^{N,i})$ we get (we omit the $*$ in α^* for simplicity)

$$\begin{aligned} &-\partial_t U(t,x_i,m_{\boldsymbol{x}}^{N,i}) + H(x_i, \Delta^x U(t,x_i,m_{\boldsymbol{x}}^{N,i})) \\ &\qquad - \int_\Sigma D^m U(t,x_i,m_{\boldsymbol{x}}^{N,i},y)\cdot \alpha(y, \Delta^x U(t,y,m_{\boldsymbol{x}}^{N,i})) dm_{\boldsymbol{x}}^{N,i}(y) = F(x_i, m_{\boldsymbol{x}}^{N,i}), \end{aligned}$$

with the correct final condition $u^{N,i}(t,\boldsymbol{x}) = U(T,x_i,m_{\boldsymbol{x}}^{N,i}) = G(x_i,m_{\boldsymbol{x}}^{N,i})$. By definition of empirical measure we can rewrite

$$\begin{aligned} &-\partial_t U(t,x_i,m_{\boldsymbol{x}}^{N,i}) + H(x_i, \Delta^x U(t,x_i,m_{\boldsymbol{x}}^{N,i})) \\ &\quad - \frac{1}{N-1}\sum_{j=1, j\neq i}^{N} D^m U(t,x_i,m_{\boldsymbol{x}}^{N,i},x_j)\cdot \alpha(x_j, \Delta^x U(t,x_j,m_{\boldsymbol{x}}^{N,i})) = F^{N,i}(\boldsymbol{x}). \end{aligned}$$

Thanks to Proposition 1.9, we have

$$\begin{aligned} &\frac{1}{N-1}\sum_{j=1, j\neq i}^{N} D^m U(t,x_i,m_{\boldsymbol{x}}^{N,i},x_j)\cdot \alpha(x_j, \Delta^x U(t,x_j,m_{\boldsymbol{x}}^{N,i})) \\ &\qquad = \sum_{j=1, j\neq i}^{N} \Delta^j u^{N,i}(t,\boldsymbol{x})\cdot \alpha(x_j, \Delta^x U(t,x_j,m_{\boldsymbol{x}}^{N,i})) \\ &\qquad - \sum_{j=1, j\neq i}^{N} \tau^{N,i,j}(t,\boldsymbol{x})\cdot \alpha(x_j, \Delta^x U(t,x_j,m_{\boldsymbol{x}}^{N,i})) \\ &\qquad =:\ 1)\ +\ 2). \end{aligned}$$

For the first term we add and subtract the quantity $\alpha(x_j, \Delta^x U(t,x_j,m_{\boldsymbol{x}}^{N,j}))$:

$$\begin{aligned} 1) &= \sum_{j\neq i} \Delta^j u^{N,i}(t,\boldsymbol{x})\cdot \alpha(x_j, \Delta^x U(t,x_j,m_{\boldsymbol{x}}^{N,i})) - \alpha(x_j, \Delta^x U(t,x_j,m_{\boldsymbol{x}}^{N,j})) \\ &+ \sum_{j\neq i} \Delta^j u^{N,i}(t,\boldsymbol{x})\cdot \alpha(x_j, \Delta^x U(t,x_j,m_{\boldsymbol{x}}^{N,j})) \\ &= (A) + (B). \end{aligned}$$

For (A) we have, using first the Lipschitz continuity of α with respect to the second variable and then the Lipschitz continuity of $\Delta^x U$ with respect to m:

$$(A) \le \sum_{j\neq i} \Delta^j u^{N,i}(t,\boldsymbol{x})\cdot (\Delta^x U(t,x_j,m_{\boldsymbol{x}}^{N,i}) - \Delta^x U(t,x_j,m_{\boldsymbol{x}}^{N,j}))$$

$$\leq C \sum_{j \neq i} ||\Delta^j u^{N,i}|| \cdot |m_{\boldsymbol{x}}^{N,i} - m_{\boldsymbol{x}}^{N,j}|$$
$$\leq \frac{C}{N-1} \sum_{j \neq i} ||\Delta^j u^{N,i}|| \leq \frac{C}{N},$$

where the last inequality is a consequence of (1.34) and the uniform bound on $||D^m U||$ for the solution to (M). Part (B) of 1) is instead what we want to obtain in the equation for $u^{N,i}$, so we leave it as it is.

For the term 2), we simply note that α is bounded from above by definition, and thus the whole term 2) is also of order $O\left(\frac{1}{N}\right)$. □

The central part of the proof of convergence is based on comparing the optimal trajectories associated to $v^{N,i}$ with the ones associated to $u^{N,i}$. Hence, consider the processes

$$X_i(t) = Z_i + \int_0^t \int_\Xi \sum_{y \in \Sigma} (y - X_i(s^-)) \mathbf{1}_{]0, \tilde{\alpha}_y^i(s, \boldsymbol{X}_{s^-})[}(\xi_y) \mathcal{N}_i(ds, d\xi), \quad t \in [0,T] \tag{1.36}$$

where $\tilde{\alpha}_y^i(t, \boldsymbol{X}_t) := [\alpha^*(X_i(t), \Delta^i u^{N,i}(t, \boldsymbol{X}_t))]_y$. Observe that the processes $\boldsymbol{X}$ and $\boldsymbol{Y}$ are exchangeable. For future use, let us also recall the inequalities

$$\left| m_{\boldsymbol{x}}^N - m_{\boldsymbol{y}}^N \right| \leq C \boldsymbol{d}_1(m_{\boldsymbol{x}}^N, m_{\boldsymbol{y}}^N) \leq \frac{C}{N} \sum_{i=1}^N |x_i - y_i| \tag{1.37}$$

for every $\boldsymbol{x}, \boldsymbol{y} \in \Sigma^N$, where the first inequality comes from the equivalence of all the metrics in $P(\Sigma)$ and the second is well-known for the Wasserstein distance $\boldsymbol{d}_1$ (we stated in (19) in general) The result needed to prove the main theorems is the following

Theorem 1.11. *With the notation introduced above, under the assumptions of Theorem 1.7, we have*

$$\mathbb{E}\left[\sup_{t \in [0,T]} |Y_i(t) - X_i(t)|\right] \leq \frac{C}{N}, \tag{1.38}$$

$$\mathbb{E}\left[\sup_{t \in [0,T]} |m_{\boldsymbol{Y}}^N(t) - m_{\boldsymbol{X}}^N(t)|\right] \leq \frac{C}{N}, \tag{1.39}$$

$$\mathbb{E}\left[\sup_{t \in [0,T]} |u^{N,i}(t, \boldsymbol{Y}_t) - v^{N,i}(t, \boldsymbol{Y}_t)|^2 + \int_0^T \left| \Delta^i u^{N,i}(t, \boldsymbol{Y}_t) - \Delta^i v^{N,i}(t, \boldsymbol{Y}_t) \right|^2 dt \right] \leq \frac{C}{N^2}, \tag{1.40}$$

$$\frac{1}{N} \sum_{i=1}^N |v^{N,i}(0, \boldsymbol{Z}) - u^{N,i}(0, \boldsymbol{Z})| \leq \frac{C}{N} \quad \mathbb{P}\text{-}a.s. \tag{1.41}$$

Proof. In order to prove (1.40), we apply Itô's Formula to the function

$$\Psi(t, \boldsymbol{Y}_t) = (u^{N,i}(t, \boldsymbol{Y}_t) - v^{N,i}(t, \boldsymbol{Y}_t))^2,$$

$$d\Psi(t, \boldsymbol{Y}_t) = \frac{\partial \Psi(t, \boldsymbol{Y}_t)}{\partial t} + \sum_{j=1}^N \int_\Xi [\Psi(t, \tilde{\boldsymbol{Y}}_{t^-}^j) - \Psi(t, \boldsymbol{Y}_{t-})] \mathcal{N}_j(dt, d\xi),$$

where

$$\tilde{\boldsymbol{Y}}_t^j = \left(Y_{1,t}, \ldots, Y_{j-1,t}, Y_{j,t} + \sum_{y \in \Sigma} (y - Y_{j,t}) \mathbb{1}_{]0, \alpha_y^j[}(\xi_y), Y_{j+1,t}, \ldots, Y_{N,t} \right),$$

and, as above,

$$\alpha_y^j(t, \boldsymbol{Y}_t) = \left[\alpha^*(Y_{j,t}, \Delta^j v^{N,j}(t, \boldsymbol{Y}_t)) \right]_y.$$

It follows that

$$\begin{aligned} &d\Psi(t, \boldsymbol{Y}_t) \\ &\quad = 2(u^{N,i}(t, \boldsymbol{Y}_t) - v^{N,i}(t, \boldsymbol{Y}_t))(\partial_t u^{N,i} - \partial_t v^{N,i}) \\ &\quad + \sum_{j=1}^N \int_\Xi [(u^{N,i}(t, \tilde{\boldsymbol{Y}}_{t^-}^j) - v^{N,i}(t, \tilde{\boldsymbol{Y}}_{t^-}^j))^2 - (u^{N,i}(t, \boldsymbol{Y}_{t^-}) - v^{N,i}(t, \boldsymbol{Y}_{t^-}))^2] \mathcal{N}_j(dt, d\xi). \end{aligned}$$

Integrating on the time interval $[t, T]$, we get:

$$\begin{aligned} &[u^{N,i}(T, \boldsymbol{Y}_T) - v^{N,i}(T, \boldsymbol{Y}_T)]^2 \\ &\quad = [u^{N,i}(t, \boldsymbol{Y}_t) - v^{N,i}(t, \boldsymbol{Y}_t)]^2 \\ &\quad + 2 \int_t^T (u^{N,i}(s, \boldsymbol{Y}_s) - v^{N,i}(s, \boldsymbol{Y}_s))(\partial_t u^{N,i}(s, \boldsymbol{Y}_s) - \partial_t v^{N,i}(s, \boldsymbol{Y}_s)) ds \\ &\quad + \sum_{j=1}^N \int_t^T \int_\Xi [(u^{N,i}(s, \tilde{\boldsymbol{Y}}_{s^-}^j) - v^{N,i}(s, \tilde{\boldsymbol{Y}}_{s^-}^j))^2 - (u^{N,i}(s, \boldsymbol{Y}_{s^-}) - v^{N,i}(s, \boldsymbol{Y}_{s^-}))^2] \mathcal{N}_j(ds, d\xi). \end{aligned}$$

For brevity, in the remaining part of the proof we set $u_t^i := u^{N,i}(t, \boldsymbol{Y}_t)$ and $v_t^i := v^{N,i}(t, \boldsymbol{Y}_t)$. Next, we take the conditional expectation on the initial data $\boldsymbol{Z}$, denoting

$$\mathbb{E}^{\boldsymbol{Z}} = \mathbb{E}[\,\cdot\,| \boldsymbol{Y}_t = \boldsymbol{Z}].$$

Note that we are allowed to condition on such event since it has positive probability, thanks to the bound from below on the jump rates. Applying again Lemma 3 of [24], we obtain

$$\begin{aligned} \mathbb{E}^{\boldsymbol{Z}}[(u_T^i - v_T^i)^2] = \mathbb{E}^{\boldsymbol{Z}}[(u_t^i - v_t^i)^2] + 2\mathbb{E}^{\boldsymbol{Z}} \left[\int_t^T (u_s^i - v_s^i)(\partial_t u_s^i - \partial_t v_s^i) ds \right] \\ + \sum_{j=1}^N \mathbb{E}^{\boldsymbol{Z}} \left[\int_t^T \alpha^j(s, \boldsymbol{Y}_s) \cdot \Delta^j[(u_s^i - v_s^i)^2] ds \right]. \end{aligned}$$

Let us first study the term $\mathbb{E}^{\boldsymbol{Z}} \left[\int_t^T (u_s^i - v_s^i)(\partial_t u_s^i - \partial_t v_s^i) ds \right]$. Applying Equations (1.35) and (HJB), we get

$$\begin{aligned} &\mathbb{E}^{\boldsymbol{Z}} \left[\int_t^T (u_s^i - v_s^i)(\partial_t u_s^i - \partial_t v_s^i) ds \right] \\ &\quad = \mathbb{E}^{\boldsymbol{Z}} \left[\int_t^T (u_s^i - v_s^i) \left\{ \sum_{j=1, j \neq i}^N \left(-\alpha^j(Y_{j,s}, \Delta^j u_s^j) \cdot \Delta^j u_s^i + \alpha^j(Y_{j,s}, \Delta^j v_s^j) \cdot \Delta^j v_s^i \right. \right. \right. \\ &\quad \left. \left. \left. + \alpha^j \cdot \Delta^j u_s^i - \alpha^j \cdot \Delta^j u_s^i \right) - H(Y_{i,s}, \Delta^i v_s^i) + H(Y_{i,s}, \Delta^i u_s^i) - r^{N,i}(s, \boldsymbol{Y}_s) \right\} ds \right]. \end{aligned}$$

Recall that $\alpha^j(\Delta^j u_s^j) =: \tilde{\alpha}^j$. Note that we also added and subtracted $\alpha^j \cdot \Delta^j u_s^i$ in the last line so that we can use the Lipschitz properties of H, α^* and the bound on $r^{N,i}$ to get the correct estimates. Specifically, we can rewrite

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T (u_s^i - v_s^i)(\partial_t u_s^i - \partial_t v_s^i)ds\right] \\
&= \mathbb{E}^{\boldsymbol{Z}}\Bigg[\int_t^T (u_s^i - v_s^i)\Bigg\{\sum_{j=1,j\neq i}^N \Big((\alpha^j - \tilde{\alpha}^j)\cdot\Delta^j u_s^i - \alpha^j\cdot(\Delta^j u_s^i - \Delta^j v_s^i)\Big) \\
&\quad - H(Y_{i,s},\Delta^i v_s^i) + H(Y_{i,s},\Delta^i u_s^i) - r^{N,i}(s,\boldsymbol{Y}_s)\Bigg\}ds\Bigg].
\end{aligned}$$

Recollecting the above, we find

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{Z}}[(u_T^i - v_T^i)^2] \\
&= \mathbb{E}^{\boldsymbol{Z}}[(u_t^i - v_t^i)^2] + 2\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T (u_s^i - v_s^i)(\partial_t u_s^i - \partial_t v_s^i)ds\right] \\
&\qquad + \sum_{j=1}^N \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^j(s,\boldsymbol{Y}_s)\cdot\Delta^j[(u_s^i - v_s^i)^2]ds\right] \\
&= \mathbb{E}^{\boldsymbol{Z}}[(u_t^i - v_t^i)^2] + 2\mathbb{E}^{\boldsymbol{Z}}\Bigg[\int_t^T (u_s^i - v_s^i)\Bigg\{\sum_{j\neq i}^N \Big((\alpha^j - \tilde{\alpha}^j)\cdot\Delta^j u_s^i - \alpha^j\cdot(\Delta^j u_s^i - \Delta^j v_s^i)\Big) \\
&- H(Y_{i,s},\Delta^i v_s^i) + H(Y_{i,s},\Delta^i u_s^i) - r^{N,i}(s,\boldsymbol{Y}_s)\Bigg\}ds\Bigg] \\
&+ \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^i(s,\boldsymbol{Y}_s)\cdot\Delta^i[(u_s^i - v_s^i)^2]ds\right] + \sum_{j\neq i}^N \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^j(s,\boldsymbol{Y}_s)\cdot\Delta^j[(u_s^i - v_s^i)^2]ds\right] \\
&= \mathbb{E}^{\boldsymbol{Z}}[(u_t^i - v_t^i)^2] + \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^i(s,\boldsymbol{Y}_s)\cdot\Delta^i[(u_s^i - v_s^i)^2]ds\right] \\
&+ 2\mathbb{E}^{\boldsymbol{Z}}\Bigg[\int_t^T (u_s^i - v_s^i)\left\{\sum_{j\neq i}^N \Big((\alpha^j - \tilde{\alpha}^j)\cdot\Delta^j u_s^i - \alpha^j\cdot(\Delta^j u_s^i - \Delta^j v_s^i)\Big)\right\}ds \\
&+ \int_t^T \sum_{j\neq i}^N \frac{1}{2}\alpha^j\cdot\Delta^j[(u_s^i - v_s^i)^2]ds \\
&\qquad + \int_t^T (u_s^i - v_s^i)(-H(Y_{i,s},\Delta^i v_s^i) + H(Y_{i,s},\Delta^i u_s^i) - r^{N,i}(s,\boldsymbol{Y}_s))ds\Bigg].
\end{aligned}$$

On the other hand, observing that $\Delta^j[(u^i - v^i)^2] = \Delta^j(u^i - v^i)\times(\Delta^j(u^i - v^i) + 2(\mathbf{1}(u^i - v^i)))$, $\times$ being the element by element product between vectors and $\mathbf{1} = (1,\dots,1)^\dagger$, the expression

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{Z}}\Bigg[\int_t^T (u_s^i - v_s^i)\left\{\sum_{j=1,j\neq i}^N \Big(-2\alpha^j\cdot(\Delta^j u_s^i - \Delta^j v_s^i)\Big)\right\}ds \\
&\quad + \int_t^T \sum_{j=1,j\neq i}^N \Big(\alpha^j\cdot\Delta^j[(u_s^i - v_s^i)^2]\Big)\,ds\Bigg]
\end{aligned}$$

can be simplified as follows

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{Z}}\Bigg[\int_t^T (u_s^i - v_s^i)\Bigg\{\sum_{j=1,j\neq i}^N \Big(-2\alpha^j\cdot(\Delta^j u_s^i - \Delta^j v_s^i)\Big)\Bigg\}ds \\
&\qquad + \int_t^T \sum_{j=1,j\neq i}^N \Big(\alpha^j\cdot\Delta^j[(u_s^i - v_s^i)^2]\Big)\,ds\Bigg] \\
&= \sum_{j=1,j\neq i}^N \mathbb{E}^{\boldsymbol{Z}}\Bigg[\int_t^T \Big\{-2\alpha^j\cdot(u_s^i - v_s^i)(\Delta^j u_s^i - \Delta^j v_s^i) \\
&\qquad +\alpha^j\cdot(\Delta^j(u_s^i - v_s^i)\times(\Delta^j(u_s^i - v_s^i) + 2(\mathbf{1}(u_s^i - v_s^i)))\Big\}\,ds\Bigg] \\
&= \sum_{j=1,j\neq i}^N \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^j\cdot(\Delta^j(u_s^i - v_s^i))^2 ds\right].
\end{aligned}$$

Thus, we have found

$$\begin{aligned}
0 =&\mathbb{E}^{\boldsymbol{Z}}[(u_T^i - v_T^i)^2] \\
&= \mathbb{E}^{\boldsymbol{Z}}[(u_t^i - v_t^i)^2] + 2\mathbb{E}^{\boldsymbol{Z}}\Bigg[\int_t^T (u_s^i - v_s^i)\Bigg\{\sum_{j=1,j\neq i}^N \Big((\alpha^j - \tilde{\alpha}^j)\cdot\Delta^j u_s^i\Big) \\
&- H(Y_{i,s},\Delta^i v_s^i) + H(Y_{i,s},\Delta^i u_s^i) - r^{N,i}(s,\boldsymbol{Y}_s)\Bigg\}ds\Bigg] \\
&+ \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^i(s,\boldsymbol{Y}_s)\cdot\Delta^i[(u_s^i - v_s^i)^2]ds\right] + \sum_{j=1,j\neq i}^N \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^j\cdot(\Delta^j(u_s^i - v_s^i))^2 ds\right].
\end{aligned}$$

Now, using again the expression for $\Delta^i((u_s^i - v_s^i)^2)$,

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^i(s,\boldsymbol{Y}_s)\cdot\Delta^i[(u_s^i - v_s^i)^2]ds\right] \\
&\quad = \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^i(s,\boldsymbol{Y}_s)\cdot(\Delta^i(u_s^i - v_s^i))^2 ds\right] \\
&\qquad\qquad + \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^i(s,\boldsymbol{Y}_s)\cdot(\Delta^i(u_s^i - v_s^i)\times 2(\mathbf{1}(u_s^i - v_s^i)ds\right],
\end{aligned}$$

so that we can rewrite the previous as

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{Z}}[(u_t^i - u_t^i)^2] + \sum_{j=1}^N \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^j\cdot(\Delta^j(u_s^i - v_s^i))^2 ds\right] \\
&\quad = -2\mathbb{E}^{\boldsymbol{Z}}\Bigg[\int_t^T (u_s^i - v_s^i)\Bigg\{\sum_{j=1,j\neq i}^N ((\alpha^j - \tilde{\alpha}^j)\cdot\Delta^j u_s^i) - H(Y_{i,s},\Delta^i v_s^i) \\
&\quad + H(Y_{i,s},\Delta^i u_s^i) - r^{N,i}(s,\boldsymbol{Y}_s)\Bigg\}ds\Bigg] \\
&\quad - \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^i(s,\boldsymbol{Y}_s)\cdot(\Delta^i(u_s^i - v_s^i)\times 2(\mathbf{1}(u_s^i - v_s^i)))ds\right].
\end{aligned}$$

Recalling that $\alpha^j \geq 0$ (since it is a vector of transition rates), we can estimate

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{Z}}[(u_t^i - v_t^i)^2] + \sum_{j=1}^N \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^j \cdot (\Delta^j(u_s^i - v_s^i))^2 ds\right] \\
&\leq 2\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \left|u_s^i - v_s^i\right| \left\{ \sum_{j \neq i}^N \left|(\alpha^j - \tilde{\alpha}^j) \cdot \Delta^j u_s^i\right| \right.\right. \\
&\qquad \left.\left. + \left|H(Y_{i,s}, \Delta^i v_s^i) - H(Y_{i,s}, \Delta^i u_s^i)\right| + \left|r^{N,i}(s, \boldsymbol{Y}_s)\right| \right\} ds\right] \\
&\qquad + 2\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \left|u_s^i - v_s^i\right| \cdot \left|\alpha^i(s, \boldsymbol{Y}_s) \cdot \Delta^i(u_s^i - v_s^i)\right| ds\right].
\end{aligned}$$

This also implies, erasing the terms with $j \neq i$ in the left hand side,

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{Z}}[(u_t^i - v_t^i)^2] + \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \alpha^i \cdot (\Delta^i(u_t^i - v_t^i))^2 ds\right] \\
&\leq 2\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \left|u_s^i - v_s^i\right| \left\{ \sum_{j \neq i}^N \left|(\alpha^j - \tilde{\alpha}^j) \cdot \Delta^j u_s^i\right| \right.\right. \\
&\qquad \left.\left. + \left|H(Y_{i,s}, \Delta^i v_s^i) - H(Y_{i,s}, \Delta^i u_s^i)\right| + \left|r^{N,i}(s, \boldsymbol{Y}_s)\right| \right\} ds\right] \\
&\qquad + 2\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \left|u_s^i - v_s^i\right| \left|\alpha^i(s, \boldsymbol{Y}_s) \cdot \Delta^i(u_s^i - v_s^i)\right| ds\right].
\end{aligned}$$

For the boundedness of α^i from below and above (recall that the admissible controls α are such that $\alpha \in A = [\kappa, M]^d$), we get

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{Z}}[(u_t^i - v_t^i)^2] + \kappa\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \left|\Delta^i(u_s^i - v_s^i)\right|^2 ds\right] \\
&\leq 2\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \left|u_s^i - v_s^i\right| \left\{ \sum_{j \neq i}^N \left|(\alpha^j - \tilde{\alpha}^j) \cdot \Delta^j u_s^i\right| \right.\right. \\
&\qquad \left.\left. + \left|H(Y_{i,s}, \Delta^i v_s^i) - H(Y_{i,s}, \Delta^i u_s^i)\right| + \left|r^{N,i}(s, \boldsymbol{Y}_s)\right| \right\} ds\right] \\
&\qquad + 2C\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \left|u_s^i - v_s^i\right| \left|\Delta^i(u_s^i - v_s^i)\right| ds\right].
\end{aligned}$$

We now use the Lipschitz continuity of H and α^* (assumption (**H1**)) and the bounds on $||r^{N,i}|| \leq \frac{C}{N}$ and $||\Delta^j u^i|| \leq \frac{1}{N}||D^m U|| \leq \frac{C}{N}$ proved in Propositions 1.9 and 1.10 to obtain

$$\begin{aligned}
&\mathbb{E}^{\boldsymbol{Z}}[(u_t^i - v_t^i)^2] + \kappa\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \left|\Delta^i(u_s^i - v_s^i)\right|^2 ds\right] \\
&\leq 2\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \left|u_s^i - v_s^i\right| \left\{ \frac{C}{N} \sum_{j=1, j \neq i}^N \left|\Delta^j u_s^j - \Delta^j v_s^j\right| + C\left|\Delta^i(v_s^i - u_s^i)\right| + \frac{C}{N} \right\} ds\right] \\
&\qquad + 2C\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \left|u_s^i - v_s^i\right| \left|\Delta^i(u_s^i - v_s^i)\right| ds\right]
\end{aligned}$$

$$\leq \frac{C}{N}\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T |u_s^i - v_s^i| ds\right] + \frac{C}{N}\sum_{j\neq i}\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T |u_s^i - v_s^i|\left|\Delta^j(u_s^j - v_s^j)\right| ds\right]$$
$$+ C\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T |u_s^i - v_s^i|\left|\Delta^i(u_s^i - v_s^i)\right| ds\right].$$

By the convexity inequality $AB \leq \epsilon A^2 + \frac{B^2}{4\epsilon}$ we can further estimate the right hand side to get

$$\mathbb{E}^{\boldsymbol{Z}}[(u_t^i - v_t^i)^2] + \kappa\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \left|\Delta^i(u_s^i - v_s^i)\right|^2 ds\right]$$
$$\leq \frac{C}{N^2} + C\mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T |u_s^i - v_s^i|^2 ds\right] + \kappa\frac{1}{2N}\sum_{j=1}^N \mathbb{E}^{\boldsymbol{Z}}\left[\int_t^T \left|\Delta^j(u_s^j - v_s^j)\right|^2 ds\right].$$

By Gronwall's Lemma, we obtain

$$\sup_{t\in[0,T]} \mathbb{E}^{\boldsymbol{Z}}[(u_t^i - v_t^i)^2] + \kappa\mathbb{E}^{\boldsymbol{Z}}\left[\int_0^T \left|\Delta^i(u_s^i - v_s^i)\right|^2 ds\right]$$
$$\leq \frac{C}{N^2} + \frac{\kappa}{2N}\sum_{j=1}^N \mathbb{E}^{\boldsymbol{Z}}\left[\int_0^T \left|\Delta^j(u_s^j - v_s^j)\right|^2 ds\right]. \tag{1.42}$$

Taking the expectation and using the exchangeability of the processes $(Y_{j,t})_{j=1,\dots N}$ we obtain (1.40).

In order to derive (1.41), we consider (1.42) in $t = 0$ and average over $i = 1, \dots, N$, so that we get

$$\frac{1}{N}\sum_{i=1}^N \mathbb{E}^{\boldsymbol{Z}}|u^{N,i}(0,\boldsymbol{Z}) - v^{N,i}(0,\boldsymbol{Z})|^2 \leq \frac{C}{N^2},$$

which immediately implies (1.41) almost surely.

We now estimate the difference $X_i - Y_i$. Thanks to Equations (1.36) and (1.29) and the Lipschitz continuity in x and α of the dynamics given by f (see Lemma 2 in [24]), we obtain

$$\mathbb{E}\left[\sup_{s\in[0,t]} |X_{i,s} - Y_{i,s}|\right]$$
$$\leq C\mathbb{E}\left[\int_0^t \left|\alpha^*(X_{i,s}, \Delta^i u^{N,i}(\boldsymbol{X}_s)) - \alpha^*(Y_{i,s}, \Delta^i v^{N,i}(\boldsymbol{Y}_s))\right| ds\right]$$
$$+ C\mathbb{E}\left[\int_0^t |X_{i,s} - Y_{i,s}|\, ds\right]$$
$$\leq C\mathbb{E}\left[\int_0^t |X_{i,s} - Y_{i,s}| ds\right] + C\mathbb{E}\left[\int_0^T \left|\Delta^i u^{N,i}(\boldsymbol{Y}_s) - \Delta^i v^{N,i}(\boldsymbol{Y}_s)\right| ds\right]$$
$$+ C\mathbb{E}\left[\int_0^t \left|\Delta^x U\left(s, X_{i,s}, m^{N,i}_{\boldsymbol{X}_s}\right) - \Delta^x U\left(s, Y_{i,s}, m^{N,i}_{\boldsymbol{Y}_s}\right)\right| ds\right]$$
$$\leq \frac{C}{N} + C\mathbb{E}\left[\int_0^t \sup_{r\in[0,s]} |X_{i,r} - Y_{i,r}| ds\right] + C\mathbb{E}\left[\int_0^t \sup_{r\in[0,s]} \left|m^N_{\boldsymbol{X}_r} - m^N_{\boldsymbol{Y}_r}\right| ds\right], \tag{1.43}$$

where we applied (1.40) and the Lipschitz continuity in m of $\Delta^x U$ in the last inequality. Applying inequality (1.37) and the exchangeability of $(\boldsymbol{X}, \boldsymbol{Y})$ to (1.43), yields

$$\mathbb{E}\left[\sup_{s\in[0,t]} |X_{i,s} - Y_{i,s}| ds\right] \leq \frac{C}{N} + C\mathbb{E}\left[\int_0^t \sup_{r\in[0,s]} |X_{i,r} - Y_{i,r}| ds\right],$$

so that by Gronwall's inequality we get (1.38). Finally (1.38), applying again (1.37), gives (1.39). □

1.3.2 Proofs of the main results

We are now in the position to prove the main results.

Proof of Theorem 1.7. For proving (1.27), we just compute (1.41) - which can be derived for any $t_0 \in [0,T]$ when considering processes starting from t_0 - for $\boldsymbol{Z}$ uniformly distributed on Σ: this yields

$$\frac{1}{N}\sum_{i=1}^{N} |U(t_0, x_i, m_{\boldsymbol{x}}^{N,i}) - v^{N,i}(t_0, \boldsymbol{x})| \leq \frac{C}{N}.$$

Then, we can replace $U(t_0, x_i, m_{\boldsymbol{x}}^{N,i})$ with $U(t_0, x_i, m_{\boldsymbol{x}}^{N})$ using the Lipschitz continuity of U with respect to m, the additional error term being of order $1/N$.

For (1.28), we compute

$$\begin{aligned}
&||w^{N,i}(t_0, \cdot, m_0) - U(t_0, \cdot, m_0)||_{L^1(m_0)} = \\
&\quad = \sum_{x_i=1}^{d} |w^{N,i}(t_0, x_i, m_0) - U(t_0, x_i, m_0)| m_0(x_i) \\
&\quad = \sum_{x_i=1}^{d} \left| \sum_{x_1,\dots,x_{i-1},x_{i+1},\dots,x_N=1}^{d} v^{N,i}(t,\boldsymbol{x}) \prod_{j\neq i} m_0(x_j) - U(t, x_i, m_0) \right| m_0(x_i) \\
&\quad = \sum_{x_i=1}^{d} \left| \sum_{x_1,\dots,x_{i-1},x_{i+1},\dots,x_N=1}^{d} \left\{ v^{N,i}(t,\boldsymbol{x}) \prod_{j\neq i} m_0(x_j) - u^{N,i}(t,\boldsymbol{x}) \prod_{j\neq i} m_0(x_j) \right.\right. \\
&\quad \left.\left. + u^{N,i}(t,\boldsymbol{x}) \prod_{j\neq i} m_0(x_j) \right\} - U(t, x_i, m_0) \right| m_0(x_i) \\
&\quad \leq \mathbb{E}[|v^{N,i}(t,\boldsymbol{Z}) - u^{N,i}(t,\boldsymbol{Z})|] + \sum_{x_1,\dots,x_N=1}^{d} |u^{N,i}(t,\boldsymbol{x}) - U(t, x_i, m_0)| \prod_{j=1}^{N} m_0(x_j),
\end{aligned} \tag{1.44}$$

where in the last inequality the initial data $\boldsymbol{Z} = (Z_1, \dots, Z_N)$ are distributed as m_0.

By (1.40), the first term in (1.44) is of order $1/N$. For the second term we further estimate, using again the Lipschitz continuity of U with respect to m,

$$\begin{aligned}
&\sum_{x_1,\dots,x_N=1}^{d} |u^{N,i}(t,\boldsymbol{x}) - U(t, x_i, m_0)| \prod_{j=1}^{N} m_0(x_j) \\
&\qquad = \sum_{x_1,\dots,x_N=1}^{d} |U(t, x_i, m_{\boldsymbol{x}}^{N,i}) - U(t, x_i, m_0)| \prod_{j=1}^{N} m_0(x_j) \\
&\qquad \leq C\mathbb{E}\left[\boldsymbol{d}_1(m_{\boldsymbol{Z}}^{N,i}, m_0)\right] \leq \frac{C}{\sqrt{N}},
\end{aligned}$$

where in the last inequality we used that $\mathbb{E}\left[\boldsymbol{d}_1(m_{\boldsymbol{Z}}^{N}, m_0)\right] \leq \frac{C}{\sqrt{N}}$, thanks to Theorem 1 of [59], where $\boldsymbol{Z} := (Z_1 \dots, Z_N)$, the Z_i's are i.i.d. initial data, m_0-distributed, $\boldsymbol{d}_1$ is the 1-Wasserstein distance and $m_{\boldsymbol{Z}}^{N}$ is the corresponding empirical measure. Overall, we have bounded (1.44) by a term of order $1/\sqrt{N}$, and thus (1.28) is also proved. □

Finally, we get to the proof of the propagation of chaos (Theorem 1.8). Recall that the $Y_{i,t}$'s are the optimal processes, i.e. the solutions to system (1.29), the $X_{i,t}$'s are the processes associated to the functions $u^{N,i}$, i.e. they solve System (1.36), while the $\tilde{X}_{i,t}$'s - to which we would like to prove convergence - are the decoupled limit processes (they solve system (1.30)). First, we need the following lemma, whose proof can be found for example in [90]:

Lemma 1.12. *Let $\tilde{\boldsymbol{X}}_t = (\tilde{X}_{i,t})_{i\in 1,\dots,N}$ be N i.i.d. processes with values in $\mathbb{R}$, with $Law(\tilde{X}_{i,t}) = m(t)$. Then*

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left|m^{N,i}_{\tilde{\boldsymbol{X}}_t} - m_t\right|\right] \le C\mathbb{E}\left[\sup_{t\in[t_0,T]} \boldsymbol{d}_1(m^{N,i}_{\tilde{\boldsymbol{X}}_t}, m_t)\right] \le CN^{-1/9}. \tag{1.45}$$

Proof of Theorem 1.8. The assertion of the theorem is proved if we show that

$$\mathbb{E}\left[\sup_{t\in[0,T]}|X_{i,t} - \tilde{X}_{i,t}|\right] \le CN^{-1/9}. \tag{1.46}$$

Indeed, by the triangle inequality and (1.38) in Theorem 1.11 we can estimate

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in[0,T]}|Y_{i,t} - \tilde{X}_{i,t}|\right] &\le \mathbb{E}\left[\sup_{t\in[0,T]}|Y_{i,t} - X_{i,t}|\right] + \mathbb{E}\left[\sup_{t\in[0,T]}|X_{i,t} - \tilde{X}_{i,t}|\right]\\ &\le C(N^{-1} + N^{-1/9}).\end{aligned}$$

We are then left to prove (1.46). As in the proof of (1.38), we have

$$\begin{aligned}\rho(t) &:= \mathbb{E}\left[\sup_{s\in[0,t]}|X_{i,s} - \tilde{X}_{i,s}|\right]\\
&\le \mathbb{E}\Bigg[\int_0^t \left|\alpha^*(X_{i,s}, \Delta^i u^{N,i}(\boldsymbol{X}_s)) - \alpha^*(\tilde{X}_{i,s}, \Delta^x U(s, \tilde{X}_{i,s}, m(s)))\right| ds\\
&\qquad + \int_0^t \left|X_{i,s} - \tilde{X}_{i,s}\right| ds\Bigg]\\
&\le \mathbb{E}\Bigg[\int_0^t \left|\alpha^*(X_{i,s}, \Delta^x U(r, X_{i,s}, m^{N,i}_{\boldsymbol{X}_s})) - \alpha^*(X_{i,s}, \Delta^x U(s, \tilde{X}_{i,s}, m^{N,i}_{\tilde{\boldsymbol{X}}_s}))\right| ds\\
&\qquad + \int_0^t \left|X_{i,s} - \tilde{X}_{i,s}\right| ds\\
&+ \int_0^t \left|\alpha^*(X_{i,s}, \Delta^x U(s, \tilde{X}_{i,s}, m^{N,i}_{\tilde{\boldsymbol{X}}_s})) - \alpha^*(\tilde{X}_{i,s}, \Delta^x U(s, \tilde{X}_{i,s}, m(s)))\right| ds\Bigg].\end{aligned}$$

By the Lipschitz continuity of the optimal controls, and of $\Delta^x U$, we can write

$$\begin{aligned}\rho(t) &\le C\int_0^t \mathbb{E}\left[|X_{i,s} - \tilde{X}_{i,s}| + \left|m^{N,i}_{\tilde{\boldsymbol{X}}_s} - m^{N,i}_{\boldsymbol{X}_s}\right| + \left|m^{N,i}_{\tilde{\boldsymbol{X}}_s} - m(s)\right|\right] ds\\
&\le C\int_0^t \mathbb{E}\left[|X_{i,s} - \tilde{X}_{i,s}| + \frac{1}{N-1}\sum_{j\ne i}|X_{j,s} - \tilde{X}_{j,s}| + \left|m^{N,i}_{\tilde{\boldsymbol{X}}_s} - m(s)\right|\right] ds.\end{aligned}$$

Using (1.45) of Lemma 1.12 and the exchangeability of the processes, we obtain

$$\rho(t) \le C\int_0^t \left(\mathbb{E}\left[\sup_{r\in[0,s]}|X_{i,r} - \tilde{X}_{i,r}|\right] + \frac{1}{N-1}\sum_{j\ne i}\mathbb{E}\left[\sup_{r\in[0,s]}|X_{j,r} - \tilde{X}_{j,r}|\right]\right) ds$$

$$+ C\mathbb{E}\left[\sup_{r\in[0,T]}\left|m^{N,i}_{\tilde{\boldsymbol{X}}_r} - m_r\right|\right]$$
$$\leq C\int_0^t \rho(s)ds + CN^{-1/9},$$

which, by Gronwall's Lemma, concludes the proof of (1.31). Finally (1.32) follows from (1.31) and (1.45), using also (1.37). □

1.4 Fluctuations and large deviations

The convergence results, Theorem 1.7 and 1.8, allow one to derive a Central Limit Theorem and a Large Deviation Principle for the asymptotic behaviour of the empirical measure process of the N-player game optimal trajectories. First of all, we recall from Proposition 1.5 that, for any i, the value function $v^{N,i}$ of player i in the N-player game is invariant under permutations of $(x_1,\dots,x_{i-1},x_{i+1},\dots,x_N)$. This is equivalent to say that the value functions can be viewed as functions of the empirical measure of the system, i.e. there exists a map $V^N : [0,T]\times\Sigma\times P(\Sigma)$ such that

$$v^{N,i}(t,\boldsymbol{x}) = V^N(t,x_i,m^{N,i}_{\boldsymbol{x}}) \tag{1.47}$$

for any $i = 1,\dots,N$, $t\in[0,T]$ and $\boldsymbol{x}\in\Sigma^N$.

1.4.1 The empirical measure process

We consider the empirical measure process of the optimal evolution $\boldsymbol{Y}$ - defined in (1.29) - of the N-player game. If the system is in $\boldsymbol{x}$ at time t, then the rate at which player i goes from x_i to y is given, via the optimal control, by

$$\alpha^*_y(x_i,\Delta^i V^N(t,x_i,m^{N,i}_{\boldsymbol{x}})) =: \Gamma^N_{x_i,y}(t,m^N_{\boldsymbol{x}}), \tag{1.48}$$

i.e. by a function Γ^N which depends only on the empirical measure $m^N_{\boldsymbol{x}}$ and on the number of players N.

Thus the empirical measure of the system $(m^N_t)_{t\in[0,T]}$, $m^N_t := m^N_{\boldsymbol{Y}}(t) = \frac{1}{N}\sum_{i=1}^N \delta_{Y_{i,t}}$, evolves as a (time-inhomogeneous) Markov process on $[0,T]$, with values in $S_N := P(\Sigma)\cap\frac{1}{N}\mathbb{Z}^d$. The number of players in state x, when the empirical measure is m, is Nm_x. Hence the jump rate of m^N in the direction $\frac{1}{N}(\delta_y-\delta_x)$ at time t is $Nm_x\Gamma^N_{x,y}(t,m)$. Therefore the generator of the time-inhomogeneous Markov process m^N is given, at time t, by

$$\mathcal{L}^N_t g(m) := N\sum_{x,y\in\Sigma} m_x\Gamma^N_{x,y}(t,m)\left[g\left(m+\frac{1}{N}(\delta_y-\delta_x)\right)-g(m)\right], \tag{1.49}$$

for any $g : S_N \longrightarrow \mathbb{R}$. Theorem 1.8 implies that the empirical measures converge in L^1 - on the space of trajectories $\mathcal{D}([0,T];P(\Sigma))$ - to the deterministic flow of measures m which is the unique solution to the mean field game system, whose dynamics is given by the KFP ODE

$$\begin{cases}\frac{d}{dt}m(t) = \Gamma(t,m(t))^\dagger m(t)\\ m(0) = m_0,\end{cases} \tag{1.50}$$

where Γ is the matrix defined by

$$\Gamma_{x,y}(t,m) := \alpha^*_y(x,\Delta^x U(t,x,m)) \tag{1.51}$$

and U is the solution to the master equation. Viewing $m(t)$ as a Markov process - and so we will write m_t in this section -, its infinitesimal generator is given, at time t, by

$$\mathcal{L}_t g(m) := \sum_{x,y\in\Sigma} m_x \Gamma_{x,y}(t,m)[D^m g(m,x)]_y \tag{1.52}$$

for any $g : P(\Sigma) \longrightarrow \mathbb{R}$. Thanks to (1.8), the generator can be equivalently written as

$$\mathcal{L}_t g(m) := \sum_{x,y\in\Sigma} m_x \Gamma_{x,y}(t,m)[D^m g(m,1)]_y = m^\dagger \Gamma(t,m) D^m g(m,1). \tag{1.53}$$

In order to prove the asymptotic results, we will also consider the empirical measure of the process $\boldsymbol{X}$ defined in (1.36), in which each player chooses the same control $\Gamma_{x,y}$ independent of N. We denote by $\eta_t^N := \frac{1}{N}\sum_{i=1}^N \delta_{X_i(t)}$ the empirical measure process of $\boldsymbol{X}$, whose generator is given, for any $g : P(\Sigma) \longrightarrow \mathbb{R}$, by

$$\mathcal{M}_t^N g(m) := N \sum_{x,y\in\Sigma} m_x \Gamma_{x,y}(t,m) \left[g\left(m + \frac{1}{N}(\delta_y - \delta_x) \right) - g(m) \right]. \tag{1.54}$$

1.4.2 Central Limit Theorem

A natural refinement of the Law of Large Numbers (1.32) consists in studying the fluctuations around the limit, that is the asymptotic distribution of $m_t^N - m_t$.

This can be done through a functional Central Limit Theorem: we define the fluctuation flow

$$\rho_t^N := \sqrt{N}(m_t^N - m_t), \quad t \in [0,T], \tag{1.55}$$

and study its asymptotic behaviour as N tends to infinity. We follow a classical weak convergence approach based on uniform convergence of the generator of the fluctuation flow (1.55) to a limiting generator of a diffusion process to be determined; see e.g. [37] for reference. Before stating the theorem we observe that the process (1.55) has values in $P_0(\Sigma)$, which in the following we treat as a subset of $\mathbb{R}^d$.

Theorem 1.13 (Central Limit Theorem)**.** *Let U be a regular solution to the master equation and assume (**H1**). Then the fluctuation flow ρ_t^N in (1.55) converges, as $N \to +\infty$, in the sense of weak convergence of stochastic processes, to a limit Gaussian process ρ_t which is the solution of the linear SDE*

$$\begin{cases} d\rho_t = \Big(\Gamma(t,m_t)^\dagger \rho_t + b(t,m_t,\rho_t)\Big)\,dt + \sigma(t,m_t)dB_t, \\ \rho_0 = \bar{\rho}, \end{cases} \tag{1.56}$$

where $\bar\rho$ is the limit of ρ_0^N in distribution, B is a standard d-dimensional Brownian motion, Γ is the transition rate matrix in (1.51), $b \in \mathbb{R}^d$ is linear in μ and defined, for any $y \in \Sigma$ and $\mu \in P_0(\Sigma)$, by

$$b(t,m,\mu)_y := \sum_{x\in\Sigma} m_x \left[D^m \Gamma_{x,y}(t,m,1)\cdot \mu \right], \tag{1.57}$$

and $\sigma \in \mathbb{R}^{d\times d}$ is given by the relations

$$(\sigma^2)_{x,y}(t,m) = -(m_x\Gamma_{x,y}(t,m) + m_y\Gamma_{y,x}(t,m)), \ \textit{for } x \neq y, \tag{1.58}$$

$$(\sigma^2)_{x,x}(t,m) = \sum_{y\neq x}(m_y\Gamma_{y,x}(t,m) + m_x\Gamma_{x,y}(t,m)). \tag{1.59}$$

In particular the matrix σ^2 is the opposite of the generator of a Markov chain, is symmetric and positive semidefinite with one null eigenvalue, and the same properties hold for σ, meaning that $\rho_t \in P_0(\Sigma)$ for any t.

Proof. The key observation is that we can reduce ourselves to study the asymptotics of the fluctuation flow

$$\mu_t^N := \sqrt{N}(\eta_t^N - m_t), \tag{1.60}$$

which is more standard since η_t^N, whose generator $\mathcal{M}$ is defined in (1.54), is the empirical measure of an uncontrolled system of N mean-field interacting particles. Indeed, by (1.39) we have that $\sqrt{N}(m^N - \eta^N)$ tends to 0 almost surely as N goes to infinity.

Thus, it remains to prove the convergence in law of (1.60) to the solution to (1.56). The convergence of μ_0^N (and ρ_0^N) to the initial condition $\bar{\rho}$ follows from the Central Limit Theorem for the i.i.d. sequence of initial conditions Z_i in systems (22) and (28). Then, we compute the generator of (1.60) for $t \geq 0$. We note that μ_t^N is obtained from η_t^N through a time dependent, linear invertible transformation $\Phi_t : S_N \to P_0(\Sigma) \subset \mathbb{R}^d$, defined by

$$\Phi_t(\vartheta) := \sqrt{N}(\vartheta - m_t),$$

with inverse $\Phi_t^{-1}(\mu) := m_t + \frac{\mu}{\sqrt{N}}$. Thus, the generator $\mathcal{H}_t^N$ of (1.60) can be written as

$$\mathcal{H}_t^N g(\mu) = \mathcal{M}_t^N[g \circ \Phi_t](\Phi_t^{-1}(\mu)) + \frac{\partial}{\partial t}[g \circ \Phi_t](\Phi_t^{-1}(\mu)), \tag{1.61}$$

for any $g : P_0(\Sigma) \to \mathbb{R}$ regular and with compact support (we can extend the definition of g to be a smooth function in the whole space $\mathbb{R}^d$, so that the usual derivatives are well defined). We have

$$\begin{aligned}\frac{\partial}{\partial t}[g \circ \Phi_t](\Phi_t^{-1}(\mu)) &= -\sqrt{N}\nabla_\mu g(\mu) \cdot \frac{d}{dt}m_t = -\sqrt{N}\nabla_\mu g(\mu) \cdot \left(\Gamma(t,m_t)^\dagger m_t\right) \\ &= -\sqrt{N} \sum_{x,y\in\Sigma} \frac{\partial}{\partial \mu_y} g(\mu)\Gamma_{x,y}(t,m_t)(m_t)_x.\end{aligned}$$

where the second equality follows from the KFP equation for m_t. For the remaining part in (1.61), we have

$$\begin{aligned}\mathcal{M}_t^N[g \circ \Phi_t](\Phi_t^{-1}(\mu)) &= N \sum_{x,y\in\Sigma} \left(m_t + \frac{\mu}{\sqrt{N}}\right)_x \Gamma_{x,y}\left(t, m_t + \frac{\mu}{\sqrt{N}}\right) \times \\ &\quad \times \left\{[g\circ\Phi_t]\left(m_t + \frac{\mu}{\sqrt{N}} + \frac{1}{N}(\delta_y - \delta_x)\right) - [g \circ \Phi_t]\left(m_t + \frac{\mu}{\sqrt{N}}\right)\right\} \\ &= N \sum_{x,y\in\Sigma} \left(m_t + \frac{\mu}{\sqrt{N}}\right)_x \Gamma_{x,y}\left(t, m_t + \frac{\mu}{\sqrt{N}}\right) \left\{g\left(\mu + \frac{1}{\sqrt{N}}(\delta_y - \delta_x)\right) - g(\mu)\right\}.\end{aligned}$$

Thus, we have found

$$\begin{aligned}\mathcal{H}_t^N g(\mu) = N \sum_{x,y\in\Sigma} &\left(m_t + \frac{\mu}{\sqrt{N}}\right)_x \Gamma_{x,y}\left(t, m_t + \frac{\mu}{\sqrt{N}}\right) \left\{g\left(\mu + \frac{1}{\sqrt{N}}(\delta_y - \delta_x)\right) - g(\mu)\right\} \\ &- \sqrt{N} \sum_{x,y\in\Sigma} \frac{\partial}{\partial \mu_y} g(\mu)\Gamma_{x,y}(t,m_t)(m_t)_x.\end{aligned}$$

In order to perform a Taylor expansion of the generator, we first develop the term

$$\begin{aligned}&g\left(\mu + \frac{1}{\sqrt{N}}(\delta_y - \delta_x)\right) - g(\mu) \\ &\quad = \frac{1}{\sqrt{N}}\nabla_\mu g(\mu) \cdot (\delta_y - \delta_x) + \frac{1}{2N}(\delta_y - \delta_x)^\dagger D^2_{\mu\mu} g(\mu)(\delta_y - \delta_x)\end{aligned}$$

$$+ O\left(\frac{1}{N^{3/2}}\right).$$

Substituting, we get

$$\begin{aligned}\mathcal{H}_t^N g(\mu) &= \sqrt{N} \sum_{x,y\in\Sigma} \left(m_t + \frac{\mu}{\sqrt{N}}\right)_x \Gamma_{x,y}\left(t, m_t + \frac{\mu}{\sqrt{N}}\right) \nabla_\mu g(\mu)\cdot(\delta_y - \delta_x) \\ &+ \frac{1}{2}\sum_{x,y\in\Sigma} \left(m_t + \frac{\mu}{\sqrt{N}}\right)_x \Gamma_{x,y}\left(t, m_t + \frac{\mu}{\sqrt{N}}\right) (\delta_y - \delta_x)^\dagger D^2_{\mu\mu} g(\mu)(\delta_y - \delta_x) \\ &- \sqrt{N}\sum_{x,y\in\Sigma} \frac{\partial}{\partial \mu_y} g(\mu)\Gamma_{x,y}(t, m_t)(m_t)_x + O\left(\frac{1}{\sqrt{N}}\right).\end{aligned}$$

Now, we note that

$$\begin{aligned}&\sum_{x,y\in\Sigma} \left(m_t + \frac{\mu}{\sqrt{N}}\right)_x \Gamma_{x,y}\left(t, m_t + \frac{\mu}{\sqrt{N}}\right) \nabla_\mu g(\mu)\cdot(\delta_y - \delta_x) \\ &\qquad = \sum_{x,y\in\Sigma} \left(m_t + \frac{\mu}{\sqrt{N}}\right)_x \Gamma_{x,y}\left(t, m_t + \frac{\mu}{\sqrt{N}}\right) \frac{\partial}{\partial \mu_y} g(\mu),\end{aligned}$$

since $\sum_y \Gamma_{x,y} = 0$. This property allows us to rewrite

$$\begin{aligned}\mathcal{H}_t^N g(\mu) &= \sum_{x,y\in\Sigma} \mu_x \Gamma_{x,y}\left(t, m_t + \frac{\mu}{\sqrt{N}}\right) \frac{\partial}{\partial \mu_y} g(\mu) \\ &+ \sqrt{N}\sum_{x,y\in\Sigma} (m_t)_x \frac{\partial}{\partial \mu_y} g(\mu)\left[\Gamma_{x,y}\left(t, m_t + \frac{\mu}{\sqrt{N}}\right) - \Gamma_{x,y}(t, m_t)\right] \\ &\quad + \frac{1}{2}\sum_{x,y\in\Sigma} \left(m_t + \frac{\mu}{\sqrt{N}}\right)_x \Gamma_{x,y}\left(t, m_t + \frac{\mu}{\sqrt{N}}\right) (\delta_y - \delta_x)^\dagger D^2_{\mu\mu} g(\mu)(\delta_y - \delta_x) \\ &\qquad + O\left(\frac{1}{\sqrt{N}}\right).\end{aligned}$$

Then, using the Lipschitz continuity of Γ as we did in Proposition 3, we linearize the term

$$\Gamma_{x,y}\left(t, m_t + \frac{\mu}{\sqrt{N}}\right) - \Gamma_{x,y}(t, m_t) = \frac{1}{\sqrt{N}} D^m \Gamma_{x,y}(t, m_t, 1)\cdot \mu + O\left(\frac{1}{N}\right).$$

We thus deduce that

$$\lim_{N\to+\infty} \sup_{t\in[0,T]} \sup_{\mu\in P_0(\Sigma)} |\mathcal{H}_t^N g(\mu) - \mathcal{H}_t g(\mu)| = 0$$

for any g, the convergence being of order $\frac{1}{\sqrt{N}}$, where

$$\begin{aligned}\mathcal{H}_t g(\mu) :&= \sum_{x,y\in\Sigma} \mu_x \Gamma_{x,y}(t, m_t) \frac{\partial}{\partial \mu_y} g(\mu) + \sum_{x,y\in\Sigma} (m_t)_x \left[D^m \Gamma_{x,y}(t, m_t, 1)\cdot \mu\right] \frac{\partial}{\partial \mu_y} g(\mu) \\ &+ \frac{1}{2}\sum_{x,y\in\Sigma} (m_t)_x \Gamma_{x,y}(t, m_t)(\delta_y - \delta_x)^\dagger D^2_{\mu\mu} g(\mu)(\delta_y - \delta_x).\end{aligned} \tag{1.62}$$

The proof is then completed if we show that the generator (1.62) is associated to the SDE (1.56).

The drift component can be immediately identified, since

$$\sum_{x,y\in\Sigma} \mu_x \Gamma_{x,y}\left(t,m_t\right) \frac{\partial}{\partial \mu_y} g(\mu) = \left(\Gamma(t,m_t)^\dagger \mu\right) \cdot \nabla_\mu g(\mu),$$

and

$$\sum_{x,y\in\Sigma} (m_t)_x \left[D^m \Gamma_{x,y}(t,m_t,1)\cdot\mu\right] \frac{\partial}{\partial \mu_y} g(\mu) = b(t,\mu)\cdot\nabla_\mu g(\mu).$$

For the diffusion component, we first note that, for each $x,y\in\Sigma$,

$$(\delta_y-\delta_x)^\dagger D^2_{\mu\mu} g(\mu)(\delta_y-\delta_x) = \frac{\partial^2}{\partial\mu_y\mu_y} g(\mu) + \frac{\partial^2}{\partial\mu_x\mu_x} g(\mu) - \frac{\partial^2}{\partial\mu_x\mu_y} g(\mu) - \frac{\partial^2}{\partial\mu_y\mu_x} g(\mu),$$

so that

$$\begin{aligned}&\frac{1}{2}\sum_{x,y\in\Sigma} (\delta_y-\delta_x)^\dagger D^2_{\mu\mu} g(\mu)(\delta_y-\delta_x)(m_t)_x\Gamma_{x,y}(t,m_t)\\ &= \frac{1}{2}\sum_{x,y\in\Sigma}\left[\frac{\partial^2}{\partial\mu_y\mu_y} g(\mu) + \frac{\partial^2}{\partial\mu_x\mu_x} g(\mu) - \frac{\partial^2}{\partial\mu_x\mu_y} g(\mu) - \frac{\partial^2}{\partial\mu_y\mu_x} g(\mu)\right](m_t)_x\Gamma_{x,y}(t,m_t),\end{aligned}$$

which is equal to

$$\frac{1}{2}\mathrm{Tr}(\sigma^2(t,m_t)D^2_{\mu\mu}g(\mu)) = \frac{1}{2}\sum_{x,y\in\Sigma}(\sigma^2(t,m_t))_{x,y}\frac{\partial^2}{\partial\mu_x\mu_y} g(\mu),$$

if we define $(\sigma^2)_{x,y\in\Sigma}$ by the relations (1.58) and (1.59).

Finally, we observe that the limit process ρ_t defined in (1.56) takes values in $P_0(\Sigma)$, as required. Indeed, by diagonalizing σ^2 - which is symmetric and such that its rows sum to 0 - we get that all the eigenvectors, besides the constant one relative to the null eigenvalue, have components which sum to 0 (by orthogonality). The same properties hold for the square root matrix σ, so that Equation (1.56) preserves the space $P_0(\Sigma)$. □

1.4.3 Large Deviation Principle

We state the large deviation result, which is a sample path Large Deviation Principle on $\mathcal{D}([0,T];P(\Sigma))$. To define the rate function, we first introduce the local rate function $\lambda:\mathbb{R}\longrightarrow[0,+\infty]$,

$$\lambda(r) := \begin{cases} r\log r - r + 1 & r>0,\\ 1 & r=0,\\ +\infty & r<0.\end{cases} \tag{1.63}$$

For $t\in[0,T]$, $m\in P(\Sigma)$ and $\mu\in P_0(\Sigma)$, define

$$\begin{aligned}\Lambda(t,m,\mu) := \inf\Bigg\{ &\sum_{x,y\in\Sigma} m_x\Gamma_{x,y}(t,m)\lambda\left(\frac{q_{x,y}}{\Gamma_{x,y}(t,m)}\right) : q_{x,y}\geq 0,\\ &\sum_{x,y\in\Sigma} q_{x,y}(\delta_y-\delta_x) = \mu \quad \forall x,y\Bigg\},\end{aligned} \tag{1.64}$$

and set, for $\gamma:[0,T]\longrightarrow P(\Sigma)$,

$$I(\gamma) := \begin{cases}\int_0^T \Lambda(t,\gamma(t),\dot\gamma(t))dt & \text{if } \gamma \text{ is absolutely continuous and } \gamma(0)=m_0\\ +\infty & \text{otherwise.}\end{cases} \tag{1.65}$$

We are now able to state the Large Deviation Principle. We equip $\mathcal{D}([0,T];P(\Sigma))$ with the Skorokhod J_1-topology and denote by $\mathcal{B}(\mathcal{D}([0,T];P(\Sigma)))$ the associated Borel σ-algebra.

Theorem 1.14 (Large Deviation Principle)**.** *Let U be a regular solution to the master equation and assume (**H1**). Also, assume that the initial conditions $(m_0^N)_{N\in\mathbb{N}}$ are deterministic and $\lim_N m_0^N = m_0$. Then the sequence of empirical measure processes $(m^N)_{N\in\mathbb{N}}$ satisfies the sample path Large Deviation Principle on $\mathcal{D}([0,T];P(\Sigma))$ with the (good) rate function I. Specifically,*

(i) If $E \in \mathcal{B}(\mathcal{D}([0,T];P(\Sigma)))$ is closed then

$$\limsup_N \frac{1}{N}\log \mathbb{P}(m^N \in E) \le -\inf_{\gamma\in E}\{I(\gamma)\}. \tag{1.66}$$

(ii) If $E \in \mathcal{B}(\mathcal{D}([0,T];P(\Sigma)))$ is open then

$$\liminf_N \frac{1}{N}\log \mathbb{P}(m^N \in E) \ge -\inf_{\gamma\in E}\{I(\gamma)\}. \tag{1.67}$$

(iii) For any $M < +\infty$ the set

$$\{\gamma \in \mathcal{D}([0,T];P(\Sigma)) : I(\gamma) \le M\} \tag{1.68}$$

is compact.

We remark that the initial conditions are assumed to be deterministic only for simplicity, otherwise there would be another term in the rate function I. Before proving Theorem 1.14, let us give another characterization of I. For $m \in P(\Sigma)$ and $\theta \in \mathbb{R}^d$, define

$$\Psi(t,m,\theta) := \sum_{x,y} m_x \Gamma_{x,y}(t,m)\left[e^{\theta\cdot(\delta_y-\delta_x)} - 1\right] \tag{1.69}$$

and let Λ^0 be the Legendre transform of Ψ:

$$\Lambda^0(t,m,\mu) = \sup_{\theta\in\mathbb{R}^d}\left[\theta\cdot\mu - \Psi(t,m,\theta)\right]. \tag{1.70}$$

Define I^0 as in (1.65) but with Λ replaced by Λ^0. Via a standard result in convex analysis, Proposition 6.2 in [54] shows that $\Lambda = \Lambda^0$ and then $I = I^0$.

Several authors studied large deviation properties of mean field interacting processes similar to ours. However, most of them deal with the case in which the prelimit jump rates, $m_x^N\Gamma^N$, are constant and equal to the limit rates $m_x\Gamma$; see e.g. [80], [94] and [93]. We mention that in this latter paper, as in many others, it is also assumed that the jump rates of the prelimit process are bounded from below and away from 0; this does not apply to our case, since the number of agents in a state x could be 0, implying that $m_x^N\Gamma_{x,y}^N$ might also be 0.

To prove the claim, we apply the results in [54]: to our knowledge, it is the first paper which proves a Large Deviation Principle considering the jump rates of any player depending on N (and deals also with systems with simultaneous jumps). Theorem 3.4.1 in [101] shows, however, the exponential equivalence of the processes m^N and the processes η^N given by (1.54) in which the jump rates of the prelimit system $m_x^N\Gamma^N$ are replaced by $m_x\Gamma$, which does not depend on N; the proof uses a coupling of the two Markov chains. The results in [54] and [101] are derived assuming the following properties:

1. the dynamics of any agent is ergodic and the jump rates are uniformly bounded;
2. for each $x, y \in \Sigma$, the limit jump rates $\Gamma_{x,y}$ are Lipschitz continuous in m;
3. for each $x, y \in \Sigma$, given any sequence $m^N \in S_N$ such that $\lim_N m^N = m$,

$$\lim_N \sup_{0\le t\le T} |m_x^N \Gamma_{x,y}^N(t, m^N) - m_x \Gamma_{x,y}(t,m)| = 0. \tag{1.71}$$

Property (1) holds in our model since the jump rates of any player belong to $[\kappa, M]$, while (2) is true because of the regularity of the solution U to the master equation.

Proof of Theorem 1.14. The fact that I is a good rate function, i.e condition (iii), is proved for instance in Theorem 1.1 of [53]. Due to Theorem 3.9 in [54], in order to prove the claims (i) and (ii), it is enough to show (1.71). Actually [54] studies time homogeneous Markov processes, but their results still apply in the non-homogeneous case if one proves the uniform in time convergence given by (1.71).

Let $x, y \in \Sigma$, $m^N = m_{\boldsymbol{x}}^N \in S_N$, $\boldsymbol{x} = (x_1, \ldots, x_N) \in \Sigma^N$ and $m_{\boldsymbol{x}}^N \to m$. Then

$$\begin{aligned}\left|[m_{\boldsymbol{x}}^N]_x \Gamma_{x,y}^N(t, m_{\boldsymbol{x}}^N) - m_x\Gamma_{x,y}(t,m)\right| &\le \left|[m_{\boldsymbol{x}}^N]_x \Gamma_{x,y}^N(t, m_{\boldsymbol{x}}^N) - [m_{\boldsymbol{x}}^N]_x \Gamma_{x,y}(t, m_{\boldsymbol{x}}^N)\right| \\ &\quad + \left|[m_{\boldsymbol{x}}^N]_x \Gamma_{x,y}(t, m_{\boldsymbol{x}}^N) - m_x \Gamma_{x,y}(t,m)\right| =: A + B.\end{aligned}$$

The first term goes to zero, uniformly over time, thanks to (1.27):

$$\begin{aligned} A &= \left| \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{x_i = x\}} \alpha_y^*(x_i, \Delta^i V^N(t, x_i, m_{\boldsymbol{x}}^{N,i})) - \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{x_i = x\}} \alpha_y^*(x_i, \Delta^x U(t, x_i, m_{\boldsymbol{x}}^N))\right| \\ &\le C\frac{1}{N}\sum_{i=1}^N \left|\Delta^i V^N(t, x_i, m_{\boldsymbol{x}}^{N,i}) - \Delta^x U(t, x_i, m_{\boldsymbol{x}}^N)\right| \\ &\le C \sup_{\boldsymbol{x}\in\Sigma^N} \frac{1}{N}\sum_{i=1}^N \left| v^{N,i}(t,\boldsymbol{x}) - U(t, x_i, m_{\boldsymbol{x}}^N)\right| \le \frac{C}{N}. \end{aligned}$$

While B converges to 0, uniformly over t, for the regularity of U:

$$\begin{aligned} B &= \left| [m_{\boldsymbol{x}}^N]_x \alpha_y^*(x, \Delta^x U(t,x,m_{\boldsymbol{x}}^N)) - m_x \alpha_y^*(x, \Delta^x U(t,x,m))\right| \\ &\le |\alpha_y^*(x, \Delta^x U(t,x,m_{\boldsymbol{x}}^N))| |[m_{\boldsymbol{x}}^N]_x - m_x| + C|m_x| |\Delta^x U(t,x,m_{\boldsymbol{x}}^N) - \Delta^x U(t,x,m)| \\ &\le C|m_{\boldsymbol{x}}^N - m|, \end{aligned}$$

which tends to 0 by assumption. □

1.5 The master equation: well-posedness and regularity

In this section we study the well-posedness of Equation (M) under the assumptions of monotonicity and regularity for F, G, H we already introduced **(Mon)**, **(RegFG)**, **(RegH)**. A preliminary remark is that, thanks to Proposition 1 in [61], if H is differentiable (and this is indeed the case of our assumptions) then

$$\alpha_x^*(y,p) = -\frac{\partial}{\partial p_x} H(y,p). \tag{1.72}$$

For this reason, we will in the following use α^* interchangeably with $-D_pH$.

Theorem 1.15. *Assume **(Mon)**, **(RegFG)** and **(RegH)**. Then there exists a unique classical solution to* (M) *in the sense of Definition 1.6. Moreover, it is regular.*

The proof exploits the renowned method of characteristics, which consists in proving that

$$U(t_0, x, m_0) := u(t_0, x) \tag{1.73}$$

solves (M), u being the solution of the mean field game system (MFG) with initial time t_0 and initial distribution m_0. In order to perform the computations, we have to prove the regularity in m of the function $U(t_0, x, m)$ defined above. In particular, we have to show that $D^m U$ exists and is bounded. For this, we follow the strategy shown in [15] - which is developed in infinite dimension - adapting it to our discrete setting. The idea consists in studying the well-posedness and regularity properties of the linearized version of System (MFG), whose solution will end up coinciding with $D^m U \cdot \mu_0$, for all possible directions $\mu_0 \in P_0(\Sigma)$. In the remaining part of this section, C will denote any constant which does not depend on t_0, m_0, and is allowed to change from line to line.

1.5.1 Estimates on the mean field game system

We start by proving the well-posedness of System (MFG)

$$\begin{cases} -\frac{d}{dt}u(t,x) + H(x, \Delta^x u(t,x)) = F(x, m(t)), \\ \frac{d}{dt}m_x(t) = \sum_y m_y(t)\alpha^*_x(y, \Delta^y u(t,y)), \\ u(T,x) = G(x, m(T)), \\ m_x(t_0) = m_{x,0}, \end{cases}$$

and a useful a priori estimate on its solution (u, m). The existence of solutions follows from a standard fixed point argument: see Proposition 4 of [61]. Let us remark that any flow of measures m lies in the space

$$\left\{ m \in C^0\left([t_0, T], P(\Sigma)\right) : |m(t) - m(s)| \le 2\nu(\Xi)\sqrt{d}|t-s| \right\},$$

which is a compact and convex subset of the space of continuous functions, endowed with the uniform norm (Lemma 4 of [24]). On the other hand the uniqueness of solution, under our assumptions, is a consequence of the following a priori estimates. Before stating the proposition, we recall the notation $||u|| := \sup_{t\in[t_0,T]} \max_{x\in\Sigma} |u(t,x)|$.

Proposition 1.16. *Assume **(Mon)**, **(RegFG)** and **(RegH)**. Let (u_1, m_1) and (u_2, m_2) be two solutions to* (MFG) *with initial conditions $m_1(t_0) = m_0^1$ and $m_2(t_0) = m_0^2$. Then*

$$||u_1 - u_2|| \le C|m_0^1 - m_0^2|, \tag{1.74}$$

$$||m_1 - m_2|| \le C|m_0^1 - m_0^2|. \tag{1.75}$$

Proof. Without loss of generality, let us set $t_0 = 0$. Let $u := u_1 - u_2$ and $m := m_1 - m_2$. The proof is carried out in three steps.

Step 1. Use of Monotonicity. The couple (u, m) solves

$$\begin{cases} -\frac{d}{dt}u(t,x) + H(x, \Delta^x u_1(t,x)) - H(x, \Delta^x u_2(t,x)) = F(x, m_1(t)) - F(x, m_2(t)) \\ \frac{d}{dt}m(t,x) = \sum_y \left[m_1(t,y)\alpha^*_x(y, \Delta^y u_1(t,y)) - m_2(t,y)\alpha^*_x(y, \Delta^y u_2(t,y))\right] \\ u(T,x) = G(x, m_1(T)) - G(x, m_2(T)) \\ m(0,x) = m_0^1 - m_0^2. \end{cases} \tag{1.76}$$

Since $\frac{d}{dt}\sum_x m(x)u(x) = \sum_x m(x)\frac{du}{dt}(x) + \sum_x \frac{dm}{dt}(x)u(x)$, integrating over $[0,T]$ we have

$$\begin{aligned}
&\sum_x [m(T,x)u(T,x) - m(0,x)u(0,x)] \\
&= \int_0^T \sum_x [H(x,\Delta^x u_1) - H(x,\Delta^x u_2) - F(x,m_1) + F(x,m_2)] (m_1(x) - m_2(x))dt \\
&+ \int_0^T \sum_x \sum_y [m_1(y)\alpha^*_x(y,\Delta^y u_1) - m_2(y)\alpha^*_x(y,\Delta^y u_2)] (u_1(x) - u_2(x))dt.
\end{aligned}$$

Using the fact that $\sum_x \alpha^*_x(y) = 0$ and the initial-final data, we can rewrite

$$\begin{aligned}
&\sum_x [G(x,m_1) - G(x,m_2)](m_1(x) - m_2(x)) \\
&+ \int_0^T \sum_x [F(x,m_1) - F(x,m_2)] (m_1(x) - m_2(x))dt \\
&= \sum_x (m_0^1(x) - m_0^2(x))(u_1(0,x) - u_2(0,x)) \\
&\quad + \int_0^T \sum_x \{[H(x,\Delta^x u_1) - H(x,\Delta^x u_2)](m_1(x) - m_2(x)) \\
&\quad +\Delta^x u \cdot [m_1(x)\alpha^*(x,\Delta^x u_1) - m_2(x)\alpha^*(x,\Delta^x u_2)]\} \, dt.
\end{aligned}$$

We now apply the monotonicity of F and G in the first line and the uniform convexity of H in the last two lines. In fact, recalling that $\alpha^*_y(x,p) = -\frac{\partial}{\partial p_y}H(x,p)$, by **(RegH)** we have that, for each x,

$$\begin{aligned}
H(x,\Delta^x u_1) - H(x,\Delta^x u_2) - \Delta^x u \cdot \frac{\partial}{\partial p} H(x,\Delta^x u_1) \leq -C^{-1}|\Delta^x u|^2 \\
H(x,\Delta^x u_2) - H(x,\Delta^x u_1) + \Delta^x u \cdot \frac{\partial}{\partial p} H(x,\Delta^x u_2) \leq -C^{-1}|\Delta^x u|^2.
\end{aligned}$$

Hence we obtain

$$\int_0^T \sum_x |\Delta^x u(x)|^2 (m_1(x) + m_2(x))dt \leq C(m_0^1 - m_0^2) \cdot (u_1(0) - u_2(0)). \tag{1.77}$$

Step 2. Estimate on Kolmogorov-Fokker-Planck equation. Integrating the second equation in (1.76) over $[0,t]$, we get

$$m(t,x) = m(0,x) + \int_0^t \sum_y [m_1(s,y)\alpha^*_x(y,\Delta^y u_1(s,y)) - m_2(s,y)\alpha^*_x(y,\Delta^y u_2(s,y))] \, ds.$$

The boundedness and Lipschitz continuity of the rates give

$$\max_x |m(t,x)| \leq C|m_0^1 - m_0^2| + C\int_0^t \max_x |m(s,x)|ds + C\int_0^t \sum_x |\Delta^x u(s,x)| m_1(s,x)ds$$

and hence, by Gronwall's Lemma,

$$||m|| \leq C|m_0^1 - m_0^2| + C\int_0^T \sqrt{\sum_x |\Delta^x u(t,x)|^2 m_1(x)} dt. \tag{1.78}$$

This, together with inequality (1.77), yields

$$||m|| \leq C(|m_0^1 - m_0^2| + |m_0^1 - m_0^2|^{1/2}||u||^{1/2}). \tag{1.79}$$

Step 3. Estimate on Hamilton-Jacobi-Bellman equation. Integrating the first equation in (1.76) over $[t,T]$, we get

$$u(t,x) = G(x,m_1(T)) - G(x,m_2(T)) + \int_t^T \left[F(x,m_1) - F(x,m_2) + H(x,\Delta^x u_2) - H(x,\Delta^x u_1)\right] ds.$$

Using the Lipschitz continuity of F, G, H and the bound

$$\max_x |\Delta^x u(x)| \le C \max_x |u(x)|,$$

we obtain

$$\max_x |u(t,x)| \le C|m_1(T) - m_2(T)| + C\int_t^T |m_1(s) - m_2(s)| ds + C\int_t^T \max_x |u(s,x)| ds.$$

Then, Gronwall's Lemma gives

$$||u|| \le C||m||. \tag{1.80}$$

This bound (1.80) and estimate (1.79) yield claim (1.75), using the convexity inequality $AB \le \varepsilon A^2 + \frac{1}{4\varepsilon}B^2$ for $A, B > 0$. Again (1.80) finally proves claim (1.74). □

1.5.2 Linearized MFG system

For proving Theorem 1.15, we introduce the linearized version of System (MFG) around its solutions and then prove that it provides the derivative of $u(t_0,x)$ with respect to the initial condition m_0.

