



Propagation of chaos: a review of models, methods and applications. I. Models and methods

Louis-Pierre Chaintron, Antoine Diez

► To cite this version:

Louis-Pierre Chaintron, Antoine Diez. Propagation of chaos: a review of models, methods and applications. I. Models and methods. 2022. hal-03585065v3

HAL Id: hal-03585065

<https://hal.science/hal-03585065v3>

Preprint submitted on 4 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Propagation of chaos: a review of models, methods and applications.

I. Models and methods

Louis-Pierre CHAINTRON¹ and Antoine DIEZ²

¹DMA, École Normale Supérieure
45 rue d'Ulm
75005 Paris, France
`lchaintron@clipper.ens.fr`

²Department of Mathematics,
Imperial College London, South Kensington Campus,
London, SW7 2AZ, UK
`antoine.diez18@imperial.ac.uk`

May 4, 2022

Abstract

The notion of propagation of chaos for large systems of interacting particles originates in statistical physics and has recently become a central notion in many areas of applied mathematics. The present review describes old and new methods as well as several important results in the field. The models considered include the McKean-Vlasov diffusion, the mean-field jump models and the Boltzmann models. The first part of this review is an introduction to modelling aspects of stochastic particle systems and to the notion of propagation of chaos. The second part presents concrete applications and a more detailed study of some of the important models in the field.

Keywords: Kac's chaos, McKean-Vlasov, Boltzmann models, mean-field limit, particle system

AMS subject classification: 82C22, 82C40, 35Q70, 65C35, 92-10

Contents

1	Introduction	3
2	Models and properties	9
2.1	Particle systems, setting and conventions	9
2.2	Mean-field models	14
2.2.1	Abstract mean-field generators and mean-field limits	14
2.2.2	McKean-Vlasov diffusion	15
2.2.3	Mean-field jump processes and PDMPs	21
2.3	Boltzmann models	27
2.3.1	General form	27
2.3.2	Parametric Boltzmann models	34
2.3.3	Classical models in collisional kinetic theory	37
3	Notions about chaos	43
3.1	Topology reminders: metrics and convergence for probability measures . .	43
3.1.1	Distances on the space of probability measures	43
3.1.2	Convergence in the space of probability measures	46
3.1.3	Entropic convergence	51
3.2	Representation of symmetric particle systems	52
3.2.1	Finite particle systems	52
3.2.2	Infinite particle systems and random measures	57
3.3	Kac's chaos	60
3.3.1	Definition and characterisation	60
3.3.2	Quantitative versions of Kac's chaos	62
3.4	Propagation of chaos	64
3.4.1	Quantitative and uniform in time propagation of chaos.	65
3.4.2	Pointwise from pathwise.	66
3.4.3	Propagation of chaos via the empirical process.	66
4	Proving propagation of chaos	69
4.1	Coupling methods	70
4.1.1	Definition	70
4.1.2	Synchronous coupling.	71
4.1.3	Reflection coupling for the McKean-Vlasov diffusion.	72
4.1.4	Optimal jumps	74
4.1.5	Analysis in Wasserstein spaces: optimal coupling	76
4.2	Compactness methods	77
4.2.1	Martingale methods.	77
4.2.2	Gradient flows.	78
4.3	A pointwise study of the empirical process	79
4.3.1	The empirical generator.	80
4.3.2	A toy-example using the measure-valued formalism	81
4.3.3	The limit semi-group and the nonlinear measure-valued process . .	85

4.3.4	More on the limit generator	87
4.3.5	The abstract theorem	90
4.4	Large Deviation Related Methods	93
4.4.1	Chaos through Large Deviation Principles	94
4.4.2	Chaos from entropy bounds	96
4.4.3	Tools for concentration inequalities	102
4.5	Tools for Boltzmann interactions	104
4.5.1	Series expansions	104
4.5.2	Interaction graph	106
A	Probability reminders	109
A.1	Convergence of probability measures	109
A.2	Skorokhod's topology and tightness on the Skorokhod space	110
A.3	Stochastic processes and martingales	112
A.3.1	Martingales	112
A.3.2	Local martingales and quadratic variation	113
A.3.3	Semimartingales	115
A.3.4	D -semimartingales	116
A.4	Markov processes and Markov representation for PDEs	117
A.4.1	Time-homogeneous Markov processes and linear PDEs	118
A.4.2	Time-inhomogeneous Markov processes	124
A.4.3	Non-linear Markov processes	127
A.5	Large Deviation Principles and Sanov theorem	128
A.6	Girsanov transform	129
A.7	Poisson random measures	129
B	A strengthened version of Hewitt-Savage theorem and its partial system version	132
	References	135

1 Introduction

When Boltzmann published his most famous article [24] one century and a half ago, the study of large systems of interacting particles was entirely motivated by the microscopic modelling of thermodynamic systems. Although it was far from being an accepted idea at that time, Boltzmann postulated that since a macroscopic volume of gas contains a myriad of elementary particles, it is both hopeless and needless to keep track of each particle and one should rather seek a statistical description. He thus derived the equation that now bears his name and which gives the time evolution of the continuum probability distribution (in the phase space) of a typical particle. With the H-theorem, he also extended and justified the pioneering works of Maxwell and Clausius for equilibrium thermodynamic systems, paving the way alongside Gibbs for a consistent kinetic theory

of gases. The Boltzmann equation is derived from first principles under a crucial assumption, called *molecular chaos*. This assumption was already known from Maxwell and is often called the *Stosszahlansatz* since Ehrenfest. Informally, it translates the idea that, despite the multitude of interactions, two particles taken at random should be statistically independent when the total number of particles grows to infinity. It is not so clear how the appearance of probability theory should be interpreted in this context. In the following years, the *Stosszahlansatz* and its consequences (the H-theorem) were the object of a fierce debate among physicists as they seem to break the microscopic reversibility. Beyond the scientific debate, it has raised metaphysical and philosophical questions about the profound nature of time and randomness.

The rigorous justification of the work of Boltzmann and the status of molecular chaos became true mathematical questions when Hilbert addressed them in his Sixth Problem at the Paris International Congress of Mathematicians in 1900. Quoting Hilbert, the problem which motivates the present work is to “[develop] mathematically the limiting processes [...] which lead from the atomistic view to the laws of motion of continua”. Our starting point will be the seminal article of Kac [104]. More than half a century after Hilbert, Kac gave the first rigorous mathematical definition of chaos and introduced the idea that for time-evolving systems, chaos should be propagated in time, a property therefore called the *propagation of chaos*. Kac was still motivated by the mathematical justification of the classical collisional kinetic theory of Boltzmann for which he developed a simplified probabilistic model. Soon after Kac, McKean [125] introduced a class of diffusion models which were not originally part of Boltzmann theory but which satisfy Kac’s propagation of chaos property. In the classical kinetic theory of Boltzmann, the problem is the derivation of continuum models starting from deterministic, Newtonian, systems of particles. In comparison, the fundamental contribution of Kac and McKean is to have shown that the classical equations of kinetic theory also have a natural stochastic interpretation. This philosophical shift is addressed in the enlightening introduction of Kac [105] written for the centenary of the Boltzmann equation.

Kac and McKean introduced a new mathematical formalism, gave many insights on the stochastic modelling in kinetic theory and proved the two building block theorems (Theorem II-3.1 and Theorem II-4.1). Their works have stimulated the development of a rich and still active *mathematical kinetic theory*. Keeping strong connections with the original theory of Boltzmann, some fundamental questions raised several decades ago have been answered only recently (see for instance [17, 83, 130]). On the other hand, systems of interacting particles are ubiquitous in many applications now and over the last two decades, the tools and concepts developed in kinetic theory have somehow escaped the realm of pure statistical physics. This review paper is motivated by the growing number of models in applied mathematics where the notion of chaos plays a central role. Some recent new domains of applications include the following ones. In mathematical biology and social sciences, self-organization models describe systems of indistinguishable particles (birds, insects, bacteria, crowds. . .) with a behaviour which can hardly be predicted at the microscopic scale but which are (sometimes) well explained by continuum models derived within the framework of mathematical kinetic theory [8, 9, 134, 132, 59, 2, 157]. In another context, the recent theory of mean-field games studies the asymptotic properties

of games with a large numbers of players [30, 29, 36, 37]. Even more recently, systems of particles have been used to model complex phenomena in data sciences, with applications in Markov Chain Monte Carlo theory [63, 64, 65], in optimization [139, 155, 90, 39], or for the training of neural networks [126, 142, 147, 47, 56]. Compared to the models in statistical physics, many aspects should be reconsidered. To cite a few examples: the basic conservation laws (of momentum, energy...) do not always hold for biological systems and may be replaced by other types of constraints (optimization constraints, geometrical constraints...); the intrinsic randomness (or uncertainty) of the models in applied sciences is often a crucial modelling assumption; the complexity of the interaction mechanisms entails new analytical tools etc. These differences have motivated many new techniques, new insights on the question of propagation of chaos and in the end, new results.

This review article on propagation of chaos is not the first one on the subject. The course of Sznitman at Saint-Flour [151] studies many of the most important historical probabilistic models. The probabilistic methods are explained in details in the book [152] (in particular the courses of Méléard [127] and Pulvirenti [140]). More recently, the review of Jabin and Wang [98] focuses on McKean mean-field systems and PDE applications. By its nature, the notion of chaos lies in the interplay between probability theory and Partial Differential Equations. The present review discusses both analytic and probabilistic methods and includes many (very) recent results. We also refer to the article by Hauray and Mischler [94] which is to our knowledge, the most complete reference on Kac's chaos (without propagation of chaos). For deterministic systems which will not be considered here, we refer to the very thorough reviews [97, 85].

Outline.

The article is organised as follows.

Section 2 introduces the setting and the conventions that will be used throughout the article. A gallery of the models which will be studied is presented; we will distinguish McKean's mean-field jump and diffusion models (Section 2.2) and Boltzmann-Kac models (Section 2.3).

Section 3 is devoted to the description of the fundamental tools and concepts needed in the study of exchangeable particle systems. The central notions of chaos (Section 3.3) and propagation of chaos (Section 3.4) are defined in this section.

Section 4 is a review of the methods used to prove propagation of chaos. Several probabilistic and analytical techniques are described as well as abstract theorems which will be applied to specific models in the second part of this review.

Finally, for the reader's convenience, we collect in Appendix A useful notions and results in Probability theory regarding stochastic processes in the Skorokhod space, Markov processes, martingale methods, large deviations, the Girsanov transform and the theory of Poisson random measures. Some corollaries of the quantitative Hewitt-Savage theorem are gathered in Appendix B.

The second part of this review will be devoted on the one hand to the classical models introduced by Kac and McKean and their recent developments and on the other hand to a gallery of recent applications in applied mathematics and beyond. Throughout this first

part, references to the second part are indicated by “II-” (for instance Section II-2 refers to the second section of the second part).

Notations and conventions

Sets

$C(I, E)$	The set of continuous functions from a time interval $I = [0, T]$ to a set E , endowed with the uniform topology.
$C_b(E), C_b^k(E)$	Respectively the set of real-valued bounded continuous functions and the set of functions with $k \geq 1$ bounded continuous derivatives on a set E .
$C_c(E)$	The set of real-valued continuous functions with compact support on a locally compact space E .
$C_0(E)$	The set of real-valued continuous functions vanishing at infinity on a locally compact space E , i.e. $\varphi \in C_0(E)$ when for all $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset E$ such that $ \varphi(x) < \varepsilon$ for all $x \in E$ outside K_ε .
$D(I, E)$	The space of functions which are right continuous and have left limit everywhere from a time interval $I = [0, T]$ to a set E , endowed with the Skorokhod $J1$ topology. This is the space of the <i>càdlàg</i> functions. This space is also called the <i>Skorokhod space</i> or the <i>path space</i> .
$L^p(E)$ or $L_\mu^p(E)$	The set of measurable functions φ defined almost everywhere on a measured space (E, μ) such that the $ \varphi ^p$ is integrable for $p \geq 1$. When $p = +\infty$, this is the set of functions with a bounded essential supremum. We do not specify the dependency in μ when no confusion is possible.
$\mathcal{M}_d(\mathbb{R})$	The set of d -dimensional square real matrices.
$\mathcal{M}(E)$	The set of signed measures on a measurable space E .
$\mathcal{M}^+(E)$	The set of positive measures on a measurable space E .
$\mathcal{P}(E)$	The set of probability measures on a space E .
$\mathcal{P}_p(E)$	The set of probability measures with bounded moment of order $p \geq 1$ on a space E .
$\hat{\mathcal{P}}_N(E)$	The set of empirical measures of size N over a set E , that is measures of the form $\mu = \frac{1}{N} \sum_{i=1}^N \delta_{x^i}$, where $x^i \in E$.
$\mathbb{R}_+$	The set $[0, +\infty)$.
$\mathfrak{S}_N$	The permutation group of the set $\{1, \dots, N\}$.
$\mathbb{S}^{d-1}$	The sphere of dimension $d - 1$.

Generic elements and operations

C	A generic nonnegative constant, the value of which may change from line to line.
-----	--

$C(a_1, \dots, a_n)$	A generic nonnegative constant which depends on some fixed parameters denoted by $a_1, \dots, a_n$. Its value may change from line to line.
$\text{diag}(x)$	The d -dimensional diagonal matrix whose diagonal coefficients $x_1, \dots, x_d$ are the components of the d -dimensional vector x .
$\nabla \cdot V$	The divergence of a vector field $V : \mathbb{R}^d \rightarrow \mathbb{R}^d$ or of a matrix field $V : \mathbb{R}^d \rightarrow \mathcal{M}_d(\mathbb{R})$, respectively defined by $\nabla \cdot V = \sum_{i=1}^d \partial_{x_i} V_i$ or componentwise by $(\nabla \cdot V)_i = \sum_{j=1}^d \partial_{x_j} V_{ij}$.
$A : B$ and $\ A\ $	The Frobenius inner product of two matrices $A, B \in \mathcal{M}_d(\mathbb{R})$ defined by $A : B := \sum_{i=1}^d \sum_{j=1}^d A_{ij} B_{ij}$ and the associated norm $\ A\ := \sqrt{A : A}$.
$\nabla^2 V$	The Hessian matrix of a scalar field $V : \mathbb{R}^d \rightarrow \mathbb{R}$ defined componentwise by $(\nabla^2 V)_{ij} = \partial_{x_i, x_j}^2 V$.
I_d	The d -dimensional identity matrix.
Id	The identity operator on a vector space.
$\langle x, y \rangle$ or $x \cdot y$	The Euclidean inner product of two vectors $x, y \in \mathbb{R}^d$ defined by $\langle x, y \rangle \equiv x \cdot y := \sum_{i=1}^d x^i y^i$. One notation or the other may be preferred for typographical reasons in certain cases.
M_{ij}	The (i, j) (respectively row and column indexes) component of a matrix M .
$P(u)$	The projection matrix $P(u) := I_d - \frac{u \otimes u}{ u ^2}$ on the plane orthogonal to a vector $u \in \mathbb{R}^d$.
$\varphi \in C_b(E)$	A generic bounded continuous test function on E .
$\varphi_N \in C_b(E^N)$	A generic bounded continuous test function on the product space E^N .
$\Phi \in C_b(\mathcal{P}(E))$	A generic bounded continuous test function on the set of probability measures on E .
$u \otimes v, \mu \otimes \nu$ or $\varphi \otimes \psi$	Respectively, the matrix tensor product of two vectors $u, v \in \mathbb{R}^d$ defined componentwise by $(u \otimes v)_{ij} = u_i v_j$; the product measure on $E \times F$ of two measures μ, ν respectively on E and F ; the product function on $E \times F$ defined by $(\varphi \otimes \psi)(x, y) = \varphi(x) \psi(y)$ for two real-valued function φ, ψ respectively on E and F .
$\text{Tr } M$	The trace of the matrix M .
M^T	The transpose of the matrix M .
$\mathbf{x}^N = (x^1, \dots, x^N)$	A generic element of a product space E^N . The components are indexed with a superscript.
$\mathbf{x}^{M,N} = (x^1, \dots, x^M)$	The M -dimensional vector in E^M constructed by taking the M first components of $\mathbf{x}^N$.
$x = (x_1, \dots, x_d)^T$ and $ x $	A generic element of a d -dimensional space and its norm. The coordinates are indexed with a subscript. The norm of x denoted by $ x $ is the Euclidean norm.

Probability and measures

$K \star \mu$	The convolution of a function $K : E \times F \rightarrow G$ with a measure μ on F defined as the function $K \star \mu : x \in E \mapsto \int_F K(x, y) \mu(dy) \in G$. When $E = F = G = \mathbb{R}^d$ and $K : \mathbb{R}^d \rightarrow \mathbb{R}^d$, we write $K \star \mu(x) = \int_{\mathbb{R}^d} K(x - y) \mu(dy)$.
δ_x	The Dirac measure at the point x .
$\mu_{\mathbf{x}^N}$	The empirical measure defined by $\mu_{\mathbf{x}^N} = \frac{1}{N} \sum_{i=1}^N \delta_{x^i}$ where $\mathbf{x}^N = (x^1, \dots, x^N)$.
$\mathbb{E}_\mu[\varphi]$	Alternative expression for $\langle \mu, \varphi \rangle$ when μ is a probability measure. When $\mu = \mathbb{P}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$, the expectation is simply denoted by $\mathbb{E}$.
$H(\nu \mu)$	The relative entropy (or Kullback-Leibler divergence) between two measures μ, ν , see Definition 3.10.
$\langle \mu, \varphi \rangle$	The integral of a measurable function φ with respect to a measure μ .
$\text{Law}(X)$	The law of a random variable X as an element of $\mathcal{P}(E)$ where X takes its value in the space E .
$(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$	A filtered probability space. Unless otherwise stated, all the random variables are defined on this set. The expectation is denoted by $\mathbb{E}$.
$\sigma(X^1, X^2, \dots)$	The σ -algebra generated by the random variables $X^1, X^2, \dots$.
$T_\# \mu$	The pushforward of the measure μ on a set E by the measurable map $T : E \rightarrow F$. This is a measure on the set F defined by $T_\# \mu(\mathcal{A}) = \mu(T^{-1}(\mathcal{A}))$ for any measurable set $\mathcal{A}$ of F .
$\ \cdot\ _{\text{TV}}$	The Total Variation (TV) norm for measures.
W_p	The Wasserstein- p distance between probability measures (see Definition 3.1).
$X \sim \mu$	It means that the law of the random variable X is μ .
$(X_t)_t$ or $(Z_t)_t$	The canonical process on the path space $D(I, E)$ defined by $X_t(\omega) = \omega(t)$.
$(\mathbf{X}_t^N)_t$ or $(\mathbf{Z}_t^N)_t$	The canonical process on the product space $D(I, E)^N$ with components $\mathbf{X}_t^N = (X_t^1, \dots, X_t^N)$.

Systems of particles and operators

E	The state space of the particles, assumed to be at least a Polish space.
f_t^N	The N -particle distribution in $\mathcal{P}(E^N)$ at time $t \geq 0$.
$f_t^{k,N}$	The k -th marginal of f_t^N .
f_I^N	The N -particle distribution on the path space in $\mathcal{P}(D(I, E^N))$ or $\mathcal{P}(C(I, E^N))$ for a time interval $I = [0, T]$. We identify $D(I, E^N) \simeq D(I, E)^N$.
f_t	The limit law in $\mathcal{P}(E)$ at time $t \geq 0$.
f_I	The limit law on the path space in $\mathcal{P}(D(I, E))$ or $\mathcal{P}(C(I, E))$.
F_t^N	The law of the empirical process in $\mathcal{P}(\mathcal{P}(E))$ at time $t \geq 0$.
$F_I^{\mu,N}$	The weak pathwise law of the empirical process in $\mathcal{P}(D(I, \mathcal{P}(E)))$ on the time interval $I = [0, T]$.

F_I^N	The strong pathwise law of the empirical process in $\mathcal{P}(\mathcal{P}(D(I, E)))$ on the time interval $I = [0, T]$.
$\mathcal{L}_N$	The N -particle generator acting on (a subset of) $C_b(E^N)$.
$\mathcal{L}^N$	The N -particle generator acting on $\mathcal{P}(E^N)$ defined as the formal adjoint of $\mathcal{L}_N$.
$L \diamond_i \varphi_N$	The action of an operator L on (a subset of) $C_b(E)$ against the i -th variable of a function φ_N in $C_b(E^N)$, defined as the function in (a subset of) $C_b(E^N)$ $L \diamond_i \varphi_N : (x^1, \dots, x^N) \mapsto L[x \mapsto \varphi_N(x^1, \dots, x^{i-1}, x, x^{i+1}, \dots, x^N)](x^i)$. The definition readily extends to the case of an operator $L^{(2)}$ acting on $C_b(E^2)$ and two indexes $i < j$ in which case we write $L^{(2)} \diamond_{ij} \varphi_N$.
$(\mathcal{X}_t^N)_t$	The N -particle process, with components $\mathcal{X}_t^N = (X_t^{1,N}, \dots, X_t^{N,N}) \in E^N$. Often we write $X_t^{i,N} \equiv X_t^i$ and $(\mathcal{X}_t^N)_t \equiv \mathcal{X}_{[0,T]}^N$.
$(\mathcal{Z}_t^N)_t$	An alternative notation for the N -particle process with $\mathcal{Z}_t^N = (Z_t^{1,N}, \dots, Z_t^{N,N})$. Often used for Boltzmann particle systems or kinetic systems.

2 Models and properties

2.1 Particle systems, setting and conventions

The starting point of this review is a system of N particles

$$\mathcal{X}_I^N \equiv (\mathcal{X}_t^N)_{t \in I} \equiv (X_t^{1,N}, \dots, X_t^{N,N})_{t \in I},$$

where each particle $(X_t^{i,N})_{t \in I}$ is a stochastic process with values in the state space E which is at least Polish (i.e. separable and completely metrizable) and defined on a time interval $I = [0, T]$ with $T \in (0, +\infty]$. When no confusion is possible, we drop the N superscript and only write $X_t^i \equiv X_t^{i,N}$ for the i -th particle. The N -particles are not independent; they are said to *interact*.

Throughout this review, $(\mathcal{X}_t^N)_t$ is a nice stochastic process in E^N which satisfies the strong Markov property and which has càdlàg sample paths (the related topology is the J_1 Skorokhod topology, see definition A.5). Several examples will be given in the next sections but to fix ideas, the particle system will either be a Feller diffusion process in $E = \mathbb{R}^d$ (or in a Borel subset or in a manifold) or a jump process in a more general state space which satisfies the C_b -Feller property (i.e. the transition operator is strongly continuous and maps $C_b(E^N)$ to $C_b(E^N)$). One may also consider mixed jump-diffusion processes. The N -particle process will often be given as the solution of a Stochastic Differential Equation (SDE) but in full generality, it will be described by its generator denoted by $\mathcal{L}_N$ acting on a domain $\text{Dom}(\mathcal{L}_N) \subset C_b(E^N)$ or $\text{Dom}(\mathcal{L}_N) \subset C_0(E^N)$ (a detailed description of the analytical setting for Markov processes and their generators is provided in Appendix A.4, see in particular Definitions A.25, definition A.27 and Theorem A.29). The generator determines what is called the *interaction mechanism*. The only but crucial assumption that is made on this interaction mechanism is the symmetry or *exchangeability*.

Definition 2.1 (Exchangeability). A family $(X^i)_{i \in I}$ of random variables is said to be *exchangeable* when the law of $(X^i)_{i \in I}$ is invariant under every permutation of a finite number of indexes $i \in I$.

In a dynamical setting, the *pathwise exchangeability* is assumed in the sense that exchangeability holds for the family of processes $(X_I^{i,N})_{1 \leq i \leq N}$, at the level of trajectories. Taking the time-coordinate (i.e. the push-forward of the family law by the map $\omega \mapsto \omega(t)$), this implies the *pointwise exchangeability* i.e. exchangeability for the position vector $(\mathcal{X}_t^{1,N}, \dots, \mathcal{X}_t^{N,N})$ at any time $t \geq 0$. Formally, $\mathcal{X}_t^N$ can be seen as an element of $E^N/\mathfrak{S}_N$, where $\mathfrak{S}_N$ denotes the group of all permutations of $\{1, \dots, N\}$, although for simplicity we will keep using E^N as the state space. Such a particle system will be called *exchangeable*.

The statistical description.

In statistical physics, the previous description in terms of stochastic process is sometimes called the *microscopic scale* because the trajectory of each individual particle is recorded. When N is *large*, the microscopic scale contains too much information and a *statistical description* is sought. There are at least three statistical points of view on the particle system, detailed below.

1. The easiest one, is simply given by the N -particle distribution $f_t^N \in \mathcal{P}(E^N)$ at time $t \in I$. From the general theory of Markov processes (see Appendix A.4), f_t^N satisfies the forward Kolmogorov equation written in weak form:

$$\forall \varphi_N \in \text{Dom}(\mathcal{L}_N), \quad \frac{d}{dt} \langle f_t^N, \varphi_N \rangle = \langle f_t^N, \mathcal{L}_N \varphi_N \rangle. \quad (1)$$

Here and throughout this review, the bracket notation $\langle \cdot, \cdot \rangle$ denotes the integral of a test function (here φ_N) against a probability measure (here f_t^N).

Remark 1. Note that

$$\langle f_t^N, \varphi_N \rangle = \langle f_0^N, u_N(t, \cdot) \rangle,$$

where for $\mathbf{x}^N \in E^N$, $u_N \equiv u_N(t, \mathbf{x}^N) := \mathbb{E}[\varphi_N(\mathcal{X}_t^N) | \mathcal{X}_0^N = \mathbf{x}^N]$ solves the backward Kolmogorov equation

$$\partial_t u_N = \mathcal{L}_N u_N.$$

The equation (1) thus describes the dynamics of an *observable* of the system.

Equation (1) is also called the *master equation* in a probabilistic context and is better known as the *Liouville equation* in classical (deterministic) kinetic theory. In this review, we follow this latter terminology and Equation (1) will be called the *(weak) Liouville equation*. The forward Kolmogorov equation, or (strong) Liouville equation, reads

$$\partial f_t^N = \mathcal{L}^N f_t^N,$$

where $\mathcal{L}^N \equiv \mathcal{L}_N^*$ is the dual operator of $\mathcal{L}_N$. In general, no explicit expression for $\mathcal{L}^N$ is available and it is thus easier to focus on the weak point of view. From the weak

Liouville equation, it is possible to compute the time evolution of any *observable* of the particle system, that is of any of the averaged quantities $\langle f_t^N, \varphi_N \rangle$ for a test function φ_N . The drawback is that $f_t^N \in \mathcal{P}(E^N)$ belongs to a high dimensional space (since N is large). However, by the exchangeability assumption, the law f_t^N is a symmetric probability measure and it is thus possible to define for any $k \in \mathbb{N}$ the k -th marginal $f_t^{k,N} \in \mathcal{P}(E^k)$ by:

$$\forall \varphi_k \in C_b(E^k), \quad \langle f_t^{k,N}, \varphi_k \rangle := \langle f_t^N, \varphi_k \otimes 1^{\otimes(N-k)} \rangle.$$

The exchangeability ensures that the term on the right-hand side does not depend on the indexes of the k variables; for instance, it would be equivalent to take $1^{\otimes(N-k)} \otimes \varphi_k$ as a test function instead of $\varphi_k \otimes 1^{\otimes(N-k)}$. Each marginal distribution satisfies a Liouville equation obtained from (1) by taking φ_k as a test function. This equation may not be closed in the sense that, depending on $\mathcal{L}_N$, the right-hand side may depend on f_t^N or on the other marginals.

Remark 2. Note that with a slight abuse we take φ_k in the space $C_b(E^k)$ although we should say that φ_k belongs to a subset of $C_b(E^k)$ and is such that $\varphi_k \otimes 1^{\otimes(N-k)}$ belongs to $\text{Dom}(\mathcal{L}_N)$. We will often keep doing that in the following.

2. From the point of view of stochastic analysis, the particle system $\mathcal{X}_I^N$ can be seen as a random element of $D(I, E^N) \simeq D(I, E)^N$ so its law is a probability measure on the path space, denoted by $f_I^N \in \mathcal{P}(D(I, E^N))$. This *pathwise law* is generally given as the unique solution of the following martingale problem (probability reminders can be found in Appendix A).

Definition 2.2 (Particle martingale problem). Let $T \in (0 + \infty]$. A pathwise law $f_{[0,T]}^N \in \mathcal{P}(D([0,T], E^N))$ is said to be a solution of the martingale problem associated to the particle system issued from $f_0^N \in \mathcal{P}(E^N)$ whenever for all $\varphi_N \in \text{Dom}(\mathcal{L}_N)$,

$$M_t^{\varphi_N} := \varphi_N(\mathbf{X}_t^N) - \varphi_N(\mathbf{X}_0^N) - \int_0^t \mathcal{L}_N \varphi_N(\mathbf{X}_s^N) ds, \quad (2)$$

is a $f_{[0,T]}^N$ -martingale, where $(\mathbf{X}_t^N)_{t \geq 0}$ denotes the canonical process on the Skorokhod space $D([0,T], E^N)$ defined for $\omega \in D([0,T], E^N)$ by $\mathbf{X}_t^N(\omega) = \omega(t)$.

Note that the time marginal or *pointwise law* is given by $f_t^N = (\mathbf{X}_t^N)_\# f_I^N$. The weak Liouville equation (1) can be recovered by taking the expectation in (2). This description is called *pathwise* and the previous one *pointwise*.

3. Finally an exchangeable particle system can also be described by its empirical measure:

$$\mu_{\mathcal{X}_t^N} := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}} \in \mathcal{P}(E). \quad (3)$$

Contrary to f_t^N , the measure $\mu_{\mathcal{X}_t^N}$ is a *random object*: it can be seen as a measure-valued random variable whose law is the push-forward measure of f_t^N by the application $\boldsymbol{\mu}_N : \mathbf{x}^N \in E^N \mapsto \mu_{\mathbf{x}^N} \in \mathcal{P}(E)$. Thanks to the exchangeability, the $\mathcal{P}(E)$ -valued process $(\mu_{\mathcal{X}_t^N})_{t \geq 0}$ is a Markov process. Moreover the expectation of the empirical measure gives the first marginal of f_t^N :

$$\forall \varphi \in C_b(E), \quad \mathbb{E} \langle \mu_{\mathcal{X}_t^N}, \varphi \rangle = \langle f_t^{1,N}, \varphi \rangle.$$

It follows that from the empirical measure, it is possible to reconstruct the law of any individual particle. As we shall see later, in fact, it characterises the full N -particle distribution f_t^N . Its pathwise version is the empirical measure on the path space:

$$\mu_{\mathcal{X}_I^N} := \frac{1}{N} \sum_{i=1}^N \delta_{X_I^i} \in \mathcal{P}(D(I, E)),$$

where each particle X_I^i is seen as a $D(I, E)$ -valued process.

The mesoscopic scale.

The main concern of this review is the description of the limit dynamics when $N \rightarrow +\infty$. This will be given by a *nonlinear* object which describes the *average* behaviour of the system. Various points of view may be adopted: from the previous discussion, a natural idea is to study the limit $N \rightarrow +\infty$ of the statistical objects $f_t^{k,N}$, for all $k \in \mathbb{N}$, or $\mu_{\mathcal{X}_t^N}$. The central notion of this review is the *Propagation of Chaos* property which states that for all $t \in I$ there exists $f_t \in \mathcal{P}(E)$ such that,

$$\forall k \in \mathbb{N}, \quad f_t^{k,N} \xrightarrow[N \rightarrow +\infty]{} f_t^{\otimes k}, \quad (4)$$

for the weak convergence of probability measures and provided that this property holds true when $t = 0$. As we will see in the following, the property (4) is equivalent to

$$\mu_{\mathcal{X}_t^N} \xrightarrow[N \rightarrow +\infty]{} f_t, \quad (5)$$

for the convergence in law (note that the limit is a deterministic object). The term *propagation* introduced by Kac [104] refers to the idea that the above convergence at $t = 0$ is sufficient to prove the convergence at a later time. As we shall see in the following, the status of the time variable is a central question: one may try to quantify the convergence speed with respect to t , analyse its behaviour when $I = [0, +\infty)$ is infinite or study the more general question of the pathwise convergence of the trajectories. All these aspects will define as many notions of propagation of chaos.

In (4), the tensor product indicates that any k particles taken among the N become statistically independent when $N \rightarrow +\infty$. Any subsystem of the N -particle system thus behaves as a system of i.i.d processes with common law f_t (note that the particles are always identically distributed by the exchangeability assumption). This translates the physical idea that for large systems, the correlations between two (or more) given particles

which are due to the interactions become negligible. By looking at the whole system, only an averaged behaviour can be observed instead of the detailed correlated trajectories of each particle. This level of description is called the *mesoscopic scale* in statistical physics.

The central question is therefore the description of the limit law f_t . It turns out that in most cases, it is relatively easy to see that f_t is formally the solution of one of the following nonlinear problems.

1. The solution of a (nonlinear) Partial Differential Equation (PDE) obtained from the Liouville equation by some closure assumption. In some cases, f_t can also be seen as the law of a nonlinear Markov process in the sense of McKean, typically the law of a nonlinear SDE (these notions will be properly defined later). Just as in the classical case, the two approaches are linked by Itô's formula.
2. The solution of a (nonlinear) martingale problem. This description is more detailed than the previous one, since it gives the probability distribution on the path space $f_I \in \mathcal{P}(D(I, E))$ with time marginals $(f_t)_t$.

For each of the particle systems considered in this review, the program is thus the following:

- (1) Prove that some limit point (4) or (5) exists in a suitable topology.
- (2) Identify the limit as the solution of a nonlinear problem.
- (3) Prove that the nonlinear problem is wellposed so that the limit is uniquely defined.

Note that the three steps can be carried out in any order. In this review the main concern will be the first step, which is the core of the propagation of chaos property. This also provides an existence result for the nonlinear problem of the second step. The third step is often proved beforehand. Actually, in many cases, the nonlinear problem has its own dedicated literature; many of its properties are known and may be useful to carry out the first step.

