UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MATEMÁTICA

QUANTITATIVE LARGE POPULATION APPROXIMATIONS FOR STOCHASTIC MODELS WITH INTERACTION OR VARYING ENVIRONMENT

TESIS PARA OPTAR AL GRADO DE DOCTOR EN CIENCIAS DE LA INGENIERÍA MENCIÓN MODELACIÓN MATEMÁTICA
EN COTUTELA CON EL INSTITUT POLYTECHNIQUE DE PARIS

FELIPE ANDRÉS MUÑOZ HERNÁNDEZ

PROFESORES GUÍA:
JOAQUÍN FONTBONA TORRES
VINCENT BANSAYE

MIEMBROS DE LA COMISIÓN:
NICOLAS CHAMPAGNAT
ANTON WAKOLBINGER
DANIEL REMENIK ZISIS
THIERRY BODINEAU
ANSGAR JÜNGEL

Tesis financiada por CONICYT-PFCHA/Doctorado Nacional/2017-21171912
Este trabajo ha sido parcialmente financiado por CMM ANID PIA AFB170001,
CMM ANID BASAL ACE210010 y CMM ANID BASAL FB210005

SANTIAGO DE CHILE
2021

RESUMEN DE LA MEMORIA PARA OPTAR AL TÍTULO DE: Doctor en Ciencias de la Ingeniería, Mención Modelación Matemática
POR: Felipe Andrés Muñoz Hernández
FECHA: 26/11/2021
PROFESORES GUÍA: Joaquín Fontbona Torres, Vincent Bansaye

Aproximaciones cuantitativas de grandes poblaciones para modelos estocásticos con interacción o medio variable

Esta tesis se concentra en el estudio de modelos estocásticos de poblaciones compuestas de individuos interactuando entre ellos o con su medio.

En una primera parte consideramos sistemas de difusión cruzada para dos especies. Desarrollamos un enfoque de dualidad que permite obtener estimaciones cuantitativas de estabilidad. También introducimos un modelo estocástico basado en individuos sobre un espacio discreto. Los individuos siguen marchas aleatorias y son sensibles al número de individuos de la otra especie en el mismo sitio, con una dependencia lineal en sus tasas de movimiento. Establecimos la convergencia en ley del modelo estocástico hacia los sistemas de difusión cruzada cuando el número de individuos por sitio es más grande que el cuadrado del número de sitios, suponiendo condiciones iniciales pequeñas.

En una segunda parte obtenemos una tasa de convergencia explícita para ciertos sistemas de difusiones con interacción de tipo campo medio con ramificación binaria logística hacia las soluciones de sistemas de auto-difusión no local con crecimiento de masa logístico, que describen sus aproximaciones de grandes poblaciones. La demostración se apoya en un argumento de acoplamiento para difusiones con ramificación binaria basado en transporte óptimo, el cual nos permite aproximar la trayectoria de la población ramificante e interactuante por un sistema de partículas independientes con nacimientos espacio-temporales aleatorios y convenientemente distribuidos.

Finalmente, en una tercera parte, consideramos el árbol reducido asociado a procesos de nacimiento y muerte en medios variables que da la estructura genealógica de la población. Describimos geométricamente este objeto utilizando la construcción lookdown introducida por Kurtz y Rodrigues. Introduciendo un acoplamiento y una distancia adaptados, aproximamos la genealogía en grandes poblaciones.

RESUMEN DE LA MEMORIA PARA OPTAR AL TÍTULO DE: Doctor en Ciencias de la Ingeniería, Mención Modelación Matemática
POR: Felipe Andrés Muñoz Hernández
FECHA: 26/11/2021
PROFESORES GUÍA: Joaquín Fontbona Torres, Vincent Bansaye

Quantitative large population approximations for stochastic models with interaction or varying environment

This thesis focuses on the study of stochastic population models composed of individuals interacting between them or with the environment.

In a first part, we consider cross-diffusion systems for two species. We develop a duality approach which allows to obtain quantitative stability estimates. We also introduce a stochastic individual-based model on a discrete space. The individuals follow random walks and they are sensitives to the number of individuals of the other species on the same site, with a linear dependence in their rates of motion. We stablish the convergence in law of the stochastic model towards the cross-diffusion systems when the number of individuals per site is greater than the square of the number of sites, assuming small initial conditions.

In a second part, we obtain an explicit rate of convergence for some systems of mean-field interacting diffusions with logistic binary branching towards the solutions of non-local self-diffusion systems with logistic mass growth, that describe their large population approximations. The proof relies on a coupling argument for binary branching diffusions based on optimal transport, which allows us to approximate the trajectory of the interacting branching population by a system of independent particles with suitably distributed random space-time births.

Finally, in a third part, we consider the reduced tree associated with birth and death processes in varying environments that gives the genealogical structure of the population. We describe geometrically this object by using the lookdown construction introduced by Kurtz and Rodrigues. By introducing a suitable coupling and distance, we approximate the genealogy in the large population regime.

To my comrade Javiera ♡.

To my mother and father Susana and Hugo,
and to my brothers Marcelo and Cristian.

Acknowledgements

In these lasts years I have crossed paths with an enormous amount of people and in every interaction I had, I always learnt or kept something. Therefore, I would like first to apologise to everyone that helped me in one way or another and whose name is not mentioned in these lines.

I could not start if not by thanking my thesis directors, Joaquín Fontbona and Vincent Bansaye. Without them and their continuous support none of this would have been possible. Thank you for the trust that both of you placed in me, for all the time, patience and effort you put on this thesis, for introducing me to interesting topics, for giving me the freedom to wander in the vast world of mathematics, and also for pointing me in the right directions when this turned out to be overwhelming. I learned a lot from you and each discussion was an infinite source of motivation. I will always be amazed about how all of this started, as if consequences of randomness. For all of this and more, I am very grateful and will always be indebted.

I would also like to thank Anton Wakolbinger and Nicolas Champagnat who took the time to carefully read and review this manuscript and whose remarks helped to improve it. Thanks also to Amandine Véber, Thierry Bodineau, Ansgar Jüngel and Daniel Remenik for accepting to be part of the jury for this thesis, it is quite an honour. During this thesis I also had the pleasure and privilege of collaborating with Ayman Moussa, whom I cannot thank enough for all the time and effort invested in explaining ideas and techniques while remaining pedagogical.

Je souhaite remercier les membres du GdT PEIPS pour tous les exposés et moments de convivialité partagés, en particulier merci à Sylvie Méléard et Carl Graham, et aussi à la chaire MMB. Merci à l'équipe administrative du CMAP et tout particulièrement à Nassera Náar, Alexandra Noiret et Pierre Straebler pour leur gentillesse et disponibilité. Merci également à tous et toutes les doctorant.e.s du laboratoire qui ont fait de mon passage par le CMAP une expérience très agréable, en particulier Fedor Goncharov, Josué Tchouanti, Kevish Napal, Heythem Farhat, Benoît Dagallier, Vianney Debavelaere et Paulin Jacquot (dessolé pour le bruit). Je remercie Adrien Prodhomme pour les rires, les promenades et nos discussions autour des questions existentielles et les probabilités. Parcourir ces années de thèse ensemble m'a été d'une grande aide.

Quisiera agradecer también a los investigadores y a las investigadoras que estuvieron basades (y no) en Chile y que conocí durante estos años: Héctor Olivero, Luis Fredes, Nikolas Tapia, Amitai Linker, Avelio Sepúlveda, Roberto Cortez, Claire Delplancke, Elsa Cazelles, Felipe Campos, Mauricio Duarte, Manuel Cabezas y Santiago Saglietti, por todas las charlas, discusiones y conferencias en las que compartimos. De igual manera me gustaría agradecer al personal administrativo del DIM: Silvia Mariano, María Cecilia Cea, Gladys Cavallone, Luis Mella, Oscar Mori, Eterin Jaña y Karen Hernández. En particular me gustaría agradecer a Natacha Astromujoff por su apoyo y por siempre buscar una solución al problema que sea, y también a Georgina Paredes por su gentileza y constante preocupación. Gracias también a quienes fueron mis compañeres de oficina: Sebastián Tapia, Ariel Perez, Diana Narváez, María Eugenia Martínez, Claudia Soto y Alexis Fuentes. Gracias a Álvaro Bustos y a Emilio Molina por compartir el sufrimiento burocrático.

Gracias también a las personas que hicieron que la estadía en Francia sobrepasara todas las expectativas: José Tomás Méndez, Marcela Meza, Lorena Araos, Milena Rojas, Bérengère Même, Felipe Garrido, Sebastián Riffo, Carlos Alvear, Daniela Cortés, Alfonso Montes, Mariana Huepe, Amparo Montes, Pablo Ugalde, Vito Roppo, Grégoire, Karine, Zéna y por supuesto a la república popular de Igny.

Gracias también a la gente que siempre me ha acompañado independientemente de las circunstancias y con la que he compartido (y no) en este último tiempo: Felipe San Martín, Camilo Iturra, Angélica Toledo, Marcelo Salazar, Felipe Salas, Fernanda Lira, José Manuel Zolezzi, Alfredo Villalobos, Uriel Troncoso, Gian Luca Carniglia, Francisco Arana, Rodrigo Bravo, Benjamín Ruíz y Josu Armendariz.

Sería fácil creer que todo este trabajo es solo mío porque la portada tiene mi nombre, pero la verdad es que hay una coautora: Javiera. Sin tu apoyo nada hubiese sido posible y ciertamente no sería la persona que soy hoy. Gracias, gracias y más gracias, por acompañarme siempre y empujarme a hacer lo que me gusta, por tu infinita motivación y energía. Si todo esto resultó bien es gracias a ti.

No podría tampoco omitir el rol que mi familia ha jugado desde siempre, cuyo esfuerzo se vió reflejado en todas las oportunidades que pudieron entregarme y que me llevaron al privilegio y lujo de poder seguir este camino. Gracias a mi madre y padre, Susana y Hugo, y a mis hermanos, Marcelo y Cristian, por su apoyo incondicional.

Finalmente agradezco a todas las personas que participaron en la revuelta social iniciada en Chile en Octubre de 2019 y particularmente a la clase trabajadora.

Contents

CHAPTER 1

Introduction

Mathematical models can be understood as an abstracted, idealized and approximated representation of reality by means of mathematical concepts and language. They lie at the core of almost every basic discipline, such as physics, biology, chemistry and computer science, as they serve to represent ideas and formalize observations of systems. Because of this, they turn out to be indispensable scientific tools, as they help in the generation of insight, explanation and prediction, and thus giving a basis for theoretical and empirical understanding. Therefore, the importance of having precise mathematical models, while having in mind at the same time the trade-off between realism and tractability.

This thesis focuses in the study of mathematical models arising from biological motivations. Mathematical biology has been an area of wide interest in recent decades, in particular the last years have seen a very rich interplay between these two areas of science. In this context, mathematical models are used to investigate the principles that govern the structure, development and behavior of systems of living organisms. The modeling of complex biological processes became a fundamental tool for creating analytical and computational approaches to many different bio-inspired problems, coming from different branches such as population dynamics, cell biology, genetics and epidemiology, to name but a few. The main objective being to get biological insight as a result of mathematical analysis and thus helping to understand fundamental questions about nature.

In particular, the models considered fall under the scope of what is known as dynamics and evolution within the field of population models. These two branches correspond to the study of how certain quantities of interest in a population evolve in time and the evolutionary process behind the diversity of populations. Broadly speaking, the study of populations in ecology includes understanding, explaining, and predicting species behaviour. Why do species inhabit particular areas, and how are they prevented from establishing beyond their range limits? How do they grow? What are the genealogical relationships along their evolutionary history and how does different factors affect this information? In particular such range questions have become popular in the last decade or so in response to concerns about climate change, and in particular their mathematical modeling, as ecologist effectively rely on models.

Constructing precise population models is a problem that has attracted attention since long time ago. We have seen through decades numerous efforts aiming in this direction, starting with Robert Malthus proposing the unbounded exponential growth of the size of a population in the presence of abundant resources. Then, with the logistic equation in where the self-limiting effect that a population has over its growth and the availability of resources is considered. Along with these lines, Lotka [89] and Volterra [116] developed a model for a predator-prey or competing species dynamics as simultaneous nonlinear differential equations, representing one of the major advances in modern mathematical ecology. Since then, a whole variety of models taking into account interaction between species have emerged. For example we have predator-prey models with complex density dependent interactions, useful for the study of functional responses for example, or models based on partial differential equations for structured populations dynamics. All of these models are deterministic descriptions of population dynamics. Nevertheless, when considering ecological models it is also natural to think in some source of randomness, since many biological processes are stochastic. This led to models incorporating randomness. In the last years this has been one of the principal approaches of mathematical ecology.

The central aspect which is explored in this thesis is the microscopic origin of macroscopic behaviour and its explicit approximation. This comes motivated from the idea that the huge number of interactions in real ecosystems difficult any attempt to create a precise model. It becomes natural then to approximate by a macroscopic behaviour, in order to obtain more tractable models. In this sense, the precise quantification of these approximations is of importance since then we can measure the error in which we incur by doing this simplification. Also, this approach yields simplifications even when we consider microscopic systems evolving stochastically, since it is often claimed heuristically that stochasticity at individual level can be ignored in the study of large groups.

Another notion that lies at the core of this work, is the role that interactions play in the dynamics of a population. In a very broad sense interaction can be understood as a dependence between two objects, where this dependence can be sideways or in only one sense. Given a population we have for example interaction between the genetic material of the individuals, interaction of an species with its environment through their reproductive dynamics or interaction between species at the level of their displacements. In this thesis some of these mechanisms are explored.

In the first part, we study spatial interaction within a population model composed of two differents species. The kind of interaction considered is known as local interaction, which means that individuals affect each other when they are present in the same spatial point. Starting from a stochastic microscopic model having local interaction, we show that in the regime of large-population approximation there is a deterministic macroscopic evolution belonging to the class of cross-diffusion systems. Moreover, we quantify this convergence obtaining an optimal scale.

In the second part, we study another type of spatial interaction known as non-local interaction. This means that individuals are allowed to interact with a whole region around them by averaging the effect of the individuals within. We consider a previously introduced stochastic microscopic model that approximates a deterministic behaviour showing this type of interaction in its large-population approximation, and we focus on the case in where there is only one species that interacts in this way with itself. For this model we prove quantitative estimates for the approximation of the microscopic system by the macroscopic dynamics.

Lastly, in the third part, we are interested in a different type of model. At this point we focus our attention on models regarding the evolution of the size of a population, while considering effects of the environment through the demographic rates. In particular, we study the genealogical information associated to such models through a previously introduced particle representation, which allows to approximate the genealogy when we have a very large number of individuals by the genealogy behind a macroscopic dynamics.

In what follows we introduce further the main ideas and problems handled in this thesis and we informally explain the results addressing each chapter in order.

1.1 Cross-diffusion models

In this section we will focus on understanding how a model showing segregation effects arises from dynamics in which we consider local interaction between individuals, and the questions that naturally come along, such as the approximation of these models.

1.1.1 Local interaction

Suppose that we have a population composed of only one species. In order to understand how a model for the evolution of its spatial distribution arises, we can start by considering the following equation

$$\partial_t u - \Delta(au) = 0, \tag{1.1.1}$$

where u is the unknown and a is a function given beforehand. From a probabilistic point of view, such equation can be obtained as the macroscopic behaviour of a stochastic microscopic particle system. Indeed, consider for example a collection of particles or individuals placed in a finite number of sites with a periodic boundary. Suppose that each individual performs a centered random walk and such that at each site x, it waits an amount of time exponentially distributed with parameter $a(x)$ before giving the next step. Given this dynamics, it can be shown that when the number of sites and individuals go to infinity (with this last parameter growing faster than the former), the properly rescaled empirical measure associated with this system approximates the solution of (1.1.1). This allows to understand u as a density or concentration per site of individuals (see for example [76] for the formalism of this approach).

Another interpretation for (1.1.1) comes from an heuristic argument. Suppose that u represents the concentration of particles in some medium. By expanding the laplacian, we have two different contributions to the temporal variation of u

$$-\Delta(au) = -\nabla \cdot (a\nabla u) - \nabla \cdot (u\nabla a). \tag{1.1.2}$$

First, we have a diffusion term $-\nabla \cdot (a\nabla u)$, which Fick's laws of diffusion describe as the contribution to the change with respect to time of the concentration due to the movement of particles from regions of high concentration to lower concentration ones. This, with a diffusivity a determined by the medium, which in turn is related to the speed at which this occurs. Secondly, we have a transport term with direction $-\nabla a$, describing the transport of particles towards regions in where a is smaller.

Given these ways of reasoning, from a population dynamics point of view, (1.1.1) can be thought of as a model for a species u evolving along with a species a of fixed concentration over a common environment. Individuals of the species u will tend to leave sites faster as more individuals of a are present at the same site. Also, they will avoid regions in where the concentration of species a is high. Since the changes in concentration for the species u at a given site depend on the presence of a at that same site, we say that the species interact locally, with this interaction being only in one direction for the moment, since a is fixed.

Consider now that we have a population composed by two different species u and v. Assuming that the species prefer the same environment to live and that they are under the influence of intra and inter-specific pressures, using a similar heuristic as the one described in the previous paragraphs, Shigesada, Kawasaki and Teramoto introduced in [108] the following system of equations for modeling the temporal evolution of the spatial distribution of the population

$$\begin{aligned} \partial_t u - \Delta\Big[(d_1 + a_{11}u + a_{12}v)u\Big] &= (r_1 - s_{11}u(t) - s_{12}v(t))u(t), \\ \partial_t v - \Delta\Big[(d_2 + a_{21}u + a_{22}v)v\Big] &= (r_2 - s_{21}u(t) - s_{22}v(t))v(t), \end{aligned} \tag{SKT}$$

where d_1 and d_2 are the intrinsic diffusion coefficients, a_{11} and a_{22} the self-diffusion coefficients and a_{12} and a_{21} the cross-diffusion coefficients, all of which are non-negative constants. On the right hand side we have a Lotka-Volterra-type of evolution governing the demographics of the model, where r_1 and r_2 represent the intrinsic growth rate of each species and s_{ij} the competition for resources for $i, j = 1, 2$.

This system of strongly coupled parabolic equations is one of the best known models in the class of cross-diffusion systems, namely those presenting nonlinearities of the form $-\nabla \cdot (A(u)\nabla u)$ for vector-valued u and a diffusion matrix A. From an ecological point of view, the system (SKT) has an interesting property: under certain conditions on the parameters, there exists non constant steady states which represents the phenomena of pattern formation or segregation. Furthermore, in this model intuition correspond to theory, in the sense that large intrinsic diffusion coefficients prevent pattern formation, while having large cross-diffusion coefficients helps in the formation of patterns. In fact, the main motivation in [108], was to propose a model showing non constant steady states, since for example if we consider a diffusive Lotka-Volterra system (that is, $a_{ij} = 0$ for $i, j = 1, 2$), then one can show that all of its steady states are constant, which is not so convincing from a modeling point of view.

Several efforts have been made to understand and answer the natural questions arising from the system (SKT), ranging from the basic questions regarding the existence and uniqueness of solutions, to the analysis of steady states and the different regimes arising from different sets of parameters. It is so rich in structure that it remains almost on its own as an active area of research.

Analytical development of cross-diffusion systems

To summarize the development of the theory behind the system of two species, we start by mentioning the first global existence result in the one dimensional setting, neglecting self-diffusion and assuming that all the remaining diffusion coefficients are equal, obtained in [74]. Significant progress was made then by Amann in [4, 5], obtaining general results concerning parabolic quasi-linear systems, which in particular allow to conclude that local regular solutions exists for (SKT). Next, there are results with restrictive structural assumptions, such as the triangular case (that is, $a_{11} = a_{21} = a_{22} = 0$ and $a_{12} \neq 0$ in (SKT)) or assuming smallness of the coefficients (see for example [91] and [43]). The first global existence results without structural restrictions were obtained first in the one dimensional case in [63], and then generalized to arbitrary dimension in [29] and [30]. Regarding the analysis of steady states of (SKT), one of the firsts studies was carried out in [96], showing spatial segregation for a set of parameters different than the one originally analyzed in [108]. An important work was made in [90], by describing the interplay between diffusion and cross-diffusion in pattern formation. For more on this last subject we refer to [21] and references therein.

Concerning the generalization to the case with more than two species, we have fewer results since, as expected, the system is more complicated to analyze. The first time that the system (SKT) was derived and formulated for n different species was done in [117]. Then, the global existence result for this n species system was obtained in [31]. In general, the formulation for n different species in the case in where there are no reaction terms can be written as

$$\partial_t u_i - \Delta\left(d_i u_i + \sum_{j=1}^{n} a_{ij} u_i u_j \right) = 0, \tag{1.1.3}$$

for $u = (u_1, \dots, u_n)$.

Much of the theory used for showing the existence of weak solutions to (1.1.3), is based in the boundedness-by-entropy method [71]. The key idea of this approach is to find a priori estimates through a Lyapunov functional of the form

$$H(u) = \int h(u)\,\mathrm{d}x,$$

for a very suitable choice of h. Indeed, and as we would expect from this Lyapunov functional approach, under some conditions on the function h, it can be shown that this functional decreases along the trajectories of (1.1.3), that is, $\frac{\mathrm{d}}{\mathrm{d}t} H(u) \leq 0$. Moreover, this method provides more information than just the monotonicity of this entropy functional. In fact, the existence of an entropy structure is intimately tied to the existence of a transformation that yields the system (1.1.3) symmetric (specifically its diffusion matrix). This allows the use of more standard techniques from the analysis of parabolic equations. As shown in [31], by setting

$$H(u) = \int \sum_{i=1}^{n} \pi_i \Big(u_i(x) \log(u_i(x)) - u_i(x) + 1 \Big)\,\mathrm{d}x, \tag{1.1.4}$$

for coefficients $\pi_i > 0$ satisfying the detailed balance condition

$$\pi_i a_{ij} = \pi_j a_{ji}, \quad \text{for } i, j = 1, \dots, n, \tag{1.1.5}$$

gives the monotonicity of the functional and also an a priori control over the gradient of both $\sqrt{u_i}$ and u_i, being this the key for obtaining the global existence result through an approximation procedure. Nevertheless, it is conjectured in [31] that the entropy for this system is bounded for all times, all non-negative coefficients and all non-negative initial conditions, as well as for all coefficients $a_{ij} > 0$.

Derivations from other systems

A central question that has gathered attention through the last decade, is the derivation of models in the family of cross-diffusion systems by means of properly scaled models. This, in order to justify in some sense that these systems of equations arise as natural limiting objects for reasonable microscopic dynamics, since their conception in [108] was based purely in heuristics.

The first time that the system (SKT) was derived following this objective was done in [69]. There, the authors formally showed that the solution of the triangular system can be approximated by means of properly scaled reaction-diffusion systems. This approach was later formalized in [42].

A different formal derivation of the system can be obtained by following for example [99] and [117]. There, a procedure for recovering a model in the class of (SKT) from a random walk inspired lattice model was proposed, which we describe next. Let $(x_j)_{j\in\mathbb{Z}}$ be the lattice under consideration, where $h = x_j - x_{j-1} > 0$, and $u_i(x_j)$ be the proportion of the i-th species on x_j. Suppose that the species jump from site j towards $j \pm 1$ with rate $r_i^{j,\pm}$, that the particles from $j-1$ jump to j at rate $r_i^{j-1,+}$, and that the particles from $j+1$ jump to j at rate $r_i^{j+1,-}$. This dynamics yields, at a formal level, the following master equation for the evolution in time of the proportion u_i at the site x_j

$$\partial_t u_i(x_j) = r_i^{j-1,+} u_i(x_{j-1}) + r_i^{j+1,-} u_i(x_{j+1}) - (r_i^{j,+} + r_i^{j,-}) u_i(x_j),$$

for $i = 1, \dots, n$ and $j \in \mathbb{Z}$. The transition rates are defined by

$$r_i^{j,\pm} = \sigma p_i(u(x_j)) q_i(u_{n+1}(x_{j\pm1})),$$

where $u(x_j) = (u_1(x_j), \dots, u_n(x_j))$, and also supposing that $u_{n+1}(x_j) = 1 - \sum_{k=1}^{n} u_k(x_j)$. This assumption on the structure of the rates aims to model the following effect: if a site is crowded or the neighbors are less occupied, then the species will tend to leave the site. Specifically, p_i measures the tendency of the species i to leave the j-th site and q_i measures the movement from the neighboring sites into the site j. In particular, we understand this kind of model as if $u_i(x_j)$ represents a volume fraction of occupancy and u_{n+1} the volume fraction not occupied by the species, yielding the effect known as volume-filling. By considering that $q_i(u_{n+1}) = 1$ for all $i = 1, \dots, n$ (no volume-filling effect) and taking $n = 3$ and $p_i(u) = d_i + a_{i1}u_1 + a_{i2}u_2$, it was shown (formally) that one recovers the system (SKT) without reaction terms when $h \to 0$. This procedure already shines a light for developing a microscopic approximation.

Another well-known approach to approximate a nonlinear partial differential equation is by means of a system of stochastic differential equations describing a many particle system. A derivation of (SKT) in this spirit was obtained partially in [28] and then extended in [27] recovering the full model. There, the authors approximated the limiting system by performing a two-step limit procedure. Starting from an interacting particle system, in where interactions occur in a non-local way, they approached first an intermediary non-local cross-diffusion system, when the number of particles goes to infinity, and then they studied the limit towards local interaction, all of this seen through a single particle and the equation satisfied by its law. We will revisit this terminology and approach in the next section.

Making ends meet

Following the approach of obtaining the system (SKT) as the limit of suitable scaled microscopic models and motivated by understanding the origin of the entropy (1.1.4) and the detailed balance condition (1.1.5), the authors of [37] introduced therein a stochastic microscopic model and showed a link between its entropy structure (in the classical sense of entropy for Markov chains) and (1.1.4) by means of a two-step limit procedure. This allowed them to exhibit the detailed balance condition (1.1.5) as the detailed balance condition for finite-state Markov chains. More precisely, given $M \in \mathbb{N}^*$, the model introduced is a stochastic particle system in where each particle evolves on the state space

$$\Omega_M = \{x_k : k = 0, \dots, M-1\},$$

where $x_k = k/M$ and with periodic boundary. Taking into account the relative fractions $\pi_1, \dots, \pi_n$, with $\pi_i > 0$, of n species, the system is defined in a way such that there are $\lfloor \pi_i N \rfloor$ individuals of the i-th species, for $i = 1, \dots, n$. Thus, a state of this process is given by

$$\boldsymbol{x} = \left(x_1^1, \dots, x_1^{\lfloor \pi_1 N \rfloor}, \dots, x_n^1, \dots, x_n^{\lfloor \pi_n N \rfloor}\right) \in \Omega_M^{\otimes(\lfloor \pi_1 N \rfloor + \cdots + \lfloor \pi_n N \rfloor)}.$$

Also, the system is endowed with a dynamics such that the configuration of particles evolves as a continuous-time Markov chain that we describe next. Assuming that the particles are indistinguishable, consider the following four transitions

$$\left.\begin{array}{l} \boldsymbol{x} \mapsto \boldsymbol{x} + \mathbf{e}_i^a + \mathbf{e}_j^b \\ \boldsymbol{x} \mapsto \boldsymbol{x} - \mathbf{e}_i^a - \mathbf{e}_j^b \end{array}\right\} \text{ at rate } \delta_{(i,a)\neq(j,b)} \delta_{x_i^a = x_j^b} \frac{d_{ij}}{N}, \quad \text{and} \quad \left.\begin{array}{l} \boldsymbol{x} \mapsto \boldsymbol{x} + \mathbf{e}_i^a \\ \boldsymbol{x} \mapsto \boldsymbol{x} - \mathbf{e}_i^a \end{array}\right\} \text{ at rate } d_i,$$

for $i, j = 1, \dots, n$, and where $\mathbf{e}_i^a \in \Omega_M^{\otimes(\lfloor \pi_1 N \rfloor + \cdots + \lfloor \pi_n N \rfloor)}$ denotes the vector whose entries are zero everywhere except for the a-th particle of species i, with $a = 1, \dots, \lfloor \pi_i N \rfloor$, in where its value is $1/M$, and similarly for $\mathbf{e}_j^b$. Only one of these transitions occurs at a jump time. The process defined in this way turns out to be a reversible Markov process and having the uniform distribution as invariant measure, meaning that each state has probability $M^{-(\lfloor \pi_1 N \rfloor + \cdots + \lfloor \pi_n N \rfloor)}$. Furthermore, by denoting $\mathbb{P}_t^N$ the time marginal of the process for given $t > 0$, which is a measure over the discrete set $\Omega_M^{\otimes(\lfloor \pi_1 N \rfloor + \cdots + \lfloor \pi_n N \rfloor)}$, and defining the entropy functional

$$\widetilde{H}\left(\mathbb{P}_t^N\right) = \sum_{\boldsymbol{x} \in \Omega_M^{\otimes(\lfloor \pi_1 N \rfloor + \cdots + \lfloor \pi_n N \rfloor)}} \mathbb{P}_t^N(\boldsymbol{x}) \log\left(\frac{\mathbb{P}_t^N(\boldsymbol{x})}{M^{(\lfloor \pi_1 N \rfloor + \cdots + \lfloor \pi_n N \rfloor)}}\right), \tag{1.1.6}$$

it can be shown that this quantity decreases with respect to time, i.e. $\frac{\mathrm{d}}{\mathrm{d}t} \widetilde{H}(\mathbb{P}_t^N) \leq 0$. This last fact can be seen also as a consequence of a much general statement proved in [93], which ensures that the law of such continuous-time Markov chain evolves as the gradient flow of the entropy.

Under the assumption that as N goes to infinity particles become independent, stated as

$$\mathbb{P}^N(x_1^1, \dots, x_1^{\lfloor \pi_1 N \rfloor}, \dots, x_n^1, \dots, x_n^{\lfloor \pi_n N \rfloor}) \approx u_1(x_1^1) \cdots u_1(x_1^{\lfloor \pi_1 N \rfloor}) \cdots u_n(x_n^1) \cdots u_n(x_n^{\lfloor \pi_n N \rfloor}),$$

it was shown formally in [37] that the marginals u_i evolve as

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} u_i(x) &= d_i(u_i(x+h) + u_i(x-h) - 2u_i(x)) \\ &\quad + \sum_{j=1}^n d_{ij} \pi_j (u_i(x+h) u_j(x+h) + u_i(x-h) u_j(x-h) - 2u_i(x) u_j(x)), \end{aligned} \tag{1.1.7}$$

for $i = 1, \dots, n$, and that $\frac{1}{N} \widetilde{H}(\mathbb{P}_t^N)$ converges towards $\sum_{i=1}^n \pi_i \sum_{j=0}^{M-1} u_i(x_j) \log(t, u_i(x_j)/M)$ (which also decreases), making the connection of the entropy structure of both objects, the process and the semi-discrete system (1.1.7), which eventually leads to (1.1.4) when the discretization step M^{-1} goes to zero. The independence assumption is based in that as the interaction between two particles is scaled by N^{-1}, this implies that the correlation between them should also be scaled by the same factor, thus becoming independent in the limit $N \to \infty$. This assumption is known as propagation of chaos, notion that will be revisited later on.

1.1.2 Main question and answer

We have seen that the derivations of models in the family of cross-diffusion systems, from models showing local interaction, have been done at a purely formal level. This takes us to consider the natural question concerning the possibility of defining a random individual-based model showing local interaction and having a dynamics reflecting plausible interactions, such that in the macroscopic limit it behaves as a cross-diffusion system.

Motivated by the previous question, we introduced a random microscopic model composed of particles that perform repulsive random walks, and such that it has a cross-diffusion system without self-diffusion nor reaction terms as its scaling limit. We moreover quantify this approximation through the application of analytic techniques that are not so common in the probabilistic setting, obtaining polynomial rates. These same techniques also allow to obtain a stability result for bounded solutions of the cross-diffusion system considered. The approach employed does not involve the use of an entropy structure and the approximation of the microscopic model by the macroscopic behaviour is done in a one-step limit.

1.1.3 Our contribution

We focused on the study of the following system on the one dimensional torus $\mathbb{T}$

$$\begin{aligned} &\partial_t u - \Delta\Big(d_1 u + a_{12} uv\Big) = 0, \\ &\partial_t v - \Delta\Big(d_2 v + a_{21} uv\Big) = 0, \end{aligned} \tag{1.1.8}$$

where $d_i > 0$ and $a_{ij} > 0$ for $i, j = 1, 2$. This system does not show self-diffusion nor reaction terms, with the idea of dealing only with the main difficulty arising from the cross-diffusion terms.

Given this system, we introduced a discrete microscopic model representing a population composed by two species spatially distributed among $M \in \mathbb{N}^*$ sites. Specifically, the approximation is done through the convergence of the number of individuals of each species at each site renormalized by a factor N, represented by the process $(\boldsymbol{U}^{M,N}(t), \boldsymbol{V}^{M,N}(t))_{t\geq 0}$ taking values in $(\mathbb{N}/N)^M \times (\mathbb{N}/N)^M$. This approximation takes place in the regime in where the number of individuals and the number of sites is very large.

The transitions of this process are as follows. For any vector of configurations $(\boldsymbol{U}, \boldsymbol{V}) \in (\mathbb{N}/N)^M \times (\mathbb{N}/N)^M$, where $\boldsymbol{U} = (U_i)_{i=1}^M$ and $\boldsymbol{V} = (V_i)_{i=1}^M$, we have that

$$\begin{aligned} &\boldsymbol{U} \mapsto \boldsymbol{U} + (\mathbf{e}_{i+\theta} - \mathbf{e}_i)N^{-1} \quad \text{at rate} \quad 2M^2 N U_i(d_1 + a_{12} V_i), \\ &\boldsymbol{V} \mapsto \boldsymbol{V} + (\mathbf{e}_{i+\theta} - \boldsymbol{e}_i)N^{-1} \quad \text{at rate} \quad 2M^2 N V_i(d_2 + a_{21} U_i), \end{aligned}$$

where $(\boldsymbol{e})_{1\leq j\leq M}$ is the canonical vector of $\mathbb{R}^M$, $\boldsymbol{e}_0 = \boldsymbol{e}_M$, $\boldsymbol{e}_{M+1} = \boldsymbol{e}_1$ and $\theta \in \{-1, 1\}$. This transitions reflect two behaviours. First, that the particles perform independent random walks.

Secondly, that they are also under a localized effect as they tend to leave faster a given site if the presence of particles of the other species is higher in that same site.

Through an analysis of the infinitesimal generator of the process previously defined, we formally deduced that when the number of particles is big enough, leaving the number of sites fixed, the system behaves like the solution $(\boldsymbol{u}^M(t), \boldsymbol{v}^M(t))_{t\geq 0}$ of the semi-discrete system of ordinary differential equations

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{u}^M - \Delta_M\big(d_1\boldsymbol{u}^M + a_{12}\boldsymbol{u}^M \odot \boldsymbol{v}^M\big) = 0, \\ \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{v}^M - \Delta_M\Big(d_2\boldsymbol{v}^M + a_{21}\boldsymbol{u}^M \odot \boldsymbol{v}^M\Big) = 0, \end{aligned} \tag{1.1.9}$$

where Δ_M is the periodic laplacian matrix defined by

$$\Delta_M := M^2 \begin{pmatrix} -2 & 1 & 0 & \cdots & 1 \\ 1 & -2 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & -2 & 1 \\ 1 & \cdots & 0 & 1 & -2 \end{pmatrix},$$

and $\odot$ denotes the component wise product of vectors. Notice that this semi-discrete system coincides with (1.1.7).

Given $T > 0$, we proved in a first instance that the process $(\boldsymbol{U}^{M,N}(t), \boldsymbol{V}^{M,N}(t))_{t\in[0,T]}$ converges towards $(\boldsymbol{u}^M(t), \boldsymbol{v}^M(t))_{t\in[0,T]}$, solution of the previous system. Furthermore, we quantify this convergence showing that it is necessary to consider

$$N \gg M^4 \exp(CM^4T),$$

in order to pass to the limit in one step. Thus, it is concluded that the convergence towards (1.1.8) following this approach demands a superexponential number of particles, which also depends on T, with respect to the spatial discretization.

We next derived an alternative approach by considering another spatial discretization of (1.1.8), denoted by $(\widehat{\boldsymbol{u}}^M(t), \widehat{\boldsymbol{v}}^M(t))_{t\geq 0}$, yielding in this way a new semi-discrete system to compare with the process $(\boldsymbol{U}^{M,N}(t), \boldsymbol{V}^{M,N}(t))_{t\geq 0}$. Moreover, the structure of the difference between these two objects takes a form that is analogous to the discrete version of the continuous equation

$$\partial_t z - \Delta(\mu z) = \Delta f + r, \tag{1.1.10}$$

where z plays the role of the difference between the two objects. This analogy led to the development of what are known as duality lemmas. In the continuous setting these tools allow to control in a suitable norm the solution of Kolmogorov-type equations by means of the initial data and the diffusivity. In a first step, a tool of this kind was developed for the continuous equation (1.1.10), assuming that the diffusivity μ is uniformly bounded, $r = 0$ and also imposing an integrability condition over f. An application of this tool in the setting of bounded solutions for the system (1.1.8) yields one of the main statements, which is a stability result. Considering the norm

$$|||\cdot|||_T := \Big(\|\cdot\|^2_{L^\infty(0,T;H^{-1}(\mathbb{T}^d))} + \|\cdot\|^2_{L^2(Q_T)}\Big)^{1/2},$$

where $\mathbb{T}^d$ is the periodic d-dimensional torus and $Q_T = [0,T] \times \mathbb{T}^d$ the periodic cylinder, the result reads as follows.

Theorem. *Let $T > 0$. Consider a couple (u,v) and $(\overline{u},\overline{v})$ of non-negative uniformly bounded weak solutions of (1.1.8) in dimension d, respectively initialized by (u_0,v_0) and $(\overline{u}_0,\overline{v}_0)$, both bounded. If the smallness condition $\|\overline{u}\|_{L^\infty(Q_T)}\|\overline{v}\|_{L^\infty(Q_T)} < \frac{d_1 d_2}{a_{12}a_{21}}$, is satisfied, then the following stability estimate holds*

$$\|\!|u-\overline{u}|\!\|_T^2 + \|\!|v-\overline{v}|\!\|_T^2 \lesssim \|u_0-\overline{u}_0\|^2_{H^{-1}(\mathbb{T}^d)} + \|v_0-\overline{v}_0\|^2_{H^{-1}(\mathbb{T}^d)} + T\Big([u_0-\overline{u}_0]^2_{\mathbb{T}^d}\|\mu_1(v_0)\|_{L^1(\mathbb{T}^d)} + [v_0-\overline{v}_0]^2_{\mathbb{T}^d}\|\mu_2(u_0)\|_{L^1(\mathbb{T}^d)}\Big),$$

where the constant behind $\lesssim$ depends on the parameters of the model. In particular, if a bounded non-negative solution satisfies the smallness condition, then there is no other bounded non-negative solution starting from the same initial data.

Finally, we translated the continuous duality lemma to a discrete setting in order to obtain quantitative estimates for the comparison between the process $(\boldsymbol{U}^{M,N}(t), \boldsymbol{V}^{M,N}(t))_{t\geq 0}$ and the alternative semi-discrete system $(\widehat{\boldsymbol{u}}^M(t), \widehat{\boldsymbol{v}}^M(t))_{t\geq 0}$. This allowed us to obtain a better time and space scaling relation for the convergence to hold. By considering the piecewise linear interpolation $\pi_M(\boldsymbol{u})$ of a vector $\boldsymbol{u} \in \mathbb{R}^M$ and the system (1.1.8) on the one dimensional torus, this second main result reads as follows.

Theorem. *Assume the existence of a non-negative solution $(\overline{u},\overline{v})$ of C^1 regularity in time and C^4 regularity in space of the system (1.1.8), initialized by $(\overline{u}_0,\overline{v}_0) \in C^4(\mathbb{T})$ and satisfying the same smallness assumption as before. Assuming the existence of a constant C_0 such that,*

$$\|\boldsymbol{U}^{M,N}(0)\|_{1,M} + \|\boldsymbol{V}^{M,N}(0)\|_{1,M} \leq C_0, \quad \textit{almost surely,}$$

for all $M,N \in \mathbb{N}^*$, we have that for any $T > 0$,

$$\mathbb{E}\Big[\|\!|\pi_M(\boldsymbol{U}^{M,N}) - \overline{u}|\!\|_T^2 + \|\!|\pi_M(\boldsymbol{V}^{M,N}) - \overline{v}|\!\|_T^2\Big] \lesssim \mathbb{E}\Big[\|\pi_M(\boldsymbol{U}^{M,N}(0)) - \overline{u}_0\|^2_{H^{-1}(\mathbb{T})} + \|\pi_M(\boldsymbol{V}^{M,N}(0)) - \overline{v}_0\|^2_{H^{-1}(\mathbb{T})}\Big] + M^{-4} + \frac{M^2}{N},$$

where the symbol $\lesssim$ depends on the parameters of the model.

We recall that the existence of local regular solutions for (1.1.8) is ensured by [4, 5].

1.2 Non-local self-diffusion models

Since understanding the spatial behaviour of a population interacting locally turns out to be a challenging question, in this section we will consider another type of interaction which aims to be weaker. To this end we will focus on models in where the particles are allowed to interact spatially within a region or neighborhood centered in the particles' location.

1.2.1 Non-local interaction

We saw in the previous section that a pertinent macroscopic model for local interaction is the one given by the system (SKT). Now, coming back to Kolmogorov's equation (1.1.1), we can modify this equation to derive heuristically, in the same way as before, another model in where we relax the local interaction. Indeed, in order to take into account the effect that the particles do not interact locally, we consider an interaction kernel ρ that will serve to regularize the diffusivity at a given point. This yields the modified Kolmogorov equation

$$\partial_t u - \Delta((\rho * \mu)u) = 0,$$

where μ is a given diffusivity coefficient. Recalling the decomposition (1.1.2), we can see again this equation as the composition of two different behaviours, with the difference being that now all the effects produced locally by μ are averaged according to ρ, yielding a non-local type of interaction.

Models in this class, yet not thoroughly studied as local systems, have also been subject of some attention in the recent years. For example, we have [62, 22, 23] and [46], to name but a few. In general, works involving non-local interactions tend to treat the case in where we have only the effect of transport of individuals or particles according to a non-local field of velocities, that is, only taking into account a term of the form $-\nabla \cdot (u\nabla(\rho * \mu))$ in the spatial evolution.

In a general setting, by considering a non-local effect one could argue in the same spirit of [108], in order to obtain what can be seen as the translation of (SKT) to its non-local version, namely the system

$$\begin{aligned} \partial_t u - \Delta\Big[(d_1 + G_{11} * u + G_{12} * v)u\Big] &= (r_1 - C_{11} * u(t) - C_{12} * v(t))u(t), \\ \partial_t v - \Delta\Big[(d_2 + G_{21} * u + G_{22} * v)v\Big] &= (r_2 - C_{21} * u(t) - C_{22} * v(t))v(t), \end{aligned} \tag{1.2.1}$$

where G_{ij} are interaction kernels and C_{ij} competition kernels, for $i, j = 1, 2$. In what follows, we will be interested in models belonging to this class and its generalizations.

Derivation from a stochastic individual-based approach

As already stated, there is an interest in the derivation of models through the approximation by microscopic models suitably scaled. A classical approach for doing this has been the probabilistic method. This consists in introducing a random microscopic model such that in the limit when the number of particles goes to infinity, known as large population approximation, the empirical particle density of the system approximates, in a suitable sense, a deterministic dynamics specified by a macroscopic equation. This approach turns out to be useful for specified existence of weak solutions for the macroscopic dynamics. The usual assumption in this framework is that the population is large enough such that the law of large numbers makes the random fluctuations negligible in the limit.

This approach, generally known as stochastic individual-based modeling in the ecological context, spurred from the seminal paper of Fournier and Méléard [61], and since then it has showed to be a very useful technique for deriving asymptotic results for microscopic models when the number of individuals is very large. It started a fruitful branch of research and gave rise to models and macroscopic dynamics that give useful insights of biological or ecological systems (see for example [24, 26, 25] and [7]). Within this approach, the interest is focused in giving a coherent dynamics to each specific individual and then on the equation that arises when we consider the aggregated behaviour of the individuals, which is seen through an empirical measure, rather than the particular laws that the particles follow in the infinite limit. One of the central motivations for pursuing this approach comes from the idea that the macroscopic dynamics is far more simple to analyze than the many particle dynamics when the number of particles is very large, which usually yields an untractable system.

Nonetheless, and in a very general sense, the idea of approximating non-linear partial differential equations through the convergence of an empirical measure associated with an interacting particle system with mean-field interactions can be traced back to Sznitman [110] (and references therein). An underlying notion that is present in this context is the propagation of chaos property, which can be formally stated as follows. Consider E a measurable metric space and $(P^N)_{N\in\mathbb{N}}$ a sequence of symmetric (exchangeable) probability measures over E^N. We say that this sequence is μ-chaotic if, for any $k \in \mathbb{N}^*$, the k-marginal P_k^N of P^N converges weakly to $\mu^{\otimes k}$ as $N \to \infty$, or equivalently, if the empirical measure associated to P^N, that is, the empirical measure constructed using the entries of the vector sampled according to P^N, converges in law to the deterministic measure μ as N goes to infinity. We can understand this notion of chaos as asymptotic independence of particles. In its seminal article, Kac [72] introduced the notion of propagation of chaos while studying a process of colliding particles. In this work, Kac showed that the convergence of the many particle system followed from the propagation of chaos property.

In this direction, system (1.2.1) arises in a somewhat natural way, as it turns out to be impossible to directly define an interacting particle system, composed of independent particles evolving in a diffusive way, in where the particles undergo local interaction, since two independent particles will never encounter. Thus, by the introduction of an interaction kernel, this standard approach for approximating second order parabolic equations can be used for generating weak solutions.

Following this and the individual-based approach, Fontbona and Méléard introduced in [58] a stochastic particle model such that its large population limit satisfies a general version of (1.2.1) and that we loosely describe in what follows. Consider a population composed of n species, in where each one possesses its own spatial dynamics depending on the distribution of the whole system, and demographically they are under competitive pressure. The spatial configuration of the species $i = 1, \dots, n$, is described by the empirical measure

$$\mu_t^{i,K} = \frac{1}{K} \sum_{n=1}^{N_t^{K,i}} \delta_{X_t^{n,i}}, \tag{1.2.2}$$

where $K \in \mathbb{N}^*$ is the charge capacity, $N_t^{K,i} := K\langle \mu_t^{i,K}, 1\rangle$ the number of particles alive at time t and $X_t^{n,i} \in \mathbb{R}^d$ their positions in space. Each particle has two independent exponential clocks, a reproduction clock of parameter r_i and a mortality clock of parameter $\sum_{j=1}^n C^{i,j} * \mu_t^{j,K}$, both being functions of the particle's position. During their lifetimes, the particles follow a diffusion process with diffusion matrix $a^i(\,\cdot\,, G^{i,1} * \mu_t^{1,K}, \dots, G^{i,n} * \mu_t^{n,K})$ and drift vector $b^i(\,\cdot\,, H^{i,1} * \mu_t^{1,K}, \dots, H^{i,n} * \mu_t^{n,K})$. The choice of coefficients reflect the effect that the spatial density of the different species has through the interaction kernels.

One of the main result of [58] is the convergence in law, when $K \to \infty$, of the former empirical measure towards the weak solution of the system

$$\begin{aligned} \partial_t u^i = \frac{1}{2} \sum_{k,l=1}^{d} \partial^2_{x_k x_l} \Big(a_i^{(kl)}(\,\cdot\,, G^{i,1} * u^1, \dots, G^{i,n} * u^n) u^i \Big) \\ - \sum_{k=1}^{d} \partial_{x_k} \Big(b_i^{(k)}(\,\cdot\,, H^{i,1} * u^1, \dots, H^{i,n} * u^n) u^i \Big) + \Big(r_i - \sum_{j=1}^{n} C^{i,j} * u^j \Big) u^i, \end{aligned} \tag{1.2.3}$$

for $i = 1, \dots, n$, where u^i represents the density of the i-th species. This system is a generalization of (1.2.1) to multiple species.

From non-local interaction towards local interaction

Systems presenting non-local cross-diffusion effects tend to simplify the analysis of solutions, since the non-linearities of (1.2.1) are more tractable than those of (SKT), as the convolution kernels, being smooth functions, tend to regularize the behaviour of the unknown.

Following this observation and a question that was left open in [58], regarding the convergence of the non-local interaction kernels towards local interaction (i.e. $G \to \delta$), it was shown first in [97] that the triangular non-local model associated with (1.2.3) converges towards the model with local interaction, in the case without self-diffusion. This result was then extended in [47] recovering the full model. This last results was obtained under a symmetry condition on the interaction kernels that yields the existence of an entropy structure, which allows to handle the problem in the sense discussed in the previous section. Finally, in this direction, it is worth mentioning again the works [28] and [27], that make use of the non-local limiting system (1.2.3) as an intermediate system to show convergence towards local interaction.

1.2.2 Main question and answer

Given the weak convergence result obtained in [58], relating the stochastic microscopic model (1.2.2) in its large population approximation with the non-local system (1.2.3), a natural question arises in this approximation context: is it possible to quantify, in some suitable distance, how close are the empirical measure of the interacting particle system and the weak solution of the limiting equation?

To answer this question, we focused in the case in where there is one species, showing only a self-diffusive spatial behaviour and having a demographic evolution of logistic type. We developed a probabilistic approach to obtain quantitative estimates for a distance compatible with the objects considered. Particularly, this was done using techniques inspired from optimal transport, namely by the construction of an optimal coupling. Furthermore, the estimates derived allowed us to obtain a propagation of chaos property. Finally, this approach also provides a procedure for constructing optimal couplings in non-conservative systems, which is expected to generalize to more complex models.

1.2.3 Our contribution

We centered on the analysis of a single species in where its density evolves as a non-local self-diffusion equation showing logistic growth, namely an evolution of the form

$$\partial_t \mu_t = \frac{1}{2} \sum_{i,j=1}^{d} \partial^2_{x_i x_j} \Big(a^{(ij)}(\,\cdot\,, G * \mu_t)\mu_t \Big) - \sum_{i=1}^{d} \partial_{x_i} \Big(b^{(i)}(\,\cdot\,, H * \mu_t)\mu_t \Big) + (r - c\langle \mu_t, 1\rangle)\mu_t, \quad (1.2.4)$$

understood in the weak sense, where b and σ are given drift and diffusion coefficients respectively, with $a := \sigma\sigma^{\mathrm{t}}$ and a given initial condition μ_0.

Together with this equation, we considered the stochastic process introduced in [58] that approximates its solution, which is represented by the empirical measure

$$\mu_t^K = \frac{1}{K} \sum_{n=1}^{N_t^K} \delta_{X_t^{n,K}}.$$

In this particle system, each particle follows a diffusion process of the form

$$\mathrm{d}X_t^n = b\Big(X_t^n, H * \mu_t^K(X_t^n)\Big)\,\mathrm{d}t + \sigma\Big(X_t^n, G * \mu_t^K(X_t^n)\Big)\,\mathrm{d}B_t^n,$$

and at the same time it produces an offspring at its current position at a constant rate $r > 0$. On the other hand, particles in this system independently die at rate cN_t^K/K, for $c > 0$, as a result of competition.

By considering $\mathrm{BL}(\mathbb{R}^d)$ the space of Lipschitz-continuous bounded functions on $\mathbb{R}^d$ endowed with the norm

$$\|\varphi\|_{\mathrm{BL}} = \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{\|x - y\|} + \sup_{x} |\varphi(x)|,$$

one has the corresponding dual norm $\|\cdot\|_{\mathrm{BL}^*}$ on the space $\mathcal{M}(\mathbb{R}^d)$ of finite signed measures, which is given by

$$\|\mu - \nu\|_{\mathrm{BL}^*} = \sup_{\|\varphi\|_{\mathrm{BL}} \leq 1} |\langle \mu - \nu, \varphi \rangle|.$$

Also, for every $\mu \in \mathcal{M}^+(\mathbb{R}^d)$ we will denote by $\bar{\mu}$ the normalization by its mass.

Motivated by the comparison of the weak solution of (1.2.4) with the approximating empirical measure under the previous distance, we first showed that the following general relation holds for any $\mu, \nu \in \mathcal{M}^+(\mathbb{R}^d)$

$$\|\mu - \nu\|_{\mathrm{BL}^*} \leq \langle \mu, 1\rangle W_1(\bar{\mu}, \bar{\nu}) + |\langle \mu, 1\rangle - \langle \nu, 1\rangle|, \quad (1.2.5)$$

where the p-Wasserstein distance $W_p(\bar{\mu}, \bar{\nu})$ between two probability measures $\bar{\mu}, \bar{\nu} \in \mathcal{P}(\mathbb{R}^d)$ is defined by

$$W_p(\bar{\mu}, \bar{\nu}) = \left(\inf_{\pi \in \Pi(\bar{\mu}, \bar{\nu})} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \; \pi(\mathrm{d}x, \mathrm{d}y) \right)^{\frac{1}{p}},$$

with $\Pi(\bar{\mu}, \bar{\nu})$ being the set of probability measures over $\mathbb{R}^d \times \mathbb{R}^d$ that have $\bar{\mu}$ and $\bar{\nu}$ respectively as first and second marginals.

In parallel, by writing in a more succinct way (1.2.4), one has

$$\frac{\partial \mu_t}{\partial t} = L^*_{\mu_t}\mu_t + \Big(r - c\langle \mu_t, 1\rangle\Big)\mu_t, \tag{1.2.6}$$

where L_μ is a generator defined for $\phi \in C^2(\mathbb{R}^d)$ by

$$L_\mu\phi(x) = \frac{1}{2}\mathrm{Tr}\left(a(x, G * \mu(x))\mathrm{Hess}(\phi)(x)\right) + b(x, H * \mu(x)) \cdot \nabla\phi(x),$$

and $L^*_{\mu_t}$ its formal adjoint. This took us to consider the equation satisfied by the renormalized solution of (1.2.6), namely the evolution

$$\frac{\partial \bar{\mu}_t}{\partial t} = L^*_{\mu_t}\bar{\mu}_t,$$

where $\bar{\mu}_t = \frac{\mu_t}{\langle \mu_t, 1\rangle}$. This holds only when r and c do not depend on space, which is the case treated here. We identified the previous equation as the evolution of the law of a diffusion process of the form

$$\mathrm{d}Y_t = b(Y_t, H * \mu_t(Y_t))\,\mathrm{d}t + \sigma(Y_t, G * \mu_t(Y_t))\,\mathrm{d}W_t. \tag{1.2.7}$$

Based on the relation (1.2.5) and the previous observation, we considered the process $(Y_t)_{t\geq 0}$, solution of the stochastic differential equation (1.2.7), as an analogous of the nonlinear process in the McKean-Vlasov framework. Then, by following the ideas developed in [35] for the construction of an optimal coupling between an empirical measure and a flow of probability measures in a conservative system, and the results obtained in [60], concerning the quantification of the convergence of the empirical measure of an i.i.d. sample with common law towards such law, we developed an optimal coupling construction in this non-conservative setting. This construction allows to obtain explicit estimates yielding thus the following main result.

Theorem. *Under some assumptions on b, σ, G and H, and given the convergence in law of the initial conditions, if $\sup_K \mathbb{E}(\langle \mu_0^K, 1\rangle^p) < \infty$ for some $p \geq 4$ and the initial condition μ_0 of (1.2.6) has finite q-moments for some $q > 2$, then for all $K \in \mathbb{N}^*$ and $T > 0$ we have that*

$$\sup_{t\in[0,T]} \mathbb{E}\Big(\|\mu_t^K - \mu_t\|_{\mathrm{BL}^*}\Big) \leq C_T \Psi_{d,q}(K),$$

where the explicit rate function $\Psi_{d,q}(K) \to 0$ as $K \to \infty$, with a speed that depends on the dimension and on the moments that μ_0 has, and C_T is a constant depending on T, p, q and the data of the model.

As a consequence of this result, we also obtain a propagation of chaos property for the system.

Definition. *Let $(N^K)_{K\in\mathbb{N}^*}$ be random variables in $\mathbb{N}$ going in law to ∞ as $K \to \infty$. We say a family $((Y^{1,K}, \ldots, Y^{N^K,K}))_{K\in\mathbb{N}^*}$ of random vectors, $(\mathbb{R}^d)^{N^K}$-valued and exchangeable conditionally on N^K for each K, is conditionally P-chaotic given $(N^K)_{K\in\mathbb{N}^*}$ if for some $P \in \mathcal{P}(\mathbb{R}^d)$ and every $j \in \mathbb{N}^*$ the (random) conditional laws $(\mathcal{L}(Y^{1,K}, \ldots, Y^{j\wedge N^K,K} | N^K))_{K\in\mathbb{N}^*}$ given N^K and the event $\{N^K \geq j\}$ converge in distribution in $\mathcal{P}((\mathbb{R}^d)^j)$ to $P^{\otimes j}$ as $K \to \infty$.*

Corollary. *For each $t \geq 0$ the family $((X_t^{1,K}, \ldots, X_t^{N_t^K,K}))_{K\in\mathbb{N}^*}$ is conditionally P-chaotic given $(N_t^K)_{K\in\mathbb{N}^*}$ with $P = \mu_t/\langle \mu_t, 1\rangle$.*

1.3 Genealogical constructions of branching processes

We have seen in the previous sections how interaction affects the evolution of a population. We now focus in what can be seen, in an abstract way, as another type of interaction by considering a population evolving in an environment that changes. Here we have a one way interaction in the sense that the environment affects the evolution and not the other way around. More precisely, we will be centered on models for the growth of the size of a population in an heterogeneous environment and on the genealogical relations between its individuals.

1.3.1 Branching processes and their genealogies

When modeling the evolution of the size of a population, under some broad assumptions on the reproductive dynamics, branching processes turn out to be the classical choice. One of the most classical models in this spirit are Galton-Watson processes ([6] being the classical reference on these objects). These discrete time and space models rely on the assumptions that all individuals are of a single type, they do not affect the reproduction of each other and such that the offspring distribution does not change in time (or generations). Their continuous time and space counterpart are continuous-state branching processes, which first appeared in [56]. A few years later, they were formally introduced by Jiřina in [70] and have been studied thoroughly since then. More precisely, we say that a $[0,\infty]$-valued strong Markov process $(X_t)_{t\geq 0}$, together with a family of laws $(\mathbb{P}_x)_{x\geq 0}$, is a continuous-state branching process if it is càdlàg and satisfies the branching property: for all $\theta \geq 0$ and $x, y \geq 0$

$$\mathbb{E}_{x+y}\Big(e^{-\theta X_t}\Big) = \mathbb{E}_x\Big(e^{-\theta X_t}\Big)\mathbb{E}_y\Big(e^{-\theta X_t}\Big), \quad \forall t \geq 0.$$

This can be understood as that the sum of the size of two independent populations starting from sizes x and y is equal in distribution to the size of a population starting from size $x+y$. In fact, this identity characterize the law of the process. Furthermore, there is a limiting relation [85, 64] between continuous-state branching processes and Galton-Watson processes, which allows us to understand that the former processes may model the evolution of renormalized large populations on a large time scale.

Further generalization of these models have emerged since their introduction, such as multi-type branching processes, models with immigration or considering competition effects. One generalization of particular interest is the case in where the underlying offspring distribution is allowed to vary according to another process, which is seen as the environment in where the population lives. From a modeling point of view, this class yields an even more realistic approach. They were first studied by Smith and Wilkinson [109] in the Galton-Watson case. Their scaling limits were analyzed in [77] and more recently in [11] in a very general setting. Their continuous time and space analogue, known as continuous-state branching processes in random environments, have also been subject of interest in the lasts years (see for example [19] for the continuous paths case and [66, 100] for a more general framework). Given the rapidly changing environment in which we live, due to climate change and related effects, one special case needs to be pointed out from the chain of generalizations and it is the case in where catastrophes occur. This scenario, which was first explored in [12] and then studied more profoundly in [10], deals with the existence of dramatical punctual events that kill a fraction of the population.

One of the questions that naturally arises when modeling the demographic evolution of a population, is how to describe the genealogical structure behind the successive births and deaths. In this direction, we have the famous continuum real tree introduced by Aldous ([1, 2] and [3]) and its connections with the underlying genealogy branching processes, through the coding of the tree via excursion theory (we refer to the survey [86] on this subject). For general continuous-state branching processes a construction in this manner was done later in [87], and it proved to be useful for studying the genealogy of branching processes with immigration [81]. Another kind of genealogical construction was introduced by Bertoin and Le Gall via stochastic flows of bridges ([13, 14, 15] and [16]), which allowed them to provide a notion of the genealogy for a measure-valued branching process. Other approaches include splitting trees [82] and tree-valued processes [41].

In general, from an individual-based point of view, the structure of the genealogical tree is usually implicit in the description of the corresponding population model, but once we pass to the diffusive limit or large population approximation, along with losing the notion of a single individual in the population, we also often lose their genealogical information. Because of this, it is of great help to have an approach that allows us to pass to the limit while preserving the genealogical relations between individuals. This can be done by representing the limiting population model by means of a more tractable system.