As a preliminary step, we study a related linear system of ODE's, which will come useful several times.

$$\begin{cases} -\frac{d}{dt}z(t,x) - \alpha^*(x,\Delta^x u)\cdot\Delta^x z(t,x) = D^m F(x,m(t),1)\cdot\rho(t) + b(t,x) \\ \frac{d}{dt}\rho(t,x) = \sum_y \rho_y \alpha^*_x(y,\Delta^y u) + \sum_y m_y(t) D_p\alpha^*_x(y,\Delta^x u)\cdot\Delta^y z + c(t,x) \\ z(T,x) = D^m G(x,m(T),1)\cdot\rho(T) + z_T(x) \\ \rho(t_0,\cdot) = \rho_0, \end{cases} \tag{1.81}$$

The unknowns are z and ρ, while b, c, z_T, ρ_0 are given measurable functions, with $c(t) \in P_0(\Sigma)$, and (u,m) is the solution to (MFG). We state an immediate but useful estimate regarding the first of the two equations in (1.81).

Lemma 1.17. *If **(RegFG)** holds then the equation*

$$\begin{cases} -\frac{d}{dt}z(t,x) - \alpha^*(x,\Delta^x u)\cdot\Delta^x z(t,x) = D^m F(x,m(t),1)\cdot\rho(t) + b(t,x) \\ z(T,x) = D^m G(x,m(T),1)\cdot\rho(T) + z_T(x) \end{cases} \tag{1.82}$$

has a unique solution for each final condition $z_T(x)$ and satisfies

$$||z|| \le C\left[\max_x |z_T(x)| + ||\rho|| + ||b||\right]. \tag{1.83}$$

Proof. The well-posedness of the equation is immediate from classical ODE's theory. Integrating over the time interval $[t,T]$ and using that

$$\alpha^*(x,\Delta^x u)\cdot\Delta^x z(t,x) = \sum_y \alpha^*_y(x,\Delta^x u) z_y(t),$$

we find

$$z(t,x) - z(T,x) - \int_t^T \sum_y \alpha_y^*(x,\Delta^x u) z_y(s) ds = \int_t^T D^m F \cdot \rho(s) ds + \int_t^T b(s,x) ds.$$

Substituting the expression for $z(T,x)$, and using the bound on the control and on the derivatives of F and G we can estimate

$$\begin{aligned}\max_x |z(t,x)| \le \max_x |z_T(x)| + C \max_x |\rho(T,x)| \\ + C\int_t^T \max_x |z(s,x)| ds + C\int_t^T \max_x |\rho(s,x)| ds + \int_t^T \max_x |b(s,x)| ds\end{aligned}$$

and thus, applying Gronwall's Lemma and taking the supremum on t, we get (1.83). □

In the next result we prove the well-posedness of System (1.81) together with useful a priori estimates on its solution.

Proposition 1.18. *Assume* ***(RegH)***, ***(Mon)*** *and* ***(RegFG)***. *Then, for any (measurable) b, c, z_T, the linear system* (1.81) *has a unique solution $(z,\rho) \in C^1([0,T];\mathbb{R}^d \times P_0(\Sigma))$. Moreover it satisfies*

$$||z|| \le C(|z_T| + ||b|| + ||c|| + |\rho_0|) \tag{1.84}$$

$$||\rho|| \le C(|z_T| + ||b|| + ||c|| + |\rho_0|). \tag{1.85}$$

Proof. Without loss of generality we assume $t_0 = 0$. We use a fixed-point argument to prove the existence of a solution to (1.81). Uniqueness will be then implied by estimates (1.84) and (1.85), thanks to the linearity of the system.

We define the map $\Phi : C^0([0,T];P_0(\Sigma)) \to C^0([0,T];P_0(\Sigma))$ as follows: for a fixed $\rho \in C^0([0,T];P_0(\Sigma))$ we consider the solution $z = z(\rho)$ to Equation (1.82), and define $\Phi(\rho)$ to be the solution of the second equation in (1.81) with $z = z(\rho)$. In order to prove the existence of a fixed point of Φ, which is clearly a solution to (1.81), we apply Leray-Schauder Fixed Point Theorem. We remark the fact that more standard fixed point theorems are not applicable to this situation since we cannot assume that ρ belongs to a compact subspace of $C^0([0,T];P_0(\Sigma))$, since $P_0(\Sigma)$ is not compact. First of all, we note that $C^0([0,T];P_0(\Sigma))$ is convex and that the map Φ is trivially continuous, because of the linearity of the system. Moreover, using the equation for ρ in System (1.81), it is easy to see that Φ is a compact map, i.e. it sends bounded sets of $C^0([0,T];P_0(\Sigma))$ into bounded sets of $C^1([0,T];P_0(\Sigma))$. Thus, to apply Leray-Schauder Theorem it remains to prove that the set $\{\rho \,:\, \rho = \lambda\Phi(\rho) \text{ for some } \lambda \in [0,1]\}$ is bounded in $C^0([0,T];P_0(\Sigma))$.

Let us fix a ρ such that $\rho = \lambda\Phi(\rho)$. Then the couple (z,ρ) solves

$$\begin{cases} -\frac{d}{dt} z(t,x) - \alpha^*(x,\Delta^x u)\cdot \Delta^x z(t,x) = \lambda\left(D^m F(x,m(t),1)\cdot\rho(t) + b(t,x)\right) \\ \frac{d}{dt}\rho(t,x) = \sum_y \rho_y \alpha_x^*(y,\Delta^y u) + \lambda\left(\sum_y m_y(t) D_p \alpha_x^*(y,\Delta^x u)\cdot \Delta^y z + c(t,x)\right) \\ z(T,x) = \lambda\left(D^m G(x,m(T),1)\cdot \rho(T) + z_T(x)\right) \\ \rho(t_0,\cdot) = \lambda\rho_0. \end{cases}$$

First, we note that we can restrict to $\lambda > 0$, since otherwise $\rho = 0$. Therefore, we can use the equations (for brevity we omit the dependence of α^* on the second variable) to get

$$\frac{d}{dt}\sum_x z(t,x)\rho_x(t) = -\lambda \sum_x \rho(t,x)[D^m F(x,m(t),1)\cdot\rho(t) + b(t,x)]$$

$$
\begin{aligned}
&-\sum_{x,y}\rho_x(t)\alpha^*_y(x)[z(t,y)-z(t,x)]+\sum_{x,y}\rho_y(t)\alpha^*_x(y)z(t,x)\\
&+\lambda\sum_{x,y}m_y z(t,x)D_p\alpha^*_x(y)\cdot\Delta^y z+\lambda\sum_x c(t,x)z(t,x).
\end{aligned}
$$

The second line is 0, using the fact that $\sum_x \rho_x(t)=0$ and changing x and y in the second double sum. Integrating over $[0,T]$ and using the expression for $z(T,x)$ we obtain

$$
\begin{aligned}
\lambda\sum_x &\rho_x(T)\left[D^m G(x,m(T),1)\cdot\rho(T)+z_T(x)\right]-\lambda z(0)\cdot\rho_0\\
&=-\lambda\int_0^T\sum_x\rho_x(t)[D^mF(x,m(t),1)\cdot\rho(t)+b(t,x)]dt\\
&\quad+\lambda\int_0^T\sum_{x,y}m_yD_p\alpha^*_x(y)\cdot\Delta^y z(z(t,x)-z(t,y))dt\\
&\quad+\lambda\int_0^T\sum_x c(t,x)z(t,x)dt-\lambda\int_0^T\rho(t,x)D^mG(x,m(T),1)\cdot\rho(T)dt,
\end{aligned}
$$

where in the second term of the sum we have also used that $\sum_{x,y}[m_yD_p\alpha^*_x(y)\cdot\Delta^y z]z(t,y)=0$.

Dividing by $\lambda>0$ and bringing the terms with F and G on the left hand side, together with the term in m and $D_p\alpha^*$, we can rewrite

$$
\begin{aligned}
-\int_0^T\sum_{x,y}&m_y\Delta^y zD_p\alpha^*_x(y)\cdot\Delta^y z dt+\int_0^T\sum_x\rho(t,x)[D^mF(x,m(t),1)\cdot\rho(t)]dt\\
&+\sum_x\rho(T,x)D^mG(x,m(T),1)\cdot\rho(T)\\
&=-\sum_x z_T(x)\rho(T,x)+\sum_x z(0,x)\rho_0(x)-\int_0^T\sum_x\rho(t,x)b(t,x)dt\\
&\quad+\int_0^T\sum_x c(t,x)z(t,x)dt.
\end{aligned}
$$

We observe that, by **(Mon)** and **(RegFG)**, we have

$$
\sum_x\rho(t,x)[D^mF(x,m(t),1)\cdot\rho(t)]\geq 0, \tag{1.86}
$$

$$
\sum_x\rho(T,x)[D^mG(x,m(T),1)\cdot\rho(T)]\geq 0. \tag{1.87}
$$

Furthermore assumption (1.12) yields

$$
-\int_0^T\sum_{x,y}m_y\Delta^y zD_p\alpha^*_x(y)\cdot\Delta^y z dt\geq C^{-1}\int_0^T\sum_x m_x|\Delta^x z|^2dt,
$$

so that we can estimate the previous equality by

$$
\begin{aligned}
C^{-1}\int_0^T\sum_x m_x|\Delta^x z|^2dt &\leq |z_T\cdot\rho(T)|+|z(0)\cdot\rho_0|+\int_0^T|c(t)\cdot z(t)|\,dt\\
&\quad+\int_0^T|\rho(t)\cdot b(t)|\,dt\\
&\leq |z_T||\rho(T)|+|z(0)||\rho_0|+\int_0^T|c(t)||z(t)|dt+\int_0^T|\rho(t)||b(t)|dt.
\end{aligned} \tag{1.88}
$$

On the other hand, by the equation for ρ we have

$$\rho(t,x) = \rho_0(x) + \int_0^t \sum_y \rho(s,y)\alpha_x^*(y)ds + \int_0^t \left[\sum_y m_y D_p\alpha_x^*(y)\cdot\Delta^y z + c(x)\right] ds,$$

and thus

$$|\rho(t,x)| \le |\rho_0(x)| + M\int_0^t \sum_y |\rho_y|ds + C\int_0^t \left[\sum_y m_y|\Delta^y z| + |c(x)|\right] ds,$$

so that, by Gronwall's Lemma and taking the sum for $x\in\Sigma$ and the sup over $t\in[0,T]$,

$$\begin{aligned}
||\rho|| &\le C|\rho_0| + C\int_0^T \sum_x \sqrt{m_x}\sqrt{m_x}|\Delta^x z|dt + C||c||\\
&\le C|\rho_0| + C\int_0^T \sqrt{\sum_x (\sqrt{m_x})^2}\sqrt{\sum_x m_x|\Delta^x z|^2}dt + C||c||\\
&= C|\rho_0| + C\int_0^T \sqrt{\sum_x m_x|\Delta^x z|^2}dt + C||c||\\
&\le C|\rho_0| + C\sqrt{\int_0^T \sum_x m_x|\Delta^x z|^2 dt} + C||c||.
\end{aligned}$$

Now, we use estimate (1.88) on $\int_0^T \sum_x m_x|\Delta^x z|^2$ that we found above to get

$$\begin{aligned}
||\rho|| \le C||c|| + C|\rho_0| + C\left(|\rho_0||z(0)| + |z_T||\rho(T)| + \int_0^T |c(t)||z(t)| + \int_0^T |\rho(t)||b(t)|\right)^{\frac{1}{2}}\\
\le C||c|| + C|\rho_0|\\
+ C\left(|z(0)|^{1/2}|\rho_0|^{1/2} + |z_T|^{1/2}|\rho(T)|^{1/2}| + ||c||^{1/2}||z||^{1/2} + ||\rho||^{1/2}||b||^{1/2}\right).
\end{aligned}$$

We further estimate the right hand side using bound (1.83):

$$\begin{aligned}
||\rho|| &\le C(||c|| + |\rho_0|)\\
&\quad + C\Big[|z_T|^{1/2}|\rho(T)|^{1/2}\\
&\quad\quad + (||c||^{1/2} + |\rho_0|^{1/2})(|z_T|^{1/2} + ||\rho||^{1/2} + ||b||^{1/2}) + ||\rho||^{1/2}||b||^{1/2}\Big].
\end{aligned}$$

Using the inequality $AB \le \varepsilon A^2 + \frac{1}{4\varepsilon}B^2$ for $A, B > 0$, we obtain

$$||\rho|| \le C(||c|| + |z_T| + ||b|| + |\rho_0|) + \frac{1}{2}||\rho||,$$

which implies (1.85). Then (1.84) follows from (1.83). □

Given the solution (u,m) to System (MFG), with initial condition m_0 for m and final condition G for u, we introduce the linearized system:

$$\begin{cases}
-\frac{d}{dt}v(t,x) - \alpha^*(x,\Delta^x u(t,x))\cdot\Delta^x v(t,x) = D^m F(x,m(t),1)\cdot\mu(t)\\
\frac{d}{dt}\mu_x(t) = \sum_y \mu_y(t)\alpha_x^*(y,\Delta^y u(t,y)) + \sum_y m_y D_p\alpha_x^*(y,\Delta^y u)\cdot\Delta^y v(t,x)\\
v(T,x) = D^m G(x,m(T),1)\cdot\mu(T)\\
\mu(t_0) = \mu_0 \in P_0(\Sigma).
\end{cases} \tag{LIN}$$

Note that in the right hand side of the first equation

$$D^m F(x, m(t), 1) \cdot \mu(t) = D^m F(x, m(t), y) \cdot \mu(t)$$

for every $y \in \Sigma$, using identity (1.8) and the fact that $\mu(t) \in P_0(\Sigma)$ for every t (i.e. identity (1.10)). For this reason we just fixed the choice to $D^m F(x, m(t), 1)$ and $D^m G(x, m(T), 1)$ in System (LIN).

The existence and uniqueness of a solution $(v, \mu) \in C^1([0, T]; \mathbb{R}^d \times P_0(\Sigma))$ is ensured by Proposition 1.18. The aim is to show that the solution (v, μ) to System (LIN) satisfies

$$v(t_0, x) = D^m U(t_0, x, m_0, 1) \cdot \mu_0. \tag{1.89}$$

This proves that the solution U defined via (1.73) is differentiable with respect to m_0 in any direction μ_0, with derivative given by (1.89), and also that $D^m U$ is continuous in m. Equality (1.89) is implied by the following

Theorem 1.19. *Assume **(RegH)**, **(Mon)** and **(RegFG)**. Let (u, m) and $(\hat{u}, \hat{m})$ be the solutions to* (MFG) *respectively starting from (t_0, m_0) and $(t_0, \hat{m}_0)$. Let (v, μ) be the solution to* (LIN) *starting from (t_0, μ_0), with $\mu_0 := \hat{m}_0 - m_0$. Then*

$$||\hat{u} - u - v|| + ||\hat{m} - m - \mu|| \leq C|m_0 - \hat{m}_0|^2. \tag{1.90}$$

Proof. Set $z := \hat{u} - u - v$ and $\rho := \hat{m} - m - \mu$, they solve (1.81)

$$\begin{cases} -\frac{d}{dt} z(t, x) - \alpha^*(x, \Delta^x u) \cdot \Delta^x z(t, x) = D^m F(x, m(t), 1) \cdot \rho(t) + b(t, x) \\ \frac{d}{dt} \rho(t, x) = \sum_y \rho_y \alpha_x^*(y, \Delta^y u) + \sum_y m_y(t) D_p \alpha_x^*(y, \Delta^x u) \cdot \Delta^y z + c(t, x) \\ z(T, x) = D^m G(x, m(T), 1) \cdot \rho(T) + z_T(x) \\ \rho(t_0, \cdot) = 0, \end{cases}$$

with

$$\begin{aligned} b(t, x) &:= A(t, x) + B(t, x) \\ A(t, x) &:= -\int_0^1 [D_p H(x, \Delta^x u + s(\Delta^x \hat{u} - \Delta^x u)) - D_p H(x, \Delta^x u)] \cdot (\Delta^x \hat{u} - \Delta^x u) ds \\ B(t, x) &:= \int_0^1 [D^m F(x, m + s(\hat{m} - m), 1) - D^m F(x, m, 1)] \cdot (\hat{m} - m) ds \\ c(t, x) &:= \sum_y (\hat{m}_y - m_y) D_p \alpha_x^*(y, \Delta^y u) \cdot (\Delta^y \hat{u} - \Delta^y u) \\ &\quad + \sum_y \hat{m}_y \int_0^1 [D_p \alpha_x^*(y, \Delta^x u + s(\Delta^x \hat{u} - \Delta^x u)) - D_p \alpha_x^*(y, \Delta^y u)] \cdot (\Delta^y \hat{u} - \Delta^y u) ds \\ z_T(x) &:= \int_0^1 [D^m G(x, m(T) + s(\hat{m}(T) - m(T)), 1) - D^m G(x, m(t), 1)] \times \\ &\qquad \times (\hat{m}(T) - m(T)) ds. \end{aligned}$$

Using the assumptions, namely the Lipschitz continuity of $D_p H$, $D^2_{pp} H$, $D^m F$ and $D^m G$, and the bound $\max_x |\Delta^x u| \leq C|u|$, we estimate

$$\begin{aligned} ||b|| &\leq ||A|| + ||B|| \\ ||A|| &\leq C||\hat{u} - u||^2 \\ ||B|| &\leq C||\hat{m} - m||^2 \end{aligned}$$

$$|z_T| \leq C|\hat{m}(T) - m(T)|^2$$
$$||c|| \leq C||\hat{m} - m|| \cdot ||\hat{u} - u|| + C||\hat{u} - u||^2.$$

Applying (1.84) and (1.85) to the above system and then (1.74) and (1.75), we obtain

$$\begin{aligned} ||z|| + ||\rho|| &\leq C(|z_T| + ||b|| + ||c||) \\ &\leq C\left(||\hat{u} - u||^2 + ||\hat{m} - m||^2 + ||\hat{m} - m|| \cdot ||\hat{u} - u||\right) \\ &\leq C|m_0 - \hat{m}_0|^2. \end{aligned}$$

□

1.5.3 Proof of Theorem 1.15

We are finally in the position to prove the main theorem of this section.

1.5.3.1 Existence

Let U be the function defined by (1.73), i.e. $U(t_0, x, m_0) := u(t_0, m_0)$. We have shown in the above Theorem 1.19 that U is C^1 in m, while the fact that it is C^1 in t is clear. We compute the limit, as h tends to 0, of

$$\begin{aligned} &\frac{U(t_0 + h, x, m_0) - U(t_0, x, m_0)}{h} \\ &= \frac{U(t_0 + h, x, m_0) - U(t_0 + h, x, m(t_0 + h))}{h} + \frac{U(t_0 + h, x, m(t_0 + h)) - U(t_0, x, m_0)}{h}. \end{aligned} \tag{1.91}$$

For the first term, we have, for any $y \in \Sigma$,

$$\begin{aligned} U(t_0 + h, x, &m(t_0 + h)) - U(t_0 + h, x, m(t_0)) \\ &= [m_s := m(t_0) + s(m(t_0 + h) - m(t_0))] \\ &= \int_0^1 \frac{\partial}{\partial(m(t_0 + h) - m(t_0))} U(t_0 + h, x, m_s, y) ds \\ &= \int_0^1 D^m U(t_0 + h, x, m_s, y) \cdot (m(t_0 + h) - m(t_0)) ds \\ &= \int_0^1 ds \int_{t_0}^{t_0+h} D^m U(t_0 + h, x, m_s, y) \cdot \left(\sum_{k=1}^d m_k(t) \alpha^*(k, \Delta^k u(t)) \right) dt \\ &= \int_0^1 ds \int_{t_0}^{t_0+h} \sum_{z=1}^d \sum_{k=1}^d m_k(t) \left[D^m U(t_0 + h, x, m_s, y) \right]_z \alpha_z^*(k, \Delta^k u(t)) dt. \end{aligned}$$

Using identity (1.8), we obtain

$$\begin{aligned} U(t_0 + h, x, &m(t_0 + h)) - U(t_0 + h, x, m(t_0)) \\ &= \int_0^1 ds \int_{t_0}^{t_0+h} \sum_{z=1}^d \sum_{k=1}^d m_k(t) \left[D^m U(t_0 + h, x, m_s, k) \right]_z \alpha_z^*(k, \Delta^k u(t)) dt \\ &+ \int_0^1 ds \int_{t_0}^{t_0+h} \sum_{z=1}^d \sum_{k=1}^d m_k(t) \left[D^m U(t_0 + h, x, m_s, y) \right]_k \alpha_z^*(k, \Delta^k u(t)) dt \end{aligned}$$

$$= \int_0^1 ds \int_{t_0}^{t_0+h} \sum_{z=1}^{d} \sum_{k=1}^{d} m_k(t) \left[D^m U(t_0+h, x, m_s, k)\right]_z \alpha_z^*(k, \Delta^k u(t))dt,$$

where the last equality follows from

$$\sum_{z=1}^{d} \sum_{k=1}^{d} m_k(t) \left[D^m U(t_0+h, x, m_s, y)\right]_k \alpha_z^*(k, \Delta^k u(t))$$
$$= \sum_{k=1}^{d} m_k(t) \left[D^m U(t_0+h, x, m_s, y)\right]_k \sum_{z=1}^{d} \alpha_z^*(k, \Delta^k u(t)) = 0,$$

since $\sum_{z=1}^{d} \alpha_z^* = 0$, as $\alpha_k^*(k) = -\sum_{z\neq k} \alpha_z^*(k)$.

Summarizing, we have found that,

$$U(t_0+h, x, m(t_0+h)) - U(t_0+h, x, m(t_0))$$
$$= \int_0^1 ds \int_{t_0}^{t_0+h} dt \int_\Sigma D^m U(t_0+h, x, m_s, y) \cdot \alpha^*(y, \Delta^y u(t)) m(t)(dy).$$

Dividing by h and letting $h \to 0$, we get

$$\lim_{h\to 0} \frac{U(t_0+h, x, m(t_0+h)) - U(t_0+h, x, m(t_0))}{h}$$
$$= \int_\Sigma D^m U(t_0, x, m_0, y) \cdot \alpha^*(y, \Delta^y u(t_0)) dm_0(y)$$
$$= \int_\Sigma D^m U(t_0, x, m_0, y) \cdot \alpha^*(y, \Delta^x U(t_0, y, m_0)) dm_0(y),$$

using the continuity of $D^m U$ in time and dominate convergence to take the limit inside the integral in ds.

The second term in (1.91), for $h > 0$, is instead

$$U(t_0+h, x, m(t_0+h)) - U(t_0, x, m_0) = u_x(t_0+h) - u_x(t_0) = h\frac{d}{dt}u_x(t_0) + o(h),$$

and thus

$$\lim_{h\to 0^+} \frac{U(t_0+h, x, m(t_0+h)) - U(t_0, x, m_0)}{h} = \frac{d}{dt}u_x(t_0).$$

Finally, we can rewrite (1.91), after taking the limit $h \to 0$, to obtain

$$\partial_t U(t_0, x, m_0) = -\int_\Sigma D^m U(t_0, x, m_0, y) \cdot \alpha^*(y, \Delta^x U(t_0, y, m_0)) dm_0(y)$$
$$+ \frac{d}{dt}u_x(t_0) = [\text{using the equation for } u]$$
$$= -\int_\Sigma D^m U(t_0, x, m_0, y) \cdot \alpha^*(y, \Delta^x U(t_0, y, m_0)) dm_0(y)$$
$$+ H(x, \Delta^x U(t_0, x, m_0)) - F(x, m_0),$$

and thus

$$-\partial_t U(t_0, x, m_0) + H(x, \Delta^x U(t_0, x, m_0))$$
$$- \int_\Sigma D^m U(t_0, x, m_0, y) \cdot \alpha^*(y, \Delta^y U) dm_0(y) = F(x, m_0),$$

which is exactly (M) computed in (t_0, m_0).

1.5.3.2 Uniqueness

Let us consider another solution V of (M). Since $||D^mV|| \leq C$, we know that V is Lipschitz with respect to m, and so is $\Delta^x V$. From this remark and the Lipschitz continuity of α^* with respect to p , it follows that the equation

$$\begin{cases} \frac{d}{dt}\tilde{m}(t) = \sum_y \tilde{m}_y(t)\alpha^*(y, \Delta^y V(t, y, \tilde{m}(t))) \\ \tilde{m}(t_0) = m_0 \end{cases}$$

admits a unique solution in $[t_0, T]$.

If we now set $\tilde{u}(t, x) := V(t, x, \tilde{m}(t))$, we can compute (using for e.g. $D^mV(\cdot, \cdot, \cdot, 1)$)

$$\begin{aligned} \frac{d}{dt}\tilde{u}(t,x) &= \partial_t V(t,x,\tilde{m}(t)) + D^mV(t,x,\tilde{m}(t),1)\cdot \frac{d}{dt}\tilde{m}(t) \\ &= [\text{using the equation for } \tilde{m}] \\ &= \partial_t V(t,x,\tilde{m}(t)) + D^mV(t,x,\tilde{m}(t),1)\cdot\left(\sum_y \tilde{m}_y(t)\alpha^*(y,\Delta^y V(t,y,\tilde{m}(t)))\right) \\ &= [\text{using identity (1.8) on } D^mV(\cdot,\cdot,\cdot,1)] \\ &= \partial_t V(t,x,\tilde{m}(t)) + \int_\Sigma D^mV(t,x,\tilde{m}(t),y)\cdot\alpha^*(y,\Delta^y V(t,y,\tilde{m}(t)))\tilde{m}(t)(dy) \\ &= [\text{using the equation for } V] \\ &= H(x,\Delta^x V(t,x,\tilde{m}(t))) - F(x,\tilde{m}) = H(x,\Delta^x\tilde{u}(t,x)) - F(x,\tilde{m}(t)), \end{aligned}$$

and thus the pair $(\tilde{u}(t), \tilde{m}(t))$ satisfies

$$\begin{cases} -\frac{d}{dt}\tilde{u}(t,x) + H(x,\Delta^x\tilde{u}(t,x)) = F(x,\tilde{m}(t)), \\ \frac{d}{dt}\tilde{m}_x(t) = \sum_j \tilde{m}_y(t)\alpha^*_x(y,\Delta^y\tilde{u}(t,y)), \\ \tilde{u}(T,x) = V(T,x,\tilde{m}(T)) = G(x,\tilde{m}(T)), \\ \tilde{m}(t_0) = m_0. \end{cases}$$

Namely, $(\tilde{u}, \tilde{m})$ solves System (MFG), whose solution is unique thanks to Proposition 1.16, so that we can conclude $V(t_0, x, m_0) = U(t_0, x, m_0)$ for each (t_0, x, m_0), and thus the uniqueness of solutions to (M) follows.

1.5.3.3 Regularity

It remains to prove that the unique classical solution defined via (1.73) is regular, in the sense of Definition 1.6, i.e. that D^mU is Lipschitz continuous with respect to m, uniformly in t, x.

So let (u_1, m_1) and (u_2, m_2) be two solution to (MFG) with initial conditions $m_1(t_0) = m_0^1$ and $m_2(t_0) = m_0^2$, respectively. Let also (v_1, μ_1) and (v_2, μ_2) be the associated solutions to (LIN) with $\mu_1(t_0) = \mu_2(t_0) = \mu_0$. Recall from Equation (1.89) that $v_1(t_0, x) = D^mU(t_0, x, m_0^1, 1) \cdot \mu_0$ and $v_2(t_0, x) = D^mU(t_0, x, m_0^2, 1) \cdot \mu_0$, thus we have to estimate the norm $||v_1 - v_2||$.

Set $z := v_1 - v_2$ and $\rho := \mu_1 - \mu_2$. They solve the linear system (1.81) with $\rho_0 = 0$ and

$$\begin{aligned} b(t,x) &:= [D^mF(x,m_1,1) - D^mF(x,m_2,1)]\cdot\mu_2 + [\alpha^*(x,\Delta^x u_1) - \alpha^*(x,\Delta^x u_2)]\cdot\Delta^x v_2 \\ c(t,x) &:= \sum_y \mu_{2,y}\left[\alpha^*_x(y,\Delta^y u_1) - \alpha^*_x(y,\Delta^y u_2)\right] \end{aligned}$$

$$+ \sum_y [m_{1,y} D_p \alpha^*_x(y, \Delta^y u_1) - m_{2,y} D_p \alpha^*_x(y, \Delta^y u_2)] \cdot \Delta^x v_2$$
$$z_T(x) := [D^m G(x, m_1(T), 1) - D^m G(x, m_2(t), 1)] \cdot \mu_2.$$

Using the Lipschitz continuity of D_pH, $D^2_{pp}H$, D^mF and D^mG, applying the bounds (1.84) to v_2 and (1.85) to μ_2 and also (1.74) and (1.75), we estimate

$$||b|| \le C||m_1 - m_2|| \cdot ||\mu_2|| + C||u_1 - u_2|| \cdot ||v_2|| \le C|m^1_0 - m^2_0| \cdot |\mu_0|$$
$$||c|| \le C||u_1 - u_2|| \cdot ||\mu_2|| + C||m_1 - m_2|| \cdot ||v_2|| + C||u_1 - u_2|| \cdot ||v_2|| \le C|m^1_0 - m^2_0| \cdot |\mu_0|$$
$$|z_T| \le C||m_1 - m_2|| \cdot ||\mu_2|| \le C|m^1_0 - m^2_0| \cdot |\mu_0|.$$

Then (1.84) gives

$$||z|| \le C(||b|| + ||c|| + |z_T|) \le C|m^1_0 - m^2_0| \cdot |\mu_0|,$$

which, since $z(t_0, x) = (D^m U(t_0, x, m^1_0, 1) - D^m U(t_0, x, m^2_0, 1)) \cdot \mu_0$, yields

$$\begin{aligned}\max_x |D^m U(t_0, &x, m^1_0, 1) - D^m U(t_0, x, m^2_0, 1)| \\ &\le C \max_x \sup_{\mu_0 \in P_0(\Sigma)} \frac{|(D^m U(t_0, x, m^1_0, 1) - D^m U(t_0, x, m^2_0, 1)) \cdot \mu_0|}{|\mu_0|} \\ &\le C|m^1_0 - m^2_0|.\end{aligned}$$

1.6 Conclusions

Let us summarize the results we have obtained. The two sets of assumptions are given in Section 2.2 and verified in Example 2.1.

1. If **(H1)** holds and there exists a regular solution U to the master equation (M), in the sense of Definition 1.6, then the value functions of the N-player game converge to U (Theorem 1.7) and the optimal trajectories (1.29) satisfy a propagation of chaos property, i.e they converge to the limit i.i.d. solution to (1.30) (Theorem 1.8);

2. Under the assumptions required for convergence, the empirical measures processes (1.49) associated with the optimal trajectories satisfy a Central limit Theorem (Theorem 1.13) and a Large Deviation Principle with rate function I in (1.65) (Theorem 1.14);

3. Assuming **(RegH)**, **(Mon)** and **(RegFG)**, there exists a unique classical solution to (M) and it is also regular in the sense of Definition 1.6.

CHAPTER 2

The convergence problem in a two state model without uniqueness

In this chapter we consider finite state N-player and mean field games, restricting the framework of Chapter 1 to the case where the position of each agent belongs to a binary state space $\{-1, 1\}$. If there is uniqueness of mean field game solutions, e.g. under monotonicity assumptions, then the results of Chapter 1 apply: the master equation possesses a smooth solution which can be used to prove convergence of the value functions and of the feedback Nash equilibria of the N-player game, as well as a propagation of chaos property for the associated optimal trajectories.

Here instead, we study an example with anti-monotonic costs, and show that the mean field game has exactly three solutions, which are found explicitly. We prove that the value functions converge to the entropy solution of the master equation, which in this case can be written as a scalar conservation law in one space dimension, and that the optimal trajectories admit a limit: they select one mean field game soution, so there is propagation of chaos (except for a critical case, as we shall see, in which the limit is random). Moreover, viewing the mean field game system as the necessary conditions for optimality of a deterministic control problem, we show that the N-player game selects the optimizer of this problem. A two-state non-uniqueness example was first considered in [64, 65], where the master equation was studied formally and numerical evidence on the convergence behavior was presented; our example should also be compared to the "illuminating example" in [77, Sect. 3.3]) and to the example in [3, Sect. 3.3], both in the diffusion setting. In the infinite time horizon and finite state case, an example of non-uniqueness is studied in [38], via numerical simulations, where periodic orbits emerge as solutions to the mean field game.

Different recent results are related to our example: among others, we cite [87], where the authors address the convergence problem for a class of mean field games of optimal stopping. The limit model there possesses multiple solutions, which are grouped into three classes according to a qualitative criterion characterizing the proportion of players that have stopped at any given time. Solutions in one of the three classes will always arise as limit points of N-player Nash equilibria, solutions in the second class may be selected in the limit, while solutions in the third class cannot be reached through N-player Nash equilibria. In [78], the author attacks the convergence problem in Markov feedback strategies by probabilistic methods. For a class of games with non-degenerate Brownian dynamics that may exhibit non-uniqueness, the author shows that all limit points of the

N-player feedback Nash equilibria are concentrated, as in the open-loop case, on weak solutions of the mean field game. These solutions are more general than randomizations of ordinary ("strong") solutions of the mean field game; their flows of measures, in particular, are allowed to be stochastic containing additional randomness. Still, uniqueness in ordinary solutions implies uniqueness in weak solutions, which permits to partially recover the results in [15]. In this direction we acknowledge the recent result [17], which deals with the ergodic setting, where the authors show that the limit of Nash equilibria in generalized Markov strategies is not necessarily a mean field game equilibrium. The result is thus in sharp contrast with the finite horizon case of [78].

The question of which weak mean field game solutions can appear as limits of feedback Nash equilibria in a situation of non-uniqueness seems to be mainly open. In [42], a class of linear-quadratic mean field games with multiple solutions is studied in the diffusion setting. They prove that by adding a common noise to the limit dynamics uniqueness of solutions is re-established. As a converse to this regularization by noise result, they identify the mean field game solutions that are selected when the common noise tends to zero as those induced by the (unique weak) entropy solution of the master equation of the original problem. The interpretation of the master equation as a scalar conservation law works in their case thanks to a one-dimensional parametrization of an a priori infinite dimensional problem. Limit points of N-player Nash equlibria are also considered in [42], but in stochastic open-loop strategies. Again, the mean field game solutions that are selected are those induced by the entropy solution of the master equation. Interestingly, these solutions are not minimal cost solutions; indeed, the solution which minimizes the cost of the representative player in the mean field game is shown to be different from the ones selected by the limit of the Nash equilibria. In [42], the N-player limit and the vanishing common noise limit both select two solutions of the original mean field game with equal probability. This is due to the fact that in [42] the initial distribution for the state trajectories is chosen to sit at the discontinuity of the unique entropy solution of the master equation. In our case, we expect to see the same behavior if we started at the discontinuity, see Section 2.3 below. We also mention [6], where the authors consider a two-state example without uniqueness with an anti-monotonic cost, where a running cost term is also added in the interaction. As in our example, they show that the entropy solution of the master equation is of particular importance as it is the one which gets selected in the limit. Some of the above models, including ours, can be framed into the class of *submodular* mean field games. In this regard, we mention the recent result [48], where the authors prove that the set of solutions in these types of models enjoys an ordered lattice structure.

It is worth mentioning that the opposite framework to the one treated here is considered in the examples presented in [49] and in [19, Sect. 7.2.5]. In these examples, uniqueness of mean field game solutions holds, but there are multiple feedback Nash equilibria for the N-player game. This is due to the fact that in both cases the authors consider a finite action set (while for us it is continuous), so that in particular the Nash system is not well-posed. They prove that there is a sequence of (feedback) Nash equilibria which converges to the mean field game limit, but also a sequence that does not converge.

The rest of this chapter is organized as follows. In Section 2.1, we briefly recall the notation of Chapter 1 for mean field and N-player games with finite state space. Section 2.2 presents the two-state example, starting from the limit model, analyzed first in terms of the mean field game system (Subsection 2.2.1), then in terms of its master equation (Subsection 2.2.2). In Subsections 2.2.4 and 2.2.5 we show that the N-player Nash equilibria converge to the unique entropy solution of the master equation; cf. Theorems 2.7

and 2.10 below for convergence of value functions and propagation of chaos, respectively. The qualitative property of the Nash equilibria used in the proofs of convergence is in Subsection 2.2.3. Subsection 2.2.7 gives the variational characterization of the solution that is selected by the Nash equilibria. Concluding remarks are in Section 2.3.

2.1 Mean field games with finite state space

In this section we briefly recall the notation and the equations in play for finite state mean field games, introduced in Sections 1.1 and 1.2.1 of Chapter 1, to which the reader can refer for more details.

2.1.1 The N-player game

As in Chapter 1, we consider the continuous time evolution of the states $(X_i(t))_{i=1,\dots,N}$ of N players; the state of each player belongs to a given finite set Σ. Players are allowed to control, via an arbitrary *feedback*, their jump rates. For $i = 1, 2, \dots, N$ and $y \in \Sigma$, we denote by $\alpha^i_y : [0, T] \times \Sigma^N \to [0, +\infty)$ the rate at which player i jumps to the state $y \in \Sigma$. Let $\boldsymbol{\alpha}^N \in \mathcal{A}^N$ denote the controls of all players, to which we refer also as *strategy vector*. Recall $P(\Sigma)$ to be the simplex of probability measures on Σ. To every $\boldsymbol{x} \in \Sigma^N$ we associate the element of $P(\Sigma)$

$$m^{N,i}_{\boldsymbol{x}} := \frac{1}{N-1} \sum_{j=1, j\neq i}^{N} \delta_{x_j}. \tag{2.1}$$

Thus, $m^{N,i}_{\boldsymbol{X}}(t) := m^{N,i}_{\boldsymbol{X}_t}$ is the empirical measure of all the players except the i-th. Recall the *cost* associated to the i-th player

$$J^N_i(\boldsymbol{\alpha}^N) := \mathbb{E}\left[\int_0^T \left[L(X_i(t), \alpha^i(t, \boldsymbol{X}_t)) + F\left(X_i(t), m^{N,i}_{\boldsymbol{X}}(t)\right)\right] dt + G\left(X_i(T), m^{N,i}_{\boldsymbol{X}}(T)\right)\right].$$

At the N-player level, the concept of solution is that of a Nash equilibrium, given by Definition 1.2. We work under the same assumptions of Chapter 1 (see Section 1.2.2, and in particular (1.11)) that guarantee the existence and uniquess of the Nash equilibrium for the N-player game. Within this framework, the search for the Nash equilibrium is equivalent to solving System (HJB), a system of $N|\Sigma|^N$ coupled ODE's, indexed by $i \in \{1, \dots, N\}$ and $\boldsymbol{x} \in \Sigma^N$, whose well-posedness for all $T > 0$ can be proved through standard ODEs techniques under regularity assumptions which guarantee that a^* and H are uniformly Lipschitz in their second variable (i.e. Assumption **(H1)** of Section 1.2.2 - which we assume valid throughout the chapter). Under these conditions, the N-player game has a unique Nash equilibrium given by the feedback strategy vector $\boldsymbol{\alpha}^N \in \mathcal{A}^N$ defined by

$$\alpha^{i,N}(t, \boldsymbol{x}) := a^*(x_i, \Delta^i v^{N,i}(t, \boldsymbol{x})) \qquad i = 1, \dots, N.$$

2.1.2 The macroscopic limit: the mean field game and the master equation

We recall that the limit as $N \to +\infty$ of the N-player game admits two alternative descriptions, illustrated in Section 1.2.4. On the one hand we have the mean field game system (MFG), implemented by coupling the HJB equation of the control problem with cost (1.25) for a fixed deterministic flow of probabilities m, with the forward Kolmogorov equation for the distribution of the optimal evolution of the representative agent, which must coincide with m, yielding the mean field equilibrium condition (1.26). It is known, and largely exemplified in this chapter, that well-posedness of (HJB) does not imply uniqueness of solution to (MFG).

An alternative description of the macroscopic limit stems from the ansatz, justified by Proposition 1.5, that the solution to the N-player Hamilton-Jacobi-Bellman system (HJB) is of the form

$$v^{N,i}(t, \boldsymbol{x}) = v^N(t, x_i, m_{\boldsymbol{x}}^{N,i}),$$

for some $v^N : [0,T] \times \Sigma \times P(\Sigma) \to \mathbb{R}$. Assuming v^N admits a limit U as $N \to +\infty$, we formally obtain that U solves the *master equation* (M), where we recall the definition of derivative $D^m U : [0,T] \times \Sigma \times P(\Sigma) \times \Sigma \to \mathbb{R}^\Sigma$ with respect to $m \in P(\Sigma)$

$$[D^m U(t,x,m,y)]_z := \lim_{s \downarrow 0} \frac{U(t,x,m + s(\delta_z - \delta_y)) - U(t,x,m)}{s}. \tag{2.2}$$

We conclude this section by recalling that uniqueness in both (MFG) and (M) is guaranteed if the cost functions F and G are *monotone* in the Lasry-Lions sense, i.e. for every $m, m' \in P(\Sigma)$,

$$\sum_{x \in \Sigma} (F(x,m) - F(x,m'))(m_x - m'_x) \geq 0, \tag{2.3}$$

and the same for the final cost G. We are interested here in examples that violate this monotonicity condition.

2.2 An example of non uniqueness

We consider now a special example within the class of finite state models described above. We let $\Sigma := \{-1, 1\}$ be the state space. An element $m \in P(\Sigma)$ can be identified with its mean $m_1 - m_{-1}$; so from now we write $m \in [-1,1]$ to denote the mean, while the element of $P(\Sigma)$ will be denoted only in vector form (m_1, m_{-1}). We also write $\alpha^i(t, \boldsymbol{x})$ for $\alpha^i_{-x_i}(t, \boldsymbol{x})$, i.e. the rate at which player i *flips* its state from x_i to $-x_i$. Moreover we choose

$$L(x,a) := \frac{a^2}{2}, \qquad F(x,m) \equiv 0, \qquad G(x,m) := -mx.$$

Observe that the final cost G favors alignment with the majority, while the running cost is a simple quadratic cost. Compared to condition (2.3), note that the final cost is *anti-monotonic*, as

$$\sum_{x \in \Sigma} (G(x,m) - G(x,m'))(m_x - m'_x) = -(m - m')^2 \leq 0.$$

The associated Hamiltonian is given by

$$H(x,p) = \sup_{a \geq 0} \left\{ a p_{-x} - \frac{a^2}{2} \right\} = \frac{(p^-_{-x})^2}{2}, \tag{2.4}$$

with $a^*(x,p) = p^-_{-x}$, where p^- denotes the negative part of p. From now on, we identify p with $p_{-x} \in \mathbb{R}$ and $\Delta^x u$ with its non-zero component $u(-x) - u(x)$.

2.2.1 The mean field game system

The first equation in (MFG), i.e the HJB equation for the value function $u(t,x)$, reads, using (2.4),

$$\begin{cases} -\frac{d}{dt}u(t,x) + \frac{1}{2}\left[(\Delta^x u(t,x))^-\right]^2 = 0 \\ u(T,x) = -m(T)x \end{cases} \tag{2.5}$$

Now define $z(t) := u(t,-1) - u(t,1)$. Subtracting the equations (2.5) for $x = \pm 1$ and observing that

$$\left[(\Delta^x u(t,-1))^-\right]^2 - \left[(\Delta^x u(t,1))^-\right]^2 = z|z|,$$

we have that $z(t)$ solves

$$\begin{cases} \dot{z} = \frac{z|z|}{2} \\ z(T) = 2m(T). \end{cases} \tag{2.6}$$

This equation must be coupled with the forward Kolmogorov equation, i.e. the second equation in (MFG), that reads $\dot{m} = -m|z| + z$. The mean field game system takes therefore the form:

$$\begin{cases} \dot{z} = \frac{z|z|}{2} \\ \dot{m} = -m|z| + z \\ z(T) = 2m(T) \\ m(0) = m_0. \end{cases} \tag{2.7}$$

Proposition 2.1. *Let $T(m_0)$ be the unique solution in $T \in \left[\frac{1}{2}, 2\right]$ to the equation*

$$|m_0| = \frac{(2T-1)^2(T+4)}{27T}. \tag{2.8}$$

Then, for every $m_0 \in [-1,1] \setminus \{0\}$, System (2.7) admits

(i) a unique solution for $T < T(m_0)$;

(ii) two distinct solutions for $T = T(m_0)$;

(iii) three distinct solutions for $T > T(m_0)$.

If $m_0 = 0$, then $T(0) = 1/2$ and (2.7) admits

(i) a unique solution for $T \le 1/2$;

(ii) three distinct solutions for $T > 1/2$: the constant zero solution, (z_+, m_+), and (z_-, m_-), where $m_+(t) = -m_-(t) > 0$ for every $t \in (0,T]$.

Proof. Note that (2.6) can be solved as a final value problem, giving

$$z(t) = \frac{2m(T)}{|m(T)|(T-t)+1}. \tag{2.9}$$

This can then be inserted in the forward Kolmogorov equation $\dot{m} = -m|z| + z$, giving as unique solution

$$m(t) = (m_0 - \operatorname{sgn}(m(T)))\left(\frac{|m(T)|(T-t)+1}{|m(T)|T+1}\right)^2 + \operatorname{sgn}(m(T)). \tag{2.10}$$

These are actually solutions of (2.7) if and only if the consistency relation obtained by setting $t = T$ in (2.10) holds, i.e. if and only if $m(T) = M$ solves

$$T^2 M^3 + T(2-T)M|M| + (1-2T)M - m_0 = 0. \tag{2.11}$$

Moreover, distinct solutions of (2.11) correspond to distinct solutions of (2.7). We first look for nonnegative solutions of (2.11). Set

$$f(M) := T^2 M^3 + T(2-T)M^2 + (1-2T)M - m_0.$$

Note that

$$f'(M) < 0 \quad \Longleftrightarrow \quad M \in \left(-\frac{1}{T}, \frac{2T-1}{3T}\right).$$

If $T \leq \frac{1}{2}$ then f is strictly increasing in $(0, +\infty)$, so the equation $f(M) = 0$ admits a unique nonnegative solution if $m_0 \geq 0$, otherwise there is no nonnegative solution. If $T > \frac{1}{2}$, then f restricted to $(0, +\infty)$ has a global minimum at $M^* = \frac{2T-1}{3T}$. If $m_0 > 0$ then there is still a unique nonnegative solution, while for $m_0 = 0$ there are two nonnegative solution, one of which is zero. If, instead, $m_0 < 0$, so that $f(0) > 0$, the equation $f(M) = 0$ has zero, one or two nonnegative solutions, depending on whether $f(M^*) > 0$, $f(M^*) = 0$ or $f(M^*) < 0$ respectively. Observing that

$$f(M^*) = -m_0 - \frac{(2T-1)^2(T+4)}{27T},$$

we see that those three alternatives occur if $T < T(m_0)$, $T = T(m_0)$ and $T > T(m_0)$ respectively. The case $M \leq 0$ is treated similarly.

□

2.2.2 The master equation

Identifying again a probability on Σ with its mean m, using the expression for H and its minimizer given in (2.4), Equation (M) takes the form

$$\begin{cases} -\frac{\partial U}{\partial t}(t,x,m) + \frac{1}{2}\left[(\Delta^x U(t,x,m))^-\right]^2 - D^m U(t,x,m,1)\,(\Delta^x U(t,1,m))^-\,\frac{1+m}{2} \\ \qquad\qquad -D^m U(t,x,m,-1)\,(\Delta^x U(t,-1,m))^-\,\frac{1-m}{2} = F(x,m), \\ U(T,x,m) = G(x,m), \quad (x,m) \in \{-1,1\} \times [-1,1]. \end{cases} \tag{2.12}$$

In (2.12), the derivative $D^m U$ is still intended in the sense introduced in (2.2), but identifying the resulting vector with its non-zero component (e.g. $D^m U(t,x,m,1) = [D^m U(t,x,m,1)]_{-1} = \frac{\partial}{\partial(m_{-1}-m_1)} U(t,x,m)$). Similarly, we identify the vector $\Delta^x U$ with its non-zero component. Setting

$$Z(t,m) := U(T-t,-1,m) - U(T-t,1,m),$$

we easily derive a closed equation for Z:

$$\begin{cases} \frac{\partial Z}{\partial t} + \frac{\partial}{\partial m}\left(m\frac{Z|Z|}{2} - \frac{Z^2}{2}\right) = 0, \\ Z(0,m) = 2m, \end{cases} \tag{2.13}$$

where $\frac{\partial}{\partial m}$ is denoting the differentiation in the usual sense with respect to $m \in [-1,1]$. In particular, observe that $\frac{\partial}{\partial m} = \frac{1}{2}\frac{\partial}{\partial(m_{-1}-m_1)}$.

Note that this equation has the form of a scalar *conservation law*

$$\begin{cases} \frac{\partial Z}{\partial t}(t,m) + \frac{\partial}{\partial m}\mathfrak{g}(m, Z(t,m)) = 0 \\ Z(0,m) = \mathfrak{f}(m). \end{cases} \tag{2.14}$$

Scalar conservation laws typically possess unique smooth solutions for small time, but develop singularities in finite time: weak solutions exist but uniqueness may fail. To recover uniqueness the notion of *entropy solution* is introduced (see Appendix A for a very brief, non-exhaustive overview). A simple sufficient condition can be given for piecewise smooth functions (see [34]):

Proposition 2.2. *Let $Z(t,m)$ be a piecewise $\mathcal{C}^1$ function, which is $\mathcal{C}^1$ outside a $\mathcal{C}^1$ curve $m = \gamma(t)$, and assume the following conditions hold:*

(i) Z solves (2.14) *in the classical sense outside the curve $m = \gamma(t)$.*

(ii) The initial condition $Z(0,m) = \mathfrak{f}(m)$ holds for every m.

(iii) Denoting

$$Z_+(t) := \lim_{m \downarrow \gamma(t)} Z(t,m), \quad Z_-(t) := \lim_{m \uparrow \gamma(t)} Z(t,m),$$

we have that, for every $t \geq 0$ and every c strictly between $Z_-(t)$ and $Z_+(t)$,

$$\dot{\gamma}(t) = \frac{\mathfrak{g}(\gamma(t), Z_-(t)) - \mathfrak{g}(\gamma(t), Z_+(t))}{Z_-(t) - Z_+(t)}, \tag{2.15}$$

$$\frac{\mathfrak{g}(\gamma(t), c) - \mathfrak{g}(\gamma(t), Z_+(t))}{c - Z_+(t)} < \dot{\gamma}(t) < \frac{\mathfrak{g}(\gamma(t), c) - \mathfrak{g}(\gamma(t), Z_-(t))}{c - Z_-(t)}. \tag{2.16}$$

Then, Z is the unique entropy solution to (2.14)*.*

Condition (2.15) is called the *Rankine-Hugoniot condition*, while (2.16) is called the *Lax condition*. When specialized to the case $\mathfrak{g}(m,z) := m\frac{z|z|}{2} - \frac{z^2}{2}$ and $\gamma(t) \equiv 0$ we simply obtain

$$Z_+(t) = -Z_-(t) \geq 0. \tag{2.17}$$

For Equation (2.13), the entropy solution can be explicitly found. Let

$$g(M,t,m) := t^2 M^3 + t(2-t)M|M| + (1-2t)M - m \tag{2.18}$$

and $M(t,m)$ denote the unique solution to $g(M,t,m) = 0$ with the same sign of m, if $m \neq 0$; M is defined for any time and let $M(t,0) \equiv 0$. Define

$$Z(t,m) := \frac{2M(t,m)}{t|M(t,m)| + 1}. \tag{2.19}$$

Such function has a unique discontinuity in $m = 0$, for $t > 1/2$, and is $\mathcal{C}^1$ outside. However, observe that Equation (2.13) must be solved in the finite interval $t \in [0,T]$, where T is the final time appearing in (2.12). Thus, for $T < 1/2$ the solution is regular.

Theorem 2.3. *The function Z defined in* (2.19) *is the unique entropy admissible weak solution to* (2.13)*.*

Proof. From the properties of $g(M,t,m)$, it follows that

$$\lim_{m\downarrow 0} M(t,m) = -\lim_{m\uparrow 0} M(t,m) \geq 0,$$

for any time. These limits correspond to the solutions m_+ and m_- of Proposition 2.1, evaluated at the terminal time. Therefore (2.17) is satisfied. We remark that the conservation law is set in the domain $[-1,1]$ without any boundary condition, but this is not a problem as we have invariance of the domain under the action of the characteristics. □

Remark 2.4. *We observe that to the entropy solution* (2.19) *of* (2.13) *there corresponds a unique solution of* (2.12). *It can be constructed via the method of characteristic curves, in terms of a specific solution to the mean field game system for the couple* (u,m)*, the one that corresponds to the solution to* (2.7) *employed in the definition of* (2.19).

It is known that, if there were a regular solution to the master equation (2.13), thus Lipschitz in m, then this solution would provide a unique solution to the mean field game system (2.7), since the KFP equation would be well posed for any initial condition, when using $z(t) = Z(T-t, m(t))$ induced by the solution to the master equation:

$$\begin{cases} \dot{m} = -m|Z(T-t,m)| + Z(T-t,m) \\ m(0) = m_0. \end{cases} \tag{2.20}$$

In our example there are no regular solutions to the master equation; however the entropy solution still induces a unique mean field game solution, if $m_0 \neq 0$.

Proposition 2.5. *Let* Z *be the entropy solution defined in* (2.19). *Then* (2.20) *admits a unique solution* m^**, for any* T*, if* $m_0 \neq 0$*: it is the unique solution which does not change sign, for any time.*

Proof. Let $m_0 > 0$. If t and $|m - m_0|$ are small then $Z(T-t,m)$ is regular (Lipschitz-continuous) and remains positive. So we have a unique solution to (2.20), for small time $t \in [0,t_0]$; moreover it is such that $\dot{m} > 0$ and hence in particular $m(t_0) > m_0$. Thus we can iterate this procedure starting from $m(t_0) > 0$: we end up with the required solution, which is positive and such that $m(t) > m_0$ for any time. This solution is unique (for any T) since $Z(t,m)$ is Lipschitz for $m \in [m_0, 1]$. In fact the other two solutions described in Proposition 2.1 would require the vector field Z in (2.20) to be negative for any time, and this is not possible when considering the entropy solution Z. The same argument gives the claim when $m_0 < 0$. □

2.2.3 Properties of the $N+1$-player game

We consider now the game played by $N+1$ players, labeled by the integers $\{0,1,\ldots,N\}$. By symmetry, we can interpret the player with label 0 as the *representative player.* Let

$$\mu^N_{\boldsymbol{x}} := \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i=1} \in \left\{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N-1}{N}, 1\right\}$$

be the fraction of the "other" players having state 1. Comparing with the notations in (2.1), note that $\mu^N_{\boldsymbol{x}} = \frac{1+m^{N+1,0}_{\boldsymbol{x}}}{2}$. In what follows, we use N rather than $N+1$ as apex in all objects related to the $N+1$-player game. By symmetry again, the value function $v^{N,0}(t,\boldsymbol{x})$ introduced in (HJB) is of the form

$$v^{N,0}(t,\boldsymbol{x}) = V^N(t, x_0, \mu^N_{\boldsymbol{x}}),$$

where $V^N : [0,T] \times \{-1,1\} \times \left\{0, \frac{1}{N}, \frac{2}{N}, \dots, \frac{N-1}{N}, 1\right\} \to \mathbb{R}$. Since the model we are considering, besides permutation invariance, is invariant by the sign change of the state vector, it follows that

$$V^N(t,1,\mu^N_{\boldsymbol{x}}) = V^N(t,-1,1-\mu^N_{\boldsymbol{x}}). \tag{2.21}$$

We can therefore redefine $V^N(t,\mu) := V^N(t,1,\mu)$; from System (HJB) we derive the following closed equation for V^N:

$$\begin{cases} -\frac{d}{dt}V^N(t,\mu) + H(V^N(t,1-\mu) - V^N(t,\mu)) \\ \quad = N\mu \left[V^N(t,1-\mu) - V^N(t,\mu)\right]^- \left[V^N\left(t,\mu-\frac{1}{N}\right) - V^N(t,\mu)\right] \\ \qquad +N(1-\mu)\left[V^N\left(t,\mu+\frac{1}{N}\right) - V^N\left(t,1-\mu-\frac{1}{N}\right)\right]^- \left[V^N\left(t,\mu+\frac{1}{N}\right) - V^N(t,\mu)\right], \\ V^N(T,\mu) = -(2\mu-1), \end{cases} \tag{2.22}$$

with $H(p) = \frac{(p^-)^2}{2}$. It is easy to check that, when imposing a final datum $V^N(T,\mu) \in [-1,1]$, any solution to System (2.22) is such that $V^N(t,\mu) \in [-1,1]$ for any $t < T$. The locally Lipschitz property of the vector field is thus enough to conclude the existence and uniqueness of solution for any $T > 0$ for the above system with $|V^N(t,\mu)| \leq 1$. Such solution allows to obtain the unique Nash equilibrium, given by the feedback strategy

$$\alpha^{0,N}(t,\boldsymbol{x}) = \begin{cases} \left[V^N(t,1-\mu^N_{\boldsymbol{x}}) - V^N(t,\mu^N_{\boldsymbol{x}})\right]^- & \text{for } x_0 = 1, \\ \left[V^N(t,1-\mu^N_{\boldsymbol{x}}) - V^N(t,\mu^N_{\boldsymbol{x}})\right]^+ & \text{for } x_0 = -1. \end{cases} \tag{2.23}$$

We now set

$$Z^N(t,\mu) := V^N(t,1-\mu) - V^N(t,\mu).$$

The following result, that will be useful later, shows that if the representative player agrees with the majority, i.e. $x_0 = 1$ and $\mu^N_{\boldsymbol{x}} \geq \frac{1}{2}$, or $x_0 = -1$ and $\mu^N_{\boldsymbol{x}} \leq \frac{1}{2}$, then she/he keeps her/his state by applying the control zero.

Theorem 2.6. *For any $\mu \in S_N = \left\{0, \frac{1}{N}, \dots, 1\right\}$, we have*

$$Z^N(t,\mu) \geq 0 \quad (\alpha^N(t,1,\mu) = 0) \quad \text{if } \mu \geq \frac{1}{2}, \tag{2.24}$$

$$Z^N(t,\mu) \leq 0 \quad (\alpha^N(t,-1,\mu) = 0) \quad \text{if } \mu \leq \frac{1}{2}. \tag{2.25}$$

Proof. We prove (2.24), the proof of (2.25) is similar. For any N even, observe that $Z^N(\frac{1}{2}) = 0$, so that it is enough to prove the claim for $\mu \geq \frac{1}{2} + \frac{1}{N}$. Define

$$W^N(t,\mu) := V^N(t,\mu) - V^N(t,\mu+\frac{1}{N}).$$

By (2.22),

$$\begin{aligned} \frac{d}{dt}Z^N(t,\mu) &= H(-Z^N(t,\mu)) - H(Z^N(t,\mu)) \\ &\quad + N\mu\left\{\left(Z^N(t,\mu)\right)^- W^N\left(t,\mu-\frac{1}{N}\right)\left(Z^N\left(t,\mu-\frac{1}{N}\right)\right)^- W^N(t,1-\mu)\right\} \\ &\quad - N(1-\mu)\left\{\left(Z^N\left(t,\mu+\frac{1}{N}\right)\right)^+ W^N(t,\mu) \right. \\ &\qquad \left. + \left(Z^N(t,\mu)\right)^+ W^N\left(t,1-\mu-\frac{1}{N}\right)\right\} \end{aligned} \tag{2.26}$$

and

$$
\begin{aligned}
\frac{d}{dt}W^N(t,\mu) &= H(Z^N(t,\mu)) - H\left(Z^N\left(t,\mu+\frac{1}{N}\right)\right) \\
&\quad - N\mu\left(Z^N(t,\mu)\right)^- W^N\left(t,\mu-\frac{1}{N}\right) \\
&\quad + N\left(\mu+\frac{1}{N}\right)\left(Z^N\left(t,\mu+\frac{1}{N}\right)\right)^- W^N(t,\mu) \\
&\quad + N(1-\mu)\left(Z^N\left(t,\mu+\frac{1}{N}\right)\right)^+ W^N(t,\mu) \\
&\quad - N\left(1-\mu-\frac{1}{N}\right)\left(Z^N\left(t,\mu+\frac{2}{N}\right)\right)^+ W^N\left(t,\mu+\frac{1}{N}\right).
\end{aligned} \tag{2.27}
$$

Note that, for $\mu > \frac{1}{2}$, $Z^N(T,\mu) = 4\mu - 2 > 0$ and $W^N(T,\mu) = \frac{2}{N} > 0$. So, set

$$
s := \sup\left\{t \le T : Z^N(t,\nu) \le 0 \text{ or } W^N(t,\nu) \le 0 \text{ for some } \nu > \frac{1}{2}\right\}.
$$

We complete the proof by showing that $s = -\infty$. Assume $s > -\infty$. For $t \in [s,T]$ we have $Z^N(t,\mu) \ge 0$ and $W^N(t,\mu) \ge 0$ for all $\mu > \frac{1}{2}$, so, from (2.26), observing that the terms in $\left(Z^N\right)^-$ disappear,

$$
\begin{aligned}
\frac{d}{dt}Z^N(t,\mu) &\le H(-Z^N(t,\mu)) + N(1-\mu)Z^N(t,\mu)W^N\left(t,1-\mu-\frac{1}{N}\right) \\
&= Z^N(t,\mu)\left[\frac{1}{2}Z^N(t,\mu) + N(1-\mu)W^N\left(t,1-\mu-\frac{1}{N}\right)\right].
\end{aligned}
$$

Since the control zero is suboptimal, it follows that $|V^N(t,\mu)| \le 1$ for all t,μ, so that $|Z^N(t,\mu)| \le 2$ and $|W^N(t,\mu)| \le 2$. Therefore, for $t \in [s,T]$, $Z^N(t,\mu)$ is bounded from below by the solution to

$$
\begin{cases}
\frac{d}{dt}z(t) = z(t)\left[1 + 2N(1-\mu)\right] \\
z(T) = 4\mu - 2,
\end{cases} \tag{2.28}
$$

which is strictly positive for all times. In particular $Z^N(s,\mu) > 0$. Similarly, for $t \in [s,T]$, from (2.27)

$$
\frac{d}{dt}W^N(t,\mu) \le N(1-\mu)Z^N\left(t,\mu+\frac{1}{N}\right)W^N(t,\mu) \le 2N(1-\mu)W^N(t,\mu),
$$

which implies that also $W^N(s,\mu) > 0$; by continuity in time, this contradicts the definition of s. Finally, observe that in the proof we fixed N even. The proof for N odd can be easily adapted with a bit of care, noting that $\mu = \frac{1}{2}$ cannot hold. □

2.2.4 Convergence of the value functions

We now consider the value function V^N, the unique solution to Equation (2.22), and study its limit as $N \to +\infty$. We show that its limit corresponds to the entropy solution of the Master Equation (2.12). More precisely, let U be the solution to (2.12) corresponding to the entropy solution Z of (2.13). Define, for $\mu \in [0,1]$

$$
U^*(t,\mu) := U\left(t,1,2\mu-1\right).
$$

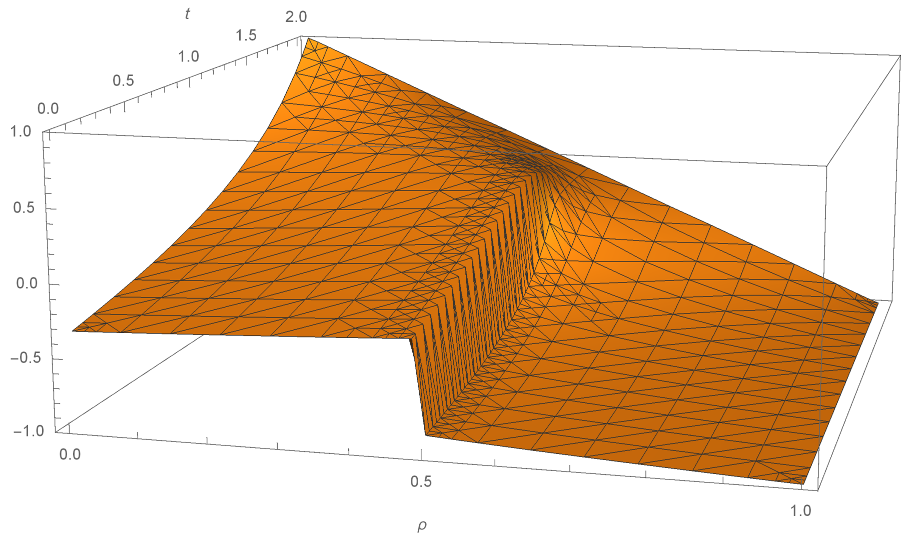


Figure 2.1: Simulation of the N-player dynamics (2.22). We plot the value function $V^N(t,\rho)$ for $N = 1000, T = 2$, $\rho \in \left\{0, \frac{1}{N}, \dots, 1\right\}$, the fraction of players in state 1.

Note that, for $T > \frac{1}{2}$, $U^*(t,\cdot)$ is discontinuous at $\mu = \frac{1}{2}$, but it is smooth elsewhere. In Figure 2.1 we indeed see the formation of a shock at the N-player level in the discontinuity point $\mu = \frac{1}{2}$, while there is smoothness elsewhere. The main result of this chapter establishes that V^N converges to U^* uniformly outside any neighborhood of $\mu = \frac{1}{2}$. In what follows, $S_N := \left\{0, \frac{1}{N}, \frac{2}{N}, \dots, 1\right\}$.

Theorem 2.7 (Convergence of value functions)**.** *For any $\varepsilon > 0$, $t \in [0,T]$ and $\mu \in S^N \setminus \left(\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon\right)$ we have*

$$|V^N(t,\mu) - U^*(t,\mu)| \le \frac{C_\varepsilon}{N}, \tag{2.29}$$

where C_ε does not depend on N nor on t, μ, but $\lim_{\varepsilon\to 0} C_\varepsilon = +\infty$.

The proof of Theorem 2.7 is based on the arguments developed in Chapter 1. We first slightly extend the above notation, letting, for $x \in \{-1, 1\}$

$$U^*(t,x,\mu) := U(t,x,2\mu - 1).$$

Moreover, let

$$v^{N,i}(t,\boldsymbol{x}) = V^N(t,x_i,\mu_{\boldsymbol{x}}^{N,i}), \qquad u^{N,i}(t,\boldsymbol{x}) = U^*(t,x_i,\mu_{\boldsymbol{x}}^{N,i})$$

for $i = 0,\dots,N$, where $\mu_{\boldsymbol{x}}^{N,i} = \frac{1}{N}\sum_{j=0,j\neq i}^N \delta_{\{x_i=1\}}$ is the fraction of the other players in 1. Let also $S_N^\varepsilon := S_N \setminus (\frac{1}{2} - \varepsilon, \frac{1}{2} + \varepsilon)$. The following results are the adaptations of Propositions 1.9 and 1.10 of Chapter 1. The first provides a bound for $\Delta^j u^{N,i}(t,\boldsymbol{x})$, while the second shows that U^* restricted to S_N^ε is "almost" a solution of (2.22).

Proposition 2.8. *For any $t \in [0,T]$, $\varepsilon > 0$ and any $\boldsymbol{x}$ such that $\mu_{\boldsymbol{x}}^{N,i} \in S_N^\varepsilon$, if $N \ge \frac{2}{\varepsilon}$, we have*

$$\Delta^j u^{N,i}(t,\boldsymbol{x}) = -\frac{1}{N}\frac{\partial}{\partial\mu}U(t,x_i,\mu_{\boldsymbol{x}}^{N,i}) + \tau^{N,i,j}(t,\boldsymbol{x}), \tag{2.30}$$

for any $j \neq i$, with $\left|\tau^{N,i,j}(t,\boldsymbol{x})\right| \le \frac{C_\varepsilon}{N^2}$. The constant C_ε is proportional to the Lipschitz constant of the master equation outside the discontinuity, which behaves like $\varepsilon^{-\frac{2}{3}}$.

Proposition 2.9. *For any $t \in [0,T]$, any $\varepsilon > 0$ and any μ such that either $\mu \in [\frac{1}{2}+\varepsilon, 1]$ or $\mu \in [0, \frac{1}{2}-\varepsilon]$, the function $U^*(t,\mu)$ satisfies*

$$
\begin{aligned}
&-\frac{d}{dt}U^*(t,\mu) + H(U^*(t,1-\mu) - U^*(t,\mu)) \qquad (2.31)\\
&\quad = N\mu\,[U^*(t,1-\mu) - U^*(t,\mu)]^- \left[U^*\left(t,\mu-\frac{1}{N}\right) - U^*(t,\mu)\right]\\
&\quad\quad + N(1-\mu)\left[U^*\left(t,\mu+\frac{1}{N}\right) - U^*\left(t,1-\mu-\frac{1}{N}\right)\right]^- \left[U^*\left(t,\mu+\frac{1}{N}\right) - U^*(t,\mu)\right]\\
&\quad\quad\quad + r^N(t,\mu), \qquad (2.32)
\end{aligned}
$$

with $\left|r^N(t,\mu)\right| \le \frac{C_\varepsilon}{N}$, where C_ε is as in Proposition 2.8.