We conclude this introductory section by a brief overview of the models studied in this work and which will be detailed in the following subsections.

Section 2.2.2 is devoted to the description of various diffusion processes, starting from the prototypical example introduced in the seminal work of McKean [125, 123]. At the microscopic scale, the particle system is defined as a system of interacting Itô processes, where the interaction depends only on (observables of) the empirical measure (3). Physically, it means that each particle interacts with a single averaged quantity in which the other particles contribute with a weight of the order $1/N$. This type of interaction is called *mean-field interaction* and the propagation of chaos is a particular instance of a *mean-field limit*. In section 2.2.3 the diffusion interaction is replaced by a jump mechanism.

Section 2.3 presents another class of models which extends the work of Kac [104] on the Boltzmann equation. The N -particle process is driven by time-discrete pairwise interactions which update the state of only two particles at each time. In classical kinetic theory, the particles are said to *collide*.

Remark 3 (Notational convention). We will adopt the following notational convention: the N -particle mean-field processes are denoted by the letter $\mathcal{X}$ and the Boltzmann N -particle processes by the letter $\mathcal{Z}$. Historically, many of the Boltzmann models that we are going to study are spatially homogeneous versions of a Boltzmann kinetic system. We thus also use the letter $\mathcal{Z}$ for kinetic systems, that is systems where each particle Z_t^i is defined by its position and its velocity, respectively denoted by the letters X_t^i and V_t^i .

2.2 Mean-field models

2.2.1 Abstract mean-field generators and mean-field limits

A *mean-field particle system* is a system of N particles characterised by a generator of the form

$$\mathcal{L}_N \varphi_N(\mathbf{x}^N) = \sum_{i=1}^N L_{\mu_{\mathbf{x}^N}} \diamond_i \varphi_N(\mathbf{x}^N), \quad (6)$$

where given a probability measure $\mu \in \mathcal{P}(E)$, L_μ is the generator of a Markov process on E . Throughout this review, the notation $L \diamond_i \varphi_N$ denotes the action of an operator L defined on (a subset of) $C_b(E)$ against the i -th variable of a function $\varphi_N \in C_b(E^N)$; in other words, $L \diamond_i \varphi_N$ is defined as the function:

$$L \diamond_i \varphi_N : (x^1, \dots, x^N) \in E^N \mapsto L[x \mapsto \varphi_N(x^1, \dots, x^{i-1}, x, x^{i+1}, \dots, x^N)](x^i) \in \mathbb{R}.$$

There are two main classes of mean-field models, depending on the form of the generator L_μ .

1. In Section 2.2.2, L_μ is the generator of a diffusion process, and the associated N -particle system is called a *McKean-Vlasov diffusion*.
2. In Section 2.2.3, L_μ is the generator of a jump process and the N -particle system is called a *mean-field jump process*. When L_μ is the sum of a pure jump generator and the generator of a deterministic flow, the process is called a mean-field Piecewise Deterministic Markov Process (PDMP for short).

It is also possible to consider *mixed processes* when L_μ is the sum of a diffusion generator and a jump generator.

It is classically assumed that the domain of the generator L_μ does not depend on μ . This domain will be denoted by $\mathcal{F} \subset C_b(E)$.

In that case, it is easy to guess the form of the associated nonlinear system obtained when $N \rightarrow +\infty$. Taking a test function of the form

$$\varphi_N(\mathbf{x}^N) = \varphi(x^1),$$

where $\varphi \in \mathcal{F}$, one obtains the one-particle Kolmogorov equation:

$$\frac{d}{dt} \langle f_t^{1,N}, \varphi \rangle = \int_{E^N} L_{\mu_{\mathbf{x}^N}} \varphi(x^1) f_t^N(d\mathbf{x}^N).$$

Note that the right-hand side depends on the N -particle distribution. As already discussed in the introduction, if the limiting system exists then, its law f_t at time $t \geq 0$ is typically obtained as the limit of the empirical measure process: when $\mathcal{X}_t^N \sim f_t^N$,

$$\mu_{\mathcal{X}_t^N} \xrightarrow{N \rightarrow +\infty} f_t.$$

This also implies $f_t^{1,N} \rightarrow f_t$. Reporting formally in the previous equation, it follows that f_t should satisfy

$$\forall \varphi \in \mathcal{F}, \quad \frac{d}{dt} \langle f_t, \varphi \rangle = \langle f_t, L_{f_t} \varphi \rangle. \quad (7)$$

This is the weak form of an equation that is called *the (nonlinear) evolution equation*. Note that the evolution equation is nonlinear due to the dependency of L on the measure argument f_t . With a slight abuse, we will often simply write $C_b(E)$ instead of $\mathcal{F}$ in the following.

This is a very analytical derivation. Its probabilistic counterpart is the following nonlinear martingale problem.

Definition 2.3 (Nonlinear mean-field martingale problem). Let $T \in (0, +\infty]$ and let us write $I = [0, T]$. A pathwise law $f_I \in \mathcal{P}(D([0, T], E))$ is said to be a solution of the nonlinear mean-field martingale problem issued from $f_0 \in \mathcal{P}(E)$ whenever for all $\varphi \in \mathcal{F}$,

$$M_t^\varphi := \varphi(\mathbf{X}_t) - \varphi(\mathbf{X}_0) - \int_0^t L_{f_s} \varphi(\mathbf{X}_s) ds, \quad (8)$$

is a f_I martingale, where $(\mathbf{X}_t)_t$ is the canonical process and for $t \geq 0$, $f_t := (\mathbf{X}_t)_\# f_I$. The natural filtration of the canonical process is denoted by $\mathcal{F}$.

Note that f_I contains much more information than the evolution equation (7) and as the notation implies, $f_t = (\mathbf{X}_t)_\# f_I \in \mathcal{P}(E)$ solves the evolution equation. If the nonlinear martingale problem is wellposed then the canonical process $(\mathbf{X}_t)_t$ is a time inhomogeneous Markov process on the probability space $(D([0, T], E), \mathcal{F}, f_I)$. This may seem a little bit abstract for now but in what follows, we will see that most often, given the usual abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$, one can define a process on Ω such that its (pathwise) law is a solution of the nonlinear martingale problem. Such process is called *nonlinear in the sense of McKean* or simply *nonlinear* for short.

2.2.2 McKean-Vlasov diffusion

Let be given two functions

$$b : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d, \quad \sigma : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathcal{M}_d(\mathbb{R}) \quad (9)$$

respectively called the *drift* vector and the *diffusion* matrix. For a fixed $\mu \in \mathcal{P}(\mathbb{R}^d)$, the following generator is the generator of a diffusion process in $\mathbb{R}^d$:

$$\forall \varphi \in C_b^2(\mathbb{R}^d), \quad L_\mu \varphi(x) := b(x, \mu) \cdot \nabla \varphi + \frac{1}{2} \sum_{i,j=1}^d a_{ij}(x, \mu) \partial_{x_i} \partial_{x_j} \varphi, \quad (10)$$

where $a(x, \mu) := \sigma(x, \mu)\sigma(x, \mu)^T$. The N -particle generator (6) associated to this class of diffusion generators defines a process called a *McKean-Vlasov diffusion process*. The associated N -particle process is governed by the following system of SDEs:

$$\forall i \in \{1, \dots, N\}, \quad dX_t^{i,N} = b(X_t^{i,N}, \mu_{X_t^N})dt + \sigma(X_t^{i,N}, \mu_{X_t})dB_t^i \quad (11)$$

where $B_t^1, \dots, B_t^N$ are N independent Brownian motions.

Remark 4 (On the assumptions on b and σ). Note that for simplicity, here, we have implicitly assumed that Equation (11) has a strong solution: it may require precise assumptions on the coefficients b and σ (typically Lipschitz assumptions as in Proposition 1 below). This will also be the case for Equation (12) below. However we stress out that these assumptions are not satisfied in many important cases. In this introductory section, such strong formulations are only used for better readability but in the following and in particular in the second part of this review, the assumptions on b and σ will play a crucial role. Example 1 below present some important models where b and σ are singular and for which Equations (11) and (12) are not defined in strong form. These models will be thoroughly reviewed in the second part.

Remark 5. Note that in the system (11) there are actually dN independent one-dimensional Brownian motions. This remark may be helpful in cases where the Brownian motions in the different directions are different. In particular, for *kinetic* particles defined by their positions and velocities, the noise is often added on the velocity variable only (this case is nevertheless covered by (11) with a block-diagonal matrix σ with a vanishing block on the position variable).

In this case, the evolution equation (7) can be written in a strong form (at least formally) and reads:

$$\partial_t f_t(x) = -\nabla_x \cdot \{b(x, f_t)f_t\} + \frac{1}{2} \sum_{i,j=1}^d \partial_{x_i} \partial_{x_j} \{a_{ij}(x, f_t)f_t\} \quad (12)$$

This is a nonlinear Fokker-Planck equation which is used in many important modelling problems (see Example 1). This equation was obtained (formally) previously using only the generators when $N \rightarrow +\infty$. Here, there is an alternative way to derive the limiting system: looking at the SDE system (11), the empirical measure can be formally replaced by its expected limit f_t . Since all the particles are exchangeable, this can be done in any of the N equations. The result is a process $(\bar{X}_t)_t$ which solves the SDE:

$$d\bar{X}_t = b(\bar{X}_t, f_t)dt + \sigma(\bar{X}_t, f_t)dB_t. \quad (13)$$

where B_t is a Brownian motion and $\bar{X}_0 \sim f_0$. Moreover, since for all i , X_t^i has law $f_t^{1,N}$ and since it is expected that $f_t^{1,N} \rightarrow f_t$, the process $\bar{X}_t$ and the distribution f_t should be linked by the relation

$$f_t = \text{Law}(\bar{X}_t).$$

The dependency on its law of the solution of a SDE is a special case of what is called a nonlinear process in the sense of McKean. It would now be desirable to prove that the

process (13) is well defined or (equivalently) that the PDE (12) or the martingale problem (8) are wellposed. The following result gives the reference framework in which all these objects are well defined.

Proposition 1. *Let us assume that the functions b and σ are globally Lipschitz: there exists $C > 0$ such that for all $x, y \in \mathbb{R}^d$ and for all $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ it holds that:*

$$|b(x, \mu) - b(y, \nu)| + |\sigma(x, \mu) - \sigma(y, \nu)| \leq C(|x - y| + W_2(\mu, \nu)),$$

where W_2 denotes the Wasserstein-2 distance (see Definition 3.1). Assume that $f_0 \in \mathcal{P}_2(\mathbb{R}^d)$. Then for any $T > 0$ the SDE (13) has a unique strong solution on $[0, T]$ and consequently, its law is the unique weak solution to the Fokker-Planck equation (12) and the unique solution to the nonlinear martingale problem (8).

The proof of this proposition is fairly classical. In some special linear cases (see below), it can be found in [125, Section 3], [151, Theorem 1.1] or [127, Theorem 2.2]. For the most general case which includes the above proposition, we refer to [35, Theorem 1.7]. The proof is based on a fixed point argument that is sketched below.

Proof. Let us define the map:

$$\Psi : \mathcal{P}_2(C([0, T], \mathbb{R}^d)) \rightarrow \mathcal{P}_2(C([0, T], \mathbb{R}^d)), \quad m \mapsto \Psi(m),$$

where for $m \in \mathcal{P}_2(C([0, T], \mathbb{R}^d))$, $\Psi(m)$ is the law (on the path space) of the solution $(X_t^m)_{0 \leq t \leq T}$ of the following SDE:

$$dX_t^m = b(X_t^m, m_t)dt + \sigma(X_t^m, m_t)dB_t.$$

Note that the map $t \in [0, T] \mapsto m_t \in \mathcal{P}_2(\mathbb{R}^d)$ is continuous for the W_2 -distance where m_t is the time marginal of m . The goal is to prove that Ψ admits a unique fixed point. Let $m, m' \in \mathcal{P}_2(C([0, T], \mathbb{R}^d))$ and let $t \in [0, T]$. Then by the Burkholder-Davis-Gundy inequality (see Proposition 16) and the Lipschitz assumptions on b and σ , one can prove that there exists a constant $C > 0$ such that:

$$\begin{aligned} \mathbb{E} \left[\sup_{0 \leq s \leq t} |X_s^m - X_s^{m'}|^2 \right] \\ \leq CT \left(\int_0^t \mathbb{E} \left[\sup_{0 \leq r \leq s} |X_r^m - X_r^{m'}|^2 \right] ds + \int_0^t W_2^2(m_s, m'_s) ds \right). \end{aligned}$$

A similar computation will be detailed in the proof of Theorem II-3.1. By Gronwall lemma, we obtain that for a constant $C(T)$ it holds that:

$$\mathbb{E} \left[\sup_{0 \leq s \leq t} |X_s^m - X_s^{m'}|^2 \right] \leq C(T) \int_0^t W_2^2(m_s, m'_s) ds.$$

Let us denote by $W_{d_{C_t}, 2}$ the Wasserstein-2 distance on the space of probability measures on a path space of the form $\mathcal{C}_t = C([0, t], \mathbb{R}^d)$ endowed with the uniform topology (see Definition 3.5) for a given $t \in [0, T]$. For any $t \in [0, T]$, we also write $m_{[0, t]} \in \mathcal{P}_2(C([0, t], \mathbb{R}^d))$ for

the restriction of m on $[0, t]$. In particular, note that if $s \leq t$ then $W_{d_{c_s}, 2}(m_{[0, s]}, m'_{[0, s]}) \leq W_{d_{c_t}, 2}(m_{[0, t]}, m'_{[0, t]})$. Then by definition of Ψ and $W_{d_{c_t}, 2}$, we conclude that:

$$\begin{aligned} W_{d_{c_t}, 2}^2(\Psi(m_{[0, t]}), \Psi(m'_{[0, t]})) &\leq C(T) \int_0^t W_2^2(m_s, m'_s) ds \\ &\leq C(T) \int_0^t W_{d_{c_s}, 2}^2(m_{[0, s]}, m'_{[0, s]}) ds. \end{aligned}$$

By iterating this inequality, the k -th iterate Ψ^k of Ψ satisfies:

$$\begin{aligned} W_{d_{c_T}, 2}^2(\Psi^k(m), \Psi^k(m')) &\leq c(T)^k \int_0^T \frac{s^k}{(k-1)!} W_{d_{c_s}, 2}^2(m_{[0, s]}, m'_{[0, s]}) ds \\ &\leq \frac{(c(T)T)^k}{k!} W_{d_{c_T}, 2}^2(m, m'), \end{aligned}$$

from which it can be seen that Ψ^k is a contraction and thus admits a unique fixed point for k large enough. $\square$

Example 1. Depending on the form of the drift and diffusion coefficients, the McKean-Vlasov diffusion can be used in a wide range of modelling problems. Some examples are gathered below and many other will be given in Section II-5.

1. The first case is obtained when b and σ depend linearly on the measure argument. Namely, for $n, m \in \mathbb{N}$, let us consider two functions

$$K_1 : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^m, \quad K_2 : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^n,$$

and let us take

$$b(x, \mu) = \tilde{b}(x, K_1 \star \mu(x)), \quad \sigma(x, \mu) = \tilde{\sigma}(x, K_2 \star \mu(x)),$$

where $\tilde{b} : \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^d$ and $\tilde{\sigma} : \mathbb{R}^d \times \mathbb{R}^n \rightarrow \mathcal{M}_d(\mathbb{R})$. When K_1, K_2 and $\tilde{b}, \tilde{\sigma}$ are Lipschitz and bounded, the propagation of chaos result is the given by McKean's theorem (Theorem II-3.1).

In many applications, σ is a constant diffusion matrix, $K_1(x, y) \equiv K(y - x)$ for a fixed symmetric radial kernel $K : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $b(x, \mu) = K \star \mu(x)$. The case where K has a singularity is much more delicate (see Section II-3.1.2, Section II-3.4, Section II-3.2.2) but contains many important cases. For instance, in fluid dynamics, when K is the Biot-Savart kernel $K(x) = x^\perp/|x|^2$ in dimension $d = 2$ (defining $(x_1, x_2)^\perp = (-x_2, x_1)$) and $\sigma(x, \mu) \equiv \sqrt{2\sigma}I_2$ for a fixed $\sigma > 0$, the limit Fokker-Planck equation reads:

$$\partial_t f_t + \nabla \cdot (f_t K \star f_t) = \sigma \Delta f_t, \tag{14}$$

By invariance by translation, the quantity $\omega = f_t - 1$ is the solution of the famous *vorticity equation* which can be shown to be equivalent to the 2D incompressible Navier-Stokes system (see [99]).

In biology, still in dimension $d = 2$ but with $K(x) = x/|x|^2$, the equation (14) is an example of the Patlak-Keller-Segel model for chemotaxis (see [28]). Other examples in mathematical physics and mathematical biology are presented in Section II-5.1.

Another important model of this form is the Kuramoto model obtained when $E = \mathbb{R}$ and $K(\theta) = K_0 \sin(\theta)$ for a given $K_0 > 0$. In this case, the particles model the frequencies of a system of *oscillators* which tends to synchronize, see for instance [1, 120, 13, 14] and Section II-5.2.1.

2. The case of *gradient systems* is a sub-case of the previous one when $\sigma(x, \mu) \equiv \sigma I_d$ for a constant $\sigma > 0$ and

$$b(x, \mu) = -\nabla V(x) - \int_{\mathbb{R}^d} \nabla W(x - y) \mu(dy), \quad (15)$$

where V, W are two symmetric potentials on $\mathbb{R}^d$ respectively called the confinement potential and the interaction potential. The limit Fokker-Planck equation

$$\partial f_t = \frac{\sigma^2}{2} \Delta f_t + \nabla \cdot (f_t \nabla (V + W \star f_t)),$$

is called the granular-media equation and will be studied in Section II-3.1.3. An important issue is the long-time behaviour of gradient systems which is often studied under convexity assumptions on the potentials (see Section II-3.1.3, Section II-3.2.1, Section II-3.2.3).

3. A kinetic particle $Z_t^{i,N} = (X_t^{i,N}, V_t^{i,N}) \in \mathbb{R}^d \times \mathbb{R}^d$ is a particle defined by two arguments, its position $X_t^{i,N}$ and its velocity $V_t^{i,N}$ defined as the time derivative of the position. The evolution of a system of kinetic particles is usually governed by Newton's laws of motion. In a random setting, the typical system of SDEs is thus the following: for $i \in \{1, \dots, N\}$,

$$\begin{cases} dX_t^{i,N} &= V_t^{i,N} dt \\ dV_t^{i,N} &= F(X_t^{i,N}, V_t^{i,N}, \mu_{\mathcal{X}_t^N}) dt + \sigma(X_t^i, V_t^{i,N}, \mu_{\mathcal{X}_t^N}) dB_t^i, \end{cases}$$

where $F : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathcal{M}_d(\mathbb{R})$. Note that it is often assumed that the force field induced by the interactions between the particles depends only on their positions, which is why we have written

$$\mu_{\mathcal{X}_t^N} := \frac{1}{N} \sum_{i=1}^N \delta_{X_t^{i,N}} \in \mathcal{P}(\mathbb{R}^d)$$

instead of $\mu_{Z_t^N} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$. This special case of the McKean-Vlasov diffusion in $E = \mathbb{R}^d \times \mathbb{R}^d$ is also often called a *second order system* by opposition to the *first order systems* when $E = \mathbb{R}^d$. Note that when $\sigma = 0$, the limit Equation (12) is the renowned Vlasov equation which is historically one of the first and most important models in plasma physics and celestial mechanics. In the following, we

will nevertheless most often consider stochastic models although some of the results still apply in this deterministic case (in particular the important Theorem II-3.1). For a detailed account of the Vlasov equation in this context, we refer to the review article [97]. Several examples of stochastic kinetic particle systems are given in Section II-5.2.2 which deals with swarming models. For instance, the (stochastic) Cucker-Smale model [51, 93, 38, 42] describes a system of bird-like particles which interact by aligning their velocities to the ones of their neighbours:

$$dV_t^i = \frac{1}{N} \sum_{j \neq i} K(|X_t^j - X_t^i|)(V_t^j - V_t^i)dt + \sigma dB_t^i,$$

where $\sigma \geq 0$ is a noise parameter and $K : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is a smooth nonnegative function vanishing at infinity which models the vision of the particles. Other classical models of this form include the attraction-repulsion models [70, 41] or the stochastic Vicsek models [62, 60, 61].

4. The general case (9), where b and σ have a possibly nonlinear dependence on μ can be extended to even more general cases. A simple extension is the case of time-dependent functions b and σ . They may also be random themselves and the x and μ arguments may be replaced respectively by a full trajectory on $C([0, T], \mathbb{R}^d)$ and a pathwise probability distribution on $\mathcal{P}(C([0, T], \mathbb{R}^d))$. The most general setting is thus:

$$b : [0, T] \times \Omega \times C([0, T], \mathbb{R}^d) \times \mathcal{P}(C([0, T], \mathbb{R}^d)) \rightarrow \mathbb{R}^d.$$

Such cases may be very difficult to handle but have recently been used in the theory of mean-field games [30, 108, 36, 37, 29]. Under strong Lipschitz assumptions (in the appropriate topology), some very general results can be obtained by a relatively simple adaptation of the proofs valid in the linear case (see Section II-3.6.1). For more general systems, we will only briefly mention some existing results in Section II-3.6.2 and Section II-3.6.3.

Remark 6 (Martingale measures). Starting from an arbitrary nonlinear Fokker-Planck with operator (10), one may wonder if it can always be written (at least formally) as the limit of a particle system. The McKean-Vlasov diffusion positively answers when the diffusion matrix in the Fokker-Planck equation is of the form $a(x, \mu) = \sigma(x, \mu)\sigma(x, \mu)^T$. For more general matrices a , the situation is more complicated. For instance the Landau equation would correspond to a matrix of the form $a(x, \mu) = \int_{\mathbb{R}^d} \sigma(x, y)\sigma(x, y)^T \mu(dy)$. In this case, the problem has been studied with a stochastic point of view in [82] and later in [128] and [77] where an explicit approximating particle system is given (see also Section II-5.1.4). The N -particle system is characterised as the solution of a system of N SDEs similar to (11) but where the N Brownian motions are replaced by N *martingale measures* with intensity $\mu_{\mathcal{X}_t^N}(dy) \otimes dt$. The notion of martingale measure which originates in the Stochastic PDE literature is studied for instance in [74]. Except for the cases investigated in the aforementioned works and although it seems to generalise many of the models presented in this review, there is, to the best of our knowledge, no general theory of propagation of chaos for particle systems driven by martingale measures.

2.2.3 Mean-field jump processes and PDMPs

In this section E is any Polish space. Let us be given a family of probability measures called the *jump measures*:

$$P : E \times \mathcal{P}(E) \rightarrow \mathcal{P}(E), \quad (x, \mu) \mapsto P_\mu(x, dy),$$

and a positive function, called the *jump rate*:

$$\lambda : E \times \mathcal{P}(E) \rightarrow \mathbb{R}_+, \quad (x, \mu) \mapsto \lambda(x, \mu),$$

For a given $\mu \in \mathcal{P}(E)$ the following generator is the generator of a *pure jump process*:

$$L_\mu \varphi(x) = \lambda(x, \mu) \int_E \{\varphi(y) - \varphi(x)\} P_\mu(x, dy).$$

We will also consider the case of a PDMP when

$$L_\mu \varphi(x) = a \cdot \nabla \varphi(x) + \lambda(x, \mu) \int_E \{\varphi(y) - \varphi(x)\} P_\mu(x, dy), \quad (16)$$

where, with a slight abuse of notation, $a \cdot \nabla$ denotes the transport flow associated to a function $a : E \rightarrow E$. Using the family of generators (16), the N -particle system with mean-field generator (6) can be constructed as follows (see [66, Theorem 13.2.5]).

- To each particle $i \in \{1, \dots, N\}$ is attached a Poisson clock with jump rate $\lambda(\mu_{\mathcal{X}_t^N}, X_t^i)$. The jump times of particle i are denoted by $(T_n^i)_n$. The jump times can be constructed recursively by taking $T_0^i = 0$ and

$$T_{n+1}^i = \inf \left\{ t \geq T_n^i, \int_{T_n^i}^t \lambda(\mu_{\mathcal{X}_s^N}, X_s^i) ds \geq E_n^i \right\},$$

where the random variables $(E_n^i)_{n \geq 0}$ are i.i.d. and distributed according to an exponential law with parameter 1.

- Between two jump times, the motion of a particle is purely deterministic:

$$\forall n \in \mathbb{N}, \quad \forall t \in [T_n^i, T_{n+1}^i), \quad dX_t^i = a(X_t^i) dt, \quad (17)$$

- At each time T_n^i , a new state is sampled from the jump measure on E :

$$X_{T_n^i}^i \sim P_{\mu_{\mathcal{X}_{T_n^i-}^N}} \left(X_{T_n^i-}^i, dy \right) \in \mathcal{P}(E). \quad (18)$$

One expects that in the limit $N \rightarrow +\infty$, the law f_t of a particle will satisfy the evolution equation (7) which, in this case, reads:

$$\frac{d}{dt} \langle f_t, \varphi \rangle = \langle f_t, a \cdot \nabla_x \varphi \rangle + \iint_{E \times E} \lambda(x, f_t) \{\varphi(y) - \varphi(x)\} P_{f_t}(x, dy) f_t(dx), \quad (19)$$

for all $\varphi \in C_b(E)$. Two important cases are given in the following examples.

Example 2 (Nanbu particle system). Let us take $a = 0$ and $\lambda = 1$ for simplicity. When the jump measure is linear in μ , i.e. is of the form:

$$P_\mu(x, dy) = \int_{z \in E} \Gamma^{(1)}(x, z, dy) \mu(dz),$$

where $\Gamma^{(1)} : E \times E \rightarrow \mathcal{P}(E)$, then the mean-field generator (6) describes a N -particle system where at each jump, a particle with state x chooses uniformly another particle, say which has a state z , and sample a new state according to the law $\Gamma^{(1)}(x, z, dy)$. In [89], this particle system is called a *Nanbu particle system* in honour of Nanbu who introduced a similar system in [135] and used it as an approximation scheme for the Boltzmann equation of rarefied gas dynamics (38). This equation will be described more thoroughly in Section 2.3.3. When $\Gamma^{(1)}$ is an abstract law, the associated mean-field jump particle system generalizes the one introduced by Nanbu and the limit equation is the following *general Boltzmann equation* (written in weak form):

$$\frac{d}{dt} \langle f_t, \varphi \rangle = \int_{E \times E \times E} \varphi(y) \Gamma^{(1)}(x, z, dy) f_t(dx) f_t(dz) - \langle f_t, \varphi \rangle,$$

for all $\varphi \in C_b(E)$. A more classical derivation of the general Boltzmann equation will be given in Section 2.3 and the subsequent Example 7 provides an alternative point of view on the Nanbu particle system.

Example 3 (BGK type model). In kinetic theory the state space is $E = \mathbb{R}^d \times \mathbb{R}^d$ and the N particles are given by $Z_t^i = (X_t^i, V_t^i)$ with X_t^i the position and V_t^i the velocity of particle i at time t . Without external force, it is natural to expect that the particles evolve deterministically and continuously between two jumps as

$$dX_t^i = V_t^i dt, \quad dV_t^i = 0.$$

Moreover, the post-jump distribution and the jump rate often do not depend specifically on the pre-jump velocity of the jumping particle but only on its position and on the distribution of particles. Thus we take:

$$P_\mu((x, v), dx', dv') = \delta_x(dx') \otimes \mathcal{M}_{\mu, x}(v') dv',$$

where given $\mu \in \mathcal{P}(E)$ and $x \in \mathbb{R}^d$, $\mathcal{M}_{\mu, x}$ is a probability density function. In this case, Equation (7) becomes:

$$\frac{d}{dt} \langle f_t, \varphi \rangle = \langle f_t, v \cdot \nabla_x \varphi \rangle + \iint_{\mathbb{R}^d \times \mathbb{R}^d} \lambda(f_t, x) \{ \varphi(x, v') - \varphi(x, v) \} \mathcal{M}_{f_t, x}(v') dv' f_t(dx, dv),$$

and its strong form reads:

$$\partial_t f_t(x, v) + v \cdot \nabla_x f_t(x, v) = \lambda(x, f_t) \left(\rho_{f_t}(x) \mathcal{M}_{f_t, x}(v) - f_t(x, v) \right),$$

where the spatial density of the particles at time t is defined by:

$$\rho_{f_t}(x) := \int_{\mathbb{R}^d} f_t(x, v) dv.$$

When $\mathcal{M}_{f_t,x}$ is the Maxwellian distribution

$$\mathcal{M}_{f_t,x}(v) = \frac{\rho_f}{(2\pi T)^{d/2}} \exp\left(-\frac{|v-u|^2}{2T}\right),$$

with $(\rho_f u, \rho_f |u|^2 + \rho_f T) = \int_{\mathbb{R}^d} (v, |v|^2) f_t(x, v) dv$, then this equation is called the Bhatnagar-Gross-Krook (BGK) equation [15]. It is used in mathematical physics as a simplified model of rarefied gas dynamics (for a detailed account of the subject, we refer the interested reader to the reviews [58] and [159] or to the book [45]).

In this review, we found useful to distinguish a class of mean-field jump models that we call *parametric models* which are defined by a jump measure of the form

$$P_\mu(x, dy) = \left(\psi(x, \mu, \cdot)_\# \nu\right)(dy),$$

where $\nu \in \mathcal{P}(\Theta)$ is a probability measure on a fixed parameter space Θ and

$$\psi : E \times \mathcal{P}(E) \times \Theta \rightarrow E.$$

In this case, for all test function $\varphi \in C_b(E)$,

$$\int_E \varphi(y) P_\mu(x, dy) = \int_\Theta \varphi(\psi(x, \mu, \theta)) \nu(d\theta).$$

The N -particle process associated to a parametric model admits a SDE representation using the formalism of Poisson random measure which is briefly recalled in Appendix A.7.

Example 4 (SDE representation for parametric models). Let us assume for all $\theta \in \Theta$, the function $\psi(\cdot, \cdot, \theta) : E \times \mathcal{P}(E) \rightarrow E$ is Lipschitz for the distance on E and the Wasserstein-1 distance on $\mathcal{P}(E)$, with a Lipschitz constant $L(\theta) > 0$ and a function $L \in L^1_\nu(\Theta)$. This (classical) hypothesis will ensure the wellposedness of the SDE representations of both the particle system and its nonlinear limit, see [4, Section 3.1] and [87, Theorem 1.2 and Theorem 2.1].

To each particle $i \in \{1, \dots, N\}$ is attached an independent Poisson random measures $\mathcal{N}^i(ds, du, d\theta)$ on $[0, +\infty) \times [0, +\infty) \times \Theta$ with intensity measure $ds \otimes du \otimes \nu(d\theta)$ where dt and du denote the Lebesgue measure. The N independent random measures $\mathcal{N}^i$ play a comparable role to the N independent Brownian motions which define a McKean-Vlasov diffusion in (11). In the present case, the mean-field jump N -particle process is the solution of the following system of SDEs driven by the measures $\mathcal{N}^i$

$$\begin{aligned} X_t^i &= X_0^i + \int_0^t a(X_s^i) ds \\ &+ \int_0^t \int_0^{+\infty} \int_\Theta \left\{ \psi\left(X_{s-}^i, \mu_{\mathcal{X}_{s-}^N}, \theta\right) - X_{s-}^i \right\} \mathbb{1}_{(0, \lambda(X_{s-}^i, \mu_{\mathcal{X}_{s-}^N}))}(u) \mathcal{N}^i(ds, du, d\theta). \end{aligned} \quad (20)$$

Remark 7. Parametric models are reminiscent but different from the so-called *disordered models* where random i.i.d. parameters θ^i (called *disorder*) are attached to each particle.

For these models, various notions of propagation of chaos can be defined, for instance one can prove a classical propagation of chaos result for each realization of the disorder (*quenched* behavior) or the convergence of the joint empirical measure of the processes and the disorder (*quenched averaged* behavior). These problems are of particular interest from the point of view of Large Deviations. The main case of application is the renowned Kuramoto model [1, 120, 52] which will be briefly discussed in Section II-5.2.1.

In neurosciences, the variable X_t^i represents the membrane potential of a neuron indexed by i at time t and the Poisson random measures model the interactions between the neurons due to the chemical synapses. A random jump is called a *spike*. In the model introduced by [79], the effect of the spikes is to reset the potential of the membrane to a fixed value, fixed to 0. In Equation (20) this corresponds to the simple case where $X_t^i \in \mathbb{R}_+$ and $\psi \equiv 0$. Note that in this particular case there is no need to consider a parameter space Θ and $\mathcal{N}^i$ is a Poisson random measures on $[0, +\infty) \times [0, +\infty)$ only. Note that in [79], the deterministic drift $a \equiv a(x, \mu)$ also depends on the empirical measure of the system: it models the effect of electrical synapses which tends to relax the membrane potential of the neurons towards the average potential of the system. An additional interaction mechanism is described in the subsequent example.

Example 5 (Simultaneous jumps). The neurons models [57, 79, 4] extend the (parametric) mean-field jump model (20) to allow *simultaneous jumps* at each jump time T_n^i . It models the effect that at each spiking event of a neuron i , the membrane potential of all the other neurons $j \neq i$ is also increased by a small amplitude.

In the parametric setting with $E = \mathbb{R}^d$ (or more generally when E has a vector space structure), the mean-field jump model with simultaneous jumps is defined by the following objects.