Lookdown construction: a first approach

Motivated by the study of the Fleming-Viot process, a measure-valued process that corresponds to the large population limit of Moran-type models (one of the most well-known discrete genetic models), the authors of [48] constructed therein a particular infinite particle system such that its empirical measure corresponds in distribution to a Fleming-Viot process. The advantage of this construction is that this countable representation gives more information about the underlying genealogy of this last process. We describe this construction in what follows.

Consider a population in where each individual is characterised by a trait x belonging to some space E. Starting with N individuals, each one is endowed with a *level*, which in this case will refer to an index ranging from 1 to N, and this assignment will be uniform between all possible assignations. The dynamics of this process is then defined as follows: to each pair of levels (i, j) we append a Poisson process of parameter λ such that when its associated exponential clock rings, the individual with the highest level (i or j) is removed and replaced with a copy of the individual with the lower level. Given this evolution, the generator of the process is then defined for $f \in B(E^N)$ by

$$A^N f(x) = \sum_{1\leq i<j\leq N} \lambda\Big(f(\Phi_{ij}(x)) - f(x)\Big),$$

where $\Phi_{ij}(x)$ is obtained from x by replacing x_j by x_i. Furthermore, we can think formally in its extension to an infinite number particles given by the generator

$$Af(x) = \sum_{1\leq i<j} \lambda\Big(f(\Phi_{ij}(x)) - f(x)\Big),$$

defined in this case for $f \in \cup_{N\geq 0} B(E^N)$, noticing that when $f \in B(E^N)$ we have that $Af = A^N f$. This particle process first appeared in [40] with the goal of studying the support of a

Fleming-Viot process. The name lookdown for this construction comes from the fact that a level j waits a period of time exponentially distributed with parameter $\lambda(j-1)$ and then *looks down* at a level uniformly distributed between the first $j-1$ levels, adopts its value and then continues its evolution.

In [48] it was shown that in fact one can couple this construction with a Moran-type model for a population of size N, such that the empirical measure associated with the first N levels $(X^1, \dots, X^N)$ coincides with the empirical measure associated to this N-Moran model. Furthermore, it was shown that the infinite particle system $(X^1, X^2, \dots)$ turns out to be exchangeable, and thus, by de Finetti's theorem for infinite exchangeable sequences [67], its empirical measure exists. Since the identification of the first N particles with a Moran-type model holds, the de Finetti measure should correspond to a Fleming-Viot process. Also, this gives the conclusion that the genealogy of these first particles is governed by Kingman's coalescent [75]. This is one of the main results obtained in [48], proving that the limiting empirical measure exists as a process and that it corresponds indeed with a Fleming-Viot process.

Behind this infinite particle model we have also a projective property, in the sense that when $N > M$ the M-particle model is embedded in the N-particle model, which simplifies passing to the limit. Moreover, from the construction and this last observation, it can be seen that the genealogy of the first $n < N$ particles does not change when N grows, thus showing the preservation of genealogy property of this construction.

This construction was later improved in [49], and then in [50] in order to cover a broader class of models. Since their development, lookdown-type constructions have served to study different processes from a genealogical point of view. For example, in [17] the lookdown construction is used for relating the genealogical structure of a particular class of branching processes with coalescent processes. Also, we have constructions of this type for the non spatial Λ-Fleming-Viot in [18] and for its spatial version in [113].

The Markov mapping theorem and a refined approach

The seemingly equivalence in the martingale problems for the Moran model and the lookdown construction introduced in [48] led to the development of what is known as the Markov mapping theorem in [78] (see also [79]).

Recall that a process $(X_t)_{t\geq 0}$ is a solution of the martingale problem for the generator A if there is a filtration $(\mathcal{F}_t)_{t\geq 0}$ such that $(X_t)_{t\geq 0}$ is $\mathcal{F}_t$-adapted and satisfies that

$$f(X_t) - f(X_0) - \int_0^t Af(X_s)\,\mathrm{d}s,$$

is a $\mathcal{F}_t$-martingale for each $f \in D(A)$, the domain of A. Let us suppose that we have another filtration $(\mathcal{G}_t)_{t\geq 0}$ such that $\mathcal{G}_t \subset \mathcal{F}_t$ for $t \geq 0$, and let π_t be the conditional distribution of X_t given $\mathcal{G}_t$, i.e., $\pi_t(\mathrm{d}x) = \mathbb{P}(X_t \in \mathrm{d}x \,|\, \mathcal{G}_t)$. Then, a classical observation regarding this is that for every $f \in D(A)$

$$\pi_t f - \pi_0 f - \int_0^t \pi_s Af\,\mathrm{d}s,$$

is a $\mathcal{G}_t$-martingale, where $\pi_t f = \mathbb{E}(f(X_t) \,|\, \mathcal{G}_t)$. Under some technical assumptions on A and its domain $D(A)$, the Markov mapping theorem yields a converse for this last observation. This type of tool allows the study of what is called a filtered martingale problem [79].

More precisely, given a Markov process $(X_t)_{t\geq 0}$ with generator A and letting $Y_t := \gamma(X_t)$ for some measurable transformation γ, the filtered martingale problem in this case refers to the martingale problem satisfied by the conditional law of X_t with respect to the process Y up to time t, in the sense of the previous paragraph. As it was shown in [78], and following the conclusions of [79], as a corollary of a more general filtering result we have that by considering α to be a transition function satisfying $\int h(\gamma(z))\alpha(y, \mathrm{d}z) = h(y)$ for all bounded measurable h, and the transformation

$$C = \left\{ \left(\int f(z)\alpha(\cdot, \mathrm{d}z), \int Af(z)\alpha(\cdot, \mathrm{d}z) \right) : f \in D(A) \right\},$$

with $\nu_0 = \int \alpha(y, \cdot)\mu_0(\mathrm{d}y)$ for some probability measure μ_0, then the following holds: if there exists a solution $\widetilde{Y}$ for the (C, μ_0) martingale problem, then there exists a solution X for the (A, ν_0) martingale problem such that $\widetilde{Y}$ and $Y = \gamma(X)$ have the same distribution. Moreover, we have that $\mathbb{P}(X_t \in \Gamma \,|\, (Y_s)_{s\leq t}) = \alpha(Y_t, \Gamma)$. Finally, if uniqueness holds for the (A, ν_0) martingale problem, then uniqueness holds for the (C, μ_0) martingale problem. The main advantage of the Markov mapping theorem is that it simplifies proofs of equivalence of seemingly different martingale problems.

Following this direction, we can obtain some of the conclusions of [48] by means of this tool. Indeed, for given $N > 0$, by considering $\gamma(x_1, \ldots, x_N) = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}$ and the corresponding transition kernel α, the conclusion that follows from an application of the previous theorem is that when we forget the particular labelling of the first N particles in the lookdown construction from [48], we obtain the Moran model for a population of size N that is related to the construction, and since existence holds for the martingale problem associated to this model, we can conclude the existence of a solution for the martingale problem related to the lookdown construction.

This powerful tool motivated the development of a general lookdown construction in [80], covering a wider range of models, notably ones that present a branching structure that depends on spatial positions of individuals in the model. More recently, this type of construction was extended further in [52], giving in this way a very rich toolbox for constructing in a genealogical way different mechanisms that can be considered when building a population model such as: multiple births and deaths, immigration, mutations and spatial motion among others. By composing these mechanisms it is possible to construct more complex population models with this approach.

1.3.2 Main question and answer

Given the full genealogical tree of a population, a subtree which is of interest is the so-called reduced tree, also known as reconstructed tree or coalescent tree. This tree corresponds to the genealogy of the living individuals in the population at a given time, neglecting the branches associated with individuals that are no longer present at this particular time. A natural question concerning this object is whether we can describe the reduced genealogy of a birth and death process in varying environment and approximate it when we have a very large number of individuals.

Using the approach given by the genealogical constructions introduced in [80], we described the reduced tree for a birth and death process in varying environment. Moreover, thanks to the nested property of the construction used, we approximated this tree in the regime of large population by the tree associated with the Feller diffusion in varying environment. The approach used provides an explicit coupling construction that is expected to yield quantitative estimates for the approximation.

1.3.3 Our contribution

We focused on the description of the genealogy of a birth and death process in varying environment, particularly on the case in where punctual catastrophic events occur. Such process is approximated in the large population regime by the solution of

$$X_t = X_0 + \int_0^t b(s)X_s \,\mathrm{d}s + \int_0^t \sqrt{2\sigma(s)X_s}\,\mathrm{d}B_s + \sum_{j\geq 1,\, t\geq t_j} \Big(\beta(m_{t_j})^{-1} - 1\Big) X_{t_j^-}, \tag{1.3.1}$$

where b, σ, β, $(t_j)_{j\in\mathbb{N}}$ and $(m_{t_j})_{j\in\mathbb{N}}$ are parameters of the model. Here, t_j represents a catastrophe time and m_{t_j} its intensity, which in turn is modulated by the function $\beta \geq 1$. In particular, we studied the reduced tree associated with the genealogy of these processes. To this end, we used the lookdown construction introduced in [80], which we recall next.

Given $K \in \mathbb{N}^*$, consider the state space $E^K = \bigcup_{n=0}^{\infty}[0,K]^n$. For $u = (u_1, \ldots, u_n)$, let $f(u) = \prod_{i=1}^n g(u_i)$, where g is a continuously differentiable function satisfying $0 \leq g \leq 1$ and $g(u_i) = 1$ for each $u_i > K$. Define the following generator for f as before

$$A_t^K f(u) = f(u)\sum_{i=1}^n 2\sigma(t)\int_{u_i}^K (g(v)-1)\mathrm{d}v + f(u)\sum_{i=1}^n \Big(\sigma(t)u_i^2 - b(t)u_i\Big)\frac{g'(u_i)}{g(u_i)}. \tag{1.3.2}$$

Under the conditions $\sigma(t) \geq 0$ and $K\sigma(t) - b(t) \geq 0$, the stochastic process described by this generator is a particle system in where each particle is characterized by a real value, which is called *level*. The level of each particle evolves according to

$$u(t)' = \sigma(t)u(t)^2 - b(t)u(t).$$

A particle with level u at time t will produce a new particle at rate $2\sigma(t)(K-u)$, and the level of its offspring will be uniformly distributed in $[u, K]$. When the level of a particle reaches the value K, it is removed. Also, we consider another mechanism in the evolution of the process, which will be understood as catastrophes. This is specified by $(t_j)_{j\in\mathbb{N}}$, the sequence of catastrophe times and $(m_{t_j})_{j\in\mathbb{N}}$ their respective intensities. Starting from a given initial condition, we let the process evolve according to the generator (1.3.2) until a catastrophe time arrives. At the catastrophic event occurring at time t_j, we multiply the level of each particle by $\beta(m_{t_j})$, and then we let the process evolve according to the generator (1.3.2) starting with an updated initial condition. We denote the resulting process by $(U_t^K)_{t\geq 0}$.

Following the results obtained in [80], by considering the number of particles with level less than K, we obtain a quantity that evolves in law as a birth and death process with rates $K\sigma(t)$ and $K\sigma(t) - b(t)$ respectively, and such that at each catastrophe time t_j, each particle is removed independently with probability $\beta(m_{t_j})^{-1}$.

On the other hand, when $K \to \infty$ we can heuristically obtain another generator from (1.3.2). The process encoded by this new generator follows a similar dynamics than the previously described. The difference is that now a particle with level u at time t will produce a new particle with uniformly distributed level in the interval $[u+\ell_1, u+\ell_2]$, for $0 \leq \ell_1 \leq \ell_2$, at rate $2\sigma(t)(\ell_2-\ell_1)$. Also, when the level of a particle hits infinity, it is removed. Concerning the catastrophes, the dynamics stays the same, i.e., the level of each particle is amplified at each catastrophic event. We denote this process by $(U_t)_{t\geq 0}$.

Similarly as before, by following [80], we have that $\lim_{N\to\infty} \frac{1}{N}\sum_i \mathbf{1}_{[0,N]}(U_t^i)$ is equal in distribution to a process that evolves according to (1.3.1).

Both of these representations yield an explicit way of constructing a branching process together with its genealogy, and in particular their reduced genealogy. Indeed, given $T>0$ and the processes $(U_t^K)_{t\geq 0}$ and $(U_t)_{t\geq 0}$, the genealogy of the individuals alive at time T in each process is given by the collection of particles whose levels remain below K and infinity respectively, until that time. Moreover, to determine which particles remain below a given value, it suffices to look at the deterministic evolution of the levels, which in turn is the same for all particles.

Specifically, in order to endow the processes with a genealogical structure, we start by enumerating the particles at time 0 according to the increasing order of their levels, and we label them according to this numbering. Then for each particle we consider the product set of its label times its lifetime, which is $[0,T]$, and then we take the union of all of these sets. When a new particle is created at time t', we label it by following the Ulam-Harris-Neveu formalism [98] using the set $\mathcal{U} = \bigcup_{n\geq 0}(\mathbb{N}^*)^n$, and again we consider the product set of its label times its lifetime, which in this case is $[t', T]$, for then taking the union of this set with what we already have. Iterating this procedure yields a chronological tree, which turns out to be a subset of $\mathcal{U}\times[0,T]$.

Given this procedure, we define $\mathbf{R}_T^K$ as the chronological tree constructed from $(U_t^K)_{t\in[0,T]}$ considering only the particles with levels below K, and $\mathbf{R}_T$ the chronological tree constructed from $(U_t)_{t\in[0,T]}$ considering only the particles with levels that do not reach infinity.

In order to compare these two objects, we introduced the following distance

$$d_g^T(\mathbf{T},\mathbf{T}') = \sum_{u\in\mathcal{U}} \int_0^T |\mathbf{1}_{(u,s)\in\mathbf{T}} - \mathbf{1}_{(u,s)\in\mathbf{T}'}| g(s)\,\mathrm{d}s,$$

where the function $g\colon [0,T]\to\mathbb{R}_+$ represents a temporal weight that allows taking into account the fact that the number of branches in the trees go to infinity as we get closer to T, since the reproductive rates explode at this time. Given this, g satisfies an integrability condition to ensure the finiteness of the distance.

Our main result is an approximation under this distance of $\mathbf{R}_T^K$ by $\mathbf{R}_T$.

Theorem. *Suppose that b and σ are continuous, bounded and such that $b(t)\leq 0$, $\sigma(t)\geq 0$ and σ is bounded away from zero. Then, we have that*

$$\mathbb{E}\Big(d_g^T(\mathbf{R}_T,\mathbf{R}_T^K)\Big) \to 0, \quad \textit{as } K\to\infty.$$

CHAPTER 2

Stability of a cross-diffusion system and approximation by repulsive random walks: a duality approach

Contents

This chapter is based on [8], written in collaboration with Vincent Bansaye and Ayman Moussa.

2.1 Introduction and notation

Approximations of interacting large populations is motivated by physics, chemistry, biology and ecology. A famous macroscopic model was introduced by Shigesada, Kawasaki and Teramoto in [108] to describe competing species which diffuse *with* local repulsion. In the case of two species, it writes

$$\begin{cases} \partial_t u - \Delta\Big(d_1 u + a_{11}u^2 + a_{12}uv\Big) = u(r_1 - s_{11}u - s_{12}v), \\ \partial_t v - \Delta\Big(d_2 v + a_{21}uv + a_{22}v^2\Big) = v(r_2 - s_{21}u - s_{22}v), \end{cases}$$

where u and v are the densities of the two species and d_i, r_i, a_{ij} and s_{ij} are non-negative real numbers. Completed by initial and boundary conditions, this system (that we simply refer to as the *SKT system*) offers a model for the spreading of two interacting species which mutually influence their propensity to diffuse, through the cross-diffusion terms a_{ij}. The other coefficients represent either natural diffusion (d_i coefficients), reproduction (r_i coefficients) or competition (s_{ij} coefficients). The main motivation of [108] was to propose a population dynamics model able to detect segregation, that is the existence of non-constant steady states $\overline{u}$ and $\overline{v}$ having disjoint superlevel sets of low threshold value. As a consequence of this motivation, the first mathematical results dealing with this system focused on sufficient conditions for the coefficients to ensure existence of non-constant steady states, with a careful study of the stability of the latter. This study of possible segregation states is still active and we refer to the introduction of [21] for a nice state of the art. It is a striking fact that during its first years of existence within the mathematical community, the SKT system has not been studied through the prism of its Cauchy problem. As a matter of fact, existence of solutions has been tackled only a few years later: the first paper dealing with this issue is [74] and explores the system under very restrictive conditions. Several attempt followed, but only with partial results. A substantial progress was achieved by Amann [4, 5], who proposed a rather abstract approach to study generic quasilinear parabolic systems. The scope of this technology goes far beyond the sole case of cross-diffusion systems. In the specific case of the SKT system, it offers existence of local (regular) solutions, together with a criteria of explosion to decide if the existence is global or not. This fundamental result of Amann has been then used by several authors to establish existence of global solutions for particular forms of the SKT system. This is done, in general, under a strong constraint on the coefficients. For instance, [92] treats the case of equal diffusion rates in low dimension and [68], settles the one of triangular systems (that is, for two species, when $a_{12}a_{21} = 0$). However, the general question of existence of global solution for the complete system remains open, even in low dimension.

Another way to produce a global solution is to sacrifice the regularity of the solutions, and deal with only weak ones. This strategy relies on the so-called *entropic structure* of the system: SKT systems as the one previously introduced, admit Lyapunov functionals which decay along time and whose dissipation allows to control the gradient of the solution. This method has been used successfully in [30] to prove, for the first time, existence of global weak solutions for the SKT system, without restrictive assumptions on its coefficients. After it first discovery in [63], this entropic structure has been explored and generalized to several systems, allowing for the construction of global weak solutions for variants of the original SKT system (see [71] and the references therein). With this low level of regularity for the solutions, uniqueness becomes an issue in itself. It has been studied either under simplifying assumptions on the system like in [104, 32] or in the weak-strong setting thanks to the use of a relative entropy (see [33]).

2.1.1 Objectives and state of the art

This work is initally motivated by yet another mathematical challenge offered by the SKT system: its rigorous derivation. The diffusion operator used in the SKT system is specific. We focus in this paper on the main difficulty raised by this operator, which is the non-linearity of diffusion term. The initial goal of the work is to approximate the conservative SKT system, without self-diffusion, that is the following one

$$\begin{cases} \partial_t u - \Delta(d_1 u + a_{12} uv) = 0, \\ \partial_t v - \Delta(d_2 v + a_{21} uv) = 0, \end{cases} \tag{2.1.1}$$

where u and v are densities and all the coefficients d_i and a_{ij} are assumed to be positive. Whereas (possibly heterogeneous) diffusion of lifeless matter (*e.g.* ink or any type of chemical substance) uses the Fick diffusion operator $-\text{div}(\mu\nabla\cdot)$ to express the spread, SKT systems rely on the (more singular) operator $-\Delta(\mu\,\cdot)$. As it was already explained in [108], this choice of diffusion operator is at the core of the repulsive mechanism allowing the segregation to appear. However, the justification proposed in [108] was rather formal, leaving open the question of the rigorous justification of SKT systems. As far as our knowledge goes, there exist mainly three approaches for the derivation of SKT systems

(i) The first path was proposed in [69], where an SKT model is obtained as an asymptotic limit of a family of reaction-diffusion systems. In this approach the idea is that one of the two species exists in two states (stressed or not), and switch from one to the other with a reaction rate which diverges. This was used in [69] to obtain formally a triangular cross diffusion system. This strategy has been followed with a rigorous analysis, mainly to produce triangular systems (see [111] and references therein) and more recently for a family of "full" systems in [38] which, however, do not include the SKT one.

(ii) Another strategy was proposed by Fontbona and Méléard in [58]. The idea is to start from a stochastic population model in continuous space where the individuals' displacements depend on the presence of concurrents. Then, the large population limit (under adequate scaling) leads to a non-local cross-diffusion model. In comparison with the system (2.1.1), the limit model rigorously derived in [58] is a lot less singular, because of several convolution kernels. It was explicitly asked in [58], whether letting the convolution kernels vanish to the Dirac mass was handable limit or not. A first partial answer was given in [97], but applied for only specific triangular systems. More recently, it was discovered [47] that even for the non-local systems, it is possible to ensure the persistence of the entropy structure, allowing to answer fully to the question of Fontbona and Méléard, at least for the standard SKT system. Let us mention also [28, 27] which also use a non-local model as an intermediate to derive variants of the SKT system.

(iii) The third path was proposed in [37] and justifies the SKT model through a semi-discrete one. The latter is itself derived from a stochastic population model in discrete space where individuals are assumed to move by pair, in order to ensure reversibility of the process and the existence of an entropy for the limit model. In [37] the link to the stochastic was done formally whereas the asymptotic analysis linking the semi-discrete model to the SKT system was proved rigorously, relying on a compactness argument which is allowed thanks to the existence of the Lyapunov functional for the semi-discrete system.

In this paper, we are interested in connections between microscopic random individual-based models (or particle system) and such macroscopic deterministic dynamics, in the spirit of strategies (ii) and (iii) described above. We do not use any non-local approximated system, being inspired instead by the semi-discrete approach proposed in [37]. We consider also a discrete space and that each species moves randomly and is only sensitive to the local size of the other species. Let us comment the main differences and novelties of this work compared to [37]. First, we prove rigorously that the suitably scaled stochastic process converges in law in Skorokhod space to SKT system (2.1.1) and we perform this space and time scaling limit at once. Besides, individuals of each species move independently with a rate proportional to the number of individuals of the other species, on the same site. We do not need to make them move by pair, which may be hard to justify regarding phenomenon at stake. Indeed, we do not need a reversibility property and do not use the entropic structure. The main difficulty to prove convergence of the stochastic process at once lies in the control of the cumulative quadratic rates due to local interactions when the number of sites becomes large. As far as we have seen, entropy structure does not provide the suitable control of these non-linear terms and a way to get tightness and identification in general. We use a different approach based on generalized duality. This provides quantitative estimates in terms of space discretization and size of population. Moreover, at the level of the PDE system, it implies a local uniqueness result for bounded solutions of the SKT system. The duality approach allows to compare locally the stochastic process with its semi discrete deterministic approximation. It is optimal in the sense that it provides the good time space scaling for such an approximation.

Let us describe now the stochastic individual-based model. The population is spatially distributed among M sites. The process under consideration is a continuous time Markov chain $(\boldsymbol{U}(t), \boldsymbol{V}(t))_{t\geq 0}$ taking values in $\mathbb{N}^M \times \mathbb{N}^M$. The two coordinates count the number of individuals of each species at each site, for each time $t \geq 0$. Each individual of each species follows a random walk and its jumps rate increases linearly with respect to the number of individuals of the other species. The dynamic is defined by the jump rates as follows. For any vector of configurations $(\boldsymbol{u}, \boldsymbol{v}) \in \mathbb{N}^M \times \mathbb{N}^M$, the transitions are

$$\begin{aligned} \boldsymbol{u} &\mapsto \boldsymbol{u} + \Big(\mathbf{e}_{i+\theta} - \mathbf{e}_i\Big) \quad \text{at rate} \quad 2u_i(d_1 + a_{12}v_i), \\ \boldsymbol{v} &\mapsto \boldsymbol{v} + \Big(\mathbf{e}_{i+\theta} - \mathbf{e}_i\Big) \quad \text{at rate} \quad 2v_i(d_2 + a_{21}u_i), \end{aligned}$$

where $(\mathbf{e}_j)_{1\leq j\leq M}$ is the canonical basis of $\mathbb{R}^M$, $\mathbf{e}_0 = \mathbf{e}_M$, $\mathbf{e}_{M+1} = \mathbf{e}_1$ and $\theta \in \{-1, 1\}$ with both values equally likely. Let us mention that hydrodynamic limits of other stochastic models with repulsive species have been considered, in particular in the context of exclusion processes, see e.g. [106]. In that case, local densities are bounded so difficulties and limits are different. In an other direction, stochastic versions of the limiting SKT systems have been considered, see e.g. [44, 45].

This work contains two main results which at first sight can appear unrelated in their formulation. The first result is a quantitative stability estimate on the SKT system which bounds the distance between two solutions in terms of their initial distance. This result is based on a new duality lemma and applies for bounded solutions, only if one of them is small enough. As a by-product of this stability estimate, we prove uniqueness of (small) bounded solutions of the conservative SKT system. This result is valid in arbitrary dimension and is, as far as our knowledge goes, new. Uniqueness theorems for (only) bounded solutions of the full SKT system are missing in the current literature [32, 33, 104].

The second main result is the convergence of the properly scaled sequence of processes $(\boldsymbol{U}^{M,N}, \boldsymbol{V}^{M,N})_{M,N\in\mathbb{N}}$ to the SKT system. We obtain quantitative estimates of the gap between the trajectories of this process extended to the continuous space and the solution of SKT system, in a large population and diffusive regime. This analysis is performed in a one dimensional setting for the space variable. The strategy is to insert the semi-discrete model proposed in [37] and estimate separately the gap between our stochastic process and this semi-discrete system and then, estimate (with enough uniformity) the distance between the semi-discrete system and the continuous SKT limit. Following this plan, we first propose a general estimate, which rely on naive bounds of the quadratic diffusion term. Roughly, we simply bound locally the size of the population by the (constant) total number of individuals. These bounds allow for convergence with a fixed number of sites but lead to an unreasonable assumption of a superexponential number of individuals per site when the number of sites increases. When we faced this difficulty, we tried to obtain an estimate as sharp as possible to capture the good scales and compare the semi-discrete system and the continuous one. It's during this step that we discovered the stability estimate described above, which is interesting for its own sake. A nice feature of this stability estimate is that we can transfer it onto the semi-discrete and stochastic setting. We obtain then the convergence of the stochastic model towards the SKT system, with sharp estimates and relevant size scales. This asymptotic study shares a similar limitation as the previous paragraph: it holds only under the assumption of small regular solution of the SKT system, which is ensured by Amann's theorem [4, 5].

The paper is organized as follows. In the end of this section, we collect several notations which will be used throughout the paper. In Section 2.2 we define the (sequence of) stochastic processes we consider, we recover the semi-discrete system introduced in [37] and state our two main results. In Section 2.3 we show the convergence in law in path space of the stochastic process towards the semi-discrete system when the number of individuals goes to infinity but the number of sites remains fixed. We provide a quantification of this convergence. It implies the general (no restriction on the limiting SKT system) but naive (in terms of scales) convergence discussed above. Then, Section 2.4 is dedicated to the duality estimates with source terms and their consequences. These duality estimates account for the interacting system when one of the population is seen as an exogenous environment, which amounts to decouple the two species. In a first short paragraph (Subsection 2.4.1) we state and prove the generalized duality lemma and its application to the stability estimate of the SKT system in the continuous setting. This paragraph is the only one of the study in which we work in arbitrary dimension for the space variable. Then, the rest of Section 2.4 focuses on the translation of these estimates in the semi-discrete setting. This includes the definition of reconstruction operators, the study of the discrete laplacian matrix and the translation of classical function spaces into the discrete setting. Eventually in Section 2.5, we apply the previous machinery to the difference between the stochastic process and the approximated system that solutions of (2.1.1) solve when looked at a semi-discrete level. We then deduce our main asymptotic theorem by controlling some martingales and approximation errors. In a short appendix, we also give a dictionary which gives the correspondence of different objects in the discrete and continuous settings.

2.1.2 Notation

Finite-dimensional vectors

Throughout the article, vectors will always be written in bold letters and if not stated otherwise, the components of the vector $\boldsymbol{u} \in \mathbb{R}^M$ are $(u_i)_{1\leq i\leq M}$. The canonical basis of $\mathbb{R}^M$ will be denoted $(\mathbf{e}_j)_{1\leq j\leq M}$. Due to the periodic boundary condition that we will use, we will frequently use the convention $\mathbf{e}_0 = \mathbf{e}_M$ and $\mathbf{e}_{M+1} = \mathbf{e}_1$.

Given $M \in \mathbb{N}$ and $p \in [1, \infty]$ we denote by $\|\cdot\|_p = \left(\sum_{i=1}^M |x_i|^p\right)^{1/p}$ the usual ℓ^p norm on $\mathbb{R}^M$ and $\|\cdot\|_{p,M}$ the rescaled norm defined for $\boldsymbol{x} \in \mathbb{R}^M$ by

$$\|\boldsymbol{x}\|_{p,M} \coloneqq \left(\frac{1}{M}\sum_{i=1}^M |x_i|^p\right)^{1/p} \qquad \text{for } p < \infty, \text{ and} \quad \|\boldsymbol{x}\|_\infty \coloneqq \max_{1\leq i\leq M} |x_i|.$$

Similarly, the corresponding (rescaled) euclidean inner-product of $\mathbb{R}^M$ is denoted $(\cdot|\cdot)_M$:

$$(\boldsymbol{x}|\boldsymbol{y})_M = \frac{1}{M}\sum_{i=1}^M x_i y_i,$$

so that $\|\cdot\|_{2,M}^2 = (\cdot|\cdot)_M$.

The symbol $\odot$ is the internal Hadamard product on $\mathbb{R}^M$, that is $(\boldsymbol{x} \odot \boldsymbol{y})_i = x_i y_i$. We will also often use (when it makes sense) the operator $\boldsymbol{x} \oslash \boldsymbol{y}$ defined by $(\boldsymbol{x} \oslash \boldsymbol{y})_i = x_i/y_i$ and the "vectorial" square-root $\boldsymbol{x}^{1/2}$ whose components are $(\sqrt{x_i})_{1\leq i\leq M}$.

The arithmetic average of all the components of a vector $\boldsymbol{x}$ will be denoted

$$[\boldsymbol{x}]_M \coloneqq \frac{1}{M}\sum_{i=1}^M x_i.$$

The vector of $\mathbb{R}^M$ for which every component equals 1 is denoted $\mathbf{1}_M$. The orthogonal projection onto $\operatorname{Span}_{\mathbb{R}}(\mathbf{1}_M)^\perp$ is denoted with a tilde, that is: $\widetilde{\boldsymbol{x}} = \boldsymbol{x} - [\boldsymbol{x}]_M \mathbf{1}_M$.

For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^M$ we write $\boldsymbol{x} \geq \boldsymbol{y}$ whenever $\boldsymbol{x} - \boldsymbol{y} \in \mathbb{R}_+^M$.

Functions

We will manipulate random and deterministic functions which may depend on the time variable $t \in \mathbb{R}_+$ and the space variable $x \in \mathbb{T}^d$, where $\mathbb{T} \coloneqq \mathbb{R}/\mathbb{Z}$ is the flat periodic torus. We will rely on the following convention for functions: uppercase letters will be reserved for random elements whereas lowercase letters will represent deterministic functions. Accordingly to the previous paragraph, vector valued functions will be denoted in bold whereas scalar valued functions will be denoted with the normal font.

Quite often results will be stated on a fixed time interval $[0,T]$. For this reason, we introduce the periodic cylinder $Q_T := [0,T] \times \mathbb{T}^d$. For any function space E defined on $\mathbb{T}^d$ or Q_T, the corresponding norm will be denoted $\|\cdot\|_E$, *e.g.* $\|\cdot\|_{L^2(\mathbb{T}^d)}$. In case of a Hilbert structure, the inner-product will be denoted by $(\cdot|\cdot)_E$, *e.g.* $(\cdot|\cdot)_{L^2(\mathbb{T}^d)}$. We will use frequently two Sobolev spaces on $\mathbb{T}^d$, the definition of which we briefly recall for the reader's convenience.

Any distribution $\varphi \in \mathscr{D}'(\mathbb{T}^d)$ decomposes

$$\varphi = \sum_{k\in\mathbb{Z}^d} c_k(\varphi) e_k,$$

where $e_k(x) := e^{2i\pi k\cdot x}$, and $c_k(\varphi) := \langle \varphi, e_k\rangle$. For $s \in \mathbb{R}$ we define $H^s(\mathbb{T}^d)$ as the subspace of $\mathscr{D}'(\mathbb{T}^d)$ whose elements φ satisfy

$$\sum_{k\in\mathbb{Z}^d} |c_k(\varphi)|^2 (1+|k|^2)^s < +\infty,$$

equipped with the norm

$$\|\varphi\|_{H^s(\mathbb{T}^d)} = \left\{ \sum_{k\in\mathbb{Z}^d} |c_k(\varphi)|^2 (1+|k|^2)^s \right\}^{1/2}.$$

By analogy with notation for the average of the previous paragraph, for any integrable function φ defined on $\mathbb{T}^d$, we denote

$$[\varphi]_{\mathbb{T}^d} := \int_{\mathbb{T}^d} \varphi,$$

in general and $[\varphi]_{\mathbb{T}^d} = c_0(\varphi)$ if φ is merely a distribution. The expression

$$\|\varphi\|_{\dot{H}^s(\mathbb{T}^d)} := \left\{ \sum_{k\in\mathbb{Z}^d} |c_k(\varphi)|^2 |k|^{2s} \right\}^{1/2},$$

is only a semi-norm on $H^s(\mathbb{T}^d)$ and is a norm on the homogeneous Sobolev space $\dot{H}^s(\mathbb{T}^d)$ constituted of those elements φ belonging to $H^s(\mathbb{T}^d)$ and having a vanishing mean, *i.e.* for which $[\varphi]_{\mathbb{T}} = c_0(\varphi) = 0$. We use mainly these spaces for $s = 1$ and $s = -1$.

Finally, for any metric space X, $D([0,T],X)$ denotes the space of càdlàg functions from $[0,T]$ to X endowed with the Skorokhod topology.

2.2 Main objects and results

Before stating our main results, we need to define precisely the objects that we aim at considering.

2.2.1 Repulsive random walks and scaling

Let us define the stochastic process by means of a trajectorial representation using Poisson point measures. We consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ satisfying the usual conditions. We introduce a family of independent Poisson point measure $(\mathcal{N}^j)_{j\in\mathbb{N}}$ on $\mathbb{R}_+ \times \mathbb{R}_+ \times \{-1,1\}$ with common intensity $\mathrm{d}s \otimes \mathrm{d}\rho \otimes \beta(\mathrm{d}\theta)$, where β is the law of a Bernoulli$\left(\frac{1}{2}\right)$ random variable. Almost surely the initial data $(\boldsymbol{U}(0), \boldsymbol{V}(0))$ belongs to $\mathbb{N}^M \times \mathbb{N}^M$, and the corresponding process $(\boldsymbol{U}(t), \boldsymbol{V}(t))_{t\geq 0}$ is then defined as the unique strong solution in $D([0,\infty), \mathbb{N}^{2M})$ of the following system of stochastic differential equations (SDEs) driven by the aforementioned measures

$$\begin{cases} \boldsymbol{U}(t) = \boldsymbol{U}(0) + \displaystyle\int_0^t \int_{\mathbb{R}_+\times\{-1,1\}} \sum_{j=1}^M \mathbf{1}_{\rho \leq 2U_j(s^-)(d_1 + a_{12}V_j(s^-))} \Big(\mathbf{e}_{j+\theta} - \mathbf{e}_j\Big) \mathcal{N}^j(\mathrm{d}s, \mathrm{d}\rho, \mathrm{d}\theta), \\ \boldsymbol{V}(t) = \boldsymbol{V}(0) + \displaystyle\int_0^t \int_{\mathbb{R}_+\times\{-1,1\}} \sum_{j=1}^M \mathbf{1}_{\rho \leq 2V_j(s^-)(d_2 + a_{21}U_j(s^-))} \Big(\mathbf{e}_{j+\theta} - \mathbf{e}_j\Big) \mathcal{N}^j(\mathrm{d}s, \mathrm{d}\rho, \mathrm{d}\theta), \end{cases}$$

where the jump rates d_1, d_2, a_{12} and a_{21} are the one of (2.1.1). Uniqueness and existence for the previous system of SDEs are obtained easily from a classical inductive construction. Indeed, the total population size of each species is constant along time: $\|\boldsymbol{U}(t)\|_{1,M} = \|\boldsymbol{U}(0)\|_{1,M}$, $\|\boldsymbol{V}(t)\|_{1,M} = \|\boldsymbol{V}(0)\|_{1,M}$. Therefore, conditionally on the initial value $(\boldsymbol{U}(0), \boldsymbol{V}(0))$, the process $(\boldsymbol{U}(t), \boldsymbol{V}(t))_{t\geq 0}$ is a pure jump Markov process on a finite state space with bounded rates.

We are interested in the approximation (hydrodynamic limit) when the population size and the number of sites tend to infinity. Informally, we consider $(\boldsymbol{U}(M^2t)/N, \boldsymbol{V}(M^2t)/N)_{t\geq 0}$ and interaction now occurs through the local density of individuals. The scaling parameter $N \in \mathbb{N}^*$ yields the normalization of the population per site and provides a limiting density when N goes to infinity. The initial population per site is of order of magnitude N and each species' motion rate is an affine function of the density of the other species on the same site. The motion of each individual is centered and we consider the diffusive regime. As a consequence, we accelerate the time by the factor of M^2, which amounts to multiply the transition rates by M^2.

We denote the renormalized process by $(\boldsymbol{U}^{M,N}(t), \boldsymbol{V}^{M,N}(t))_{t\geq 0}$. Moreover, for $u, v \in \mathbb{R}$ and $i, j = 1, 2$ we set

$$\eta_{1,j}^{M,N}(t) \coloneqq 2M^2 N U_j^{M,N}(t)\Big(d_1 + a_{12}V_j^{M,N}(t)\Big),$$
$$\eta_{2,j}^{M,N}(t) \coloneqq 2M^2 N V_j^{M,N}(t)\Big(d_2 + a_{21}U_j^{M,N}(t)\Big).$$

For a given initial condition $(\boldsymbol{U}^{M,N}(0), \boldsymbol{V}^{M,N}(0))$, the process $(\boldsymbol{U}^{M,N}(t), \boldsymbol{V}^{M,N}(t))_{t\geq 0}$ is the unique solution in $D([0,\infty), \mathbb{R}_+^{2M})$ of the following system of SDEs

$$\begin{cases} \boldsymbol{U}^{M,N}(t) = \boldsymbol{U}^{M,N}(0) + \displaystyle\int_0^t \int_{\mathbb{R}_+\times\{-1,1\}} \sum_{j=1}^M \mathbf{1}_{\rho \leq \eta_{1,j}^{M,N}(s^-)} \frac{\mathbf{e}_{j+\theta} - \mathbf{e}_j}{N} \mathcal{N}^j(\mathrm{d}s, \mathrm{d}\rho, \mathrm{d}\theta), \\ \boldsymbol{V}^{M,N}(t) = \boldsymbol{V}^{M,N}(0) + \displaystyle\int_0^t \int_{\mathbb{R}_+\times\{-1,1\}} \sum_{j=1}^M \mathbf{1}_{\rho \leq \eta_{2,j}^{M,N}(s^-)} \frac{\mathbf{e}_{j+\theta} - \mathbf{e}_j}{N} \mathcal{N}^j(\mathrm{d}s, \mathrm{d}\rho, \mathrm{d}\theta). \end{cases} \tag{2.2.1}$$

2.2.2 The intermediate (semi-discrete) system

To estimate the gap between the discrete stochastic process (2.2.1) and the SKT system (2.1.1), we are going to use a third system on which our asymptotic analysis will pivot

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{u}^M(t) - \Delta_M(d_1\boldsymbol{u}^M(t) + a_{12}\boldsymbol{u}^M(t)\odot\boldsymbol{v}^M(t)\Big) = 0, \\ \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{v}^M(t) - \Delta_M\Big(d_2\boldsymbol{v}^M(t) + a_{21}\boldsymbol{u}^M(t)\odot\boldsymbol{v}^M(t)\Big) = 0, \end{cases} \tag{2.2.2}$$

where the unknowns are the vector valued curves $\boldsymbol{u}^M, \boldsymbol{v}^M : \mathbb{R}_+ \to \mathbb{R}^M$, and the matrix Δ_M is the periodic laplacian matrix, that is

$$\Delta_M \coloneqq M^2 \begin{pmatrix} -2 & 1 & 0 & \cdots & 1 \\ 1 & -2 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & -2 & 1 \\ 1 & \cdots & 0 & 1 & -2 \end{pmatrix} \in \mathrm{M}_M(\mathbb{R}). \tag{2.2.3}$$

This semi-discrete system corresponds to a large population approximation but fixed number of sites M. Existence and uniqueness for (2.2.2) can be proven using the standard Picard-Lindelöf theorem, as this is done in [37] where this semi-discrete system has been introduced.

2.2.3 Formal insight

Before stating our main results, let us give an informal argument to see how the stochastic process (2.2.1) can be linked with the SKT system (2.1.1), through the semi-discrete system (2.2.2).

We first introduce the infinitesimal generator $L^{M,N}$ of the process (2.2.1). For this purpose, we define the translation operator τ_a for any vector $a \in \mathbb{R}^M$. It acts on any function $G \colon \mathbb{R}^M \to \mathbb{R}$ by the formula $\tau_a G(\cdot) \coloneqq G(\cdot + a)$. Then, for $1 \le j \le M$, we define the operator

$$\mathcal{L}_j^N = \tau_{N^{-1}(\boldsymbol{e}_{j+1}-\boldsymbol{e}_j)} + \tau_{N^{-1}(\boldsymbol{e}_{j-1}-\boldsymbol{e}_j)} - 2\mathrm{Id},$$

for $G \colon \mathbb{R}^M \to \mathbb{R}$. We recall here the periodic convention: $\mathbf{e}_0 = \mathbf{e}_M$ and $\mathbf{e}_{M+1} = \mathbf{e}_1$. Then, for any measurable and bounded function $F \colon \mathbb{R}_+^{2M} \to \mathbb{R}$, we define for $(\boldsymbol{u}, \boldsymbol{v}) \in \mathbb{R}_+^{2M}$

$$L^{M,N}F\Big(\boldsymbol{u}, \boldsymbol{v}\Big) = \sum_{j=1}^{M} \frac{1}{2}\left\{\eta_{1,j}^{M,N}(u_j, v_j)\mathcal{L}_j^N[F\Big(\cdot, \boldsymbol{v}\Big)](\boldsymbol{u}) + \eta_{2,j}^{M,N}(u_j, v_j)\mathcal{L}_j^N[F\Big(\boldsymbol{u}, \cdot\Big)](\boldsymbol{v})\right\}.$$

For N going to infinity and F differentiable, Taylor's approximation ensures that $L^{M,N}F$ converges to

$$L^M F(\boldsymbol{u}, \boldsymbol{v}) = \Big(\Delta_M(d_1\boldsymbol{u} + a_{12}\boldsymbol{u}\odot\boldsymbol{v}) \,\Big|\, \nabla_{\boldsymbol{u}} F(\boldsymbol{u}, \boldsymbol{v})\Big) + \Big(\Delta_M(d_2\boldsymbol{v} + a_{21}\boldsymbol{v}\odot\boldsymbol{u}) \,\Big|\, \nabla_{\boldsymbol{v}} F(\boldsymbol{u}, \boldsymbol{v})\Big),$$

where $(\cdot|\cdot)$ is the inner product on $\mathbb{R}^M$ and Δ_M is the discrete laplacian matrix defined in (2.2.3). Roughly, this ensures that for a fixed number of sites, the stochastic model can be approximated in large population by the semi-discrete system (2.2.2). Then, as M goes to infinity, the discrete laplacian represented by Δ_M is expected to be formally replaced by the laplacian, thus the components of $\boldsymbol{u}^M$ and $\boldsymbol{v}^M$ are expected to approach the values of u and v on a uniform grid of step $\frac{1}{M}$, yielding the cross-diffusion system (2.1.1).

2.2.4 Statements

Our first main result is a stability estimate for the conservative SKT system (2.1.1). As far as our knowledge goes, this result is new in the context of weak solutions for the SKT system. To measure the distance between two solutions on a time interval $[0,T]$, we introduce the following norm

$$|||\cdot|||_T := \Big(\|\cdot\|^2_{L^\infty([0,T];H^{-1}(\mathbb{T}^d))} + \|\cdot\|^2_{L^2(Q_T)}\Big)^{1/2}. \tag{2.2.4}$$

We define also the affine functions $\mu_i\colon \mathbb{R}\to\mathbb{R}$ for $i=1,2$, by $\mu_i(x) := d_i + a_{ij}x$ with $\{i,j\} = \{1,2\}$.

Theorem 2.1. *Let $T>0$ and consider a couple $(u,v)\in L^\infty(Q_T)^2$ and $(\overline{u},\overline{v})\in L^\infty(Q_T)^2$ of non-negative bounded weak solutions of the SKT system (2.1.1), respectively initialized by $(u_0,v_0)\in L^\infty(\mathbb{T}^d)^2$ and $(\overline{u}_0,\overline{v}_0)\in L^\infty(\mathbb{T}^d)^2$. If the following smallness condition*

$$\|\overline{u}\|_{L^\infty(Q_T)}\|\overline{v}\|_{L^\infty(Q_T)} < \frac{d_1 d_2}{a_{12}a_{21}}, \tag{2.2.5}$$

is satisfied, then we have the stability estimate

$$\begin{aligned}|||u-\overline{u}|||_T^2 + |||v-\overline{v}|||_T^2 \lesssim \|u_0-\overline{u}_0\|^2_{H^{-1}(\mathbb{T}^d)} + \|v_0-\overline{v}_0\|^2_{H^{-1}(\mathbb{T}^d)}\\ + T\Big([u_0-\overline{u}_0]^2_{\mathbb{T}^d}\|\mu_1(v_0)\|_{L^1(\mathbb{T}^d)} + [v_0-\overline{v}_0]^2_{\mathbb{T}^d}\|\mu_2(u_0)\|_{L^1(\mathbb{T}^d)}\Big),\end{aligned}$$

where the constant behind $\lesssim$ depends only on $a_{ij}, d_i, \|\bar{u}\|_{L^\infty(Q_T)}, \|\bar{v}\|_{L^\infty(Q_T)}$, and $|||\cdot|||_T$ is defined by (2.2.4). In particular, if a bounded non-negative solution satisfies (2.2.5) then, there is no other bounded non-negative solution sharing the same initial data.

Remark 2.2.1. *In case of equality in the smallness condition (2.2.5), uniqueness remains but the stability estimate controls only the H^{-1} part of the $|||\cdot|||_T$ norm.*

The proof of Theorem 2.1 relies on a generalized duality lemma presented in Subsection 2.4.1 and on the concept of *dual solutions* developed in [97], for the Kolmogorov equation. The uniqueness result contained in Theorem 2.1 is conditional: *if* there exists a bounded (non-negative) solution $(\overline{u},\overline{v})$ satisfying (2.2.5), then it is unique in the class of bounded weak solutions. The existence of *global* bounded solutions for the SKT system is a long standing challenge in the context of cross-diffusion systems. Partial results are known, in the wake of the quest of even more regular solutions (which are in particular bounded), like [68] or [92] that we already cited. In the weak solutions setting, the paper [117] gives sufficient –yet restrictive– conditions on the coefficients of the SKT system to ensure boundedness. Since the previous results are rather constraining on the coefficients, we prefer to rely on Amann's theory [4, 5] and understand Theorem 2.1 as a *local* result which holds for sufficiently small initial data. Indeed, Amann's theory proves existence of regular solutions, which exist at least in a neighborhood of the origin. Starting from an initial data satisfying (2.2.5), we recover in this way a small interval on which the estimates remains valid. As the proof of Theorem 2.1 (which is done in Subsection 2.4.1) is totally insensitive to the dimension d, it is here stated in full generality. However, the remaining part of the paper (which deals with the approximation of the SKT system by stochastic processes) will focus on the case $d = 1$.

Before stating our second main result, let us comment briefly the Section 2.3 in which we propose a first approach to estimate the gap between the stochastic process defined by (2.2.1) and the semi-discrete system (2.2.2) on a fixed interval $[0, T]$. The methodology at stake in this paragraph, which is quite rough, allows for asymptotic quadratic closeness between these two objects, *provided that*, as $N, M \to +\infty$, we have the following

$$N \gg M^4 \exp(cM^4 T), \tag{2.2.6}$$

where c is some constant which will become more explicit in the next section. Combining this fact with the compactness result [37, Theorem 8], we obtain convergence (up to a subsequence) of our stochastic process towards a weak solution of the SKT system. The result is general in terms of parameters and form of the solution. However, the drawbacks of this approach are twofold. First, this necessitates a self-diffusion term in the system (which tends indeed to regularize the solution) in order to use the compactness result of [37]. Second, and most importantly, the scaling condition (2.2.6) involves a superexponential and time dependent number of individuals per site in order to make the law of large numbers to hold on each site and to be able to sum local estimates. As we will see, and as we can guess from the form of quadratic variations, it is too restrictive.

We propose instead a different approach, based on the discrete translation of Theorem 2.1. This alternative method does not rely on [37], so that self-diffusion is not needed in the system. The convergence result is obtained by means of a quantitative estimate which bounds the expectation of the $|||\cdot|||_T$-norm of the gap between the stochastic processes and the solution of the SKT system. In particular, there are no compactness tools used and the entropy of the system is not needed. Convergence is then guaranteed only with a quadratic number of individuals per site. This corresponds to the expected scaling for having local control of the stochastic process by its semi-discrete approximation, since beyond this scaling quadratic variations do not vanish. The main disadvantage of this new method is that, like for Theorem 2.1, it works only in a perturbative setting: it needs the existence of a small regular solution.

In order to state the following result, we need to introduce, for any integer $M \geq 1$, the discretization of the flat (one dimensional) torus $\mathbb{T}$

$$\mathbb{T}_M := \{x_1, x_2, \cdots, x_M\}, \quad \text{with } x_k = \frac{k}{M}, \text{ for } 1 \leq k \leq M. \tag{2.2.7}$$

Given a vector $\boldsymbol{u} \in \mathbb{R}^M$, classically there exists exactly one continuous piecewise linear function defined on $\mathbb{T}$ for which its value on each point x_k of $\mathbb{T}_M$ is given by u_k; we denote this function $\pi_M(\boldsymbol{u})$. We adapt the same notation if instead of $\boldsymbol{u}$ one considers a vector valued map $\boldsymbol{U}$ (which could depend on the event ω or the time t for instance), so that $\pi_M(\boldsymbol{U})$ becomes a real-valued map.

Theorem 2.2. *In the one dimensional case $d = 1$, assume the existence of a non-negative solution $(\overline{u}, \overline{v})$ of C^1 regularity in time and C^4 regularity in space of the system (2.1.1), initialized by $(\overline{u}_0, \overline{v}_0) \in C^4(\mathbb{T})$ and satisfying the smallness assumption (2.2.5). Consider the stochastic processes $(\boldsymbol{U}^{M,N}, \boldsymbol{V}^{M,N})$ defined by (2.2.1) and assume the existence of a constant C_0 such that for all $M, N \in \mathbb{N}$,*

$$\|\boldsymbol{U}^{M,N}(0)\|_{1,M} + \|\boldsymbol{V}^{M,N}(0)\|_{1,M} \leq C_0, \quad \textit{almost surely.} \tag{2.2.8}$$

Then, for any $(M,N) \in \mathbb{N}^2$ such that N/M^2 is large enough, for any $T > 0$,

$$\mathbb{E}\Big[|||\pi_M\Big(\boldsymbol{U}^{M,N}\Big) - \overline{u}|||_T^2 + |||\pi_M\Big(\boldsymbol{V}^{M,N}\Big) - \overline{v}|||_T^2\Big]$$
$$\lesssim \mathbb{E}\Big[\|\pi_M\Big(\boldsymbol{U}^{M,N}(0)\Big) - \overline{u}_0\|_{H^{-1}(\mathbb{T})}^2 + \|\pi_M\Big(\boldsymbol{V}^{M,N}(0)\Big) - \overline{v}_0\|_{H^{-1}(\mathbb{T})}^2\Big] + M^{-4} + \frac{M^2}{N}, \quad (2.2.9)$$

where $|||\cdot|||_T$ is defined (2.2.4) and the symbol $\lesssim$ depends on $C, T, d_i, a_{ij}, \|\overline{u}\|_{L^\infty(Q_T)}, \|\overline{v}\|_{L^\infty(Q_T)}$.

This immediately implies the following convergence for the $|||\cdot|||_T$-norm.

Corollary 2.2.1. *Let $T > 0$. Under the assumptions of Theorem 2.2, consider an extraction function $\phi\colon \mathbb{N} \to \mathbb{N}$ such that $M^2 = \mathrm{o}(\phi(M))$. If the initial positions of the individuals are well-prepared in the sense that*

$$\mathbb{E}\Big[\|\pi_M\Big(\boldsymbol{U}^{M,\phi(M)}(0)\Big) - \overline{u}_0\|_{H^{-1}(\mathbb{T})}^2 + \|\pi_M\Big(\boldsymbol{V}^{M,\phi(M)}(0)\Big) - \overline{v}_0\|_{H^{-1}(\mathbb{T})}^2\Big] \underset{M\to+\infty}{\longrightarrow} 0,$$

then we have

$$\lim_{M\to\infty} \mathbb{E}\Big[|||\pi_M\Big(\boldsymbol{U}^{M,\phi(M)}\Big) - \overline{u}|||_T^2 + |||\pi_M\Big(\boldsymbol{V}^{M,\phi(M)}\Big) - \overline{v}|||_T^2\Big] = 0.$$

Similarly to Theorem 2.1, we still have a smallness condition (2.2.5) on the target solution. In some sense, this restriction is not so surprising. Even though it is a bit more hidden in this asymptotic context, the estimate (2.2.9) already contains a kind of uniqueness property for the target solution $(\overline{u}, \overline{v})$, just as the quantitative estimate of Theorem 2.1. At the very least, (2.2.9) states that among all possible weak solutions, $(\overline{u}, \overline{v})$ is the one which "attracts" such stochastic processes. And then, a natural way to select such a solution is to ensure uniqueness by means of sufficient regularity. These two differences come from the fact that, contrary to the previous result, Theorem 2.2 estimates the distance between a vector-valued stochastic process and a deterministic function which is defined on the whole torus $\mathbb{T}$. This obliges to consider corrector terms. The first one consists in the martingale term which measures locally the gap between the stochastic process and the semi-discrete deterministic approximation. Here, we observe that the estimates are sharp and the scales obtained for convergence are optimal: when $N = \phi(M)$ is of order M^2, the local behavior of the size of the population in the individual based model will remain stochastic at the limit. This limiting stochastic regime should be interesting for future works. The second correction term consists in replacing $\overline{u}$ by a continuous piecewise linear function in order to be able to compare it to the semi-discrete system and thus with $\pi_M(\boldsymbol{U}^{M,N})$. As a matter of fact, the proof of Theorem 2.2 relies on a careful translation of the (idealized) functional setting of Theorem 2.1 to the discrete level, together with the treatment of those corrective terms. This analysis necessitates, among other things, discrete duality lemmas including potential singular error terms. These are stated and proved in Subsection 2.4.4. Let us end up with a remark and perspectives. Another appraoch for future works would be to prove ℓ^∞ estimates for the semi-discrete system such that it is independent of M. With this one could show that the semi-discrete system is not far from verifying the limiting equation, and from here evoke the continuous version of the duality estimates in order to quantify the convergence. Also, the results obtained can be extended to the case in where the system (2.1.1) presents self-diffusion and a source term (which would correspond to adding births and deaths in the stochastic process).

2.3 A general and rough estimate

The trajectorial representation (2.2.1) yields for each coordinate of $\boldsymbol{U}^{M,N}$

$$U_i^{M,N}(t) = U_i^{M,N}(0) - \frac{1}{N}\int_0^t \int_{\mathbb{R}_+\times\{-1,1\}} \mathbf{1}_{\rho\leq \eta_{1,i}^{M,N}(s^-)}\,\mathcal{N}^i(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta)$$
$$+ \frac{1}{N}\int_0^t \int_{\mathbb{R}_+\times\{-1,1\}} \mathbf{1}_{\rho\leq \eta_{1,i-1}^{M,N}(s^-)}\mathbf{1}_{\theta=1}\,\mathcal{N}^{i-1}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta)$$
$$+ \frac{1}{N}\int_0^t \int_{\mathbb{R}_+\times\{-1,1\}} \mathbf{1}_{\rho\leq \eta_{1,i+1}^{M,N}(s^-)}\mathbf{1}_{\theta=-1}\,\mathcal{N}^{i+1}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta). \tag{2.3.1}$$

By compensating the Poisson point measure, we obtain the semimartingale decomposition

$$\boldsymbol{U}^{M,N}(t) = \boldsymbol{A}^{M,N}(t) + \boldsymbol{\mathcal{M}}^{M,N}(t), \tag{2.3.2}$$

where $\boldsymbol{A}^{M,N} = (A_i^{M,N})_{1\leq i\leq M}$ is a continuous process defined by

$$\boldsymbol{A}^{M,N}(t) = \boldsymbol{U}^{M,N}(0) + \int_0^t d_1\Delta_M \boldsymbol{U}^{M,N}(s)\,\mathrm{d}s + \int_0^t a_{12}\Delta_M\Big(\boldsymbol{U}^{M,N}(s)\odot \boldsymbol{V}^{M,N}(s)\Big)\,\mathrm{d}s,$$

with Δ_M as defined in (2.2.3), and $\mathcal{M}_i^{M,N}$ is a square integrable martingale whose predictable quadratic variation is given by

$$\Big\langle \mathcal{M}_i^{M,N}\Big\rangle(t) = \frac{M^2}{N}\int_0^t d_1\Big(2U_i^{M,N}(s) + U_{i+1}^{M,N}(s) + U_{i-1}^{M,N}(s)\Big)\mathrm{d}s \tag{2.3.3}$$
$$+ \frac{M^2}{N}\int_0^t a_{12}\Big(2U_i^{M,N}(s)V_i^{M,N}(s) + U_{i+1}^{M,N}(s)V_{i+1}^{M,N}(s) + U_{i-1}^{M,N}(s)V_{i-1}^{M,N}(s)\Big)\mathrm{d}s.$$

The analogous decomposition holds for the coordinates of $(\boldsymbol{V}^{M,N}(t))_{t\geq 0}$, the second species.

Let us give first estimates of the gap between the stochastic process and its approximation in large population for a fixed number of sites. Let

$$\boldsymbol{\mathcal{U}}^{M,N}(t) = \boldsymbol{U}^{M,N}(t) - \boldsymbol{u}^M(t), \quad \boldsymbol{\mathcal{V}}^{M,N}(t) = \boldsymbol{V}^{M,N}(t) - \boldsymbol{v}^M(t).$$

Proposition 2.3.1. *Assume that there exists $C_0 > 0$ such that almost surely, for any $M, N \geq 1$,*

$$\max(\|\boldsymbol{U}^{M,N}(0)\|_{1,M}, \|\boldsymbol{V}^{M,N}(0)\|_{1,M}, \|\boldsymbol{u}^M(0)\|_{1,M}, \|\boldsymbol{v}^M(0)\|_{1,M}) \leq C_0.$$

Then, for any $T \geq 0$, there exist $c_1, c_2 > 0$ such that for any $M, N \geq 1$,

$$\mathbb{E}\bigg(\sup_{t\in[0,T]}\Big\|\boldsymbol{\mathcal{U}}^{M,N}(t)\Big\|_{2,M}^2 + \sup_{t\in[0,T]}\Big\|\boldsymbol{\mathcal{V}}^{M,N}(t)\Big\|_{2,M}^2\bigg)$$
$$\leq \bigg(\mathbb{E}\Big(\Big\|\boldsymbol{\mathcal{U}}^{M,N}(0)\Big\|_{2,M}^2 + \Big\|\boldsymbol{\mathcal{V}}^{M,N}(0)\Big\|_{2,M}^2\Big) + c_1\Big(\frac{M^2}{\sqrt{N}} + T\frac{M^3}{N}\Big)\bigg)e^{c_2\left(M^4+\frac{M^2}{\sqrt{N}}\right)T},$$

where c_1 only depends on the diffusion parameters and the initial bounds and c_2 only depends on the diffusion parameters.

In particular, this estimate guarantees that the normalized stochastic process converges to the semi discrete SKT system when the population size becomes large and the number of sites is fixed. As evoked in the introduction, this is a first step for convergence to the continuous SKT system, when the semi discrete system itself indeed converges to the expected continuous limit.