We now use the information provided by Theorem 2.6. Set

$$
\Sigma_N^\varepsilon := \left\{ \boldsymbol{x} \in \{-1,1\}^{N+1} : \sum_{i=0}^{N} \delta_{x_i=1} \notin \left(\frac{N}{2} - N\varepsilon, \frac{N}{2} + N\varepsilon + 1\right) \right\}. \qquad (2.33)
$$

If $\boldsymbol{x} \in \Sigma_N^\varepsilon$, then $\mu_{\boldsymbol{x}}^{N,i} \in S_N^\varepsilon$ for all i. Denote by $\boldsymbol{Y}_s$ the state at time s of the $N+1$ players corresponding to the Nash equilibrium. By Theorem 2.6 it follows that, if $\boldsymbol{Y}_t \in \Sigma_N^\varepsilon$ for some $t < T$, then $\boldsymbol{Y}_s \in \Sigma_N^\varepsilon$ for all $s \in [t,T]$. Computing V^N (or U^*) in the optimal trajectories $\boldsymbol{Y}_s$ when starting the dynamics at time t in Σ_N^ε, we get

$$
v^{N,i}(s,\boldsymbol{Y}_s) = V^N(s, Y_i(s), \mu^{N,i}(s)) = \begin{cases} V^N(s,1,\mu^{N,i}(s)), & Y_i(s) = 1,\\ V^N(s,1,1-\mu^{N,i}(s)), & Y_i(s) = -1, \end{cases} \qquad (2.34)
$$

in light of the invariance property (2.21). Thus, the dynamics is such that it keeps being either on the right or on the left of the strip centered in the discontinuity. In particular we obtain

$$
v^{N,i}(s,\boldsymbol{Y}_s) \le \max_{\mu^N \in S_N^\varepsilon} V^N(s,\mu^N), \qquad (2.35)
$$

$$
|v^{N,i}(s,\boldsymbol{Y}_s) - u^{N,i}(s,\boldsymbol{Y}_s)| \le \max_{\mu^N \in S_N^\varepsilon} |V^N(s,\mu^N) - U^*(s,\mu^N)|, \qquad (2.36)
$$

for every $s \in [t,T]$, almost surely, and

$$
\max_{\boldsymbol{x} \in \Sigma_N^\varepsilon} |v^{N,i}(s,\boldsymbol{x}) - u^{N,i}(s,\boldsymbol{x})| = \max_{\mu^N \in S_N^\varepsilon} |V^N(s,\mu^N) - U^*(s,\mu^N)|. \qquad (2.37)
$$

Moreover, we note that

$$
\begin{aligned}
&|\Delta^i v^{N,i}(s,\boldsymbol{Y}_s) - \Delta^i v^{N,i}(s,\boldsymbol{Y}_s)|\\
&\qquad = |V^N(s,-Y_i(s),\mu_{\boldsymbol{Y}}^{N,i}(s)) - U(s,-Y_i(s),\mu_{\boldsymbol{Y}}^{N,i}(s))\\
&\qquad\qquad - V^N(s,Y_i(s),\mu_{\boldsymbol{Y}}^{N,i}(s)) + U(s,Y_i(s),\mu_{\boldsymbol{Y}}^{N,i}(s))|\\
&\qquad \le 2\max_{\mu^N \in S_N^\varepsilon} |V^N(s,\mu^N) - U(s,\mu^N)|.
\end{aligned} \qquad (2.38)
$$

Proof of Theorem 2.7. We choose a deterministic initial condition $\boldsymbol{Y}_t \in \Sigma_N^\varepsilon$, at time $t \in [0,T)$. As in the proof of Theorem 1.7 of Chapter 1, we exploit the characterization of the N-player dynamics in terms of SDEs driven by Poisson random measures, and we apply Itô's formula to the squared difference between the functions $u_t^{N,i}$ and $v_t^{N,i}$, both computed in the optimal trajectories $(\boldsymbol{Y}_s)_{s\in[t,T]}$. [1] Using equations (2.31) and (2.22), we then find

$$\mathbb{E}[(u_t^{N,i}-v_t^{N,i})^2]+\sum_{j=0}^N\mathbb{E}\left[\int_t^T\alpha^j(s,\boldsymbol{Y}_s)\Big(\Delta^j[u_s^{N,i}-v_s^{N,i}]\Big)^2ds\right] \tag{2.39}$$
$$=-2\mathbb{E}\Bigg[\int_t^T(u_s^{N,i}-v_s^{N,i})\bigg\{-r^N(s,\mu_{\boldsymbol{Y}}^{N,i}(s))+H(\Delta^iu_s^{N,i})-H(\Delta^iv_s^{N,i})$$
$$+\sum_{j=0,j\neq i}^N(\alpha^j-\overline{\alpha}^j)\Delta^ju^{N,i}+\alpha^i(\Delta^iu_s^{N,i}-\Delta^iv_s^{N,i})\bigg\}ds\Bigg],$$

where α^i is the Nash equilibrium played by player i, $\overline{\alpha}^i$ is the control induced by U and all the functions are evaluated on the optimal trajectories, e.g. $v_s^{N,i}:=v^{N,i}(s,\boldsymbol{Y}_s)$. We erase all the positive sum on the lhs and estimate the rhs using the Lipschitz properties of H, the bounds on r^N and Δ^ju^i given by Proposition 2.8, and the bound on α^j given by the fact that $Z^N(t,\mu)\le 2$, to get, for $N\ge\frac{2}{\varepsilon}$,

$$\mathbb{E}[(u_t^{N,i}-v_t^{N,i})^2]$$
$$\le\frac{C}{N}\mathbb{E}\left[\int_t^T|u_s^{N,i}-v_s^{N,i}|ds\right]+C\mathbb{E}\left[\int_t^T|u_s^{N,i}-v_s^{N,i}||\Delta^iu_s^{N,i}-\Delta^iv_s^{N,i}|ds\right]$$
$$+\frac{C}{N}\sum_{j=0,j\neq i}^N\mathbb{E}\left[\int_t^T|u_s^{N,i}-v_s^{N,i}||\Delta^ju_s^{N,j}-\Delta^jv_s^{N,j}|ds\right],$$

which can be further estimated via the convexity inequality $ab\le\frac{1}{2}a^2+\frac{1}{2}b^2$ yielding

$$\mathbb{E}[(u_t^{N,i}-v_t^{N,i})^2]\le\frac{C}{N^2}+C\mathbb{E}\left[\int_t^T\left|u_s^{N,i}-v_s^{N,i}\right|^2ds\right]+C\mathbb{E}\left[\int_t^T\left|\Delta^iu_s^{N,i}-\Delta^iv_s^{N,i}\right|^2ds\right]$$
$$+\frac{C}{N}\sum_{j=0}^N\mathbb{E}\left[\int_t^T|\Delta^ju_s^{N,j}-\Delta^jv_s^{N,j}|^2ds\right].$$

Here C denotes any constant which may depend on ε, and is allowed to change from line to line. Since all the functions are evaluated on the optimal trajectories, we apply (2.36) and (2.38) to obtain

$$|u^{N,i}(t,\boldsymbol{Y}_t)-v^{N,i}(t,\boldsymbol{Y}_t)|^2\le\frac{C}{N^2}+C\int_t^T\max_{\mu\in S_N^\varepsilon}|U(s,\mu)-V^N(s,\mu)|^2ds$$

for any deterministic initial condition $\boldsymbol{Y}_t\in\Sigma_N^\varepsilon$. Therefore (2.37) gives

$$\max_{\mu\in S_N^\varepsilon}|U(t,\mu)-V^N(t,\mu)|^2\le\frac{C}{N^2}+C\int_t^T\max_{\mu\in S_N^\varepsilon}|U(s,\mu)-V^N(s,\mu)|^2ds \tag{2.40}$$

and thus Gronwall's lemma applied to the quantity $\max_{\mu\in S_N^\varepsilon}|U(s,\mu)-V^N(s,\mu)|^2$ allows to conclude that

$$\max_{\mu\in S_N^\varepsilon}|U(t,\mu)-V^N(t,\mu)|^2\le\frac{C}{N^2}, \tag{2.41}$$

[1] We remark that in Chapter 1 the controls are assumed to be bounded below away from zero. Nevertheless, this fact is not used to derive the analogous identity to (2.39). A proof of the convergence results with no lower bound on the controls can be found in Section 3.1 of [23], if the master equation possesses a classical solution. Moreover, see Section 2.2.6 for a proof of Theorem 2.7 in the case of bounded below transition rates.

which immediately implies (2.29), but only if $N \geq \frac{2}{\varepsilon}$. Changing the value of $C = C_\varepsilon$, the thesis follows for any N. □

2.2.5 Propagation of chaos

The next result gives the propagation of chaos property for the optimal trajectories. Consider the initial datum (in $t = 0$) $\boldsymbol{\xi}$ i.i.d. with $P(\xi_i = 1) = \mu_0$ and $\mathbb{E}[\xi_i] = m_0 = 2\mu_0 - 1$, and denote by $\boldsymbol{Y}_t = (Y_0(t), Y_1(t), \ldots, Y_N(t))$ the optimal trajectories of the $N+1$-player game, i.e. when agents play the Nash equilibrium given by (2.23). Also, denote by $\widetilde{\boldsymbol{X}}_t$ the i.i.d. process in which players choose the local control $\widetilde{\alpha}(t, \pm 1) := [Z(t, m^*(t))]^{\mp}$, where Z is the entropy solution to (2.13) and m^* is the unique mean field game solution induced by Z, if $m_0 \neq 0$ ($\mu_0 \neq \frac{1}{2}$), that is the one which does not change sign (see Proposition 2.5). The propagation of chaos consists in proving the convergence of $\boldsymbol{Y}_t$ to the i.i.d. process $\widetilde{\boldsymbol{X}}_t$.

Theorem 2.10 (Propagation of chaos). *If $\mu_0 \neq \frac{1}{2}$ then, for any N and $i = 0, \ldots, N$,*

$$\mathbb{E}\left[\sup_{t \in [0,T]} |Y_i(t) - \widetilde{X}_i(t)|\right] \leq \frac{C_{\mu_0}}{\sqrt{N}}, \tag{2.42}$$

where C_{μ_0} does not depend on N, and $\lim_{\mu_0 \to \frac{1}{2}} C_{\mu_0} = \infty$.

Denote by $X_i(t)$ the dynamics of the i-th player when choosing the control

$$\bar{\alpha}^i(t, \boldsymbol{x}) = [\Delta^i U(t, x_i, \mu_{\boldsymbol{x}}^{N,i})]^- \tag{2.43}$$

induced by the master equation. We use $\boldsymbol{X}_t$ as an intermediate process for obtaining the propagation of chaos result. In fact, $\boldsymbol{X}_t$ can be treated as a mean field interacting system of particles (since the rate in (2.43) depends on N only through the empirical measure), for which propagation of chaos results are more standard. Next result shows the proximity of the optimal dynamics to the intermediate process just introduced.

Theorem 2.11. *If $\mu_0 \neq \frac{1}{2}$ then, for any N and $i = 0, \ldots, N$,*

$$\mathbb{E}\left[\sup_{t \in [0,T]} |Y_i(t) - X_i(t)|\right] \leq \frac{C_{\mu_0}}{N}, \tag{2.44}$$

where C_{μ_0} does not depend on N, and $\lim_{\mu_0 \to \frac{1}{2}} C_{\mu_0} = +\infty$.

Proof. Let $\mu_0 = \frac{1}{2} + 2\varepsilon$ and consider the event A where both $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$ belong to Σ_N^ε, for any time. Exploting the probabilistic representation of the dynamics in terms of Poisson random measures, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{s \in [0,t]} |X_i(s) - Y_i(s)|\right] \\
&\leq C\mathbb{E}\left[\int_0^t \left[\left|\alpha^*(X_{i,s}, \Delta^i u^{N,i}(s, \boldsymbol{X}_s)) - \alpha^*(Y_{i,s}, \Delta^i v^{N,i}(s, \boldsymbol{Y}_s))\right| + |X_{i,s} - Y_{i,s}|\right] ds\right] \\
&\leq C\mathbb{E}\left[\int_0^t \left[|X_i(s) - Y_i(s)| + |\Delta^i u^{N,i}(s, \boldsymbol{X}_s) - \Delta^i v^{N,i}(s, \boldsymbol{Y}_s)|\right] ds\right] \\
&\leq C\mathbb{E}\left[\int_0^t |X_i(s) - Y_i(s)| ds\right] + C\mathbb{E}\left[\mathbb{1}_A \int_0^t |\Delta^i u^{N,i}(s, \boldsymbol{Y}_s) - \Delta^i v^{N,i}(s, \boldsymbol{Y}_s)| ds\right] \\
&+ C\mathbb{E}\left[\mathbb{1}_A \int_0^t |\Delta^i u^{N,i}(s, \boldsymbol{X}_s) - \Delta^i u^{N,i}(s, \boldsymbol{Y}_s)| ds\right] + CP(A^c).
\end{aligned}$$

and now we apply (2.29) together with (2.38), the Lipschitz continuity of U in Σ_N^ε and the exchangeability of the processes to get, if $N \geq \frac{2}{\varepsilon}$,

$$\begin{aligned}
\mathbb{E}\left[\sup_{s\in[0,t]} |X_i(s) - Y_i(s)|\right] &\leq \frac{C}{N} + C\int_0^t \mathbb{E}|X_i(s) - Y_i(s)|ds + P(A^c) \\
&+ C\mathbb{E}\Bigg[\mathbb{1}_A \int_0^t \Big[|U(s, X_i(s), \mu_{\boldsymbol{X}}^{N,i}(s)) - U(s, X_i(s), \mu_{\boldsymbol{Y}}^{N,i}(s))| \\
&\qquad + |U(s, -X_i(s), \mu_{\boldsymbol{X}}^{N,i}(s)) - U(s, -X_i(s), \mu_{\boldsymbol{Y}}^{N,i}(s))|\Big]\, ds\Bigg] \\
&\leq \frac{C}{N} + C\int_0^t \mathbb{E}|X_i(s) - Y_i(s)|ds + P(A^c) + C\mathbb{E}\left[\mathbb{1}_A \int_0^t \frac{1}{N}\sum_{j\neq i} |X_j(s) - Y_j(s)|ds\right] \\
&\leq \frac{C}{N} + C\int_0^t \mathbb{E}|X_i(s) - Y_i(s)|ds + P(A^c). \qquad (2.45)
\end{aligned}$$

We can bound the probability of A^c by considering the process in which the transition rates are equal to 0, for any time, i.e. the constant process equal to the initial condition $\boldsymbol{\xi}$. Thanks to the shape of the Nash equilibrium, which prevents the dynamics from crossing the discontinuity, and of the control induced by the solution to the master equation, we have

$$P(A^c) = P(\exists t: \text{ either } \boldsymbol{X}_t \text{ or } \boldsymbol{Y}_t \notin \Sigma_N^\varepsilon) \leq 2P(\boldsymbol{\xi} \notin \Sigma_N^\varepsilon). \qquad (2.46)$$

For the latter, we have

$$\begin{aligned}
P(\boldsymbol{\xi} \notin \Sigma_N^\varepsilon) &= P\left(\sum_{i=0}^N \xi_i \in \left(\frac{N}{2} - N\varepsilon, \frac{N}{2} + N\varepsilon + 1\right)\right) \\
&\leq P\left(\sum_{i=0}^N \xi_i \leq \frac{N}{2} + N\varepsilon + 1\right) \leq P\left(\mu_{\boldsymbol{\xi}}^N \leq \frac{1}{2} + \varepsilon_N\right),
\end{aligned}$$

denoting

$$\varepsilon_N := \frac{\frac{N}{2} + N\varepsilon + 1}{N+1} - \frac{1}{2}. \qquad (2.47)$$

Observing that $(N+1)\mu_{\boldsymbol{\xi}}^N \sim \text{Bin}(N+1, \frac{1}{2} + 2\varepsilon)$ (recall $\mu_0 = \frac{1}{2} + 2\varepsilon$), we can further estimate, by standard Markov inequality,

$$\begin{aligned}
P(\boldsymbol{\xi} \notin \Sigma_N^\varepsilon) &\leq P\left(\left|\mu_{\boldsymbol{\xi}}^N - \frac{1}{2} - 2\varepsilon\right| \geq 2\varepsilon - \varepsilon_N\right) \leq \frac{\text{Var}\left[\mu_{\boldsymbol{\xi}}^N\right]}{(2\varepsilon - \varepsilon_N)^2} \\
&= \frac{1}{N+1} \frac{\left(\frac{1}{2} + 2\varepsilon\right)\left(\frac{1}{2} - 2\varepsilon\right)}{\left(2\varepsilon - \frac{N}{N+1}\left(\frac{1}{2} + \varepsilon\right) - \frac{1}{N+1} + \frac{1}{2}\right)^2} \leq \frac{C}{N\varepsilon} \qquad (2.48)
\end{aligned}$$

if $N \geq \frac{2}{\varepsilon}$, so that $2\varepsilon - \varepsilon_N \geq \frac{\varepsilon}{4}$.

Putting estimate (2.48) into (2.45), and denoting $\varphi(t) := \mathbb{E}\left[\sup_{s\in[0,t]} |X_i(s) - Y_i(s)|\right]$, we obtain

$$\varphi(t) \leq \frac{C}{N\varepsilon} + C\int_0^t \varphi(s)ds \qquad (2.49)$$

which, by Gronwall's lemma, gives (2.44), but only if $N \geq \frac{2}{\varepsilon}$. By changing the value of $C = C_\varepsilon$, the claim follows for any N. □

We are now in the position to prove Theorem 2.10. Thanks to (2.44), it is enough to show that

$$\mathbb{E}\left[\sup_{t\in[0,T]} |X_i(t) - \widetilde{X}_i(t)|\right] \leq \frac{C_{\mu_0}}{\sqrt{N}}, \tag{2.50}$$

Recall that the $\widetilde{X}_i$'s are i.i.d. and $\mathrm{Law}(\widetilde{X}_i(t)) = m^*(t)$; also, set $m = m^*$ and $\mu = \frac{m+1}{2}$. Moreover, we know that $(N+1)\mu^N_{\widetilde{\boldsymbol{X}}}(t) \sim \mathrm{Bin}(N+1, \mu(t))$. The rate of convergence follows from the estimate

$$\mathbb{E}\left|\mu^N_{\widetilde{\boldsymbol{X}}}(t) - \mu(t)\right| \leq \frac{C}{\sqrt{N}}, \tag{2.51}$$

for any time, by Cauchy-Schwarz inequality.

Proof of Theorem 2.10. Let $\mu_0 = \frac{1}{2} + 2\varepsilon$ and consider the event A where both $\boldsymbol{X}_t$ and $\widetilde{\boldsymbol{X}}_t$ belong to Σ^ε_N, for any time. Arguing as in the proof of Theorem 2.11, we obtain

$$\begin{aligned}
\mathbb{E}\left[\sup_{s\in[0,t]} |X_i(s) - \widetilde{X}_i(s)|\right] &\leq C\int_0^t \mathbb{E}|X_i(s) - \widetilde{X}_i(s)|ds + P(A^c) \\
&\quad + C\mathbb{E}\bigg[\mathbb{1}_A \int_0^t |U(s, X_i(s), \mu^{N,i}_{\boldsymbol{X}}(s)) - U(s, X_i(s), \mu^{N,i}_{\widetilde{\boldsymbol{X}}}(s))| \\
&\qquad + |U(s, -X_i(s), \mu^{N,i}_{\widetilde{\boldsymbol{X}}}(s)) - U(s, -X_i(s), \mu(s))| ds\bigg] \\
&\leq C\int_0^t \mathbb{E}|X_i(s) - \widetilde{X}_i(s)|ds + P(A^c) \\
&\quad + C\mathbb{E}\left[\mathbb{1}_A \int_0^t \frac{1}{N}\sum_{j\neq i} |X_j(s) - \widetilde{X}_j(s)|ds\right] + C\sup_{t\in[0,T]} \mathbb{E}\left|\mu^N_{\widetilde{\boldsymbol{X}}}(t) - \mu(t)\right| \\
&\leq \frac{C}{\sqrt{N}} + C\int_0^t \mathbb{E}|X_i(s) - \widetilde{X}_i(s)|ds + P(A^c).
\end{aligned}$$

We can bound the probability of A^c as before and thus Gronwall's Lemma allows to conclude. □

2.2.6 A modified example

This section was developed in a previous, unpublished version of the author's work [25], and can also be found in the PhD Dissertation [23]. We consider a modified framework allowing only for controls bounded from below (i.e. $\alpha(t,x) \geq \kappa > 0$). Most of the results are analogous to the previous setting, so we just sketch them, but the convergence proof is different: it involves a large deviation principle, and so it could be of interest itself. It still relies on a characterization of the Nash equilibrium as in Theorem 2.6, but unfortunately we have not managed to prove it. Moreover, it is interesting to note that the insertion of a lower bound on the transition rates can restore uniqueness of solutions to the mean field game (see Proposition 2.12 below).

In the modified setting, we leave the final cost unchanged and consider the Lagrangian

$$L_\kappa(a) = \frac{|a-\kappa|^2}{2},$$

so that the running cost is still zero if a player chooses the control equal to the minimum. The Hamiltonian of the problem is

$$H_\kappa(p) := \sup_{a\geq\kappa}\left\{-ap - \frac{(a-\kappa)^2}{2}\right\} = -\kappa p + \frac{(p^-)^2}{2}, \tag{2.52}$$

whose argmax is given by $a^*_\kappa(p) := \kappa + p^-$. The mean field game system becomes

$$\begin{cases} \dot z = z\left(\frac{|z|}{2} + 2\kappa\right) \\ \dot m = -m(|z|+2\kappa) + z \\ z(T) = 2m(T) \\ m(0) = m_0. \end{cases} \tag{2.53}$$

In order to solve System (2.53), we again suppose $m(T) = M$ is given so that we can find $z(t)$. As one can check via computation,

$$z(t) := \frac{4\kappa M}{(2\kappa + |M|)e^{(T-t)2\kappa} - |M|}. \tag{2.54}$$

Substituting this expression in the KFP equation, we find

$$m(t) = \frac{e^{2\kappa t}(|M| - e^{2\kappa(T-t)}(2\kappa+|M|))^2\left(m_0 + \frac{(-1+e^{2\kappa t})M(2e^{2\kappa T}(1+e^{2\kappa t})\kappa + (-2e^{2\kappa t}+e^{2\kappa T}+e^{2\kappa(t+T)})|M|}{(e^{2\kappa t}|M| - e^{2\kappa T}(2\kappa+|M|))^2}\right)}{(|M| - e^{2\kappa T}(2\kappa+|M|))^2}. \tag{2.55}$$

By imposing the mean field condition $m(T) = M$ we can characterize the MFG solutions via the solutions in M to

$$-M + \frac{4e^{2\kappa T}\kappa^2\left[m_0 + \frac{(-1+e^{2\kappa T})M(2e^{2\kappa T}(1+e^{2\kappa T})\kappa + (-e^{2\kappa T}+e^{4\kappa T})|M|)}{(e^{2\kappa T}|M| - e^{2\kappa T}(2\kappa+|M|))^2}\right]}{(|M| - e^{2\kappa T}(2\kappa+|M|))^2} = 0. \tag{2.56}$$

Note that this is a generalization of the case $\kappa = 0$: indeed, for $\kappa \to 0$ we recover the previous mean field condition, given by (2.11). The above equation can be rewritten as

$$\begin{aligned} M^3(e^{2\kappa T}-1)^2 - M|M|(e^{2\kappa T}-1)[(1-4\kappa)e^{2\kappa T}-1] \\ + 2\kappa M[e^{4\kappa T}(2\kappa-1)+1] - 4e^{2\kappa T}\kappa^2 m_0 = 0. \end{aligned} \tag{2.57}$$

We can now state the analogous of Proposition 2.1.

Proposition 2.12. *If $\kappa \geq \frac{1}{2}$ the MFG system* (2.53) *has a unique solution for any T and m_0; if $\kappa < \frac{1}{2}$ and $T \leq T_\kappa := \frac{-\log(1-2\kappa)}{4\kappa}$ the MFG system* (2.53) *has a unique solution for any m_0.*

Moreover, let $T_\kappa(m_0)$ *be the unique solution in* $T \in [T_\kappa, +\infty[$ *to*

$$
\begin{aligned}
|m_0| = \\
&\frac{1}{4e^{2\kappa T}\kappa^2}\left\{\frac{1}{3(-1+e^{2\kappa T})^2}2\kappa\left(1+e^{4\kappa T}(-1+2\kappa)\right)\left[-1+e^{2\kappa T}(2-4\kappa)+e^{4\kappa T}(-1+4\kappa)\right.\right.\\
&\left.-\sqrt{(-1+e^{2\kappa T})^2(1-6\kappa+e^{2\kappa T}(-2+8\kappa)+e^{4\kappa T}(1-2\kappa+4\kappa^2)}\right]\\
&+\frac{1}{9(-1+e^{2\kappa T})^3}(1+e^{2\kappa T}(-1+4\kappa))\left[1+e^{4\kappa T}(1-4\kappa)+e^{2\kappa T}(-2+4\kappa)\right.\\
&\left.+\sqrt{(-1+e^{2\kappa T})^2(1-6\kappa+e^{2\kappa T}(-2+8\kappa)+e^{4\kappa T}(1-2\kappa+4\kappa^2)}\right]^2\\
&+\frac{1}{27(-1+e^{2\kappa T})^4}\left[1+e^{4\kappa T}(1-4\kappa)+e^{2\kappa T}(-2+4\kappa)\right.\\
&\left.\left.+\sqrt{(-1+e^{2\kappa T})^2(1-6\kappa+e^{2\kappa T}(-2+8\kappa)+e^{4\kappa T}(1-2\kappa+4\kappa^2))}\right]^3\right\}.
\end{aligned}
\tag{2.58}
$$

Then, for any $m_0 \in [-1, 1]$*, the MFG system* (2.53) *possesses*

(i) a unique solution for $T < T_\kappa(m_0)$*;*

(ii) two solutions if $T = T_\kappa(m_0)$*;*

(iii) three distinct solutions for $T > T_\kappa(m_0)$*.*

Note that $\lim_{\kappa\downarrow 0} T_\kappa = \frac{1}{2}$, as in Proposition 2.1, and $\lim_{\kappa\uparrow\frac{1}{2}} T_\kappa = +\infty$. In fact, the insertion of a lower bound κ increases the time for which there is uniqueness of solutions. Moreover the three distinct solutions, when they exist, possess the same properties as for $\kappa = 0$. Namely, if $m_0 \neq 0$ there is a unique solution, denoted by (z^*_κ, m^*_κ), which does not change sign, and is the one that exists for any T. If $m_0 = 0$ instead, the three solutions are: the constant 0, the one always positive and the one always negative, if $T > T_\kappa$.

The master equation and the Nash system have the same shape as in (2.12) and (2.22), where the Hamiltonian is replaced by H_κ and p^- by a^*_κ. The master equation can still be written as a scalar conservation law, whose entropy solution, denoted by $Z^*_\kappa(t,m)$, has the same properties as before: it has a shock at $m = 0$, for $t > T_k$, and is smooth elsewhere. If we show that the solution to the Nash system enjoys the same properties, then we are able to prove the convergence of the value functions as well as a propagation of chaos for $m_0 \neq 0$.

From now on, we thus fix $0 < \kappa < 1/2$ and $T > T_\kappa$. Denote $V^N_\kappa(t,\mu) = V^N_\kappa(t,1,m)$ and $Z^N_\kappa(t,\mu) = V^N_\kappa(t,1-\mu) - V^N_\kappa(t,\mu)$, so that the Nash equilibrium is given by

$$\alpha^N_\kappa(t,\pm 1,\mu) = \kappa + Z^N_\kappa(t,\mu)^{\mp}.$$

Let $U_\kappa(t,x,m)$ be the solution to the master equation corresponding to the the entropy solution $Z^*_\kappa(t,m)$ and define $U^*_\kappa(t,\mu) = U_\kappa(t,1,2\mu-1)$. Let also $\boldsymbol{Y}^\kappa$, $\boldsymbol{X}^\kappa$ and $\tilde{\boldsymbol{X}}^\kappa$ be the analogue of the processes defined in Section 2.2.5.

Theorem 2.13. *Fix* $N \geq 1$ *and* $0 < \kappa < \frac{1}{2}$*. Assume that for any* $\mu \in S_N = \left\{0, \frac{1}{N}, \dots, 1\right\}$

$$Z^N_\kappa(t,\mu) \geq 0 \qquad \text{if } \mu \geq \frac{1}{2}, \tag{2.59}$$

$$Z^N_\kappa(t,\mu) \leq 0 \qquad \text{if } \mu \leq \frac{1}{2}. \tag{2.60}$$

Then, for any $t \in [0,T]$, $\varepsilon > 0$ and $\mu \in S^N \setminus \left]\frac{1}{2}-\varepsilon, \frac{1}{2}+\varepsilon\right[$, we have

$$|V_\kappa^N(t,\mu) - U_\kappa^*(t,\mu)| \leq \frac{C_{\varepsilon,\kappa}}{N}, \tag{2.61}$$

where $C_{\varepsilon,\kappa}$ does not depend on N nor on t, μ. Moreover, if $\mu_0 \neq \frac{1}{2}$

$$\mathbb{E}\left[\sup_{t\in[0,T]} |Y_i^\kappa(t) - \widetilde{X}_i^\kappa(t)|\right] \leq \frac{C_{\mu_0,\kappa}}{\sqrt{N}}, \tag{2.62}$$

$$\mathbb{E}\left[\sup_{t\in[0,T]} |Y_i^\kappa(t) - X_i^\kappa(t)|\right] \leq \frac{C_{\mu_0,\kappa}}{N}. \tag{2.63}$$

Proof. We start by proving (2.61), omitting for simplicity the κ from the notation. Let $\varepsilon > 0$ be fixed and consider a deterministic initial condition $\boldsymbol{\xi} = \boldsymbol{Y}_t$ at time t such that $\mu_{\boldsymbol{\xi}}^N \in \Sigma_N^\varepsilon$, where Σ_N^ε is defined by (2.33). Let $\bar{\varepsilon} = \bar{\varepsilon}(T,\kappa,\varepsilon) := \frac{\varepsilon}{2}e^{-2\kappa T}$, fix $N \geq \frac{2}{\varepsilon}$ and consider the set

$$A_\varepsilon := \left\{\boldsymbol{Y}_s \in \Sigma_N^{\bar{\varepsilon}} \quad \forall s \in [t,T]\right\}.$$

We first bound the probability of A_ε^C. For the purpose, consider the process $\widetilde{\boldsymbol{Y}}$ in which the transition rates of each $\widetilde{Y}_i$ are all constant and equal to the minimum κ, with the same initial condition $\boldsymbol{Y}_t$. Thanks to the properties of the Nash equilibrium (2.59) and (2.60), we have $P(A_\varepsilon^C) \leq P(\widetilde{A}_\varepsilon^C)$, where $\widetilde{A}_\varepsilon$ is the set where $\widetilde{\boldsymbol{Y}}_s \in \Sigma_N^{\bar{\varepsilon}}$ for any $s \in [t,T]$. The fraction of particles in state 1 of this process, denoted by $\left(\widetilde{\mu}^N(s)\right)_{s\in[t,T]}$, has a non-zero probability of crossing the discontinuity, due to $\kappa > 0$, thus we cannot argue as for $\kappa = 0$.

We are allowed to consider a sequence of deterministic initial conditions such that

$$\lim_{N\to\infty} \mu_{\boldsymbol{\xi}}^N =: \mu_t^* \in [0,1] \setminus]1/2-\varepsilon, 1/2+\varepsilon[=: S^\varepsilon; \tag{2.64}$$

in particular the limit exists. We have that the $\widetilde{Y}_i$'s are independent processes (even if not identically distributed), and the sequence of processes $\left(\widetilde{\mu}^N(s)\right)_{s\in[t,T]}$ satisfies a sample path large deviation principle on $D([0,T];[0,1])$, thanks to a version of Sanov's Theorem; see e.g. [45] and [52]. We actually need only the upper bound:

$$\limsup_{N\to\infty} \frac{1}{N} \log P(\widetilde{A}_\varepsilon^C) \leq -\mathcal{I}_{T,\kappa,\varepsilon}, \tag{2.65}$$

where $\mathcal{I}$ is a good rate functional, $\mathcal{I}_{T,\kappa,\varepsilon} := \inf_{\lambda \in \overline{B_{T,\kappa,\varepsilon}}} \mathcal{I}(\lambda)$, with

$$B_{T,\kappa,\varepsilon} := \left\{\lambda \in D([0,T];[0,1]) : \lambda(s) \notin \left]\frac{1}{2}-\bar{\varepsilon}, \frac{1}{2}-\bar{\varepsilon}\right[\quad \forall s \in [t,T]\right\}.$$

Thanks to (2.64), the sequence of processes $\left(\widetilde{\mu}^N(s)\right)_{s\in[t,T]}$ satisfies a propagation of chaos property with the limit given by $\mu^*(s) = \frac{1}{2} + \left(\mu_t^* - \frac{1}{2}\right)e^{-2\kappa(s-t)}$ for $t \leq s \leq T$: it is provided by the solution to the KFP equation when $z = 0$. It is well known that the rate functional is always positive and, if the propagation of chaos holds, $\mathcal{I}(\lambda) = 0$ if and only if $\lambda = \mu^*$. Therefore we can conclude that $\mathcal{I}_{T,\kappa,\varepsilon} > 0$, because of the choice of $\bar{\varepsilon}$: indeed, $|\mu^*(s) - 1/2| \geq 2\bar{\varepsilon}$ for all $s \in [t,T]$ and for any choice of t and $\mu_t^* \in S^\varepsilon$. Thus, μ^* does not belong to the closure of $B_{T,\kappa,\varepsilon}$. This implies that

$$P(A_\varepsilon^C) \sim e^{-N\mathcal{I}_{T,\kappa,\varepsilon}}. \tag{2.66}$$

Moreover, the solution U^* to the master equation is smooth outside $[1/2-\overline{\varepsilon}, 1/2+\overline{\varepsilon}]$ and so the conclusions of Proposition 2.8 follow in the same way for $N \geq 2/\overline{\varepsilon}$. With the same steps as in the proof of Theorem 2.7, we obtain Equation (2.39):

$$\begin{aligned}
&\mathbb{E}[(u_t^{N,i}-v_t^{N,i})^2] + \sum_{j=0}^N \mathbb{E}\left[\int_t^T \alpha^j(s,\mathbf{Y}_s)\Big(\Delta^j[u_s^{N,i}-v_s^{N,i}]\Big)^2 ds\right] \\
&= -2\mathbb{E}\left[\int_t^T (u_s^{N,i}-v_s^{N,i})\Big\{-r^N(s,\mathbf{Y}_s) + H(\Delta^i u_s^{N,i}) - H(\Delta^i v_s^{N,i}) \\
&+ \sum_{j=0,j\neq i}^N (\alpha_s^j - \overline{\alpha}_s^j)\Delta^j u_s^{N,i} + \alpha_s^i(\Delta^i u_s^{N,i} - \Delta^i v_s^{N,i})\Big\} ds\right]
\end{aligned}$$

with the same notation. This time, we split the expectation in $\mathbb{E}[\mathbb{1}_{A_\varepsilon}\ldots] + \mathbb{E}[\mathbb{1}_{A_\varepsilon^C}\ldots]$. The second term is bounded by

$$\mathbb{E}[\mathbb{1}_{A_\varepsilon^C}\ldots] \leq CNP(A_\varepsilon^C) \sim CNe^{-CN} \leq \frac{C}{N^2}$$

for $N \geq N_\varepsilon$ large enough. For the first term instead, we note that under the event A_ε we can use Lipschitz properties of H_κ and a_κ^* and the bounds on $r^{N,i}$ and $\Delta^j u^{N,i}$. On the left hand side we also erase the positive sum $\sum_{j\neq i}$ and estimate $\alpha^i \geq \kappa$. After the above procedures we end up with

$$\begin{aligned}
&\mathbb{E}[(u_t^{N,i}-v_t^{N,i})^2] + \kappa\mathbb{E}\left[\mathbb{1}_{A_\varepsilon}\int_t^T \left|\Delta^i u_s^{N,i} - \Delta^i v_s^{N,i}\right|^2 ds\right] \leq \\
&\leq \frac{C}{N}\mathbb{E}\left[\mathbb{1}_{A_\varepsilon}\int_t^T |u_s^{N,i}-v_s^{N,i}| ds\right] + C\mathbb{E}\left[\mathbb{1}_{A_\varepsilon}\int_t^T |u_s^{N,i}-v_s^{N,i}||\Delta^i u_s^{N,i} - \Delta^i v_s^{N,i}| ds\right] \\
&+ \frac{C}{N+1}\sum_{j=0,j\neq i}^N \mathbb{E}\left[\mathbb{1}_{A_\varepsilon}\int_t^T |u_s^{N,i}-v_s^{N,i}||\Delta^j u_s^{N,j} - \Delta^j v_s^{N,j}| ds\right] + CNP(A_\varepsilon^C).
\end{aligned}$$

The right hand side can be further bounded using the inequality $ab \leq \delta a^2 + \frac{b^2}{4\delta}$, so that we can write

$$\begin{aligned}
&\mathbb{E}[(u_t^{N,i}-v_t^{N,i})^2] + \kappa\mathbb{E}\left[\mathbb{1}_{A_\varepsilon}\int_t^T \left|\Delta^i u_s^{N,i} - \Delta^i v_s^{N,i}\right|^2 ds\right] \\
&\leq \frac{C}{N^2} + C\mathbb{E}\left[\mathbb{1}_{A_\varepsilon}\int_t^T \left|u_s^{N,i}-v_s^{N,i}\right|^2 ds\right] \\
&\qquad + \frac{\kappa}{2(N+1)}\sum_{j=0}^N \mathbb{E}\left[\mathbb{1}_{A_\varepsilon}\int_t^T |\Delta^j u_s^{N,j} - \Delta^j v_s^{N,j}|^2 ds\right] \\
&\leq \frac{C}{N^2} + C\mathbb{E}\left[\int_t^T \left|u_s^{N,i}-v_s^{N,i}\right|^2 ds\right] + \frac{\kappa}{2(N+1)}\sum_{j=0}^N \mathbb{E}\left[\mathbb{1}_{A_\varepsilon}\int_t^T |\Delta^j u_s^{N,j} - \Delta^j v_s^{N,j}|^2 ds\right].
\end{aligned} \tag{2.67}$$

Taking the averages $\frac{1}{N+1}\sum_{i=0}^N$ of the above, we obtain

$$\frac{1}{N+1}\sum_{i=0}^N \mathbb{E}[(u_t^{N,i}-v_t^{N,i})^2] + \frac{\kappa}{2(N+1)}\sum_{j=0}^N \mathbb{E}\left[\mathbb{1}_{A_\varepsilon}\int_t^T |\Delta^j u_s^{N,j} - \Delta^j v_s^{N,j}|^2 ds\right]$$

$$\leq \frac{C}{N^2} + C\int_t^T \frac{1}{N+1}\sum_{i=0}^N \mathbb{E}\left[|u_s^{N,i} - v_s^{N,i}|^2\right] ds$$

and thus Gronwall's Lemma, applied to the quantity $\frac{1}{N+1}\sum_{i=0}^N \mathbb{E}\left[|u_s^{N,i} - v_s^{N,i}|^2\right]$ yields, erasing the second (positive) term of the lhs,

$$\sup_{t\leq s\leq T}\left\{\frac{1}{N+1}\sum_{i=0}^N \mathbb{E}\left[\left|u^{N,i}(s,\boldsymbol{Y}_s) - v^{N,i}(s,\boldsymbol{Y}_s)\right|^2\right]\right\} \leq \frac{C}{N^2},$$

which also implies

$$\frac{\kappa}{2(N+1)}\sum_{j=0}^N \mathbb{E}\left[\mathbb{1}_{A_\varepsilon}\int_t^T |\Delta^j u_s^{N,j} - \Delta^j v_s^{N,j}|^2 ds\right] \leq \frac{C}{N^2}. \tag{2.68}$$

Applying (2.68) to the rhs of (2.67) and using Gronwall's Lemma again, we get

$$|u^{N,i}(t,\boldsymbol{\xi}) - v^{N,i}(t,\boldsymbol{\xi})|^2 \leq \frac{C}{N^2} \tag{2.69}$$

for any deterministic $\boldsymbol{\xi} \in \Sigma_N^\varepsilon$, which immediately gives (2.61), in light of (2.37).

To prove (2.62), we first observe that (2.68) can be derived in the same way for more general non-deterministic initial conditions. Indeed, assuming now that the initial time is 0 and the initial condition $\boldsymbol{\xi}$ is i.i.d. with $P(\xi_i = 1) = \frac{1}{2} + 2\varepsilon$, the same argument we used above yields $P(A_\varepsilon^C) \leq CN^{-2}$ and thus, by summing on both sides of (2.68) the same quantity appearing on the lhs, but with A_ε replaced by A_ε^C, and then using the exchangeability of the process $\boldsymbol{Y}$, we deduce

$$\mathbb{E}\left[\int_0^T \left|\Delta^i v^{N,i}(s,\boldsymbol{Y}_s) - \Delta^i u^{N,i}(s,\boldsymbol{Y}_s)\right| ds\right] \leq \frac{C}{N}. \tag{2.70}$$

Consider now the set E_ε where both $\boldsymbol{X}_t$ and $\boldsymbol{Y}_t$ belong to $\Sigma_N^{\bar\varepsilon}$, for any time. We can bound

$$\begin{aligned} P(E_\varepsilon^C) &= P\left(\exists t:\ \text{either } \mu_{\boldsymbol{X}}^N(t) \text{ or } \mu_{\boldsymbol{Y}}^N(t) \notin \Sigma_N^{\bar\varepsilon}\right) \leq 2P\left(\exists t: \mu_{\tilde{\boldsymbol{Y}}}^N(t) \notin \Sigma_N^{\bar\varepsilon}\right) \\ &\leq 2P(\tilde{A}_\varepsilon^C) \leq \frac{C}{N}. \end{aligned}$$

Proceeding as in the proof of (2.44), applying (2.70), the Lipschitz continuity of U^* in E_ε and the exchangeability of the processes, we find

$$\begin{aligned} &\mathbb{E}\left[\sup_{s\in[0,t]}|X_i(s) - Y_i(s)|\right] \leq C\mathbb{E}\left[\int_0^t |X_i(s) - Y_i(s)| + |\Delta^i u^{N,i}(s,\boldsymbol{X}_s) - \Delta^i v^{N,i}(s,\boldsymbol{Y}_s)| ds\right] \\ &\leq C\mathbb{E}\left[\int_0^t |X_i(s) - Y_i(s)| ds\right] + C\mathbb{E}\left[\mathbb{1}_{E_\varepsilon}\int_0^t |\Delta^i u^{N,i}(s,\boldsymbol{Y}_s) - \Delta^i v^{N,i}(s,\boldsymbol{Y}_s)| ds\right] \\ &+ C\mathbb{E}\left[\mathbb{1}_{E_\varepsilon}\int_0^t |\Delta^i u^{N,i}(s,\boldsymbol{X}_s) - \Delta^i u^{N,i}(s,\boldsymbol{Y}_s)| ds\right] + CP(E_\varepsilon^C) \\ &\leq C\int_0^t \mathbb{E}|X_i(s) - Y_i(s)| ds + \frac{C}{N} + C\mathbb{E}\left[\mathbb{1}_{E_\varepsilon}\int_0^t \frac{1}{N}\sum_{j\neq i}|X_j(s) - Y_j(s)| ds\right] + \frac{C}{N} \\ &\leq \frac{C}{N} + C\int_0^t \mathbb{E}\left[\sup_{0\leq r\leq s}|X_i(r) - Y_i(r)|\right] ds \end{aligned}$$

and thus Gronwall's inequality gives (2.62).

Finally, (2.63) derives from (2.62) as in Theorem 2.10. Actually, we obtained the claims only for N large enough, but by changing the value of $C = C_\varepsilon$ the thesis follows for any N. □

2.2.7 Potential mean field game

We give here another characterization of the solutions to the MFG system (2.7). For a more detailed introduction on potential mean field games in the finite state space see [23, Sect. 1.4.1]. We show that System (2.7) can be viewed as the necessary conditions for optimality, given by the Pontryagin maximum principle, of a *deterministic* optimal control problem in $\mathbb{R}^2$. We show that the N-player game, in the limit as $N \to +\infty$ selects exactly the global minimizer of this problem when it is unique, i.e. when $m_0 \neq 0$.

The notation is slightly different in this section. Consider the controlled dynamics, representing the KFP equation,

$$\begin{cases} \dot{m}_1 = m_{-1}\alpha_{-1} - m_1\alpha_1 \\ \dot{m}_{-1} = m_1\alpha_1 - m_{-1}\alpha_{-1} \\ m(0) = m_0. \end{cases} \tag{2.71}$$

The state variable is $m(t) = (m_1(t), m_{-1}(t))$. Note that, in the previous notation, we had $m_1 = \mu$ and $m = m_1 - m_{-1}$. Here the control is $\alpha(t) = (\alpha_1(t), \alpha_{-1}(t))$, deterministic and open-loop, taking values in

$$A = \{(a_1, a_{-1}) : a_1, a_{-1} \geq 0\}.$$

Clearly, if $m_0 = (m_{0,1}, m_{0,-1})$ belongs to the simplex

$$P(\{1,-1\}) := \{(m_1, m_{-1}) : m_1 + m_{-1} = 1, m_1, m_{-1} \geq 0\},$$

then, for any choice of the control α, the dynamics remains in $P(\{1,-1\})$ for any time.

The cost to be minimized is

$$\mathcal{J}(\alpha) = \int_0^T \left(m_1(t)\frac{\alpha_1(t)^2}{2} + m_{-1}(t)\frac{\alpha_{-1}(t)^2}{2} \right) dt + \mathcal{G}(m(T)), \tag{2.72}$$

where $\mathcal{G}(m_1, m_{-1}) := -\frac{(m_1 - m_{-1})^2}{2}$ is such that

$$\frac{\partial}{\partial m_1}\mathcal{G}(m) = -(m_1 - m_{-1}) =: G(1, m)$$
$$\frac{\partial}{\partial m_{-1}}\mathcal{G}(m) = m_1 - m_{-1} =: G(-1, m),$$

whereas $G(x, m) = -x(m_1 - m_{-1})$, for $x = \pm 1$, is the terminal cost. This structure is called *potential* mean field game, since we have $\nabla \mathcal{G}(m) = G(\cdot, m)$.

The Hamiltonian of this problem is

$$\begin{aligned} \mathcal{H}(m, u) &= \sup_{a \in A} \left\{ -b(m, a) \cdot u - m_1 \frac{a_1^2}{2} - m_{-1}\frac{a_{-1}^2}{2} \right\} \\ &= m_1 \frac{[(u_{-1} - u_1)^-]^2}{2} + m_{-1}\frac{[(u_1 - u_{-1})^-]^2}{2}, \end{aligned}$$

where $b_x(m, a) = m_{-x}a_{-x} - m_x a_x$, for $x = \pm 1$, is the vector field in (2.71), and the argmax of the Hamiltonian is

$$a_1^*(u) = (u_{-1} - u_1)^-,$$
$$a_{-1}^*(u) = (u_1 - u_{-1})^-.$$

Thus, the HJB equation of the control problem reads

$$\begin{cases} -\frac{\partial \mathcal{U}}{\partial t} + \mathcal{H}(m, \nabla_m \mathcal{U}) = 0 & t \in [0,T), m \in \mathcal{P}(\{1,-1\}) \\ \mathcal{U}(T,m) = \mathcal{G}(m), \end{cases} \tag{2.73}$$

and its characteristics curves are given by the MFG system

$$\begin{cases} -\dot{u}_1 + \frac{[(u_{-1}-u_1)^-]^2}{2} = 0 \\ -\dot{u}_{-1} + \frac{[(u_1-u_{-1})^-]^2}{2} = 0 \\ \dot{m}_1 = m_{-1}a^*_{-1}(u) - m_1 a^*_1(u) \\ \dot{m}_{-1} = m_1 a^*_1(u) - m_{-1}a^*_{-1}(u) \\ u_{\pm 1}(T) = G(\pm 1, m(T)), \quad m(0) = m_0. \end{cases} \tag{2.74}$$

Lemma 2.14. *The following claims hold:*

1. *There exists an optimum of the control problem* (2.71)-(2.72).
2. *The MFG system* (2.74) *represents the necessary conditions for optimality, given by the Pontryagin maximum principle.*

Proof. The first claim follows from Theorem 5.2.1 p. 94 in [13], which can be applied since the dynamics is linear in α and the running cost is convex in α. Conclusion *2* is standard. □

We know that, if T is large enough, there are three solutions to the MFG system. The control problem (2.71)-(2.72) has a minimum, so we wonder which of these solutions is indeed a minimizer.

First, we need to investigate some property of the roots of (2.11). Let $T > T(m_0)$ be fixed. Let $M_1(m_0) < M_2(m_0) < M_3(m_0)$ be the three solutions to (2.11). If $m_0 = 0$ denote $M_- = M_1(0) < 0$, $M_+ = M_3(0) > 0$; we have $M_2(0) = 0$ and $M_+ = M_-$. If $m_0 > 0$ then, by Proposition 2.1, $M_3(m_0) > 0$ and $M_1(m_0), M_2(m_0) < 0$; if $m_0 < 0$ then $M_3(m_0) < 0$ and $M_1(m_0), M_2(m_0) > 0$.

Lemma 2.15. *Let $m_0 > 0$ and $T > T(m_0)$ be fixed. Then*

1. *The function $[0, m_0] \ni m \mapsto M_3(m) \in [0,1]$ is increasing, $M_2(m)$ is decreasing and $M_1(m)$ is increasing. In particular for any $m \in [0, m_0]$*

$$M_3(m) > M_+ = |M_-| > |M_1(m)| > |M_2(m)| > M_2(0) = 0 \tag{2.75}$$

2. *We have $M_1(m) < -\frac{2T-1}{3T} < M_2(m) < 0$ and for any $m \in [0, m_0]$*

$$\left| M_2(m) + \frac{2T-1}{3T} \right| > \left| M_1(m) + \frac{2T-1}{3T} \right|. \tag{2.76}$$

The case $m_0 < 0$ is symmetric.

Proof. Claim (1) derives from the proof of Proposition 2.1. For claim (2), $M_1(m)$ and $M_2(m)$ are the two negative roots of $f(M) = T^2M^3 - T(2-T)M^2 + (1-2T)M - m = 0$. The roots of $f'(M)$ are $q := -\frac{2T-1}{3T}$ and $\frac{1}{T}$. Hence $M_1 < q < M_2 < 0$, $f(q) > 0$ and we have, by Taylor's formula (which here is actually a change of variable),

$$f(q+\varepsilon) = f(q) + f'(q)\varepsilon + \frac{f''(q)}{2}\varepsilon^2 + \frac{f'''(q)}{6}\varepsilon^3 = f(q) + \frac{f''(q)}{2}\varepsilon^2 + T^2\varepsilon^3$$

$$f(q-\varepsilon) = f(q) - f'(q)\varepsilon + \frac{f''(q)}{2}\varepsilon^2 - \frac{f'''(q)}{6}\varepsilon^3 = f(q) + \frac{f''(q)}{2}\varepsilon^2 - T^2\varepsilon^3$$

for any $\varepsilon > 0$. Thus $f(q+\varepsilon) - f(q-\varepsilon) = 2T^2\varepsilon^3 > 0$ for any $\varepsilon > 0$, which implies (2.76). □

For $i = 1, 2, 3$, denote by $m_i, z_i, \alpha_i, m_i, u_i$ the solution to the MFG system corresponding to M_i.

Theorem 2.16. *Let $m_0 > 0$ and $T > T(m_0)$ be fixed. Then for any $m \in [0, m_0]$ and $i = 1, 2, 3$ we have $\mathcal{J}(\alpha_i) = \varphi(M_i(m))$, where $\varphi : [-1, 1] \to [-1, 1]$,*

$$\varphi(M) := M^2\left(T - \frac{1}{2} - T|M|\right). \tag{2.77}$$

Moreover, for any $m \in \ (0, m_0]$,

$$\varphi(M_+) = \varphi(M_-) < \varphi(0) = 0, \tag{2.78}$$
$$\varphi(M_3(m)) < \varphi(M_+) < \varphi(M_1(m)), \tag{2.79}$$
$$\varphi(M_1(m)) < \varphi(M_2(m)) > 0, \tag{2.80}$$

meaning that α_+ and α_- are both optimal if $m = 0$ and $\alpha \equiv 0$ is not, while α_3 is the unique minimizer if $m > 0$, with

$$\mathcal{J}(\alpha_3) < \mathcal{J}(\alpha_1) < \mathcal{J}(\alpha_2). \tag{2.81}$$

Proof. The first claim and (2.77) follow directly from (2.72) and (2.10).

We continue by proving (2.79). The roots of φ' are 0 and $\pm q$, with $q := -\frac{2T-1}{3T}$. The function φ is then increasing if either $M < q$ or $0 < M < -q$. Thus (2.79) follows from (2.75) and the fact that $\varphi(M_+) = \varphi(M_-)$, as $\varphi(M)$ only depends on $|M|$.

Next, we show that $\varphi(M_+) < 0 = \varphi(0)$. Since M_+ solves $T^2M^2 + T(2-T)M + 1 - 2T = 0$, we obtain, for $M = M_+$,

$$\varphi(M) = \frac{M^2}{2}(2T - 1 - 2TM) = \frac{M^2}{2}(T^2M^2 - T^2M) = \frac{T^2M^3}{2}(M-1) < 0$$

because $M_+ < 1$.

To prove (2.80), we first note that we have just showed that it holds in $m = 0$: $\varphi(M_1(0)) = \varphi(M-) = \varphi(M_+) < 0 = \varphi(0) = \varphi(M_2(0))$. We also know that $\varphi(M_1(m)) > \varphi(M_1(0))$ and $\varphi(M_2(m)) > \varphi(M_2(0))$, thanks to the monotonicity behavior of φ and Lemma 2.15. Hence suppose by contradiction that there exists $m \in]0, m_0]$ such that $\varphi(M_1(m)) = \varphi(M_2(m)) = c$, for some $c > 0$. This implies that both $M_1(m)$ and $M_2(m)$ are negative roots of $\varphi(M) - c$. Thus they are also negative roots of

$$\psi(M) := T\varphi(M) - Tc - f(M) = \frac{3}{2}TM^2 - (1 - 2T)M + m - Tc = 0$$

and $\psi'(q) = 0$, where $q = -\frac{2T-1}{3T}$ as above. Since ψ has degree 2, it follows that $|M_2(m) - q| = |M_1(m) - q|$, but this contradicts (2.76). Therefore there is no m for which $\varphi(M_1(m)) = \varphi(M_2(m))$, and then if (2.80) holds for $m = 0$ (which is (2.78)) then it is true for any $m \in [0, m_0]$. □

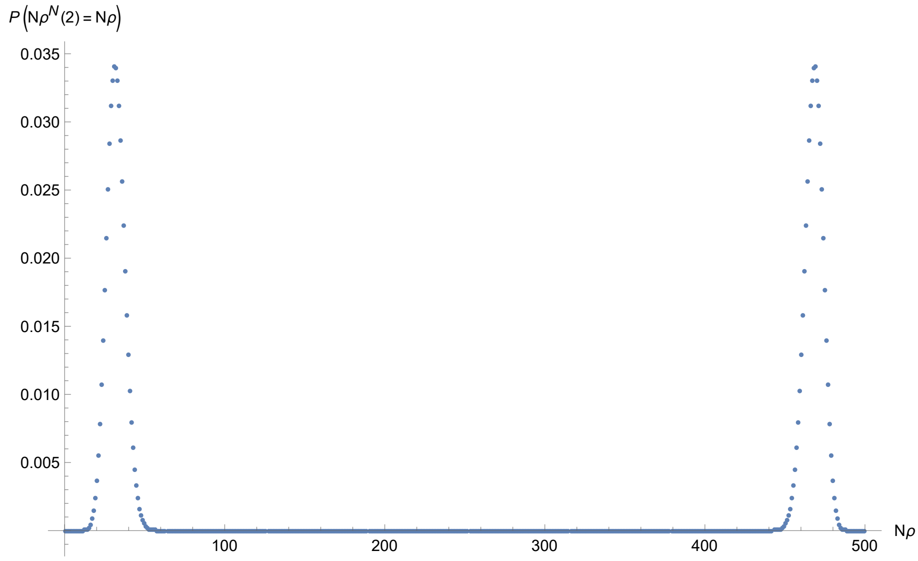


Figure 2.2: Simulation of the N-player dynamics. We plot the empirical distribution of $\rho^N(t)$ at the final time $t = T = 2$, $N = 500, \rho \in \{0, 1/N, \dots, 1\}$ for an initial datum concentrated in $\frac{1}{2}$.

Note that the results in this section imply that the N-player game selects, in the limit as $N \to +\infty$, the global minimizer of the control problem (2.72), when it is unique. Moreover, the sequence of the N-player value functions V^N converges to the derivative of the value function of such control problem, as the latter is constructed by using the same characteristic curves used for constructing the solution (2.19) to the master equation. We remark that the value function of the control problem (2.72) can also be characterized as the unique viscosity solution to (2.73).

2.3 Conclusions

Let us summarize the main results we have obtained for this two state model with anti-monotonic terminal cost:

1. the mean field game possesses exactly 3 solutions, if $T > 2$ (Proposition 2.1);
2. the N-player value functions converge to the entropy solution to the master equation (Theorem 2.7);
3. the N-player optimal trajectories converge to one mean field game solution, if $m_0 \neq 0$ (Theorem 2.10);
4. viewing the mean field game system as the necessary conditions for optimality of a deterministic control problem, the N-player game selects the global minimizer of this problem, when it is unique, i.e. $m_0 \neq 0$ (Theorem 2.16).

We remark that in the convergence proof we did not make use of the characterization of the right solution to the master equation as the entropy admissible one; the key point was to show that the N-player optimal trajectories do not cross the discontinuity. Neither did we use the potential structure of the problem: these are properties which might allow to extend the convergence results to more general models.

Observe that solutions of the MFG system, whether selected by the limit of N-player Nash equilibria or not, always yield approximate Nash equilibria in decentralized symmetric feedback strategies; see, for instance, [4] and [24] in the finite state setting. In this sense the other two solutions still have a physical meaning.

What is left to prove for this model is a propagation of chaos result when $m_0 = 0$. Let m_+, resp. m_-, be the mean field game solution always positive, resp. always negative. What is evident from the simulations (see Figure 2.2) is the following

Conjecture 2.1. *Let $m_0 = 0$ and m^N be the empirical mean related to the optimal trajectories of the N-player game, viewed as a random variable in $D([0,T],[-1,1])$. Then*

$$\lim_N Law(m^N) = \frac{1}{2}\delta_{m_+} + \frac{1}{2}\delta_{m_-}. \tag{2.82}$$

The limit of the empirical measures is not deterministic: in this sense there is no propagation of chaos when $m_0 = 0$, i.e. the initial point is exactly in the discontinuity. Unfortunately we did not manage to prove this result for our model, since it is difficult to track the Nash system in a neighborhood of the discontinuity. We remark that a similar result, in the regime of open-loop controls, was recently obtained in [42] for a linear-quadratic mean field game in dimension 1. We finally remark that the claim of Conjecture 2.1 should remain true also for the modified example of Section 2.2.6, due to the numerical simulations which show the same behaviour as in Figure 2.2 .

Part II

Non-Markovian interacting spin systems

CHAPTER 3

The mean field case

In this chapter we analyze two examples of non-Markovian mean field interacting spin systems. In both cases we consider dynamics of spin-flip type, related to the Curie–Weiss model. In the first example we relax the Markovianity assumption by replacing the memoryless distribution of the waiting times of a classical spin-flip dynamics with a distribution with memory. The resulting stochastic evolution for a single particle is a spin-valued *renewal* process, an example of two states semi-Markov process. As we shall see, we can associate to the individual dynamics an equivalent Markovian description, which is the subject of our analysis. We study a corresponding interacting particle system, where a mean field interaction is introduced as a time scaling, depending on the overall magnetization of the system, on the waiting times between two successive particle's jumps.

In the second model instead, the non-Markovianity follows by an *augmentation of state* procedure, where we double the state space assigning to each microscopic spin another spin-valued variable which produces frustration in the system. The resulting model is finite-dimensional, allowing for a deeper analysis of the phase-space diagram of the macroscopic limit equations.

Interestingly, we show that the above models belong to the same universality class: they both feature the presence of a unique stable neutral phase for values of the parameters corresponding to high temperatures, the emergence of periodic orbits in an intermediate range of the parameter values, and a subsequent ferromagnetic ordered phase for increasingly lower temperatures. In particular both dynamics can generate self-sustained oscillations: in the first case this seems to be a global phenomenon (even though we were not able to prove it), while in the finite-dimensional model we show that the cycles appear only starting the dynamics in a specific area of the phase-space.

Emerging periodic behavior in complex systems with a large number of interacting units is a commonly observed phenomenon in a variety of life science applications such as neuroscience ([55]) and ecology ([99]), but also in socioeconomics ([27, 100]), whose mathematical modelization has raised an interest in the community of probabilists and physicists working on interacting particle systems, and in particular on mean field models, due to their analytical tractability. With the term self-sustained periodic behavior we refer to systems where each individual particle has no natural tendency to behave periodically, but the oscillations are rather an effect of self-organization, visible in the macroscopic limit when the number of particles tends to infinity. One of the goals of the mathematical theory in this field is to understand which types of microscopic interactions and mechanisms can lead to or enhance the above self-organization. Among others, we cite noise ([36], [92],

[98]), dissipation in the interaction potential ([1], [29], [30], [35]), delay in the transmission of information and/or frustration in the interaction network ([31], [50], [97]). Specifically, in [50] the authors consider non-Markovian dynamics, studying systems of interacting nonlinear Hawkes processes for modeling neurons.

Although not proved in general, a strong belief in the literature is that, at least for Markovian dynamics, self-sustained periodic behavior cannot emerge if one does not introduce some time-irreversible phenomenon in the dynamics, as it is the case in all the above cited works (see e.g. [10], [60]). While the finite-dimensional model treated in Section 3.3 falls within the above examples (due to the presence of frustration), the model of Section 3.2, in which the limit dynamics is still reversible with respect to the stationary distribution around which cycles emerge (see Remark 3.3 below), suggests that this paradigm could be false for the non-Markovian case.

Before proceeding with the analysis of the two models, we briefly review some preliminary concepts in Section 3.1. Finally, we point out that in this chapter we proceed less rigorously than in the rest of the Dissertation, often relying on numerical evidence.

3.1 Preliminaries

As we already mentioned, the models considered in this chapter can be seen as proper modifications of the Curie–Weiss dynamics. When we refer to the latter, we mean a spin-flip type Markovian dynamics for a system of N interacting spins $\sigma_i \in \{-1,1\}$, $i = 1,\dots,N$, which is *reversible* with respect to the equilibrium Gibbs probability measure on the space of configurations $\{-1,1\}^N$,

$$P_{N,\beta}(\boldsymbol{\sigma}) := \frac{1}{Z_N(\beta)} \exp\left[-\beta H(\boldsymbol{\sigma})\right], \tag{3.1}$$

with $\boldsymbol{\sigma} := (\sigma_1,\dots,\sigma_N) \in \{-1,1\}^N$, $\beta > 0$ (*ferromagnetic* case), and $Z_N(\beta)$ is a normalizing constant. In statistical mechanics, the function H is called *Hamiltonian*: it specifies the energy of each spin configuration $\boldsymbol{\sigma} \in \{-1,1\}^N$, and in the Curie–Weiss case it is given by

$$H_N(\boldsymbol{\sigma}) := -\frac{1}{2N}\left(\sum_{i=1}^N \sigma_i\right)^2. \tag{3.2}$$

Define also the empirical *magnetization* as

$$m^N := \frac{1}{N}\sum_{i=1}^N \sigma_i.$$

Note that the distribution (3.1) gives higher probability to the configurations with minimal energy, which by (3.2) are the ones where the individual spins are aligned in the same state. The equilibrium model undergoes a *phase transition* tuned by the interaction parameter $\beta > 0$ (referred to as inverse temperature in the literature), which can be recognized by proving a Law of Large Numbers for the equilibrium empirical magnetization

$$\mathrm{Law}(m^N) \xrightarrow{N\to+\infty} \begin{cases} \delta_0, & \text{if } \beta \le 1, \\ \frac{1}{2}\delta_{+m_\beta} + \frac{1}{2}\delta_{-m_\beta}, & \text{if } \beta > 1, \end{cases} \tag{3.3}$$

where $m_\beta > 0$ is the so-called *spontaneous magnetization.*

When we turn to the dynamics, different choices can be made in order to satisfy the above-mentioned reversibility with respect to (3.1). The prototype is a continuous-time spin-flip dynamics defined in terms of the infinitesimal generator L, applied to a function $f : \{-1,1\}^N \to \mathbb{R}$,

$$Lf(\boldsymbol{\sigma}) = \sum_{i=1}^{N} e^{-\beta\sigma_i m^N} \left[f(\boldsymbol{\sigma}^i) - f(\boldsymbol{\sigma}) \right], \tag{3.4}$$

where $\boldsymbol{\sigma}^i \in \{-1,1\}^N$ is obtained from $\boldsymbol{\sigma}$ by *flipping* the i-th spin, i.e.

$$\boldsymbol{\sigma}^i_k := \begin{cases} -\sigma_k, & \text{if } k = i, \\ \sigma_k & \text{if } k \neq i. \end{cases}$$

Note that in Chapter 4 we consider the different spin-flip rates

$$1 - \tanh(\beta\sigma_i m^N),$$

for which the reversibility with respect to (3.1) still holds. Dynamics (3.4) induces a continuous-time Markovian evolution for the empirical magnetization process $m^N(t)$, which is given in terms of a generator $\mathcal{L}$ applied to a function $g : [-1,1] \to \mathbb{R}$:

$$\mathcal{L}^N g(m) = N\frac{1+m}{2} e^{-\beta m} \left[g\left(m - \frac{2}{N}\right) - g(m) \right] + N\frac{1-m}{2} e^{\beta m} \left[g\left(m + \frac{2}{N}\right) - g(m) \right]. \tag{3.5}$$

The above generator can be obtained by observing that, when a spin σ_i flips from σ_i to $-\sigma_i$ at time t, the empirical magnetization changes by a quantity $-\frac{2\sigma_i}{N}$. The factors $N\frac{1+m}{2}$ and $N\frac{1-m}{2}$ represent the number of spins in state 1 and -1 respectively. It is easy to obtain the weak limit of the sequence of processes $\left(m^N(t)\right)_{t\in[0,T]}$, by studying the uniform convergence of the generator (3.5) as $N \to +\infty$. The limit process $m(t)$ is deterministic and solves the Curie–Weiss ODE

$$\begin{cases} \dot{m}(t) = 2\sinh(\beta m(t)) - 2m(t)\cosh(\beta m(t)), \\ m(0) = m_0 \in [-1,1]. \end{cases} \tag{3.6}$$

The presence of the phase transition highlighted in (3.3) can be recognized as well in the out-of-equilibrium dynamical model (3.6). Indeed, studying the long-term behavior of (3.6), one finds that:

- for $\beta \leq 1$, (3.6) possesses a unique stationary solution, globally attractive, constantly equal to 0;
- for $\beta > 1$, 0 is still stationary but it is unstable; two other symmetric stationary locally attractive solutions, $\pm m_\beta$, appear: the two non-zero solutions to

  $$m = \tanh(\beta m).$$

  The dynamics $m(t)$ gets attracted for $t \to +\infty$ to the polarized stationary state which has the same sign as the initial magnetization m_0.

Another concept which we refer to in Section 3.2 is that of a *renewal* process, a generalization of the Poisson process. As the latter, it is a stochastic process for events that occur randomly in time. For our purposes, we identify a renewal process with the sequence of its *interarrival* times (also commonly referred to as sojourn times or waiting times in the literature) $\{T_n\}_{n=1}^{\infty}$, i.e. the holding times between the occurrences of two consecutive events. The Poisson process (see [12, Ch. 8] for a thorough introduction) is characterized by having independent and identically distributed interarrival times, where each T_i is exponentially distributed. In particular, the following *memoryless* property is satisfied for any $i = 1, 2, \dots$

$$\mathbb{P}(T_i > s + t | T_i > t) = \mathbb{P}(T_i > s),$$

for any $s, t \geq 0$. The interarrival times of a renewal process are still independent and identically distributed, but their distribution is not required to be exponential. We recall that a continuous-time homogeneous Markov chain can be identified by a Poisson process, modeling the jump times, and a stochastic transition matrix, identifying the possible arrival states at each jump time. Due to the lack of the memoryless property, when one replaces the Poisson process in the definition of the spin-flip dynamics with a more general renewal process, the resulting evolution is thus non-Markovian. In the literature, the associated dynamics is referred to as *semi-Markov* process, first introduced by Levy in [81].

3.2 Oscillatory behavior in a model of mean field interacting renewal processes

In this section we consider a non-Markovian, infinite-dimensional modification of the Curie–Weiss model and exhibit some partial evidence of its oscillatory behavior.

In order to introduce the model, we start by observing that the Curie–Weiss dynamics (3.4), as any spin-flip Glauber dynamics, can be obtained by adding interaction to a system of independent spin-flips: at the times of a Poisson process of intensity 1, the spin in a given site flips; different sites have independent Poisson processes. Our aim here is to replace Poisson processes by more general renewal processes, otherwise keeping the structure of the interaction. For the moment we focus on a single spin $\sigma(t) \in \{-1, 1\}$. If driven by a Poisson process of intensity 1, its dynamics has infinitesimal generator

$$\mathcal{L}f(\sigma) = f(-\sigma) - f(\sigma), \tag{3.7}$$

$f : \{-1, 1\} \to \mathbb{R}$. If the Poisson process is replaced by a renewal process, the spin dynamics is not Markovian. In what follows, we refer to the resulting dynamics as a spin-valued renewal process, that is an example of two-states semi-Markov process. We can associate a Markovian description to the latter: define $y(t)$ as the time elapsed since the last spin-flip occured up to time t. Suppose that the waiting times τ (interchangeably referred to as interarrival times) of the renewal satisfy

$$\mathbb{P}(\tau > t) = \varphi(t), \tag{3.8}$$

for some smooth function $\varphi : [0, +\infty) \to \mathbb{R}$. Then, the pair $(\sigma(t), y(t))_{t \geq 0}$ is Markovian with generator

$$\mathcal{L}f(\sigma, y) = \frac{\partial f}{\partial y}(\sigma, y) + F(y)[f(-\sigma, 0) - f(\sigma, y)], \tag{3.9}$$

for $f : \{-1, 1\} \times \mathbb{R}^+ \to \mathbb{R}$, with

$$F(y) := -\frac{\varphi'(y)}{\varphi(y)}. \tag{3.10}$$

This is equivalent to say that the couple $(\sigma(t), y(t))_{t \geq 0}$ evolves according to

$$\begin{cases} (\sigma(t), y(t)) \mapsto (-\sigma(t), 0), & \text{with rate } F(y(t)), \\ dy(t) = dt, & \text{otherwise.} \end{cases} \tag{3.11}$$

Expression (3.10) for the jump rate follows by observing that, for an interarrival time τ of the jump process $\sigma(t)$, we have

$$\mathbb{P}(\sigma(t+h) = -\sigma | \sigma(t) = \sigma) = 1 - \mathbb{P}(\tau > t+h | \tau > t) = 1 - \frac{\varphi(t+h)}{\varphi(t)},$$

for any $h > 0$. Observe that when the τ's are exponentially distributed $F(y) \equiv 1$, so we get back to dynamics (3.7).

Dynamics (3.9) can be perturbed by allowing the distribution of the waiting time for a spin-flip to depend on the current spin value σ; the simplest way is to model this dependence as a time scaling:

$$\mathbb{P}(\tau > t | \sigma) = \varphi(a(\sigma)t). \tag{3.12}$$

Under this distribution for the waiting times the generator of $(\sigma(t), y(t))_{t \geq 0}$ becomes:

$$\mathcal{L}f(\sigma, y) = \frac{\partial f}{\partial y}(\sigma, y) + a(\sigma) F(a(\sigma) y)[f(-\sigma, 0) - f(\sigma, y)].$$

The rest of this section is organized as follows: in Section 3.2.1 we introduce the mean field model; in Section 3.2.2 we study the linearized Fokker-Planck equation around a neutral equilibrium, for two different choices of renewal dynamics. We determine the discrete spectrum of the linearized operator in terms of the zeros of two holomorphic functions, which we then study numerically as functions of the interaction parameters. The results are then compared in Section 3.2.3 with the ones obtained by simulating the finite particle system, finding a precise accordance between the two approaches.

3.2.1 The mean field model

On the basis of what seen above, it is rather simple to define a system of mean-field interacting spins with non-exponential waiting times. For a collection of N pairs $(\sigma_i(t), y_i(t))_{i=1,\dots,N}$, we set $m^N(t) := \frac{1}{N}\sum_{i=1}^N \sigma_i(t)$ to be the magnetization of the system at time t, and a parameter $\beta > 0$ tuning the interaction between the particles. The interacting dynamics is

$$\begin{cases} (\sigma_i(t), y_i(t)) \mapsto (-\sigma_i(t), 0), & \text{with rate} \quad F\left(y_i(t) e^{-\beta\sigma_i(t) m^N(t)}\right) e^{-\beta\sigma_i(t) m^N(t)}, \\ dy_i(t) = dt, & \text{otherwise}. \end{cases} \tag{3.13}$$

Denoting $\boldsymbol{\sigma} := (\sigma_1, \dots, \sigma_N) \in \{-1, 1\}^N$, $\boldsymbol{y} := (y_1, \dots, y_N) \in (\mathbb{R}^+)^N$, $m^N := \frac{1}{N}\sum_{i=1}^N \sigma_i$, the associated infinitesimal generator is

$$\mathcal{L}^N f(\boldsymbol{\sigma}, \boldsymbol{y}) = \sum_{i=1}^N \frac{\partial f}{\partial y_i}(\boldsymbol{\sigma}, \boldsymbol{y}) + \sum_{i=1}^N F\left(y_i e^{-\beta\sigma_i m^N}\right) e^{-\beta\sigma_i m^N} \left[f(\boldsymbol{\sigma}^i, \boldsymbol{y}^i) - f(\boldsymbol{\sigma}, \boldsymbol{y})\right], \tag{3.14}$$

where $\boldsymbol{\sigma}^i$ is obtained from $\boldsymbol{\sigma}$ by flipping the i-th spin, while $\boldsymbol{y}^i$ by setting to zero the i-th coordinate. The additional factor $e^{-\beta\sigma_i(t) m^N(t)}$ in the jump rate in (3.13) follows from the observation we made in (3.12) and the definition of $F(y) = -\frac{\varphi'(y)}{\varphi(y)}$. Note that, for $F \equiv 1$, we retrieve the Curie–Weiss dynamics (3.4) for the spins.

The macroscopic limit and propagation of chaos for this class of models should be standard, although some difficulties may arise for general choices of F not globally Lipschitz. For computational reasons which will be made clear in the next section, we focus on the case $F(y) = y^\gamma$, for $\gamma \in \mathbb{N}$, which corresponds to considering, in the single spin model, the tails of the distribution of the interarrival times to be $\varphi(t) \propto e^{-\frac{t^{\gamma+1}}{\gamma+1}}$. In Appendix B, we study rigorously the well-posedeness of the pre-limit and limit dynamics and the propagation of chaos under this choice of rate function.