- The jump rate function:

$$\lambda : (x, \mu) \in E \times \mathcal{P}(E) \mapsto \lambda(x, \mu) \in \mathbb{R}_+.$$

- A symmetric probability measure ν_N on the N -fold product of the parameter space Θ^N . We also assume that there exists a symmetric probability measure on $\Theta^{\mathbb{N}}$ such that ν_N coincides with the projection of ν on the first N coordinates. This assumption is natural to be able to take the limit $N \rightarrow +\infty$. In [4], the parameter space is $\Theta = [0, 1]$.
- The main jump measure

$$P_\mu(x, dy) = \left(\psi(x, \mu, \cdot)_{\#} \nu \right)(dy),$$

where

$$\psi : E \times \mathcal{P}(E) \times \Theta \rightarrow E, \quad (x, \mu, \theta) \mapsto x + \alpha(x, \mu, \theta),$$

and $\alpha : E \times \mathcal{P}(E) \times \Theta \rightarrow E$ is the jump amplitude.

- The collateral jump measures

$$\tilde{P}_\mu^N(x, z, dy) = \left(\tilde{\psi}^N(x, z, \mu, \cdot)_{\#} \nu_2 \right)(dy),$$

where

$$\tilde{\psi}^N : E \times E \times \mathcal{P}(E) \times \Theta^2 \rightarrow E, \quad (x, z, \mu, \theta_1, \theta_2) \mapsto x + \frac{\tilde{\alpha}(x, z, \mu, \theta_1, \theta_2)}{N},$$

and $\tilde{\alpha} : E \times E \times \mathcal{P}(E) \times \Theta^2 \rightarrow E$ is the collateral jump amplitude. It satisfies $\tilde{\alpha}(x, x, \mu, \theta_1, \theta_2) = 0$ for all $x \in E$, $\mu \in \mathcal{P}(E)$ and $\theta_1, \theta_2 \in \Theta$. In [79], the amplitude is fixed $\tilde{\alpha}(x, z, \mu, \theta_1, \theta_2) \equiv 1$ for $x \neq z$.

The N -particle process can be defined as before by an algorithmic description. At each time T_n^i , a parameter $\theta \sim \nu_N$ is drawn and then the state of particle i is updated by adding the jump amplitude

$$\alpha \left(X_{T_n^i}^i, \mu_{\mathcal{X}_{T_n^i}^N}, \theta_i \right).$$

But in this case, at time T_n^i , all the other particles $j \neq i$ also jumps by the amplitude

$$\frac{\tilde{\alpha} \left(X_{T_n^i}^j, X_{T_n^i}^i, \mu_{\mathcal{X}_{T_n^i}^N}, \theta_j, \theta_i \right)}{N}.$$

When the parameters $\alpha, \tilde{\alpha}$ satisfy the Lipschitz integrability conditions of [4, Section 3.1], a SDE representation of the particle system can also be given. As before, let $\mathcal{N}^i(ds, du, d\theta)$ be a set of N independent Poisson random measures on $[0, +\infty) \times [0, +\infty) \times \Theta^N$ with intensity $ds \otimes du \otimes \nu$, where ds and du denote the Lebesgue measure. The SDE representation of the N -particle system is given by the following system of SDEs:

$$\begin{aligned} X_t^i &= X_0^i + \int_0^t a(X_s^i) ds \\ &+ \int_0^t \int_0^{+\infty} \int_{\Theta^N} \left\{ \psi \left(X_{s-}^i, \mu_{\mathcal{X}_{s-}^N}, \theta_i \right) - X_{s-}^i \right\} \mathbb{1}_{(0, \lambda(X_{s-}^i, \mu_{\mathcal{X}_{s-}^N}))}](u) \mathcal{N}^i(ds, du, d\theta) \\ &+ \sum_{j \neq i} \int_0^t \int_0^{+\infty} \int_{\Theta^N} \left\{ \tilde{\psi}^N \left(X_{s-}^i, X_{s-}^j, \mu_{\mathcal{X}_{s-}^N}, \theta_i, \theta_j \right) - X_{s-}^i \right\} \times \\ &\quad \times \mathbb{1}_{(0, \lambda(X_{s-}^j, \mu_{\mathcal{X}_{s-}^N}))}](u) \mathcal{N}^j(ds, du, d\theta). \end{aligned} \quad (21)$$

Compared to the previous framework, in addition to the main jump operator (16), each particle is also subject to the *collateral jump generator* defined for all $\varphi \in C_b(E)$ and $x \in E$ by:

$$\tilde{L}_\mu^N \varphi(x) := N \int \int_{E \times E} \lambda(z, \mu) \{ \varphi(y) - \varphi(x) \} \tilde{P}_\mu^N(x, z, dy) \mu(dz). \quad (22)$$

Note that this generator depends on N but it satisfies the weak limit: for all $\varphi \in C_b^1(E)$, $x \in E$ and $\mu \in \mathcal{P}(E)$,

$$\tilde{L}_\mu^N \varphi(x) \xrightarrow{N \rightarrow +\infty} \tilde{L}_\mu \varphi(x) := \iint_{E \times \Theta^2} \lambda(z, \mu) \tilde{\alpha}(x, z, \mu, \theta_1, \theta_2) \cdot \nabla \varphi(x) \mu(dz) \nu_2(d\theta_1, d\theta_2).$$

The N -particle system is thus defined by the mean-field generator (6) which takes the form:

$$\forall \varphi_N \in C_b(E^N), \quad \mathcal{L}_N \varphi_N := \sum_{i=1}^N \left\{ L_\mu \diamond_i \varphi_N + \tilde{L}_\mu^N \diamond_i \varphi_N \right\}.$$

In the limit $N \rightarrow +\infty$, the nonlinear evolution equation (7) is expected to become:

$$\forall \varphi \in C_b^1(E), \quad \frac{d}{dt} \langle f_t, \varphi \rangle = \langle f_t, L_{f_t} \varphi \rangle + \langle f_t, \tilde{L}_{f_t} \varphi \rangle.$$

Note that this is the equation satisfied by the law of the solution of the following nonlinear SDE:

$$\begin{aligned} \bar{X}_t &= \bar{X}_0 + \int_0^t a(\bar{X}_s) ds \\ &+ \int_0^t \int_0^{+\infty} \int_{\Theta^{\mathbb{N}}} \alpha(\bar{X}_{s-}, f_s, \theta_1) \mathbb{1}_{(0, \lambda(\bar{X}_{s-}, f_s)]}(u) \mathcal{N}(ds, du, d\theta) \\ &+ \int_0^t \iint_{E \times \Theta^2} \lambda(z, f_s) \tilde{\alpha}(\bar{X}_s, z, f_s, \theta_1, \theta_2) f_s(dz) \nu_2(d\theta_1, d\theta_2) ds, \end{aligned} \quad (23)$$

with $\text{Law}(\bar{X}_s) = f_s$. In the last equation, $\mathcal{N}(ds, du, d\theta)$ is a Poisson random measure on $[0, +\infty) \times [0, +\infty) \times \Theta^{\mathbb{N}}$ with intensity $ds \otimes du \otimes \nu(d\theta)$ where ds and du denote the Lebesgue measure.

Mean-field jump processes and PDMPs are not so common in the literature compared to the McKean-Vlasov diffusion models or the Boltzmann models (Section 2.3). The Nanbu particle system serves as a simplified Boltzmann model (see Example 7). Mean-field jump processes can also be used as an approximation of a McKean-Vlasov diffusion. Since the dynamics only relies on a sampling mechanism on the state space E , compared to diffusion processes, it allows more flexibility and avoids some technicalities for instance when E has a more complex geometrical structure, typically when E is a manifold. In applications, mean-field jump processes model a motion called *run and tumble* which is classical in the study of the dynamics of populations of bacteria. As already mentioned, Example 5 corresponds to a toy example of neuron model. More realistic examples often consider a combination of (simultaneous) jumps and a diffusive behaviour, see [4] and the references therein. The nonlinear martingale problem associated to mixed jump-diffusion models is studied in [88] where the wellposedness is proved under classical Lipschitz and boundedness assumptions on the parameters (see also Section II-3.3). Other neuron models based on a mean-field jump process will be described in Section II-5.2.3.

2.3 Boltzmann models

2.3.1 General form

Given a Polish space E , a Boltzmann model is a N -particle system with an infinitesimal generator acting on $\varphi_N \in C_b(E^N)$ of the form:

$$\mathcal{L}_N \varphi_N = \sum_{i=1}^N L^{(1)} \diamond_i \varphi_N + \frac{1}{N} \sum_{i < j} L^{(2)} \diamond_{ij} \varphi_N, \quad (24)$$

where $\varphi_N \equiv \varphi_N(z^1, \dots, z^N)$ is a test function on the product space E^N . The operator $L^{(2)}$ acts on two-variable test functions and stands for binary interactions between particles. The operator $L^{(1)}$ acts on one-variable test functions and describes the individual flow of each particle (and possibly the boundary conditions). More explicitly, let us recall the notations, for $(z^1, \dots, z^N) \in E^N$ and $i < j$,

$$L^{(1)} \diamond_i \varphi_N(z^1, \dots, z^N) = L^{(1)}[u \mapsto \varphi_N(z^1, \dots, z^{i-1}, u, z^{i+1}, \dots, z^N)](z^i)$$

and

$$\begin{aligned} L^{(2)} \diamond_{ij} \varphi_N(z^1, \dots, z^N) \\ = L^{(2)}[(u, v) \mapsto \varphi_N(z^1, \dots, z^{i-1}, u, z^{i+1}, \dots, z^{j-1}, v, z^{j+1}, \dots, z^N)](z^i, z^j). \end{aligned}$$

These models are called *Boltzmann models* in reference to the famous Boltzmann equation of rarefied gas dynamics which is a fundamental equation for mathematicians, physicists and philosophers. It will be explained at the end of this section (see Equation (38)) how it can be obtained as the limit of a general particle system of the form (24). The specificity of Boltzmann models is that the particles interact only at random times by pair and not individually with an average of all the other particles as in mean-field models. In full generality, the state space E is an abstract space. In classical kinetic theory, $E = \mathbb{R}^d \times \mathbb{R}^d$ is the phase space of positions and velocities and two particles interact when they are close enough: they are said to *collide* and by analogy, we will keep this terminology to refer to an interaction between two particles even in an abstract space. In addition to these pairwise interactions, each particle is also subject to an individual flow prescribed by the operator $L^{(1)}$. Typical examples in kinetic theory include

- (Free transport) $L^{(1)}\varphi(x, v) = v \cdot \nabla_x \varphi$,
- (Space diffusion) $L^{(1)}\varphi(x, v) = \Delta_x \varphi$.
- (Velocity diffusion) $L^{(1)}\varphi(x, v) = \Delta_v \varphi$.

When two particles collide, the effect of the collision is prescribed by the operator $L^{(2)}$. In kinetic theory, this operator acts on the velocity variable only but in full generality, in an abstract space E , it will be assumed to satisfy the following assumptions.

Assumption 1. *The operator $L^{(2)}$ satisfies the following properties.*

(1) The domain of the operator $L^{(2)}$ is $C_b(E^2)$.

(2) There exist a continuous map called the post-collisional distribution

$$\Gamma^{(2)} : (z_1, z_2) \in E \times E \mapsto \Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2) \in \mathcal{P}(E \times E),$$

and a symmetric function called the collision rate

$$\lambda : (z_1, z_2) \in E \times E \mapsto \lambda(z_1, z_2) \in \mathbb{R}_+,$$

such that for all $\varphi_2 \in C_b(E^2)$ and all $z_1, z_2 \in E$,

$$L^{(2)}\varphi_2(z_1, z_2) = \lambda(z_1, z_2) \iint_{E \times E} \{\varphi_2(z'_1, z'_2) - \varphi_2(z_1, z_2)\} \Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2). \quad (25)$$

(3) For all $z_1, z_2 \in E$, the post-collisional distribution is symmetric in the sense that

$$\Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2) = \Gamma^{(2)}(z_2, z_1, dz'_2, dz'_1). \quad (26)$$

(4) The function λ is measurable on $\{(z_1, z_2) \in E^2, z_1 \neq z_2\}$ and for all $z \in E$, $\lambda(z, z) = 0$.

Remark 8. The assumption that λ is a (measurable) function prevents from considering the true classical Boltzmann inhomogeneous case in kinetic theory $\lambda(z_1, z_2) = \delta_{x_1=x_2}$ (that is, two particles collide when they are exactly at the same position), which is beyond the scope of this review (see however Example 15). The collision rate is often assumed to be uniformly bounded

$$\sup_{z_1, z_2 \in E} \lambda(z_1, z_2) \leq \Lambda < \infty.$$

This *cutoff* assumption is unfortunately not physically relevant for many models where an infinite number of collisions may happen in finite time.

Note that when E is a locally compact Polish space, the Riesz-Markov-Kakutani theorem states that any linear operator on the space of two-variable test functions in $C_c(E \times E)$ can be written in the form (25). The third assumption ensures that the law f_t^N defined by the backward Kolmogorov equation remains symmetric for all time provided that f_0^N is symmetric. This follows from the observation that under (26), the action of any transposition $\tau \in \mathfrak{S}_N$ on $C_b(E^N)$ commutes with $\mathcal{L}_N$:

$$\tau^{-1} \mathcal{L}_N \tau = \mathcal{L}_N.$$

Example 6 (Jump amplitude). When $E = \mathbb{R}^d$ (or more generally when E has a vector space structure), the interaction law $\Gamma^{(2)}$ is often given in terms of jump amplitudes. Given the law $\widehat{\Gamma}^{(2)}$ of the jump amplitudes of the form:

$$\widehat{\Gamma}^{(2)} : (z_1, z_2) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto \widehat{\Gamma}^{(2)}(z_1, z_2, dh, dk) \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d),$$

the post-collisional law $\Gamma^{(2)}$ is the image measure of $\widehat{\Gamma}^{(2)}$ by the translation

$$(h, k) \in \mathbb{R}^d \times \mathbb{R}^d \mapsto (z_1 + h, z_2 + k) \in \mathbb{R}^d \times \mathbb{R}^d,$$

so that

$$\begin{aligned} \iint_{\mathbb{R}^2 \times \mathbb{R}^d} \varphi_2(z'_1, z'_2) \Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2) \\ = \iint_{\mathbb{R}^d \times \mathbb{R}^d} \varphi_2(z_1 + h, z_2 + k) \widehat{\Gamma}^{(2)}(z_1, z_2, dh, dk). \end{aligned}$$

This is the case investigated in [127, 89].

The N -particle generator can be interpreted as a generator of a Piecewise Deterministic Markov Process in the space E^N . Consequently, and similarly to the case of a mean-field PDMP process (Section 2.2.3), following [66, Chapters 12 and 13, Theorem 13.2.5], a simple N -particle process with generator $\mathcal{L}_N$ of the form (24) is given in the following proposition.

Proposition 2. *Let $\mathcal{Z}_t^N = (Z_t^1, \dots, Z_t^N)$ be the N -particle process defined by the three following rules.*

- (i) *For each (unordered) pair of particles (i, j) , consider an independent non homogeneous Poisson process with rate $\lambda(Z_t^i, Z_t^j)/N$. The jump times $(T_n^{ij})_n$ can be constructed recursively by taking $T_0^{ij} = 0$ and*

$$T_{n+1}^{ij} = \inf \left\{ t \geq T_n^{ij}, \frac{1}{N} \int_{T_n^{ij}}^t \lambda(Z_s^i, Z_s^j) ds \geq E_n^{ij} \right\},$$

where $(E_n^{ij})_n$ are i.i.d. random variables which follow an exponential law with parameter 1.

- (ii) *Between two jump times, the particles evolve independently according to $L^{(1)}$.*

- (iii) *At each jump time T_n^{ij} of a pair (i, j) , update the states of the particles by:*

$$\left(Z_{T_n^{ij}}^i, Z_{T_n^{ij}}^j \right) \sim \Gamma^{(2)} \left(Z_{(T_n^{ij})-}^i, Z_{(T_n^{ij})-}^j, dz'_1, dz'_2 \right). \quad (27)$$

Then the generator of $(\mathcal{Z}_t^N)_t$ is $\mathcal{L}_N$ given by (24) under Assumption 1.

Note that if $\lambda(z_1, z_2)$ remains of order 1, the factor $1/N$ on the right-hand side of (24) ensures that any particle undergoes on average $\mathcal{O}(1)$ collisions per unit of time, which is crucial to take the limit $N \rightarrow +\infty$. Let us now describe what this limit may look like. Ultimately, the goal is to describe the limiting behaviour of the one-particle distribution function $f_t^{1,N}$. More generally, taking a test function of the form

$$\varphi_N = \varphi_s \otimes 1^{N-s},$$

for $s < N$ and $\varphi_s \in C_b(E^s)$, the weak Liouville equation (1) becomes:

$$\begin{aligned} \frac{d}{dt} \langle f_t^{s,N}, \varphi_s \rangle &= \sum_{i=1}^s \langle f_t^{s,N}, L^{(1)} \diamond_i \varphi_s \rangle + \frac{1}{N} \sum_{1 \leq i < j \leq s} \langle f_t^{s,N}, L^{(2)} \diamond_{ij} \varphi_s \rangle \\ &\quad + \frac{N-s}{N} \sum_{i=1}^s \langle f_t^{s+1,N}, L^{(2)} \diamond_{i,s+1} (\varphi_s \otimes 1) \rangle. \end{aligned}$$

This equation is not closed and involves the $(s+1)$ -marginal. This hierarchy of equations is called the BBGKY hierarchy (see Section 3.2.1). The nonlinear model associated to the Boltzmann particle system is obtained by taking the closure:

$$\forall t \geq 0, \exists f_t \in \mathcal{P}(E), \forall s \in \mathbb{N}, \quad f_t^{s,N} \xrightarrow{N \rightarrow +\infty} f_t^{\otimes s}, \quad (28)$$

which is called the *chaos assumption*. The fundamental question in this review is to justify when this property holds. If the chaos assumption holds, then taking $s = 1$ in the Liouville equation shows formally that the one-particle distribution converges towards the weak measure solution f of:

$$\frac{d}{dt} \langle f_t, \varphi \rangle = \langle f_t, L^{(1)} \varphi \rangle + \langle f_t^{\otimes 2}, L^{(2)} (\varphi \otimes 1) \rangle,$$

which is called the *general Boltzmann equation*. Using Assumption 1, this equation can be rewritten

$$\begin{aligned} \frac{d}{dt} \langle f_t, \varphi \rangle &= \langle f_t, L^{(1)} \varphi \rangle \\ &\quad + \int_{E^3} \lambda(z_1, z_2) \{ \varphi(z'_1) - \varphi(z_1) \} \Gamma^{(2)}(z_1, z_2, dz'_1, E) f_t(dz_1) f_t(dz_2), \end{aligned} \quad (29)$$

or in a more symmetric form, using (26):

$$\begin{aligned} \frac{d}{dt} \langle f_t, \varphi \rangle &= \langle f_t, L^{(1)} \varphi \rangle \\ &\quad + \frac{1}{2} \int_{E^4} \lambda(z_1, z_2) \{ \varphi(z'_1) + \varphi(z'_2) - \varphi(z_1) - \varphi(z_2) \} \Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2) f_t(dz_1) f_t(dz_2) \end{aligned} \quad (30)$$

All the Boltzmann type equations in this review are special instances of this general equation for a specific choice of λ and $\Gamma^{(2)}$. Note that the general Boltzmann equation (29) is written in weak form. Examples of $\Gamma^{(2)}$ which lead to more classical Boltzmann type equations used in the modelling of rarefied gas dynamics and written in strong form are given in Section 2.3.3. Here, we can only formally write the dual version of (29):

$$\partial_t f_t = L^{(1)*} f_t + Q(f_t, f_t),$$

where Q is called the *collision operator* which is a (symmetric) quadratic operator on $\mathcal{P}(E) \times \mathcal{P}(E) \rightarrow \mathcal{M}(E)$, defined weakly, for $\varphi \in C_b(E)$, by:

$$\begin{aligned} \langle Q(\mu_1, \mu_2), \varphi \rangle &= \\ &\quad \frac{1}{2} \int_{E^4} \lambda(z_1, z_2) \{ \varphi(z'_1) + \varphi(z'_2) - \varphi(z_1) - \varphi(z_2) \} \Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2) \mu_1(dz_1) \mu_2(dz_2). \end{aligned}$$

Example 7 (Nanbu particle system, continuation of Example 2). The general Boltzmann equation (29) only depends on the marginals of $\Gamma^{(2)}$. In other words, the detail of the interaction mechanism at the particle level is lost in the limit. As a consequence, one can construct different mechanisms which lead to the same Boltzmann equation. For instance, let the marginal of $\Gamma^{(2)}$ be denoted by:

$$\forall (z_1, z_2) \in E^2, \quad \Gamma^{(1)}(z_1, z_2, dz'_1) := \Gamma^{(2)}(z_1, z_2, dz'_1, E).$$

Let us consider the new post-collisional law:

$$\begin{aligned} \tilde{\Gamma}^{(2)}(z_1, z_2, dz'_1, dz'_2) \\ = \frac{1}{2} \left(\Gamma^{(1)}(z_1, z_2, dz'_1) \otimes \delta_{z_2}(dz'_2) + \Gamma^{(1)}(z_2, z_1, dz'_2) \otimes \delta_{z_1}(dz'_1) \right), \end{aligned}$$

and let us denote by $\tilde{\mathcal{L}}_N$ the new corresponding N -particle generator (with $L^{(1)}$ unchanged). This is the generator associated to a particle system such that when a collision occurs, only one particle among the two updates its state (according to the law $\Gamma^{(1)}$) while the state of the other particle remains unchanged. Such mechanism is called a Nanbu interaction mechanism following the terminology of [89, 135]. Nevertheless, one can check that the Boltzmann equation associated to this process is exactly (29) with an interaction rate λ replaced by $\lambda/2$. In the limit $N \rightarrow +\infty$, one cannot distinguish this system from the system where both the particles update their states after a collision.

Note that as explained in Example 2 the Nanbu particle system is also a special case of a mean-field jump process (see Section 2.2.3) with:

$$\lambda(z, \mu) = \int_E \lambda(z, z') \mu(dz') \equiv (\lambda \star \mu)(z),$$

and

$$P_\mu(z, dz') = \frac{\int_{z'' \in E} \lambda(z, z'') \Gamma^{(1)}(z, z'', dz') \mu(dz'')}{\int_{z'' \in E} \lambda(z, z'') \mu(dz'')}.$$

The two following examples are two variants of the Boltzmann model.

Example 8 (External clock). The authors of [34] consider a model where the time between two collisions is given by a Poisson process with fixed rate ΛN , $\Lambda > 0$, independently of the particles. When a collision occurs, the pair (i, j) of particles which interact is chosen among all the pairs of particles with probability $p_{i,j}(\mathcal{Z}_t^N)$, normalised so that for all $\mathcal{Z}^N \in E^N$

$$\sum_{i < j} p_{i,j}(\mathcal{Z}^N) = 1.$$

In this case,

$$\mathcal{L}_N \varphi_N(\mathcal{Z}^N) = \Lambda N \sum_{i < j} p_{i,j}(\mathcal{Z}^N) L^{(2)} \diamond_{ij} \varphi_N(\mathcal{Z}^N).$$

The situation differs from the previous case where a collision rate is attached to each pair of particles and no normalisation constraint is imposed. Dropping the normalisation constraint and taking $p_{i,j}(\mathcal{Z}^N) = \lambda(Z^i, Z^j)/\Lambda$ would give exactly (24). In the case

$$\forall i < j, \quad p_{i,j}(\mathcal{Z}^N) = \frac{2}{N(N-1)},$$

that is when all the pairs are chosen with the same probability, then the situation is equivalent to the previous case with all the collision rates equal to the constant Λ . In [34], propagation of chaos is proved in this case only. It is believable that propagation of chaos does not hold when interactions are driven by a clock independent of the particles.

Example 9 (Non cutoff models). In this review, we made the choice to distinguish the jump rate λ and the post-collisional distribution $\Gamma^{(2)}$. An alternative choice in the literature (see for instance [127]) is to consider $\lambda \equiv 1$ and a *collision kernel*

$$\Gamma^{(2)} : (z_1, z_2) \in E \times E \mapsto \Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2) \in \mathcal{M}^+(E \times E),$$

which is a positive measure but not necessarily a probability distribution. The collision rate λ is thus directly encoded in the total mass of the collision kernel. Two cases may happen, for given $z_1, z_2 \in E$, either

$$\Gamma^{(2)}(z_1, z_2, E, E) < +\infty$$

or

$$\Gamma^{(2)}(z_1, z_2, E, E) = +\infty.$$

The first case is called the cutoff case. This case is equivalent to the previous case (24) with Assumption 1 and the following post-collisional law and collision rate:

$$\tilde{\Gamma}^{(2)}(z_1, z_2, dz'_1, dz'_2) = \frac{\Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2)}{\Gamma^{(2)}(z_1, z_2, E, E)}, \quad \tilde{\lambda}(z_1, z_2) = \Gamma^{(2)}(z_1, z_2, E, E).$$

In the second case, called the non-cutoff case, the lack of integrability means that there are an infinite number of collisions in finite time. Such system therefore cannot be simulated by a particle system as in Proposition 2. Nevertheless it still makes sense to consider the abstract Markov process defined by the generator $\mathcal{L}_N$. Non-cutoff models are historically important as explained in Section 2.3.3. Non-cutoff models are often handled by approximating them by cutoff models. In this review we implicitly consider cutoff models but we will occasionally specify when a technique can be extended to non-cutoff cases.

The nonlinear limit can also be defined as the solution of a more general martingale problem.

Definition 2.4 (Nonlinear Boltzmann martingale problem). Let $T > 0$ and $f_0 \in \mathcal{P}(E)$. We write $I = [0, T]$. We say that $f_I \in \mathcal{P}(D([0, T], E))$ is a solution to the nonlinear Boltzmann martingale problem with initial law f_0 when for any test function $\varphi \in \text{Dom}(L^{(1)})$,

$$M_t^\varphi = \varphi(Z_t) - \varphi(Z_0) - \int_0^t \{L^{(1)}\varphi(Z_s) + K\varphi(Z_s, f_s)\} ds,$$

is a f_I -martingale, where $(Z_t)_t$ is the canonical process, $f_s = (Z_s)_\# f_I$ and for $\mu \in \mathcal{P}(E)$ and $z_1 \in E$,

$$K\varphi(z_1, \mu) := \iint_{E \times E} \lambda(z_1, z_2) \{\varphi(z'_1) - \varphi(z_1)\} \Gamma^{(1)}(z_1, z_2, dz'_1) \mu(dz_2).$$

Existence and uniqueness for the nonlinear Boltzmann martingale problem holds under classical Lipschitz and boundedness assumptions on the parameters. It is a special case of the model studied in [88]. Note that this martingale problem is also a special case of the nonlinear mean-field martingale problem (Definition 2.3) with

$$L_\mu \varphi(z) = L^{(1)} \varphi(z) + K\varphi(z, \mu).$$

This translates the fact that the Boltzmann equation is obtained as the limit of both the Boltzmann model and the Nanbu particle system which is a special case of mean-field jump process.

We end this section with a classical useful proposition which states that when the collision rate λ is uniformly bounded, then the situation is essentially the same as when it is constant.

Proposition 3 (Uniform clock trick). *Assume that*

$$\sup_{z_1, z_2 \in E} \lambda(z_1, z_2) \leq \Lambda < \infty, \quad (31)$$

and let $(\tilde{Z}_t^N)_t$ be the process defined by the three following rules.

- (i) *To each pair of particles is attached an independent Poisson process with rate Λ/N .*
- (ii) *Between two jump times, the particles evolve independently according to $L^{(1)}$.*
- (iii) *When the clock of the pair (i, j) rings at time T_{ij} , then the states of the particles is updated with probability $\lambda(\tilde{Z}_t^i, \tilde{Z}_t^j)/\Lambda$ by:*

$$\left(\tilde{Z}_{T_{ij}^+}^i, \tilde{Z}_{T_{ij}^+}^j \right) \sim \Gamma^{(2)} \left(\tilde{Z}_{T_{ij}^-}^i, \tilde{Z}_{T_{ij}^-}^j, dz'_1, dz'_2 \right),$$

and with probability $(1 - \lambda(\tilde{Z}_t^i, \tilde{Z}_t^j)/\Lambda)$, nothing happens (this case is called a fictitious collision).

Then the law of $(\tilde{Z}_t^N)_t$ is equal to the law of the process constructed in Proposition 2.

Proof. Let us compute the generator $\tilde{\mathcal{L}}_N$ of the process $(\tilde{Z}_t^N)_t$. It holds that

$$\tilde{\mathcal{L}}_N \varphi_N = \sum_{i=1}^N L^{(1)} \diamond_i \varphi_N + \frac{1}{N} \sum_{i < j} \tilde{L}^{(2)} \diamond_{ij} \varphi_N,$$

with, given $\varphi_2 \in C_b(E^2)$,

$$\begin{aligned}
\tilde{L}^{(2)}\varphi_2(z_1, z_2) &= \Lambda \int_0^1 \left\{ \mathbb{1}_{\eta \leq \frac{\lambda(z_1, z_2)}{\Lambda}} \left(\iint_{E \times E} \varphi_2(z'_1, z'_2) \Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2) \right) \right. \\
&\quad \left. + \mathbb{1}_{\eta \geq \frac{\lambda(z_1, z_2)}{\Lambda}} \varphi_2(z_1, z_2) \right\} d\eta - \Lambda \varphi_2(z_1, z_2) \\
&= \Lambda \times \frac{\lambda(z_1, z_2)}{\Lambda} \iint_{E \times E} \varphi_2(z'_1, z'_2) \Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2) \\
&\quad + \Lambda \left(1 - \frac{\lambda(z_1, z_2)}{\Lambda} \right) \varphi_2(z_1, z_2) - \Lambda \varphi_2(z_1, z_2) \\
&= L^{(2)}\varphi_2(z_1, z_2),
\end{aligned}$$

and thus $\mathcal{L}_N = \tilde{\mathcal{L}}_N$ and the two processes are equal in law. $\square$

2.3.2 Parametric Boltzmann models

In many applications, the post-collisional distribution is explicitly given as the image measure of a known parameter space (Θ, ν) endowed with a probability measure ν (or a positive measure with infinite mass in the non cutoff case). Analogously to the case of mean-field jump models (see Example 4), in this review, we distinguish this particular class of models and we call them *parametric Boltzmann models*.

