Preface

"... It also happens...
that branches which were thought
to be completely disparate
are suddenly seen to be related ..."

Michiel Hazewinkel, 1977

During September 4–9, 2006 the Stefan Banach International Center (BC) of the Institute of Mathematics of the Polish Academy of Sciences and the (Italian) International Summer Institute for Mathematics (CIME) jointly organized at the Mathematical Research and Conference Center (MRCC), Będlewo, Poland, the School *From a Microscopic to Macroscopic Description of Complex Systems*. In addition to the main speakers, whose contributions are included in this volume, there were a significant number of participants from 11 countries. In parallel with the School, a workshop on *Modelling Cellular Systems with Applications to Tumor Growth* was organized in the framework of the activity of the EU MCRTN "MRTN-CT-2004-503661". The courses were targeted at Ph.D. students and young researchers and have had an educational character, whereas the workshop offered presentations on particular applications to modelling tumour growth phenomena.

The aim of the School has been to offer a broad presentation of updated methods suitable to provide a mathematical framework for the development of a hierarchy of models of complex systems in the natural sciences, with special attention to biology and medicine. The mastering of complexity implies the sharing of different tools which require a much higher level of communication between different mathematical and scientific schools, for solving classes of problems of the same nature. Nowadays, more than ever, one of the most important challenges derives from the bridging of parts of a system evolving at different time and space scales, especially with respect to computational affordability. Therefore, the courses have had a rather general character and

method; the main role is played here by stochastic processes, positive semigroups, asymptotic analysis, continuum theory and game theory.

For many biological systems only non-negative states or solutions make sense. The theory of Banach lattices and positive operators, developed in the series of lectures *Positivity in the Natural Sciences"* by Jacek Banasiak, provides a mathematical framework to address such problems. The lectures show how the interplay of positivity and compactness yields very strong results in many fields ranging from well-posedness of the problem at hand, through long time behaviour of solutions (including emergence of chaos), to asymptotic analysis of systems displaying multiple scale phenomena. Theoretical results are applied to a variety of specific problems occurring in natural sciences, including birth-and-death type models that describe the development of drug resistance in cancer cells, blood cells' evolution equation, singularly perturbed models of sole migration, or diffusion approximation of the Fokker–Planck equation.

As a paradigmatic microcosm for all of biology, i.e. as an observable system where mutation and evolution take place, in the course on *Cancer* by Mark Chaplain, the different aspects of the growth phases of a tumour are described(raise query?) solid tumours (the most frequent of all cancers) progress through several key stages of growth from a single transformed/mutated cell, to a multicell spheroid (avascular growth), vascular growth in connection with blood vessels, and finally invasive growth of the local tissue and metastasis to distant sites where secondary tumours occur. Modelling these growth phases involves a mixture of continuum models (ordinary, delay and partial differential equations – reaction–diffusion–taxis equations) and individual-based models (cellular automata, discrete modelling techniques). The lecture notes present a range of mathematical techniques to examine a family of models (qualitative analysis of DEs, asymptotic analysis, numerical analysis and computation) and provide a general framework for developing quantitative and predictive models.

The mathematical framework for searching for links between solutions related to equation modelling at the microscopic, mesoscopic and macroscopic levels is the topic of the course *Links between microscopic and macroscopic description* by Mirosław Lachowicz. Usually, the description of biological populations is carried out on a macroscopic level of interacting sub-populations. The mathematical structures are deterministic reaction–diffusion equations. They describe the (deterministic) evolution of densities of subpopulations rather than the interactions between their individual entities. However, in many cases the description on a micro-scale of interacting entities (e.g. cells) seems to be more appropriate. The problem of relationships between the various scales of description seems to be one of the most important problems of the mathematical modelling of complex systems, e.g. in the modeling of tumour growth. The following strategy can be applied. One starts with the deterministic macroscopic model for which the identification of parameters by an experiment is easier. Then one provides the theoretical framework for modelling at the

microscopic scale in such a way that the corresponding models at the macro- and micro-scales are asymptotically equivalent, i.e. the solutions are close to each other in a properly chosen norm. Then, if the microscopic model is chosen suitably, one may hope that it covers not only the macroscopic behaviour of the system in question, but also some of its microscopic features. The microscopic model by its nature is richer and it may describe a larger variety of phenomena. In mathematical terms, we are interested in the links between the following mathematical structures: at the micro-scale of stochastically interacting entities (cells, individuals,...), in terms of continuous stochastic semigroups, the meso-scale of statistical entities, in terms of continuous nonlinear semigroups related to the solutions of Boltzmann-type nonlocal kinetic equations, and the macroscale of densities of interacting entities in terms of dynamical systems related to reaction–diffusion equations.

The notes on *Rescaling Stochastic Processes: Asymptotics*, by Vincenzo Capasso and Daniela Morale, investigate the links among different scales, from a more probabilistic point of view. As already mentioned, particular attention is being paid to the mathematical modelling of the social behaviour of interacting individuals in a biological population, on the one hand because there is an intrinsic interest in the dynamics of population herding, and on the other hand, as agent-based models are being used in complex optimization problems. Among other interesting features, these systems lead to self-organization phenomena, which exhibit interesting spatial patterns. Here, we show how properties on the macroscopic level depend on interactions at the microscopic level; in particular, suitable laws of large numbers are shown to imply the convergence of the evolution equations for empirical spatial distributions of interacting individuals to nonlinear reaction–diffusion equations for a so-called mean field, as the total number of individuals becomes sufficiently large. As a working example, an interacting particle system modelling social behaviour has been proposed, based on a system of stochastic differential equations, driven by both aggregating/repelling and external forces. To support a rigorous derivation of the asymptotic nonlinear integro-differential equation, compactness criteria for convergence in metric spaces of measures, and problems of existence of a weak/entropic solution have been analyzed. Further the temporal asymptotic behaviour of the stochastic system of a fixed number of interacting particles has been discussed. This leads to the problem of the existence of nontrivial invariant probability measures.

These microscopic interactions between individuals can often be described within game-theoretic models. This theme has been discussed in the notes on *Evolutionary Game Theory and Population Dynamics* by Jacek Miękisz. In such evolutionary models, individuals adapt to a changing environment and are subject to selection pressure and mutations. We will present deterministic and stochastic models of adaptive dynamics and discuss the stability of equilibria in appropriate dynamical systems such as time-delay equations and Markov chains.

As Co-Directors, we are pleased to thank both institutions, BC and CIME, in particular the former Director of the Institute of Mathematics and founder of MRCC, Professor Bogdan Bojarski, and the Director of CIME Professor Pietro Zecca, for letting us organize the joint school, and for their continuous support. It has had a special meaning at the time of extension of Europe towards the East; in this way a concrete occasion has been offered to the young participants to contribute actively in building a common European Research Area. They have had the chance of experiencing directly that different Schools in Europe may actively contribute to making the European Union a highly competitive scientific community.

Our special thanks are due to our Colleagues, Professors Banasiak, Chaplain and Miękisz, for their careful preparation and stimulating presentation of the material, both at the school and in these Lecture Notes. We thank Dr Gabriela Lorelai Litcanu for her work in preparation of the workshop organized in connection with the courses. All the participants contributed to the creation of an exceptionally friendly atmosphere which also characterized the various social events organized in the beautiful environment of the Będlewo Palace. We thank the Director and the whole staff of MRCC in Będlewo for their warm and efficient hospitality. Thanks are due to Dr Daniela Morale for her editorial assistance during the preparation of this volume.

Milan, Warsaw,
March 2007

Vincenzo Capasso
Mirosław Lachowicz

Rescaling Stochastic Processes: Asymptotics

Vincenzo Capasso and Daniela Morale

Department of Mathematics, University of Milano, Via C. Saldini, 50, Milano, Italy
vincenzo.capasso@unimi.it, daniela.morale@mat.unimi.it

Summary. In this chapter the authors investigate the links among different scales, from a probabilistic point of view. Particular attention is being paid to the mathematical modelling of the social behavior of interacting individuals in a biological population, on one hand because there is an intrinsic interest in dynamics of population herding, on the other hand since agent based models are being used in complex optimization problems. Among other interesting features, these systems lead to phenomena of self-organization, which exhibit interesting spatial patterns. Here we show how properties on the macroscopic level depend on interactions at the microscopic level; in particular suitable laws of large numbers are shown to imply convergence of the evolution equations for empirical spatial distributions of interacting individuals to nonlinear reaction–diffusion equations for a so called mean field, as the total number of individuals becomes sufficiently large. As a working example, an interacting particle system modelling social behavior has been proposed, based on a system of stochastic differential equations, driven by both aggregating/repelling and external "forces". In order to support a rigorous derivation of the asymptotic nonlinear integro-differential equation, compactness criteria for convergence in metric spaces of measures, and problems of existence of a weak/entropic solution have been analyzed. Further the temporal asymptotic behavior of the stochastic system of a fixed number of interacting particles has been discussed. This leads to the problem of the existence of nontrivial invariant probability measure.

1 Introduction

In group-living animals, a wide range of behaviors like resting, foraging or moving are usually performed collectively. Observational and empirical evidence show that animal groups move across the landscape quite cohesively, which strongly suggests that a collective decision has been taken. Thus, it could be assumed that individual decisions lead to a common decision, allowing the group to remain cohesive [38–40]. Such self-organized processes allow groups to carry out collective actions in various environments without any lead, external control or central coordination.

Over the past couple of decades, a large amount of literature has been devoted to the mathematical modelling of self-organizing populations, based on the concepts of short-range/long-range "social interaction" among different individuals of a biological population. The aim of the modelling is to catch the main features of the interaction at the lower scale of single individuals that are responsible, at a larger scale, for a more complex behavior that leads to the formation of aggregating patterns.

A classical widespread approach has been based on either linear or nonlinear PDE's [24,27]. Such kind of models are often called *Eulerian models*; they describe the evolution of population densities by means of typically deterministic nonlinear partial differential equations of the advection-reaction-diffusion type

$$\rho_t + \nabla \cdot (\mathbf{v}\rho) = \nabla \cdot (D\nabla\rho) + \nu(\rho);$$

ρ is the population density, $\mathbf{v}$ is a velocity field and $\nu(\rho)$ is a possible additive reaction term which may include birth and death processes. The advection term may describe the interaction mechanisms among individuals (via the velocity $\mathbf{v}$), while the non-convective (diffusive) flux takes into account the spatial (random) dispersal of the population.

The advantages of the continuum approach are those of ease of analysis; it is useful in the case of large and dense populations [13]. The disadvantages of this approach include especially the fact that the identity of individuals is compromised. In many situations it is more appropriate to use discrete individual-based models, in which a finite number of individuals is considered and only a finite sequence of decisions is made by individuals. As pointed out by Durrett and Levin [13] and by Grünbaum and Okubo [18], an individual-based approach is also useful in deriving the correct limiting equation, as N increases to infinity, also in the case when the use of a continuum model can be justified.

As already mentioned, a fruitful approach suggested since long by various authors [4,13,25,26,29,30] is based on the modelling of the given population as a system of interacting individuals; each individual "particle" is embedded in the total population of N similar particles (the so called individual based model - IBM). Each of them is treated as a discrete particle subject to simple rules of movement. This is known as *Lagrangian approach*: particles are followed in their individual motion. Possible randomness may be included, so that the variation in time of the random location of the k-th individual in the group at time $t \geq 0$, $X_N^k(t) \in \mathbb{R}^d, k = 1, \ldots, N$ is described by a system of stochastic differential equations (SDEs). On the other hand particles are subject to specific forces of interaction which are included in the advection term. In other words, from a Lagrangian point of view, the state of a system of N particles may be described as a stochastic process $\{X_N^k(t)\}_{t\in\mathbb{R}_+}$ defined on a suitable probability space $(\Omega, \mathcal{F}, P)$ and valued in $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$, where $\mathcal{B}_{\mathbb{R}^d}$ is the Borel σ-algebra generated by intervals. If the number of particles is kept

constant, the time evolution of particle locations $X_N^k(t), k = 1, \ldots, N$ are subject to a system of stochastic differential equations (SDEs) of the following type

$$dX_N^k(t) = H_N^k(X_N^1(t), \ldots, X_N^N(t), t)\, dt + \sigma \left(X_N^1(t), \ldots, X_N^N(t), t\right)\, dW^k(t),$$

where, for $k = 1, \ldots, N$, the function H_N^k defined on $\mathbb{R}^{Nd} \times \mathbb{R}_+$, describes the deterministic force acting on the k-th particle, and randomness has been included via a family of independent Wiener processes. Let us remark that the choice of an additive noise, described by a Wiener process allows a rigorous mathematical analysis based on Itô calculus. This will be clarified later in these lecture notes.

We will show a possible procedure to model interaction among individuals, how to introduce and distinguish among different scales. Indeed, a mathematical way to consider different scales, in the above sense, is based on the choice of a "scaling" parameter, depending upon N; if we consider a system of N particles located in $\mathbb{R}^d$, in the macroscopic space-time coordinates the typical distance between neighboring particles is $O(N^{-1/d})$ and the order of the size of the whole space $O(1)$. In this respect, we may distinguish three main types of interactions:

a. *McKean-Vlasov interaction* (macroscale): any particle interacts with $O(N)$ other particles; collective long-range forces are predominant and the particles are weakly interacting; the range of interaction gets very large in comparison with the typical distance between neighboring particles, and its strength decreases fast, like $1/N$.
b. *hydrodynamic interaction* (microscale): any particle interacts with $O(1)$ other particles in a very small neighborhood with volume $O(1/N)$. The interaction gets short-ranged and rather strong for large N.
c. *moderate interaction* (mesoscale) [28]: any particle interacts with many $O\left(N^{1-\beta}\right)$ other particles in a small volume $O\left(N^{-\beta}\right)$.

By an Eulerian approach, the collective behavior of the discrete (in the number of particles) system, may be recovered in terms of the spatial distribution of particles at time t, expressed, in term of an empirical measure in the space of the probability measures on $\mathbb{R}^d$,

$$X_N(t) = \frac{1}{N} \sum_{k=1}^{N} \epsilon_{X_N^k(t)} \in \mathcal{M}_P(\mathbb{R}^d),$$

which measures the spatial relative frequency of the particles at time t. Again by methods of stochastic calculus, we may derive the time evolution for the empirical measure, and then analyze its asymptotic behavior as the number of individuals N tends to infinity. This is equivalent to study a "law of large numbers" for measures. In particular we may look for sufficient conditions for having a unique limit measure, which is absolutely continuous with respect to the Lebesgue measure, and determine the evolution equation for the density.

In conclusion, the Lagrangian and Eulerian approaches describe the system at different scales: the microscale, the finer scale description based on the stochastic behavior of individuals and the macroscale, the larger scale description based on the continuum behavior of population densities. "The central problem is to determine how information is transferred across scales, and what detail at fine scales is exactly necessary and sufficient for understanding patterns on averaged scales" [13].

Nowadays, one of the interesting mathematical field regards the analysis for providing a mathematically rigorous framework for bridging the gap between the micro and the macro scale.

In these lecture notes the authors discuss some aspects of the asymptotics of rescaled stochastic processes, in particular paying attention to the deterministic approximation of stochastic systems (see Section 4). In Section 4.1 they discuss a time change technique applied to a jump process, while in Sections 4.2 and 4.3 a general description is given of a particle systems via a system of SDEs, and, via an application of Itô's formula, obtain a weak formulation of a stochastic evolution equation for the empirical measure $X_N(t)$, and investigate the properties of the stochastic component. Then they introduce and discuss the concept of the weak convergence in the space $\mathcal{M}_P(\mathbb{R}^d)$. The compactness properties of $\mathcal{M}_P(\mathbb{R}^d)$ and, consequently, the space of the trajectories of the stochastic process X_N, i.e. $\mathcal{M}(C([0,T],\mathcal{M}(\mathbb{R}^d)))$ are presented. An alternative approach to handle multiple scales, including the analysis at a mesoscales is presented in [21].

In Section 5 a specific model for a stochastic interacting particle is introduced [25, 26]. It is shown a mathematical way to describe the interaction, via some interaction kernels which characterize both a McKean-Vlasov and a moderate interaction. The advection term also includes an external flow. The asymptotics of a system of N stochastic differential equations is analyzed, as the population size N increase to infinity. In this way a nonlinear partial integral differential equation is obtained for the asymptotic mean field. Then, equations for the path of each individual particle are given, driven by such mean field.

In Section 6 the long time behavior of the system of SDEs s analyzed, for a fixed N. The interest here is to study mechanisms that are responsible for stable aggregation. The concept of invariant measure is introduced. Both the pure interacting particle system and the system with both the interacting term and a suitable "confining" potential are considered; it is shown how, in the first case, the system cannot admit a nontrivial invariant distribution; in the latter, under suitable conditions on the "localizing" potential U, the system does admit a nontrivial invariant distribution to which the system converges. We notice that, by applying recent results by Veretennikov [35,36], the requirement on U are less restrictive about its convexity with respect to the requirements of previous literature [9, 22, 36]. Our interest about these topics has been addressed by a wide-through-years literature by several authors interested in modelling aggregation behavior, and studying existence and convergence to an invariant distribution [5, 22, 36, 37].

In order to make the lecture notes self consistent, in Sections 2 and 3 we introduce the stochastic processes of our interest, i.e. Markov processes and in particular diffusion, the Brownian motion and the stochatic Itô calculus.

Now we consider some simple examples of scaling of stochastic process very well know since the first courses in probability.

1.1 First Examples of Rescaling

In the theory of stochastic processes there are some very well-known rescalings which are useful to describe and introduce the fundamental stochastic processes, i.e. the Binomial, the Poisson and the Wiener Processes. Here we introduce the basic ideas.

The Binomial Process

Take $T = \mathbb{N}$. Let $(Y_j)_{j\in\mathbb{N}\setminus\{0\}}$ be a sequence of independent and identically distributed (i.i.d.) Bernoulli random variables, having common distribution $B(1,p)$, the Binomial distribution with parameters 1 and $p \in [0,1]$. The Bernoulli process can be defined as the discrete time process $(X(n))_{n\in\mathbb{N}}$, such that

1. $X(0) = 0$;
2. $X(n) = \sum_{j=1}^{n} Y_j$, for $n \geq 1$.

As a consequence we know that $X(n)$ has a $B(n,p)$ distribution, for any $n \in \mathbb{N}$, and the following properties hold

i) $X(0) = 0$;
ii) $X(n+r) - X(n)$ is independent of $\mathcal{F}_n = \sigma(X(1), \ldots, X(n))$, the σ-algebra generated by the process till time n, for any $n, r \in \mathbb{N}$; (property of independent increments);
iii) $X(n+r) - X(n)$ has a $B(r,p)$ distribution, i.e. a Binomial distribution with parameters r and p.

Rescaling the Binomial Process: The Poisson Process

Let $t \in \mathbb{R}_+$, $t > 0$, and let $n \in \mathbb{N} \setminus \{0\}$. We may rescale the Binomial process, $B(n,p)$ by choosing for the probability p the following one

$$p = \lambda \frac{t}{n} + o\left(\frac{1}{n}\right),$$

with $\lambda > 0$.

Consider the process

1. $Y^{(n)}(0) = 0$;
2. $Y^{(n)}(t) = \sum_{j=1}^{n} Y_j$,

where now $(Y_j)_{j\in\mathbb{N}\setminus\{0\}}$ is a sequence of i.i.d. Bernoulli random variables, having common distribution $B\left(1, \lambda\frac{t}{n} + o\left(\frac{1}{n}\right)\right)$.

It is well known that, for n tending to infinity, i.e. by considering an infinite sum of Bernoulli trials during the time interval $[0,t]$, the process $(Y^{(n)}(t))_{t\in\mathbb{R}+}$ "converges" to a Poisson process $(Y(t))_{t\in\mathbb{R}+}$, such that

i) $Y(0) = 0$;
ii) $Y(t+s) - Y(s)$ is independent of $\mathcal{F}_s$ for any $s, t \in \mathbb{R}+$; (property of independent increments);
iii) $Y(t+s) - Y(s)$ has a $P(\lambda t)$ distribution for any $s, t \in \mathbb{R}+$, that is a Poisson distribution with parameter λt.