Proof. First, using the fact that the total number of individuals is constant along time, we observe that under our assumptions

$$\max(\|\boldsymbol{U}^{M,N}(t)\|_{1,M}, \|\boldsymbol{V}^{M,N}(t)\|_{1,M}) = \max(\|\boldsymbol{U}^{M,N}(0)\|_{1,M}, \|\boldsymbol{V}^{M,N}(0)\|_{1,M}) \le C_0, \tag{2.3.4}$$

almost surely for any $M, N \ge 1$, and

$$\max(\|\boldsymbol{u}^{M}(t)\|_{1,M}, \|\boldsymbol{v}^{M}(t)\|_{1,M}) = \max(\|\boldsymbol{u}^{M}(0)\|_{1,M}, \|\boldsymbol{v}^{M}(0)\|_{1,M}) \le C_0, \tag{2.3.5}$$

for any $M \ge 1$. Combining (2.3.2) and (2.2.2), we notice that the process $\boldsymbol{\mathcal{U}}^{M,N}(t)$ has finite variations and satisfies

$$\begin{aligned}\boldsymbol{\mathcal{U}}^{M,N}(t) = \boldsymbol{\mathcal{U}}^{M,N}(0) &+ \int_0^t d_1 \Delta_M \boldsymbol{\mathcal{U}}^{M,N}(s)\,\mathrm{d}s \\ &+ \int_0^t a_{12}\Delta_M\Big(\boldsymbol{U}^{M,N}(s)\odot\boldsymbol{V}^{M,N}(s) - \boldsymbol{u}^M(s)\odot\boldsymbol{v}^M(s)\Big)\,\mathrm{d}s + \boldsymbol{\mathcal{M}}^{M,N}(t).\end{aligned}$$

Consider now the square of its coordinates

$$\mathcal{U}_i^{M,N}(t)^2 = \mathcal{U}_i^{M,N}(0)^2 + \int_0^t 2\,\mathcal{U}_i^{M,N}(s^-)\,\mathrm{d}\mathcal{U}_i^{M,N}(s) + R_i^{M,N}(t),$$

for $i = 1, \dots, M$, where

$$\begin{aligned}R_i^{M,N}(t) &= \sum_{0<s\le t}\Big\{\mathcal{U}_i^{M,N}(s)^2 - \mathcal{U}_i^{M,N}(s^-)^2 - 2\,\mathcal{U}_i^{M,N}(s^-)\Big(\mathcal{U}_i^{M,N}(s) - \mathcal{U}_i^{M,N}(s^-)\Big)\Big\} \\ &= \left(\frac{1}{N}\right)^2 \sum_{0<s\le t} \mathbf{1}_{U_i^{M,N}(s)\neq U_i^{M,N}(s-)},\end{aligned}$$

since the jumps of $\mathcal{U}_i^{M,N}$ and $U_i^{M,N}$ are of size $1/N$. Putting the two expressions together yields

$$\begin{aligned}\mathcal{U}_i^{M,N}(t)^2 = \mathcal{U}_i^{M,N}(0)^2 &+ 2d_1\int_0^t \mathcal{U}_i^{M,N}(s)\Big(\Delta_M\boldsymbol{\mathcal{U}}^{M,N}(s)\Big)_i\,\mathrm{d}s \\ &+ 2a_{12}\int_0^t \mathcal{U}_i^{M,N}(s)\Big(\Delta_M\Big(\boldsymbol{U}^{M,N}(s)\odot\boldsymbol{V}^{M,N}(s) - \boldsymbol{u}^M(s)\odot\boldsymbol{v}^M(s)\Big)\Big)_i\,\mathrm{d}s \\ &+ 2\int_0^t \mathcal{U}_i^{M,N}(s^-)\,\mathrm{d}\mathcal{M}_i^{M,N}(s) + R_i^{M,N}(t).\end{aligned}$$

Given $\boldsymbol{u} \in \mathbb{R}^M$ let us introduce the discrete gradient vector $\nabla_M^+\boldsymbol{u} = (M(u_{i+1} - u_i))_{1\le i\le M}$ (recalling the periodic convention). Summing over all the sites $i \in \{1, \dots, M\}$ and using discrete integration by parts in the second and third terms of the right hand side yields

$$\begin{aligned}\Big\|\boldsymbol{\mathcal{U}}^{M,N}(t)\Big\|_2^2 &= \Big\|\boldsymbol{\mathcal{U}}^{M,N}(0)\Big\|_2^2 - 2d_1\int_0^t \Big\|\nabla_M^+\boldsymbol{\mathcal{U}}^{M,N}(s)\Big\|_2^2\,\mathrm{d}s \\ &- 2a_{12}\int_0^t\sum_{i=1}^M\Big(\nabla_M^+\boldsymbol{\mathcal{U}}^{M,N}(s)\Big)_i\Big(\nabla_M^+\Big(\boldsymbol{U}^{M,N}(s)\odot\boldsymbol{V}^{M,N}(s) - \boldsymbol{u}^M(s)\odot\boldsymbol{v}^M(s)\Big)\Big)_i\,\mathrm{d}s \\ &+ 2\sum_{i=1}^M\int_0^t \mathcal{U}_i^{M,N}(s^-)\,\mathrm{d}\mathcal{M}_i^{M,N}(s) + \Big\|\boldsymbol{R}^{M,N}(t)\Big\|_1.\end{aligned}$$

Dropping the second term which is negative, taking absolute value in the third term and using $2|ab| \le |a|^2 + |b|^2$ ensures that

$$\left\|\boldsymbol{\mathcal{U}}^{M,N}(t)\right\|_2^2 \le \left\|\boldsymbol{\mathcal{U}}^{M,N}(0)\right\|_2^2 + a_{12}\int_0^t \left\|\nabla_M^+\boldsymbol{\mathcal{U}}^{M,N}(s)\right\|_2^2 \mathrm{d}s + a_{12}\int_0^t \left\|\nabla_M^+\Big(\boldsymbol{U}^{M,N}(s)\odot\boldsymbol{V}^{M,N}(s) - \boldsymbol{u}^M(s)\odot\boldsymbol{v}^M(s)\Big)\right\|_2^2 \mathrm{d}s + 2\sum_{i=1}^M\int_0^t \mathcal{U}_i^{M,N}(s^-)\,\mathrm{d}\mathcal{M}_i^{M,N}(s) + \left\|\boldsymbol{R}^{M,N}(t)\right\|_1.$$

Let us observe that $\left\|\boldsymbol{R}^{M,N}(t)\right\|_1$ is given by the number of jumps before time t

$$\mathbb{E}\left(\left\|\boldsymbol{R}^{M,N}(t)\right\|_1\right) = 2N^{-2}\mathbb{E}(\#\{t\ge 0 : \boldsymbol{U}^{M,N}(s) \ne \boldsymbol{U}^{M,N}(s^-)\}).$$

Moreover, the total jump rate in the scaled process $\boldsymbol{\mathcal{U}}^{M,N}$, when the number of individuals of each species in site i is equal to (u_i, v_i), is

$$2M^2\sum_{i=1}^M u_i\Big(d_1 + a_{12}\frac{v_i}{N}\Big) \le 2M^2\|\boldsymbol{u}\|_1\Big(d_1 + a_{12}\frac{\|\boldsymbol{v}\|_1}{N}\Big) \le C_0' M^3 N(1+M),$$

where $C_0' = 2(d_1 + a_{12})C_0$, by (2.3.4). Then we get

$$\mathbb{E}\left(\left\|\boldsymbol{R}^{M,N}(t)\right\|_1\right) \le 2C_0'\,t\,\frac{M^3}{N}(1+M).$$

Lets us now deal with the third and fourth terms. We notice that

$$\Big(\nabla_M^+\boldsymbol{\mathcal{U}}^{M,N}(s)\Big)_i^2 = M^2\Big(\mathcal{U}_{i+1}^{M,N}(s) - \mathcal{U}_i^{M,N}(s)\Big)^2 \le 2M^2\left(\mathcal{U}_{i+1}^{M,N}(s)^2 + \mathcal{U}_i^{M,N}(s)^2\right),$$

Similarly, using also $|ab - cd| \le |a-c|b + c|b-d|$ to deal with the difference of products of positive terms, we get

$$\begin{aligned}
&\Big(\nabla_M^+(\boldsymbol{U}^{M,N}(s)\odot\boldsymbol{V}^{M,N}(s) - \boldsymbol{u}^M(s)\odot\boldsymbol{v}^M(s))\Big)_i^2 \\
&\quad\le 4M^2\Big(\|\boldsymbol{u}^M(0)\|_1^2\,\mathcal{V}_{i+1}^{M,N}(s)^2 + \|\boldsymbol{u}^M(0)\|_1^2\,\mathcal{V}_i^{M,N}(s)^2 \\
&\qquad + \|\boldsymbol{V}^{M,N}(0)\|_1^2\,\mathcal{U}_{i+1}^{M,N}(s)^2 + \|\boldsymbol{V}^{M,N}(0)\|_1^2\,\mathcal{U}_i^{M,N}(s)^2\Big) \\
&\quad\le 4C_0^2M^4\Big(\mathcal{V}_{i+1}^{M,N}(s)^2 + \mathcal{V}_i^{M,N}(s)^2 + \mathcal{U}_{i+1}^{M,N}(s)^2 + \mathcal{U}_i^{M,N}(s)^2\Big),
\end{aligned}$$

using (2.3.4) and (2.3.5). Gathering these bounds, taking supremum and then expectation gives us

$$\begin{aligned}
&\mathbb{E}\Big(\sup_{s\in[0,t]}\|\boldsymbol{\mathcal{U}}^{M,N}(s)\|_2^2\Big) \\
&\quad\le \mathbb{E}\Big(\|\boldsymbol{\mathcal{U}}^{M,N}(0)\|_2^2\Big) + 4a_{12}M^2\int_0^t \mathbb{E}\Big(\|\boldsymbol{\mathcal{U}}^{M,N}(s)\|_2^2\Big)\,\mathrm{d}s \\
&\qquad + 8C_0^2a_{12}M^4\left(\int_0^t \mathbb{E}\Big(\|\boldsymbol{\mathcal{V}}^{M,N}(s)\|_2^2\Big)\,\mathrm{d}s + \int_0^t \mathbb{E}\Big(\|\boldsymbol{\mathcal{U}}^{M,N}(s)\|_2^2\Big)\,\mathrm{d}s\right) \\
&\qquad + 2\sum_{i=1}^M \mathbb{E}\Big(\sup_{s\in[0,t]}\int_0^s \mathcal{U}_i^{M,N}(r^-)\,\mathrm{d}\mathcal{M}_i^{M,N}(r)\Big) + 2C_0'\,T\,\frac{M^3}{N}(1+M).
\end{aligned}$$

For the martingale part, we use Cauchy-Schwarz and Burkholder-Davis-Gundy inequalities which together with (2.3.3) and (2.3.4) yield

$$\begin{aligned}
\mathbb{E}\bigg(\sup_{s\in[0,t]}\int_0^s \mathcal{U}_i^{M,N}(r^-)\,\mathrm{d}\mathcal{M}_i^{M,N}(r)\bigg)^2
&\leq \mathbb{E}\bigg(\sup_{s\in[0,t]}\bigg|\int_0^s \mathcal{U}_i^{M,N}(r^-)\,\mathrm{d}\mathcal{M}_i^{M,N}(r)\bigg|^2\bigg)\\
&\leq \mathbb{E}\bigg(\int_0^t \mathcal{U}_i^{M,N}(r^-)^2\,\mathrm{d}\big\langle\mathcal{M}_i^{M,N}\big\rangle(r)\bigg)\\
&\leq 2\frac{M^2}{N}\mathbb{E}\bigg(\Big\|\boldsymbol{U}^{M,N}(0)\Big\|_1\Big(d_1 + a_{12}\,\Big\|\boldsymbol{V}^{M,N}(0)\Big\|_1\Big)\int_0^t \mathcal{U}_i^{M,N}(s)^2\,\mathrm{d}s\bigg)\\
&\leq C_0'\frac{M^3}{N}(1+M)\int_0^t \mathbb{E}\Big(\mathcal{U}_i^{M,N}(s)^2\Big)\,\mathrm{d}s.
\end{aligned}$$

Using that $\sqrt{1+x}\leq 1+x$ for all $x\geq 0$, we obtain

$$\mathbb{E}\bigg(\sup_{s\in[0,t]}\int_0^s \mathcal{U}_i^{M,N}(r^-)\,\mathrm{d}\mathcal{M}_i^{M,N}(r)\bigg)\leq\sqrt{2C_0''}\frac{M^2}{\sqrt{N}}\bigg(1+\int_0^t \mathbb{E}\Big(\mathcal{U}_i^{M,N}(s)^2\Big)\,\mathrm{d}s\bigg).$$

Putting everything together and using again (2.3.4) yields

$$\begin{aligned}
\mathbb{E}\bigg(\sup_{s\in[0,t]}\|\boldsymbol{\mathcal{U}}^{M,N}(s)\|_2^2\bigg) \leq \mathbb{E}\Big(\|\boldsymbol{\mathcal{U}}^{M,N}(0)\|_2^2\Big) + 2\sqrt{2C_0''}\frac{M^3}{\sqrt{N}} + 2C_0'T\frac{M^4}{N}&\\
+\bigg(8C_0a_{12}M^4 + 2\sqrt{2C_0''}\frac{M^2}{\sqrt{N}}\bigg)\int_0^t \mathbb{E}\bigg(\sup_{r\in[0,s]}\|\boldsymbol{\mathcal{U}}^{M,N}(r)\|_2^2\bigg)\,\mathrm{d}s&\\
+ 8C_0a_{12}M^4\int_0^t \mathbb{E}\bigg(\sup_{r\in[0,s]}\|\boldsymbol{\mathcal{V}}^{M,N}(r)\|_2^2\bigg)\,\mathrm{d}s,&
\end{aligned}$$

for some $C_0''>0$. In a similar way we can obtain analogous bounds for $\boldsymbol{V}^{M,N}$. Adding the two inequalities and then applying Gronwall's lemma leads us to the desired conclusion. □

The proof above is general in the sense that we have no conditions on the limiting SKT system. But as explained in the previous sections, convergence with a large number of sites requires a superexponential number of individuals per site. The bounds in the previous proof are indeed not sharp at several steps. In particular, we have controlled the quadratic terms by bounding the local size of one species by the total number of individuals, which is fixed and thus controlled quantity. Similarly, the gradient term has been dominated by brute force since we have summed the components. To go beyond these estimates and deal with the quadratic term, we develop a duality approach. This will bring stability property and allow us to compare the terms involved in the stochastic process to those of the targeted SKT limit. The stochastic process will then appear as a stable perturbation of this SKT system.

2.4 Duality estimates

2.4.1 The continuous setting

The duality lemma is a tool first introduced by Martin, Pierre and Schmitt [94, 105], in the context of reaction-diffusion systems. It consists in an *a priori* estimate for solutions of the Kolmogorov equation. The strength of the estimate is that it requires very low regularity on the diffusivity (merely integrability), which allows its use when dealing with rather weak solutions. We propose below a small generalization of the duality lemma, which was suggested in [97, Remark 7]. As a matter of fact, we will not directly use the duality lemma presented in this paragraph, but rather translate it in a discrete setting (see Subsection 2.4.4 below). The purpose of this paragraph is then twofold. First, prove Theorem 2.1. Second, explain, avoiding several technicalities inherent to the discrete setting, the core ideas that will be used in Subsection 2.4.4. Below, we call a *weak solution* a solution in the distributional sense. During (and only in) this whole paragraph, we work in arbitrary dimension d.

Lemma 2.4.1. *Consider $\mu \in L^\infty(Q_T)$ such that $\alpha := \inf_{Q_T} \mu > 0$, $z_0 \in H^{-1}(\mathbb{T}^d)$ and $f \in L^2(Q_T)$. Then, there exists a unique $z \in L^2(Q_T)$ that solves weakly the Kolmogorov equation*

$$\begin{cases} \partial_t z - \Delta(\mu z) = \Delta f, \\ z(0,\cdot) = z_0. \end{cases} \tag{2.4.1}$$

Furthermore, this solution z belongs to $C([0,T]; H^{-1}(\mathbb{T}^d))$ and satisfies the duality estimate

$$\|z(T)\|^2_{H^{-1}(\mathbb{T}^d)} + \int_{Q_T} \mu z^2 \le \|z_0\|^2_{H^{-1}(\mathbb{T}^d)} + [z_0]^2_{\mathbb{T}^d} \int_{Q_T} \mu + \frac{1}{\alpha} \int_{Q_T} f^2. \tag{2.4.2}$$

Remark 2.4.1. *This duality estimate is stronger than the one stated in [97]: it contains a (singular) source term and allows a uniform-in-time control of the $H^{-1}(\mathbb{T}^d)$ norm.*

Proof. The proof of existence and uniqueness is exactly the same as [97, Theorem 3]: following the naming of this article, z is the unique dual solution of (2.4.1). For this z, the regularity $C([0,T]; H^{-1}(\mathbb{T}^d))$ is obtained classically from the belongings $z \in L^2(Q_T)$ and $\partial_t z \in L^2([0,T]; H^{-2}(\mathbb{T}^d))$, which come from the equation itself. We can thus focus here on the duality estimate which needs to be proven only in the case when every function involved in (2.4.2) is smooth, in the sense that they are C^∞. Indeed, the assumptions on the data give us a smooth sequence $(\mu^n, z_0^n, f_n)_{n\in\mathbb{N}}$ converging to (μ, z_0, f) in $L^1(Q_T) \times H^{-1}(\mathbb{T}^d) \times L^2(Q_T)$, with a uniform bound for the first component. Let's call $(z^n)_{n\in\mathbb{N}}$ the corresponding sequence of solutions. Note that, by parabolic regularity, the z^n's are also smooth. Then, if the duality estimate (2.4.2) is proved in the smooth setting, we get (up to some subsequence) weak(-$\star$) convergence of $(z^n)_{n\in\mathbb{N}}$, in $L^\infty([0,T]; H^{-1}(\mathbb{T}^d)) \cap L^2(Q_T)$. But, by uniqueness of the target equation, the only possible limit point is precisely z, the solution of (2.4.1). The whole sequence $(z^n)_n$ converges therefore weakly(-$\star$) towards z, and (2.4.2) is recovered by the usual semi-continuity argument for weak convergence.

So, without loss of generality, we assume now that μ, z_0, f and z are smooth. This allows to justify rigorously the following computations. For any function w defined on $\mathbb{T}^d$ and having zero average there exists a unique function ϕ of zero average satisfying $\Delta\phi = w$ (which is easily seen *via* the Fourier coefficients). In particular, for any $t \in [0,T]$ there exists a unique $\phi(t)$ of vanishing mean such that $-\Delta\phi(t) = z(t) - [z(t)]_{\mathbb{T}^d}$. By integrating the Kolmogorov equation we get

$$\frac{\mathrm{d}}{\mathrm{d}t}[z(t)]_{\mathbb{T}^d} = 0,$$

so that $[z(t)]_{\mathbb{T}^d} = [z_0]_{\mathbb{T}^d}$ and $-\partial_t\Delta\phi = \partial_t z$. In particular, we have by integration by parts

$$\int_{\mathbb{T}^d} \phi(t)\, \partial_t z(t) = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{T}^d} |\nabla\phi(t)|^2.$$

Therefore, multiplying equation (2.4.1) by ϕ and using integration by parts

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{T}^d} |\nabla\phi(t)|^2 + \int_{\mathbb{T}^d} \mu z(z - [z_0]_{\mathbb{T}^d}) = -\int_{\mathbb{T}^d}(z - [z_0]_{\mathbb{T}^d})f.$$

Integrating in time and using Young's inequality for the right hand side, we get

$$\begin{aligned}\frac{1}{2}\int_{\mathbb{T}^d} |\nabla\phi(T)|^2 + \int_{Q_T} \mu z^2 \leq \int_{Q_T} \mu z[z_0]_{\mathbb{T}^d} + \frac{1}{2}\int_{\mathbb{T}^d} |\nabla\phi(0)|^2 \\ + \frac{1}{2}\int_{Q_T} (z - [z_0]_{\mathbb{T}^d})^2\mu + \frac{1}{2}\int_{Q_T} \frac{f^2}{\mu},\end{aligned}$$

and thus, using $\mu \geq \alpha > 0$,

$$\int_{\mathbb{T}^d} |\nabla\phi(T)|^2 + \int_{Q_T} \mu z^2 \leq \int_{\mathbb{T}^d} |\nabla\phi(0)|^2 + [z_0]_{\mathbb{T}^d}^2\int_{Q_T}\mu + \frac{1}{\alpha}\int_{Q_T} f^2.$$

Noticing that $\|z(t)\|_{\dot H^{-1}(\mathbb{T}^d)} = \|z(t) - [z_0]_{\mathbb{T}^d}\|_{H^{-1}(\mathbb{T}^d)} = \|\nabla\phi(t)\|_2$, once we add $[z_0]_{\mathbb{T}^d}$ to each side of the inequality to get the full $H^{-1}(\mathbb{T}^d)$ norms, the proof is over. □

In Subsection 2.4.4, we will give (in the discrete setting) variants of the previous duality lemma which include in the r.h.s. some error term, which is possibly singular in the time variable. Being able to take into account those error terms will be crucial in the final asymptotic limit studied in Section 2.5. However, already in its current form, the previous duality lemma is a valuable piece of information. We highlight this with an application of this lemma: the proof of Theorem 2.1, which applies to the conservative SKT system (2.1.1) that we consider here with (u_0, v_0) as initial data. We recall the definition of the affine functions $\mu_i(x) := d_i + a_{ij}x$ for $i,j = 1,2$, so that (2.1.1) rewrites

$$\begin{cases}\partial_t u - \Delta(\mu_1(v)u) = 0,\\ \partial_t v - \Delta(\mu_2(u)v) = 0.\end{cases}$$

In particular, we recover the framework of Lemma 2.4.1, as soon as v and u are bounded and non-negative.

Proof of Theorem 2.1. Let's introduce $z := \overline{u} - u$ and $w := \overline{v} - v$, so that, by substraction

$$\begin{aligned}\partial_t z - \Delta(\mu_1(v)z) = \Delta f,\\ \partial_t w - \Delta(\mu_2(u)w) = \Delta g,\end{aligned}$$

where $f := a_{12}\overline{u}(\overline{v} - v)$ and $g := a_{21}\overline{v}(\overline{u} - u)$. Since u and v are bounded and non-negative, we recover the structure of Lemma 2.4.1 and we get

$$\|z(T)\|_{H^{-1}(\mathbb{T}^d)}^2 + d_1 \int_{Q_T} z^2 \le \|z_0\|_{H^{-1}(\mathbb{T}^d)}^2 + [z_0]_{\mathbb{T}^d}^2 \int_{Q_T} \mu_1(v) + \frac{a_{12}^2}{d_1} \|\overline{u}\|_{L^\infty(Q_T)}^2 \int_{Q_T} w^2,$$

$$\|w(T)\|_{H^{-1}(\mathbb{T}^d)}^2 + d_2 \int_{Q_T} w^2 \le \|w_0\|_{H^{-1}(\mathbb{T}^d)}^2 + [w^0]_{\mathbb{T}^d}^2 \int_{Q_T} \mu_2(u) + \frac{a_{21}^2}{d_2} \|\overline{v}\|_{L^\infty(Q_T)}^2 \int_{Q_T} z^2,$$

since $\inf_{Q_T} \mu_i \ge d_i$, $|f| \le a_{12}|w|\|\overline{u}\|_{L^\infty(Q_T)}$ and $|g| \le a_{21}|z|\|\overline{v}\|_{L^\infty(Q_T)}$. By combining the two inequalities we infer

$$\begin{aligned}\|z(T)\|_{H^{-1}(\mathbb{T}^d)}^2 + d_1 \int_{Q_T} z^2 \le \|z_0\|_{H^{-1}(\mathbb{T}^d)}^2 + [z_0]_{\mathbb{T}^d}^2 \int_{Q_T} \mu_1(v) \\ + \frac{a_{12}^2}{d_1 d_2} \|\overline{u}\|_{L^\infty(Q_T)}^2 \Big(\|w_0\|_{H^{-1}(\mathbb{T}^d)}^2 + [w_0]_{\mathbb{T}^d}^2 \int_{Q_T} \mu_2(u) \Big) \\ + d_1 \left(\frac{a_{12}a_{21}}{d_1 d_2} \right)^2 \|\overline{u}\|_{L^\infty(Q_T)}^2 \|\overline{v}\|_{L^\infty(Q_T)}^2 \int_{Q_T} z^2.\end{aligned}$$

In particular, if we want to absorb the last term of the r.h.s. in the l.h.s. the inequality that we need is exactly the smallness condition (2.2.5). If the later is satisfied, and if we allow the symbol $\lesssim$ to depend on $d_i, a_{ij}, \|\overline{u}\|_{L^\infty(Q_T)}$ and $\|\overline{v}\|_{L^\infty(Q_T)}$, we have actually established

$$\begin{aligned}\|z(T)\|_{H^{-1}(\mathbb{T}^d)}^2 + \int_{Q_T} z^2 \lesssim \|z_0\|_{H^{-1}(\mathbb{T}^d)}^2 + \|w^0\|_{H^{-1}(\mathbb{T}^d)}^2 \\ + [z_0]_{\mathbb{T}^d}^2 \int_{Q_T} \mu_1(v) + [w_0]_{\mathbb{T}^d}^2 \int_{Q_T} \mu_2(u).\end{aligned}$$

Since the previous computation is still valid replacing T by any $t \in [0, T]$, we have in fact

$$|||z|||_T^2 \lesssim \|z_0\|_{H^{-1}(\mathbb{T}^d)}^2 + \|w_0\|_{H^{-1}(\mathbb{T}^d)}^2 + [z_0]_{\mathbb{T}^d}^2 \int_{Q_T} \mu_1(v) + [w_0]_{\mathbb{T}^d}^2 \int_{Q_T} \mu_2(u).$$

Exchanging the roles $(z, \overline{u}, \overline{v}, u, v) \leftrightarrow (w, \overline{v}, \overline{u}, v, u)$, the previous right hand side remains unchanged: we have exactly the same estimate for $\|w\|_T^2$ on the left hand side. The proof is over once we notice that $\int_{Q_T} \mu_1(v) = T \int_{\mathbb{T}^d} \mu_1(v_0)$ and $\int_{Q_T} \mu_2(v) = T \int_{\mathbb{T}^d} \mu_2(u_0)$, since the space integrals of u and v are conserved through time. □

2.4.2 Reconstruction operators

As explained in the previous paragraph, we plan now to transfer the previous duality and stability estimates into a discrete setting. The purpose is to be able to use these results on the semi-discrete system (2.2.2). We will have to manipulate several norms on $\mathbb{R}^M$, reminiscent of classical function spaces of the continuous variable. As the number of points M of the discretization will be sent to infinity, it will be crucial to have estimates which do not depend on this parameter. In particular, the following notion of uniform equivalence will be relevant.

Definition 2.4.1. Given norms $P_{1,M}$ and $P_{2,M}$ on $\mathbb{R}^M$, we say that $P_{1,M}$ and $P_{2,M}$ are uniformly equivalent if there exists $\alpha, \beta > 0$ such that

$$\forall M \in \mathbb{N}, \quad \forall \boldsymbol{u} \in \mathbb{R}^M, \quad \alpha P_{1,M}(\boldsymbol{u}) \le P_{2,M}(\boldsymbol{u}) \le \beta P_{1,M}(\boldsymbol{u}).$$

If this is satisfied, we write $P_{1,M} \sim P_{2,M}$.

Given a discretization like (2.2.7), there are several ways to build a function defined on the whole torus $\mathbb{T}$. The generic approach is to fix a profile θ (generally compactly supported) and consider

$$x \mapsto \sum_{k=1}^{M} \theta\left(M(x-x_k)\right) u_k. \tag{2.4.3}$$

Definition 2.4.2. For $\boldsymbol{u} \in \mathbb{R}^M$ and $\theta \coloneqq \mathbf{1}_{(-1,0]}$, the function defined by (2.4.3) is a step function that we denote $\sigma_M(\boldsymbol{u})$. For $\boldsymbol{u} \in \mathbb{R}^M$ and $\theta(z) \coloneqq (1-|z|)^+$, the function defined by (2.4.3) is a piecewise linear function that we denote $\pi_M(\boldsymbol{u})$. The corresponding vector space of functions (step and continuous piecewise linear functions respectively) are denoted

$$\mathfrak{s}_M \coloneqq \left\{\sigma_M(\boldsymbol{u}) : \boldsymbol{u} \in \mathbb{R}^M\right\} \quad \text{and} \quad \mathfrak{p}_M \coloneqq \left\{\pi_M(\boldsymbol{u}) : \boldsymbol{u} \in \mathbb{R}^M\right\}.$$

If $t \mapsto \boldsymbol{u}(t)$ is a map from $[0,T]$ to $\mathbb{R}^M$, we simply denote by $\sigma_M(\boldsymbol{u})$ and $\pi_M(\boldsymbol{u})$ the respective maps from $[0,T]$ to $\mathfrak{s}_M$ and $\mathfrak{p}_M$ respectively.

Proposition 2.4.1. *For $\boldsymbol{u} \in \mathbb{R}^M$ we have $\|\boldsymbol{u}\|_\infty = \|\sigma_M(\boldsymbol{u})\|_{L^\infty(\mathbb{T})} = \|\pi_M(\boldsymbol{u})\|_{L^\infty(\mathbb{T})}$ and for $p < \infty$ we have $\|\boldsymbol{u}\|_{p,M} = \|\sigma_M(\boldsymbol{u})\|_{L^p(\mathbb{T})} \geq \|\pi_M(\boldsymbol{u})\|_{L^p(\mathbb{T})}$. For $\boldsymbol{u} \in \mathbb{R}^M_+$ we have furthermore $\|\sigma_M(\boldsymbol{u})\|_{L^p(\mathbb{T})} \sim \|\pi_M(\boldsymbol{u})\|_{L^p(\mathbb{T})}$ (with a small abuse of notation, because the uniform equivalence holds only on a positive cone).*

Proof. We first notice $\mathbf{1}_{[-1,0]} \star \mathbf{1}_{[0,1]}(x) = \int_{-1}^0 \mathbf{1}_{[0,1]}(x-y)\,\mathrm{d}y = (1-|x|)^+$. In particular, we infer for $\varphi(x) = (1-|x|)^+$

$$\begin{aligned}
\varphi_{k,M}(x) \coloneqq \varphi\left(M(x-x_k)\right) &= \int \mathbf{1}_{[-1,0]}\left(M(x-x_k)-y\right)\mathbf{1}_{[0,1]}(y)\,\mathrm{d}y \\
&= M\int \mathbf{1}_{[-1,0]}\left(M(x-z-x_k)\right)\mathbf{1}_{[0,1]}(Mz)\,\mathrm{d}z \\
&= \theta_{k,M} \star \rho_M(x),
\end{aligned}$$

where $\theta_{k,M}(x) = \mathbf{1}_{[-1,0]}\left(M(x-x_k)\right)$ and $\rho_M(x) = M\mathbf{1}_{[0,1]}(Mx)$. We have thus established $\pi_M(\boldsymbol{u}) = \sigma_M(\boldsymbol{u}) \star \rho_M$ where, $(\rho_M)_M$ is an approximation of the identity. Therefore, we have $\|\pi_M(\boldsymbol{u})\|_{L^p(\mathbb{T})} \leq \|\sigma_M(\boldsymbol{u})\|_{L^p(\mathbb{T})}$.

Conversely, assume $\boldsymbol{u} \geq 0$. By definition we have

$$\pi_M(\boldsymbol{u}) = \sum_{k=1}^{M} u_k \varphi_{k,M},$$

with $\varphi_{k,M}(x) = \varphi(M(x-x_k))$ and $\varphi(x) = (1-|x|)^+$. Recall that for any vector $\boldsymbol{w} \in \mathbb{R}^M$, one has $M^{1/p}\|\boldsymbol{w}\|_{p,M} \leq M\|\boldsymbol{w}\|_{1,M}$. In particular, using $u_k \geq 0$, we infer at any point $x \in \mathbb{T}$

$$\pi_M(\boldsymbol{u})(x) = \sum_{k=1}^{M} u_k\varphi_{k,M} \geq \left(\sum_{k=1}^{M} u_k^p \varphi_{k,M}^p(x)\right)^{1/p},$$

from where we conclude

$$\|\pi_M(\boldsymbol{u})\|_{L^p(\mathbb{T})} \geq M\|\boldsymbol{u}\|_{p,M}^p \|\varphi\|_{L^p(\mathbb{T})}^p = \|\sigma_M(\boldsymbol{u})\|_{L^p(\mathbb{T})}^p \frac{2}{p+1},$$

using that $\|\varphi_{k,M}\|_{L^p(\mathbb{T})}^p = \frac{1}{M}\|\varphi\|_{L^p(\mathbb{T})}^p = \frac{1}{M}\frac{2}{p+1}$. ∎

We end this paragraph with an estimate that belongs to the folklore of the finite element method and omit the proof. It is usually proved using the Bramble-Hilbert lemma, but since here we focus here on the one dimensional case, it is also possible to give a direct, elementary proof.

Lemma 2.4.2. *For $\varphi \in H^2(\mathbb{T})$ and $M \in \mathbb{N}^*$ there exists a unique $\iota_M(\varphi) \in \mathfrak{p}_M$ matching the values of φ on the grid $(x_k)_{1\le k\le M}$. It satisfies*

$$\|\varphi - \iota_M(\varphi)\|_{\dot{H}^{-1}(\mathbb{T})} \lesssim M^{-2}\|\varphi\|_{\dot{H}^2(\mathbb{T})},$$
$$\|\varphi - \iota_M(\varphi)\|_{L^2(\mathbb{T})} \lesssim M^{-2}\|\varphi\|_{\dot{H}^2(\mathbb{T})},$$
$$\|\varphi - \iota_M(\varphi)\|_{\dot{H}^1(\mathbb{T})} \lesssim M^{-1}\|\varphi\|_{\dot{H}^2(\mathbb{T})},$$

where the symbol $\lesssim$ means that the inequality holds up to a constant independent of φ and M.

2.4.3 Prerequisites on the discrete Laplacian matrix

We give in this paragraph several useful properties linked to the discrete periodic Laplacian matrix introduced in (2.2.3). This matrix Δ_M is not invertible, we have the relations $\mathrm{Ker}(\Delta_M) = \mathrm{Span}_{\mathbb{R}}(\mathbf{1}_M)$ and $\mathrm{Ran}(\Delta_M) = \mathrm{Ker}(\Delta_M)^\perp = \{\boldsymbol{u} \in \mathbb{R}^M : [\boldsymbol{u}]_M = 0\}$. We refer to Subsection 2.1.2 for the definition of $\mathbf{1}_M$ and $[\cdot]_M$.

Definition 2.4.3. For each $\boldsymbol{u} \in \mathrm{Ran}(\Delta_M)$ there exists a unique $\Phi \in \mathrm{Ran}(\Delta_M)$ such that $\boldsymbol{u} = \Delta_M\Phi$. By a small abuse of notation we write then $\Phi = \Delta_M^{-1}\boldsymbol{u}$.

Proposition 2.4.2. *The matrix $-\Delta_M$ is symmetric non-negative and admits therefore a unique symmetric non-negative square root that we denote $\sqrt{-\Delta_M}$.*

Proof. The proof is standard and we simply note that the spectrum of $-\Delta_M$ is given by

$$\left\{M^2\left(2-2\cos\left(\frac{2\pi k}{M}\right)\right) : 0 \le k \le M-1\right\} = \left\{4M^2\sin^2\left(\frac{\pi k}{M}\right) : 0 \le k \le M-1\right\} \subset \mathbb{R}_+,$$

which establishes the non-negativeness. □

Proposition 2.4.3. *For any $\mathbf{\Phi} \in \mathbb{R}^M$ we have the estimate $\|\mathbf{\Phi} - [\mathbf{\Phi}]_M\|_{2,M} \le \|\Delta_M\mathbf{\Phi}\|_{2,M}$.*

Remark 2.4.2. *This is the discrete counterpart of the following consequence of the Poincaré-Wirtinger inequality $\|\varphi - [\varphi]_\mathbb{T}\|_{L^2(\mathbb{T})} \lesssim \|\Delta\varphi\|_{L^2(\mathbb{T})}$, for $\varphi \in H^2(\mathbb{T})$.*

Proof. Using the identity $\sin(\pi k/M) = \sin(\pi(M-k)/M)$, the spectrum of $-\Delta_M$ that we identified in the proof of Proposition 2.4.2 rewrites

$$\left\{4M^2\sin^2\left(\frac{\pi k}{M}\right) : 0 \le k \le \frac{M-1}{2}\right\}.$$

In particular, using the inequality $\sin(x) \ge \frac{2}{\pi}x$ valid on $[0,\pi/2]$ we see that apart from 0 all the eigenvalues of $-\Delta_M$ are lower-bounded by 16. $-\Delta_M$ being symmetric, its diagonalization can be written in an orthonormal basis of $\mathbb{R}^M$ that we denote $(\boldsymbol{w}_k)_{0\le k\le M-1}$, with $\boldsymbol{w}_0$ being the (only) element of this set belonging to $\mathrm{Ker}(\Delta_M)$. We have therefore

$$\|\mathbf{\Phi} - [\mathbf{\Phi}]_M\|_{2,M}^2 = \frac{1}{M}\sum_{k=1}^{M-1}|(\mathbf{\Phi}|\boldsymbol{w}_k)|^2 \le \frac{1}{M}\frac{1}{16^2}\sum_{k=1}^{M-1}\lambda_k^2|(\mathbf{\Phi}|\boldsymbol{w}_k)|^2 = \frac{1}{16^2}\|\Delta_M\mathbf{\Phi}\|_{2,M}^2.$$ □

Before introducing an analog of the negative Sobolev norm, we recall a standard computation linked with the Lagrange finite elements method for which we need to introduce the following matrix

$$B_M := \begin{pmatrix} \frac{2}{3} & \frac{1}{6} & 0 & \cdots & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{6} & \cdots & 0 & \frac{1}{6} & \frac{2}{3} \end{pmatrix}. \tag{2.4.4}$$

Proposition 2.4.4. *For $\boldsymbol{w} \in \mathbb{R}^M$ we have*

$$-(\boldsymbol{w}|\Delta_M \boldsymbol{w})_M = \int_{\mathbb{T}} |\nabla \pi_M(\boldsymbol{w})(x)|^2 \, \mathrm{d}x, \tag{2.4.5}$$

where we recall that $(\cdot|\cdot)_M$ denotes the rescaled inner product on $\mathbb{R}^M$ (see Subsection 2.1.2). Furthermore, for any $\boldsymbol{u} \in \mathbb{R}^M$ we have

$$B_M \boldsymbol{u} = -\Delta_M \boldsymbol{w} \iff \forall \psi \in \mathfrak{p}_M, \int_{\mathbb{T}} \psi(x) \, \pi_M(\boldsymbol{u})(x) \, \mathrm{d}x = \int_{\mathbb{T}} \nabla \psi(x) \cdot \nabla \pi_M(\boldsymbol{w})(x) \, \mathrm{d}x. \tag{2.4.6}$$

Proof. $\mathfrak{p}_M$ is the vector space spanned by the functions $\varphi_{k,M}(x) := \varphi(M(x - x_k))$ where $\varphi(x) := (1 - |x|)^+$, so the r.h.s. of the equivalence (2.4.6) boils down to

$$\int_{\mathbb{T}} \varphi_{k,M}(x) \pi_M(\boldsymbol{u})(x) \, \mathrm{d}x = \int_{\mathbb{T}} \nabla \varphi_{k,M}(x) \cdot \nabla \pi_M(\boldsymbol{w})(x) \, \mathrm{d}x,$$

for $k \in \{1, \ldots, M\}$, and one checks that

$$\int_{\mathbb{T}} \varphi_{i,M}(x) \varphi_{j,M}(x) \, \mathrm{d}x = \frac{1}{M} \left(\frac{2}{3} \mathbf{1}_{i=j} + \frac{1}{6} \mathbf{1}_{|i-j|=1} \right),$$
$$\int_{\mathbb{T}} \nabla \varphi_{i,M}(x) \cdot \nabla \varphi_{j,M}(x) \, \mathrm{d}x = M(2\mathbf{1}_{i=j} - \mathbf{1}_{|i-j|=1}),$$

where the equality $|i - j| = 1$ has to be understood modulo M. Expanding $\pi_M(\boldsymbol{u})$ and $\pi_M(\boldsymbol{w})$ on the basis $(\varphi_{k,M})_{1 \le k \le M}$, we get the equivalence (2.4.6). Formula (2.4.5) is obtained in the same fashion, expanding $\pi_M(\boldsymbol{w})$ on the basis. □

We observe that $\boldsymbol{u} \mapsto -(\boldsymbol{u}|\Delta_M^{-1} \boldsymbol{u})_M$ is non-negative, due to the symmetry and non-negativity of $-\Delta_M$ (see Proposition 2.4.2). For $\boldsymbol{u} \in \mathbb{R}^M$, recalling that $\widetilde{\boldsymbol{u}} = \boldsymbol{u} - [\boldsymbol{u}]_M \mathbf{1}_M$, we have then $-(\widetilde{\boldsymbol{u}}|\Delta_M^{-1} \widetilde{\boldsymbol{u}})_M \ge 0$. This enables us to introduce the following norm $\|\cdot\|_{-1,M}$, which is a discrete counterpart of the $H^{-1}(\mathbb{T})$ norm.

Definition 2.4.4. For $\boldsymbol{u} \in \mathbb{R}^M$, we define

$$\|\boldsymbol{u}\|_{-1,M} := \sqrt{-(\widetilde{\boldsymbol{u}}|\Delta_M^{-1} \widetilde{\boldsymbol{u}})_M + [\boldsymbol{u}]_M^2}.$$

This is a norm on $\mathbb{R}^M$.

Proposition 2.4.5. *We have the equivalence*

$$M\|\pi_M(\cdot)\|_{H^{-1}(\mathbb{T})} + \|\pi_M(\cdot)\|_{L^2(\mathbb{T})} \sim M\|\cdot\|_{-1,M} + \|\pi_M(\cdot)\|_{L^2(\mathbb{T})}. \tag{2.4.7}$$

Moreover for any $\boldsymbol{u} \in \mathbb{R}^M$,

$$\|\boldsymbol{u}\|_{-1,M} \ \le\ \|\boldsymbol{u}\|_{2,M}. \tag{2.4.8}$$

Remark 2.4.3. *The above definition is reminiscent of the equality*

$$\|\varphi - [\varphi]_{\mathbb{T}}\|^2_{H^{-1}(\mathbb{T})} = -\int_{\mathbb{T}} (\varphi - [\varphi]_{\mathbb{T}})\psi,$$

where ψ *is the unique solution of* $-\Delta\psi = \varphi - [\varphi]_{\mathbb{T}}$.

Proof. We first observe the uniform equivalences

$$\begin{aligned} \|\pi_M(\boldsymbol{u})\|_{L^2(\mathbb{T})} &\sim \|\pi_M(\widetilde{\boldsymbol{u}})\|_{L^2(\mathbb{T})} + |[\boldsymbol{u}]_M|, \\ \|\pi_M(\boldsymbol{u})\|_{H^{-1}(\mathbb{T})} &\sim \|\pi_M(\widetilde{\boldsymbol{u}})\|_{H^{-1}(\mathbb{T})} + |[\boldsymbol{u}]_M|, \\ \|\boldsymbol{u}\|_{-1,M} &\sim \|\widetilde{\boldsymbol{u}}\|_{-1,M} + |[\boldsymbol{u}]_M|. \end{aligned}$$

Without loss of generality we can therefore establish the uniform equivalence (2.4.7) under the assumption $[\boldsymbol{u}]_M = 0$.

We have $\|\boldsymbol{u}\|^2_{-1,M} = -(\boldsymbol{u}|\Delta_M^{-1}\boldsymbol{u})_M = -(\Delta_M\boldsymbol{\Phi}, \boldsymbol{\Phi})_M$ where $\boldsymbol{\Phi} := -\Delta_M^{-1}\boldsymbol{u}$. Thanks to Proposition 2.4.4 we have therefore

$$\|\boldsymbol{u}\|^2_{-1,M} = \|\nabla\pi_M(\boldsymbol{\Phi})\|^2_{L^2(\mathbb{T})}. \tag{2.4.9}$$

The matrix B_M defined by (2.4.4) satisfies $6B_M = M^{-2}\Delta_M + 6\mathrm{I}_M$, so it commutes with Δ_M. In particular, the equation $\boldsymbol{u} = -\Delta_M\boldsymbol{\Phi}$ is strictly equivalent to

$$B_M\boldsymbol{u} = -\Delta_M\boldsymbol{w},$$

where $\boldsymbol{w} := B_M\boldsymbol{\Phi}$. We obtain from Proposition 2.4.4 that this last equation is exactly equivalent to

$$\forall\psi \in \mathfrak{p}_M, \quad \int_{\mathbb{T}} \psi(x)\,\pi_M(\boldsymbol{u})(x)\,\mathrm{d}x = \int_{\mathbb{T}} \nabla\psi(x)\cdot\nabla\pi_M(\boldsymbol{w})(x)\,\mathrm{d}x.$$

Since we assumed $[\boldsymbol{u}]_M = 0$, we also have that $[\pi_M(\boldsymbol{u})]_{\mathbb{T}} = 0$ and we can therefore solve $-\Delta\varphi_M = \pi_M(\boldsymbol{u})$, for a unique $\varphi_M \in \dot{H}^2(\mathbb{T})$. We have then, by integration by parts,

$$\forall\psi \in \mathfrak{p}_M, \quad \int_{\mathbb{T}} \psi(x)\,\pi_M(\boldsymbol{u})(x)\,\mathrm{d}x = \int_{\mathbb{T}} \nabla\psi(x)\cdot\nabla\varphi_M(x)\,\mathrm{d}x.$$

In particular, we have established

$$\forall\psi \in \mathfrak{p}_M, \quad \int_{\mathbb{T}} \nabla\psi(x)\cdot(\nabla\pi_M(\boldsymbol{w})(x) - \nabla\varphi_M(x))\,\mathrm{d}x = 0,$$

and this equality holds in particular for $\psi = \pi_M(\boldsymbol{w})$. We deduce that for each $\psi \in \mathfrak{p}_M$

$$\begin{aligned}
&\int_{\mathbb{T}} |\nabla \pi_M(\boldsymbol{w})(x) - \nabla \varphi_M(x)|^2 \,\mathrm{d}x \\
&\quad = \int_{\mathbb{T}} (\nabla \pi_M(\boldsymbol{w})(x) - \nabla \varphi_M(x) + \nabla \psi(x) - \nabla \pi_M(\boldsymbol{w})(x)) \cdot (\nabla \pi_M(\boldsymbol{w})(x) - \nabla \varphi_M(x)) \,\mathrm{d}x \\
&\quad = \int_{\mathbb{T}} (\nabla \psi(x) - \nabla \varphi_M(x)) \cdot (\nabla \pi_M(\boldsymbol{w})(x) - \nabla \varphi_M(x)) \,\mathrm{d}x,
\end{aligned}$$

and we get by the Cauchy-Schwarz inequality

$$\|\nabla \pi_M(\boldsymbol{w}) - \nabla \varphi_M\|_{L^2(\mathbb{T})} \leq \inf_{\psi \in \mathfrak{p}_M} \|\nabla \psi - \nabla \varphi_M\|_{L^2(\mathbb{T})}.$$

Taking $\psi = \iota_M(\varphi)$ and using successively $\|\nabla f\|_{L^2(\mathbb{T})} = \|f\|_{\dot{H}^1(\mathbb{T})}$, by the third estimate of Lemma 2.4.2, we get

$$\begin{aligned}
\|\nabla \pi_M(\boldsymbol{w}) - \nabla \varphi_M\|_{L^2(\mathbb{T})} &\leq \|\nabla \iota_M(\varphi) - \nabla \varphi_M\|_{L^2(\mathbb{T})} \\
&= \|\iota_M(\varphi) - \varphi_M\|_{\dot{H}^1(\mathbb{T})} \\
&\lesssim \frac{1}{M} \|\varphi_M\|_{\dot{H}^2(\mathbb{T})},
\end{aligned}$$

where we refer to Subsection 2.1.2 for the definition of the homogeneous norms $\|\cdot\|_{\dot{H}^s(\mathbb{T})}$. Recalling that $-\Delta \varphi_M = \pi_M(\boldsymbol{u})$, we have the equalities $\|\pi_M(\boldsymbol{u})\|_{\dot{H}^{-1}(\mathbb{T})} = \|\nabla \varphi_M\|_{L^2(\mathbb{T})}$ and $\|\varphi_M\|_{\dot{H}^2(\mathbb{T})} = \|\Delta \varphi_M\|_{L^2(\mathbb{T})} = \|\pi_M(\boldsymbol{u})\|_{L^2(\mathbb{T})}$. All in all, using the reversed triangular inequality we have established

$$\Big| \|\nabla \pi_M(\boldsymbol{w})\|_{L^2(\mathbb{T})} - \|\pi_M(\boldsymbol{u})\|_{\dot{H}^{-1}(\mathbb{T})} \Big| \lesssim \frac{1}{M} \|\pi_M(\boldsymbol{u})\|_{L^2(\mathbb{T})}.$$

To conclude, due to (2.4.9), it is sufficient to prove that $\|\nabla \pi_M(\boldsymbol{w})\|_{L^2(\mathbb{T})} \sim \|\nabla \pi_M(\Phi)\|_{L^2(\mathbb{T})}$, where we recall $\boldsymbol{w} = B_M \Phi$. This last equality implies in particular

$$\pi_M(\boldsymbol{w}) = \frac{2}{3} \pi_M(\Phi) + \frac{1}{6} \tau_{\frac{1}{M}} \pi_M(\Phi) + \frac{1}{6} \tau_{-\frac{1}{M}} \pi_M(\Phi),$$

where we recall for $a \in \mathbb{R}$ the translation operator τ_a defined by $\tau_a f(x) = f(x + a)$. We have therefore

$$\nabla \pi_M(\boldsymbol{w}) = \frac{2}{3} \nabla \pi_M(\Phi) + \frac{1}{6} \tau_{\frac{1}{M}} \nabla \pi_M(\Phi) + \frac{1}{6} \tau_{-\frac{1}{M}} \nabla \pi_M(\Phi). \tag{2.4.10}$$

Both $\nabla \pi_M(\boldsymbol{w})$ and $\nabla \pi_M(\Phi)$ belong to $\mathfrak{s}_M(\mathbb{T})$ *i.e.* are respectively equal to some functions $\sigma_M(\boldsymbol{\lambda})$ and $\sigma_M(\boldsymbol{\gamma})$, for some $\boldsymbol{\lambda}, \boldsymbol{\gamma} \in \mathbb{R}^M$.

Note that B_M is uniformly well-conditioned: the spectral radii of B_M and B_M^{-1} are bounded independently of M. This can be seen writing $B_M = \frac{2}{3} I_M + \frac{1}{6} J_M$, where J_M is the matrix

$$J_M := \begin{pmatrix} 0 & 1 & 0 & \cdots & 1 \\ 1 & 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & 1 \\ 1 & \cdots & 0 & 1 & 0 \end{pmatrix}.$$

The eigenvalues of J_M are $\left\{2\cos\left(\frac{2\pi k}{M}\right)\ :\ k\in\{0,\dots,M-1\}\right\}$, so the spectrum of B_M lies within $[1/3,1]$.

The identity (2.4.10) shows that $\boldsymbol{\lambda}=B_M\boldsymbol{\gamma}$, and we have just controlled the euclidean subordinate norms of B_M and B_M^{-1}: we have $\|\boldsymbol{\gamma}\|_{2,M}\sim\|B_M\boldsymbol{\gamma}\|_{2,M}$, and therefore we obtain $\|\nabla\pi_M(\boldsymbol{w})\|_{L^2(\mathbb{T})}\sim\|\nabla\pi_M(\Phi)\|_{L^2(\mathbb{T})}$, thanks to Proposition 2.4.1, concluding the proof of (2.4.7).

Let us turn to the proof of (2.4.8). Using Proposition 2.4.3, $\|\Delta_M^{-1}\widetilde{\boldsymbol{u}}\|_{2,M}\le\|\widetilde{\boldsymbol{u}}\|_{2,M}$ and Cauchy-Schwarz inequality entails that $-(\widetilde{\boldsymbol{u}}|\Delta_M^{-1}\widetilde{\boldsymbol{u}})_M\le\|\widetilde{\boldsymbol{u}}\|_{2,M}^2$. By Pythagore's identity, we obtain (2.4.8), since $\boldsymbol{u}=\widetilde{\boldsymbol{u}}+[\boldsymbol{u}]_M\mathbf{1}_M$ and $\|[\boldsymbol{u}]_M\mathbf{1}_M\|_{2,M}^2=[\boldsymbol{u}]_M^2$. □

Proposition 2.4.6. *For $\boldsymbol{w}\in C^1([0,T];\mathrm{Ran}(\Delta_M))$, we have*

$$-(\Delta_M^{-1}\boldsymbol{w}(t)|\boldsymbol{w}'(t))_M=\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{w}(t)\|_{-1,M}^2,$$

where as usual $(\cdot|\cdot)_M$ denotes the rescaled inner-product on $\mathbb{R}^M$.

Proof. If $\boldsymbol{v}(t):=-\Delta_M^{-1}\boldsymbol{w}(t)$, we have $\Delta_M\boldsymbol{v}(t)=-\boldsymbol{w}(t)$ and therefore $\Delta_M\boldsymbol{v}'(t)=-\boldsymbol{w}'(t)$, with still $[\boldsymbol{v}'(t)]_M=0$. We then have $\boldsymbol{v}'(t)=-\Delta_M^{-1}\boldsymbol{w}'(t)$. We infer, by symmetry of $\sqrt{-\Delta_M}$,

$$\begin{aligned}
-(\Delta_M^{-1}\boldsymbol{w}(t)|\boldsymbol{w}'(t))_M&=-\Big(\boldsymbol{v}(t)|\Delta_M\boldsymbol{v}'(t)\Big)_M\\
&=\Big(\sqrt{-\Delta_M}\boldsymbol{v}(t)|\sqrt{-\Delta_M}\boldsymbol{v}'(t)\Big)_M\\
&=\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\Big(\sqrt{-\Delta_M}\boldsymbol{v}(t)|\sqrt{-\Delta_M}\boldsymbol{v}(t)\Big)_M\\
&=-\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{v}(t)|\Delta_M\boldsymbol{v}(t))_M=\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{w}(t)\|_{-1,M}^2.
\end{aligned}$$ □

2.4.4 The discrete duality lemma

We are now all set to state and prove the discrete duality lemmas. These estimates will apply to linear differential equations with source terms. We first consider the case when the source term is continuous and then the case when it is not regular, respectively Lemma 2.4.3 and 2.4.4. We need to combine them to deal with the approximation of the stochastic process and this is achieved in Proposition 2.4.7.

Lemma 2.4.3. *Consider $\boldsymbol{\mu}\in C([0,T];\mathbb{R}^M_{>0})$ so that each component is uniformly (w.r.t. to time and index) lower bounded by a positive constant $\alpha>0$. Assume that $\boldsymbol{z}\in C^1([0,T];\mathbb{R}^M)$ solves*

$$\boldsymbol{z}'(t)=\Delta_M\Big[\boldsymbol{z}(t)\odot\boldsymbol{\mu}(t)+\boldsymbol{f}(t)\Big]+\boldsymbol{r}(t),$$

where $\boldsymbol{f}$ and $\boldsymbol{r}$ are two functions in $C([0,T];\mathbb{R}^M)$. Then we have the following estimate, for any *parameter* $a > 0$

$$\sup_{t\in[0,T]} \|\boldsymbol{z}(t)\|_{-1,M}^2 + \int_{Q_T} \sigma_M(\boldsymbol{z}\odot\boldsymbol{\mu}^{1/2})(s,x)^2\,\mathrm{d}s\,\mathrm{d}x$$
$$\le (1+a)\Big[\|\boldsymbol{z}(0)\|_{-1,M}^2 + [\boldsymbol{z}(0)]_M^2 \int_0^T [\boldsymbol{\mu}(s)]_M\,\mathrm{d}s + \frac{1}{\alpha}\int_{Q_T}\sigma_M(\boldsymbol{f})(s,x)^2\,\mathrm{d}s\,\mathrm{d}x\Big]$$
$$+ (1+a^{-1})\left(T + T\int_0^T [\boldsymbol{\mu}(s)]_M\,\mathrm{d}s + \frac{1}{\alpha}\right)\int_{Q_T}\sigma_M(\boldsymbol{r})(s,x)^2\,\mathrm{d}s\,\mathrm{d}x, \quad (2.4.11)$$

where the Hadamard product $\odot$ *and the square-root* $\boldsymbol{\mu}^{1/2}$ *are defined in Subsection 2.1.2.*

This is a counterpart of Lemma 2.4.1. In the case, $\boldsymbol{r} = 0$, one can get rid of a.

Proof. We follow the proof of the continuous case, Lemma 2.4.1. Since $\mathrm{Ran}(\Delta_M) \subseteq \mathbf{1}_M^\perp$, we claim

$$[\boldsymbol{z}(t)]_M' = \frac{1}{M}(\boldsymbol{z}'(t), \mathbf{1}_M) = [\boldsymbol{r}(t)]_M,$$

and therefore

$$[\boldsymbol{z}(t)]_M = [\boldsymbol{z}(0)]_M + \int_0^t [\boldsymbol{r}(s)]_M\,\mathrm{d}s. \quad (2.4.12)$$

Recalling the definition $\tilde{\boldsymbol{z}}(t) := \boldsymbol{z}(t) - [\boldsymbol{z}(t)]_M$ we also have

$$\boldsymbol{z}'(t) = \tilde{\boldsymbol{z}}'(t) + [\boldsymbol{r}(t)]_M.$$

Now, taking the inner-product with the vector $\Delta_M^{-1}\tilde{\boldsymbol{z}}(t)$ in the differential equation solved by $\boldsymbol{z}$, we get, using the symmetry of Δ_M and the fact $\Delta_M^{-1}\tilde{\boldsymbol{z}}(t) \in \mathrm{Span}_{\mathbb{R}}(\mathbf{1}_M)^\perp$ (see Subsection 2.4.3),

$$-\Big(\Delta_M^{-1}\tilde{\boldsymbol{z}}(t)\Big|\tilde{\boldsymbol{z}}'(t)\Big)_M + \Big(\tilde{\boldsymbol{z}}(t)\Big|\boldsymbol{z}(t)\odot\boldsymbol{\mu}(t)\Big)_M = -\Big(\tilde{\boldsymbol{z}}(t)\Big|\boldsymbol{f}(t) + \Delta_M^{-1}\tilde{\boldsymbol{r}}(t)\Big)_M.$$

We use Proposition 2.4.6 to identify the first term of the l.h.s. and get

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\tilde{\boldsymbol{z}}(t)\|_{-1,M}^2 + \Big(\tilde{\boldsymbol{z}}(t)\Big|\boldsymbol{z}(t)\odot\boldsymbol{\mu}(t)\Big)_M = -\Big(\tilde{\boldsymbol{z}}(t)\Big|\boldsymbol{f}(t) + \Delta_M^{-1}\tilde{\boldsymbol{r}}(t)\Big)_M. \quad (2.4.13)$$

Using that the entries of $\boldsymbol{\mu}(t)$ are all lower-bounded by $\alpha > 0$ we have the following inequality (see Subsection 2.1.2 for the notation $\oslash$), for any vector $\boldsymbol{g} \in \mathbb{R}^M$

$$\begin{aligned}
\left|\Big(\tilde{\boldsymbol{z}}(t)\Big|\boldsymbol{g}\Big)_M\right| &= \left|\Big(\tilde{\boldsymbol{z}}(t)\odot\boldsymbol{\mu}(t)^{1/2}\Big|\boldsymbol{g}\oslash\boldsymbol{\mu}(t)^{1/2}\Big)_M\right| \\
&\le \|\tilde{\boldsymbol{z}}(t)\odot\boldsymbol{\mu}(t)^{1/2}\|_{2,M}\|\boldsymbol{g}\oslash\boldsymbol{\mu}(t)^{1/2}\|_{2,M} \\
&\le \frac{1}{\sqrt{\alpha}}\|\tilde{\boldsymbol{z}}(t)\odot\boldsymbol{\mu}(t)^{1/2}\|_{2,M}\|\boldsymbol{g}\|_{2,M} \\
&\le \frac{1}{2}\|\tilde{\boldsymbol{z}}(t)\odot\boldsymbol{\mu}(t)^{1/2}\|_{2,M}^2 + \frac{1}{2\alpha}\|\boldsymbol{g}\|_{2,M}^2 \\
&= \frac{1}{2}\Big(\tilde{\boldsymbol{z}}(t)|\tilde{\boldsymbol{z}}(t)\odot\boldsymbol{\mu}(t)\Big)_M + \frac{1}{2\alpha}\|\boldsymbol{g}\|_{2,M}^2,
\end{aligned}$$

where we used Young's inequality. We use this estimate in (2.4.13) with $\boldsymbol{g} \coloneqq \boldsymbol{f}(t) + \Delta_M^{-1}\widetilde{\boldsymbol{r}}(t)$

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\widetilde{\boldsymbol{z}}(t)\|_{-1,M}^2 + \Big(\widetilde{\boldsymbol{z}}(t)\Big|\boldsymbol{z}(t)\odot\boldsymbol{\mu}(t)\Big)_M \\ \leq \frac{1}{2}\Big(\widetilde{\boldsymbol{z}}(t)\Big|\widetilde{\boldsymbol{z}}(t)\odot\boldsymbol{\mu}(t)\Big)_M + \frac{1}{2\alpha}\|\boldsymbol{f}(t) + \Delta_M^{-1}\widetilde{\boldsymbol{r}}(t)\|_{2,M}^2,$$

which, after expanding the definition $\widetilde{\boldsymbol{z}}(t) \coloneqq \boldsymbol{z}(t) - [\boldsymbol{z}(t)]_M$, becomes

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\widetilde{\boldsymbol{z}}(t)\|_{-1,M}^2 + \frac{1}{2}\Big(\boldsymbol{z}(t)\Big|\boldsymbol{z}(t)\odot\boldsymbol{\mu}(t)\Big)_M \\ \leq \frac{1}{2}[\boldsymbol{z}(t)]_M^2[\boldsymbol{\mu}(t)]_M + \frac{1}{2\alpha}\|\boldsymbol{f}(t) + \Delta_M^{-1}\widetilde{\boldsymbol{r}}(t)\|_{2,M}^2.$$

Using Proposition 2.4.3 to infer $\|\Delta_M^{-1}\widetilde{\boldsymbol{r}}(t)\|_{2,M} \leq \|\widetilde{\boldsymbol{r}}(t)\|_{2,M} \leq \|\boldsymbol{r}(t)\|_{2,M}$ and the convex inequality $(x+y)^2 \leq (1+a)x^2 + (1+a^{-1})y^2$ we eventually get, after integration in time

$$\|\widetilde{\boldsymbol{z}}(t)\|_{-1,M}^2 + \int_0^t \|\boldsymbol{z}(s)\odot\boldsymbol{\mu}(s)^{1/2}\|_{2,M}^2\,\mathrm{d}s \\ \leq \|\widetilde{\boldsymbol{z}}(0)\|_{-1,M}^2 + \int_0^t [\boldsymbol{z}(s)]_M^2[\boldsymbol{\mu}(s)]_M\,\mathrm{d}s \\ + \frac{1+a}{\alpha}\int_0^t \|\boldsymbol{f}(s)\|_{2,M}^2\,\mathrm{d}s + \frac{1+a^{-1}}{\alpha}\int_0^t \|\boldsymbol{r}(s)\|_{2,M}^2\,\mathrm{d}s. \quad (2.4.14)$$

On the other hand, using once more the above convex inequality, we claim from (2.4.12) and Cauchy-Schwarz inequality that

$$[\boldsymbol{z}(t)]_M^2 \leq (1+a)[\boldsymbol{z}(0)]_M^2 + (1+a^{-1})T\int_0^T [\boldsymbol{r}(s)]_M^2\,\mathrm{d}s.$$

Summing the two last inequalities we obtain (2.4.11) since for any vector $\boldsymbol{u} \in \mathbb{R}^M$ we have $\|\boldsymbol{u}\|_{2,M} = \|\sigma_M(\boldsymbol{u})\|_{L^2(\mathbb{T})}$. □

Lemma 2.4.4. *Consider $\boldsymbol{\mu} \in C([0,T];\mathbb{R}_{>0}^M)$ so that each component is uniformly (w.r.t. to time and index) lower bounded by a positive constant $\alpha > 0$. Assume that $\boldsymbol{z}_d \colon [0,T] \to \mathbb{R}^M$ solves*

$$\boldsymbol{z}_d(t) = \int_0^t \Delta_M\Big[\boldsymbol{z}_d(s)\odot\boldsymbol{\mu}(s)\Big]\,\mathrm{d}s + \boldsymbol{x}_d(t), \quad (2.4.15)$$

where $\boldsymbol{x}_d$ is any càdlàg $\mathbb{R}^M$ valued function over $[0,T]$. Then we have the following estimate

$$\sup_{t\in[0,T]}\|\boldsymbol{z}_d(t)\|_{-1,M}^2 + \int_{Q_T}\sigma_M(\boldsymbol{z}_d\odot\boldsymbol{\mu}^{1/2})(s,x)^2\,\mathrm{d}s\,\mathrm{d}x \\ \lesssim \sup_{t\in[0,T]}\|\boldsymbol{x}_d(t)\|_{-1,M}^2 + \int_0^T [\boldsymbol{\mu}(s)]_M[\boldsymbol{x}_d(s)]_M^2\,\mathrm{d}s, \quad (2.4.16)$$

where the constant behind $\lesssim$ is universal and $\boldsymbol{\mu}^{1/2}$ denotes the vector map whose entries are the square-roots of the ones of $\boldsymbol{\mu}$.