Definition 2.5 (Parametric Boltzmann model). Let be given two measurable functions

$$\psi_1 : E \times E \times \Theta \rightarrow E, \quad \psi_2 : E \times E \times \Theta \rightarrow E,$$

which satisfy the symmetry assumption

$$\forall (z_1, z_2) \in E^2, \quad (\psi_1, \psi_2)(z_1, z_2, \cdot)_{\#} \nu = (\psi_2, \psi_1)(z_2, z_1, \cdot)_{\#} \nu. \quad (32)$$

Let the function ψ be defined by

$$\psi : E \times E \times \Theta \rightarrow E^2, \quad (z_1, z_2, \theta) \mapsto (\psi_1(z_1, z_2, \theta), \psi_2(z_1, z_2, \theta)).$$

A parametric Boltzmann model with parameters (Θ, ψ) is a Boltzmann model of the form (24) with Assumption 1 and a post-collisional distribution of the form:

$$\forall (z_1, z_2) \in E^2, \quad \Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2) = \psi(z_1, z_2, \cdot)_{\#} \nu.$$

The symmetry assumption (32) is the equivalent of (26) in this special case. In particular, for any two-variable test function $\varphi_2 \in C_b(E^2)$ and any $(z_1, z_2) \in E \times E$,

$$\begin{aligned}
\iint_{E \times E} \varphi_2(z'_1, z'_2) \Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2) &= \int_{\Theta} \varphi_2(\psi_1(z_1, z_2, \theta), \psi_2(z_1, z_2, \theta)) \nu(d\theta) \\
&= \int_{\Theta} \varphi_2(\psi_2(z_2, z_1, \theta), \psi_1(z_2, z_1, \theta)) \nu(d\theta),
\end{aligned}$$

where the last equality follows from (32). A sufficient condition for (32) to hold is the case investigated in [149] with:

$$\psi_2(z_1, z_2, \theta) = \psi_1(z_2, z_1, \theta).$$

In terms of particle systems, following Proposition 2, in a parametric model, when a collision occurs, a parameter $\theta \sim \nu$ is sampled first and then the states of the particle is updated by:

$$\left(Z_{T_{ij}^+}^i, Z_{T_{ij}^+}^j \right) = \psi \left(Z_{T_{ij}^-}^i, Z_{T_{ij}^-}^j, \theta \right).$$

Example 10 (Symmetrization). Wagner [162] treats the case of particle systems with a generator of the form: for $\mathbf{z}^N = (z^1, \dots, z^N) \in E^N$ and $\varphi_N \in C_b(E^N)$,

$$\begin{aligned} \mathcal{L}_N \varphi_N(\mathbf{z}^N) &= \sum_{i=1}^N L^{(1)} \diamond_i \varphi_N(\mathbf{z}^N) \\ &\quad + \frac{1}{2N} \sum_{i \neq j} \tilde{\lambda}(z^i, z^j) \int_{\tilde{\Theta}} \{ \varphi_N(\mathbf{z}^N(i, j, \tilde{\theta})) - \varphi_N(\mathbf{z}^N) \} \tilde{\nu}(d\tilde{\theta}), \end{aligned} \quad (33)$$

where $\tilde{\lambda} : E \times E \rightarrow \mathbb{R}_+$, $\tilde{\Theta}$ is a parameter set endowed with a probability measure $\tilde{\nu}$ and $\mathbf{z}^N(i, j, \tilde{\theta})$ is the N dimensional vector whose k component is equal to

$$z^k(i, j, \tilde{\theta}) = \begin{cases} z^k & \text{if } k \neq i, j \\ \tilde{\psi}_1(z^i, z^j, \tilde{\theta}) & \text{if } k = i \\ \tilde{\psi}_2(z^i, z^j, \tilde{\theta}) & \text{if } k = j \end{cases},$$

for two given functions $\tilde{\psi}_1, \tilde{\psi}_2 : E \times E \times \tilde{\Theta} \rightarrow E$. The main difference with the generator (24) is that Wagner distinguishes the pairs (i, j) and (j, i) while in (24) we consider unordered pairs of particles but add the symmetry assumption (26). Consequently, the double sum in (33) runs over all indices $i, j = 1, \dots, N$ while in the sum (24), it runs over the indices $i < j$. Nevertheless, using a simple symmetrization procedure, the model (33) fits into the previous general framework with

$$\Theta = \tilde{\Theta} \times [0, 1], \quad \nu(d\theta) = \tilde{\nu}(d\tilde{\theta}) \otimes d\sigma,$$

where $\theta = (\tilde{\theta}, \sigma) \in \Theta$, $d\sigma$ is the uniform probability measure on $[0, 1]$ and for $z_1, z_2 \in E$ we define

$$\begin{aligned} \lambda(z_1, z_2) &= \frac{\tilde{\lambda}(z_1, z_2) + \tilde{\lambda}(z_2, z_1)}{2}, \\ \psi_1(z_1, z_2, \theta) &= \mathbb{1}_{\sigma \leq \frac{\tilde{\lambda}(z_1, z_2)}{2\lambda(z_1, z_2)}} \tilde{\psi}_1(z_1, z_2, \tilde{\theta}) + \mathbb{1}_{\sigma > \frac{\tilde{\lambda}(z_1, z_2)}{2\lambda(z_1, z_2)}} \tilde{\psi}_2(z_2, z_1, \tilde{\theta}), \\ \psi_2(z_1, z_2, \theta) &= \mathbb{1}_{\sigma \leq \frac{\tilde{\lambda}(z_1, z_2)}{2\lambda(z_1, z_2)}} \tilde{\psi}_2(z_1, z_2, \tilde{\theta}) + \mathbb{1}_{\sigma > \frac{\tilde{\lambda}(z_1, z_2)}{2\lambda(z_1, z_2)}} \tilde{\psi}_1(z_2, z_1, \tilde{\theta}). \end{aligned}$$

One can check that the functions ψ_1 and ψ_2 satisfy (32) and that the generator (24) of the associated parametric model (Definition 2.5) is equal to (33). In this case, the Boltzmann equation (29) reads

$$\begin{aligned} \frac{d}{dt}\langle f_t, \varphi \rangle &= \langle f_t, L^{(1)}\varphi \rangle + \frac{1}{2} \int_{\tilde{\Theta} \times E^2} \left\{ \tilde{\lambda}(z_1, z_2) [\varphi(\tilde{\psi}_1(z_1, z_2, \tilde{\theta})) - \varphi(z_1)] \right. \\ &\quad \left. + \tilde{\lambda}(z_2, z_1) [\varphi(\tilde{\psi}_2(z_2, z_1, \tilde{\theta})) - \varphi(z_1)] \right\} \tilde{\nu}(d\tilde{\theta}) f_t(dz_1) f_t(dz_2), \end{aligned}$$

or equivalently after the change of variables $(z_1, z_2) \mapsto (z_2, z_1)$,

$$\begin{aligned} \frac{d}{dt}\langle f_t, \varphi \rangle &= \langle f_t, L^{(1)}\varphi \rangle + \frac{1}{2} \int_{\tilde{\Theta} \times E^2} \tilde{\lambda}(z_1, z_2) \left\{ \varphi(\tilde{\psi}_1(z_1, z_2, \tilde{\theta})) + \varphi(\tilde{\psi}_2(z_1, z_2, \tilde{\theta})) \right. \\ &\quad \left. - \varphi(z_1) - \varphi(z_2) \right\} \tilde{\nu}(d\tilde{\theta}) f_t(dz_1) f_t(dz_2). \quad (34) \end{aligned}$$

The introductory section of [162] contains many examples of such models, in particular models (in Russian) due to Leontovich in the 30's and Skorokhod in the 80's that we did not manage to find. A more recent example inspired by economic models of wealth distribution [122] is given in [48]. The authors assume $E = \mathbb{R}$ with $L^{(1)} = 0$, $\lambda = 1$, $\tilde{\Theta} = \mathbb{R}^4$ and

$$\tilde{\psi}_1(z_1, z_2, (L, R, \tilde{L}, \tilde{R})) = Lz_1 + Rz_2,$$

and

$$\tilde{\psi}_2(z_1, z_2, (L, R, \tilde{L}, \tilde{R})) = \tilde{L}z_2 + \tilde{R}z_1.$$

In this model, the state of a particle represents the wealth of an individual and the parameters $(L, R, \tilde{L}, \tilde{R})$ specify how a trade between two individuals affect their wealth. This model generalises a famous model due to Kac [104] which will be discussed in the next section. From the modelling point of view, it is more natural to use generators of the form (33); several additional examples will be given in particular in Section II-5.2.4 for socio-economic models. On the other hand, the generator form (24) will simplify some computations in Section II-4.

In the parametric framework, the particle system can be advantageously written as the solution of a system of SDEs driven by Poisson measures. In a famous article, Tanaka [153] proposed a SDE approach to study the nonlinear Boltzmann system of rarefied gas dynamics (which will be presented in the next section). He introduced a class of nonlinear SDEs driven by Poisson random measures which will be described in Section II-4.4. As we shall see, although it is relatively easy to write a system of coupled SDEs which describes the particle system, its relationship with Tanaka's SDE is not completely straightforward. Around the same time, Murata [133] tackled the question and proved the propagation of chaos (for a specific model) using a coupling argument between the two systems of SDEs. The idea is of course reminiscent of the well-known McKean's theorem and all the works reviewed in Section II-3.1 for McKean-Vlasov systems. Note however that Murata's work is among the first ones which use the very fruitful idea of coupling to prove propagation of chaos. His argument is based on a clever but not so easy optimal coupling argument which seems to have been largely forgotten in the subsequent literature. A recent series

of articles [81, 48, 49] has proposed a more contemporary point of view on the question. The arguments are very similar to Murata's but take advantage of the development of the theory of optimal transport. Let us also mention that these articles seem to be based on [77] which also introduces an optimal coupling argument reminiscent of Murata's but in a different context, namely the derivation of the Landau equation from a system of interacting diffusion processes. We will continue this discussion in Section II-4.4.

Example 11 (Semi-parametric model). A natural extension of the parametric model would consider a measure on Θ which depends on the state of the particles, for instance one can consider a post-collisional distribution of the form

$$\begin{aligned} \iint_{E \times E} \varphi_2(z'_1, z'_2) \Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2) \\ = \int_{\Theta} \varphi_2(\psi_1(z_1, z_2, \theta), \psi_2(z_1, z_2, \theta)) q(z_1, z_2, \theta) \nu(d\theta) \end{aligned} \quad (35)$$

where for all $z_1, z_2 \in E$, $q(z_1, z_2, \cdot)$ is a probability density function with respect to the measure $\nu \in \mathcal{M}^+(E)$. Wagner [162] considered such model that will be called *semi-parametric* in this review. If there exists $M > 0$ and $q_0(\theta)$ a probability density function with respect to ν such that

$$\forall z_1, z_2 \in E, \forall \theta \in \Theta, \quad q(z_1, z_2, \theta) \leq M q_0(\theta), \quad (36)$$

then the situation can be reduced to the parametric case thanks to an accept-reject scheme similar to the one in Proposition 3. Namely in the extended parameter space

$$\tilde{\Theta} = \Theta \times [0, 1],$$

endowed with the probability measure $q_0(\theta) d\theta \otimes d\eta$, let us define the function

$$\tilde{\psi}(z_1, z_2, (\theta, \eta)) = \begin{cases} (\psi_1(z_1, z_2, \theta), \psi_2(z_1, z_2, \theta)) & \text{if } \eta \leq \frac{q(z_1, z_2, \theta)}{M q_0(\theta)} \\ (z_1, z_2) & \text{if } \eta > \frac{q(z_1, z_2, \theta)}{M q_0(\theta)} \end{cases}.$$

Then up to a time rescaling $t \rightarrow tM$, the parametric model $(\tilde{\Theta}, \tilde{\psi})$ is equivalent in law to the semi-parametric model. Note that (36) automatically holds when Θ is compact and q bounded.

2.3.3 Classical models in collisional kinetic theory

The foundations of kinetic theory lie in the seminal work of Boltzmann and Maxwell who attempted to understand the large scale behaviour of a gas of particles defined in the phase space $E = \mathbb{R}^d \times \mathbb{R}^d$ by their position and velocity. Many interactions mechanisms can be considered, depending on the physical assumptions. The starting point is the Newton equations satisfied by the N -particle system $(\mathcal{Z}_t^N)_t$, for $i \in \{1, \dots, N\}$,

$$\begin{cases} \frac{dX_t^i}{dt} = V_t^i \\ \frac{dV_t^i}{dt} = - \sum_{j=1}^N \nabla V(|X_t^j - X_t^i|) \end{cases}, \quad (37)$$

where V is a (smooth) repulsive potential, typically an inverse power law. Another important system is the hard-sphere system which will be described in Example 15. Without any other assumption, it is not clear that this set of equations defines a binary collision process. In fact, it is more reminiscent of a mean-field system without the (crucial) $1/N$ scaling in front of the sum. Boltzmann and Maxwell considered the case of dilute gases (also called *rarefied gas*), that is gases where the density of particles is so small that in the sum in (37), there is typically no more than one non-zero term. The dynamics of each particle is therefore mainly driven by the free transport until the particle comes very close to another particle which induces a deviation of its trajectory (as well as the trajectory of the other particle) depending on the potential V . During this process, everything is deterministic and the only source of randomness comes from the initial condition. The probabilistic interpretation presented in this section is due to Kac. Boltzmann derived the equation satisfied by the one-particle distribution when $N \rightarrow +\infty$. In its most general form, the Boltzmann equation of rarefied gas dynamics reads (in strong form):

$$\begin{aligned} \partial_t f_t(x, v) + v \cdot \nabla_x f_t \\ = \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B(v - v_*, \sigma) \left(f_t(x, v'_*) f_t(x, v') - f_t(x, v_*) f_t(x, v) \right) dv_* d\sigma, \end{aligned} \quad (38)$$

where

$$\begin{cases} v' &= \frac{v + v_*}{2} + \frac{|v - v_*|}{2} \sigma \\ v'_* &= \frac{v + v_*}{2} - \frac{|v - v_*|}{2} \sigma \end{cases}, \quad (39)$$

are the post-collisional velocities. The parameter $\sigma \in \mathbb{S}^{d-1}$ is often called the scattering angle. This transformation preserves energy and momentum. The function $B : \mathbb{R}^d \times \mathbb{S}^{d-1} \rightarrow \mathbb{R}_+$ is of the form

$$B(u, \sigma) = \Phi(|u|) \Sigma(\theta), \quad (40)$$

with $\cos \theta = \frac{u}{|u|} \cdot \sigma$, $\theta \in [0, \pi]$. The function Φ is called the velocity cross-section and the function Σ is called the angular cross-section. The function B is referred as the collision kernel (in the literature, it is also sometimes called the cross-section). It is customary to write $B(u, \sigma) \equiv B(|u|, \cos \theta)$. Depending on the choice of the potential V , some of the most important collision kernels derived by Maxwell are listed below.

- (Hard-sphere)

$$\Phi(|u|) = |u|, \quad \Sigma(\theta) = 1. \quad (41)$$

- (Inverse-power law potentials)

$$\Phi(|u|) = |u|^\gamma, \quad \gamma = \frac{s - (2d - 1)}{s - 1}, \quad s > 2,$$

and Σ has a non-integrable singularity when $\theta \rightarrow 0$, so that

$$\int_0^\pi \Sigma(\theta) d\theta = +\infty$$

- (Maxwell molecules)

$$\Phi(|u|) = 1, \quad \int_0^\pi \Sigma(\theta) d\theta = +\infty. \quad (42)$$

- (Maxwell molecules with Grad's cutoff)

$$\Phi(|u|) = 1, \quad \int_0^\pi \Sigma(\theta) d\theta < +\infty. \quad (43)$$

We will not go further into the description of the Boltzmann equation. The interested reader will find a thorough discussion and analysis of these different models in the reviews [159, 58] or in the classical books [43, 45]. We also mention the book [44] which contains a very interesting biography of Ludwig Boltzmann as well as a scientific discussion of the physics of his time and of his legacy.

This review is focused on the rigorous derivation of the Boltzmann equation from a system of particles. On the right-hand side of (38), the variable x (position) only appears as a parameter: this is the limit where collisions between two particles happen only when the two particles are at the same position. From a mathematical point of view, this purely local interaction mechanism makes the derivation very difficult if not impossible with stochastic tools (see [127]). A deterministic example (Lanford's theorem) is nevertheless given in Example 15 and Section II-4.6. Apart from this result we will focus on simplified mechanisms: either spatially homogeneous *Kac models* (Example 12 and Example 13) or kinetic mollified models (Example 14). Both cases have a natural probabilistic interpretation which fits into the framework of Section 2.3.1.

Example 12 (Spatially homogeneous Boltzmann equation). In the case of a spatially homogeneous problem, the difficulty due to the local interaction does not appear. In this case the spatially-homogeneous Boltzmann equation of rarefied gas dynamics describes a gas of particles defined by their velocity only:

$$\partial_t f(t, v) = \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B(v - v_*, \sigma) \left(f_t(v'_*) f_t(v') - f_t(v_*) f_t(v) \right) dv_* d\sigma, \quad (44)$$

This last equation can be shown to be the strong form of the Boltzmann equation (29) with a parametric post-collisional distribution given by the cross-section B . The N -particle stochastic process associated to this equation is given by Proposition 2. More precisely, the (spatially-homogeneous) hard-sphere model and the (spatially-homogeneous) model of Maxwell molecules with Grad's cutoff fit into the framework of Section 2.3.2 with:

$$\psi_1(v, v_*, \theta) = v', \quad \psi_2(v, v_*, \theta) = v'_*,$$

and

$$\lambda(v, v_*) = \Phi(|v - v_*|) \int_0^\pi \Sigma(\theta) d\theta, \\ \Gamma(v, v_*, dz', dv'_*) = \psi(v, v_*, \cdot)_\# \left(\frac{\Sigma}{\int_0^\pi \Sigma(\theta) d\theta} \right).$$

To be more precise, with this particular choice of the parameters, the weak-form of the general Boltzmann equation (30) reads:

$$\begin{aligned} \frac{d}{dt} \langle f_t, \varphi \rangle &= \frac{1}{2} \int_{\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{S}^{d-1}} \left\{ \varphi(v') + \varphi(v'_*) - \varphi(v) - \varphi(v_*) \right\} f_t(v) f_t(v_*) B(v - v_*, \sigma) dv dv_* d\sigma. \end{aligned}$$

The strong form (44) is obtained thanks to the following classical involutive unit Jacobian changes of variables which allow to exchange (v, v_*) and (v', v'_*) :

$$(v, v_*, \sigma) \rightarrow (v', v'_*, \vec{k}), \quad (v, v_*) \rightarrow (v_*, v),$$

with $\vec{k} = (v - v_*)/|v - v_*|$ (see [159, Chapter 1, Section 4.5]). The collision kernel B is invariant by these changes of variables so we can keep its arguments unchanged.

The non-cutoff cases are more difficult to handle due to the non-integrability of the angular cross-section (see Example 9).

The following example describes the famous Kac model, which is a one dimensional caricature of a gas of Maxwellian molecules. Beyond the relative simplicity of the model, the seminal article of Kac [104] is of particular importance because it introduces the mathematical definition of propagation of chaos.

Example 13 (Kac model). Within the framework of Section 2.3.2, the Kac model is defined in $E = \mathbb{R}$ by $L^{(1)} = 0$ and the post-collisional distribution

$$\begin{aligned} \iint_{\mathbb{R} \times \mathbb{R}} \varphi_2(z'_1, z'_2) \Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2) \\ := \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi_2(z_1 \cos \theta + z_2 \sin \theta, -z_1 \sin \theta + z_2 \cos \theta) d\theta \end{aligned}$$

and the collision rate $\lambda(z_1, z_2) = \nu = \text{constant}$. Then the weak Boltzmann equation becomes

$$\frac{d}{dt} \langle \varphi, f_t \rangle = \frac{\nu}{2\pi} \iint_{\mathbb{R} \times \mathbb{R}} \int_{-\pi}^{\pi} \{ \varphi(z_1 \cos \theta + z_2 \sin \theta) - \varphi(z_1) \} f_t(dz_1) f_t(dz_2).$$

With the change of variable (with θ fixed)

$$(z'_1, z'_2) = (z_1 \cos \theta + z_2 \sin \theta, -z_1 \sin \theta + z_2 \cos \theta),$$

followed by $\theta \mapsto -\theta$ (both changes of variable have unit jacobian), Kac obtained the following equation in strong form:

$$\partial_t f_t(z_1) = \frac{\nu}{2\pi} \int_{\mathbb{R}} \int_{-\pi}^{\pi} \{ f(z'_1) f(z'_2) - f_t(z_1) f_t(z_2) \} dz_2 d\theta.$$

We refer the reader to [32, 129] for a thorough analysis and discussion of the Kac model and its generalisations in kinetic theory. Keeping a collision rate λ constant the authors

of [34] generalised the arguments of the proof of the propagation of chaos to a larger class of models. This generalised result, that we will call *Kac's theorem*, will be discussed in Section II-4.1. As already mentioned in Section 2.3.2, the Kac model is also a special instance of the model studied in [48] within a framework which will be described in Section II-4.4.

The work of Kac had a very strong influence on the literature so that Boltzmann models with $L^{(1)} = 0$ are sometimes called Kac models, or also *homogeneous* Boltzmann models.

Example 14 (Mollified Boltzmann models). The interaction mechanism of a kinetic particle system is said to be *purely local* when, within the framework of Section 2.3.1, the collision rate is taken equal to

$$\lambda((x_1, v_1), (x_2, v_2)) = \delta_{x_1, x_2}.$$

This indicates that two particles interact if and only if they are exactly at the same position. As explained in [127], the probabilistic interpretation of purely local models is extremely difficult and one can rather consider the smoothened version:

$$\lambda((x_1, v_1), (x_2, v_2)) = K(|x_1 - x_2|),$$

where K is a smooth mollifier with fixed radius (i.e. a non negative radial function which tends to zero at infinity and which integrates to one). The post-collisional distribution is unchanged and acts only on the velocity variable:

$$\Gamma^{(2)}(z_1, z_2, dz'_1, dz'_2) \equiv \Gamma^{(2)}(v_1, v_2, dv'_1, dv'_2) \otimes \delta_{x_1}(dx'_1) \otimes \delta_{x_2}(dx'_2).$$

This model is called a mollified Boltzmann model. Its probabilistic treatment is discussed in [89] and [127]. Note that with a general state space E , all the models in Section 2.3.1 and in particular the one in Proposition 2 are implicitly mollified Boltzmann models. Most of the models reviewed in Section II-4 are mollified models. Purely local Boltzmann models can be recovered by letting the mollifier converge to a Dirac delta $K \rightarrow \delta_0$ (formally or with a quantitative control). Another example of purely local Boltzmann model is the hard-sphere system defined in the next Example 15.

Example 15 (Hard-sphere system). A *hard-sphere* is a spherical particle defined by its position, its velocity and its diameter $\varepsilon > 0$. Moreover, it is assumed that two hard-spheres cannot overlap. A system of N hard-spheres is thus defined on the domain:

$$\mathcal{D}_N := \{\mathbf{z}^N = (x^i, v^i)_{i \in \{1, \dots, N\}} \in (\mathbb{R}^d \times \mathbb{R}^d)^N, \forall i \neq j, |x^i - x^j| \geq \varepsilon\}.$$

The dynamics of the hard-sphere system is a special degenerate case of (37) with a vanishing potential but with an additional boundary condition which tells what happens on the boundary of $\mathcal{D}_N$, that is when two particles are at a distance ε (the term collision is here self-explanatory). The collision of two hard-spheres is an elastic collision which preserves energy and momentum. Starting with a pair of pre-collisional velocities (v^i, v^j) ,

writing down the conservation laws leads to the following formula for the post-collisional velocities:

$$\begin{aligned} v^{i*} &= v^i - \nu^{i,j} \cdot (v^i - v^j) \nu^{i,j} \\ v^{j*} &= v^j + \nu^{i,j} \cdot (v^i - v^j) \nu^{i,j} \end{aligned} \quad (45)$$

where $\nu^{i,j} := (x^i - x^j)/|x^i - x^j| \in \mathbb{S}^{d-1}$. This representation is not the same as the representation (39) but it can be shown that they are actually equivalent [159, Chapter 1, Section 4.6]. Pre-collisional means that (v^i, v^j) are such that $(v^i - v^j) \cdot \nu^{i,j} < 0$. It can also be checked that the post-collisional velocities satisfy $(v^{i*} - v^{j*}) \cdot \nu^{i,j} > 0$. Note that this transformation is an involution in the sense that if $(v^i - v^j) \cdot \nu^{i,j} > 0$ (that is the v^i and v^j are in a post-collisional configuration), then (45) gives the pre-collisional velocities. Note also that this dynamical system is completely deterministic.

The large scale behaviour when $N \rightarrow +\infty$ and $\varepsilon \rightarrow 0$ is given by the Boltzmann equation (38) with the hard-sphere cross-section. Under the chaotic assumption (28) at time $t = 0$, Lanford's theorem [110] states that in the Boltzmann-Grad limit $N\varepsilon^{d-1} \rightarrow 1$, (28) also holds for later time. This scaling was introduced by Grad in [86]. The proof of Lanford's theorem is extremely difficult. We will briefly review the main ideas in Section II-4.6. Our presentation will follow closely [83].

3 Notions about chaos

3.1 Topology reminders: metrics and convergence for probability measures

Since propagation of chaos is about the convergence of probability measures, we first need to present the topological tools that will be constantly used in the following. The content of this section is fairly classical, most of the results specific to our topic can be found in [94], see also [97, Section 3.4], [130, Section 2.5], [131, Section 3] or [158]. A more general overview of the topology of the space of probability measures can be found in the classical books [161, 16, 138].

3.1.1 Distances on the space of probability measures

Let $(\mathcal{E}, d_{\mathcal{E}})$ be a Polish space. For $p \geq 1$, a measure μ in $\mathcal{P}(\mathcal{E})$ admits a *finite p -th moment* when there exists $x_0 \in \mathcal{E}$ such that

$$\mathbb{E}_{X \sim \mu} [d_{\mathcal{E}}(X, x_0)^p] := \int_{\mathcal{E}} d_{\mathcal{E}}(x, x_0)^p \mu(dx) < +\infty.$$

This property does not depend on x_0 . The space of probability measures with finite p -th moment is denoted by $\mathcal{P}_p(\mathcal{E})$. The Wasserstein distance on $\mathcal{P}_p(\mathcal{E})$ will be the most important one in the following.

Definition 3.1 (Wasserstein distances). For $p \geq 1$, the Wasserstein- p distance between the probability measures μ and ν in $\mathcal{P}_p(\mathcal{E})$ is defined by

$$W_{d_{\mathcal{E}}, p}(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{\mathcal{E} \times \mathcal{E}} d_{\mathcal{E}}(x, y)^p \pi(dx, dy) \right)^{1/p} = \inf_{\substack{X \sim \mu \\ Y \sim \nu}} \mathbb{E} [d_{\mathcal{E}}(X, Y)^p]^{1/p}$$

where $\Pi(\mu, \nu)$ is the set of all couplings of μ and ν , that is to say, the set of probability measures on $E \times E$ with first and second marginals respectively equal to μ and ν .

The total variation distance can be understood as a Wasserstein-1 distance with the trivial distance $d_{\mathcal{E}}(x, y) = \delta_{x, y}$.

Definition 3.2 (Total variation norm). The total variation distance between two probability measures μ and ν in $\mathcal{P}(\mathcal{E})$ is defined by

$$\|\mu - \nu\|_{\text{TV}} = 2 \inf_{\substack{X \sim \mu \\ Y \sim \nu}} \mathbb{P}(X \neq Y).$$

Since $\mathcal{P}(\mathcal{E})$ can be seen as a subset of the dual space $C_b(\mathcal{E})^*$, natural strong norms on $\mathcal{P}(\mathcal{E})$ are induced by usual norms on functional spaces. The following proposition links these distances to dual norms:

Proposition 4 (Duality formulae). *The total variation and Wasserstein-1 distances satisfy:*

$$\|\mu - \nu\|_{\text{TV}} = \sup_{\|\varphi\|_{\infty} \leq 1} \left\{ \int_{\mathcal{E}} \varphi(x) \mu(\mathrm{d}x) - \int_{\mathcal{E}} \varphi(x) \nu(\mathrm{d}x) \right\}$$

and

$$W_{d,1}(\mu, \nu) = \sup_{\|\varphi\|_{\text{Lip}, d_{\mathcal{E}}} \leq 1} \left\{ \int_{\mathcal{E}} \varphi(x) \mu(\mathrm{d}x) - \int_{\mathcal{E}} \varphi(x) \nu(\mathrm{d}x) \right\}$$

where

$$\|\varphi\|_{\text{Lip}, d_{\mathcal{E}}} := \sup_{x \neq y} \frac{|\varphi(x) - \varphi(y)|}{d_{\mathcal{E}}(x, y)}.$$

Proof. See [160, Theorem 1.14] □

Another important class of dual norms is given by the negative Sobolev norms $W^{-s,p}$. Let us emphasize two special cases.

Definition 3.3 (Some negative Sobolev norms). When $\mathcal{E} = \mathbb{R}^d$ we define the following norms.

- For $\mu, \nu \in \mathcal{P}(\mathcal{E})$ and $s > \frac{d}{2}$

$$\|\mu - \nu\|_{H^{-s}}^2 := \int_{\mathbb{R}^d} |\hat{\mu}(\xi) - \hat{\nu}(\xi)|^2 \frac{\mathrm{d}\xi}{(1 + |\xi|^2)^s},$$

where $\hat{\mu}$ is the Fourier transform of μ .

- The dual norm of the Euclidean Lipschitz semi-norm

$$\|\mu - \nu\|_{W^{-1,\infty}} := \sup_{\|\varphi\|_{W^{1,\infty}} \leq 1} \langle \mu - \nu, \varphi \rangle$$

where the $W^{1,\infty}$ Sobolev norm is $\|\varphi\|_{W^{1,\infty}} = \|\varphi\|_{\infty} + \|\nabla \varphi\|_{\infty}$.

An important property of the negative Sobolev norm H^{-s} is its polynomial structure (see [94, Lemma 2.9]), in the sense that the definition of this norm only involves the integration of some functions against tensor products of measures.

Lemma 3.4. *The negative Sobolev norm H^{-s} , $s > d/2$ on $\mathbb{R}^d$ satisfies for any $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$,*

$$\|\mu - \nu\|_{H^{-s}}^2 = \int_{\mathbb{R}^{2d}} \Phi_s(x - y) (\mu^{\otimes 2} - \mu \otimes \nu)(\mathrm{d}x, \mathrm{d}y) \tag{46}$$

$$+ \int_{\mathbb{R}^{2d}} \Phi_s(x - y) (\nu^{\otimes 2} - \nu \otimes \mu)(\mathrm{d}x, \mathrm{d}y) \tag{47}$$

$$= \int_{\mathbb{R}^{2d}} \Phi_s(x - y) (\mu - \nu)^{\otimes 2}(\mathrm{d}x, \mathrm{d}y), \tag{48}$$

where $\Phi_s(z) := \int_{\mathbb{R}^d} e^{-iz \cdot \xi} (1 + |\xi|^2)^{-s} \mathrm{d}\xi$.

Remark 9. A similar distance has also been recently considered in particular in the work of Serfaty [146] where it is called a *modulated energy*.

Proposition 5 (Comparison of distances). *Assume the distance $d_{\mathcal{E}}$ to be bounded. The following uniform topological equivalences hold.*

- The TV distance dominates the Wassertein-1 distance W_1 .
- $\|\cdot\|_{\text{Lip}, d_{\mathcal{E}}}$ is equivalent to $\|\cdot\|_{W^{1,\infty}}$ (see [94, Equations (2.4) and (2.5)]) and this implies the same for W_1 and $\|\cdot\|_{W^{-1,\infty}}$ distances.
- The W_2 -distance dominates the W_1 -distance, and for $s > \frac{d+1}{2}$ the W_1 -distance dominates the square of the H^{-s} -distance.
- For measures in $\mathcal{P}_p(\mathcal{E})$ with $p > 0$ and $s \geq 1$, the H^{-s} distance dominates the W_1 distance up to a positive exponent.
- For measures in $\mathcal{P}_p(\mathcal{E})$ with $p > 2$, the W_1 -distance dominates the W_2 -distance up to a positive exponent.

Proof. See [94, Lemma 2.1], which gives a quantitative version of this. $\square$

Finally, let $\mathcal{F} = \{\varphi^k, k \in \mathbb{N}\}$ be a countable and separating subset of $C_b(\mathcal{E})$ (see Definition 3.7 below) and such that $\|\varphi^k\|_{\infty} \leq 1$ for all $k \in \mathbb{N}$. Then the following expression defines a distance on $\mathcal{P}(E)$ for any $p \geq 1$

$$D_p(\mu, \nu) := \left(\sum_{k=1}^{+\infty} \frac{1}{2^k} |\langle \mu - \nu, \varphi^k \rangle|^p \right)^{1/p}. \quad (49)$$

In the literature, the most encountered distances are D_1 and D_2 which are used as a convenient tool to metricise the notion of weak convergence defined below (see Example 16). To conclude, we summarise the main cases of interest for the Wasserstein distance and other related distances.

Definition 3.5. In problems related to propagation of chaos, the Wasserstein distances are often used in the following cases.

- When $\mathcal{E} = E$ is the state space of the particles, endowed with a distance d_E , we do not specify the dependency in d_E :

$$W_{d_E, p} \equiv W_p$$

When $E = \mathbb{R}^d$, the bounded moment assumption can be removed by using the bounded distance $\tilde{d}_E(x, y) := \inf(|x - y|, 1)$.

- When $\mathcal{E} = E^k$, $k \in \mathbb{N}$ is a product space of the state space (E, d_E) , unless otherwise specified we follow [94] and use the normalised distance: for $\mathbf{x}^k = (x^1, \dots, x^k)$ and $\mathbf{y}^k = (y^1, \dots, y^k)$,

$$d_{E^k}(\mathbf{x}^k, \mathbf{y}^k) := \frac{1}{k} \sum_{i=1}^k d_E(x^i, y^i),$$

and we simply write $W_{d_{E^k}, p} \equiv W_p$. When we use the non-normalised distance

$$\widetilde{d}_{E^k}(\mathbf{x}^k, \mathbf{y}^k) := \sum_{i=1}^k d_E(x^i, y^i),$$

we write $W_{\widetilde{d}_{E^k}, p} \equiv \widetilde{W}_p$. In the special case $E = \mathbb{R}^d$ endowed with the Euclidean norm, we will rather use the normalised distance,

$$d_{E^k}^p(\mathbf{x}^k, \mathbf{y}^k) := \frac{1}{k} \sum_{i=1}^k |x^i - y^i|^p,$$

and the non-normalised one

$$\widetilde{d}_{E^k}^p(\mathbf{x}^k, \mathbf{y}^k) := \sum_{i=1}^k |x^i - y^i|^p,$$

so that $\widetilde{W}_p^p(\mu, \nu) = kW_p^p(\mu, \nu)$.

- The continuous path space $\mathcal{E} = C([0, T], \mathcal{F})$ is endowed with the uniform topology

$$d_{\mathcal{E}}((X_t)_t, (Y_t)_t) := \sup_{t \in [0, T]} d_{\mathcal{F}}(X_t, Y_t)$$

Two important cases are $\mathcal{F} = E$ and $\mathcal{F} = E^k$ for $k \in \mathbb{N}$ and is useful to note that

$$C([0, T], E^k) \simeq C([0, T], E)^k$$

- The Skorokhod space $\mathcal{E} = D([0, T], \mathcal{F})$ is endowed with the Skorokhod distance (see Section A.2). It is often more convenient to use the uniform topology although it does not make the space complete. However, as the uniform topology is stronger than the Skorokhod topology, any estimate in Wasserstein distance for the uniform topology implies the same estimate for the Skorokhod topology, see [4, Section 3].
- When $\mathcal{E} = \mathcal{P}(\mathcal{F})$ is a space of probability measures over a space $\mathcal{F}$ which is typically one of the aforementioned spaces, we will mainly encounter three cases:

$$\mathcal{W}_p := W_{W_p, p} \quad \mathcal{W}_{D_1} := W_{D_1, 1} \quad \mathcal{W}_{H^{-s}} := W_{H^{-s}, 2}$$

3.1.2 Convergence in the space of probability measures

Since $\mathcal{P}(\mathcal{E})$ is a subset of $C_b(\mathcal{E})^*$, a weak topology is induced by the weak- $\star$ topology on $C_b(\mathcal{E})^*$.

Definition 3.6. (Weak convergence) The weak convergence of a sequence of probability measures $(\mu_N)_N$ towards $\mu \in \mathcal{P}(\mathcal{E})$ is defined as the related weak- $\star$ convergence in $C_b(\mathcal{E})^*$.

More precisely, a sequence of probability measures $(\mu_N)_N$ is said to converge weakly towards μ when

$$\forall \varphi \in C_b(\mathcal{E}), \quad \langle \mu_N, \varphi \rangle \xrightarrow{N \rightarrow +\infty} \langle \mu, \varphi \rangle.$$

The corresponding topology is the *weakest* topology which makes the evaluation maps $\nu \mapsto \langle \nu, \varphi \rangle$ measurable. In probability theory, the related convergence for μ_N -distributed random variables is also called *convergence in law* or *convergence in distribution* (see for instance [16, Section 3]).

In many examples, the set of continuous bounded test functions is too large and it is necessary to work with a smaller subspace, for instance the domain of a generator. The minimal needed assumptions on a subspace of test functions are given by the following definition (see [75, p.112]).