When $F(y) = y^\gamma$, (3.13) becomes

$$\begin{cases} (\sigma_i(t), y_i(t)) \mapsto (-\sigma_i(t), 0), & \text{with rate} \quad y_i^\gamma(t) e^{-(\gamma+1)\beta\sigma_i(t) m^N(t)}, \\ dy_i(t) = dt, & \text{otherwise.} \end{cases} \tag{3.15}$$

As for the Curie–Weiss model, dynamics (3.15) is subject to a cooperative-type interaction: the spin-flip rate is larger for particles which are not aligned with the majority. Assuming propagation of chaos, at the macroscopic limit $N \to +\infty$ the representative particle $(\sigma(t), y(t))$ has a mean-field dynamics

$$\begin{cases} (\sigma(t), y(t)) \mapsto (-\sigma(t), 0), & \text{with rate} \quad y^\gamma(t) e^{-(\gamma+1)\beta\sigma(t) m(t)}, \\ dy(t) = dt, & \text{otherwise,} \end{cases} \tag{3.16}$$

with $m(t) = \mathbb{E}[\sigma(t)]$. To this dynamics we can associate (see [76]) the non-linear infinitesimal generator

$$\mathcal{L}(m(t)) f(\sigma, y) = \frac{\partial f}{\partial y}(\sigma, y) + y^\gamma e^{-(\gamma+1)\beta\sigma m(t)} \left[f(-\sigma, 0) - f(\sigma, y) \right], \tag{3.17}$$

where the non-linearity is due to the dependence of the generator on $m(t)$, a function of the joint law at time t of the processes $(\sigma(t), y(t))$.

3.2.2 Local analysis of the Fokker-Planck

In this section we perform a local analysis on the Fokker-Planck equation for the mean-field limit dynamics (3.16) with $\gamma = 1$ and $\gamma = 2$. Our approach is the following: we find a neutral stationary solution of interest, we linearize formally the dynamics around that equilibrium and we compute the discrete spectrum of the associated linearized operator, which we show to be given by the zeros of an explicit holomorphic function $H_{\beta,\gamma}(\lambda)$. We then study numerically the character of the eigenvalues when β varies: for both $\gamma = 1, 2$, we find that for all $\beta < \beta_c(\gamma)$ all eigenvalues have negative real part; at $\beta_c(\gamma)$ two eigenvalues are conjugate and purely imaginary, suggesting the possible presence of a Hopf bifurcation in the limit dynamics. These critical values of β are then compared to the ones obtained by simulating the finite particle system in Section 3.2.3.

The Fokker-Planck equation associated to (3.16) is a PDE describing the time evolution of the density function $f(t, \sigma, y)$ of the limit process $(\sigma(t), y(t))$. It is given by

$$\begin{cases} \frac{\partial}{\partial t} f(t, \sigma, y) + \frac{\partial}{\partial y} f(t, \sigma, y) + y^\gamma e^{-(\gamma+1)\beta\sigma m(t)} f(t, \sigma, y) = 0, \\ f(t, \sigma, 0) = \int_0^{+\infty} y^\gamma e^{(\gamma+1)\beta\sigma m(t)} f(t, -\sigma, y) dy, \\ m(t) = \int_0^\infty [f(t, 1, y) - f(t, -1, y)] dy, \\ 1 = \int_0^\infty [f(t, 1, y) + f(t, -1, y)] dy, \\ f(0, \sigma, y) = f_0(\sigma, y), \text{ for } \sigma \in \{-1, 1\}, \ y \in \mathbb{R}^+. \end{cases} \tag{3.18}$$

A general study of (3.18) is beyond the scope of this work. Here we just observe that (3.18) can be seen as a system of two quasilinear PDEs (one for $\sigma = 1$ and another for $\sigma = -1$), where the non-linearity enters in an integral form through $m(t)$ in the exponent of the rate function. Moreover, the boundary integral condition in the second line poses additional challenges.

Remark 3.1. *While the other equations in* (3.18) *are derived in a standard way from the expression of the generator* (3.17)*, the boundary integral condition might need to be motivated. In words, it is a mass-balance between the spins that have just jumped (thus having $y = 0$). We reason heuristically by discretizing the state space $[0, +\infty)$ in small intervals of amplitude ε. The discretized version of $y(t)$ takes values in $\{n\varepsilon \, : \, n \in \mathbb{N}\}$. The associated generator is, for $n \in \mathbb{N}$ and $\sigma \in \{-1, 1\}$,*

$$\begin{aligned}\mathcal{L}_\varepsilon f(\sigma, n\varepsilon) = \frac{1}{\varepsilon}\big[f(\sigma, (n+1)\varepsilon) - f(\sigma, n\varepsilon)\big] \\ + (n\varepsilon)^\gamma e^{-(\gamma+1)\beta\sigma m(t)}\big[f(-\sigma, 0) - f(\sigma, n\varepsilon)\big].\end{aligned}$$

Denoting $f(t, \sigma, 0)$ the density of the discretized process in $(\sigma, 0)$ at time t, it follows from the expression of $\mathcal{L}_\varepsilon$,

$$\frac{d}{dt} f(t, \sigma, 0) = \sum_{n \in \mathbb{N}} (n\varepsilon)^\gamma e^{-(\gamma+1)\beta\sigma m(t)} f(t, -\sigma, n\varepsilon) - \frac{1}{\varepsilon} f(t, \sigma, 0),$$

that is the discretized version of the integral condition in (3.17)*.*

It is easy to exhibit a particular stationary solution to (3.18):

Proposition 3.2. *The function*

$$f^*(\sigma, y) = \frac{1}{2\Lambda} e^{-\frac{y^{\gamma+1}}{\gamma+1}}, \tag{3.19}$$

with $\Lambda := \int_0^{+\infty} e^{-\frac{y^{\gamma+1}}{\gamma+1}}$, is a stationary solution to System (3.18) *with $m = 0$.*

Proof. Setting $m = 0$ in the above system, the stationary version of the first equation becomes

$$\frac{\partial}{\partial y} f(\sigma, y) + y^\gamma f(\sigma, y) = 0, \tag{3.20}$$

whose solution is of the form $f^*(\sigma, y) = c(\sigma) f(\sigma, 0) e^{-\frac{y^{\gamma+1}}{\gamma+1}}$. Denoting $\Lambda := \int_0^{+\infty} e^{-\frac{y^{\gamma+1}}{\gamma+1}}$, it is easy to see that the integral conditions imply $c(\sigma) = c(-\sigma) = \frac{1}{\Lambda}$ and $f(\sigma, 0) = f(-\sigma, 0) = \frac{1}{2}$. □

Remark 3.3. *Let $g^*(\sigma)$ be the marginal of $f^*(\sigma, y)$ with respect to the first coordinate. Then, $g^*(\sigma)$ is a stationary reversible distribution for the limit renewal process $(\sigma(t))_{t \geq 0}$. Indeed, by choosing $\sigma(0) \sim g^*$, $g^*(1) = g^*(-1) = \frac{1}{2}$, we have that $m(t) \equiv 0$ and $(\sigma(t))_{t \geq 0}$ is a renewal process with interarrival times τ such that $\mathbb{P}(\tau > t) \propto e^{-\frac{t^{\gamma+1}}{\gamma+1}}$ independently of the value of σ, so its law is invariant by time reversal.*

3.2.2.1 Linearized stationary system

We now compute formally the linearization of the operator associated to System (3.18) around the solution (3.19) with $m = 0$. Namely, if we write the first equation in (3.18) in operator form

$$\frac{\partial}{\partial t} f(t, \sigma, y) - \mathcal{L}^{nl}_\gamma f(t, \sigma, y) = 0,$$

with $\mathcal{L}^{nl}_\gamma f(t, \sigma, y) := -\frac{\partial}{\partial y} f(t, \sigma, y) - y^\gamma e^{-(\gamma+1)\beta\sigma m(t)} f(t, \sigma, y)$, we want to find the linearized version of the operator $\mathcal{L}^{nl}_\gamma$.

For the purpose, we express a generic stationary solution to (3.18) as

$$f(\sigma, y) = f^*(\sigma, y) + \varepsilon g(\sigma, y),$$

imposing

$$\int_0^\infty [g(1, y) + g(-1, y)]dy = 0, \tag{3.21}$$

so that $\int_0^\infty [f(1, y) + f(-1, y)]dy = 1$ is satisfied. We also denote $m_f := \int_0^\infty [f(1, y) - f(-1, y)]dy$, which by the above consideration satisfies

$$m_f = 2\varepsilon \int_0^\infty g(1, y)dy =: \varepsilon k. \tag{3.22}$$

The stationary version of the first equation in (3.18) becomes

$$\frac{\partial}{\partial y} f^*(\sigma, y) + \varepsilon \frac{\partial}{\partial y} g(\sigma, y) + y^\gamma e^{-\beta\sigma\varepsilon k(\gamma+1)} [f^*(\sigma, y) + \varepsilon g(\sigma, y)] = 0.$$

By expanding at the first order in ε the term $e^{-\beta\sigma\varepsilon k(\gamma+1)} \approx 1 - (\gamma + 1)\beta\sigma\varepsilon k$, and by considering only the resulting linear terms in ε, we get

$$\frac{\partial}{\partial y} f^*(\sigma, y) + \varepsilon \frac{\partial}{\partial y} g(\sigma, y) + y^\gamma f^*(\sigma, y) + y^\gamma \varepsilon g(\sigma, y) - y^\gamma (\gamma + 1)\beta\sigma\varepsilon k f^*(\sigma, y) = 0.$$

Finally, using that f^* solves (3.20) and substituting its expression (3.19), we get

$$\frac{\partial}{\partial y} g(\sigma, y) + y^\gamma g(\sigma, y) - \frac{\beta\sigma k(\gamma + 1)}{2\Lambda} y^\gamma e^{-\frac{y^{\gamma+1}}{\gamma+1}} = 0.$$

We can define the linearized operator as

$$\mathcal{L}_\gamma^{\text{lin}} g(\sigma, y) := -\frac{\partial}{\partial y} g(\sigma, y) - y^\gamma g(\sigma, y) + \frac{\beta\sigma k(\gamma + 1)}{2\Lambda} y^\gamma e^{-\frac{y^{\gamma+1}}{\gamma+1}}. \tag{3.23}$$

We proceed with the linearization of the integral condition in the second line of System (3.18):

$$\begin{aligned} f^*(\sigma, 0) + \varepsilon g(\sigma, 0) &= \int_0^\infty [f^*(-\sigma, y) + \varepsilon g(-\sigma, y)] y^\gamma e^{\beta\sigma\varepsilon k(\gamma+1)} \\ &\approx \int_0^\infty f^*(-\sigma, y) y^\gamma (1 + \beta\sigma\varepsilon k(\gamma + 1)) + \varepsilon \int_0^\infty g(-\sigma, y) y^\gamma (1 + \beta\sigma\varepsilon k(\gamma + 1)) \\ &\approx \int_0^\infty f^*(-\sigma, y) y^\gamma + \beta\sigma\varepsilon k(\gamma + 1) \int_0^\infty f^*(-\sigma, y) y^\gamma + \varepsilon \int_0^\infty g(-\sigma, y) y^\gamma. \end{aligned}$$

Using again that f^* solves (3.20) and its expression in (3.19), we get

$$g(\sigma, 0) = \frac{\beta\sigma k(\gamma + 1)}{2\Lambda} + \int_0^\infty g(-\sigma, y) y^\gamma dy. \tag{3.24}$$

In order to gain indications on the stability properties of the stationary solution to (3.18) with $m = 0$, we study the discrete spectrum of $\mathcal{L}_\gamma^{\text{lin}}$ defined in (3.23), i.e., we search for the eigenfunctions g and the eigenvalues $\lambda \in \mathbb{C}$, satisfying the linearized integral conditions (3.21) and (3.24) found above, and such that

$$\mathcal{L}_\gamma^{\text{lin}} g(\sigma, y) = \lambda g(\sigma, y), \tag{3.25}$$

which is equivalent to

$$\frac{\partial}{\partial y}g(\sigma,y)+y^{\gamma}g(\sigma,y)-\frac{\beta\sigma k(\gamma+1)}{2\Lambda}y^{\gamma}e^{-\frac{y^{\gamma+1}}{\gamma+1}}=-\lambda g(\sigma,y). \tag{3.26}$$

The eigen-system around $m=0$ is thus given by

$$\begin{cases}\frac{\partial}{\partial y}g(\sigma,y)+y^{\gamma}g(\sigma,y)-\frac{\beta\sigma k(\gamma+1)}{2\Lambda}y^{\gamma}e^{-\frac{y^{\gamma+1}}{\gamma+1}}=-\lambda g(\sigma,y),\\ g(\sigma,0)=\frac{\beta\sigma k(\gamma+1)}{2\Lambda}+\int_0^{\infty}g(-\sigma,y)y^{\gamma}dy,\\ \int_0^{\infty}[g(\sigma,y)+g(-\sigma,y)]dy=0,\quad (\sigma,y)\in\{-1,1\}\times\mathbb{R}^+,\end{cases} \tag{3.27}$$

where, recall by (3.22), $k=2\int_0^{\infty}g(1,y)dy$, and $\Lambda=\int_0^{\infty}e^{-\frac{y^{\gamma+1}}{\gamma+1}}dy$. We work out the computations for the two cases $\gamma=1$, $\gamma=2$.

Remark 3.4. *The derivation of the linearized operator* (3.25) *was formal. One could think to define it more rigorously, by indicating an Hilbert space where $\mathcal{L}_{\gamma}^{lin}$ acts on. The natural choice appears to be (a subspace of) $\left(L^2_{\mu_\gamma}(\mathbb{R}^+)\right)^2$ satisfying conditions* (3.21) *and* (3.24)*, where the outer square comes from the explicitation of the spin variable $\sigma=\pm1$, and the measure μ_γ is defined as*

$$\mu_{\gamma}(dy):=f^*(\sigma,y)dy=\frac{1}{2\Lambda}e^{-\frac{y^{\gamma+1}}{\gamma+1}}dy. \tag{3.28}$$

As in what follows we do not use the particular choice of domain of the operator or its properties, we do not investigate further on this.

3.2.2.2 Case $\gamma=1$

In this case, $\Lambda=\sqrt{\frac{\pi}{2}}$, and the eigen-system (3.27) becomes

$$\begin{cases}\frac{\partial}{\partial y}g(\sigma,y)+yg(\sigma,y)+\lambda g(\sigma,y)=\beta\sigma k\left(\sqrt{\frac{\pi}{2}}\right)^{-1}ye^{-\frac{y^2}{2}}\\ g(\sigma,0)=\beta\sigma k\left(\sqrt{\frac{\pi}{2}}\right)^{-1}+\int_0^{\infty}yg(-\sigma,y)dy,\\ \int_0^{\infty}[g(\sigma,y)+g(-\sigma,y)]dy=0,\end{cases} \tag{3.29}$$

where $k=2\int_0^{\infty}g(1,y)dy$.

Proposition 3.5. *The solutions in $\lambda\in\mathbb{C}$ to* (3.29) *are the zeros of the holomorphic function*

$$H_{\beta,1}(\lambda):=H_1(\lambda)\left[-4\beta-\lambda^3\sqrt{\frac{\pi}{2}}\right]+\sqrt{2\pi}\lambda^2-4\beta\lambda+2\beta\sqrt{2\pi}, \tag{3.30}$$

with

$$H_1(\lambda):=\int_0^{\infty}e^{-\frac{y^2}{2}}e^{-\lambda y}. \tag{3.31}$$

Moreover, it holds

$$H_1(\lambda)=\sqrt{\frac{\pi}{2}}\sum_{m=0}^{\infty}\frac{\lambda^{2m}}{(2m)!!}-\lambda\sum_{m=0}^{\infty}\frac{(2\lambda)^{2m}m!}{(2m+1)!}\frac{1}{2^m}. \tag{3.32}$$

Proof. In order to solve the first equation in (3.29), we set

$$h(\sigma, y) := g(\sigma, y)e^{\frac{y^2}{2}}.$$

It holds

$$\frac{\partial}{\partial y}h(\sigma, y) = -\lambda h(\sigma, y) + \frac{y\beta\sigma k}{\sqrt{\frac{\pi}{2}}},$$

whose solution is

$$h(\sigma, y) = e^{-\lambda y}\left[h(\sigma, 0) + \frac{\beta\sigma k}{\sqrt{\frac{\pi}{2}}}\int_0^y ue^{\lambda u}du\right].$$

Noting that $\int_0^y ue^{\lambda u}du = \frac{1}{\lambda^2} - \frac{e^{\lambda y}}{\lambda^2} + \frac{e^{\lambda y}}{\lambda}y$, we obtain

$$g(\sigma, y) = e^{-\frac{y^2}{2}}e^{-\lambda y}\left[g(\sigma, 0) + \frac{\beta\sigma k}{\sqrt{\frac{\pi}{2}}}\left(\frac{1}{\lambda^2} - \frac{e^{\lambda y}}{\lambda^2} + \frac{e^{\lambda y}}{\lambda}y\right)\right]. \tag{3.33}$$

We now impose the integral conditions. First, we note that $\int_0^\infty [g(\sigma, y) + g(-\sigma, y)]dy = 0$ is equivalent to $g(\sigma, y) + g(-\sigma, y) = 0$ for every $y \in \mathbb{R}^+$ because of expression (3.33). For the computation of k, recalling notation (3.31), we find

$$\begin{aligned} k &= 2\int_0^\infty g(1, y)dy \\ &= 2g(1, 0)H_1(\lambda) + 2\frac{\beta k}{\sqrt{\frac{\pi}{2}}}\frac{1}{\lambda^2}H_1(\lambda) - 2\frac{\beta k}{\lambda^2}\frac{1}{\sqrt{\frac{\pi}{2}}}\sqrt{\frac{\pi}{2}} + 2\frac{\beta k}{\sqrt{\frac{\pi}{2}}}\frac{1}{\lambda}, \end{aligned}$$

so that

$$k = \frac{2g(1, 0)H_1(\lambda)}{1 - 2\frac{\beta}{\lambda\sqrt{\frac{\pi}{2}}} - 2\frac{\beta H_1(\lambda)}{\lambda^2\sqrt{\frac{\pi}{2}}} + 2\frac{\beta}{\lambda^2}}. \tag{3.34}$$

The integral condition in the second line of (3.29) gives

$$\begin{aligned} g(\sigma, 0) &= \frac{\beta\sigma k}{\sqrt{\frac{\pi}{2}}} + \int_0^\infty y\left[e^{-\frac{y^2}{2}}e^{-\lambda y}\left(g(-\sigma, 0) - \frac{\beta\sigma k}{\sqrt{\frac{\pi}{2}}}\left(\frac{1}{\lambda^2} - \frac{e^{\lambda y}}{\lambda^2} + \frac{e^{\lambda y}}{\lambda}y\right)\right)\right] \\ &= \frac{\beta\sigma k}{\sqrt{\frac{\pi}{2}}} - g(\sigma, 0)(1 - \lambda H_1(\lambda)) - \frac{\beta\sigma k}{\sqrt{\frac{\pi}{2}}}\frac{(1 - \lambda H_1(\lambda))}{\lambda^2} \\ &\qquad + \frac{1}{\lambda^2}\frac{\beta\sigma k}{\sqrt{\frac{\pi}{2}}} - \frac{1}{\lambda}\frac{\beta\sigma k}{\sqrt{\frac{\pi}{2}}}\int_0^\infty y^2 e^{-\frac{y^2}{2}}dy \\ &= \frac{\beta\sigma k}{\sqrt{\frac{\pi}{2}}} - g(\sigma, 0)(1 - \lambda H_1(\lambda)) - \frac{(1 - \lambda H_1(\lambda))}{\lambda^2}\frac{\beta\sigma k}{\sqrt{\frac{\pi}{2}}} + \frac{1}{\lambda^2}\frac{\beta\sigma k}{\sqrt{\frac{\pi}{2}}} - \frac{1}{\lambda}\beta\sigma k. \end{aligned}$$

In the second equality we have used that $\int_0^\infty ye^{-\frac{y^2}{2}}e^{-\lambda y} = 1 - \lambda H_1(\lambda)$ which can be obtained by an integration by parts. Solving for $g(1, 0)$ in the above

$$g(1, 0)[2 - \lambda H_1(\lambda)] = \beta k\left[\frac{1}{\sqrt{\frac{\pi}{2}}} - \frac{(1 - \lambda H_1(\lambda))}{\lambda^2}\frac{1}{\sqrt{\frac{\pi}{2}}} + \frac{1}{\lambda^2}\frac{1}{\sqrt{\frac{\pi}{2}}} - \frac{1}{\lambda}\right].$$

Substituting the value of k we found in (3.34), we get

$$g(1,0)[2-\lambda H_1(\lambda)] = \frac{2\beta g(1,0)H_1(\lambda)}{1-2\frac{\beta}{\lambda\sqrt{\frac{\pi}{2}}}-2\frac{\beta H_1(\lambda)}{\lambda^2\sqrt{\frac{\pi}{2}}}+2\frac{\beta}{\lambda^2}}\left[\frac{1}{\sqrt{\frac{\pi}{2}}}-\frac{(1-\lambda H_1(\lambda))}{\lambda^2}\frac{1}{\sqrt{\frac{\pi}{2}}}+\frac{1}{\lambda^2}\frac{1}{\sqrt{\frac{\pi}{2}}}-\frac{1}{\lambda}\right],$$

which is equivalent to

$$2-\lambda H_1(\lambda) = \frac{2\beta H_1(\lambda)\left[\lambda^2+\lambda H_1(\lambda)-\lambda\sqrt{\frac{\pi}{2}}\right]}{\lambda^2\sqrt{\frac{\pi}{2}}-2\beta\lambda-2\beta H_1(\lambda)+2\beta\sqrt{\frac{\pi}{2}}}. \tag{3.35}$$

As a polynomial in λ, (3.35) can be written as

$$-\lambda^3 H_1(\lambda)\sqrt{\frac{\pi}{2}}+\lambda^2\sqrt{2\pi}-4\beta\lambda-4\beta H_1(\lambda)+2\sqrt{2\pi}\beta=0,$$

or, grouping for $H_1(\lambda)$,

$$H_1(\lambda)\left[-4\beta-\lambda^3\sqrt{\frac{\pi}{2}}\right]+\sqrt{2\pi}\lambda^2-4\beta\lambda+2\beta\sqrt{2\pi}=0,$$

i.e. the zeros of $H_{\beta,1}(\lambda)$, provided we prove expression (3.32) for $H_1(\lambda)$. In fact, as defined in (3.31), $H_1(\lambda)$ is a holomorphic function on $\mathbb{C}$, whose expression in series is

$$H_1(\lambda) = \int_0^\infty e^{-\frac{y^2}{2}}e^{-\lambda y}dy = \sum_{n=0}^{\infty}(-1)^n\frac{\lambda^n}{n!}\int_0^\infty y^n e^{-\frac{y^2}{2}}dy.$$

The latter integral is known

$$\int_0^\infty y^n e^{-\frac{y^2}{2}}dy = 2^{\frac{1}{2}(n-1)}\Gamma\left(\frac{n+1}{2}\right), \tag{3.36}$$

where $\Gamma(\cdot)$ is the Gamma function. When $n=2m+1$, for the properties of the Gamma function on $\mathbb{N}$, (3.36) reduces to

$$\int_0^\infty y^n e^{-\frac{y^2}{2}}dy = 2^{\frac{1}{2}(n-1)}\Gamma\left(\frac{n+1}{2}\right) = 2^m m!.$$

For $n=2m$ instead we have, by the property $\Gamma\left(l+\frac{1}{2}\right)=\frac{(2l-1)!!}{2^l}\sqrt{\pi}$ for any $l\in\mathbb{N}$,

$$\int_0^\infty y^n e^{-\frac{y^2}{2}}dy = 2^{\frac{1}{2}(n-1)}\Gamma\left(\frac{n+1}{2}\right) = \sqrt{\frac{\pi}{2}}(2m-1)!!.$$

We use these equalities, and reorder the terms of the absolutely convergent series of $H_1(\lambda)$ to finally get

$$H_1(\lambda) = \sqrt{\frac{\pi}{2}}\sum_{m=0}^{\infty}\frac{\lambda^{2m}}{(2m)!!}-\lambda\sum_{m=0}^{\infty}\frac{(2\lambda)^{2m}m!}{(2m+1)!}\frac{1}{2^m},$$

i.e. expression (3.32). □

3.2.2.3 Case $\gamma = 2$

In this case the eigen-system is given by

$$\begin{cases} \frac{\partial}{\partial y} g(\sigma,y) + y^2 g(\sigma,y) + \lambda g(\sigma,y) = \frac{3}{2\Lambda}\beta\sigma k y^2 e^{-\frac{y^3}{3}}, \\ g(\sigma,0) = \frac{3}{2\Lambda}\beta\sigma k + \int_0^\infty y^2 g(-\sigma,y)dy, \\ \int_0^\infty [g(\sigma,y)+g(-\sigma,y)]dy = 0, \end{cases} \tag{3.37}$$

where $\Lambda = \int_0^\infty e^{-\frac{y^3}{3}} = \frac{\Gamma(\frac{1}{3})}{3^{2/3}}$ and $k = 2\int_0^\infty g(1,y)dy$.

Proposition 3.6. *The solutions in* $\lambda \in \mathbb{C}$ *to* (3.37) *are the zeros of the holomorphic function*

$$\begin{aligned} H_{\beta,2}(\lambda) := H_2(\lambda)\Bigg[&12\beta - \lambda^4\Lambda + 6\beta\lambda\Lambda - 6\beta\lambda 3^{1/3}\Gamma(4/3) \\ &+ 3\beta\lambda^2 3^{2/3}\Gamma(5/3) - 6\beta\lambda^2\frac{\Gamma(2/3)}{3^{1/3}}\Bigg] + \\ &\left[2\Lambda\lambda^3 - 12\beta\Lambda + 12\beta\frac{\Gamma(2/3)}{3^{1/3}}\lambda - 6\beta\lambda^2\right], \end{aligned} \tag{3.38}$$

with

$$H_2(\lambda) := \int_0^\infty e^{-\lambda y}e^{-\frac{y^3}{3}}dy. \tag{3.39}$$

Moreover, it holds

$$H_2(\lambda) = \sum_{n=0}^\infty (-1)^n \frac{\lambda^n}{n!} 3^{\frac{1}{3}(n-2)}\Gamma\left(\frac{n+1}{3}\right). \tag{3.40}$$

Proof. We proceed as in the previous case, by setting

$$h(\sigma,y) = g(\sigma,y)e^{\frac{y^3}{3}},$$

so that

$$\frac{\partial}{\partial y}h(\sigma,y) = -\lambda h(\sigma,y) + \frac{3}{2\Lambda}\beta\sigma k y^2.$$

Thus,

$$h(\sigma,y) = e^{-\lambda y}\left[h(\sigma,0) + \frac{3\beta\sigma k}{2\Lambda}\int_0^y u^2 e^{\lambda u}du\right].$$

Since $\int_0^y u^2 e^{\lambda u}du = \frac{1}{\lambda^3}[-2+e^{\lambda y}(2+\lambda y(-2+\lambda y))]$, we can write

$$g(\sigma,y) = e^{-\frac{y^3}{3}}e^{-\lambda y}\left[g(\sigma,0) + \frac{3\beta\sigma k}{2\Lambda}\frac{1}{\lambda^3}\left(-2+2e^{\lambda y} - 2\lambda y e^{\lambda y} + \lambda^2 y^2 e^{\lambda y}\right)\right]. \tag{3.41}$$

Recalling notation (3.39), we compute

$$\begin{aligned} k = 2\int_0^\infty g(1,y)dy &= 2H_2(\lambda)g(1,0) - 2H_2(\lambda)\frac{3\beta k}{\Lambda\lambda^3} + 2\frac{3\beta k}{\Lambda\lambda^3}\int_0^\infty e^{-\frac{y^3}{3}}dy \\ &- 2\frac{3\beta k}{\Lambda\lambda^2}\int_0^\infty y e^{-\frac{y^3}{3}}dy + \frac{3\beta k}{\Lambda\lambda}\int_0^\infty y^2 e^{-\frac{y^3}{3}} \\ &= 2H_2(\lambda)g(1,0) - 2H_2(\lambda)\frac{3\beta k}{\Lambda\lambda^3} + 2\frac{3\beta k}{\lambda^3} - 2\frac{3\beta k}{\Lambda\lambda^2}\frac{\Gamma(2/3)}{3^{1/3}} + \frac{3\beta k}{\Lambda\lambda}, \end{aligned}$$

which gives

$$k = \frac{2g(1,0)H_2(\lambda)}{1 + 2H_2(\lambda)\frac{3\beta}{\Lambda\lambda^3} - 2\frac{3\beta}{\lambda^3} + 2\frac{3\beta}{\Lambda\lambda^2}\frac{\Gamma(2/3)}{3^{1/3}} - \frac{3\beta}{\Lambda\lambda}}. \tag{3.42}$$

As before, the condition $\int_0^\infty [g(\sigma,y) + g(-\sigma,y)]dy = 0$ in (3.37) is equivalent to $g(\sigma,y) + g(-\sigma,y) = 0$ for every $y \in \mathbb{R}^+$ because of (3.41). Using this observation for $y = 0$ in the other integral condition, we compute

$$\begin{aligned} g(\sigma,0) = \frac{3\beta\sigma k}{2\Lambda} + \int_0^\infty y^2 \Bigg[e^{-\frac{y^3}{3}} e^{-\lambda y} \Bigg(- g(\sigma,0) \\ - \frac{3\beta\sigma k}{2\Lambda}\frac{1}{\lambda^3}(-2 + 2e^{\lambda y} - 2\lambda y e^{\lambda y} + \lambda^2 y^2 e^{\lambda y}) \Bigg) \Bigg] dy. \end{aligned}$$

Observing that, by integration by parts,

$$\int_0^\infty y^2 e^{-\frac{y^3}{3}} e^{-\lambda y} dy = 1 - \lambda H_2(\lambda),$$

we find

$$\begin{aligned} g(\sigma,0) = \frac{3\beta\sigma k}{2\Lambda} - (1 - \lambda H_2(\lambda))g(\sigma,0) + \frac{3\beta\sigma k}{\Lambda\lambda^3}(1 - \lambda H_2(\lambda)) - \frac{3\beta\sigma k}{\Lambda\lambda^3} \\ + \frac{3\beta\sigma k}{\Lambda\lambda^2} 3^{1/3}\Gamma(4/3) - \frac{3\beta\sigma k}{2\Lambda\lambda} 3^{2/3}\Gamma(5/3). \end{aligned}$$

Computing in $\sigma = 1$ and grouping for $g(1,0)$,

$$\begin{aligned} g(1,0)[2 - \lambda H_2(\lambda)] &= k \left[\frac{3\beta}{2\Lambda} + \frac{3\beta}{\Lambda\lambda^3}(1 - \lambda H_2(\lambda)) - \frac{3\beta}{\Lambda\lambda^3} \right. \\ &\quad \left. + \frac{3\beta}{\Lambda\lambda^2} 3^{1/3}\Gamma(4/3) - \frac{3\beta}{2\Lambda\lambda} 3^{2/3}\Gamma(5/3) \right] \\ &= k \left[\frac{3\beta}{2\Lambda} - \frac{3\beta}{\Lambda\lambda^2} H_2(\lambda) + \frac{3\beta}{\Lambda\lambda^2} 3^{1/3}\Gamma(4/3) - \frac{3\beta}{2\Lambda\lambda} 3^{2/3}\Gamma(5/3) \right]. \end{aligned}$$

Plugging expression (3.42) for k,

$$\begin{aligned} 2 - \lambda H_2(\lambda) = \frac{2H_2(\lambda)}{1 + 2H_2(\lambda)\frac{3\beta}{\Lambda\lambda^3} - 2\frac{3\beta}{\lambda^3} + 2\frac{3\beta}{\Lambda\lambda^2}\frac{\Gamma(2/3)}{3^{1/3}} - \frac{3\beta}{\Lambda\lambda}} \left[\frac{3}{2\Lambda}\beta - \frac{3\beta}{\Lambda\lambda^2} H_2(\lambda) \right. \\ \left. + \frac{3\beta}{\Lambda\lambda^2} 3^{1/3}\Gamma(4/3) - \frac{3\beta}{2\Lambda\lambda} 3^{2/3}\Gamma(5/3) \right]. \end{aligned}$$

This gives

$$2 - \lambda H_2(\lambda) = \frac{2\lambda H_2(\lambda)\left[\frac{3}{2}\beta\lambda^2 - 3\beta H_2(\lambda) + 3\beta 3^{1/3}\Gamma(4/3) - \frac{3}{2}\beta\lambda 3^{2/3}\Gamma(5/3)\right]}{\Lambda\lambda^3 + 6\beta H_2(\lambda) - 6\beta\Lambda + 6\beta\lambda\frac{\Gamma(2/3)}{3^{1/3}} - 3\beta\lambda^2},$$

which is equivalent to

$$\begin{aligned} &2\Lambda\lambda^3 + 12\beta H_2(\lambda) - 12\beta\Lambda + 12\beta\lambda\frac{\Gamma(2/3)}{3^{1/3}} - 6\beta\lambda^2 - \lambda^4\Lambda H_2(\lambda) \\ &\quad + 6\beta\lambda H_2(\lambda)\Lambda - 6\beta\lambda^2\frac{\Gamma(2/3)}{3^{1/3}} H_2(\lambda) = 6\beta\lambda H_2(\lambda) 3^{1/3}\Gamma(4/3) - 3\beta\lambda^2 H_2(\lambda) 3^{2/3}\Gamma(5/3). \end{aligned}$$

As a polynomial in λ, this is

$$-\Lambda H_2(\lambda)\lambda^4 + 2\Lambda\lambda^3 + \lambda^2\left[-6\beta - 6\beta\frac{\Gamma(2/3)}{3^{1/3}}H_2(\lambda) + 3\beta H_2(\lambda)3^{2/3}\Gamma(5/3)\right]$$
$$+\lambda\left[6\beta\Lambda H_2(\lambda) - 6\beta H_2(\lambda)3^{1/3}\Gamma(4/3) + 12\beta\frac{\Gamma(2/3)}{3^{1/3}}\right] + 12\beta H_2(\lambda) - 12\beta\Lambda = 0.$$

Equivalently, in terms of $H_2(\lambda)$ we have

$$H_2(\lambda)[12\beta - \lambda^4\Lambda + 6\beta\lambda\Lambda - 6\beta\lambda 3^{1/3}\Gamma(4/3) + 3\beta\lambda^2 3^{2/3}\Gamma(5/3) - 6\beta\lambda^2\frac{\Gamma(2/3)}{3^{1/3}}]$$
$$+\left[2\Lambda\lambda^3 - 12\beta\Lambda + 12\beta\frac{\Gamma(2/3)}{3^{1/3}}\lambda - 6\beta\lambda^2\right] = 0,$$

i.e. the zeros of $H_{\beta,2}(\lambda)$ in (3.38), provided we show the validity of expression (3.40) for $H_2(\lambda)$. As defined in (3.39), $H_2(\lambda)$ is a holomorphic function on $\mathbb{C}$, which can be expressed in series as

$$H_2(\lambda) = \int_0^\infty e^{-\lambda y}e^{-\frac{y^3}{3}}dy = \sum_{n=0}^\infty(-1)^n\frac{\lambda^n}{n!}\int_0^\infty y^n e^{-\frac{y^3}{3}}dy$$
$$= \sum_{n=0}^\infty(-1)^n\frac{\lambda^n}{n!}3^{\frac{1}{3}(n-2)}\Gamma\left(\frac{n+1}{3}\right),$$

which is expression (3.40), where we have used the known formula for

$$\int_0^\infty y^n e^{-\frac{y^3}{3}}dy = 3^{\frac{1}{3}(n-2)}\Gamma\left(\frac{n+1}{3}\right).$$

□

3.2.2.4 Numerical evidence on the eigenvalues

We studied numerically the two eigenvalues' equations

$$H_{\beta,1}(\lambda) = 0, \tag{3.43}$$

and

$$H_{\beta,2}(\lambda) = 0. \tag{3.44}$$

We used a numerical root finding built-in function of the software `Mathematica`, specifically `FindRoot`, starting the search from different initial points of the complex plane and from different values of β. Here we report the results:

- Case $\gamma = 1$:

  (1.1) we find two conjugate purely imaginary solutions to (3.43), for

  $$\lambda = \pm\lambda_c(1) := \pm i(1.17055)$$

  and

  $$\beta = \beta_c(1) := 0.768834; \tag{3.45}$$

  (1.2) iterating the search around $(\beta_c(1), \lambda_c(1))$, the resulting complex eigenvalue goes from having a negative real part for $\beta < \beta_c(1)$ to a positive real part for $\beta > \beta_c(1)$;

(1.3) no other purely immaginary solution $\lambda = \pm ix$ is found for $0 \le x \le 500$ and $0 \le \beta \le 20$;

(1.4) for $\beta < \beta_c(1)$ all the eigenvalues $\lambda = ix + y$ are such that $y < 0$. This was verified for $-100 \le x \le 100$, $-100 \le y \le 100$.

- Case $\gamma = 2$:

(2.1) we find two conjugate purely imaginary solutions to (3.44), for

$$\lambda = \pm\lambda_c(2) := \pm i(1.97765)$$

and

$$\beta = \beta_c(2) := 0.362275; \tag{3.46}$$

(2.2) analogous to (1.2);

(2.3) analogous to (1.3), verified for $0 \le x \le 10$ and $0 \le \beta \le 5$;

(2.4) analogous to (1.4), verified for $-25 \le x \le 25$, $-25 \le y \le 25$;

(2.5) apart from being sensibly slower, the numerical root finding for $\gamma = 2$ suffers from numerical instability issues. This is why we were able to check the results for much smaller intervals in this case.

3.2.3 Finite particle system simulations

We made several simulations of the particle system (N large but finite, $N = 1500$) for $\gamma = 1, 2$, which seem in accordance with the above numerical results on the eigenvalues (compare with (3.45) and (3.46)). This is a description of the evidences:

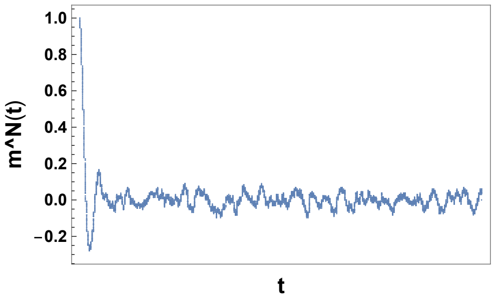
- For β small the system is stable, in particular the magnetization goes to zero regardless of the initial datum (Figure 3.1).

- There is a critical β (around 0.75 for $\gamma = 1$, 0.35 for $\gamma = 2$) above which the magnetization starts oscillating. Close to the critical points oscillations (Figure 3.2) do not look very regular (corrupted by noise?), but they soon become very regular if β is not too close to the critical value. We also made joint plots of the magnetization with the empirical mean of the y_i's (Figure 3.3). A limit cycle seems to emerge.

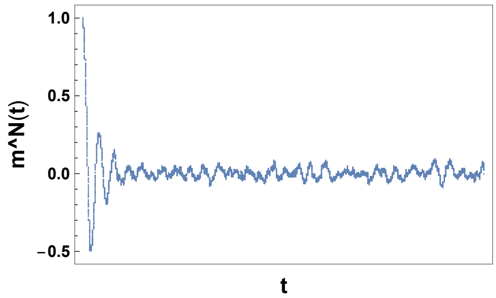
- As β increases, the amplitude of the oscillation of the magnetization increases (Figure 3.4), while the period looks nearly constant. As β crosses another critical value (around 1.3 for $\gamma = 1$, 1.65 for $\gamma = 2$) oscillations disappear, and the sistem magnetizes, i.e. the magnetization stabilizes to a non-zero value, actually close to ± 1 (Figure 3.5).

- The oscillations are lasting for a wider interval of β's for $\gamma = 2$ (from $\beta \approx 0.35$ until $\beta \approx 1.65$) than $\gamma = 1$ (from $\beta \approx 0.75$ until $\beta \approx 1.3$). The period is instead smaller for $\gamma = 2$ than for $\gamma = 1$.

- For both $\gamma = 1, 2$, the appearance of the oscillations does not seem to depend on the initial data for the dynamics, suggesting the possible presence of a *global* Hopf bifurcation.

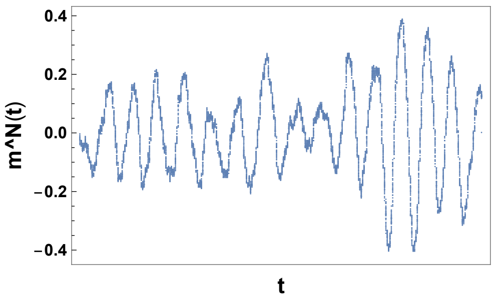


(a) $\beta = 0.25$, $\gamma = 1$

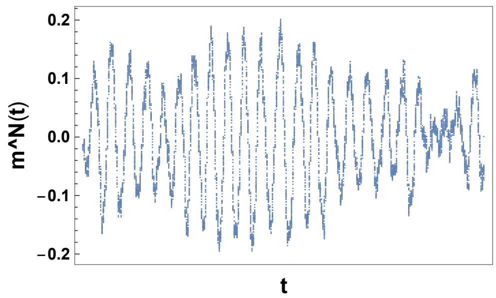


(b) $\beta = 0.1$, $\gamma = 2$

Figure 3.1: Simulation of the finite particle system's dynamics for $\gamma = 1$ (left) and $\gamma = 2$ (right), with number of spins $N = 1500$. We plot the empirical magnetization, with initial data $\sigma_i(0) = 1$ for every $i = 1, \ldots, N$.

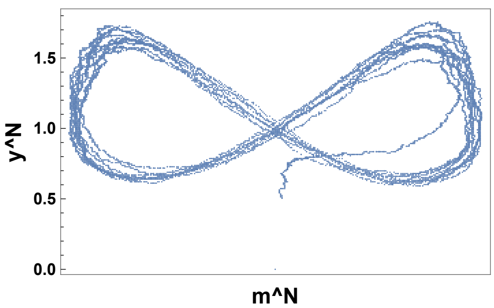


(a) $\beta = 0.75$, $\gamma = 1$

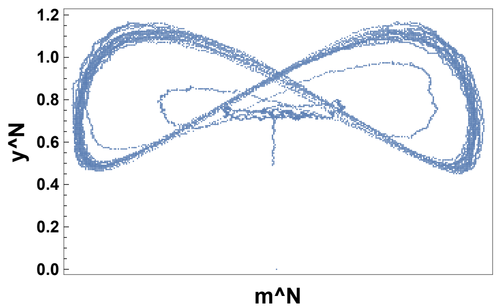


(b) $\beta = 0.35$, $\gamma = 2$

Figure 3.2: Simulation of the finite particle system's dynamics for $\gamma = 1$ (left) and $\gamma = 2$ (right), with number of spins $N = 1500$. We plot the empirical magnetization.

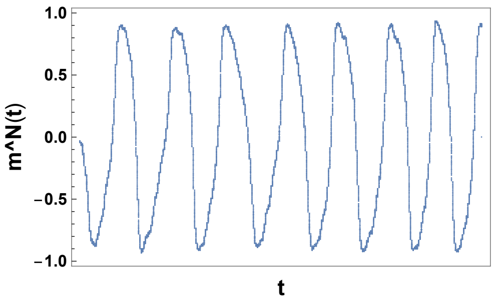


(a) $\beta = 1.1$, $\gamma = 1$

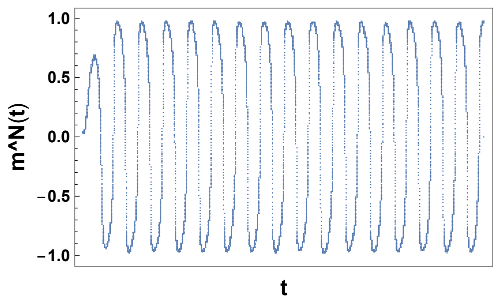


(b) $\beta = 0.6$, $\gamma = 2$

Figure 3.3: Simulation of the finite particle system's dynamics for $\gamma = 1$ (left) and $\gamma = 2$ (right), with number of spins $N = 1500$. We plot the empirical magnetization of the spins against the empirical mean of the y_i's.

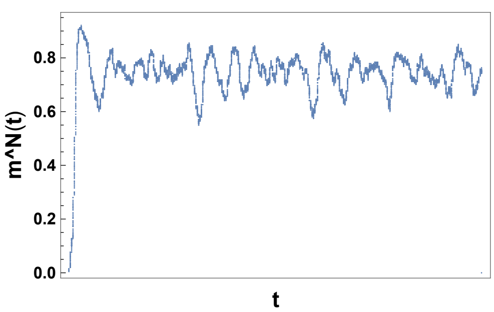


(a) $\beta = 1.2$, $\gamma = 1$

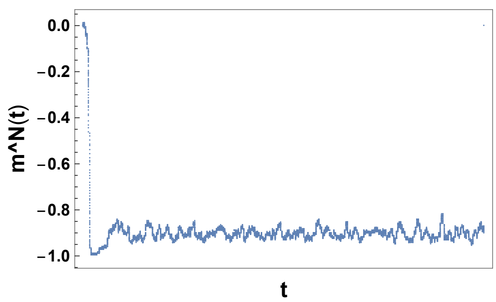


(b) $\beta = 0.9$, $\gamma = 2$

Figure 3.4: Simulation of the finite particle system's dynamics for $\gamma = 1$ (left) and $\gamma = 2$ (right), with number of spins $N = 1500$. We plot the empirical magnetization.



(a) $\beta = 1.3$, $\gamma = 1$



(b) $\beta = 1.65$, $\gamma = 2$

Figure 3.5: Simulation of the finite particle system's dynamics for $\gamma = 1$ (left) and $\gamma = 2$ (right), with number of spins $N = 1500$. We plot the empirical magnetization.

3.3 Oscillatory behavior in a model of mean field interacting spins with frustration

This section is devoted to the study of a finite-dimensional non-Markovian modification of the Curie–Weiss spin-flip dynamics which belongs to the same universality class of the model considered in Section 3.2, as we already remarked in the beginning of the chapter. However, it must be noted that the properties of reversibility of the previous model (Remark 3.3) are not holding here. Indeed, the periodic behavior is here fostered by the presence of frustration in the dynamics, which brings time-irreversibility into play. Nevertheless, the finite dimensionality of the model allows one to characterize the phase diagram to a much larger extent than in the previous case. In particular, even though we do not prove any global result, we give strong numerical evidence on the way in which cycles disappear above a certain threshold of the parameters' values. We mention the work [97], which looks related to our model. Besides performing a much more general analysis (temporal delay in the transmission of information and quenched disorder are also considered), in that case the population consists of two types of individuals: mainstreams, following the majority, and hipsters. In the author's terminology, our toy model could then be referred to that of a single population where each individual has both a conformist and an anti-conformist side in his/her personality.

In the formulation of the model, the state of the i-th particle in the system is identified by a pair of spin-valued variables $(x_i, y_i) \in \{-1,1\}^2$. The dynamics is given in terms of a continuous time spin-flip type Markov chain on the augmented state space $\{-1,1\}^{2N}$, where each particle flips one component of its state independently conditioned on the current state of the population, with rates

$$\begin{cases} x_i \to -x_i & \text{with rate} \quad (1-\varepsilon x_i y_i) e^{-\beta x_i m_x^N}, \\ y_i \to -y_i & \text{with rate} \quad e^{\gamma y_i m_x^N}, \end{cases} \tag{3.47}$$

where $\gamma, \beta \geq 0$, $0 \leq \varepsilon \leq 1$, and $m_x^N := \frac{1}{N}\sum_{i=1}^N x_i$ is the magnetization of the spins x_i's. Note that, if $\varepsilon = 0$, and if we restrict the model to the spins x_i's we recover the Curie–Weiss model (3.4). Dynamics (3.47) can be thought of as a perturbation of the Curie–Weiss (the strength of the perturbation being governed by the parameter ε), under the additional presence of frustration, which we here introduced through the variables y_i's, whose tendency is to disalign with the state of the majority of the x_i's. The strength of the alignment and disalignment tendencies is governed by the two parameters β and γ respectively. Moreover, when the private states of an individual are aligned, the interaction with the rest of the population is mitigated (the intensity of this phenomenon being tuned by ε). Denoting $(\boldsymbol{x}, \boldsymbol{y}) := (x_1, \dots, x_N, y_1, \dots, y_N)$, $\boldsymbol{x}^i := (x_1, \dots, x_{i-1}, -x_i, x_{i+1}, \dots, x_N)$ and $\boldsymbol{y}^i := (y_1, \dots, y_{i-1}, -y_i, y_{i+1}, \dots, y_N)$, the infinitesimal generator associated to dynamics (3.47) is, for $f : \{-1,1\}^{2N} \to \mathbb{R}$,

$$\begin{aligned} L^N f(\boldsymbol{x}, \boldsymbol{y}) = &\sum_{i=1}^N (1-\varepsilon x_i y_i) e^{-\beta x_i m_x^N} \left[f(\boldsymbol{x}^i, \boldsymbol{y}) - f(\boldsymbol{x}, \boldsymbol{y}) \right] \\ &+ \sum_{i=1}^N e^{\gamma y_i m_x^N} \left[f(\boldsymbol{x}, \boldsymbol{y}^i) - f(\boldsymbol{x}, \boldsymbol{y}) \right]. \end{aligned} \tag{3.48}$$

The rest of the chapter is organized as follows: in Section 3.3.1 we introduce the macroscopic limit system, proving its well-posedness; in Section 3.3.2 we perform a linear analysis around the disordered equilibrium, studying the local phase-diagram when the interaction parameters vary; in Section 3.3.3 we find numerically all the other equilibria of the system; in Section 3.3.4 we study numerically the local character of the previously found equilibria; finally, in Section 3.3.5 we give detailed illustrations of the dynamics and of the global phase-diagram, via numerical simulations of the macroscopic equations and resorting to the previous analyses.

3.3.1 Macroscopic limit

In addition to the empirical magnetization m_x^N of the x_i's, we also define the analogous quantity m_y^N for the y_i's and

$$m_{xy}^N := \frac{1}{N}\sum_{i=1}^N x_i y_i.$$

We denote $\left((m_x^N(t), m_y^N(t), m_{xy}^N(t))\right)_{t\geq 0}$ the process obtained by evaluating the empirical magnetization's functions on the particles dynamics. Let $E_N \subseteq [-1,1]^3$ be the image of $\{-1,1\}^{2N}$ under the map $\Phi^N : (\boldsymbol{x}, \boldsymbol{y}) \mapsto (m_x^N, m_y^N, m_{xy}^N)$. It is easy to check that dynamics (3.47) induces a Markovian evolution on E_N for the process $\left((m_x^N(t), m_y^N(t), m_{xy}^N(t))\right)_{t\geq 0}$. Indeed, generator (3.48) is closed in the three empirical magnetizations variables, as we verify by applying it separately to the three functions (m_x^N, m_y^N, m_{xy}^N):

$$\begin{aligned}
L^N m_x^N &= \sum_{i=1}^N (1-\varepsilon x_i y_i) e^{-\beta x_i m_x^N}\left[-\frac{2}{N}x_i\right] \\
&= -\frac{2}{N}\sum_{i=1}^N x_i e^{-\beta x_i m_x^N} + \frac{2\varepsilon}{N}\sum_{i=1}^N y_i e^{-\beta x_i m_x^N} \\
&= -\frac{2}{N}\sum_{i=1}^N \frac{1+x_i}{2} e^{-\beta m_x^N} + \frac{2}{N}\sum_{i=1}^N \frac{1-x_i}{2} e^{\beta m_x^N} \\
&+ \frac{2\varepsilon}{N}\sum_{i=1}^N \frac{1+x_i}{2} y_i e^{-\beta m_x^N} + \frac{2\varepsilon}{N}\sum_{i=1}^N y_i \frac{1-x_i}{2} e^{\beta m_x^N} \\
&= -e^{-\beta m_x^N} - m_x^N e^{-\beta m_x^N} + e^{\beta m_x^N} - m_x^N e^{\beta m_x^N} + \varepsilon m_y^N e^{-\beta m_x^N} \\
&+ \varepsilon m_{xy}^N e^{-\beta m_x^N} + \varepsilon m_y^N e^{\beta m_x^N} + \varepsilon m_{xy}^N e^{\beta m_x^N} \\
&= -2 m_x^N \cosh(\beta m_x^N) + 2\sinh(\beta m_x^N) + 2\varepsilon m_y^N \cosh(\beta m_x^N) - 2\varepsilon m_{xy}^N \sinh(\beta m_x^N).
\end{aligned}$$

For m_y^N we get

$$\begin{aligned}
L^N m_y^N &= \sum_{i=1}^N e^{\gamma y_i m_x^N}\left[-\frac{2}{N}y_i\right] \\
&= -2\sinh(\gamma m_x^N) - 2 m_y^N \cosh(\gamma m_x^N),
\end{aligned}$$

while for m_{xy}^N,

$$\begin{aligned}
L^N m_{xy}^N &= \sum_{i=1}^N (1-\varepsilon x_i y_i) e^{-\beta m_x^N x_i}\left[-\frac{2}{N}x_i y_i\right] + \sum_{i=1}^N e^{\gamma y_i m_x^N}\left[-\frac{2}{N}x_i y_i\right] \\
&= -\frac{2}{N}\sum_{i=1}^N \frac{1+x_i}{2} y_i e^{-\beta m_x^N} + \frac{2}{N}\sum_{i=1}^N \frac{1-x_i}{2} y_i e^{\beta m_x^N}
\end{aligned}$$

$$
\begin{aligned}
&+\frac{2\varepsilon}{N}\sum_{i=1}^{N}\frac{1+x_i}{2}e^{-\beta m_x^N}+\frac{2\varepsilon}{N}\sum_{i=1}^{N}\frac{1-x_i}{2}e^{\beta m_x^N}\\
&-\frac{2}{N}\sum_{i=1}^{N}x_i\frac{1+y_i}{2}e^{\gamma m_x^N}+\frac{2}{N}\sum_{i=1}^{N}x_i\frac{1-y_i}{2}e^{-\gamma m_x^N}\\
&=2m_y^N\sinh(\beta m_x^N)-2m_{xy}^N\cosh(\beta m_x^N)+2\varepsilon\cosh(\beta m_x^N)\\
&-2\varepsilon m_x^N\sinh(\beta m_x^N)-2m_x^N\sinh(\gamma m_x^N)-2m_{xy}^N\cosh(\gamma m_x^N).
\end{aligned}
$$

The process $\big((m_x^N(t), m_y^N(t), m_{xy}^N(t))\big)_{t\geq 0}$ is an *order parameter*, in the sense that its evolution is Markovian. The associated infinitesimal generator can be derived by applying the change of variables prescribed by the map Φ^N. In this setting, the proof of a propagation of chaos property for the sequence of the order parameters should be standard, by studying the uniform convergence of the generators and by applying the results in [56] to obtain (weak) convergence to a limiting deterministic process. We do not treat this problem here, but we rather focus on studying the limit deterministic process $\big((x(t), y(t), w(t))\big)_{t\geq 0}$ to which $\big((m_x^N(t), m_y^N(t), m_{xy}^N(t))\big)_{t\geq 0}$ should converge. By the above computations on L^N, it necessarily satisfies

$$
\begin{cases}
\dot{x}(t)=-2x(t)\cosh(\beta x(t))+2\sinh(\beta x(t))+2\varepsilon y(t)\cosh(\beta x(t))-2\varepsilon w(t)\sinh(\beta x(t)),\\
\dot{y}(t)=-2y(t)\cosh(\gamma x(t))-2\sinh(\gamma x(t)),\\
\dot{w}(t)=-2w(t)\cosh(\gamma x(t))-2x(t)\sinh(\gamma x(t))-2w(t)\cosh(\beta x(t))+2y(t)\sinh(\beta x(t))\\
\qquad\qquad +2\varepsilon\cosh(\beta x(t))-2\varepsilon x(t)\sinh(\beta x(t)),\\
x(0)=x_0,\quad y(0)=y_0,\quad w(0)=w_0.
\end{cases}
\tag{3.49}
$$

Proposition 3.7 (Well-posedeness)**.** *For any $\gamma, \beta \geq 0$, $0 \leq \varepsilon \leq 1$, System* (3.49) *has a unique global solution such that $(x(t), y(t), w(t)) \in [-1, 1]^3$ for any $t \geq 0$, provided $(x_0, y_0, w_0) \in [-1, 1]^3$ is such that, if $(x_0, y_0) = (\pm 1, \mp 1)$, then $w_0 = -1$ and if $(x_0, y_0) = (\pm 1, \pm 1)$, then $w_0 = 1$.*

Proof. We study the sign of the vector field at the borders of $(-1, 1)^3$ for each variable separately. First of all, we note that, by the vector field of the second equation, if $y(t) = 1$ (resp. $y(t) = -1$), then $\frac{d}{dt}y(t) \leq 0$ (resp. $\frac{d}{dt}y(t) \geq 0$), for any choice of the parameters. By looking at the first equation, we have that, for $x(t) = 1$,

$$
\frac{d}{dt}x(t)\Big|_{x(t)=1}=-2\cosh\beta+2\sinh\beta+2\varepsilon y(t)\cosh\beta-2\varepsilon w(t)\sinh\beta.
$$

When $y(t) = 1$, we must have $w(t) = 1$, due to the assumption on the initial datum, the definition of $m_{xy}^N(t)$ (for which the property is satisfied) and the supposed convergence $(m_x^N(t), m_y^N(t), m_{xy}^N(t)) \to (x(t), y(t), w(t))$. Thus, when $y(t) = 1$

$$
\begin{aligned}
\frac{d}{dt}x(t)\Big|_{x(t)=1}&=-2\cosh\beta+2\sinh\beta+2\varepsilon\cosh\beta-2\varepsilon\sinh\beta\\
&=-2(1-\varepsilon)\cosh\beta+2(1-\varepsilon)\sinh\beta\leq 0,
\end{aligned}
$$

for any $0 \leq \varepsilon \leq 1$. Similarly, when $y(t) = -1$ we have $w(t) = -1$, and thus

$$
\frac{d}{dt}x(t)\Big|_{x(t)=1}=-2(1+\varepsilon)\cosh\beta+2(1+\varepsilon)\sinh\beta\leq 0,
$$

for any $\varepsilon \geq 0$.

For $x(t) = -1$,

$$\frac{d}{dt}x(t)\Big|_{x(t)=-1} = 2\cosh\beta - 2\sinh\beta + 2\varepsilon y(t)\cosh\beta + 2\varepsilon w(t)\sinh\beta.$$

When $y(t) = -1$ (thus $w(t) = 1$), we have

$$\begin{aligned}\frac{d}{dt}x(t)\Big|_{x(t)=-1} &= 2\cosh\beta - 2\sinh\beta - 2\varepsilon\cosh\beta + 2\varepsilon\sinh\beta \\ &= 2(1-\varepsilon)\cosh\beta - 2(1-\varepsilon)\sinh\beta \geq 0,\end{aligned}$$

for any $0 \leq \varepsilon \leq 1$, while for $y(t) = 1$ (and $w(t) = -1$),

$$\begin{aligned}\frac{d}{dt}x(t)\Big|_{x(t)=-1} &= 2\cosh\beta - 2\sinh\beta + 2\varepsilon\cosh\beta - 2\varepsilon\sinh\beta \\ &= 2(1+\varepsilon)\cosh\beta - 2(1+\varepsilon)\sinh\beta \geq 0,\end{aligned}$$

for any $\varepsilon \geq 0$. With analogous considerations on the third equation we get the assertion for $w(t)$. Indeed, note that the terms of the vector field of $w(t)$ which depend on γ have always the right signs at the borders.

For the well-posedeness, by the Lipschitz properties of the vector field in $[-1,1]^3$, we can conclude by the classic theorems of global existence and uniqueness for ODEs. □

Remark 3.8. *We refer to System* (3.49) *as the macroscopic limit of the particle system* (3.47)*. Observe that the disordered state* $\left(0,0,\frac{\varepsilon}{2}\right)$ *is an equilibrium for every choice of the parameters* β *and* γ*. In the following we always assume* $\varepsilon \leq 1$*, which is not restrictive for the resulting phase diagram.*

3.3.2 Local analysis

In this section we perform a local analysis by linearizing the dynamics around the disordered equilibrium. Denoting with $\overrightarrow{f}(x,y,w)$ the vector field of System (3.49), the linearized system around $\left(0,0,\frac{\varepsilon}{2}\right)$ is given by

$$\frac{d}{dt}\begin{pmatrix}\tilde{x}(t)\\ \tilde{y}(t)\\ \tilde{w}(t)\end{pmatrix} = A\begin{pmatrix}\tilde{x}(t)\\ \tilde{y}(t)\\ \tilde{w}(t)\end{pmatrix}, \tag{3.50}$$

with $A = D\overrightarrow{f}(x,y,w)|_{(x,y,w)=\left(0,0,\frac{\varepsilon}{2}\right)}$,

$$A = \begin{pmatrix}-2+2\beta-\varepsilon^2\beta & 2\varepsilon & 0\\ -2\gamma & -2 & 0\\ 0 & 0 & -4\end{pmatrix}.$$

With this notation, we are able to prove the following:

Proposition 3.9. *In a neighborhood of* $\left(0,0,\frac{\varepsilon}{2}\right)$*, System* (3.49) *possesses:*

- *a single pitchfork bifurcation in* $\beta_c^* := \frac{2(\varepsilon\gamma+1)}{2-\varepsilon^2}$ *for* $\varepsilon\gamma \leq 1$*;*
- *when* $\varepsilon\gamma > 1$*, the model features:*

1. *a Hopf bifurcation in* $\beta_c^{**} := \frac{4}{2-\varepsilon^2}$;
2. *a pitchfork bifurcation in* $\beta_c^* = \frac{2(\varepsilon\gamma+1)}{2-\varepsilon^2}$.

Proof. Denoting with A' the upper left submatrix of A,

$$A' = \begin{pmatrix} -2+2\beta-\varepsilon^2\beta & 2\varepsilon \\ -2\gamma & -2 \end{pmatrix},$$

we have that the characteristic polynomial of A is

$$P(\lambda) := (-4-\lambda)\left[\lambda^2 - \mathrm{Tr}(A')\lambda + \Delta_{A'}\right], \tag{3.51}$$

with $\mathrm{Tr}(A') = -4+2\beta-\varepsilon^2\beta$ and $\Delta_{A'} = 4-4\beta+2\varepsilon^2\beta+4\varepsilon\gamma$. We know that the disordered state is linearly stable if $\mathrm{Tr}(A') < 0$ and $\Delta_{A'} > 0$, which is equivalent to $\beta < \frac{4}{2-\varepsilon^2}$ and $\beta < \frac{2(\varepsilon\gamma+1)}{2-\varepsilon^2}$. When $\varepsilon\gamma \le 1$, this is equivalent to $\beta < \frac{2(\varepsilon\gamma+1)}{2-\varepsilon^2}$, since $\frac{2(\varepsilon\gamma+1)}{2-\varepsilon^2} \le \frac{4}{2-\varepsilon^2}$, while for $\varepsilon\gamma > 1$ we have a linearly stable disordered state for $\beta < \frac{4}{2-\varepsilon^2}$.

A pitchfork bifurcation arises, associated with a change of stability of the fixed point $(0,0,\varepsilon/2)$, when $\Delta_{A'} = 0$ and $\mathrm{Tr}(A') \le 0$. The first condition is true for $\beta = \beta_c^* = \frac{2(\varepsilon\gamma+1)}{2-\varepsilon^2}$, while the second holds for $\beta \le \frac{4}{2-\varepsilon^2}$. In order for them to be true at the same time we need to have $\frac{2(\varepsilon\gamma+1)}{2-\varepsilon^2} \le \frac{4}{2-\varepsilon^2}$, which is equivalent to $\varepsilon\gamma \le 1$.

Another pitchfork bifurcation is present, this time not associated with a change of stability of the fixed point, when $\Delta_{A'} = 0$ and $\mathrm{Tr}(A') > 0$, which is equivalent to $\beta > \frac{4}{2-\varepsilon^2}$ and $\beta = \beta_c^* = \frac{2(\varepsilon\gamma+1)}{2-\varepsilon^2}$, which is possible if and only if $\varepsilon\gamma > 1$.

Finally, the condition for having a Hopf bifurcation is $\mathrm{Tr}(A') = 0$ and $\Delta_{A'} > 0$, which is true for $\beta = \beta_c^{**} = \frac{4}{2-\varepsilon^2}$ and $\varepsilon\gamma > 1$. □

In the Hopf bifurcation case we can go a step further by computing the Lyapunov coefficient around the fixed point to deduce that

Proposition 3.10. *When $\varepsilon\gamma > 1$, a unique stable cycle emerges after $\beta_c^{**} = \frac{4}{2-\varepsilon^2}$, i.e. the Hopf bifurcation is supercritical.*

Proof (non-rigorous). First, we note that locally around the fixed point $(0,0,\varepsilon/2)$ the dynamics is two-dimensional, since the linear dynamics is such that $\frac{d}{dt}\tilde{w}(t) = -4\tilde{w}(t)$. Thus, in order to compute the normal form, we restrict System (3.49) to the first two equations, and we substitute $w(t) \equiv \frac{\varepsilon}{2}$ (this linear approximation step can be made rigorous by applying the methodology of the center manifold detailed in Theorem 1 of [89, Sect. 2.12]). We then obtain

$$\begin{cases} \dot{x}(t) = -2x(t)\cosh(\beta x(t)) + 2\sinh(\beta x(t)) + 2\varepsilon y(t)\cosh(\beta x(t)) - \varepsilon^2\sinh(\beta x(t)) \\ \dot{y}(t) = -2y(t)\cosh(\gamma x(t)) - 2\sinh(\gamma x(t)). \end{cases} \tag{3.52}$$

Expanding up to the third order around $(0,0)$, we get the topologically equivalent system

$$\begin{cases} \dot{x}(t) = (-2+2\beta-\varepsilon^2\beta)x(t) + 2\varepsilon y(t) - \beta^2 x^3(t) + \frac{\beta^3}{3}x^3(t) + \varepsilon\beta^2 y(t)x^2(t) - \varepsilon^2\frac{\beta^3}{6}x^3(t) \\ \dot{y}(t) = -2\gamma x(t) - 2y(t) - \gamma^2 x^2(t)y(t) - \frac{\gamma^3}{3}x^3(t). \end{cases} \tag{3.53}$$

Applying Remark 1 in [89, Sect. 4.4], the computation of the Lyapunov coefficient σ in our case gives:

$$\sigma = \frac{3\pi}{2\Delta_{A'}^{3/2}}(4-4\varepsilon\gamma)(\beta^2(1+\beta+2\beta\varepsilon^2)+\gamma^2),$$

which is negative for $\varepsilon\gamma > 1$, so that we can conclude by Theorem 1 in [89, Sect. 4.4] for planar systems. □

3.3.3 Numerical evidence on the global phase diagram

Performing a global analysis, and specifically proving global results on the existence of periodic solutions to (3.49), does not seem easy. The main reason for this is that System (3.49) is three-dimensional and it does not appear to have a lower dimensional or more tractable representation. In fact, while for planar ODEs some global results on periodic orbits are available (see for e.g. [89, Ch. 3]) , a general global theory for higher dimensions is not developed. For this reason, at the global level we resort to a numerical analysis which allows one to describe the phase portrait to a large extent. First of all, we characterize all the fixed points of System (3.49), which are the solutions to

$$\begin{cases} -2x\cosh(\beta x) + 2\sinh(\beta x) + 2\varepsilon y\cosh(\beta x) - 2\varepsilon w\sinh(\beta x) = 0, \\ -2y\cosh(\gamma x) - 2\sinh(\gamma x) = 0, \\ -2w\cosh(\gamma x) - 2x\sinh(\gamma x) - 2w\cosh(\beta x) + 2y\sinh(\beta x) + 2\varepsilon\cosh(\beta x) \\ \qquad -2\varepsilon x\sinh(\beta x) = 0, \end{cases} \tag{3.54}$$

which can be rewritten as

$$\begin{cases} x &= (1-\varepsilon w)\tanh(\beta x) + \varepsilon y, \\ y &= -\tanh(\gamma x), \\ w &= -x\frac{\sinh(\gamma x)}{\cosh(\gamma x)+\cosh(\beta x)} + y\frac{\sinh(\beta x)}{\cosh(\gamma x)+\cosh(\beta x)} \\ & \quad +\varepsilon\frac{\cosh(\beta x)}{\cosh(\gamma x)+\cosh(\beta x)} - \varepsilon x\frac{\sinh(\beta x)}{\cosh(\gamma x)+\cosh(\beta x)}. \end{cases} \tag{3.55}$$

Substituting the values of y and w at the equilibrium in the equation for x we find

$$\begin{aligned} f(x) := {} & \tanh(\beta x) - \varepsilon\tanh(\gamma x) \\ & + \frac{\varepsilon\tanh(\gamma x)\sinh(\beta x)\tanh(\beta x)}{\cosh(\gamma x)+\cosh(\beta x)} - \frac{\varepsilon^2\sinh(\beta x)}{\cosh(\gamma x)+\cosh(\beta x)} \\ & - x\left[1 - \frac{\varepsilon\sinh(\gamma x)\tanh(\beta x)}{\cosh(\gamma x)+\cosh(\beta x)} - \frac{\varepsilon^2\sinh(\beta x)\tanh(\beta x)}{\cosh(\gamma x)+\cosh(\beta x)}\right] = 0. \end{aligned} \tag{3.56}$$

The scalar function $f(x)$ just defined has the same sign of the first component of the vector field in (3.49), and identifies the equilibria of the dynamics. Let us consider first the case $\varepsilon\gamma < 1$. A plot is shown in Figure 3.6, for $\varepsilon = 0.5$ and $\gamma = 1$. The critical value of the parameter β is in this case $\beta_c^* \approx 1.714$. As we see from the plot, for subcritical values of the parameters we have only one equilibrium, corresponding to the disordered state. Then, we see the appearance of two other equilibria of ferromagnetic type above β_c^*. From the linear analysis performed above, we know that - at least locally - the disordered state is stable until β_c^*, where it inverts its stability, with the two emerging polarized equilibria being stable.

The case $\varepsilon\gamma > 1$ is substantially different and richer. We report the plots separately in the two Figures 3.7 and 3.8, fixing in both $\varepsilon = 0.5$ and $\gamma = 7$. In Figure 3.7, it is shown that for $\beta < 2$ only the equilibrium corresponding to the disordered state is present. In $\beta = a_c \approx 2$ two other equilibria emerge, each of which then splits into two other equilibria, resulting in a total of five fixed points for the dynamics. We remark that the critical value a_c where the other two equilibria emerge was not found through the linear analysis on

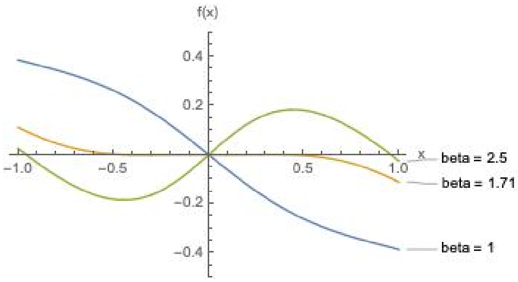


Figure 3.6: Plot of the function $f(x)$ for $\varepsilon = 0.5$, $\gamma = 1$. The parameter β takes three different values, $\beta = 1$, prior to the pitchfork bifurcation, $\beta = 1.71$ which is approximately the value β_c^* for which the pitchfork bifurcation arises, and $\beta = 2.5$.

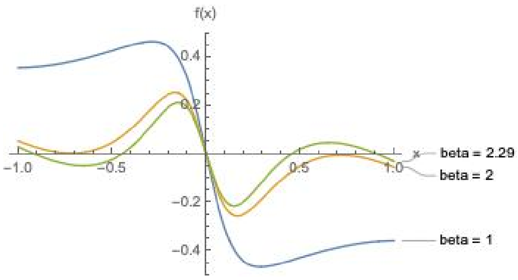


Figure 3.7: Plot of the function $f(x)$ for $\varepsilon = 0.5$, $\gamma = 7$. The parameter β takes three different values, $\beta = 1$, $\beta = 2$ for which we see the appearance of other two equilibria, and $\beta = 2.5$ where the equilibria have become five in total.

the bifurcations we performed above: it is presumably the result of a global phenomenon. Figure 3.8 shows the plot of $f(x)$ for other three values of the parameter β. In $\beta = 3$ (blue curve) the five equilibria are still present, even though we see that the two intermediate ones are decreasing towards 0. Indeed, in $\beta = \beta_c^* \approx 5.14$ the two intermediate equilibria disappear by collapsing at 0. This value of β is the one corresponding to the pitchfork bifurcation we found above. Finally, in $\beta = 7$ we have three equilibria remaining: the disordered state and two polarized equilibria.

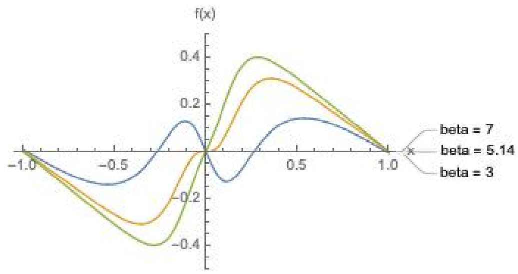


Figure 3.8: Plot of the function $f(x)$ for $\varepsilon = 0.5$, $\gamma = 7$. The parameter β takes three different values, $\beta = 3$, where the five equilibria are still present, $\beta = 5.14$ which is approximately the value β_c^* for which the pitchfork bifurcation arises, and $\beta = 7$, where only three equilibria survive.