Rescaling the Poisson Process: The Wiener Process

Take the standard Poisson process $(Y(t))_{t\in\mathbb{R}+}$, $(\lambda = 1)$, that is let $Y(t)$ have a $P(t)$ distribution. Let us re-scale both the time and the jump of this process by $N \in \mathbb{N} \setminus \{0\}$ so to obtain the process

$$\tilde{Y}(t) = \frac{1}{N} Y(Nt), \quad t \in \mathbb{R}_+.$$

The strong law of large numbers implies that

$$\lim_{N\to\infty} \tilde{Y}(t) - t = \lim_{N\to\infty} \frac{1}{N}[Y(Nt) - Nt] = 0, \quad \text{a.s.}$$

Indeed, thanks to the martingale properties of the Poisson process, and consequently to Doob's inequality, we have both a functional law of large numbers,

$$\lim_{N\to\infty} \sup_{u\le v} \frac{1}{N}|Y(Nt) - Nt| = 0, \quad \text{a.s.},$$

and a functional Central Limit Theorem; i.e. the rescaled process

$$\left(\frac{1}{\sqrt{N}}[Y(Nt) - Nt]\right)_{t\in\mathbb{R}+},$$

for N tending to infinity, "converges" to the standard Wiener process $(W(t))_{t\in\mathbb{R}+}$, which is defined by the following

i) $W(0) = 0$;
ii) $W(t+s) - W(s)$ is independent of $\mathcal{F}_s$ for any $s, t \in \mathbb{R}+$; (property of independent increments);
iii) $W(t)$ has a $N(0,t)$ distribution, that is a zero mean normal distribution with variance t, for any $t \in \mathbb{R}+$.

The Wiener process can also be obtained by rescaling another discrete time process of great interest in applications.

The Simple Random Walk

Let $(Y_j)_{j\in\mathbb{N}\setminus\{0\}}$ be a sequence of i.i.d. dichotomic random variables, having common distribution $P(Y_j = 1) = P(Y_j = -1) = 1/2$. The simple random walk is defined as the discrete time process $(S_n)_{n\in\mathbb{N}}$, such that

1. $S_0 = 0$;
2. $S_n = \sum_{j=1}^{n} Y_j$, for $n \geq 1$.

Clearly, for each $n \in \mathbb{N}$, we have that $E[X_n] = 0$ and $Var[X_n] = n$.

Rescaling the Simple Random Walk

Let $t \in \mathbb{R}_+$, $t > 0$, and let $n \in \mathbb{N} - \{0\}$. Let $S_n, n \geq 1$ and $Y_j, j = 1, \ldots, n$ be defined as above. Define now the following rescaled and linearly interpolated process

$$X_n(t) := \frac{1}{\sqrt{n}} S_{[nt]} + (nt - [nt]) \frac{1}{\sqrt{n}} Y_{[nt]+1}, \quad t \in \mathbb{R}_+.$$

Donsker's theorem [33] states that the process $(X_n(t))_{t\in\mathbb{R}+}$, for n tending to infinity, "converges" to the standard Wiener process $(W(t))_{t\in\mathbb{R}+}$.

Our aim below is to provide a framework that makes the above statements mathematically rigorous. In particular the construction of continuous-time stochastic processes, valued in metric spaces, from their finite dimensional distributions; and later the convergence of stochastic processes on the relevant metric spaces.

2 Stochastic Processes

Stochastic processes generalize the notion of (finite-dimensional) vectors of random variables to the case of any family of random variables indexed in a general set T. Typically, the latter represents "time" and is an interval of $\mathbb{R}$ (in the continuous case) or $\mathbb{N}$ (in the discrete case).

Definition 1. Let $(\Omega, \mathcal{F}, P)$ be a probability space, T an index set, and (E, B) a measurable space. An $(E, \mathcal{B})$-valued *stochastic process* on $(\Omega, \mathcal{F}, P)$ is a family $(X_t)_{t\in T}$ of random variables $X_t : (\Omega, \mathcal{F}) \to (E, \mathcal{B})$ for $t \in T$.

The triple $(\Omega, \mathcal{F}, P)$ is called the underlying *probability space* of the process $(X_t)_{t\in T}$, while $(E, \mathcal{B})$ is known as the *state space* or *phase space*. For $t \in T$, the random variable X_t is the *state of the process at "time" t*. Moreover, for all $\omega \in \Omega$, the mapping $X(\cdot, \omega) : t \in T \to X_t(\omega) \in E$ is called the *trajectory* or *path of the process* corresponding to ω. Any trajectory $X(\cdot, \omega)$ of the process belongs to the space E^T of functions defined in T and valued in E.

In order to introduce a suitable probability space $(E^T, \mathcal{B}^T, P^T)$ for the trajectories, we consider the σ-algebra $\mathcal{B}^T$ on E^T defined as the σ-algebra generated by the algebra of the set of cylinders with finite-dimensional base, i.e.

$$\mathcal{B}^T = \sigma\left(\pi_{ST}^{-1}(A) :\ A \in \mathcal{B}^S = \bigotimes_{t\in S} \mathcal{B}_t, \mathcal{S} = \{S \subset T |\ \text{S is finite}\}\right),$$

where π_{ST} is the canonical projection of E^T on E^S [8, 23].

By the Carathéodory's extension theorem [33], it is possible to extend on $\mathcal{B}^T$, in a unique way, a probability measure defined on an algebra which generates $\mathcal{B}^T$, i.e. on the family of cylinders with the finite rectangles $B_1 \times \cdots \times B_n \in \mathcal{B}^n$ as bases, by

$$P^T(\pi_{ST}^{-1}(B_1 \times \cdots \times B_n)) = P^S(B_1 \times \cdots \times B_n) = P(X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n).$$

As a consequence $\pi_{SS'}(P^{S'}) = P^S$, $S \subset S'$ and then $\left(P^S\right)_{S\in\mathcal{S}}$ is a *compatible system* of measures, as defined below.

Definition 2. If, for all $(S, S') \in \mathcal{S} \times \mathcal{S}'$, with $S \subset S'$, we have that $\pi_{SS'}(P^{S'}) = P^S$, then $(E^S, \mathcal{B}^S, P^S, \pi_{SS'})_{S,S'\in\mathcal{S};S\subset S'})$ is called a *projective system* of measurable spaces and $(P^S)_{S\in\mathcal{S}}$ is called a *compatible system* of measures on the finite products $(E^S, \mathcal{B}^S)_{S\in\mathcal{S}}$.

We may mention this fundamental result, which characterizes the probability measures P^T [8, 23].

Theorem 1 (Kolmogorov–Bochner). *Let $(E_t, \mathcal{B}_t)_{t\in T}$ be a family of Polish spaces (i.e., metric, complete, separable) endowed with their respective Borel σ-algebras, and let $\mathcal{S}$ be the collection of finite subsets of T and, for all $S \in \mathcal{S}$ with $W^S = \prod_{t\in S} E_t$ and $\mathcal{B}^S = \bigotimes_{t\in S} \mathcal{B}_t$, let μ_S be a finite measure on $(W^S, \mathcal{B}^S)$. Under these assumptions the following two statements are equivalent:*

1. *there exists a μ_T measure on $(W^T, \mathcal{B}^T)$ such that for all $S \in \mathcal{S}$: $P^S = \pi_{ST}(P^T)$;*
2. *the system $(W^S, \mathcal{B}^S, P^S, \pi_{SS'})_{S,S'\in\mathcal{S};S\subset S'}$ is projective.*

Moreover, in both cases, P^T, as defined in 1, is unique.

The unique measure P^T of Theorem 1 is called the *projective limit* of the projective system $(W^S, \mathcal{B}^S, P^S, \pi_{SS'})_{S,S'\in\mathcal{S};S\subset S'}$.

Example 1. Let $(X_t)_{t\in T}$ be a family of independent random variables defined on $(\Omega, \mathcal{F}, P)$ and valued in $(E, \mathcal{B})$. (In fact, in this case, it is sufficient to assume that all finite families of $(X_t)_{t\in T}$ are independent.) We know that for all $t \in T$ the probability $P_t = X_t(P)$ is defined on $(E, \mathcal{B})$. Then

$$\forall S = \{t_1, \ldots, t_r\} \in \mathcal{S}: \qquad P^S = \bigotimes_{k=1}^{r} P_{t_k}, \text{ for some } r \in \mathbb{N}^*,$$

and the system $(P^S)_{S\in\mathcal{S}}$ is compatible with the finite products $(E^S, \mathcal{B}^S)_{S\in\mathcal{S}}$. In fact, if B is a rectangle in $\mathcal{B}^S$, i.e., $B = B_{t_1} \times \cdots \times B_{t_r}$, and if $S \subset S'$, where $S, S' \in \mathcal{S}$, then

$$\begin{aligned} P^S(B) &= P^S(B_{t_1} \times \cdots \times B_{t_r}) = P_{t_1}(B_{t_1}) \cdot \cdots \cdot P_{t_r}(B_{t_r}) \\ &= P_{t_1}(B_{t_1}) \cdot \cdots \cdot P_{t_r}(B_{t_r}) P_{t_{r+1}}(E) \cdot \cdots \cdot P_{t_{r'}}(E) \\ &= P^{S'}(\pi_{SS'}^{-1}(B)); \end{aligned}$$

i.e. that $P^S = \pi_{SS'}(P^{S'})$.

Definition 3. A real-valued stochastic process $(X_t)_{t\in\mathbb{R}_+}$ is *continuous in probability* if

$$P - \lim_{s\to t} X_s = X_t, \qquad s, t \in \mathbb{R}_+.$$

Definition 4. A *filtration* $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is an increasing family of sub-algebras of $\mathcal{F}$. The filtration $\mathcal{F}_t = \sigma(X(s), 0 \le s \le t)$, $t \in \mathbb{R}_+$ is called the *generated* or *natural* filtration of the process X_t.

We may refer to the natural filtration of the process X_t, also as the history of the process.

Definition 5. A stochastic process $(X_t)_{t\in\mathbb{R}_+}$ is *right-(left-)continuous* if its trajectories are right-(left-)continuous almost surely.

Definition 6. A stochastic process $(X_t)_{t\in\mathbb{R}_+}$ is said to be *right-continuous with left limits* (RCLL) or *continu à droite avec limite à gauche* (càdlàg) if, almost surely, it has trajectories that are RCLL. The latter is denoted $X_{t^-} = \lim_{s\uparrow t} X_s$.

Definition 7. A filtered complete probability space $(\Omega, \mathcal{F}, P, (\mathcal{F}_t)_{t\in\mathbb{R}_+})$ is said to satisfy the *usual hypotheses* if

1. $\mathcal{F}_0$ contains all the P-null sets of $\mathcal{F}$,
2. $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$, for all $t \in \mathbb{R}_+$; i.e., the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is right-continuous.

Henceforth we will always assume that the usual hypotheses hold, unless specified otherwise.

Definition 8. The process $(X_t)_{t\in\mathbb{R}_+}$ is said to be *progressively measurable* with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, if, for all $t \in \mathbb{R}_+$, the mapping $(s,\omega) \in [0,t] \times \Omega \to X(s,\omega) \in E$ is $(\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t)$-measurable.

Definition 9. A random variable T defined on Ω (endowed with the σ-algebra $\mathcal{F}$) and valued in $\bar{\mathbb{R}}_+$ is called a *stopping time* (or *Markov time*) with respect to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, or simply an $\mathcal{F}_t$*-stopping time*, if

$$\forall t \in \mathbb{R}_+ : \qquad \{\omega \in \Omega | T(\omega) \le t\} \in \mathcal{F}_t.$$

The stopping time is said to be finite if $P(T = \infty) = 0$.

In the following, we consider some very important stochastic processes, which will be used in the main part of these lecture notes.

2.1 Processes with Independent Increments

The stochastic process $(\Omega, \mathcal{F}, P, (X_t)_{t\in\mathbb{R}_+})$, with state space $(E, \mathcal{B})$, is called a *process with independent increments* if, for all $n \in \mathbb{N}$ and for all $(t_1, \ldots, t_n) \in \mathbb{R}_+^n$, where $t_1 < \cdots < t_n$, the random variables $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent.

Let us call $\mu_{t,s} = P_{X_t - X_s}$, the law of the increment $X_t - X_s$. It is possible to construct a compatible system of probability laws $(P^S)_{S\in\mathcal{S}}$, where $\mathcal{S} = \{t_1, \ldots, t_n\}$ is a collection of finite subsets of the index set by

$$\begin{aligned} P^S(B) &= P((X_{t_1}, \ldots, X_{t_n}) \in B) = E[I_B(X_{t_1}, \ldots, X_{t_n})] \\ &= \int I_B(y_0 + y_1, \ldots, y_0 + \cdots + y_n) \\ &\qquad\qquad \mu_0 \otimes \mu_{0,t_1} \otimes \cdots \otimes \mu_{t_{n-1},t_n}(dy_0, \ldots, dy_n). \end{aligned}$$

A process with independent increments is called *time-homogeneous* if

$$\mu_{s,t} = \mu_{s+h,t+h} \qquad \forall s, t, h \in \mathbb{R}_+, s < t. \tag{1}$$

If $(\Omega, \mathcal{F}, P, (X_t)_{t\in\mathbb{R}_+})$ is a homogeneous process with independent increments, then in particular we have

$$\mu_{s,t} = \mu_{0,t-s}, \qquad \forall s, t \in \mathbb{R}_+, s < t.$$

2.2 Martingales

Let $(X_t)_{t\in\mathbb{R}_+}$ be a real-valued family of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$ and let $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ be a filtration. The stochastic process $(X_t)_{t\in\mathbb{R}_+}$ is said to be *adapted* to the family $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if, for all $t \in \mathbb{R}_+$, X_t is $\mathcal{F}_t$-measurable.

The stochastic process $(X_t)_{t\in\mathbb{R}_+}$, adapted to the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, is a *martingale* with respect to this filtration, provided the following conditions hold:

1. X_t is P-integrable, for all $t \in \mathbb{R}_+$;
2. for all $(s,t) \in \mathbb{R}_+ \times \mathbb{R}_+, s < t : E[X_t|\mathcal{F}_s] = X_s$ almost surely.

$(X_t)_{t\in\mathbb{R}_+}$ is said to be a *submartingale* (*supermartingale*) with respect to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if, in addition to condition 1 and instead of condition 2, we have:

2′. for all $(s,t) \in \mathbb{R}_+ \times \mathbb{R}_+, s < t : E[X_t|\mathcal{F}_s] \geq X_s$ $(E[X_t|\mathcal{F}_s] \leq X_s)$ almost surely.

Remark 1. When the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is not specified, it is understood to be the increasing σ-algebra generated by the random variables of the process $(\sigma(X_s, 0 \leq s \leq t))_{t\in\mathbb{R}_+}$ (suitable extended). In this case we can write $E[X_t|X_r, 0 \leq r \leq s]$, instead of $E[X_t|\mathcal{F}_s]$.

Example 2. The evolution of a gambler's wealth in a game of chance, the latter specified by the sequence of real-valued random variables $(X_n)_{n\in\mathbb{N}}$, can serve as a descriptive example of the above definitions. Suppose that two players flip a coin and the loser pays the winner (who guessed head or tail correctly) the amount α after every round. If $(X_n)_{n\in\mathbb{N}}$ represents the cumulative fortune of player 1, then after n throws he holds

$$X_n = \sum_{i=0}^{n} \Delta_i.$$

The random variables Δ_i (just like every flip of the coin) are independent and take values α and $-\alpha$ with probabilities p and q, respectively. Therefore, we see that

$$\begin{aligned} E[X_{n+1}|X_0,\ldots,X_n] &= E[\Delta_{n+1} + X_n|X_0,\ldots,X_n] \\ &= X_n + E[\Delta_{n+1}|X_0,\ldots,X_n]. \end{aligned}$$

Since Δ_{n+1} is independent of every $\sum_{i=0}^{k} \Delta_i, k = 0,\ldots,n$, we obtain

$$E[X_{n+1}|X_0,\ldots,X_n] = X_n + E[\Delta_{n+1}] = X_n + \alpha(p-q).$$

- If the game is fair, then $p = q$ and $(X_n)_{n\in\mathbb{N}}$ is a martingale.
- If the game is in player 1's favor, then $p > q$ and $(X_n)_{n\in\mathbb{N}}$ is a submartingale.
- If the game is to the disadvantage of player 1, then $p < q$ and $(X_n)_{n\in\mathbb{N}}$ is a supermartingale.

Example 3. Let $(X_t)_{t\in\mathbb{R}_+}$ be (for all $t \in \mathbb{R}_+$) a P-integrable stochastic process on $(\Omega, \mathcal{F}, P)$ with independent increments. Then $(X_t - E[X_t])_{t\in\mathbb{R}_+}$ is a martingale. In fact:[1]

$$E[X_t|\mathcal{F}_s] = E[X_t - X_s|\mathcal{F}_s] + E[X_s|\mathcal{F}_s], \qquad s < t,$$

and recalling that both X_s is $\mathcal{F}_s$-measurable and $(X_t - X_s)$ is independent of $\mathcal{F}_s$, we obtain that

$$E[X_t|\mathcal{F}_s] = E[X_t - X_s] + X_s = X_s, \qquad s < t.$$

Proposition 1. [8] *Let $(X_t)_{t\in\mathbb{R}_+}$ be a real-valued martingale. If the function $\phi : \mathbb{R} \to \mathbb{R}$ is both convex and measurable and such that*

$$\forall t \in \mathbb{R}_+, \qquad E[\phi(X_t)] < +\infty,$$

then $(\phi(X_t))_{t\in\mathbb{R}_+}$ is a submartingale.

Proposition 2. [8] *Let $(X_n)_{n\in\mathbb{N}\setminus\{0\}}$ be a sequence of real random variables defined on the probability space $(\Omega, \mathcal{F}, P)$, and X_n^+ the positive part of X_n.*

1. *If $(X_n)_{n\in\mathbb{N}\setminus\{0\}}$ is a submartingale, then*

$$P\left(\max_{1\le k\le n} X_k > \lambda\right) \le \frac{1}{\lambda} E[X_n^+], \qquad \lambda > 0, n \in \mathbb{N} \setminus \{0\}.$$

2. *If $(X_n)_{n\in\mathbb{N}\setminus\{0\}}$ is a martingale and if, for all $n \in \mathbb{N} \setminus \{0\}$, $X \in L^p(P)$, $p > 1$, then*

$$E\left[\left(\max_{1\le k\le n} |X_k|\right)^p\right] \le \left(\frac{p}{p-1}\right)^p E[|X_n|^p], \qquad n \in \mathbb{N} \setminus \{0\}.$$

(Points 1 and 2 are called Doob's inequalities*.)*

Corollary 1. [8] If $(X_n)_{n\in\mathbb{N}\setminus\{0\}}$ is a martingale such that $X_n \in L^p(P)$ for all $n \in \mathbb{N} \setminus \{0\}$, then

$$P\left(\max_{1\le k\le n} |X_k| > \lambda\right) \le \frac{1}{\lambda^p} E[|X_n|^p], \qquad \lambda > 0.$$

[1] For simplicity, but without loss of generality, we will assume that $E[X_t] = 0$, for all t. In the case where $E[X_t] \neq 0$, we can always define a variable $Y_t = X_t - E[X_t]$, so that $E[Y_t] = 0$. In that case $(Y_t)_{t\in\mathbb{R}_+}$ will again be a process with independent increments, so that the analysis is analogous.

Proposition 3. [8] *Let $(X_t)_{t\in\mathbb{R}_+}$ be a stochastic process on $(\Omega,\mathcal{F},P)$ valued in $\mathbb{R}$.*

1. *If $(X_t)_{t\in\mathbb{R}_+}$ is a submartingale, then*

$$P\left(\sup_{0\le s\le t} X_s > \lambda\right) \le \frac{1}{\lambda}E[X_t^+], \qquad \lambda>0, t\ge 0.$$

2. *If $(X_t)_{t\in\mathbb{R}_+}$ is a martingale such that, for all $t\ge 0$, $X_t\in L^p(P)$, $p>1$, then*

$$E\left[\sup_{0\le s\le t}|X_s|^p\right] \le \left(\frac{p}{p-1}\right)^p E[|X_t|^p].$$

Definition 10. *A stochastic process $X=(X_n,\mathcal{F}_n)$ is a* local martingale *if there is a sequence $(\tau_k)_{k\ge1}$ of stopping times such that $\tau_k\le\tau_{k+1}$ $(\mathbb{P}-a.s.)$, and $\tau_k\to\infty$ $(\mathbb{P}-a.s.)$ as $k\to\infty$, and every stopped sequence $X^{\tau_k}=(X_{\min(\tau_k,n)}I_{\{\tau_k\ge0\}},\mathcal{F}_n)$ is a martingale.*

2.3 Markov Processes

Let $(X_t)_{t\in\mathbb{R}_+}$ be a stochastic process on a probability space, valued in $(E,\mathcal{B})$ and adapted to the increasing family $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ of σ-algebras of subsets of $\mathcal{F}$. $(X_t)_{t\in\mathbb{R}_+}$ is a *Markov process* with respect to $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ if for any $B\in\mathcal{B}$, and for any $(s,t)\in\mathbb{R}_+\times\mathbb{R}_+, s<t$, the following condition is satisfied:

$$P(X_t\in B|\mathcal{F}_s) = P(X_t\in B|X_s) \quad \text{a.s.} \tag{2}$$

If, for all $t\in\mathbb{R}_+$, $\mathcal{F}_t=\sigma(X_r,0\le r\le t)$, then condition (2) becomes

$$P(X_t\in B|X_r,0\le r\le s) = P(X_t\in B|X_s) \quad \text{a.s.}$$

for all $B\in\mathcal{B}$, for all $(s,t)\in\mathbb{R}_+\times\mathbb{R}_+$, and $s<t$.