Definition 3.7 (Separating and convergence determining class). A subset $\mathcal{F} \subset C_b(\mathcal{E})$ is called *separating* whenever for all $\mu, \nu \in \mathcal{P}(\mathcal{E})$, the condition

$$\forall \varphi \in \mathcal{F}, \quad \langle \mu, \varphi \rangle = \langle \nu, \varphi \rangle,$$

implies that $\mu = \nu$. The subset $\mathcal{F}$ is said to be *convergence determining* whenever for any sequence $(\mu_N)_N$ in $\mathcal{P}(\mathcal{E})$ and $\mu \in \mathcal{P}(\mathcal{E})$, the condition

$$\forall \varphi \in \mathcal{F}, \quad \langle \mu_N, \varphi \rangle \xrightarrow{N \rightarrow +\infty} \langle \mu, \varphi \rangle,$$

implies that $\mu_N \rightarrow \mu$ weakly. Note that a convergence determining set is also separating (the converse is false in general).

Example 16. The following sets are convergence determining.

- By the Portmanteau theorem [16, Theorem 2.1], the set $UC_b(\mathcal{E})$ of bounded uniformly continuous functions on $\mathcal{E}$ is convergence determining (for any equivalent metric on $\mathcal{E}$).
- When $\mathcal{E}$ is locally compact, the space $C_c(\mathcal{E})$ of continuous functions with compact support is convergence determining [75, Chapter 3, Proposition 4.4]. The space $C_0(\mathcal{E})$ of continuous functions vanishing at infinity is thus also convergence determining. Note that the space $\mathcal{P}(\mathcal{E})$ is not a closed subspace of $C_c(\mathcal{E})^*$ (nor of $C_0(\mathcal{E})^*$).
- When $\mathcal{E}$ is locally compact, the Stone-Weierstrass theorem implies that $C_0(\mathcal{E})$ is separable. Thus, any dense countable subset $\mathcal{F} = \{\varphi^k, k \in \mathbb{N}\} \subset C_0(\mathcal{E})$ is convergence determining. Without loss of generality we can assume that $\|\varphi^k\|_\infty \leq 1$ for all $k \in \mathbb{N}$. Consequently, the distance (49) metricises the weak-convergence. Indeed, since each term of the series (49) is bounded by 2^{-k} , the convergence $D_p(\mu_N, \mu) \rightarrow 0$ as $N \rightarrow +\infty$ is equivalent to $\langle \varphi^k, \mu_N \rangle \rightarrow \langle \mu, \varphi^k \rangle$ for all $k \in \mathbb{N}$. It is also possible to take φ^k Lipschitz with a Lipschitz constant bounded by 1 for all $k \in \mathbb{N}$ and vanishing at infinity.

- In general $C_b(\mathcal{E})$ is not separable so there is no obvious other countable convergence determining set. There exists nevertheless another classical choice when $\mathcal{E}$ is only separable. By a theorem due to Urysohn, any separable metric space can be topologically imbedded in $[0, 1]^{\mathbb{N}}$ and it is therefore possible to construct on $\mathcal{E}$ an equivalent metric $\tilde{d}_{\mathcal{E}}$ which makes $(\mathcal{E}, \tilde{d}_{\mathcal{E}})$ a totally bounded set. The completion $\tilde{\mathcal{E}}$ of this space is therefore compact and the set $UC_b(\mathcal{E})$ under this metric is isomorphic to the set $C_b(\tilde{\mathcal{E}})$ which is separable since $\tilde{\mathcal{E}}$ is compact (by Stone-Weierstrass theorem). In conclusion, there exists a countable dense subset $\mathcal{F} = \{\varphi^k, k \in \mathbb{N}\}$ in $UC_b(\mathcal{E})$. Up to replacing φ^k by $\varphi^k / \|\varphi^k\|_{\infty}$ one can assume that the φ^k are bounded by 1 and the distance (49) thus metricises the weak convergence, see [138, Theorem 6.6] and [148, Theorem 1.1.2]. Note that $(\mathcal{P}(\mathcal{E}), D_1)$ is separable and D_1 is equivalent to a complete metric, see the remark which follows [148, Theorem 1.1.2] and [54, Remark 3.2.2].
- Since the space $\text{Lip}(\mathcal{E})$ is dense in the space $C_b(\mathcal{E})$, the functions φ^k in the above examples can be taken Lipschitz (with a Lipschitz constant bounded by 1).

The weak convergence is thus metricised by a D_1 distance. Since this distance is weaker than the Wasserstein-1 distance (it can be seen by Proposition 4), this implies that the topology induced by the Wasserstein distance is stronger than the topology induced by the weak convergence. The topology induced by the Wasserstein distance is described by the following theorem.

Theorem 3.8 (Wassertein topology). *Let $(\mathcal{E}, d_{\mathcal{E}})$ be a Polish space and $p \geq 1$. The Wasserstein distance $W_{d_{\mathcal{E}}, p}$ metricises the weak convergence in $\mathcal{P}_p(\mathcal{E})$, defined as the convergence against bounded continuous test functions and the convergence of the p -th moments.*

Proof. [161, Theorem 6.9] □

In the following, an important case is the case $\mathcal{E} = \mathcal{P}(E)$. Weak convergence of measures in $\mathcal{P}(\mathcal{P}(E))$ is thus defined as the convergence against test functions in $C_b(\mathcal{P}(E))$. Their representation is not intuitive, except for *linear* test functions of the kind $\mu \mapsto \langle \mu, \varphi \rangle$ where φ belongs to $C_b(E)$. The following results (stated in a more probabilistic framework) show that these functions are sufficient to prove weak convergence.

Proposition 6 (Measure-valued convergence in $\mathcal{W}_{D_1}$). *Let D_1 be a distance given by (49) and Example 16 which metricises the weak convergence on $\mathcal{P}(E)$. Consider a sequence $(\mu_N)_N$ of $\mathcal{P}(E)$ -valued random variables and another random probability measure μ . The following properties hold.*

- (i) *If $\mathcal{W}_{D_1}(\text{Law}(\mu_N), \text{Law}(\mu)) \rightarrow 0$ as $N \rightarrow +\infty$ then $(\mu_N)_N$ converges in law towards μ .*
- (ii) *If $\mathbb{E}|\langle \mu_N - \mu, \varphi \rangle| \rightarrow 0$ as $N \rightarrow +\infty$ for all $\varphi \in UC_b(E)$, then it holds that $\mathcal{W}_{D_1}(\text{Law}(\mu_N), \text{Law}(\mu)) \rightarrow 0$ and $(\mu_N)_N$ converges in law towards μ .*

Proof. Let us recall [138, Theorem 6.1] that the space of bounded uniformly continuous functions is convergence determining. Thus, let $\Phi \in C_b(\mathcal{P}(E))$ be a function which is uniformly continuous for the metric D_1 . For any $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that for any $\mu, \nu \in \mathcal{P}(E)$,

$$D_1(\mu, \nu) \leq \delta(\varepsilon) \Rightarrow |\Phi(\mu) - \Phi(\nu)| \leq \varepsilon.$$

The first point then directly stems from the Markov inequality:

$$\begin{aligned} |\langle \text{Law}(\mu_N) - \text{Law}(\mu), \Phi \rangle| &\leq \mathbb{E}|\Phi(\mu_N) - \Phi(\mu)| \leq \varepsilon + 2\|\Phi\|_\infty \mathbb{P}(|\Phi(\mu_N) - \Phi(\mu)| \geq \varepsilon) \\ &\leq \varepsilon + \frac{2\|\Phi\|_\infty}{\delta(\varepsilon)} \mathbb{E}D_1(\mu_N, \mu). \end{aligned}$$

Since this is true for any $\text{Law}(\mu_N), \text{Law}(\mu)$ -distributed random variables μ_N, μ this finally gives

$$|\langle \text{Law}(\mu_N) - \text{Law}(\mu), \Phi \rangle| \leq \varepsilon + \frac{2\|\Phi\|_\infty}{\delta(\varepsilon)} \mathcal{W}_{D_1}(\text{Law}(\mu_N), \text{Law}(\mu)),$$

and the conclusion follows. For the second point, using the expression (49) for $D_1(\mu_N, \mu)$, the monotonic convergence theorem gives

$$\mathcal{W}_{D_1}(\text{Law}(\mu_N), \text{Law}(\mu)) \leq \sum_{k=1}^{+\infty} \frac{1}{2^k} \mathbb{E}|\langle \mu_N - \mu, \varphi^k \rangle|,$$

and then the dominated convergence theorem concludes the proof. $\square$

Remark 10 (Comparison to $\mathcal{W}_1$). For E locally compact, it has been proven at the same time that

$$\mathcal{W}_{D_1}(\text{Law}(\mu_N), \text{Law}(\mu)) \leq \sup_{\|\varphi\|_{\text{Lip}} \leq 1} \mathbb{E}|\langle \mu_N - \mu, \varphi \rangle|.$$

Since

$$\sup_{\|\varphi\|_{\text{Lip}} \leq 1} \mathbb{E}|\langle \mu_N - \mu, \varphi \rangle| \leq \mathbb{E} \left[\sup_{\|\varphi\|_{\text{Lip}} \leq 1} |\langle \mu_N - \mu, \varphi \rangle| \right] = \mathbb{E}W_1(\mu_N, \mu),$$

this pinpoints, taking the infimum on the $\text{Law}(\mu_N), \text{Law}(\mu)$ -distributed random variables μ_N, μ , that $\mathcal{W}_1$ is stronger than $\mathcal{W}_{D_1}$ and both are stronger than the weak convergence on $\mathcal{P}(\mathcal{P}(E))$.

Corollary 1 (Sufficient conditions in a deterministic case). *With the same assumptions as above, if μ is a deterministic $\mathcal{P}(E)$ -valued random variable (i.e. $\text{Law}(\mu) \in \mathcal{P}(\mathcal{P}(E))$ is a Dirac mass), then the following assertions are equivalent.*

(i) $\mathcal{W}_{D_1}(\text{Law}(\mu_N), \text{Law}(\mu)) \rightarrow 0$ as $N \rightarrow +\infty$

(ii) For all bounded uniformly continuous function φ on E , $\mathbb{E}|\langle \mu_N - \mu, \varphi \rangle| \rightarrow 0$ as $N \rightarrow +\infty$.

The second assertion is also equivalent to $\mathbb{E}|\langle \mu_N - \mu, \varphi \rangle|^2 \rightarrow 0$ as $N \rightarrow +\infty$ for all bounded uniformly continuous function φ on E .

Proof. The direct implication uses the fact $\mathcal{W}_{D_1}$ metricises the convergence in law of measure-valued random variables and that $\nu \mapsto |\langle \nu - \mu, \varphi \rangle|$ is continuous for the weak- $\star$ topology on $\mathcal{P}(E)$ when μ is deterministic. The converse implication is the second point of the previous proposition. $\square$

The previous results can be found in [63, Section 2] or in [158, Section 5, Lemma 10] for an equivalent argument in $E = \mathbb{R}^d$. In the previous lemma, only *linear* test functions are used. This notion can be generalised by considering the algebra of *polynomials* on $\mathcal{P}(E)$. Its definition and main properties are stated in the following lemma.

Lemma 3.9. *Let E be a Polish space. For $k \in \mathbb{N}$ and $\varphi_k \in C_b(E^k)$, the monomial function of order k on $\mathcal{P}(E)$ is defined by:*

$$R_{\varphi_k} : \mathcal{P}(E) \rightarrow \mathbb{R}, \quad \mu \mapsto \langle \mu^{\otimes k}, \varphi_k \rangle.$$

The linear span of the set of monomial functions is called the algebra of polynomial functions on $\mathcal{P}(E)$. The following properties hold.

- (i) *Every monomial is bounded and continuous on $\mathcal{P}(E)$ for the weak topology.*
- (ii) *The algebra of polynomial functions is a convergence determining subset of $C_b(\mathcal{P}(E))$.*
- (iii) *If E is compact then the algebra of polynomials is dense in $C_b(\mathcal{P}(E))$.*

Proof. (i) First it is clear that every monomial and thus every polynomial is bounded. Let $(\mu_N)_N$ be a sequence in $\mathcal{P}(E)$ and $\mu \in \mathcal{P}(E)$ such that $\mu_N \rightarrow \mu$ as $N \rightarrow +\infty$. For any $k \in \mathbb{N}$, and any tensorized test function $\varphi_k \in \varphi^1 \otimes \dots \otimes \varphi^k \in C_b(E^k)$, it holds that

$$\langle \mu_N^{\otimes k}, \varphi_k \rangle = \prod_{j=1}^k \langle \mu_N, \varphi^j \rangle \xrightarrow{N \rightarrow +\infty} \langle \mu^{\otimes k}, \varphi_k \rangle.$$

Then using [75, Chapter 3, Proposition 4.6], the set $C_b(E)^{\otimes k} \subset C_b(E^k)$ is convergence determining and thus $\mu_N^{\otimes k} \rightarrow \mu^{\otimes k}$. It implies that for all $\varphi_k \in C_b(E^k)$, $R_{\varphi_k}(\mu_N) \rightarrow R_{\varphi_k}(\mu)$ and R_{φ_k} is therefore continuous.

- (ii) From the first point, the algebra of polynomials is a subset of $C_b(\mathcal{P}(E))$. Let D_1 be a metric of the form (49) such that $(\mathcal{P}(E), \tilde{D}_1)$ is a Polish space for a metric $\tilde{D}_1$ which is equivalent to D_1 . A set of functions $\mathcal{F} \subset C_b(\mathcal{P}(E))$ is said to strongly separates points when for every $\mu \in \mathcal{P}(E)$, and $\delta > 0$, there exists a finite set $\{\Phi^1, \dots, \Phi^k\} \subset \mathcal{F}$ such that

$$\inf_{\nu: D_1(\mu, \nu) \geq \delta} \max_{1 \leq j \leq k} |\Phi^j(\nu) - \Phi^j(\mu)| > 0.$$

Since D_1 is equivalent to a complete metric, [75, Chapter 3, Theorem 4.5] states that $\mathcal{F}$ is convergence determining if $\mathcal{F}$ strongly separates points. It is thus enough to prove that the algebra of polynomial functions contains a subset which strongly separates points. Let $(\varphi^k)_k$ be the sequence of functions in $C_b(E)$ which defines

D_1 . Then the set $\{R_{\varphi^k}, k \in \mathbb{N}\} \subset C_b(\mathcal{P}(E))$ strongly separates points. Indeed, let $\mu \in \mathcal{P}(E)$, let $\delta > 0$ and let $m \in \mathbb{N}$ be such that $2^{-m} < \delta/4$. For any $\nu \in \mathcal{P}(E)$ such that $D_1(\mu, \nu) \geq \delta$, it holds that

$$\sum_{k=1}^m |\langle \mu, \varphi^k \rangle - \langle \nu, \varphi^k \rangle| \geq \frac{\delta}{2},$$

and hence $\max_{1 \leq k \leq m} |\langle \mu, \varphi^k \rangle - \langle \nu, \varphi^k \rangle| \geq \delta/(2m)$. The conclusion follows.

(iii) This follows from the Stone-Weierstrass theorem, since $\mathcal{P}(E)$ is compact in this case. $\square$

3.1.3 Entropic convergence

Powerful tools to compare measures are also given by the Kullback-Leibler divergence, which is traditionally called the *relative entropy* in our context, and the related Fisher information.

Definition 3.10 (Entropy, Fisher information and entropic convergence). Let $\mathcal{E}$ be a Polish space. Given two probability measures $\mu, \nu \in \mathcal{P}(\mathcal{E})$ (or more generally two measures), the relative entropy is defined by

$$H(\nu|\mu) := \int_{\mathcal{E}} \frac{d\nu}{d\mu} \log \left(\frac{d\nu}{d\mu} \right) d\mu,$$

where $d\nu/d\mu$ is the Radon-Nikodym derivative. When the two measures are mutually singular, by convention, the relative entropy is set to $+\infty$ (the same holds for the Fisher information below). If moreover E is endowed with a smooth manifold structure, the Fisher information can be defined as

$$I(\nu|\mu) := \int_{\mathcal{E}} \left| \nabla \log \left(\frac{d\nu}{d\mu} \right) \right|^2 d\nu,$$

with the same conventions. These quantities are dimensionally super-additive, equality being achieved only for tensorized distributions, in the sense that given $\nu \in \mathcal{P}(\mathcal{E} \times \mathcal{E})$ with marginals $\nu_1, \nu_2 \in \mathcal{P}(\mathcal{E})$ and $\mu \in \mathcal{P}(\mathcal{E})$, then

$$H(\nu|\mu \otimes \mu) \geq H(\nu_1|\mu) + H(\nu_2|\mu),$$

equality being achieved if only if $\nu = \nu_1 \otimes \nu_2$. Moreover $H(\nu|\mu) \geq 0$ and $H(\nu|\mu) = 0$ if and only if $\mu = \nu$. The *entropic convergence* of a sequence $(\mu_N)_N$ in $\mathcal{P}(\mathcal{E})$ towards μ is defined by the convergence of the relative entropy:

$$H(\mu_N|\mu) \xrightarrow{N \rightarrow +\infty} 0.$$

The relative entropy between two probability measures is also called the *Kullback-Leibler divergence*.

Remark 11 (Towards dimension free quantities). For $\mu^N, \nu^N \in \mathcal{P}(\mathbb{E}^N)$, the normalized entropy $H_N(\nu^N|\mu^N) := \frac{1}{N}H(\nu^N|\mu^N)$ can be handful, since it leads to estimates which do not depend on N when $\mu^N = \mu^{\otimes N}$ is tensorized; the same holds for $W_1(\mu^{\otimes N}, \nu^{\otimes N})$ when W_1 is defined using the normalized distance on E^N , see [94, Proposition 2.6]. An extension to random measures $\pi \in \mathcal{P}(\mathcal{P}(\mathcal{E}))$ is provided in [94] setting $\mathcal{H}(\pi) = \mathbb{E}_{\nu \sim \pi} H(\nu|\mu)$ for a given $\mu \in \mathcal{P}(\mathcal{E})$.

The entropic convergence is stronger than the strongest distance.

Proposition 7 (Pinsker inequality). *The following inequality implies that the entropy convergence is stronger than the convergence in total variation norm:*

$$\|\mu - \nu\|_{\text{TV}}^2 \leq 2H(\nu|\mu). \quad (50)$$

The link with the Wassertein-2 distance can be recovered through the following proposition.

Proposition 8 (HWI inequality [137]). *If $d\nu = e^{-\Psi(x)}dx$ is a probability measure on $\mathbb{R}^d$ with finite second order moment, C^2 -regularity for Ψ and $D^2\Psi \geq \lambda I_d$ for some real λ , then for any $\mu \in \mathcal{P}(\mathbb{R}^d)$ which is absolutely continuous with respect to ν , it holds that*

$$H(\nu|\mu) \leq W_2(\mu, \nu) \sqrt{I(\nu|\mu)} - \frac{\lambda}{2} W_2^2(\mu, \nu).$$

Further results which link the relative entropy and the distances on $\mathcal{P}(\mathcal{E})$ will be given in Section 4.4.3. The relative entropy will play an important role in Section 4.4.2.

3.2 Representation of symmetric particle systems

This section introduces the various points of view to describe a system of particles. So far, we have mainly discussed the case of finite systems described by the N -particle distribution function f_t^N at time t . The first Section 3.2.1 will detail more of its properties. Then, as the goal is to deal with the limit $N \rightarrow +\infty$, a framework for infinite particle systems is needed, this will be described in Section 3.2.2.

3.2.1 Finite particle systems

Let $N \in \mathbb{N}$ be a fixed finite number of particles. In full generality, there is only one property of the N -particle distribution function that is always true: at any time and for any of the models considered, it is a symmetric probability measure on E^N (the particle system is said to be exchangeable). Let us therefore consider in this section a symmetric probability measure $f^N \in \mathcal{P}_{\text{sym}}(E^N)$ (in a static framework, it does not depend on the time). There exist two main representations of f^N which are based on this symmetry assumption.

The marginal distributions and the BBGKY hierarchy.

The symmetry assumption implies that for any $k \leq N$, we can define the k -th marginal distribution on E^k by:

$$\forall \varphi_k \in C_b(E^k), \quad \langle f^{k,N}, \varphi_k \rangle = \langle f^N, \varphi_k \otimes 1^{\otimes(N-k)} \rangle,$$

and $f^{k,N} \in \mathcal{P}_{\text{sym}}(E^k)$ is itself a symmetric probability measure. The N -th marginal is of course the measure f^N itself. However, keeping in mind that the final goal is to take $N \rightarrow +\infty$, one can consider for any fixed $k \in \mathbb{N}$ the limit of $f^{k,N}$ in $\mathcal{P}(E^k)$, which is not possible for f^N directly since it belongs to a space which depends on N . As we shall see in the following, it is often enough to treat the case $k = 2$.

In a dynamic framework, when f_t^N solves the Liouville equation (1), for each given $k \in \mathbb{N}$, a natural idea is to derive an equation for the k -th marginal distribution by considering a test function in (1) of the form $\varphi_N = \varphi_k \otimes 1^{\otimes(N-k)}$ with $\varphi_k \in C_b(E^k)$. For Boltzmann models, this computation as already been sketched in Section 2.3.1 and gave:

$$\begin{aligned} \frac{d}{dt} \langle f_t^{k,N}, \varphi_k \rangle &= \sum_{i=1}^k \langle f_t^{k,N}, L^{(1)} \diamond_i \varphi_k \rangle + \frac{1}{N} \sum_{1 \leq i < j \leq k} \langle f_t^{k,N}, L^{(2)} \diamond_{ij} \varphi_k \rangle \\ &\quad + \frac{N-k}{N} \sum_{i=1}^k \langle f_t^{k+1,N}, L^{(2)} \diamond_{i,k+1} (\varphi_k \otimes 1) \rangle. \end{aligned} \quad (51)$$

For mean-field systems, let us look at the special linear case:

$$\forall \varphi \in \mathcal{F}, \quad L_\mu \varphi(x) = \int_E \tilde{L}_x \varphi(y) \mu(dy),$$

where for any $x \in E$, $\tilde{L}_x$ is a Markov operator on $\mathcal{F}$ (such that for all $\varphi \in \mathcal{F}$ and $y \in E$, the map $x \mapsto \tilde{L}_x \varphi(y)$ is measurable). Then one can check similarly (using the symmetry of f_t^N) that the k -th marginal of the Liouville equation satisfies:

$$\begin{aligned} \frac{d}{dt} \langle f_t^{k,N}, \varphi_k \rangle &= \frac{1}{N} \sum_{1 \leq i, j \leq k} \int_{E^k} \tilde{L}_{x^i} \diamond_i \varphi_k(\hat{\mathbf{x}}^{i,j}) f_t^{k,N}(\mathbf{d}\mathbf{x}^k) \\ &\quad + \frac{N-k}{N} \sum_{i=1}^k \int_{E^{k+1}} \tilde{L}_{x^i} \diamond_i \varphi_k(\hat{\mathbf{x}}^{i,k+1}) f_t^{k+1,N}(\mathbf{d}\mathbf{x}^{k+1}), \end{aligned} \quad (52)$$

where we recall the notation $\mathbf{x}^k = (x^1, \dots, x^k) \in E^k$ and for $i \leq k$, $\hat{\mathbf{x}}^{i,j}$ denotes the vector in E^k where the i -th element is replaced by x^j .

In both equations (51) and (52), the important point to notice is that the leading term (in N) on the right-hand side depends on the $(k+1)$ -th marginal. Since $f_t^{k,N}$ depends on $f_t^{k+1,N}$ for any $k < N$, this gives a hierarchy of N equations, the N -th one being the Liouville equation itself. This hierarchy is called the BBGKY hierarchy, from the names of the mathematicians Bogolioubov, Born, Green, Kirkwood and Yvon. It is more customary to write the BBGKY in the forward form. For the classical Boltzmann model

of Section 2.3.3, using the notations of (38) and (39), the first equation of the hierarchy reads,

$$\begin{aligned} \partial_t f_t^{1,N}(x, v) + v \cdot \nabla_x f_t^{1,N} \\ = \frac{N-1}{N} \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} B(v - v_*, \sigma) \left(f_t^{2,N}(x, v'_*, x, v') - f_t^{2,N}(x, v_*, x, v) \right) dv_* d\sigma. \end{aligned}$$

We refer to the classical reference [45] for a more detailed derivation of the BBGKY hierarchy associated to this model and to [33] for another class of Boltzmann models.

For the mean-field case, let us consider the diffusion operator in $E = \mathbb{R}^d$ given by:

$$\tilde{L}_x \varphi(y) = K(x, y) \cdot \nabla_y \varphi(y) + \Delta_y \varphi(y),$$

where $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a symmetric function with $K(x, x) = 0$ for all $x \in \mathbb{R}^d$. Then the first equation of the BBGKY hierarchy in forward form reads:

$$\partial_t f_t^{1,N}(x) = -\frac{N-1}{N} \nabla_x \cdot \left\{ \int_{\mathbb{R}^d} K(x, z) f_t^{2,N}(x, z) dz \right\} + \Delta_x f_t^{1,N}(x).$$

Note that in both cases, if

$$f_t^{2,N} = f_t^{1,N} \otimes f_t^{1,N}, \quad (53)$$

then, up to the factor $(N-1)/N$, the first marginal $f_t^{1,N}$ solves the nonlinear limit problem, respectively the Boltzmann equation (38) and the Fokker-Planck equation (12) (with $b(x, \mu) = K \star \mu(x)$ and $\sigma = \sqrt{2}I_d$). The relation (53) is called a *closure assumption* because under this assumption, the marginals (here the first one) satisfy a closed equation. The question of Kac's chaos and the propagation of chaos is precisely to justify this closure assumption in the asymptotic limit $N \rightarrow +\infty$. Indeed, the relation (53) is never true since it means that any two particles are statistically independent (which is not the case since they interact).

The BBGKY hierarchy is useful only in the linear cases described above. For general mean-field models with an operator L_μ which has a more complicated dependence in μ , it is not possible to derive a BBGKY hierarchy: this procedure would only say that $f_t^{k,N}$ depends on the whole distribution f_t^N which is not informative. For the Boltzmann model described above, the proof of the propagation of chaos and the justification of the closure assumption (53) are reviewed in Section II-4.6 (this result is the renown Lanford's theorem). We also mention that, beyond propagation of chaos, different closure assumptions than (53) can be considered as an approximating procedure in a numerical perspective, see for instance [12] for the mean-field model described above.

The empirical measure.

With a more probabilistic point of view, a symmetric measure $f^N \in \mathcal{P}(E^N)$ means that any system $\mathcal{X}^N = (X^1, \dots, X^N) \in E^N$ of f^N -distributed random variables is invariant in law under any permutation of the indexes. Such an exchangeable system is equivalently described by its (random) empirical measure

$$\mu_{\mathcal{X}^N} = \frac{1}{N} \sum_{i=1}^N \delta_{X^i} \in \mathcal{P}(E), \quad (54)$$

as this measure contains all the statistical information up to the particle numbering (a quantitative version is stated in the Lemma 3.11 below). One can immediately see the advantage of such representation: it is possible to work with only one element which belongs to the fixed space $\mathcal{P}(E)$, in contrast to $f^N \in \mathcal{P}(E^N)$ or to the N marginal distributions. To be completely rigorous, one should work in the quotient space $E^N/\mathfrak{S}_N$, whose elements $\bar{\mathbf{x}}^N$ gather all the permutations of the vector $\mathbf{x}^N \in E^N$. There is the one to one mapping:

$$\boldsymbol{\mu}_N : E^N/\mathfrak{S}_N \rightarrow \widehat{\mathcal{P}}_N(E), \quad \bar{\mathbf{x}}^N \mapsto \mu_{\mathbf{x}^N}, \quad (55)$$

where $\widehat{\mathcal{P}}_N(E)$ denotes the space of empirical measures of size N on E . Since $\mu_{\mathbf{x}^N} \in \widehat{\mathcal{P}}_N(E) \subset \mathcal{P}(E)$ is a random element, a somehow unfortunate complication arises for the space of observables $C_b(\mathcal{P}(E))$: in this framework, test functions are continuous bounded functions on (a subset of) the set of probability measures (endowed with the weak topology). This is clearly more difficult to handle than usual test functions on E^N or E^k .

Remark 12. Note that the two sets $C_b(E^N/\mathfrak{S}_N)$ and $C_b(\widehat{\mathcal{P}}_N(E))$ are naturally identified by taking the composition with the previous map. Moreover, since all the measures considered are symmetric, integration on $E^N/\mathfrak{S}_N$ is equivalent to integration on E^N . This is why, with a slight abuse, the test functions always belong to $C_b(E^N)$.

From the point of view of measure theory, the representation (54) means that the law f^N is replaced by its push-forward by the map (55) (seen as a map $E^N \rightarrow \mathcal{P}(E)$), defined by:

$$F^N := (\boldsymbol{\mu}_N)_\# f^N \in \mathcal{P}(\mathcal{P}(E)).$$

The following lemma shows that F^N is enough to characterise f^N .

Lemma 3.11 (Approximation rate of marginals). *For $1 \leq k \leq N$, let the moment measure $F^{k,N} \in \mathcal{P}(E^k)$ be defined by:*

$$\forall \varphi_k \in C_b(E^k), \quad \langle F^{k,N}, \varphi_k \rangle = \int_{\mathcal{P}(E)} \langle \nu^{\otimes k}, \varphi_k \rangle F^N(d\nu).$$

Then it holds that:

$$\|f^{k,N} - F^{k,N}\|_{\text{TV}} \leq 2 \frac{k(k-1)}{N}. \quad (56)$$

Coming back to the probabilistic point of view, F^N is the law of the random measure $\mu_{\mathcal{X}^N}$ where $\mathcal{X}^N \sim f^N$. The moment measures can thus be written

$$F^{k,N} = \mathbb{E}[\mu_{\mathcal{X}^N}^{\otimes k}],$$

where this expression is understood in the weak sense, for all $\varphi_k \in C_b(E^k)$,

$$\langle F^{k,N}, \varphi_k \rangle = \langle \mathbb{E}[\mu_{\mathcal{X}^N}^{\otimes k}], \varphi_k \rangle = \mathbb{E}[\langle \mu_{\mathcal{X}^N}^{\otimes k}, \varphi_k \rangle].$$

Proof. Given a test function $\varphi_k \in C_b(E^k)$, using the symmetry of $f^{k,N}$, it holds that:

$$\langle \mathbb{E}_{\mathcal{X}^N} \mu_{\mathcal{X}^N}^{\otimes k}, \varphi_k \rangle = \int_{E^N} \langle \mu_{\mathbf{x}^N}^{\otimes k}, \varphi_k \rangle f^N(d\mathbf{x}^N),$$

and

$$\langle f^{k,N}, \varphi_k \rangle = \int_{E^N} \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \varphi_k(x^{\sigma(1)}, \dots, x^{\sigma(k)}) f^N(d\mathbf{x}^N).$$

Consequently,

$$\left| \langle f^{k,N} - \mathbb{E}_{\mathcal{X}^N} \mu_{\mathcal{X}^N}^{\otimes k}, \varphi_k \rangle \right| \leq \sup_{\mathbf{x}^N \in E^N} \left| \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \varphi_k(x^{\sigma(1)}, \dots, x^{\sigma(k)}) - \langle \mu_{\mathbf{x}^N}^{\otimes k}, \varphi_k \rangle \right|.$$

Moreover,

$$\frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \varphi_k(x^{\sigma(1)}, \dots, x^{\sigma(k)}) = \frac{1}{A_N^k} \sum_{\substack{i_1, \dots, i_k \\ \text{pairwise distinct}}} \varphi_k(x^{i_1}, \dots, x^{i_k}),$$

and

$$\begin{aligned} \langle \mu_{\mathbf{x}^N}^{\otimes k}, \varphi_k \rangle &= \frac{1}{N^k} \sum_{i_1, \dots, i_k} \varphi_k(x^{i_1}, \dots, x^{i_k}) \\ &= \frac{1}{N^k} \sum_{\substack{i_1, \dots, i_k \\ \text{pairwise distinct}}} \varphi_k(x^{i_1}, \dots, x^{i_k}) + R_{k,N}, \end{aligned}$$

where $A_N^k := N!/(N-k)!$. The number of pairwise distinct tuples $(x^{i_1}, \dots, x^{i_k})$ of integers between 1 and N being A_N^k , this leads $|R_{k,N}| \leq \|\varphi_k\|_\infty (1 - A_N^k/N^k)$. Consequently, using the same argument

$$\left| \frac{1}{N!} \sum_{\sigma \in \mathfrak{S}_N} \varphi_k(x^{\sigma(1)}, \dots, x^{\sigma(k)}) - \langle \mu_{\mathbf{x}^N}^{\otimes k}, \varphi_k \rangle \right| \leq 2\|\varphi_k\|_\infty \left(1 - \frac{A_N^k}{N^k} \right).$$

The conclusion follows by noticing that

$$1 - \frac{A_N^k}{N^k} \leq 1 - \left(1 - \frac{k-1}{N} \right)^k \leq \frac{k(k-1)}{N},$$

and using Proposition 4. □

This (elementary) lemma is known at least since [91] where it was used to prove propagation of chaos (see Section 4.3). This lemma can also be seen as finite system version of the de Finetti theorem, see [68, Theorem 13]. The case of infinite systems is discussed in the following section. Note that the result of [68, Theorem 13] is actually an *existence* result for a measure $F^N \in \mathcal{P}(\mathcal{P}(E))$ which satisfies (56). Finally the empirical measure map is an isometry for the Wasserstein distance as shown in the following proposition.

Proposition 9 (Proposition 2.14 in [94]). *Let f^N, g^N be two symmetric probability measures on E^N and let $F^N = (\mu_N)_\# f^N$ and $G^N = (\mu_N)_\# g^N$ be the associated empirical law in $\mathcal{P}(\mathcal{P}(E))$. Then it holds that*

$$W_1(f^N, g^N) = \mathcal{W}_1(F^N, G^N).$$

This result also holds for the Wasserstein-2 distance [40, Lemma 11].