Remark 2.4.4. *In this lemma we consider the (discrete) Kolmogorov equation with a singular source term $\boldsymbol{x}_d$. The mere integrability of this term forbids to differentiate in time this equation, so we cannot proceed as we have done in the proof Lemma 2.4.3.*

Proof. Using (2.4.15), we first remark that $[\boldsymbol{z}_d]_M = [\boldsymbol{x}_d]_M$ and therefore

$$\widetilde{\boldsymbol{z}}_d(t) = \int_0^t \Delta_M \Big[\boldsymbol{z}_d(s) \odot \boldsymbol{\mu}(s)\Big] \,\mathrm{d}s + \widetilde{\boldsymbol{x}}_d(t).$$

We take as usual the inner product with $-\Delta_M^{-1}\widetilde{\boldsymbol{z}}_d(t)$ and use symmetry to write

$$-\Big(\Delta_M^{-1}\widetilde{\boldsymbol{z}}_d(t)\Big|\widetilde{\boldsymbol{z}}_d(t)\Big)_M + \int_0^t \Big(\widetilde{\boldsymbol{z}}_d(s)\Big|\boldsymbol{z}_d(s)\odot\boldsymbol{\mu}(s)\Big)_M \,\mathrm{d}s = -\Big(\Delta_M^{-1}\widetilde{\boldsymbol{z}}_d(t), \widetilde{\boldsymbol{x}}_d(t)\Big)_M.$$

Using the definition of the $\|\cdot\|_{-1,M}$ norm (see Proposition 2.4.5) and that $\widetilde{\boldsymbol{z}}_d = \boldsymbol{z}_d - [\boldsymbol{z}_d]_M$ we infer

$$\begin{aligned}\|\widetilde{\boldsymbol{z}}_d(t)\|_{-1,M}^2 &+ \int_0^t \|\boldsymbol{z}_d(s)\odot\boldsymbol{\mu}(s)^{1/2}\|_{2,M}^2 \,\mathrm{d}s \\ &= \int_0^t [\boldsymbol{z}_d(s)]_M\Big(\mathbf{1}_M|\boldsymbol{z}_d(s)\odot\boldsymbol{\mu}(s)\Big)_M \,\mathrm{d}s - \Big(\Delta_M^{-1}\widetilde{\boldsymbol{z}}_d(t), \widetilde{\boldsymbol{x}}_d(t)\Big)_M.\end{aligned}$$

The first term of the r.h.s. can be handled using Young's inequality to absorb a part of it in the l.h.s. and get

$$\begin{aligned}\|\widetilde{\boldsymbol{z}}_d(t)\|_{-1,M}^2 &+ \int_0^t \|\boldsymbol{z}_d(s)\odot\boldsymbol{\mu}(s)^{1/2}\|_{2,M}^2 \,\mathrm{d}s \\ &\lesssim \int_0^t [\boldsymbol{z}_d(s)]_M^2\|\boldsymbol{\mu}(s)^{1/2}\|_{2,M}^2 \,\mathrm{d}s - \Big(\Delta_M^{-1}\widetilde{\boldsymbol{z}}_d(t), \widetilde{\boldsymbol{x}}_d(t)\Big)_M.\end{aligned} \tag{2.4.17}$$

Now, defining $\boldsymbol{\Phi}_M \coloneqq \Delta_M^{-1}\widetilde{\boldsymbol{z}}_d$ and $\boldsymbol{\Psi}_M \coloneqq \Delta_M^{-1}\widetilde{\boldsymbol{x}}_d$ we have that using Cauchy-Schwarz's inequality, the definition of the $\|\cdot\|_{-1,M}$ norm and the symmetry of the discrete laplacian matrix

$$\begin{aligned}-\Big(\Delta_M^{-1}\widetilde{\boldsymbol{z}}_d(t)\;\Big|\;\widetilde{\boldsymbol{x}}_d(t)\Big)_M &= -\Big(\boldsymbol{\Phi}_M(t)\;\Big|\;\Delta_M\boldsymbol{\Psi}_M(t)\Big)_M \\ &= \Big(\sqrt{-\Delta_M}\boldsymbol{\Phi}_M(t)\;\Big|\;\sqrt{-\Delta_M}\boldsymbol{\Psi}_M(t)\Big)_M \\ &\le \|\sqrt{-\Delta_M}\boldsymbol{\Phi}_M(t)\|_{2,M}\|\sqrt{-\Delta_M}\boldsymbol{\Psi}_M(t)\|_{2,M} \\ &= \|\widetilde{\boldsymbol{z}}_d(t)\|_{-1,M}\|\widetilde{\boldsymbol{x}}_d(t)\|_{-1,M}.\end{aligned}$$

Pluging this estimate in (2.4.17), we have

$$\|\widetilde{\boldsymbol{z}}_d(t)\|_{-1,M}^2 + \int_0^t \|\boldsymbol{z}_d(s)\odot\boldsymbol{\mu}(s)^{1/2}\|_{2,M}^2 \,\mathrm{d}s \lesssim \int_0^t [\boldsymbol{z}_d(s)]_M^2\|\boldsymbol{\mu}(s)^{1/2}\|_{2,M}^2 \,\mathrm{d}s + \|\widetilde{\boldsymbol{x}}_d(t)\|_{-1,M}^2.$$

Recalling that $[\boldsymbol{z}_d]_M^2 = [\boldsymbol{x}_d]_M^2$ and adding this term to the inequality, we get (2.4.16). □

Proposition 2.4.7. *Consider $\boldsymbol{\mu} \in C([0,T];\mathbb{R}^M_{>0})$ so that each component is uniformly (w.r.t. to time and index) lower bounded by a positive constant $\alpha > 0$. Assume that $\boldsymbol{z}\colon [0,T] \to \mathbb{R}^M$ solves*

$$\boldsymbol{z}(t) = \boldsymbol{z}(0) + \int_0^t \Delta_M\Big[\boldsymbol{z}(s) \odot \boldsymbol{\mu}(s) + \boldsymbol{f}(s)\Big]\,\mathrm{d}s + \boldsymbol{x}(t),$$

where $\boldsymbol{f}$ is a function in $C([0,T];\mathbb{R}^M)$ and $\boldsymbol{x} = \boldsymbol{x}_r + \boldsymbol{x}_d$, with the regular component $\boldsymbol{x}_r \in C^1([0,T],\mathbb{R}^M)$ and the singular component $\boldsymbol{x}_d$ is any càdlàg $\mathbb{R}^M$ valued function over $[0,T]$. Then we have the following estimate, for any $a > 0$,

$$\begin{aligned}\sup_{t\in[0,T]} \|\boldsymbol{z}(t)\|^2_{-1,M} &+ \int_{Q_T} \sigma_M(\boldsymbol{z}\odot\boldsymbol{\mu}^{1/2})(s,x)^2\,\mathrm{d}s\,\mathrm{d}x \\ &\leq (1+a)^2\left[\|\boldsymbol{z}(0)\|^2_{-1,M} + [\boldsymbol{z}(0)]^2_M \int_0^T [\boldsymbol{\mu}(s)]_M \mathrm{d}s + \frac{1}{\alpha}\int_{Q_T}\sigma_M(\boldsymbol{f})(s,x)^2\,\mathrm{d}s\,\mathrm{d}x\right] \\ &\quad + (1+a)(1+a^{-1})\left(T + T\int_0^T [\boldsymbol{\mu}(s)]_M \mathrm{d}s + \frac{1}{\alpha}\right)\int_{Q_T}\sigma_M(\boldsymbol{x}_r')(s,x)^2\,\mathrm{d}s\,\mathrm{d}x \\ &\quad + C(1+a^{-1})\Big[\sup_{t\in[0,T]}\|\boldsymbol{x}_d(t)\|^2_{-1,M} + \int_0^T [\boldsymbol{\mu}(s)]_M[\boldsymbol{x}_d(s)]^2_M\,\mathrm{d}s\Big], \end{aligned} \tag{2.4.18}$$

where $C > 0$ is a constant.

Proof. Let's define $\boldsymbol{z}_r \in C^1([0,T];\mathbb{R}^M)$ as the unique solution of

$$\boldsymbol{z}_r(t) = \boldsymbol{z}(0) + \int_0^t \Delta_M\Big[\boldsymbol{z}_r(s) \odot \boldsymbol{\mu}(s) + \boldsymbol{f}(s)\Big]\,\mathrm{d}s + \boldsymbol{x}_r(t),$$

which, since $\boldsymbol{x}_r$ is continuously differentiable, is equivalent to the Cauchy problem

$$\boldsymbol{z}_r'(t) = \Delta_M\Big[\boldsymbol{z}_r(t)\odot\boldsymbol{\mu}(t) + \boldsymbol{f}(t)\Big]\,\mathrm{d}s + \boldsymbol{x}_r'(t), \tag{2.4.19}$$

$$\boldsymbol{z}_r(0) = \boldsymbol{z}(0). \tag{2.4.20}$$

Now, defining $\boldsymbol{z}_d := \boldsymbol{z} - \boldsymbol{z}_r$, one readily checks that it solves

$$\boldsymbol{z}_d(t) = \int_0^t \Delta_M\Big[\boldsymbol{z}_d(s)\odot\boldsymbol{\mu}(s)\Big]\,\mathrm{d}s + \boldsymbol{x}_d(t).$$

The Cauchy problem (2.4.19) – (2.4.20) is exactly the one of Lemma 2.4.3, with $\boldsymbol{r}(t) := \boldsymbol{x}_r'(t)$, we therefore infer from this very lemma, for any $a > 0$

$$\begin{aligned}\sup_{t\in[0,T]} \|\boldsymbol{z}_r(t)\|^2_{-1,M} &+ \int_{Q_T} \sigma_M(\boldsymbol{z}_r\odot\boldsymbol{\mu}^{1/2})(s,x)^2\,\mathrm{d}s\,\mathrm{d}x \\ &\leq (1+a)\left[\|\boldsymbol{z}(0)\|^2_{-1,M} + [\boldsymbol{z}(0)]^2_M \int_0^T [\boldsymbol{\mu}(s)]_M\,\mathrm{d}s + \frac{1}{\alpha}\int_{Q_T}\sigma_M(\boldsymbol{f})(s,x)^2\,\mathrm{d}s\,\mathrm{d}x\right] \\ &\quad + (1+a^{-1})\left(T + T\int_0^T [\boldsymbol{\mu}(s)]_M\,\mathrm{d}s + \frac{1}{\alpha}\right)\int_{Q_T}\sigma_M(\boldsymbol{x}_r')(s,x)^2\,\mathrm{d}s\,\mathrm{d}x. \end{aligned}$$

Now, since $\boldsymbol{z} = \boldsymbol{z}_d + \boldsymbol{z}_r$, combining the triangular inequality and the convex inequality $(x+y)^2 \leq (1+a)x^2 + (1+a^{-1})y^2$, implies

$$\begin{aligned}
&\sup_{t\in[0,T]} \|\boldsymbol{z}(t)\|_{-1,M}^2 + \int_{Q_T} \sigma_M(\boldsymbol{z}\odot\boldsymbol{\mu}^{1/2})(s,x)^2 \,\mathrm{d}s\,\mathrm{d}x \\
&\qquad \leq (1+a)\left[\sup_{t\in[0,T]} \|\boldsymbol{z}_r(t)\|_{-1,M}^2 + \int_{Q_T} \sigma_M(\boldsymbol{z}_r\odot\boldsymbol{\mu}^{1/2})(s,x)^2 \,\mathrm{d}s\,\mathrm{d}x\right] \\
&\qquad\qquad + (1+a^{-1})\left[\sup_{t\in[0,T]} \|\boldsymbol{z}_d(t)\|_{-1,M}^2 + \int_{Q_T} \sigma_M(\boldsymbol{z}_d\odot\boldsymbol{\mu}^{1/2})(s,x)^2 \,\mathrm{d}s\,\mathrm{d}x\right],
\end{aligned}$$

so that the proof follows from Lemma 2.4.4, which focuses on the non-regular component. □

2.5 Quantitative estimates and proof of Theorem 2.2

Let u, v be a $C^1([0,T];C^4(\mathbb{T}))$ solution of the system (2.1.1). We have, by Taylor expansion, for any $h > 0$ and $C^4(\mathbb{T})$ function f

$$\begin{aligned}
\tau_h f &= f + hf' + \frac{h^2}{2!}f'' + \frac{h^3}{3!}f''' + \mathrm{O}_{h\to 0}(h^4), \\
\tau_{-h} f &= f - hf' + \frac{h^2}{2!}f'' - \frac{h^3}{3!}f''' + \mathrm{O}_{h\to 0}(h^4),
\end{aligned}$$

where $\mathrm{O}_{h\to 0}$ refers to the $L^\infty(Q_T)$ norm. We have therefore

$$\frac{\tau_h f + \tau_{-h} f - 2f}{h^2} = f'' + \mathrm{O}_{h\to 0}(h^2).$$

In particular, denoting by $\widehat{\boldsymbol{u}}^M(t)$ and $\widehat{\boldsymbol{v}}^M(t)$ the respectives values of u and v at the points (t, x_k) for $k = 1, \ldots, M$, we have the following discrete system:

$$\begin{aligned}
\partial_t \widehat{\boldsymbol{u}}^M(t) &= \Delta_M\Big[d_1\widehat{\boldsymbol{u}}^M(t) + a_{12}\widehat{\boldsymbol{u}}^M(t)\odot\widehat{\boldsymbol{v}}^M(t)\Big] + \boldsymbol{r}^M(t), \\
\partial_t \widehat{\boldsymbol{v}}^M(t) &= \Delta_M\Big[d_1\widehat{\boldsymbol{v}}^M(t) + a_{21}\widehat{\boldsymbol{v}}^M(t)\odot\widehat{\boldsymbol{u}}^M(t)\Big] + \boldsymbol{s}^M(t),
\end{aligned}$$

with

$$\|\boldsymbol{r}^M(t)\|_\infty + \|\boldsymbol{s}^M(t)\|_\infty \lesssim M^{-2}, \tag{2.5.1}$$

uniformly for t on compact intervals.

On the other hand, we recall that our stochastic process satisfies

$$\begin{aligned}
\boldsymbol{U}^{M,N}(t) &= \boldsymbol{U}^{M,N}(0) + \int_0^t \Delta_M\Big(d_1\boldsymbol{U}^{M,N}(s) + a_{12}\boldsymbol{U}^{M,N}(s)\odot\boldsymbol{V}^{M,N}(s)\Big)\,\mathrm{d}s + \boldsymbol{\mathcal{M}}^{M,N}(t), \\
\boldsymbol{V}^{M,N}(t) &= \boldsymbol{V}^{M,N}(0) + \int_0^t \Delta_M\Big(d_2\boldsymbol{V}^{M,N}(s) + a_{21}\boldsymbol{U}^{M,N}(s)\odot\boldsymbol{V}^{M,N}(s)\Big)\,\mathrm{d}s + \boldsymbol{\mathcal{N}}^{M,N}(t),
\end{aligned}$$

where $\boldsymbol{\mathcal{M}}^{M,N}$ is square integrable martingale whose quadratic variation is given by (2.3.3) and $\boldsymbol{\mathcal{N}}^{M,N}$ satisfies similar properties. By symmetry, we can focus on the first species $\boldsymbol{U}^{M,N}$. Denoting

$$\boldsymbol{Z}^{M,N}(t) = \widehat{\boldsymbol{u}}^M(t) - \boldsymbol{U}^{M,N}(t), \quad \boldsymbol{X}^{M,N}(t) = \int_0^t \boldsymbol{r}^M(s)\,\mathrm{d}s - \boldsymbol{\mathcal{M}}^{M,N}(t),$$

we have yet another system satisfied by these quantities

$$\boldsymbol{Z}^{M,N}(t) = \boldsymbol{Z}^{M,N}(0) + \int_0^t \Delta_M\Big(\boldsymbol{Z}^{M,N}(s) \odot \boldsymbol{\Lambda}^{M,N}(s) + \boldsymbol{F}^{M,N}(s)\Big)\,\mathrm{d}s + \boldsymbol{X}^{M,N}(t), \tag{2.5.2}$$

where

$$\begin{aligned}
\boldsymbol{\Lambda}^{M,N}(t) &= d_1\mathbf{1}_M + a_{12}\boldsymbol{V}^{M,N}(t),\\
\boldsymbol{W}^{M,N}(t) &= \widehat{\boldsymbol{v}}^M(t) - \boldsymbol{V}^{M,N}(t),\\
\boldsymbol{F}^{M,N}(t) &= a_{12}\widehat{\boldsymbol{u}}^M \odot \boldsymbol{W}^{M,N}(t).
\end{aligned}$$

We can now apply the discrete duality lemma developed in the previous section to control the gap $\boldsymbol{Z}^{M,N}$. This is the core of the next result, which yields Theorem 2.2. For $\boldsymbol{z}\colon [0,T] \to \mathbb{R}^M$, let

$$|||\boldsymbol{z}|||_{T,M} \coloneqq \left(\sup_{t\in[0,T]} \|\boldsymbol{z}(t)\|_{-1,M}^2 + \|\sigma_M(\boldsymbol{z})\|_{L^2(Q_T)}^2\right)^{1/2}.$$

Proposition 2.5.1. *Let u, v be a solution of C^1 regularity in time and C^4 regularity in space of the system (2.1.1). If*

$$\frac{a_{12}a_{21}}{d_1 d_2}\|u\|_{L^\infty(Q_T)}\|v\|_{L^\infty(Q_T)} < 1,$$

then for any $(M,N) \in \mathbb{N}^2$ such that N/M^2 is large enough

$$\begin{aligned}
&\mathbb{E}\Big(|||\boldsymbol{Z}^{M,N}|||_{T,M}^2 + |||\boldsymbol{W}^{M,N}|||_{T,M}^2\Big)\\
&\quad\lesssim |||\boldsymbol{Z}^{M,N}(0)|||_{T,M}^2(1 + [\boldsymbol{V}^{M,N}(0)]_M) + |||\boldsymbol{W}^{M,N}(0)|||_{T,M}^2(1 + [\boldsymbol{U}^{M,N}(0)]_M)\\
&\qquad + \Big(1 + T^2 + T^2[\boldsymbol{U}^{M,N}(0) + \boldsymbol{V}^{M,N}(0)]_M\Big)\,M^{-4} + TM^2N^{-1}.
\end{aligned} \tag{2.5.3}$$

Proof. We first observe that $t \mapsto [\boldsymbol{\Lambda}^{M,N}(t)]_M$ is constant and we set

$$\begin{aligned}
\boldsymbol{\lambda}_T^{M,N} &= T + T\int_0^T [\boldsymbol{\Lambda}^{M,N}(s)]_M\mathrm{d}s + \frac{1}{d_1}\\
&= T + T^2(d_1 + a_{12}[\boldsymbol{V}^{M,N}(0)]_M) + \frac{1}{d_1}.
\end{aligned}$$

By applying Proposition 2.4.7 with $\boldsymbol{x}_d \coloneqq -\boldsymbol{\mathcal{M}}^{M,N}$ and

$$\boldsymbol{x}_r\colon t \mapsto \int_0^t \boldsymbol{r}^M(\sigma)\,\mathrm{d}\sigma,$$

we obtain for any $a > 0$ that

$$
\begin{aligned}
&\sup_{t\in[0,T]} \|\boldsymbol{Z}^{M,N}(t)\|_{-1,M}^2 + \int_{Q_T} \sigma_M\Big(\boldsymbol{Z}^{M,N} \odot (\boldsymbol{\Lambda}^{M,N})^{1/2}\Big)(s,x)^2 \,\mathrm{d}s\mathrm{d}x \\
&\qquad \le (1+a)^2\Big(\|\boldsymbol{Z}^{M,N}(0)\|_{-1,M}^2 + [\boldsymbol{Z}^{M,N}(0)]_M^2 \int_0^T [\boldsymbol{\Lambda}^{M,N}(s)]_M \,\mathrm{d}s \\
&\qquad\quad + \frac{1}{d_1}\int_{Q_T} \sigma_M(\boldsymbol{F}^{M,N})(s,x)^2 \,\mathrm{d}s\mathrm{d}x\Big) + (1+a)(1+a^{-1})\boldsymbol{\lambda}_T^{M,N} \int_{Q_T} \sigma_M(\boldsymbol{r}^M)(s,x)^2 \,\mathrm{d}s\mathrm{d}x \\
&\qquad\quad + C(1+a^{-1})\Big(\sup_{t\in[0,T]} \|\boldsymbol{\mathcal{M}}^{M,N}(t)\|_{-1,M}^2 + \int_0^T [\boldsymbol{\Lambda}^{M,N}(s)]_M [\boldsymbol{\mathcal{M}}^{M,N}(s)]_M^2 \,\mathrm{d}s\Big), \quad (2.5.4)
\end{aligned}
$$

for some constant $C > 0$. Moreover, since we have that $\Lambda_i^{M,N} \ge d_1$ and $|\sigma_M(\boldsymbol{F}^{M,N})(s,x)| \le a_{12}\|u\|_{L^\infty(Q_T)}|\sigma_M(\boldsymbol{W}^{M,N})(s,x)|$, as $\hat{\boldsymbol{u}}^M$ takes the values of u in the grid, we obtain

$$
\begin{aligned}
&\frac{1}{d_1}\sup_{t\in[0,T]} \|\boldsymbol{Z}^{M,N}(t)\|_{-1,M}^2 + \int_{Q_T} \sigma_M(\boldsymbol{Z}^{M,N})(s,x)^2 \,\mathrm{d}s\mathrm{d}x \\
&\qquad \le \frac{(1+a)^2}{d_1}\Big(\|\boldsymbol{Z}^{M,N}(0)\|_{-1,M}^2 + T[\boldsymbol{Z}^{M,N}(0)]_M^2[\boldsymbol{\Lambda}^{M,N}(0)]_M \\
&\qquad\quad + \frac{(a_{12}\|u\|_{L^\infty(Q_T)})^2}{d_1}\int_{Q_T} \sigma_M(\boldsymbol{W}^{M,N})(s,x)^2 \,\mathrm{d}s\mathrm{d}x\Big) \\
&\qquad\quad + \frac{1}{d_1}(1+a)(1+a^{-1})\boldsymbol{\lambda}_T^{M,N} \int_{Q_T} \sigma_M(\boldsymbol{r}^M)(s,x)^2 \,\mathrm{d}s\mathrm{d}x \\
&\qquad\quad + \frac{C}{d_1}(1+a^{-1})\Big(\sup_{t\in[0,T]} \|\boldsymbol{\mathcal{M}}^{M,N}(t)\|_{-1,M}^2 + \int_0^T [\boldsymbol{\Lambda}^{M,N}(s)]_M[\boldsymbol{\mathcal{M}}^{M,N}(s)]_M^2 \,\mathrm{d}s\Big).
\end{aligned}
$$

As the roles of $\boldsymbol{Z}^{M,N}$ and $\boldsymbol{W}^{M,N}$ are symmetric in the previous inequality, we have a similar estimate for $\boldsymbol{W}^{M,N}$. Thus, by setting

$$
\boldsymbol{\Gamma}^{M,N}(t) = d_2 + a_{21}\boldsymbol{U}^{M,N}(t),
$$

and

$$
\begin{aligned}
\boldsymbol{\gamma}_T^{M,N} &= T + T\int_0^T [\boldsymbol{\Gamma}^{M,N}(s)]_M \mathrm{d}s + \frac{1}{d_2} \\
&= T + T^2(d_2 + a_{21}[\boldsymbol{U}^{M,N}(0)]_M) + \frac{1}{d_2},
\end{aligned}
$$

we get

$$
\begin{aligned}
&\frac{1}{d_2}\sup_{t\in[0,T]} \|\boldsymbol{W}^{M,N}(t)\|_{-1,M}^2 + \int_{Q_T} \sigma_M(\boldsymbol{W}^{M,N})(s,x)^2 \,\mathrm{d}s\mathrm{d}x \\
&\qquad \le \frac{(1+a)^2}{d_2}\Big(\|\boldsymbol{W}^{M,N}(0)\|_{-1,M}^2 + T[\boldsymbol{W}^{M,N}(0)]_M^2[\boldsymbol{\Gamma}^{M,N}(0)]_M \\
&\qquad\quad + \frac{(a_{21}\|v\|_{L^\infty(Q_T)})^2}{d_2}\int_{Q_T} \sigma_M(\boldsymbol{Z}^{M,N})(s,x)^2 \,\mathrm{d}s\mathrm{d}x\Big) \\
&\qquad\quad + \frac{1}{d_2}(1+a)(1+a^{-1})\boldsymbol{\gamma}_T^{M,N} \int_{Q_T} \sigma_M(\boldsymbol{s}^M)(s,x)^2 \,\mathrm{d}s\mathrm{d}x \\
&\qquad\quad + \frac{C}{d_2}(1+a^{-1})\Big(\sup_{t\in[0,T]} \|\boldsymbol{\mathcal{N}}^{M,N}(t)\|_{-1,M}^2 + \int_0^T [\boldsymbol{\Gamma}^{M,N}(s)]_M[\boldsymbol{\mathcal{N}}^{M,N}(s)]_M^2 \,\mathrm{d}s\Big).
\end{aligned}
$$

Plugging now this inequality in the estimate for $\boldsymbol{Z}^{M,N}$ gives us

$$\begin{aligned}
&\frac{1}{d_1}\sup_{t\in[0,T]}\|\boldsymbol{Z}^{M,N}(t)\|_{-1,M}^2+\int_{Q_T}\sigma_M(\boldsymbol{Z}^{M,N})(s,x)^2\,\mathrm{d}s\mathrm{d}x\\
&\quad\le\frac{(1+a)^2}{d_1}\Big(\|\boldsymbol{Z}^{M,N}(0)\|_{-1,M}^2+T[\boldsymbol{Z}^{M,N}(0)]_M^2[\boldsymbol{\Lambda}^{M,N}(0)]_M\Big)\\
&\qquad+\frac{(1+a)^4}{d_2}\Big(\frac{a_{12}\|u\|_{L^\infty(Q_T)}}{d_1}\Big)^2\Big(\|\boldsymbol{W}^{M,N}(0)\|_{-1,M}^2+T[\boldsymbol{W}^{M,N}(0)]_M^2[\boldsymbol{\Gamma}^{M,N}(0)]_M\Big)\\
&\qquad+(1+a)^4\Big(\frac{a_{12}a_{21}\|u\|_{L^\infty(Q_T)}\|v\|_{L^\infty(Q_T)}}{d_1d_2}\Big)^2\int_{Q_T}\sigma_M(\boldsymbol{Z}^{M,N})(s,x)^2\,\mathrm{d}s\mathrm{d}x\\
&\qquad+\frac{(1+a)^2}{d_2}\Big(\frac{a_{12}\|u\|_{L^\infty(Q_T)}}{d_1}\Big)^2(1+a^{-1})\Big((1+a)\boldsymbol{\gamma}_T^{M,N}\int_{Q_T}\sigma_M(\boldsymbol{s}^M)(s,x)^2\,\mathrm{d}s\mathrm{d}x\\
&\qquad+C\Big(\sup_{t\in[0,T]}\|\boldsymbol{\mathcal{N}}^{M,N}(t)\|_{-1,M}^2+\int_0^T[\boldsymbol{\Gamma}^{M,N}(s)]_M[\boldsymbol{\mathcal{N}}^{M,N}(s)]_M^2\,\mathrm{d}s\Big)\Big)\\
&\qquad+\frac{1}{d_1}(1+a)(1+a^{-1})\boldsymbol{\lambda}_T^{M,N}\int_{Q_T}\sigma_M(\boldsymbol{r}^M)(s,x)^2\,\mathrm{d}s\mathrm{d}x\\
&\qquad+\frac{C}{d_1}(1+a^{-1})\Big(\sup_{t\in[0,T]}\|\boldsymbol{\mathcal{M}}^{M,N}(t)\|_{-1,M}^2+\int_0^T[\boldsymbol{\Lambda}^{M,N}(s)]_M[\boldsymbol{\mathcal{M}}^{M,N}(s)]_M^2\,\mathrm{d}s\Big).
\end{aligned}$$

By using our assumption on the bound of $\|u\|_{L^\infty(Q_T)}\|v\|_{L^\infty(Q_T)}$ and then fixing a to be small enough, we can absorb the third term on the r.h.s. of the previous inequality in the l.h.s. in order to recover the definition of $|||\cdot|||_{T,M}$. Thus, letting $\lesssim$ to depend on these parameters, this yields

$$\begin{aligned}
|||\boldsymbol{Z}^{M,N}|||_{T,M}^2\lesssim{}&\|\boldsymbol{Z}^{M,N}(0)\|_{-1,M}^2+T[\boldsymbol{Z}^{M,N}(0)]_M^2[\boldsymbol{\Lambda}^{M,N}(0)]_M\\
&+\|\boldsymbol{W}^{M,N}(0)\|_{-1,M}^2+T[\boldsymbol{W}^{M,N}(0)]_M^2[\boldsymbol{\Gamma}^{M,N}(0)]_M\\
&+\Big(\boldsymbol{\lambda}_T^{M,N}+\boldsymbol{\gamma}_T^{M,N}\Big)\int_{Q_T}\Big(\sigma_M(\boldsymbol{r}^M)(s,x)^2+\sigma_M(\boldsymbol{s}^M)(s,x)^2\Big)\,\mathrm{d}s\mathrm{d}x\\
&+\sup_{t\in[0,T]}\|\boldsymbol{\mathcal{M}}^{M,N}(t)\|_{-1,M}^2+\int_0^T[\boldsymbol{\Lambda}^{M,N}(s)]_M[\boldsymbol{\mathcal{M}}^{M,N}(s)]_M^2\,\mathrm{d}s\\
&+\sup_{t\in[0,T]}\|\boldsymbol{\mathcal{N}}^{M,N}(t)\|_{-1,M}^2+\int_0^T[\boldsymbol{\Gamma}^{M,N}(s)]_M[\boldsymbol{\mathcal{N}}^{M,N}(s)]_M^2\,\mathrm{d}s.
\end{aligned}$$

The previous r.h.s. is again invariant with respect to the roles of $\boldsymbol{Z}^{M,N}$ and $\boldsymbol{W}^{M,N}$. Then using the uniform bounds on $\sigma_M(\boldsymbol{r}^M)$ and $\sigma_M(\boldsymbol{s}^M)$ from (2.5.1) and taking expectation, we get

$$\begin{aligned}
&\mathbb{E}\Big(|||\boldsymbol{Z}^{M,N}|||_{T,M}^2+|||\boldsymbol{W}^{M,N}|||_{T,M}^2\Big)\\
&\quad\lesssim\|\boldsymbol{Z}^{M,N}(0)\|_{-1,M}^2+T[\boldsymbol{Z}^{M,N}(0)]_M^2[\boldsymbol{\Lambda}^{M,N}(0)]_M\\
&\qquad+\|\boldsymbol{W}^{M,N}(0)\|_{-1,M}^2+T[\boldsymbol{W}^{M,N}(0)]_M^2[\boldsymbol{\Gamma}^{M,N}(0)]_M+\Big(\boldsymbol{\lambda}_T^{M,N}+\boldsymbol{\gamma}_T^{M,N}\Big)M^{-4}\\
&\qquad+\mathbb{E}\Big(\sup_{t\in[0,T]}\|\boldsymbol{\mathcal{M}}^{M,N}(t)\|_{-1,M}^2\Big)+[\boldsymbol{\Lambda}^{M,N}(0)]_M\int_0^T\mathbb{E}\Big([\boldsymbol{\mathcal{M}}^{M,N}(s)]_M^2\Big)\,\mathrm{d}s\\
&\qquad+\mathbb{E}\Big(\sup_{t\in[0,T]}\|\boldsymbol{\mathcal{N}}^{M,N}(t)\|_{-1,M}^2\Big)+[\boldsymbol{\Gamma}^{M,N}(0)]_M\int_0^T\mathbb{E}\Big([\boldsymbol{\mathcal{N}}^{M,N}(s)]_M^2\Big)\,\mathrm{d}s.
\end{aligned}\tag{2.5.5}$$

We are left then with controlling the local martingale terms that appear at the end. Since

$$[\boldsymbol{\mathcal{M}}^{M,N}(s)]_M^2 = \Big(\frac{1}{M}\sum_{i=1}^M \mathcal{M}_i^{M,N}(s)\Big)^2$$
$$\leq \frac{1}{M}\sum_{i=1}^M \mathcal{M}_i^{M,N}(s)^2 \leq \frac{1}{M}\sum_{i=1}^M \sup_{t\in[0,T]} \mathcal{M}_i^{M,N}(t)^2,$$

and recalling that $[\boldsymbol{\Lambda}^{M,N}(0)]_M = [d_1 + a_{12}\boldsymbol{V}^{M,N}(0)]_M$, we have

$$[\boldsymbol{\Lambda}^{M,N}(0)]_M \int_0^T \mathbb{E}\Big([\boldsymbol{\mathcal{M}}^{M,N}(s)]_M^2\Big)\,\mathrm{d}s$$
$$\leq (d_1 + a_{12}\|\boldsymbol{V}^{M,N}(0)\|_{1,M})\frac{T}{M}\sum_{i=1}^M \mathbb{E}\Big(\sup_{t\in[0,T]} \mathcal{M}_i^{M,N}(t)^2\Big).$$

Besides, we also have using (2.4.8)

$$\mathbb{E}\Big(\sup_{t\in[0,T]} \|\boldsymbol{\mathcal{M}}^{M,N}(t)\|_{-1,M}^2\Big) \leq \mathbb{E}\Big(\sup_{t\in[0,T]} \|\boldsymbol{\mathcal{M}}^{M,N}(t)\|_{2,M}^2\Big)$$
$$\leq \mathbb{E}\Big(\sup_{t\in[0,T]} \Big(\frac{1}{M}\sum_{i=1}^M \mathcal{M}_i^{M,N}(t)^2\Big)\Big)$$
$$\leq \frac{1}{M}\sum_{i=1}^M \mathbb{E}\Big(\sup_{t\in[0,T]} \mathcal{M}_i^{M,N}(t)^2\Big).$$

Now, Doob's inequality ensures that $\mathbb{E}\Big(\sup_{t\in[0,T]} \mathcal{M}_i^{M,N}(t)^2\Big) \lesssim \mathbb{E}(\langle \mathcal{M}_i^{M,N}\rangle(T))$, where the expression of the quadratic variation $\langle \mathcal{M}_i^{M,N}\rangle$ is found in (2.3.3). Moreover, since

$$\sum_{i=1}^M \Big(2U_i^{M,N}(s)V_i^{M,N}(s) + U_{i+1}^{M,N}(s)V_{i+1}^{M,N}(s) + U_{i-1}^{M,N}(s)V_{i-1}^{M,N}(s)\Big)$$
$$\lesssim \sum_{i=1}^M \Big(U_i^{M,N}(s)^2 + V_i^{M,N}(s)^2\Big)$$
$$= \|\boldsymbol{U}^{M,N}(t)\|_2^2 + \|\boldsymbol{V}^{M,N}(t)\|_2^2,$$

we get

$$\frac{1}{M}\sum_{i=1}^M \mathbb{E}\Big(\sup_{t\in[0,T]} \mathcal{M}_i^{M,N}(t)^2\Big)$$
$$\lesssim \frac{1}{M}\mathbb{E}\Big(\frac{M^2}{N}\int_0^T \|\boldsymbol{U}^{M,N}(s)\|_1\,\mathrm{d}s + \frac{M^2}{N}\int_0^T \Big(\|\boldsymbol{U}^{M,N}(s)\|_2^2 + \|\boldsymbol{V}^{M,N}(s)\|_2^2\Big)\,\mathrm{d}s\Big)$$

Moreover $\|\boldsymbol{U}^{M,N}(s)\|_1 = \|\boldsymbol{U}^{M,N}(0)\|_1$ a.s. and we recall that $\boldsymbol{U}^{M,N}(t) = \widehat{\boldsymbol{u}}^M(t) - \boldsymbol{Z}^{M,N}(t)$ and $\boldsymbol{V}^{M,N}(t) = \widehat{\boldsymbol{v}}^M(t) - \boldsymbol{W}^{M,N}(t)$ for any $s \geq 0$. Adding that boundedness assumption on the solution of the SKT system and (2.3.4) ensure that

$$T\frac{M^2}{N}\|\boldsymbol{U}^{M,N}(0)\|_{1,M} + \frac{M^2}{N}\int_{Q_T} \sigma_M\Big(\widehat{\boldsymbol{u}}^M\Big)^2 + \sigma_M\Big(\widehat{\boldsymbol{v}}^M\Big)^2 = T\mathcal{O}\Big(\frac{M^2}{N}\Big),$$

we finally have

$$\begin{aligned}
&\frac{1}{M}\sum_{i=1}^{M}\mathbb{E}\Big(\sup_{t\in[0,T]}\mathcal{M}_i^{M,N}(t)^2\Big)\\
&\qquad\lesssim \frac{M}{N}\int_0^T \mathbb{E}\Big(\|\boldsymbol{Z}^{M,N}(s)\|_2^2+\|\boldsymbol{W}^{M,N}(s)\|_2^2\Big)\,\mathrm{d}s+T\frac{M^2}{N}\\
&\qquad\lesssim \frac{M^2}{N}\int_{Q_T}\mathbb{E}\left(\sigma_M\Big(\boldsymbol{Z}^{M,N}\Big)(s,x)^2+\sigma_M\Big(\boldsymbol{W}^{M,N}\Big)(s,x)^2\right)\,\mathrm{d}s\mathrm{d}x+T\frac{M^2}{N}\\
&\qquad\lesssim \frac{M^2}{N}|||\boldsymbol{Z}^{M,N}|||_{T,M}^2+\frac{M^2}{N}|||\boldsymbol{W}^{M,N}|||_{T,M}^2+T\frac{M^2}{N}.
\end{aligned}$$

By symmetry we have bounds of the same order for the terms involving $(\boldsymbol{\mathcal{N}}^{M,N}(t))_{t\geq 0}$. We plug these bounds in (2.5.5) and gather the terms $|||\boldsymbol{Z}^{M,N}|||_{T,M}^2$ and $|||\boldsymbol{W}^{M,N}|||_{T,M}^2$ in the left hand side. For N/M^2 large enough, we can contol the left hand side and get

$$\begin{aligned}
&\mathbb{E}(|||\boldsymbol{Z}^{M,N}|||_{T,M}^2+|||\boldsymbol{W}^{M,N}|||_{T,M}^2)\\
&\qquad\lesssim \|\boldsymbol{Z}^{M,N}(0)\|_{-1,M}^2+T[\boldsymbol{Z}^{M,N}(0)]_M^2[\boldsymbol{\Lambda}^{M,N}(0)]_M\\
&\qquad\quad+\|\boldsymbol{W}^{M,N}(0)\|_{-1,M}^2+T[\boldsymbol{W}^{M,N}(0)]_M^2[\boldsymbol{\Gamma}^{M,N}(0)]_M\\
&\qquad\quad+\left(\boldsymbol{\lambda}_T^{M,N}+\boldsymbol{\gamma}_T^{M,N}\right)M^{-4}+T\left(1+[\boldsymbol{\Lambda}^{M,N}(0)+\boldsymbol{\Gamma}^{M,N}(0)]_M\right)\frac{M^2}{N}.
\end{aligned}$$

Using that $T[\boldsymbol{u}]_M^2\leq\|\sigma_M(\boldsymbol{u})\|_{L^2(Q_T)}^2$ for any $\boldsymbol{u}\in\mathbb{R}^M$, by rearranging the terms we conclude the proof. □

Now we can prove the remaining main result.

Proof of Theorem 2.2. We have

$$\begin{aligned}
\zeta^{M,N}&:=\pi_M(\boldsymbol{U}^{M,N})-u\\
&=\pi_M(\boldsymbol{U}^{M,N}-\widehat{\boldsymbol{u}}^M)+\pi_M(\widehat{\boldsymbol{u}}^M)-u=\pi_M(\boldsymbol{Z}^{M,N})+\iota_M(u)-u,
\end{aligned}$$

where the interpolation operator ι_M is the one used in Lemma 2.4.2. Using the triangular inequality, we infer

$$\begin{aligned}
&\mathbb{E}\Big[\sup_{t\in[0,T]}\|\zeta^{M,N}(t)\|_{\dot H^{-1}(\mathbb{T})}^2+\|\zeta^{M,N}\|_{L^2(Q_T)}^2\Big]\\
&\qquad\leq\mathbb{E}\Big[\sup_{t\in[0,T]}\|\pi_M(\boldsymbol{Z}^{M,N})(t)\|_{\dot H^{-1}(\mathbb{T})}^2+\|\pi_M(\boldsymbol{Z}^{M,N})\|_{L^2(Q_T)}^2\Big]\\
&\qquad\quad+\sup_{t\in[0,T]}\|\iota_M(u)-u\|_{\dot H^{-1}(\mathbb{T})}^2+\|\iota_M(u)-u\|_{L^2(Q_T)}^2.
\end{aligned}\tag{2.5.6}$$

Now, using Proposition 2.4.1 we have that $\|\pi_M(\boldsymbol{Z}^{M,N})\|_{L^2(Q_T)}\leq\|\sigma_M(\boldsymbol{Z}^{M,N})\|_{L^2(Q_T)}$, and using the equivalence (2.4.7) of Proposition 2.4.5 we get for all $t\in[0,T]$

$$\|\pi_M(\boldsymbol{Z}^{M,N})(t)\|_{H^{-1}(\mathbb{T})}\lesssim\|\boldsymbol{Z}^{M,N}(t)\|_{-1,M}+M^{-1}\|\pi_M(\boldsymbol{Z}^{M,N}(t))\|_{L^2(\mathbb{T})}.$$

This means that the expectation term in the r.h.s. of (2.5.6) satisfies the following bound for $M \geq 1$

$$\mathbb{E}\Big[\sup_{t\in[0,T]} \|\pi_M(\boldsymbol{Z}^{M,N})(t)\|^2_{\dot{H}^{-1}(\mathbb{T})} + \|\pi_M(\boldsymbol{Z}^{M,N})\|^2_{L^2(Q_T)}\Big] \lesssim_T \mathbb{E}\Big[|||\boldsymbol{Z}^{M,N}|||^2_{T,M}\Big].$$

All in all, using Proposition 2.5.1, we get

$$\mathbb{E}\Big[\sup_{t\in[0,T]} \|\pi_M(\boldsymbol{Z}^{M,N})(t)\|^2_{\dot{H}^{-1}(\mathbb{T})} + \|\pi_M(\boldsymbol{Z}^{M,N})\|^2_{L^2(Q_T)}\Big] \lesssim_T \varepsilon_{M,N},$$

where $\varepsilon_{M,N}$ is the r.h.s. of (2.5.3). Getting back to (2.5.6), we still have to control the second expectation term of its r.h.s., for which invoke Lemma 2.4.2 which allow us to write

$$\sup_{t\in[0,T]} \|\iota_M(u) - u\|^2_{\dot{H}^{-1}(\mathbb{T})} + \|\iota_M(u) - u\|^2_{L^2(Q_T)} \lesssim M^{-4}\|u\|^2_{L^\infty\cap L^2([0,T];H^2(\mathbb{T}))}.$$

Gathering all the terms leads to the conclusion. □

A Appendix: discrete–continuous dictionary

Discrete	Continuous
Δ_M	Δ
$\|\cdot\|_{p,M}$	$\|\cdot\|_{L^p(\mathbb{T})}$
$(\cdot\|\cdot)_M$	$(\cdot\|\cdot)_{L^2(\mathbb{T})}$
$\|\cdot\|_{-1,M}$	$\|\cdot\|_{H^{-1}(\mathbb{T})}$
$\|\|\|\cdot\|\|\|_{T,M}$	$\|\|\|\cdot\|\|\|_T$
$[\cdot]_M$	$[\cdot]_{\mathbb{T}}$

CHAPTER 3

Quantitative large-population asymptotics for mean-field interacting branching diffusions via optimal transport

Contents

This chapter is based on [59], written in collaboration with Joaquín Fontbona.

3.1 Introduction

Mathematical models of interacting and randomly evolving populations have been intensively studied the last decades through probabilistic and analytic approaches. Both points of view are able to integrate several biologically or ecologically meaningful features, including individuals' displacements, reproduction and deaths, competition for resources, selection, and dispersive or attractive interactions. While PDE and analysis methods can provide aggregate deterministic descriptions of the collective or macroscopic behavior of such populations (see [108, 23, 22, 46, 62] and [31], to name but a few works), probabilistic methods have successfully been employed to describe the random behaviors and interactions of individuals at the microscopic (or finite population) level, and to justify, in a rigorous way, that solutions of certain nonlinear evolution PDEs are the limits in law of the empirical measures of some individual-based models, when the population size goes to infinity (see for example [61, 7, 58, 28] and [27]). Nevertheless, although it is clear that certain law of large numbers for exchangeable systems lies beneath the passage from the microscopic to the macroscopic scale here as in other settings, the speed of this convergence is not explicitly known for branching population models, even in the simple case of pure binary branching diffusions.

In this work, we develop a probabilistic approach to obtain quantitative convergence estimates for the large population limit of a class of spatially branching diffusions with logistic growth and mean-field interactive spatial dynamics. Specifically, we assume that during its life-span, each individual's position evolves in $\mathbb{R}^d$ following the SDE

$$\mathrm{d}X_t^n = b\Big(X_t^n, H * \mu_t^K(X_t^n)\Big)\,\mathrm{d}t + \sigma\Big(X_t^n, G * \mu_t^K(X_t^n)\Big)\,\mathrm{d}B_t^n, \qquad n = 1, \ldots, N_t^K, \tag{3.1.1}$$

where $(B^n)_{n\geq 1}$ are independent Brownian motions in $\mathbb{R}^d$, and σ and b are regular coefficients depending on the position X_t^n and of the empirical measure $\mu_t^K = \frac{1}{K}\sum_{k=1}^{N_t^K} \delta_{X_t^k}$ of all N_t^K individuals alive at time t. The functions H and G are regular kernels controlling the strength and the range of the interaction of an individual with the population, through the empirical density of the latter (precise assumptions will be given later on). The parameter K measures the population size, and can be interpreted as the carrying capacity of the underlying environment (see [7]).

Furthermore, each individual gives birth to one offspring at its current position at constant rate $r > 0$ independently and, as a result of global competition, dies at rate cN_t^K/K, with $c \geq 0$ a fixed parameter. Additionally, a random number N_0^K of individuals can also be given birth at time 0 at random positions $X_0^n, n = 1, \ldots, N_0^K$, such that the corresponding empirical measure μ_0^K converges in law to some deterministic finite measure μ_0 on $\mathbb{R}^d$ as $K \to \infty$.

This model corresponds to a subclass of some non local Lotka-Volterra cross-diffusion systems, introduced in [58] as a microscopic, individual-based counterpart of the celebrated Shigesada-Kawasaki-Teramoto cross-diffusion system [108]. More precisely, in the model of [58], the competition for resources on one hand, and the spatial dispersion resulting from individuals repulsions or environmental conditions on the other, can take place at different macroscopic spatial ranges, and heterogeneously in space. In order to address the question of quantifying the large-population limit, we consider a simplified setting of one single species with self-interactions in the displacements, and we moreover assume that the demographic parameters determining the individual births and deaths are spatially homogeneous. In particular, the competition kernel of [58] is a constant here. This amounts to say that the competitive pressure is exerted on each individual simultaneously by the whole population alive, proportionally to its total size.

Following [58], when K goes to infinity, the empirical measure process $(\mu_t^K)_{t\in[0,T]}$ converges in law for each $T > 0$ (in the Skorokhod space of finite measure valued paths on $[0,T]$) towards a deterministic continuous measure-valued function $(\mu_t)_{t\in[0,T]}$, that is the unique weak solution of the non-local self-diffusion equation

$$\partial_t \mu_t = \frac{1}{2}\sum_{i,j=1}^d \partial^2_{x_i x_j}\Big(a^{(ij)}(\,\cdot\,, G * \mu_t)\mu_t\Big) - \sum_{i=1}^d \partial_{x_i}\Big(b^{(i)}(\,\cdot\,, H * \mu_t)\mu_t\Big) + (r - c\langle \mu_t, 1\rangle)\mu_t, \tag{3.1.2}$$

with $a = \sigma\sigma^{\mathrm{t}}$, and the initial condition μ_0. In this case, the total mass $n_t \coloneqq \langle \mu_t, 1\rangle$ of the finite measure μ_t evolves in time logistically: $\partial_t n_t = (r - cn_t)n_t$.

It is well known that convergence of the empirical probability distribution of N exchangeable particles to some deterministic probability measure, when N is a non-random integer that goes to infinity, is equivalent to the property of propagation of chaos, or asymptotic independence of the particles ([110, 95]). In order to establish an explicit convergence rate for the measure process $(\mu_t^K)_{t\in[0,T]}$ as K goes to infinity, we will extend to the branching populations setting probabilistic techniques developed to quantify propagation of chaos in particle systems arising in kinetic theory ([35, 36]). A natural strategy to establish bounds in that framework is by coupling the interacting particle system to a certain auxiliary system of particles with less (typically without) interactions between them, suitably constructed in the same probability space. It is then possible to derive non asymptotic bounds from quantitative estimates available for independent objects.

Generalizing this idea to interacting branching populations, we will construct a coupling of the system of interacting particles with logistic branching, with certain system of independent diffusions, with random births and deaths suitably distributed in time and space. The fundamental feature of this coupling, allowing us to put in place the above mentioned strategy, is that the random birth positions in the auxiliary system "mimic" the branching positions in the original system in the best possible way, in the sense of mean quadratic error. We are able to do this adapting to this setting optimal transport based techniques, developed in [35] to prove quantitative chaos estimates for particle systems with binary jumps. Our construction will thus allow us to transfer the rate of convergence in Wasserstein-2 distance of empirical measures of N i.i.d. samples, established in [60], to an analogous (in terms of K) convergence rate for the dual bounded-Lipschitz distance of the empirical processes $(\mu_t^K)_{t\in[0,T]}$. The ideas and techniques developed can in principle be refined and extended to more general systems of interacting branching populations, including the general setting of [58]. Nevertheless, this requires to deal with significant additional technicalities, and we have chosen to focus here on the basic ideas, leaving possible extensions for future work.

We next provide a detailed description of the population model we will consider, referring to [58] for additional background. We then state our assumptions and main result, and outline the paper's organization.

3.2 Model, notations and main result

The population and its evolution are described by a right-continuous measure-valued Markov process $(\mu_t^K)_{t\geq 0}$, taking values, for each fixed $K\in\mathbb{N}^*$, in the space of weighted finite point measures over $\mathbb{R}^d$:

$$\mathcal{M}^K(\mathbb{R}^d) := \left\{ \frac{1}{K}\sum_{n=1}^{N} \delta_{x^n} : x^n \in \mathbb{R}^d, N\in\mathbb{N} \right\} \subseteq \mathcal{M}^+(\mathbb{R}^d).$$

The notation $\mathcal{M}^+(\mathbb{R}^d)$ stands for the space of finite nonnegative measures on $\mathbb{R}^d$, endowed with the weak topology. Its subspace of probability measures is denoted by $\mathcal{P}(\mathbb{R}^d)$. The measure μ_t^K describing the population at time $t\geq 0$ is denoted by

$$\mu_t^K = \frac{1}{K}\sum_{n=1}^{N_t^K} \delta_{X_t^{n,K}}, \tag{3.2.1}$$

where $N_t^K := K\langle \mu_t^K, 1\rangle \in \mathbb{N}$ is the number of living individuals at time $t \geq 0$ and the variables $X_t^{1,K}, \dots, X_t^{N_t^K,K}$ are their positions in $\mathbb{R}^d$. In the sequel, we will simply write $\left(X_t^1, \dots, X_t^{N_t^K}\right) = \left(X_t^{1,K}, \dots, X_t^{N_t^K,K}\right)$ when working with fixed $K \geq 1$ and no ambiguity is possible.

The labeling $1, \dots, N_t^K$ of the atoms of μ_t^K will be assigned according to some dynamic rule, to be made explicit later (cf. Section 3.5), and will be such that the random vector $(X_t^1, \dots, X_t^{N_t^K})$ is exchangeable conditionally on N_t^K. The generator and thus the law of the measure valued Markov process $(\mu_t^K)_{t\geq 0}$ does nevertheless not depend on the chosen labeling rule.

The dynamics of $(\mu_t^K)_{t\geq 0}$ is summarized as follows:

- The initial population is represented by a random measure $\mu_0^K \in \mathcal{M}^K(\mathbb{R}^d)$.
- Each living individual carries at each instant $t > 0$ two clocks, independent between them and of the rest of the system: one reproduction clock, exponential of parameter $r > 0$, and one mortality clock, conditionally exponential of parameter cN_t^K/K given the population size N_t^K. If the reproduction clock of a particle rings at time t when at position x, it gives birth to a new particle at that same position. If the mortality clock rings the particle disappears. Equivalently, the process jumps from μ_{t-}^K to $\mu_t^K = \mu_{t-}^K + K^{-1}\delta_x$ in the first case and to $\mu_t^K = \mu_{t-}^K - K^{-1}\delta_x$ in the second.
- Between birth or death events, each individual $X_t^n, n = 1, ..., N_t$ evolves according to the diffusion processes (3.1.1), where $(B^n)_{n\geq 1}$ are Brownian motions in $\mathbb{R}^d$, independent between them and independent of μ_0^K and of the birth and deaths events.

The following conditions will be enforced in what follows.

Hypothesis $(\mathbf{H})$:

1. $(\langle \mu_0^K, 1\rangle)_K$ converges in law as $K \to \infty$ to some deterministic value in $(0, \infty)$. Moreover, for each $K \geq 1$, conditionally on $\langle \mu_0^K, 1\rangle$ the $N_0^K = K\langle \mu_0^K, 1\rangle$ atoms of μ_0^K are i.i.d. random variables with common law $\bar{\mu}_0 \in \mathcal{P}(\mathbb{R}^d)$ not depending on K.
2. The functions $\sigma \colon \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^{d\otimes d}$ and $b \colon \mathbb{R}^d \times \mathbb{R}_+ \to \mathbb{R}^d$ are Lipschitz. Moreover, there exists $C_\sigma > 0$ such that for each $x \in \mathbb{R}^d$ and $v \in \mathbb{R}_+$,
$$|\sigma(x, v)| \leq C_\sigma(1 + |v|).$$
3. The functions $G, H \colon \mathbb{R}^d \to \mathbb{R}$ are nonnegative, bounded and Lipschitz continuous.

Write $a := \sigma\sigma^{\mathrm{t}}$ and, given $\mu \in \mathcal{M}^+(\mathbb{R}^d)$, define a generator acting on $C^2(\mathbb{R}^d)$ functions ϕ by

$$L_\mu\phi(x) = \frac{1}{2}\mathrm{Tr}\left(a(x, G * \mu(x))\mathrm{Hess}(\phi)(x)\right) + b(x, H * \mu(x)) \cdot \nabla\phi(x).$$

Under assumption $(\mathbf{H})$, the process $(\mu_t^K)_{t\geq 0}$ has finitely many jumps in each finite time interval. Moreover, $(\mu_t^K)_{t\geq 0}$ is Markov with infinitesimal generator $\mathcal{L}^K$ given by

$$\mathcal{L}^K := \mathcal{L}_D^K + \mathcal{L}_J^K, \tag{3.2.2}$$

where $\mathcal{L}_D^K$ is the diffusion operator defined by

$$\mathcal{L}_D^K F(\nu) = \langle \nu, L_\nu \phi \rangle f'(\langle \nu, \phi \rangle) + \langle \nu, a(\cdot, G * \nu)[\phi']^2 \rangle f''(\langle \nu, \phi \rangle),$$

on functions $F\colon \mathcal{M}^K(\mathbb{R}^d) \to \mathbb{R}$ with the form $F(\nu) = f(\langle \nu, \phi \rangle)$, for $f \in C_b^2(\mathbb{R})$ and $\phi \in C^2(\mathbb{R}^d)$, and $\mathcal{L}_J^K$ is the jump operator

$$\mathcal{L}_J^K F(\nu) = rK \int \nu(dx) \left(F(\nu + K^{-1}\delta_x) - F(\nu) \right) \\ + c\langle \nu, 1 \rangle K \int \nu(dx) \left(F(\nu - K^{-1}\delta_x) - F(\nu) \right).$$

The above class of functions F is a core for the generator , and the law of $(\mu_t^K)_{t\geq 0}$ is uniquely determined by the operator $\mathcal{L}^K$. We refer to [58] for further details on the model and to [39] for background on measure-valued Markov processes.

Assumption $(\mathbf{H})\, 1)$ implies that $(\mu_0^K)_K$ converges in probability to the deterministic finite measure

$$\mu_0 := \lim_{K\to\infty} \langle \mu_0^K, 1 \rangle \bar{\mu}_0 \in \mathcal{M}^+(\mathbb{R}^d),$$

(see Section 3.3), and holds for instance if $K\mu_0^K$ is for each K a Poisson point measure of intensity $K\mu_0$, with $\mu_0 \in \mathcal{M}^+(\mathbb{R}^d)$ given. As a particular case of Theorem 3.1 in [58], we have the following result.