Condition (2) states that for a Markov process the future depends only on the present and not on the past history.

Theorem 2. *Every real valued stochastic process $(X_t)_{t\in\mathbb{R}_+}$ with independent increments is a Markov process.*

A Markov process $(X_t)_{t\in[t_0,T]}$ on $\mathbb{R}^d$ is well-defined by an *initial distribution* P_0, the distribution of $X(t_0)$ and by a *Markov transition distribution*, that is a non negative function $p(s,x,t,A)$, defined for $0\le s<t<\infty, x\in\mathbb{R}, A\in\mathcal{B}_{\mathbb{R}}$ such that it satisfies the following conditions

1. for all $0\le s<t<\infty$, for all $A\in\mathcal{B}_{\mathbb{R}}$, $p(s,\cdot,t,A)$ is $\mathcal{B}_{\mathbb{R}}$-measurable;
2. for all $0\le s<t<\infty$, for all $x\in\mathbb{R}$, $p(s,x,t,\cdot)$ is a probability measure on $\mathcal{B}_{\mathbb{R}}$;

3. p satisfies the Chapman–Kolmogorov equation (compatibility):

$$p(s,x,t,A) = \int_{\mathbb{R}} p(s,x,r,dy)p(r,y,t,A) \qquad \forall x \in \mathbb{R}, s < r < t. \tag{3}$$

Indeed, Theorem 3 holds [1,14], from which we can deduce that

$$p(s,x,t,A) = P(X_t \in A | X_s = x), \qquad 0 \le s < t < \infty, x \in \mathbb{R}, A \in \mathcal{B}_{\mathbb{R}}.$$

Theorem 3. *[1, 14] Let E be a Polish space endowed with the σ-algebra $\mathcal{B}_E$ of its Borel sets, P_0 a probability measure on $\mathcal{B}_E$, and $p(r,x,s,A), t_0 \le r < s \le T, x \in E, A \in \mathcal{B}_E$ a Markov transition probability function. Then there exists a unique (in the sense of equivalence) Markov process $(X_t)_{t\in[t_0,T]}$ valued in E, with P_0 as its initial distribution and p as its transition probability.*

A Markov process $(X_t)_{t\in[t_0,T]}$ is *homogeneous* if the transition probability functions $p(s,x,t,A)$ depend on t and s only through their difference $t-s$. Therefore, for all $(s,t) \in [t_0,T]^2$, $s < t$, for all $u \in [0, T-t]$, for all $A \in \mathcal{B}_{\mathbb{R}}$, and for all $x \in \mathbb{R}$:

$$p(s,x,t,A) = p(s+u,x,t+u,A) \qquad \text{a.s.}$$

Semigroups Associated with Markov Transition Probability Functions.

To a Markov transition probability function $p(s,x,t,A)$ (or with its corresponding Markov process), one may associate a semigroup of operators $\{T_{s,t}\}_{0\le s\le t\le T}$ such that for any $0 \le s < t \le T$, $T_{s,t} : C_b(\mathbb{R}) \to C_b(\mathbb{R})$, is defined by assigning, for all $f \in C_b(\mathbb{R})$,

$$(T_{s,t}f)(x) = \int_{\mathbb{R}} f(y)p(s,x,t,dy) = E\left[f(X_t)|X_s = x\right], \quad x \in \mathbb{R}, \tag{4}$$

where $C_b(\mathbb{R})$ is the space of all continuous and bounded functions on $\mathbb{R}$, endowed with the norm $\|f\| = \sup_{x\in\mathbb{R}} |f(x)| (< \infty)$. It is clear that (4) is a semigroup; indeed

i) if $s = t$, then $T_{t,t} = I$ (identity), because

$$p(s,x,s,A) = \begin{cases} 1 \text{ if } x \in A, \\ 0 \text{ if } x \notin A; \end{cases}$$

ii) moreover, $T_{s,t}T_{t,u} = T_{s,u},\ 0 \le s < t < u$. In fact, if $f \in C_b(\mathbb{R})$ and $x \in \mathbb{R}$,

$$\begin{aligned}
&(T_{s,t}(T_{t,u}f))(x) \\
&= \int_{\mathbb{R}} (T_{t,u}f)(y)p(s,x,t,dy) \\
&= \int\int_{\mathbb{R}^2} f(z)p(t,y,u,dz)p(s,x,t,dy) \\
&= \int_{\mathbb{R}} f(z) \int_{\mathbb{R}} p(t,y,u,dz)p(s,x,t,dy) \text{ (by Fubini's theorem)} \\
&= \int_{\mathbb{R}} f(z)p(s,x,u,dz) \text{ (by the Chapman–Kolmogorov equation)} \\
&= (T_{s,u}f)(x).
\end{aligned}$$

□

If $(X_t)_{t\in\mathbb{R}_+}$ is a Markov process with transition probability function p and associated semigroup $\{T_{s,t}\}$, then the operator

$$\mathcal{A}_s f = \lim_{h\downarrow 0} \frac{T_{s,s+h}f - f}{h}, \qquad s \geq 0, f \in C_b(\mathbb{R})$$

is called the *infinitesimal generator of the Markov process* $(X_t)_{t\geq 0}$. Its domain $\mathcal{D}_{\mathcal{A}_s}$ consists of all $f \in C_b(\mathbb{R})$ for which the above limit exists uniformly (and therefore in the norm of $C_b(\mathbb{R})$) [16]. We may observe that

$$(\mathcal{A}_s f)(x) = \lim_{h\downarrow 0} \frac{1}{h} \int_{\mathbb{R}} [f(y) - f(x)] p(s, x, s+h, dy).$$

The above results may be extended to more general, possibly uncountable, state spaces [2]. In particular, we will assume that E is a subset of $\mathbb{R}^d$ for $d \in \mathbb{N} \setminus \{0\}$. If we consider the time-homogeneous case, a Markov process $(X_t)_{t\in\mathbb{R}_+}$ on $(E, \mathcal{B}_E)$ will be defined in terms of a transition kernel $p(t, x, B)$ for $t \in \mathbb{R}_+$, $x \in E$, $B \in \mathcal{B}_E$, such that

$$p(h, X_t, B) = P(X_{t+h} \in B | \mathcal{F}_t) \qquad \forall t, h \in \mathbb{R}_+, B \in \mathcal{B}_E,$$

given that $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$ is the natural filtration of the process. Equivalently we may state

$$E[g(X_{t+h})|\mathcal{F}_t] = \int_E g(y) p(h, X_t, dy) \qquad \forall t, h \in \mathbb{R}_+, g \in C_b(E).$$

In this case the transition semigroup of the process is a one-parameter contraction semigroup $(T(t), t \in \mathbb{R}_+)$ on $C_b(E)$ defined by

$$(T(t)g)(x) := \int_E g(y) p(t, x, dy) = E[g(X_t)|X_0 = x], \qquad x \in E,$$

for any $g \in C_b(E)$. The infinitesimal generator will be time independent. It is defined as

$$\mathcal{A}g = \lim_{t\to 0+} \frac{1}{t}(T(t)g - g)$$

for $g \in \mathcal{D}(\mathcal{A})$, the subset of $C_b(E)$ for which the above limit exists, in $C_b(E)$, with respect to the sup norm. Given the above definitions, it is obvious that for all $g \in \mathcal{D}(\mathcal{A})$,

$$\mathcal{A}g(x) = \lim_{t\to 0+} \frac{1}{t} E[g(X_t)|X_0 = x], \qquad x \in E.$$

If $(T(t), t \in \mathbb{R}_+)$ is the contraction semigroup associated with a Markov process, it is not difficult to show that the mapping $t \to T(t)g$ is right-continuous

in $t \in \mathbb{R}_+$ provided that $g \in C_b(E)$ is such that the mapping $t \to T(t)g$ is right continuous in $t = 0$. Then, for all $g \in \mathcal{D}(\mathcal{A})$ and $t \in \mathbb{R}_+$,

$$\int_0^t T(s)gds \in \mathcal{D}(\mathcal{A})$$

and

$$T(t)g - g = \mathcal{A}\int_0^t T(s)gds = \int_0^t \mathcal{A}T(s)gds = \int_0^t T(s)\mathcal{A}gds$$

by considering Riemann integrals. The following, so-called *Dynkin's formula*, establishes a fundamental link between Markov processes and martingales [8, 15].

Theorem 4. *Assume $(X_t)_{t\in\mathbb{R}_+}$ is a Markov process on $(E, \mathcal{B}_E)$, with transition kernel $p(t, x, B)$, $t \in \mathbb{R}_+$, $x \in E$, $B \in \mathcal{B}_E$. Let $(T(t), t \in \mathbb{R}_+)$ denote its transition semigroup and $\mathcal{A}$ its infinitesimal generator. Then, for any $g \in \mathcal{D}(\mathcal{A})$, the stochastic process*

$$M(t) := g(X_t) - g(X_0) - \int_0^t \mathcal{A}g(X_s)ds$$

is an $\mathcal{F}_t$-martingale.

The next proposition shows that a Markov process is indeed characterized by its infinitesimal generator via a martingale problem [8, 15, 32].

Theorem 5. (Martingale problem for Markov processes). *If an RCLL Markov process $(X_t)_{t\in\mathbb{R}_+}$ is such that*

$$g(X_t) - g(X_0) - \int_0^t \mathcal{A}g(X_s)ds$$

is an $\mathcal{F}_t$-martingale for any function $g \in \mathcal{D}(\mathcal{A})$, where $\mathcal{A}$ is the infinitesimal generator of a contraction semigroup on E, then X_t is equivalent to a Markov process having $\mathcal{A}$ as its infinitesimal generator.

Example 4. A Poisson process (see the following section for more details) is an integer-valued Markov process $(N_t)_{t\in\mathbb{R}_+}$, a so called "jump process". If its intensity parameter is $\lambda > 0$, the process $(X_t)_{t\in\mathbb{R}_+}$, defined by $X_t = N_t - \lambda t$, is a stationary Markov process with independent increments. The transition kernel of X_t is

$$p(h, x, B) = \sum_{k=0}^{\infty} \frac{(\lambda h)^k}{k!} e^{-\lambda h} I_{\{x+k-\lambda h\in B\}} \text{ for } x \in \mathbb{N}, h \in \mathbb{R}_+, B \subset \mathbb{N}.$$

Its transition semigroup is then

$$T(h)g(x) = \sum_{k=0}^{\infty} \frac{(\lambda h)^k}{k!} e^{-\lambda h} g(x + k - \lambda h) \quad \text{for } x \in \mathbb{N}, g \in C_b(\mathbb{R}).$$

The infinitesimal generator is then

$$\mathcal{A}g(x) = \lambda(g(x+1) - g(x)) - \lambda g'(x+).$$

According to previous theorems,

$$M(t) = g(X_t) - \int_0^t ds(\lambda(g(X_s + 1) - g(X_s)) - \lambda g'(X_s+)$$

is a martingale for any $g \in C_b(\mathbb{R})$, such that $g(0) = 0$.

Examples of Markov Processes

Markov Diffusion Processes

A very important class of Markov processes is the one of the *diffusion processes*, that are Markov processes on $\mathbb{R}$ with transition probability function $p(s, x, t, A)$ which satisfies the following properties

1. for all $\epsilon > 0$, for all $t \geq 0$, and for all $x \in \mathbb{R}$

$$\lim_{h \downarrow 0} \frac{1}{h} \int_{|x-y|>\epsilon} p(t, x, t + h, dy) = 0;$$

2. there exist $a(t, x)$ and $b(t, x)$ such that, for all $\epsilon > 0$, for all $t \geq 0$, and for all $x \in \mathbb{R}$,

$$\lim_{h \downarrow 0} \frac{1}{h} \int_{|x-y|<\epsilon} (y - x) p(t, x, t + h, dy) = a(t, x),$$
$$\lim_{h \downarrow 0} \frac{1}{h} \int_{|x-y|<\epsilon} (y - x)^2 p(t, x, t + h, dy) = b(t, x).$$

$a(t, x)$ is the *drift coefficient* and $b(t, x)$ the *diffusion coefficient* of the process.

Proposition 4. *If $(X_t)_{t \in \mathbb{R}_+}$ is a diffusion process with transition probability function p and drift and diffusion coefficients $a(x, t)$ and $b(x, t)$, respectively, and if $\mathcal{A}_s$ is the infinitesimal generator associated with p, then we have that*

$$(\mathcal{A}_s f)(x) = \frac{\partial f}{\partial x} a(s, x) + \frac{1}{2} \frac{\partial^2 f}{\partial x^2} b(s, x), \tag{5}$$

provided that f is bounded and twice continuously differentiable.

Markov Jump Processes

Consider a Markov process $(X_t)_{t\in\mathbb{R}_+}$ valued in a countable set E (say, $\mathbb{N}$ or $\mathbb{Z}$). In such a case it is sufficient (with respect to Theorem 3) to provide the so-called *one-point* transition probability function

$$p_{ij}(s,t) := p(s,i,t,j) := P(X_t = j | X_s = i)$$

for $t_0 \le s < t$, $i,j \in E$. It follows from the general structure of Markov processes that the one-point transition probabilities satisfy the following relations:

(a) $p_{ij}(s,t) \ge 0$,
(b) $\sum_{j\in E} p_{ij}(s,t) = 1$,
(c) $p_{ij}(s,t) = \sum_{k\in E} p_{ik}(s,r)p_{kj}(r,t)$,

provided $t_0 \le s \le r \le t$, in $\mathbb{R}_+$, and $i,j \in E$. To these three conditions we need to add

(d)
$$\lim_{t\to s+} p_{ij}(s,t) = p_{ij}(s,s) = \delta_{ij} = \begin{cases} 1 & \text{for } i = j, \\ 0 & \text{for } i \ne j. \end{cases}$$

The time-homogeneous case gives transition probabilities $(\tilde{p}_{ij}(t))_{t\in\mathbb{R}_+}$, such that

$$p_{ij}(s,t) = \tilde{p}_{ij}(t-s), \qquad s \le t.$$

From now on we shall limit our analysis to the time-homogeneous case, whose transition probabilities will be denoted $(p_{ij}(t))_{t\in\mathbb{R}_+}$. We may obtain the following result [8, 17, 32]

Theorem 6. *The limits*

$$\lim_{t\to 0+} \frac{p_{ij}(t)}{t} = p'_{ij}(0) =: q_{ij} < +\infty$$

always exist (finite) for any $i \ne j$. The limits

$$q_{ii} = -\lim_{h\to 0+} \frac{1 - p_{ii}(h)}{h} \le +\infty$$

always exists (finite or not).

Consider the family of matrices $(P(t))_{t\in\mathbb{R}_+}$, with entries $(p_{ij}(t))_{t\in\mathbb{R}_+}$, for $i,j \in E$. We may rewrite conditions (c) and (d) in matrix form as follows:

(c′) $P(s+t) = P(s)P(t)$ for any $s,t \ge 0$;
(d′) $\lim_{h\to 0+} P(h) = P(0) = I$.

A family of stochastic matrices fulfilling conditions (c′) and (d′) is called a *matrix transition function*. The matrix $Q = (q_{ij})_{i,j\in E}$ is called the *intensity matrix*. The transition and the intensity matrices satisfy the following system, which is the matrix form of the Kolmogorov backward equations

$$P'(t) = QP(t), \qquad t > 0,$$

subject to

$$P(0) = I.$$

If Q is a finite-dimensional matrix, the function $exp\{tQ\}$ for $t > 0$ is well defined.

Theorem 7. *[8, 20] If E is finite, the matrix transition function can be represented in terms of its intensity matrix Q via*

$$P(t) = e^{tQ}, \qquad t \geq 0.$$

Consider a time-homogeneous Markov jump process on a countable state space E with intensity matrix $Q = (q_{ij})_{i,j\in E}$. The matrix Q can be seen as a functional operator on E as follows: For any $f : E \to \mathbb{R}_+$ define

$$Q : f \to Q(f) = \sum_{j\in E} q_{ij} f(j) = \sum_{j\neq i} q_{ij}(f(j) - f(i)).$$

From [32] we obtain the following theorem.

Theorem 8. *For any function $g \in C^{1,0}(\mathbb{R}_+ \times E)$ such that the mapping*

$$t \to \frac{\partial}{\partial t} g(t, x)$$

is continuous for all $x \in E$, the process

$$\left(g(t, X(t)) - g(0, X(0)) - \int_0^t \left(\frac{\partial g}{\partial t} + Q(g(s, \cdot)) \right)(s, X(s)) ds \right)_{t\in\mathbb{R}_+}$$

is a local martingale.

2.4 Brownian Motion and the Wiener Process

A real-valued process $(W_t)_{t\in\mathbb{R}_+}$ is a *Wiener process* if it satisfies the following conditions:

1. $W_0 = 0$ almost surely;
2. $(W_t)_{t\in\mathbb{R}_+}$ is a process with independent increments;
3. $W_t - W_s$ is normally distributed with $N(0, t - s)$, $(0 \leq s < t)$.

From the definition it turns easily out that the Wiener process is both a Markov process and, since it has independent increments, a martingale. Furthermore, by simple calculations, one can easily show that $E[W_t] = 0$ and $Cov[W_t, W_s] = \min\{s, t\}$, $s, t \in \mathbb{R}_+$.

Now we deal with almost sure properties of the Brownian sample path, in particular its continuity and nowhere differentiability [8].

Theorem 9. *Let $(W_t)_{t\in\mathbb{R}_+}$ be a real-valued Wiener process. Then*

1. *it has continuous trajectories almost surely;*
2. *$P(\sup_{t\in\mathbb{R}_+} W_t = +\infty) = 1$,and $P(\inf_{t\in\mathbb{R}_+} W_t = -\infty) = 1$.*
3. *for all $h > 0$, $P\left(\max_{0\le s\le h} W_s > 0\right) = P\left(\min_{0\le s\le h} W_s < 0\right) = 1$; moreover, for almost every $\omega \in \Omega$ the process $(W_t)_{t\in\mathbb{R}_+}$ has a zero (i.e., crosses the spatial axis) in $]0, h]$, for all $h > 0$;*
4. *almost every trajectory of $(W_t)_{t\in\mathbb{R}_+}$ is differentiable almost nowhere.*

The real-valued process $(W_1(t), \ldots, W_n(t))'_{t\ge 0}$ is said to be an *n- dimensional Wiener process (or Brownian motion)* if:

1. for all $i \in \{1, \ldots, n\}$, $(W_i(t))_{t\ge 0}$ is a Wiener process,
2. the processes $(W_i(t))_{t\ge 0}$, $i = 1, \ldots, n$, are independent

(thus the σ-algebras $\sigma(W_i(t), t \ge 0)$, $i = 1, \ldots, n$, are independent).

Proposition 5. If $(W_1(t), \ldots, W_n(t))'_{t\ge 0}$ is an n-dimensional Brownian motion, then it can be shown that:

1. $(W_1(0), \ldots, W_n(0)) = (0, \ldots, 0)$ almost surely;
2. $(W_1(t), \ldots, W_n(t))'_{t\ge 0}$ has independent increments;
3. $(W_1(t), \ldots, W_n(t))' - (W_1(s), \ldots, W_n(s))', 0 \le s < t$, has a multivariate normal distribution $N(\mathbf{0}, (t - s)I)$ (where $\mathbf{0}$ is the null-vector of order n and I is the $n \times n$ identity matrix).

3 Itô Calculus

3.1 The Itô Integral

In classical calculus the primary components were the use of differentiation to describe rates of change, the use of integration and then the fundamental theorem of calculus. From them, the concept of ordinary differential equations came out.

In stochastic calculus, the history turn the other way round: in order to make meaningful the ordinary differential equations involving continuous stochastic processes and in particular the Brownian motion which is nowhere differentiable, it was necessary to introduce first the stochastic integral and later the stochastic differential.