3.2.2 Infinite particle systems and random measures

In the previous section, finite exchangeable particle systems are described either by the marginal distributions or by the empirical measure. In this section the framework to take the limit $N \rightarrow +\infty$ is presented. An infinite set of exchangeable random variables $(\bar{X}^1, \bar{X}^2, \dots)$ is described by one of the two following objects.

1. The infinite hierarchy of marginals distributions $f^k \in \mathcal{P}_{\text{sym}}(E^k)$, $k \in \mathbb{N}$ such that

$$f^k = \text{Law}(\bar{X}^1, \dots, \bar{X}^k).$$

They satisfy the compatibility relation: for every $1 \leq j \leq k$,

$$\forall \varphi_j \in C_b(E^j), \quad \langle f^k, \varphi_j \otimes 1^{\otimes(k-j)} \rangle = \langle f^j, \varphi_j \rangle. \quad (57)$$

In other words, the j -particle marginal of f^k is f^j .

2. The infinite sequence of random empirical measures of size N , $N \in \mathbb{N}$,

$$\mu_{\bar{X}^N} = \frac{1}{N} \sum_{i=1}^N \delta_{\bar{X}^i}. \quad (58)$$

The two representations are linked by the de Finetti and Hewitt-Savage theorems stated below. Let us first state some preliminary useful results.

Given an infinite system of exchangeable particles $(\bar{X}^i)_{i \geq 1}$, important measurable events are given by two particular σ -algebras.

Definition 3.12 (Symmetric and asymptotic σ -algebras). Let $C_{\text{sym}}(E^N)$ denote the set of symmetric continuous $\mathbb{R}$ -valued functions on E^N which are invariant under permutations of their arguments.

- The σ -algebra of *exchangeable* events (i.e. events which do not depend on any finite permutation of the $\bar{X}^i$) is defined by:

$$\mathcal{S}_\infty := \bigcap_{k \geq 1} \sigma\left(\sigma(\varphi_k(\bar{X}^1, \dots, \bar{X}^k), \varphi_k \in C_{\text{sym}}(E^k)), \bar{X}^{k+1}, \bar{X}^{k+2}, \dots\right),$$

where we recall that $\sigma(X^1, X^2, \dots)$ is the σ -algebra generated by the random variables $X^1, X^2, \dots$.

- The *asymptotic* σ -algebra (whose events do not depend on any finite number of the $\bar{X}^i$) is defined by:

$$\mathcal{A}_\infty := \bigcap_{k \geq 1} \sigma(\bar{X}^{k+1}, \bar{X}^{k+2}, \dots).$$

The fundamental result for exchangeable systems is the following proposition.

Proposition 10 ([117]). *For exchangeable systems, the following equality holds*

$$\mathcal{S}_\infty = \mathcal{A}_\infty.$$

Corollary 2 (Hewitt-Savage 0-1 law). *In the special case where the $\bar{X}^i$ are i.i.d. variables (and then automatically exchangeable), then any event in the σ -algebra $\mathcal{S}_\infty$ or in the σ -algebra $\mathcal{A}_\infty$ has measure 0 or 1. This is known as the Kolmogorov 0-1 law for $\mathcal{A}_\infty$ and the Hewitt-Savage 0-1 law for $\mathcal{S}_\infty$.*

Since the empirical measures (58) are random measures, a criteria for the convergence in law in $\mathcal{P}(\mathcal{P}(E))$ is often needed; this motivate the following results. A thorough discussion of the theory of random measures can be found in [54]. An important notion is the notion of moment measure already introduced earlier and properly defined below.

Definition 3.13 (Moment measures). For $k \in \mathbb{N}$, the k -th moment measure of a measure $\pi \in \mathcal{P}(\mathcal{P}(E))$ is defined by:

$$\pi^k := \int_{\mathcal{P}(E)} \nu^{\otimes k} \pi(d\nu) = \mathbb{E}_{\nu \sim \pi} [\nu^{\otimes k}] \in \mathcal{P}(E^k).$$

This definition is understood in the weak sense, so that $\langle \pi^k, \varphi_k \rangle = \mathbb{E}_{\nu \sim \pi} \langle \nu^{\otimes k}, \varphi_k \rangle$ for any φ_k in $C_b(E^k)$.

Note that the sequence of moment measures $(\pi^k)_k$ satisfies the compatibility property. They also characterise the convergence in $\mathcal{P}(\mathcal{P}(E))$.

Lemma 3.14 (Convergence of random measures). *A sequence $(\pi_N)_N$ of random measures in $\mathcal{P}(\mathcal{P}(E))$ converges weakly towards $\pi \in \mathcal{P}(\mathcal{P}(E))$ if and only if*

$$\forall k \geq 1, \quad \pi_N^k \xrightarrow{N \rightarrow +\infty} \pi^k,$$

where the convergence is the weak convergence in $\mathcal{P}(E^k)$.

Proof. The direct implication stems from the fact the maps $\pi \mapsto \pi^k$ are continuous for the respective weak- $\star$ topologies. For the converse, the weak convergence of $(\pi_N^k)_N$ towards π^k implies that for all $\varphi_k \in C_b(E^k)$, $\langle \pi_N, R_{\varphi_k} \rangle \rightarrow \langle \pi, R_{\varphi_k} \rangle$ where R_{φ_k} is the monomial function:

$$R_{\varphi_k} : \nu \in \mathcal{P}(E) \mapsto \int_{E^k} \varphi_k(x^1, \dots, x^k) \nu^{\otimes k}(dx^1, \dots, dx^k) \in \mathbb{R}.$$

The conclusion follows from Lemma 3.9. □

The following lemma is a useful tightness criterion in $\mathcal{P}(\mathcal{P}(E))$; it can be found in [151, Proposition 2.2 (2.5)], where the first moment measures π^1 is referred as the *intensity measure* related to π (this terminology reminiscent of the intensity of a Poisson random measure).

Lemma 3.15 (Tightness for random measures). *The tightness of a sequence $(\pi_N)_N$ in $\mathcal{P}(\mathcal{P}(E))$ is equivalent to the tightness of the sequence $(\pi_N^1)_N$ in $\mathcal{P}(E)$.*

Proof. The direct implication stems from the fact the map $\pi \mapsto \pi^1$ is continuous for the respective weak- $\star$ topologies. For the converse, assume the $(\pi_N^1)_N$ is tight. For every

$\varepsilon > 0$, there exists a compact subset $K_\varepsilon^c \subset E$ such that $\pi_N^1(K_\varepsilon^c) \leq \varepsilon$ for every N . By the Markov inequality, for every $k \geq 1$ and every $N \geq 1$, it holds that

$$\pi_N \left(\left\{ \nu \in \mathcal{P}(E), \nu \left(K_{\varepsilon(k2^k)-1}^c \right) \geq \frac{1}{k} \right\} \right) \leq k \pi^1 \left(K_{\varepsilon(k2^k)-1}^c \right) \leq \frac{\varepsilon}{2^k},$$

so that

$$\pi_N \left(\bigcap_{k \geq 1} \left\{ \nu \in \mathcal{P}(E), \nu \left(K_{\varepsilon(k2^k)-1}^c \right) \leq \frac{1}{k} \right\} \right) \geq 1 - \sum_{k \geq 1} \frac{\varepsilon}{2^k} = 1 - \varepsilon.$$

Since the intersection at the last line is a compact subset of $\mathcal{P}(E)$, the sequence $(\pi_N)_N$ is tight. $\square$

Example 17 (The case of empirical measures). This lemma is particularly interesting for random empirical measures $\mu_{\mathcal{X}^N}$, since it reduces the question of tightness of $(\mu_{\mathcal{X}^N})_N$ in $\mathcal{P}(\mathcal{P}(E))$ to tightness of $(X^{1,N})_N$ in $\mathcal{P}(E)$.

The following theorems are the two main results of this section. The first one states that (the law of) a random measure can always be represented by an infinite exchangeable particle system. This theorem is due to de Finetti and can be found in [54, Theorem 11.2.1].

Theorem 3.16 (De Finetti representation theorem for random measures). *Let $\pi \in \mathcal{P}(\mathcal{P}(E))$. Then there exists a sequence $(\bar{X}^i)_{i \geq 1}$ of E -valued exchangeable random variables such that the following properties hold.*

- (1) *For any $k \geq 1$, $(\bar{X}^1, \dots, \bar{X}^k)$ has joint distribution π^k .*
- (2) *The weak limit*

$$\mu = \lim_{k \rightarrow +\infty} \frac{1}{k} \sum_{i=1}^k \delta_{\bar{X}^i} \in \mathcal{P}(E),$$

exists almost surely and μ is π -distributed.

- (3) *The random measure μ is $\mathcal{S}_\infty$ -measurable, and conditionally on $\mathcal{S}_\infty$ the random variables $\bar{X}^i$ are independent and μ -distributed.*

Example 18. A famous example is given in [54]: a de Finetti representation of the Fleming-Viot measure-valued process is given by the Moran particle system.

Note that the last property says that exchangeability implies *conditional* independence and thus exchangeable particles are not so far from i.i.d. variables.

Conversely, an infinite exchangeable particle system is always associated to a unique element in $\mathcal{P}(\mathcal{P}(E))$. The following theorem is also due to de Finetti in the case of Bernoulli random variables. It has been generalised to any exchangeable Borel measurable variables in a Polish space by Hewitt and Savage. The following quantitative version of the Hewitt-Savage theorem is due to [94, Theorem 5.1].

Theorem 3.17 (De Finetti, Hewitt-Savage). *Let E be a locally compact Polish space. Let $(f^N)_N$ be an infinite sequence of symmetric probability measures on E^N , $N \in \mathbb{N}$, which satisfy the compatibility relation (57). Then the following properties hold.*

(1) *There exists a unique $\pi \in \mathcal{P}(\mathcal{P}(E))$ such that:*

$$f^N = \pi^N := \int_{\mathcal{P}(E)} \nu^{\otimes N} \pi(d\nu).$$

(2) *When $E \subset \mathbb{R}^d$ is a Borel set, for any $s > d/2$, the sequence $(\text{Law}(\mu_{\mathcal{X}^N}))_{N \geq 1}$ is a Cauchy sequence in $\mathcal{P}(\mathcal{P}(E))$ for the distance $\mathcal{W}_{H^{-s}}$ (see Definition 3.5) : for any $N, M \geq 1$,*

$$\mathcal{W}_{H^{-s}}^2(\text{Law}(\mu_{\mathcal{X}^N}), \text{Law}(\mu_{\mathcal{X}^M})) \leq 2\|\Phi_s\|_\infty \left(\frac{1}{N} + \frac{1}{M} \right),$$

where Φ_s is defined by (46) and $\mathcal{X}^N \sim f^N$. The limit of this sequence is the measure π characterised above.

Proof (some ideas). The original argument of Hewitt and Savage is based on the Krein-Milman theorem and the fact that tensorised measures are extreme points of the convex set $\mathcal{P}_{\text{sym}}(E)$. A constructive quantitative approach due to Diaconis and Freedman is based on the approximation Lemma 3.11, see [68, Theorem 14] in the compact case. An alternative argument based on the density of polynomial functions in $\mathcal{P}(E)$ (thanks to the Stone-Weierstrass theorem) is due to Pierre-Louis Lions. We refer the interested reader to [143, Section 2.1] and the references therein. For the second point proved in [94, Theorem 5.1], the Cauchy-estimates relies on the polynomial structure of the H^{-s} -norm (46) combined with the observation that f^{N+M} is a transference plan between f^M and f^N (by the compatibility property). It turns the problem into controlling $\mathbb{E}\|\mu_{\mathcal{X}^N} - \mu_{\mathcal{X}^M}\|_{H^{-s}}^2$ for $(\mathcal{X}^N, \mathcal{X}^M) \sim f^N \otimes f^M$. Once convergence is shown, the limit is identified by the moment measures and Lemma 3.11. Convergence can be obtained in stronger metrics than $\mathcal{W}_{H^{-s}}^2$, see Corollary 5 in the appendix. $\square$

3.3 Kac's chaos

3.3.1 Definition and characterisation

The notion of chaos was introduced in the seminal article of Mark Kac [104].

Definition 3.18 (Kac's chaos). Let $f \in \mathcal{P}(E)$. A sequence $(f^N)_{N \geq 1}$ of symmetric probability measures on E^N is said to be *f-chaotic* when for any $k \in \mathbb{N}$ and any function $\varphi_k \in C_b(E^k)$,

$$\lim_{N \rightarrow +\infty} \langle f^N, \varphi_k \otimes 1^{\otimes N-k} \rangle = \langle f^{\otimes k}, \varphi_k \rangle. \quad (59)$$

It means that for all $k \in \mathbb{N}$, the k -th marginal satisfies $f^{k,N} \rightarrow f^{\otimes k}$ for the weak topology. Kac's chaos can be equivalently defined by considering only tensorized test functions $\varphi_k = \varphi^1 \otimes \dots \otimes \varphi^k$, since the algebra of tensorized functions in $C_b(E)$ is a convergence-determining class according to [75, Chapter 3, Theorem 4.5 and Proposition 4.6, pp.113-115].

Interpreting f^N as the law of an exchangeable system of N particles, the property (59) means that for any group of k particles, the particles become statistically independent as N tends to $+\infty$, hence the terminology of chaos. The results of the previous sections on finite and infinite exchangeable systems lead to the following useful characterization of Kac's chaos.

Lemma 3.19. *Each of the following assertions is equivalent to Kac's chaos.*

- (i) *There exists $k \geq 2$ such that $f^{k,N}$ converges weakly towards $f^{\otimes k}$.*
- (ii) *The random measure $\mu_{\mathcal{X}^N}$ with $\mathcal{X}^N \sim f^N$ converges in law towards the deterministic measure f .*

This classical result can be found in [151, Proposition 2.2].

Proof. Clearly, Kac's chaos implies (i). Then using Proposition 6, it can be proved that (i) $\Rightarrow$ (ii). Let $\mathcal{X}^N \sim f^N$. It is enough to prove that for any $\varphi \in C_b(E)$,

$$\mathbb{E}|\langle \mu_{\mathcal{X}^N} - f, \varphi \rangle|^2 \xrightarrow{N \rightarrow +\infty} 0.$$

Assume (59) with $k = 2$; it then also holds for $k = 1$. Using the symmetry of f^N , it holds that

$$\begin{aligned} \mathbb{E}|\langle \mu_{\mathcal{X}^N} - f, \varphi \rangle|^2 &= \frac{1}{N^2} \sum_{i,j=1}^N \mathbb{E}[\varphi(X^i)\varphi(X^j)] - \frac{2}{N} \langle f, \varphi \rangle \sum_{i=1}^N \mathbb{E}[\varphi(X^i)] + \langle f, \varphi \rangle^2 \\ &= \frac{1}{N} \mathbb{E}[\varphi(X^1)^2] + \frac{N-1}{N} \mathbb{E}[\varphi(X^1)\varphi(X^2)] - 2\langle f, \varphi \rangle \mathbb{E}[\varphi(X^1)] \\ &\quad + \langle f, \varphi \rangle^2, \end{aligned}$$

where the symmetry of f^N has been used. Since φ is bounded, the first term goes to 0 as $N \rightarrow \infty$. The remaining expression vanishes using (59) with $k = 1, 2$. This proves (ii). Then the condition (ii) implies Kac's chaos. If $F^N := \text{Law}(\mu_{\mathcal{X}^N}) \rightarrow \delta_f$, then according to Lemma 3.14, the k -th moment measure $F^{k,N}$ converges weakly towards $f^{\otimes k}$ for every $k \geq 1$. The approximation Lemma 3.11 implies that $f^{k,N}$ converges weakly towards $f^{\otimes k}$ for every $k \geq 1$. $\square$

Remark 13 (Chaos as a limit of de Finetti representations). Kac's chaos tells that the marginals $f^{k,N}$ converge towards the marginals of an infinite system $(\bar{X}^i)_{i \geq 1}$ of i.i.d. f -distributed particles. By the de Finetti and Hewitt-Savage theorems, the sequence of empirical measures of this latter system converges towards a random measure μ which is $\mathcal{S}_\infty = \mathcal{A}_\infty$ -measurable. By the Hewitt-Savage 0-1 law, this σ -algebra is trivial so that μ is a deterministic measure. The last part of de Finetti representation Theorem 3.16 tells that conditionally on $\mathcal{S}_\infty$, the $\bar{X}^i$ are μ -distributed so this allows to conclude $\text{Law}(\mu) = \delta_f$.

Remark 14 (Chaos as a law of large numbers). Fix φ in $C_b(E)$. Given a bounded continuous function $\theta : \mathbb{R} \rightarrow \mathbb{R}$, the function $\nu \mapsto \theta(\langle \nu, \varphi \rangle)$ is still bounded and weakly- $\star$ continuous on $\mathcal{P}(E)$. The convergence in law of $\mu_{\mathcal{X}^N}$ towards f thus implies that

$$\frac{\varphi(X^{1,N}) + \dots + \varphi(X^{N,N})}{N} - \mathbb{E}[\varphi(X^{1,N})] = \langle \mu_{\mathcal{X}^N}, \varphi \rangle - \langle f^{1,N}, \varphi \rangle \xrightarrow{N \rightarrow +\infty} 0,$$

where the convergence is the convergence in law. This relation is reminiscent of the law of large numbers. If the X^i were moreover i.i.d. (in this case no need to write $X^{i,N}$, X^i is enough) the law of large numbers would state

$$\langle \mu_{\mathcal{X}^N}, \varphi \rangle \xrightarrow{N \rightarrow +\infty} \mathbb{E}[\varphi(X^1)] \quad \text{a.s.},$$

so that almost surely $\mu_{\mathcal{X}^N} \rightarrow \text{Law}(X^1)$ weakly. In the general case where particles $X^{i,N}$ are only exchangeable (no more i.i.d.), Kac's chaos states an analogous but weaker result since the convergence of $\mu_{\mathcal{X}^N}$ towards f is only weak; but it however differs since f is the law of a typical particle in the limit system, and not the law of $X^{1,N}$ as in the i.i.d. case, because $X^{1,N}$ still depends on N (i.e. on the other particles). Fluctuations of $\langle \mu_{\mathcal{X}^N}, \varphi \rangle$ in the law of large numbers are described through the central limit theorem; the same can be done for chaos with concentration inequalities and large deviation principles (see Section II-5.4.1).

Remark 15 (Chaos, limit hierarchy and moment measures). Taking (formally) the limit $N \rightarrow \infty$ in the BBGKY hierarchy (Section 3.2.1) gives an infinite set of coupled equations on $(f_t^k)_{k \geq 1}$ which satisfy the compatibility relation (57). This system is Kac's chaotic when this limit hierarchy has the *factorisation property*, that is to say $f_t^k = f_t^{\otimes k}$ for every $k \geq 1$; this implies that the related $\mathcal{P}(\mathcal{P}(E))$ -representation of the system is δ_{f_t} . This infinite hierarchy is called the Boltzmann hierarchy in kinetic theory (see [45] and Section II-4.6). By the Hewitt-Savage Theorem 3.17 it is uniquely associated to an element $\pi \in \mathcal{P}(\mathcal{P}(E))$. We thus point out that the Boltzmann hierarchy coincides with the system of moment measures of π : this object is also commonly used, in another context, in the study of measure-valued processes [54].

The following property will be useful for time-dependent systems, its proof is straightforward, see [151, Proposition 2.4].

Proposition 11 (Chaos transportation). *Let $(f^N)_N$ be a f -chaotic sequence and let $\mathsf{T} : E \rightarrow F$ be a f -almost surely continuous map between Polish spaces. Then the sequence $(\mathsf{T}_\# f^N)_N$ is $\mathsf{T}_\# f$ -chaotic.*

3.3.2 Quantitative versions of Kac's chaos

Kac's chaos is a non quantitative property which relies only on the weak convergence. Quantitative (stronger) versions can naturally be defined using the topological framework of Section 3.1. The following definitions of quantitative chaos can be found in [94]. A strating point is the notion of chaos in Wasserstein distance. In practise, W_2 is well-adapted to the study of diffusion processes, while W_1 is often used for jump processes.

Definition 3.20 (Chaos in Wasserstein- p distance). Let $p \in \mathbb{N}$, let $(f^N)_N$ be a sequence of symmetric measures on E^N and let $f \in \mathcal{P}(E)$. The following three notions of chaos in Wasserstein- p distance were introduced in [94]:

- (Wasserstein- p Kac's chaos). For all $k \in \mathbb{N}$,

$$\Omega_k(f^N, f) := W_p(f^{k,N}, f^{\otimes k}) \xrightarrow{N \rightarrow +\infty} 0. \quad (60)$$

- (Infinite dimensional Wasserstein- p chaos).

$$\Omega_N(f^N, f) := W_p(f^N, f^{\otimes N}) \xrightarrow{N \rightarrow +\infty} 0. \quad (61)$$

- (Wasserstein- p empirical chaos). For $\mathcal{X}^N \sim f^N$,

$$\Omega_\infty(f^N, f) := \mathcal{W}_p(\text{Law}(\mu_{\mathcal{X}^N}), \delta_f) \xrightarrow{N \rightarrow +\infty} 0, \quad (62)$$

where $\mathcal{W}_p$ is a Wasserstein- p distance on $\mathcal{P}(\mathcal{P}(E))$ (see Definition 3.5 for the conventions used when $p = 1, 2$).

When moment bounds are available and $p = 1$, the three notions of chaos (60), (61) and (62) are actually equivalent. Such a result would not hold if the Wasserstein-1 distance were replaced by another Wasserstein- p distance.

Theorem 3.21 (Equivalence in Wasserstein-1 distance). *Let $E = \mathbb{R}^d$ and let $q \geq 1$ such that the sum of moments of order q of f and $f^{1,N}$ are bounded by a constant $\mathcal{M}_q \in (0, \infty)$. Then for any constant $\gamma < (d + 1 + d/q)^{-1}$, there exists $C = C(d, q, \gamma) \in (0, \infty)$ such that for any $k, \ell \in \{1, \dots, N\} \cup \{\infty\}$ with $\ell \neq 1$:*

$$\Omega_k(f^N, f) \leq C \mathcal{M}_q^{1/q} \left(\Omega_\ell(f^N, f) + \frac{1}{N} \right)^\gamma,$$

where Ω_k and Ω_ℓ are defined in Definition 3.20 with $p = 1$.

Proof (some ideas). See [94, Theorem 1.2] and [94, Theorem 2.4]. In particular, the link between (62) and (61) stems from Proposition 9 which states that

$$W_1(f^N, f^{\otimes N}) = \mathcal{W}_1(\text{Law}(\mu_{\mathcal{X}^N}), \text{Law}(\mu_{\bar{\mathcal{X}}^N})),$$

where $\bar{\mathcal{X}}^N \sim f^{\otimes N}$ -distributed. Given such $\bar{\mathcal{X}}^N$, $\mathcal{W}_1(\text{Law}(\mu_{\bar{\mathcal{X}}^N}), \delta_f) \leq \mathbb{E}W_1(\mu_{\bar{\mathcal{X}}}, f)$ and the quantitative laws of large numbers from [78] can be applied. $\square$

Definition 3.22 (Strong entropic and TV chaos). Stronger notions can also be defined using stronger norms.

- $(f^N)_N$ is f -TV chaotic when for every $k \geq 1$, $\|f^{k,N} - f^{\otimes k}\|_{\text{TV}} \rightarrow 0$ as $N \rightarrow +\infty$.
- $(f^N)_N$ is f -strong entropic chaotic when for every $k \geq 1$, $H(f^{k,N}|f^{\otimes k}) \rightarrow 0$ as $N \rightarrow +\infty$.

The second one is stronger than the first one by Pinsker's inequality (50).

In Definition (3.22) and in (60), a stronger convergence can be obtained when the fixed $k \in \mathbb{N}$ is replaced by a function $k \equiv k(N)$ which depends on N . In that case, the chaos is said to hold *for blocks of size $k(N)$* . The infinite dimensional chaos (62) corresponds to the case $k(N) = N$.

When $E = \mathbb{R}^d$ (or $E \subset \mathbb{R}^d$) is endowed with the Lebesgue measure denoted by σ below, other stronger versions of Kac's chaos can also be defined using the notions of entropy and Fisher information. The following notions can be found in [94].

Definition 3.23 (Entropy and Fisher chaos). Let σ^N denote the Lebesgue measure on E^N .

- $(f^N)_N$ is *f-entropy chaotic* when $f^{1,N} \rightarrow f$ weakly and $\frac{H(f^N|\sigma^N)}{N} \rightarrow H(f|\sigma)$.
- $(f^N)_N$ is *f-Fisher chaotic* when $f^{1,N} \rightarrow f$ weakly and $\frac{I(f^N|\sigma^N)}{N} \rightarrow I(f|\sigma)$.

Using sharp versions of the HWI inequality, these notions are classified in a quantitative way in [94].

Proposition 12. *Each of the below assertions implies the following.*

- $(f^N)_N$ is *f-Fisher chaotic*.
- $(f^N)_N$ is *f-Kac chaotic* with $\left(\frac{I(f^N|\sigma^N)}{N}\right)_N$ bounded.
- $(f^N)_N$ is *f-entropy chaotic*.
- $(f^N)_N$ is *f-Kac chaotic*.

In classical kinetic theory, another important notion of quantitative chaos arises when the N particles are constrained to evolve on the Kac's sphere:

$$E_N = \{\mathbf{v}^N \in \mathbb{R}^N, |v^1|^2 + \dots |v^N|^2 = N\} \subset \mathbb{R}^N,$$

or on the Boltzmann's sphere:

$$E_N = \{\mathbf{v}^N \in (\mathbb{R}^3)^N, |v^1|^2 + \dots |v^N|^2 = N, v^1 + \dots + v^N = 0\} \subset (\mathbb{R}^3)^N.$$

In these cases, the adapted notions of entropy chaos and Fisher chaos using a dedicated sequence of reference measures $(\sigma^N)_N$ are defined in [94] and [32].

3.4 Propagation of chaos

This section finally presents the central concept of this review, the notion of propagation of chaos, which is a dynamical version of Kac's chaos. Let us fix a final time $T \in [0, +\infty]$ and let us write $I = [0, T]$. Let $\mathcal{X}_I^N = (\mathcal{X}_t^N)_{t \in I}$ be a time-evolving (stochastic) càdlàg system of N -exchangeable particles in E with a f_0 -chaotic initial distribution $f_0^N \in \mathcal{P}(E^N)$ where $f_0 \in \mathcal{P}(E)$. One aims to compare the law of a typical particle with a limit flow of measures $(f_t)_{t \in I}$, where $f_t \in \mathcal{P}(E)$. The propagation of chaos property is said to hold when the initial chaos is propagated at later times. This property can hold either at the level of the law or at the level of trajectories.

Definition 3.24 (Pointwise and pathwise propagation of chaos). Let $f_0^N \in \mathcal{P}(E^N)$ be the initial f_0 -chaotic distribution of $\mathcal{X}_0^N$ at time $t = 0$.

- *Pointwise propagation of chaos* holds towards a flow of measures $(f_t)_t \in C(I, \mathcal{P}(E))$ when the law $f_t^N \in \mathcal{P}(E^N)$ of $\mathcal{X}_t^N$ is f_t -chaotic for every time $t \in I$. Note that the flow of measures is continuous in time as it is the solution of a PDE, but the (random) trajectories of the particles are càdlàg.
- *Pathwise propagation of chaos* holds towards a distribution $f_I \in \mathcal{P}(D(I, E))$ on the path space when the law $f_I^N \in \mathcal{P}(D(I, E)^N)$ of the process $\mathcal{X}_I^N$ (seen as a random element in $D(I, E)^N$) is f_I -chaotic.

The pointwise level is the analytical point of view where $(f_t)_t$ is the solution of a PDE. At the pathwise level, the limit distribution $f_{[0,T]}$ is often identified as the solution of a nonlinear martingale problem.

3.4.1 Quantitative and uniform in time propagation of chaos.

As in Section 3.3.2, it is possible to define quantitative versions of the propagation of chaos by using any of the quantitative notions of Kac's chaos. Then, one can wonder if the propagation of chaos holds uniformly in time, i.e. independently on T and with which rate of convergence in N . For instance, a typical quantitative pointwise propagation of chaos estimate reads:

$$\delta(f_t^{k,N}, f_t^{\otimes k}) \leq \varepsilon(N, k, T) \left(1 + \delta(f_0^{k,N}, f_0^{\otimes k})\right), \quad (63)$$

where δ is any of the distances on E^k defined in Section 3.1, $k \in \mathbb{N}$ is fixed, $t \in [0, T]$ and $\varepsilon(N, k, T) \rightarrow 0$ as $N \rightarrow +\infty$.

Propagation of chaos is said to hold uniformly in time when $\varepsilon(N, k, T)$ does not depend on T . As we shall see, it is usually possible to prove propagation of chaos uniformly in time only for physical models which enjoy some conservation properties. A closely related question when propagation of chaos holds on $I = [0, +\infty)$ is the ergodicity of the process as $t \rightarrow +\infty$. For instance, one may wonder if it possible to take the double limit $N \rightarrow +\infty$ and $t \rightarrow +\infty$ in (63). It would be possible for instance if propagation of chaos held uniformly in time and f_t converged towards an equilibrium f_∞ as $t \rightarrow +\infty$. This would give a relaxation estimate on $\delta(f_t^{k,N}, f_\infty^{\otimes k})$. This question is of particular importance in the study of the Boltzmann equation (38) in view of the famous H-theorem. Relaxation towards equilibrium at the particle level, will be mentioned in Section II-4.5 for the spatially homogeneous Boltzmann equation (44) and in Section II-3.1.3 for diffusion processes associated to the granular media and Vlasov-Fokker-Planck equations. On the other hand, it is sometimes possible to prove that propagation of chaos does *not* hold uniformly in time, an example is given in [33, 34].

Finally, for any quantitative version of propagation of chaos, it has been recently shown in [109] that the optimal rate of convergence is

$$\forall k \in \mathbb{N}, \forall \varphi_k \in C_b(E^k), \quad |\langle f_t^{k,N}, \varphi_k \rangle - \langle f_t^{\otimes k}, \varphi_k \rangle| = \mathcal{O}(k/N).$$

In particular, as we shall see in the following, many proofs and in particular the widely known proof of McKean's Theorem II-3.1 via coupling arguments (see Section 4.1) only yields the sub-optimal rate of convergence $\sqrt{k/N}$.

3.4.2 Pointwise from pathwise.

Pathwise propagation of chaos is more general since it keeps tracks of the whole trajectory of the particles. When pathwise propagation of chaos holds, it implies pointwise propagation of chaos: since the coordinate maps are continuous, this directly stems from Proposition 11 (it is also a consequence of the convergence of the finite dimensional distributions [75, Chapter 3, Theorem 7.8]). Note however, that this does not preserve the convergence rates. The converse does not always hold (see the counterexample below). In general, pathwise results are more difficult to obtain and can be proved only on finite time intervals. Pointwise propagation of chaos provides also more flexibility since it allows to work on $C(I, \mathcal{S})$, where $\mathcal{S}$ may be a subset of $\mathcal{P}(E)$ or a larger topological space. For instance, useful spaces to study fluctuations are the class of tempered distributions in [55] or negative weighted Sobolev spaces in [127]. In a more analytical perspective, when $(f_t)_t$ solves a known PDE, $\mathcal{S}$ is more naturally identified to a functional space, for instance a Sobolev space [103, 131].

3.4.3 Propagation of chaos via the empirical process.

The characterisation of Kac's chaos via the empirical measure given in Lemma 3.19 implies that pointwise propagation of chaos is equivalent to the convergence in law of the random measure $\mu_{\chi_t^N}$ towards the deterministic measure f_t for any $t \in [0, T]$. At the pathwise level, there are two notions of pathwise empirical propagation of chaos which are presented below. To begin with, a slightly more general definition of the empirical measure map is needed. Given a set $\mathcal{E}$ the empirical measure map is defined by:

$$\mu_N^{\mathcal{E}} : \mathcal{E}^N \rightarrow \mathcal{P}(\mathcal{E}), \quad \mathbf{x}^N \mapsto \mu_{\mathbf{x}^N}.$$

In the following, $\mathcal{E}$ is a Polish space; it will be either the state space E (in which case we may omit the superscript E as in (55)) or the path space. The following maps link the pathwise and pointwise properties.