Theorem 3.1. *Assume* $(\mathbf{H})$ *and that for some* $p \geq 3$, $\sup_K \mathbb{E}(\langle \mu_0^K, 1 \rangle^p) < +\infty$. *The sequence of processes* $(\mu^K)_K$ *converges in law in* $D([0,T], \mathcal{M}^+(\mathbb{R}^d))$ *as* $K \to \infty$ *to the unique (deterministic) continuous finite measure valued function* $(\mu_t)_{t\in[0,T]}$ *solution of*

$$\langle \mu_t, f(t,\cdot) \rangle = \langle \mu_0, f(0,\cdot) \rangle + \int_0^t \Big\langle \mu_s, \partial_s f(s,\cdot) + L_{\mu_s} f(s,\cdot) + (r - c\langle \mu_s, 1 \rangle) f(s,\cdot) \Big\rangle \, \mathrm{d}s,$$

$\forall t \in [0,T]$ *and every* $f \in C_b^{1,2}([0,T] \times \mathbb{R}^d)$ *such that* $\sup_{(t,x)\in[0,T]\times\mathbb{R}^d} (1 + |x|)|\nabla f(t,x)| < \infty$.

Let now $\mathrm{BL}(\mathbb{R}^d)$ denote the space of Lipschitz-continuous bounded functions in $\mathbb{R}^d$ endowed with the norm

$$\|\varphi\|_{\mathrm{BL}} = \sup_{x\neq y} \frac{|\varphi(x) - \varphi(y)|}{\|x - y\|} + \sup_x |\varphi(x)|.$$

The corresponding dual norm $\|\cdot\|_{\mathrm{BL}^*}$ on the space $\mathcal{M}(\mathbb{R}^d)$ of finite signed measures on $\mathbb{R}^d$ induces the distance

$$\|\mu - \nu\|_{\mathrm{BL}^*} = \sup_{\|\varphi\|_{\mathrm{BL}} \leq 1} |\langle \mu - \nu, \varphi \rangle|,$$

on $\mathcal{M}^+(\mathbb{R}^d)$, which generates the weak convergence topology. Given a measure $\mu \in \mathcal{M}^+(\mathbb{R}^d)$, we denote its q-th moment for $q \in [1, \infty)$ by

$$M_q(\mu) = \int_{\mathbb{R}^d} |x|^q \, \mu(\mathrm{d}x).$$

For each $K \geq 1$ and $p \geq 1$, we define also

$$I_p(K) = \mathbb{E}\Big(\Big| \langle \mu_0^K, 1 \rangle - \langle \mu_0, 1 \rangle \Big|^p \Big)^{\frac{1}{p}}.$$

The following is our main result.

Theorem 3.2. *Assume* $(\mathbf{H})$*, that* $\sup_K \mathbb{E}(\langle \mu_0^K, 1\rangle^p) < \infty$ *for some* $p \geq 4$ *and that* $M_q(\mu_0) < \infty$ *for some* $q > 2$*. Then, for all* $K \geq 1$ *and* $T > 0$ *one has*

$$\sup_{t\in[0,T]} \mathbb{E}\Big(\big\|\mu_t^K - \mu_t\big\|_{\mathrm{BL}^*}\Big) \leq C_T \begin{cases} \Big(I_4(K) + K^{-\frac{1}{4}} + K^{-\frac{(q-2)}{2q}}\Big), & \text{if } d < 4 \text{ and } q \neq 4, \\ \Big(I_4(K) + K^{-\frac{1}{4}}(\log(1+K))^{\frac{1}{2}} + K^{-\frac{(q-2)}{2q}}\Big), & \text{if } d = 4 \text{ and } q \neq 4, \\ \Big(I_4(K) + K^{-\frac{1}{d}} + K^{-\frac{(q-2)}{2q}}\Big), & \text{if } d > 4 \text{ and } q \neq d/(d-2), \end{cases}$$

where C_T *is a constant depending on* T, p, q *and the data of the model.*

We will denote by $\Psi_{d,q}(K)$ the function of K appearing on the right hand side of the bound in Theorem 3.2.

The convergence rate $\Psi_{d,q}(K)$ in dual bounded-Lipschitz distance thus depend non-increasingly on the dimension d, and on how many finite moments the measure μ_0 has, provided that the initial total masses $\langle \mu_0^K, 1\rangle$ converge at least as fast as $K^{-1/4}$ in L^4 to $\langle \mu_0, 1\rangle$. This requirement is not too stringent (it holds e.g. in the aforementioned Poissonian setting, see Lemma 3.3.3 below), and can be relaxed if the particles' spatial dynamics do not interact (see Section 3.6). For modeling purposes, the most relevant setting is $d = 3$, in which case $\Psi_{d,q}(K)$ is equivalent under the previous condition to $K^{-1/4}$ if $q \in [4, +\infty)$, or to the slower rate $K^{-\frac{(q-2)}{2q}}$ if $q \in (2, 4)$.

The conditional independence in Assumption $(\mathbf{H})$ 1) can be relaxed to a conditional exchangeability and chaoticity condition, to be made precise in Section 3.3, at the price of an additional term in Theorem 3.2 associated with the initial empirical distribution. See also Section 3.7 in that direction. Also, the same result, with the natural modification of the limiting PDE, can be obtained in the case that each individual of the population additionally carries an independent, autonomous exponential killing clock of a fixed parameter.

Definition 3.2.1. Let $(N^K)_{K\in\mathbb{N}^*}$ be random variables in $\mathbb{N}$ going in law to ∞ as $K \to \infty$. We say a family $((Y^{1,K}, \dots, Y^{N^K,K}))_{K\in\mathbb{N}^*}$ of random vectors, $(\mathbb{R}^d)^{N^K}$-valued and exchangeable conditionally on N^K for each K, is conditionally P-*chaotic* given $(N^K)_{K\in\mathbb{N}^*}$ if for some $P \in \mathcal{P}(\mathbb{R}^d)$ and every $j \in \mathbb{N}^*$ the (random) conditional laws $(\mathcal{L}(Y^{1,K}, \dots, Y^{j\wedge N^K,K}|N^K))_{K\in\mathbb{N}^*}$ given N^K and the event $\{N^K \geq j\}$ converge in distribution in $\mathcal{P}((\mathbb{R}^d)^j)$ to $P^{\otimes j}$ as $K \to \infty$.

In the case that $N^K = K$ is deterministic for all $K \in \mathbb{N}^*$, one recovers the well known notion of P-chaoticity [110, 95]. Under the same assumptions of Theorem 3.2 we deduce the following result, proved at the end of Section 3.7.

Corollary 3.2.1. *For each* $t \geq 0$ *the family* $((X_t^{1,K}, \dots, X_t^{N_t^K,K}))_{K\in\mathbb{N}^*}$ *is conditionally* P*-chaotic given* $(N_t^K)_{K\in\mathbb{N}^*}$ *with* $P = \mu_t/\langle \mu_t, 1\rangle$.

Structure of the paper

In Section 3.3 we recall some basic facts regarding distances on finite measures and probability measures, we state quantitative estimates in the Wasserstein−2 distance established for i.i.d. samples in [60], and we show how they translate into estimates for random measures μ_0^K satisfying assumption $(\mathbf{H})$ 1). We also make some complementary remarks on this assumption.

In Section 3.4 we explain the core of the proof of Theorem 3.2, namely the construction of a coupling of $(\mu_t^K)_{\geq 0}$ with an auxiliary particle system $(\nu_t^K)_{\geq 0}$, which has the structure described in assumption $(\mathbf{H})\, 1)$ at all times. We introduce condition $(\mathbf{C})$ gathering three properties that the auxiliary system and the coupling with it must satisfy, in order that the asserted bounds for system $(\mu_t^K)_{\geq 0}$ can be deduced, and we explain in details how and why optimal transport must be used to do so.

In Section 3.5 we explicitly construct the coupled particle systems $(\mu_t^K)_{\geq 0}$ and $(\nu_t^K)_{\geq 0}$ in terms of Brownian motions and a Poisson point measure, with help of a measurable construction in [35]. The latter allows us to dynamically sample from the atoms of the point measure the realizations of optimal transport plans between a continuous flow of probability laws and a predictable flow of random empirical distributions. We also check that the two simplest properties in condition $(\mathbf{C})$ are verified by this construction.

In Section 3.6 we prove that the coupling satisfies the third and last property of condition $(\mathbf{C})$, in the simpler case of pure binary branching processes (i.e. with no mean-field interaction between the particles nor competition). We then deduce Theorem 3.2 in that setting, with slightly better bounds.

Finally, in Section 3.7 we prove that the last required condition on the coupling also holds in the general case, and we deduce the proofs of the main results. We then end the paper commenting on potential extensions of the developed ideas and results to more general branching population models.

3.3 Preliminaries

Recall that, for $p \in [1, \infty)$, the p-Wasserstein distance $W_p(\mu, \nu)$ between two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ is defined by

$$W_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p \; \pi(\mathrm{d}x, \mathrm{d}y) \right)^{\frac{1}{p}},$$

where $\Pi(\mu, \nu)$ is the set of probability measures over $\mathbb{R}^d \times \mathbb{R}^d$ that have μ and ν respectively as first and second marginals. A coupling $\pi \in \Pi(\mu, \nu)$ realizing the infimum always exists and is called an *optimal coupling* between μ and ν for the transport cost $c(x, y) = |x-y|^p$. The quantity $W_p(\mu, \nu)$ defines a complete distance if restricted to the space of probability measure with finite p-th moment, and is therein equivalent to the weak topology, strengthened with the convergence of p-th moments. See [114] for background.

For every $\mu \in \mathcal{M}^+(\mathbb{R}^d)$, we will throughout denote by $\bar{\mu}$ the probability measure on $\mathbb{R}^d$ obtained from it by normalization:

$$\bar{\mu} := \frac{1}{\langle \mu, 1 \rangle} \mu \in \mathcal{P}(\mathbb{R}^d).$$

We next state some simple but useful basic relations between finite measures and their normalizations:

Lemma 3.3.1. *Let $\mu,\nu \in \mathcal{M}^+(\mathbb{R}^d)$. We have that*

$$\|\mu-\nu\|_{\mathrm{BL}^*} \le \langle \mu,1\rangle \|\bar\mu-\bar\nu\|_{\mathrm{BL}^*} + \Big|\langle \mu,1\rangle - \langle \nu,1\rangle\Big|,$$

and

$$\|\bar\mu-\bar\nu\|_{\mathrm{BL}^*} \le \inf_{\pi\in\Pi(\bar\mu,\bar\nu)} \int |x-y|\wedge 2\,\pi(\mathrm{d}x,\mathrm{d}y) \le W_1(\bar\mu,\bar\nu).$$

Proof. Since $\|\bar\nu\|_{\mathrm{BL}^*} = \langle\bar\nu,1\rangle = 1$, we have

$$\begin{aligned}\|\mu-\nu\|_{\mathrm{BL}^*} =&\ \|\langle \mu,1\rangle\,(\bar\mu-\bar\nu) + \bar\nu\,(\langle \mu,1\rangle - \langle \nu,1\rangle)\|_{\mathrm{BL}^*}\\ &\le \langle \mu,1\rangle \|\bar\mu-\bar\nu\|_{\mathrm{BL}^*} + \Big|\langle \mu,1\rangle - \langle \nu,1\rangle\Big|.\end{aligned}$$

Now, for any $\mu,\nu\in\mathcal{P}(\mathbb{R}^d)$, $\|\mu-\nu\|_{\mathrm{BL}^*} = \sup_{\|\varphi\|_{\mathrm{BL}}\le 1}\Big|\int_{\mathbb{R}^d\times\mathbb{R}^d}(\varphi(x)-\varphi(y))\,\pi(\mathrm{d}x,\mathrm{d}y)\Big|$ for each coupling $\pi\in\mathcal{P}(\mathbb{R}^{2d})$ of μ and ν. Using the fact that $|\varphi(x)-\varphi(y)|\le|x-y|\wedge 2$ when $\|\varphi\|_{\mathrm{BL}}\le 1$ and taking infimum over all $\pi\in\Pi(\mu,\nu)$, we conclude that

$$\|\mu-\nu\|_{\mathrm{BL}^*} \le \inf_{\pi\in\Pi(\mu,\nu)} \int |x-y|\wedge 2\,\pi(\mathrm{d}x,\mathrm{d}y) \le W_1(\mu,\nu). \qquad \square$$

Let us now recall in the case $p=2$ the quantitative bounds in p-Wasserstein distance for empirical measures of i.i.d. samples proved in [60], on which our main result relies.

Theorem 3.3. *Let $\bar\mu\in\mathcal{P}(\mathbb{R}^d)$ and $(X^n)_{n\in\mathbb{N}}$ be an i.i.d. collection of random variables with law $\bar\mu$. Assume that $M_q(\bar\mu)<\infty$ for some $q>2$. There exists a constant $C_{d,q}>0$ depending only on d and q such that, for all $N\ge 1$,*

$$\mathbb{E}\bigg(W_2^2\bigg(\frac{1}{N}\sum_{n=1}^N \delta_{X^n},\bar\mu\bigg)\bigg) \le C_{d,q} M_q^{\frac{2}{q}}(\bar\mu)\,R_{d,q}(N),$$

where $R_{d,q}\colon\mathbb{N}^*\to\mathbb{R}_+$ is defined by

$$R_{d,q}(N) := \begin{cases} N^{-\frac12} + N^{-\frac{(q-2)}{q}}, & \text{if } d<4 \text{ and } q\neq 4,\\ N^{-\frac12}\log(1+N) + N^{-\frac{(q-2)}{q}}, & \text{if } d=4 \text{ and } q\neq 4,\\ N^{-\frac2d} + N^{-\frac{(q-2)}{q}}, & \text{if } d>4 \text{ and } q\neq \frac{d}{d-2}.\end{cases}$$

We deduce analogous estimates for random empirical measures in $\mathcal{M}^K(\mathbb{R}^d)$ satisfying condition $(\mathbf{H})\,1)$.

Lemma 3.3.2. *Let $\mu\in\mathcal{M}^+(\mathbb{R}^d)$ be such that $M_q(\mu)<\infty$ for some $q>2$ and let (N,ν^K) be a random variable in $\mathbb{N}\times\mathcal{M}^K(\mathbb{R}^d)$ such that $\mathbb{E}(N)<\infty$ and, conditionally on N, ν^K is supported on N atoms that are i.i.d. random variables of law $\bar\mu$. Then, there exists a constant $C_{d,q}>0$ that depends only on d,q such that*

$$\mathbb{E}\Big(\frac{N}{K}W_2^2\left(\bar\nu^K,\bar\mu\right)\Big) \le C_{d,q} M_q^{\frac{2}{q}}(\bar\mu)\,\mathbb{E}(1\vee(N/K))R_{d,q}(K). \tag{3.3.1}$$

Proof. Write $\alpha = \frac{1}{2}$ when $d < 4$ or $\alpha = \frac{2}{d}$ when $d > 4$. Thanks to Theorem 3.3, for some $C_{d,q} > 0$,

$$\begin{aligned}\mathbb{E}\Big(\frac{N}{K}W_2^2\left(\bar{\nu}^K,\bar{\mu}\right)\Big) &= \mathbb{E}\Big(\frac{N}{K}\mathbb{E}\Big(W_2^2\left(\bar{\nu}^K,\bar{\mu}\right)\Big| N\Big)\Big)\\ &\le C_{d,q}M_q^{\frac{2}{q}}(\bar{\mu})\,\mathbb{E}\Big(\frac{N}{K}\Big(N^{-\alpha}+N^{-\frac{q-2}{q}}\Big)\Big)\\ &= C_{d,q}M_q^{\frac{2}{q}}(\bar{\mu})\Big(K^{-\alpha}\,\mathbb{E}\Big(\Big(\frac{N}{K}\Big)^{1-\alpha}\Big)+K^{-\frac{q-2}{q}}\mathbb{E}\Big(\Big(\frac{N}{K}\Big)^{\frac{2}{q}}\Big)\Big)\\ &\le C_{d,q}M_q^{\frac{2}{q}}(\bar{\mu})\Big(K^{-\alpha}\mathbb{E}\Big(\frac{N}{K}\Big)^{1-\alpha}+K^{-\frac{q-2}{q}}\mathbb{E}\Big(\frac{N}{K}\Big)^{\frac{2}{q}}\Big),\end{aligned}$$

using Jensen's inequality in the last line. This implies the result for $d \neq 4$. When $d = 4$ we get the bounds

$$\begin{aligned}\mathbb{E}\Big(\frac{N}{K}W_2^2\left(\bar{\nu}^K,\bar{\mu}\right)\Big) &\le C_{d,q}M_q^{\frac{2}{q}}(\bar{\mu})\Big(K^{-\frac{1}{2}}\,\mathbb{E}\Big(\Big(\frac{N}{K}\Big)^{\frac{1}{2}}\log(1+N)\Big)+K^{-\frac{q-2}{q}}\mathbb{E}\Big(\frac{N}{K}\Big)^{\frac{2}{q}}\Big)\\ &\le C_{d,q}M_q^{\frac{2}{q}}(\bar{\mu})\Big(K^{-\frac{1}{2}}\,\mathbb{E}\Big(\frac{N}{K}\Big)^{\frac{1}{2}}\mathbb{E}\Big(\log^2(e+N)\Big)^{\frac{1}{2}}+K^{-\frac{q-2}{q}}\mathbb{E}\Big(\frac{N}{K}\Big)^{\frac{2}{q}}\Big).\end{aligned}$$

The function $x \in [e,\infty) \mapsto \log^2(x)$ being concave, we can extend it linearly on $(-\infty, e)$ to get a C^1 concave function on $\mathbb{R}$. Jensen's inequality then yields

$$\mathbb{E}\Big(\log^2(e+N)\Big)^{\frac{1}{2}} \le \log(e+K\mathbb{E}(N/K)) \le 1+\log(1+K)+\log(1\vee\mathbb{E}(N/K)),$$

and the case $d = 4$ follows. □

We end this section gathering some remarks on $(\mathbf{H})\,1)$ and related properties.

Lemma 3.3.3. *a) Under $(\mathbf{H})\,1)$, $(\mu_0^K)_K$ converges in law to the deterministic finite measure*

$$\mu_0 := \lim_{K\to\infty}\langle\mu_0^K,1\rangle\bar{\mu}_0.$$

b) The same conclusion as in a) holds if $(\langle\mu_0^K,1\rangle)_K$ converges in law as $K\to\infty$ to a constant in $(0,\infty)$ and there exists a $\bar{\mu}_0-$chaotic family of exchangeable random vectors $((Y^{1,N},\dots,Y^{N,N}) : N \in \mathbb{N})$ such that for all K, conditionally on $K\langle\mu_0^K,1\rangle = N$ the set of atoms of μ_0^K has the same law as $(Y^{1,N},\dots,Y^{N,N})$.

c) $(\mathbf{H})\,1)$ holds if $K\mu_0^K$ is for each K a Poisson point measure on $\mathbb{R}^d$ of intensity $K\nu_0$ with $\nu_0 \in \mathcal{M}^+(\mathbb{R}^d)$ fixed. In this case, μ_0 defined in a) is equal to ν_0. Moreover, in that case we have $I_4(K) \le CK^{-1/2}$.

The proof is simple and is given in the Appendix for completeness.

Remark 3.3.1. *If instead of $(\mathbf{H})\,1)$ one assumes that the initial condition μ_0^K satisfies only the condition in Lemma 3.3.3 b), Theorem 3.2 still holds but with an additional term on the r.h.s. of generic form: $C_T\mathbb{E}\Big(\frac{1}{K}\sum_{n=1}^{N_0^K}\|X_0^n-Y_0^n\|^2\Big)$ where, conditionally on the event $\{N_0^K = N\}$, $\Big((X_0^1,\dots,X_0^N),(Y_0^1,\dots,Y_0^N)\Big)$ is for each $N,K\in\mathbb{N}$ a coupling of the N atoms of μ_0^K and an i.i.d. sample of size N of the law $\bar{\mu}_0$. See Remark 3.7.1 for details and for the optimal value of this term.*

3.4 Strategy of the proof

The basis to obtain quantitative estimates for $W_2^2(\mu_t^K, \mu_t)$ will be Lemma 3.3.2. However, the conditional independence property required to apply that result holds only when $t = 0$, by assumption $(\mathbf{H})\,1)$, and is lost as soon as $t > 0$, even in the case of binary branching diffusions without any interactions.

The core of the proof will thus be the construction of an auxiliary system of particles in $\mathcal{M}^K(\mathbb{R}^d)$ denoted

$$\nu_t^K := \frac{1}{K}\sum_{n=1}^{N_t^K} \delta_{Y_t^n}\,,\; t \geq 0,$$

and defined in the same probability space as $(\mu_t^K)_{t\geq 0}$, such that the following condition holds.

Condition $(\mathbf{C})$:

1) $\nu_0^K = \mu_0^K$ and $K\langle \nu_t^K, 1\rangle = K\langle \mu_t^K, 1\rangle = N_t^K$ for all $t \geq 0$ almost surely.

2) For each $t \geq 0$, conditionally on $\langle \nu_t^K, 1\rangle$, the atoms of ν_t^K are i.i.d. random variables of law $\bar{\mu}_t$.

3) For each $T > 0$ there is a constant $C_T > 0$ depending on T and on the data of Theorem 3.2 such that
$$\mathbb{E}\Big(\frac{N_t^K}{K} W_2^2\left(\bar{\nu}_t^K, \bar{\mu}_t^K\right)\Big) \leq C_T\,(R_{d,q}(K) + I_4^2(K)).$$

Under the assumptions of Theorem 3.2 and **Condition** $(\mathbf{C})$ we have the following result.

Lemma 3.4.1. *There is a finite constant $C_T > 0$ as above such that for all $t \in [0, T]$:*

$$\mathbb{E}\left(\|\mu_t^K - \mu_t\|_{\mathrm{BL}^*}\right) \leq C_T\Big(R_{d,q}^{\frac{1}{2}}(K) + I_4(K)\Big). \tag{3.4.1}$$

The following bounds are needed in the proof of Lemma 3.4.1 and are proved in Section 3.7.

Lemma 3.4.2. *For each $T > 0$ and $p \geq 1$ there is a constant $C_{T,p} > 0$ such that*

$$\sup_K \mathbb{E}\Bigg(\sup_{t\in[0,T]} \langle \mu_t^K, 1\rangle^p\Bigg) < C_{T,p} \sup_K \mathbb{E}(\langle \mu_0^K, 1\rangle^p).$$

Moreover, if $\sup_K \mathbb{E}(\langle \mu_0^K, 1\rangle^2) < \infty$, for all $T > 0$ we have

$$\mathbb{E}\Big(\Big(\langle \mu_t^K, 1\rangle - \langle \mu_t, 1\rangle\Big)^2\Big) \leq C_T\Big(I_2^2(K) + K^{-1}\Big).$$

Lemma 3.4.3. *For each $T > 0$ and $q \geq 2$ there is a constant $C_T' > 0$ such that*

$$\sup_{t\in[0,T]} M_q(\bar{\mu}_t) < C_T'(1 + M_q(\bar{\mu}_0)).$$

Proof of Lemma 3.4.1. Since $\langle \mu_t^K, 1\rangle = \frac{N_t^K}{K}$, applying Lemma 3.3.1 and the triangle inequality for W_1 we get

$$\begin{aligned}
\mathbb{E}\left(\|\mu_t^K - \mu_t\|_{\mathrm{BL}^*}\right) &\leq \mathbb{E}\Big(\frac{N_t^K}{K} W_1\left(\bar{\nu}_t^K, \bar{\mu}_t^K\right)\Big) + \mathbb{E}\Big(\frac{N_t^K}{K} W_1\left(\bar{\nu}_t{}^K, \bar{\mu}_t\right)\Big) \\
&\quad + \mathbb{E}\Big(\Big|\langle \mu_t^K, 1\rangle - \langle \mu_t, 1\rangle\Big|\Big) \\
&\leq \left(\mathbb{E}\Big(\frac{N_t^K}{K} W_2^2\left(\bar{\nu}_t^K, \bar{\mu}_t\right)\Big)^{\frac{1}{2}} + \mathbb{E}\Big(\frac{N_t^K}{K} W_2^2\left(\bar{\nu}_t^K, \bar{\mu}_t^K\right)\Big)^{\frac{1}{2}}\right)\mathbb{E}\Big(\frac{N_t^K}{K}\Big)^{\frac{1}{2}} \\
&\quad + \mathbb{E}\Big(\Big(\langle \mu_t^K, 1\rangle - \langle \mu_t, 1\rangle\Big)^2\Big)^{1/2} \\
&\leq C_T\Bigg(\mathbb{E}\Big(\frac{N_t^K}{K} W_2^2\left(\bar{\nu}_t^K, \bar{\mu}_t\right)\Big)^{\frac{1}{2}} + \mathbb{E}\Big(\frac{N_t^K}{K} W_2^2\left(\bar{\nu}_t^K, \bar{\mu}_t^K\right)\Big)^{\frac{1}{2}} \\
&\quad + I_2(K) + K^{-1/2}\Bigg).
\end{aligned}$$

We have also used the Cauchy-Schwarz inequality and the inequality $W_1^2 \leq W_2^2$ in the second line, and the bounds in Lemma 3.4.2 in the last one. Thanks to the first bound in Lemma 3.4.2, Lemma 3.4.3 and conditions $(\mathbf{C})$ 1) and 2), we can apply Lemma 3.3.2 to $\bar{\nu} = \bar{\nu}_t^K$, $N = N_t^K$ and $\bar{\mu} = \bar{\mu}_t^K$, to bound the first term in the square parentheses by $R_{d,q}^{\frac{1}{2}}(K)$. The second term is bounded by $C_T(R_{d,q}^{\frac{1}{2}}(K) + I_4(K))$, due to $(\mathbf{C})$ 3). Since $I_2(K) \leq I_4(K)$ and $K^{-1/2} \leq R_{d,q}^{\frac{1}{2}}$, the proof is complete. ☐

Before providing a detailed pathwise construction of the coupling, let us briefly explain how condition $(\mathbf{C})$ will be accomplished through it. Condition $(\mathbf{C})$ 1) will be automatically granted since the system $(\nu_t^K)_{t\geq 0}$ will be constructed in such a way that its birth and death events are simultaneous with those of $(\mu_t^K)_{t\geq 0}$. Moreover, we will see that $(\mathbf{C})$ 2) can be ensured by letting each atom Y_t^n of ν_t^K evolve, during its lifespan, independently of the others and of the births and deaths events, following a specific diffusion process, defined in Proposition 3.5.1, which has the law $\bar{\mu}_t$ at each time instant t from its random birth-time on.

The most important condition and the most difficult one to ensure is $(\mathbf{C})$ 3). Assuming that $I_4(K)$ vanishes not too slowly, this roughly requires $\mathbb{E}\Big(\frac{N_t^K}{K} W_2^2\left(\bar{\nu}_t^K, \bar{\mu}_t^K\right)\Big)$ to be, over all $t \in [0, T]$, of similar order in K as $\mathbb{E}\Big(\frac{N_t^K}{K} W_2^2\left(\bar{\nu}_t^K, \bar{\mu}_t\right)\Big)$. Notice that

$$\mathbb{E}\Big(\frac{N_t^K}{K} W_2^2\left(\bar{\nu}_t^K, \bar{\mu}_t^K\right)\Big) \leq \mathbb{E}\Big(\frac{1}{K}\sum_{n=1}^{N_t^K} \|X_t^n - Y_t^n\|^2\Big),$$

since $W_2^2\left(\bar{\nu}_t^K, \bar{\mu}_t^K\right) \leq \frac{1}{N_t^K}\sum_{n=1}^{N_t^K} \|X_t^n - Y_t^n\|^2$, and so we are led to define the pairings (X_t^n, Y_t^n) of atoms of the two systems in such a way that the squared distance $\|X_t^n - Y_t^n\|^2$ is small in average, at all times. We can partially achieve this by pairing particles with equal birth and death times in the two systems, using the same Brownian motion for the two of them, and relying on the Lipschitz character of the coefficients to control their distance during their common life-span, in terms of their distance at their birth-time. Coupling efficiently the birth positions of two paired particles in the two systems w.r.t. the squared Euclidean distance will therefore be crucial. This is where optimal transport comes into play.

Indeed, recall that, on one hand, each new particle in the system $(\nu_t^K)_{t\geq 0}$, given birth at time s, is sampled in $\mathbb{R}^d$ according to the law $\bar{\mu}_s$. On the other hand, as a result of the branching dynamics, each new particle in the system $(\mu_t^K)_{t\geq 0}$ is given birth at time s, at the position of one of the N_{s-}^K atoms of $\bar{\mu}_{s-}^K$, each of which is equally likely to branch at a given time. A simple but key remark is that such branching event is equivalent to sampling a new particle in $\mathbb{R}^d$ according to the empirical law $\bar{\mu}_{s-}^K$. Thus, the best way to couple a new pair of atoms of $(\nu_t^K)_{t\geq 0}$ and $(\mu_t^K)_{t\geq 0}$, in the sense of mean quadratic distance, is by sampling at each branching time s of the latter a pair in $(\mathbb{R}^d)^2$ with (random) distribution given by the optimal coupling between the laws $\bar{\mu}_s$ and $\bar{\mu}_{s-}^K$ for the quadratic transport cost. This sampling must be done in a measurable way in terms of the state of the process right before branching, which requires a non-trivial construction carried out in [35] and adapted to our setting in Lemma 3.5.1 below. We will then see that the coupling thus constructed ensures that $\mathbb{E}\left(\frac{N_t^K}{K} W_2^2\left(\bar{\nu}_t^K, \bar{\mu}_t^K\right)\right)$ has the required order in K.

3.5 Pathwise constructions and coupling through optimal transport

We will construct systems $(\mu_t^K = \frac{1}{K}\sum_{n=1}^{N_t^K} \delta_{X_t^n})_{t\geq 0}$ and $(\nu_t^K = \frac{1}{K}\sum_{n=1}^{N_t^K} \delta_{Y_t^n})_{t\geq 0}$ from the following set of independent stochastic inputs defined in a common, complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$:

- A sequence $(W^j)_{j\geq 1}$ of independent Brownian motions in $\mathbb{R}^d$.
- A Poisson point measure $\mathcal{N}(\mathrm{d}s, \mathrm{d}\rho, d\theta)$ on $[0,\infty)\times[0,\infty)\times[0,\infty)$, with intensity $\mathrm{d}s \otimes \mathrm{d}\rho \otimes \mathrm{d}\theta$.
- A sequence $(Z_0^j)_{j\geq 1}$ of i.i.d. random vectors of law $\bar{\mu}_0$.
- A random variable N_0^K in $\mathbb{N}$.

We will also make use of a diffusion process considered in [58], which can be seen as a nonlinear process in the sense of McKean [95]. In the particular case considered here, we study it in more detail in next result.

Proposition 3.5.1. *Let $(\mu_t)_{t\geq 0}$ be the unique weak solution in $\mathcal{M}^+(\mathbb{R}^d)$ of the nonlinear equation*

$$\frac{\partial \mu_t}{\partial t} = L_{\mu_t}^* \mu_t + \Big(r - c\langle \mu_t, 1\rangle\Big)\mu_t. \tag{3.5.1}$$

given by Theorem 3.1, with initial condition μ_0. Let W be a $d-$dimensional Brownian motion and Y_0 an independent random variable in $\mathbb{R}^d$ with law $\bar{\mu}_0$. There is pathwise existence and uniqueness for the SDE

$$Y_t = Y_0 + \int_0^t b(Y_s, H * \mu_s(Y_s))\,\mathrm{d}s + \int_0^t \sigma(Y_s, G * \mu_s(Y_s))\,\mathrm{d}W_s. \tag{3.5.2}$$

Moreover, the flow of time-marginal laws of $(Y_t)_{t\geq 0}$ *is the unique weak solution* $(\bar{\mu}_t)_{t\geq 0}$ *in* $\mathcal{P}(\mathbb{R}^d)$ *of the (linear, non-homogeneous in time) Fokker-Planck equation*

$$\frac{\partial \bar{\mu}_t}{\partial t} = L^*_{\mu_t} \bar{\mu}_t, \tag{3.5.3}$$

with respect to test functions as in Theorem 3.1, and we have $\bar{\mu}_t = \frac{\mu_t}{\langle \mu_t, 1\rangle}$ *for all* $t \geq 0$*. Last, for every bounded measurable function* $f\colon \mathbb{R}^d \to \mathbb{R}$ *we have* $\langle \mu_t, f\rangle = \mathbb{E}(f(Y_t)n_t)$*, where* n_t *is the unique solution with* $n_0 = \langle \mu_0, 1\rangle$ *of the logistic equation*

$$\mathrm{d}n_t = \Big(r - cn_t\Big) n_t \,\mathrm{d}t. \tag{3.5.4}$$

Proposition 3.5.1 is proved in Section 3.7.

Remark 3.5.1. *a) The pathwise properties of the SDE (3.5.2) stated in Proposition 3.5.1 imply that if* Y'_τ *is a random variable of law* $\bar{\mu}_\tau$ *for fixed* $\tau > 0$*, independent of* W*, then the solution* $(Y'_t)_{t\geq\tau}$ *of the SDE*

$$Y'_t = Y'_\tau + \int_\tau^t b(Y'_s, H * \mu_s(Y'_s))\,\mathrm{d}s + \int_\tau^t \sigma(Y'_s, G * \mu_s(Y'_s))\,\mathrm{d}W_s.$$

has the same law as $(Y_t)_{t\geq\tau}$*. In particular,* Y'_t *has the law* $\bar{\mu}_t$ *for all* $t \geq \tau$*.*

b) When σ *and* b *depend only on the position and not on* μ*, the process (3.5.2) is the standard diffusion associated with the generator*

$$Lf(x) = \frac{1}{2}\mathrm{Tr}\Big(a(x)\mathrm{Hess} f(x)\Big) + b(x)\cdot\nabla f(x), \tag{3.5.5}$$

which in that case also drives each of the particles of the branching system $(\mu_t^K)_{t\geq 0}$*.*

Last, the following construction adapted from [35] will be used to efficiently couple the births events of the two systems.

Lemma 3.5.1. *Let* $\mathbf{i}\colon \mathbb{R} \to \mathbb{N}$ *denote the function defined by*

$$\rho \mapsto \mathbf{i}(\rho) = \lfloor \rho \rfloor + 1,$$

and N *be a positive integer. Let also* $(\bar{\mu}_t)_{t\geq 0}$ *be a flow of probability measures with finite second order moments that is weakly continuous. There exists a measurable mapping*

$$\Lambda^N\colon \mathbb{R}_+ \times (\mathbb{R}^d)^N \times [0, N) \to \mathbb{R}^d, \qquad (t, \mathbf{x}, \rho) \mapsto \Lambda^N_t(\mathbf{x}, \rho),$$

with the following property: for every $t \geq 0$ *and* $\mathbf{x} = (x^1, \ldots, x^N) \in (\mathbb{R}^d)^N$*, if* ρ *is uniformly chosen from* $[0, N)$*, then the pair* $(\Lambda^N_t(\mathbf{x}, \rho), x^{\mathbf{i}(\rho)})$ *is an optimal coupling between* $\bar{\mu}_t$ *and* $\frac{1}{N}\sum_{i=1}^N \delta_{x^i}$ *with respect to the cost function* $(u, v) \mapsto |u - v|^2$*. Moreover, if* $\mathbf{Y}$ *is any exchangeable random vector in* $(\mathbb{R}^d)^N$*, then* $\mathbb{E}\Big(\int_{j-1}^j \phi(\Lambda^N_t(\mathbf{Y}, \tau))\mathrm{d}\tau\Big) = \langle \bar{\mu}_t, \phi\rangle$ *for any* $j \in \{1, \ldots, N\}$*, and any bounded measurable function* ϕ*. Finally, the function* $\Lambda\colon \mathbb{N}\times\mathbb{R}_+\times\Big(\bigcup_{N\in\mathbb{N}\setminus\{0\}}(\mathbb{R}^d)^N\Big)\times\mathbb{R}_+ \to \mathbb{R}^d$ *given by*

$$\Lambda(N, t, \mathbf{x}, \rho) = \Lambda^N_t\Big((x^n)_{n=1}^N, \rho\wedge N\Big),$$

if $\mathbf{x} = (x^n)_{n=1}^N \in (\mathbb{R}^d)^N$*, and* $0 \in \mathbb{R}^d$ *otherwise, is measurable.*

Proof. Everything is proved in Lemma 3 of [35] except for the last assertion, which follows noting that $\Lambda^{-1}(A) = \bigcup_{N\neq 0}\{N\} \times (\Lambda^N)^{-1}(A)$ is a measurable set for any Borel set $A \in \mathbb{R}^d$ such that $0 \notin A$, and

$$\Lambda^{-1}(\{0\}) = \Big(\bigcup_{N\neq 0}\{N\} \times \mathbb{R}_+ \times \cup_{n\neq N}(\mathbb{R}^d)^n \times \mathbb{R}_+\Big) \cup \Big(\bigcup_{N\neq 0}\{N\} \times (\Lambda^N)^{-1}(\{0\})\Big).$$

□

Coupling algorithm

Before giving the algorithm, we also introduce a sequence of *labelling processes*

$$(j_t(n) : t \geq 0)_{n\geq 1},$$

taking values in the positive integers, that will be dynamically defined to select from $(W^j)_{j\geq 1}$ the Brownian motions driving each coupled pairs of particles (X^n_t, Y^n_t), in between reproduction or death events.

The systems $(\mu^K_t = \frac{1}{K}\sum_{n=1}^{N^K_t} \delta_{X^n_t})_{t\geq 0}$ and $(\nu^K_t = \frac{1}{K}\sum_{n=1}^{N^K_t} \delta_{Y^n_t})_{t\geq 0}$ are then constructed simultaneously, through the following algorithm.

Algorithm (A):

0. We set $Y^n_0 = X^n_0 = Z^n_0$ for $n \in \{1, \ldots, N^K_0\}$ and $\mu^K_0 = \nu^K_0 = \frac{1}{K}\sum_{n=1}^{N^K_0} \delta_{Z^n_0}$. We also set two counters: $\overline{N}^K_0 = N^K_0$ and $m = 0$, and we define $T_0 = 0$. Last, we initialize $j_0(n) = n$ for all $n \geq 1$.

1. For $t \geq T_m$, we set $j_t(n) = j_{T_m}(n)$ and $\mathrm{d}B^n_t = \mathrm{d}W^{j_t(n)}_t$, $n \geq 1$, and we define the dynamics of the two populations by:

   $$X^n_t = X_{T_m} + \int_{T_m}^t b\Big(X^n_s, H * \mu^K_s(X^n_s)\Big)\,\mathrm{d}s + \int_{T_m}^t \sigma\Big(X^n_s, G * \mu^K_s(X^n_s)\Big)\mathrm{d}B^n_s,\ n = 1, \ldots, N^K_{T_m},$$

   and

   $$Y^n_t = Y_{T_m} + \int_{T_m}^t b\Big(Y^n_s, H * \mu_s(Y^n_s)\Big)\,\mathrm{d}s + \int_{T_m}^t \sigma\Big(Y^n_s, G * \mu_s(Y^n_s)\Big)\mathrm{d}B^n_s, \quad n = 1, \ldots, N^K_{T_m},$$

   until the first time $t > T_m$ with (t, ρ, θ) an atom of $\mathcal{N}$, such that

   $$\rho < N^K_{T_m} \text{ and } \theta < r + cN^K_{T_m}/K.$$

   We then set $T_{m+1} = t$.

2. For $(t, \rho, \theta) = (T_{m+1}, \rho, \theta)$ as before,

   – If $\theta < r$, we update $N^K_t := N^K_{t-} + 1$ and $\overline{N}^K_t := \overline{N}^K_{t-} + 1$, then we define:

   $$X^{N^K_t}_t := X^{\mathbf{i}(\rho)}_{t-} \text{ and } Y^{N^K_t}_t := \Lambda^{N^K_{t-}}_t\Big((X^n_{t-})_{n=1}^{N^K_{t-}}, \rho\Big).$$

– If $r \leq \theta < cN_{T_m}^K/K$, we update $N_t^K := N_{t-}^K - 1$, then we redefine:

$$\left(X_t^{\mathbf{i}(\rho)}, X_t^{\mathbf{i}(\rho)+1}, \ldots, X_t^{N_t^K}\right) := \left(X_{t-}^{\mathbf{i}(\rho)+1}, X_{t-}^{\mathbf{i}(\rho)+2}, \ldots, X_{t-}^{N_{t-}^K}\right),$$
$$\left(Y_t^{\mathbf{i}(\rho)}, Y_t^{\mathbf{i}(\rho)+1}, \ldots, Y_t^{N_t^K}\right) := \left(Y_{t-}^{\mathbf{i}(\rho)+1}, Y_{t-}^{\mathbf{i}(\rho)+2}, \ldots, Y_{t-}^{N_{t-}^K}\right),$$

and we set $j_t(n) := j_{t-}(n+1)$ for all $n \geq \mathbf{i}(\rho)$.

3. We increase m by one and go to Step 1.

Let us explain in words how the algorithm works. The systems $(\mu_t^K)_{t\geq 0}$ and $(\nu_t^K)_{t\geq 0}$ start at time $t = 0$ from the same empirical measure, and pairs of particles are given birth or die in the two systems simultaneously from then on. The variable N_t^K counts the current number of living particles in each system at time t. The variable $\overline{N}_t^K$ in turn counts how many particles have been alive in each of the two systems or, equivalently, how many Brownian motions from $(W^j)_{j\geq 1}$ have been used, during the whole time interval $[0, t]$. The usefulness of this counter will come clear shortly.

Now, given an atom (t, ρ, θ), its coordinate t is used to sample a proposal of a birth or death time, and θ an "action" among those two, according to whether $\theta < r$ or $r \leq \theta < r + cN_{t-}^K/K$ respectively.

In a birth event, $\rho < N_{t-}^K$ samples two positions in space, one distributed according to $\bar{\mu}_{t-}^K$ for the system μ^K and one according to $\bar{\mu}_t$ for the system ν^K, which are optimally coupled as explained before. The pair of newborn particles picks upon birth at time t a new, common driving Brownian motion $(W_s^{\overline{N}_t^K})_{s\geq t}$ that is independent of the past of the systems.

In a death event, $\rho < N_{t-}^K$ samples a uniformly distributed atom from $\bar{\mu}_{t-}^K$ for the system μ^K and from $\bar{\nu}_{t-}^K$ for the system ν^K, with equal index $\mathbf{i}(\rho)$. The two corresponding particles are then removed, and their common driving Brownian motion, which corresponds to some W^j with $j \leq \overline{N}_t^K$, is discarded forever. The indexes of the particles in the two systems are then updated, as well as the Brownian motions from $(W^j)_{j\geq 1}$ labelled $B^{\mathbf{i}(\rho)}, B^{\mathbf{i}(\rho)+1}$, ..., in order that the particles still alive remain indexed by a full discrete interval of the form $\{1, \ldots, N_t^K\}$, and that the underlying Brownian motion W^j driving each pair is preserved. Notice that, due to this updating rule, for all times $t \geq 0$ we have $j_t(N_t^K) = \overline{N}_t^K$.

We will denote by $(\mathcal{F}_t)_{t\geq 0}$ the complete filtration generated by all the random objects effectively employed in the algorithm until each time:

$$\mathcal{F}_t := \overline{\sigma\left(N_0^K, (Z_0^n)_{n\in\{1,\ldots,N_0^K\}}, (\mathcal{N}((0,s],\cdot,\cdot) : s \leq t), (B_s^n : s \leq t)_{n\in\{1,\ldots,N_t^K\}}\right)},$$

and by $(\mathcal{G}_t)_{t\geq 0}$ its subfiltration

$$\mathcal{G}_t := \overline{\sigma\left(N_s^K : s \leq t\right)}.$$

Notice that $\mathcal{N}$ is an $(\mathcal{F}_t)_{t\geq 0}$-Poisson process, and that the processes $(N_t^K)_{t\geq 0}$, $(\overline{N}_t^K)_{t\geq 0}$ and $(j_t(n) : t \geq 0)$, $n \geq 1$ are adapted to $(\mathcal{G}_t)_{t\geq 0}$.

Remark 3.5.2. *Thanks to Lemma 3.5.1, the mapping*

$$(t, \omega, \rho) \mapsto \left(\Lambda_t^{N_{t-}^K}\left((X_{t-}^n)_{n=1}^{N_{t-}^K}, \rho\right), X_{t-}^{\mathbf{i}(\rho)}\right) = \left(\Lambda(N_{t-}^K, t, (X_{t-}^n)_{n=1}^{N_{t-}^K}, \rho \wedge N_{t-}^K), X_{t-}^{\mathbf{i}(\rho)}\right),$$

is measurable with respect to $\mathcal{P}red(\mathcal{F}_t) \otimes \mathcal{B}(\mathbb{R})$, with $\mathcal{P}red(\mathcal{F}_t) \subseteq \mathcal{B}(\mathbb{R}) \otimes \mathcal{F}$ the predictable sigma-field associated with $(\mathcal{F}_t)_{t\geq 0}$.

The system $(\nu_t^K)_{t\geq 0}$ satisfies $(\mathbf{C})$ 1) by construction. The following lemma will be useful to check, in the next paragraph, that it also satisfies $(\mathbf{C})$ 2).

Lemma 3.5.2. *Let $(\overline{T}_j)_{j\geq 1}$ denote the sequence of consecutive birth times in $(0,\infty)$ of one new particle in the system $(\nu_t^K)_{t\geq 0}$, constructed with algorithm $(\mathbf{A})$, and $(\overline{T}_j, \rho_j)$ be the first two coordinates of the atom (t,ρ,θ) corresponding to $t = \overline{T}_j$. Then, conditionally on $\mathcal{F}_{\overline{T}_j-}$ and $\left\{\rho_j < N^K_{\overline{T}_j-}\right\}$, $Y_{\overline{T}_j}^{N^K_{\overline{T}_j}} = \Lambda^{N^K_{\overline{T}_j}}\left((X^n_{t-})_{n=1}^{N^K_{t-}}, \rho_j\right)$ has law $\bar{\mu}_{\overline{T}_j}$.*

Proof. Let $f\colon \mathbb{R}^d \to \mathbb{R}$ be a bounded measurable function and $(U_t)_{t\geq 0}$ a bounded $(\mathcal{F}_t)_{t\geq 0}$-predictable process. We have

$$
\begin{aligned}
&f\left(Y_{\overline{T}_j}^{N^K_{\overline{T}_j}}\right)\mathbf{1}_{\{\rho_j < N^K_{\overline{T}_j-}\}} U_{\overline{T}_j} \\
&\quad = \int_0^\infty \int_0^\infty \int_0^\infty f\left(\Lambda_t^{N^K_{t-}}\left((X^n_{t-})_{n=1}^{N^K_{t-}}, \rho\right)\right)\mathbf{1}_{\{\rho < N^K_{t-},\, \overline{N}^K_{t-} = N^K_0 + j - 1,\, \theta < r\}} U_t \,\mathcal{N}(\mathrm{d}t, \mathrm{d}\rho, \mathrm{d}\theta).
\end{aligned}
$$

By Remark 3.5.2, we can use the compensation formula with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$, and deduce with Lemma 3.5.1 that

$$
\begin{aligned}
&\mathbb{E}\left(f\left(Y_{\overline{T}_j}^{N^K_{\overline{T}_j}}\right)\mathbf{1}_{\{\rho_j < N^K_{\overline{T}_j-}\}} U_{\overline{T}_j}\right) \\
&\qquad = \int_0^\infty \int_0^\infty \mathbb{E}\left(\langle \bar{\mu}_t, f\rangle N_t^K \mathbf{1}_{\{\overline{N}^K_t = N^K_0 + j - 1,\, \theta < r\}} U_t\right) \mathrm{d}\theta \mathrm{d}t \\
&\qquad = \mathbb{E}\left(\int_{[0,\infty)^3} \langle \bar{\mu}_t, f\rangle \mathbf{1}_{\{\rho < N^K_{t-},\, \overline{N}^K_{t-} = N^K_0 + j - 1,\, \theta < r\}} U_t \,\mathcal{N}(\mathrm{d}t, \mathrm{d}\rho, \mathrm{d}\theta)\right) \\
&\qquad = \mathbb{E}\left(\langle \bar{\mu}_{\overline{T}_j}, f\rangle \mathbf{1}_{\{\rho_j < N^K_{\overline{T}_j-}\}} U_{\overline{T}_j}\right).
\end{aligned}
$$

Since any bounded random variable measurable w.r.t. $\mathcal{F}_{\overline{T}_j-}$ can be written as $U_{\overline{T}_j}$ for some predictable process $(U_t)_{t\geq 0}$, the statement is proved. □

Verification of condition (C) 2)

Lemma 3.5.3. *For each $t \geq 0$, conditionally on $\langle \nu_t^K, 1\rangle$, the atoms of ν_t^K are i.i.d. random variables of law $\bar{\mu}_t$.*

Proof. The proof will be done constructing an alternative system $(\widehat{\nu}_t^K = \frac{1}{K}\sum_{n=1}^{N_t^K} \delta_{\widehat{Y}_t^n})_{t\geq 0}$ with the same law as $(\nu_t^K)_{t\geq 0}$, for which the required property is easily checked. This system is defined on the same probability space as $(\nu_t^K)_{t\geq 0}$, by means of a variant of the construction of $(\nu_t^K)_{t\geq 0}$ in algorithm $(\mathbf{A})$. The algorithm is as follows:

0. Define for all $j \geq 1$:

$$Z_t^j = Z_0^j + \int_0^t b\Big(Z_s^j, H * \mu_s(Z_s^j)\Big)\,\mathrm{d}s + \int_0^t \sigma\Big(Z_s^j, G * \mu_s(Z_s^j)\Big)\,\mathrm{d}W_t^j\,, \qquad t \geq 0.$$

Set $\widehat{Y}_0^n = Z_0^n$ for $n \in \{1, \dots, N_0^K\}$ and $\widehat{\nu}_0^K = \frac{1}{K}\sum_{n=1}^{N_0^K} \delta_{\widehat{Y}_0^n}$. As before, we set the same counters $\overline{N}_0^K = N_0^K$ and $m = 0$, we define $T_0 = 0$ and we initialize $j_0(n) = n$ for all $n \geq 1$.

1. For $t \geq T_m$, we set $j_t(n) = j_{T_m}(n)$ and $\mathrm{d}B_t^n = \mathrm{d}W_t^{j_t(n)}$, $n \geq 1$, and we take

$$\widehat{Y}_t^n = Z_t^{j_t(n)}, \qquad n = 1, \dots, N_{T_m}^K,$$

until the first time $t > T_m$, with (t, ρ, θ) an atom of $\mathcal{N}$ such that $\rho < N_{T_m}^K$ and $\theta < r + cN_{T_m}^K/K$. We then set $T_{m+1} = t$.

2. For $(t, \rho, \theta) = (T_{m+1}, \rho, \theta)$ as before,

   – If $\theta < r$, we update $N_t^K := N_{t-}^K + 1$ and $\overline{N}_t^K := \overline{N}_{t-}^K + 1$, then we define:

$$\widehat{Y}_t^{N_t^K} := Z_t^{\overline{N}_t^K}.$$

   – If $r \leq \theta < cN_{T_m}^K/K$, we update $N_t^K := N_{t-}^K - 1$, and we redefine:

$$\Big(\widehat{Y}_t^{\mathbf{i}(\rho)}, \widehat{Y}_t^{\mathbf{i}(\rho)+1}, \dots, \widehat{Y}_t^{N_t^K}\Big) := \Big(\widehat{Y}_{t-}^{\mathbf{i}(\rho)+1}, \widehat{Y}_{t-}^{\mathbf{i}(\rho)+2}, \dots, \widehat{Y}_{t-}^{N_{t-}^K}\Big),$$

   and $j_t(n) := j_{t-}(n+1)$ for all $n \geq \mathbf{i}(\rho)$.

3. We increase m by one and go to Step 1.

Plainly, instead of sampling at each birth time $\overline{T}_j$ the position of a new independent particle $Y^{N_{\overline{T}_j}^K}$ from the atom $(\overline{T}_j, \rho, \theta)$ of $\mathcal{N}$ as in $(\mathbf{A})$, we now add a new particle $\widehat{Y}^{N_{\overline{T}_j}^K}$ to the system by "turning on" at that time the nonlinear diffusion process $Z^{\overline{N}_{\overline{T}_j}^K} = Z^{N_0^K + j}$, which has evolved independently since time $t = 0$, driven by the same Brownian motion $W^{N_0^K + j}$ that drives the process $\left(Y_t^{N_{\overline{T}_j}^K} : t \geq \overline{T}_j\right)$ in the construction $(\mathbf{A})$. Call now

$$\widehat{\mathcal{F}}_t := \overline{\sigma\left(\mathcal{F}_t \vee \left(Z_{\overline{T}_k}^{N_0^K + k} : N_0^K + k \leq \overline{N}_t^K\right)\right)},$$

the filtration containing the information effectively employed to construct the process $(\widehat{\nu}_t^K)$, and let $(V_t)_{t\geq 0}$ be a bounded left continuous process adapted to $(\widehat{\mathcal{F}}_t)_{t\geq 0}$. Conditionally on N_0^K, $V_{\overline{T}_j}$ depends only on $\mathcal{N}$ and (W^k, Z_0^k) for $k < N_0^K + j$, while $\left(Z_t^{N_0^K + j}\right)_{t\geq 0}$ is independent of them. Therefore, we have

$$\begin{aligned}\mathbb{E}\left(f\left(\widehat{Y}_{\overline{T}_j}^{N_{\overline{T}_j}^K}\right)\mathbf{1}_{\{\rho_j < N_{\overline{T}_j-}^K\}} V_{\overline{T}_j}\right) &= \mathbb{E}\Big(f\left(Z_{\overline{T}_j}^{N_0^K + j}\right)\mathbf{1}_{\{\rho_j < N_{\overline{T}_j-}^K\}} V_{\overline{T}_j}\Big)\\ &= \mathbb{E}\Big(\langle \bar{\mu}_{\overline{T}_j}, f\rangle \mathbf{1}_{\{\rho_j < N_{\overline{T}_j-}^K\}} V_{\overline{T}_j}\Big),\end{aligned}$$

by Remark 3.5.1 a). This implies that, conditionally on $\widehat{\mathcal{F}}_{\overline{T}_j-}$ and $\{\rho_j < N^K_{\overline{T}_j-}\}$, the random variable $\widehat{Y}^{N^K_{\overline{T}_j}}_{\overline{T}_j}$ has the law $\bar{\mu}_{\overline{T}_j}$. Comparing this to the setting in Lemma 3.5.2, one can check by induction on j that the processes $(\nu^K_t)_{t\geq 0}$ and $(\widehat{\nu}^K_t)$ have the same law on each of their (common) time intervals $[0, \overline{T}_j]$, hence over all $[0,\infty)$.

To conclude, notice that the i.i.d processes $(Z^j_t)_{t\geq 0}, j \geq 1$ have law $\bar{\mu}_t$ at each $t \geq 0$, and they are independent of the filtration $(\mathcal{G}_t)_{t\geq 0}$ with respect to which the process $(N^K_t)_{t\geq 0}$ is measurable. Moreover, for each $t \geq 0$, $\{\widehat{Y}^1_t, \dots, \widehat{Y}^{N^K_t}_t\} = \{Z^{j_t(1)}_t, \dots, Z^{j_t(N^K_t)}_t\}$ is a random subset of $\{Z^1_t, \dots, Z^{\overline{N}^K_t}_t\}$, selected in a way that is measurable w.r.t. $\mathcal{G}_t$. This readily implies that, conditionally on $N^K_t = N$, $\{\widehat{Y}^1_t, \dots, \widehat{Y}^N_t\}$ are N i.i.d. random variables of law $\bar{\mu}_t$, as required. □

3.6 Proof of Theorem 3.2: the pure binary branching case

We consider in this section the case where interactions take place only through the reproduction events, that is, due only to the fact that the position of a newborn individual coincides at its birth with that of its parent (after which all individuals evolve completely independently). Since convergence bounds are neither available in this basic setting, we provide the complete proof for this case, as it might be of independent interest, and since it is useful to illustrate directly the main arguments.

We thus assume in what follows that coefficients $\sigma\colon \mathbb{R}^d \to \mathbb{R}^{d\otimes d}$ and $b\colon \mathbb{R}^d \to \mathbb{R}^d$ do not depend on μ^K_t and, moreover, that they are Lipschitz continuous, with σ bounded (for simplicity). We will also assume that the individual instantaneous birth and death rates are time inhomogeneous, and specified respectively by two measurable functions $r, c\colon [0,T] \to \mathbb{R}_+$, bounded respectively by some positive constants $\bar{r}$ and $\bar{c}$.

The analog of Theorem 3.1 is standard in this scenario (or can be proved by the same techniques used in [58]), and the limit in law of the process $(\mu^K_t)_{t\geq 0}$ is given by the unique weak solution in $\mathcal{M}^+(\mathbb{R}^d)$ to the linear evolution equation

$$\langle \mu_t, f(t,\cdot)\rangle = \langle \mu_0, f(0,\cdot)\rangle + \int_0^t \langle \mu_s, \partial_s f(s,\cdot) + Lf(s,\cdot) + (r(s) - c(s))f(s,\cdot)\rangle \, \mathrm{d}s, \tag{3.6.1}$$

for each $t \in [0,T]$ and every $f \in C^{1,2}([0,T]\times\mathbb{R}^d)$, where L is the time-homogeneous operator defined in (3.5.5).

The construction of the coupling with the auxiliary system is essentially the same as in Section 3.5, using algorithm $(\mathbf{A})$ with two minor modifications:

- Step 1 is carried out until the first time $t > T_m$, where (t,ρ,θ) is an atom of $\mathcal{N}$ such that $\rho < N^K_{T_m}$ and $\theta < r(t) + c(t)$, at which one sets $T_{m+1} = t$.
- The updates in Step 2 are carried out according to whether $\theta < r(t)$ or otherwise $r(t) \leq \theta < r(t) + c(t)$.

In between birth or deaths events, individuals in the system $(\mu^K_t)_{t\geq 0}$ evolve according to the SDEs

$$\mathrm{d}X^n_t = b(X^n_t)\,\mathrm{d}t + \sigma(X^n_t)\,\mathrm{d}B^n_t, \qquad n = 1, \dots, N^K_t, \tag{3.6.2}$$

as also do the individuals in the system $(\nu_t^K)_{t\geq 0}$.

We next state the analog of Lemma 3.4.2 valid in the current setting.

Lemma 3.6.1. *For each $T > 0$ and $p \geq 1$ there is a constant $C_{T,p} > 0$ such that*

$$\sup_K \mathbb{E}\left(\sup_{t\in[0,T]} \langle \mu_t^K, 1\rangle^p\right) < C_{T,p} \sup_K \mathbb{E}(\langle \mu_0^K, 1\rangle^p).$$

Moreover, if $\sup_K \mathbb{E}(\langle \mu_0^K, 1\rangle) < \infty$, for all $T > 0$ we have

$$\mathbb{E}\Big(\Big|\langle \mu_t^K, 1\rangle - \langle \mu_t, 1\rangle\Big|\Big) \leq C_T\Big(I_1(K) + K^{-\frac{1}{2}}\Big).$$

Proof. The first claim is shown as in [58], Lemma 3.3. For the second assertion, we write the dynamics of the number of particles in the system in terms of the Poisson point measure $\mathcal{N}$ used in algorithm $(\mathbf{A})$. We obtain

$$\begin{aligned} N_t^K &= N_0^K + \int_0^t \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \mathbf{1}_{\rho \leq N_{s-}^K} \left(\mathbf{1}_{\theta \leq r(s-)} - \mathbf{1}_{r(s-)<\theta\leq r(s-)+c(s-)}\right) \mathcal{N}(\mathrm{d}s, \mathrm{d}\rho, \mathrm{d}\theta) \\ &= N_0^K + \int_0^t (r(s) - c(s)) N_s^K \,\mathrm{d}s + M_t^K, \end{aligned}$$

where $(M_t^K)_{t\geq 0}$ is a martingale since, for all $t \geq 0$,

$$\mathbb{E}\left(\int_0^t \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \left|\mathbf{1}_{\rho \leq N_s^K} \left(\mathbf{1}_{\theta \leq r(s)} - \mathbf{1}_{r(s)<\theta\leq r(s)+c(s)}\right)\right| \mathrm{d}s\mathrm{d}\rho\mathrm{d}\theta\right) \leq (\bar{r} + \bar{c})\mathbb{E}\left(\int_0^t N_s^K \,\mathrm{d}s\right) < \infty,$$

by the first part and the assumption on first moments. Comparing this evolution to the ODE (3.6.4) satisfied by the total mass of the limiting measure, we get the estimate

$$\mathbb{E}\left(\left|\frac{N_t^K}{K} - \langle \mu_t, 1\rangle\right|\right) \leq \mathbb{E}\left(\left|\frac{N_0^K}{K} - \langle \mu_0, 1\rangle\right|\right) + (\bar{r} + \bar{c}) \int_0^t \mathbb{E}\left(\left|\frac{N_s^K}{K} - \langle \mu_s, 1\rangle\right|\right) \mathrm{d}s + \mathbb{E}\left(\frac{|M_t^K|}{K}\right).$$

The last term is controlled using the Burkholder-Davis-Gundy (BDG) inequality as follows:

$$\begin{aligned} \mathbb{E}\left(\frac{\left|M_t^K\right|}{K}\right) &\leq \frac{1}{K}\,\mathbb{E}\left(\int_0^t \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \mathbf{1}_{\{\rho \leq N_{s-}^K, \theta \leq r(s-)+c(s-)\}} \mathcal{N}(\mathrm{d}s, \mathrm{d}\rho, \mathrm{d}\theta)\right)^{\frac{1}{2}} \\ &= \frac{\mathbb{E}((\int_0^t (r(s) + c(s))N_s^K \,\mathrm{d}s)^{\frac{1}{2}}}{K} \\ &\leq \frac{C_T}{\sqrt{K}} \left(\sup_K \mathbb{E}(\langle \mu_0^K, 1\rangle)(\bar{r} + \bar{c})e^{\bar{r}} t\right)^{\frac{1}{2}}, \end{aligned}$$

for all $t \in [0, T]$. We conclude by Gronwall's lemma that

$$\mathbb{E}\left(\left|\frac{N_t^K}{K} - \langle \mu_t, 1\rangle\right|\right) \leq C_T\left(\mathbb{E}\left(\left|\frac{N_0^K}{K} - \langle \mu_0, 1\rangle\right|\right) + \frac{1}{\sqrt{K}}\right). \quad \square$$

Notice that in the case dealt within this section, Lemma 3.4.3 is a standard propagation of moments result for diffusion processes under Lipschitz conditions. The analogue of Proposition 3.5.1 in this setting is rather elementary too, yet illustrative for the general case, so we state it in detail next.