For example, let us consider the classical exponential growth model

$$\frac{dN(t)}{dt} = r(t)N(t)$$

and perturb the growth rate $r(t)$ by a some stochastic noise $\theta(t)$

$$\frac{dN(t)}{dt} = [r_0 + r_1\theta(t)]N(t),$$

so that the variation follows the following equation

$$\frac{dN(t)}{N(t)} = r_0 dt + r_1\theta(t)dt.$$

If we consider a Brownian model, i.e. $\theta(t)dt \simeq dW_t$, such that $\Delta W_t \sim \mathcal{N}(0, \Delta t)$, we get

$$dN_t = r_0 N_t dt + r_1 N_t dW_t.$$

More in general

$$dN(t) = a(t, N(t))dt + b(t, N(t))dW_t, \tag{6}$$

which, however, in the current form does not make sense, because the trajectories of $(W_t)_{t\geq 0}$ are not differentiable. Instead, we will try to interpret it in the form

$$\forall \omega \in \Omega: \qquad u(\omega, t) - u(\omega, 0) = \int_0^t a(s, u(\omega, s))ds + \int_0^t b(s, u(\omega, s))dW_s,$$

which requires us to give meaning to an integral $\int_a^b f(t)dW_t$ that is not of Lebesgue–Stieltjes neither of Riemann–Stieltjes type [8].

Definition 11. Let $(W_t)_{t\geq 0}$ be a Wiener process defined on the probability space $(\Omega, \mathcal{F}, P)$ and $\mathcal{C}$ the set of functions $f(t, \omega) : [a, b] \times \Omega \to \mathbb{R}$ satisfying the following conditions:

1. f is $\mathcal{B}_{[a,b]} \otimes \mathcal{F}$-measurable;
2. for all $t \in [a, b]$, $f(t, \cdot) : \Omega \to \mathbb{R}$ is $\mathcal{F}_t$-measurable, where $\mathcal{F}_t = \sigma(W_s, 0 \leq s \leq t)$;
3. for all $f \in L^2([a, b] \times \Omega)$ and

$$\int_a^b E\left[|f(t)|^2\right]^2 dt < \infty. \tag{7}$$

.

Condition 2 of Definition 11 stresses the non-anticipatory nature of f through the fact that it only depends on the present and the past history of the Brownian motion, but not on its future.

Let $f \in \mathcal{C}$ and consider a sequence $(\pi_n)_{n\in\mathbb{N}}$ of the partitions $\pi_n : a = t_0^{(n)} < t_1^{(n)} < \cdots < t_n^{(n)} = b$ of the interval $[a, b]$ such that

$$|\pi_n| = \sup_{k\in\{0,\ldots,n\}} \left| t_{k+1}^{(n)} - t_k^{(n)} \right| \xrightarrow{n} 0.$$

If for every $\omega \in \Omega$ $f(t, \cdot)$ is continuous, we define the *(stochastic) Itô integral of the process* f as follows

$$P - \lim_{n\to\infty} \sum_{k=0}^{n-1} f\left(t_k^{(n)}\right) \left(W_{t_{k+1}^{(n)}} - W_{t_k^{(n)}} \right) = \int_a^b f(t) dW_t. \tag{8}$$

For a more general definition (the case f non continuous everywhere), please refer to [8].

Note that for the classical Lebesgue integral results in $\int_a^b t dt = \frac{b^2-a^2}{2}$. In the stochastic Itô integral we obtain $\int_a^b W_t dW_t = \frac{1}{2}(W_b^2 - W_a^2) - \frac{b-a}{2}$, i.e. we have an additional term $(-\frac{b-a}{2})$.

An important property of the stochastic integrals is the martingality. Indeed one can prove [8] that $(X_t = \int_a^t f(s)dW_s)_{t\in[a,b]}$ is a zero mean $\mathcal{L}^2$-martingale with respect to $\mathcal{F}_t = \sigma(W_s, 0 \le s \le t)$.

The *martingale representation theorem* establishes the relationship between a martingale and the existence of a process, vice versa. Let $(M_t)_{t\in[0,T]}$ be an $\mathcal{L}^2$ martingale with respect to the Wiener process $(W_t)_{t\in[0,T]}$ and $(\mathcal{F}_t)_{t\in[0,T]}$ its natural filtration. Then there exists a unique process $(f_t)_{t\in[0,T]} \in \mathcal{C}([0, T])$ so that

$$\forall t \in [0, T], \qquad M(t) = M(0) + \int_0^t f(s)dW_s \qquad \text{a.s.} \tag{9}$$

holds.

Let $\mathcal{C}_1$ be the set of functions $f : [a, b] \times \Omega \to \mathbb{R}$ such that the conditions 1 and 2 of the characterization of the class $\mathcal{C}$ are satisfied, but, instead of condition 3, we have

$$P\left(\int_a^b |f(t)|^2 dt < \infty \right) = 1. \tag{10}$$

It is obvious that $\mathcal{C} \subset \mathcal{C}_1$ and thus $\mathcal{S} \subset \mathcal{C}_1$. It is possible to define the stochastic integral for a function $f \in \mathcal{C}_1$ as in (8).

3.2 The Stochastic Differential

Let $(u(t))_{0\le t\le T}$ be a process such that for every $(t_1, t_2) \in [0, T]\times[0, T], t_1 < t_2$:

$$u(t_2) - u(t_1) = \int_{t_1}^{t_2} a(t)dt + \int_{t_1}^{t_2} b(t)dW_t, \tag{11}$$

where $a \in \mathcal{C}_1([0,T])$ and $b \in \mathcal{C}_1([0,T])$. Then $u(t)$ is said to have the *stochastic differential*

$$du(t) = a(t)dt + b(t)dW_t \tag{12}$$

on $[0,T]$. Hence

1. the trajectories of $(u(t))_{0\le t\le T}$ are continuous almost everywhere;
2. for $t \in [0,T]$, $u(t)$ is $\mathcal{F}_t = \sigma(W_s, 0 \le s \le t)$-measurable, thus $u(t) \in \mathcal{C}_1([0,T])$.

Itô's Formula

As one of the most important topics on Brownian motion, Itô's formula represents the stochastic equivalent of Taylor's theorem about the expansion of functions. It is the key concept that connects classical and stochastic theory.

If $du(t) = a(t)dt + b(t)dW_t$ and if $f(t,x) : [0,T] \times \mathbb{R} \to \mathbb{R}$ is continuous with the derivatives f_x, f_{xx}, and f_t, then the stochastic differential of the process $f(t,u(t))$ is given by

$$\begin{aligned} df(t,u(t)) = &\left(f_t(t,u(t)) + \frac{1}{2} f_{xx}(t,u(t))b^2(t) + f_x(t,u(t))a(t) \right) dt \\ &+ f_x(t,u(t))b(t)dW_t. \end{aligned} \tag{13}$$

Examples:

1. $dW_t^2 = dt + 2W_t dW_t$;
2. $d(tW_t) = W_t dt + t dW_t$;
3. If $f \in C^2(\mathbb{R})$, then $df(W_t) = f'(W_t)dW_t + \frac{1}{2} f''(W_t)dt$.
4. If $u(t,x) : [0,T] \times \mathbb{R} \to \mathbb{R}$ is continuous with the derivatives u_x, u_{xx}, and u_t, then

$$du(t,W_t) = \left(u_t(t,W_t) + \frac{1}{2} u_{xx}(t,W_t) \right) dt + u_x(t,W_t)dW_t. \tag{14}$$

3.3 Stochastic Differential Equations

Let $(W_t)_{t\in\mathbb{R}_+}$ be a Wiener process on the probability space $(\Omega, \mathcal{F}, P)$, equipped with the filtration $(\mathcal{F}_t)_{t\in\mathbb{R}_+}$, $\mathcal{F}_t = \sigma(W_s, 0 \le s \le t)$. Furthermore, let $a(t,x)$, $b(t,x)$ be measurable functions in $[0,T] \times \mathbb{R}$ and $(u(t))_{t\in[0,T]}$ a stochastic process. Now $u(t)$ is said to be the *solution of the stochastic differential equation*

$$du(t) = a(t,u(t))dt + b(t,u(t))dW_t, \tag{15}$$

with the initial condition

$$u(0) = u^0 \text{ a.s. } (u^0 \text{ a random variable}), \tag{16}$$

if

1. $u(0)$ is $\mathcal{F}_0$-measurable;
2. $|a(t,u(t))|^{\frac{1}{2}}, b(t,u(t)) \in \mathcal{C}_1([0,T])$;
3. $u(t)$ is differentiable and $du(t) = a(t,u(t))dt + b(t,u(t))dW_t$, thus $u(t) = u(0) + \int_0^t a(s,u(s))ds + \int_0^t b(s,u(s))dW_s$, $t \in]0,T]$.

Theorem 10 (Existence and Uniqueness of the Solution). *Resorting to the notation of the preceding definition, if the following conditions are satisfied:*

1. *for all* $t \in [0,T]$ *and all* $(x,y) \in \mathbb{R}\times\mathbb{R}$: $|a(t,x)-a(t,y)|+|b(t,x)-b(t,y)| \leq K^*|x-y|$;
2. *for all* $t \in [0,T]$ *and all* $x \in \mathbb{R}$: $|a(t,x)| \leq K(1+|x|), |b(t,x)| \leq K(1+|x|)$ *(K^*, K constants);*
3. $E[|u^0|^2] < \infty$;
4. u^0 *is independent of* $\mathcal{F}_T$ *(which is equivalent to requiring* u^0 *to be* $\mathcal{F}_0$*-measurable),*

then there exists a unique $(u(t))_{t\in[0,T]}$*, solution of (15), (16), such that*

a. $(u(t))_{t\in[0,T]}$ *is continuous almost surely (thus almost every trajectory is continuous);*
b. $(u(t))_{t\in[0,T]} \in \mathcal{C}([0,T])$.

Remark 2. If $(u_1(t))_{t\in[0,T]}$ and $(u_2(t))_{t\in[0,T]}$ are two solutions of (15), (16), belonging to $\mathcal{C}([0,T])$, then the uniqueness of a solution is understood in the sense that

$$P\left(\sup_{0\leq t\leq T}|u_1(t)-u_2(t)| = 0\right) = 1.$$

Examples:

1. $$\begin{cases} u_0(t) = u^0, \\ du(t) = g(t)u(t)dW_t \end{cases}$$

   has the solution

   $$u(t) = u^0 \exp\left\{\int_0^t g(s)dW_s - \frac{1}{2}\int_0^t g^2(s)ds\right\}.$$

2. geometric Brownian motion

   $$\begin{cases} u_0(t) = u^0, \\ du(t) = au(t)dt + bu(t)dW_t; \end{cases}$$

   has the solution

   $$u(t) = u^0 \exp\left\{\left(a - \frac{1}{2}b^2\right)(t-t_0) + b(W_t - W_{t_0})\right\}.$$

3. (mean-reverting) Ornstein–Uhlenbeck process

$$\begin{cases} u_0(t) = u^0, \\ du(t) = (a - bu(t))dt + cdW_t. \end{cases}$$

has the solution

$$u(t) = \frac{a}{b}\exp\{bt_0\} + u^0\exp\{-b(t-t_0)\} + c\int_{t_0}^{t}\exp\{-b(t-s)\}dW_s.$$

Property of Solutions

Now, let $t_0 \geq 0$ and c be a random variable with $u(t_0) = c$ almost surely and, moreover, c be independent of $\mathcal{F}_{t_0,T} = \sigma(W_t - W_{t_0}, t_0 \leq t \leq T)$ as well as $E[c^2] < +\infty$. Under Conditions 1 and 2 of Theorem 10 the following properties are satisfied by the unique solution $(u(t))_{t\in[t_0,T]}$ of (15)-(16).

1. *Markov Property:* The solution $(u(t))_{t\in[t_0,T]}$ is a Markov Process. Moreover its transition probability p is such that, for all $B \in \mathcal{B}_{\mathbb{R}}$ and all $t_0 \leq s < t \leq T$ and all $x \in \mathbb{R}$

$$p(s,x,t,B) = P(u(t) \in B | u(s) = x) = P(u(t,s,x) \in B),$$

where $\{u(t,s,x), t \geq s\}$ is the unique solution of (15) subject to the initial condition $u(s) = x$.
So what we know is that every stochastic differential equation generates Markov processes in the sense that every solution is a Markov process.
2. *Diffusion Property:* If $a(t,x)$ and $b(t,x)$ are continuous functions in $(t,x) \in [0,\infty]\times\mathbb{R}$, then the solution $u(t)$ is a diffusion process with drift coefficient $a(t,x)$ and diffusion coefficient $b^2(t,x)$.

3.4 Kolmogorov and Fokker-Planck Equations

Let $u(s,t,x), t < s$ be the solution of the following stochastic differential equation

$$du(t) = a(t,u(t))dt + b(t,u(t))dW_t \tag{17}$$

$$u(s,s,x) = x \text{ a.s. } (x \in \mathbb{R}) \tag{18}$$

and suppose that the coefficients a and b satisfy the assumptions of existence and uniqueness of Theorem 10. If $f : \mathbb{R} \to \mathbb{R}$ is a twice continuously differentiable function and if there exist $C > 0$ and $m > 0$ such that

$$|f(x)| + |f'(x)| + |f''(x)| \leq C(1 + |x|^m), \qquad x \in \mathbb{R},$$

then the function

$$q(t,x) \equiv E[f(u(s,t,x))], \qquad 0 < t < s, \qquad x \in \mathbb{R}, s \in (0,T), \tag{19}$$

satisfies the equation

$$\frac{\partial}{\partial t}q(t,x) + a(t,x)\frac{\partial}{\partial x}q(t,x) + \frac{1}{2}b^2(t,x)\frac{\partial^2}{\partial x^2}q(t,x) = 0, \tag{20}$$

with the boundary condition

$$\lim_{t\uparrow s} q(t,x) = f(x). \tag{21}$$

Equation (20) is called *backward Kolmogorov's differential equation.*

From the above we can give the solution of a Cauchy problem via an average of realizations of a Markov process.

Proposition 6. *Consider the Cauchy problem:*

$$\begin{cases} L_0[q] + \frac{\partial q}{\partial t} = 0 \quad \text{in } [0,T)\times\mathbb{R}, \\ \lim\limits_{t\uparrow T} q(t,x) = \phi(x) \text{ in } \mathbb{R}, \end{cases} \tag{22}$$

where $L_0[\cdot] = \frac{1}{2}b^2(t,x)\frac{\partial^2}{\partial x^2} + a(t,x)\frac{\partial}{\partial x}$, *and suppose that*
(A_1) $a(t,x)$ *is strictly positive for all* $(t,x)\in[0,T]\times\mathbb{R}$;
(B_1) $\phi(x)$ *is continuous in* $\mathbb{R}$, *and* $A>0, a>0$ *exist such that* $|\phi(x)|\le A(1+|x|^a)$;
(B_2) a *and* b *are bounded in* $[0,T]\times\mathbb{R}$ *and uniformly Lipschitz in* (t,x) *on compact subsets of* $[0,T]\times\mathbb{R}$;
(B_3) b *is Hölder continuous in* x *and uniform with respect to* (t,x) *on* $[0,T]\times\mathbb{R}$;
then the Cauchy problem (22) admits a unique solution $q(t,x)$ *in* $[0,T]\times\mathbb{R}$ *such that*

$$q(t,x) = E[\phi(u(T,t,x))], \tag{23}$$

where $u(T,t,x)$ *is a solution of* (17) *at time* T, *with initial location in* x *at time* t.

Something interesting is that under the conditions (A_1) and (B_1), the transition probability $p(s,x,t,A) = P(u(t,s,x)\in A)$ of the Markov process $u(t,s,x)$ (the solution of the differential equation (17)) is endowed with density $f(x,s;y,t)$, i.e.

$$p(s,x,t,A) = \int_A f(x,s;y,t)dy \qquad (s<t), \text{ for all } A\in\mathcal{B}_{\mathbb{R}}. \tag{24}$$

The density, called *transition density* of the solution $u(t)$ is the solution of the backward Kolmogorov's equation

$$\begin{cases} L_0[f] + \frac{\partial}{\partial t}f = 0, \\ \lim\limits_{t\to T} f(x,s;y,T) = \delta(x-y). \end{cases} \tag{25}$$

If one requests further regularity on the transition density, i.e. there exist continuous derivatives

$$\frac{\partial f}{\partial t}(s,x,t,y), \qquad \frac{\partial}{\partial y}(a(t,y)f(s,x,t,y)), \qquad \frac{\partial^2}{\partial y^2}(b(t,y)f(s,x,t,y)),$$

then $f(s,x,t,y)$, as a function of t and y, satisfies the equation

$$\frac{\partial f}{\partial t}(s,x,t,y) + \frac{\partial}{\partial y}(a(t,y)f(s,x,t,y)) - \frac{\partial^2}{\partial y^2}(b(t,y)f(s,x,t,y)) = 0 \qquad (26)$$

in the region $t \in (s,T]$, $y \in \mathbb{R}$. Equation (26) is known as the *forward Kolmogorov's equation* or *Fokker–Planck equation.* It is worth pointing out that while the forward equation has a more intuitive interpretation than the backward equation, the regularity conditions on the functions a and b are more stringent than those needed in the backward case. The problem of existence and uniqueness of the solution of the Fokker–Planck equation is not of an elementary nature, especially in presence of boundary conditions. This suggests that the backward approach is more convenient than the forward approach from the viewpoint of analysis.

3.5 The Multidimensional Case

If we have n independent Wiener Processes $(W_j)_{1\le j\le n}$, we may define the Itô integral for an $f : [a,b] \times \Omega \to \mathbb{R}^{m\times n}$, with each component $f_{ij} \in \mathcal{C}_{1,W_j}$, as follows

$$\int_a^b f(t)dW_t = \left[\sum_{j=1}^{n}\int_a^b f_{ij}(t)dW_j(t)\right]_{1\le i\le m}.$$

In this case the stochastic differential for an m-dimensional process $(\mathbf{u}_t)_{0\le t\le T}$ is

$$du_i(t) = a_i(t)dt + \sum_{j=1}^{n}(b_{ij}(t)dW_j(t)), \quad i = 1,\dots,m,$$

or in vector form

$$d\mathbf{u}(t) = \mathbf{a}(t)dt + b(t)d\mathbf{W}(t), \qquad (27)$$

with

$$\mathbf{a} : [0,T] \times \Omega \to \mathbb{R}^m, \mathbf{a} \in \mathcal{C}_{1\mathbf{W}}([0,T]),$$
$$b : [0,T] \times \Omega \to \mathbb{R}^{mn}, b \in \mathcal{C}_{1\mathbf{W}}([0,T]).$$

If $f(t,\mathbf{x}) : \mathbb{R}_+ \times \mathbb{R}^m \to \mathbb{R}$ is a continuous function with the derivatives $f_{x_i}, f_{x_ix_j}$, and $\mathbf{u}(t)$ is an m-dimensional process, endowed with the stochastic differential (27), then $f(t,\mathbf{u}(t))$ has the stochastic differential

$$df(t,\mathbf{u}(t)) = Lf(t,\mathbf{u}(t))dt + \nabla_{\mathbf{x}}f(t,\mathbf{u}(t)) \cdot b(t)d\mathbf{W}(t), \qquad (28)$$

where $\nabla_{\mathbf{x}} f(t, \mathbf{u}(t)) \cdot b(t) d\mathbf{W}(t)$ is the scalar product of two m-dimensional vectors, $a_{ij} = (bb')_{ij}$, $i, j = 1, \ldots, m$:

$$L = \frac{1}{2} \sum_{i,j=1}^{m} a_{ij} \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{m} a_i \frac{\partial}{\partial x_i} + \frac{\partial}{\partial t}$$

and $\nabla_{\mathbf{x}}$ is the gradient operator. Furthermore let $\mathbf{a}(t, \mathbf{x}) = (a_1(t, \mathbf{x}), \ldots, a_m(t, \mathbf{x}))'$ and $b(t, \mathbf{x}) = (b_{ij}(t, \mathbf{x}))_{i=1,\ldots,m, j=1,\ldots,n}$ be measurable functions with respect to $(t, \mathbf{x}) \in [0, T] \times \mathbb{R}^n$. An *m-dimensional stochastic differential equation* is of the form

$$d\mathbf{u}(t) = \mathbf{a}(t, \mathbf{u}(t))dt + b(t, \mathbf{u}(t))d\mathbf{W}(t), \tag{29}$$

with the initial condition

$$\mathbf{u}(0) = \mathbf{u}^0 \text{ a.s.}, \tag{30}$$

where $\mathbf{u}^0$ is a fixed m-dimensional random vector. The entire theory of the one-dimensional case translates to the multidimensional case.