- **(The evaluation map).** For any $t \in [0, T]$,

$$\mathbf{X}_t^{\mathcal{E}} : D([0, T], \mathcal{E}) \rightarrow \mathcal{E}, \quad \omega \mapsto \omega(t).$$

- **(The projection map).**

$$\Pi^{\mathcal{E}} : \mathcal{P}(D([0, T], \mathcal{E})) \rightarrow D([0, T], \mathcal{P}(\mathcal{E})), \quad \mu \mapsto ((\mathbf{X}_t^{\mathcal{E}})_{\#} \mu)_{0 \leq t \leq T}.$$

The pathwise and pointwise N -particle distributions are linked by

$$f_t^N = (\mathbf{X}_t^{E^N})_{\#} f_{[0, T]}^N = \Pi^{E^N}(f_{[0, T]}^N)(t).$$

The *empirical measure process* is the measure-valued process defined by:

$$(\mu_{\mathcal{X}_t^N})_t \equiv (\boldsymbol{\mu}_N^E(\mathcal{X}_t^N))_t.$$

Since this is the image of a process this readily defines the laws for all $t \in [0, T]$,

$$F_t^N := (\boldsymbol{\mu}_N^E)_\# f_t^N \in \mathcal{P}(\mathcal{P}(E)),$$

and the pathwise version:

$$F_{[0,T]}^{\mu,N} := (\boldsymbol{\mu}_N^E \circ)_\# f_{[0,T]}^N \in \mathcal{P}(D([0, T], \mathcal{P}(E))),$$

where $\boldsymbol{\mu}_N^E \circ$ is the natural extension of $\boldsymbol{\mu}_N^E$ on the path space defined by:

$$\boldsymbol{\mu}_N^E \circ : D([0, T], E^N) \rightarrow D([0, T], \mathcal{P}(E)), \quad \omega \mapsto \boldsymbol{\mu}_N^E \circ \omega.$$

Hence, it holds that

$$F_t^N = \Pi^{\mathcal{P}(E)} \left(F_{[0,T]}^{\mu,N} \right) (t) = (\mathbf{X}_t^{\mathcal{P}(E)})_\# F_{[0,T]}^{\mu,N}.$$

But there is another choice: it is also possible to define the pathwise empirical distribution as the push-forward of the N -particle pathwise distribution by the empirical map:

$$F_{[0,T]}^N := (\boldsymbol{\mu}_N^{\mathcal{D}})_\# f_{[0,T]}^N \in \mathcal{P}(\mathcal{P}(D([0, T], E))),$$

where we write $\mathcal{D} = D([0, T], E)$. This probability distribution is linked to F_t^N and $F_{[0,T]}^{\mu,N}$ by:

$$F_{[0,T]}^{\mu,N} = (\Pi^E)_\# F_{[0,T]}^N, \quad F_t^N = (\mathbf{X}_t^{\mathcal{P}(E)} \circ \Pi^E)_\# F_{[0,T]}^N.$$

In summary, there are three levels of description of the empirical process. In the following diagram, each space on the top row is a probability space endowed with a probability measure which is the law of the specific version of the random empirical process in the bottom row. The spaces are linked by the maps Π^E and $\mathbf{X}_t^{\mathcal{P}(E)}$.

$$\begin{array}{ccccc} \left(\mathcal{P}(D([0, T], E)), F_{[0,T]}^N \right) & \xrightarrow{\Pi^E} & \left(D([0, T], \mathcal{P}(E)), F_{[0,T]}^{\mu,N} \right) & \xrightarrow{\mathbf{X}_t^{\mathcal{P}(E)}} & \left(\mathcal{P}(E), F_t^N \right) \\ \mu_{\mathcal{X}_{[0,T]}^N} & \longmapsto & (\mu_{\mathcal{X}_t^N})_{0 \leq t \leq T} & \longmapsto & \mu_{\mathcal{X}_t^N} \end{array}.$$

This gives three notions of empirical propagation of chaos.

Definition 3.25 (Empirical propagation of chaos). Let $(f_t)_t \in C([0, T], \mathcal{P}(E))$ be a flow of measures and let $f_{[0,T]} \in \mathcal{P}(D([0, T], E))$ be such that for any $t \in [0, T]$,

$$f_t = \Pi^E(f_{[0,T]})(t).$$

There are three notions of empirical propagation of chaos defined below (the convergence is the weak convergence).

1. (*Pointwise empirical propagation of chaos*). For all $t \in [0, T]$, the law F_t^N satisfies

$$F_t^N \xrightarrow{N \rightarrow +\infty} \delta_{f_t} \in \mathcal{P}(\mathcal{P}(E)). \quad (64)$$

2. (*Functional law of large numbers*). The law $F_{[0,T]}^{\mu,N}$ satisfies

$$F_{[0,T]}^{\mu,N} \xrightarrow{N \rightarrow +\infty} \delta_{\Pi^E(f_{[0,T]})} \equiv \delta_{(f_t)_t} \in \mathcal{P}(D([0, T], \mathcal{P}(E))). \quad (65)$$

3. (*(Strong) pathwise empirical propagation of chaos*). The law $F_{[0,T]}^N$ satisfies

$$F_{[0,T]}^N \xrightarrow{N \rightarrow +\infty} \delta_{f_{[0,T]}} \in \mathcal{P}(\mathcal{P}(D([0, T], E))). \quad (66)$$

The (strong) pathwise property (66) is stronger than the functional law of large numbers (65) which is stronger than the pointwise property (64).

Remark 16. The functional law of large numbers (65) is also a pathwise property which is weaker than (66). To distinguish it with the pathwise empirical propagation chaos (66) we may occasionally call the latter property *strong pathwise empirical propagation of chaos*.

The implication (66) $\Rightarrow$ (65) is not straightforward because the map Π^E is not continuous everywhere. The result holds because the limit is a Dirac mass, this is proved in [127, Theorem 4.7] using a result of Léonard [114, Lemma 2.8]. The implication (65) $\Rightarrow$ (64) is more classical, this is the convergence of the finite dimensional distributions stated in [75, Chapter 3, Theorem 7.8].

The converse implications do not hold in general. In fact, since the map Π^E is not injective the strong pathwise property is meaningless when only the flow of measures $(f_t)_t$ is known and $f_{[0,T]}$ is not specified. A counterexample is given in [151, Chapter 1, Section 3(e)]. Sznitman builds two one-dimensional processes $\mathcal{X}_{[0,T]}^N$ and $\tilde{\mathcal{X}}_{[0,T]}^N$ such that for both processes, the strong pathwise empirical propagation chaos holds but with two different limits $f_{[0,T]}$ and $\tilde{f}_{[0,T]}$ and with the equality of the flows of time-marginals $\Pi^E(f_{[0,T]}) = \Pi^E(\tilde{f}_{[0,T]})$. The process $\tilde{\mathcal{X}}_{[0,T]}^N$ is obtained by a re-ordering procedure from $\mathcal{X}_{[0,T]}^N$ which thus ensures the equality of the empirical measure processes:

$$\forall t \in [0, T], \quad \mu_{\mathcal{X}_t^N} = \mu_{\tilde{\mathcal{X}}_t^N}.$$

Finally, by Lemma 3.19, the pointwise empirical propagation of chaos is equivalent to the pointwise propagation of chaos in the sense of Definition 3.24. Similarly, the strong pathwise empirical propagation of chaos is equivalent to the pathwise propagation of chaos. On the other hand the functional law of large numbers is an intermediate notion in between pointwise and pathwise propagation of chaos.

4 Proving propagation of chaos

Several methods are available to prove propagation of chaos. The choice of the method depends on several aspects, including the following ones.

- **How is the particle system defined?** A SDE representation allows to control directly the trajectory of each particle. If the system is an abstract Markov process defined by its generator only, one has to pass to the limit inside a “statistical object”: for instance the Liouville equation (1) or a martingale problem (Definition 2.2).
- **How well is the limit process known?** As mentioned in Section 2.1, it is sometimes possible to prove at the same time the propagation of chaos and an existence result for the limit object. Many “historical” proofs are based on this idea and exploit at the particle level a property of completeness (as in McKean’s original proof, Section II-3.1.1), a compactness criterion based on a martingale formulation (Section II-3.3) or an explicit series expansion for the solution of the Liouville equation in the case of a Boltzmann problem (as in Kac’s original proof, Section II-4.1). Over the years, the study of the limit problem, in parallel to the question of propagation of chaos, has stimulated the development of new techniques where wellposedness results or regularity properties of the limit problem are used to control the particle system. Ultimately, a trade-off has always to be done between regularity of the N -particle system and regularity of the limit process. The key idea is to write the N -particle system and the limit process in a common framework which allows to compare them.
- **Which kind of propagation of chaos?** In Sections 3.3 and 3.4, several notions of chaos and propagation of chaos are introduced. The first distinction to keep in mind is between pathwise and pointwise properties. Then one may seek quantitative estimates. Pathwise chaos is stronger and it is often simpler to get pointwise quantitative estimates.

Keeping these aspects in mind, the present section is organised as follows. Section 4.1 is devoted to an introduction of coupling methods in several cases. These methods (or most of them) exploit a SDE representation of the particle system (it therefore requires some regularity at the microscopic level and often a wellposedness result for the limit system) and lead to the quantitative pathwise or pointwise propagation of chaos. Section 4.2 introduces some ideas to prove the tightness (and thus the compactness) of the law of the empirical process. This leads to non-quantitative pathwise propagation of chaos results, but as it is only based on the properties of the generator of the N -particles process, it remains valid for a wide class of models. A pointwise study of the empirical process via the asymptotic analysis of its generator is described in Section 4.3. This leads to a quantitative abstract theorem with a comparable range of applications as the compactness methods. In Section 4.4, some ideas related to large deviations are presented. It leads to strong (non quantitative) abstract results which go beyond but include the propagation of chaos. Although these results are often too strong or too abstract to be used in practise, the ideas can be reinterpreted to prove propagation of chaos for particle systems with

a very weak regularity or with a complex interaction mechanism which are difficult to handle with other methods. Finally, in the case of Boltzmann models, specific tools can be used as described in Section 4.5.

We recall that several applications of these methods will be presented in the second part of this review, in particular in Sections II-3 and II-4.

4.1 Coupling methods

4.1.1 Definition

Definition 4.1 (Chaos by coupling the trajectories). Let be given a final time $T \in (0, \infty]$, a distance d_E on E and $p \in \mathbb{N}$. Propagation of chaos holds by coupling the trajectories when for all $N \in \mathbb{N}$ there exist

- a system of particles $(\mathcal{X}_t^N)_t$ with law $f_t^N \in \mathcal{P}(E^N)$ at time $t \leq T$,
- a system of *independent* processes $(\bar{\mathcal{X}}_t^N)_t$ with law $f_t^{\otimes N} \in \mathcal{P}(E^N)$ at time $t \leq T$,
- a number $\varepsilon(N, T) > 0$ such that $\varepsilon(N, T) \xrightarrow{N \rightarrow +\infty} 0$,

such that (pathwise case)

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left[\sup_{t \leq T} d_E(X_t^i, \bar{X}_t^i)^p \right] \leq \varepsilon(N, T), \quad (67)$$

or (pointwise case)

$$\frac{1}{N} \sum_{i=1}^N \sup_{t \leq T} \mathbb{E} [d_E(X_t^i, \bar{X}_t^i)^p] \leq \varepsilon(N, T). \quad (68)$$

By definition of the Wasserstein- p distance (Definition 3.1) and Jensen inequality, the bounds (67) and (68) imply the infinite dimensional chaos (Definition 3.20), respectively:

$$W_p \left(f_{[0,T]}^N, f_{[0,T]}^{\otimes N} \right) \xrightarrow{N \rightarrow +\infty} 0, \quad \sup_{t \leq T} W_p(f_t^N, f_t^{\otimes N}) \xrightarrow{N \rightarrow +\infty} 0,$$

where we recall that the Wasserstein- p distance is associated to the normalised distance on E^N (Definition 3.5). The definition can also be weakened by assuming that only the first $k(N) < N$ particles are coupled instead of the whole system of the N particles. This would imply propagation of chaos of the form (60).

Remark 17. Note that by exchangeability of the particles all the expectations in the sums (67) and (68) are equal and the assertions therefore imply the convergence of the one-particle distribution in Wasserstein- p distance on $\mathcal{P}(E)$.

The coupling between the trajectories also implies the quantitative empirical chaos stated in the following lemma in the pointwise case. This is a simple application of [78].

Lemma 4.2. *Let $E = \mathbb{R}^d$ and assume that f_t has a bounded moment of order $q > p$. If there exists a (pointwise) coupling as in Definition 4.1 then*

$$\sup_{t \leq T} \mathcal{W}_p^p(F_t^N, \delta_{f_t}) \leq \varepsilon(N, T) + \beta_d(N),$$

where $\beta_d(N)$ is given by:

$$\begin{aligned} \beta_d(N) &= C(p, q) \begin{cases} N^{-1/2} + N^{-(q-p)/q} & \text{if } p > d/2 \text{ and } q \neq 2p \\ N^{-1/2} \log(1 + N) + N^{-(q-p)/q} & \text{if } p = d/2 \text{ and } q \neq 2p \\ N^{-p/d} + N^{-(q-p)/q} & \text{if } p < d/2 \text{ and } q \neq d/(d-p), \end{cases} \end{aligned}$$

for a constant $C(p, q) > 0$ which depends only on p and q .

Proof. By the triangle inequality, for all $t \leq T$,

$$\begin{aligned} \mathcal{W}_p^p(f_t^{\mu, N}, \delta_{f_t}) &\leq \mathbb{E} W_p^p(\mu_{\mathcal{X}_t^N}, f_t) \\ &\leq \mathbb{E} W_p^p(\mu_{\mathcal{X}_t^N}, \mu_{\bar{\mathcal{X}}_t^N}) + \mathbb{E} W_p^p(\mu_{\bar{\mathcal{X}}_t^N}, f_t). \end{aligned}$$

The second term on the right-hand side of the last inequality is bounded by $\beta_d(N)$ using [78, Theorem 1]. Moreover the bound (68) implies that

$$\mathbb{E} W_p^p(\mu_{\mathcal{X}_t^N}, \mu_{\bar{\mathcal{X}}_t^N}) \leq \varepsilon(N, T).$$

□

Note that the convergence rate depends on the dimension. When moments of sufficiently high order are available, $\beta_d(N)$ is of the order $N^{-1/2}$ for $p > d/2$ and $N^{-p/d}$ for $p < d/2$.

We now summarise the most common methods used to construct a coupling in the sense of Definition 4.1.

4.1.2 Synchronous coupling.

When the particle system can be written as the solution of a system of SDEs as in (11) or (20), a simple coupling choice consists in constructing N nonlinear processes by taking respectively the same Brownian motions and Poisson random measures as those defining the particle system. This choice is called *synchronous coupling*.

For the McKean-Vlasov diffusion, the synchronous coupling is thus based on the N independent processes $\bar{\mathcal{X}}_t^N = (\bar{X}_t^1, \dots, \bar{X}_t^N)$ defined as the solutions of the N SDEs:

$$d\bar{X}_t^i = b(\bar{X}_t^i, f_t)dt + \sigma(\bar{X}_t^i, f_t)dB_t^i, \quad (69)$$

for $i \in \{1, \dots, N\}$ where $(B_t^i)_t$ is the same Brownian motion as in (11) and where we recall that $f_t = \text{Law}(\bar{X}_t^i)$. Since the Brownian motions B_t^i are independent, this gives N independent copies of (13). The Theorem II-3.1 (and Theorem II-3.20) in Section II-3.1.1

will show that this coupling choice leads to an optimal convergence rate (in N) for any $T > 0$ but with a constant which depends exponentially in T . This comes from the fact that comparing the trajectories of (69) and (11) is similar to a stability analysis of the N processes $X_t^i - \bar{X}_t^i$. When the coefficients are globally Lipschitz, the classical Gronwall-based methods imply the stability on any time interval but with a constant which grows exponentially with the time variable.

The synchronous coupling is by far the most popular choice of coupling method in the literature since [151]. We point out that this was not the original choice of McKean: in the seminal work [125], McKean uses a synchronous coupling between the N -particle system and the subsystem of the N first particles of a system of $M > N$ particles. The proof thus does not necessitate to prove the well-posedness of the nonlinear SDE (13) as a preliminary step (since it is never used). It is actually a proof of existence which constructs a solution of (69) by a completeness argument together with a probabilistic reasoning (based on Hewitt-Savage 0-1 law) to recover the independence.

Similar ideas can be applied for parametric mean-field jump processes (with or without simultaneous jumps). Given the N SDEs (20), the synchronous coupling is defined by:

$$\begin{aligned} \bar{X}_t^i &= X_0^i + \int_0^t a(\bar{X}_s^i) ds \\ &\quad + \int_0^t \int_0^{+\infty} \int_{\Theta} \{ \psi(\bar{X}_{s-}^i, f_s, \theta) - \bar{X}_{s-}^i \} \mathbb{1}_{(0, \lambda(\bar{X}_{s-}^i, f_s)]}(u) \mathcal{N}^i(ds, du, d\theta), \end{aligned} \quad (70)$$

for $i \in \{1, \dots, N\}$, where $f_t = \text{Law}(\bar{X}_t^i)$ and where the Poisson random measures $\mathcal{N}^i$ are the same as in (20). Equation (70) can be extended straightforwardly to the case of simultaneous jumps (Example 5). It is then possible to prove similar results as in the case of the McKean-Vlasov diffusion. A complete analysis can be found in [4] (see also Section II-3.6).

To end this section, let us also mention the recent coupling method introduced in [95] which reverses the role of the empirical particle system and the nonlinear law. The author introduces the particle system defined conditionally on $\mathcal{X}_t^N$ by:

$$d\tilde{X}_t^i = b(\tilde{X}_t^i, \mu_{\mathcal{X}_t^N}) dt + \sigma(\tilde{X}_t^i, \mu_{\mathcal{X}_t^N}) d\tilde{B}_t^i.$$

Note that the processes $\tilde{X}_t^i$ are not independent. More details on how to close the argument (using a generalised Glivenko-Cantelli theorem for SDEs) will be given in Section II-3.2.2. The main advantage is that it allows more singular interactions, namely only Hölder instead of Lipschitz.

Remark 18. Coupling methods should still be possible when no SDE is available, since it is always possible to write an evolution equation for an observable $\varphi(\mathcal{X}_t^N)$ using the general Itô's formula for Markov processes (see Section A.3.4). The analog of an SDE can then be recovered taking for φ some coordinate functions.

4.1.3 Reflection coupling for the McKean-Vlasov diffusion.

Except in specific cases (see Section II-3.1.3), it is usually not possible or difficult to get uniform in time estimates using a synchronous coupling. One reason is that the strategy

can be seen as a stability analysis for the nonlinear system (13) which classically leads to Gronwall type estimates with a constant which depends exponentially in T . The recently developed *reflection coupling* [72, 73] tries to make a better use of the diffusion part to get (hopefully) uniform in times estimates under mild assumptions. This type of coupling was originally introduced in [118] to study the ergodic properties of diffusion processes in total variation norm. In this context, the coupling time of two diffusion processes with same transition probabilities but different initial distributions is defined as the first time where the two processes hit the same state. The reflection coupling method provides an explicit construction of the processes which (hopefully) ensures that their coupling time is almost surely finite. By letting the two processes being equal after the coupling time, this fact shows that the two processes are asymptotically equal in law. The idea of reflection coupling is easily understood for two 1D Brownian motions starting from the initial conditions $a > 0$ and $-a$: it is sufficient to take one Brownian motion equal to the opposite of the other one. More generally, the solution (X_t, Y_t) of the following system of SDEs in $\mathbb{R}^d$ is called a coupling by reflection:

$$\begin{aligned} dX_t &= b(X_t)dt + \sigma dB_t \\ dY_t &= b(Y_t)dt + \sigma(I_d - 2e_t e_t^T)dB_t, \end{aligned}$$

where $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is the (locally Lipschitz) drift function, $\sigma > 0$ is a constant and

$$e_t = (X_t - Y_t)/|X_t - Y_t|.$$

Moreover, after the coupling time $T := \inf\{t \geq 0, X_t = Y_t\}$, for $t \geq T$ the processes are set to $X_t = Y_t$. In a series of recent works, [72, 73], Eberle et al. have proved a quantitative contraction result in Wasserstein distance for the law of the two processes. Let b satisfy for all $x, y \in \mathbb{R}^d$,

$$\langle x - y, b(x) - b(y) \rangle \leq -\frac{\sigma^2}{2} \kappa(|x - y|) |x - y|^2,$$

where $\kappa : \mathbb{R}_+ \rightarrow \mathbb{R}$ satisfies $\liminf_{r \rightarrow +\infty} \kappa(r) > 0$. In order to measure the discrepancy between the two processes $r_t := |X_t - Y_t|$, for any fixed smooth function f , Itô's formula gives:

$$df(r_t) = r_t^{-1} \langle X_t - Y_t, b(X_t) - b(Y_t) \rangle f'(r_t) dt + 2\sigma^2 f''(r_t) dt + dM_t,$$

where M_t is a martingale. Compared to what the synchronous coupling would give, thanks to Itô's correction, the coupling by reflection adds a new term in the drift. Using the assumptions on b , the drift term is now bounded by

$$2\sigma^2 \left(f''(r_t) - \frac{1}{4} r_t \kappa(r_t) f'(r_t) \right).$$

If f is such that there exists $c > 0$ such that for all $r \geq 0$

$$f''(r) - \frac{1}{4} r \kappa(r) f'(r) \leq -\frac{c}{2\sigma^2} f(r), \quad (71)$$

then it gives:

$$\mathbb{E}[f(r_t)] \leq e^{-ct} \mathbb{E}[f(r_0)]. \quad (72)$$

The idea of [72] is to introduce a positive concave function f so that $d_f(x, y) := f(|x - y|)$ defines a distance on $\mathbb{R}^d$ and such that the bound (71) holds. From (72), this finally gives the following exponential “contraction bound” in Wasserstein distance:

$$W_{1,d_f}(\mu_t, \nu_t) \leq e^{-ct} W_{1,d_f}(\mu_0, \nu_0), \quad (73)$$

where $\mu_t, \nu_t \in \mathcal{P}(E)$ are the laws of X_t, Y_t at time $t \geq 0$. This strategy is successfully applied in [72, 73] to get quantitative contraction and convergence rates for linear and nonlinear gradient McKean-Vlasov systems, even in non convex settings.

Coming back to particle systems, in [71, 119], the authors have shown that this idea can be applied to McKean-Vlasov systems (11) (which can be seen as a classical diffusion equation in the high-dimensional space $\mathbb{R}^{dN}$). They use a componentwise reflection coupling between a particle system and a system of N independent nonlinear McKean-Vlasov systems. The analog of the exponential contraction rate (73) thus provides a proof of the uniform in time propagation of chaos for gradient systems with milder assumptions than the ones in Section II-3.1.3 obtained with a synchronous coupling. This will be reviewed in Section II-3.2.1.

Remark 19 (Extension to more general diffusions). This idea is more natural but not restricted to the case where the diffusion matrix is constant. It can be extended to more general diffusion matrices by “twisting” the metric in $\mathbb{R}^d$ to recover a constant diffusion matrix in a modified metric. See [72] for additional details as well as [121] for a similar reasoning in a different context.

4.1.4 Optimal jumps

For mean-field jump processes and Boltzmann models, the particles interact only at discrete (random) times. They update their state according to a sampling mechanism with respect to a known measure which depends on the empirical measure of the system (see (18) and (27)). The strategy adopted in [69] for mean-field jump processes (see Section 2.2.3) consists in constructing a trajectorial representation of the particle and the nonlinear systems in which the jumps are coupled optimally. Taking the same sequence of jump times $(T_n^i)_n$ for the particle X_t^i and its coupled nonlinear version $\bar{X}_t^i$, a post-jump state is sampled for the nonlinear process first:

$$\bar{X}_{T_n^i}^i \sim P_{f_{T_n^i}}(\bar{X}_{T_n^{i-}}^i, dy),$$

and then the post-jump state for the particle is defined as the image:

$$X_{T_n^i}^i = \mathsf{T}(\bar{X}_{T_n^i}^i), \quad (74)$$

where T is an optimal transfer map for the W_1 distance between the jump measures:

$$\mathsf{T}_\# P_{f_{T_n^i}}(\bar{X}_{T_n^{i-}}^i, dy) = P_{\mu_{X_{T_n^{i-}}^N}}(X_{T_n^i}^i, dy),$$

and

$$\mathbb{E}\left[\left|\bar{X}_{T_n^i}^i - X_{T_n^i}^i\right|\middle|\mathcal{F}_{T_n^{i-}}\right] = W_1\left(P_{f_{T_n^i}}\left(\bar{X}_{T_n^{i-}}^i, dy\right), P_{\mu_{\mathcal{X}_{T_n^{i-}}^N}}\left(X_{T_n^{i-}}^i, dy\right)\right),$$

where $\mathcal{F}_t$ denotes the filtration generated by all the processes $(X_t^i)_t$ and $(\bar{X}_t^i)_t$ up to time t and $\mathcal{F}_{T_n^{i-}} = \sigma(\bigcup_{t < T_n^i} \mathcal{F}_t)$. The existence of the optimal transfer map T is a classical result in optimal transport theory; it holds under mild assumptions, see [46, 76] and the references therein. Under Lipschitz assumptions on the jump measure, it can be deduced that

$$\mathbb{E}\left[\left|\bar{X}_{T_n^i}^i - X_{T_n^i}^i\right|\middle|\mathcal{F}_{T_n^{i-}}\right] \leq C\left(\left|\bar{X}_{T_n^{i-}}^i - X_{T_n^{i-}}^i\right| + W_1\left(f_{T_n^i}, \mu_{\mathcal{X}_{T_n^{i-}}^N}\right)\right). \quad (75)$$

This crude estimate may be refined when the jump measure has a known expression. Note that in the case of a parametric model (Example 4), the synchronous coupling of the Poisson random measures (70) gives an alternative coupling and an explicit transfer map:

$$\bar{X}_{T_n^i}^i = \psi\left(\bar{X}_{T_n^{i-}}^i, f_{T_n^i}, \theta\right), \quad X_{T_n^i}^i = \psi\left(X_{T_n^{i-}}^i, \mu_{\mathcal{X}_{T_n^{i-}}^N}, \theta\right), \quad (76)$$

where $\theta \sim \nu(d\theta)$ is the same random variable for the two post-jump states. This coupling is not necessarily optimal but under Lipschitz assumptions on ψ it still implies

$$\begin{aligned} \mathbb{E}\left[\left|\bar{X}_{T_n^i}^i - X_{T_n^i}^i\right|\middle|\mathcal{F}_{T_n^{i-}}\right] &= \int_{\Theta} \left|\psi\left(\bar{X}_{T_n^{i-}}^i, f_{T_n^i}, \theta\right) - \psi\left(X_{T_n^{i-}}^i, \mu_{\mathcal{X}_{T_n^{i-}}^N}, \theta\right)\right| \nu(d\theta) \\ &\leq C\left(\left|\bar{X}_{T_n^{i-}}^i - X_{T_n^{i-}}^i\right| + W_1\left(f_{T_n^i}, \mu_{\mathcal{X}_{T_n^{i-}}^N}\right)\right). \end{aligned}$$

Both couplings (74) and (76) ensure that the N nonlinear processes $(\bar{X}_t^i)_t$ remain independent, which is crucial. At each jumping time T_n^i , the error between X_t^i and $\bar{X}_t^i$ due to the jump is controlled by (75). A discrete stability analysis then ensures the propagation of the error exponentially in time. Between the jumps, the trajectories are either deterministic or can be controlled by a standard synchronous coupling. This proves the propagation of chaos on any time interval but with a (very) bad behaviour with respect to time.

For Boltzmann models, the situation is more difficult because two particles “jump” at the same time. As discussed in Section 2.3.2, it is also not completely straightforward to build a SDE representation of the nonlinear process. Moreover, contrary to the mean-field jump processes where the jump measures are typically assumed to have a smooth density, for Boltzmann models, the jumps are obtained by sampling directly from the empirical measure of the system. Since this measure is singular but the solution of the Boltzmann equation is not, an optimal transfer map may not exist. The strategy adopted in [133] and then in [48, 49, 81] is based on the analysis of the optimal transfer plan (which exists) between the empirical measure of the particle system and the law of the nonlinear system at each jump time. This strategy also needs a kind of synchronous coupling for the jump times. As in the previous cases, propagation of chaos then results from a Gronwall type estimate with a bad behaviour in time, and in this case only for marginals (or block) of size $k(N) = o(N)$. This will be discussed more thoroughly in Section II-4.4.

4.1.5 Analysis in Wasserstein spaces: optimal coupling

The Liouville equation associated to a McKean-Vlasov process with interaction parameters

$$b(x, \mu) \equiv \int_{\mathbb{R}^d} b(x, y) \mu(dy), \quad \sigma(x, \mu) \equiv \sigma I_d, \quad \sigma > 0,$$

can be written:

$$\partial_t f_t^N = -\nabla \cdot (\mathbf{b}^N f_t^N) + \frac{\sigma^2}{2} \Delta f_t^N, \quad (77)$$

where

$$\mathbf{b}^N : \mathbf{x}^N \in \mathbb{R}^{dN} \mapsto \left(\frac{1}{N} \sum_{i=1}^N b(x^1, x^i), \dots, \frac{1}{N} \sum_{i=1}^N b(x^N, x^i) \right) \in \mathbb{R}^{dN}.$$

This equation can be rewritten as a continuity equation:

$$\partial_t f_t^N = -\nabla \cdot \left(\left(\mathbf{b}^N - \frac{\sigma^2}{2} \nabla \log f_t^N \right) f_t^N \right). \quad (78)$$

Similarly, the associated nonlinear McKean-Vlasov equation is a continuity equation in $\mathbb{R}^d$ with velocity vector

$$b \star f_s - \frac{\sigma^2}{2} \nabla \log f_s.$$

Continuity equations are strongly linked to the theory of gradient flows [3, 53], which in turn provides new insights on the study of McKean-Vlasov processes. In addition to the present section, see also Section II-3.3.2.

The two recent works [145, 67] are based on a classical result in gradient flow theory (see for instance [161, Theorem 23.9]) which gives an explicit dissipation rate between the solutions of two continuity equations. In the present case, it gives an explicit control of the time derivative of $W_2(f_t^N, f_t^{\otimes N})$ in terms of the well-known maximizing Kantorovich potential ψ_t^N which links the two laws by:

$$(\nabla \psi_t^N)_\# f_t^{\otimes N} = f_t^N, \quad W_2^2(f_t^N, f_t^{\otimes N}) = \int_{\mathbb{R}^{dN}} |\nabla \psi_t^N(\mathbf{x}^N) - \mathbf{x}^N|^2 f_t^{\otimes N}(d\mathbf{x}^N).$$

The existence of ψ_t^N is ensured by Brenier's theorem [161, Theorem 9.4]. This approach of propagation of chaos follows the work of [19, 20] where the authors have derived explicit contraction rates in Wasserstein-2 distance for linear Fokker-Planck equations and for the nonlinear granular media equation (which is the nonlinear mean-field limit associated to the gradient system (15)). Starting from the same result [161, Theorem 23.9], the authors of [19, 20] introduced a new transportation inequality, the so-called WJ inequality which is then exploited at the particle level in [145, 67]. This analysis provides uniform in time propagation of chaos and convergence to equilibrium results for gradient systems in non globally convex settings. The work of [145] also provides a new unifying analytical vision of previous coupling approaches. These techniques will be detailed in Section II-3.2.3.

4.2 Compactness methods

In this section the main ideas to prove propagation of chaos via compactness arguments are presented. The main advantage of this approach is its wide range of applicability as it can be adapted to jump, diffusion, Boltzmann or even mixed models. The main drawback is that it does not provide any convergence rate.

Compactness methods are based on the empirical representation of the process described in Section 3.4. It therefore reduces the problem of the convergence of a sequence of probability measures on a space which does not depend on N but which in turn is much more delicate to handle than E as it is itself a probability space. The first approach described below is the stochastic analysis approach which is by now classical; the second approach is a more recent analytical approach based on the theory of gradient-flows.

4.2.1 Martingale methods.

Starting from the martingale characterisation of the particle system (Definition 2.2), it is possible to prove the functional law of large numbers (65) and the strong pathwise empirical propagation of chaos (66) using the traditional sequence of arguments in stochastic analysis (see for instance [101]).

- (1) First prove the tightness of the sequence $(F_{[0,T]}^{\mu,N})_N$ (functional law of large numbers) or $(F_{[0,T]}^N)_N$ (strong pathwise). This will come from usual tightness criteria (see Section A.2 and Section II-C) but requires some care regarding the spaces (namely, the path space with values in a set of probability measures). The classical and more advanced tools which are used are reminded in Appendix A. By Prokhorov theorem, it is then possible to extract a converging subsequence towards a limit, respectively $\pi \in \mathcal{P}(D([0,T], \mathcal{P}(E)))$ (functional law of large numbers) or $\pi \in \mathcal{P}(\mathcal{P}(D([0,T], E)))$ (strong pathwise).
- (2) Then identify the π -distributed limit points as solutions respectively of the weak limit PDE or the limit martingale problem. Again, this may require some care, in particular for càdlàg processes due to the topology of the Skorokhod space.
- (3) Finally prove the uniqueness of the solution of the previous problem. Usually well-posedness (that is existence and uniqueness) can be proved beforehand although existence is not required (it is automatically provided by the tightness). In conclusion, π is a Dirac delta at the desired limit point.

Strong pathwise empirical chaos (66) is not significantly harder to prove than the weaker functional law of large numbers (65) but it requires the uniqueness property of the limit martingale problem which is a stronger assumption than the corresponding one for the weak limit PDE. The strong pathwise case is detailed in Méléard's course [127, Section 4].

This approach is historically linked to the study of the spatially homogeneous Boltzmann equation of rarefied gas dynamics (44): Tanaka [154] proved weak pathwise empirical chaos for hard-spheres and inverse power Maxwellian molecules. Strong pathwise

empirical chaos is proved in [149] for a class of parametric Boltzmann models which includes the stochastic hard-sphere model. See also [162] for a functional law of large numbers applied to a large class of parametric Boltzmann models. The martingale method is exploited to treat the case of the McKean-Vlasov diffusion in [150] (with boundary conditions) and in [136, 84] (functional law of large numbers with general interaction functions), see also [113]. Strong pathwise empirical chaos is proved for a mixed jump-diffusion mean-field model in [89, 127]. See also [80, Theorem 4.1] for a strong pathwise result on a cutoff approximation of a non-cutoff Boltzmann model.

The corresponding results will be detailed in Section II-3.3.1 in the mean-field case and in Section II-4.3 for Boltzmann models.

4.2.2 Gradient flows.

This second approach gives a pointwise version of the empirical propagation of chaos and is restricted to the McKean-Vlasov gradient system (15). It is entirely analytical and exploits recent results of the theory of gradient flows [3, 53, 161]. We briefly recall below one definition of gradient-flows in a metric space.

Definition 4.3 (Gradient flows in $(\mathcal{E}, d_{\mathcal{E}})$). Let $(\mathcal{E}, d_{\mathcal{E}})$ be a geodesic metric space.

1. **(Absolutely continuous curves and metric derivative).** A $\mathcal{E}$ -valued continuous curve $\mu : (a, b) \subset \mathbb{R} \rightarrow \mathcal{E}$ is said to be absolutely continuous whenever there exists $m \in L^1_{\text{loc}}(a, b)$ such that

$$\forall a < s \leq t < b, \quad d_{\mathcal{E}}(\mu_s, \mu_t) \leq \int_s^t m(r) dr.$$

In this case, the limit

$$|\mu'| (t) = \lim_{s \rightarrow t} \frac{d_{\mathcal{E}}(\mu_s, \mu_t)}{|t - s|},$$

exists for almost every $t \in (a, b)$ and is called the metric derivative of μ at the point t .