3.3.4 Numerical linear analysis on the other equilibria

Even though the zeros of the function $f(x)$ cannot be found analytically, we can approximate them numerically with arbitrary precision for any choice of the parameters ε, γ and β. The corresponding values of y and w can be then retrieved from the expressions in (3.55). We observe that $f(x)$ is an odd function of x; the same holds for the expression of y in terms of x in (3.55), while w is even in x. Moreover, we remark that x and y at the equilibria have always opposite signs (except trivially at the disordered one). Thus, denoting with $(x^-_{\varepsilon,\gamma,\beta}, y^-_{\varepsilon,\gamma,\beta}, w^-_{\varepsilon,\gamma,\beta})$ the ferromagnetic polarized equilibrium with $x^-_{\varepsilon,\gamma,\beta} < 0$, we have that the corresponding positive one is $(x^+_{\varepsilon,\gamma,\beta}, y^+_{\varepsilon,\gamma,\beta}, w^+_{\varepsilon,\gamma,\beta}) = (-x^-_{\varepsilon,\gamma,\beta}, -y^-_{\varepsilon,\gamma,\beta}, w^-_{\varepsilon,\gamma,\beta})$. The other two intermediate equilibria appearing for $\varepsilon\gamma > 1$ for some values of β are denoted with $(x^{*,-}_{\varepsilon,\gamma,\beta}, y^{*,-}_{\varepsilon,\gamma,\beta}, w^{*,-}_{\varepsilon,\gamma,\beta})$ and $(x^{*,+}_{\varepsilon,\gamma,\beta}, y^{*,+}_{\varepsilon,\gamma,\beta}, w^{*,+}_{\varepsilon,\gamma,\beta})$, for which again we have $(x^{*,+}_{\varepsilon,\gamma,\beta}, y^{*,+}_{\varepsilon,\gamma,\beta}, w^{*,+}_{\varepsilon,\gamma,\beta}) = (-x^{*,-}_{\varepsilon,\gamma,\beta}, -y^{*,-}_{\varepsilon,\gamma,\beta}, w^{*,-}_{\varepsilon,\gamma,\beta})$.

In order to (locally) characterize the nature of these equilibria, we compute the Jacobian (and its eigenvalues) of the vector field in (3.49) at the roots of $f(x) = 0$ found numerically and the corresponding values of y and w. We report the results obtained via `Mathematica`:

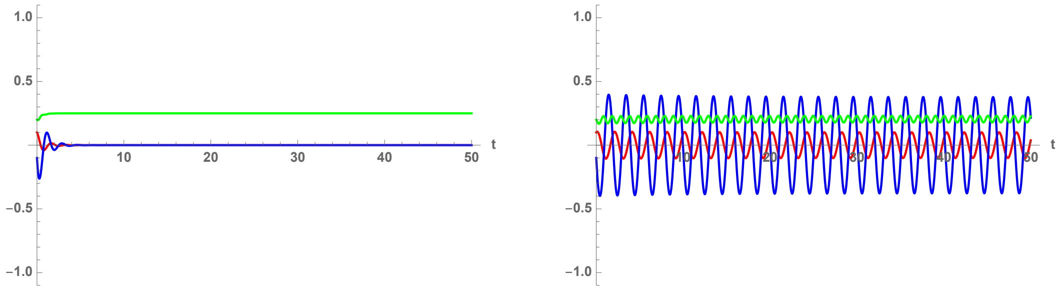
- for $\varepsilon\gamma < 1$, the two polarized equilibria $(x^\pm_{\varepsilon,\gamma,\beta}, y^\pm_{\varepsilon,\gamma,\beta}, w^\pm_{\varepsilon,\gamma,\beta})$ which emerge after the pitchfork bifurcation at β^*_c are linearly stable;
- for $\varepsilon\gamma > 1$:
  1. for the range of β's where the intermediate equilibria $(x^{*,\pm}_{\varepsilon,\gamma,\beta}, y^{*,\pm}_{\varepsilon,\gamma,\beta}, w^{*,\pm}_{\varepsilon,\gamma,\beta})$ exist ($a_c < \beta < \beta^*_c$), they are linearly unstable, with the Jacobian having two negative eigenvalues and a positive one;
  2. the two polarized equilibria $(x^\pm_{\varepsilon,\gamma,\beta}, y^\pm_{\varepsilon,\gamma,\beta}, w^\pm_{\varepsilon,\gamma,\beta})$ emerging for $\beta > a_c$ are always linearly stable.

3.3.5 Simulations and vector field projections

To investigate further on the global phase portrait, and particularly on the emergence of the cycles, we performed several simulations of System (3.49). Here, we restrict the parameters' values to the more interesting case $\varepsilon\gamma > 1$. In particular, in what follows we fix $\varepsilon = 0.5$ and $\gamma = 7$. Figures 3.9 and 3.10 show what happens when we let the dynamics start close to the disordered state. For small values of β (Figure 3.9A) the disordered state attracts the trajectories; for $\beta \approx \beta^{**}_c$ (Figure 3.9B), the value of the Hopf bifurcation, we see the emergence of periodic orbits, whose amplitude expands up to $\beta = 2.8$ (Figure 3.10A), which is approximately the maximal amplitude's point, since for $\beta = 2.9$ (Figure 3.10B) the periodic orbits disappear and everything gets attracted to the polarized equilibria.

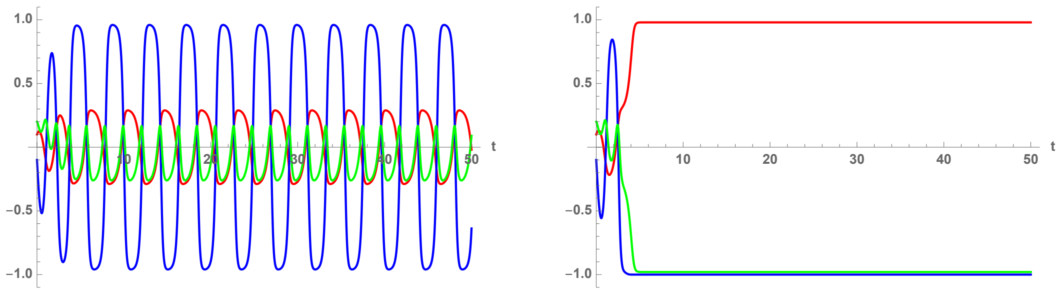
The picture is different when we start the dynamics far from the disordered state, as shown in Figures 3.11 and 3.12. As before, for small values of β the disordered state is a global attractor for the dynamics (Figure 3.11A); for intermediate values of β, right before the Hopf bifurcation (Figure 3.11B, where we considered $\beta = 2.1$, while $\beta^{**}_c \approx 2.285$), the system starts to oscillate, expecially in the y variable, but does not manage to reach periodic configurations before getting attracted to the polarized equilibria for values right above β^{**}_c: see Figure 3.12 for the case $\beta = 2.3$.

Summing up, these pictures highlight the global attractiveness for small values of β of the disordered state, the global attractiveness of the polarized states for big enough values of β, and the local nature of the presence of the stable cycle, which is visible only by starting the dynamics close to the disordered state, and for an intermediate range of values



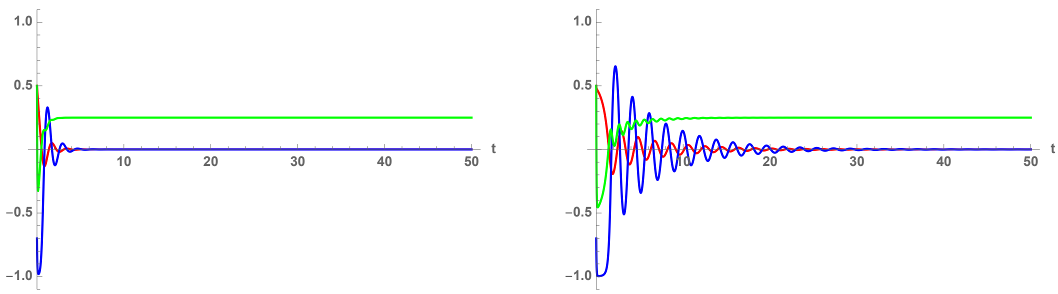
(a) $\beta = 1$ (b) $\beta = 2.3$

Figure 3.9: Plots of the solution $(x(t), y(t), w(t))$ of System (3.49) for $t \in [0, 50]$, starting close to the disordered state, $(x(0), y(0), w(0)) = (0.1, -0.1, 0.2)$, for $\varepsilon = 0.5$, $\gamma = 7$, $\beta = 1$ (left), $\beta = 2.3 \approx \beta_c^{**}$ (right). The red curve is $x(t)$, the blue one is $y(t)$ and the green is $w(t)$.



(a) $\beta = 2.8$ (b) $\beta = 2.9$

Figure 3.10: Plots of the solution $(x(t), y(t), w(t))$ of System (3.49) for $t \in [0, 50]$, starting close to the disordered state, $(x(0), y(0), w(0)) = (0.1, -0.1, 0.2)$, for $\varepsilon = 0.5$, $\gamma = 7$, $\beta = 2.8$ (left), $\beta = 2.9$ (right). The red curve is $x(t)$, the blue one is $y(t)$ and the green is $w(t)$.



(a) $\beta = 1$ (b) $\beta = 2.1$

Figure 3.11: Plots of the solution $(x(t), y(t), w(t))$ of System (3.49) for $t \in [0, 50]$, starting far from the disordered state, $(x(0), y(0), w(0)) = (0.5, -0.7, 0.5)$, for $\varepsilon = 0.5$, $\gamma = 7$, $\beta = 1$ (left), $\beta = 2.1$ (right). The red curve is $x(t)$, the blue one is $y(t)$ and the green is $w(t)$.

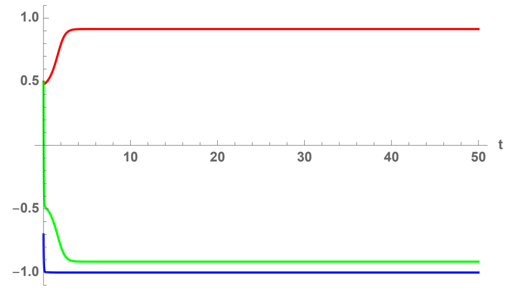


Figure 3.12: Plot of the solution $(x(t), y(t), w(t))$ of System (3.49) for $t \in [0, 50]$, starting far from the disordered state, $(x(0), y(0), w(0)) = (0.5, -0.7, 0.5)$ for $\varepsilon = 0.5$, $\gamma = 7$ and $\beta = 2.3$. The red curve is $x(t)$, the blue one is $y(t)$ and the green is $w(t)$.

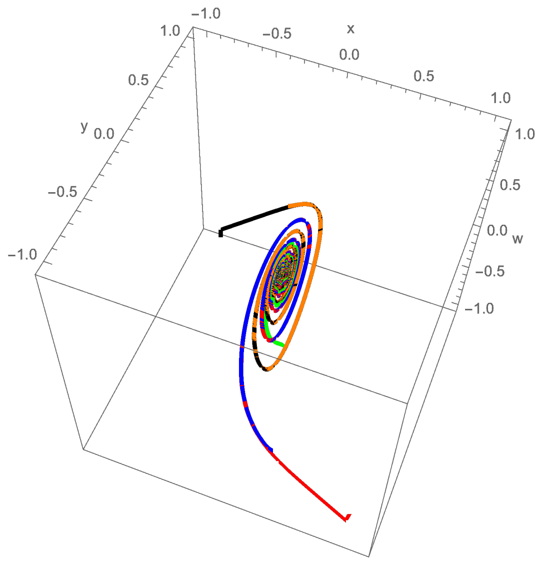


Figure 3.13: Parametric three-dimensional plot of trajectories starting at different inital points, for $\beta = 1.9$, right below the point where the other four equilibria appear.

of β: the picture is thus different from that of the model studied in Section 3.2, where the cycle seemed to have a global character. Moreover, we have found another critical point, which we call z_c, where the cycle disappears. For the chosen numerical values $\varepsilon = 0.5$ and $\gamma = 7$, we find $z_c \approx 2.817$. An interesting problem is to determine the way in which the cycle disappears: the fact that the cycle's amplitude increases with β and simultaneously the intermediate equilibria get closer and closer to the disordered state, strongly suggests that the point z_c could be identified as the value of β in which the cycle falls onto the stable manifold of one of the hyperbolic intermediate equilibria. This is indeed what is happening.

In Figures 3.13, 3.14, 3.15, 3.16, 3.17 and 3.18 we show some parametric three-dimensional plots of five trajectories of System (3.49), produced by starting the dynamics in different initial points, focusing on values of β around a_c, β_c^{**}, z_c and β_c^*. The trajectory in green was obtained by starting close to the disordered state, the blue and the orange ones in two intermediate states, while the red and the black trajectories are starting close to the polarized states. We see that, for values of $\beta < a_c$ (Figure 3.13), the disordered

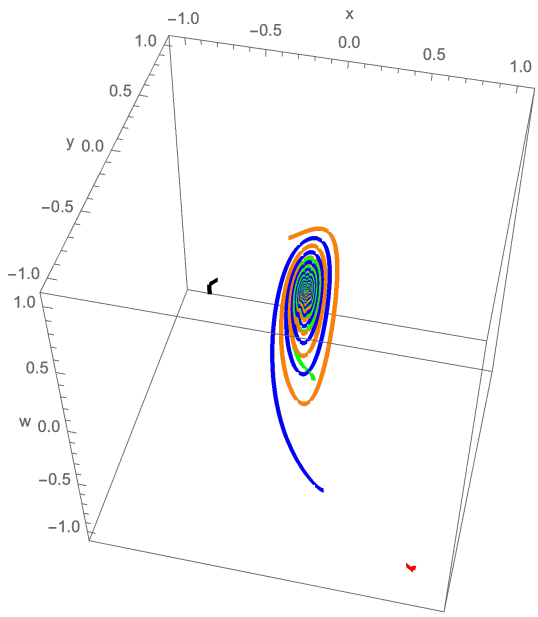


Figure 3.14: Parametric three-dimensional plot of trajectories starting at different inital points, for $\beta = 2.1$, i.e. right after the appearance of the other four equilibria.

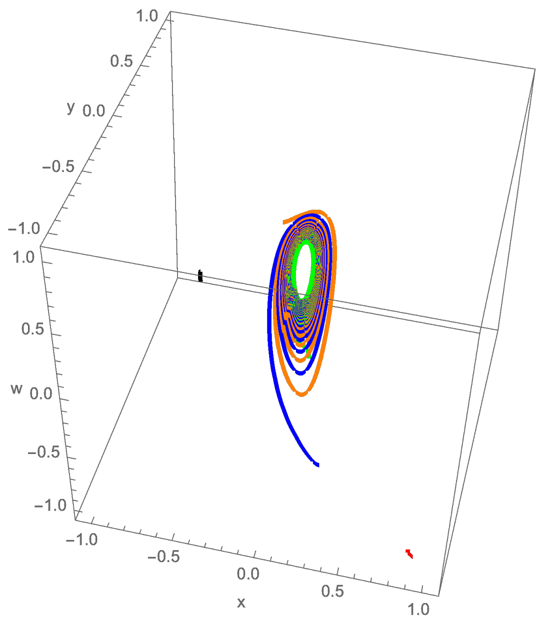


Figure 3.15: Parametric three-dimensional plot of trajectories starting at different inital points, for $\beta = 2.3$, right after the value of β of the Hopf bifurcation.

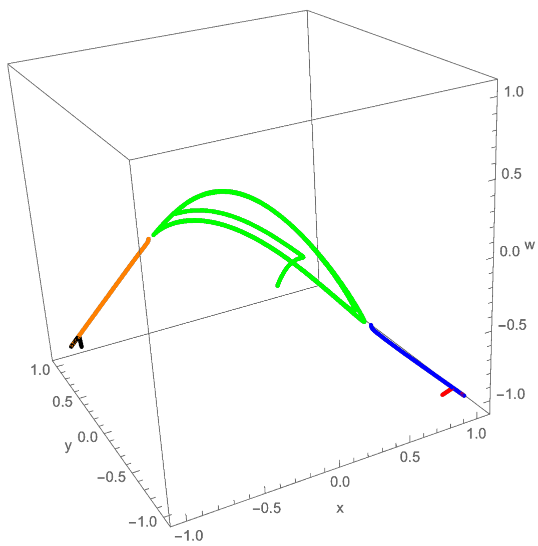


Figure 3.16: Parametric three-dimensional plot of trajectories starting at different inital points, for $\beta = 2.8$, i.e. right below the critical point z_c where the periodic cycle disappears.

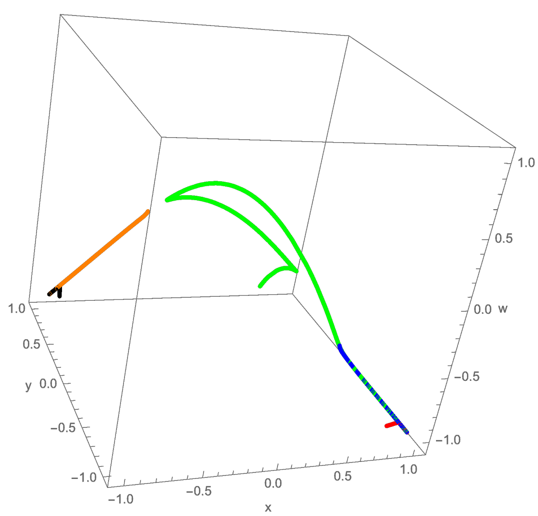


Figure 3.17: Parametric three-dimensional plot of trajectories starting at different inital points, for $\beta = 2.9$, i.e. right above the critical point z_c where the periodic cycle disappears.

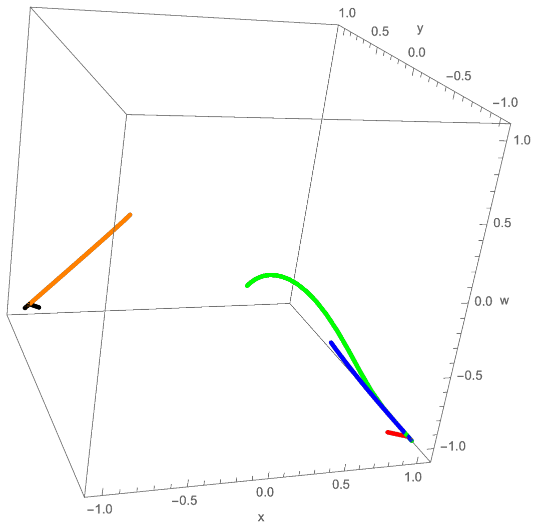
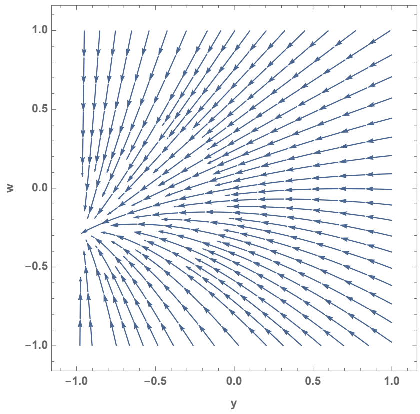


Figure 3.18: Parametric three-dimensional plot of trajectories starting at different inital points, for $\beta = 5.3$, where the unstable intermediate equilibria have disappeared, and the polarized equilibria remain the only attractive equilibria in the dynamics.

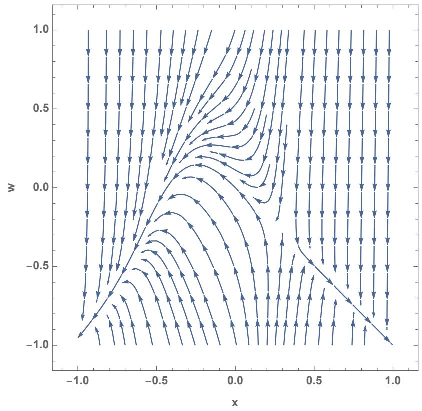
state attracts all the five trajectories, and it is a focus. When we increase β right above the value a_c where the other four equilibria appear, we see that (Figure 3.14) the red and black trajectories are now well distinguished and each tends towards the closer polarized state, while the solutions in blue and orange, which started close to the unstable intermediate equilibria, get attracted to the disordered state, which is still locally stable. Of course, if we started these two solutions closer to the polarized states we would have seen an attraction towards the other equilibria. In Figure 3.15 we have chosen a β right above the critical value β_c^{**} of the Hopf bifurcation: we here indeed see the appearance of a cycle (the one in green), and that the two intermediate solutions get attracted to the cycle. The cycle increases its amplitude until $\beta \approx 2.8$, which is shown in Figure 3.16. Here, even though the cycle is still stable, the two intermediate solutions in orange and blue are now attracted to the polarized states: indeed, the actual value of the unstable intermediate equilibria is decreasing towards the disordered state as β increases, but we are keeping fixed the initial "intermediate" data so that after a while it falls within the domain of attraction of the extreme equibria.

Figure 3.17 was realized by choosing a β right above the critical value z_c where the cycle disappears: as expected, the trajectory in green touches one of the two intermediate solutions (in this case the blue one), and consequently gets attracted towards the polarized state. As we saw from the numerical linear analysis of Section 3.3.4, locally around the intermediate equilibria we have a two dimensional stable manifold and a one dimensional unstable one. Necessarily then, the cycle eventually hits the two dimensional stable manifold, the resulting trajectory escapes through the one dimensional unstable curve, and finally it gets attracted to the polarized state. A confirmation of this is shown in Figure 3.19, where we plotted the vector field projected onto the three coordinates x, y and w, for a value of β right below z_c. As we see from the pictures, the plane $x = \bar{x} \approx x^{*,+}_{\varepsilon,\gamma,\beta}$ is (approximately) the stable two dimensional manifold associated to the intermediate equilibrium.

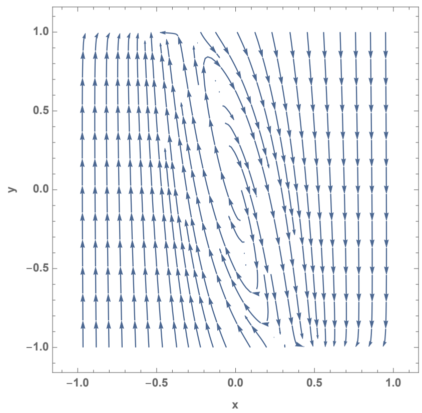
Finally, Figure 3.18 shows the three-dimensional simulation for a big value of β, after the pitchfork bifurcation in β_c^* where the two intermediate unstable equilibria vanish. Here,



(a) Plane $x = \bar{x}$



(b) Plane $y = \bar{y}$



(c) Plane $w = \bar{w}$

Figure 3.19: The vector field projected onto the planes $x = \bar{x} \approx x^{*,+}_{\varepsilon,\gamma,\beta}$ (Figure A), $y = \bar{y} \approx y^{*,-}_{\varepsilon,\gamma,\beta}$ (Figure B), $w = \bar{w} \approx w^{*,-}_{\varepsilon,\gamma,\beta}$ (Figure C), for $\beta = 2.8$.

the green trajectory oscillates less and less and gets attracted to the polarized state, which is the only stable equilibrium in the dynamics and thus attracts all the trajectories.

CHAPTER 4

Beyond the mean field case: a hierarchical mean field model of interacting spins

This chapter is devoted to the study of a model of interacting spins with a hierarchical mean field structure. As the models of Chapter 3, it can be viewed as an example of non-Markovian spin system, where the non-Markovianity is realized via a *state-augmentation* procedure, as it was for the model studied in Section 3.3. At the individual level, the full Markovian state is indeed given by a pair of variables: the spin and a continuous variable which evolves in a diffusive way. The main goal of our study is to obtain macroscopic limits at various spatio-temporal scales for both the mean field and the two-level hierarchical formulation of the model, analyzing the presence of phase transitions in the system.

In the literature, hierarchical models were often employed for applications in population dynamics and genetics, where individuals naturally dispose in groups with a hierarchical structure (families, clans, villages, colonies, populations and so on). A series of papers from the '90s - '00s (initiated with [39] and [40] among others), nicely reviewed in [70], deals with different types of hierarchical mean field linearly interacting diffusions (the prototype being linear Wright-Fisher diffusions), where in most cases the macroscopic limits are retrieved at every spatio-temporal scale, and a renormalization map can be defined, allowing one to pass from one hierarchical level to the other. Moreover, the study of the fixed points of the renormalization map is in some cases fully worked out. Two crucial ingredients which allow for an iterative renormalization procedure are the linearity of the interactions, which in the above works is realized by considering linear drifts, of imitative type, which scale with the hierarchical distance, and some ergodicity properties of the individual dynamics. The motivation for focusing on diffusive dynamics as building blocks for the hierarchical models stems from the fact that, with their choices, each individual non-interacting dynamics can itself be obtained as a continuum limit of a corresponding finite state space model of interacting particles: for example, the discrete prelimit counterpart of the Wright-Fisher diffusion is the *voter model* (see e.g. [33]).

When working directly on finite state models fewer results are known, due to the non-linearity of the microscopic interactions. Hierarchical Ising-type models for spin systems were introduced in [41]. Since then, a literature on the hierarchical group and renormalization theory for spin systems was developed (e.g. [11], [46], [47], [69], [74]), but always studying equilibrium models. On the finite state space dynamics, we acknowledge the work [2], which studies contact processes on the hierarchical group, with a focus on deriving sufficient conditions on the speed of decay of the infection rates for obtaining a phase transition between extinction and survival.

In our case, we define hierarchical dynamics of spin-flip type with a ferromagnetic mean field interaction, coupled with a system of linearly interacting diffusions of Ornstein-Uhlenbeck type. In particular, the diffusive variables enter in the spin-flip rates, effectively acting as dynamical magnetic fields. In absence of the diffusions, the spin-flip dynamics can be thought of as a hierarchical version of the Curie–Weiss model (3.4) (except for considering alternative, but equivalent, transition rates). Note that in our model the interaction between the spins is highly non-linear. However, as we shall see, the linear diffusions *drive* the system of spins, eventually allowing for a separation of spatio-temporal scales for the spin dynamics as well.

It should also be noted that the model treated in this chapter is somehow related to the model of interacting renewal processes of Section 3.2. Indeed, in the zero-temperature limit and for some range of the diffusion parameters, the block averages of the spins behave as macro-spins themselves, with random jump times which are non-exponentially distributed (see Remark 4.9 for the single particle case). However, the structure of the interactions in this model is definitely different from that of Section 3.2, where in the latter it was introduced as a time scaling on the waiting times of each particle's jumps depending on the magnetization of the spins. Nevertheless, we cannot rule out the possible emergence of oscillating behavior for some particular parameters values also for this model, even though we were not able to experience it with simulations.

The chapter is organized as follows: in Section 4.1 we formulate the dynamics for a general interaction graph, which we then specify to the two contexts of our interest: the mean field case - analyzed in Section 4.2, and the two-level hierarchical case - analyzed in Section 4.3. In particular, in Section 4.2 we derive rigorously the macroscopic limit at the two characteristic timescales of the model for any value of the parameters, highlighting the presence of a phase transition, and studying the resulting effects on the dynamics at each timescale. In Section 4.3, we study rigorously the macroscopic limits at the three different timescales of the two-level hierarchical model, restricting ourselves to a range of interaction parameters which we refer to as *subcritical*. We also formulate a generalization of these results to the k-level hierarchical version of the model, for any $k \in \mathbb{N}$ finite (Section 4.3.5). In the *supercritical* region we focus on the zero-temperature limit (Section 4.3.6), where we give a description of the limit dynamics supported by numerics and heuristic arguments, allowing for a comparison with the mean field scenario.

4.1 Introducing the model

Consider a set V (possibly countably infinite), indexing individuals in a population. Each individual $r \in V$ is identified with a pair of variables (σ_r, x_r): a spin variable $\sigma_r \in \{-1, 1\}$, and a continuous one $x_r \in \mathbb{R}$, modeling some summary statistics of the remaining characteristics of the individual, and thus being naturally normally distributed by a central limit theorem. The interaction between each pair of spin variables $\sigma_r, \sigma_s \in V$ is encoded in a (possibly random) variable $J_{rs} \in \mathbb{R}$. Analogously, x_r and x_s interact with a strength proportional to some variables $J'_{rs} \in \mathbb{R}$. The particles $(\sigma_r, x_r)_{r \in V}$ follow stochastic dynamics given by

$$\begin{cases} \sigma_r \mapsto -\sigma_r, & \text{with rate} \quad 1 + \tanh\left[-\sigma_r \sum_{s \in V} J_{rs}(\sigma_s + x_s)\right], \\ dx_r = -\sum_{s \in V} J'_{rs}(x_r - x_s)dt + \sigma dW_r(t), \end{cases} \tag{4.1}$$

where $W_r(t)$'s are $|V|$ independent Brownian motions, and $\sigma > 0$ is the diffusion coefficient. The choice of the rate function $1 + \tanh(\cdot)$ in (4.1) might seem unusual. Note that it is alternative to the more common choice $e^{-\sigma_r \sum_{s\in V} J_{rs}(\sigma_s + x_s)}$, which we discussed briefly in Section 3.1 in the ferromagnetic mean field case without diffusions. As the latter, in the case without diffusions, it defines a Glauber-type spin-flip dynamics with respect to which the Gibbs measure

$$\pi(\sigma) \propto 1 + \tanh\left(\sum_{r,s\in V} J_{rs}\sigma_r\sigma_s\right)$$

is reversible. The main reason for the alternative choice $1+\tanh(\cdot)$ is technical, as the boundedness of the transition rates is convenient for the proofs, even though we believe it is not an essential ingredient.

We focus on two different choices for V and (deterministic) interaction parameters J_{rs} and J'_{rs}:

- *Ferromagnetic mean field case:*

$$\begin{aligned} V &:= \{1,\dots,N\}, \\ J_{rs} &= \frac{\beta}{N}, \\ J'_{rs} &= \frac{\alpha}{N}, \end{aligned} \tag{4.2}$$

  with $\alpha, \beta \geq 0$.

- *Ferromagnetic two-level hierarchical case:*

$$V := \{1,\dots,N\} \times \{1,\dots,N\}, \qquad \begin{cases} J_{rs} = \frac{\beta_1}{N}, \quad J'_{rs} = \frac{\alpha_1}{N}, & \text{if } |r-s| \leq 1, \\ J_{rs} = \frac{\beta_2}{N^2}, \quad J'_{rs} = \frac{\alpha_2}{N^3}, & \text{if } |r-s| = 2, \end{cases} \tag{4.3}$$

  with $\alpha_1, \alpha_2, \beta_1, \beta_2 \geq 0$, where the distance $|\cdot|$ between $r := (i,j)$ and $s := (k,l)$ is defined by

$$|r-s| := \begin{cases} 0, & \text{if } i=k,\ j=l \\ 1, & \text{if } i\neq k,\ j=l \\ 2, & \text{otherwise.} \end{cases} \tag{4.4}$$

The two-level hierarchical case can be thought of as a model for a collection of N interacting populations, each of which is itself a mean field interacting particle system with N particles. In the definition (4.4) of the hierarchical distance $|r-s|$, the first index i refers to the individual, while the index j identifies the j-th population. Two individuals $r=(i,j)$ and $s=(k,l)$ are thus said to be at distance 1 if $j=l$ (i.e. they belong to the same population); otherwise, they are at distance 2. The choices in (4.3) are such that the strength of the interaction is inversely proportional to the hierarchical distance. This construction can be reiterated a finite number of times to define a k-level hierarchical model, where $V := \{1,\dots,N\}^k$, $J_{rs} \propto \frac{1}{N^l}$, $J'_{rs} \propto \frac{1}{N^{2l-1}}$ for $|r-s| = l$, with $l = 1,\dots,k$. See Section 4.3.5 for details. The main goal of this chapter is to obtain a limit description of both the mean field and the two-level hierarchical formulation of dynamics (4.1) at different spatio-temporal scales, analyzing the possible presence of phase transitions in the system.

4.2 A starter: the mean field model

In this section we study the mean field version of the model, i.e. the case of a single population of N individuals with a mean field type interaction. We denote by $(\boldsymbol{\sigma}, \boldsymbol{x}) := (\sigma_j, x_j)_{j=1,\dots,N} \in (\{-1,1\} \times \mathbb{R})^N$ a configuration of the entire population. In the following, we interchangeably use the coordinates (σ_i, λ_i) and (σ_i, x_i), where $\lambda_i := \sigma_i + x_i$. We denote by

$$m^N(t) := \frac{1}{N}\sum_{i=1}^N \sigma_i(t)$$

the magnetization of the spin variables at time t, and by $\lambda^N(t)$ (resp. $x^N(t)$) the analogous quantity for the $\lambda_i(t)$'s (resp. $x_i(t)$'s). The dynamics is such that, at time t, the i-th spin flips with rate

$$\sigma_i \mapsto -\sigma_i, \quad \text{with rate} \quad 1 + \tanh(-\beta\sigma_i(t)\lambda^N(t)),$$

where $\lambda^N(t)$ and $x^N(t)$ satisfy

$$\begin{cases} d\lambda^N(t) = dm^N(t) + dx^N(t), \\ dx^N(t) = \frac{\sigma}{\sqrt{N}} dW^N(t), \end{cases} \tag{4.5}$$

where W^N is a Brownian motion and $\sigma > 0$ the diffusion coefficient. Indeed, by substituting the mean field coupling constants (4.2) in the general dynamics (4.1) we retrieve

$$\begin{aligned} 1 + \tanh\left[-\sigma_i(t)\sum_{j=1}^N J_{ij}(\sigma_j(t) + x_j(t))\right] &= 1 + \tanh\left(-\beta\sigma_i(t)(m^N(t) + x^N(t))\right) \\ &= 1 + \tanh\left(-\beta\sigma_i(t)\lambda^N(t)\right), \\ dx_i(t) = -\sum_{j=1}^N \frac{\alpha}{N}(x_i(t) - x_j(t))dt + \sigma dW_i(t) &= -\alpha(x_i(t) - x^N(t))dt + \sigma dW_i(t), \end{aligned}$$

and thus, by averaging the second line over $i = 1, \dots, N$,

$$dx^N(t) = \frac{\sigma}{\sqrt{N}} dW^N(t),$$

where $W^N := \frac{1}{\sqrt{N}}\sum_{i=1}^N W_i$ is a Brownian motion. We stress that the law of W^N does not depend on N, but we keep the notation W^N to refer to the specific Brownian motion obtained by the aggregation of the single W_i's.

From the definition of the spin-flip rates, we obtain the transition rates at time t for the order parameter m^N

$$\begin{aligned} m^N &\mapsto m^N + \frac{2}{N}, \quad && \text{with rate} \quad N\frac{1 - m^N(t)}{2}\left[1 + \tanh(\beta\lambda^N(t))\right] \\ m^N &\mapsto m^N - \frac{2}{N}, \quad && \text{with rate} \quad N\frac{1 + m^N(t)}{2}\left[1 - \tanh(\beta\lambda^N(t))\right]. \end{aligned} \tag{4.6}$$

We assume i.i.d. initial data for the single variables $x_i(0) \sim \mathcal{N}(x_0, \sigma^2)$, and $\sigma_i(0) \sim \text{Ber}(p)$, for some $p \in [0,1]$.

The infinitesimal generator associated to the dynamics (4.5) and (4.6), applied to a function $f : \mathbb{R} \times [-1,1] \to \mathbb{R}$, is thus given by

$$\begin{aligned}\mathcal{L}^N f(\lambda, m) &= N\frac{1-m}{2}\left[1+\tanh(\beta\lambda)\right]\left[f\left(\lambda+\frac{2}{N}, m+\frac{2}{N}\right) - f(\lambda,m)\right] \\ &+ N\frac{1+m}{2}\left[1-\tanh(\beta\lambda)\right]\left[f\left(\lambda-\frac{2}{N}, m-\frac{2}{N}\right) - f(\lambda,m)\right] + \frac{\sigma^2}{2N}\frac{\partial^2}{\partial\lambda^2} f(\lambda,m).\end{aligned} \tag{4.7}$$

In the alternative variables (x, m), with $x = \lambda - m$, the generator takes the form

$$\begin{aligned}\mathcal{L}^N f(x, m) &= N\frac{1-m}{2}\left[1+\tanh(\beta(x+m))\right]\left[f\left(x, m+\frac{2}{N}\right) - f(x,m)\right] \\ &+ N\frac{1+m}{2}\left[1-\tanh(\beta(x+m))\right]\left[f\left(x, m-\frac{2}{N}\right) - f(x,m)\right] + \frac{\sigma^2}{2N}\frac{\partial^2}{\partial x^2} f(x,m).\end{aligned} \tag{4.8}$$

The rest of this section on the mean field case is organized as follows. In the next two subsections we motivate the expected limit behavior at the two different timescales characterizing the model: in Section 4.2.1 we deduce the order 1 timescale *deterministic* limit dynamics for $N \to +\infty$, while in Section 4.2.2 we introduce the problem of studying the *fluctuations* around the deterministic limit at an accelerated timescale of order N. An easy but important property of the accelerated dynamics is then given in Proposition 4.2, which will turn out to be very useful for generalizing some results to the two-level hierarchical case. We finally address rigorously the convergence problem in the so-called *subcritical* regime in Section 4.2.3, and in the *supercritical* regime in Section 4.2.4.

4.2.1 Deterministic mean field limit

At times of order 1, where the fluctuations terms (i.e. the terms which tend to 0 for $N \to +\infty$ in the generator (4.7)) become negligible for $N \gg 0$, the dynamics of the system is well approximated by the following system of two ODEs

$$\begin{cases}\dot\lambda(t) = 2\tanh(\beta\lambda(t)) - 2m(t) \\ \dot m(t) = 2\tanh(\beta\lambda(t)) - 2m(t) \\ \lambda(0) = \lambda_0 \in \mathbb{R}, \\ m(0) = m_0 \in [-1,1],\end{cases} \tag{4.9}$$

which can be thought of as the mean field limit of the dynamics introduced at the beginning. Note that if we choose initial conditions such that $\lambda_0 = m_0$, the above system restricts to the Curie–Weiss model (3.6), except for a missing multiplicative term in the vector field which does not modify the qualitative behavior of the dynamics. System (4.9) is such that its equilibria form a one-dimensional curve of fixed points, given by

$$m = \tanh\beta\lambda,$$

corresponding to the points (λ, m) for which $(\dot\lambda, \dot m) = (0,0)$. By studying the sign of the two-dimensional vector field in (4.9), which has a constant slope of 1 since its components are equal, one can get convinced that the equilibrium curve is a global attractor for the dynamics. However, we can distinguish two regimes, depending on the value of the parameter β.

Figures 4.1 and 4.2 should highlight the qualitative behavior of the dynamics: for $\beta < 1$, when the slope of the invariant curve is always smaller than the one of the vector

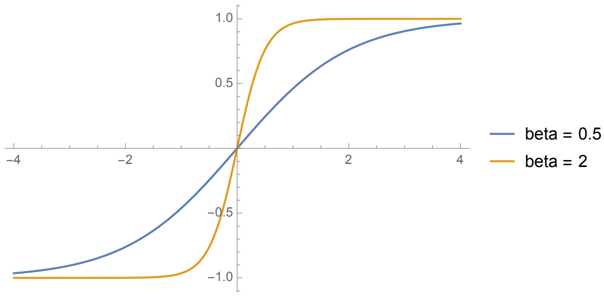


Figure 4.1: Invariant manifold for different values of β

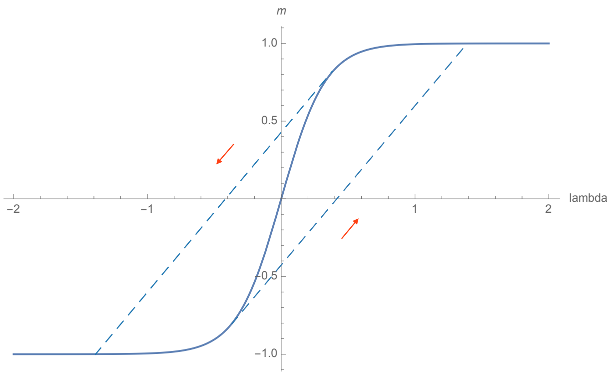


Figure 4.2: Qualitative behavior for $\beta > 1$

field, the whole curve is a stable manifold; for $\beta > 1$ instead, the curve is stable in the two disjoint external intervals where the slope is less than 1, while it shows an unstable behavior in the internal interval where the slope of the curve is greater than 1. For $\beta > 1$, we denote the critical points where the curve has a slope equal to 1 as $(\pm\lambda_a(\beta), \pm m_a(\beta))$, where

$$\begin{aligned} \lambda_a(\beta) &= \frac{1}{\beta}\operatorname{arctanh}\left(\sqrt{1-\frac{1}{\beta}}\right), \\ m_a(\beta) &= \sqrt{1-\frac{1}{\beta}}. \end{aligned} \tag{4.10}$$

Thus, for some initial conditions close enough to the critical points, the dynamics will be soon attracted to the other branch of the curve, as shown in Figure 4.2, where the vector field lines are also drawn in red. Consequently one can expect that, at the larger timescales where the fluctuations are not negligible, the corresponding N-particle system might show an oscillating behavior between the two stable intervals, where the fluctuations play a role in driving the order parameters close enough to the endpoints of the stable intervals, thus determining a sudden change in the macroscopic variables.

The limit system (4.9) is easily derived by observing that the generator (4.7) uniformly converges to

$$\begin{aligned} \mathcal{L}f(\lambda, m) = (1-m)\left[1+\tanh\beta\lambda\right]\left[\frac{\partial}{\partial\lambda}f(\lambda,m) + \frac{\partial}{\partial m}f(\lambda,m)\right] \\ -(1+m)\left[1-\tanh(\beta\lambda)\right]\left[\frac{\partial}{\partial\lambda}f(\lambda,m) + \frac{\partial}{\partial m}f(\lambda,m)\right], \end{aligned}$$

which can be rewritten as

$$\mathcal{L}f(\lambda, m) = (2\tanh(\beta\lambda) - 2m)\left[\frac{\partial}{\partial\lambda}f(\lambda, m) + \frac{\partial}{\partial m}f(\lambda, m)\right]. \tag{4.11}$$

From the uniform convergence of the generators we obtain the weak convergence of the stochastic processes $(\lambda^N(t), m^N(t))_{t\in[0,T]}$ satisfying dynamics (4.5) and (4.6) to the limit deterministic process $(\lambda(t), m(t))_{t\in[0,T]}$, for which system (4.9) holds (see [56] for a classic reference).

4.2.2 Fluctuations around the deterministic limit

In order to rigorously understand the oscillating behavior with jumps which we qualitatively described in the previous section, we are led to study the fluctuations of the N-particle dynamics around its deterministic limit (4.9).This boils down to accelerating the dynamics and studying what happens at the timescale where the fluctuations are not negligible anymore, where one expects to see some limiting diffusive motion across the equilibria. In the following, for ease of notation, we still denote as

$$(\lambda^N(t), m^N(t))$$

the accelerated dynamics at a timescale of order N, i.e.

$$(\lambda^N(t), m^N(t)) := (\lambda^N(Nt), m^N(Nt)),$$

with the latter being the original process at a timescale of order 1 (and the same notation $(x^N(t), m^N(t))$ for the alternative variables).

To motivate the presence of a limiting diffusive behavior at the accelerated timescale, it is instructive to make a preliminary computation. We develop the jump terms in the generator (4.7) at the second order, as follows:

$$\begin{aligned}
f\Big(\lambda + \frac{2}{N}, &m + \frac{2}{N}\Big) - f(\lambda, m)\\
&= f\left(\lambda + \frac{2}{N}, m + \frac{2}{N}\right) - f\left(\lambda, m + \frac{2}{N}\right) + f\left(\lambda, m + \frac{2}{N}\right) - f(\lambda, m)\\
&\approx \frac{2}{N}\frac{\partial}{\partial\lambda}f\Big(\lambda, m + \frac{2}{N}\Big) + \frac{2}{N^2}\frac{\partial^2}{\partial\lambda^2}f\Big(\lambda, m + \frac{2}{N}\Big)\\
&+ \frac{2}{N}\frac{\partial}{\partial m}f(\lambda, m) + \frac{2}{N^2}\frac{\partial^2}{\partial m^2}f(\lambda, m) + O\Big(\frac{1}{N^3}\Big)\\
&\approx \frac{4}{N^2}\frac{\partial^2}{\partial m\partial\lambda}f(\lambda, m) + \frac{2}{N}\frac{\partial}{\partial\lambda}f(\lambda, m) + \frac{2}{N^2}\frac{\partial^2}{\partial\lambda^2}f(\lambda, m)\\
&+ \frac{2}{N}\frac{\partial}{\partial m}f(\lambda, m) + \frac{2}{N^2}\frac{\partial^2}{\partial m^2}f(\lambda, m) + O\Big(\frac{1}{N^3}\Big),
\end{aligned}$$

and, with analogous computations,

$$\begin{aligned}
f\Big(\lambda - \frac{2}{N}, m - \frac{2}{N}\Big) - f(\lambda, m) &\approx \frac{4}{N^2}\frac{\partial^2}{\partial m\partial\lambda}f(\lambda, m) - \frac{2}{N}\frac{\partial}{\partial\lambda}f(\lambda, m) + \frac{2}{N^2}\frac{\partial^2}{\partial\lambda^2}f(\lambda, m)\\
&- \frac{2}{N}\frac{\partial}{\partial m}f(\lambda, m) + \frac{2}{N^2}\frac{\partial^2}{\partial m^2}f(\lambda, m) + O\Big(\frac{1}{N^3}\Big).
\end{aligned}$$

Thus, without considering the remainder terms of higher orders, we have

$$
\begin{aligned}
\mathcal{L}^N f(\lambda, m) &\approx (2\tanh(\beta\lambda) - 2m)\left[\frac{\partial}{\partial\lambda} f(\lambda, m) + \frac{\partial}{\partial m} f(\lambda, m)\right] \\
&+ N\frac{1-m}{2}\left[1+\tanh(\beta\lambda)\right]\left[\frac{4}{N^2}\frac{\partial^2}{\partial m\partial\lambda} f(\lambda, m) + \frac{2}{N^2}\frac{\partial^2}{\partial\lambda^2} f(\lambda, m) + \frac{2}{N^2}\frac{\partial^2}{\partial m^2} f(\lambda, m)\right] \\
&+ N\frac{1+m}{2}\left[1-\tanh(\beta\lambda)\right]\left[\frac{4}{N^2}\frac{\partial^2}{\partial m\partial\lambda} f(\lambda, m) + \frac{2}{N^2}\frac{\partial^2}{\partial\lambda^2} f(\lambda, m) + \frac{2}{N^2}\frac{\partial^2}{\partial m^2} f(\lambda, m)\right] \\
&+ \frac{\sigma^2}{2N}\frac{\partial^2}{\partial\lambda^2} f(\lambda, m) \\
&= (2\tanh(\beta\lambda) - 2m)\left[\frac{\partial}{\partial\lambda} f(\lambda, m) + \frac{\partial}{\partial m} f(\lambda, m)\right] \\
&+ \frac{1}{N}(2 - 2m\tanh(\beta\lambda))\left[2\frac{\partial^2}{\partial m\partial\lambda} f(\lambda, m) + \frac{\partial^2}{\partial\lambda^2} f(\lambda, m) + \frac{\partial^2}{\partial m^2} f(\lambda, m)\right] \\
&+ \frac{\sigma^2}{2N}\frac{\partial^2}{\partial\lambda^2} f(\lambda, m).
\end{aligned}
\tag{4.12}
$$

In this approximation, the bidimensional diffusion process, which we denote as

$$
(\tilde{\lambda}^N(t), \tilde{m}^N(t)),
$$

associated to the approximation of the generator $N\mathcal{L}^N$ (i.e. in the accelerated timescale of order N) is

$$
\begin{cases}
d\tilde{\lambda}^N(t) = N(2\tanh(\beta\tilde{\lambda}^N(t)) - 2\tilde{m}^N(t))dt + \sigma_{11}(\tilde{\lambda}^N(t), \tilde{m}^N(t))dB_1(t) \\
\qquad\qquad +\sigma_{12}(\tilde{\lambda}^N(t), \tilde{m}^N(t))dB_2(t), \\
d\tilde{m}^N(t) = N(2\tanh(\beta\tilde{\lambda}^N(t)) - 2\tilde{m}^N(t))dt + \sigma_{21}(\tilde{\lambda}^N(t), \tilde{m}^N(t))dB_1(t) \\
\qquad\qquad +\sigma_{22}(\tilde{\lambda}^N(t), \tilde{m}^N(t))dB_2(t),
\end{cases}
\tag{4.13}
$$

where

$$
\begin{aligned}
\frac{1}{2}(\sigma^2)_{11}(\tilde{\lambda}^N(t), \tilde{m}^N(t)) &= \frac{\sigma^2}{2} + (2 - 2\tilde{m}^N(t)\tanh(\beta\tilde{\lambda}^N(t))), \qquad (4.14)\\
\frac{1}{2}(\sigma^2)_{12}(\tilde{\lambda}^N(t), \tilde{m}^N(t)) &= (2 - 2\tilde{m}^N(t)\tanh(\beta\tilde{\lambda}^N(t))), \\
\frac{1}{2}(\sigma^2)_{22}(\tilde{\lambda}^N(t), \tilde{m}^N(t)) &= (2 - 2\tilde{m}^N(t)\tanh(\beta\tilde{\lambda}^N(t))).
\end{aligned}
$$

Note that the above diffusive approximation (4.12) of the initial generator (4.7) is correct for $N \to +\infty$, even for the accelerated dynamics. Indeed, the remainder third-order terms account for an error of order $O\left(\frac{1}{N^2}\right)$ at times of order 1, becoming of order $O\left(\frac{1}{N}\right)$ at times of order N. In other words, we have that

Remark 4.1. *The accelerated N-particle dynamics $(\lambda^N(t), m^N(t))_{t\geq 0}$ and its diffusive approximation $(\tilde{\lambda}^N(t), \tilde{m}^N(t))_{t\geq 0}$ have the same (weak) limit $(\lambda(t), m(t))_{t\geq 0}$ (provided it exists).*

The system of two diffusive SDEs (4.13) features a strong drift which grows with N in both variables, and a bidimensional diffusion term which is of order 1. Intuitively, the limit process $(\lambda(t), m(t))_{t\geq 0}$ should thus satisfy the equation $m(t) = \tanh(\beta\lambda(t))$, so that the strong drift component vanishes. Equivalently stated, the strong drift in (4.13) fastly attracts the dynamics towards the curve $m(t) = \tanh(\beta\lambda(t))$, on which the diffusive part then acts on a larger timescale. In the limit $N \to +\infty$, one is then expecting to see an effective one-dimensional diffusive motion onto the curve $m(t) = \tanh(\beta\lambda(t))$, which can be parametrized by using any of the two variables. Because of the difference in the stability properties of the curve for different values of the parameters, additional care must be put in the case when $\beta > 1$, where one should expect to retrieve a diffusive motion on the two stable intervals of the curve, $(-\infty, -\lambda_a(\beta))$ and $(\lambda_a(\beta), +\infty)$ (with $\lambda_a(\beta)$ as in (4.10)), with jumps from one to the other component when the dynamics hits the critical points, as we shortly described in the previous section. Moreover, as we prove below, the arrival points of the jumps are also deterministic, and they are given by the intersection of the invariant curve with the tangent line passing through the critical points (see Figure 4.2).

An easy computation shows that our intuition, motivated by the form of the approximate dynamics (4.13), is indeed correct. The accelerated N-particles *exact* dynamics $(\lambda^N(t), m^N(t))$ contracts the distance between m and the invariant curve $\tanh(\beta\lambda)$, but only in the stable intervals $(-\infty, -\lambda_a(\beta))$ and $(\lambda_a(\beta), +\infty)$ when $\beta > 1$. Specifically, if we denote

$$y^N(t) := m^N(t) - \tanh(\beta\lambda^N(t)), \tag{4.15}$$

we have the following

Proposition 4.2. *Let $y^N(t)$ be as in* (4.15)*. Then, for any $T > 0$, $k > 0$, $\beta < 1$,*

$$\mathbb{E}\left[\sup_{t\in[0,T]} |y^N(t)|^k\right] \to 0, \tag{4.16}$$

for $N \to +\infty$.

Proof. If we apply the generator (4.7) in the accelerated timescale to any power k of the distance $|y^N(t)|$, we obtain

$$\begin{aligned} N\mathcal{L}^N|y^N(t)|^k &= N\mathcal{L}^N|m^N(t) - \tanh(\beta\lambda^N(t))|^k \\ &\le -2kN(m^N(t) - \tanh(\beta\lambda^N(t)))|m^N(t) - \tanh(\beta\lambda^N(t))|^{k-1}\times \\ &\quad\times \operatorname{sign}(m^N(t) - \tanh(\beta\lambda^N(t)))\left[-\frac{d}{d\lambda}(\tanh(\beta\lambda^N(t))) + 1\right] + O(1) \\ &= -2kN|m^N(t) - \tanh(\beta\lambda^N(t))|^k\left[-\beta(1 - \tanh^2(\beta\lambda^N(t))) + 1\right] + O(1), \end{aligned}$$

where in the equality we have used $x \cdot \operatorname{sign}(x) = |x|$. The $O(1)$ terms are estimated by exploiting the diffusive approximation (4.12). Observing that, for $\beta < 1$, the function $1 - \beta(1 - \tanh^2(\beta\lambda))$ has a global minimum in 0 given by $1 - \beta$, we have found

$$N\mathcal{L}^N|y^N(t)|^k \le -C(\beta, k)N|y^N(t)|^k + O(1), \tag{4.17}$$

with $C(\beta, k) := 2k(1 - \beta) > 0$. By definition of $\mathcal{L}^N$, (4.17) implies

$$\frac{d}{dt}\mathbb{E}\Big[|y^N(t)|^k\Big] \le -C(\beta, k)N\mathbb{E}\Big[|y^N(t)|^k\Big] + O(1),$$

which, integrating both sides gives

$$\mathbb{E}\Big[|y^N(t)|^k\Big] \le e^{-C_1Nt}\mathbb{E}\Big[|y^N(0)|^k\Big] - \frac{C_2}{N}e^{-C_1Nt} + \frac{C_2}{N}.$$

Thus, $\sup_{t\geq 0}\mathbb{E}\Big[|y^N(t)|^k\Big] \leq \mathbb{E}\Big[|y^N(0)|^k\Big] + \frac{C}{N}$. Note that by the assumptions on the initial data we have by a LLN that $\mathbb{E}\Big[|y^N(0)|^k\Big] \to 0$ for $N \to +\infty$. For getting the stronger convergence (4.16) we refer to Section 4 of [28] for the diffusive case and to the Appendix of [32] for a general proof for jump processes, where their results imply here that, for any $\delta > 0$,

$$\mathbb{P}\left[\sup_{t\in[0,T]} |y^N(t)|^k > \delta\right] \to 0,$$

for $N \to +\infty$. Since $|y^N(t)|^k$ is uniformly bounded (4.16) follows. □

Remark 4.3. *For $\beta > 1$, when $\frac{d}{d\lambda}(\tanh(\beta\lambda)) < 1$ we can repeat the previous arguments to obtain an estimate as* (4.17)*. To be more precise, for any $\delta > 0$ we can find an $\varepsilon > 0$ such that $\frac{d}{d\lambda}\left[\tanh(\beta(\lambda_a(\beta)+\delta))\right] = \frac{d}{d\lambda}\left[\tanh(-\beta(\lambda_a(\beta)+\delta))\right] = 1-\varepsilon$, and $\frac{d}{d\lambda}\left[\tanh(\beta\lambda)\right] < 1-\varepsilon$ for any $\lambda \in (-\infty, -\lambda_a(\beta)-\delta) \cup (\lambda_a(\beta)+\delta, +\infty)$. Then, for any (λ, m) satisfying the above conditions we have, denoting $y := m - \tanh(\beta\lambda)$,*

$$N\mathcal{L}^N|y|^k \leq -C(\delta,\beta,k,\varepsilon)N|y|^k + O(1), \tag{4.18}$$

with $C(\delta,\beta,k,\varepsilon) > 0$ if and only if $\lambda \in (-\infty, -\lambda_a(\beta)-\delta) \cup (\lambda_a(\beta)+\delta, +\infty)$.

4.2.3 The subcritical case: $\beta < 1$

In this section we employ the result of Proposition 4.2 to obtain the convergence of the sequence of the accelerated processes $(\lambda^N(t), m^N(t))_{t\geq 0}$ to some limit random process $(\lambda(t), m(t))_{t\geq 0}$ in the subcritical case $\beta < 1$. For convenience and coherence with the further analyses, we state the main result of the section (Proposition 4.4) for the variables $(x^N(t), m^N(t))_{t\geq 0}$, whose infinitesimal accelerated generator is given by $N\mathcal{L}^N$, with $\mathcal{L}^N$ as in (4.8). To be precise, $(x^N(t))_{t\geq 0}$ satisfies

$$\begin{cases} dx^N(t) = \sigma dW^N(t), \\ x^N(0) \sim \mathcal{N}\left(x_0, \frac{1}{N}\sigma^2\right), \end{cases} \tag{4.19}$$

with W^N the Brownian motion $W^N(t) := \frac{1}{\sqrt{N}}\sum_{i=1}^N W_i(t)$, while $(m^N(t))_{t\geq 0}$ is given as in (4.6) but with rates multiplied by N, i.e.

$$\begin{cases} m^N(t) \mapsto m^N(t) \pm \frac{2}{N} \quad \text{rate} \quad N^2\frac{1\mp m^N(t)}{2}\Big(1 \pm \tanh\Big(\beta(x^N(t)+m^N(t))\Big)\Big), \\ m^N(0) = \frac{1}{N}\mathrm{Bin}(Np). \end{cases} \tag{4.20}$$

We show below that the limit process for the sequence $(x^N(t), m^N(t))_{t\geq 0}$ is given by

$$\begin{cases} m(t) = \tanh(\beta(x(t)+m(t))), \\ dx(t) = \sigma dW(t), \\ m(0) = m_0 \in [-1,1], \\ x(0) = x_0 \in \mathbb{R}, \end{cases} \tag{4.21}$$

with $m_0 = 2p-1$ and W a Brownian motion. In the subcritical case, Equation (4.21) is well-posed. Indeed, for $\beta < 1$, the relation $m(t) = \tanh(\beta(x(t)+m(t)))$ can be made explicit so that $m(t) = \varphi(x(t))$ for some function $\varphi : \mathbb{R} \to [-1,1]$ (see also Proposition 4.5 below).

Proposition 4.4 (Subcritical order N mean field limit dynamics)**.** *Let $T > 0$ and $\beta < 1$. Then, $(x^N(t), m^N(t))_{t\in[0,T]}$ converges for $N \to +\infty$, in the sense of weak convergence of stochastic processes, to $(x(t), m(t))_{t\in[0,T]}$, the solution to* (4.21)*.*

Proof. We plug in the definition (4.19) of $x^N(t)$ the *same* Brownian motion $W(t)$ appearing in the definition (4.21) of $x(t)$. We then prove, for the resulting processes

$$\mathbb{E}\left[\sup_{t\in[0,T]} |m^N(t) - m(t)|\right] \to 0, \tag{4.22}$$

$$\mathbb{E}\left[\sup_{t\in[0,T]} |x^N(t) - x(t)|\right] \to 0, \tag{4.23}$$

for $N \to +\infty$. Since $W^N \overset{\mathcal{D}}{=} W$ for every N, as they are both Brownian motions, (4.22) and (4.23) imply the desired convergence in distribution between the processes. Limit (4.23) is trivial, since the dynamics of $x^N(t)$ in the accelerated scale is

$$x^N(t) = x^N(0) + \sigma \int_0^t dW(t),$$

and $x^N(0) \to x(0)$ by a LLN. For (4.22), we estimate

$$\begin{aligned}\mathbb{E}\left[\sup_{t\in[0,T]} \left|m^N(t) - m(t)\right|\right] &\leq \mathbb{E}\left[\sup_{t\in[0,T]} \left|m^N(t) - \tanh(\beta(x^N(t) + m^N(t)))\right|\right] \\ &+ \mathbb{E}\left[\sup_{t\in[0,T]} \left|\tanh(\beta(x^N(t) + m^N(t))) - m(t)\right|\right].\end{aligned}$$

The first term in the right hand side tends to 0 thanks to (4.16) for $k = 1$. For the second term, using Equation (4.21) for $m(t)$, we have

$$\begin{aligned}&\mathbb{E}\left[\sup_{t\in[0,T]} \left|\tanh(\beta(x^N(t) + m^N(t))) - m(t)\right|\right] \\ &= \mathbb{E}\left[\sup_{t\in[0,T]} \left|\tanh(\beta(x^N(t) + m^N(t))) - \tanh(\beta(x(t) + m(t)))\right|\right] \\ &\leq \beta\mathbb{E}\left[\sup_{t\in[0,T]} |x^N(t) - x(t)|\right] + \beta\mathbb{E}\left[\sup_{t\in[0,T]} |m^N(t) - m(t)|\right],\end{aligned}$$

where in the inequality we have used the global Lipschitz continuity of $\tanh(\cdot)$. Thus, recollecting the above estimates

$$(1-\beta)\mathbb{E}\left[\sup_{t\in[0,T]} |m^N(t) - m(t)|\right] \leq \beta\mathbb{E}\left[\sup_{t\in[0,T]} |x^N(t) - x(t)|\right] \to 0,$$

for $N \to +\infty$. □

We conclude this section by noting that, in the subcritical regime $\beta < 1$, we can furthermore obtain an explicit one-dimensional description of the limit process $m(t)$. Indeed, in the dynamics (4.21), the only randomness is due to the diffusion $x(t)$, while $m(t)$ is slaved to be onto the invariant curve. A standard application of Itô's formula shows that

Proposition 4.5 (Limit diffusion)**.** *The process* $(m(t))_{t\geq 0}$ *defined in* (4.21) *is a strong solution to*

$$\begin{cases} dm(t) = -\frac{\beta^2\sigma^2 m(t)\left(1-m^2(t)\right)}{(1-\beta(1-m^2(t)))^3}dt + \frac{\sigma\beta(1-m^2(t))}{1-\beta(1-m^2(t))}dW(t), \\ m(0) = m_0 \in [-1,1]. \end{cases} \tag{4.24}$$

Proof. By Equation (4.21), $m(t)$ can be written as an *explicit* function of $x(t)$, and thus its dynamics must be of the form

$$dm(t) = a(t, m(t))dt + b(t, m(t))dW(t)$$

for some functions $a, b : [0,\infty) \times [-1,1] \to \mathbb{R}$ to be determined, and $W(t)$ is the same Brownian motion appearing in the dynamics of $x(t)$. By applying Itô's formula to the function $\tanh(\beta(x(t)+m(t))$, we find

$$\begin{aligned} dm(t) &= d\left\{\tanh\beta(x(t)+m(t))\right\} \\ &= \beta[1-\tanh^2\beta(x(t)+m(t))](dx(t)+dm(t)) \\ &- \beta^2\tanh\beta(x(t)+m(t))[1-\tanh^2\beta(x(t)+m(t))](b(t,m(t))+\sigma)^2dt \\ &= \beta(1-m^2(t))(dx(t)+dm(t)) - \beta^2 m(t)(1-m^2(t))(b(t,m(t))+\sigma)^2dt \\ &= \beta(1-m^2(t))(\sigma dW(t) + a(t,m(t))dt + b(t,m(t))dW(t)) \\ &- \beta^2 m(t)(1-m^2(t))(b(t,m(t))+\sigma)^2dt \\ &= \left[\beta(1-m^2(t))a(t,m(t)) - \beta^2 m(t)(1-m^2(t))(b(t,m(t))+\sigma)^2\right]dt \\ &+ \beta(1-m^2(t))\left[\sigma + b(t,m(t))\right]dW(t). \end{aligned}$$

By reading the diffusion coefficient from the last line, we must have

$$b(t,m(t)) = \beta(1-m^2(t))[\sigma + b(t,m(t))],$$

and thus

$$b(t,m(t)) = b(m(t)) = \frac{\sigma\beta(1-m^2(t))}{1-\beta(1-m^2(t))}.$$

For the drift term instead

$$a(t,m(t)) = \beta(1-m^2(t))a(t,m(t)) - \beta^2 m(t)(1-m^2(t))[(b(t,m(t))+\sigma)^2]. \tag{4.25}$$

Using the expression found for $b(t,m(t))$, we have that

$$\begin{aligned} (b(t,m(t))+\sigma)^2 &= b^2(t,m(t)) + \sigma^2 + 2\sigma b(t,m(t)) \\ &= \frac{\sigma^2\beta^2(1-m^2(t))^2}{(1-\beta(1-m^2(t)))^2} + \sigma^2 + \frac{2\sigma^2\beta(1-m^2(t))}{1-\beta(1-m^2(t))} \\ &= \frac{\sigma^2\beta^2(1-m^2(t))^2 + \sigma^2(1-\beta(1-m^2(t)))^2 + 2\sigma^2\beta(1-m^2(t))(1-\beta(1-m^2(t))}{(1-\beta(1-m^2(t)))^2} \\ &= \frac{\sigma^2}{(1-\beta(1-m^2(t)))^2}, \end{aligned}$$

and thus, reading from (4.25),

$$a(t,m(t))(1-\beta(1-m^2(t))) = -\beta^2 m(t)(1-m^2(t))\frac{\sigma^2}{(1-\beta(1-m^2(t)))^2},$$

so that we can conclude. □

Remark 4.6. *For $\beta < 1$, the SDE* (4.24) *is well-posed. Existence follows by Proposition 4.5. Uniqueness follows by the Lipschitz properties of the drift and diffusion functions in* $[-1, 1]$. *Indeed, note that Equation* (4.24) *defines a dynamics in* $[-1, 1]$, *due to the sign of the drift at the borders of* $(-1, 1)$ *and to the fact that the diffusion is zero at the borders of* $(-1, 1)$.

4.2.4 The supercritical case: $\beta > 1$

In this section we deal with the analysis of the supercritical case $\beta > 1$. The main result is the following convergence theorem:

Theorem 4.7 (Supercritical order N mean field limit dynamics)**.** *Fix $T > 0$, $\beta > 1$, and let $(x^N(t), m^N(t))_{t\in[0,T]}$ be the accelerated processes defined in* (4.19) *and* (4.20)*, with $x^N(0) \overset{\mathcal{D}}{\to} x_0 > \lambda_a(\beta) - m_a(\beta)$ and $m^N(0) \overset{\mathcal{D}}{\to} m_0 > m_a(\beta)$, or $x^N(0) \overset{\mathcal{D}}{\to} x_0 < m_a(\beta) - \lambda_a(\beta)$ and $m^N(0) \overset{\mathcal{D}}{\to} m_0 < -m_a(\beta)$, with $(\lambda_a(\beta), m_a(\beta))$ as in* (4.10)*. Then, the accelerated sequence of processes $(m^N(t))_{t\in[0,T]}$ converges weakly in the sense of stochastic processes, for $N \to +\infty$, to the process which solves the following SDE*

$$dm(t) = \mathbb{1}_{|m(t)|>m_a}\left(-\frac{\beta^2\sigma^2 m(t)\left(1-m^2(t)\right)}{\left(1-\beta(1-m^2(t))\right)^3}dt + \frac{\sigma\beta(1-m^2(t))}{1-\beta(1-m^2(t))}dW(t)\right) \tag{4.26}$$
$$+ (m_b + m_a)\mathbb{1}_{m(t)=-m_a} - (m_b + m_a)\mathbb{1}_{m(t)=m_a},$$

with $m(0) = m_0$, and $m_b := m_b(\beta)$ is the solution in y to

$$g(y) := 2\beta y - 2\beta(m_a(\beta) - \lambda_a(\beta)) - \log(1 + y) + \log(1 - y) = 0. \tag{4.27}$$

In particular, in Section 4.2.4.1 we derive heuristically the limit dynamics, while in Section 4.2.4.2 we address the rigorous proof of convergence.

4.2.4.1 Heuristic limit

In this section we describe on a heuristic level the effective motion taking place on the invariant curve for the supercritical regime $\beta > 1$. As highlighted in Remark 4.3, in this case the N-particle dynamics is contractive only in the union of the two intervals where $1 - \beta(1 - \tanh^2(\beta\lambda)) > 0$, i.e. for $\lambda > \lambda_a(\beta)$ or $\lambda < -\lambda_a(\beta)$, which we refer to as the *stable* components of the invariant curve. For the heuristic argument, we consider again the approximate diffusive system (4.13) for $(\tilde{\lambda}^N, \tilde{m}^N)$ at times of order 1, which is of the form

$$\begin{cases} dx_1(t) = -2(x_2(t) - f(x_1(t)))dt + \frac{1}{\sqrt{N}}\left(\sigma_{11}dB_1(t) + \sigma_{12}dB_2(t)\right) \\ dx_2(t) = -2(x_2(t) - f(x_1(t)))dt + \frac{1}{\sqrt{N}}\left(\sigma_{21}dB_1(t) + \sigma_{22}dB_2(t)\right), \end{cases} \tag{4.28}$$

denoting $(x_1, x_2) := (\tilde{\lambda}^N, \tilde{m}^N)$. Recall that, in our case, $f(x_1) = \tanh(\beta x_1)$, and the diffusion coefficients can be read off from (4.14). We want to derive a limit one-dimensional diffusion for each variable, which also contains the jump components illustrated in Figure 4.2 for $\beta > 1$. In fact, when the dynamics hits the critical points, we expect to see an *instantaneous* jump to the point given by the intersection between the vector field line passing through the critical point, and the invariant curve.

In order to derive an expression for the drift and diffusion coefficients, we follow the approach highlighted in [88], where they analyze the case of a *globally attractive* invariant manifold for the dynamics, under diffusive fluctuations of smaller order. Their approach should thus work in our case only for the stable components of the invariant curve in the supercritical case, and rigorously in the whole space in the subcritical case.

Remark 4.8. *Because of Remark 4.1, one can apply the same arguments below to the subcritical case $\beta < 1$ to obtain an alternative proof of Proposition 4.4 and Proposition 4.5 for the shape of the limiting subcritical dynamics.*

As in [88], if we take the point $\boldsymbol{x} = (x_1, x_2)$ to be the current location of the process governed by equation (4.28), the limit SDE on the invariant manifold is obtained by applying Itô's formula to:

$$\pi(\boldsymbol{x}) := \lim_{t\to\infty} \xi_{\boldsymbol{x}}(t),$$

where $\xi_{\boldsymbol{x}}(t)$ is the deterministic trajectory solving System (4.9), with initial datum $(\lambda_0, m_0) = \boldsymbol{x}$. The point $\pi(\boldsymbol{x})$ is in our case given by the intersection between the vector field line passing through $\boldsymbol{x}$ and the invariant curve.

Itô's formula applied to each component $\pi_i(\boldsymbol{x})$, $i = 1, 2$, yields

$$d\pi_i(\boldsymbol{x}) = \frac{1}{2N}\sum_{j,k,l=1}^{2} \sigma_{jl}(\boldsymbol{x})\sigma_{kl}(\boldsymbol{x})\frac{\partial^2 \pi_i(\boldsymbol{x})}{\partial x_j \partial x_k}dt + \frac{1}{\sqrt{N}}\sum_{j,l=1}^{2} \sigma_{jl}(\boldsymbol{x})\frac{\partial \pi_i(\boldsymbol{x})}{\partial x_j}dB_l(t).$$

Note that, in order this to be fully rigorous, one should find an equation closed in a variable $\tilde{\boldsymbol{x}}$ which lives on the invariant curve. For the details we again refer to [88].

In the case of a one-dimensional invariant manifold, one can explicitly compute the coefficients of the limit diffusion written above by performing a second-order approximation of the $\pi_i(\boldsymbol{x})$'s (see [88, pp. 6-9]). Denote by $\gamma(x_1) = (x_1, \tanh(\beta x_1))$ the points on the invariant curve parametrised in the first coordinate. Then, one can compute

$$\begin{aligned}
\frac{\partial}{\partial x_1}\pi_1(\boldsymbol{x})\bigg|_{\boldsymbol{x}=\gamma(x_1)} &= \frac{1}{1-\beta(1-\tanh^2(\beta x_1))},\\
\frac{\partial}{\partial x_2}\pi_1(\boldsymbol{x})\bigg|_{\boldsymbol{x}=\gamma(x_1)} &= -\frac{1}{1-\beta(1-\tanh^2(\beta x_1))},\\
\frac{\partial^2}{\partial x_1\partial x_1}\pi_1(\boldsymbol{x})\bigg|_{\boldsymbol{x}=\gamma(x_1)} &= \frac{1}{1-\beta(1-\tanh^2(\beta x_1))}\left(-\frac{2\beta^2\tanh(\beta x_1)(1-\tanh^2(\beta x_1))}{(1-\beta(1-\tanh^2(\beta x_1)))^2}\right)\\
\frac{\partial^2}{\partial x_1\partial x_2}\pi_1(\boldsymbol{x})\bigg|_{\boldsymbol{x}=\gamma(x_1)} &= -\frac{\partial^2}{\partial x_1\partial x_1}\pi_1(\boldsymbol{x})\bigg|_{\boldsymbol{x}=\gamma(x_1)}\\
\frac{\partial^2}{\partial x_2\partial x_2}\pi_1(\boldsymbol{x})\bigg|_{\boldsymbol{x}=\gamma(x_1)} &= \frac{\partial^2}{\partial x_1\partial x_1}\pi_1(\boldsymbol{x})\bigg|_{\boldsymbol{x}=\gamma(x_1)}.
\end{aligned}$$

Finally, observing that, for a point $\boldsymbol{x} = \gamma(x_1)$ on the invariant curve, the covariance matrix $(\sigma^2)_{ij}$ written in (4.14) reduces to

$$\sigma^2(\gamma(x_1)) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & 0 \end{pmatrix},$$

we get that the limit diffusion in the accelerated first variable $x_1(t) \in \mathbb{R}$, corresponding to $\lambda(t)$, is given by

$$dx_1(t) = -\frac{\sigma^2\beta^2\tanh(\beta x_1(t))(1-\tanh^2(\beta x_1(t)))}{(1-\beta(1-\tanh^2(\beta x_1(t))))^3}dt + \frac{\sigma}{1-\beta(1-\tanh^2(\beta x_1(t)))}dB(t). \tag{4.29}$$

Similarly, parametrising the points on the invariant curve with respect to the second coordinate, i.e. $\boldsymbol{x} = \gamma(x_2) = (\frac{1}{\beta}\text{arctanh}(x_2), x_2)$, one gets the limit diffusion for $x_2(t) \in [-1, 1]$, corresponding to the magnetization $m(t)$,

$$dx_2(t) = -\frac{\sigma^2\beta^2 x_2(t)(1-x_2^2(t))}{(1-\beta(1-x_2^2(t)))^3}dt + \frac{\sigma\beta(1-x_2^2(t))}{1-\beta(1-x_2^2(t))}dB(t), \tag{4.30}$$

which is indeed the same SDE as (4.24). Note that both the drift and diffusion coefficients explode in the critical points of the invariant curve.