4 Deterministic Approximation of Stochastic Systems

4.1 Continuous Approximation of Jump Population Processes

In Section 2.3 we have shown that a jump Markov process can be described in terms of an intensity matrix Q, made of the rates of the relevant transitions.

For a wide class of epidemic models the total population $(N_t)_{t \in \mathbb{R}_+}$ includes three subclasses: the class of susceptibles $(S_t)_{t \in \mathbb{R}_+}$, the infectives $(I_t)_{t \in \mathbb{R}_+}$, and the class of removals $(R_t)_{t \in \mathbb{R}_+}$. For a constant total population size N it is sufficient to provide a model only for the bivariate jump Markov process $(S_t, I_t)_{t \in \mathbb{R}_+}$, which is valued in $E = \mathbb{N}^2$.

For a typical model (general stochastic epidemic) the relevant elements of the intensity matrix Q are given by

- $q_{(s,i),(s,i-1)} = \delta i = \delta N \dfrac{i}{N}$, removal rate of infectives;
- $q_{(s,i),(s-1,i+1)} = \dfrac{\kappa}{N} si = N\kappa \dfrac{s}{N} \dfrac{i}{N}$, infection rate of susceptibles,

thus taking into account a rescaling with respect to the size of the total population [7].

Both transition rates are of the form $q^{(N)}_{k,k+l} = N\beta_l \left(\dfrac{k}{N} \right)$, for $k = (s, i)$, and $k + l = (s, i - 1)$, or $(s - 1, i + 1)$.

It can be shown [8,15] that we can write the evolution equation for $\hat{X}^{(N)} = (s(t), i(t))'$ as follows

$$\hat{X}^{(N)}(t) = \hat{X}^{(N)}(0) + \sum_{l \in \mathbb{Z}^d} l Y_l \left(N \int_0^t \beta_l \left(\frac{\hat{X}^{(N)}(s)}{N} \right) ds \right),$$

where the Y_l are independent standard Poisson processes.

We may rescale the process itself $\hat{X}^{(N)}(t)$ by setting

$$X_N = \frac{1}{N} \hat{X}^{(N)},$$

for which we have

$$X_N(t) = X_N(0) + \frac{1}{N} \sum_{l \in \mathbb{Z}^d} l Y_l \left(N \int_0^t \beta_l \left(X_N(s) \right) ds \right),$$

Let now

$$\begin{aligned} \tilde{Y}_l(u) &= Y_l(u) - u \\ F(x) &= \sum_{l \in \mathbb{Z}^d} l \beta_l(x), \qquad x \in \mathbb{R}^d, \end{aligned}$$

so that

$$\begin{aligned} X_N(t) = X_N(0) &+ \int_0^t F(X_N(s)) ds \\ &+ \frac{1}{N} \sum_{l \in \mathbb{Z}^d} l \, \tilde{Y}_l \left(N \int_0^t \beta_l \left(X_N(s) \right) ds \right). \end{aligned} \tag{31}$$

By Doob's inequality for martingales, we get a uniform strong law of large numbers, i.e.

$$\lim_{N \to \infty} \sup_{u \le v} \left| \frac{1}{N} \tilde{Y}_l(Nu) \right| = 0, \qquad \text{a.s.}$$

for any $v \ge 0$. This is the fundamental reason why the limit evolution for large N tends to a deterministic one. Indeed, the following theorem holds ([15] page 456).

Theorem 11. *Suppose that for each compact $K \subset E$,*

$$\sum_{l \in \mathbb{Z}^d} |l| \sup_{x \in K} \beta_l(x) < +\infty,$$

and there exists an $M_K > 0$ such that

$$|F(x) - F(y)| \le M_K |x - y|, \qquad x, y \in K;$$

suppose X_N satisfies equation (31) *above, with $\lim_{N \to \infty} X^{(N)}(0) = x_0 \in \mathbb{R}^d$.*

Then, for every $t \geq 0$,

$$\lim_{N\to\infty} \sup_{s\leq t} |X_N(s) - x(s)| = 0 \qquad a.s.,$$

where $x(t)$, $t \in \mathbb{R}_+$ *is the unique solution of*

$$x(t) = x_0 + \int_0^t F(x(s))ds, \qquad t \geq 0,$$

wherever it exists.

In differential form

$$\frac{dx}{dt}(t) = F(x(t)),$$

subject to the initial condition

$$x(0) = x_0.$$

4.2 Continuous Approximation of Stochastic Interacting Particle Systems

Suppose a population is composed of $N \in \mathbb{N} \setminus \{0\}$ individuals; the random location of the k-th individual out of N be described by a stochastic process $\{X_N^k(t)\}_{t\in\mathbb{R}_+}$ defined on a suitable probability space $(\Omega, \mathcal{F}, P)$ and valued in $\left(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d}\right)$. In another way, it may be modelled as a random Dirac-measure $\epsilon_{X_N^k(t)} \in \mathcal{M}_P(\mathbb{R}^d)$, defined as follows

$$\epsilon_{X_N^k(t)}(B) = \begin{cases} 1 \ \text{ if } X_N^k(t) \in B \\ \\ 0 \ \text{ if } X_N^k(t) \notin B \end{cases} \qquad \forall B \in \mathcal{B}_{\mathbb{R}^d}.$$

For any sufficiently smooth $f : \mathbb{R}^d \to \mathbb{R}$ the Dirac measure is such that

$$\int_{\mathbb{R}^d} f(y)\epsilon_{X_N^k(t)}(dy) = f\left(X_N^k(t)\right).$$

Correspondingly, a global description of the spatial distribution of the system of N particles at time t may be given in terms of the random measure on $\mathbb{R}^d$

$$X_N(t) = \frac{1}{N}\sum_{k=1}^{N} \epsilon_{X_N^k(t)} \in \mathcal{M}(\mathbb{R}^d), \tag{32}$$

which is known as the *empirical measure*, while the process $X_N = \{X_N(t)\}_{t\in\mathbb{R}_+}$ is called the *empirical process*.

The social evolution of the system of N individuals may be expressed via the evolution equation of each individual $k = 1, \ldots, N$,

$$dX_N^k(t) = F_N[X_N(t)](X_N^k(t))\, dt + \sigma_N\left(X_N(t)\right)\, dW^k(t). \tag{33}$$

This is an Itô type SDE in which both the drift F_N and the diffusion coefficient σ_N^2 may depend upon the empirical measure $X_N(t)$, i.e. upon the spatial distribution of the N particles. In this way the individual behavior depends upon the distribution of the system. We suppose that W^k, $k = 1, \ldots, N$, is a family of independent standard Wiener processes which are responsible of the stochastic fluctuations in the motion of each individual. The modeller will express the specific case by introducing a suitable mathematical model for F_N and σ_N^2 in terms of X_N. In order to avoid further technical difficulties, we will consider $\sigma_N^2(X_N(t)) = \sigma_N^2$, constant in time.

An evolution equation for the empirical process $(X_N(t))_{t\in\mathbb{R}_+}$ can be obtained thanks to a nice application of the Itô's formula. Given a regular function $f \in C_b^{2,1}(\mathbb{R}^d \times \mathbb{R}_+)$, we have

$$\begin{aligned} f\left(X_N^k(t), t\right) = f\left(X_N^k(0), 0\right) &+ \int_0^t F_N[X_N(s)]\left(X_N^k(s)\right)\nabla f\left(X_N^k(s), s\right) ds \\ &+ \int_0^t \left[\frac{\partial}{\partial s} f\left(X_N^k(s), s\right) + \frac{\sigma_N^2}{2}\triangle f\left(X_N^k(s), s\right)\right] ds \\ &+ \sigma_N \int_0^t \nabla f\left(X_N^k(s), s\right) dW^k(s). \end{aligned} \tag{34}$$

Correspondingly, for the empirical process $(X_N(t))_{t\in\mathbb{R}_+}$, we get the following weak formulation of its evolution equation. For any $f \in C_b^{2,1}(\mathbb{R}^d \times \mathbb{R}_+)$ we have

$$\begin{aligned} \langle X_N(t), f(\cdot, t)\rangle = \langle X_N(0), f(\cdot, 0)\rangle &+ \int_0^t \langle X_N(s), F_N[X_N(s)](\cdot)\nabla f(\cdot, s)\rangle\, ds \\ &+ \int_0^t \left\langle X_N(s), \frac{\sigma_N^2}{2}\triangle f(\cdot, s) + \frac{\partial}{\partial s} f(\cdot, s)\right\rangle ds \\ &+ \frac{\sigma_N}{N}\int_0^t \sum_k \nabla f\left(X_N^k(s), s\right) dW^k(s). \end{aligned} \tag{35}$$

In the previous expressions, we have used the notation $\langle \mu, f\rangle = \int f(x)\mu(dx)$, for any measure μ on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ and any (sufficiently smooth) function $f : \mathbb{R}^d \to \mathbb{R}$. The last term

$$M_N(f, t) \;=\; \frac{\sigma_N}{N}\int_0^t \sum_k \nabla f(X_N^k(s), s) dW^k(s);$$

is the only explicit source of stochasticity in the equation. It is a martingale with respect to the natural filtration of the process $\{X_N(t), t \in \mathbb{R}_+\}$.

Equation (35) shows how, when the number of particles N is large but still finite, also from the Eulerian point of view the system keeps the stochasticity which characterizes each individual. This is not true anymore when the

size of the system tends to infinity. The main reason is that the martingale term $M_N(f,t)$ vanishes in probability. Indeed we may apply Doob's inequality to obtain

$$E\left[\sup_{t\le T}|M_N(f,t)|\right]^2 \le E\left[\sup_{t\le T}|M_N(f,t)|^2\right] \le 4E\left[|M_N(f,T)|^2\right]$$
$$\le \frac{4\sigma_N^2}{N^2}\sum_{k=1}^{N} E\left[\int_0^T |\nabla f(X_N^k(s),s)|^2 ds\right]$$
$$\le \frac{4\sigma_N^2\|\nabla f\|_\infty^2 T}{N}. \tag{36}$$

Hence for the zero-mean martingale $M_N(f,t)$, the quadratic variation (36) vanishes in the limit $N\to\infty$. This implies convergence to zero in probability. This is the substantial reason of the deterministic limiting behavior of the process, as $N\to\infty$, since in this limit the evolution equation of the process will not contain the Brownian noise anymore.

4.3 Convergence of the Empirical Measure

The problem we want to deal with now is the convergence, for N tending to infinity, of the empirical process $X_N(t)\in\mathcal{M}_P(\mathbb{R}^d), t\ge 0$. So we need to introduce a concept of convergence in suitable spaces of probability measures.

Weak Convergence on a Metric Space (S,d)

Let (S,d) be a metric space, and let $\mathcal{S}$ be the σ-algebra of Borel subsets generated by the topology induced by d. Let $P,P_1,P_2,\ldots$ be probability measures on $(S,\mathcal{S})$.

By definition, a sequence of probability measures $\{P_n\}_{n\in\mathbb{N}}$ converges weakly to the probability measure P (notation $P_n \overset{\mathcal{W}}{\to} P$) if

$$\int_E f dP_n \to \int_E f dP$$

for every function $f\in C_b(S)$.

With respect to random variables we may state the following: a sequence (X_n) of random variables with values in the common measurable space $(S,\mathcal{S})$ converges in distribution to the random variable X,

$$X_n \overset{\mathcal{D}}{\to} X$$

if the probability laws $P_n=\mathcal{L}(X_n)$ of the X_n's weakly converge to the probability law $P=\mathcal{L}(X)$ of X

$$\mathcal{L}(X_n)\overset{\mathcal{W}}{\to}\mathcal{L}(X).$$

From the weak convergence of probability measures one can deduce also the almost sure convergence of particular random variables, as stated in the following important theorem [3].

Theorem 12 (Skorohod representation theorem). *Consider a sequence* $(P_n)_{n\in\mathbb{N}}$ *of probability measures and a probability measure* P *on a separable metric space* $(S,\mathcal{S})$*, such that* $P_n \xrightarrow[n\to\infty]{\mathcal{W}} P$*. Then there exists a sequence of* $S-$*valued random variables* $(Y_n)_{n\in\mathbb{N}}$ *and an* $S-$*valued random variable* Y *defined on a common (suitably extended) measurable space* $(\Omega,\mathcal{F})$*, such that* Y_n *has probability law* P_n*,* Y *has probability law* P*, and*

$$Y_n \xrightarrow[n\to\infty]{a.s.} Y.$$

Metrics on $\mathcal{M}_P(S)$ *induced by d*

Let us define the following distance on $\mathcal{M}_P(S)$

$$d_P(Q,P) = \inf\{\epsilon > 0 : Q(A) \le P(A^\epsilon) + \epsilon, P(A) \le Q(A^\epsilon) + \epsilon, \forall A \in \mathcal{S}\}, \quad (37)$$

where $A^\epsilon = \{d(x,A) < \epsilon\}$. The function d_P is a metric on $\mathcal{M}_P(S)$ (induced by d), called *Prokhorov's metric.*

Instead of Prokhorov's metric, in the next sections we consider the *bounded Lipschitz metric* proposed by Dudley [11], as it is easier to work with.

Denote by $Lip_b(S)$ the set of bounded and Lipschitz real function on S. For any $P,Q \in \mathcal{M}(S)$, define the *bounded Lipschitz metric* as follows

$$d_{BL}(Q,P) = \sup\left\{\left|\int f dP - \int f\, dQ\right| : f \in Lip_b(S), \|f\|_{Lip} \le 1\right\}, \quad (38)$$

where

$$\|f\|_{Lip} = \|f\|_\infty + \sup_{x\neq} \frac{f(x)-f(y)}{d(x,y)}, f \in Lip_b(S).$$

Recall that the metric space (S,d) is called *separable* if it has a countable dense subset, that is, there are $\{x_1,x_2,\dots,\}$ in S such that its closure is equal to S, $\overline{\{x_1,x_2,\dots,\}} = S$.

Both d_P-convergence and d_{BL}-convergence imply weak convergence of probability measures [12, 34]. In the case the metric space is separable one can get an equivalence.

Furthermore, if (S,d) is a separable (and complete) metric space, then so is $\mathcal{M}_P(S)$ with the induced either Prokhorov metric and bounded Lipschitz metric. Moreover in the case of S separable, a sequence in $\mathcal{M}_P(S)$ converges in the Prokhorov metric (BL-metric) if and only if it converges weakly and then in both senses to the same limit [12, 34].

Compactness Properties

In the study of the limit behavior of stochastic processes one often needs to know when a sequence of random variables is convergent in distribution or, at least, admits a subsequence that converges in distribution. This requires a good description of the sequences in $\mathcal{M}_P(S)$ that admit a convergent subsequence, i.e. of the relatively compact sets of $\mathcal{M}_P(S)$. Recall that a subset A of a metric space is called relatively compact if its closure $\bar{A}$ is compact.

First let us recall some important definitions.

A probability measure $P \in \mathcal{M}_P(S)$ is called *tight* if for every $\epsilon > 0$ there exists a compact set $K \subset S$ such that $P(X \setminus K) < \epsilon$. Of course, if (S, d) is a compact metric space, then every measure on S is tight. If (S, d) is a complete separable metric space (usually called a Polish space), then every probability measure on S is tight.

We may extend the property of tightness to a family of probability measures. A family Π of probability measures on the general metric space $(S, \mathcal{S})$ is said to be *(uniformly) tight* if, for all $\epsilon > 0$, there exists a compact set K such that

$$P(K) > 1 - \epsilon \qquad \forall P \in \Pi.$$

Π is said to be *relatively compact* if every sequence of elements of Π contains a weakly convergent subsequence; i.e., for every sequence $\{P_n\}$ in Π there exists a subsequence $\{P_{n_k}\}$ and a probability measure P (defined on $(S, \mathcal{S})$, but not necessarily an element of Π) such that $P_{n_k} \xrightarrow{\mathcal{W}} P$.

The following theorem by Prokhorov offers a useful equivalence of the relatively compactness in $\mathcal{M}_P(S)$, in case S is separable and complete.

Theorem 13. (Prohorov) *[3] Let Π be a family of probability measures on the probability space $(S, \mathcal{S})$. Then*

1. *if Π is tight, then it is relatively compact;*
2. *let S be separable and complete; if Π is relatively compact, then it is tight.*

One can characterize the weak convergence of a relatively compact sequence of probability measures [3, 15].

Theorem 14. *In a metric space $(S, \mathcal{S})$, let $\{P_n\}$ be a relatively compact sequence of probability measures, and P an additional probability measure on $\mathcal{S}$. Then the following propositions are equivalent*

a) $P_n \Rightarrow P$;
b) all weakly converging subsequences of $\{P_n\}$ weakly converge to P.

Corollary 2. *In a Polish (complete and separable) space $(S, \mathcal{S})$, let $\{P_n\}$ be a relatively compact (tight) sequence of probability measures, and P an additional probability measure on $\mathcal{S}$. Then the following propositions are equivalent*

a) $P_n \Rightarrow P$;
b) all finite dimensional subsequences of $\{P_n\}$ weakly converge to P.

The Space of Trajectories of the Empirical Process X_N

Let $(S,\mathcal{S}) = (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$; given $X_N(t) \in \mathcal{M}_P(\mathbb{R}^d)$ by (32), the empirical process $X_N = \{X_N(t)\}_{t\in[0,T]} \in C([0,T], \mathcal{M}_P(\mathbb{R}^d))$.

In $\mathcal{M}_P(\mathbb{R}^d)$, we consider the Prokhorov metric or the BL-metric, while in the space $C([0,T], \mathcal{M}_P(\mathbb{R}^d))$ we take the uniform metric with respect to $t \in [0,T]$, so that the distance between $f, g \in C([0,T], \mathcal{M}_P(\mathbb{R}^d))$ is given by

$$d_1(f,g) := \sup_{0\le t\le T} \rho(f(t), g(t)), \tag{39}$$

where ρ is either the Prokhorov or the BL metric.

From the separability and completeness of $\mathbb{R}^d$, we can deduce the same properties for the space of probability measures $\mathcal{M}_P(C([0,T], \mathcal{M}_P(\mathbb{R}^d))$. Indeed,

Theorem 15. *Since $\mathbb{R}^d$ is a Polish space, then $\mathcal{M}_P(\mathbb{R}^d)$ is a Polish space, so that $C([0,T], \mathcal{M}_P(\mathbb{R}^d))$ is a Polish space, and consequently $\mathcal{M}_P(C([0,T], \mathcal{M}(\mathbb{R}^d)))$ is a Polish space .*

The measure-valued process X_N lives in $C = C([0,T], \mathcal{M}_P(\mathbb{R}^d))$. Although, in general, weak convergence in C does not follow from weak convergence of the finite-dimensional distributions, it does in presence of relative compactness thanks to Corollary 2 [3]. We have the following result.

Theorem 16. *Let P_n, P be probability measures on $C([0,T], \mathcal{M}_P(\mathbb{R}^d)$. If the finite-dimensional distributions of P_n weakly converge to those of P, and if $\{P_n\}$ is tight, then $P_n \xrightarrow{\mathcal{W}} P$.*

To use this theorem we have to characterize the relative compactness on the space $C([0,T], \mathcal{M}_P(\mathbb{R}^d)$. Let us consider first $\mathcal{M}_P(\mathbb{R}^d)$.

Lemma 1. [15] *A subset $K \subset \mathcal{M}_P(\mathbb{R}^d)$ is relatively compact if and only if*

1. $\sup\{\|\mu\|_0 : \mu \in K\} < \infty$, *with* $\|\mu\|_0 = \sup\{\langle \mu, f\rangle, f \in C_b(\mathbb{R}^d), \|f\| \le 1\}$;
2. $\lim_{n\to\infty} \sup\{\|\mathbb{I}_{B_n^c}\mu\|_0 : \mu \in K\} = 0$, *with* $B_n^c := \{x \in \mathbb{R}^d : \|x\| \le n\}^c$.

The condition of compactness may be stated in terms of the following modulus of continuity

$$w_f(\delta) = w(f,\delta) = \sup_{|s-t|<\delta} |f(s) - f(t)|, \quad 0 < \delta \le 1,$$

thanks to a generalization of the Ascoli-Arzelá characterization. Indeed the Ascoli-Arzelá theorem states that a sequence of continuous functions on a metric space is tight if and only it is uniformly bounded and equicontinuous; uniform boundedness and equicontinuity are equivalent to (40) and (41) below.