Proposition 3.6.1. *Let $(\mu_t)_{t\geq 0}$ be the unique weak solution in $\mathcal{M}^+(\mathbb{R}^d)$ of the linear equation*

$$\frac{\partial \mu_t}{\partial t} = L^* \mu_t + (r(t) - c(t))\mu_t, \tag{3.6.3}$$

with initial condition μ_0 (given as a particular case of Theorem 3.1), and $(Y_t)_{t\geq 0}$ be the unique pathwise solution to the SDE

$$Y_t = Y_0 + \int_0^t b(Y_s)\,\mathrm{d}s + \int_0^t \sigma(Y_s)\,\mathrm{d}W_s,$$

where W is a d-dimensional Brownian motion and Y_0 and independent random variable in $\mathbb{R}^d$ with law $\bar{\mu}_0$. Then, the flow $(\bar{\mu}_t)_{t\geq 0}$ of time-marginal laws of $(Y_t)_{t\geq 0}$ is the unique weak solution of the Fokker-Planck equation

$$\frac{\partial \bar{\mu}_t}{\partial t} = L^* \bar{\mu}_t,$$

and satisfies $\bar{\mu}_t = \frac{\mu_t}{\langle \mu_t, 1\rangle}$ for all $t \geq 0$. In particular, for all bounded real function f we have $\langle \mu_t, f\rangle = \mathbb{E}(f(Y_t)n_t)$, where n_t is the unique solution with $n_0 = \langle \mu_0, 1\rangle$ of the linear differential equation

$$\mathrm{d}n_t = (r(t) - c(t))n_t\,\mathrm{d}t. \tag{3.6.4}$$

Proof. The first claim is standard and easily seen using Itô's formula (uniqueness is also standard using for example the Feynman-Kac formula). The relation between the law of Y_t and μ_t for all $t \geq 0$ is easily shown considering the function $h(t,x) = \langle \mu_t, 1\rangle f(t,x)$ and computing

$$\begin{aligned}
&\langle \bar{\mu}_t, h(t,\cdot)\rangle \\
&\quad= \langle \bar{\mu}_0, h(0,\cdot)\rangle + \int_0^t \langle \bar{\mu}_s, \partial_s h(s,\cdot) + Lh(s,\cdot)\rangle\,\mathrm{d}s \\
&\quad= \langle \langle \mu_0, 1\rangle \bar{\mu}_0, f(0,\cdot)\rangle + \int_0^t \langle \bar{\mu}_s, f(s,\cdot)\partial_s\langle \mu_s, 1\rangle + \langle \mu_s, 1\rangle \partial_s f(s,\cdot) + \langle \mu_s, 1\rangle Lf(s,\cdot)\rangle\,\mathrm{d}s \\
&\quad= \langle \langle \mu_0, 1\rangle \bar{\mu}_0, f(0,\cdot)\rangle + \int_0^t \langle \langle \mu_s, 1\rangle \bar{\mu}_s, \partial_s f(s,\cdot) + Lf(s,\cdot) + (r(s) - c(s))f(s,\cdot)\rangle\,\mathrm{d}s.
\end{aligned}$$

This means that $(\langle \mu_t, 1\rangle \bar{\mu}_t)_{t\geq 0}$ satisfies equation (3.6.1). Uniqueness for that equation yields $\langle \mu_t, 1\rangle \bar{\mu}_t = \mu_t$ for all $t \geq 0$ as claimed. This means that

$$\langle \mu_t, f\rangle = \mathbb{E}(\langle \mu_t, 1\rangle f(Y_t)),$$

for all bounded f, and the fact that $(\langle \mu_t, 1\rangle)_{t\geq 0}$ satisfies (3.6.4) is immediate. □

In order to prove that Condition (**C**) 3) holds, one last additional control is needed.

Lemma 3.6.2. *Let $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ be two diffusion processes with generator L driven by the same Brownian motion B. For each $T > 0$ there exists $C_T > 0$ such that for all $0 < u < t < T$*

$$\mathbb{E}(\|X_t - Y_t\|^2 - \|X_u - Y_u\|^2) \leq C_T \int_u^t \mathbb{E}(\|X_s - Y_s\|^2)\,\mathrm{d}s.$$

Proof. Let $(\tau_n)_{n\in\mathbb{N}}$ be the sequence defined by $\tau_n := \inf\{s \geq 0 : \|X_s\|^2 + \|Y_s\|^2 > n\}$, which localizes the local martingale parts of X and Y. We first establish a control on the running suprema of the processes. Using the fact that b is Lipschitz we obtain

$$\sup_{u\in[0,t\wedge\tau_n]} \|X_u\|^2 \leq 2\|X_0\|^2 + C_T + C_T \int_0^t \sup_{u\in[0,s\wedge\tau_n]} \|X_u\|^2 \,\mathrm{d}s + 2\sum_{i,j=1}^{d} \left(\sup_{u\in[0,t\wedge\tau_n]} \left| \int_0^u \sigma^{(ij)}(X_s)\,\mathrm{d}B_s^{(j)} \right| \right)^2.$$

With the BDG inequality and the fact that σ is also Lipschitz we then get

$$\mathbb{E}\left(\sup_{u\in[0,t\wedge\tau_n]} \|X_u\|^2 \right) \leq 2\mathbb{E}\left(\|X_0\|^2\right) + C_T + C_T \int_0^t \mathbb{E}\left(\sup_{u\in[0,s\wedge\tau_n]} \|X_u\|^2 \right) \mathrm{d}s.$$

Applying Gronwall's lemma and then Fatou's lemma upon letting $n \to \infty$ we deduce

$$\mathbb{E}\left(\sup_{s\in[0,T]} \|X_t\|^2 \right) \leq C_T(\mathbb{E}(\|X_0\|^2) + 1), \tag{3.6.5}$$

and a similar estimate for the process Y. Now, Itô's formula shows that

$$\|X_t - Y_t\|^2 = \|X_u - Y_u\|^2 + \int_u^t 2(X_s - Y_s)^{\mathrm{t}}(b(X_s) - b(Y_s))\,\mathrm{d}s + \int_u^t 2(X_s - Y_s)^{\mathrm{t}}(\sigma(X_s) - \sigma(Y_s))\,\mathrm{d}B_s + \sum_{i,j=1}^{d} \int_u^t (\sigma^{(ij)}(X_s) - \sigma^{(ij)}(Y_s))^2\,\mathrm{d}s.$$

The sequence $(\tau_n)_n$ localizes the local martingale on the right hand side. Taking expectation for the stopped process and using the Lipschitz character of b and σ leads to

$$\mathbb{E}(\|X_{t\wedge\tau_n} - Y_{t\wedge\tau_n}\|^2) \leq \mathbb{E}(\|X_u - Y_u\|^2) + C\int_u^t \mathbb{E}(\|X_{s\wedge\tau_n} - Y_{s\wedge\tau_n}\|^2)\,\mathrm{d}s.$$

By dominated convergence using the bound (3.6.5), we can take $n \to \infty$ and conclude. □

Now we can state the bound leading to condition $(\mathbf{C})$ 3) and to the proof of the main result, in the case of pure binary branching.

Lemma 3.6.3. *There exists a constant $C_T > 0$ depending on d and q, such that for all $K \in \mathbb{N}$ and $t \in [0,T]$:*

$$\mathbb{E}\left(\frac{1}{K} \sum_{n=1}^{N_t^K} \|X_t^n - Y_t^n\|^2 \right) \leq C_T \int_0^t \mathbb{E}\left(\frac{N_s^K}{K} W_2^2(\bar{\nu}_s^K, \bar{\mu}_s) \right) \mathrm{d}s.$$

Proof. Consider the product empirical measure $\eta_t^K := \frac{1}{K}\sum_{n=1}^{N_t^K} \delta_{(X_t^n, Y_t^n)}$ and the sequence of jump times $(T_m)_{m\in\mathbb{N}}$ of the process $(N_t^K)_{t\geq 0}$, defined through algorithm $(\mathbf{A})$. We decompose the evolution of η_t^K in terms of $(T_m)_{m\in\mathbb{N}}$ as follows:

$$\eta_t^K = \eta_t^K + \sum_{m=1}^{\infty} \left(\mathbf{1}_{t\geq T_m}\left(\eta_{T_m}^K - \eta_{T_{m-1}}^K\right) - \mathbf{1}_{T_{m+1}>t>T_m}\eta_{T_m}^K \right) + \eta_0^K.$$

Defining $A_t^K \coloneqq \sum_{s\le t} |\Delta N_s^K|$, where $\Delta N_s^K = N_s^K - N_{s-}^K$, we can rewrite the previous equality as

$$\eta_t^K = \eta_0^K + \eta_t^K - \eta_{T_{A_t^K}}^K + \sum_{m=1}^{\infty} \mathbf{1}_{t\ge T_m} \left(\eta_{T_m}^K - \eta_{T_m^-}^K + \eta_{T_m^-}^K - \eta_{T_{m-1}}^K\right).$$

The aim of this decomposition is to control separately what happens in between jumps and at the jump instants. Integrating the function $d_2(x,y) \coloneqq \|x-y\|^2$ and taking expectation yields

$$\begin{aligned}\mathbb{E}(\langle \eta_t^K, d_2\rangle) = \mathbb{E}\Big(\langle \eta_0^K, d_2\rangle\Big) + \mathbb{E}\Bigg(\sum_{m=1}^{\infty} \mathbf{1}_{t\ge T_m}\Big(\langle \eta_{T_m}^K, d_2\rangle - \langle \eta_{T_m^-}^K, d_2\rangle\Big)\Bigg)\\ + \mathbb{E}\Bigg(\Big\langle \eta_t^K, d_2\Big\rangle - \Big\langle \eta_{T_{A_t^K}}^K, d_2\Big\rangle + \sum_{m=1}^{\infty} \mathbf{1}_{t\ge T_m}\left(\langle \eta_{T_m^-}^K, d_2\rangle - \langle \eta_{T_{m-1}}^K, d_2\rangle\right)\Bigg).\end{aligned} \tag{3.6.6}$$

By Lemma 3.6.2, and since the evolution of η_t^K is independent of the sigma field $(\mathcal{G}_t)_{t\ge0}$ on each interval $[T_{m-1}, T_m)$, we get

$$\begin{aligned}&\mathbb{E}\Big(\mathbf{1}_{t\ge T_m}(\langle \eta_{T_m^-}^K, d_2\rangle - \langle \eta_{T_{m-1}}^K, d_2\rangle)\ \Big|\ \mathcal{G}_t\Big)\\ &\qquad = \mathbb{E}\Bigg(\frac{1}{K}\sum_{n=1}^{N_{T_{m-1}}^K} \|X_{T_m^-}^n - Y_{T_m^-}^n\|^2 - \|X_{T_{m-1}}^n - Y_{T_{m-1}}^n\|^2\ \Bigg|\ \mathcal{G}_t\Bigg)\mathbf{1}_{t\ge T_m}\\ &\qquad \le \frac{1}{K}\sum_{n=1}^{N_{T_{m-1}}^K} C\int_{T_{m-1}}^{T_m^-} \mathbb{E}\Big(\|X_s^n - Y_s^n\|^2\ \Big|\ \mathcal{G}_t\Big)\,\mathrm{d}s\mathbf{1}_{t\ge T_m}\\ &\qquad = C\int_{T_{m-1}}^{T_m^-} \mathbb{E}\Big(\langle \eta_s^K, d_2\rangle\ \Big|\ \mathcal{G}_t\Big)\,\mathrm{d}s\mathbf{1}_{t\ge T_m},\end{aligned} \tag{3.6.7}$$

and similarly, for the remaining time interval,

$$\mathbb{E}\Bigg(\mathbb{E}\Bigg(\Big\langle \eta_t^K, d_2\Big\rangle - \Big\langle \eta_{T_{A_t^K}}^K, d_2\Big\rangle\ \Bigg|\ \mathcal{G}_t\Bigg)\Bigg) \le C\int_{T_{A_t^K}}^{t} \mathbb{E}\Big(\langle \eta_s^K, d_2\rangle\ \Big|\ \mathcal{G}_t\Big)\,\mathrm{d}s.$$

Recalling Step 2 of the variant of algorithm $(\mathbf{A})$ used in this section, the term involving the jumps of the processes can be written as

$$\begin{aligned}&\mathbb{E}\Bigg(\sum_{m=1}^{\infty}\mathbf{1}_{t\ge T_m}\left(\langle \eta_{T_n}^K, d_2\rangle - \langle \eta_{T_n^-}^K, d_2\rangle\right)\Bigg)\\ &\quad = \mathbb{E}\Bigg(\frac{1}{K}\int_{[0,t]\times\mathbb{R}_+\times\mathbb{R}_+}\Bigg(\mathbf{1}_{\rho\le N_{s-}^K}\mathbf{1}_{\theta\le r(s-)}\Big\|X_s^{N_s^K} - Y_s^{N_s^K}\Big\|^2\\ &\qquad - \mathbf{1}_{\rho\le N_{s-}^K}\mathbf{1}_{r(s-)<\theta\le r(s-)+c(s-)}\Big\|X_{s-}^{\mathbf{i}(\rho)} - Y_{s-}^{\mathbf{i}(\rho)}\Big\|^2\Bigg)\mathcal{N}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta)\Bigg)\\ &\quad \le \mathbb{E}\Bigg(\int_{[0,t]\times\mathbb{R}_+\times\mathbb{R}_+}\frac{1}{K}\mathbf{1}_{\rho\le N_{s-}^K}\mathbf{1}_{\theta\le r(s-)}\Big\|X_{s-}^{\mathbf{i}(\rho)} - \Lambda_s^{N_{s-}^K}\Big((X_{s-}^n)_{n=1}^{N_{s-}^K},\rho\Big)\Big\|^2\mathcal{N}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta)\Bigg)\\ &\quad = \mathbb{E}\Bigg(\int_0^t\frac{N_s^K}{K}r(s)W_2^2\left(\bar{\mu}_s^K,\bar{\mu}_s\right)\,\mathrm{d}s\Bigg),\end{aligned} \tag{3.6.8}$$

where we used Lemma 3.5.1 and Remark 3.5.2 in the last equality. Since $\mathbb{E}(\langle \eta_0^K, d_2\rangle) = 0$, by combining the two previous estimates and writing C for some constant that may change from line to line, we deduce

$$\begin{aligned}\mathbb{E}(\langle \eta_t^K, d_2\rangle) &\leq C\int_0^t \mathbb{E}(\langle \eta_s^K, d_2\rangle)\,\mathrm{d}s + \mathbb{E}\bigg(\int_0^t \frac{N_s^K}{K} r(s) W_2^2\left(\bar{\mu}_s^K, \bar{\mu}_s\right)\,\mathrm{d}s\bigg)\\ &\leq C\int_0^t \mathbb{E}(\langle \eta_s^K, d_2\rangle)\,\mathrm{d}s + C\int_0^t \mathbb{E}\Big(\frac{N_s^K}{K} W_2^2\left(\bar{\nu}_s^K, \bar{\mu}_s\right)\Big)\,\mathrm{d}s\\ &\qquad + C\int_0^t \mathbb{E}\Big(\frac{N_s^K}{K} W_2^2\left(\bar{\mu}_s^K, \bar{\nu}_s^K\right)\Big)\,\mathrm{d}s\\ &\leq C\int_0^t \mathbb{E}(\langle \eta_s^K, d_2\rangle)\,\mathrm{d}s + C\int_0^t \mathbb{E}\Big(\frac{N_s^K}{K} W_2^2\left(\bar{\nu}_s^K, \bar{\mu}_s\right)\Big)\,\mathrm{d}s,\end{aligned}$$

where in the last inequality, we used the fact that

$$\mathbb{E}\Big(\frac{N_t^K}{K} W_2^2\left(\bar{\mu}_t^K, \bar{\nu}_t^K\right)\Big) \leq \mathbb{E}\bigg(\frac{1}{K}\sum_{n=1}^{N_t^K} \|X_t^n - Y_t^n\|^2\bigg), \tag{3.6.9}$$

since $W_2^2\left(\bar{\mu}_t^K, \bar{\nu}_t^K\right) \leq \frac{1}{N_t^K}\sum_{n=1}^{N_t^K} \|X_t^n - Y_t^n\|^2$. We conclude by Gronwall's lemma. □

Corollary 3.6.1. *Condition* $(\mathbf{C})$ 3) *holds for this model with the improved bound:* $C_T R_{d,q}(K)$.

Proof. Combine inequality (3.6.9) with Lemma 3.6.3 and apply then Lemma 3.3.2. □

Finally, under the assumptions of Theorem 3.2 and condition $(\mathbf{C})$, the main result in this case reads as follows.

Theorem 3.4. *There exists a finite constant* $C_T > 0$ *such that for all* $t \in [0, T]$

$$\sup_{t\in[0,T]} \mathbb{E}\Big(\left\|\mu_t^K - \mu_t\right\|_{\mathrm{BL}^*}\Big) \leq C_T \begin{cases}\Big(I_1(K) + K^{-\frac{1}{4}} + K^{-\frac{(q-2)}{2q}}\Big), & \text{if } d < 4 \text{ and } q \neq 4,\\ \Big(I_1(K) + K^{-\frac{1}{4}}(\log(1+K))^{\frac{1}{2}} + K^{-\frac{(q-2)}{2q}}\Big), & \text{if } d = 4 \text{ and } q \neq 4,\\ \Big(I_1(K) + K^{-\frac{1}{d}} + K^{-\frac{(q-2)}{2q}}\Big), & \text{if } d > 4 \text{ and } q \neq d/(d-2).\end{cases}$$

Proof. As in the proof of Lemma 3.4.1 we show that

$$\begin{aligned}\mathbb{E}\left(\|\mu_t^K - \mu_t\|_{\mathrm{BL}^*}\right) &\leq C_T\bigg(\mathbb{E}\Big(\frac{N_t^K}{K} W_2^2\left(\bar{\nu}_t^K, \bar{\mu}_t\right)\Big)^{\frac{1}{2}} + \mathbb{E}\Big(\frac{N_t^K}{K} W_2^2\left(\bar{\nu}_t^K, \bar{\mu}_t^K\right)\Big)^{\frac{1}{2}}\\ &\qquad + \mathbb{E}\Big(\Big|\langle \mu_t^K, 1\rangle - \langle \mu_t, 1\rangle\Big|\Big)\bigg)\\ &\leq C_T\Big(R_{d,q}(K)^{\frac{1}{2}} + I_1(K) + K^{-\frac{1}{2}}\Big),\end{aligned}$$

using Lemma 3.3.2 and Lemma 3.6.1, and we conclude noting that $K^{-\frac{1}{2}} \leq R_{d,q}(K)^{\frac{1}{2}}$. □

3.7 Proof of Theorem 3.2: the general case

We now address general processes $(\mu_t^K)_{t\geq 0}$. We start by proving Proposition 3.5.1, which relates the solution $(\mu_t)_{t\geq 0}$ of equation (3.5.1) with a non-linear process of McKean-Vlasov type.

Proof of Proposition 3.5.1. Pathwise existence and uniqueness for the SDE (3.5.2) comes from the fact that the coefficients are Lipschitz functions. In order to characterize the flow of time-marginal laws of $(Y_t)_{t\geq 0}$, consider a function $f \in C^{1,2}([0,T]\times\mathbb{R}^d)$ satisfying the conditions in Theorem 3.1. By Itô's formula we obtain

$$f(t,Y_t) = f(0,Y_0) + \int_0^t \frac{\partial f(s,Y_s)}{\partial s}\,\mathrm{d}s + \int_0^t \nabla f(s,Y_s)^{\mathrm{t}} b(Y_s, H*\mu_s(Y_s))\,\mathrm{d}s$$
$$+ \int_0^t \nabla f(s,Y_s)^{\mathrm{t}}\sigma(Y_s, G*\mu_s(Y_s))\,\mathrm{d}W_s + \frac{1}{2}\int_0^t \mathrm{Tr}(a(Y_s, G*\mu_s(Y_s))\mathrm{Hess} f(s,Y_s))\,\mathrm{d}s.$$

Taking expectation in this equation shows that the law of the time-marginal is a weak solution of (3.5.3) with respect to that set of test functions. Now, consider the function $h(t,x) = \langle \mu_t, 1\rangle f(t,x)$. By equation (3.5.3) we get

$$\begin{aligned}
&\langle \bar{\mu}_t, h(t,\cdot)\rangle \\
&= \langle \bar{\mu}_0, h(0,\cdot)\rangle + \int_0^t \langle \bar{\mu}_s, \partial_s h(s,\cdot) + L_{\mu_s} h(s,\cdot)\rangle\,\mathrm{d}s \\
&= \langle \langle \mu_0, 1\rangle\bar{\mu}_0, f(0,\cdot)\rangle + \int_0^t \langle \bar{\mu}_s, f(s,\cdot)\partial_s\langle \mu_s,1\rangle + \langle \mu_s,1\rangle\partial_s f(s,\cdot) + \langle \mu_s,1\rangle L_{\mu_s} f(s,\cdot)\rangle\,\mathrm{d}s \\
&= \langle \langle \mu_0, 1\rangle\bar{\mu}_0, f(0,\cdot)\rangle + \int_0^t \langle \langle \mu_s,1\rangle\bar{\mu}_s, \partial_s f(s,\cdot) + L_{\mu_s} f(s,\cdot) + (r - c\langle \mu_s,1\rangle) f(s,\cdot)\rangle\,\mathrm{d}s,
\end{aligned}$$

which implies that $(\tilde{\mu}_t)_{t\geq 0} := (\langle \mu_t, 1\rangle\bar{\mu}_t)_{t\geq 0}$ satisfies the following "linearized" version of equation (3.1.2)

$$\langle \tilde{\mu}_t, f(t,\cdot)\rangle = \langle \mu_0, f(0,\cdot)\rangle + \int_0^t \Big\langle \tilde{\mu}_s, \partial_s f(s,\cdot) + L_{\mu_s} f(s,\cdot) + (r - c\langle \mu_s,1\rangle) f(s,\cdot)\Big\rangle\,\mathrm{d}s.$$

With similar (indeed simpler) arguments as in the uniqueness part of Theorem 3.1 (see Section 4 in [58]) one can show that uniqueness of weak solutions (with respect to the same class of test functions) of this equation holds. Since $(\tilde{\mu}_t)_{t\geq 0} = (\mu_t)_{t\geq 0}$ also is a solution, we deduce that $\langle \mu_t, 1\rangle\bar{\mu}_t = \mu_t$ for all $t \geq 0$.

The previous identity yields $\langle \mu_t, f\rangle = \mathbb{E}(\langle \mu_t, 1\rangle f(Y_t))$ for every bounded measurable f, and the fact that $(\langle \mu_t, 1\rangle)_{t\geq 0}$ is the unique solution of equation (3.5.4) is readily obtained by taking $f = 1$ in Theorem 3.1, recalling that the local Lipschitz character of the ODE's coefficient ensures uniqueness for it. □

We next prove Lemmas 3.4.2 and 3.4.3, which are needed to obtain the estimate in Lemma 3.4.1.

Proof of Lemma 3.4.2. For the first part concerning the bounds on the moments of the total mass we refer to Lemma 3.3 in [58]. For the second part, we resort to algorithm $(\mathbf{A})$ to represent the dynamics of the number of particles by

$$\begin{aligned}N_t^K &= N_0^K + \int_0^t \int \mathbf{1}_{\rho \le N_{s-}^K} \left(\mathbf{1}_{\theta \le r} - \mathbf{1}_{r < \theta \le r + c\frac{N_{s-}^K}{K}} \right) \mathcal{N}(\mathrm{d}s, \mathrm{d}\rho, \mathrm{d}\theta) \\ &= N_0^K + \int_0^t \left(r - c\frac{N_s^K}{K} \right) N_s^K \,\mathrm{d}s + M_t^K,\end{aligned}$$

$(M_t^K)_{t\ge 0}$ is a martingale since, for all $t \ge 0$,

$$\mathbb{E}\left(\int_0^t \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} \left| \mathbf{1}_{\rho \le N_s^K} \left(\mathbf{1}_{\theta \le r(s)} - \mathbf{1}_{r < \theta \le r + c\frac{N_{s-}^K}{K}} \right) \right| \mathrm{d}s \,\mathrm{d}\rho \,\mathrm{d}\theta \right) \le (r+c)\mathbb{E}\left(\int_0^t (N_s^K)^2 \,\mathrm{d}s \right) < \infty,$$

by the previous part and the assumption on second moments. Taking into account that

$$\langle \mu_t, 1 \rangle = \langle \mu_0, 1 \rangle + \int_0^t (r - c\langle \mu_s, 1 \rangle) \langle \mu_s, 1 \rangle \,\mathrm{d}s,$$

we obtain with Itô 's formula that

$$\begin{aligned}\left(\frac{N_t^K}{K} - \langle \mu_t, 1 \rangle \right)^2 &= \left(\frac{N_0^K}{K} - \langle \mu_0, 1 \rangle \right)^2 + \int_0^t 2\left(\frac{N_{s-}^K}{K} - \langle \mu_{s-}, 1 \rangle \right) \mathrm{d}\left(\frac{M_s^K}{K} \right) \\ &\quad + \int_0^t \left[2r\left(\frac{N_s^K}{K} - \langle \mu_s, 1 \rangle \right)^2 - \left(\frac{N_s^K}{K} - \langle \mu_s, 1 \rangle \right)^2 \left(\frac{N_s^K}{K} + \langle \mu_s, 1 \rangle \right) \right] \mathrm{d}s \\ &\quad + \int_0^t \int \mathbf{1}_{\rho \le N_{s-}^K} \mathbf{1}_{r < \theta \le r + c\frac{N_{s-}^K}{K}} \left(\frac{1}{K} \right)^2 \mathcal{N}(\mathrm{d}s, \mathrm{d}\rho, \mathrm{d}\theta) \\ &\quad + \int_0^t \int \mathbf{1}_{\rho \le N_{s-}^K} \mathbf{1}_{\theta \le r} \left(\frac{1}{K} \right)^2 \mathcal{N}(\mathrm{d}s, \mathrm{d}\rho, \mathrm{d}\theta).\end{aligned}$$

Neglecting the negative term in the second line gives us the bound

$$\begin{aligned}&\left(\frac{N_t^K}{K} - \langle \mu_t, 1 \rangle \right)^2 \\ &\quad \le \left(\frac{N_0^K}{K} - \langle \mu_0, 1 \rangle \right)^2 + \int_0^t 2r\left(\frac{N_s^K}{K} - \langle \mu_s, 1 \rangle \right)^2 \mathrm{d}s + \int_0^t \frac{r}{K}\left(\frac{N_s^K}{K} \right) \mathrm{d}s \\ &\qquad + \int_0^t \frac{c}{K}\left(\frac{N_s^K}{K} \right)^2 \mathrm{d}s + \int_0^t 2\left(\frac{N_{s-}^K}{K} - \langle \mu_{s-}, 1 \rangle \right) \mathrm{d}\left(\frac{M_s^K}{K} \right) + \bar{M}_t^K + \tilde{M}_t^K \\ &\quad \le \left(\frac{N_0^K}{K} - \langle \mu_0, 1 \rangle \right)^2 + \int_0^t 2r\left(\frac{N_s^K}{K} - \langle \mu_s, 1 \rangle \right)^2 \mathrm{d}s + \frac{rT}{K} \sup_{s\in[0,T]} \langle \mu_s^K, 1 \rangle \\ &\qquad + \frac{cT}{K} \sup_{s\in[0,T]} \langle \mu_s^K, 1 \rangle^2 + \int_0^t 2\left(\frac{N_{s-}^K}{K} - \langle \mu_{s-}, 1 \rangle \right) \mathrm{d}\left(\frac{M_s^K}{K} \right) + \bar{M}_t^K + \tilde{M}_t^K, \qquad (3.7.1)\end{aligned}$$

where $(\bar{M}_t^K)_{t\ge 0}$ and $(\tilde{M}_t^K)_{t\ge 0}$ are compensated Poisson integrals. Let now $(\tau_m)_m$ be the sequence of stopping times defined by $\tau_m = \inf\{t > 0 : \overline{N}_t^K > m\}$ for $m \ge 1$ and $\tau_0 = 0$. Since $\overline{N}_s^K$ is increasing by one and $\overline{N}^K_{r-} \geq m+1 = \overline{N}^K_{\tau_m} > \overline{N}^K_{s-}$ for all $r > \tau_m \geq s$, we have

$$\int_0^{t\wedge\tau_m} 2\left(\frac{N^K_{s-}}{K} - \langle \mu_{s-}, 1\rangle\right) \mathrm{d}\left(\frac{M^K_s}{K}\right) = 2\int_0^t \mathbf{1}_{\{\overline{N}^K_{s-}\leq m\}}\left(\frac{N^K_{s-}}{K} - \langle \mu_{s-}, 1\rangle\right) \mathrm{d}\left(\frac{M^K_s}{K}\right)$$
$$= 2\int_0^t \int \phi(s,\rho,\theta)\tilde{\mathcal{N}}(\mathrm{d}s, \mathrm{d}\rho, \mathrm{d}\theta),$$

with $\tilde{\mathcal{N}}$ the compensated measure associated with $\mathcal{N}$ and ϕ the predictable process

$$\phi(s,\rho,\theta) = \mathbf{1}_{\overline{N}^K_{s-}\leq m}\mathbf{1}_{\rho\leq N^K_{s-}}\frac{1}{K}\left(\mathbf{1}_{\theta\leq r} - \mathbf{1}_{r<\theta\leq r+c\frac{N^K_{s-}}{K}}\right)\left(\frac{N^K_{s-}}{K} - \langle \mu_{s-}, 1\rangle\right).$$

The inequality $N^K_s \leq \overline{N}^K_s$ implies that

$$\mathbb{E}\left(\int_0^t\int_0^\infty\int_0^\infty |\phi(s,\rho,\theta)|\,\mathrm{d}s\,\mathrm{d}\rho\,\mathrm{d}\theta\right)$$
$$\leq \mathbb{E}\left(\int_0^t \mathbf{1}_{\overline{N}^K_s\leq m}(s)\frac{N^K_s}{K}\left(r + c\frac{N^K_s}{K}\right)\left(\frac{N^K_s}{K} + \langle \mu_s, 1\rangle\right)\mathrm{d}s\right)$$
$$\leq \mathbb{E}\left(\int_0^t \frac{m}{K}\left(r + c\frac{m}{K}\right)\left(\frac{m}{K} + \langle \mu_s, 1\rangle\right)\mathrm{d}s\right)$$
$$\leq C_{T,K,m}\left(1 + \sup_{s\in[0,T]}\langle \mu_s, 1\rangle\right),$$

and so the integral w.r.t. $\mathrm{d}\left(\frac{M^K_s}{K}\right)$ in (3.7.1) is a martingale. By similar reasonings, the stopped processes $(\bar{M}^K_{t\wedge\tau_m})_{t\geq 0}$ and $(\tilde{M}^K_{t\wedge\tau_m})_{t\geq 0}$ are also seen to be martingales. Taking expectation in (3.7.1) we get

$$\mathbb{E}\left(\left(\frac{N^K_{t\wedge\tau_m}}{K} - \langle \mu_{t\wedge\tau_m}, 1\rangle\right)^2\right)$$
$$\leq \mathbb{E}\left(\left(\frac{N^K_0}{K} - \langle \mu_0, 1\rangle\right)^2\right) + \mathbb{E}\left(\int_0^{t\wedge\tau_m} 2r\left(\frac{N^K_s}{K} - \langle \mu_s, 1\rangle\right)^2\mathrm{d}s\right) + \frac{C_T}{K}$$
$$\leq \mathbb{E}\left(\left(\frac{N^K_0}{K} - \langle \mu_0, 1\rangle\right)^2\right) + \int_0^t 2r\mathbb{E}\left(\left(\frac{N^K_{s\wedge\tau_m}}{K} - \langle \mu_{s\wedge\tau_m}, 1\rangle\right)^2\right)\mathrm{d}s + \frac{C_T}{K}.$$

Using Gronwall's lemma we obtain

$$\mathbb{E}\left(\left(\frac{N^K_{t\wedge\tau_m}}{K} - \langle \mu_{t\wedge\tau_m}, 1\rangle\right)^2\right) \leq \left(\mathbb{E}\left(\left(\frac{N^K_0}{K} - \langle \mu_0, 1\rangle\right)^2\right) + \frac{C_T}{K}\right)e^{2rT}. \tag{3.7.2}$$

from which the conclusion follows, applying Fatou's lemma. □

Proof of Lemma 3.4.3. We will use the diffusion process $(Y_t)_{t\geq 0}$ considered in Proposition 3.5.1, proven therein to satisfy $\mathbb{E}(\|Y_t\|^q) = M_q(\bar{\mu}_t)$. Applying Itô's formula to $\|Y_t\|^q$ for $q \geq 2$ yields

$$\begin{aligned}\|Y_t\|^q = \|Y_0\|^q + \int_0^t q\|Y_s\|^{q-2} Y_s^{\mathrm{t}} b(Y_s, H * \mu_s(Y_s))\,\mathrm{d}s + \int_0^t q\|Y_s\|^{q-2} Y_s^{\mathrm{t}} \sigma(Y_s, G * \mu_s(Y_s))\,\mathrm{d}B_s \\ + \frac{1}{2}\sum_{i,j=1}^d \sum_{k=1}^d \int_0^t \Big(q(q-2)\|Y_s\|^{q-4} |Y_s^{(i)}||Y_s^{(j)}| + \delta_{ij}\|Y_s\|^{q-2} \Big) \\ \times \sigma^{(ik)}(Y_s, G * \mu_s(Y_s))\sigma^{(jk)}(Y_s, G * \mu_s(Y_s)))\,\mathrm{d}s. \quad (3.7.3)\end{aligned}$$

Since b is Lipschitz, we have $\|b(Y_s, H * \mu_s(Y_s))\| \leq C\big(1 + \|Y_s\| + |H * \mu_s(Y_s)|\big)$ with $|H * \mu_s(Y_s)| = |\int H(x - Y_s)\mu_s(\mathrm{d}x)| \leq \|H\|_\infty \sup_{t\in[0,T]} |\langle \mu_s, 1\rangle|$, and similarly for σ and G. We thus get that

$$\|b(Y_s, H * \mu_s(Y_s))\| \leq C\Big(1 + \|X_s\|\Big) \text{ and } \|\sigma(X_s, G * \mu_s(X_s))\| \leq C\Big(1 + \|X_s\|\Big).$$

Using this in (3.7.3) gives us the bound

$$\begin{aligned}\|Y_t\|^q \leq \|Y_0\|^q + C\int_0^t \|Y_s\|^{q-2}\,\mathrm{d}s + C\int_0^t \|Y_s\|^{q-1}\,\mathrm{d}s + C\int_0^t \|Y_s\|^q\,\mathrm{d}s \\ + \int_0^t q\|Y_s\|^{q-2} Y_s^{\mathrm{t}}\sigma(Y_s, G * \mu_s(Y_s))\,\mathrm{d}B_s.\end{aligned}$$

Let now $(\tau_n)_{n\in\mathbb{N}}$ be a localizing sequence for the local martingale in the right hand side. Taking expectation of the stopped process yields

$$\begin{aligned}\mathbb{E}(\|Y_{t\wedge\tau_n}\|^q) \leq \mathbb{E}(\|Y_0\|^q) + C\int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^{q-2})\,\mathrm{d}s + C\int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^{q-1})\,\mathrm{d}s \\ + C\int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^q)\,\mathrm{d}s.\end{aligned}$$

Notice that, by Hölder's inequality, one gets

$$\int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^{q-1})\,\mathrm{d}s \leq \int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^q)^{\frac{q-1}{q}}\,\mathrm{d}s \leq C_T + C\int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^q)\,\mathrm{d}s,$$

and a similar bound holds for the term of order $q-2$. Combined with the previous, this entails

$$\mathbb{E}(\|Y_{t\wedge\tau_n}\|^q) \leq \mathbb{E}(\|Y_0\|)^q + C_T + C\int_0^t \mathbb{E}(\|Y_{s\wedge\tau_n}\|^q)\,\mathrm{d}s,$$

from where Gronwall's lemma yields

$$\mathbb{E}(\|Y_{t\wedge\tau_n}\|^q) \leq C_T(\mathbb{E}(\|Y_0\|)^q + 1).$$

We conclude with Fatou's lemma taking $n \to \infty$. □

Everything left to do is to show that condition (**C**) 3) holds. We will need two additional bounds stated in the next two results.

Lemma 3.7.1. *Let N and $K \in \mathbb{N}^*$ be fixed, and consider the diffusion processes $(X^n)_{n=1}^N$ in $(\mathbb{R}^d)^N$ evolving according to*

$$\mathrm{d}X_t^n = b(X_t^n, H * \mu_t^K(X_t^n))\,\mathrm{d}t + \sigma(X_t^n, G * \mu_t^K(X_t^n))\,\mathrm{d}B_t^n, \qquad t \geq 0,$$

where $(B^n)_{n=1}^N$ are independent Brownian motions in $\mathbb{R}^d$ and μ_t^K stands for the empirical measure $\mu_t^K = \frac{1}{K}\sum_{n=1}^N \delta_{X_t^n}$. Consider also N i.i.d. copies $(Y^n)_{n=1}^N$ of the process (3.5.2),

$$\mathrm{d}Y_t^n = b(Y_t^n, H * \mu_t(Y_t^n))\,\mathrm{d}t + \sigma(Y_t^n, G * \mu_t(Y_t^n))\,\mathrm{d}B_t^n, \qquad t \geq 0,$$

driven by the same Brownian motions. For each $T > 0$, there is $C_T > 0$ not depending on K nor on N such that for all $0 < u < t < T$ and each $n = 1, \ldots, N$,

$$\mathbb{E}(\|X_t^n - Y_t^n\|^2 - \|X_u^n - Y_u^n\|^2) \leq C_T \int_u^t \mathbb{E}(\|X_s^n - Y_s^n\|^2)\,\mathrm{d}s + \int_u^t \mathbb{E}\Big(\big\|\mu_s^K - \mu_s\big\|_{\mathrm{BL}^*}^2\Big)\,\mathrm{d}s.$$

Proof. We first check that the running supremum of each process (X^n) is square integrable. Using similar bounds as in the proof of Lemma 3.4.3, we get for each $t \in [0, T]$,

$$\begin{aligned}
\|X_t^n\|^2 &\leq \|X_0^n\|^2 + \int_0^t 2\|X_s^n\|\|b(X_s^n, H * \mu_s^K(X_s^n))\|\,\mathrm{d}s \\
&\quad + \int_0^t 2(X_s^n)^{\mathrm{t}}\sigma(X_s^n, G * \mu_s^K(X_s^n))\,\mathrm{d}B_s + \int_0^t \|\sigma(X_s^n, G * \mu_s^K(X_s^n))\|^2\,\mathrm{d}s \\
&\leq \|X_0^n\|^2 + C_T + C\int_0^t \|X_s^n\|\,\mathrm{d}s + C\int_0^t \|X_s^n\|^2\,\mathrm{d}s + C\int_0^t \|X_s^n\||H * \mu_s^K(X_s^n)|\,\mathrm{d}s \\
&\quad + C\int_0^t |G * \mu_s^K(X_s^n)|^2\,\mathrm{d}s + \int_0^t 2(X_s^n)^{\mathrm{t}}\sigma(X_s^n, G * \mu_s^K(X_s^n))\,\mathrm{d}B_s \\
&\leq \|X_0^n\|^2 + C_T + CT\|H\|_\infty^2\left(\frac{N}{K}\right)^2 + CT\|G\|_\infty^2\left(\frac{N}{K}\right)^2 + C\int_0^t \|X_s^n\|^2\,\mathrm{d}s \\
&\quad + \int_0^t 2(X_s^n)^{\mathrm{t}}\sigma(X_s^n, G * \mu_s^K(X_s^n))\,\mathrm{d}B_s,
\end{aligned}$$

since, in the present lemma's setting, $\langle \mu_s^K, 1\rangle = \frac{N}{K}$ for all $s \geq 0$. Let $(\tau_n)_{n\in\mathbb{N}}$ be a localizing sequence for the local martingale in the previous inequality. As in the proof of Lemma 3.6.2 we localize and then we take supremum until time $t \wedge \tau_n$ on both sides, obtaining in this way that

$$\begin{aligned}
\sup_{u\in[0,t\wedge\tau_n]} \|X_u^n\|^2 &\leq \|X_0^n\|^2 + C_T + CT\|H\|_\infty^2 \sup_{s\in[0,T]} \langle \mu_s^K, 1\rangle^2 \\
&\quad + CT\|G\|_\infty^2 \sup_{s\in[0,T]} \langle \mu_s^K, 1\rangle^2 + C\int_0^t \sup_{u\in[0,s\wedge\tau_n]} \|X_u^n\|^2\,\mathrm{d}s \\
&\quad + \sum_{i,j=1}^d \left(\sup_{u\in[0,t\wedge\tau_n]}\left|\int_0^u 2(X_s^n)^{(i)}\sigma^{(ij)}(X_s^n, G * \mu_s^K(X_s^n))\,\mathrm{d}B_s^{(j)}\right|\right).
\end{aligned}$$

The expectation of the last term is controlled using the BDG inequality by

$$\begin{aligned}
\sum_{i,j=1}^{d} \mathbb{E}\bigg(\sup_{u\in[0,t\wedge\tau_n]} \bigg| \int_0^u 2(X_s^n)^{(i)} \sigma^{(ij)}(X_s^n, G*\mu_s^K(X_s^n))\, \mathrm{d}B_s^{(j)} \bigg| \bigg) \\
&\le \sum_{i,j=1}^{d} \mathbb{E}\bigg(\bigg(\int_0^{t\wedge\tau_n} 4\Big((X_s^n)^{(i)} \sigma^{(ij)}(X_s^n, G*\mu_s^K(X_s^n)) \Big)^2 \mathrm{d}s \bigg)^{\frac{1}{2}} \bigg) \\
&\le C\mathbb{E}\bigg(\bigg(\int_0^{t\wedge\tau_n} \|X_s^n\|^2 \|\sigma(X_s^n, G*\mu_s^K(X_s^n))\|^2 \,\mathrm{d}s \bigg)^{\frac{1}{2}} \bigg) \\
&\le C\mathbb{E}\bigg(\bigg(1 + \|G\|_\infty^2 \Big(\frac{N}{K}\Big)^2 \bigg)^{\frac{1}{2}} \bigg(\int_0^t \|X_{s\wedge\tau_n}^n\|^2 \,\mathrm{d}s \bigg)^{\frac{1}{2}} \bigg) \\
&\le C_T + C_T \int_0^t \mathbb{E}(\|X_{s\wedge\tau_n}^n\|^2)\,\mathrm{d}s,
\end{aligned}$$

where we used the finiteness of the second order moment of the total mass. This allows us to obtain that

$$\mathbb{E}\bigg(\sup_{u\in[0,t\wedge\tau_n]} \|X_u^n\|^2 \bigg) \le \mathbb{E}\Big(\|X_0^n\|^2 \Big) + C_T + \int_0^t \mathbb{E}\bigg(\sup_{u\in[0,s\wedge\tau_n]} \|X_u^n\|^2 \bigg)\,\mathrm{d}s,$$

which implies by Gronwall's lemma and then monotone convergence that

$$\mathbb{E}\bigg(\sup_{t\in[0,T]} \|X_t^n\|^2 \bigg) < \infty. \tag{3.7.4}$$

A similar argument can be applied to the process $(Y_t^n)_{t\ge0}$ in order to obtain the same conclusion.

We now apply Itô's formula for fixed n as in (3.7.3) to get

$$\begin{aligned}
\|X_t^n - Y_t^n\|^2 = \|X_u^n - Y_u^n\|^2 + \int_u^t 2(X_s^n - Y_s^n)^{\mathrm{t}} \Big(b(X_s^n, H*\mu_s^K(X_s^n)) - b(Y_s^n, H*\mu_s(Y_s^n)) \Big)\,\mathrm{d}s \\
+ \int_u^t 2(X_s^n - Y_s^n)^{\mathrm{t}} \Big(\sigma(X_s^n, G*\mu_s^K(X_s^n)) - \sigma(Y_s^n, G*\mu_s(Y_s^n)) \Big)\,\mathrm{d}B_s^n \\
+ \sum_{i,j=1}^{d} \int_u^t \Big(\sigma^{(ij)}(X_s^n, G*\mu_s^K(X_s^n)) - \sigma^{(ij)}(Y_s^n, G*\mu_s(Y_s^n)) \Big)^2 \mathrm{d}s.
\end{aligned}$$

Using the Lipschitz character of the coefficients we get the bound

$$\begin{aligned}
\|X_t^n - Y_t^n\|^2 \le \|X_u^n - Y_u^n\|^2 + C\int_u^t \Big(\|X_s^n - Y_s^n\|^2 + \|X_s^n - Y_s^n\| |H*\mu_s^K(X_s^n) - H*\mu_s(Y_s^n)| \Big)\,\mathrm{d}s \\
+ C\int_u^v \Big(\|X_s^n - Y_s^n\|^2 + |G*\mu_s^K(X_s^n) - G*\mu_s(Y_s^n)|^2 \Big)\,\mathrm{d}s \\
+ \int_u^t 2(X_s^n - Y_s^n)^{\mathrm{t}} (\sigma(X_s^n, G*\mu_s^K(X_s^n)) - \sigma(Y_s^n, G*\mu_s(Y_s^n)))\,\mathrm{d}B_s^n.
\end{aligned}$$

Recalling that the function $H(\cdot - x)$ is bounded and Lipschitz for each $x \in \mathbb{R}^d$, we see that

$$\begin{aligned}
\Big| H*\mu_s^K(X_s^n) - H*\mu_s(Y_s^n) \Big| &\le \Big| H*\mu_s^K(X_s^n) - H*\mu_s(X_s^n) \Big| + |H*\mu_s(X_s^n) - H*\mu_s(Y_s^n)| \\
&\le C\|\mu_s^K - \mu_s\|_{\mathrm{BL}^*} + C\|\mu_s\|_{\mathrm{BL}^*} \|X_s^n - Y_s^n\|,
\end{aligned}$$

and similarly for the terms involving G. The uniform bound on the mass of $(\mu_t)_{t\geq 0}$ on finite time intervals allows us to get for all $0 < u < t < T$ that

$$\begin{aligned}\|X_t^n - Y_t^n\|^2 \leq \|X_u^n - Y_u^n\|^2 &+ C\int_u^t \Big(\|X_s^n - Y_s^n\|^2 + \|X_s^n - Y_s^n\|\|\mu_s^K - \mu_s\|_{\mathrm{BL}^*}\Big)\,\mathrm{d}s \\ &+ C\int_u^v \Big(\|X_s^n - Y_s^n\|^2 + \|\mu_s^K - \mu_s\|_{\mathrm{BL}^*}^2\Big)\,\mathrm{d}s \\ &+ \int_u^t 2(X_s^n - Y_s^n)^{\mathrm{t}}(\sigma(X_s^n, G*\mu_s^K(X_s^n)) - \sigma(Y_s^n, G*\mu_s(Y_s^n)))\,\mathrm{d}B_s^n \\ \leq \|X_u^n - Y_u^n\|^2 &+ C\int_u^t \Big(\|X_s^n - Y_s^n\|^2 + \|\mu_s^K - \mu_s\|_{\mathrm{BL}^*}^2\Big)\,\mathrm{d}s \\ &+ \int_u^t 2(X_s^n - Y_s^n)^{\mathrm{t}}(\sigma(X_s^n, G*\mu_s^K(X_s^n)) - \sigma(Y_s^n, G*\mu_s(Y_s^n)))\,\mathrm{d}B_s^n,\end{aligned}$$

where we used Young's inequality for the second inequality, and where C is a constant not depending on K nor on N that might change from line to line. By considering a localizing sequence $(\tau_n)_n$ for the local martingale on the right hand side, we can take expectation of the stopped process to obtain

$$\begin{aligned}\mathbb{E}(\|X_{t\wedge\tau_n}^n - Y_{t\wedge\tau_n}^n\|^2) \leq \mathbb{E}(\|X_u^n - Y_u^n\|^2) &+ C\int_u^t \mathbb{E}(\|X_{s\wedge\tau_n}^n - Y_{s\wedge\tau_n}^n\|^2)\,\mathrm{d}s \\ &+ \int_u^t \mathbb{E}(\left\|\mu_{s\wedge\tau_n}^K - \mu_{s\wedge\tau_n}\right\|_{\mathrm{BL}^*}^2)\,\mathrm{d}s,\end{aligned}$$

for all $0 < u < t < T$. Thanks to the second moments controls on the running suprema of X^n and Y^n, and since the total mass of $\mu_t^K$is constant in the context of the present lemma, we can use dominated convergence to take $n \to \infty$ and conclude the proof. □

The following additional convergence estimate for moments of the total mass will also be needed to establish condition (**C**) 3), due to the nonlinearities coming from the interactions in the dynamics.

Lemma 3.7.2. *Suppose that* $\sup_{K\in\mathbb{N}} \mathbb{E}(\langle \mu_0^K, 1\rangle^4) < \infty$. *Then, we have*

$$\mathbb{E}\bigg(\Big(\frac{N_t^K}{K} - \langle \mu_t, 1\rangle\Big)^4\bigg) \leq C_T\Big(I_4^4(K) + \frac{1}{K}\Big),$$

where $C_T > 0$ *is a constant that depends on* r *and* T.

Proof. Proceeding as in the proof of Lemma 3.4.2, we see that

$$\begin{aligned}
\left(\frac{N_t^K}{K}-\langle\mu_t,1\rangle\right)^4 &= \left(\frac{N_0^K}{K}-\langle\mu_0,1\rangle\right)^4+\int_0^t 4\left(\frac{N_{s^-}^K}{K}-\langle\mu_{s^-},1\rangle\right)^3\mathrm{d}\left(\frac{M_s^K}{K}\right)\\
&\quad+\int_0^t\left[4r\left(\frac{N_s^K}{K}-\langle\mu_s,1\rangle\right)^4-4\left(\frac{N_s^K}{K}-\langle\mu_s,1\rangle\right)^4\left(\frac{N_s^K}{K}+\langle\mu_s,1\rangle\right)\right]\mathrm{d}s\\
&\quad+\int_0^t\int \mathbf{1}_{\rho\le N_{s-}^K}\mathbf{1}_{\theta\le r}\Bigg[\left(\frac{N_{s-}^K}{K}-\langle\mu_{s-},1\rangle+\frac{1}{K}\right)^4\\
&\quad-\left(\frac{N_{s-}^K}{K}-\langle\mu_{s-},1\rangle\right)^4-4\left(\frac{N_{s-}^K}{K}-\langle\mu_{s-},1\rangle\right)^3\frac{1}{K}\Bigg]\mathcal{N}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta)\\
&\quad+\int_0^t\int \mathbf{1}_{\rho\le N_{s-}^K}\mathbf{1}_{r<\theta\le r+c\frac{N_{s-}^K}{K}}\Bigg[\left(\frac{N_{s-}^K}{K}-\langle\mu_{s-},1\rangle-\frac{1}{K}\right)^4\\
&\quad-\left(\frac{N_{s-}^K}{K}-\langle\mu_{s-},1\rangle\right)^4+4\left(\frac{N_{s-}^K}{K}-\langle\mu_{s-},1\rangle\right)^3\frac{1}{K}\Bigg]\mathcal{N}(\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta).
\end{aligned}$$

Neglecting the negative term in the second line and compensating the integrals with respect to the Poisson point measure gives us

$$\begin{aligned}
&\left(\frac{N_t^K}{K}-\langle\mu_t,1\rangle\right)^4\\
&\qquad\le\left(\frac{N_0^K}{K}-\langle\mu_0,1\rangle\right)^4+\int_0^t 4r\left(\frac{N_s^K}{K}-\langle\mu_s,1\rangle\right)^4\mathrm{d}s\\
&\qquad\quad+\int_0^t rN_s^K\left(6\left(\frac{N_s^K}{K}-\langle\mu_s,1\rangle\right)^2\frac{1}{K^2}+4\left(\frac{N_s^K}{K}-\langle\mu_s,1\rangle\right)\frac{1}{K^3}+\frac{1}{K^4}\right)\mathrm{d}s\\
&\qquad\quad+\int_0^t cN_s^K\frac{N_s^K}{K}\left(6\left(\frac{N_s^K}{K}-\langle\mu_s,1\rangle\right)^2\frac{1}{K^2}-4\left(\frac{N_s^K}{K}-\langle\mu_s,1\rangle\right)\frac{1}{K^3}+\frac{1}{K^4}\right)\mathrm{d}s\\
&\qquad\quad+\int_0^t 4\left(\frac{N_{s^-}^K}{K}-\langle\mu_{s^-},1\rangle\right)^3\mathrm{d}\left(\frac{M_s^K}{K}\right)+R_t^K+\bar{R}_t^K,
\end{aligned}$$

where $(R_t^K)_{t\ge0}$ and $(\bar{R}_t^K)_{t\ge0}$ are compensated Poisson integrals. Using Young's inequality we deduce that

$$\begin{aligned}
&\left(\frac{N_t^K}{K}-\langle\mu_t,1\rangle\right)^4\\
&\qquad\le\left(\frac{N_0^K}{K}-\langle\mu_0,1\rangle\right)^4+C\int_0^t\left(\frac{N_s^K}{K}-\langle\mu_s,1\rangle\right)^4\mathrm{d}s+\frac{C}{K^2}\int_0^t\left(\frac{N_s^K}{K}-\langle\mu_s,1\rangle\right)^2\mathrm{d}s\\
&\qquad\quad+\frac{C_T}{K^3}\sup_{s\in[0,T]}\langle\mu_s^K,1\rangle+\frac{C_T}{K}\sup_{s\in[0,T]}\langle\mu_s^K,1\rangle^2+\frac{C_T}{K}\sup_{s\in[0,T]}\langle\mu_s^K,1\rangle^4\\
&\qquad\quad+\int_0^t 4\left(\frac{N_{s^-}^K}{K}-\langle\mu_{s^-},1\rangle\right)^3\mathrm{d}\left(\frac{M_s^K}{K}\right)+R_t^K+\bar{R}_t^K. \qquad (3.7.5)
\end{aligned}$$

Proceeding as in the proof of Lemma 3.4.2, we can verify again that the last three processes are martingales if stopped at $\tau_n = \inf\{t > 0 : \overline{N}_t^K > n\}$. Thus, stopping the inequality (3.7.5) and taking expectation yields

$$\begin{aligned}\mathbb{E}\bigg(\Big(\frac{N^K_{t\wedge\tau_n}}{K} - \langle \mu_{t\wedge\tau_n}, 1\rangle\Big)^4\bigg) &\leq I_4^4(K) + \frac{C_T}{K} + C\int_0^t \mathbb{E}\bigg(\Big(\frac{N^K_{s\wedge\tau_n}}{K} - \langle \mu_{s\wedge\tau_n}, 1\rangle\Big)^4\bigg)\,\mathrm{d}s \\ &\quad + \frac{C}{K^2}\int_0^t \mathbb{E}\bigg(\Big(\frac{N^K_{s\wedge\tau_n}}{K} - \langle \mu_{s\wedge\tau_n}, 1\rangle\Big)^2\bigg)\,\mathrm{d}s \\ &\leq I_4^4(K) + \frac{C_T}{K} + C\int_0^t \mathbb{E}\bigg(\Big(\frac{N^K_{s\wedge\tau_n}}{K} - \langle \mu_{s\wedge\tau_n}, 1\rangle\Big)^4\bigg)\,\mathrm{d}s \\ &\quad + \frac{C_T T}{K^2}\Big(I_2^2(K) + \frac{1}{K}\Big),\end{aligned}$$

where we also used (3.7.2) to obtain the second inequality. Gronwall's inequality and then Fatou's lemma yield at last

$$\mathbb{E}\bigg(\Big(\frac{N^K_t}{K} - \langle \mu_t, 1\rangle\Big)^4\bigg) \leq I_4^4(K) + C_T\Big(\frac{1}{K} + \frac{I_2^2(K)}{K^2}\Big),$$

and we conclude noting that $I_2^2(K) \leq \sqrt{I_4^4(K)} \leq 1 + I_4^4(K)$. □

We can finally state a bound allowing us to ensure condition (**C**) 3).

Lemma 3.7.3. *For $t \in [0,T]$*

$$\mathbb{E}\Big(\frac{1}{K}\sum_{n=1}^{N_t^K} \|X_t^n - Y_t^n\|^2\Big) \leq C_T\Big[I_4^2(K) + K^{-\frac{1}{2}} + \int_0^T \mathbb{E}\Big(\frac{N_s^K}{K} W_2^2(\bar{\nu}_s^K, \bar{\mu}_s)\Big)\,\mathrm{d}s\Big].$$

where $C_T > 0$ is a constant that depends on the parameters of the model.