Denoting with $c(\cdot)$ the drift function and with $\sqrt{g(\cdot)}$ the diffusion coefficient, we get that the global limiting accelerated one-dimensional dynamics, written in either of the two variables x_1 or x_2, should be of the form

$$dX(t) = \mathbb{1}_{|X(t)|>a}\left(\sqrt{g(X(t))}dW(t) + c(X(t))dt\right) + (b+a)\mathbb{1}_{X(t)=-a} - (b+a)\mathbb{1}_{X(t)=a}, \tag{4.31}$$

where the point $a = a(\beta)$ is the critical (positive) point on the invariant curve, and the point $b = b(\beta)$ (resp. $-b$) is the intersection between the curve and the vector field line passing through $-a$ (resp. a).

Remark 4.9 (Limit case $\beta \to \infty$). *When $\beta \to \infty$, the limit dynamics for the accelerated magnetization $m(t)$ is expected to be a spin-valued jump process $m(t) \in \{-1, 1\}$ with non-exponentially distributed random interarrival jump times, with their distribution being the one of the hitting times of a Brownian motion with diffusion coefficient $\sigma > 0$. Indeed, the critical points (see Eq. (4.10)) tend to ± 1 in the m-variable, and to 0 in the λ-variable, while the diagonal line $x(t) = \lambda(t) - m(t)$, determining when the process jumps, still evolves according to a Brownian motion with diffusion coefficient $\sigma > 0$. This observation served as a further motivation for studying the model of interacting spin-valued renewal processes of Chapter 3, Section 3.2, which was originally thought of as a two-level hierarchical model for the zero-temperature regime $\beta = \infty$.*

4.2.4.2 The convergence argument

We now address the full proof of convergence to the limit dynamics for $\beta > 1$, given in Theorem 4.7. As we did above for the subcritical case, we consider the dynamics in the alternative variables (x^N, m^N), whose generator, we recall, is given by (4.8). Recall that the variable x^N, the intersection between the diagonal line (at 45 degrees) passing through the point (λ^N, m^N) and the λ-axis, follows a Brownian motion, while m^N is a jump process depending on x^N: if we think of the latter as being deterministic and fixed, such motion is a unidimensional continuous-time Markov chain taking place along the diagonal line parametrized by the fixed value $x^N = x$, which is attractive towards the invariant curve. The limit dynamics is thus the projection of the combination of these two motions on the invariant curve. We divide the proof of Theorem 4.7 in three lemmas. In the following proofs we assume that $x_0 > \lambda_a - m_a$ and $m_0 > m_a$. For the symmetry of the problem the case $x_0 < m_a - \lambda_a$ and $m_0 < -m_a$ is analogous.

Lemma 4.10. *Let*

$$T^{\varepsilon}_{m_a} := \inf\left\{t \geq 0 : x^N(t) = \lambda_a - m_a - \varepsilon\right\}, \tag{4.32}$$

for $\varepsilon \in \mathbb{R}$. Then,

$$\mathbb{P}(T^{\varepsilon}_{m_a} < \infty) = 1. \tag{4.33}$$

Proof. Recall that x^N evolves as in (4.19). For the proof, we assume for simplicity that $x^N(0) = x_0$ (otherwise, we just add an additional term in the variance at time t, accounting for the initial variance - which is small in N).We thus have that $x^N(t) \sim \mathcal{N}(0, \sigma^2 t)$, and we can get explicitly the distribution of $T^{\varepsilon}_{m_a}$ in a classic way, using the reflection principle for the Brownian motion. Indeed, we have that for any $t \geq 0$,

$$\begin{aligned}\mathbb{P}(T^{\varepsilon}_{m_a} \leq t) &= \mathbb{P}(\inf_{0\leq s\leq t} x^N(s) \leq \lambda_a - m_a - \varepsilon) = 2\mathbb{P}(x^N(t) \leq \lambda_a - m_a - \varepsilon) \\ &= \frac{2}{\sqrt{2\pi\sigma^2 t}} \int_{-\infty}^{\lambda_a - m_a - \varepsilon} e^{-\frac{(x-x_0)^2}{2\sigma^2 t}} dx = \left[z = \frac{x - x_0}{\sqrt{t}}\right] \\ &= \frac{2}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\frac{\lambda_a - m_a - \varepsilon - x_0}{\sqrt{t}}} e^{-\frac{z^2}{2\sigma^2}} dz.\end{aligned}$$

By taking the derivative with respect to t of the previous expression we get that $T^{\varepsilon}_{m_a}$ has density

$$f_{T^{\varepsilon}_{m_a}}(t) = \frac{(\lambda_a - m_a - \varepsilon - x_0)}{\sqrt{2\pi\sigma^2}} \frac{1}{t^{3/2}} e^{-\frac{(\lambda_a - m_a - \varepsilon - x_0)^2}{2\sigma^2 t}},$$

and, as one can check

$$\mathbb{P}(T^{\varepsilon}_{m_a} < \infty) = \int_0^{\infty} f_{T^{\varepsilon}_{m_a}}(t) dt = 1,$$

so that (4.33) is verified. □

Lemma 4.10 tells us that, almost surely, the process $x^N(t)$ reaches in a finite time the point $\lambda_a - m_a - \varepsilon$, which corresponds - up to an ε error - to the critical point on the invariant curve we discussed in the previous section. The following two lemmas respectively describe the limit equation for the times preceding and following the hitting time $T^{\varepsilon}_{m_a}$. For $t < T^{-\delta}_{m_a}$, for some $\delta > 0$, we can proceed similarly as in Propositions 4.4 and 4.5 since the contraction estimates of Remark 4.3 are holding, while for $t > T^{\varepsilon}_{m_a}$ for some $\varepsilon > 0$ we capture the jumps via a direct estimate. We then conclude by the continuity with respect to ε and δ of the hitting times distributions $T^{\varepsilon}_{m_a}$, $T^{-\delta}_{m_a}$.

Lemma 4.11. *Fix $T, \delta > 0$. Let $T^{-\delta}_{m_a} := \inf\left\{t \geq 0 : x^N(t) = \lambda_a - m_a + \delta\right\}$. Let*

$$\left(m^N(t \wedge T^{-\delta}_{m_a})\right)_{t\in[0,T]}$$

denote the accelerated stopped process, with initial conditions as in Theorem 4.7. Then, $\left(m^N(t \wedge T^{-\delta}_{m_a})\right)_{t\in[0,T]}$ converges weakly in the sense of stochastic processes, for $N \to +\infty$, to $\left(m(t \wedge T^{-\delta}_{m_a})\right)_{t\in[0,T]}$, with $(m(t))_{t\geq 0}$ the solution to (4.24) with the same initial conditions as in Theorem 4.7, and $\left(m(t \wedge T^{-\delta}_{m_a})\right)_{t\in[0,T]}$ its stopped version.

Proof. As in the proof of Proposition 4.4, we plug in the definition (4.19) of $x^N(t)$ the *same* Brownian motion $W(t)$ appearing in the definition (4.21) of $x(t)$. Let $T^{-\delta}_{m_a}$ be the resulting stopping time: we prove, for the resulting processes

$$\mathbb{E}\left[\sup_{t\in[0,T]} \left| m^N(t \wedge T^{-\delta}_{m_a}) - m(t \wedge T^{-\delta}_{m_a}) \right| \right] \to 0, \tag{4.34}$$

for $N \to +\infty$, which implies the result in distribution by reasoning as in Proposition 4.4. When $t < T^{-\delta}_{m_a}$ we have that $x^N(t) > \lambda_a - m_a + \delta$. Thus, we are in the stable component of the invariant curve.

From (4.23), it follows that

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left|x^N(t\wedge T_{m_a}^{-\delta})-x(t\wedge T_{m_a}^{-\delta})\right|\right]\to 0, \tag{4.35}$$

for $N\to+\infty$. For (4.34), denoting the event

$$A:=\left\{\min_{t\in[0,T]}\lambda^N(t\wedge T_{m_a}^{-\delta})>\lambda_a+\delta\right\},$$

we estimate,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]}\left|m^N(t\wedge T_{m_a}^{-\delta})-m(t\wedge T_{m_a}^{-\delta})\right|\right] \qquad (4.36)\\
&=\mathbb{E}\left[\sup_{t\in[0,T]}\left|m^N(t\wedge T_{m_a}^{-\delta})-m(t\wedge T_{m_a}^{-\delta})\right|\mathbb{1}_A\right]\\
&\quad+\mathbb{E}\left[\sup_{t\in[0,T]}\left|m^N(t\wedge T_{m_a}^{-\delta})-m(t\wedge T_{m_a}^{-\delta})\right|\mathbb{1}_{\exists t\in[0,T]:\lambda^N(t\wedge T_{m_a}^{-\delta})<\lambda_a+\delta}\right]\\
&\le\mathbb{E}\left[\sup_{t\in[0,T]}\left|m^N(t\wedge T_{m_a}^{-\delta})-\tanh(\beta(x^N(t\wedge T_{m_a}^{-\delta})+m^N(t\wedge T_{m_a}^{-\delta})))\right|\mathbb{1}_A\right]\\
&\quad+\mathbb{E}\left[\sup_{t\in[0,T]}\left|\tanh(\beta(x^N(t\wedge T_{m_a}^{-\delta})+m^N(t\wedge T_{m_a}^{-\delta})))-m(t\wedge T_{m_a}^{-\delta})\right|\mathbb{1}_A\right].\\
&\qquad+2\mathbb{P}\Big(\exists t\in[0,T]:\lambda^N(t\wedge T_{m_a}^{-\delta})<\lambda_a+\delta\Big),
\end{aligned}$$

where in the last line we have used the boundedness of the integrands. The first term in the right hand side of the above inequality tends to 0 thanks to estimate (4.18) of Remark 4.3 for $k=1$, which can be applied for any $\lambda>\lambda_a+\delta$, and to the same argument used for the proof of Proposition 4.2. For the second term in the right hand side of inequality (4.36), using Equation (4.21) for $m(t\wedge T_{m_a}^{-\delta})$, we have

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]}\left|\tanh(\beta(x^N(t\wedge T_{m_a}^{-\delta})+m^N(t\wedge T_{m_a}^{-\delta})))-m(t\wedge T_{m_a}^{-\delta})\right|\mathbb{1}_A\right]\\
&=\mathbb{E}\left[\sup_{t\in[0,T]}\left|\tanh(\beta(x^N(t\wedge T_{m_a}^{-\delta})+m^N(t\wedge T_{m_a}^{-\delta})))-\right.\right.\\
&\qquad\left.\left.-\tanh(\beta(x(t\wedge T_{m_a}^{-\delta})+m(t\wedge T_{m_a}^{-\delta})))\right|\mathbb{1}_A\right]\\
&\le(1-\varepsilon)\mathbb{E}\left[\sup_{t\in[0,T]}\left|x^N(t\wedge T_{m_a}^{-\delta})-x(t\wedge T_{m_a}^{-\delta})\right|\mathbb{1}_A\right]\\
&\qquad+(1-\varepsilon)\mathbb{E}\left[\sup_{t\in[0,T]}\left|m^N(t\wedge T_{m_a}^{-\delta})-m(t\wedge T_{m_a}^{-\delta})\right|\mathbb{1}_A\right]\\
&\le(1-\varepsilon)\mathbb{E}\left[\sup_{t\in[0,T]}\left|x^N(t\wedge T_{m_a}^{-\delta})-x(t\wedge T_{m_a}^{-\delta})\right|\right]\\
&\qquad+(1-\varepsilon)\mathbb{E}\left[\sup_{t\in[0,T]}\left|m^N(t\wedge T_{m_a}^{-\delta})-m(t\wedge T_{m_a}^{-\delta})\right|\right],
\end{aligned}$$

where in the first inequality we have used that, by the properties of $\tanh(\cdot)$ and by definition of λ_a, there exists an $\varepsilon > 0$ such that $\frac{d}{d\lambda}\tanh(\beta\lambda) < 1-\varepsilon$ for every $\lambda > \lambda_a + \delta$. Finally, the third term in the right hand side of (4.36) can be estimated as follows

$$\begin{aligned} & 2\mathbb{P}\Big(\exists t \in [0,T] : \lambda^N(t\wedge T^{-\delta}_{m_a}) < \lambda_a + \delta\Big) \qquad (4.37)\\ &= 2\mathbb{P}\Big(\exists t \in [0,T] : x^N(t\wedge T^{-\delta}_{m_a}) + m^N(t\wedge T^{-\delta}_{m_a}) < \lambda_a + \delta\Big)\\ &= 2\mathbb{P}\Big(\exists t \in [0,T] : x^N(t\wedge T^{-\delta}_{m_a}) < \lambda_a - m^N(t\wedge T^{-\delta}_{m_a}) + \delta\Big)\\ &\leq 2\mathbb{P}\Big(\exists t \in [0,T] : m^N(t\wedge T^{-\delta}_{m_a}) < m_a\Big), \end{aligned}$$

where the inequality follows by the definition of $T^{-\delta}_{m_a}$. To bound the latter, we introduce an auxiliary process $(\tilde{m}^N(t))_{t\in[0,T]}$, coupled with $(x^N(t), m^N(t))_{t\in[0,T]}$, with dynamics

$$\begin{cases} \tilde{m}^N(t) \mapsto \tilde{m}^N(t) \pm \frac{2}{N} \quad \text{rate} \quad N^2\frac{1\mp\tilde{m}^N(t)}{2}\Big(1\pm\tanh\Big(\beta(\lambda_a - m_a + \delta + \tilde{m}^N(t))\Big)\Big), \\ \tilde{m}^N(0) = m^N(0), \end{cases}$$

and consider its stopped version $\Big(\tilde{m}^N(t\wedge T^{-\delta}_{m_a})\Big)_{t\in[0,T]}$. Since, by definition of $T^{-\delta}_{m_a}$, it holds $x^N(t\wedge T^{-\delta}_{m_a}) \geq \lambda_a - m_a + \delta$, we have that the rate of increase of $m^N(t\wedge T^{-\delta}_{m_a})$ is bigger than the rate of increase of $\tilde{m}^N(t\wedge T^{-\delta}_{m_a})$; symmetrically, the rate of decrease of $m^N(t\wedge T^{-\delta}_{m_a})$ is smaller than the rate of decrease of $\tilde{m}^N(t\wedge T^{-\delta}_{m_a})$. We thus have, for any $t\in[0,T]$, $N\in\mathbb{N}$, $\overline{m}\in[-1,1]$,

$$\mathbb{P}\Big(m^N(t\wedge T^{-\delta}_{m_a}) < \overline{m}\Big) \leq \mathbb{P}\Big(\tilde{m}^N(t\wedge T^{-\delta}_{m_a}) < \overline{m}\Big). \qquad (4.38)$$

Moreover, note that $\tilde{m}^N(t\wedge T^{-\delta}_{m_a})$ is a jump process with rates independent of x^N, starting above m_a with probability tending to 1 for $N\to+\infty$, and that it gets fastly attracted, for $N\to+\infty$, to the point m^* on the invariant curve identified by

$$\begin{cases} x = \lambda_a - m_a + \delta, \\ m = \tanh(\beta(x+m)), \end{cases}$$

for which it holds by construction $m_a < m^*$. Thus

$$\mathbb{P}\Big(\exists t\in[0,T] : \tilde{m}^N(t\wedge T^{-\delta}_{m_a}) < m_a\Big) \leq C(N),$$

for $C(N)\to 0$, when $N\to+\infty$. By the above observation (4.38), this implies the same bound for m^N in the last line of the right hand side of (4.37).

Thus, recollecting the above estimates from (4.36),

$$\begin{aligned} \mathbb{E}\left[\sup_{t\in[0,T]}\left|m^N(t\wedge T^{-\delta}_{m_a}) - m(t\wedge T^{-\delta}_{m_a})\right|\right] &\leq \frac{(1-\varepsilon)}{\varepsilon}\mathbb{E}\left[\sup_{t\in[0,T]}\left|x^N(t\wedge T^{-\delta}_{m_a}) - x(t\wedge T^{-\delta}_{m_a})\right|\right]\\ &+ C(N) \leq \frac{C(N)}{\varepsilon} \to 0, \end{aligned}$$

for $N\to+\infty$, where $C(N)$ is allowed to change from line to line. □

The next lemma deals with the times which follow the hitting time $T^{\varepsilon}_{m_a}$. Using the strong Markov's property, we can restart the dynamics from the point reached at the hitting time, assuming that we are above the invariant curve.

Lemma 4.12. *Fix $\varepsilon > 0$ such that $\lambda_a - \varepsilon > 0$. Let $(x^N(t), m^N(t))_{t\geq 0}$ be the accelerated processes, with initial data $(x^N(0), m^N(0)) = (x_0, m_0)$, such that $x_0 = \lambda_a - m_a - \varepsilon$ and $m_0 > \tanh\beta(x_0 + m_a)$. Let*

$$T_{\varepsilon/2} := \inf\left\{t > 0 : x^N(t) = \lambda_a - m_a - \frac{\varepsilon}{2}\right\},$$

and $T_{m_b} := \inf\left\{t > 0 : m^N(t) = m_b\right\}$, with m_b as in (4.27). Then,

$$\lim_{N\to\infty} \mathbb{P}(T_{m_b} < T_{\varepsilon/2}) = 1. \tag{4.39}$$

Proof. The proof makes extensive use of $(\nu^N(t))_{t\geq 0}$, an auxiliary CTMC - coupled with $(m^N(t))_{t\geq 0}$ - with the same initial datum m_0, whose transition rates are given by

$$\begin{aligned} \nu^N &\mapsto \nu^N + \frac{2}{N} \quad \text{with rate} \quad N^2\frac{1-\nu^N(t)}{2}\left[1 + \tanh(\beta(\nu^N(t) + \lambda_a - m_a - \varepsilon/2))\right] \\ \nu^N &\mapsto \nu^N - \frac{2}{N} \quad \text{with rate} \quad N^2\frac{1+\nu^N(t)}{2}\left[1 - \tanh(\beta(\nu^N(t) + \lambda_a - m_a - \varepsilon/2))\right]. \end{aligned} \tag{4.40}$$

Note that $(\nu^N(t))_{t\geq 0}$ is independent of $(x^N(t))_{t\geq 0}$. Moreover, setting

$$\tilde{T}_{m_b} := \inf\left\{t > 0 : \nu^N(t) = m_b\right\},$$

we have

$$\mathbb{P}(T_{m_b} < T_{\varepsilon/2}) \geq \mathbb{P}(\tilde{T}_{m_b} < T_{\varepsilon/2}). \tag{4.41}$$

Indeed, it is easy to check that for $t \leq T_{\varepsilon/2}$, for which $x^N(t) \leq \lambda_a - m_a - \varepsilon/2$, the rate of increase in the dynamics of $\nu^N(t)$ is greater than that of $m^N(t)$, while the opposite is true for the rate of decrease. Since $m_b < m_0$, (4.41) follows.

Consider now the slowed version of the process $\nu^N(t)$, i.e. $\tilde{\nu}^N(t) := \nu^N(tN^{-1})$, whose generator is

$$\begin{aligned} \mathcal{L}^N f(\tilde{\nu}) := {} & N\frac{1+\tilde{\nu}}{2}\left[1 - \tanh(\beta(\tilde{\nu} + \lambda_a - m_a - \varepsilon/2))\right]\left[f\left(\tilde{\nu} - \frac{2}{N}\right) - f(\tilde{\nu})\right] \\ & + N\frac{1-\tilde{\nu}}{2}\left[1 + \tanh(\beta(\tilde{\nu} + \lambda_a - m_a - \varepsilon/2))\right]\left[f\left(\tilde{\nu} + \frac{2}{N}\right) - f(\tilde{\nu})\right]. \end{aligned}$$

Expanding it to the first order, we find, up to terms of order $O\left(\frac{1}{N}\right)$,

$$\mathcal{L}^N f(\tilde{\nu}) \approx \left[-2\tilde{\nu} + 2\tanh(\beta(\tilde{\nu} + \lambda_a - m_a - \varepsilon/2))\right] f'(\tilde{\nu}).$$

This implies that, in the limit $N \to +\infty$, the process $(\tilde{\nu}^N(t))_{t\geq 0}$ weakly converges to the solution of the following ODE

$$\begin{cases} \frac{d}{dt}\overline{m}(t) = v(\overline{m}) = -2\overline{m}(t) + 2\tanh(\beta(\overline{m}(t) + \lambda_a - m_a - \varepsilon/2)) \\ \overline{m}(0) = m_0. \end{cases} \tag{4.42}$$

The vector field $v(m)$ in (4.42) is positive if and only if

$$f(m) := 2\beta m - 2\beta(m_a - \lambda_a) - \log(1+m) + \log(1-m) - \beta\varepsilon > 0. \tag{4.43}$$

Indeed, $v(m)$ is positive if and only if

$$m < \tanh(\beta(m + \lambda_a - m_a - \varepsilon/2)),$$

which is equivalent to

$$\frac{1}{\beta}\operatorname{arctanh}(m) < m + \lambda_a - m_a - \varepsilon/2.$$

By using the identity $\operatorname{arctanh}(m) = \frac{1}{2}\log\left(\frac{1+m}{1-m}\right)$ we get the desired inequality (4.43). With analogous steps we get the definition of g given by (4.27), for $\varepsilon = 0$. Recall that by our choice $m_0 > 0$. First of all, it is easy to see that $f(m) < 0$ whenever $m \geq 0$. Indeed, $f(0) < 0$, f has a local maximum in $m = m_a = \sqrt{1 - \frac{1}{\beta}}$ for which $f\left(\sqrt{1-\frac{1}{\beta}}\right) = -\beta\varepsilon < 0$, and $f(m) \to -\infty$ for $m \to +1$. Moreover, recalling the expression for $g(\cdot)$ in (4.27), we see that

$$f(m) = g(m) - \beta\varepsilon,$$

so that $f(m) < g(m)$ for all $m \in [-1, 1]$. Since $f'(m) = g'(m) = 2\beta - \frac{1}{1+m} - \frac{1}{1-m}$ we have that g has a local maximum at $m = m_a$, for which we have $g(m_a) = 0$, while $g(m) < 0$ for all $m > 0,\ m \neq m_a$. We also observe that:

- $\exists!\ m^*_{f,b}$ such that $f(m^*_{f,b}) = 0$;
- $g(m_b) = 0$ and $g(m) \neq 0 \ \ \forall m \neq m_a, m_b$;
- $g(m) > 0$ if $m < m_b$, $g(m) < 0$ if $m > m_b$;
- $f(m) > 0$ if $m < m^*_{f,b}$, $f(m) < 0$ if $m > m^*_{f,b}$;
- $m^*_{f,b} < m_b$;
- $m_b \to -1$ when $\beta \to \infty$.

In order to check the remarks, we note that, when $m \leq 0$,

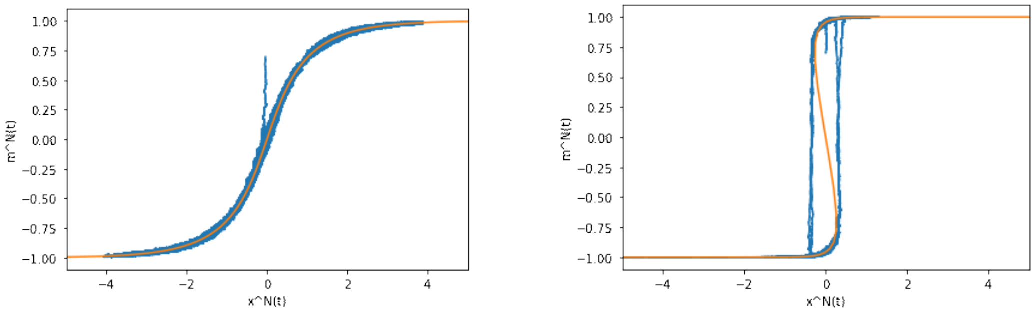
$$f'(m) = g'(m) > 0 \text{ iff } m < -m_a = -\sqrt{1 - \frac{1}{\beta}},$$

and $-m_a$ is a local minimum, for which $f(-m_a),\ g(-m_a) < 0$. Moreover, $f(m),\ g(m) \to +\infty$ for $m \to -1$. Combining these with the above considerations for $m \geq 0$, we deduce the first four bullet points. For the fact that $f(m) < g(m)$ we get the fifth remark, while for the last it is sufficient to observe that $m_b < -\sqrt{1 - \frac{1}{\beta}} \to -1$ for $\beta \to \infty$.

The above remarks and the convergence of $(\tilde{\nu}^N(t))_{t \geq 0}$ to the deterministic process $(\overline{m}(t))_{t \geq 0}$ imply that, if we define $\bar{T}_{m_b} := \inf\left\{t > 0 : \tilde{\nu}^N(t) = m_b\right\}$ and $\bar{T}_{m^*_{f,b}} := \inf\left\{t > 0 : \tilde{\nu}^N(t) = m^*_{f,b}\right\}$, we have that there exists a $C > 0$, independent of N, such that

$$\mathbb{P}(\bar{T}_{m_b} \leq C) \geq \mathbb{P}(\bar{T}_{m^*_{f,b}} \leq C) \to 1, \tag{4.44}$$

for $N \to +\infty$. Indeed, for the deterministic process $\overline{m}(t)$ the arrival time in $m^*_{f,b}$ (which is greater than the one for arriving in m_b) is for sure limited by a constant, because of the sign of the vector field of (4.42).



(a) $(x^N(t), m^N(t))$ subcritical case. (b) $(x^N(t), m^N(t))$ supercritical case.

Figure 4.3: Simulation of the finite N dynamics, for $N = 2000$, $\sigma = 2$, $\beta = 0.5$ (left), and $\beta = 2$ (right).

If we now consider the original auxiliary process $(\nu^N(t))_{t\geq 0}$, i.e. the sped up version of $\tilde{\nu}^N(t)$, we get that, defining $\tilde{T}_{m^*_{f,b}} := \inf\left\{t > 0 : \nu^N(t) = m^*_{f,b}\right\}$,

$$\mathbb{P}(\tilde{T}_{m_b} \leq C(N)) \geq \mathbb{P}(\tilde{T}_{m^*_{f,b}} \leq C(N)) \to 1, \tag{4.45}$$

for $N \to +\infty$, with $C(N) \to 0$, by means of (4.44).

We can finally conclude the proof of (4.39), by estimating

$$\mathbb{P}(T_{m_b} < T_{\varepsilon/2}) \geq \mathbb{P}(\tilde{T}_{m_b} < T_{\varepsilon/2}) \geq \mathbb{P}(\tilde{T}_{m^*_{f,b}} < T_{\varepsilon/2}) \to 1,$$

as $N \to +\infty$. The last limit is deduced by (4.45) and by the fact that $T_{\varepsilon/2}$ has an explicit distribution - independent of N - which can be found through the reflection principle for the Brownian motion, in the same way we did in Lemma 4.10, for which we have

$$\mathbb{P}(T_{\varepsilon/2} \leq \delta) \to 0,$$

as $\delta \to 0$. □

Proof of Theorem 4.7. Apply Lemmas 4.10, 4.11 and 4.12 for fixed $\varepsilon, \delta > 0$. Observe that the density of $T_{\varepsilon/2}$ is smooth with respect to ε, and of course $T_{\varepsilon/2} \to 0$ for $\varepsilon \to 0$. Indeed, repeating analogous computations as in Lemma 4.10, we find, for $t \geq 0$,

$$\mathbb{P}(T_{\varepsilon/2} \leq t) = \frac{\varepsilon}{2\sqrt{2\pi\sigma^2}} \frac{1}{t^{3/2}} e^{-\frac{\varepsilon^2}{8\sigma^2 t}}.$$

The same is true for both $T^{\varepsilon}_{m_a}, T^{-\delta}_{m_a} \to T^0_{m_a}$, when $\varepsilon, \delta \to 0$. Sending first $N \to +\infty$ and then $\varepsilon, \delta \to 0$, we get the convergence in distribution for all the times $t \leq T_{m_b}$. Once we are in m_b, we can restart the dynamics by the strong Markov property and repeat the arguments above for the symmetric negative component of the invariant curve. Inductively, we can find a sequence of almost surely finite stopping times $(T_k)_{k\in\mathbb{N}}$ (the alternate arrival times in the two symmetric critical points), such that $[0, T] = \cup_k \{[T_k, T_{k+1}] \cap [0, T]\}$. This is enough to deduce the weak convergence of $(m^N(t))_{t\in[0,T]}$ to the process with instantaneous deterministic jumps described by SDE (4.26). □

In Figure 4.3 we show a comparison between two prelimit trajectories in the subcritical and supercritical case for the same initial conditions, where we used the coordinates (x, m) instead of (λ, m), which were instead employed in Figures 4.1 and 4.2. These plots will come useful for a qualitative comparison with the two-level hierarchical case.

4.3 The hierarchical model

In this section we study the two-level hierarchical version of the previous model. We consider N interacting populations, each of which consists of N mean field interacting particles. We denote with a subscript (i,j) the i-th individual in the j-th population, with $i,j = 1,\dots,N$. We thus have a collection of N^2 pairs of variables (x_{ij}, σ_{ij}) (equivalently $(\lambda_{ij}, \sigma_{ij})$), where the σ_{ij}'s are the spins, and the x_{ij}'s represent the aggregated remaining characteristics of the individual. As above, we define

$$m_j^N(t) := \frac{1}{N}\sum_{i=1}^N \sigma_{ij}(t),$$

the magnetization of the j-th population, and the analogous definition for $x_j^N(t)$ and $\lambda_j^N(t)$. Moreover, we define the two-level magnetization as

$$M^N(t) := \frac{1}{N^2}\sum_{i,j=1}^N \sigma_{ij}(t) = \frac{1}{N}\sum_{j=1}^N m_j^N(t),$$

and the analogous quantities $X^N(t) := \frac{1}{N^2}\sum_{ij} x_{ij}(t) = \frac{1}{N}\sum_{j=1}^N x_j^N(t)$ (resp. $\Lambda^N(t)$) for the x (resp. λ) variables. Ideally, we want to describe the dynamics at the different hierarchical levels as a projection of a diffusion process onto an invariant curve, as we did for the one population scenario.

With the choices specified in (4.3), the stochastic dynamics (4.1) becomes

$$\begin{cases} \sigma_{ij} \mapsto -\sigma_{ij} \quad \text{rate} \quad 1+\tanh\Big[-\beta_1\sigma_{ij}(t)(x_j^N(t)+m_j^N(t)) - \beta_2\sigma_{ij}(t)(X^N(t)+M^N(t))\Big], \\ dx_{ij}(t) = \sigma dW_{ij}(t) - \alpha_1\Big[x_{ij}(t) - x_j^N(t)\Big]dt - \frac{\alpha_2}{N}\Big[x_{ij}(t) - X^N(t)\Big]dt, \\ \sigma_{ij}(0) \sim \text{Ber}(p), \\ x_{ij}(0) \sim \mathcal{N}(0,1), \end{cases} \tag{4.46}$$

for $\beta_1, \beta_2, \sigma, \alpha_1, \alpha_2 > 0$, with the $W_{ij}(t)$'s being N^2 independent one-dimensional Brownian motions. In terms of the alternative variables $(\sigma_{ij}, \lambda_{ij})$ and their corresponding macroscopic quantities, the above can be rewritten as

$$\begin{cases} \sigma_{ij} \mapsto -\sigma_{ij} \quad \text{with rate} \quad 1+\tanh\Big[-\beta_1\sigma_{ij}(t)\lambda_j^N(t) - \beta_2\sigma_{ij}(t)\Lambda^N(t)\Big], \\ d\lambda_{ij}(t) = d\sigma_{ij}(t) + \sigma dW_{ij}(t) - \alpha_1\Big[(\lambda_{ij}(t)-\sigma_{ij}(t)) - \Big(\lambda_j^N(t) - m_j^N(t)\Big)\Big]dt \\ \qquad\qquad -\frac{\alpha_2}{N}\Big[(\lambda_{ij}(t)-\sigma_{ij}(t)) - \Big(\Lambda^N(t) - M^N(t)\Big)\Big]dt, \\ \sigma_{ij}(0) \sim \text{Ber}(p), \\ \lambda_{ij}(0) \sim \text{Ber}(p) * \mathcal{N}(0,1), \end{cases}$$

where the $*$ in the initial conditions for λ_{ij} denotes the convolution between the two distributions. Thanks to the linearity of the dynamics for the x_{ij}'s, it follows directly from (4.46) that

$$\begin{cases} dx_j^N(t) = -\frac{\alpha_2}{N}\Big[x_j^N(t) - X^N(t)\Big]dt + \frac{\sigma}{\sqrt{N}}dW_j^N(t), \\ x_j^N(0) \sim \mathcal{N}\left(0, \frac{1}{N}\right). \end{cases} \qquad \begin{cases} dX^N(t) = \frac{\sigma}{N}dW^N(t), \\ X^N(0) \sim \mathcal{N}\left(0, \frac{1}{N^2}\right), \end{cases} \tag{4.47}$$

where

$$W_j^N := \frac{1}{\sqrt{N}}\sum_{i=1}^N W_{ij}$$

are N independent Brownian motions, and

$$W^N := \frac{1}{\sqrt{N}} \sum_{j=1}^{N} W_j^N$$

is another Brownian motion. Note that the laws of $(W_j^N(t))_{t\geq 0}$ and $(W^N(t))_{t\geq 0}$ are independent of N, but we keep the dependency on N in the notation to refer to the specific Brownian motions. As we did for the mean field case, we describe each population through the order parameters $(m_j^N(t), x_j^N(t))_{t\geq 0}$. The collective behavior of the system can be studied in terms of the infinitesimal generator of the dynamics applied to a function $f = f((m_1, x_1), (m_2, x_2), \dots, (m_N, x_N)) =: f(\boldsymbol{m}, \boldsymbol{x})$, $f : [-1,1]^N \times \mathbb{R}^N \to \mathbb{R}$, which is given by

$$\begin{aligned}
\mathcal{L}^N f(\boldsymbol{m}, \boldsymbol{x}) = \sum_{j=1}^{N} \Bigg\{ & N \frac{1+m_j}{2} \left(1 - \tanh\left[\beta_1(x_j + m_j) + \beta_2(X^N + M^N)\right]\right) \times \\
& \times \left[f\left(x_j, m_j - \frac{2}{N}\right) - f(x_j, m_j) \right] \\
& + N \frac{1-m_j}{2} \left(1 + \tanh\left[\beta_1(x_j + m_j) + \beta_2(X^N + M^N)\right]\right) \times \\
& \times \left[f\left(x_j, m_j + \frac{2}{N}\right) - f(x_j, m_j) \right] \\
& + \frac{1}{2N} \sigma^2 \frac{\partial^2}{\partial x_j^2} f(x_j, m_j) - \frac{\alpha_2}{N} \left(x_j - X^N\right) \frac{\partial}{\partial x_j} f(x_j, m_j) \Bigg\}.
\end{aligned} \tag{4.48}$$

The rest of the chapter is organized as follows: in Section 4.3.1 we develop some heuristics to present the expected limit behaviors; in Section 4.3.2 we study the convergence at times of order 1; we then restrict to the subcritical regime for studying rigorously the convergence to the limit dynamics at times of order N and N^2 (respectively addressed in Sections 4.3.3 and 4.3.4); in Section 4.3.5 we generalize the results giving a conjecture on the k-level hierarchical case, for any k finite; finally, in Section 4.3.6 we study heuristically, with the help of numerics, the zero-temperature limit case $\beta_1 = \beta_2 = +\infty$, highlighting the presence of a phase transition tuned by the diffusion parameters.

4.3.1 Heuristics

At the first hierarchical level we are interested in describing the limit behavior of the order parameters of each population, i.e. the convergence of the sequences $(m_j^N(t), x_j^N(t))_{t\geq 0}$, both at a timescale of order 1 and N. At times of order 1, by (4.47) it follows that $dx_j^N(t) \to 0$ and thus $x_j^N(t) \to 0$ for $N \to +\infty$, that is the mean of the initial condition. The same holds for the sequence $X^N(t) \to 0$. Expanding the generator (4.48) at the first order in the variables m_j's, similarly to what we did for the one population case, we find that $m_j^N(0) \to m(0) = 2p-1$, $m_j^N(t) \to m(t)$, and $M^N(t) \to m(t)$ for $N \to +\infty$, where $(m(t), x(t))_{t\geq 0}$ solves the ODE

$$\begin{cases} \dot{m}(t) = 2\tanh((\beta_1 + \beta_2)m(t)) - 2m(t), \\ \dot{x}(t) = 0, \\ m(0) = 2p - 1, \\ x(0) = 0. \end{cases} \tag{4.49}$$

Equation (4.49) is (except for a missing multiplicative term in the vector field which does not modify the qualitative behavior of the dynamics) the mean field equation for the Curie–Weiss model (3.6) with inverse temperature parameter $\beta_1+\beta_2$. The equilibria of the above ODE are either just one ($m=0$), when $\beta_1+\beta_2 \leq 1$, or three when $\beta_1+\beta_2>1$: two stable (the polarized ones) and one unstable (the disordered one), where the asymptotic one is one of the two polarized states, determined by the sign of the initial magnetization.

At times of order N, the diffusions x_j^N's are now subject to non-trivial dynamics. Indeed, denoting again - with an abuse of notation - the sped up processes as $x_j^N(t) := x_j^N(Nt)$, $X^N(t) := X^N(Nt)$, equations (4.47) become

$$\begin{cases} dx_j^N(t) = -\alpha_2 \left[x_j^N(t) - X^N(t)\right] dt + \sigma dW_j^N(t), \\ x_j^N(0) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}\frac{1}{N}\right). \end{cases} \quad \begin{cases} dX^N(t) = \frac{\sigma}{\sqrt{N}} dW^N(t), \\ X^N(0) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}\frac{1}{N^2}\right), \end{cases} \tag{4.50}$$

where the initial data are given by the long-time limit of the diffusions at the timescale of order 1. In this timescale we thus find $x_j^N(t) \to x(t)$, $X^N(t) \to 0$, where $x(t)$ follows the Ornstein-Uhlenbeck dynamics

$$\begin{cases} dx(t) = -\alpha_2 x(t)dt + \sigma dW(t), \\ x(0) = 0, \end{cases}$$

with W a Brownian motion. As in the mean field case, the accelerated approximate diffusive generator can give us intuition on the limit dynamics for the magnetization processes at a timescale of order N. Indeed, expanding up to the second order the jump terms of the dynamics in m_j in (4.48), we get

$$\begin{aligned} N\mathcal{L}^N f(m_j, x_j) &\approx \\ &\approx N\left[2\tanh(\beta_1(x_j+m_j) + \beta_2(X^N+M^N)) - 2m_j\right] \frac{\partial}{\partial m_j} f(m_j, x_j) \\ &+ \left[2 - 2m_j \tanh(\beta_1(x_j+m_j) + \beta_2(X^N+M^N))\right] \frac{\partial^2}{\partial m_j^2} f(x_j, m_j) \\ &+ \frac{\sigma^2}{2}\frac{\partial^2}{\partial x_j^2} f(x_j, m_j) - \alpha_2\left(x_j - X^N\right)\frac{\partial}{\partial x_j} f(x_j, m_j). \end{aligned} \tag{4.51}$$

Assuming that a propagation of chaos property holds, the presence of the strong drift in the above generator should be such that the limit of the magnetizations processes $m_j^N(t)$'s is a (mean field) process laying on the curve $m = \tanh(\beta_1(x+m)) + \beta_2 M)$, where the dynamics is driven by the evolution of the Ornstein-Uhlenbeck limit process $x(t)$. Moreover, the limit mean field $M(t)$ should be proved to be the mean of $m(t)$ with respect to the distribution of $x(t)$. Specifically, denoting with $\mu_t(dx)$ the distribution of the O-U process at time t, we should find that each pair of accelerated processes $(x_j^N(t), m_j^N(t))_{t\geq 0}$, for $j=1,\dots,N$, at times of order N, converges to

$$\begin{cases} m(t)(x) = \tanh[\beta_1(x+m(t)(x)) + \beta_2 M(t)], \\ dx(t) = \sigma dW(t) - \alpha_2 x(t)dt, \\ m(0) = 2p-1, \\ x(0) = 0, \\ M(t) = \int_{\mathbb{R}} m(t)(x)\mu_t(dx). \end{cases} \tag{4.52}$$

The study of (4.52) is hard to perform for general choices of the parameters. Indeed, the behavior of the limiting dynamics can drastically change, depending on $\beta_1, \beta_2, \alpha_2, \sigma$ and the initial conditions. By analogy with the mean field case, one can expect to recognize a radical difference between the case where one has uniqueness of the equilibrium for the dynamics at order 1 (4.49), and the case where multiple equilibria appear.

At the second hierarchical level, we write the infinitesimal generator for a function $f(M, X)$ by averaging over the different populations,

$$\begin{aligned}
&\mathcal{L}^N f(M,X) = \\
&\quad N\sum_{j=1}^{N} \frac{1+m_j}{2}\left(1-\tanh\left[\beta_1(x_j+m_j)+\beta_2(X+M)\right]\right)\times \\
&\quad \times \left[f\left(M-\frac{2}{N^2},X\right)-f(M,X)\right] \\
&\quad + N\sum_{j=1}^{N} \frac{1-m_j}{2}\left(1+\tanh\left[\beta_1(x_j+m_j)+\beta_2(X+M)\right]\right)\times \\
&\quad \times \left[f\left(M+\frac{2}{N^2},X\right)-f(M,X)\right]+\frac{1}{2}\frac{\sigma^2}{N^2}\frac{\partial^2}{\partial X^2}f(M,X).
\end{aligned}$$

With analogous expansions as above for the jump components, we find

$$\begin{aligned}
&\mathcal{L}^N f(M,X) \approx \\
&\approx \frac{1}{N}\sum_{j=1}^{N}\Big[2\tanh\left[\beta_1(x_j+m_j)+\beta_2(X+M)\right]-2m_j\Big]\frac{\partial}{\partial M}f(M,X) \\
&+\frac{1}{N^3}\sum_{j=1}^{N}\Big[2-2m_j\tanh\left[\beta_1(x_j+m_j)+\beta_2(X+M)\right]\Big]\frac{\partial^2}{\partial M^2}f(M,X) \\
&+\frac{1}{2N^2}\sigma^2\frac{\partial^2}{\partial X^2}f(M,X).
\end{aligned}$$

In the drift component we can recognize the empirical average of the drifts of the single magnetizations. It is reasonable to ask for a description of the limit dynamics of $M^N(t)$ at any timescale. As we already motivated heuristically, at a timescale of order 1 the limit $M(t)$ of the macroscopic magnetization is the same as the magnetization of each population, which follows a Curie–Weiss ODE. For long times (but still of order 1), the value of $M^N(t)$ should converge to the stable equilibrium of the C–W ODE, which, depending on the value of $\beta_1 + \beta_2$ may be the disordered or a polarized state. Once we consider a scale of order N, we expect the single magnetizations to be close to their invariant curves. However, the evolution of $M^N(t)$ can change drastically depending on the interaction and diffusion parameters. We expect to find a regime of the parameters for which $M^N(t)$ does not move much from the equilibrium reached at times of order 1, eventually starting to move only at a scale of order N^2, when the macroscopic diffusion $X^N(t)$ starts to evolve non-trivially. At least in this regime, we expect the N^2 accelerated second-level process $M^N(t)$, conditionally on $X^N(t) \approx X$, to converge, for every fixed $t \geq 0$, to the deterministic value

$$\begin{cases} M(t) = \int_{\mathbb{R}} \tanh(\beta_1(x+m(t)(x))+\beta_2(X+M(t)))\mu_\infty(dx;X), \\ M(0) = 2p-1, \end{cases} \tag{4.53}$$

where $\mu_\infty(dx; X)$ is the stationary distribution of the process

$$dx(\xi) = -\alpha_2(x(\xi)-X)d\xi + \sigma dW(\xi),$$

where X enters as a parameter (it must be intended as the *current* fixed value of $X(t)$), and $m(t)(x)$ is the solution to

$$m(t)(x) = \tanh(\beta_1(x + m(t)(x)) + \beta_2(X + M(t))).$$

In turns, the limit process $X(t)$, $X^N(t) \to X(t)$, evolves as

$$\begin{cases} dX(t) = \sigma dB(t), \\ X(0) = 0, \end{cases} \tag{4.54}$$

where B is a Brownian motion. In order to obtain a full description of the law of the limit process $M(t)$, one then needs to consider a combination of the conditional dynamics (4.53) and (4.54), which takes into account the diffusive motion of $X(t)$ (see Section 4.3.4 for details).

For a rigorous treatment (Sections 4.3.2-4.3.4) we restrict to the subcritical case $\beta_1 + \beta_2 < 1$ (except for the order 1 timescale, analyzed in Section 4.3.2, where the argument works for any choice of the parameters), while we give solid heuristics and numerics for the supercritical zero-temperature limit regime $\beta_1 = \beta_2 \to +\infty$, analyzing the relevance of the diffusion parameters α_2 and σ for obtaining a phase transition already at a timescale of order N (see Section 4.3.6 below). Moreover, in Section 4.3.5 we conjecture a generalization of the results on the subcritical regime to the k-level hierarchical version of the model.

4.3.2 Propagation of chaos at times of order 1

In this section we prove the convergence of the empirical processes $(m_j^N(t), x_j^N(t))_{j=1,\dots,N}$ to the deterministic limit dynamics given by (4.49), for any choice of the parameters. Our proof works as well for random i.i.d. initial data $x_j^N(0) \sim \mu(dx)$, when $\mu(dx)$ is a normal distribution $\mathcal{N}(0, (\sigma^*)^2)$ (in our particular case we have $\sigma^* = \frac{1}{\sqrt{N}}$, so that randomness is deleted in the limit), with the resulting modification of the limit dynamics,

$$\begin{cases} \dot{m}(t)(x) = 2\tanh(\beta_1(x + m(t)(x)) + \beta_2 M(t)) - 2m(t)(x), \\ m(0)(x) \equiv 2p - 1, \\ M(t) = \int_{\mathbb{R}} m(t)(x)\mu(dx). \end{cases} \tag{4.55}$$

Considering random initial data also for the limit dynamics will be useful for the analyses of the longer timescales. For clarity we recall the dynamics of the empirical processes $(x_j^N(t), m_j^N(t))_{j=1,\dots,N}$,

$$\begin{cases} m_j^N \mapsto m_j^N \pm \frac{2}{N} \text{ rate } N\frac{1\mp m_j^N(t)}{2}\left(1 \pm \tanh\left[\beta_1(x_j^N(t) + m_j^N(t)) + \beta_2(X^N(t) + M^N(t))\right]\right), \\ m_j^N(0) = m_j \sim \frac{1}{N}\text{Bin}(Np), \\ dx_j^N(t) = -\frac{\alpha_2}{N}\left[x_j^N(t) - X^N(t)\right]dt + \frac{\sigma}{\sqrt{N}}dW_j^N(t), \\ x_j^N(0) = x_j \sim \mathcal{N}\left(0, (\sigma^*)^2\right). \end{cases} \tag{4.56}$$

Since the magnetizations are not appearing in the diffusion dynamics, the propagation of chaos property for the $x_j^N(t)$'s is trivially true for any finite time interval. Indeed, every diffusion is converging to its initial datum due to the decaying factors in front of the drift and diffusion coefficients. The i.i.d. processes $(\tilde{m}_j(t))_{j=1,\dots,N}$ to which the $m_j^N(t)$'s will be proved to converge are defined by

$$\tilde{m}_j(t) := m(t)(x_j),$$

where the x_j's coincide with the initial data for the diffusions, and $m(t)(x)$ is the solution to (4.55).

Theorem 4.13 (Propagation of chaos at order 1)**.** *Fix $T > 0$. For any $\beta_1, \beta_2, \alpha_1, \alpha_2, \sigma > 0$, and any $j = 1, \dots, N$, we have*

$$\lim_{N\to\infty} \mathbb{E}\left[\sup_{t\in[0,T]} \left| m_j^N(t) - \tilde{m}_j(t) \right| \right] = 0. \tag{4.57}$$

Before proving Theorem 4.13 we need to assess the well-posedness of Equation (4.55). We rewrite the dynamics with a generic initial datum

$$\begin{cases} \dot{m}(t)(x) = 2\tanh(\beta_1(x + m(t)(x)) + \beta_2 M(t)) - 2m(t)(x), \\ m(0)(x) = m_0(x), \\ M(t) = \int_{\mathbb{R}} m(t)(x)\mu(dx), \end{cases} \tag{4.58}$$

with $m_0 : \mathbb{R} \to [-1, 1]$, $m_0 \in C(\mathbb{R})$.

Proposition 4.14 (Well-posedness at order 1)**.** *For any $T > 0$, Equation* (4.58) *has a unique solution $m : [0, T] \times \mathbb{R} \to [-1, 1]$ such that $m(t)(\cdot) \in C(\mathbb{R})$ for any $t \in [0, T]$.*

Proof. The vector field $f : \mathbb{R} \times C(\mathbb{R}) \to C(\mathbb{R})$,

$$f(x, m) := 2\tanh(\beta_1(x + m) + \beta_2 M) - 2m \tag{4.59}$$

is globally Lipschitz continuous for any $\beta_1, \beta_2 > 0$, thus existence and uniqueness of a solution to (4.58), with $m(t)(\cdot) \in C(\mathbb{R})$ for any $t \in [0, T]$, is standard. Moreover, studying the sign of the vector field (4.59), we see that (4.58) defines a dynamics such that $m(t) : \mathbb{R} \to [-1, 1]$, provided the initial datum $m_0 : \mathbb{R} \to [-1, 1]$ has the same property. Indeed, at a point $\overline{x} \in \mathbb{R}$ for which $m(t)(\overline{x}) = 1$, we have that $\frac{d}{dt} m(t)(x)\Big|_{x=\overline{x}} \le 0$, and symmetrically if $m(t)(\overline{x}) = -1$ it holds $\frac{d}{dt} m(t)(x)\Big|_{x=\overline{x}} \ge 0$. □

For the proof of Theorem 4.13, we make use of a representation of the jump processes $m_j^N(t)$'s in terms of SDEs, by employing Poisson random measures (see [66]), as we did repeatedly in the previous chapters of this Dissertation. We fix a time horizon $T > 0$ independent of N and study the processes up to T. We then write the magnetization processes as

$$m_j^N(t) = m_j^N(0) + \int_0^t \int_\Xi f(m_j^N(s^-), \xi, M^N(s^-), x_j^N(s), X^N(s)) \mathcal{N}_j(ds, d\xi), \tag{4.60}$$

for $j = 1, \dots, N$, where each $m_j^N(t)$ takes values in $\Sigma = \left\{-1, -1 + \frac{2}{N}, \dots, 1 - \frac{2}{N}, 1\right\}$; the $\mathcal{N}_j$'s are N i.i.d. stationary Poisson random measures on $[0, T] \times \Xi$ with intensity measure ν on $\Xi := [0, \infty)^{|\Sigma|} \subset \mathbb{R}^{|\Sigma|}$ given by

$$\nu(E) := \sum_{i=1}^{|\Sigma|} \ell(E \cap \Xi_i), \tag{4.61}$$

for any E in the Borel σ-algebra $\mathcal{B}(\Xi)$ of Ξ, where $\Xi_j := \{u \in \Xi \ : \ u_i = 0 \ \ \forall\, i \ne j\}$ is viewed as a subset of $\mathbb{R}$, and ℓ is the Lebesgue measure on $\mathbb{R}$. We fix a probability space

$(\Omega, \mathcal{F}, \mathbb{P})$ and denote by $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ the filtration generated by the Poisson measures. The function f, modeling the possible jumps of the process, is given by

$$f(m, \xi, M, x, X) := \sum_{y\in\Sigma}(y-m)\mathbb{1}_{]0,\lambda_{my}[}(\xi_y),$$

where λ_{my} denotes the rate of jumping from state m to state y. Denoting by

$$\lambda_{\pm}(m, M, x, X) := N\frac{1\mp m}{2}\left(1 \pm \tanh\left[\beta_1(x+m) + \beta_2(X+M)\right]\right)$$

the rate of going from m to $m \pm \frac{2}{N}$, in our case the function f further simplifies to

$$f(m, \xi, M, x, X) = \frac{2}{N}\mathbb{1}_{]0,\lambda_+[}(\xi_{m+\frac{2}{N}}) - \frac{2}{N}\mathbb{1}_{]0,\lambda_-[}(\xi_{m-\frac{2}{N}}), \tag{4.62}$$

since the only possible jumps are the ones from m to $m \pm \frac{2}{N}$ with rates $\lambda_\pm$. The above definitions of f and ν ensure that $\lambda_\pm$ are exactly the transition rates of the continuous time Markov chains $m_j^N(t)$'s, and that $\pm\frac{2}{N}$ are the only possible jumps allowed at every time. Indeed, it is easy to prove that with our choices (4.60) is equivalent to

$$\begin{aligned}&\mathbb{P}\left[m_j^N(t+h) = m \pm \frac{2}{N}\middle| m_j^N(t) = m, M^N(t) = M, x_j^N(t) = x, X^N(t) = X\right]\\&= \lambda_\pm(m, M, x, X)h + o(h).\end{aligned} \tag{4.63}$$

By the smoothing formula of Poisson calculus (see [12, Ch. 9]), we have

$$\begin{aligned}\mathbb{E}\left[m_j^N(t)\right] &= \mathbb{E}\left[m_j^N(0)\right] + \mathbb{E}\left[\int_0^t\int_\Xi f(m_j^N(s^-), \xi, M^N(s^-), x_j^N(s), X^N(s))ds\nu(d\xi)\right]\\&= \mathbb{E}\left[m_j^N(0)\right] + \mathbb{E}\left[\int_0^t\int_\Xi\left[\frac{2}{N}\mathbb{1}_{]0,\lambda_+[}(\xi_{m+\frac{2}{N}}) - \frac{2}{N}\mathbb{1}_{]0,\lambda_-[}(\xi_{m-\frac{2}{N}})\right]ds\nu(d\xi)\right]\\&= \mathbb{E}\left[m_j^N(0)\right] + \mathbb{E}\Bigg[\int_0^t\Big[2\tanh\Big(\beta_1(x_j^N(s)+m_j^N(s)) + \beta_2(X^N(s)+M^N(s))\Big)\\&\quad - 2m_j^N(s)\Big]ds\Bigg].\end{aligned} \tag{4.64}$$

Proof of Theorem 4.13. First, we observe that, by the dynamics (4.60) with the choice (4.62) for f, we can write

$$\begin{aligned}&\sup_{s\in[0,t]}|m_j^N(s) - \tilde{m}_j(s)|\\&\quad= \sup_{s\in[0,t]}\left|\int_0^s\int_\Xi f(m_j^N(r^-), \xi, M^N(r^-), x_j^N(r), X^N(r))\mathcal{N}_j(dr, d\xi) - \tilde{m}_j(s)\right|.\end{aligned}$$

Taking the expectation and using formula (4.64) and the limit dynamics (4.55), we can estimate

$$\begin{aligned}&\mathbb{E}\left[\sup_{s\in[0,t]}\left|m_j^N(s) - \tilde{m}_j(s)\right|\right] \le\\&\quad\le \mathbb{E}\Big[\left|m_j^N(0) - (2p-1)\right|\Big] + \mathbb{E}\Bigg[\int_0^t\Big|2\tanh\Big(\beta_1(x_j^N(s)+m_j^N(s)) + \beta_2(X^N(s)+M^N(s))\Big)\end{aligned}$$

$$- 2\tanh\left(\beta_1(x_j + \tilde{m}_j(s)) + \beta_2 M(s)\right)\Big|ds\Big] + \mathbb{E}\Big[\int_0^t \Big|2m_j^N(s) - 2\tilde{m}_j(s)\Big|ds\Big]$$

$$\leq \mathbb{E}\Big[\big|m_j^N(0) - (2p-1)\big|\Big] + \mathbb{E}\Big[\int_0^t \Big|2\tanh\Big(\beta_1(x_j^N(s) + m_j^N(s)) + \beta_2(X^N(s) + M^N(s))\Big)$$

$$- 2\tanh\left(\beta_1(x_j + \tilde{m}_j(s)) + \beta_2 M(s)\right)\Big|ds\Big] + C\mathbb{E}\Big[\int_0^t \sup_{r\in[0,s]} \Big|m_j^N(r) - \tilde{m}_j(r)\Big|ds\Big].$$

By LLN on the initial data we have

$$\mathbb{E}\Big[\big|m_j^N(0) - (2p-1)\big|\Big] \leq C(N),$$

with $C(N) \to 0$ for $N \to +\infty$. We now focus on estimating the first of the two integrals. Using the globally Lipschitz continuity of $\tanh(\cdot)$, we have

$$\mathbb{E}\Big[\int_0^t \Big|2\tanh\Big(\beta_1(x_j^N(s) + m_j^N(s)) + \beta_2(X^N(s) + M^N(s))\Big)$$

$$- 2\tanh\left(\beta_1(x_j + \tilde{m}_j(s)) + \beta_2 M(s)\right)\Big|ds\Big]$$

$$\leq C\mathbb{E}\Big[\int_0^t \Big|x_j^N(s) - x_j\Big|ds\Big] + C\mathbb{E}\Big[\int_0^t \Big|X^N(s)\Big|ds\Big]$$

$$+ C\mathbb{E}\Big[\int_0^t \Big|m_j^N(s) - \tilde{m}_j(s)\Big|ds\Big] + C\mathbb{E}\Big[\int_0^t \Big|M^N(s) - M(s)\Big|ds\Big]$$

$$\leq C\mathbb{E}\Big[\int_0^T \Big|x_j^N(s) - x_j\Big|ds\Big] + C\mathbb{E}\Big[\int_0^T \Big|X^N(s)\Big|ds\Big]$$

$$+ C\mathbb{E}\Big[\int_0^t \sup_{r\in[0,s]} \Big|m_j^N(r) - \tilde{m}_j(r)\Big|ds\Big] + C\mathbb{E}\Big[\int_0^t \sup_{r\in[0,s]} \Big|M^N(r) - M(r)\Big|ds\Big],$$

where the constants are allowed to change from line to line. By the propagation of chaos for the diffusions, we have

$$\mathbb{E}\Big[\int_0^T \Big|x_j^N(s) - x_j\Big|ds\Big] + \mathbb{E}\Big[\int_0^T \Big|X^N(s)\Big|ds\Big] \leq C(N),$$

for some $C(N) \to 0$ when $N \to +\infty$. For the last integral, denoting $\tilde{M}^N(t) := \frac{1}{N}\sum_{i=1}^N \tilde{m}_i(t)$, we estimate

$$\mathbb{E}\Big[\int_0^t \sup_{r\in[0,s]} \Big|M^N(r) - M(r)\Big|ds\Big]$$

$$\leq \mathbb{E}\Big[\int_0^t \sup_{r\in[0,s]} \Big|M^N(r) - \tilde{M}^N(r)\Big|ds\Big] + \mathbb{E}\Big[\int_0^t \sup_{r\in[0,s]} \Big|\tilde{M}^N(r) - M(r)\Big|ds\Big]$$

$$\leq C\frac{1}{N}\sum_{i=1}^N \mathbb{E}\Big[\int_0^t \sup_{r\in[0,s]} \Big|m_i^N(r) - \tilde{m}_i(r)\Big|ds\Big] + C(N)$$

$$= C\mathbb{E}\left[\int_0^t \sup_{r\in[0,s]} \left|m_j^N(r) - \tilde{m}_j(r)\right| ds\right] + C(N),$$

where the $C(N) \to 0$ when $N \to +\infty$ by LLN, and the last equality is a consequence of the exchangeability of the processes $(m_i^N(t), \tilde{m}_i(t))_{i=1,\dots,N}$. Recollecting all the above observations and estimates, we have found

$$\mathbb{E}\left[\sup_{s\in[0,t]} |m_j^N(s) - \tilde{m}_j(s)|\right] \leq C(N) + C\mathbb{E}\left[\int_0^t \sup_{r\in[0,s]} \left|m_j^N(r) - \tilde{m}_j(r)\right| ds\right],$$

with $C(N)$ going to 0 for $N \to +\infty$. Denoting $\varphi(t) := \mathbb{E}\left[\sup_{s\in[0,t]} |m_j^N(s) - \tilde{m}_j(s)|\right]$, the last estimate implies

$$\varphi(t) \leq C(N) + \int_0^t \varphi(s) ds.$$

Thus, the propagation of chaos follows by the Gronwall's lemma. □

Remark 4.15. *Note that the strong convergence* (4.57) *implies the convergence (in e.g.* 1-*Wasserstein distance) of the associated empirical measures* $\mu^N(t) := \frac{1}{N}\sum_{j=1}^N \delta_{m_j^N(t)}$ *and* $\tilde{\mu}^N(t) := \frac{1}{N}\sum_{j=1}^N \delta_{\tilde{m}_j(t)}$ *to the deterministic measure* $\mu(t)$*, the distribution of the i.i.d. processes* $\tilde{m}_j(t)$*. Indeed, by inequality* (19) *and the convergence* (4.57)*,* $||\mu^N - \tilde{\mu}^N||_{\boldsymbol{d_1}} \to 0$ *as* $N \to +\infty$*, while* $||\tilde{\mu}^N - \mu||_{\boldsymbol{d_1}} \to 0$ *as* $N \to +\infty$ *is standard (by LLN). This in turns implies the propagation of chaos in the classic sense.*

The following proposition assesses the long-time behavior of the deterministic limit dynamics. Specifically, we show the convergence to a unique symmetric stationary profile $\overline{m}(x)$, regardless of the initial datum $m_0(x)$.

Proposition 4.16 (Long-time subcritical limit behavior)**.** *For* $\beta_1 + \beta_2 < 1$*, the solution* $m(t)(\cdot)$ *to* (4.58) *is such that*

$$\mathbb{E}\left[|m(t)(\xi) - \overline{m}(\xi)|^2\right] \to 0, \tag{4.65}$$

for $t \to \infty$*, with* $\xi \sim \mathcal{N}(0, \sigma^*)$ *and* $\overline{m}(\cdot)$ *is the unique solution to*

$$\overline{m}(x) = \tanh(\beta_1(x + \overline{m}(x))). \tag{4.66}$$

Proof. The uniqueness of solution to Equation (4.66) follows by considering any two solutions $m(x), n(x)$ and observing that

$$|m(x) - n(x)| \leq \beta_1 |m(x) - n(x)| \leq \cdots \leq \beta_1^k |m(x) - n(x)|,$$

for any $x \in \mathbb{R}$, so that we can conclude by a contraction argument. For the proof of (4.65), consider any two solutions $m(t)$ and $n(t)$ with different initial data. It holds

$$\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}} (m(t)(x) - n(t)(x))^2 \mu(dx) \leq -2(1 - (\beta_1 + \beta_2)) \int_{\mathbb{R}} (m(t)(x) - n(t)(x))^2 \mu(dx), \tag{4.67}$$

which is negative for $\beta_1 + \beta_2 < 1$, thus implying (4.65) because of the well-posedness of (4.58). Indeed $\overline{m}(x)$, the unique solution to Equation (4.66), is always a solution to (4.58) with initial datum $m_0(x) = -m_0(-x)$ and $M(t) = 0$ for every t.

In order to verify (4.67), we use Equation (4.58) to compute

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\int_{\mathbb{R}}(m(t)(x)-n(t)(x))^2\mu(dx) &= \int_{\mathbb{R}}(\dot m(t)(x)-\dot n(t)(x))(m(t)(x)-n(t)(x))\mu(dx)\\
&= -2\int_{\mathbb{R}}(m(t)(x)-n(t)(x))^2\mu(dx)\\
&+2\int_{\mathbb{R}}\Big[\tanh(\beta_1(m(t)(x)+x)+\beta_2 M(t)) - \tanh(\beta_1(n(t)(x)+x)+\beta_2 N(t))\Big]\times\\
&\times (m(t)(x)-n(t)(x))\mu(dx)\\
&\le -2\int_{\mathbb{R}}(m(t)(x)-n(t)(x))^2\mu(dx) + 2(\beta_1+\beta_2)\int_{\mathbb{R}}(m(t)(x)-n(t)(x))^2\mu(dx),
\end{aligned}$$

where in the last step we have used the Lipschitz properties of $\tanh(\cdot)$ and the definitions of $M(t)$ and $N(t)$. □

Remark 4.17. *Theorem 4.13 and Propositions 4.14, 4.16 can be generalized to the case of Gaussian initial data not centered around zero. The limit equation becomes*

$$\begin{cases}\dot m(t)(x) = 2\tanh(\beta_1(x+m(t)(x))+\beta_2(\overline{X}+M(t))) - 2m(t)(x),\\ m(0)(x) = m_0(x),\\ M(t) = \int_{\mathbb{R}} m(t)(x)\mu(dx;\overline{X}),\end{cases} \tag{4.68}$$

with $\mu(dx;\overline{X}) = \mathcal{N}\left(\overline{X},\rho^2\right)$*. The equilibrium solution to* (4.68) *is given by*

$$\begin{cases}\overline{m}_{\overline{X}}(x) = \tanh\left(\beta_1(x+\overline{m}_{\overline{X}}(x))+\beta_2(\overline{X}+\overline{M})\right),\\ \overline{M} = \int_{\mathbb{R}} \overline{m}_{\overline{X}}(x)\mu(dx;\overline{X}),\end{cases} \tag{4.69}$$

whose well-posedness can be obtained by a contraction argument as in Proposition 4.16.

We conclude the section noting that the processes x_j^N's and m_j^N's are close to their i.i.d. limits for any fixed time ranging in an interval which is allowed to grow with N with a certain speed.

Theorem 4.18 (Long-time subcritical particles behavior)**.** *For any* $T>0$, $\beta_1+\beta_2<1$, $\varepsilon>0$ *and* $j=1,\ldots,N$, *we have*

(i) For any $A\in\mathcal{B}(\mathbb{R})$,

$$\sup_{t\in[0,TN^{2-\varepsilon}]}\left|\mathbb{P}\Big(x_j^N(t)\in A\Big) - \mathbb{P}\Big(x_j(t)\in A\Big)\right| \to 0$$

for $N\to+\infty$.

(ii) For any $A\in\mathcal{B}([-1,1])$,

$$\sup_{t\in[0,TN^{2/3-\varepsilon}]}\left|\mathbb{P}\Big(m_j^N(t)\in A\Big) - \mathbb{P}\Big(\tilde m_j(t)\in A\Big)\right| \to 0$$

for $N\to+\infty$,

where $\tilde{m}_j(t) := m(t)(x_j(t))$, *with*

$$\begin{cases} dx_j^N(t) = -\frac{\alpha_2}{N}(x_j^N(t) - X^N(t))dt + \frac{\sigma}{\sqrt{N}}dW_j^N(t), \\ x_j^N(0) = x_j \sim \mathcal{N}\left(0, \frac{1}{N}\right), \end{cases} \tag{4.70}$$

and

$$\begin{cases} dx_j(t) = -\frac{\alpha_2}{N}(x_j(t) - \mathbb{E}[x_j(t)])dt + \frac{\sigma}{\sqrt{N}}dW_j(t), \\ x_j(0) = x_j \sim \mathcal{N}\left(0, \frac{1}{N}\right), \end{cases} \tag{4.71}$$

with $X^N(t) := \frac{1}{N}\sum_{k=1}^N x_k^N(t)$ *and* $W_j(t)$ *is a Brownian motion.*

Proof. We realize the process $x_j^N(t)$ by plugging in (4.70) the *same* Brownian motion $W_j(t)$ of the definition of $x_j(t)$ in (4.71). Then, for the resulting processes we prove

$$\sup_{0\le t\le TN^{2-\varepsilon}} \mathbb{E}\left[\left(x_j^N(t) - x_j(t)\right)^2\right] \to 0, \tag{4.72}$$

$$\sup_{0\le t\le TN^{2/3-\varepsilon}} \mathbb{E}\left[\left|m_j^N(t) - \tilde{m}_j(t)\right|\right] \to 0, \tag{4.73}$$

for $N \to +\infty$, which imply the limits in distribution (i) and (ii). First of all we observe that, for any $t \ge 0$, we have

$$\mathbb{E}[x_j(t)] = 0,$$
$$X^N(t) = \frac{\sigma}{N}W(t),$$

with $W(t) := \frac{1}{\sqrt{N}}\sum_{k=1}^N W_k(t)$. For (4.72), by Itô's formula, we compute

$$\begin{aligned}
\mathbb{E}\Big[(x_j^N(t) - x_j(t))^2\Big] &= \mathbb{E}\Big[(x_j^N(0) - x_j(0))^2\Big] - \frac{2\alpha_2}{N}\int_0^t \mathbb{E}\Big[(x_j^N(s) - x_j(s))^2\Big]ds \\
&- \frac{2\alpha_2}{N}\int_0^t \mathbb{E}\Big[(x_j^N(s) - x_j(s))X^N(s)\Big]ds \\
&\le \mathbb{E}\Big[(x_j^N(0) - x_j(0))^2\Big] - \frac{2\alpha_2}{N}\int_0^t \mathbb{E}\Big[(x_j^N(s) - x_j(s))^2\Big]ds \\
&+ \frac{2\alpha_2}{N}\int_0^t \mathbb{E}\Big[|x_j^N(s) - x_j(s)||X^N(s)|\Big]ds \\
&\le \mathbb{E}\Big[(x_j^N(0) - x_j(0))^2\Big] - \frac{2\alpha_2}{N}\int_0^t \mathbb{E}\Big[(x_j^N(s) - x_j(s))^2\Big]ds \\
&+ \frac{\alpha_2}{N}\int_0^t \mathbb{E}\Big[(x_j^N(s) - x_j(s))^2\Big]ds + \frac{\alpha_2}{N}\int_0^t \mathbb{E}\Big[(X^N(s))^2\Big]ds,
\end{aligned}$$

where in the last estimate we have used $ab \le \frac{a^2}{2} + \frac{b^2}{2}$. By definition, we have

$$\int_0^t \mathbb{E}[(X^N(s))^2]ds = \frac{\sigma^2}{N^2}\int_0^t \mathbb{E}[(W(s))^2]ds = \frac{1}{N^2}\sigma^2\frac{t^2}{2}.$$

Recollecting the above expressions, we have found

$$\mathbb{E}[(x_j^N(t) - x_j(t))^2] \le \mathbb{E}[(x_j^N(0) - x_j(0))^2] - \frac{\alpha_2}{N}\int_0^t \mathbb{E}[(x_j^N(s) - x_j(s))^2]ds + \frac{\alpha_2}{N^3}\sigma^2\frac{t^2}{2},$$

which, denoting with $c(t) := \mathbb{E}[(x_j^N(t) - x_j(t))^2]$, in differential form reads

$$\dot{c}(t) \leq -\frac{\alpha_2}{N} c(t) + \frac{\alpha_2}{N^3}\sigma^2 t.$$

By solving the differential equation on the right hand side of the inequality, we deduce

$$c(t) \leq e^{-\frac{\alpha_2}{N}t} c(0) + \frac{\sigma^2}{N\alpha_2}(e^{-\frac{\alpha_2}{N}t} - 1) + \frac{\sigma^2}{N^2} t. \tag{4.74}$$

Note that $c(0) = 0$ because of our choices of initial data. When we take the supremum over t in the above expression the dominant term is $\frac{\sigma^2}{N^2}t$, which still tends to 0 with N going to infinity, if the supremum is taken over $0 \leq t \leq TN^{2-\varepsilon}$, so that (4.72) is proved.