Theorem 17 (Ascoli-Arzelà). [3] *A subset $A \in C$ is relative compact if and only if*

$$\sup_{f\in A} |f(0)| < \infty \tag{40}$$

e

$$\lim_{\delta\to 0} \sup_{f\in A} w_f(\delta) = 0. \tag{41}$$

From the previous theorem it is possible to characterize the tightness on $\mathcal{M}_P(C)$, where C is the space of continuous function on $[0,T]$.

Theorem 18. *[3] A sequence $\{P_n\}$ on C is tight if and only if these two conditions hold:*

i) for each positive η, there exists an a such that

$$P_n\{f : |f(0)| > a\} \le \eta, \quad n \ge 1;$$

ii) for each positive ϵ and η, there exists a δ, with $0 < \delta < 1$, and an integer n_0 such that

$$P_n\{f : w_f(\delta) \ge \epsilon\} \le \eta, \quad n \ge n_0.$$

A useful sufficient condition for tightness is offered by the following theorem.

Theorem 19. *[3] A sequence $\{P_n\}$ is tight if these two conditions are satisfied:*

i) for each positive η, there exists an a such that

$$P_n\{x : |x(0)| > a\} \le \eta \quad n \ge 1.$$

ii) for each positive ϵ and η, there exists a δ, with $0 < \delta < 1$, and an integer n_0 such that

$$\frac{1}{\delta} P_n\{x : \sup_{t\le s\le t+\delta} |x(s) - x(t)| \ge \epsilon\} \le \eta, \quad n \ge n_0,$$

for all t.

Let us now specialize the previous sufficient condition on the space of probability measures on $C([0,T], \mathcal{M}(\mathbb{R}^d))$, on which we consider the distance d_1, defined by (39).

Theorem 20. *Consider a sequence $(X_N)_{N\in\mathbb{N}}$ of measure-valued stochastic processes in $C([0,T], \mathcal{M}(\mathbb{R}^d))$. Suppose that*

(i) the initial sequence $(X_N(0))_{N\in\mathbb{N}}$ is tight in $\mathcal{M}(\mathbb{R}^d)$;

(ii) for any real positive ϵ and η there exists a $\delta \in (0,1)$, and a positive integer n_0 such that

$$\frac{1}{\delta} P\left(\sup_{t \le s \le t+\delta} d_1(X_N(s), X_N(t)) \ge \epsilon \right) \le \eta$$

for $N \ge n_0$ and $0 \le t \le T$.

Then the sequence of laws $(\mathcal{L}(X_N))_{N\in\mathbb{N}}$ in $\mathcal{M}\left(C([0,T], \mathcal{M}(\mathbb{R}^d))\right)$ is tight.

Condition *i)* states the initial tightness of the process, while condition *ii)* states the small variation of the process during small time intervals.

A convenient, though stronger, sufficient condition for compactness is given by Ethier-Kurtz [15].

Theorem 21. *Consider a sequence $(X_N)_{N\in\mathbb{N}}$ of measure-valued stochastic processes in $C([0,T], \mathcal{M}(\mathbb{R}^d))$, and let $\mathcal{F}_t^N := \sigma\{X_N(s) | s \le t\}$ be the natural filtration associated with $\{X_N(t), t \in [0,T]\}$. Suppose that*

(i) for any real positive ϵ and for any nonnegative rational t, a compact $\Gamma_{t,\epsilon}$ exists such that

$$\inf_N P(X_N(t) \in \Gamma_{t,\epsilon}) > 1 - \epsilon;$$

(ii) let $\alpha > 0$; for any real $\delta \in (0,1)$ a sequence $(\gamma_N^T(\delta))_{N\in\mathbb{N}}$ of nonnegative real random variables exists such that

$$\lim_{\delta\to 0} \limsup_{N\to\infty} \mathbb{E}[\gamma_N^T(\delta)] = 0$$

and, for any $t \in [0,T]$,

$$\mathbb{E}[d_1(X_N(t+\delta), X_N(t)))^\alpha | \mathcal{F}_t^N] \le \mathbb{E}[\gamma_N^T(\delta) | \mathcal{F}_t^N].$$

Then $(\mathcal{L}(X_N))_{N\in\mathbb{N}}$ is tight.

Condition *(i)* implies a pointwise control of compactness , while condition *(ii)* is again a control of the variation of the process during small time intervals. Indeed, it states that the mean distance of the process at two close times t and $t+\Delta t$ is small, and asymptotically in N, it converges to zero as the length Δt of the time interval tends to zero.

Existence and Uniqueness of a Limit Measure

The compactness properties of the laws of the process $\{X(t)\}_{t\in\mathbb{R}_+}$ allow us to prove the existence of limit measures of subsequences. In order to prove the existence of a unique limit, by the characterization given by Theorem 14, we need to prove that all the limits are equal. Usually this can be shown by characterizing the limit measure as the unique solution of a measure-valued differential equation.

As a consequence the main steps for proving the existence and uniqueness of a limit of a sequence of the laws of X_N are the following

a) prove relative compactness of the sequences $(\mathcal{L}(X_N))_{N\in\mathbb{N}}$, so to prove the existence of a limiting measure-valued process $\{X_\infty(t),\, t \in \mathbb{R}_+\}$.
b) identify any possible limit $\{X_\infty(t),\, t \in \mathbb{R}_+\}$ as the solution of a deterministic measure-valued differential equation;
c) possibly prove its absolute continuity with respect to the usual Lebesgue measure on $\mathbb{R}^d$, and identify the density $\rho(x,t)$ of $X_\infty(t)$ as a solution of a deterministic partial differential equation;
d) show uniqueness by proving the uniqueness of the solution of the deterministic partial differential equation.

In the next section we carry on the programme for a particularly interesting case.

5 A Specific Model for Interacting Particles

Here we want to present a possible model for the drift F_N in the system of stochastic differential equations (33). In particular we consider each individual subject to both an individual motion and an interaction with other individuals. Interaction is due to long range aggregation and short range repulsion forces [4, 26].

Modelling interaction at different ranges may be obtained in terms of an appropriate rescaling of a given reference kernel K. A mathematical way to distinguish among different scales is based on the choice of a "scaling" parameter in the kernel describing the interaction among individuals. As already mentioned in the introduction, if we consider a system of N particles located in $\mathbb{R}^d$, in the macroscopic space-time coordinates the *typical distance between neighboring particles* is $O(N^{-1/d})$, while the order of the size of the whole space is $O(1)$.

Let K be a sufficiently regular function, and assume that the interaction between two particles, out of N, located in x and y respectively, is modelled by

$$\frac{1}{N}K_N(x-y), \quad \text{where} \quad K_N(z) = N^\beta K(N^{\beta/d}z), \tag{42}$$

which expresses the rescaling of K with respect to the total number N of individuals, in terms of the scaling parameter $\beta \in [0,1]$. The force exerted on the k-th single particle located at $X_N^k(t)$, due to its interaction with all the others in the population, is given by

$$\begin{aligned} I^k \equiv (X_N(t) * V_N)(X_N^k(t)) &= \sum_{i=1}^N \frac{1}{N} V_N\left(X_N^k(t) - X_N^i(t)\right) \\ &= \sum_{i=1}^N N^{\beta-1} V_1\left(N^{\beta/d}\left(X_N^k(t) - X_N^i(t)\right)\right). \end{aligned} \tag{43}$$

We have a McKean-Vlasov (long range) interaction if $\beta = 0$, i.e. the range and the strength of the interaction do not depend on the scaling parameter N, and a moderate (short range) interaction if $\beta \in (0,1)$. In the latter case, as N tends to infinity, the range of the interaction kernel tends to zero, while its strength tends to infinity. For the long range aggregation we consider a symmetric kernel $G : \mathbb{R}^d \longrightarrow \mathbb{R}_+$, such that $G(x) = \tilde{G}(|x|), \quad x \in \mathbb{R}^d$, where $\tilde{G}$ is an increasing real valued function defined on $\mathbb{R}_+$; for the short range repulsion we consider a symmetric function V_N rescaled by N via a symmetric (with respect to zero) probability density V_1, i.e.

$$V_N(z) = N^{\beta} V_1(N^{\beta/d} z), \quad \beta \in (0,1). \tag{44}$$

So the system of SDE's becomes

$$\begin{aligned} dX_N^k(t) = &- \left[\gamma_1 \nabla U(X_N^k(t)) + \gamma_2 \left(\nabla \left(V_N - G\right) * X_N\right)(X_N^k(t))\right] dt \\ &+ \sigma dW^k(t), \qquad k = 1, \ldots, N, \end{aligned} \tag{45}$$

where $U : \mathbb{R}^d \to \mathbb{R}_+ \in C^2(\mathbb{R}^d)$ is a non negative smooth even potential and $\gamma_1, \gamma_2 \in \mathbb{R}_+$ are suitable weights. From the modelling point of view, a transport term including U may represent some external information from the environment, which drives any individual along the gradient of U. Figures 1, 2 and 3 show the behavior of the system for different choices of the parameters and the kernels.

Let us consider the following assumptions about the regularity of the kernels involved in system (45)

$$G, V_1 \in C_b^2(\mathbb{R}^d, \mathbb{R}_+) \cap L^1(\mathbb{R}^d, \mathbb{R}_+) \tag{46}$$

$$U \in C^2(\mathbb{R}^d, \mathbb{R}_+) \tag{47}$$

$$|\nabla U(\mathbf{x}) - \nabla U(\mathbf{y})|^2 \leq k|\ \mathbf{x} - \mathbf{y}|^2, \quad \forall (\mathbf{x}, \mathbf{y}) \in \mathbb{R}^d \times \mathbb{R}^d \tag{48}$$

$$|\nabla U(\mathbf{x})|^2 \leq k^*(|\ \mathbf{x}|^2 + 1), \quad \forall \mathbf{x} \in \mathbb{R}^d \tag{49}$$

where k and k^* are positive constants. As a consequence, System (45) admits a unique solution.

By (35), the evolution equation of the empirical measure (32), is, for any $f \in C_b^{2,1}(\mathbb{R}^d \times \mathbb{R}_+)$

$$\begin{aligned} \langle X_N(t), f(\cdot,t)\rangle = &\langle X_N(0), f(\cdot,0)\rangle \\ &- \int_0^t \langle X_N(s), \left[\gamma_1 \nabla U + \gamma_2 \left(\nabla \left(V_N - G\right) * X_N\right)\right](\cdot) \nabla f(\cdot, s)\rangle\, ds \\ &+ \int_0^t \left\langle X_N(s), \frac{\sigma_N^2}{2} \triangle f(\cdot, s) + \frac{\partial}{\partial s} f(\cdot, s) \right\rangle ds \\ &+ \sigma_N \int_0^t \langle X_N(s), \nabla f\,(\cdot, s)\rangle\, dW^k(s). \end{aligned} \tag{50}$$

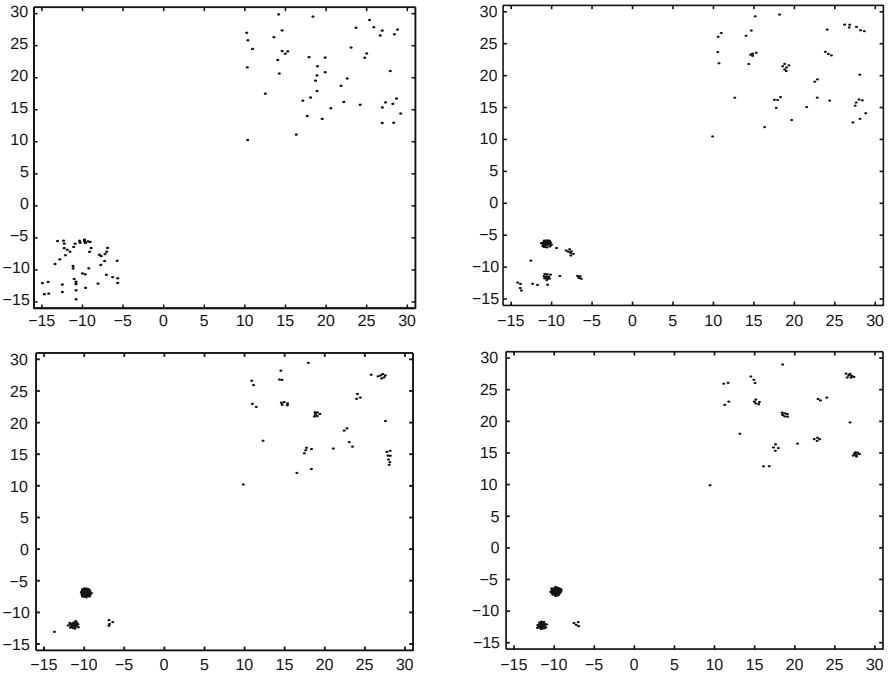

Fig. 1: Configuration of 100 particles for parameters values $\gamma_1 = 0$, $\gamma_2 = 1$, $\sigma = 0.02$, $\beta = 0.5$: (up left) $T = 0$, (up right) $T = 500$, (down left) $T = 1000$, (down right) $T = 2000$

5.1 Asymptotic Behavior of the System for Large Populations: A Heuristic Derivation

Suppose that indeed the empirical process $\{X_N(t),\ t \in \mathbb{R}_+\}$ tends, as $N \to \infty$, to a deterministic process $\{X(t),\ t \in \mathbb{R}_+\}$, and further that it admits for any $t \in \mathbb{R}_+$, a density $\rho(x,t)$ with respect to the Lebesgue measure on $\mathbb{R}^d$, so that

$$\lim_{N\to+\infty} \langle X_N(t), f(\cdot,t)\rangle = \langle X(t), f(\cdot,t)\rangle = \int_{\mathbb{R}^d} f(x,t)\rho(x,t)dx.$$

As a formal consequence, we get

$$\lim_{N\to+\infty} (X_N(t) * V_N)(x) = g_N(x) = \rho(x,t),$$

$$\lim_{N\to+\infty} (X_N(t) * \nabla V_N)(x) = \nabla g_N(x) = \nabla\rho(x,t),$$

$$\lim_{N\to+\infty} (X_N(t) * \nabla G_a(\cdot,t))(x) = (X(t) * \nabla G_a(\cdot,t))(x) = (\rho * \nabla G_a(\cdot,t))(x).$$

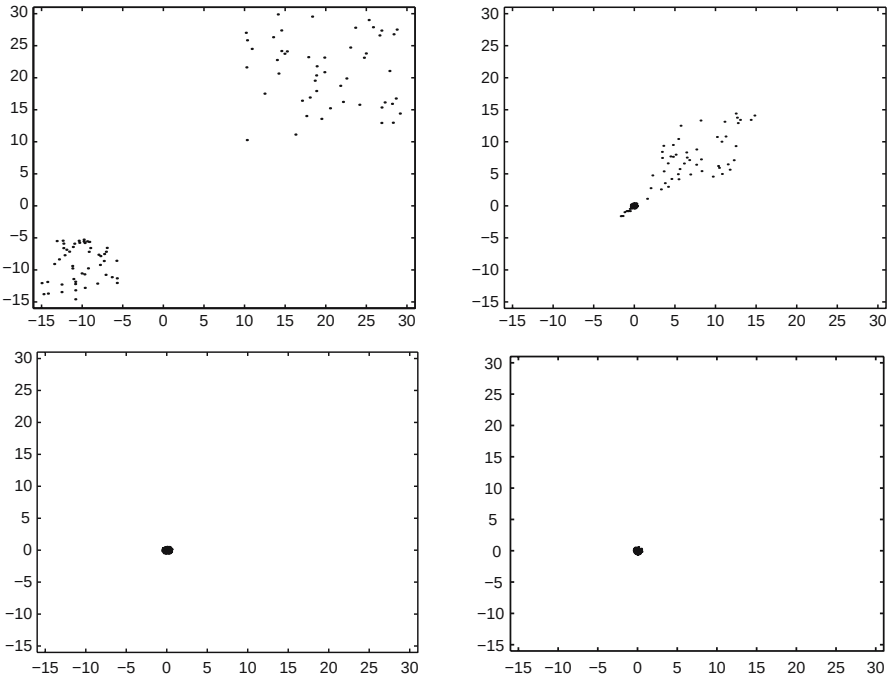

Fig. 2: Configuration of 100 particles for parameters values $\gamma_1 = 1$, $\gamma_2 = 1$, $\sigma = 0.02$, $\beta = 0.5$ with potential U such that $\nabla U(x) = x/(1+|x|)$: (up left) $T = 0$, (up right) $T = 200$, (down left) $T = 2000$, (down right) $T = 4000$

Hence, by applying the above limits, from (36) and (50) we get

$$\begin{aligned}\int_{\mathbb{R}^d} f(x,t)\rho(x,t)dx \;=\; & \int_{\mathbb{R}^d} f(x,0)\rho(x,0)dx \\ & + \int_0^t ds \int_{\mathbb{R}^d} dx\,[-\gamma_1 \nabla U(x) + \gamma_2(\nabla G_a * \rho(\cdot,s))(x) \\ & \qquad -\nabla\rho(x,s)] \cdot \nabla f(x,s)\rho(x,s) \\ & + \int_0^t ds \int_{\mathbb{R}^d} dx \left[\frac{\partial}{\partial s} f(x,s)\rho(x,s) + \frac{\sigma_\infty^2}{2} \triangle f(x,s)\rho(x,s)\right],\end{aligned} \tag{51}$$

where

$$\lim_{N\to\infty} \sigma_N = \sigma_\infty.$$

We recognize that (51) is the weak version of the following equation for the spatial density $\rho(x,t)$, for $x \in \mathbb{R}^d, t \geq 0$

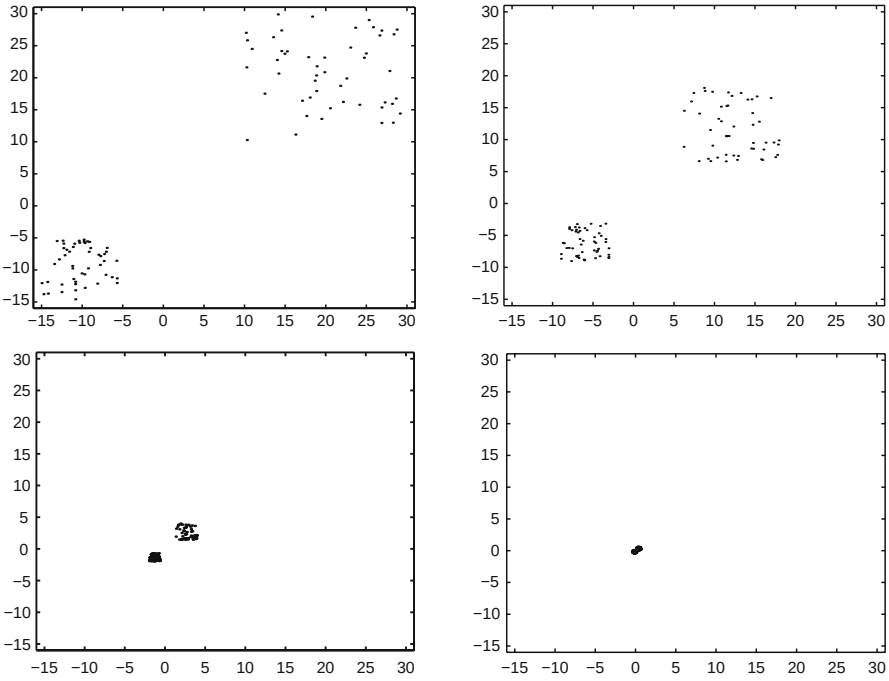

Fig. 3: Configuration of 100 particles for parameters values $\gamma_1 = 1$, $\gamma_2 = 1$, $\sigma = 0.02$, $\beta = 0.5$ with potential U such that $\nabla U(x) = |x|^2$: (up left) $T = 0$, (up right) $T = 5$, (down left) $T = 50$, (down right) $T = 100$

$$\frac{\partial}{\partial t}\rho(x,t) = \frac{\sigma_\infty^2}{2}\triangle\rho(x,t) + \gamma_1\nabla\cdot(\rho(x,t)\nabla U(x)) + \gamma_2\nabla\cdot(\rho(x,t)\left[\nabla\rho(x,t)\right) - \ (\nabla G_a * \rho(\cdot,t))(x)\right], \tag{52}$$

$$\rho(x,0) \ = \ \rho_0(x), \quad x \in \mathbb{R}^d. \tag{53}$$

From (52), if $\sigma_\infty > 0$ the dynamics of the density is smoothed by the diffusive term. This is due to the memory of the fluctuations existing when the number N of particles is finite. This also means that the dynamics of each individual particle is still stochastic. Indeed, it can be shown that for any k, we have $X_k(t) \sim Y^k(t)$, this last one satisfying the following SDE,

$$dY^k(t) = -\left[\nabla G_a * \rho(\cdot,t)(Y^k(t)) - \nabla\rho(Y^k(t)) - \nabla U(Y^k(t))\right]dt + \sigma_\infty dW^k(t),$$

subject to the initial condition $Y^k(0) = X_N^k(0)$. The Brownian stochasticity of the movement of each particle is preserved.