Proof. As in the proof of Lemma 3.6.3 we consider the product empirical measure $\eta_t^K := \frac{1}{K}\sum_{n=1}^{N_t^K} \delta_{(X_t^n, Y_t^n)}$ and decompose again

$$\mathbb{E}\Big(\frac{1}{K}\sum_{n=1}^{N_t^K} |X_t^n - Y_t^n|^2\Big) = \mathbb{E}(\langle \eta_t^K, d_2\rangle),$$

in terms of the sequence of jump times $(T_m)_{m\in\mathbb{N}}$, as in (3.6.6). We can proceed in a similar way as in (3.6.7) to control the evolution between jumps, now with help of Lemma 3.7.1, and control the contributions in the jump instants in the same way as in (3.6.8), to obtain

$$\begin{aligned}\mathbb{E}(\langle \eta_t^K, d_2\rangle) \leq C\int_0^t \mathbb{E}(\langle \eta_s^K, d_2\rangle)\,\mathrm{d}s + C\int_0^t \mathbb{E}\Big(\frac{N_s^K}{K} W_2^2(\bar{\nu}_s^K, \bar{\mu}_s)\Big)\,\mathrm{d}s \\ + C\int_0^t \mathbb{E}\Big(\frac{N_s^K}{K}\|\mu_s^K - \mu_s\|^2_{\mathrm{BL}^*}\Big)\,\mathrm{d}s,\end{aligned}$$

where C is a positive constant. Thus, with respect to the case dealt with in the previous section, incorporating interactions at the level of the dynamics only results in the addition of the last term. In order to bound this new term, we use Lemma 3.3.1 to get

$$\begin{aligned}
\mathbb{E}\Big(\frac{N_s^K}{K}\left\|\mu_s^K-\mu_s\right\|_{\mathrm{BL}^*}^2\Big) &\leq \mathbb{E}\Big(\frac{N_s^K}{K}\Big(\langle\mu_s,1\rangle\|\bar{\mu}_s^K-\bar{\mu}_s\|_{\mathrm{BL}^*}+\Big|\frac{N_s^K}{K}-\langle\mu_s,1\rangle\Big|\Big)^2\Big)\\
&\leq 2\sup_{u\in[0,T]}\langle\mu_u,1\rangle^2\mathbb{E}\Big(\frac{N_s^K}{K}\|\bar{\mu}_s^K-\bar{\mu}_s\|_{\mathrm{BL}^*}^2\Big)\\
&\qquad+2\mathbb{E}\Big(\frac{N_s^K}{K}\Big|\frac{N_s^K}{K}-\langle\mu_s,1\rangle\Big|^2\Big)\\
&\leq C\mathbb{E}\Big(\frac{N_s^K}{K}\|\bar{\mu}_s^K-\bar{\nu}_s^K\|_{\mathrm{BL}^*}^2\Big)+C\mathbb{E}\Big(\frac{N_s^K}{K}\|\bar{\mu}_s-\bar{\nu}_s^K\|_{\mathrm{BL}^*}^2\Big)\\
&\qquad+2\mathbb{E}\Big(\frac{N_s^K}{K}\Big|\frac{N_s^K}{K}-\langle\mu_s,1\rangle\Big|^2\Big),
\end{aligned}\tag{3.7.6}$$

where the control on the mass of the solution to equation (3.5.1) on finite time intervals is used. To control the first term of the right hand side we relate it to the Wasserstein distance using again Lemma 3.3.1, obtaining

$$\mathbb{E}\Big(\frac{N_s^K}{K}\|\bar{\mu}_s^K-\bar{\nu}_s^K\|_{\mathrm{BL}^*}^2\Big)\leq\mathbb{E}\Big(\frac{N_s^K}{K}W_2^2(\bar{\mu}_s^K,\bar{\nu}_s^K)\Big)\leq\mathbb{E}(\big\langle\eta_s^K,d_2\big\rangle),$$

and we do the same with the second term to get

$$\mathbb{E}\Big(\frac{N_s^K}{K}\|\bar{\mu}_s-\bar{\nu}_s^K\|_{\mathrm{BL}^*}^2\Big)\leq\mathbb{E}\Big(\frac{N_s^K}{K}W_2^2(\bar{\mu}_s,\bar{\nu}_s^K)\Big).$$

We thus obtain the inequality

$$\begin{aligned}
\mathbb{E}(\langle\eta_t^K,d_2\rangle)\leq C\int_0^t\mathbb{E}(\langle\eta_s^K,d_2\rangle)\,\mathrm{d}s+C\int_0^t\mathbb{E}\Big(\frac{N_s^K}{K}W_2^2(\bar{\nu}_s^K,\bar{\mu}_s)\Big)\,\mathrm{d}s&\\
+2\int_0^t\mathbb{E}\Big(\frac{N_s^K}{K}\Big|\frac{N_s^K}{K}-\langle\mu_s,1\rangle\Big|^2\Big)\,\mathrm{d}s,&
\end{aligned}\tag{3.7.7}$$

where only the last term needs to be controlled. Using Hölder's inequality yields

$$\mathbb{E}\Big(\frac{N_s^K}{K}\Big|\frac{N_s^K}{K}-\langle\mu_s,1\rangle\Big|^2\Big)\leq\mathbb{E}\Big(\Big(\frac{N_s^K}{K}\Big)^2\Big)^{\frac{1}{2}}\mathbb{E}\Big(\Big|\frac{N_s^K}{K}-\langle\mu_s,1\rangle\Big|^4\Big)^{\frac{1}{2}},$$

where the first factor on the r.h.s. is controlled by Lemma 3.4.2. Thanks to Lemma 3.7.2 we obtain that

$$\mathbb{E}(\langle\eta_t^K,d_2\rangle)\leq C\int_0^t\mathbb{E}(\langle\eta_s^K,d_2\rangle)\,\mathrm{d}s+C\int_0^t\mathbb{E}\Big(\frac{N_s^K}{K}W_2^2(\bar{\nu}_s^K,\bar{\mu}_s)\Big)\,\mathrm{d}s+C_T\Big(I_4^2(K)+\frac{1}{\sqrt{K}}\Big).$$

Finally, Gronwall's lemma yields

$$\mathbb{E}(\langle\eta_t^K,d_2\rangle)\leq C_T\Big[I_4^2(K)+\frac{1}{\sqrt{K}}+\int_0^T\mathbb{E}\Big(\frac{N_s^K}{K}W_2^2(\bar{\nu}_s^K,\bar{\mu}_s)\Big)\,\mathrm{d}s\Big]e^{CT}.$$

□

We deduce the following result.

Corollary 3.7.1. *Condition* $(\mathbf{C})$ *3) holds.*

Proof. Applying Lemma 3.7.3, Lemma 3.3.2 and noting that $\frac{1}{\sqrt{K}} \leq CR_{d,q}(K)$, we obtain the bound

$$\mathbb{E}(\langle \eta_t^K, d_2 \rangle) \leq C_T\Big(I_4^2(K) + R_{d,q}(K)\Big). \tag{3.7.8}$$

Combining this with the inequality $\mathbb{E}\big(\frac{N_t^K}{K} W_2^2\left(\bar{\mu}_t^K, \bar{\nu}_t^K\right)\big) \leq \mathbb{E}\big(\frac{1}{K}\sum_{n=1}^{N_t^K} \|X_t^n - Y_t^n\|^2\big)$ yields the conclusion. ☐

Proof of Theorem 3.2. Conditions $(\mathbf{C})$ 1), 2) and 3) being established, it suffices to apply Lemma 3.4.1. ☐

We next provide the proof of the conditional propagation of chaos property stated in Corollary 3.2.1.

Proof of Corollary 3.2.1. Let $\Psi_{d,q}(K)$ denote the function of K appearing on the right hand side of the bound in Theorem 3.2. By exchangeability of $\Big((X_t^1, Y_t^1), \dots, \big(X_t^{N_t^K}, Y_t^{N_t^K}\big)\Big)$ conditionally on N_t^K, for all $t \geq 0$ we get

$$\mathbb{E}\bigg(\frac{N_t^K}{K}\|X_t^1 - Y_t^1\|^2\bigg) = \mathbb{E}\bigg(\frac{1}{K}\sum_{n=1}^{N_t^K}\|X_t^n - Y_t^n\|^2\bigg) \leq C_t \Psi_{d,q}^2(K), \tag{3.7.9}$$

thanks to (3.7.8). By Proposition 3.5.3, we have $\mathcal{L}\big(Y_t^1, \dots, Y_t^j \mid N_t^K\big) = \bar{\mu}_t^{\otimes j}$ on the event $\{j \leq N_t^K\}$. Now, letting $c_t := \langle \mu_t, 1\rangle \in (0, \infty)$ denote the limit in law of N_t^K/K, and using the second inequality of Lemma 3.3.1 in the third bound below we get, for all $\varepsilon > 0$, that

$$\begin{aligned}
&\mathbb{P}\bigg(\Big\|\mathcal{L}\Big(X_t^1, \dots, X_t^{j\wedge N_t^K} \,\Big|\, N_t^K\Big) - \bar{\mu}_t^{\otimes j}\Big\|_{\mathrm{BL}^*} > \varepsilon,\, N_t^K \geq j\bigg) \\
&\qquad \leq \mathbb{P}\Bigg(\frac{N_t^K}{K}\Big\|\mathcal{L}\Big(X_t^1, \dots, X_t^j \,\Big|\, N_t^K\Big) - \bar{\mu}_t^{\otimes j}\Big\|_{\mathrm{BL}^*}\bigg(\frac{N_t^K}{K}\bigg)^{-1} > \frac{\varepsilon c_t}{2}\frac{2}{c_t},\, N_t^K \geq j\Bigg) \\
&\qquad \leq \mathbb{P}\Bigg(\frac{N_t^K}{K}\Big\|\mathcal{L}\Big(X_t^1, \dots, X_t^j \,\Big|\, N_t^K\Big) - \bar{\mu}_t^{\otimes j}\Big\|_{\mathrm{BL}^*} > \frac{\varepsilon c_t}{2},\, N_t^K \geq j\Bigg) \\
&\qquad\qquad + \mathbb{P}\bigg(\frac{N_t^K}{K} < \frac{c_t}{2}\bigg) \\
&\qquad \leq \frac{2}{\varepsilon c_t}\mathbb{E}\Bigg(\frac{N_t^K}{K}\mathbb{E}\bigg(\sum_{n=1}^{j}\|X_t^n - Y_t^n\| \,\bigg|\, N_t^K\bigg)\mathbf{1}_{\{N_t^K \geq j\}}\Bigg) + \mathbb{P}\bigg(\frac{N_t^K}{K} < \frac{c_t}{2}\bigg) \\
&\qquad \leq \frac{2j}{\varepsilon c_t}\mathbb{E}\bigg(\frac{N_t^K}{K}\|X_t^1 - Y_t^1\|\bigg) + \mathbb{P}\bigg(\frac{N_t^K}{K} < \frac{c_t}{2}\bigg) \\
&\qquad \leq \frac{2j}{\varepsilon c_t} C_t' \Psi_{d,q}^2(K) + \mathbb{P}\bigg(\frac{N_t^K}{K} < \frac{c_t}{2}\bigg),
\end{aligned}$$

using also the Cauchy-Schwarz inequality, the estimate (3.7.9) and the fact that $\mathbb{E}(N_t^K/K)^{1/2} < \infty$ in the last inequality. Since $N_t^K/K \to c_t$ in law, the terms in the last line go to 0 when $K \to \infty$. The convergence $\mathbb{P}(N_t^K \geq j) \to 1$ then yields

$$\mathbb{P}\Big(\Big\|\mathcal{L}\Big(X_t^1, \ldots, X_t^{j\wedge N_t^K} \;\Big|\; N_t^K\Big) - \bar{\mu}_t^{\otimes j}\Big\|_{\mathrm{BL}^*} > \varepsilon \;\Big|\; N_t^K \geq j\Big) \longrightarrow 0$$

as $K \to \infty$ and the statement follows. □

We finish with some remarks regarding possible extensions of our approach, and the technical issues that must be solved in order to establish similar results in some related, more general settings.

Remark 3.7.1. *If instead of* $(\mathbf{H})$ 1) *it is assumed that the initial data* μ_0^K *satisfies the condition in Lemma 3.3.3 b), the arguments and construction leading to the proof of Theorem 3.2 must be modified, along the following lines:*

- *In condition* $(\mathbf{C})$ *1),* $\nu_0^K = \mu_0^K$ *is not enforced, but* $K\langle \nu_t^K, 1\rangle = K\langle \mu_t^K, 1\rangle = N_t^K$ *is kept.*
- *In the construction of the coupling using algorithm* $(\mathbf{A})$*, the random variables* $(Y^k)_{k\geq 1}$ *are chosen as before while, for any* K *and* N*, the random vectors* $(X_0^1, \ldots, X_0^N)$ *are chosen on the event* $\{N_0^K = N\}$*, suitably coupled with* $(Y_0^1, \ldots, Y_0^N)$*. This results in an extra term of the form* $\mathbb{E}(\langle \eta_0^K, d_2\rangle)$ *on the r.h.s. of the bounds in the statement and proof of Lemma 3.7.3 which in turn translates into an additional term* $C_T\mathbb{E}(\langle \eta_0^K, d_2\rangle)^{1/2}$ *on the r.h.s. of the bound in Theorem 3.2.*
- *In order to minimize the value of this additional term, the coupling of* $(X_0^1, \ldots, X_0^N)$ *and* $(Y_0^1, \ldots, Y_0^N)$ *must be chosen on each event* $\{N_0^K = N\}$ *so as to realize the squared Wasserstein−2 distance between the laws of* $(X_0^1, \ldots, X_0^N)$ *and* $\bar{\mu}_0^{\otimes N}$ *in* $(\mathbb{R}^d)^N$*. Denoting*

$$\widetilde{W}_2^2(\mathcal{L}(X_0^1, \ldots, X_0^N), \bar{\mu}_0^{\otimes N}) = \frac{1}{N} W_2^2(\mathcal{L}(X_0^1, \ldots, X_0^N), \bar{\mu}_0^{\otimes N}),$$

the normalized squared Wasserstein-2 distance, the additional term $\mathbb{E}(\langle \eta_0^K, d_2\rangle)^{1/2}$ *then writes*

$$\mathbb{E}\left(\frac{N_0^K}{K}\widetilde{W}_2^2\left(\mathcal{L}(X_0^1, \ldots, X_0^{N_0^K})|N_0^K), \bar{\mu}_0^{\otimes N_0^K}\right)\right)^{1/2}.$$

The following possible generalizations are left for future work:

- The extension of the ideas here developed, to populations with spatially or density depending birth or death events, as in the more general setting studied in [58], seems to be possible but presents one major additional difficulty, namely that the jump times are correlated with the spatial dynamics. The main consequence of this is that, in any coupling with some auxiliary system of conditionally independent (or less dependent) particles, the jump times cannot be expected to happen simultaneously. However, under the condition of spatial Lipschitz continuity of the reproduction rate and the competition kernel, it should be possible to keep at least some subsystems effectively coupled on finite time intervals, while controlling explicitly the discrepancy between jump times in the two systems, in terms of the distance of the empirical measures of the systems themselves, and in such a way that the discrepancies asymptotically vanish as the population size goes to infinity.

- A further desirable generalization regards the case of branching events more general than binary ones. The natural extension of the argument used here would consist in coupling all the offspring of a branching particle in the original system, with a set of equally many independent new particles given birth at the same time in the auxiliary system. However it is not clear how to make compatible the use of optimal transport plans to couple the branching particle and the positions of the new particles in the auxiliary system, with the independence requirement in the auxiliary system. A possible way of coping with this problem could be to make a two-steps coupling construction: first, between the branching particle in the original system and the positions of new particles in the auxiliary system (which would define an exchangeable random vector of particles in any case) and, in a second step, coupling those positions with independent particles with the required law.

B Appendix

Proof of Lemma 3.3.3. Since condition $(\mathbf{H})\ 1)$ assumed in a) is a particular case of the assumptions in b), it is enough to prove b) to get both parts. Taking $\mu = \mu_0$ and $\nu = \mu_0^K$ in Lemma 3.3.1, we get

$$\limsup_K \mathbb{P}(\|\mu_0 - \mu_0^K\|_{\mathrm{BL}^*} \geq \varepsilon) \leq \limsup_K \mathbb{P}(\|\bar{\mu}_0 - \bar{\mu}_0^K\|_{\mathrm{BL}^*} \geq \varepsilon/(2\langle \mu_0, 1\rangle)), \tag{B.1}$$

with $\bar{\mu}_0^K = \frac{1}{N_0^K}\sum_{i=1}^{N_0^K} \delta_{X_0^i}$. On the other hand, for each $\delta > 0$ and $M > 0$,

$$\begin{aligned}\mathbb{P}(\|\bar{\mu}_0 - \bar{\mu}_0^K\|_{\mathrm{BL}^*} \geq \delta) &\leq \sum_{N \geq M} \mathbb{E}\left[\mathbb{P}(\|\bar{\mu}_0 - \bar{\mu}_0^K\|_{\mathrm{BL}^*} \geq \delta | N_0^K = N)\mathbf{1}_{N_0^K = N}\right] + \mathbb{P}(N_0^K < M) \\ &\leq \sup_{N\geq M} \mathbb{P}\left(\left\|\bar{\mu}_0 - \frac{1}{N}\sum_{i=1}^N \delta_{Y^{i,N}}\right\|_{\mathrm{BL}^*} \geq \delta\right) + \mathbb{P}(\langle \mu_0^K, 1\rangle < M/K).\end{aligned}$$

Since $\langle \mu_0^K, 1\rangle$ converges weakly to a non null constant, the last term goes to 0 when $K \to \infty$. On the other hand, it is well known that the assumed $\bar{\mu}_0-$chaoticity is equivalent to the convergence in distribution of the random probability $\frac{1}{N}\sum_{i=1}^N \delta_{Y^{i,N}}$ to $\bar{\mu}_0$ as $N \to \infty$. If follows that $\limsup_{K\to\infty} \mathbb{P}(\|\bar{\mu}_0 - \bar{\mu}_0^K\| \geq \delta) = 0$ which entails the claim in view of (B.1).

c) The r.v. $N_0^K = K\langle \mu_0^K, 1\rangle$ is Poisson of parameter $K\langle \nu_0, 1\rangle$ and equals in law the partial sum $\sum_{i=1}^K N^i$ of i.i.d. Poisson r.v. $(N^i)_{i\in\mathbb{N}}$ of parameter $\langle \nu_0, 1\rangle$. By the law of large numbers, $\langle \mu_0^K, 1\rangle = N_0^K/K$ converges in law to the constant $\langle \nu_0, 1\rangle$. It is immediate from basic properties of Poisson point measures that the N_0^K atoms of μ_0^K are i.i.d. of law $\bar{\nu}_0$ given $\langle \mu_0^K, 1\rangle$, and we necessarily have $\mu_0 = \langle \nu_0, 1\rangle\bar{\nu}_0 = \nu_0$. Lastly, the r.v. N_0^K is Poisson of parameter $K\langle \mu_0, 1\rangle$ and so we have $I_4^4(K) = K^{-3}\left(\langle \mu_0, 1\rangle + 3K\langle \mu_0, 1\rangle^2\right) \leq CK^{-2}$. $\square$

CHAPTER 4

Large population approximation of the genealogy of branching processes in varying environments

Contents

4.1 Introduction

Models for the growth of the size of a population are of great importance in population dynamics, since they serve for example to study the development of population under abundance of resources, the probability of fixation in biological invasion or the different scenarios that arise in epidemiological settings. These models rely on independence of individuals and the so-called branching property. They range from discrete state versions, classically known as Galton-Watson processes, to their continuous state counterpart arising in their large population approximations on large time scales (see for example [7]). These lasts processes have been intensively studied in the last decades (we refer to [88] or [102] for their treatment). Also, their generalizations to varying or random environments have received attention, starting in [77] and more recently in [11], to name but a few. An important class of these models are the ones that have been constructed as solutions of stochastic differential equations (see [100, 66] and [55]). In general, these models allow to take into account the variability of the environment into the dynamics. This gives a general framework for studying questions of great importance in ecology and population dynamics, such as long term behaviour [101] or the effects of catastrophic events and extinction (see for example [10] and [9]).

Another question which is of great interest in population dynamics, and in particular in models having reproductive dynamics, is to understand the genealogical structure behind the branching process modeling its growth. In the case of finite population models or discrete state process this information is implicit since we can trace the genealogical history of each individual, but this becomes a more complex procedure when we pass to large population approximations, as we loose the notion of a single individual. Throughout the years different approaches have been developed in order to understand the genealogy behind branching processes. Starting with the famous continuum random tree introduced in [1], which can be obtained as the scaling limit of Galton-Watson trees, and later with its generalizations such as Lévy trees, which are random trees models that correspond to general branching processes [51], or the flow of bridges representation introduced in [14] and developed further in [15, 16]. Another approaches are spinal constructions [65], when there is a spatial component, or the splitting trees approach [82].

One particular and useful approach for population models, in what respects to the description of their genealogical structure, is the one given by lookdown constructions. The original lookdown construction was first introduced in [48] (later improved in [50]) as a way of constructing an infinite (yet countable) particle system having the same distribution as a Fleming-Viot process. This construction provides new insight about the genealogical structure of the process given this countable representation, simplifying its analysis. Later, in [80] a construction in this spirit was given for producing countable representations of measure-valued branching processes, allowing the branching rates to depend on the particles' spatial positions. A useful property of lookdown constructions is that when passing from finite population models to their high density limits, the genealogies are preserved by a projective property of the associated martingale problem. More recently, in [52] a more complete toolbox was given in order to construct representations of models presenting a variety of mechanisms of evolution, in particular for models showing interaction between its individuals. In general, the lookdown approach has been proved to be a powerful way for constructing processes augmented with a genealogical structure (see for example [103, 18, 113] and references therein).

The aim of this work is to describe the tree spanned by the population alive at a certain time, which is known as the *reduced tree*, in a large population regime for branching processes in varying environments. This object is also known as the reconstructed tree in phylogenetics or as the coalescent tree, seen as the tree generated backwards by the population alive at the present time (related works concerning this object are [84] and [83]). An associated object to this tree that is worth mentioning is its width process, called the reduced process, which has also been object of study (see [57, 20, 115, 112] and [73]).

The description of the tree is achieved by means of a lookdown construction, which allows to exploit a coupling in a Poissonian framework and quantify approximations. Using the lookdown construction introduced in [80], we give an approximation of the reduced tree for a birth and death process in varying environment by its large population approximation, namely a Feller diffusion in varying environment. Specifically, we develop an approach to approximate the reduced tree of a birth and death process in the large population regime, by the reduced tree associated with the branching process characterized by the following stochastic differential equation

$$X_t = X_0 + \int_0^t b(s)X_s \,\mathrm{d}s + \int_0^t \sqrt{2\sigma(s)X_s}\,\mathrm{d}B_s - \sum_{j\geq 1,\, t\geq t_j} F_{t_j} X_{t_j^-}, \tag{4.1.1}$$

where $(t_j)_{j\geq 1}$ are given catastrophes times and F_{t_j} the fraction of the population that dies at time t_j, and this for $b(t) \leq 0$ and $\sigma(t)$ bounded away from zero. This last process arises as the large population approximation of a birth and death process in where individuals have a small mass and also reproduce and die very fast. Moreover, since the rates are big, the variance of the increments persists in the limit, thus making appear the diffusive term. We have also the effect of the environment through the reproduction rates of individuals, affecting the mean behaviour and the diffusivity, and by punctual catastrophic events. With this structure and parameters, the process belongs to the family of critical or subcritical branching processes in varying environments.

By considering the lookdown representation of the birth and death process that approximates the previous object, we proceed by filtrating each birth event in the particle process by the probability of survival of the offspring until a given time $T > 0$. Using this procedure, we construct a random chronological tree which takes values in a particular space endowed with a distance that is tailored for the analysis of these objects. Then, by similar arguments we construct a limit candidate and we show the convergence under the distance previously introduced. This is done by a coupling argument that evokes the nature of the lookdown representation.

We next provide a detailed description of the models under consideration, define the main objects and state our main result.

4.2 Models and main result

In what follows we recall the framework introduced in [80].

4.2.1 Lookdown constructions

Consider the state space $E = \bigcup_{n=0}^{\infty}[0,K]^n$, the domain

$$D(A^K) = \left\{ f(u) = \prod_{i=1}^{n} g(u_i) : g \in C^1([0,K]), 0 \leq g \leq 1, g(K) = 1, g'(K) = 0 \right\}, \tag{4.2.1}$$

where $u = (u_1, \ldots, u_n)$, and the generator

$$A_t^K f(u) = f(u) \sum_{i=1}^{n} 2\sigma(t) \int_{u_i}^{K} (g(v) - 1)\mathrm{d}v + f(u) \sum_{i=1}^{n} \Big(\sigma(t) u_i^2 - b(t) u_i \Big) \frac{g'(u_i)}{g(u_i)}, \tag{4.2.2}$$

for $\sigma(t) \geq 0$ and $b(t)$ two bounded functions. This operator represents the following stochastic dynamics. We start with a collection of particles where each one has a real value assigned, called *level.* The level associated to each particle evolves in time according to

$$u_i'(t) = \sigma(t)u_i(t)^2 - b(t)u_i(t). \tag{4.2.3}$$

A particle with level u at time t, produces a new particle with uniformly distributed level in the interval $[u, K]$ at rate $2\sigma(t)(K-u)$. When a particle's level reaches the value K, it is removed from the system.

This process is the lookdown construction for a birth and death process, introduced in [80]. The justification of this representation comes from the Markov mapping theorem (see Theorem 4.2 in Appendix C). Indeed, let $\alpha_K(n, \mathrm{d}u)$ be the joint distribution of n i.i.d. uniform $[0, K]$ random variables and $\hat{f} = \int f(u)\,\alpha_K(n, \mathrm{d}u)$. By defining $C_t^K \hat{f}(n) := \int A_t f(u)\,\alpha_K(n, \mathrm{d}u)$, it can be shown that

$$C_t^K \hat{f}(n) = K\sigma(t)n(\hat{f}(n+1) - \hat{f}(n)) + (K\sigma(t) - b(t))n(\hat{f}(n-1) - \hat{f}(n)). \tag{4.2.4}$$

Under the condition $K\sigma(t) - b(t) \geq 0$ for each $t \geq 0$, we obtain the generator of a birth and death process. Thanks to the Markov mapping theorem, since there is existence of solutions for the martingale problem associated with this generator, we have existence for the martingale problem associated with (4.2.2). Another consequence of this theorem is that if we have uniqueness for the martingale problem associated with (4.2.2), we obtain uniqueness for the martingale problem associated to (4.2.4). Finally, the theorem also allows us to conclude that the number of particles whose levels are below K is equal in distribution to the solution of the martingale problem associated with (4.2.4).

We notice that the solution of the martingale problem associated with (4.2.4), once renormalized, can be approximated by the Feller diffusion. This remark leads us to analyze the generator (4.2.2) when $K \to \infty$, which gives us another generator that we introduce next.

Let now $E = [0,\infty)^\infty \cup \bigcup_{n=0}^\infty [0,\infty)^n$ and consider the domain

$$D(A) = \left\{ f(u) = \prod_{i>0} g(u_i) : g \in C^1([0,\infty)), 0 \leq g \leq 1, \exists v_g \text{ such that } g(v) = 1, \forall v \geq v_g \right\},$$

and the operator defined by

$$A_t f(u) = f(u) \sum_i 2\sigma(t) \int_{u_i}^{v_g} \Big(g(v) - 1 \Big)\,\mathrm{d}v + f(u) \sum_i \Big(\sigma(t)u_i^2 - b(t)u_i \Big) \frac{g'(u_i)}{g(u_i)}. \tag{4.2.5}$$

This operator represents a slightly different dynamics than the one previously introduced. A particle with level u at time t, will produce a new particle with uniformly distributed level in the interval $[u + \ell_1, u + \ell_2]$ for $0 \leq \ell_1 \leq \ell_2$, at rate $2\sigma(t)(\ell_2 - \ell_1)$. This time, a particle is removed from the system when its level reach the value infinite.

In this case, by letting $\alpha(z, \cdot)$ be the distribution of a Poisson process on $[0, \infty)$ with intensity z and by defining $C_t \hat{f}(z) := \int A_t f(y)\,\alpha(z, \mathrm{d}y)$, we obtain that

$$C_t \hat{f}(x) = \sigma(t) x \hat{f}''(x) + b(t) x \hat{f}'(x). \tag{4.2.6}$$

Then, by checking the existence of solutions for the martingale problem associated with this last generator, which can be done by explicitly constructing a process with such generator, we obtain as a consequence of the Markov mapping theorem that

$$\lim_{N\to\infty} \frac{1}{N} \sum_i \mathbf{1}_{[0,N]}\left(V_t^i\right) \overset{(d)}{=} X_t,$$

where $(X_t)_{t\geq 0}$ is a process that evolves according to the generator (4.2.6) and $(V_t)_{t\geq 0}$ the solution of the martingale problem associated to (4.2.5). Moreover, thanks to the same theorem, we also obtain that at time t the distribution of the levels is conditionally Poisson with mean X_t.

The connection between the two particle processes is explicit. By considering the solution of the martingale problem for (4.2.5) and taking only into account the particles with levels below K, we recover a solution of the martingale problem for (4.2.2), simply by restricting the domain of the operator. As a consequence, the genealogies are embedded in a projective way, which is one of the key features of the lookdown construction.

In this work we will also consider the effect of catastrophes in the evolution of the two particle processes. For this, we let $(t_j)_{j\geq 1}$ and $(m_{t_j})_{j\geq 1}$ be given catastrophes times and intensities, and $\beta(m) \geq 1$ a modulating function for the catastrophes. Given the initial conditions, we let the processes evolve according to (4.2.5) and (4.2.2) respectively. When we arrive at a catastrophe time t_j we amplify the level of each particle in both processes by a factor $\beta(m_{t_j})$. Then, we restart the dynamics specified by the generators (4.2.5) and (4.2.2) with the current states as initial conditions. Following this construction, we denote the two resulting processes by $(U_t)_{t\geq 0}$ and $(U_t^K)_{t\geq 0}$ respectively.

With this construction we obtain that until the first jump time (catastrophic event), the number of particles in $(U_t^K)_{t\geq 0}$ and $(U_t)_{t\geq 0}$, renormalized in this last case, follow the evolution given by the generators (4.2.4) and (4.2.6) respectively. At a catastrophic event, the update rule used translates into removing each particle independently with probability $\beta(m_{t_j})^{-1} \leq 1$ in the birth and death process being represented by $(U_t^K)_{t\geq 0}$, as it was shown in [80]. At the same time, in the branching process represented by $(U_t)_{t\geq 0}$, a catastrophic event corresponds to multiply by $\beta(m_{t_j})^{-1} \leq 1$ the size of the population, yielding the effect of a dramatic event that kills a fraction of the population. Then, since in the lookdown construction we can start from any initial condition, by restarting the dynamics the same construction ensures that we obtain a representation for the inhomogeneous branching processes after the jump. We thus obtain a process $(U_t)_{t\geq 0}$, in where its renormalized size is equal in distribution to (4.1.1), with $F_{t_j} := -(\beta(m_{t_j})^{-1} - 1)$. This can be seen as the quenched version of the model proposed in [80] for random catastrophes.

4.2.2 The reduced tree

Recalling that $b, \sigma\colon \mathbb{R} \to \mathbb{R}$ are continuous and bounded functions, and that $\beta\colon \mathbb{R} \to \mathbb{R}_+$ such that $\beta > 1$, we further assume that $b(t) \leq 0 < \sigma(t)$ for each $t \geq 0$.

Consider $T > 0$. Given the processes $(U_t^K)_{t\geq 0}$ and $(U_t)_{t\geq 0}$ defined in the previous section, we construct their reduced trees at time T as follows. First, let u_T^K and u_T^∞ be the solutions of

$$u(t) = u(s) + \int_s^t \sigma(r)u(r)^2 \,\mathrm{d}r - \int_s^t b(r)u(r)\,\mathrm{d}r + \sum_{j\geq 1,\, t\geq t_j\geq s} (\beta(m_{t_j}) - 1)u(t_j^-), \tag{4.2.7}$$

for each $0 \leq s \leq t < T$, with terminal conditions $u_T^K(T) = K$ and $\lim_{t\to T} u_T^\infty(t) = \infty$, respectively. The existence of these solutions is ensured by the condition $b(t) \leq 0 < \sigma(t)$.

We then define $\mathbf{R}_T^K$ as the tree generated by the particles starting with levels below $u_T^K(0)$ and evolving according to (4.2.7). In this tree, a particle with level u at time t will produce a new particle at rate $2\sigma(t)(u_T^K(t) - u)$ with uniformly distributed level on $[u, u_T^K(t)]$. Similarly, $\mathbf{R}_T$ will be defined as the tree generated by the particles whose levels are below $u_T^\infty(0)$ at $t = 0$. Here, a particle with level u at time t will produce a new particle at rate $2\sigma(t)(u_T^\infty(t) - u)$ with uniformly distributed level on $[u, u_T^\infty(t)]$. These definitions come from filtrating each birth event in the processes $(U_t^K)_{t\geq 0}$ and $(U_t)_{t\geq 0}$ by their probability of survival until time T.

To endow this objects with a genealogical structure we start by considering the set $\mathcal{U} = \bigcup_{n\geq 0}(\mathbb{N}^*)^n$. At time 0 we enumerate the particles according to the increasing order of their levels and we label them according to this numbering. We then consider for each particle the product set of its label times its lifetime, which is $[0, T]$, and we take the union of all of these sets. When a new particle appears at time t', we label it following the Ulam-Harris-Neveu formalism [98] and we consider the product set of its label times its lifetime, which is $[t', T]$, and finally we take the union of this set with the set that we already had. Iterating this procedure will yield a random chronological tree that is a subset of $\mathcal{U} \times [0, T]$. These objects are also known as (inhomogeneous) splitting trees (see [34] for more on these objects). Furthermore, since we are interested in the closeness of the two objects, the structure of the trees involved motivates the introduction of the following distance in order to compare them

$$d_g^T\Big(\mathbf{T}, \mathbf{T'}\Big) = \sum_{u\in\mathcal{U}} \int_0^T \Big|\mathbf{1}_{(u,s)\in\mathbf{T}} - \mathbf{1}_{(u,s)\in\mathbf{T'}}\Big| g(s)\,\mathrm{d}s,$$

where $g\colon [0, T] \to \mathbb{R}_+$ is a function satisfying the integrability condition

$$\int_0^T e^{\int_0^s 2\sigma(r)u_T^\infty(r)\,\mathrm{d}r} g(s)\,\mathrm{d}s < \infty. \tag{4.2.8}$$

This distance can be seen as a weighted total variation distance between splitting trees.

Our main result is then an approximation of $\mathbf{R}_T^K$ by $\mathbf{R}_T$ for large K, representing the approximation of the reduced tree for a birth and death process in varying environment by an object that plays the role of the reduced tree for the Feller diffusion in varying environment. Assuming that the process (4.1.1) starts from $x \in \mathbb{R}_+$, the result is stated as follows.

Theorem 4.1. *Let $T > 0$. Suppose that the functions b and σ are continuous, bounded and such that $b(t) \leq 0$, $\sigma(t) \geq 0$ and σ is bounded away from zero. Then, we have that*

$$\mathbb{E}\Big(d_g^T\Big(\mathbf{R}_T, \mathbf{R}_T^K\Big)\Big) \to 0, \quad \text{as } K \to \infty.$$

Given the assumptions, this result holds for critical and subcritical branching processes since we need to assume $b(t) \leq 0$ in order to ensure the required properties of (4.2.7). Also, we are forced to chose g satisfying (4.2.8) since as we approach the final time T the number of individuals in the trees explode, which leads to having to control an infinity of small time intervals when computing the distance.

The proof of Theorem 4.1 relies on a coupling between the two objects, which allows to exploit the projective property of the lookdown construction that we make use of. This approach sets the ground for in a next step take advantage of the Poissonian framework in order to obtain quantitative estimates. Also, Theorem 4.1 is expected to generalize to the case of random catastrophes, in time and intensity, following a Poissonian law.

The rest of the chapter is organized as follows. In Section 4.3, we give preliminary definitions and set the formalism of chronological trees, specifically the space in which we are going to consider the trees to be elements of and the distance that we use. In Section 4.4, we give the proper construction of the trees and we construct a coupling between them in order to prove the main result. Finally, in the Appendix C we recall the Markov mapping theorem and how it applies in the inhomogeneous setting.

4.3 Preliminaries on trees

In this section, we introduce the necessary tools for giving mathematical sense to the genealogical information arising in our context.

We start by defining the notion of a discrete tree. Such object can be coded by the Ulam-Harris-Neveu [98] formalism. Let

$$\mathcal{U} = \bigcup_{n\geq 0} (\mathbb{N}^*)^n,$$

be the set of finite sequences of positive integers, where $(\mathbb{N}^*)^0 = \{\varnothing\}$. The root of the tree is denoted by $\varnothing$ and each vertex of the tree is represented by a finite sequence of the form $v = (v_1, \dots, v_m) \in (\mathbb{N}^*)^m$. We denote the i-th child of v by vi, were vw denotes the concatenation for $v, w \in \mathcal{U}$. A discrete tree $\mathbb{T}$ is a subset of $\mathcal{U}$ that satisfies

(i) $\varnothing \in \mathbb{T}$.
(ii) If $vj \in \mathbb{T}$, where $j \in \mathbb{N}^*$, then $v \in \mathbb{T}$.
(iii) If $v \in \mathbb{T}$, then $vj \in \mathbb{T}$ if and only if $1 \leq j \leq \delta_v(\mathbb{T})$, for a positive integer $\delta_v(\mathbb{T})$.

We will say that w is an ancestor of v when there is a sequence $z \in \mathcal{U}$ such that $v = wz$. We denote this relation by $w \prec v$. Furthermore, we denote by $v \wedge w$ the most recent common ancestor of v and w, which is the longest sequence $z \in \mathcal{U}$ such that $z \prec v$ and $z \prec w$.

Next, we are interested in augmenting the information encoded in the tree. If we want to take into account the lifetime of individuals, the framework given by chronological trees is the adequate. This kind of trees are particular cases of $\mathbb{R}$-trees, which in turn are abstract complete metric spaces satisfying two properties that characterize the natural notion of tree: there is an unique path between two points and there are no loops (see for example [54]).

Definition 4.3.1. A chronological tree $\mathbf{T}$ is a subset of $\mathcal{U} \times \mathbb{R}_+$ such that

(i) $\rho := (\varnothing, 0) \in \mathbf{T}$.
(ii) $\mathbb{T} := P_{\mathcal{U}}(\mathbf{T}) = \{v \in \mathcal{U} : \exists s \geq 0, (v, s) \in \mathbf{T}\}$ is a discrete tree.
(iii) $\forall v \in \mathbb{T}$, there exists $0 \leq b(v) < d(v) \leq \infty$ such that $(v, t) \in \mathbf{T}$ if and only if $t \in (b(v), d(v)]$.
(iv) $\forall v \in \mathbb{T}$ and $i \in \mathbb{N}^*$ such that $vi \in \mathbb{T}$, we have that $b(vi) \in (b(v), d(v))$.
(v) $\forall v \in \mathbb{T}$ and $i, j \in \mathbb{N}^*$ such that $vi, vj \in \mathbb{T}$, $i \neq j \Rightarrow b(vi) \neq b(vj)$.

See Figure 4.1 for a graphical representation of a chronological tree.

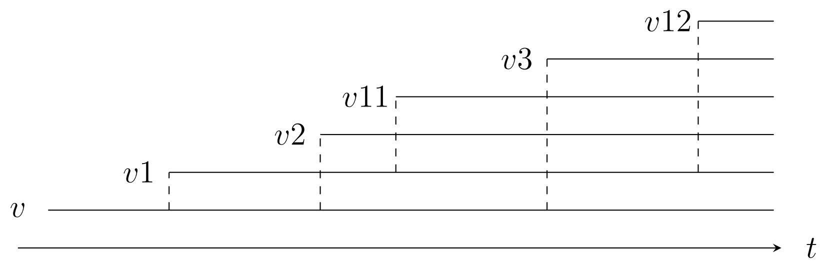

Figure 4.1: Example of chronological labelling for the progeny of a given label $v \in \mathcal{U}$.

Given a chronological tree $\mathbf{T}$, we define the set of individuals living at time t as

$$V_t(\mathbf{T}) = \{v \in \mathcal{U} : (v,t) \in \mathbf{T}\}.$$

On the other hand, for any (v,s) and (w,t) belonging to $\mathbf{T}$, we say that (v,s) is an ancestor of (w,t), denoted $(v,s) \prec (w,t)$ as before, if $v \prec w$, and $s \le t$ if $v = w$ or $s \le b(vj)$ if $v \neq w$, where j is the unique integer such that $vj \prec w$.

As we are interested in trees at a given time $T > 0$, we will restrict the analysis to trees that are subsets of $\mathcal{U} \times [0,T]$. In our context it will be also important to be able to quantify the distance between two chronological trees. In order to do this we start by defining the set in which we are going to be looking the trees as elements of. For this, we consider the set

$$\mathcal{T}^g := \left\{ \mathbf{T} \subset \mathcal{U} \times [0,T] : \int_0^T N_{\mathbf{T}}(s) g(s)\, \mathrm{d}s < \infty \right\},$$

where $N_{\mathbf{T}}(t) := |V_t(\mathbf{T})|$ represents the number of individuals or labels in the tree at time t and $g \colon [0,T] \to \mathbb{R}_+$ is a given function.

Definition 4.3.2. Given two chronological trees $\mathbf{T}, \mathbf{T}' \in \mathcal{T}^g$, their g-weighted chronological distance is defined by

$$d_g^T\Big(\mathbf{T}, \mathbf{T}'\Big) = \sum_{u \in \mathcal{U}} \int_0^T \Big| \mathbf{1}_{(u,s) \in \mathbf{T}} - \mathbf{1}_{(u,s) \in \mathbf{T}'} \Big| g(s)\, \mathrm{d}s.$$

The idea behind introducing this notion of distance is that we want to compare the trees by taking into account the differences between the labels of individuals in each tree and their living times, which can be obtained through the evolution of their levels. Moreover, we remark that $(\mathcal{T}^g, d_g^T)$ is a metric space.

In order to see how strong is the distance that we introduced, we can compare it to a known distance between metric spaces. The natural framework in this setting is the one given by the Gromov-Hausdorff distance. Recall that given a metric space $(\mathcal{X}, d_{\mathcal{X}})$, the Hausdorff distance between two closed subsets $A, B \subset \mathcal{X}$ is defined by

$$d_{\mathrm{H}}(A,B) := \inf\{\varepsilon > 0 : A \subset B^\varepsilon \text{ and } B \subset A^\varepsilon\},$$

where $U^\varepsilon := \{x \in \mathcal{X} : d_{\mathcal{X}}(x,U) < \varepsilon\}$ denotes the ε-neighborhood of $U \subset \mathcal{X}$. The Gromov-Hausdorff distance between two metric spaces is then defined by

$$d_{\mathrm{GH}}\Big((\mathcal{Y}, d_{\mathcal{Y}}), (\mathcal{Z}, d_{\mathcal{Z}})\Big) := \inf\Big\{ d_{\mathrm{H}}(\phi(\mathcal{Y}), \psi(\mathcal{Z}))\Big\}$$

where the infimum is taken over all isometric embeddings $\phi : \mathcal{Y} \to \mathcal{X}$ and $\psi : \mathcal{Z} \to \mathcal{X}$.

Given a chronological tree $\mathbf{T}$ we can endow it with a natural distance. Let us consider a function $q : [0,T] \to \mathbb{R}_+$ and let $p : \mathcal{U} \times [0,T] \to \mathbb{R}_+$ be given by $p((v,t)) = \int_0^t q(s)\mathrm{d}s$. By defining the metric

$$d(w,v) = p(w) + p(v) - 2p(w \wedge v),$$

$(\mathbf{T}, d)$ turns out to be a complete metric space.

By setting $q \coloneqq g$, we have the following comparison between metrics.

Lemma 4.3.1. *Let $\varepsilon > 0$. If $d_g^T\Big(\mathbf{T}, \mathbf{T}'\Big) \le \varepsilon$ then $d_{\mathrm{GH}}\Big(\mathbf{T}, \mathbf{T}'\Big) \le \varepsilon$.*

Proof. Consider the space $(\mathbf{T} \cup \mathbf{T}', d)$. Given any element $\boldsymbol{v} = (v,t) \in \mathbf{T}'$, we need to show that $\boldsymbol{v} \in \mathbf{T}^\varepsilon$. Indeed, suppose that $\boldsymbol{v} \notin \mathbf{T}$. The distance from $\boldsymbol{v}$ to its progenitor $\hat{\boldsymbol{v}}$ is given by $\int_{b(v)}^t g(s)\,\mathrm{d}s$, which is smaller than ε since $d_g^T\Big(\mathbf{T}, \mathbf{T}'\Big) \le \varepsilon$. If $\hat{\boldsymbol{v}} \in \mathbf{T}$, then it is clear that $\boldsymbol{v} \in \mathbf{T}^\varepsilon$, since we can add the branch $\{v\} \times [b(v), t]$ without enlarging the tree further than ε under the distance d. If on the contrary $\hat{\boldsymbol{v}} \notin \mathbf{T}$, then we can add the two branches $\{\hat{v}\} \times [b(\hat{v}), t] \cup \{v\} \times [b(v), t]$, and since $\int_{b(\hat{v})}^t g(s)\,\mathrm{d}s + \int_{b(v)}^t g(s)\,\mathrm{d}s < \varepsilon$ thanks to $d_g^T\Big(\mathbf{T}, \mathbf{T}'\Big) \le \varepsilon$, the same reasoning as before applies, yielding thus that $\boldsymbol{v} \in \mathbf{T}^\varepsilon$. Repeating this argument inductively until arriving to the root gives the inclusion, and the same applies for the other direction of the inclusion for the same ε. □

Poissonian construction of a chronological tree

We end this section by giving a construction of a chronological tree from a collection of Poisson point measures.

Let $(\mathcal{N}_v)_{v \in \mathcal{U}}$ be an i.i.d. collection of Poisson random measures with common intensity $\kappa(t)\mathrm{d}s$. Using these measures, we can recursively define a sequence of trees whose union is a chronological tree in where each individuals lives up to time T. Starting with $\mathbf{T}_1 = \{\varnothing\} \times [0,T]$ and $b(\varnothing) = 0$, we define

$$\mathbf{T}_n = \bigcup_{v \in P_{\mathcal{U}}(\mathbf{T}_{n-1})} \bigcup_{i \ge 1}^{\mathcal{N}_v((b(v),T]} \{vi\} \times (b(vi), T] \subset \mathcal{U} \times \mathbb{R}_+,$$

where $P_{\mathcal{U}}(\cdot)$ is the projection on $\mathcal{U}$. The birth times are given by

$$\begin{aligned} b(v1) &= \inf\{t > 0 \mid \mathcal{N}_v((b(v), b(v)+t]) > 0\}, \quad \text{if } \mathcal{N}_v((b(v),T]) > 0, \\ b(vi) &= \inf\{t > b(v(i-1)) \mid \mathcal{N}_v((b(v), b(v)+t]) > i\}, \quad 2 \le i \le \mathcal{N}_v((b(v),T]). \end{aligned}$$

By defining $\mathbf{T} = \bigcup_{n \ge 1} \mathbf{T}_n$ we obtain a chronological tree.

4.4 Construction of the trees and coupling

In this section, we describe in detail the construction of the reduced trees behind the particle models involved. Then, we construct a coupling between the trees which allows us to prove the main result.

In what follows we consider σ, b, β and $(t_j, m_{t_j})_{j\geq 1}$ as given in Section 4.2.

Following for example the results of [57], in order to obtain the reduced process associated with a discrete branching process observed until a given time $T > 0$, that is the process that at each time $t < T$ counts the number of individuals having progeny alive at time T, at each birth event we need to filtrate each new particle according to its probability of survival until time T. With this approach, recovering the reduced tree from the lookdown construction for a birth and death process with catastrophes turns out to be an explicit procedure. Indeed, given $K \in \mathbb{N}^*$, if we consider $(U_t^K)_{t\geq 0}$ the process defined by (4.2.2), filtrating a new particle in this system according to its survival probability corresponds to look at its starting level in order to see if it will die before time T. Specifically, we had that in this process the threshold value for the level of the particles is K, meaning that when the level of a particle reaches this value the particle is removed from the system. Then, in order to keep only the particles that are alive at time T, we need to look at the solution of (4.2.3), including the upward jumps due to catastrophes, taking the value K exactly at time T, and consider all the particles whose levels' remain below this curve. This procedure is then repeated analogously for the process defined by (4.2.5) and the catastrophes, using this time the solution of (4.2.3) including jumps and reaching ∞ at time T.

In what follows we will state some preliminary facts about the evolution of the levels that will be used for the construction of the trees.

4.4.1 Level evolution and threshold

Recall that the level of each particle satisfies the equation

$$u(t) = u(s) + \int_s^t \sigma(r)u(r)^2 \,\mathrm{d}r - \int_s^t b(r)u(r)\,\mathrm{d}r + \sum_{j\geq 1,\, t\geq t_j\geq s} (\beta(m_{t_j}) - 1)u(t_j^-), \tag{4.4.1}$$

which in between jumps translates as the evolution given by

$$u'(t) = \sigma(t)u(t)^2 - b(t)u(t). \tag{4.4.2}$$

Lemma 4.4.1. *Given $0 \leq s$ and $\overline{u} \geq 0$, the solution of equation (4.4.2) such that $u(s) = \overline{u}$ is given by*

$$u(t) = \frac{\overline{u}e^{-\int_s^t b(r)\,\mathrm{d}r}}{1 - \overline{u}\int_s^t e^{-\int_s^r b(w)\,\mathrm{d}w}\sigma(r)\,\mathrm{d}r}, \tag{4.4.3}$$

for each $s \leq t < T(s, \overline{u})$, where $T(s, \overline{u})$ is the explosion time determined by

$$\int_s^{T(s,\overline{u})} e^{-\int_s^r b(w)\,\mathrm{d}w}\sigma(r)\,\mathrm{d}r = \frac{1}{\overline{u}}.$$

Moreover, $u(t)$ is increasing, nonnegative and the flow associated to (4.4.2) is strictly monotone with respect to the initial conditions.

Proof. By direct computation it can be checked that (4.4.3) satisfies (4.4.2). On the other hand, given two non-negative initial conditions $u^1(0)$ and $u^2(0)$ such that $u^1(0) \leq u^2(0)$, thanks to the monotony of $x \mapsto 1/(1-x)$, it can be deduced that $u^1(t) \leq u^2(t)$. Furthermore, we obtain the strict monotonicity by local uniqueness of the equation, since if two solutions coincide at some point they are forced to be equal in all previous times and until right before the explosion time. Finally, since the constant function $u(t) = 0$ is a solution of (4.4.2), we obtain the positivity. $\square$

From here we notice that the condition $b(t) \leq 0 < \sigma(t)$ ensures two things, first that the solutions of (4.4.2) are increasing, and secondly, that the explosion time is finite. Furthermore, notice that the explosion time is monotone with respect to the initial condition.

Comparing the two solutions associated to these initial conditions yields the following result.

Lemma 4.4.2. *Given $0 \leq \overline{u}_1 \leq \overline{u}_2$ and $s \geq 0$, consider the two solutions $u_1(t)$ and $u_2(t)$ of (4.4.2) associated with these initial conditions given by (4.4.3). For each $t \in [s, T(s, u_2))$ we have*

$$u_2(t) - u_1(t) \leq \Big(\overline{u}_2 - \overline{u}_1\Big) e^{\int_s^t \left(|b(r)| + 2|\sigma(r)| u_2(r)\right) \mathrm{d}r}. \tag{4.4.4}$$

Furthermore, if $\overline{u}_1 \nearrow \overline{u}_2$ then $u_1(t)$ converges uniformly to $u_2(t)$ in each compact subset of $[s, T(s, u_2))$.

Proof. Thanks to the monotony of solutions we have that $T(u_2, s) \leq T(u_1, s)$, so both functions are well defined in the stated interval. Now, given $t \in [s, T(s, u_2))$ we have that

$$\begin{aligned} u_2(t) - u_1(t) &= \overline{u}_2 - \overline{u}_1 + \int_0^t \sigma(s)(u_2(s)^2 - u_1(s)^2) \, \mathrm{d}s - \int_0^t b(s)(u_2(s) - u_1(s)) \, \mathrm{d}s \\ &\leq \overline{u}_2 - \overline{u}_1 + \int_0^t \sigma(s)(u_2(s) + u_1(s))(u_2(s) - u_1(s)) \, \mathrm{d}s \\ &\qquad + \int_0^t |b(s)|(u_2(s) - u_1(s)) \, \mathrm{d}s \\ &\leq \overline{u}_2 - \overline{u}_1 + \int_0^t \Big(|b(s)| + 2|\sigma(s)| u_2(s)\Big)(u_2(s) - u_1(s)) \, \mathrm{d}s. \end{aligned}$$

An application of Gronwall's lemma yields the bound. In particular, this implies that $u_1(t) \nearrow u_2(t)$ as $\overline{u}_1 \nearrow \overline{u}_2$. Thanks to Dini's theorem we obtain the complete statement. $\square$

Since we need to look which evolutions of (4.4.1) hit K and ∞ at a given time $T > 0$, we use the previous results in order to construct these solutions explicitly.

Given $T > 0$, denote by $(t_j)_{i=1}^M$ the collection of catastrophes times that fall in the interval $[0, T]$. We will work backwards in order to construct the two functions that we need. First, consider the time t_M and define $u_T^K(t)$ and $u_T^\infty(t)$ for $t \in [t_M, T]$, by (4.4.3) with initial conditions

$$\begin{aligned} u_T^K(t_M) &= \left(\frac{e^{-\int_{t_M}^T b(r) \, \mathrm{d}r}}{K} + \int_{t_M}^T e^{-\int_{t_M}^s b(r) \, \mathrm{d}r} \sigma(s) \, \mathrm{d}s \right)^{-1}, \\ u_T^\infty(t_M) &= \left(\int_{t_M}^T e^{-\int_{t_M}^s b(r) \, \mathrm{d}r} \sigma(s) \, \mathrm{d}s \right)^{-1}. \end{aligned}$$

With this, we have that $u_T^K(T) = K$ and $\lim_{t \to T} u_T^\infty(T) = \infty$ as desired. Now, for $t \in [t_{M-1}, t_M)$ we define $u_T^K(t)$ and $u_T^\infty(s)$ as the solutions of (4.4.2) with terminal conditions $\beta(m_{t_M})^{-1} u_T^K(t_M)$ and $\beta(m_{t_M})^{-1} u_T^\infty(t_M)$ respectively, which in turn can actually be obtained explicitly by (4.4.3). Repeating this procedure in each time interval $[t_j, t_{j+1})$ gives us the two functions $u_T^K(t)$ and $u_T^\infty(t)$ defined for each $t \in [0, T)$.

Remark 4.4.1. *When there are no catastrophic events, for each* $t \in [0, T)$ *we have that* $u_T^K(t)$ *and* $u_T^\infty(t)$ *are simply given by* (4.4.3) *with initial conditions*

$$u_T^K(0) = \left(\frac{e^{-\int_0^T b(r)\,\mathrm{d}r}}{K} + \int_0^T e^{-\int_0^s b(r)\,\mathrm{d}r} \sigma(s)\,\mathrm{d}s \right)^{-1},$$

$$u_T^\infty(0) = \left(\int_0^T e^{-\int_0^s b(r)\,\mathrm{d}r} \sigma(s)\,\mathrm{d}s \right)^{-1}.$$

Lemma 4.4.3. u_T^K *converges uniformly to* u_T^∞ *in every compact interval of* $[0, T)$.

Proof. Given $\varepsilon > 0$ such that $t_M \leq T - \varepsilon$, we will show that u_T^K converges uniformly to u_T^∞ on $[0, T-\varepsilon]$. Indeed, thanks to (4.4.4) and from the definition of $u_T^K(t_M)$ and $u_T^\infty(t_M)$, we have that in $[t_M, T-\varepsilon]$, u_T^K converges uniformly to u_T^∞. Next, on the interval $[t_{M-1}, t_M]$, we have that since in (4.4.3) we can determine the initial condition that gives rise to a given final condition, namely by computing

$$u(t_{M-1}) = \left(\frac{e^{-\int_{t_{M-1}}^{t_M} b(r)\,\mathrm{d}r}}{u(t_M^-)} + \int_{t_{M-1}}^{t_M} e^{-\int_{t_{M-1}}^{s} b(r)\,\mathrm{d}r} \sigma(s)\,\mathrm{d}s \right)^{-1},$$

we can determine $u_T^K(t_{M-1})$ and $u_T^\infty(t_{M-1})$ explicitly as we have $u_T^K(t_M^-) \coloneqq \beta(m_{T_M})^{-1} u_T^K(t_M)$ and $u_T^\infty(t_M^-) \coloneqq \beta(m_{T_M})^{-1} u_T^\infty(t_M)$ respectively. Finally, as the value of both functions at t_M depend continuously on their respective values at t_M^-, thanks to (4.4.4) we obtain convergence on the aforementioned interval. Iterating this procedure yields the result. □

4.4.2 Trees construction

For what follows we set $\Psi(t, u) \coloneqq \sigma(t)u^2 - b(t)u$.

Construction of R_T^K

Recall the particle process defined by the generator (4.2.2) and the catastrophes. In this process, a particle with level u at time t gives birth to a new particle at rate $2\sigma(t)(K-u)$, where the level of the new particle is uniformly distributed in $[u, K]$. On the other hand, when we arrive at a catastrophe time t_j we amplify the level of each particle by a factor $\beta(m_{t_j})$, and we let the process evolve with the new configuration of particles.

In order to obtain the reduced tree associated with the process previously described, we need to filtrate each birth event by the probability of survival until time T. From the definition of the process we have a straightforward way to determine if a particle will remain alive until the final time. Lets suppose that we look at a particle with level u at time t that gives birth to a new particle at that time. We have that if the starting level of its offspring falls in the interval $[u_T^K(t), K]$, this particle will reach the value K before time T, since u_T^K is exactly the curve that is equal to K at time T and the evolution of the levels is monotone in the initial condition. On the other hand, if the starting level of its offspring falls in $[u, u_T^K(t)]$, the particle's level will remain under $u_T^K(t)$ in the posterior times, and thus will be below K at time T.

We state the following result in order to formalize the previous observation.

Lemma 4.4.4. *The survival probability* $p_T^K(t)$ *of a particle born with uniformly distributed level in* $[u, K]$ *at time* $t \in [0, T]$ *is given by*

$$p_T^K(t) = \frac{u_T^K(t) - u \wedge u_T^K(t)}{K - u}.$$

Proof. The proof is given by the last paragraph, since we have that

$$p_T^K = \mathbb{P}(u(T) < u_T^K(T)) = \mathbb{P}(u < u_T^K(t)) = \frac{u_T^K(t) - u \wedge u_T^K(t)}{K - u},$$

where $u(s)$ for $s \in [t, T]$, is the evolution of a level starting from u. □

Using this, we have that filtrating the birth rate $2\sigma(t)(K - u)$ according to the survival probability gives us the new rate

$$2\sigma(t)\Big(u_T^K(t) - u \wedge u_T^K(t)\Big).$$

Since we know that a uniform random variable conditioned to land in a smaller interval is uniformly distributed in the smaller interval, we propose the following construction for the reduced tree $\mathbf{R}_T^K$.

Assume that we start with levels $U_0^K = \Big(U_0^{1,K}, \ldots, U_0^{N^K,K}\Big)$, each one smaller than $u_T^K(0)$. We rearrange the entries of this vector in an increasing way and we denote it the same. We then assign a label $i \in \mathcal{U}$ to each entry, where i matches its index in the vector, that is $i = 1, \ldots, N^K$.

For each label $i = 1, \ldots, N^K$, we define $\ell^K(i, \cdot)$ as the solution of

$$\ell^K(i, t) = U_0^{i,K} + \int_0^t \Psi\Big(s, \ell^K(i, s)\Big)\, \mathrm{d}s + \sum_{j \geq 1,\, t \geq t_j} (\beta(m_{t_j}) - 1)\ell^K(i, t^-), \quad \forall t \in [0, T].$$

Each one of these particles will produce a new particle according to a Poisson process of parameter $2\sigma(t)\Big(u_T^K(t) - \ell^K(i, t)\Big)$. At each jump event of this process, we sample a uniform random variable over the interval $[\ell^K(i, t), u_T^K(t)]$, and we set this value as the starting level of the new offspring. For example, lets suppose that the j-th birth of the particle with label i happens at time t^*. The new particle will have label ij and its level $\ell^K(ij, \cdot)$ will be defined for all $t \geq t^*$ as the solution of the problem

$$\ell^K(ij, t) = \ell^K(i, t^*) + \theta_{t^*}^{i,K} + \int_{t^*}^t \Psi\Big(s, \ell^K(ij, s)\Big)\, \mathrm{d}s + \sum_{j \geq 1,\, t \geq t_j \geq t^*} (\beta(m_{t_j}) - 1)\ell^K(ij, t^-),$$

for each $t \in [t^*, T]$, where $\theta_{t^*}^{i,K}$ is a uniform random variable over $[\ell^K(i, t^*), u_T^K(t^*)]$. This particle will have appended a Poisson process of parameter $2\sigma(t)\Big(u_T^K(t) - \ell^K(ij, t)\Big)$ for producing new particles. Iterating this procedure gives us the process representing the reduced tree $\mathbf{R}_T^K$.

Construction of $\mathbf{R}_T$

Recalling now the particle process represented by the generator (4.2.5) and the catastrophes, we had that a particle with level u at time t produces a new particle with level in $[u + \ell_1, u + \ell_2]$ at rate $2\sigma(t)(\ell_2 - \ell_1)$. By following the exact same procedure as in Lemma 4.4.4, we have that in this model filtrating the birth events according to the survival probability yields the rate

$$2\sigma(t)\Big(u_T^\infty(t) - u \wedge u_T^\infty(t)\Big).$$

Given this, we construct the tree $\mathbf{R}_T$ associated to the process $(U_t)_{t\geq 0}$ determined by the generator (4.2.5) and the catastrophes as follows.

Starting with levels $U_0 = (U_0^1, \dots, U_0^N)$, each one smaller than $u_T^\infty(0)$, we rearrange the vector as before and we assign the corresponding label to each particle. For each $i = 1, \dots, N$, we denote by $\ell(i, \cdot)$, the solution of the following equation

$$\ell(i,t) = U_0^i + \int_0^t \Psi(\ell(i,s))\,\mathrm{d}s + \sum_{j\geq 1,\, t\geq t_j} (\beta(m_{t_j}) - 1)\ell(i,t^-), \quad \forall t \in [0,T].$$

Each one of these particles will give birth to a new particle according to a Poisson process of parameter $2\sigma(t)\Big(u_T^\infty(t) - \ell(i,t)\Big)$. When this process jumps, we sample a uniform random variable over the interval $[\ell(i,t), u_T^\infty(t)]$ and we set this value as the starting level of the new particle. This last particle will follow the same evolution than its progenitor, that is the one given by Ψ and the catastrophic jumps. Iterating this procedure as before yields the process behind the tree $\mathbf{R}_T$.

Both trees defined in this way can be seen as elements of the same space. Indeed, let $g\colon [0,T] \to \mathbb{R}$ be a function satisfying the integrability condition (4.2.8). In particular, one could consider the explicit choice $g(s) := \exp(-\int_0^s 2\sigma(r)u_T^\infty(r)\,\mathrm{d}r)h(s)$, with h any function in $L^1([0,T])$. We then have the following result.