Moreover, we have

$$\int_0^t \mathbb{E}\left[|x_j^N(s) - x_j(s)|\right] ds \leq \frac{C}{N} t^{\frac{3}{2}}. \tag{4.75}$$

Indeed, by Jensen's and Hölder's inequalities and by (4.74), we estimate

$$\begin{aligned}
&\left(\int_0^t \mathbb{E}\left[\left|x_j^N(s) - x_j(s)\right|\right] ds\right)^2 = \left(\frac{t}{t}\int_0^t \mathbb{E}\left[\left|x_j^N(s) - x_j(s)\right|\right] ds\right)^2 \\
&\leq t\int_0^t \mathbb{E}\left[\left|x_j^N(s) - x_j(s)\right|^2\right] ds \leq t^2 \sup_{s\in[0,t]} \left[\frac{\sigma^2}{N\alpha_2}(e^{-\frac{\alpha_2}{N}s} - 1) + \frac{\sigma^2}{N^2} s\right] \\
&\leq \frac{\sigma^2}{N^2} t^3
\end{aligned}$$

so that (4.75) follows by taking the square root. Note also that

$$\int_0^t \mathbb{E}\left[|X^N(s)|\right] ds = \frac{\sigma}{N}\int_0^t \mathbb{E}[|W(s)|] ds \leq \frac{C}{N} t^{\frac{3}{2}},$$

since $|X^N(s)| = \frac{1}{N}|W(s)|$, and $\mathbb{E}[|W(s)|] \leq C\sqrt{s}$.

Finally, for proving (4.73) we compute (using $\text{sign}(x) \cdot x = |x|$),

$$\begin{aligned}
\mathbb{E}\left[\left|m_j^N(t) - \tilde{m}_j(t)\right|\right] &= \mathbb{E}\left[\left|m_j^N(0) - \tilde{m}_j(0)\right|\right] - 2\int_0^t \mathbb{E}\left[\left|m_j^N(s) - \tilde{m}_j(s)\right|\right] ds \\
&+ 2\int_0^t \mathbb{E}\Big[\text{sign}(m_j^N(s) - \tilde{m}_j(s))\Big(\tanh(\beta_1(x_j^N(s) + m_j^N(s)) + \beta_2(M^N(s) + X^N(s)) - \\
&- \tanh(\beta_1(x_j(s) + \tilde{m}_j(s)) + \beta_2 M(s))\Big)\Big] ds.
\end{aligned}$$

Using the Lipschitz properties of $\tanh(\cdot)$ and the boundedness of the magnetizations processes we can estimate

$$\begin{aligned}
&\mathbb{E}\left[\left|m_j^N(t) - \tilde{m}_j(t)\right|\right] \\
&\quad\leq \mathbb{E}\left[\left|m_j^N(0) - \tilde{m}_j(0)\right|\right] - 2(1-\beta_1)\int_0^t \mathbb{E}\left[\left|m_j^N(s) - \tilde{m}_j(s)\right|\right] ds
\end{aligned}$$

$$+ 2\beta_1 \int_0^t \mathbb{E}\left[\left|x_j^N(s) - x_j(s)\right|\right] ds + 2\beta_2 \int_0^t \mathbb{E}\left[\left|M^N(s) - M(s)\right|\right] ds$$
$$+ 2\beta_2 \int_0^t \mathbb{E}\left[\left|X^N(s)\right|\right] ds.$$

Denoting $\tilde{M}^N(t) := \frac{1}{N}\sum_{j=1}^N \tilde{m}_j(t)$, and $\mu^N(x_1, \dots, x_N) := \frac{1}{N}\sum_{j=1}^N \delta_{x_j}$, we have

$$\mathbb{E}\left[\left|\tilde{M}^N(t) - M(t)\right|\right] = \mathbb{E}\left[\left|\int_{\mathbb{R}} m(t)(x)(\mu^N - \mu)(dx)\right|\right]$$
$$\leq ||\mu^N - \mu||_{\boldsymbol{d_1}} \leq \frac{C}{\sqrt{N}},$$

where $\boldsymbol{d_1}$ is the 1-Wasserstein metric, and the estimate follows by LLN. Furthermore, we have

$$\mathbb{E}\left[\left|M^N(s) - M(s)\right|\right] \leq \mathbb{E}\left[\left|M^N(s) - \tilde{M}^N(s)\right|\right] + \mathbb{E}\left[\left|\tilde{M}^N(s) - M(s)\right|\right]$$
$$\leq \mathbb{E}\left[\left|m_j^N(s) - \tilde{m}_j(s)\right|\right] + \mathbb{E}\left[\left|\tilde{M}^N(s) - M(s)\right|\right],$$

where in the last estimate we have used the exchangeability of the magnetizations processes. Finally, we can collect all the previous estimates to get

$$\mathbb{E}\left[\left|m_j^N(t) - \tilde{m}_j(t)\right|\right]$$
$$\leq \mathbb{E}\left[\left|m_j^N(0) - \tilde{m}_j(0)\right|\right] - 2(1 - \beta_1 - \beta_2) \int_0^t \mathbb{E}\left[\left|m_j^N(s) - \tilde{m}_j(s)\right|\right] ds$$
$$+ \frac{C_1}{N} t^{3/2} + C_2 \frac{t}{\sqrt{N}}.$$

In differential form, with $c(t) := \mathbb{E}\left[\left|m_j^N(t) - \tilde{m}_j(t)\right|\right]$, $k := 2(1 - \beta_1 - \beta_2) > 0$, the previous estimate reads

$$\dot{c}(t) \leq -kc(t) + \frac{C_1}{N} t^{1/2} + \frac{C_2}{\sqrt{N}},$$

implying

$$c(t) \leq e^{-kt} c(0) + \frac{C}{N} t^{3/2} + \frac{C}{\sqrt{N}}.$$

Recalling that $c(0) \to 0$ for $N \to +\infty$ by a LLN, we obtain claim (4.73) when we take the supremum for $0 \leq t \leq TN^{2/3-\varepsilon}$. □

Remark 4.19. *We observe that the results proved in this section are slightly more general than what was needed. Indeed, the initial data for the limit diffusion processes x_j's should have been set to $x_j(0) = 0$ for any $j = 1, \dots, N$. Clearly, every result obtained above holds true under this framework as well. As far as the other timescales are considered, at a timescale of order N we still consider trivial initial data for the limit diffusions (thanks to the previous theorem), while at a scale of order N^2 the initial data for the diffusions are provided by the long-time limit of the diffusion processes at a timescale of order N, i.e. they are i.i.d. normally distributed random variables, where the parameters of the distribution are given by the ergodic limit of the Ornstein-Uhlenbeck process.*

4.3.3 Propagation of chaos at times of order N: the subcritical case

In this section we adapt the proof of the propagation of chaos to times of order N for the case $\beta_1+\beta_2<1$. Thanks to Theorem 4.18, in this scale we can assume that the initial data for the processes are given by the long-time limit at the previous timescale of order 1. For the diffusions it holds $x_j^N(0)=x_j\sim\mathcal{N}\left(0,\frac{1}{N}\frac{\sigma^2}{2\alpha_2}\right)$ for any $j=1,\dots,N$, while the magnetizations are starting the dynamics in the long-time limit symmetric equilibrium $\overline{m}(x)$. For ease of notation we still denote the sped up processes by

$$x_j^N(t):=x_j^N(Nt), \qquad m_j^N(t):=m_j^N(Nt).$$

They evolve according to:

$$\begin{cases} m_j^N\mapsto m_j^N\pm\frac{2}{N} \text{ rate } N^2\frac{1\mp m_j^N(t)}{2}\left(1\pm\tanh\left[\beta_1(x_j^N(t)+m_j^N(t))+\beta_2(X^N(t)+M^N(t))\right]\right),\\ m_j^N(0)=\overline{m}(x_j),\\ dx_j^N(t)=-\alpha_2\left[x_j^N(t)-X^N(t)\right]dt+\sigma dW_j^N(t),\\ x_j^N(0)=x_j\sim\mathcal{N}\left(0,\frac{1}{N}\frac{\sigma^2}{2\alpha_2}\right). \end{cases} \tag{4.76}$$

The limit i.i.d. processes to which the sped up processes at order N will be proved to converge are denoted as

$$(\tilde{x}_j(t),\tilde{m}_j(t))_{j=1,\dots,N},$$

where $\tilde{m}_j(t):=m(t)(\tilde{x}_j(t))$, with

$$\begin{cases} d\tilde{x}_j(t)=-\alpha_2\tilde{x}_j(t)dt+\sigma dW_j(t),\\ \tilde{x}_j(0)=0, \end{cases} \tag{4.77}$$

with W_j's N independent Brownian motions, and $m(t)(x)$ solves

$$\begin{cases} m(t)(x)=\tanh\left(\beta_1(x+m(t)(x))+\beta_2 M(t)\right),\\ m(0)(x)\equiv\overline{m}(x),\\ M(t)=\int_{\mathbb{R}} m(t)(x)\mu_t(dx), \end{cases} \tag{4.78}$$

where $\mu_t(dx)$ is the distribution at time t of the Ornstein-Uhlenbeck i.i.d. processes $\tilde{x}_j(t)$'s, and $\overline{m}(x)$ is the solution to Equation (4.66). Once again, the propagation of chaos for the diffusion processes is standard at this scale (for any fixed interval of time). What we need to prove is the same property for the magnetizations processes,

Theorem 4.20 (Propagation of chaos at order N)**.** *Fix $T>0$. For any $\beta_1+\beta_2<1$, $\alpha_1,\alpha_2,\sigma>0$, and any $j=1,\dots,N$, $\left(m_j^N(t)\right)_{t\in[0,T]}$ converges weakly in the sense of stochastic processes, for $N\to+\infty$, to $\left(\tilde{m}_j(t)\right)_{t\in[0,T]}$.*

Before addressing the proof, we must check that Equation (4.78) is well-posed. In fact, the limit dynamics (4.78) is trivial at this scale.

Proposition 4.21 (Well-posedness at order N)**.** *For any $\beta_1+\beta_2<1$, Equation* (4.78) *has a unique classical solution $m:[0,T]\times\mathbb{R}\to[-1,1]$ such that $m(t)(\cdot)\in C(\mathbb{R})$ for any $t\in[0,T]$. Moreover, we have $m(t)(x)=\overline{m}(x)$ and $M(t)=0$ for any $t\in[0,T]$.*

Proof. The non-explosiveness of Equation (4.78) is obvious by construction. Indeed, $m(t)(x) \in [-1, 1]$ for any $t \in [0, T]$, $x \in \mathbb{R}$. For the uniqueness, define

$$F(m)(t)(x) := \tanh\left(\beta_1(x + m(t)(x)) + \beta_2 M(t)\right),$$

and consider two solutions $m(t)(\cdot), m'(t)(\cdot) \in C(\mathbb{R})$. Then, we have

$$\begin{aligned}|F(m) - F(m')|(t)(x) &\leq \max_{\xi \in \mathbb{R}} |1 - \tanh^2(\xi)| \Big[\beta_1 |m(t)(x) - m'(t)(x)| + \beta_2 |M(t) - M'(t)|\Big] \\ &\leq \beta_1 |m(t)(x) - m'(t)(x)| + \beta_2 |M(t) - M'(t)|.\end{aligned}$$

By taking the sup over $x \in \mathbb{R}$,

$$||F(m)(t) - F(m')(t)||_\infty \leq (\beta_1 + \beta_2)||m(t) - m'(t)||_\infty,$$

since $|M(t) - M'(t)| \leq \int_{\mathbb{R}} |m(t)(x) - m'(t)(x)| \mu_t(dx) \leq ||m(t) - m'(t)||_\infty$. Thus, we can conclude the uniqueness of solution by a contraction argument when $\beta_1 + \beta_2 < 1$. Moreover, the triviality of the dynamics is due to the symmetry around zero of the distribution $\mu_t(dx) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}(1 - e^{-2\alpha_2 t})\right)$, for which we have that $M(t) \equiv 0$ for any t, and thus that $m(t)(x) \equiv \overline{m}(x)$ is the unique solution to the dynamics in this regime. □

Even though the limit deterministic dynamics is trivial, we still have to prove the convergence of the sped up dynamics to the limit chaotic processes. Moreover, the techniques employed in the proof will come useful for the further timescales analyses. While the requirement $\beta_1 + \beta_2 < 1$ ensures the uniqueness of solution to the limit dynamics of order N, the crucial observation - working for $\beta_1 < 1$ independently of β_2 - which allows to adapt the previous proof is the following

Proposition 4.22 (Contraction estimates). *Let $(x_j^N(t), m_j^N(t))_{j=1,\dots,N}$ the empirical sped up processes at a timescale of order N. Let*

$$y_j(t) := m_j^N(t) - \tanh\left(\beta_1(x_j^N(t) + m_j^N(t)) + \beta_2(X^N(t) + M^N(t))\right).$$

Then, for any $\beta_1 < 1$, $k > 0$, $j = 1, \dots, N$,

$$N\mathcal{L}^N |y_j(t)|^k \leq -CN|y_j(t)|^k + O(1), \tag{4.79}$$

for some $C := C(\beta_1, k) > 0$, where $O(1)$ is uniform in time and space and $\mathcal{L}^N$ is given by (4.48).

Proof. The proof uses analogous arguments to the ones used in the mean field case for obtaining (4.17). For simplicity, we use the coordinates (λ_j, m_j) instead of (x_j, m_j). Applying the accelerated generator in the other coordinates to the function $y_j^k(t)$, and expanding to the second order in (m_j, λ_j), we get

$$\begin{aligned}N\mathcal{L}^N |y_j(t)|^k &\leq -2N\left[m_j^N(t) - \tanh\left(\beta_1 \lambda_j^N(t) + \beta_2 \Lambda^N(t)\right)\right]\left[\frac{\partial}{\partial m_j}|y_j(t)|^k + \frac{\partial}{\partial \lambda_j}|y_j(t)|^k\right] \\ &\quad + O(1) \\ &= -2Ny_j(t)\left[\frac{\partial}{\partial m_j}|y_j(t)|^k + \frac{\partial}{\partial \lambda_j}|y_j(t)|^k\right] + O(1).\end{aligned}$$

The $O(1)$ follows from the fact that both y_j and the coefficients appearing in the higher order terms of the generator are uniformly bounded by some constant C not depending on time nor space. Indeed, the dominating remainder terms of the development are the second order terms, which in the accelerated timescale of order N are of order 1. Computing

$$\frac{\partial}{\partial m_j}|y_j(t)|^k + \frac{\partial}{\partial \lambda_j}|y_j(t)|^k$$
$$= k\,|y_j(t)|^{k-1}\,\mathrm{sign}(y_j(t))\left[1 - \left(\beta_1 + \frac{\beta_2}{N}\right)\left(1 - \tanh^2\left(\beta_1\lambda_j^N(t) + \beta_2\Lambda^N(t)\right)\right)\right],$$

we see that the factor $\frac{\beta_2}{N}$ can be included in the terms of order $O(1)$. Thus, using that $x \cdot \mathrm{sign}(x) = |x|$, we have

$$N\mathcal{L}^N|y_j(t)|^k \leq -2kN|y_j(t)|^k\left[1 - \beta_1\left(1 - \tanh^2\left(\beta_1\lambda_j^N(t) + \beta_2\Lambda^N(t)\right)\right)\right] + O(1).$$

Finally, observing that the function $f(\lambda_j) := \left[1 - \beta_1\left(1 - \tanh^2\left(\beta_1\lambda_j^N + \beta_2\Lambda^N\right)\right)\right]$ is always positive for $\beta_1 < 1$ and has a unique minimum for $\lambda_j^* = -\frac{k}{\beta_1 + \frac{\beta_2}{N}}$, with $k = \beta_2\frac{1}{N}\sum_{k\neq j}\lambda_k$ such that $f(\lambda_j^*) = 1 - \beta_1$, we can conclude by choosing $C(\beta_1, k) := k(1 - \beta_1)$. □

Remark 4.23. *Proposition 4.22 can be trivially generalized to any timescale of order $N^m t$, yielding*

$$N^m\mathcal{L}^N|y_j^m(t)|^k \leq -CN^m|y_j^m(t)|^k + O(N^{m-1}),$$

with $y_j^m(t) := m_j^N(N^m t) - \tanh\Big(\beta_1(x_j^N(N^m t) + m_j^N(N^m t)) + \beta_2(X^N(N^m t) + M^N(N^m t))\Big)$.

Corollary 4.24. *Let $y_j^m(t)$ be defined as in Remark 4.23. Then, for any $T > 0$, $k > 0$, $m = 1, 2$*

$$\mathbb{E}\left[\sup_{t\in[0,T]}|y_j^m(t)|^k\right] \leq C(N, m, k), \tag{4.80}$$

with $C(N, m, k) \to 0$ for $N \to +\infty$.

Proof. Observing that the infinitesimal generator of the processes (x_j^N, m_j^N) at a timescale of order N^m is $N^m\mathcal{L}^N$, from the contraction estimates (4.79) generalized as in Remark 4.23 it follows

$$\frac{d}{dt}\mathbb{E}\Big[|y_j(t)|^k\Big] \leq -CN^m\mathbb{E}\Big[|y_j(t)|^k\Big] + O(N^{m-1}).$$

Integrating both sides with respect to time we then get claim for any time $t \in [0, T]$, provided that the assertion is true for the initial datum. More precisely, the previous estimate implies

$$\mathbb{E}\Big[|y_j(t)|^k\Big] \leq e^{-C_1N^m t}\mathbb{E}\Big[|y_j(0)|^k\Big] - C_2\frac{N^{m-1}}{N^m}e^{-C_1N^m t} + C_2\frac{N^{m-1}}{N^m}$$
$$= e^{-C_1N^m t}\mathbb{E}\Big[|y_j(0)|^k\Big] - \frac{C_2}{N}e^{-C_1N^m t} + \frac{C_2}{N}.$$

Thus, $\sup_{t\geq 0}\mathbb{E}\Big[|y_j(t)|^k\Big] \leq \mathbb{E}\Big[|y_j(0)|^k\Big] + \frac{C}{N}$. Note that by the assumptions on the initial data we have by a LLN that $\mathbb{E}\Big[|y_j(0)|^k\Big] \to 0$ for $N \to +\infty$. This works both at a timescale of order N and N^2. For getting the stronger convergence (4.80) we again refer to Section 4 of [28] for the diffusive case and to the Appendix of [32] for a general proof for jump processes. We can then conclude as we did in the proof of Proposition 4.2 for the mean field case. □

Proof of Theorem 4.20. As we repeatedly did above, we plug in the definition of the sped up diffusions $x_j^N(t)$ the *same* Brownian motion $W_j(t)$ appearing in the definition of the limit process $\tilde{x}_j(t)$ in (4.77). The weak convergence in distribution is then implied by

$$\lim_{N\to\infty} \mathbb{E}\left[\sup_{t\in[0,T]} |m_j^N(t) - \tilde{m}_j(t)|\right] = 0, \tag{4.81}$$

for the resulting processes, since $W_j^N \overset{\mathcal{D}}{=} W_j$ for $j = 1, \dots, N$. First, we estimate

$$\begin{aligned}
&\mathbb{E}\left[\sup_{s\in[0,t]} |m_j^N(s) - \tilde{m}_j(s)|\right] \\
&\qquad \le \mathbb{E}\left[\sup_{s\in[0,t]} \left|m_j^N(s) - \tanh\Big(\beta_1(x_j^N(s) + m_j^N(s)) + \beta_2(X^N(s) + M^N(s))\Big)\right|\right] \\
&\qquad + \mathbb{E}\left[\sup_{s\in[0,t]} \Big|\tanh\Big(\beta_1(x_j^N(s) + m_j^N(s)) + \beta_2(X^N(s) + M^N(s))\Big)\right. \\
&\qquad\quad \left. - \tanh\left(\beta_1(\tilde{x}_j(s) + \tilde{m}_j(s)) + \beta_2 M(s)\right)\Big|\right].
\end{aligned}$$

The first term in the right hand side of the above inequality is dealt with the contraction estimates of Corollary 4.24 for $m = k = 1$. For the other term we use the global Lipschitz continuity of $\tanh(\cdot)$ in the following way:

$$\begin{aligned}
&\mathbb{E}\left[\sup_{s\in[0,t]} \Big|\tanh\Big(\beta_1(x_j^N(s) + m_j^N(s)) + \beta_2(X^N(s) + M^N(s))\Big)\right. \\
&\quad \left. - \tanh\left(\beta_1(\tilde{x}_j(s) + \tilde{m}_j(s)) + \beta_2 M(s)\right)\Big|\right] \\
&\quad \le \beta_1 \mathbb{E}\left[\sup_{s\in[0,t]} \left|x_j^N(s) - \tilde{x}_j(s)\right|\right] + \beta_1 \mathbb{E}\left[\sup_{s\in[0,t]} \left|m_j^N(s) - \tilde{m}_j(s)\right|\right] \\
&\quad + \beta_2 \mathbb{E}\left[\sup_{s\in[0,t]} \left|X^N(s)\right|\right] + \beta_2 \mathbb{E}\left[\sup_{s\in[0,t]} \left|M^N(s) - M(s)\right|\right].
\end{aligned}$$

For standard arguments of propagation of chaos for the interacting diffusions we have

$$\mathbb{E}\left[\sup_{s\in[0,t]} \left|x_j^N(s) - \tilde{x}_j(s)\right|\right] \le C(N),$$

and

$$\mathbb{E}\left[\sup_{s\in[0,t]} \left|X^N(s)\right|\right] \le C(N),$$

with $C(N) \to 0$ for $N \to +\infty$. For the term $\mathbb{E}\left[\sup_{s\in[0,t]} \left|M^N(s) - M(s)\right|\right]$ we proceed by a coupling as in the proofs of Theorem 4.13, to get

$$\mathbb{E}\left[\sup_{s\in[0,t]} \left|M^N(s) - M(s)\right|\right] \le C(N) + \mathbb{E}\left[\sup_{s\in[0,t]} |m_j^N(s) - \tilde{m}_j(s)|\right].$$

Recollecting all the estimates, we have found

$$(1-\beta_1-\beta_2)\mathbb{E}\left[\sup_{s\in[0,t]}\left|m_j^N(s)-\tilde{m}_j(s)\right|\right] \\ \leq C(N),$$

with $C(N)\to 0$ for $N\to+\infty$. Thanks to the hypothesis $\beta_1+\beta_2<1$ we get (4.81). □

Remark 4.25. *As we did in Remark 4.15, we note that the strong convergence* (4.81) *implies the convergence (e.g. in* 1*-Wasserstein distance) of the associated empirical measures* $\mu^N(t) := \frac{1}{N}\sum_{j=1}^N \delta_{m_j^N(t)}$ *and* $\tilde{\mu}^N(t) := \frac{1}{N}\sum_{j=1}^N \delta_{\tilde{m}_j(t)}$ *to the theoretical distribution of the* $\tilde{m}_j$*'s, again by* (19)*. This in turns implies the propagation of chaos in the classic sense.*

In words, we have found that in the subcritical regime $\beta_1+\beta_2<1$ the equilibrium that the dynamics reaches for long times of order 1 is the same as the equilibrium of the dynamics at long times of order N. The limit dynamics is thus a process moving across the equilibria, due to the movement of the limit diffusion $x(t)$. In particular, define the limit order N dynamics as the pair of processes $(x(t),m(t))_{t\geq 0}$ satisfying

$$\begin{cases} m(t)=\tanh(\beta_1(x(t)+m(t))+\beta_2 M(t)), \\ dx(t)=-\alpha_2 x(t)+\sigma dW(t), \\ M(t)=\mathbb{E}[m(t)], \\ m(0)=0, \\ x(0)=0, \end{cases} \tag{4.82}$$

for which it holds $\left(\tilde{m}_j(t)\right)_{t\in[0,T]} \overset{\mathcal{D}}{=} \left(m(t)\right)_{t\in[0,T]}$ for any $j=1,\dots,N$. Then, we have the analogous of Proposition 4.5:

Proposition 4.26. *The process* $\left(m(t)\right)_{t\geq 0}$ *defined in* (4.82) *is a strong solution to*

$$\begin{cases} dm(t)=\left[-\frac{\alpha_2\beta_1(1-m^2(t))\left(\frac{1}{\beta_1}\operatorname{arctanh}(m(t))-m(t)\right)}{1-\beta_1(1-m^2(t))}-\frac{\beta_1^2\sigma^2 m(t)\left(1-m^2(t)\right)}{\left(1-\beta_1(1-m^2(t))\right)^3}\right]dt \\ \qquad\qquad +\frac{\sigma\beta_1(1-m^2(t))}{1-\beta_1(1-m^2(t))}dW(t), \\ m(0)=0. \end{cases} \tag{4.83}$$

Proof. By Proposition 4.21 it follows that $M(t)\equiv 0$. Thus, by Equation (4.82) we have that $m(t)$ can be written as an *explicit* function of $x(t)$. We can then perform analogous computations as in the proof of Proposition 4.5, by noting that $m(t)$ must be of the form

$$dm(t)=a(t,m(t))dt+b(t,m(t))dW(t)$$

for some functions $a,b:[0,\infty)\times[-1,1]\to\mathbb{R}$ to be determined, and $W(t)$ is the same Brownian motion appearing in the dynamics of $x(t)$ as in (4.83). By applying Itô's formula to the function $\tanh(\beta_1(x(t)+m(t)))$, we find

$$\begin{aligned} dm(t) &= d\left\{\tanh\beta_1(x(t)+m(t))\right\} \\ &= \beta_1[1-\tanh^2\beta_1(x(t)+m(t))](dx(t)+dm(t)) \\ &\quad -\beta_1^2\tanh\beta_1(x(t)+m(t))[1-\tanh^2\beta_1(x(t)+m(t))](b(t,m(t))+\sigma)^2dt \end{aligned}$$

$$
\begin{aligned}
&= \beta_1(1-m^2(t))(dx(t)+dm(t)) - \beta_1^2 m(t)(1-m^2(t))(b(t,m(t))+\sigma)^2 dt \\
&= \beta_1(1-m^2(t))(-\alpha_2 x(t)dt + \sigma dW(t) + a(t,m(t))dt + b(t,m(t))dW(t)) \\
&- \beta_1^2 m(t)(1-m^2(t))(b(t,m(t))+\sigma)^2 dt \\
&= \left[\beta_1(1-m^2(t))(a(t,m(t)) - \alpha_2 x(t)) - \beta_1^2 m(t)(1-m^2(t))(b(t,m(t))+\sigma)^2\right] dt \\
&+ \beta_1(1-m^2(t))\left[\sigma + b(t,m(t))\right] dW(t).
\end{aligned}
$$

Observe that $x(t) = \frac{1}{\beta_1}\operatorname{arctanh}(m(t)) - m(t)$. By reading the diffusion coefficient from the last line, we must have

$$
b(t,m(t)) = \beta_1(1-m^2(t))[\sigma + b(t,m(t))],
$$

and thus

$$
b(t,m(t)) = b(m(t)) = \frac{\sigma\beta_1(1-m^2(t))}{1-\beta_1(1-m^2(t))}.
$$

For the drift term instead

$$
\begin{aligned}
a(t,m(t)) = \beta_1(1-m^2(t))\left(a(t,m(t)) - \alpha_2\left(\frac{1}{\beta_1}\operatorname{arctanh}(m(t)) - m(t)\right)\right) \\
- \beta_1^2 m(t)(1-m^2(t))[(b(t,m(t))+\sigma)^2].
\end{aligned}
\tag{4.84}
$$

As in the proof of Proposition 4.5, we have that

$$
(b(t,m(t))+\sigma)^2 = \frac{\sigma^2}{(1-\beta_1(1-m^2(t)))^2},
$$

and thus, reading from (4.84),

$$
\begin{aligned}
a(t,m(t))(1-\beta_1(1-m^2(t))) = -\alpha_2\beta(1-m^2(t))\left(\frac{1}{\beta_1}\operatorname{arctanh}(m(t)) - m(t)\right) \\
- \beta_1^2 m(t)(1-m^2(t))\frac{\sigma^2}{(1-\beta_1(1-m^2(t)))^2},
\end{aligned}
$$

so that we can conclude. □

Remark 4.27. *The analogous statement to Remark 4.6 holds: for* $\beta_1 < 1$*, the SDE* (4.83) *is well-posed. Existence follows by Proposition 4.26. Uniqueness follows by the Lipschitz properties of the drift and diffusion functions in* $[-1,1]$*. Indeed, note that Equation* (4.83) *differs from* (4.24) *only by an additional drift which is regular and tends to* 0 *at the borders of* $(-1,1)$ *(observe that* $(1-x^2)\operatorname{arctanh}(x) \to 0$ *when* $x \to \pm 1$*).*

Remark 4.28. *Analogously to what we did for the order* 1 *dynamics in Remark 4.17, we can generalize Proposition 4.21 and Theorem 4.20 to the case where the initial data for the diffusions are centered around a point* $\overline{X} \neq 0$*, provided we start the magnetizations' dynamics around the corresponding stable point on the invariant curve (otherwise there would be an initial transient fast dynamics for reaching the corresponding equilibrium). The limit order* N *equation becomes*

$$
\begin{cases}
m(t)(x) = \tanh\left(\beta_1(x + m(t)(x)) + \beta_2(\overline{X} + M(t))\right), \\
m(0)(x) \equiv \overline{m}_{\overline{X}}(x), \\
M(t) = \int_{\mathbb{R}} m(t)(x)\mu_t(dx;\overline{X}),
\end{cases}
\tag{4.85}
$$

for some $\overline{X} \in \mathbb{R}$*, where* $\overline{m}_{\overline{X}}(x)$ *is the solution to* (4.69)*, and* $\mu_t(dx; \overline{X})$ *is a normal distribution with mean* $\overline{X}$ *and variance depending on time (the distribution of the Ornstein-Uhlenbeck diffusions).*

Note that in this case dynamics (4.85) *is not trivial:* $M(t)$ *fluctuates around an equilibrium point due to the time-dependent variance of the Ornstein-Uhlenbeck diffusions, where the equilibrium point depends both on the given* $\overline{X}$ *and on the parameters of the diffusions* σ *and* α_2*. For example, for* $\sigma \gg 1$*, studying the equation for* $M(t)$ *one finds that the equilibrium point is close to* 0 *independently of* $\overline{X}$*. In the long run,* $M(t) \to M(\infty) := \int_{\mathbb{R}} m(t)(x)\mu_\infty(dx; \overline{X})$*, with* $\mu_\infty = \mathcal{N}\left(\overline{X}, \frac{\sigma^2}{2\alpha_2}\right)$*.*

An analogous equation to (4.83) *can also be written, by adding an additional drift term following by the fact that* $x(t) = \frac{1}{\beta_1}\operatorname{arctanh}(m(t)) - m(t) - \frac{\beta_2}{\beta_1}(M(t) + \overline{X})$*, and with initial datum* $m(0) = \overline{m}_{\overline{X}}(\overline{X})$*. Due to the term* $M(t) = \mathbb{E}[m(t)]$ *the resulting equation is a diffusion of McKean-Vlasov type:*

$$\begin{cases} dm(t) = \left[-\frac{\alpha_2\beta_1(1-m^2(t))\left(\frac{1}{\beta_1}\operatorname{arctanh}(m(t))-m(t)-\frac{\beta_2}{\beta_1}(M(t)+\overline{X})\right)}{1-\beta_1(1-m^2(t))} - \frac{\beta_1^2\sigma^2 m(t)\left(1-m^2(t)\right)}{(1-\beta_1(1-m^2(t)))^3} \right] dt \\ \qquad + \frac{\sigma\beta_1(1-m^2(t))}{1-\beta_1(1-m^2(t))} dW(t), \\ m(0) = \overline{m}_{\overline{X}}(\overline{X}), \end{cases} \tag{4.86}$$

whose well-posedeness should also be standard.

4.3.4 Dynamics at times of order N^2: the subcritical case

At this timescale a refined study of the interacting diffusions is needed to describe the limit dynamics. Denoting with t the macroscopic time of order N^2, the single x_j^N's evolve at a much faster timescale with respect to the current value of their empirical mean $X^N(t)$, which is not anymore zero but evolves randomly as a Brownian motion with constant diffusion coefficient σ. Thus, one can expect that in an infinitesimal time dt of order N^2 the single diffusions become asymptotically independent and reach their equilibrium distribution *given* the current value of $X^N(t) = \overline{X}$. In turns, in the same dt the magnetization's processes are also asymptotically i.i.d. and reach an equilibrium given by a macroscopic magnetization $\overline{M}$, whose value can be read off from (4.85) in Remark 4.28, substituting μ_t with μ_∞, the ergodic measure of the Ornstein-Uhlenbeck processes. The reiteration of this procedure for any dt describes the dynamics at the order N^2. In particular, the dynamics at order N^2 does not propagate chaos, unless we condition it with respect to $X^N(t)$.

As before, we still denote the sped up processes under the same notation,

$$x_j^N(t) := x_j^N(N^2 t), \qquad m_j^N(t) := m_j^N(N^2 t),$$

using as initial data the long-time limit at the previous timescale of order N. For clarity we write them again:

$$\begin{cases} dx_j^N(t) = -N\alpha_2(x_j^N(t) - X^N(t))dt + \sqrt{N}\sigma dW_j^N(t), \\ x_j^N(0) = x_j \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}\right), \end{cases} \tag{4.87}$$

with $X^N(t) := \frac{1}{N}\sum_{k=1}^N x_k^N(t)$. The dynamics of the magnetizations is now given by

$$\begin{cases} m_j^N \mapsto m_j^N \pm \frac{2}{N} \text{ rate } N^3 \frac{1\mp m_j^N(t)}{2}\Big(1\pm\tanh\Big[\beta_1(x_j^N(t)+m_j^N(t))+\beta_2(X^N(t)+M^N(t))\Big]\Big)\,, \\ m_j^N(0) = \overline{m}(x_j). \end{cases} \tag{4.88}$$

At this level, we aim to prove that the conditional distribution of the empirical macroscopic magnetization $M^N(t)$ with respect to $X^N(t)$ converges to the conditional distribution of $M(t)$ given $X(t)$ (which is actually a delta), with

$$\begin{cases} m(t)(x) = \tanh\left(\beta_1(x+m(t)(x)) + \beta_2(X(t)+M(t))\right), \\ m(0)(x) \equiv \overline{m}_{X(0)}(x), \\ M(t) = \int_{\mathbb{R}} m(t)(x)\mu_\infty(dx; X(t)), \end{cases} \tag{4.89}$$

where $\mu_\infty(dx; X(t)) = \mathcal{N}\left(X(t), \frac{\sigma^2}{2\alpha_2}\right)$ must be intended as a conditional distribution *given* the current realization of $X(t)$, whose random evolution is

$$\begin{cases} dX(t) = \sigma dW(t), \\ X(0) = 0, \end{cases} \tag{4.90}$$

with W a Brownian motion. Moreover, denoting with

$$Q_t(0, dX) = \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{X^2}{2\sigma^2 t}} dX \tag{4.91}$$

the transition kernel's density at time t associated to the limit diffusion (4.90), we also prove the convergence of the full law of $M^N(t)$ to the law of the process $M(t)$ defined by

$$\begin{cases} m(t)(x) = \tanh\left(\beta_1(x+m(t)(x)) + \beta_2(X(t)+M(t))\right), \\ m(0)(x) \equiv \overline{m}_{X(0)}(x), \\ M(t) = \int_{\mathbb{R}} m(t)(x)\tilde{\mu}^t(dx), \end{cases} \tag{4.92}$$

with

$$\tilde{\mu}^t(\cdot) := \int_{\mathbb{R}} Q_t(0, d\overline{X})\mu_\infty(\cdot; \overline{X}). \tag{4.93}$$

In details, we have

Theorem 4.29 (Limit dynamics at order N^2)**.** *For any $T > 0$, $\beta_1, \beta_2 > 0$ such that $\beta_1 + \beta_2 < 1$ and $\alpha_1, \alpha_2, \sigma > 0$*

(i) For all the finite time dimensional distributions of the form $(t_1, \dots, t_k) \in [0,T]^k$, it holds

$$Law\Big(M^N(t_1), \dots, M^N(t_k)\Big) \to Law\Big(M(t_1), \dots, M(t_k)\Big), \tag{4.94}$$

for $N \to +\infty$, with $M(t)$ the process defined by (4.92) *and* (4.93)*.*

(ii) For every $t \in [0,T]$,

$$Law\Big(M^N(t)\Big| |X^N(t) - X| \le \varepsilon_N\Big) \to \delta_{M(t)}, \tag{4.95}$$

for $N \to +\infty$, with $M(t)$ the (deterministic) variable defined by (4.89) *with $X(t) = X$, and $\varepsilon_N \to 0$ for $N \to +\infty$.*

(iii) *(Conditional propagation of chaos) For every $t \in [0,T]$ and every k-tuple of distinct indexes $j_1, \dots, j_k \in \{1, \dots, N\}^k$, we have*

$$\begin{aligned} Law\Big(m_{j_1}^N(t), \dots, m_{j_k}^N(t) \Big| |X^N(t) - X| \le \varepsilon_N\Big) \to & Law\Big(\tilde{m}_{j_1}(t), \dots, \tilde{m}_{j_k}(t)\Big) \\ &= Law\Big(\tilde{m}_{j_1}(t)\Big)^k, \end{aligned} \tag{4.96}$$

for $N \to +\infty$, where $\tilde{m}_{j_i}(t) := m(t)(x_{j_i})$, with $m(t)(x)$ given by (4.89) with $X(t) = X$, the x_{j_i}'s are i.i.d. random variables distributed as $x \sim \mu_\infty(dx; X) = \mathcal{N}\left(X, \frac{\sigma^2}{2\alpha_2}\right)$, and $\varepsilon_N \to 0$ for $N \to +\infty$.

Note that the well-posedness of the limit dynamics (4.89) and (4.92) can be proved in the same way as we did for the order N case in Proposition 4.21, since any two solutions $m(t)$ and $n(t)$ share the same $X(t)$. Moreover, we point out that we expect property (i) to hold in the stronger sense of weak convergence of stochastic processes, though we did not work out a proof yet. The main ingredients for proving the convergence to the limit at this timescale are provided by Lemmas 4.30 and 4.31. The first establishes a handy distributional representation of the interacting diffusions in terms of a combination of (fast) stationary independent Ornstein-Uhlenbeck processes plus a (slow) independent Brownian motion and a small interaction term. Lemma 4.31 involves a sort of Law of Large Numbers/averaging property for non-linear implicit functions of the magnetizations and of the diffusions. In what follows we strongly rely on the Gaussianity of the interacting processes (4.87). Before stating the next result, we need to introduce the following processes. Let $\left(\xi_j^N(t)\right)_{j=1,\dots,N}$ be defined as,

$$\begin{cases} d\xi_j^N(t) = -\alpha_2 N \xi_j^N(t) dt + \sigma\sqrt{N} dW_j(t), \\ \xi_j^N(0) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}\right), \end{cases} \tag{4.97}$$

with $W_j(t)$ independent Brownian motions, and set $\overline{\xi}_N(t) := \frac{1}{N}\sum_{j=1}^N \xi_j(t)$. Moreover, let $\Big(U_N(t)\Big)_{t\ge 0}$ be defined as

$$\begin{cases} dU_N(t) = \sigma^2 dW(t), \\ U_N(0) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2 N}\right), \end{cases} \tag{4.98}$$

with W a Brownian motion *independent* of all the W_j's. Note that the dependence on N in $U_N(t)$ is only through the initial datum. We are now ready to introduce the following

Lemma 4.30. *Let $(x_j^N(t))_{j=1,\dots,N}$ be as in (4.87). Then, for any $T > 0$, we have that*

(i) *For every $j = 1, \dots, N$ and every $N \in \mathbb{N}$,*

$$Law\Big((x_j^N(t))_{t\in[0,T]}\Big) = Law\Big((\xi_j^N(t) - \overline{\xi}_N(t) + U_N(t))_{t\in[0,T]}\Big). \tag{4.99}$$

(ii) *For every k-tuple of distinct indexes $(j_1, \dots, j_k) \in \{1, \dots, N\}^k$ and every fixed $t \in [0,T]$,*

$$Law\Big(x_{j_1}^N(t), \dots, x_{j_k}^N(t)\Big)(dx) = \int_{\mathbb{R}} Q_t(0, dX)\mu_\infty^k(dx; X) =: \tilde{\mu}^{t,k}(dx), \tag{4.100}$$

for every $N \in \mathbb{N}$, with $\mu_\infty^k(dx; X) = \mu_\infty(dx_1; X) \times \dots \times \mu_\infty(dx_k; X)$.

(iii) (Conditional propagation of chaos) For every k-tuple of distinct indexes $(j_1, \dots, j_k) \in \{1, \dots, N\}^k$ *and every fixed* $t \in [0,T]$,

$$Law\Big(x^N_{j_1}(t), \dots, x^N_{j_k}(t)\Big|\,|X^N(t) - X| \le \varepsilon_N\Big)(dx) \to \mu^k_\infty(dx; X), \tag{4.101}$$

for $N \to +\infty$, *with* $\varepsilon_N \to 0$ *for* $N \to +\infty$.

Proof. Because of the Gaussianity of the mean-zero processes $(x^N_j(t))_{t\ge0}$, $(\xi^N_j(t))_{t\ge0}$ and $(U_N(t))_{t\ge0}$ we can check assertion (i) only looking at the covariance functions. For a fixed $t \ge 0$, denote $A(t) := \mathbb{E}[(x^N_j(t))^2]$ and $B(t) := \mathbb{E}[x^N_j(t)x^N_i(t)]$. Because of the exchangeability of the processes $(x^N_j(\cdot))_{j=1,\dots,N}$ we have that A and B do not depend on j nor i. Applying Itô's formula to $f(x^N_j(t)) = (x^N_j(t))^2$ and to $f(x^N_j(t), x^N_i(t)) = x^N_j(t)x^N_i(t)$, and then taking the expectation, we obtain that $A(t)$ and $B(t)$ must solve

$$\begin{cases} \dot{A}(t) = -2\alpha_2(N-1)A(t) + 2\alpha_2(N-1)B(t) + \sigma^2 N, \\ \dot{B}(t) = -2\alpha_2 B(t) + 2\alpha_2 A(t), \\ A(0) = \frac{\sigma^2}{2\alpha_2}, \\ B(0) = 0, \end{cases} \tag{4.102}$$

which gives

$$\begin{aligned} A(t) &= \frac{\sigma^2(1+2\alpha_2 t)}{2\alpha_2}, \\ B(t) &= \sigma^2 t. \end{aligned} \tag{4.103}$$

Now, fix any $s, t \ge 0$ with $t > s$. Denote $A_N(s,t) := \mathbb{E}[x^N_j(s)x^N_j(t)]$ and $B_N(s,t) := \mathbb{E}[x^N_j(s)x^N_i(t)]$. Clearly, we have $A_N(s,s) = A(s)$ and $B_N(s,s) = B(s)$. The evolution in t of the above quantities can be obtained by applying Itô's formula to $x^N_j(s)x^N_j(t)$ and $x^N_j(s)x^N_i(t)$ on the time interval $[s,t]$, keeping s fixed as an initial datum. We obtain the following system of ODEs in $t \in [s, +\infty)$:

$$\begin{cases} \frac{d}{dt}A_N(s,t) = -(N-1)\alpha_2 A_N(s,t) + (N-1)\alpha_2 B_N(s,t), \\ \frac{d}{dt}B_N(s,t) = -\alpha_2 B_N(s,t) + \alpha_2 A_N(s,t), \\ A_N(s,s) = A(s) = \frac{\sigma^2(1+2\alpha_2 s)}{2\alpha_2}, \\ B_N(s,s) = B(s) = \sigma^2 s, \end{cases} \tag{4.104}$$

whose solution gives

$$\begin{aligned} A_N(s,t) &= \frac{\sigma^2}{2\alpha_2 N}\left[1 - e^{-\alpha_2 N(t-s)}\right] + \frac{\sigma^2}{2\alpha_2}e^{-\alpha_2 N(t-s)} + \sigma^2 s, \\ B_N(s,t) &= A_N(s,t) - \frac{\sigma^2}{2\alpha_2}e^{-\alpha_2 N(t-s)}. \end{aligned} \tag{4.105}$$

Now, denote

$$Y_j(t) := \xi^N_j(t) - \overline{\xi}_N(t) + U_N(t).$$

For any $t \ge 0$ we have

$$\mathbb{E}[Y^2_j(t)] = \left(1 + \frac{1}{N}\right)\mathbb{E}[(\xi^N_j(t))^2] + \mathbb{E}[U^2_N(t)] - \frac{2}{N}\mathbb{E}[(\xi_j(t))^2],$$

and

$$\mathbb{E}[Y_i(t)Y_j(t)] = \mathbb{E}[Y^2_j(t)] - \mathbb{E}[(\xi^N_j(t))^2].$$

For any $t > s$ we get

$$\mathbb{E}[Y_j(s)Y_j(t)] = \left(1 - \frac{2}{N}\right) \mathbb{E}[\xi_j^N(s)\xi_j^N(t)] + \frac{1}{N}\mathbb{E}[\xi_j^N(s)\xi_j^N(t)] + \mathbb{E}[U_N(t)U_N(s)],$$

and

$$\mathbb{E}[Y_j(s)Y_i(t)] = \mathbb{E}[Y_j(s)Y_j(t)] - \mathbb{E}[\xi_j^N(s)\xi_j^N(t)].$$

Note that for the stationary Ornstein-Uhlenbeck processes $\xi_j^N(t)$ we have, for any $t \geq 0$,

$$\mathbb{E}[(\xi_j^N(t))^2] = \frac{\sigma^2}{2\alpha_2},$$

and for $t > s$,

$$\mathbb{E}[(\xi_j^N(t)\xi_j^N(s)] = \frac{\sigma^2}{2\alpha_2} e^{-\alpha_2 N(t-s)}.$$

Moreover, by the independence between the $\xi_j^N(t)$'s,

$$\mathbb{E}[(\overline{\xi}_N(t))^2] = \frac{1}{N}\mathbb{E}[(\xi_j^N(t))^2] = \frac{1}{N}\frac{\sigma^2}{2\alpha_2},$$

and

$$\mathbb{E}[\overline{\xi}_N(t)\overline{\xi}_N(s)] = \frac{1}{N}\frac{\sigma^2}{2\alpha_2} e^{-\alpha_2 N(t-s)}.$$

For $U_N(t)$ we get

$$\mathbb{E}[U_N^2(t)] = \sigma^2 t + \frac{\sigma^2}{2\alpha_2 N},$$

and, for $t > s$,

$$\mathbb{E}[U_N(t)U_N(s)] = \sigma^2 t + \frac{\sigma^2}{2\alpha_2 N}.$$

One can extend the above computations to any $t, s \geq 0$: it suffices to take the minimum between s and t in the above formulae, and multiply by $\text{sign}(t-s)$ in the exponentials. Denoting with $c_N(s,t)$ and $d_N(s,t)$ the covariance functions of $(x_j^N(t))_{t\in[0,T]}$ and $(Y_j(t))_{t\in[0,T]}$ (i.e. the process on the right hand side of (4.99)), the above computations on Y_j and the expressions (4.103) and (4.105) show that, for any $T > 0$,

$$c_N(s,t) = d_N(s,t),$$

so that (i) is proved. For the proof of (ii), recall that

$$\text{Law}\left(x_j^N(t)\right) = \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}(1 + 2\alpha_2 t)\right).$$

On the other hand, note that, integrating in dX, recalling (4.91), (4.93) and $\mu_\infty(dx;X) = \mathcal{N}\left(X, \frac{\sigma^2}{2\alpha_2}\right)$,

$$\begin{aligned}\tilde{\mu}^t(dx) &= \int_{\mathbb{R}} Q_t(0,dX)\mu_\infty(dx;X) \\ &= \left[\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{X^2}{2\sigma^2 t}} \frac{1}{\sqrt{\frac{\pi\sigma^2}{\alpha_2}}} e^{-\frac{(x-X)^2}{\sigma^2/\alpha_2}} dX\right] dx\end{aligned}$$

$$= \frac{1}{\sqrt{\frac{\pi\sigma^2}{\alpha_2}}\sqrt{1+2\alpha_2 t}} e^{-\frac{\alpha_2 x^2}{\sigma^2(1+2\alpha_2 t)}} dx = \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}(1+2\alpha_2 t)\right)(dx),$$

i.e.

$$\mathrm{Law}\Big(x_j^N(t)\Big) = \tilde{\mu}^t, \tag{4.106}$$

for every $N \in \mathbb{N}$, with $\tilde{\mu}^t$ as in (4.93).

We now check the validity of (ii) for bidimensional vectors $(x_i^N(t), x_j^N(t))$, as the assertion then follows by the Gaussianity of the processes in play. By the computations developed for the proof of (i), we know that $(x_i^N(t), x_j^N(t))$ is normally distributed, with $\mathbb{E}[x_i^N(t)] = \mathbb{E}[x_j^N(t)] = 0$, $\mathrm{Var}(x_i^N(t)) = A(t) = \frac{\sigma^2(1+2\alpha_2 t)}{2\alpha_2}$, and $\mathrm{Cov}(x_i^N(t), x_j^N(t)) = B(t) = \sigma^2 t$. Then, we just need to check that $\tilde{\mu}^{t,2}(dx)$, as defined in (ii), has the same moments. Let $(X_1, X_2) \sim \tilde{\mu}^{t,2}$. As one can check (e.g. via Mathematica):

$$\begin{aligned}
\mathbb{E}[X_1 X_2] &= \\
&= \int_{\mathbb{R}^3} x_1 x_2 \frac{1}{\sqrt{2\pi\sigma^2 t}} e^{-\frac{X^2}{2\sigma^2 t}} \left(\frac{1}{\sqrt{\frac{\pi\sigma^2}{\alpha_2}}}\right)^2 e^{-\frac{(x_1-X)^2}{\sigma^2/\alpha_2}} e^{-\frac{(x_2-X)^2}{\sigma^2/\alpha_2}} dX\, dx_1 dx_2 \\
&= \sigma^2 t,
\end{aligned}$$

while the other moments were already verified.

For the proof of (iii), we note that for fixed $j \in \{1, \dots, N\}$ and any $T > 0$ with $t \in [0, T]$,

$$\mathrm{Law}\Big(x_j^N(t) \big| |X^N(t) - X| \le \varepsilon_N\Big) = \mathrm{Law}\left(\xi_j^N(t) - \overline{\xi}_N(t) + U_N(t) \middle| \Big|U_N(t) - X\Big| \le \varepsilon_N\right),$$

since $X^N(t) \overset{\mathcal{D}}{=} U_N(t)$. By noting that

$$\mathbb{E}\Big[\xi_j^N(t) - \overline{\xi}_N(t) + U_N(t) | U_N(t)\Big] = U_N(t),$$

and

$$\mathrm{Var}\Big(\xi_j^N(t) - \overline{\xi}_N(t) + U_N(t) | U_N(t)\Big) = \left(1 - \frac{1}{N}\right)\frac{\sigma^2}{2\alpha_2},$$

we find that

$$\lim_{N\to\infty} \mathrm{Law}\Big(x_j^N(t) \big| |X^N(t) - X| \le \varepsilon_N\Big) = \lim_{N\to\infty} \mathcal{N}\left(X, \left(1 - \frac{1}{N}\right)\frac{\sigma^2}{2\alpha_2}\right) = \mu_\infty(\cdot; X).$$

Furthermore, computing

$$\begin{aligned}
&\mathrm{Cov}\Big(\xi_i^N(t) - \overline{\xi}_N(t) + U_N(t), \xi_j^N(t) - \overline{\xi}_j^N(t) + U_N(t) \Big| \overline{\xi}_N\Big) \\
&\quad = -\frac{2}{N}\mathbb{E}[(\xi_i^N(t))^2] + \mathbb{E}[\overline{\xi}_N^2(t)] = -\frac{1}{N}\frac{\sigma^2}{2\alpha_2} \to 0,
\end{aligned}$$

we can deduce the conditional law of bidimensional vectors $(x_i^N(t), x_j^N(t))$, so that (iii) is verified. □

Lemma 4.31 (Averaging property)**.** *Under the notation above, let* $f : \mathbb{R}^3 \times [-1,1] \to [-1,1]$ *be globally Lipschitz continuous in each variable. Let* L *be the Lipschitz constant with respect to its fourth argument, i.e., for any* $M, M' \in [-1,1]$*,*

$$|f(x_1,x_2,x_3,M) - f(x_1,x_2,x_3,M')| \leq L|M - M'|,$$

for every $(x_1,x_2,x_3) \in \mathbb{R}^3$*, and suppose* $L < 1$*. Let* $\mu(du) = \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}\right)(du)$*. Then, for any* $T > 0$ *we have that*

(i) For every $N \in \mathbb{N}$ *and* $t \in [0,T]$*, the equation*

$$M_N(t) = \frac{1}{N}\sum_{j=1}^{N} f(\xi_j^N(t), \overline{\xi}_N(t), U_N(t), M_N(t)) \tag{4.107}$$

has a unique solution almost surely.

(ii) Let $\Big(B(t)\Big)_{t\geq 0}$ *a Brownian motion. For every finite* k*-tuple of times* $(t_1,\dots,t_k) \in [0,T]^k$*,*

$$Law\Big(M_N(t_1),\dots,M_N(t_k)\Big) \to Law\Big(M(t_1),\dots,M(t_k)\Big), \tag{4.108}$$

for $N \to +\infty$*, where the process* $\Big(M(t)\Big)_{t\geq 0}$ *is defined by*

$$M(t) := \int_{\mathbb{R}} f(u, 0, \sigma^2 B(t), M(t))\mu(du). \tag{4.109}$$

(iii) For every fixed $t \in [0,T]$*,*

$$Law\Big(M_N(t)\Big| |U_N(t) - z| \leq \varepsilon_N\Big) \to \delta_{M(t)}, \tag{4.110}$$

for $N \to +\infty$ *and* $\varepsilon_N \to 0$ *for* $N \to +\infty$*, with*

$$M(t) := \int_{\mathbb{R}} f(u, 0, z, M(t))\mu(du). \tag{4.111}$$

Proof. The map

$$m \mapsto \frac{1}{N}\sum_{j=1}^{N} f(\xi_j^N(t), \overline{\xi}_N(t), U_N(t), m)$$

is L-Lipschitz continuous with $L < 1$. Thus, (i) follows by a contraction argument (e.g. Banach-Caccioppoli Theorem).

For the proof of (ii) we make some preliminary remarks. First, note that by definition of $\overline{\xi}_N(t)$, we have

$$\begin{cases} d\overline{\xi}_N(t) = -\alpha_2 N \overline{\xi}_N(t)dt + \sigma dW^N(t), \\ \overline{\xi}_N(0) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2 N}\right), \end{cases}$$

with $W^N(t) := \frac{1}{\sqrt{N}}\sum_{j=1}^N W_j(t)$, with the W_j's appearing in dynamics (4.97). The solution of the above equation is

$$\overline{\xi}_N(t) = \overline{\xi}_N(0)e^{-\alpha_2 N t} + \sigma \int_0^t e^{-\alpha_2 N(t-s)} dW^N(s),$$

which implies, for any $T > 0$,

$$\mathbb{E}\left[\sup_{t\in[0,T]} |\overline{\xi}_N(t)|\right] \to 0, \tag{4.112}$$

for $N \to +\infty$. Moreover, recalling Equation (4.98) for $\left(U_N(t)\right)_{t\geq 0}$ and the definition of $\left(M_N(t)\right)_{t\geq 0}$ (4.107), we have the almost sure equality between the processes $\left(M_N(t)\right)_{t\geq 0}$ and $\left(M_N^*(t)\right)_{t\geq 0}$, the latter being defined by

$$M_N^*(t) = \frac{1}{N}\sum_{j=1}^N f\left(\xi_j^N(t), \overline{\xi}_N(t), \sigma^2 W(t) + U_N(0), M_N^*(t)\right).$$

Let $\left(\hat{M}_N(t)\right)_{t\geq 0}$ be the process defined by

$$\hat{M}_N(t) = \frac{1}{N}\sum_{j=1}^N f\left(\xi_j^N(t), 0, \sigma^2 W(t), \hat{M}_N(t)\right). \tag{4.113}$$

In light of (4.112), the trivial convergence $\mathbb{E}\left[\sup_{t\in[0,T]} |U_N(t) - \sigma^2 W(t)|\right] \to 0$ for $N \to +\infty$ and the Lipschitz assumptions on f, we obtain

$$\mathbb{E}\left[\sup_{t\in[0,T]} |M_N(t) - \hat{M}_N(t)|\right] \to 0, \tag{4.114}$$

for $N \to +\infty$. In particular $\left(M_N(t)\right)_{t\in[0,T]}$ and $\left(\hat{M}_N(t)\right)_{t\in[0,T]}$ share the same limit in distribution, provided it exists.

Now we fix a $t \in [0, T]$ and prove (ii) for all the one-dimensional distributions. Let $\hat{M}_N(t)(z)$ be the unique solution to

$$\hat{M}_N(t)(z) = \frac{1}{N}\sum_{j=1}^N f(\xi_j^N(t), 0, z, \hat{M}_N(t)(z)),$$

and $M(t)(z)$

$$M(t)(z) = \int_{\mathbb{R}} f(u, 0, z, M(t)(z))\mu(du).$$

If we show that, for every $z \in \mathbb{R}$,

$$\hat{M}_N(t)(z) \to M(t)(z), \tag{4.115}$$

almost surely for $N \to +\infty$, then we have $\hat{M}_N(t) = \hat{M}_N(t)(\sigma^2 W(t)) \to M(t)(\sigma^2 W(t)) = M(t)$ almost surely for $N \to +\infty$, and thus the one-dimensional version of (ii) follows by (4.114). For the proof of (4.115), we omit for the moment the arguments 0 and z, and rewrite

$$\hat{M}_N(t) = \frac{1}{N}\sum_{j=1}^N f(\xi_j^N(t), \hat{M}_N(t)) = \int_{\mathbb{R}} f(u, \hat{M}_N(t))\mu_N(t)(du),$$

where $\mu_N(t) := \frac{1}{N}\sum_{j=1}^N \delta_{\xi_j^N(t)}$ is the empirical measure of the $\xi_j^N(t)$'s. We now set $F : [-1, 1] \times \mathcal{M}_1(\mathbb{R}) \to \mathbb{R}$ to be given by

$$F(m, \mu) := \int_{\mathbb{R}} f(u, m)\mu(du),$$

endowing $\mathcal{M}_1(\mathbb{R})$ with the BL (bounded-Lipschitz) metric

$$||\mu - \nu||_{\mathrm{BL}} = \sup \left\{ \left| \int_{\mathbb{R}} g d\mu - \int_{\mathbb{R}} g d\nu \right| : ||g||_\infty \leq 1,\ g\ 1 - \mathrm{Lip.} \right\}.$$

Note that $m \mapsto F(m, \mu)$ is L-Lipschitz, so that there exists a unique $m(\mu)$ such that

$$m(\mu) = F(m(\mu), \mu).$$

Moreover, we have

$$\begin{aligned}
|m(\mu) - m(\nu)| &= \left| \int_{\mathbb{R}} f(u, m(\mu))\mu(du) - \int_{\mathbb{R}} f(u, m(\nu))\nu(du) \right| \\
&\leq \int_{\mathbb{R}} \left| f(u, m(\mu)) - f(u, m(\nu)) \right| \mu(du) \\
&+ \left| \int_{\mathbb{R}} f(u, m(\nu))\mu(du) - \int_{\mathbb{R}} f(u, m(\nu))\nu(du) \right| \\
&\leq L|m(\mu) - m(\nu)| + ||\mu - \nu||_{\mathrm{BL}},
\end{aligned}$$

so that

$$|m(\mu) - m(\nu)| \leq \frac{||\mu - \nu||_{\mathrm{BL}}}{1 - L}.$$

In particular, $m(\mu)$ is continuous in μ. Finally, since $\hat{M}_N(t) = m(\mu_N(t))$ and by a LLN $\mu_N(t) \to \mu = \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}\right)$ almost surely, we have that, restoring the dependence on z in the previous expression,

$$\hat{M}_N(t)(z) \to m(t)(\mu) = \int_{\mathbb{R}} f(u, 0, z, m(t)(\mu))\mu(du) = M(t)(z),$$

so that (4.115) is proved. Recall that, for any t, the above implies

$$\hat{M}_N(t)(\sigma^2 W(t)) \to M(t)(\sigma^2 W(t)) \tag{4.116}$$

almost surely for $N \to +\infty$. The finite time dimensional version follows directly by (4.116) and by the continuity of the processes with respect to time, which yield

$$\mathbb{P}\left(\hat{M}_N(t)(\sigma^2 W(t)) \xrightarrow{N \to +\infty} M(t)(\sigma^2 W(t)),\quad \forall t \in [0, T]\right) = 1.$$

Assertion (iii) follows directly by (ii). It is indeed the corresponding conditional statement of the one-dimensional version of (ii) noting, as we did above, that $U_N(t) \overset{\mathcal{D}}{=} \sigma^2 W(t) + U_N(0)$ for every N, with $U_N(0) \to 0$ for $N \to +\infty$, and

$$\begin{aligned}
\mathbb{E}[\xi_j^N(t)|U_N(t)] &= \mathbb{E}[\xi_j^N(t)] = 0, \\
\mathrm{Var}(\xi_j^N(t)|U_N(t)) &= \mathrm{Var}(\xi_j^N(t)) = \frac{\sigma^2}{2\alpha_2}.
\end{aligned}$$

The limit distribution is a delta in $M(t)$ since $M(t) = \int_{\mathbb{R}} f(u, 0, z, M(t))\mu(du)$ is deterministic. □

Proof of Theorem 4.29. Consider the set of N processes $(\tilde{m}_j^N(t), \tilde{M}^N(t))_{j=1,\dots,N}$, coupled with $(m_j^N(t), M^N(t))_{j=1,\dots,N}$, defined by

$$\begin{cases} \tilde{m}_j^N(t) = \tanh\left(\beta_1(x_j^N(t) + \tilde{m}_j^N(t)) + \beta_2(X^N(t) + \tilde{M}^N(t))\right), \\ \tilde{m}_j^N(0) = m_j^N(0), \\ \tilde{M}^N(t) = \frac{1}{N}\sum_{j=1}^N \tanh\left(\beta_1(x_j^N(t) + \tilde{m}_j^N(t)) + \beta_2(X^N(t) + \tilde{M}^N(t))\right). \end{cases} \tag{4.117}$$

By the contraction estimates (4.80) for $k = 1, m = 2$, we know that both $m_j^N(t) - \tilde{m}_j^N(t)$ and $M^N(t) - \tilde{M}^N(t) \to 0$ in strong norm, for $N \to +\infty$. Indeed, (4.80) can be trivially adapted to show that $M^N(t)$ collapses onto the empirical mean of the processes laying on the invariant curve. It is then sufficient to study the convergence in distribution of $\left(\tilde{M}^N(t)\right)_{t\in[0,T]}$. We first observe that, by (i) of Lemma 4.30, for every $N \in \mathbb{N}$ it holds

$$(\tilde{m}_j^N(t), \tilde{M}^N(t))_{t\in[0,T]} \stackrel{\mathcal{D}}{=} (\hat{m}_j^N(t), \hat{M}^N(t))_{t\in[0,T]},$$

with $(\hat{m}_j^N(t), \hat{M}^N(t))_{j=1,\dots,N}$ given by

$$\begin{cases} \hat{m}_j^N(t) = \tanh\left(\beta_1(\xi_j^N(t) - \overline{\xi}_N(t) + U_N(t) + \hat{m}_j^N(t)) + \beta_2(U_N(t) + \hat{M}^N(t))\right), \\ \hat{m}_j^N(0) = m_j^N(0), \\ \hat{M}^N(t) = \frac{1}{N}\sum_{j=1}^N \tanh\Big(\beta_1(\xi_j^N(t) - \overline{\xi}_N(t) + U_N(t) + \hat{m}_j^N(t)) + \beta_2(U_N(t) + \hat{M}^N(t))\Big), \end{cases} \tag{4.118}$$

with $\xi_j^N(t)$ and $U_N(t)$ given by (4.97) and (4.98) respectively. Now, we note that the function

$$\varphi(\xi, \overline{\xi}, U, M) := \tanh\left(\beta_1(\xi - \overline{\xi} + U + \varphi(\xi, \overline{\xi}, U, M)) + \beta_2(U + M)\right)$$

satisfies the Lipschitz properties of Lemma 4.31 for any choice of $\beta_1, \beta_2 > 0$ such that $\beta_1 + \beta_2 < 1$. Indeed, the Lipschitz continuity in the first three variables follows from the regularity of $\tanh(\cdot)$. For the last argument of φ, for any $M, M' \in [-1, 1]$, we estimate

$$\left|\varphi(\xi, \overline{\xi}, U, M) - \varphi(\xi, \overline{\xi}, U, M')\right| \le \beta_1 \left|\varphi(\xi, \overline{\xi}, U, M) - \varphi(\xi, \overline{\xi}, U, M')\right| + \beta_2 |M - M'|,$$

so that

$$\left|\varphi(\xi, \overline{\xi}, U, M) - \varphi(\xi, \overline{\xi}, U, M')\right| \le \frac{\beta_2}{1 - \beta_1} |M - M'|.$$

Thus, φ is L-Lipschitz continuous in M with $L := \frac{\beta_2}{1-\beta_1} < 1$ if and only if $\beta_1 + \beta_2 < 1$. We can then apply (ii) of Lemma 4.31 to $\hat{M}^N(t)$, to get, for all the finite time dimensional distributions $(t_1, \dots, t_k) \in [0, T]^k$,

$$\mathrm{Law}\Big(\hat{M}^N(t_1), \dots, \hat{M}^N(t_k)\Big) \to \mathrm{Law}\Big(M^*(t_1), \dots, M^*(t_k)\Big),$$

for $N \to +\infty$, where

$$M^*(t) := \int_{\mathbb{R}} \varphi(u, 0, \sigma^2 W(t), M^*(t)) \mu(du),$$

with $\mu = \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}\right)$. Assertion (i) is then implied by

$$\mathrm{Law}\Big(\big(M^*(t)\big)_{t\in[0,T]}\Big) = \mathrm{Law}\Big(\big(M(t)\big)_{t\in[0,T]}\Big), \tag{4.119}$$

with $M(t)$ as in (4.92). We start by proving (4.119) for all the one-dimensional time distributions. For the purpose, we note that for $t \in [0,T]$

$$\sigma^2 W(t) \stackrel{\mathcal{D}}{=} X(t) \sim \mathcal{N}(0, \sigma^2 t),$$

with $X(t)$ the limit in distribution of $X^N(t)$. Substituting in $M^*(t)$, we have the equality in distribution

$$M^*(t) = \int_{\mathbb{R}} \tanh\Big(\beta_1(u + X(t) + \varphi(u, 0, X(t), M^*(t))) + \beta_2(X(t) + M^*(t))\Big)\mu(du).$$

Finally, with the change of variable $x := u + X(t)$, noting that, by the computations in Lemma 4.30, we have that the random variable $x(t) := \xi + \sigma^2 W(t) \stackrel{\mathcal{D}}{=} \xi + X(t)$, with $\xi \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}\right)$ is distributed according to

$$\mathrm{Law}(x(t))(dx) = \tilde{\mu}^t(dx),$$

the relation (4.119) is proved for the one-dimensional time marginal distributions. The analogous conclusion is immediately obtained for all the finite time dimensional distributions, by using properties (i) and (ii) of Lemma 4.30, which hold for any N and thus also for the limit. The equality in law for the whole process follows, since both $\Big(M^*(t)\Big)_{t\in[0,T]}$ and $\Big(M(t)\Big)_{t\in[0,T]}$ are functions of the same Gaussian process $(X(t))_{t\in[0,T]}$.

The proof of (ii) follows easily by property (iii) of Lemma 4.31 and by an analogous change of coordinates as above. In details, we know that, with the above notation

$$\mathrm{Law}\Big(\tilde{M}^N(t)\Big| |X^N(t) - X| \le \varepsilon_N\Big) = \mathrm{Law}\Big(\hat{M}^N(t)\Big| |U_N(t) - X| \le \varepsilon_N\Big),$$

for some $\varepsilon_N \to 0$ when $N \to +\infty$. By property (iii) of Lemma 4.31 and for the above couplings, this implies

$$\mathrm{Law}\Big(M^N(t)\Big| |X^N(t) - X| \le \varepsilon_N\Big) \to \delta_{M(t)},$$

where

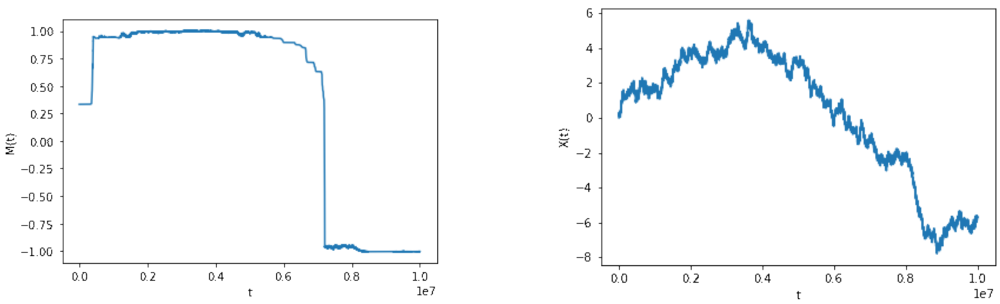
$$M(t) = \int_{\mathbb{R}} \varphi(u, X, M(t))\mu(du).$$

With the change of coordinates $x = u + X$ we get (4.95).

For the proof of (iii), consider a single process $\hat{m}_j^N(t)$, as given in (4.118), for a fixed $t \in [0,T]$. Combining assertion (iii) of Lemma 4.30 with (ii) of this theorem, it follows directly

$$\mathrm{Law}\Big(\tilde{m}_j^N(t)\Big| |X^N(t) - X(t)| \le \varepsilon_N\Big) \to \mathrm{Law}\Big(\tilde{m}_j(t)\Big).$$

The asymptotic independence among the magnetizations, i.e. (iii), follows by noting that the m_j^N's (resp. $\tilde{m}_j^N$) are functions of x_j^N, X^N, and M^N (resp. $\tilde{M}^N$). When we condition with respect to $X^N(t)$, we have that the $x_j^N(t)$'s are asymptotically independent by (iii) of Lemma 4.30, and $M^N(t)$ tends to a deterministic value by (ii). Thus the dependence among the magnetizations is deleted in the limit. □



(a) $(t, M^N(t))$ (b) $(t, X^N(t))$

Figure 4.4: Simulation of the finite particle system's dynamics at a timescale of order N^2, for $N = 200$, $\beta_1 = \beta_2 = 0.3$, $\alpha_1 = \alpha_2 = \sigma = 1$, $T = 10^7$.

Remark 4.32 (One-dimensional description)**.** *Theorem 4.29 can potentially be adapted to get rid of the dependence of the diffusions in the definition of the magnetizations dynamics. The corresponding ergodic measure (the analogous of $\mu_\infty(dx; X)$) with respect to which the order N^2 results should hold, is the conditional (with respect to M) ergodic measure associated to the McKean-Vlasov SDE* (4.86)*, which of course is much harder to find than $\mu_\infty(dx; X)$.*

Remark 4.33 (Long-time behavior)**.** *Despite the well-posedness of the limit dynamics for any finite time interval, observe that in this timescale the long-time behavior of the limit equation cannot be determined, as the process $(X(t))_{t\geq 0}$ does not admit an invariant measure on the whole space. However, it is clear that for big positive values of $X(t)$ the second-level magnetization $M(t)$ will be close to $+1$, while for big negative values it will be close to -1, as it is shown in Figure 4.4.*

Remark 4.34 (Mean field case: the conditional propagation of chaos)**.** *We note that an analogous result to the conditional propagation of chaos* (*iii*) *of Theorem 4.29 should hold for the mean field model of Section 4.2, where the single spins (resp. diffusions) σ_j (resp. x_j) replace the magnetizations m_j^N (resp. x_j^N) , m^N (resp. x^N) replaces M^N (resp. X^N) and the timescale is of order N instead of order N^2.*

4.3.5 Renormalization theory: the subcritical case

The results of the previous section are expected to be generalizable to the k-th hierarchical level for any $k > 0$ finite. In this section we state what we think should be the corresponding statement, in form of a conjecture, as we did not work out a proof yet. The goal is to define inductively a *renormalization* map φ_d, with $d = 1, \dots, k$, which allows one to describe the limit dynamics for the aggregated magnetizations at each timescale $N^d t$ in terms of the corresponding aggregated diffusions.