When $\sigma_\infty = 0$ all stochasticity disappears; this leads to a degenerate equation, for $x \in \mathbb{R}^d, t \geq 0$

$$\frac{\partial}{\partial t}\rho(x,t) = \nabla \cdot (\rho(x,t)\,[\gamma_1 \nabla U(x)) + \gamma_2(\nabla \rho(x,t) - \ (\nabla G_a * \rho(\cdot,t))(x))]\,.$$

For the individuals, for any k we have $X_k(t) \sim Y(t)$ subject to

$$dY(t) = -\left[\nabla G_a * \rho(\cdot,t)(Y(t)) - \nabla \rho(Y(t)) - \nabla U(Y(t))\right] dt,$$

and the initial condition $Y(0) = X_N^k(0)$, so that also the dynamics of an individual particle becomes fully deterministic; there is no memory of the existing fluctuations, when the number N of particles is finite.

Existence and Uniqueness of a Solution of the Limit Equation

In the viscous case, i.e. when $\sigma_\infty > 0$, a large literature is available showing existence and uniqueness of a sufficient regular solution satisfying Equation (52). On the other hand, in the non viscous case, i.e. when $\sigma_\infty = 0$, uniqueness is not a trivial problem, the lack of uniqueness being mainly due to the nonlocal transport term [10].

A major issue is to find the right notion of solution for equations like (54). It is well-known that classical solutions do not exist in general for degenerate equations, in particular for equations like (54) one has to expect that the solution is not differentiable at the boundary of the (compact) support. A usual way to overcome such difficulties for parabolic equations is to use *weak solutions*. However, for degenerate equations of the general form

$$\frac{\partial v}{\partial t} + \text{div } f(x,t,v) - \Delta a(v) = 0, \tag{54}$$

the weak solution may not be unique, and a different concept of solutions, so-called *entropy solutions*, has to be used in order to obtain uniqueness [9]. Another example of non uniqueness has been found for transport equations with nonlocal nonlinearity [10] corresponding to (54) without the diffusion term. To our knowledge the only available uniqueness result for an equation like (54) is due to [27], but it holds only in 1D for a very special convex long-range interaction kernel.

The above issues concerning uniqueness motivate the study of weak and entropy solutions for (54). In conservation laws, entropy solutions are usually obtained as vanishing viscosity limits and well motivated from a physical point of view. In the case of biological models, entropy solutions are hardly used and not well motivated so far. In [6] we have adapted this notion to our system, closely following this approach with a simple modification enforced by the nonlocal convolution operator.

5.2 Asymptotic Behavior of the System for Large Populations: A Rigorous Derivation

In this section we present a rigorous derivation of the limit measure $X(t)$, following Section 4.3. From now on we consider the case when the individual randomness does not vanish at infinity, that is

$$\lim_{N\to\infty} \sigma_N = \sigma_\infty > 0. \tag{55}$$

Let us consider the following additional assumptions on the kernels involved in the model; for $x \in \mathbb{R}^d$

$$V_N(x) = (W_N * W_N)(x) \tag{56}$$

$$W_N(x) = \chi_N^d W_1(\chi_N x), \quad \chi_N = N^{\beta/d}, \beta \in (0, d/(d+2)) \tag{57}$$

$$W_1 \in W_2^1(\mathbb{R}^d) \cap C_0(\mathbb{R}^d), U \in C_b^1(\mathbb{R}^d) \tag{58}$$

$$\left\langle -\nabla U(x), \frac{x}{|x|} \right\rangle \leq -\frac{r}{|x|}, \quad |x| \geq M_0, \tag{59}$$

where W_1 is a symmetric probability density, and define a mollified measure, i.e. a regular version of the empirical measure $X_N(t)$

$$h_N(x,t) = \left(W_N * X_N(t)\right)(x), \quad x \in \mathbb{R}^d.$$

Furthermore, we consider also the following properties for the initial state

$$\sup_{N\in\mathbb{N}} \mathbb{E}\left[\int_{\mathbb{R}^d} |x| X_N(0)(dx)\right] < \infty \tag{60}$$

$$\sup_{N\in\mathbb{N}} \mathbb{E}\left[\int_{\mathbb{R}^d} |h_N(x,0)|^2 dx\right] < \infty, \tag{61}$$

The first step for the analysis is to study the *relative compactness* of the stochastic process $\{X_N\}_{N\in\mathbb{N}}$. The main problem in the derivation of the compactness properties is due to the unboundedness of the drift term; so we have to take care of the possible explosion of the system. In order to deal with this problem, we define the following stopping time

$$\tau_N^k = \inf\{t \geq 0 \ : \ S_N(t) = ||h_N(\cdot,t)||_2^2 + A_N(t) - \int_0^t \langle X_N(u), 2|-\nabla U(\cdot) + (\nabla G_a * X_N(u))(\cdot)|^2\rangle du > k\},$$

where

$$A_N(t) = \int_0^t \langle X_N(s), 2(|\nabla g_N(\cdot,u)|^2 - \nabla g_N(\cdot,u)(-\nabla U(\cdot) + (\nabla G_a * X_N(u))(\cdot)) + |-\nabla U(\cdot) + (\nabla G_a * X_N(u))(\cdot)|^2)\rangle + \sigma_N^2 ||\nabla h_N(\cdot,u)||_2^2 du.$$

Following [28, 31] one can prove that the process

$$M_N(t) = ||h_N(\cdot,t)||_2^2 - \int_0^t \langle X_N(u),2| - \nabla U(\cdot) + (\nabla G_a * X_N(u))(\cdot)|^2\rangle du \\ + A_N(t) - c_1\sigma_N^2 t N^{\beta(d+2)/d-1}$$

is a martingale.

Nonexplosion in Finite Time

It is possible to prove the nonexplosion in a finite time, i.e. for any τ such that $0 < \tau < \infty$,

$$\lim_{k\to\infty} \inf_{N\in\mathbb{N}} \mathbb{P}\{\tau_N^k > \tau\} = 1. \tag{62}$$

Proof.

From the martingale property of the process $M_N(t)$, it derives that the process

$$t \mapsto S_N(t) = ||h_N(\cdot,t)||_2^2 + A_N(t) - \int_0^t \langle X_N(u),2| - \nabla U(\cdot) + (\nabla G_a * X_N(u))(\cdot)|^2\rangle du$$

is a submartingale. By Doob's inequality

$$\begin{aligned}
\mathbb{P}\left\{\sup_{t\le\tau} S_N(t) > k\right\} &\le \frac{1}{k}\mathbb{E}[S_N(\tau)] = \frac{1}{k}\mathbb{E}[M_N(\tau) + \tau\sigma_N^2 N^{\beta(d+2)/d-1}c_1] \\
&= \frac{1}{k}\mathbb{E}\left[\mathbb{E}[M_N(\tau)|\mathcal{F}_0] + C\tau\sigma_N^2 N^{\beta(d+2)/d-1}\right] \\
&= \frac{1}{k}\mathbb{E}[M_N(0) + C\tau\sigma_N^2 N^{\beta(d+2)/d-1}] \\
&= \frac{1}{k}\left(\mathbb{E}[||h_N(\cdot,0)||_2^2] + C\tau\sigma_N^2 N^{\beta(d+2)/d-1}\right) \\
&\le C(\tau)/k;
\end{aligned} \tag{63}$$

the last inequality comes from by (55),(58) and (61).

□

First we study the stopped process

$$X_{N,k}(t) = X_N(t \wedge \tau_N^k); \tag{64}$$

indeed the relative compactness of the stopped process (64) and condition (62) imply the relative compactness of the full process [28].

Relative Compactness for the Stopped Process $X_{N,k}(t)$

In order to prove the relative compactness of the laws of $X_{N,k}(t)$, we use the characterization of Theorem 21. By using the martingale property of the

process (62) and Doob's inequality, one can prove that, for any $\epsilon > 0$, there exists a compact K_ϵ^k in $(\mathcal{M}_\mathcal{P}(\mathbb{R}^d), d_{BL})$ such that

$$\inf_{N\in\mathbb{N}} \mathbb{P}\{X_{N,k}(t) \in K_\epsilon^k,\ \forall t \in [0,T]\} \geq 1-\epsilon,$$

i.e. the first sufficient condition of Theorem 21 is satisfied. Then, by the properties of the dynamics of the system, one can prove the property of small fluctuations during small time intervals, and in particular that for any $0 < \delta < 1$, there exists a sequence $\{\gamma_n^T(\delta)\}_{n\in\mathbb{N}}$ of non negative random variables such that

$$\mathbb{E}\left[d_{BL}(X_{N,k}(t+\delta), X_{N,k}(t))^4\right] \leq \mathbb{E}\left[\gamma_n^T(\delta)\right] \quad 0 \leq t \leq T,$$

and $\lim_{\delta\to 0} \limsup_{n\to\infty} \mathbb{E}[\gamma_n^T(\delta)] = 0$.

From the previous analysis we may then state that the sequence of laws $\{\mathcal{L}(X_N(\cdot \wedge \tau_N^k))\}_{N\in\mathbb{N}}$ is relatively compact in $\mathcal{M}_\mathcal{P}(C([0,T], \mathcal{M}_\mathcal{P}(\mathbb{R}^d)))$.

As already mentioned, the relative compactness of the sequence $\{\mathcal{L}(X_N)\}_{N\in\mathbb{N}}$ in $\mathcal{M}_\mathcal{P}(C([0,T], \mathcal{M}_\mathcal{P}(\mathbb{R}^d)))$ follows from the non explosion of the stopping time τ_N^k,.

Finally from Skorokhod's Theorem we may assert that for the possible unique limit law we can get also an almost sure convergence, i.e.

$$\lim_{N\to\infty} \sup_{t\leq T} d_{BL}(X_N(t), X(t)) = 0 \quad \mathbb{P}-a.s. \tag{65}$$

Regularity of the Limit Measure

The next step is to study the regularity properties of the limit measure. It is possible to show that the mollified measure h_N, given by (60) is an $L^2([0,T], \mathcal{M}_P(\mathbb{R}^d))$ Cauchy sequence [28,31]. As a consequence the sequence admits a limit $\rho \in L^2(\mathbb{R}^d, \mathbb{R}_+)$

$$\lim_{N\to\infty} \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} |h_N(x,t) - \rho(x,t)|^2 dx dt\right] = 0. \tag{66}$$

From (66) and $\lim_{N\to\infty} W_N(\cdot) = \delta_0$ (in the sense of distributions), we have

$$\lim_{N\to\infty} \int_{\mathbb{R}^d} f(x,t) X_N(t)(dx) = \int_{\mathbb{R}^d} f(x,t)\rho(x,t)dx \quad f \in C_b(\mathbb{R}^d \times [0,T]) \quad \mathbb{P}-a.s.$$

As a consequence, because of the limit property (65), we get

$$\int_{\mathbb{R}^d} f(x,t) X(t)(dx) = \int_{\mathbb{R}^d} f(x,t)\rho(x,t)dx \quad f \in C_b(\mathbb{R}^d \times [0,T]), \mathbb{P}-a.s.$$

Therefore the measure $X(t)$ is absolutely continuous with respect to the Lebesgue measure with density $\rho(x,t)$, i.e. for any $f \in C_b(\mathbb{R}^d), t \in [0,T]$

$$\lim_{N\to\infty} \langle X_N(t), f(\cdot)\rangle = \langle X(t), f(\cdot)\rangle = \int_{\mathbb{R}^d} f(x)\rho(x,t)dx. \tag{67}$$

Uniqueness of the Limit

Uniqueness of the limit measure X is shown, by proving that the density ρ is the solution of the system (52)-(53), which admits a unique solution. We further assume that a law of large numbers applies to the initial condition, i.e.

$$\lim_{N\to\infty} \mathcal{L}(X_N(0)) = \delta_{X_0} \quad \text{in} \quad \mathcal{M}_{\mathcal{P}}(\mathcal{M}_{\mathcal{P}}(\mathbb{R}^d)), \tag{68}$$

where X_0 has density $\rho(x,0)$ with respect to the Lebesgue measure. As a consequence

$$\lim_{N\to\infty} \mathcal{L}(X_N(t)) = \delta_{X_N(t)} \quad \text{in} \quad \mathcal{M}_{\mathcal{P}}(\mathcal{M}_{\mathcal{P}}(\mathbb{R}^d)),$$

for any $f \in C_b^{2,1}(\mathbb{R}^d, \mathbb{R}_+)$, uniformly in $t \in [0,T]$, where ρ is the unique solution of (52)-(53), as proven in the Section A.

6 Long Time Behavior: Invariant Measure

Here we wish to analyze the long time behavior of the system (51) for a fixed N. Our interest is to analyze mechanisms that are responsible for stable aggregation. From the mathematical point of view this is equivalent to the study of the joint distribution law of the vector $\mathbf{X} = (X_N^1, \ldots, X_N^N) \in \mathbb{R}^n$, where $n = Nd$. Let now $P_N^{x_0}(t)$ denote the joint distribution of the N particles at time t, conditional upon a non random initial condition x_0.

Invariant Measure

Since $\mathbf{X}$ is a homogeneous Markov process, one can associate a homogeneous transition probability $p(x, t-s, dy)$ such that, with $B \in \mathcal{B}_{\mathbb{R}^n}$, for all $0 \le s < t < +\infty$,

$$P\{X(t) \in B | X(s) = x\} = \int_B p(x, t-s, dy).$$

If there exists a probability measure $\overline{P}_N$ on $\mathbb{R}^n$ independent of time t, as a solution of the integral equation

$$\forall t > 0, \quad \overline{P}_N(dy) = \int_{x\in\mathbb{R}^n} \overline{P}_N(dx) p(x; t, dy),$$

then, $\overline{P}_N(dx)$ is called an *invariant measure* associated with the Markov process X.

Let us consider some sufficient conditions for the *existence of an invariant measure*. In particular we consider the results the reader may find in [19], p. 118, that stipulate that a process $\mathbf{X}(t)$ has finite mean recurrence time for some bounded domain U, and within this domain all sample paths "mix sufficiently well". More precisely we consider the following assumption (H):

There exists a bounded domain $D \subset \mathbb{R}^n$ with regular boundary, having the following properties:

H1: in the domain D and some neighborhood thereof, the smallest eingenvalue of the diffusion matrix is bounded away from zero.

H2: if $x \in D^c$, the mean time τ_{U^c} at which a path issuing from x reaches the set D is finite, i.e. $E[\tau_{D^c}|\mathbf{X}(0) = x] < \infty$, and $\sup_{x\in K} E[\tau_K|\mathbf{X}(0) = x] < \infty$, for every compact $K \subset \mathbb{R}^n$.

Theorem 22. *[19] If condition (H) holds, then the Markov process* $\mathbf{X}(t)$ *has a unique invariant measure* $\overline{P}_N$.

The Purely Interacting Case

First of all we consider the case of the drift due only to interaction among particles, that is $\gamma_1 = 0$.

In Figure 4 we consider some simulation results with $V_1 = \alpha_1\mathcal{N}(0,1)$ and $G = \alpha_2\mathcal{N}(0,1)$, for different values of α_1 and α_2. We notice that, as t increases, the mean distance between two particles and consequently the radius of clusters fluctuates around an asymptotic value which is finite, but in the case of pure repulsion.

Following [22], for the position of the center of mass of the N particles,

$$\bar{X}_N(t) = \frac{1}{N}\sum_{k=1}^{N} X_N^k(t),$$

we have

$$d\bar{X}_N(t) = -\frac{1}{N^2}\sum_{k,j=1}^{N} \nabla\left(V_N - G\right)\left(X_N^k(t) - X_N^j(t)\right)dt + \sigma d\bar{W}(t), \tag{69}$$

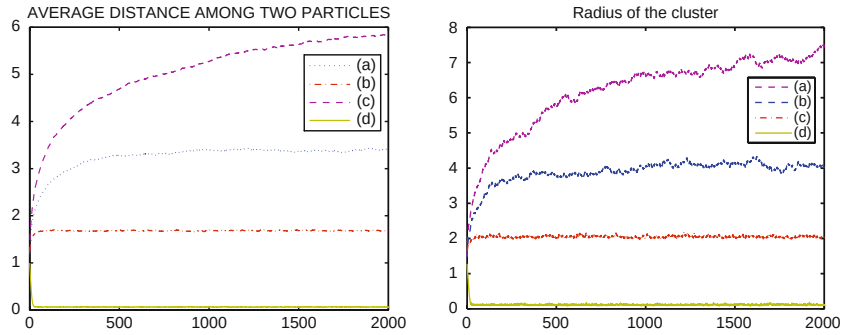

Fig. 4: Comparison among the evolution of the radius of the cluster and the average distance among particles for different values of parameters: (a) $\alpha_1 = \alpha_2 = 1$; (b) $\alpha_1 = 1; \alpha_2 = 2$; (c) $\alpha_1 = 1; \alpha_2 = 0$; (d) $\alpha_1 = 0; \alpha_2 = 1$

where $\bar{W}(t) = \frac{1}{N}\sum_{k=1}^{N} W^k(t)$; by the symmetry properties of V_1 and G, the first term on the right hand side vanishes and we get

$$d\bar{X}_N(t) = \sigma d\bar{W}(t), \tag{70}$$

i.e. $\bar{X}_N(t)$ is a Brownian motion. Hence, its distribution is

$$\mathcal{L}\left(\bar{X}_N(t)|\bar{X}_N(0)\right) = \mathcal{L}\left(\bar{X}_N(0), \sigma^2\bar{W}(t)\right) = \mathcal{N}\left(\bar{X}_N(0), \frac{\sigma^2}{N}t\right);$$

with variance $\frac{\sigma^2}{N}\,t$, which, for any fixed N, increases as t tends to infinity.

Consequently we may claim that the probability law of the system does not converge to any probability measure, since otherwise the same would happen for the law of the center of mass.

In other models [22] in addition to symmetry, strictly convexity of the interaction kernel K is added. In spite of that, once again because of the symmetry of K, the system

$$d\mathbf{X}_N(t) = -\nabla\cdot(K * X_N(t))(\mathbf{X}_N(t))dt + \sigma d\mathbf{W}(t), \quad \mathbf{X}_N(t) \in \mathbb{R}^n, \forall t,$$

with

$$\exists \lambda > 0, \forall x, v \in \mathbb{R}^d, \quad \langle \mathrm{Hess}K(x)v, v\rangle \geq \lambda\langle v, v\rangle; \tag{71}$$

does not admit an invariant distribution.

The interesting feature is that, in this case, one may show stationary properties for the reduced system for the relative coordinates, the distances of the particle position with respect to the center of mass $\bar{X}_N(t)$,

$$Y_N^j(t) = X_N^j(t) - \bar{X}_N.$$

Indeed, under the assumption (71), it has been shown [22] that the system

$$Y_N^k(t) = Y_N^k(0) - \frac{1}{N}\sum_{l=1}^{N}\int_0^t \nabla K(Y_N^k(s) - Y_N^l(s))ds$$
$$+ \sigma W^k(t) - \frac{\sigma}{N}\sum_{l=1}^{N} W^l(t) \qquad k = 1, \ldots, N,$$

is still a diffusion on the manifold

$$\mathcal{M} = \left\{(x_1, \ldots, x_n) : \sum_i x_i = 0\right\}, \tag{72}$$

and does admit an invariant measure. Roughly speaking, adding the strictly convexity, each particle, and in particular the center of mass, attracts any other one in the whole space.

Hence, if we consider the system under study, (35), with $\gamma_1 = 0$, since the kernels G and V_N are symmetric, does not admit an invariant measure. In the next section we see a way to overcome this problem and which properties to require to the added kernel U in (35) in order to get an invariant measure for the system.