Lemma 4.4.5. *Given a function g satisfying the integrability condition (4.2.8), we have that $\mathbf{R}_T^K$ and $\mathbf{R}_T$ are elements of $\mathcal{T}^g$ almost surely. Furthermore, the following bound holds*

$$\mathbb{E}\Big(N_{\mathbf{R}_T}(t)\Big) \leq \mathbb{E}\Big(N_{\mathbf{R}_T}(0)\Big)e^{\int_0^t 2\sigma(r)u_T^\infty(r)\,\mathrm{d}r}, \tag{4.4.5}$$

which is also valid for $\mathbf{R}_T^K$.

Proof. Since each particle with level u at time t produces a new one at rate $2\sigma(t)(u_T^\infty - u)$ in $\mathbf{R}_T$ and $2\sigma(t)(u_T^K - u)$ in $\mathbf{R}_T^K$, we can bound the rate of birth of each particle in each tree by $2\sigma(t)u_T^\infty(t)$, since $u_T^K(t) \leq u_T^\infty(t)$. This allows to bound $N_{\mathbf{R}_T^K}(t) := |V_t(\mathbf{R}_T^K)|$ and $N_{\mathbf{R}_T}(t) := |V_t(\mathbf{R}_T)|$ by a birth-only process with the previous rate, yielding in this way that

$$\mathbb{E}\left(\int_0^T N_{\mathbf{R}_T}(s)g(s)\,\mathrm{d}s\right) \leq \int_0^T \mathbb{E}\Big(N_{\mathbf{R}_T}(0)\Big)e^{\int_0^s 2\sigma(r)u_T^\infty(r)\,\mathrm{d}r}g(s)\,\mathrm{d}s < \infty,$$

and similarly for $\mathbf{R}_T^K$, thus obtaining the conclusion of the statement. □

We now have everything in order to prove the main result.

4.4.3 Coupling: proof of Theorem 4.1

Assume that the Feller diffusion with given catastrophes (4.1.1) that we approximate starts from $X_0 = x$.

Consider a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Over this space we consider a Poisson point measure $\mathcal{N}(\mathrm{d}s, \mathrm{d}v, \mathrm{d}\rho, \mathrm{d}\theta, \mathrm{d}\tilde{\theta})$ on $[0,\infty) \times \mathcal{U} \times [0,\infty) \times [0,1] \times [0,1]$ with intensity $\mathrm{d}s \otimes n(\mathrm{d}v) \otimes \mathrm{d}\rho \otimes \mathrm{d}\theta \otimes \mathrm{d}\tilde{\theta}$, where n is the counting measure on $\mathcal{U}$. Given $x \geq 0$, consider also an independent Poisson point measure $N(x, \mathrm{d}s)$ on $[0,\infty)$ with intensity $x\mathrm{d}s$. We consider the natural filtration $(\mathcal{F}_t)_{t\geq 0}$ associated with these random elements. In what follows we give an algorithmic construction of the coupling using these objects.

We start by considering all the atoms of $N(x, \mathrm{d}s)$ that fall under $u_T^\infty(0)$. We enumerate them as $\{U^1, \dots, U^{N_x^\infty}\}$ where $N_x^\infty := N(x, [0, u_T^\infty(0)])$, and we assign them a label in $\mathcal{U}$ according to their indices. We consider also all the atoms that fall below $u_T^K(0)$, which we denote by $\{U^1, \dots, U^{N_x^K}\}$ where as before $N_x^K := N(x, [0, u_T^K(0)])$. Moreover, we consider also a root $\{\varnothing\}$. Given the previous, we immediately notice that $\{U^1, \dots, U^{N_x^K}\} \subset \{U^1, \dots, U^{N_x^\infty}\}$.

We denote by $\mathbf{R}_T$ the tree constructed starting with all the atoms under $u_T^\infty(0)$ and by $\mathbf{R}_T^K$ the tree constructed starting with all the atoms under $u_T^K(0)$. These trees are then constructed simultaneously following the next algorithm:

1. Set $k = 0$ and $T_0 = 0$. Define $\ell^K(i,t)$ and $\ell(i,t)$ for $i = 1, \dots, N_x^K$, as the solution of the problem
$$u_i(t) = U_0^i + \int_0^t \Psi\Big(s, u_i(s)\Big)\,\mathrm{d}s + \sum_{j\geq 1,\, t\geq t_j} (\beta(m_{t_j}) - 1) u_i(t^-), \quad \forall t \in [0,T],$$
and $\ell(i,t)$ for $i = N_x^K + 1, \dots, N_x^\infty$, as the solution of the same problem for the corresponding U_0^i. Finally, initialize $V_0^K := \{1, \dots, N_x^K\}$ and $V_0 := \{1, \dots, N_x^\infty\}$.

2. At the first time $t > T_k$ with $(t, v, \rho, \theta, \tilde{\theta})$ an atom of $\mathcal{N}$ such that:

   2.1. $v \in V_{t-}^K$ and $\rho \leq 2\sigma(t^-)\Big(u_T^K(t^-) - \ell^K(v, t^-)\Big)$, we add a new particle with its corresponding chronological label $\tilde{v}$, and we define its level curve as the solution of
   $$\ell^K(\tilde{v}, t') = \ell^K(v, t^-) + \theta(u_T^\infty(t^-) - \ell^K(v, t^-)) + \int_t^{t'} \Psi(\ell^K(\tilde{v}, s))\,\mathrm{d}s + \sum_{j\geq 1,\, t'\geq t_j\geq t} (\beta(m_{t_j}) - 1)\ell^K(\tilde{v}, t'^-), \ \forall t' \in [t,T],$$
   if $\theta \leq \frac{u_T^K(t^-) - \ell^K(v,t^-)}{u_T^\infty(t^-) - \ell^K(v,t^-)}$. Otherwise, in the previous equation set the initial condition to $\ell^K(\tilde{v}, t) := \ell^K(v, t^-) + \tilde{\theta}(u_T^K(t^-) - \ell^K(v, t^-))$. Then, set
   $$V_t^K := V_{t-}^K \cup \{\tilde{v}\}.$$

   2.2. $v \in V_{t-}$ and $\rho \leq 2\sigma(t^-)\Big(u_T^\infty(t^-) - \ell(v, t^-)\Big)$, we add a new particle with its corresponding chronological label $\tilde{v}$ and define its level curve by solving
   $$\ell(\tilde{v}, t') = \ell(v, t-) + \theta(u_T^\infty(t^-) - \ell(v, t^-)) + \int_t^{t'} \Psi(\ell(\tilde{v}, s))\,\mathrm{d}s + \sum_{j\geq 1,\, t'\geq t_j\geq t} (\beta(m_{t_j}) - 1)\ell(\tilde{v}, t'^-), \quad \forall s \in [t,T].$$

Then, set

$$V_t := V_{t-} \cup \{\tilde{v}\}.$$

Finally, set $T_{k+1} := t$.

3. Increase k by one and go back to the previous step.

After iterating this algorithmic procedure until time T, we can define the filtered measures

$$\mathcal{N}_v^K(\mathrm{d}s) = \int_{\mathbb{R}_+\times[0,1]^2} \mathbf{1}_{v\in V^K_{s^-}} \mathbf{1}_{\rho\leq 2\sigma(s^-)\left(u_T^K(s^-)-\ell^K(v,s^-)\right)} \mathcal{N}(\mathrm{d}s, \{v\}, \mathrm{d}\rho, \mathrm{d}\theta, \mathrm{d}\tilde{\theta}),$$
$$\mathcal{N}_v(\mathrm{d}s) = \int_{\mathbb{R}_+\times[0,1]^2} \mathbf{1}_{v\in V_{s^-}} \mathbf{1}_{\rho\leq 2\sigma(s^-)\left(u_T^\infty(s^-)-\ell(v,s^-)\right)} \mathcal{N}(\mathrm{d}s, \{v\}, \mathrm{d}\rho, \mathrm{d}\theta, \mathrm{d}\tilde{\theta}),$$

and we can construct explicitly the chronological trees $\mathbf{R}_T^K$ and $\mathbf{R}_T$ following the procedure explained at the end of Section 4.3 with the families $(\mathcal{N}_v^K(\mathrm{d}s))_{v\in\mathcal{U}}$ and $(\mathcal{N}_v(\mathrm{d}s))_{v\in\mathcal{U}}$ respectively. Also, following the notation from section 4.3, have that $V_t^K = V_t(\mathbf{R}_T^K)$ and $V_t = V_t(\mathbf{R}_T)$.

We now pass to the proof of the main result.

Proof of Theorem 4.1. We start by noticing that the subtree issued from the particles with labels $\{N_x^K+1,\dots,N_x^\infty\}$ will contribute to the distance from the beginning. Following this, we denote by $\overline{\mathbf{R}}_T^K \subset \mathbf{R}_T$ the subtree spanned by the initial particles with labels $\{1,\dots,N_x^K\}$.

Consider $T_K \in [0,T]$ the first time that the coupling fails, meaning that the trees $\mathbf{R}_T^K$ and $\overline{\mathbf{R}}_T^K$ are equal up to time T_K^-. Given this and the observation of the previous paragraph, we can write the distance between $\mathbf{R}_T^K$ and $\mathbf{R}_T$ as

$$\begin{aligned}
d_g^T\left(\mathbf{R}_T^K, \mathbf{R}_T\right) &= \sum_{u\in\mathcal{U}} \int_{T^K}^T \left| \mathbf{1}_{(u,s)\in\mathbf{R}_T^K} - \mathbf{1}_{(u,s)\in\overline{\mathbf{R}}_T^K} \right| g(s)\,\mathrm{d}s + \int_0^T N_{\mathbf{R}_T\setminus\overline{\mathbf{R}}_T^K}(s) g(s)\,\mathrm{d}s. \\
&\leq \int_{T^K}^T N_{\mathbf{R}_T^K}(s)g(s)\,\mathrm{d}s + \int_{T^K}^T N_{\overline{\mathbf{R}}_T^K}(s)g(s)\,\mathrm{d}s \\
&\quad + \int_0^T N_{\mathbf{R}_T\setminus\overline{\mathbf{R}}_T^K}(s)g(s)\,\mathrm{d}s, \qquad (4.4.6)
\end{aligned}$$

recalling that $N_{\mathbf{T}}(t) := |V_t(\mathbf{T})|$, for a given chronological tree $\mathbf{T}$. Lets treat first the last term of the right hand side. Using the same argument as in Lemma 4.4.5 for obtaining the bound (4.4.5), we can control the number of particles in $\mathbf{R}_T \setminus \overline{\mathbf{R}}_T^K$ by a faster birth-only process starting from $N_x^\infty - N_x^K$ particles. Thus, once we take expectation, the last term is controlled by

$$\begin{aligned}
\mathbb{E}\left(\int_0^T N_{\mathbf{R}_T\setminus\overline{\mathbf{R}}_T^K}(s)g(s)\,\mathrm{d}s\right) &\leq \int_0^T \mathbb{E}\left(N_{\mathbf{R}_T\setminus\overline{\mathbf{R}}_T^K}(0)\right) e^{\int_0^s 2\sigma(r)u_T^\infty(r)\,\mathrm{d}r} g(s)\,\mathrm{d}s \\
&= \int_0^T \mathbb{E}\left(N_x^\infty - N_x^K\right) e^{\int_0^s 2\sigma(r)u_T^\infty(r)\,\mathrm{d}r} g(s)\,\mathrm{d}s \\
&= \mathbb{E}\left(N(x, [u_T^K(0), u_T^\infty(0)])\right) \int_0^T e^{\int_0^s 2\sigma(r)u_T^\infty(r)\,\mathrm{d}r} g(s)\,\mathrm{d}s \\
&\leq xC(u_T^\infty(0) - u_T^K(0)),
\end{aligned}$$

for some constant C depending on the integrability condition (4.2.8).

Next, we treat the two remaining terms in (4.4.6). Since $\mathbf{R}_T^K$ and $\mathbf{R}_T$ are elements of $\mathcal{T}^g$, it suffices to show that T^K converges to T in probability. Indeed, for any $0 \leq \eta \leq T$ we have

$$
\begin{aligned}
&\mathbb{E}\left(\int_{T^K}^T N_{\mathbf{R}_T^K}(s) g(s)\,\mathrm{d}s\right) \\
&\qquad = \mathbb{E}\left(\mathbf{1}_{|T^K - T| > \eta} \int_{T^K}^T N_{\mathbf{R}_T^K}(s) g(s)\,\mathrm{d}s + \mathbf{1}_{|T^K - T| \leq \eta} \int_{T^K}^T N_{\mathbf{R}_T^K}(s) g(s)\,\mathrm{d}s\right) \\
&\qquad \leq \mathbb{E}\left(\mathbf{1}_{|T^K - T| > \eta} \int_{T^K}^T N_{\mathbf{R}_T^K}(s) g(s)\,\mathrm{d}s\right) + \mathbb{E}\left(\int_{T-\eta}^T N_{\mathbf{R}_T^K}(s) g(s)\,\mathrm{d}s\right) \\
&\qquad \leq \mathbb{E}\left(\mathbf{1}_{|T^K - T| > \eta} \int_0^T N_{\mathbf{R}_T^K}(s) g(s)\,\mathrm{d}s\right) \\
&\qquad\qquad + x u_T^\infty(0) \int_{T-\eta}^T e^{\int_0^s 2\sigma(r) u_T^\infty(r)\,\mathrm{d}r} g(s)\,\mathrm{d}s,
\end{aligned}
$$

where we used again the argument behind the bound (4.4.5) for controlling $N_{\mathbf{R}_T^K}(s)$ by a faster process starting from $N(x, [0, u_T^K(0)])$ individuals.

In order to show the convergence in probability of T^K towards T, we notice that we can control the probability of the event $\{T - T^K > \eta\}$ by the probability of the union of two particular events. The first one, the event in where there exists a label present in both trees and having the same level curve, that produces a new particle that is in one tree but not in the other, meaning that there is a discrepancy in the birth time of this new particle. The second one, the event in where we have a label in both trees with the same levels and such that it produces a new particle but the starting levels in each tree do not match. In particular, we notice that this last event is not optimal, in the sense that the trees can continue to be coupled, but with high probability there will be a discrepancy latter created by this difference in the levels.

In order to control this probability we introduce

$$
\begin{aligned}
\gamma_T^K(v,s) &:= 2\sigma(s)(u_T^K(s) - \ell^K(v,s)), \\
\gamma_T^\infty(v,s) &:= 2\sigma(s)(u_T^\infty(s) - \ell(v,s)),
\end{aligned}
$$

to ease notation, and also

$$
\begin{aligned}
\delta_1^K(v,s) &= (\gamma_T^K(v,s) - \gamma_T^\infty(v,s))\mathbf{1}_{\gamma_T^\infty(v,s) \leq \gamma_T^K(v,s)}, \\
\delta_2^K(v,s) &= (\gamma_T^\infty(v,s) - \gamma_T^K(v,s))\mathbf{1}_{\gamma_T^K(v,s) \leq \gamma_T^\infty(v,s)},
\end{aligned}
$$

the random corrections that will help to determine the event in where a particle is born in one tree but not in the other.

The reasoning of the previous paragraph yields the following bound

$$\begin{aligned}
&\mathbb{P}(T^K < T-\eta) \\
&\leq \mathbb{P}\Bigg(\int_0^{T-\eta}\int_{\mathcal{U}\times\mathbb{R}_+\times[0,1]^2} \mathbf{1}_{v\in V^K_{s-}\cap V_{s-}}\mathbf{1}_{\ell^K(v,s-)=\ell(v,s-)} \\
&\qquad \times\Bigg[\Big(\mathbf{1}_{\gamma^K_T(v,s-)\wedge\gamma^\infty_T(v,s-)\leq\rho<\gamma^K_T(v,s-)\wedge\gamma^\infty_T(v,s-)+\delta^K_1(v,s-)} \\
&\qquad\qquad + \mathbf{1}_{\gamma^K_T(v,s-)\wedge\gamma^\infty_T(v,s-)\leq\rho<\gamma^K_T(v,s-)\wedge\gamma^\infty_T(v,s-)+\delta^K_2(v,s-)}\Big) \\
&\qquad\qquad + \mathbf{1}_{\rho\leq\gamma^K_T(v,s-)\wedge\gamma^\infty_T(v,s-)}\mathbf{1}_{\gamma^K_T(v,s-)/\gamma^\infty_T(v,s-)<\theta}\Bigg]\mathcal{N}(\mathrm{d}v,\mathrm{d}s,\mathrm{d}\rho,\mathrm{d}\theta,\mathrm{d}\tilde{\theta}) > 1\Bigg) \\
&\leq \mathbb{E}\Bigg(\int_0^{T-\eta}\sum_{v\in V^K_s\cap V_s}\mathbf{1}_{\ell^K(v,s)=\ell(v,s)}\Big(\delta^K_1(v,s)+\delta^K_2(v,s)+\gamma^K_T(v,s)\wedge\gamma^\infty_T(v,s)\Big(1-\frac{\gamma^K_T(v,s)}{\gamma^\infty_T(v,s)}\Big)\Big)\mathrm{d}s\Bigg) \\
&\leq \mathbb{E}\Bigg(\int_0^{T-\eta}\sum_{v\in V^K_s\cap V_s}\mathbf{1}_{\ell^K(v,s)=\ell(v,s)}\Big(\delta^K_1(v,s)+\delta^K_2(v,s)+2\sigma(s)|u^\infty_T(s)-u^K_T(s)|\Big)\,\mathrm{d}s\Bigg),
\end{aligned}$$

where we used Markov's inequality and the master formula for point processes ([107, Proposition 12.1.10]). Moreover, in the event in where the levels of the particles in both trees are equal we have

$$\begin{aligned}
\delta^K_1(v,s)+\delta^K_2(v,s) &\leq \gamma^K_T(v,s)\vee\gamma^\infty_T(v,s)-\gamma^K_T(v,s)\wedge\gamma^\infty_T(v,s) \\
&\leq |\gamma^K_T(v,s)-\gamma^\infty_T(v,s)| \\
&\leq 2\sigma(s)|u^\infty_T(s)-u^K_T(s)|,
\end{aligned}$$

where we used that $a\wedge b - a\wedge c \leq |b-c|$. This yields the bound

$$\begin{aligned}
\mathbb{P}(T^K < T-\eta) &\leq (2\|\sigma\|+1)\int_0^{T-\eta}\mathbb{E}\Big(|V^K_s\cap V_s|\Big)|u^\infty_T(s)-u^K_T(s)|\,\mathrm{d}s \\
&\leq (2\|\sigma\|+1)\mathbb{E}\Big(|V^K_{T-\eta}\cap V_{T-\eta}|\Big)\int_0^{T-\eta}|u^\infty_T(s)-u^K_T(s)|\,\mathrm{d}s. \qquad (4.4.7)
\end{aligned}$$

Furthermore, we can control the integral that appears in the previous inequality. Indeed, Thanks to Lemma 4.4.2 we have

$$\begin{aligned}
|u^\infty_T(s)-u^K_T(s)| &\leq \Big(u^\infty_T(t_{j(s)})-u^K_T(t_{j(s)})\Big)e^{\int_{t_{j(s)}}^s\big(|b(r)|+2\sigma(r)u^\infty_T(r)\big)\,\mathrm{d}r} \\
&\leq \beta(m_{t_{j(s)}})\Big(u^\infty_T\Big(t^-_{j(s)}\Big)-u^K_T\Big(t^-_{j(s)}\Big)\Big)e^{\int_{t_{j(s)}}^s\big(|b(r)|+2\sigma(r)u^\infty_T(r)\big)\,\mathrm{d}r}
\end{aligned}$$

where $t_{j(s)}$ is the last jump time before s. Iterating this bound yields

$$|u^\infty_T(s)-u^K_T(s)| \leq \prod_{j\geq 1, s\geq t_j}\beta(m_{t_j})\Big(u^\infty_T(0)-u^K_T(0)\Big)e^{\int_0^s\big(|b(r)|+2\sigma(r)u^\infty_T(r)\big)\,\mathrm{d}r}.$$

Coming back to (4.4.7) and using the previous inequality gives us

$$\begin{aligned}\mathbb{P}(T^K < T-\eta) &\leq C'_T\Big(u_T^\infty(0) - u_T^K(0)\Big)\mathbb{E}\Big(|V_{T-\eta}^K \cap V_{T-\eta}|\Big)e^{\int_0^{T-\eta}\Big(|b(r)|+2\sigma(r)u_T^\infty(r)\Big)\,\mathrm{d}r} \\ &\leq C'_T x u_T^\infty(0)\Big(u_T^\infty(0) - u_T^K(0)\Big)e^{2\int_0^{T-\eta}\Big(|b(r)|+2\sigma(r)u_T^\infty(r)\Big)\,\mathrm{d}r},\end{aligned}$$

where $C'_T = T(2\|\sigma\|+1)\prod_{j\geq 1, t_j\leq T}\beta(m_{t_j})$. In order to see that this last quantity converges to 0, we can restate this convergence problem as follows. Let $f\colon \mathbb{N}\to\mathbb{R}$ be a decreasing function converging to 0 and $h\colon [0,T)\to\mathbb{R}_+$ a continuous increasing function such that $\lim_{t\to T} h(t) = \infty$. In order to justify the existence of a function $\eta\colon\mathbb{N}\to\mathbb{R}$ such that $\eta(K)\to 0$ as $K\to\infty$ and

$$f(K)h(T-\eta(K)) \to 0, \quad K\to\infty,$$

we consider the generalized inverse $h^-(y) \coloneqq \inf\{x : h(x) > y\}$ and we set

$$\eta(K) = T - h^-(\tilde{\eta}(K)),$$

for another function $\tilde{\eta}:\mathbb{N}\to\mathbb{R}$ satisfying $\tilde{\eta}(K)\to\infty$ as $K\to\infty$. Since h is continuous

$$f(K)h(T-\eta(K)) = f(K)h(h^-(\tilde{\eta}(K))) = f(K)\tilde{\eta}(K),$$

thus it suffices choosing $\tilde{\eta}$ growing slow enough in order to $f(K)\tilde{\eta}(K)\to 0$ as $K\to\infty$, for example $\tilde{\eta}(K) = |\log(f(K))|$.

Following this argument we show the existence of a sequence $\eta(K)$ such that

$$\mathbb{P}(|T^K - T| > \eta(K)) \to 0, \quad \text{as } K\to\infty. \tag{4.4.8}$$

Considering this sequence, once we take expectation, (4.4.6) can be written as

$$\begin{aligned}\mathbb{E}(d_g^T\Big(\mathbf{R}_T^K,\mathbf{R}_T\Big)) \leq{}& xC(u_T^\infty(0) - u_T^K(0)) + xu_T^\infty(0)\int_{T-\eta(K)}^T e^{\int_0^s 2\sigma(r)u_T^\infty(r)\,\mathrm{d}r} g(s)\,\mathrm{d}s \\ &+ \mathbb{E}\Bigg(\mathbf{1}_{|T^K-T|>\eta(K)}\Bigg(\int_0^T N_{\mathbf{R}_T^K}(s)g(s)\,\mathrm{d}s + \int_0^T N_{\overline{\mathbf{R}}_T^K}(s)g(s)\,\mathrm{d}s\Bigg)\Bigg).\end{aligned}$$

Denoting by $I^K \coloneqq \int_0^T N_{\mathbf{R}_T^K}(s)g(s)\,\mathrm{d}s + \int_0^T N_{\overline{\mathbf{R}}_T^K}(s)g(s)\,\mathrm{d}s$, we have that the sequence $(I^K)_{K\in\mathbb{N}^*}$ is uniformly integrable, since it can be bounded uniformly in $L^1(\mathbb{P})$ using the argument of controlling by a faster birth-only process from Lemma 4.4.5 and the integrability condition (4.2.8). Moreover, thanks to this, for each $\varepsilon > 0$ there exists $\delta > 0$ such that for every measurable set A with $\mathbb{P}(A) < \delta$, we have $\sup_{K\in\mathbb{N}}\mathbb{E}(\mathbf{1}_A I^K) < \varepsilon$. Following (4.4.8), there exists K^* such that for all $K\geq K^*$, $\mathbb{P}(|T-T^K| > \eta(K)) < \delta$. Furthermore, there also exists K' such that $\int_{T-\eta(K)}^T \exp(\int_0^s 2\sigma(r)u_T^\infty(r)\,\mathrm{d}r)g(s)\,\mathrm{d}s \leq \varepsilon/(xu_T^\infty(0))$ for every $K\geq K'$ thanks to (4.2.8), and there exists another $\tilde{K}$ such that $(u_T^\infty(0) - u_T^K(0)) \leq \varepsilon/(xC)$ for each $K\geq\tilde{K}$ thanks to Lemma 4.4.3. Gathering all, we conclude that for every $K\geq\max\{K^*, K', \tilde{K}\}$, we have

$$\mathbb{E}(d_g^T\Big(\mathbf{R}_T^K,\mathbf{R}_T\Big)) \leq 3\varepsilon,$$

which terminates the proof. □

C Appendix

C.1 Markov mapping theorem

What follows is taken directly from [80] and it is presented for the sake of completeness.

Let (S, d) and (S_0, d_0) be complete, separable metric spaces, $B(S) \subset M(S)$ be the Banach space of bounded measurable functions on S, with $\|f\| = \sup_{x\in S}|f(x)|$ and $C_b(S) \subset B(S)$ be the subspace of bounded continuous functions. An operator $A \subset B(S)\times B(S)$ (following the general notation for multivalued operators from [53], which coincides with the graph of a single-valued operator) is called *dissipative* if $\|f_1 - f_2 - \varepsilon(g_1 - g_2)\| \geq \|f_1 - f_2\|$ for all $(f_1, g_1), (f_2, g_2) \in A$ and $\varepsilon > 0$. A is called a *pre-generator* if A is dissipative and there are sequences of functions $\mu_n\colon S \to \mathcal{P}(S)$ and $\lambda_n\colon S \to [0, \infty)$ such that for each $(f, g) \in A$

$$g(x) = \lim_{n\to\infty} \lambda_n(x) \int_S \Big(f(y) - f(x)\Big)\mu_n(x, dy), \tag{C.1}$$

for each $x \in S$. A is *graph-separable* if there exists a countable subset $(g_k)_{k\in\mathbb{N}} \subset D(A) \cap C_b(S)$ such that the graph of A is contained in the bounded, pointwise closure of the linear span of $(g_k, Ag_k)_{k\in\mathbb{N}}$. More precisely, we should say that there exists $(g_k, h_k)_{k\in\mathbb{N}} \subset A \cap C_b(S) \times B(S)$ such that A is contained in the bounded pointwise closure of $(g_k, h_k)_{k\in\mathbb{N}}$, but typically A is single-valued, so we use the more intuitive notation Ag_k. These two conditions are satisfied by essentially all operators A that might reasonably be thought to be generators of Markov processes. Note that A is graph-separable if $A \subset L \times L$, where $L \subset B(S)$ is separable in the sup norm topology, for example, if S is locally compact, and L is the space of continuous functions vanishing at infinity.

A collection of functions $D \subset C_b(S)$ is *separating* if $\nu, \mu \in \mathcal{P}(S)$ and $\int_S f\, d\nu = \int_S f\, d\mu$ for all $f \in D$ imply $\mu = \nu$.

For an S_0-valued, measurable process Y, $\widehat{\mathcal{F}}_t^Y$ will denote the completion of the σ-algebra: $\sigma(Y(0), \int_0^r h(Y(s))\, ds, r \leq t, h \in B(S_0))$. Notice that for almost every t, $Y(t)$ will be $\widehat{\mathcal{F}}_t^Y$-measurable, but in general, $\widehat{\mathcal{F}}_t^Y$ does not contain $\mathcal{F}_t^Y = \sigma(Y(s) : s \leq t)$. Let $\mathbf{T}^Y = \{t : Y(t) \text{ is } \widehat{\mathcal{F}}_t^Y \text{ measurable}\}$. If Y is càdlàg and has no fixed points of discontinuity (i.e., for every t, $Y(t) = Y(t^-)$ a.s.), then $\mathbf{T}^Y = [0, \infty)$. $D(S, [0, \infty))$ denotes the space of càdlàg, S-valued functions with the Skorohod topology, and $M(S, [0, \infty))$ denotes the space of Borel measurable functions, $x\colon [0, \infty) \to S$, topologized by convergence in Lebesgue measure.

Theorem 4.2. *Let (S, d) and (S_0, d_0) be complete, separable metric spaces. Let $A \subset C_b(S) \times C(S)$ and $\psi \in C(S)$, $\psi \geq 1$. Suppose that for each $f \in D(A)$ there exists $c_f > 0$ such that*

$$|Af(x)| \leq c_f\psi(x), \quad x \in S,$$

and define $A_0 f(x) = Af(x)/\psi(x)$. Suppose that A_0 is a graph-separable pre-generator, that $D(A) = D(A_0)$ is closed under multiplication and is separating. Let $\gamma\colon S \to S_0$ be Borel measurable, and let α be a transition function from S_0 into S (i.e. $y \in S_0 \to \alpha(y, \cdot) \in \mathcal{P}(S)$ is Borel measurable) satisfying $\int h \circ \gamma(z)\alpha(y, \mathrm{d}z) = h(y)$ for $y \in S_0$ and $h \in B(S_0)$, that is, $\alpha(y, \gamma^{-1}(y)) = 1$. Assume that $\widetilde{\psi}(y) := \int_S \psi(z)\alpha(y, \mathrm{d}z) < \infty$ for each $y \in S_0$, and define

$$C = \left\{ \left(\int_S f(z)\alpha(\cdot, \mathrm{d}z), \int_S Af(z)\alpha(\cdot, \mathrm{d}z) \right) : f \in D(A) \right\}.$$

Let $\mu_0 \in \mathcal{P}(S_0)$, and define $\nu_0 = \int \alpha(y,\cdot)\mu_0(\mathrm{d}y)$.

1. *If $\widetilde{Y}$ satisfies $\int_0^t \mathbb{E}(\widetilde{\psi}(\widetilde{Y}))\mathrm{d}s < \infty$ for all $t \geq 0$, and $\widetilde{Y}$ is a solution of the martingale problem for (C, μ_0), then there exists a solution X of the martingale problem for (A, ν_0) such that $\widetilde{Y}$ has the same distribution on $M([0,\infty), S_0)$ as $Y = \gamma \circ X$. If Y and $\widetilde{Y}$ are càdlàg, then Y and $\widetilde{Y}$ have the same distribution on $D([0,\infty), S_0)$.*

2. *For $t \in \mathbf{T}^Y$,*
$$\mathbb{P}(X(t) \in \Gamma | \widehat{\mathcal{F}}_t^Y) = \alpha(Y(t), \Gamma), \quad \Gamma \in \mathcal{B}(S).$$

3. *If, in addition, uniqueness holds for the martingale problem for (A, ν_0), then uniqueness holds for the $M([0,\infty), S_0)$-martingale problem for (C, μ_0). If $\widetilde{Y}$ has sample paths in $D([0,\infty), S_0)$, then uniqueness holds for the $D([0,\infty), S_0)$-martingale problem for (C, μ_0).*

4. *If uniqueness holds for the martingale problem for (A, ν_0), then Y restricted to $\mathbf{T}^Y$ is a Markov process.*

C.2 Inhomogeneous lookdown construction

In this section, we give an argument to apply the lookdown construction introduced in [80] to birth and death processes with time dependent rates, and also to the Feller diffusion with time dependent coefficients.

The lookdown construction relies on the Markov mapping theorem to ensure its existence as a solution of the martingale problem associated with the construction. Furthermore, this is obtained through the existence of solutions for the martingale problem associated with the process that is being represented. This result is stated for martingale problems that are not time dependent, thus giving rise to time homogeneous processes. In order to be able to apply it in an inhomogeneous setting, we can consider the space-time process for obtaining a time homogeneous object, which is usual procedure for translating results from the time homogeneous setting. This will be the approach of what follows.

Let $K > 0$. Consider the state space $E = \bigcup_{n=0}^{\infty} [0,K]^n$, the domain $D(A^K)$ defined in (4.2.1) and recall the operator $A^K\colon D(A^K) \subset B(E) \to B(E \times [0,\infty))$ defined for $u = (u_1, \ldots, u_n)$ by

$$A_t^K f(u) = f(u) \sum_{i=1}^{n} 2\sigma(t) \int_{u_i}^{K} (g(v)-1)\mathrm{d}v + f(u) \sum_{i=1}^{n} \Big(\sigma(t)u_i^2 - b(t)u_i\Big) \frac{g'(u_i)}{g(u_i)},$$

where b and σ are bounded continuous functions such that $\sigma(t) \geq 0$ for each $t \geq 0$. A process $(U_t)_{t\geq 0}$ defined over a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a solution of the martingale problem for A^K if for each $f \in D(A^K)$

$$f(U_t) - \int_0^t A_s^K f(U_s)\,\mathrm{d}s,$$

is an $\mathcal{F}_t^U$-martingale. This notion of solution differs from the classical one only because the domain and the range of A^K are contained in different spaces.

Now, consider the domain

$$D\Big(\overline{A}^K\Big) = \Big\{ f\zeta : f \in D(A^K), \zeta \in C_c^1([0,\infty)) \Big\},$$

and the operator $\overline{A}^K : D(\overline{A}^K) \subset B(E \times [0,\infty)) \to B(E \times [0,\infty))$ defined by

$$\overline{A}^K f\zeta(u,t) = A_t^K f(u)\zeta(t) + f(u)\zeta'(t).$$

This last formulation corresponds to the martingale problem for the space-time process $\overline{U}_t := (U_t, t)$. Furthermore, Theorem 4.7.1 in [53] ensures that solving the martingale problem for $\overline{A}^K$ (in the classical sense) is equivalent to solving the martingale problem for A^K in the sense previously defined.

Based on this last observation we have the following result that allows us to use the lookdown construction in the inhomogeneous setting for a birth and death process. For $u = (u_1, \dots, u_n) \in E$, define

$$\psi(u,t) = 1 + (|b(t)| + \sigma(t))n,$$

$\gamma(u,t) = (n,t)$ and $\alpha((n,t), \mathrm{d}(u,s)) := \widetilde{\alpha}(n, \mathrm{d}u)\delta_t(\mathrm{d}s)$, where $\widetilde{\alpha}(n,\mathrm{d}u)$ is the joint distribution of n independent uniform random variables over $[0,K]$. Recalling the definition of $\widetilde{\gamma}$ from Theorem 4.2, yields $\widetilde{\psi}(n,t) = 1 + (|b(t)| + \sigma(t))n$.

Lemma C.1. *Let $K \in \mathbb{N}^*$ and $b, \sigma : \mathbb{R}_+ \to \mathbb{R}$ two bounded continuous functions satisfying $K\sigma(t) - b(t) \geq 0$ and $\sigma(t) \geq 0$ for each $t \geq 0$. If X^K is a solution of the martingale problem for*

$$C_t^K \hat{f}(n) = K\sigma(t)n\Big(\hat{f}(n+1) - \hat{f}(n)\Big) + (K\sigma(t) - b(t))\Big(\hat{f}(n-1) - \hat{f}(n)\Big),$$

satisfying

$$\mathbb{E}\left(\int_0^t \widetilde{\psi}(X^K(s), s)\,\mathrm{d}s\right) < \infty, \quad \forall t \geq 0,$$

then there exists a solution U^K of the martingale problem for A_t^K.

Proof. First we notice that

$$\begin{aligned} |\overline{A}^K f\zeta(u,t)| &\leq \|\zeta\| |A_t^K f(u)| + \|f\| \|\zeta'\| \\ &\leq \|\zeta\| \|g'\| (K^2 + K)\psi(u,t) + \|\zeta'\| \\ &\leq c_{f\zeta}\psi(u,t), \end{aligned}$$

where $c_{f\zeta}$ is constant that depends on $f\zeta$. Then, since $\int \overline{A}^K f\zeta(u,s)\alpha((n,t),\mathrm{d}(u,s)) = C_t^K \hat{f}(n) + \hat{f}(n)\zeta'(t)$, where $\hat{f}(n) = \int f(u)\widetilde{\alpha}(n,\mathrm{d}u)$, is the space-time generator associated with the birth and death process, the result follows from the conclusions of Theorem 4.2. □

We also have the same result for the Feller diffusion with time-dependent coefficients.

Consider now the state space $E = [0,\infty)^\infty \cup \bigcup_{k=0}^\infty [0,\infty)^k$, the generator A_t from (4.2.5) and its corresponding domain $D(A)$. As before, define $D(\overline{A})$ and the space-time generator

$$\overline{A} f\zeta(u,t) = A_t f(u)\zeta(t) + f(u)\zeta'(t).$$

Also, define for $(u,t) \in E \times [0,\infty)$

$$\psi(u,t) = 1 + (|b(t)| + |\sigma(t)|)\sum_i e^{-u_i},$$

$\gamma(u,t) = (\limsup_{N\to\infty} \frac{1}{N}\sum_i \mathbf{1}_{[0,N]}(u_i), t)$ and $\alpha((y,t),\mathrm{d}(u,s)) := \widetilde{\alpha}(y,\mathrm{d}u)\delta_t(\mathrm{d}s)$, where $\widetilde{\alpha}(y,\mathrm{d}u)$ is the distribution of a Poisson process of parameter y. In this case we obtain $\widetilde{\psi}(y,t) = 1 + (|b(t) + \sigma(t)|)y$.

Lemma C.2. *Let $b, \sigma \colon \mathbb{R}_+ \to \mathbb{R}$ two bounded continuous functions such that $\sigma(t) \geq 0$ for each $t \geq 0$. If Y is a solution of the martingale problem for*

$$C_t \hat{f}(y) = \sigma(t) y \hat{f}''(y) + b(t) y \hat{f}'(y),$$

satisfying

$$\mathbb{E}\left(\int_0^t \widetilde{\psi}(Y(s), s) \, \mathrm{d}s \right) < \infty, \quad \forall t \geq 0,$$

then there exists a solution U of the martingale problem for A_t.

Proof. As before, since

$$\begin{aligned} |\overline{A} f \zeta(u,t)| &\leq \|\zeta\| |A_t f(u)| + \|f\| \|\zeta'\| \\ &\leq \|\zeta\| \|g'\| (v_g^2 + v_g) e^{v_g} \psi(u,t) + \|\zeta'\| \\ &\leq c_{f\zeta} \psi(u,t), \end{aligned}$$

and $\int \overline{A} f \zeta(u,s) \alpha((y,s), \mathrm{d}(u,s)) = C_t \hat{f}(y) + \hat{f}(y) \zeta'(t)$, where $\hat{f}(y) = \int f(u) \widetilde{\alpha}(y, \mathrm{d}u)$, the conclusion follows as in the proof of the previous Lemma. □

Bibliography

[1] D. Aldous, *The Continuum Random Tree. I*, The Annals of Probability **19**, no. 1, (1991), 1–28.

[2] D. Aldous, *The Continuum Random Tree. II. An Overview*, Stochastic Analysis: Proceedings of the Durham Symposium on Stochastic Analysis, 1990, London Mathematical Society Lecture Note Series **167**, (1991), 23–70.

[3] D. Aldous, *The Continuum Random Tree. III*, The Annals of Probability **21**, no. 1, (1993), 248–289.

[4] H. Amann, *Dynamic Theory of Quasilinear Parabolic Equations. II. Reaction-Diffusion Systems*, Differential and Integral Equations **3**, no. 1, (1990), 13–75. Zbl 0729.35062.

[5] H. Amann, *Dynamic Theory of Quasilinear Parabolic Systems. III. Global Existence (Erratum).*, Mathematische Zeitschrift **205**, no. 2, (1990), 331–332. Zbl 0719.35044.

[6] K. B. Athreya and P. E. Ney, *Branching Processes*, Die Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, (1972).

[7] V. Bansaye and S. Méléard, *Stochastic Models for Structured Populations. Scaling Limits and Long Time Behavior*, Mathematical Biosciences Institute Lecture Series, Springer, Cham, 2015.

[8] V. Bansaye, A. Moussa, and F. Muñoz-Hernández, *Stability of a Cross-Diffusion System and Approximation by Repulsive Random Walks: A Duality Approach*, Preprint, (2021). arXiv:2109.07146.

[9] V. Bansaye, J. C. Pardo, and C. Smadi, *Extinction Rate of Continuous State Branching Processes in Critical Lévy Environments*, ESAIM: Probability and Statistics **25**, (2021), 346–375.

[10] V. Bansaye, J. C. Pardo Millan, and C. Smadi, *On the Extinction of Continuous State Branching Processes with Catastrophes*, Electronic Journal of Probability **18**, (2013), 1–31.

[11] V. Bansaye and F. Simatos, *On the Scaling Limits of Galton-Watson Processes in Varying Environments*, *Electronic Journal of Probability* **20**, (2015), 1–36.

[12] V. Bansaye and V. C. Tran, *Branching Feller Diffusion for Cell Division with Parasite Infection*, ALEA: Latin American Journal of Probability and Mathematical Statistics **8**, (2011), 95–127.

[13] J. Bertoin and J.-F. Le Gall, *The Bolthausen-Sznitman Coalescent and the Genealogy of Continuous-State Branching Processes*, Probability Theory and Related Fields **117**, no. 2, (2000), 249–266.

[14] J. Bertoin and J.-F. Le Gall, *Stochastic Flows Associated to Coalescent Processes*, Probability Theory and Related Fields **126**, no. 2, (2003), 261–288.

[15] J. Bertoin and J.-F. Le Gall, *Stochastic Flows Associated to Coalescent Processes. II. Stochastic Differential Equations*, Annales de l'Institut Henri Poincaré. Probabilités et Statistiques **41**, no. 3, (2005), 307–333.

[16] J. Bertoin and J.-F. Le Gall, *Stochastic Flows Associated to Coalescent Processes. III. Limit Theorems*, Illinois Journal of Mathematics **50** no. 1-4, (2006), 147–181.

[17] M. Birkner, J. Blath, M. Capaldo, A. Etheridge, M. Möhle, J. Schweinsberg, and A. Wakolbinger, *Alpha-Stable Branching and Beta-Coalescents*, Electronic Journal of Probability **10**, no. 9, (2005), 303–325.

[18] M. Birkner, J. Blath, M. Martin, M. Steinrücken, and J. Tams, *A Modified Lookdown Construction for the Xi-Fleming-Viot Process with Mutation and Populations with Recurrent Bottlenecks*, ALEA: Latin American Journal of Probability and Mathematical Statistics **6**, (2009), 25–61.

[19] C. Böinghoff and M. Hutzenthaler, *Branching Diffusions in Random Environment*, Preprint, (2013). arXiv:1107.2773.

[20] K. A. Borovkov and V. A. Vatutin, *Reduced Critical Branching Processes in Random Environment*, Stochastic Processes and their Applications **71**, no. 2, (1997), 225–240.

[21] M. Breden, C. Kuehn, and C. Soresina, *On the Influence of Cross-Diffusion in Pattern Formation*, Journal of Computational Dynamics **8**, no. 2, (2021), 213–240.

[22] J. A. Carrillo, Y.-P. Choi, and M. Hauray, *The Derivation of Swarming Models: Mean-Field Limit and Wasserstein Distances*, Collective Dynamics from Bacteria to Crowds **553**, Springer Vienna, (2014), 1–46.

[23] J. A. Carrillo, Y. Huang, and M. Schmidtchen, *Zoology of a Nonlocal Cross-Diffusion Model for Two Species*, SIAM Journal on Applied Mathematics **78** no. 2, (2018), 1078–1104.

[24] N. Champagnat, R. Ferrière, and S. Méléard, *Unifying Evolutionary Dynamics: From Individual Stochastic Processes to Macroscopic Models*, Theoretical Population Biology **69**, no. 3, (2006), 297–321.

[25] N. Champagnat, R. Ferrière, and S. Méléard, *From Individual Stochastic Processes to Macroscopic Models in Adaptive Evolution*, Stochastic Models **24**, no. sup1, (2008), 2–44.

[26] N. Champagnat and S. Méléard, *Invasion and Adaptive Evolution for Individual-Based Spatially Structured Populations*, Journal of Mathematical Biology **55**, no. 2, (2007), 147–188.

[27] L. Chen, E. S. Daus, A. Holzinger, and A. Jüngel, *Rigorous Derivation of Population Cross-Diffusion Systems from Moderately Interacting Particle Systems*, Preprint, (2020). arXiv:2010.12389.

[28] L. Chen, E. S. Daus, and A. Jüngel, *Rigorous Mean-Field Limit and Cross-Diffusion*, Zeitschrift für angewandte Mathematik und Physik **70**, no. 4, (2019), 122.

[29] L. Chen and A. Jüngel, *Analysis of a Multidimensional Parabolic Population Model with Strong Cross-Diffusion*, SIAM Journal on Mathematical Analysis **36**, no. 1, (2004), 301–322.

[30] L. Chen and A. Jüngel, *Analysis of a Parabolic Cross-Diffusion Population Model without Self-Diffusion*, Journal of Differential Equations **224**, no. 1, (2006) 39–59.

[31] X. Chen, E. S. Daus, and A. Jüngel, *Global Existence Analysis of Cross-Diffusion Population Systems for Multiple Species*, Archive for Rational Mechanics and Analysis **227**, no. 2, (2018), 715–747.

[32] X. Chen and A. Jüngel, *A Note on the Uniqueness of Weak Solutions to a Class of Cross-Diffusion Systems*, Journal of Evolution Equations **18**, no. 2, (2018), 805–820.

[33] X. Chen and A. Jüngel, *Weak–Strong Uniqueness of Renormalized Solutions to Reaction–Cross-Diffusion Systems*, Mathematical Models and Methods in Applied Sciences **29**, no. 02, (2019) 237–270.

[34] B. Cloez and B. Henry, *Markovian Tricks for Non-Markovian Trees: Contour Process, Extinction and Scaling Limits*, Preprint, (2018). arXiv:1801.03284.

[35] R. Cortez and J. Fontbona, *Quantitative Propagation of Chaos for Generalized Kac Particle Systems*, The Annals of Applied Probability **26**, no. 2, (2016), 892–916.

[36] R. Cortez and J. Fontbona, *Quantitative Uniform Propagation of Chaos for Maxwell Molecules*, Communications in Mathematical Physics **357**, no. 3, (2018), 913–941.

[37] E. S. Daus, L. Desvillettes, and H. Dietert, *About the Entropic Structure of Detailed Balanced Multi-Species Cross-Diffusion Equations*, Journal of Differential Equations **266**, no. 7, (2019), 3861–3882.

[38] E. S. Daus, L. Desvillettes, and A. Jüngel, *Cross-Diffusion Systems and Fast-Reaction Limits*, Bulletin des Sciences Mathématiques **159**, (2020), 102824.

[39] D. Dawson, *Measure-Valued Markov Processes*, École d'Eté de Probabilités de Saint-Flour XXI–1991, Lecture Notes in Mathematics **1541**, Springer Berlin Heidelberg, (1993), 1–260.

[40] D. A. Dawson and K. J. Hochberg, *Wandering Random Measures in the Fleming-Viot Model*, The Annals of Probability **10**, no. 3, (1982), 554–580.

[41] A. Depperschmidt and A. Greven, *Tree-Valued Feller Diffusion*, Preprint, (2019). arXiv:1904.02044.

[42] L. Desvillettes and A. Trescases, *New Results for Triangular Reaction Cross Diffusion System*, Journal of Mathematical Analysis and Applications **430**, no. 1, (2015), 32–59.

[43] P. Deuring, *An Initial-Boundary-Value Problem for a Certain Density-Dependent Diffusion System*, Mathematische Zeitschrift **194**, no. 3, (1987), 375–396.

[44] G. Dhariwal, F. Huber, and A. Jüngel, *Global Martingale Solutions for a Stochastic Shigesada-Kawasaki-Teramoto Population Model*, Preprint, (2020). arXiv:2012.12765.

[45] G. Dhariwal, F. Huber, A. Jüngel, C. Kuehn, and A. Neamţu, *Global Martingale Solutions for Quasilinear SPDEs via the Boundedness-by-Entropy Method*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **57**, no. 1, (2021), 577–602.

[46] M. Di Francesco, A. Esposito, and S. Fagioli, *Nonlinear Degenerate Cross-Diffusion Systems with Nonlocal Interaction*, Nonlinear Analysis **169**, (2018), 94–117.

[47] H. Dietert and A. Moussa, *Persisting Entropy Structure for Nonlocal Cross-Diffusion Systems*, Preprint, (2021). arXiv:2101.02893.

[48] P. Donnelly and T. G. Kurtz, *A Countable Representation of the Fleming-Viot Measure-Valued Diffusion*, The Annals of Probability **24**, no. 2, (1996), 698–742.

[49] P. Donnelly and T. G. Kurtz, *Genealogical Processes for Fleming-Viot Models with Selection and Recombination*, The Annals of Applied Probability **9**, no. 4, (1999), 1091–1148.

[50] P. Donnelly and T. G. Kurtz, *Particle Representations for Measure-Valued Population Models*, The Annals of Probability **27**, no. 1, (1999), 166–205.

[51] T. Duquesne and J.-F. Le Gall, *Probabilistic and Fractal Aspects of Lévy Trees*, Probability Theory and Related Fields **131**, no. 4, (2005) 553–603.

[52] A. M. Etheridge and T. G. Kurtz, *Genealogical Constructions of Population Models*, The Annals of Probability **47**, no. 4, (2019), 1827–1910.

[53] S. N. Ethier and T. G. Kurtz, *Markov Processes*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, (1986).

[54] S. N. Evans, *Probability and Real Trees*, Lecture Notes in Mathematics **1920**, Springer, Berlin, (2008).

[55] R. Fang and Z. Li, *Construction of Continuous-State Branching Processes in Varying Environments*, Preprint, (2020). arXiv:2002.09113.

[56] W. Feller, *Diffusion Processes in Genetics*, Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, University of California Press, Berkeley and Los Angeles, (1951), 227–246.

[57] K. Fleischmann and R. Siegmund-Schultze, *The Structure of Reduced Critical Galton-Watson Processes*, Mathematische Nachrichten **79**, (1977), 233–241.

[58] J. Fontbona and S. Méléard, *Non Local Lotka-Volterra System with Cross-Diffusion in an Heterogeneous Medium*, Journal of Mathematical Biology **70**, no. 4, (2015), 829–854.

[59] J. Fontbona and F. Muñoz-Hernández, *Quantitative mean-field limit for interacting branching diffusions*, Preprint, (2021). arXiv:2101.04099.

[60] N. Fournier and A. Guillin, *On the Rate of Convergence in Wasserstein Distance of the Empirical Measure*, Probab. Theory Related Fields **162**, no. 3-4, (2015), 707–738.

[61] N. Fournier and S. Méléard, *A Microscopic Probabilistic Description of a Locally Regulated Population and Macroscopic Approximations*, The Annals of Applied Probability **14**, no. 4, (2004), 1880–1919.

[62] M. D. Francesco and S. Fagioli, *Measure Solutions for Non-Local Interaction PDEs with Two Species*, Nonlinearity **26**, no. 10, (2013), 2777–2808.

[63] G. Galiano, M. L. Garzón, and A. Jüngel, *Semi-Discretization in Time and Numerical Convergence of Solutions of a Nonlinear Cross-Diffusion Population Model*, Numerische Mathematik **93**, no. 4, (2003), 655–673.

[64] A. Grimvall, *On the Convergence of Sequences of Branching Processes*, The Annals of Probability **2**, no. 6, (1974), 1027–1045.

[65] R. Hardy and S. C. Harris, *A Spine Approach to Branching Diffusions with Applications to L^p-Convergence of Martingales*, *Séminaire de Probabilités XLII* **1979**, Springer Berlin Heidelberg, (2009), 281–330.

[66] H. He, Z. Li, and W. Xu, *Continuous-State Branching Processes in Lévy Random Environments*, Journal of Theoretical Probability **31**, no. 4, (2018), 1952–1974.

[67] E. Hewitt and L. J. Savage, *Symmetric Measures on Cartesian Products*, Transactions of the American Mathematical Society **80**, no. 2, (1955), 470–470.

[68] L. T. Hoang, T. V. Nguyen, and T. V. Phan, *Gradient Estimates and Global Existence of Smooth Solutions to a Cross-Diffusion System*, SIAM Journal on Mathematical Analysis **47**, no. 3, (2015), 2122–2177.

[69] M. Iida, M. Mimura, and H. Ninomiya, *Diffusion, Cross-Diffusion and Competitive Interaction*, Journal of Mathematical Biology **53**, no. 4, (2006), 617–641.

[70] M. Jiřina, *Stochastic Branching Processes with Continuous State Space*, Czechoslovak Mathematical Journal **8 (83)**, (1958), 292–313.

[71] A. Jüngel, *The Boundedness-by-Entropy Method for Cross-Diffusion Systems*, Nonlinearity **28**, no. 6, (2015), 1963–2001.

[72] M. Kac, *Foundations of Kinetic Theory*, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, Vol. III, University of California Press, Berkeley and Los Angeles, (1956), 171–197.

[73] G. Kersting, *On the Genealogical Structure of Critical Branching Processes in a Varying Environment*, Preprint, (2021). arXiv:2102.08324.

[74] J. U. Kim, *Smooth Solutions to a Quasi-Linear System of Diffusion Equations for a Certain Population Model*, Nonlinear Analysis: Theory, Methods & Applications **8**, no. 10, (1984), 1121–1144.

[75] J. F. C. Kingman, *The Coalescent*, Stochastic Processes and their Applications **13**, no. 3, (1982), 235–248.

[76] C. Kipnis and C. Landim, *Scaling Limits of Interacting Particle Systems*, *Grundlehren Der Mathematischen Wissenschaften* **320**, Springer Berlin Heidelberg, (1999).

[77] T. G. Kurtz, *Diffusion Approximations for Branching Processes*, Advances in Probability and Related Topics **5**, (1978), 269–292.

[78] T. G. Kurtz, *Martingale Problems for Conditional Distributions of Markov Processes*, Electronic Journal of Probability **3**, no. 9, (1998), 1–29.

[79] T. G. Kurtz and G. Nappo, *The Filtered Martingale Problem*, The Oxford Handbook of Nonlinear Filtering, Oxford University Press, (2011), 129–165.

[80] T. G. Kurtz and E. R. Rodrigues, *Poisson Representations of Branching Markov and Measure-Valued Branching Processes*, The Annals of Probability **39**, no. 3, (2011), 939–984.

[81] A. Lambert, *The Genealogy of Continuous-State Branching Processes with Immigration*, Probability Theory and Related Fields **122**, no. 1, (2002), 42–70.

[82] A. Lambert, *The Contour of Splitting Trees Is a Lévy Process*, The Annals of Probability **38**, no. 1, (2010), 348–395.

[83] A. Lambert and L. Popovic, *The Coalescent Point Process of Branching Trees*, The Annals of Applied Probability **23**, no. 1, (2013), 99–144.

[84] A. Lambert and T. Stadler, *Birth–Death Models and Coalescent Point Processes: The Shape and Probability of Reconstructed Phylogenies*, Theoretical Population Biology **90**, (2013), 113–128.

[85] J. Lamperti, *The Limit of a Sequence of Branching Processes*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **7**, (1967), 271–288.

[86] J.-F. Le Gall, *Random Trees and Applications*, Probability Surveys **2**, (2005), 245–311.

[87] J.-F. Le Gall and Y. Le Jan, *Branching Processes in Lévy Processes: The Exploration Process*, The Annals of Probability **26**, no. 1, (1998), 213–252.

[88] Z. Li, *Continuous-State Branching Processes with Immigration*, From Probability to Finance, Springer Singapore, (2020), 1–69.

[89] A. J. Lotka, *Elements of Physical Biology*, Williams & Wilkins Company, (1925).

[90] Y. Lou and W.-M. Ni, *Diffusion, Self-Diffusion and Cross-Diffusion*, Journal of Differential Equations **131**, no. 1, (1996), 79–131.

[91] Y. Lou, W.-M. Ni, and Y. Wu, *On the Global Existence of a Cross-Diffusion System*, Discrete & Continuous Dynamical Systems **4**, no. 2, (1998), 193–203.

[92] Y. Lou and M. Winkler, *Global Existence and Uniform Boundedness of Smooth Solutions to a Cross-Diffusion System with Equal Diffusion Rates*, Communications in Partial Differential Equations **40**, no. 10, (2015), 1905–1941.

[93] J. Maas, *Gradient Flows of the Entropy for Finite Markov Chains*, Journal of Functional Analysis **261**, no. 8, (2011), 2250–2292.

[94] R. Martin and M. Pierre, *Nonlinear Reaction-Diffusion Systems*, *Mathematics in Science and Engineering* **185**, Elsevier, (1992), 363–398.

[95] S. Méléard, *Asymptotic Behaviour of Some Interacting Particle Systems; McKean-Vlasov and Boltzmann Models*, Probabilistic Models for Nonlinear Partial Differential Equations **1627**, Springer Berlin Heidelberg, (1996), 42–95.

[96] M. Mimura and K. Kawasaki, *Spatial Segregation in Competitive Interaction-Diffusion Equations*, Journal of Mathematical Biology **9**, no. 1, (1980), 49–64.

[97] A. Moussa, *From Nonlocal to Classical Shigesada–Kawasaki–Teramoto Systems: Triangular Case with Bounded Coefficients*, SIAM Journal on Mathematical Analysis **52**, no. 1, (2020), 42–64.

[98] J. Neveu, *Arbres et Processus de Galton-Watson*, Annales de l'Institut Henri Poincaré. Probabilités et Statistique **22**, no. 2, (1986), 199–207.

[99] S. Ostrander, *Macroscopic Cross-Diffusion Models Derived from Spatially Discrete Continuous Time Microscopic Models*, SIAM Undergraduate Research Online **4**, (2011), 51–71.

[100] S. Palau and J. C. Pardo, *Branching Processes in a Lévy Random Environment*, Acta Applicandae Mathematicae **153**, (2018), 55–79.

[101] S. Palau, J. C. Pardo, and C. Smadi, *Asymptotic Behaviour of Exponential Functionalsof Lévy Processes with Applications to Randomprocesses in Random Environment*, Latin American Journal of Probability and Mathematical Statistics **13**, no. 1, (2016), 1235–1258.

[102] É. Pardoux, *Probabilistic Models of Population Evolution: Scaling Limits, Genealogies and Interactions*, Springer International Publishing, Cham, (2016).

[103] P. Pfaffelhuber and A. Wakolbinger, *The Process of Most Recent Common Ancestors in an Evolving Coalescent*, Stochastic Processes and their Applications **116**, no. 12, (2006), 1836–1859.

[104] D. Pham and R. Temam, *A Result of Uniqueness of Solutions of the Shigesada–Kawasaki–Teramoto Equations*, Advances in Nonlinear Analysis **8**, no. 1, (2019), 497–507.

[105] M. Pierre and D. Schmitt, *Blowup in Reaction-Diffusion Systems with Dissipation Of Mass*, SIAM Journal on Mathematical Analysis **28**, no. 2, (1997), 259–269.

[106] J. Quastel, *Diffusion of Color in the Simple Exclusion Process*, Communications on Pure and Applied Mathematics **45**, no. 6, (1992), 623–679.

[107] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, Grundlehren Der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] **293**, Springer-Verlag, Berlin, 3ed., (1999).

[108] N. Shigesada, K. Kawasaki, and E. Teramoto, *Spatial Segregation of Interacting Species*, Journal of Theoretical Biology **79**, no. 1, (1979), 83–99.

[109] W. L. Smith and W. E. Wilkinson, *On Branching Processes in Random Environments*, The Annals of Mathematical Statistics **40**, no. 3, (1969), 814–827.

[110] A.-S. Sznitman, *Topics in Propagation of Chaos*, École d'Été de Probabilités de Saint-Flour XIX—1989, *Lecture Notes in Mathematics* **1464**, Springer, Berlin, (1991), 165–251.

[111] A. Trescases, *On Triangular Reaction Cross-Diffusion Systems with Possible Self-Diffusion*, Bulletin des Sciences Mathématiques **140**, no. 7, (2016), 796–829.

[112] V. A. Vatutin, *Reduced Branching Processes in Random Environment: The Critical Case*, Theory of Probability & Its Applications **47**, no. 1, (2003), 99–113.

[113] A. Véber and A. Wakolbinger, *The Spatial Lambda-Fleming–Viot Process: An Event-Based Construction and a Lookdown Representation*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **51**, no. 2, (2015), 570–598.

[114] C. Villani, *Optimal Transport*, Grundlehren Der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] **338**, Springer-Verlag, Berlin, (2009).

[115] V. Vladimir and D. Elena, *Reduced Branching Processes in Random Environment*, Mathematics and Computer Science II, Birkhäuser Basel, Basel, (2002), 455–467.

[116] V. Volterra, *Variazioni e Fluttuazioni Del Numero d'individui in Specie Animali Conviventi*, Memorie della Regia Accademia Nazionale dei Lincei **2**, (1926), 31–113.

[117] N. Zamponi and A. Jüngel, *Analysis of Degenerate Cross-Diffusion Population Models with Volume Filling*, Annales de l'Institut Henri Poincaré C, Analyse non linéaire **34**, no. 1, (2017), 1–29.