In this case the model is defined on the set $V := \{1, \dots, N\}^k$. Any of the N^k individuals in the population is identified by a k-tuple $i = (i_1, i_2, \dots, i_k)$. For any two individuals $i, j \in V$, define the *hierarchical distance* as

$$d(i, j) := \min \left\{ d \mid 0 \leq d \leq k-1, \quad (i_{d+1}, \dots, i_k) = (j_{d+1}, \dots, j_k) \right\}. \tag{4.120}$$

If in (4.120) $(i_{d+1}, \ldots, i_k) \neq (j_{d+1}, \ldots, j_k)$ for any $0 \le d \le k-1$, then we set $d(i,j) := k$. The interaction among individuals (i,j) at distance $d(i,j) = d$ now scales as

$$\begin{aligned} J_{ij} &= \frac{\beta_d}{N^d}, \\ J'_{ij} &= \frac{\alpha_d}{N^{2d-1}}. \end{aligned} \tag{4.121}$$

For $d = 1, 2, \ldots, k-1$ and $i \in V$, set

$$i^d := (i_{d+1}, \ldots, i_k) \in \{1, 2, \ldots, N\}^{k-d},$$

and

$$i^k := \emptyset.$$

Denote for any $d < k$ the N^{k-d} *d-th level magnetizations*,

$$m^N_{i^d}(t) := \frac{1}{N^d} \sum_{j \in V : j^d = i^d} \sigma_j(t),$$

and the *k-th level magnetization*

$$m^N_{i^k}(t) := \frac{1}{N^k} \sum_{j \in V} \sigma_j(t).$$

Moreover, denote for any $d < k$ the limit *d-th level diffusion*,

$$\begin{cases} dX^{(d)}(t) = -\alpha_{d+1} X^{(d)}(t)dt + \sigma dW^{(d)}(t), \\ X^{(d)}(0) = 0, \end{cases} \tag{4.122}$$

and the limit *k-th level diffusion*

$$\begin{cases} dX^{(k)}(t) = \sigma dW(t), \\ X^{(k)}(0) = 0. \end{cases} \tag{4.123}$$

Let $Q^{(d)}_t(0, dX)$ and $Q^{(k)}_t(0, dX)$ be the transition kernels of the diffusions (4.122) and (4.123) respectively, and $\nu^{(d)}_y$ the stationary distribution of

$$\begin{cases} dz(t) = -\alpha_{d+1}(z(t) - y)dt + \sigma dB(t), \\ z(0) = 0, \end{cases} \tag{4.124}$$

where $B(t)$ is a Brownian motion. Note that, in the notation of the previous section $\nu^{(1)}_y(\cdot) = \mu_\infty(\cdot; y)$. Then, we have

Conjecture 4.1. *Assume $\beta_1 + \cdots + \beta_k < 1$ and that $x_j(0) \sim \mathcal{N}(0,1)$ i.i.d. for any $j \in V$. Then, for any i^d, $d \in \{1, 2, \ldots, k\}$ and $T > 0$,*

$$(m^N_{i^d}(N^d t))_{t \in [0,T]} \to (m^{(d)}(t))_{t \in [0,T]} \tag{4.125}$$

in the sense of weak convergence of stochastic processes, where

$$m^{(d)}(t) = \varphi_d(X^{(d)}(t), 0), \tag{4.126}$$

with $X^{(d)}$ *the solution to* (4.122) *(resp.* (4.123)*) for* $d < k$ *(resp.* $d = k$*). The function* $\varphi_d = \varphi_d(x, y)$*, for* $d = 2, \dots, k$*, is the unique solution to*

$$\varphi_d = \int_{\mathbb{R}} \varphi_{d-1}(z, \beta_d(\varphi_d + x) + y)\tilde{\mu}^{t,d}(dz), \tag{4.127}$$

with, for $d = 2, \dots, k$*,*

$$\tilde{\mu}^{t,d} := \underbrace{\int_{\mathbb{R}} \cdots \int_{\mathbb{R}}}_{d-1 \text{ times}} Q_t^{(d)}(0, dx_{d-1})\nu_{x_{d-1}}^{(d-1)}(dx_{d-2}) \cdot \ldots \cdot \nu_{x_2}^{(2)}(dx_1)\nu_{x_1}^{(1)}, \tag{4.128}$$

and $\varphi_1(x, y)$ *is the unique solution to*

$$\varphi_1 = \tanh(\beta_1(\varphi_1 + x) + y).$$

We conclude the section justifying the k-th level subcritical regime condition

$$\beta_1 + \cdots + \beta_k < 1 \tag{4.129}$$

in the above conjecture, which is in accordance with the first two hierarchical levels of the previous sections. Let L_{d-1} be the Lipschitz constant of φ_{d-1} in its second variable, with $L_{d-1} < 1$. At the d-th hierarchical level, Equation (4.127) has a unique solution, provided that the right hand side is a contraction in terms of φ_d. This is true if

$$L_{d-1}\beta_d < 1. \tag{4.130}$$

On the other hand, computing the d-th level Lipschitz constant L_d we find, for $y, y' \in \mathbb{R}$,

$$|\varphi_d(x, y) - \varphi_d(x, y')| \le L_{d-1}(1 + \beta_d L_d)|y - y'|,$$

and thus

$$L_d = \frac{L_{d-1}}{1 - L_{d-1}\beta_d}. \tag{4.131}$$

Using (4.131), the subcriticality condition (4.129) implies inductively the validity of (4.130) for $d = 1, \dots, k$, starting from $L_0 = 1$.

4.3.6 The limit case: $[\beta_1 = \beta_2 \to \infty]$

In this section we develop solid heuristics for dealing with the limit case of null temperatures. We also show some simulations which confirm our ideas. For convenience, here we use the notation

$$\begin{aligned}
\mu_0(dx) &:= \mathcal{N}(0, \rho^2), \\
\mu_0^t(dx) &:= \mathcal{N}(0, \rho^2(t)), \\
\mu_X(dx) &:= \mathcal{N}(X, \rho^2), \\
\mu_X^\infty(dx) &:= \mu_\infty(dx; X) = \mathcal{N}\left(X, \frac{\sigma^2}{2\alpha_2}\right),
\end{aligned} \tag{4.132}$$

for the normal distributions we consider, where ρ^2 and $\rho^2(t)$ depend on the diffusion parameters σ and α_2. We analyze the limit dynamics at any timescale, formally replacing $\beta_1 = \beta_2 = \infty$ in the deterministic limit equations, where the second level diffusion enters as a parameter. Substituting $\tanh(\beta_1 z + \beta_2 w)$ with $\text{sign}(z+w)$, the main focus of the section is on the study of

$$\begin{cases} \dot{m}(t)(x) = 2\text{sign}(x + m(t)(x) + X + M(t)) - 2m(t)(x), \\ m(0)(x) = m_0(x), \\ M(t) = \int_{\mathbb{R}} m(t)(x)\nu(dx), \end{cases} \tag{4.133}$$

and of its equilibria, where X has to be intended as a fixed value of the second level diffusion, and the measure $\nu(dx)$ is a normal distribution chosen among the ones in (4.132), depending on the timescale considered. We think of (4.133) as an infinitesimal time step in the dynamics at a timescale of order N^α for $\alpha > 0$, with X being the value of the second level diffusion at the current macroscopic time. The spirit of our approach is the following:

- we identify all the equilibria reached at an order 1 timescale, showing that they are provided by *staircase* functions

$$m^{x_0}(x) := \begin{cases} +1, & \forall x > x_0, \\ -1, & \forall x < x_0, \end{cases}$$

  where the discontinuity point $x_0 \in \mathbb{R}$ belongs to a certain interval which we refer to as *fixed points region*. In particular, we show that the fixed points region depends on the current value of the macroscopic quantity X and on the diffusion parameters σ and α_2 (Propositions 4.36, 4.40);

- we deduce the local stability of the above configurations (Proposition 4.38);

- we update the macroscopic time (either with an infinitesimal change at order N or N^2), and evolve X to a new value $\overline{X}$;

- we quantify the corresponding adaptation of the staircase profiles to the change in the environment, distinguishing between the order N dynamics (Section 4.3.6.2), where the first level diffusions x_j's already evolve non-trivially, and the order N^2 dynamics (Section 4.3.6.3), where the first level diffusions have reached a stationary equilibrium with the environment;

- in case the previous step brings the magnetization profile to a *non-equilibrium* configuration, we quantify the way it approaches again the fixed points region (Propositions 4.39, 4.42);

- the reiteration of the above steps allows for a heuristic description of the order N^2 dynamics;

- we show simulations of the finite particle system at any timescale, confirming the above facts and highlighting the remaining open problems.

In particular, we observe the following phenomenon: the N^2 dynamics undergoes a *phase transition*, depending on the diffusion parameters σ and α_2. Specifically, for big values of $\frac{\sigma^2}{2\alpha_2}$, $M^N(t)$ approximately evolves as a two-dimensional diffusion inside the fixed points region, where the additional dimension (with respect to the subcritical case) is a consequence of the new degree of freedom given by the position of the discontinuity point x_0. When $\frac{\sigma^2}{2\alpha_2}$ is small instead, $M^N(t)$ behaves as a two-dimensional diffusion with jumps (see Figure 4.10). The motivation for this is a loss of stability, not yet fully understood, of certain areas of the fixed points region which happens already at an order N timescale: indeed, our approach of infinitesimal time step variations does not allow for a proper justification of this phenomenon.

4.3.6.1 Order 1 dynamics

When $\beta_1 = \beta_2 \to \infty$, we have that the dynamics at order 1 is given by

$$\begin{cases} \dot{m} = 2\text{sign}(2m) - 2m, \\ m(0) = 2p - 1, \end{cases} \tag{4.134}$$

which, when $t \to \infty$ reaches the equilibrium point

$$m = \text{sign}(2m) = \text{sign}(m).$$

Clearly, Equation (4.134) has the same behavior as the low temperature limit of the Curie–Weiss model. It is easy to see that the solution $m = 0$ is unstable and the two polarized solutions $m = \pm 1$ are stable, where the one getting picked asymptotically is determined by the initial sign of $m(0)$. We denote the three equilibria of (4.134) by $m_0, m_\pm$. We now let the dynamics evolve until a time of order N, when the dynamics of the diffusions is not trivial anymore. Let this time be our new initial time, and let the system evolve again at times of order 1. The mean field equations are now

$$\begin{cases} \dot{m}(t)(x) = 2\text{sign}(x + m(t)(x) + M(t)) - 2m(t)(x), \\ m(0)(x) = m_0(x), \\ M(t) = \int_{\mathbb{R}} m(t)(x)\mu_0(dx), \end{cases} \tag{4.135}$$

with μ_0 as in the first line of (4.132), for some $\rho > 0$ which depends on the previous evolution of the diffusions and $m_0(x)$ is close (but not necessarily equal) to the function constantly equal to one of the three equilibria $m_0, m_\pm$. The study of the asymptotic profile of Equation (4.135) helps us in understanding what the dynamics at longer timescales will be. Indeed, as we already stressed, the further timescales dynamics are expected to be described by motions across the different equilibria profiles of order 1, triggered by the dynamics of the diffusions.

For studying the asymptotic profiles of the magnetization $m(t)(x)$ we make an *ansatz* on their shape, motivated by the following preliminary remark

Remark 4.35. *Any asymptotic equilibrium $m^*(x)$ of Equation (4.135) is such that*

$$m^*(x) := \begin{cases} +1, & \forall x > 2, \\ -1, & \forall x < -2. \end{cases}$$

Indeed, for any $t > 0$, we have $-2 \leq m(t)(x) + M(t) \leq 2$, and thus for $x > 2$, $\dot{m}(t)(x) > 0$ and, symmetrically, for $x < -2$, $\dot{m}(t)(x) < 0$.

Even though a full proof of the validity of the below ansatz is not established, as we expect it to hold we state it as a

Proposition 4.36 (Shape of the equilibria)**.** *Every equilibrium of Equation (4.135) is a staircase function $m^{x_0}(x)$ of the form*

$$m^{x_0}(x) := \begin{cases} +1, & \forall x > x_0, \\ -1, & \forall x < x_0, \end{cases} \tag{4.136}$$

for some $x_0 \in \mathbb{R}$ *satisfying*

$$-2\mu_0(x_0,+\infty) \le x_0 \le 2\mu_0(-\infty,x_0). \tag{4.137}$$

Proof. We restrict ourselves to prove one direction of the ansatz, which is easy. Indeed, a profile $m^{x_0}(x)$ is an equilibrium for the dynamics (4.135) if

$$m^{x_0}(x) = \mathrm{sign}(x + m^{x_0}(x) + M), \tag{4.138}$$

with

$$M = \int_{\mathbb{R}} m^{x_0}(x)\mu_0(dx) = -\mu_0(-\infty,x_0) + \mu_0(x_0,\infty) = 1 - 2\mu_0(-\infty,x_0). \tag{4.139}$$

For (4.138) to be satisfied it must be, when $x < x_0$,

$$x - 1 + M < 0,$$

while, for $x > x_0$

$$x + 1 + M > 0.$$

Using (4.139), the above inequalities become

$$x - 1 + 1 - 2\mu_0(-\infty,x_0) = x - 2\mu_0(-\infty,x_0) < 0,$$

for $x < x_0$, and

$$x + 1 + 1 - 2\mu_0(-\infty,x_0) = x + 2\mu_0(x_0,+\infty) > 0,$$

where in the second equality we have used $1 - 2\mu_0(-\infty,x_0) = -1 + 2\mu_0(x_0,+\infty)$. Because of the monotonicity of the above conditions with respect to x, they can be equivalently stated as

$$x_0 - 2\mu_0(-\infty,x_0) \le 0,$$

and

$$x_0 + 2\mu_0(x_0,+\infty) \ge 0.$$

Finally, observe that when $x_0 \ge 0$ the second inequality is trivially true and thus the right inequality of (4.137) is the equilibrium condition for this case, while for $x_0 \le 0$ the first is the trivial one, so that we obtain the left inequality of (4.137) as a necessary condition for the equilibrium. □

Remark 4.37. *Note that in* (4.137)*, both at a timescale of order* 1 *and* N*,* μ_0 *is a normal distribution centered in* 0 *with variance* ρ^2 *(possibly depending on the (macroscopic) time and on the diffusion parameters* σ *and* α_2*). Condition* (4.137) *restricts to*

$$-2 \le x_0 \le 2,$$

when $\rho \to 0$*, and to*

$$-1 \le x_0 \le 1,$$

when $\rho \to \infty$*. Moreover, the fixed points interval is monotonically decreasing with* ρ*, since* $\mu_0(-\infty,x_0)$ *is so.*

As an example of convergence to the equilibrium, let us fix a constant initial datum $m(0)(x) \equiv \overline{m}$ (with $0 < \overline{m} < \frac{1}{2}$) for (4.135) and reason heuristically by small variations of time. Since at the initial time $M(0) = \overline{m}$, for every $x > -2\overline{m}$ one has that $\frac{d}{dt}m(t)(x)\Big|_{t=0} > 0$, and symmetrically, for every $x < -2\overline{m}$, we have $\frac{d}{dt}m(t)(x)\Big|_{t=0} < 0$. One can expect that these considerations should keep being true for any $t > 0$ as the quantities inside the sign function increase/decrease monotonically with time (this is not precise because of the term $M(t)$ inside the sign). The same argument works for $-\frac{1}{2} < \overline{m} < 0$, and for the symmetric case $\overline{m} = 0$. The limit configuration, denoted by $m^*(x)$, is thus given by

$$m^*(x) := \begin{cases} -1, & \text{for } x < -2\overline{m}, \\ 0, & \text{for } x = -2\overline{m}, \\ +1, & \text{for } x > -2\overline{m}. \end{cases} \tag{4.140}$$

By integrating (4.140) over the diffusion's distribution we obtain the asymptotic value of M,

$$M = \int_{-2\overline{m}}^{2\overline{m}} \mu_0(dx). \tag{4.141}$$

Depending on the variance parameter of the distribution μ_0, the resulting asymptotic value of M can either be greater or smaller than the initial one (or equal to in the symmetric case $\overline{m} = 0$). The bigger the variance of μ_0, the more M would tend to be depolarized in this limit.

Proposition 4.36 asserts that there exists a whole region of fixed points for Equation (4.135). Concerning the stability properties of these equilibria, we have that

Proposition 4.38 (Stability of the equilibria)**.** *The equilibrium $m^{x_0}(x)$ is locally stable for the dynamics* (4.135) *if inequality* (4.137) *holds.*

Proof. The proof is non-rigorous. Fix e.g. $x_0 > 0$. Choose as initial condition for (4.135) $m_0(x) = \tilde{m}(x)$, the perturbation of $m^{x_0}(x)$ in a point $\tilde{x} > x_0$, given by

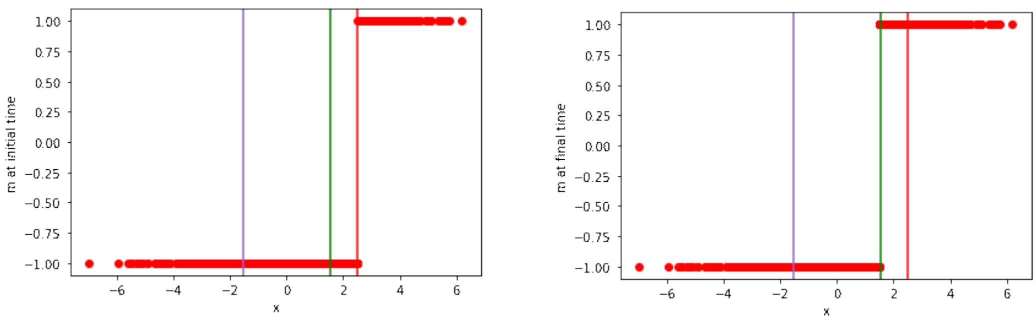
$$\tilde{m}(x) := \begin{cases} \tilde{m}(x) = m^{x_0}(x), & \forall x \neq \tilde{x} \\ \tilde{m}(\tilde{x}) = m^{x_0}(\tilde{x}) - \varepsilon. \end{cases}$$

Then we have that, heuristically, $\frac{d}{dt}m(t)(\tilde{x})\Big|_{t=0} = \varepsilon > 0$. Analogously, if $\tilde{x} < x_0$ we consider $\tilde{m}(x)$, defined as

$$\tilde{m}(x) := \begin{cases} \tilde{m}(x) = m^{x_0}(x), & \forall x \neq \tilde{x} \\ \tilde{m}(\tilde{x}) = m^{x_0}(\tilde{x}) + \varepsilon, \end{cases}$$

so that $\frac{d}{dt}m(t)(\tilde{x})\Big|_{t=0} = -\varepsilon < 0$. □

To sum up, as we saw above, when we start the dynamics with a constant initial datum $m_0(x) \equiv \overline{m}$ we soon get attracted (at times of order 1) to a staircase equilibrium $m^{x_0}(x)$ for some $x_0 \in \mathbb{R}$. The next proposition, for which we do not have a proof (but is motivated by Proposition 4.38 and supported by numerics), describes what happens at a timescale of order 1 when we start the dynamics (4.135) with a staircase initial datum $m_0(x) = m^{x_0}(x)$ with x_0 not belonging to the fixed points region given by (4.137).



(a) Initial configuration at order 1

(b) Final configuration at order 1

Figure 4.5: Simulation of the finite particle system's dynamics at a timescale of order 1, for $N = 1000$, $\beta_1 = \beta_2 = \infty$, $\alpha_1 = \alpha_2 = 1$, $\sigma = 3$. We start the dynamics with a staircase function (the red line) outside the fixed points region's band (purple and green lines). We take $x_j(0) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}\right)$.

Proposition 4.39 (Stable attractors of the dynamics)**.** *Let $m(t)(x)$ be the solution to Equation* (4.135) *with initial datum $m_0(x) = m^{x_0}(x)$, with $x_0 > 0$ (resp. $x_0 < 0$) such that $x_0 > 2\mu_0(-\infty, x_0)$ (resp. $x_0 < -2\mu_0(x_0, +\infty)$). Then, we have*

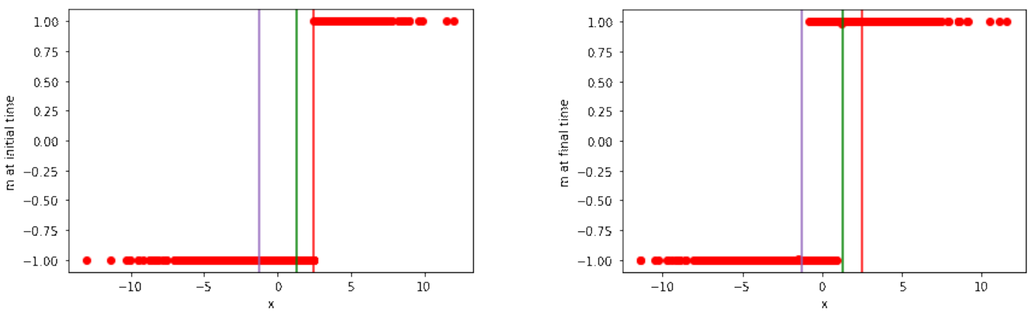
$$\lim_{t\to\infty} m(t)(x) = m^{\overline{x_0}}(x),$$

with $\overline{x_0} = 2\mu_0(-\infty, \overline{x_0})$ (resp. $\overline{x_0} = -2\mu_0(\overline{x_0}, +\infty)$).

Proposition 4.39 turns out to be very useful in describing the dynamics at order N and N^2 by infinitesimal (of order 1) variations of time. Indeed, as we shall see, the presence of a non-zero $X(t)$ can move the magnetization's profile to be outside the fixed points region. The above proposition thus quantifies how the dynamics gets attracted again towards the fixed points region, at least for times of order 1. Unfortunately we were not able to prove rigorously this result, but it is motivated heuristically by saying that the out-of-equilibrium dynamics approaches the nearest possible stable equilibrium. An illustration of this phenomenon is given in Figure 4.5.

4.3.6.2 Order N dynamics

In the timescale of order N the only additional dynamics which takes place is due to the fact that $\mu_0^t(dx)$ now depends on time (it is a normal distribution centered around 0, with variance depending on the macroscopic timescale t and proportional to $\frac{\sigma^2}{2\alpha_2}$), because of the dynamics of the Ornstein-Uhlenbeck diffusions. In this scale we thus expect to see the same staircase equilibrium previously reached, with some movement of the points close to x_0 (caused by the motion of the diffusions at order N) in between the region of fixed points described in Proposition 4.36. At the finite particle system's level indeed, the motion of the diffusions at order N should produce a coexistence of phases around x_0, with some magnetizations being $+1$ and others -1. An illustration of this is shown in Figure 4.6, the analogous of Figure 4.5 at order N. The bigger the diffusive coefficient σ (for a fixed α_2), the wider the range of the diffusions and the area with coexistence of phases are: in Figure 4.6 the coexistence area fills all the fixed points region. The deterministic limit



(a) Initial configuration at order N (b) Final configuration at order N

Figure 4.6: Simulation of the finite particle system's dynamics at a timescale of order N, for $N = 1000$, $\beta_1 = \beta_2 = \infty$, $\alpha_1 = \alpha_2 = 1$, and $\sigma = 5$. As above, we take $x_j(0) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}\right)$.

dynamics becomes

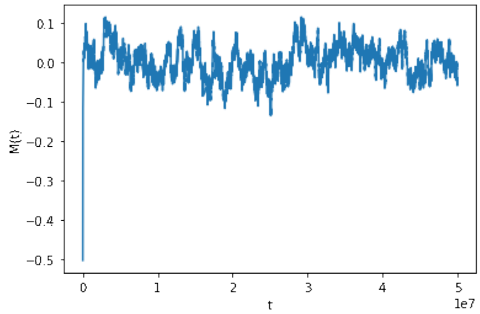
$$\begin{cases} m(t)(x) = \text{sign}(x + m(t)(x) + M(t)), \\ m(0)(x) = m^{x_0}(x), \\ M(t) = \int_{\mathbb{R}} m(t)(x)\mu_0^t(dx). \end{cases} \tag{4.142}$$

At this timescale the diffusion parameters play an important role. For $\frac{\sigma^2}{2\alpha_2}$ big, simulations suggest the presence of a very mild interaction among the magnetizations: see Figure 4.7, where we plot the path of the second level empirical magnetization relative to the same simulation of Figure 4.6. We see that, after starting from a rather polarized value, after a short time $M^N(t)$ becomes very small and from that time on it just wanders around 0, so that the single magnetization's processes are subject to a very low interaction among themselves, which could eventually tend to zero for $N \to +\infty$; in fact, the interaction among the diffusions is also tending to 0 as they propagate chaos independently of the magnetizations. In other words, the presence of a big $\sigma > 0$ (for a fixed $\alpha_2 > 0$) might render the particles asymptotically independent with $M \equiv 0$. Assuming this is the case, the limit process for each magnetization should be given by independent copies of a non-Markovian spin with jump times distributed as the hitting times of the Ornstein-Uhlenbeck. Moreover, from Figure 4.6 we see that the supposed jumps should occur precisely at the borders of the fixed points region (the purple and green lines). This regime appears then to be related to the mean field scenario of Section 4.2 for $\beta \to \infty$, highlighted in Remark 4.9, and to the single particle dynamics of the model of Section 3.2. We remark that the simulations of Figures 4.6 and 4.7 were realized by keeping fixed $X^N(t) \equiv 0$, thus ruling out the (small) fluctuations of $X^N(t)$ around 0 at a timescale of order N. Clearly, this should not change much the above picture.

When the parameter $\frac{\sigma^2}{2\alpha_2}$ is small, we instead witness the loss of stability of certain areas of the fixed points region. For a description of this we refer to the next section, where we describe the full dynamics at order N^2.

4.3.6.3 Order N^2 dynamics

In order to describe the order N^2 dynamics, we proceed as above by infinitesimal time steps, by looking at the conditional dynamics with respect to the values of the macroscopic limit diffusion $X(t)$. We fix an initial condition with $X(0) \neq 0$ and evolve the dynamics at times of order 1. The latter converges soon to some staircase equilibrium in the fixed points region and stays put for all times of order 1. At times of order N we then see some



(a) $(t, M^N(t))$

Figure 4.7: The path of the empirical 2-level magnetization for the same simulation of Figure 4.6.

diffusive behavior of the equilibrium around the fixed X, until the process $X(t)$ changes again. The reiteration of this procedure for the updated value of X is an infinitesimal time step in the order N^2 timescale. Starting with order 1 infinitesimal time steps, we are interested in the conditional dynamics

$$\begin{cases} \dot{m}(t)(x) = 2\text{sign}(x + m(t)(x) + M(t) + X) - 2m(t)(x), \\ m(0)(x) = m^{x_0}(x), \\ M(t) = \int_{\mathbb{R}} m(t)(x)\mu_X^\infty(dx), \end{cases} \tag{4.143}$$

for some fixed $X \in \mathbb{R}$ and some staircase initial condition m^{x_0} for m, which was reached at the previous timescale long-time limit. In (4.143), $\mu_X^\infty = \mathcal{N}\left(X, \frac{\sigma^2}{2\alpha_2}\right)$ is the asymptotic distribution of the (sped up) Ornstein-Uhlenbeck processes for a fixed value of X. Proposition 4.36 generalizes to

Proposition 4.40 (Shape of the equilibria)**.** *Every equilibrium of Equation* (4.143) *is a staircase function $m^{x_0}(x)$ of the form*

$$m^{x_0}(x) := \begin{cases} +1, & \forall x > x_0, \\ -1, & \forall x < x_0, \end{cases} \tag{4.144}$$

for some $x_0 \in \mathbb{R}$ satisfying

$$-2\mu_0^\infty(x_0 - X, +\infty) \le x_0 + X \le 2\mu_0^\infty(-\infty, x_0 - X), \tag{4.145}$$

Proof. The proof follows the same steps as in the proof of Proposition 4.36, observing that

$$\mu_X^\infty(-\infty, x_0) = \mu_0^\infty(-\infty, x_0 - X),$$

with $\mu_0^\infty = \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha_2}\right)$. □

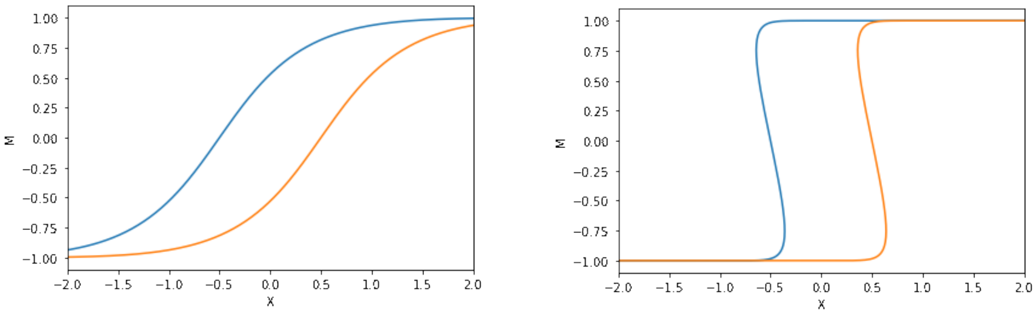
Remark 4.41. *Analogously to Remark 4.37, the fixed points region reduces to*

$$-2 \le x_0 + X \le 2,$$

when $\rho^2 := \frac{\sigma^2}{2\alpha_2} \to 0$, and to

$$-1 \le x_0 + X \le 1,$$

when $\rho^2 \to \infty$, and the fixed points interval is monotonically decreasing with ρ, since $\mu_0(-\infty, x_0 - X)$ is so.



(a) Fixed points region for large $\frac{\sigma^2}{2\alpha_2}$ (b) Fixed points region for small $\frac{\sigma^2}{2\alpha_2}$

Figure 4.8: Fixed points region for $\alpha_2 = 1$, $\sigma = 3$ (left), and for $\alpha_2 = 3$, $\sigma = 1$ (right).

Proposition 4.39 generalizes to

Proposition 4.42 (Stable attractors of the dynamics)**.** *Let $m(t)(x)$ be the solution to Equation* (4.143) *with initial datum $m_0(x) = m^{x_0}(x)$, with $x_0 + X > 0$ (resp. $x_0 + X < 0$) such that $x_0 + X > 2\mu_X^\infty(-\infty, x_0)$ (resp. $x_0 + X < -2\mu_X^\infty(x_0, +\infty)$). Then, we have*

$$\lim_{t\to\infty} m(t)(x) = m^{\overline{x_0}}(x),$$

with $\overline{x_0} = 2\mu_X^\infty(-\infty, \overline{x_0})$ (resp. $\overline{x_0} = -2\mu_X^\infty(\overline{x_0}, +\infty)$).

Remark 4.43. *It is useful to observe that the fixed points region's borders of Proposition 4.42 can be expressed in terms of (X, M), M being the asymptotically stable value of $M(t)$ in Equation* (4.143)*:*

$$\overline{x_0}(X, M) = 1 - X - M, \tag{4.146}$$

for the right border, i.e. for $\overline{x_0} + X > 0$, and

$$\overline{x_0}(X, M) = -1 - X - M, \tag{4.147}$$

for the left border, i.e. for $\overline{x_0} + X < 0$. Equations (4.146) *and* (4.147) *can be derived by using that*

$$M = 1 - 2\mu_X^\infty(-\infty, \overline{x_0}). \tag{4.148}$$

In Figure 4.8 we plot the fixed points region as a parametric function of X and M, with the two borders given by

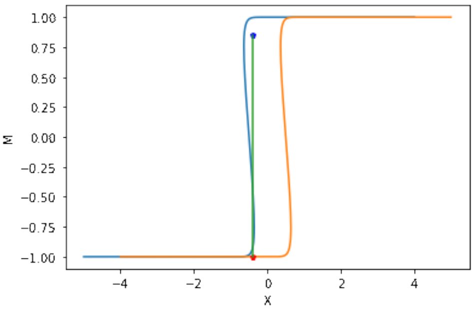
$$M = 1 - 2\mu_X^\infty(-\infty, 1 - X - M), \tag{4.149}$$

and

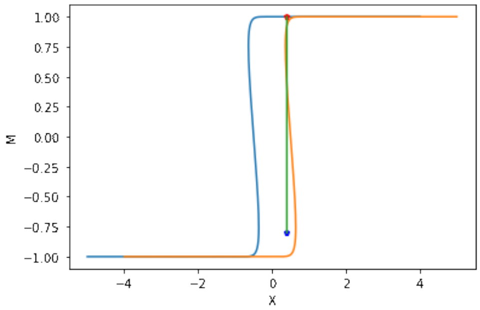
$$M = 1 - 2\mu_X^\infty(-\infty, -1 - X - M). \tag{4.150}$$

We can distinguish two regimes depending on the diffusion parameters: for large values of $\frac{\sigma^2}{2\alpha_2}$, Equations (4.149) and (4.150) define the graph of a function $M = \psi(X)$, while this is not the case when $\frac{\sigma^2}{2\alpha_2}$ is small. Unfortunately, we were not able to determine the precise value of σ and α_2 where this transition takes place, due to the implicit character of the equations in play.

Simulations suggest that the dynamics for the $M^N(t)$ is substantially different in the two cases: for big values of $\frac{\sigma^2}{2\alpha_2}$, we observe a diffusive motion onto the fixed points region,



(a) $X^N(t) \equiv X_0 = 0.4,\ M_0 = 0.802$

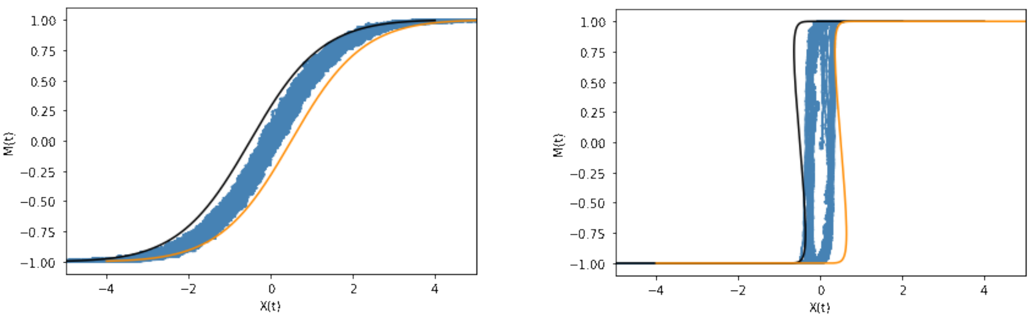


(b) $X^N(t) \equiv X_0 = -0.4,\ M_0 = -0.802$

Figure 4.9: Two simulations of the order N dynamics, with symmetric initial conditions in the two unstable regions, with $\alpha_2 = 3$, $\sigma = 1$, $\beta_1 = \beta_2 = \infty$. The blue (red) dot is the initial (final) point of each trajectory.

while for small $\frac{\sigma^2}{2\alpha_2}$ the dynamics resembles a diffusion with jumps. In both cases, as we noted for the simpler case $X^N(t) \equiv 0$, the single 1-level magnetizations $m_i^N(t)$'s should be evolving as non-Markovian spins, this time interacting since $X^N(t) \neq 0$.

At the second hierarchical level, the situation seems comparable to its mean field counterpart shown in Figure 4.3, with the diffusions' parameters playing the role of (the inverse of) β. As in the mean field case, the jumps seem to be occuring because of a loss of stability of the fixed points in certain areas of the phase-space. This loss of stability seems to originate at an order N timescale (this is also in parallel with the mean field case, where it was originating at an order 1 timescale). Indeed, for $\frac{\sigma^2}{2\alpha_2}$ small, starting the dynamics (4.143) from a staircase equilibrium which belongs to a certain area of the fixed points region, and letting it evolve for times of order N when the x_i's start their motion, we see a fast trajectory which very soon gets attracted to an area close to the opposite border from which it started. An example of this is shown in Figure 4.9, where we have kept fixed $X(t) \equiv X(0)$ to simulate the dynamics at a timescale of order N: we indeed see two fast transient trajectories starting from two initial points (the blue dots) which seem to belong to the unstable regions of fixed points. The same simulation at an order 1 timescale would have instead shown a trivial dynamics constantly equal to the initial datum, for any choice of the latter inside the two-dimensional manifold of fixed points. Finally, in Figure 4.10 we show a complete simulation of the trajectories $(X^N(t), M^N(t))$ at a timescale of order N^2, showing the different behavior depending on the value of $\frac{\sigma^2}{2\alpha_2}$. In the right plot, the white areas at the borders of the fixed points region which are not hit by any trajectory should approximate the unstable regions of fixed points.



(a) $\frac{\sigma^2}{2\alpha_2}$ large (b) $\frac{\sigma^2}{2\alpha_2}$ small

Figure 4.10: Two simulated trajectories of $(X^N(t), M^N(t))$ at the order N^2 timescale for $N = 500$ and $T = 5 \times 10^8$, with $\beta_1 = \beta_2 = \infty$. On the left $\alpha_2 = 1$, $\sigma = 5$, while on the right $\alpha_2 = 3$, $\sigma = 1$.

Appendix

APPENDIX A

Entropy solutions to scalar conservation laws

As we used some of their properties in Chapter 2, here we recall some general facts about entropy solutions to scalar conservation laws. A standard reference for what follows is [34]. Consider the Cauchy problem, for $x \in \mathbb{R}$, $t \in [0,T]$

$$\begin{cases} \partial_t u + \partial_x[f(x,u)] = 0, \\ u(0,x) = u_0(x). \end{cases} \tag{A.1}$$

The function f, called the flow, is not standard as it is space-dependent. We always assume that $f \in C^1(\mathbb{R}^2)$.

Definition A.1. *A function $u \in L^1_{loc}([0,T] \times \mathbb{R}) \cap C([0,T[, L^1_{loc}(\mathbb{R}))$ is called an* entropy solution *to* (A.1) *if*

$$\lim_{t \to 0^+} u(t) = u_0 \tag{A.2}$$

in $L^1_{loc}(\mathbb{R})$ and one of the following two equivalent conditions holds:

1. *for any entropy-entropy flux pair (η, q), that is, for any $\eta \in C^2(\mathbb{R})$ convex and $q = q(x,u)$ such that $\partial_u q(x,u) = \partial_u f(x,u)\eta'(u)$,*

$$\partial_t \eta(u) + \partial_x[q(x,u)] + \eta'(u) f_x(x,u) - q_x(x,u) \leq 0, \tag{A.3}$$

   in distribution, i.e for any $\varphi \in C^\infty_C(]0,T[\times\mathbb{R})$, $\varphi \geq 0$,

$$\int_0^T \int_{\mathbb{R}} \{\eta(u)\varphi_t + q(x,u)\varphi_x + [q_x(x,u) - \eta'(u) f_x(x,u)]\varphi\} \, dxdt \geq 0; \tag{A.4}$$

2. *for any $c \in \mathbb{R}$*

$$\partial_t |u-c| + \partial_x[\operatorname{sign}(u-c)(f(x,u) - f(x,c))] + \operatorname{sign}(u-c) f_x(x,c) \leq 0, \tag{A.5}$$

   in distribution, that is, for any $\varphi \in C^\infty_C(]0,T[\times\mathbb{R})$, $\varphi \geq 0$,

$$\int_0^T \int_{\mathbb{R}} \{|u-c|\varphi_t + \operatorname{sign}(u-c)(f(x,u) - f(x,c))\varphi_x - \operatorname{sign}(u-c) f_x(x,c)\varphi\} dxdt \geq 0. \tag{A.6}$$

Lemma A.2. *The two conditions in the above definition are equivalent and imply that u is a weak solution to* (A.1) *in the sense of distributions.*

The entropy condition can be specialized when u is a function piecewise smooth, as we already stated in Proposition 2.2:

Proposition A.3. *Let u be a function piecewice C^1 whose discontinuity points belong to the smooth curve $x = \gamma(t)$. Then u is an entropy solution to* (A.1) *if and only if*

1. *u solves* (A.1) *in the classical sense where it is smooth;*
2. *the initial condition $u(0, x) = u_0(x)$ holds in the classical sense;*
3. *denoting*
$$u_r(t) := \lim_{x\to\gamma(t)^+} u(t, x)$$
and
$$u_l(t) := \lim_{x\to\gamma(t)^-} u(t, x),$$
the right and left limits respectively, the Rankine-Hugoniot *condition holds: for all t*
$$\dot{\gamma}(t) = \frac{f(\gamma(t), u_r(t)) - f(\gamma(t), u_l(t))}{u_r(t) - u_l(t)}; \tag{RH}$$
4. *the* Lax stability condition *holds:*
$$\frac{f(\gamma(t), c) - f(\gamma(t), u_r(t))}{c - u_r(t)} < \dot{\gamma}(t) < \frac{f(\gamma(t), c) - f(\gamma(t), u_l(t))}{c - u_l(t)} \tag{L}$$
for any t and c strictly between u_l and u_r.

The Rankine-Hugoniot condition is equivalent to state that u is a weak solution to the scalar conservation law. The Lax condition can be reformulated saying that the graph of $f(\gamma(t), \cdot)$ stays above the chord joining u_r and u_l, if $u_r < u_l$, while the graph stays below the chord when $u_l < u_r$.

The main result about the theory of conservation laws is the following

Theorem A.4. *If $u_0 \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$ then there exists a unique entropy solution $u \in C([0, T]; L^1(\mathbb{R}) \cap L^\infty(\mathbb{R}))$ to* (A.1).

APPENDIX B

Propagation of chaos for the model of Section 3.2

Here we prove rigorously a propagation of chaos property for the N-particle interacting spin-valued renewal dynamics of Chapter 3, Section 3.2, to its mean-field limit, for any $\gamma \in \mathbb{N}$. Actually, we establish the proofs for $\gamma = 1$, where the rates enjoy globally Lipschitz properties, and then we generalize them to any $\gamma \in \mathbb{N}$ in Remark B.3. The generalization to non-Lipschitz rates is possible because of the a-priori bound on the variables y_i's which, by definition, are such that $0 \leq y_i \leq T$, where $T < \infty$ is the final time horizon of the dynamics. For the convenience of the reader, we write again the dynamics

$$\begin{cases} (\sigma_i(t), y_i(t)) \mapsto (-\sigma_i(t), 0), \quad \text{with rate} \quad y_i^{\gamma}(t) e^{-(\gamma+1)\beta\sigma_i(t) m^N(t)}, \\ dy_i(t) = dt, \quad \text{otherwise}, \end{cases} \tag{B.1}$$

and the mean-field version

$$\begin{cases} (\sigma(t), y(t)) \mapsto (-\sigma(t), 0), \quad \text{with rate} \quad y^{\gamma}(t) e^{-(\gamma+1)\beta\sigma(t) m(t)}, \\ dy(t) = dt, \quad \text{otherwise}, \end{cases} \tag{B.2}$$

with $m(t) = \mathbb{E}[\sigma(t)]$. The approach is analogous to the one used recurrently in the Dissertation: represent both the microscopic and the macroscopic model as solutions of certain stochastic differential equations driven by Poisson random measures, in order to apply the results in [66]. As anticipated, in the proof we restrict to a finite interval of time $[0, T]$.

To begin with, let us fix a filtered probability space $\big((\Omega, \mathcal{F}, \mathbb{P}), (\mathcal{F}_t)_{t\in[0,T]}\big)$ satisfying the usual hypotheses, rich enough to carry an inependent and identically distributed family $(\mathcal{N}_i)_{i\in\mathbb{N}}$ of stationary Poisson random measures $\mathcal{N}_i$ on $[0,T] \times \Xi$, with intensity measure ν on $\Xi := [0, +\infty)$ equal to the restriction of the Lebesgue measure onto $[0, +\infty)$. For any N, consider the system of Itô-Skorohod equations

$$\begin{cases} \sigma_i(t) = \sigma_i(0) + \int_0^t \int_\Xi f_1(\sigma_i(s^-), \xi, m^N(s^-), y_i(s^-)) \mathcal{N}_i(ds, d\xi), \\ y_i(t) = y_i(0) + t + \int_0^t \int_\Xi f_2(\sigma_i(s^-), \xi, m^N(s^-), y_i(s^-)) \mathcal{N}_i(ds, d\xi), \end{cases} \tag{B.3}$$

and the corresponding limit non-linear *reference particle*'s dynamics

$$\begin{cases} \sigma(t) = \sigma(0) + \int_0^t \int_\Xi f_1(\sigma(s^-), \xi, m(s^-), y(s^-)) \mathcal{N}(ds, d\xi), \\ y(t) = y(0) + t + \int_0^t \int_\Xi f_2(\sigma(s^-), \xi, m(s^-), y(s^-)) \mathcal{N}(ds, d\xi). \end{cases} \tag{B.4}$$

The functions $f_1, f_2 : \{-1,1\} \times \mathbb{R}^+ \times [-1,1] \times \mathbb{R}^+ \to \mathbb{R}$, modeling the jumps of the process, are given by

$$f_1(\sigma, \xi, m, y) := -2\sigma \mathbb{1}_{]0,\lambda[}(\xi), \tag{B.5}$$

and

$$f_2(\sigma, \xi, m, y) := -y \mathbb{1}_{]0,\lambda[}(\xi), \tag{B.6}$$

with $\lambda := \lambda(\sigma, m, y)$ being the rate function

$$\lambda(\sigma, m, y) = y^\gamma e^{-(\gamma+1)\beta\sigma m}.$$

Proposition B.1. *For $\gamma = 1$, Equations* (B.3) *and* (B.4) *possess a unique strong solution for $t \in [0, T]$.*

Proof. With the choices (B.5) and (B.6), the well-posedeness of Equations (B.3) and (B.4) follows by Theorems 1.2 and 2.1 in [66]. Indeed, even though the function f_2 is not globally Lipschitz continuous in y, the L^1 Lipschitz assumption of the theorem still holds, by noting that

$$\begin{aligned}
&\int_\Xi \Big| f_2(\sigma, \xi, m, y) - f_2(\tilde{\sigma}, \xi, \tilde{m}, \tilde{y}) \Big| d\xi \\
&\quad = \int_\Xi \Big| y \mathbb{1}_{]0,\lambda(\sigma,m,y)[}(\xi) - \tilde{y} \mathbb{1}_{]0,\lambda(\tilde{\sigma},\tilde{m},\tilde{y})[}(\xi) \Big| d\xi \\
&\quad \le |y| |\lambda(\sigma, m, y) - \lambda(\tilde{\sigma}, \tilde{m}, \tilde{y})| + |\lambda(\tilde{\sigma}, \tilde{m}, \tilde{y})| |y - \tilde{y}| \\
&\quad \le |y| \big[|\lambda(\sigma, m, y) - \lambda(\tilde{\sigma}, \tilde{m}, \tilde{y})| + |\lambda(\tilde{\sigma}, \tilde{m}, \tilde{y})| |y - \tilde{y}| \big] \\
&\quad \le CT \big[|m - \tilde{m}| + |y - \tilde{y}| + |\sigma - \tilde{\sigma}| \big],
\end{aligned}$$

where in the last step we have used that, by construction, the processes $y_i(t) \le T$ for every $t \in [0, T]$, so that the rates are a priori bounded and the Lipschitz properties of $ye^{-2\beta\sigma m}$ for $(y, \sigma, m) \in \mathbb{R}^+ \times \{-1, 1\} \times [-1, 1]$. □

Now, define the empirical measures

$$\mu^N := \frac{1}{N} \sum_{i=1}^{N} \delta_{(\sigma_i, y_i)},$$

and their evaluation along the paths of (B.3),

$$\mu_t^N := \frac{1}{N} \sum_{i=1}^{N} \delta_{(\sigma_i(t), y_i(t))}. \tag{B.7}$$

The measures $(\mu_t^N)_{t\in[0,T]}$ can be viewed as random variables with values in $\mathcal{P}(D)$, the space of probability measures on D, where $D := \mathcal{D}([0, T]; \{-1, 1\} \times \mathbb{R}^+)$ is the space of $\{-1, 1\} \times \mathbb{R}^+$-valued càdlàg functions equipped with the Skorohod topology.

Theorem B.2. *Fix $\gamma = 1$ and a final time $T > 0$ in* (B.3) *and* (B.4)*. Assume that the initial conditions $(\sigma_i(0), y_i(0)) = (\sigma(0), y(0))$ for dynamics* (B.3) *and* (B.4) *are μ_0-chaotic for some probability distribution μ_0 on $\{-1, 1\} \times \mathbb{R}^+$. Then, the sequence of empirical measures $(\mu_t^N)_{t\in[0,T]}$ converges in distribution (in the sense of weak convergence of probability measures) to the deterministic law $(\mu_t)_{t\in[0,T]}$ on the path space of the unique solution to Equation* (B.4) *with initial distribution μ_0.*

Proof. Consider the following system of i.i.d. processes $(\tilde{\sigma}_i(t), \tilde{y}_i(t))_{i=1,\dots,N}$, coupled with $(\sigma_i(t), y_i(t))_{i=1,\dots,N}$,

$$\begin{cases} \tilde{\sigma}_i(t) = \tilde{\sigma}_i(0) + \int_0^t \int_\Xi f_1(\tilde{\sigma}_i(s^-), \xi, m(s^-), \tilde{y}_i(s^-)) \mathcal{N}_i(ds, d\xi), \\ \tilde{y}_i(t) = \tilde{y}_i(0) + t + \int_0^t \int_\Xi f_2(\tilde{\sigma}_i(s^-), \xi, m(s^-), \tilde{y}_i(s^-)) \mathcal{N}_i(ds, d\xi), \end{cases} \tag{B.8}$$

with $m(t) = \mathbb{E}[\tilde{\sigma}_i(t)]$. Let $(\tilde{\mu}_t^N)_{t\in[0,T]}$ be the empirical measure associated to (B.8). Clearly, $(\tilde{\mu}_t^N)_{t\in[0,T]} \to (\mu_t)_{t\in[0,T]}$ in the weak convergence sense (by a functional LLN, see [66] for e.g.). We are thus left to show

$$\boldsymbol{d_1}\Big(\text{Law}((\mu_t^N)_{t\in[0,T]}), \text{Law}((\tilde{\mu}_t^N)_{t\in[0,T]})\Big) \to 0,$$

for $N \to +\infty$, with $\boldsymbol{d_1}$ being the 1-Wasserstein distance (which metrizes the weak convergence of probability measures) on $\mathcal{P}(\mathcal{P}(D))$. Since (recall (19))

$$\boldsymbol{d_1}\Big(\text{Law}((\mu_t^N)_{t\in[0,T]}), \text{Law}((\tilde{\mu}_t^N)_{t\in[0,T]})\Big) \le \frac{1}{N}\sum_{i=1}^N \mathbb{E}\big[d_{Sko}((\sigma_i, y_i), (\tilde{\sigma}_i, \tilde{y}_i))\big],$$

with d_{Sko} the Skorohod metric on D, it is enough to show that

$$\frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[\sup_{t\in[0,T]} \Big(|\sigma_i(t) - \tilde{\sigma}_i(t)| + |y_i(t) - \tilde{y}_i(t)|\Big)\right] \to 0, \tag{B.9}$$

for $N \to +\infty$. For the proof of (B.9), we estimate, using the estimates of Proposition B.1 for f_2,

$$\begin{aligned} \mathbb{E}\left[\sup_{s\in[0,t]} |y_i(s) - \tilde{y}_i(s)|\right] &\le \mathbb{E}\Big[|y_i(0) - \tilde{y}_i(0)|\Big] \\ &\quad + C\int_0^t \mathbb{E}\Big[|m^N(s) - m(s)| + |y_i(s) - \tilde{y}_i(s)| + |\sigma_i(s) - \tilde{\sigma}_i(s)|\Big] ds \\ &\le C\int_0^t \mathbb{E}\left[\sup_{r\in[0,s]} |m^N(r) - m(r)| + \sup_{r\in[0,s]} |y_i(r) - \tilde{y}_i(r)| + \sup_{r\in[0,s]} |\sigma_i(r) - \tilde{\sigma}_i(r)|\right] ds \\ &\quad + C(N), \end{aligned}$$

with $C(N) \to 0$ for $N \to +\infty$ because of the chaoticity assumption on the initial datum. Similarly for the σ_i's, using the the Lipschitz continuity of f_1, we obtain

$$\begin{aligned} \mathbb{E}\left[\sup_{s\in[0,t]} |\sigma_i(s) - \tilde{\sigma}_i(s)|\right] &\le \mathbb{E}\Big[|\sigma_i(0) - \tilde{\sigma}_i(0)|\Big] \\ &\quad + C\int_0^t \mathbb{E}\Big[|m^N(s) - m(s)| + |y_i(s) - \tilde{y}_i(s)| + |\sigma_i(s) - \tilde{\sigma}_i(s)|\Big] ds \\ &\le C\int_0^t \mathbb{E}\left[\sup_{r\in[0,s]} |m^N(r) - m(r)| + \sup_{r\in[0,s]} |y_i(r) - \tilde{y}_i(r)| + \sup_{r\in[0,s]} |\sigma_i(r) - \tilde{\sigma}_i(r)|\right] ds. \\ &\quad + C(N). \end{aligned}$$

Denoting $\tilde{m}^N(t) := \frac{1}{N}\sum_{i=1}^N \tilde{\sigma}_i(t)$, we find

$$\mathbb{E}\left[\sup_{s\in[0,t]} |m^N(s) - m(s)|\right]$$

$$\begin{aligned}
&\leq \mathbb{E}\Bigg[\sup_{s\in[0,t]}|m^N(s)-\tilde{m}^N(s)|\Bigg]+\mathbb{E}\Bigg[\sup_{s\in[0,t]}|\tilde{m}^N(s)-m(s)|\Bigg]\\
&=\frac{1}{N}\sum_{j=1}^{N}\mathbb{E}\Bigg[\sup_{s\in[0,t]}|\sigma_j(s)-\tilde{\sigma}_j(s)|\Bigg]+\mathbb{E}\Bigg[\sup_{s\in[0,t]}|\tilde{m}^N(s)-m(s)|\Bigg]\\
&=\mathbb{E}\Bigg[\sup_{s\in[0,t]}|\sigma_i(s)-\tilde{\sigma}_i(s)|\Bigg]+C(N),
\end{aligned}$$

with $C(N)\to 0$ for $N\to+\infty$ because of the chaoticity of the i.i.d. processes

$$(\tilde{\sigma}_i(t),\tilde{y}_i(t))_{i=1,\dots,N},$$

and where in the equalities we have used the exchangeability properties of the processes $(\sigma_i,\tilde{\sigma}_i)_{i=1,\dots,N}$. Recollecting the estimates, we have shown, for any $t\in[0,T]$,

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\Bigg\{\mathbb{E}\Bigg[\sup_{s\in[0,t]}|\sigma_i(s)-\tilde{\sigma}_i(s)|\Bigg]+\mathbb{E}\Bigg[\sup_{s\in[0,t]}|y_i(s)-\tilde{y}_i(s)|\Bigg]\Bigg\}\\
&\quad\leq C(N)+\int_0^t\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\Bigg[\sup_{r\in[0,s]}|\sigma_i(r)-\tilde{\sigma}_i(r)|+\sup_{r\in[0,s]}|y_i(r)-\tilde{y}_i(r)|\Bigg]ds,
\end{aligned}$$

which by the Gronwall's lemma applied to

$$\varphi(t):=\frac{1}{N}\sum_{i=1}^{N}\Bigg\{\mathbb{E}\Bigg[\sup_{s\in[0,t]}|\sigma_i(s)-\tilde{\sigma}_i(s)|\Bigg]+\mathbb{E}\Bigg[\sup_{s\in[0,t]}|y_i(s)-\tilde{y}_i(s)|\Bigg]\Bigg\},$$

implies (B.9), because $\varphi(T)$ is an upper bound for the left hand side of (B.9). □

Remark B.3. *Proposition B.1 and Theorem B.2 can be generalized to any $\gamma\in\mathbb{N}$. Indeed, the same Lipschitz L^1 estimates on the rates of Proposition B.1 (used also in Theorem B.2) hold by estimating*

$$\begin{aligned}
\big|\lambda(\sigma,m,y)-\lambda(\tilde{\sigma},\tilde{m},\tilde{y})\big| &= \big|y^\gamma e^{-(\gamma+1)\beta m\sigma}-\tilde{y}^\gamma e^{-(\gamma+1)\beta\tilde{m}\tilde{\sigma}}\big|\\
&\leq \big|y^\gamma e^{-(\gamma+1)\beta m\sigma}-\tilde{y}^\gamma e^{-(\gamma+1)\beta m\sigma}\big|+\big|\tilde{y}^\gamma e^{-(\gamma+1)\beta m\sigma}-\tilde{y}^\gamma e^{-(\gamma+1)\beta\tilde{m}\tilde{\sigma}}\big|\\
&\leq \big|e^{-(\gamma+1)\beta m\sigma}\big|\big|y^\gamma-\tilde{y}^\gamma\big|+\tilde{y}^\gamma\big|e^{-(\gamma+1)\beta m\sigma}-e^{-(\gamma+1)\beta\tilde{m}\tilde{\sigma}}\big|\\
&\leq C|y-\tilde{y}|\big|p(y,\tilde{y})\big|+\tilde{y}^\gamma\Big[C|m-\tilde{m}|+C|\sigma-\tilde{\sigma}|\Big]\\
&\leq C\Big[|y-\tilde{y}|+|m-\tilde{m}|+|\sigma-\tilde{\sigma}|\Big],
\end{aligned}$$

with $p(y,\tilde{y})$ a polynomial of degree $\gamma-1$. In the last step we have used the a priori bounds on $y\leq T$ to get $|p(y,\tilde{y})|\leq C(T)$ and the Lipschitz properties of $e^{-(\gamma+1)\beta m\sigma}$ for $(\sigma,m)\in\{-1,1\}\times[-1,1]$.

Bibliography

[1] L. Andreis, D. Tovazzi. Coexistence of stable limit cycles in a generalized Curie-Weiss model with dissipation, *J. Stat. Phys.*, 173(1), 163-181, 2018.

[2] S. R. Athreya, Jan M. Swart. Survival of contact processes on the hierarchical group, *Probab. Theory Relat. Fields*, 147:529–563, 2010.

[3] M. Bardi and M. Fischer. On non-uniqueness and uniqueness of solutions in finite-horizon Mean Field Games, *ESAIM: COCV*, pre-published on-line, `https://doi.org/10.1051/cocv/2018026`, 2018.

[4] R. Basna, A. Hilbert, and V. N. Kolokoltsov. An epsilon-Nash equilibrium for non-linear Markov games of mean-field-type on finite spaces, *Comm. Stoch. Anal.*, 8(4): 449–468, 2014.

[5] E. Bayraktar, A. Cohen. Analysis of a Finite State Many Player Game Using its Master Equation, *SIAM J. Control Optim.*, 56(5):3538-3568, 2018.

[6] E. Bayraktar, X. Zhang. On non-uniqueness in mean field games, preprint 2019, `arXiv:1908.06207`.

[7] C. Benazzoli, L. Campi, and L. Di Persio. ε-Nash equilibrium in stochastic differential games with mean-field interaction and controlled jumps, *Statistics & Probability Letters* 154, 108522, 2019.

[8] A. Bensoussan, J. Frehse, and P. Yam. Mean Field Games and Mean Field Type Control Theory, *Springer Briefs in Mathematics.* Springer, New York, 2013.

[9] A. Bensoussan, K. Sung, S. Yam, and S. Yung. Linear-quadratic mean field games, *J. Optim. Theory Appl.*, 169(2):496-529, 2016.

[10] L. Bertini, G. Giacomin, and K. Pakdaman. Dynamical aspects of mean field plane rotators and the Kuramoto model, *Journal of Statistical Physics*, 138(1-3):270-290, 2010.

[11] P.M. Bleher, P. Major. Critical phenomena and universal exponents in statistical physics. On Dyson's hierarchical model, *Ann. Probab.*, 15(2), 431–477, 1987.

[12] P. Bremaud. *Markov Chains. Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer-Verlag New York, 1998.

[13] A. Bressan, B. Piccoli. *Introduction to the Mathematical Theory of Control*, volume 2 of *AIMS Series on Applied Mathematics.* American Institute of Mathematical Sciences, Springfield (MO), 2007.

[14] P. Cardaliaguet. Notes on mean field games. *Technical report*, Université de Paris Dauphine, 2013.

[15] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions. *The master equation and the convergence problem in mean field games*, (AMS-201), Princeton University Press, 2019, `http://www.jstor.org/stable/j.ctvckq7qf`.

[16] P. Cardaliaguet, J. Graber, A. Porretta, and D. Tonon. Second order mean field games with degenerate diffusion and local coupling, *NoDEA*, 22(5): 1287–1317, 2015.

[17] P. Cardaliaguet, C. Rainer. An example of multiple mean field limits in ergodic differential games, `arXiv:1907.09785`, 2019.

[18] R. Carmona, F. Delarue. Probabilistic analysis of mean-field games, *SIAM J. Control Optim.*, 51(4):2705-2734, 2013.

[19] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I*, volumes 84 of *Probability Theory and Stochastic Modelling.* Springer, Cham, 2018.

[20] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications II*, volumes 83 of *Probability Theory and Stochastic Modelling.* Springer, Cham, 2018.

[21] R. Carmona, D. Lacker. A probabilistic weak formulation of mean field games and applications, *Ann. Appl. Prob.*, 25(3):1189-1231, 2015.

[22] R. Carmona, P. Wang. Finite State Mean Field Games with Major and Minor Players, preprint 2016, `arXiv:1610.05408`.

[23] A. Cecchin. Finite State N-player and Mean Field Games, PhD Thesis, Department of Mathematics, University of Padua, 2018.

[24] A. Cecchin, M. Fischer. Probabilistic approach to finite state mean field games, *Appl. Math. and Optim.*, `https://doi.org/10.1007/s00245-018-9488-7`, 2018.

[25] A. Cecchin, P. Dai Pra, M. Fischer, G. Pelino. On the convergence problem in mean field games: a two state model without uniqueness, *SIAM J. Control Optim.*, 57(4), 2443–2466, 2019.

[26] A. Cecchin and G. Pelino. Convergence, fluctuations and large deviations for finite state mean field games via the master equation, *Stochastic Processes Appl.*, `doi.org/10.1016/j.spa.2018.12.002`, December 2018.

[27] W.-C. Chen. Nonlinear dynamics and chaos in a fractional-order financial system, *Chaos, Solitons & Fractals*, 36(5):1305-1314, 2008.

[28] F. Collet, P. Dai Pra. The role of disorder in the dynamics of critical fluctuations of mean field models, *Electron. J. Probab.*, 17(26):1–40, 2012.

[29] F. Collet, P. Dai Pra, and M. Formentin. Collective periodicity in mean-field models of cooperative behavior, *Nonlinear Differential Equations and Applications NoDEA*, 22(5):1461-1482, 2015.

[30] F. Collet, M. Formentin. Effects of Local Fields in a Dissipative Curie-Weiss Model: Bautin Bifurcation and Large Self-sustained Oscillations, *J. Stat. Phys.* 176: 478, 2019.

[31] F. Collet, M. Formentin, and D. Tovazzi. Rhythmic behavior in a two-population mean-field Ising model, *Phys. Rev. E*, 94(4), 042139, 2016.

[32] F. Comets, Th. Eisele. Asymptotic Dynamics, Non-Critical and Critical Fluctuations for a Geometric Long-Range Interacting Model, *Commun. Math. Phys.*, 118: 531-567, 1988.

[33] T. Cox, A. Greven. On the long term behaviour of some finite particle systems, *Probab. Theory Relat. Fields*, 85, 195-237, 1990.

[34] C. M. Dafermos. *Hyperbolic Conservation Laws in Continuum Physics*, volume 325 of *Grundlehren der mathematischen Wissenschaften.* Springer, Berlin Heidelberg, 4th ed., 2016.

[35] P. Dai Pra, M. Fischer, and D. Regoli. A Curie-Weiss model with dissipation, *Journal of Statistical Physics*, 152(1):37-53, 2013.

[36] P. Dai Pra, G. Giacomin, and D. Regoli. Noise-induced periodicity: some stochastic models for complex biological systems, in *Mathematical Models and Methods for Planet Earth*, pages 25-35. Springer, 2014.

[37] P. Dai Pra, W. J. Runggaldier, E. Sartori, M. Tolotti. Large portfolio losses: a dynamic contagion model, *The annals of Applied Probability*, 19(1):347-394, 2009.

[38] P. Dai Pra, E Sartori, and M Tolotti. Climb on the Bandwagon: Consensus and periodicity in a lifetime utility model with strategic interactions. `arXiv:1804.07469 [math.OC]`, April 2018.

[39] D. A. Dawson, A. Greven. Hierarchical models of interacting diffusions: Multiple time scale phenomena, phase transition and pattern of cluster-formation, *Probab. Theory Relat. Fields*, 96:435M73, 1993.

[40] D. A. Dawson, A. Greven. Multiple time scale analysis of interacting diffusions, *Probab. Theory Relat. Fields*, 95, 467-508, 1993.

[41] F.J. Dyson. Existence of a phase transition in a one-dimensional Ising ferromagnet, *Commun. Math. Phys. 12*, 91–107, 1969.

[42] F. Delarue and R. Foguen Tchuendom. Selection of equilibria in a linear quadratic mean field game, `arXiv:1808.09137 [math.PR]`, August 2018.

[43] F. Delarue, D. Lacker, and K. Ramanan. From the master equation to mean field game limit theory: a central limit theorem, *Electronic Journal of Probability*, 24. 10.1214/19-EJP298, 2018.

[44] F. Delarue, D. Lacker, and K. Ramanan. From the master equation to mean field game limit theory: large deviations and concentration of measure, `arXiv:1804.08550 [math.PR]`, April 2018.

[45] A. Dembo, O. Zeitouni. *Large Deviations Thechniques and Applications*, volume 38 of *Applications of Mathematics.* Springer-Verlag, New York, 2nd edition, 1998.

[46] B. Derrida, L. De Seze, and C. Itzykson. Fractal Structure of Zeros in Hierarchical Models, *J. Stat. Phys.*, 33: 559, 1983.

[47] J. De Simoi, S. Marmi. Potts models on hierarchical lattices and Renormalization Group dynamics, *Journal of Physics A Mathematical and Theoretical*, 42. 10.1088/1751-8113/42/9/095001, 2007.

[48] J. Dianetti, G. Ferrari, M. Fischer, M. Nendel. Submodular Mean Field Games: Existence and Approximation of Solutions, preprint 2019, `arXiv:1907.10968`.

[49] J. Doncel, N. Gast, and B. Gaujal. Mean-field games with explicit interactions, Preprint hal-01277098, February 2016.

[50] S. Ditlevsen, E. Löcherbach. Multi-class oscillating systems of interacting neurons, *Stochastic Processes Appl.*, 127:1840-1869, 2017.

[51] M. H. Duong, G. A. Pavliotis. Mean field limits for non-Markovian interacting particles: convergence to equilibrium, GENERIC formalism, asymptotic limits and phase transitions, `arXiv:1805.04959`, 2018.

[52] P. Dupuis, R. Ellis. *A weak convergence approach to the theory of large deviations.* Wiley Series in Probability and Statistics. John Wiley & Sons, New York, 1997.

[53] P. Dupuis, R. Ellis, A. Weiss. Large deviations for Markov processes with discontinuous statistics, I: General upper bounds, *The Annals of Probability*, 19(3):1280–1297, 1991.

[54] P. Dupuis, K. Ramanan, W. Wu. Large deviation principle for finite-state mean field interacting particle systems, preprint 2016, `arXiv:1601.06219v1`.

[55] G. B. Ermentrout, D. H. Terman. *Mathematical Foundations of Neuroscience*, volume 35. Springer Science & Business Media, 2010.

[56] S. N. Ethier, T. G. Kurtz. *Markov processes, Characterization and Convergence*, Wiley, Hoboken (NJ), 1986.

[57] R. Ferreira, D. Gomes. On the convergence of finite state mean-field games through the Γ-convergence, *J. Math. Anal. Appl.*, 418(1):211-230, 2014.

[58] M. Fischer. On the connection between symmetric N-player games and mean field games, *Ann. Appl. Probab.*, 27(2):757–810, 2017.

[59] N. Fournier, A. Guillin. On the rate of convergence in Wasserstein distance of the empirical measure, *Probab. Theory Relat. Fields*, 162:707-738, 2015.

[60] G. Giacomin, C. Poquet. Noise, interaction, nonlinear dynamics and the origin of rhythmic behaviors, *Brazilian Journal of Probability and Statistics*, 29(2):460-493, 2015.

[61] D. A. Gomes, J. Mohr, and R.R. Souza. Continuous time finite state mean field games, *Appl. Math. Optim.*, 68(1):99–143, 2013.

[62] D. A. Gomes, J. Mohr, R.R. Souza. Discrete time, finite state space mean field games, *Journal de Mathématiques Pures et Appliquées*, 93(2):308-328, 2010.

[63] D. A. Gomes, J. Saúde. Monotone numerical methods for finite-state mean-field games, *Appl. Math. Optim.*, `https://doi.org/10.1007/s00245-018-9510-0`, 2018.

[64] D. A. Gomes, R.M. Velho, M.-T. Wolfram. Dual two-state mean field games, 53rd IEEE Conference on Decision and Control. Los Angeles (CA), December 15–17, 2014.

[65] D. A. Gomes, R.M. Velho, M.-T. Wolfram. Socio-economic applications of finite state mean field games, *Phil. Trans. R. Soc. A*, 372:20130405, 2014.

[66] C. Graham. McKean-Vlasov Itô-Skorokhod equations, and nonlinear diffusions with discrete jump sets, *Stochastic Processes Appl.*, 40(1):69-82, 1992.

[67] S. Méléard. Asymptotic Behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models, in *Probabilistic Models for Nonlinear Partial Differential Equations*, Springer-Verlag, 1996.

[68] O. Guéant. Existence and Uniqueness Result for Mean Field Games with Congestion Effect on Graphs, *Appl. Math. Optim.*, 72(2):291-303, 2015.

[69] T. Hara, T. Hattori, and H. Watanabe. Triviality of hierarchical Ising model in four dimensions, *Commun. Math. Phys.*, 220(1), 13–40, 2001.

[70] W.Th.F. den Hollander. Renormalization of interacting diffusions : a program and four examples, *Report Eurandom; Vol. 2005025.* Eindhoven: Eurandom, 2005.

[71] 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, *Commun. Inf. Syst.*, 6(3):221–252, 2006.

[72] N. Ikeda, S. Watanabe. Stochastic Differential Equations and Diffusion Processes, volume 24 of *North-Holland Mathematical Library*, North Holland/ Kodansha, Amsterdam, 2nd edition, 1989.

[73] M. Kac. Foundations of Kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, volume 3, 171–197, 1954.

[74] M. Kaufman, R. Griffiths. Exactly soluble Ising models on hierarchical lattices, *Physical Review B*, 24. 1-1981. 10.1103/PhysRevB.24.496, 1981.

[75] R. Z. Khasminskii. On the averaging principle for Itô stochastic differential equations, *Kybernetika (Prague)* 4, 260-279, 1968. (Russian)

[76] V.N. Kolokoltsov. *Nonlinear Markov processes and kinetic equations*, volume 182. Cambridge University Press, 2010.

[77] D. Lacker. A general characterization of the mean field limit for stochastic differential games, *Probab. Theory Related Fields*, 165(3):581–648, 2016.

[78] D. Lacker. On the convergence of closed-loop Nash equilibria to the mean field game limit, `arXiv:1808.02745 [math.PR]`, August 2018.

[79] J.-M. Lasry and P.-L. Lions. Mean field games, *Japan. J. Math.*, 2(1):229-260, 2007.

[80] C. Léonard. Large deviations for long range interacting particle systems with jumps, *Annales de l'IHP Probabilités et statistiques*, 31(2):289–323, 1995.

[81] P. Levy. Processus semi-markoviens, *Proc. Int. Cong. Math*, Amsterdam 1954, pp. 416–426.

[82] M. T. Liggett. *Interacting Particle Systems*, Springer-Verlag, New York, 1985.

[83] P.L. Lions. *Lectures at Collège de France*, `www.college-de-france.fr`

[84] R.N. Mantegna, H. Eugene Stanley. *Introduction to Econophysics*, Cambridge University Press, Cambridge, 1999.

[85] P. H. McKean. Propagation of chaos for a class of non-linear parabolic equations, *Lecture Series in Differential Equations, Catholic Univ. 7*: 41–57, 1967.

[86] P. H. McKean. A class of Markov processes associated with nonlinear parabolic equations. *Proc. Natl. Acad. Sci.*, 56 (6): 1907–1911, 1966.

[87] M. Nutz, J. San Martin, and X. Tan. Convergence to the mean field game limit: a case study, `arXiv:1806.00817 [math.OC]`, June 2018.

[88] T. L. Parsons, T. Rogers. Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: a constructive approach with examples from theoretical biology, *J. Phys. A: Math. Theor.* 50, 41, 2017.

[89] L. Perko. *Differential Equations and Dynamical Systems*, Texts in Applied Mathematics 7, Third Edition, 2001.

[90] S. T. Rachev, L. Rüschendorf. *Mass Transportation Problems*, Springer-Verlag, 1998.

[91] F. Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, Progress in Nonlinear Differential Equations and Their Applications, Vol. 87, Birkhäuser/Springer, NY, 2015.

[92] M. Scheutzow. Noise can create periodic behavior and stabilize nonlinear diffusions, *Stochastic Processes Appl.*, 20(2):323-331, 1985.

[93] A. Shwartz, A Weiss. *Large Deviations for Performance Analysis*, Chapman and Hall, 1995.

[94] A. Shwartz, A. Weiss. Large deviations with diminishing rates, *Mathematics of Operations Research*, 30(2):281–310, 2005.

[95] K. Suchecki, V.M. Eguíluz, M. San Miguel. Voter model dynamics in complex networks: role of dimensionality, disorder, and degree distribution. *Phys. Rev.* E72(8), 036132, 2005.

[96] A.-S. Sznitman. Topics in propagation of chaos, In P.-L. Hennequin, editor, Ecole d'Eté de Probabilités de Saint-Flour XIX - 1989, volume 1464 of *Lecture Notes in Math.*, pages 165–251, Berlin, Springer-Verlag, 1991.

[97] J. D. Touboul. The hipster effect: When anti-conformists all look the same, *Discret. Contin. Dyn. Syst. B*, 24, 8, 4379-4415, 2019.

[98] J. D. Touboul, G. Hermann, and O. Faugeras. Noise-induced behaviors in neural mean field dynamics, *SIAM Journal on Applied Dynamical Systems*, 11(1):49-81, 2012.

[99] P. Turchin, A. D. Taylor. Complex Dynamics in Ecological Time Series, *Ecology*, 73(1):289-305, 1992.

[100] W. Weidlich, G. Haag. *Concepts and Models of a Quantitative Sociology: The Dynamics of Interacting Populations*, volume 14. Springer Science & Business Media, 2012.

[101] W. Wu. Large Deviations and Quasipotential for Finite-state Mean Field Interacting Particle Systems, *PhD thesis, Brown University*, Providence, RI, 2014.