The Complete System

Consider now the potential P associated with a measure on $\mathbb{R}^d$, $\mu \in \mathcal{M}(\mathbb{R}^d)$

$$P(\mu)(x) = \gamma_1 U(x) - \gamma_2 \left((V_N - G) * \mu\right)(x), \quad x \in \mathbb{R}^d, \tag{73}$$

so that system (45) can be rewritten as

$$dX_N^k(t) = -\nabla P(X_N(t))(X_N^k(t)) + \sigma dW^k(t), \qquad k = 1, \ldots, N. \tag{74}$$

System (74) has been thoroughly analyzed in literature under the sufficient condition (71) of strict convexity on U; it has been shown [9, 22] that if this condition applies, system (74) does admit a nontrivial invariant distribution. From a biological point of view a strictly convex confining potential is difficult to explain; it would mean an infinite range of attraction, with an at least constant drift, even far from origin.

A weaker sufficient condition for the existence of a unique invariant measure has been more recently suggested by Veretennikov [36], following Has'minski [19],

V: there exist constants $M_0 \geq 0$ and $r > 0$ such that for $|x| \geq M_0$

$$\left(-\nabla P(\mu)(x), \frac{x}{|x|}\right) \leq -\frac{r}{|x|}. \tag{75}$$

Since the diffusion matrix is σI, where I is the identity matrix, then assumption (H1) for the existence of an invariant measure is satisfied. Assumption (75), is needed to show that $\mathbf{X}(t)$ has finite mean recurrence time for some bounded domain U, i.e. assumption (H2) is satisfied [36].

Without any further condition on the interaction kernels V_N and G, for (75) to hold it is sufficient to assume that there exist constants $M_0 \geq 0$ and $r > \frac{1}{N\gamma_1}(\frac{Nd}{2} + 1)$ such that for $|x| \geq M_0$, equation (59) hold. We wish to remark that condition (59) means that ∇D may decay to zero as $|x|$ tends to infinity, provided that its tails are sufficiently "fat".

Proposition 7. *If the confining potential U satisfies condition* (59)*, then* (74) *admits a unique invariant measure.*

Proof.

Let $\pi_i(\mathbf{x}) = x_i, i = 1, \ldots, N$ be the i-th projection of $\mathbf{x} \in (R^d)^N$, $\tilde{U}(\mathbf{x})$ and $\tilde{K}(\mathbf{x})$ the vector function defined by

$$\tilde{U}(\mathbf{x}) = (U \circ \pi_i(\mathbf{x}))_{1 \leq i \leq N}, \quad \tilde{K}(\mathbf{x}) = \left((G - V_N) * \frac{1}{N} \sum_i \epsilon_{\pi_i(\mathbf{x})} \circ \pi_i(\mathbf{x})\right)_{1 \leq i \leq N}$$

In order to apply Theorem 2 in [35], we have to prove that there exist constants $M \geq 0$ and $\tilde{r} > (\frac{Nd}{2} + 1)$ such that for all $\mathbf{x} \in (R^d)^N : |x| \geq M$

$$\left(-\gamma_1 \nabla \tilde{U}(\mathbf{x}) + \gamma_2 \nabla \tilde{K}(\mathbf{x}), \frac{\mathbf{x}}{|\mathbf{x}|}\right) \leq -\frac{\tilde{r}}{|\mathbf{x}|}. \tag{76}$$

We have

$$\begin{aligned}\left(-\gamma_1 \nabla \tilde{U}(\mathbf{x}) + \gamma_2 \nabla \tilde{K}(\mathbf{x}), \frac{\mathbf{x}}{|\mathbf{x}|}\right) &= -\gamma_1 \sum_{k=1}^{N} \nabla U(x_k) \frac{x_k}{|\mathbf{x}|} \\ &\quad + \gamma_2 \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{N} \nabla (G - V_N)(x_i - x_k) \frac{x_k}{|\mathbf{x}|} \\ &\leq -\gamma_1 \sum_{k=1}^{N} \nabla U(x_k) \frac{x_k}{|\mathbf{x}|} \\ &\quad + \gamma_2 \frac{1}{N} \sum_{k=1}^{N} \sum_{i=1}^{N} \nabla (G - V_N)(x_i - x_k) \\ &= -\gamma_1 \sum_{k=1}^{N} \nabla U(x_k) \frac{x_k}{|\mathbf{x}|} \leq -\frac{\gamma_1 r N}{|\mathbf{x}|}\end{aligned}$$

The last two inequalities derive from the symmetry of G and V_N, together with (59). So if for $\tilde{r} = \gamma_1 r N$ and condition on r in (59), we have condition (76).

□

Following [35] we may also provide an estimate of the rate of convergence. Indeed, as far as the convergence of $P_N^{x_0}(t)$, for t tending to infinity, is concerned, we may state that

Theorem 23. *Under assumption* (59)*, for any* k*,* $0 < k < \tilde{r} - \frac{Nd}{2} - 1$ *with* $m \in (2k + 2, 2\tilde{r} - Nd)$ *and* $\tilde{r} = \gamma_1 N r$*, a positive constant* c *exists such that*

$$\left|P_N^{x_0}(t) - P_N^S\right| \leq c(1 + |x_0|^m)(1 + t)^{-(k+1)},$$

where $\left|P_N^{x_0}(t) - P_N^S\right|$ *denotes the total variation distance of the two measures, i.e.*

$$\left|P_N^{x_0}(t) - P_N^S\right| = \sup_{A \in \mathcal{B}_{\mathbb{R}^d}} \left[P_N^{x_0}(t)(A) - P_N^S(A)\right].$$

and x_0 *the initial data.*

A Proof of the Identification of the Limit ρ

Since (51) is the weak form of (52), it is sufficient to show that for any $f \in C_b^{2,1}(\mathbb{R}^d, \mathbb{R}_+)$,

$$\mathbb{E}\left[\left|\langle X(t),f(\cdot,t)\rangle-\langle\mu_0,f(\cdot,0)\rangle-\int_0^t\langle\rho(\cdot,s),\frac{1}{2}\sigma_\infty^2\Delta f(\cdot,s)+\frac{\partial}{\partial s}f(\cdot,s)\right.\right.$$
$$\left.\left.+[(\nabla G_a*\rho(\cdot,s))(\cdot)-\nabla U(\cdot)-\nabla\rho(\cdot,s)]\cdot\nabla f(\cdot,s)\rangle ds\ \right|\right]=0.$$

For fixed $f\in C_b^{2,1}(\mathbb{R}^d,\mathbb{R}_+)$

$$\begin{aligned}
&\mathbb{E}\ \left[\left|\langle X(t),f(\cdot,t)\rangle-\langle\mu_0,f(\cdot,0)\rangle-\int_0^t\langle\rho(\cdot,s),\frac{1}{2}\sigma_\infty^2\Delta f(\cdot,s)+\frac{\partial}{\partial s}f(\cdot,s)\right.\right.\\
&\quad\left.\left.+[(\nabla G_a*\rho(\cdot,s))(\cdot)-\nabla U(\cdot)-\nabla\rho(\cdot,s)]\cdot\nabla f(\cdot,s)\rangle ds\ |\right]\\
&\le\mathbb{E}\,[|\langle X(t),f(\cdot,t)\rangle-\langle X_N(t),f(\cdot,t)\rangle|]\\
&+\ \mathbb{E}\,[|\langle\mu_0,f(\cdot,0)\rangle-\langle X_N(0),f(\cdot,0)\rangle|]\\
&+\ \frac{\sigma_\infty^2}{2}\mathbb{E}\left[\int_0^t|\langle-\rho(\cdot,s),\Delta f(\cdot,s)\rangle+\langle X_N(s),\Delta f(\cdot,s)\rangle|ds\right]\\
&+\ \mathbb{E}\left[\int_0^t|-\langle\rho(\cdot,s),\frac{\partial}{\partial s}f(\cdot,s)\rangle+\langle X_N(s),\frac{\partial}{\partial s}f(\cdot,s)\rangle|ds\right]\\
&+\ \mathbb{E}\left[\int_0^t|\langle\rho(\cdot,s),\nabla\rho(\cdot,s)\cdot\nabla f(\cdot,s)\rangle-\langle h_N(\cdot,s),\nabla h_N(\cdot,s)\cdot\nabla f(\cdot,s)\rangle|ds\right]\\
&+\ \mathbb{E}\left[\left|\int_0^t\langle h_N(\cdot,s),\nabla h_N(\cdot,s)\cdot\nabla f(\cdot,s)\rangle-\langle X_N(s),\nabla g_N\cdot\nabla f(\cdot,s)\rangle ds\right|\right]\\
&+\ \mathbb{E}\left[\int_0^t|-\langle\rho(\cdot,s),[(\nabla G_a*\rho(\cdot,s))(\cdot)-\nabla U(\cdot)]\cdot\nabla f(\cdot,s)\rangle\right.\\
&\quad\left.+\ \langle X_N(s),[(\nabla G_a*X_N(s))(\cdot)-\nabla U(\cdot)]\cdot\nabla f(\cdot,s)\rangle|\,ds\right]\\
&+\ \mathbb{E}\left[\left|\frac{\sigma_N}{N}\int_0^t\sum_{k=1}^N\nabla f(X_N^k(s),s)dW_k(s)\right|\right]\\
&+\ \mathbb{E}\left[\left|\langle X_N(t),f(\cdot,t)\rangle-\langle X_N(0),f(\cdot,0)\rangle\right.\right.\\
&\quad-\int_0^t\langle X_N(s),(\nabla G_a*X_N(s))\cdot\nabla f(\cdot,s)\rangle ds+\int_0^t\langle X_N(s),\nabla g_N(\cdot,s)\cdot\nabla f(\cdot,s)\rangle ds\\
&\quad+\int_0^t\langle X_N(s),\nabla U(\cdot)\cdot\nabla f(\cdot,s)\rangle ds-\int_0^t\langle X_N(s),\frac{1}{2}\sigma_N^2\Delta f(\cdot,s)+\frac{\partial}{\partial s}f(\cdot,s)\rangle ds\\
&\quad\left.\left.-\ \frac{\sigma_N}{N}\int_0^t\sum_{k=1}^N\nabla f(X_N^k(s),s)dW_k(s)\right|\right]\\
&:=\sum_{i=1}^9 I_N^i(t).
\end{aligned}\tag{77}$$

Clearly

By (67) and hypothesis (68) $\sum_{i=1}^4 I_N^i(t)=0$, by (50) $I_N^9(t)=0$ and by (36) $I_N^8(t)=0$. It remains to estimate the terms $I_N^5(t)$, $I_N^6(t)$ and $I_N^7(t)$.

$$
\begin{aligned}
I_N^5(t) &= \mathbb{E}\left[\int_0^t |\langle \rho(\cdot,s), \rho(\cdot,s)\Delta f(\cdot,s)\rangle - \langle h_N(\cdot,s), h_N(\cdot,s)\Delta f(\cdot,s)\rangle| ds\right] \\
&\leq ||\Delta f||_\infty \int_0^T \mathbb{E}\left[\int_{\mathbb{R}^d} |h_N(x,t) - \rho(x,t)||h_N(x,t) + \rho(x,t)| dx\right] dt \\
&\leq ||\Delta f||_\infty \left(\mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} |h_N(x,t) - \rho(x,t)|^2 dx dt\right]\right)^{1/2} \\
&\quad \cdot \left(\mathbb{E}\left[\int_0^T \int_{\mathbb{R}^d} |h_N(x,t) + \rho(x,t)|^2 dx dt\right]\right)^{1/2};
\end{aligned}
$$

by (63) and (66) we obtain

$$
\lim_{N\to\infty} I_N^5(t) = 0. \tag{78}
$$

By the symmetry of W_1,

$$
\begin{aligned}
I_N^6(t) &= \mathbb{E}\Bigg[\Bigg| \int_0^t \langle X_N(s), W_N * (\nabla h_N(\cdot,s) \cdot \nabla f(\cdot,s)) \\
&\qquad -(W_N * \nabla h_N(\cdot,s)) \cdot \nabla f(\cdot,s)\rangle ds\Bigg|\Bigg] \\
&= \mathbb{E}\Bigg[\Bigg| \int_0^t \Bigg(\int_{\mathbb{R}^d} X_N(s)(dx) \int_{\mathbb{R}^d} W_N(x-y) \nabla h_N(y,s) \\
&\qquad \cdot(\nabla f(y) - \nabla f(x)) dy\Bigg) ds\Bigg|\Bigg].
\end{aligned} \tag{79}
$$

By the definition of W_N and since W_1 has compact support, with $c = \text{diam}(\text{supp} W_1(\cdot))$ and $||D^2 f||_\infty = \sup_{i,j\leq d} ||\partial_{ij}^2||_\infty$, (79) is less than or equal to

$$
\begin{aligned}
& c\chi_N^{-1} ||D^2 f||_\infty \mathbb{E}\left[\int_0^t \langle X_N(s) * W_N, |\nabla h_N(\cdot,s)|\rangle ds\right] \\
&\leq c\chi_N^{-1} ||D^2 f||_\infty \left(\mathbb{E}\left[\int_0^T ||h_N(\cdot,s)||_2^2 ds\right]\right)^{1/2} \left(\mathbb{E}\left[\int_0^T ||\nabla h_N(\cdot,s)||_2^2 ds\right]\right)^{1/2} \\
&\leq c\chi_N^{-1} ||D^2 f||_\infty.
\end{aligned}
$$

It follows that

$$
\lim_{N\to\infty} I_N^6(t) = 0 \tag{80}
$$

.

$$
\begin{aligned}
I_N^7(t) = \mathbb{E}\Bigg[\int_0^t & |-\langle \rho(\cdot,s), [(\nabla G_a * \rho(\cdot,s))(\cdot) - \nabla U(\cdot)] \cdot \nabla f(\cdot,s)\rangle \\
& + \langle X_N(s), [(\nabla G_a * X_N(s))(\cdot) - \nabla U(\cdot)] \cdot \nabla f(\cdot,s)\rangle \\
& + \langle X_N(s), [(\nabla G_a * \rho(\cdot,s))(\cdot) - \nabla U(\cdot)] \cdot \nabla f(\cdot,s)\rangle \\
& - \langle X_N(s), [(\nabla G_a * \rho(\cdot,s))(\cdot) - \nabla U(\cdot)] \cdot \nabla f(\cdot,s)\rangle| \, ds \Bigg] \\
\leq \mathbb{E}\Bigg[\int_0^t & |\langle X_N(s) - \rho(\cdot,s), [(\nabla G_a * \rho(\cdot,s))(\cdot) - \nabla U(\cdot)] \cdot \nabla f(\cdot,s)\rangle| \\
& + |\langle X_N(s), [(\nabla G_a * \rho(\cdot,s))(\cdot) - (\nabla G_a * X_N(s))(\cdot)] \cdot \nabla f(\cdot,s)\rangle| \, ds \Bigg] .
\end{aligned}
$$

By (65)

$$\lim_{N\to\infty} I_N^7(t) = 0.$$

As a consequence

$$\lim_{N\to\infty} \sum_{i=1}^{9} I_N^i(t) = 0$$

uniformly in $t \in [0.T]$.

Acknowledgments

It is a pleasure to acknowledge fruitful discussions with Karl Oelschläger. Thanks also to Matteo Ortisi.

References

1. Ash R. B. & Gardner M. F., Topics in Stochastic Processes, Academic Press, London; 1975.
2. Banasiak J., Positivity in Natural Sciences. This Volume.
3. Billingsley P., Convergence of Probability Measures, Wiley, New York; 1968.
4. Boi, S., Capasso V., Morale, D. "Modeling the aggregative behavior of ants of the species Polyergus rufescens.", Nonlinear Analysis: Real World Applications, 1, 2000, p. 163-176.
5. Bodnar, M., J.J.L. Velazquez, An integro-differential equation arising as a limit of individual cell-based models, J. Diff. Eqs., 222, 2, (2006), 341380.
6. Burger M., Capasso V., Morale D., On an aggregation model with long and short range interactions, Nonlinear Anal. Real World Appl., 2006. In Press.
7. Capasso V., Mathematical Structures of Epidemic Systems, Springer-Verlag, Heidelberg; 1993.
8. Capasso V., Bakstein D. An Introduction to Continuous-Time Stochastic Processes-Theory, Models and Applications to Finance, Biology and Medicine. Birkhäuser, Boston, 2004.

9. Carrillo J., Entropy solutions for nonlinear degenerate problems, Arch. Rat. Mech. Anal., 269–361, 147; 1999.
10. Diekmann, O., M. Gyllenberg, H.R. Thieme, Lack of uniqueness in transport equations with a nonlocal nonlinearity, Math. Models and Meth. in Appl. Sciences **10** (2000), 581-592.
11. Dudley, R.M., Convergence of Baire measures, Studia Math. 27 (1966), 251-268.
12. Dudley, R.M., Real Analysis and Probability, Cambridge Studies in Advanced Mathematics 74, Cambridge University Press, Cambridge, 2002.
13. Durrett, R., Levin, S.A., "The importance of being discrete (and spatial)." Theor. Pop. Biol., **46**, 1994, 363-394.
14. Dynkin E. B., Markov Processes, Springer-Verlag, Berlin, Vols. 1–2; 1965.
15. Ethier S. N. & Kurtz T. G., Markov Processes, Characterization and Convergence, Wiley, New York; 1986.
16. Feller W., An Introduction to Probability Theory and Its Applications, Wiley, New York; 1971.
17. Gihman I. I. & Skorohod A. V., The Theory of Random Processes, Springer-Verlag, Berlin; 1974.
18. Grünbaum, D., Okubo, A. "Modelling social animal aggregations" In "Frontiers of Theoretical Biology" (S. Levin Ed.), Lectures Notes in Biomathematics, 100, Springer Verlag, New York, 1994, 296-325.
19. Has'minski, R.Z. Stochastic Stability of Differential Equations. Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands and Rockville, Maryland, USA, 1980.
20. Karlin S., Taylor H. M., A First Course in Stochastic Processes, Academic Press, New York; 1975.
21. Lachowicz M., Links Between Microscopic and Macroscopic Descriptions. This Volume.
22. Malrieu, F. Convergence to equilibrium for granular media equations and their Euler schemes. The Annals of Applied Probability, **13**, 540-560, 2003.
23. Métivier M., Notions Fondamentales de la Théorie des Probabilités, Dunod, Paris; 1968.
24. Mogilner A., L. Edelstein-Keshet, A non-local model for a swarm, J. Math. Bio. **38**, 534-549, 1999.
25. Morale, D., V. Capasso, K. Ölschlaeger, An interacting particle system modelling aggregation behavior: from individuals to populations, J. Math. Bio. **50**, 49-66(2005).
26. Morale D., Capasso V., & Oelschläger K., A rigorous derivation of the mean-field nonlinear integro-differential equation for a population of aggregating individuals subject to stochastic fluctuations, Preprint 98–38 (SFB 359), IWR, Universität Heidelberg, Juni; 1998.
27. Nagai T. & Mimura M., Some nonlinear degenerate diffusion equations related to population dynamics, J. Math. Soc. Japan, 539–561, 35; 1983.
28. Oelschläger K. A law of large numbers for moderately interacting diffusion processes. Z. Wahrscheinlichkeitstheorie verw. Gabiete 69, 279-322; 1985.
29. Oelschläger K., Large systems of interacting particles and the porous medium equation, J. Differential Equations, 294–346, 88; 1990.
30. A. Okubo, S. Levin, Diffusion and Ecological Problems : Modern Perspectives (Springer, Berlin), 2002.
31. Ortisi, M. *Limiting Behavior of an Interacting Particle Systems*, Ph.D., University of Milano, Italy, 2007.

32. Rogers L. C. G. & Williams D., Diffusions, Markov Processes and Martingales, Vol. 1, Wiley, New York; 1994.
33. Shiryaev A. N., Probability, Springer-Verlag, New York; 1995.
34. van der Vaart, Aad W. and Wellner, Jon A., Weak Convergence and Empirical Processes, with Applications to Statistics, Springer, New York, 1996.
35. Veretennikov, A.Y. On polynomial mixing bounds for stochastic differential equations. Stochastic Processes and their Applications, **70**, 115-127, 1997.
36. Veretennikov, A.Y. On polynomial mixing and convergence rate for for stochastic differential equations. Theory Probab. Appl., **44**, 361-374, 1999.
37. Veretennikov, A.Y. On subexponential mixing rate for Markov processes. Theory Probab. Appl., **49**, 110-122, 2005.
38. Conradt, L. and Roper, T.J., Group decision-making in animals, Nature, **42**, 155-158, 2003.
39. Conradt, L. and Roper, T.J., Consensus decision-making in animals, Trends Ecol. Evolut., **20**, 449-456, 2005.
40. Meunier, H., Leca, J. B., Deneubourg, J.L., and Petit,O., Group movement in capuchin monkeys: the utility of an experimental study and a mathematical model to explore the relationship between individual and collective behaviours. Preprint, 2006.