2. **(Gradient flow).** Let $T \in (0, +\infty]$ and let $\mu \in C([0, T], \mathcal{E})$ be an absolutely continuous curve. Let us consider $\lambda \in \mathbb{R}$ and a λ -convex map $\mathcal{F} : \mathcal{E} \rightarrow \mathbb{R} \cup \{+\infty\}$. Then μ is called a λ -gradient flow associated to the energy $\mathcal{F}$ whenever $\mathcal{F}(\mu_t) < +\infty$ for all $t \in [0, T)$ and μ satisfies the following *Evolution Variational Inequality* (EVI) :

$$\forall \text{a.e. } t \in [0, T), \quad \forall \nu \in \mathcal{E}, \quad \frac{1}{2} \frac{d}{dt} d_{\mathcal{E}}^2(\mu_t, \nu) + \frac{\lambda}{2} d_{\mathcal{E}}^2(\mu_t, \nu) \leq \mathcal{F}(\nu) - \mathcal{F}(\mu_t). \quad (79)$$

In the following, gradient flows will be considered in the two cases $(\mathcal{E}, d_{\mathcal{E}}) = (\mathcal{P}_2(\mathbb{R}^d), W_2)$ and $(\mathcal{E}, d_{\mathcal{E}}) = (\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d)), \mathcal{W}_2)$. The fundamental result to keep in mind is that evolutionary PDEs, as defined below, have a unique distributional solution which is a gradient-flow.

Definition 4.4 (Evolutionary PDEs). An evolutionary PDE is a PDE of the form

$$\partial_t \rho = \nabla \cdot \left(\rho \nabla \frac{\delta \mathcal{F}(\rho)}{\delta \rho} \right),$$

where $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R} \cup \{+\infty\}$ and where the first variation of $\mathcal{F}$ is defined as the unique (up to an additive constant) measurable function $\frac{\delta \mathcal{F}(\rho)}{\delta \rho} : \mathbb{R}^d \rightarrow \mathbb{R}$ such that the equality

$$\left. \frac{d}{dh} \mathcal{F}(\rho + h\chi) \right|_{h=0} = \int \frac{\delta \mathcal{F}(\rho)}{\delta \rho} d\chi,$$

holds for every measure χ such that $\rho + h\chi \in \mathcal{P}(\mathbb{R}^d)$ for small enough h .

In a recent article [40], the authors prove the pointwise empirical propagation of chaos for gradient systems (15) using a gradient flow characterisation of $(f_t^N)_t$ and $(F_t^N)_t$ seen as time continuous curves with values respectively in $\mathcal{P}_2(\mathbb{R}^{dN})$ and $\mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d))$. The argument is based on a compactness criterion in the space $C([0, T], \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d)))$ which follows from Ascoli's theorem. Under the initial chaos hypothesis, the limit of $(F_t^N)_t$ as $N \rightarrow +\infty$ is also identified as a gradient-flow and is shown to be the curve $(\delta_{f_t})_t \in C([0, T], \mathcal{P}_2(\mathcal{P}_2(\mathbb{R}^d)))$. Gradient flows are identified by the EVI (79) which plays a comparable role as the martingale characterisation in a stochastic context.

The results of [40] will be summarised in Section II-3.3.2.

4.3 A pointwise study of the empirical process

This section is devoted to an analytical pointwise study of the empirical process within the framework developed in [130, 131] following an idea of [91]. We also refer to [129] for a review of these results.

The goal is to obtain a quantitative control of the evolution of the law F_t^N of the empirical process seen as the law of a $\mathcal{P}(E)$ -valued process. This control is obtained via a careful asymptotic analysis of the infinitesimal generator of the empirical process which is shown to converge (in a sense to define) towards the generator of the flow of the limit PDE starting from a random initial condition and also seen as a measure-valued process. This method is intrinsically very abstract and can be applied to a wide range of models (in theory, at least to all the models studied in the present review). The main idea traces back to Grünbaum and his study of the spatially homogeneous Boltzmann hard-sphere model (44) [91]. However, the seminal article of Grünbaum was incomplete and based on an unproven assumption (which happened to be false in some cases). The study of measure-valued Markov processes is in general very delicate. This is mainly due to the fact that $\mathcal{P}(E)$ is only a metric space and not a vector space which causes several technical problems, the most important one being the precise definition of the notion of infinitesimal generator. A probabilistic point of view can be found in [54]. In the framework introduced by [130, 131], a new notion of differential calculus in $\mathcal{P}(E)$ is defined in order to give a rigorous definition of the limit generator. The question of propagation of chaos is then stated in a very abstract framework which leads to an abstract theorem (Theorem 4.5) which can be applied to various models after a careful check of a set of five assumptions

(Assumption 2). Most of these assumptions are related to the regularity of the nonlinear solution operator semigroup of the limit PDE. A notable application of this method is the answer to many of the questions raised by Kac in his seminal article [104] related to the spatially homogeneous Boltzmann equation (44). This will be reviewed in Section II-4.5.

The rest of this section is organised as follows. The generator and transition semi-group of the empirical process are defined next. An introductory toy example using this formalism is presented in Subsection 4.3.2. This leads to propagation of chaos for only a very small class of linear models. The next two Subsections 4.3.3 and 4.3.4 present the core of the abstract framework and the main difficulties of the approach, in particular the notion of differential calculus needed to define the limit generator. In the last Subsection 4.3.5 the five assumptions and the abstract theorem of [131] are stated.

4.3.1 The empirical generator.

In Section 3.4, the empirical process has been defined as the image of the N -particle process by the empirical measure map μ_N . This process is a $\hat{\mathcal{P}}_N(E)$ -valued Markov process and it is possible to define its transition semi-group and generator by pushing-forward those of the N -particle process. More precisely, the empirical transition semi-group is given by:

$$\hat{T}_{N,t}\Phi(\mu_{\mathbf{x}^N}) = T_{N,t}[\Phi \circ \mu_N](\mathbf{x}^N), \quad (80)$$

and the empirical generator by:

$$\hat{\mathcal{L}}_N\Phi(\mu_{\mathbf{x}^N}) = \mathcal{L}_N[\Phi \circ \mu_N](\mathbf{x}^N). \quad (81)$$

This is a consequence of the identity:

$$\hat{T}_{N,t}\Phi(\mu_{\mathcal{X}_s^N}) = \mathbb{E}\left[\Phi(\mu_{\mathcal{X}_{t+s}^N})|\mathcal{X}_s^N\right] = \mathbb{E}[(\Phi \circ \mu_N)(\mathcal{X}_{t+s}^N)|\mathcal{X}_s^N] = T_{N,t}[\Phi \circ \mu_N](\mathcal{X}_s^N).$$

Note that the empirical semi-group and generator are well defined as operators $C_b(\mathcal{P}(E)) \rightarrow C_b(\hat{\mathcal{P}}_N(E))$ and the initial law $F_0^N = (\mu_N)_\# f_0^N$ is a probability measure on $\hat{\mathcal{P}}_N(E)$. Nonetheless, in order to take the limit $N \rightarrow +\infty$, it is more convenient to look at the empirical process as a $\mathcal{P}(E)$ -valued process since $\hat{\mathcal{P}}_N(E) \subset \mathcal{P}(E)$.

Example 19 (The empirical generator for a mean-field jump process). For mean-field generators, this simply reads

$$\hat{\mathcal{L}}_N\Phi(\mu_{\mathbf{x}^N}) = \sum_{i=1}^N L_{\mu_{\mathbf{x}^N}} \diamond_i [\Phi \circ \mu_N](\mathbf{x}^N).$$

In the special case of a mean-field jump process, one has a more explicit formula:

$$\begin{aligned} \hat{\mathcal{L}}_N\Phi(\mu_{\mathbf{x}^N}) &= \sum_{i=1}^N \int_E \lambda(x^i, \mu_{\mathbf{x}^N}) \left[\Phi\left(\mu_{\mathbf{x}^N} - \frac{1}{N}\delta_{x^i} + \frac{1}{N}\delta_y\right) - \Phi(\mu_{\mathbf{x}^N}) \right] P_{\mu_{\mathbf{x}^N}}(x^i, dy) \\ &= N \left\langle \mu_{\mathbf{x}^N}, \lambda(\cdot, \mu_{\mathbf{x}^N}) \int_E \left[\Phi\left(\mu_{\mathbf{x}^N} - \frac{1}{N}\delta_\cdot + \frac{1}{N}\delta_y\right) - \Phi(\mu_{\mathbf{x}^N}) \right] P_{\mu_{\mathbf{x}^N}}(\cdot, dy) \right\rangle. \end{aligned}$$

4.3.2 A toy-example using the measure-valued formalism

The model presented in this section is a simple introduction to the measure-valued formalism. We follow an idea which was originally suggested to us by P.-E. Jabin. As we shall see, simple computations can lead to propagation of chaos in the limited case of linear models. Let us consider $\mathcal{X}_t^N = (X_t^1, \dots, X_t^N)_t$ be a PDMP (see Section 2.2.3) defined by

- a deterministic flow:

$$dX_t^i = a(X_t^i)dt,$$

where a is C^1 and globally Lipschitz,

- a jump transition kernel $P : E \times \mathcal{P}(E) \rightarrow \mathcal{P}(E), (x, \mu) \mapsto P_\mu(x, dy)$,
- a constant jump rate $\lambda \equiv 1$.

We take $E = \mathbb{R}^d$ for simplicity. From (81), the associated empirical process is a measure-valued Markov process with generator:

$$\widehat{\mathcal{L}}_N \Phi(\mu) = (a \cdot \nabla \Phi)(\mu) + N \iint_{E \times E} \left\{ \Phi\left(\mu - \frac{1}{N}\delta_x + \frac{1}{N}\delta_y\right) - \Phi(\mu) \right\} P_\mu(x, dy) \mu(dx),$$

where $\Phi \in C_b(\mathcal{P}(E)) \subset C_b(\widehat{\mathcal{P}}_N(E))$ is a test function on $\mathcal{P}(E)$ and $(a \cdot \nabla \Phi)$ is well defined when Φ is a polynomial. We recall the notation $F_0^N(d\mu) \in \mathcal{P}(\mathcal{P}(E))$ for the initial law (supported on the set of empirical measures) and $F_t^N(d\mu) \in \mathcal{P}(\mathcal{P}(E))$ for the law at time t of the measure-valued Markov process with generator $\widehat{\mathcal{L}}^N$. For all test functions $\Phi \in C_b(\mathcal{P}(E))$, one can write the evolution equation for the observables of the empirical process:

$$\begin{aligned} \frac{d}{dt} \int_{\mathcal{P}(E)} \Phi(\mu) F_t^N(d\mu) &+ \int_{\mathcal{P}(E)} \Phi(\mu) \nabla \cdot (a F_t^N)(d\mu) \\ &= N \int_{E \times E \times \mathcal{P}(E)} \left\{ \Phi\left(\mu - \frac{1}{N}\delta_x + \frac{1}{N}\delta_y\right) - \Phi(\mu) \right\} P_\mu(x, dy) \mu(dx) F_t^N(d\mu). \end{aligned} \quad (82)$$

The right-hand side is the jump operator and on the left-hand side there is a transport operator, again well defined for Φ polynomial. Let us also recall the associated nonlinear jump operator acting on $C_b(E)$:

$$L_\mu \varphi(x) := \int_E \{\varphi(y) - \varphi(x)\} P_\mu(x, dy). \quad (83)$$

Its associated carré du champ operator is denoted by Γ_μ . Our goal is to try to prove pointwise empirical propagation of chaos: namely that for any $t > 0$,

$$F_t^N \xrightarrow[N \rightarrow +\infty]{} \delta_{f_t},$$

where $f_t \in \mathcal{P}(E)$ is the solution of the following weak PDE:

$$\forall \varphi \in C_b(E), \quad \frac{d}{dt} \langle f_t, \varphi \rangle + \langle a \cdot \nabla \varphi, f_t \rangle = \langle f_t, L_{f_t} \varphi \rangle.$$

Sufficient conditions for well-posedness are C^1 regularity for a and a global Lipschitz bound, together with Lipschitz-continuity for $P_\mu(x, dy)$ in x and μ (for the Wasserstein topology). To prove propagation of chaos, we will try a “direct analytical approach” and compare directly F_t^N to its limit with a weak distance. Note that it is possible to do that because we work in $\mathcal{P}(\mathcal{P}(E))$ which is a space which does not depend on N (it is one of the main clear advantage of the approach).

Let us consider the distance $\mathcal{W}_{D_2}$, which is the Wasserstein-1 distance on the space $\mathcal{P}(\mathcal{P}(E))$ associated to the distance D_2 on $\mathcal{P}(E)$ (see Definition 3.5). Since the limit is a Dirac mass, it holds that

$$\mathcal{W}_{D_2}^2(F_t^N, \delta_{f_t}) = \int_{\mathcal{P}(E)} D_2^2(\mu, f_t) F_t^N(d\mu) = \sum_{n=1}^{+\infty} \frac{1}{2^n} \int_{\mathcal{P}(E)} \langle \mu - f_t, \varphi_n \rangle^2 F_t^N(d\mu),$$

and it is then enough to bound the quantity

$$g_N(t) := \sup_{\|\varphi\|_{\text{Lip}} \leq 1} \int_{\mathcal{P}(E)} \langle \mu - f_t, \varphi \rangle^2 F_t^N(d\mu). \quad (84)$$

Let us fix a test function $\varphi \in \text{Lip}_1(E)$. In order to control the deterministic flow, we define the (time-dependent) “modified test function” $\tilde{\varphi} \equiv \tilde{\varphi}(s, x)$ as the solution of the backward transport equation:

$$\begin{cases} \partial_s \tilde{\varphi} + a \cdot \nabla_x \tilde{\varphi} &= 0 \\ \tilde{\varphi}(s = t, x) &= \varphi(x) \end{cases} \quad (85)$$

Since a is a globally Lipschitz vector-field, the function $\tilde{\varphi}$ is Lipschitz for all $s \leq t$ with Lipschitz semi-norm:

$$\|\tilde{\varphi}\|_{\text{Lip}} \leq e^{(t-s)\|a\|_{\text{Lip}}}. \quad (86)$$

We define

$$\tilde{g}_\varphi(s) = \int_{\mathcal{P}(E)} \langle \mu - f_s, \tilde{\varphi} \rangle^2 F_s^N(d\mu), \quad (87)$$

where we do not specify the dependency in N for notational simplicity. In order to apply a Gronwall-like argument, we fix $t > 0$ and for $s < t$, thanks to (82), we compute the

time derivative:

$$\begin{aligned}
\tilde{g}'_\varphi(s) &= \frac{d}{ds} \int_{\mathcal{P}(E)} \langle \mu - f_s, \tilde{\varphi} \rangle^2 F_s^N(d\mu) \\
&= \frac{d}{ds} \left[\int_{\mathcal{P}(E)} \langle \mu, \tilde{\varphi} \rangle^2 F_s^N(d\mu) - 2 \langle f_s, \tilde{\varphi} \rangle \int_{\mathcal{P}(E)} \langle \mu, \tilde{\varphi} \rangle F_s^N(d\mu) + \langle f_s, \tilde{\varphi} \rangle^2 \right] \\
&= \int_{\mathcal{P}(E)} \langle \mu, \tilde{\varphi} \rangle^2 \partial_s F_s^N(d\mu) - 2 \langle f_s, \tilde{\varphi} \rangle \int_{\mathcal{P}(E)} \langle \mu, \tilde{\varphi} \rangle \partial_s F_s^N(d\mu) \\
&\quad - 2 \langle \partial_s f_s, \tilde{\varphi} \rangle \int_{\mathcal{P}(E)} \langle \mu, \tilde{\varphi} \rangle F_s^N(d\mu) + 2 \langle f_s, \tilde{\varphi} \rangle \langle \partial_s f_s, \tilde{\varphi} \rangle \\
&\quad + 2 \int_{\mathcal{P}(E)} \langle \mu, \tilde{\varphi} \rangle \langle \mu, \partial_s \tilde{\varphi} \rangle F_s^N(d\mu) \\
&\quad - 2 \langle f_s, \partial_s \tilde{\varphi} \rangle \int_{\mathcal{P}(E)} \langle \mu, \tilde{\varphi} \rangle F_s^N(d\mu) \\
&\quad - 2 \langle f_s, \tilde{\varphi} \rangle \int_{\mathcal{P}(E)} \langle \mu, \partial_s \tilde{\varphi} \rangle F_s^N(d\mu) + 2 \langle f_s, \tilde{\varphi} \rangle \langle f_s, \partial_s \tilde{\varphi} \rangle.
\end{aligned}$$

Using the fact that for two test functions $\varphi_1, \varphi_2 \in C_b(E)$:

$$\begin{aligned}
\int_{\mathcal{P}(E)} \langle \mu, \varphi_1 \rangle \langle \mu, \varphi_2 \rangle \partial_s F_s^N(d\mu) &= \int_{\mathcal{P}(E)} \left\{ \langle \mu, a \cdot \nabla_x \varphi_1 \rangle \langle \mu, \varphi_2 \rangle + \langle \mu, \varphi_1 \rangle \langle \mu, a \cdot \nabla_x \varphi_2 \rangle \right. \\
&\quad \left. + \langle \mu, \varphi_1 \rangle \langle \mu, L_\mu \varphi_2 \rangle + \langle \mu, L_\mu \varphi_1 \rangle \langle \mu, \varphi_2 \rangle + \frac{1}{N} \langle \mu, \Gamma_\mu(\varphi_1, \varphi_2) \rangle \right\} F_s^N(d\mu),
\end{aligned}$$

a direct computation shows that:

$$\begin{aligned}
\tilde{g}'_\varphi(s) &= \int_{\mathcal{P}(E)} \left(\langle \mu, \tilde{\varphi} \rangle - \langle f_s, \tilde{\varphi} \rangle \right) \left(\langle \mu, L_\mu \tilde{\varphi} \rangle - \langle f_s, L_{f_s} \tilde{\varphi} \rangle \right) F_s^N(d\mu) \\
&\quad + \frac{1}{N} \int_{\mathcal{P}(E)} \langle \mu, \Gamma_\mu(\tilde{\varphi}, \tilde{\varphi}) \rangle F_s^N(d\mu)
\end{aligned}$$

By the arithmetic-geometric mean inequality, we obtain:

$$\begin{aligned}
\tilde{g}'_\varphi(s) &\leq \frac{1}{2} \int_{\mathcal{P}(E)} \langle \mu - f_s, \tilde{\varphi} \rangle^2 F_s^N(d\mu) + \frac{1}{2} \int_{\mathcal{P}(E)} \left(\langle \mu, L_\mu \tilde{\varphi} \rangle - \langle f_s, L_{f_s} \tilde{\varphi} \rangle \right)^2 F_s^N(d\mu) \\
&\quad + \frac{1}{N} \int_{\mathcal{P}(E)} \langle \mu, \Gamma_\mu(\tilde{\varphi}, \tilde{\varphi}) \rangle F_s^N(d\mu).
\end{aligned}$$

The last term on the right-hand side can be controlled using a mild moment assumption:

$$\exists \gamma > 0, \quad \forall x \in E, \quad \int_E |x - y|^2 P_\mu(x, dy) \leq \gamma,$$

which implies

$$\Gamma_\mu(\tilde{\varphi}, \tilde{\varphi}) \leq \gamma \|\tilde{\varphi}\|_{\text{Lip}}^2 \leq \gamma e^{2(t-s)\|a\|_{\text{Lip}}}.$$

we deduce that:

$$\tilde{g}'_\varphi(s) \leq \frac{1}{2}\tilde{g}_\varphi(s) + \frac{\gamma e^{2(t-s)\|a\|_{\text{Lip}}}}{N} + \frac{1}{2} \int_{\mathcal{P}(E)} (\langle \mu, L_\mu \tilde{\varphi} \rangle - \langle f_s, L_{f_s} \tilde{\varphi} \rangle)^2 F_s^N(d\mu). \quad (88)$$

Unfortunately it is not possible in general to close the argument. One would like to bound the last term on the right-hand side of (88) by a quantity of the form (87) (possibly with a different φ). This is hopeless in general because (87) depends on (the square of) a linear quantity in μ but due to the nonlinearity, the collision operator

$$\langle Q(\mu), \tilde{\varphi} \rangle := \langle \mu, L_\mu \tilde{\varphi} \rangle$$

is at least quadratic in μ (or controlled by a quadratic quantity as soon as P_μ is Lipschitz in μ for a Wasserstein distance). This fact is analogous to the BBGKY hierarchy at the level of the empirical process. It is possible to close the argument only in linear cases, for instance when

$$P_\mu(x, dy) = \int_{z \in E} K(y, z) \mu(dz) dy,$$

where $K : E \times E \rightarrow \mathbb{R}_+$ is a fixed symmetric interaction kernel with $\int_E K(y, z) dy = 1$ for all $z \in E$ (note that Lipschitz continuity for K is sufficient to ensure the well-posedness). In this case the collision operator is a linear operator:

$$\langle Q(\mu), \tilde{\varphi} \rangle = \langle \mu, K \star \tilde{\varphi} \rangle - \langle \mu, \tilde{\varphi} \rangle. \quad (89)$$

Reporting into (88), we get,

$$\begin{aligned} \tilde{g}'_\varphi(s) \leq \frac{1}{2}\tilde{g}_\varphi(s) + \frac{\gamma e^{2(t-s)\|a\|_{\text{Lip}}}}{N} + \int_{\mathcal{P}(E)} \langle \mu - f_s, \tilde{\varphi} \rangle^2 F_s^N(d\mu) \\ + \int_{\mathcal{P}(E)} \langle \mu - f_s, K \star \tilde{\varphi} \rangle^2 F_s^N(d\mu). \end{aligned}$$

Since the Lipschitz norm of $\tilde{\varphi}$ is controlled by (86) and that this bound is preserved by the convolution with K , using (84), we conclude that:

$$\tilde{g}'_\varphi(s) \leq \frac{5}{2}g_N(s)e^{2(t-s)\|a\|_{\text{Lip}}} + \frac{\gamma e^{2(t-s)\|a\|_{\text{Lip}}}}{N}.$$

Integrating between 0 and t , and taking the supremum over φ on the left-hand side, we can apply Gronwall lemma and conclude that

$$\mathcal{W}_{D_2}^2(F_t^N, \delta_{f_t}) \leq g_N(t) \leq C(\gamma, a, t) \left(\mathcal{W}_{D_2}^2(F_0^N, \delta_{f_0}) + \frac{1}{N} \right). \quad (90)$$

The first term on the right-hand side depends on the initial condition and can be controlled using e.g. [78] (using their estimates, it will determine the final rate of convergence since their best convergence rate is $\mathcal{O}(N^{-1})$); sharper estimates can be available though.

Remark 20. Note that despite its relative simplicity, the model (89) has its own interest. In [34, 33] it is called the “choose the leader” model. In population dynamics, it is also a time-continuous version of the well-known Moran model [54]. It describes the “neutral” evolution of a population of individuals where the death and reproduction events happen simultaneously. The kernel K plays the role of a mutation kernel. A different scaling which leads to a different limit is presented in Section II-5.4.2.

4.3.3 The limit semi-group and the nonlinear measure-valued process

The previous approach is too coarse as it tries to compare directly the empirical law F_t^N to its limit δ_{f_t} . By looking only at the expectation of some fixed (though infinitely many) observables, we do not keep track of the detailed dynamics of the particle system. In this section we give some insights on an approach which is originally due to Grünbaum [91] but which has been made rigorous in [131, 130]. Rather than looking only at the law F_t^N , the idea is to compare the empirical process and a “nonlinear” measure-valued process through their semi-groups and generators, thus effectively keeping track of the dynamics. Note that we could define an “obvious” nonlinear empirical process by taking the image by μ_N of N i.i.d. f_t -distributed processes. This would lead us to compare $\tilde{F}_t^N = (\mu_N)_\# f_t^{\otimes N}$ to $F_t^N = (\mu_N)_\# f_t^N$ but this would not be simpler than comparing directly f_t^N to $f_t^{\otimes N}$. This could be handled by the coupling approach (Section 4.1). Instead, the approach of [130, 131] considers the nonlinear dynamics in $\mathcal{P}(E)$ from the PDE point of view. In the most abstract setting, the limiting nonlinear law f_t is the solution of

$$\partial_t f_t = Q(f_t), \quad (91)$$

where Q is a nonlinear operator. Assuming that this PDE is wellposed, this gives rise to a nonlinear time-continuous semi-group $(\bar{S}_t)_{t \geq 0}$ acting on $\mathcal{P}(E)$ such that the solution of (91) is given as:

$$f_t = \bar{S}_t(f_0),$$

where $f_0 \in \mathcal{P}(E)$ is the initial condition and

$$\bar{S}_{t+s} = \bar{S}_t \circ \bar{S}_s = \bar{S}_s \circ \bar{S}_t, \quad \partial_t \bar{S}_t = Q \circ \bar{S}_t.$$

From the stochastic point of view, the operator $\bar{S}_t$ is the dual of the transition operator $T_t^{f_0}$ of a nonlinear Markov process in the sense of McKean (see Appendix A.4). The main observation is that the deterministic dynamics (91) can be seen as a stochastic process in $\mathcal{P}(E)$, for instance, it is possible to choose a random initial condition: although it is expected to be a given $f_0 \in \mathcal{P}(E)$, it is also natural to take as initial condition the same as the one of the empirical process, that is a random empirical measure with N points sampled from f_0^N . Remember that f_0^N is assumed to be initially f_0 -chaotic. With this choice, the goal is to compare the empirical process $(\mu_{\mathcal{X}_t^N})_t$ and the nonlinear process $(\bar{S}_t(\mu_{\mathcal{X}_0^N}))_t$. The laws of these processes in $\mathcal{P}(\mathcal{P}(E))$ at time $t > 0$ are respectively given by:

$$F_t^N = (\mu_N)_\# S_t^N(f_0^N), \quad \bar{F}_t^N := (\bar{S}_t \circ \mu_N)_\# f_0^N, \quad (92)$$

where S_t^N denotes the N -particle semigroup acting on $\mathcal{P}(E^N)$, so that $f_t^N = S_t^N(f_0^N)$. The semigroups $(S_t^N)_t$ and $(\bar{S}_t)_t$ describe the forward dynamics of the probability distributions. The dynamics of the observables is described by the dual operators acting on the space of test functions on $C_b(\mathcal{P}(E))$. As explained at the beginning of this section, for the particle dynamics, everything is given in terms of $(T_{N,t})_t$ which is the semigroup acting on $C_b(E^N/\mathfrak{S}_N)$. For the nonlinear system, the following operator is defined in [131, 130]:

$$\forall \Phi \in C_b(\mathcal{P}(E)), \forall \nu \in \mathcal{P}(E), \quad T_{\infty,t} \Phi(\nu) := \Phi(\bar{S}_t(\nu)), \quad (93)$$

so that

$$\int_{E^N} T_{\infty,t} \Phi(\mu_{\mathbf{x}^N}) f_0^N(d\mathbf{x}^N) = \langle \bar{F}_t^N, \Phi \rangle = \int_{E^N} \Phi(\bar{S}_t(\mu_{\mathbf{x}^N})) f_0^N(d\mathbf{x}^N).$$

Note that the dependence on N is only in the initial condition. On the other hand, it holds that:

$$\int_{E^N} \hat{T}_{N,t} \Phi(\mu_{\mathbf{x}^N}) f_0^N(d\mathbf{x}^N) = \langle F_t^N, \Phi \rangle,$$

so very loosely speaking, the goal is to prove the convergence of the semi-groups:

$$\hat{T}_{N,t} \xrightarrow{N \rightarrow +\infty} T_{\infty,t}.$$

In a classical setting, the convergence of a sequence of semi-groups acting on a set of test functions over a Banach space is solved by Trotter [156] by proving the convergence of the generators. The generator $\hat{\mathcal{L}}_N$ associated to $\hat{T}_{N,t}$ is defined by (81), although its image is restricted to the subdomain $C_b(\hat{\mathcal{P}}_N(E)) \subset C_b(\mathcal{P}(E))$. The generator of $(T_{\infty,t})_t$ is much more delicate to define because $\mathcal{P}(E)$ is only a metric space and not a Banach space. Its precise and rigorous definition is one of the main contributions of [131, 130]. We will give insights on this later, but for now let us assume that it is possible to define a generator $\mathcal{L}_\infty$ on a sufficiently large subset of $C_b(\mathcal{P}(E))$ and such that for Φ in this subset,

$$\frac{d}{dt} T_{\infty,t} \Phi = \mathcal{L}_\infty[T_{\infty,t} \Phi] = T_{\infty,t} \mathcal{L}_\infty \Phi. \quad (94)$$

We now briefly explain how generator estimates will give an estimate on the discrepancy between F_t^N and $\bar{F}_t^N$. First, using Lemma 3.11, it is sufficient to look at the moment measures $F_t^{k,N}, \bar{F}_t^{k,N} \in \mathcal{P}(E^k)$ for all $k \in \mathbb{N}$. Let $\varphi_k \in C_b(E^k)$ be a test function and let

$$R_{\varphi_k}(\nu) \equiv \Phi_k(\nu) := \langle \nu^{\otimes k}, \varphi_k \rangle,$$

be the associated polynomial function. We recall that the operators $T_{N,t}$ and $\mathcal{L}_N$ are directly linked to their empirical versions (80), (81) by μ_N and the linear transpose map:

$$\mu_N^T : C_b(\mathcal{P}(E)) \rightarrow C_b(E^N), \quad \Phi \mapsto \Phi \circ \mu_N.$$

By definition, it holds that:

$$\langle F_t^{k,N}, \varphi_k \rangle = \langle F_t^N, \Phi_k \rangle = \langle f_0^N, T_{N,t}[\Phi_k \circ \mu_N] \rangle = \langle f_0^N, T_{N,t}[\mu_N^T \Phi_k] \rangle,$$

and

$$\langle \bar{F}_t^{k,N}, \varphi_k \rangle = \langle \bar{F}_t^N, \Phi_k \rangle = \langle f_0^N, T_{\infty,t} \Phi \circ \mu_N \rangle = \langle f_0^N, \mu_N^T [T_{\infty,t} \Phi_k] \rangle.$$

The difference between these two quantities is controlled by using the formula for $0 \leq t \leq T$:

$$\begin{aligned} T_{N,t} \mu_N^T - \mu_N^T T_{\infty,t} &= - \int_0^t \frac{d}{ds} [T_{N,t-s} \mu_N^T T_{\infty,s}] ds \\ &= \int_0^t T_{N,t-s} [\mathcal{L}_N \mu_N^T - \mu_N^T \mathcal{L}_\infty] T_{\infty,s} ds, \end{aligned}$$

which leads to the following bound:

$$\begin{aligned}
\left| \langle F_t^{k,N} \varphi_k \rangle - \langle \bar{F}_t^{k,N}, \varphi_k \rangle \right| &\leq \int_0^t \left| \langle f_0^N, T_{N,t-s} [\mathcal{L}_N \boldsymbol{\mu}_N^T - \boldsymbol{\mu}_N^T \mathcal{L}_\infty] T_{\infty,s} \Phi_k \rangle \right| ds \\
&= \int_0^t \left| \langle f_{t-s}^N, [\mathcal{L}_N \boldsymbol{\mu}_N^T - \boldsymbol{\mu}_N^T \mathcal{L}_\infty] T_{\infty,s} \Phi_k \rangle \right| ds \\
&= \int_0^t \left| \langle f_{t-s}^N, \mathcal{L}_N [T_{\infty,s} \Phi_k \circ \boldsymbol{\mu}_N] - \mathcal{L}_\infty [T_{\infty,s} \Phi_k] \circ \boldsymbol{\mu}_N \rangle \right| ds \\
&\leq T \sup_{0 \leq t \leq T} \left\| \left(\hat{\mathcal{L}}_N [T_{\infty,t} \Phi_k] - \mathcal{L}_\infty [T_{\infty,t} \Phi_k] \right) \circ \boldsymbol{\mu}_N \right\|_\infty. \quad (95)
\end{aligned}$$

At this point, in order to obtain a convergence bound in N , a generator estimate is needed to compare the behaviour of the empirical generator $\hat{\mathcal{L}}_N$ to the one of $\mathcal{L}_\infty$ against $T_{\infty,t} \Phi_k$ (the map $\boldsymbol{\mu}_N$ is just an artefact to write this comparison in E^N).

Remark 21. The estimate (95) can be made more uniform in time as soon as the particle system preserves some quantity $m : E^N \rightarrow \mathbb{R}_+$, in which case (95) becomes:

$$\begin{aligned}
\left| \langle F_t^{k,N} \varphi_k \rangle - \langle \bar{F}_t^{k,N}, \varphi_k \rangle \right| &\leq \sup_{0 \leq t \leq T} \left\| \frac{1}{m} \left(\hat{\mathcal{L}}_N [T_{\infty,t} \Phi_k] - \mathcal{L}_\infty [T_{\infty,t} \Phi_k] \right) \circ \boldsymbol{\mu}_N \right\|_\infty \int_0^T \langle f_{t-s}^N, m \rangle ds.
\end{aligned}$$

In Section 4.3.5 we will review the abstract theorem of [131] which shows how to recover propagation of chaos in the usual framework from an estimate on (95). Before doing that, we give more insights on the definition of the generator $\mathcal{L}_\infty$ within the framework of [131].

4.3.4 More on the limit generator

As an introductory example, let us consider a mean-field generator of the form (6) and a tensorized test function of order two:

$$\varphi_2 = \varphi^1 \otimes \varphi^2,$$

as well as the associated polynomial function on $\mathcal{P}(E)$ defined by $\Phi_2(\nu) = \langle \nu^{\otimes 2}, \varphi_2 \rangle$. We recall that it is not a restriction to consider tensorized test functions [75, Chapter 3, Theorem 4.5 and Proposition 4.6, pp.113-115]. Then, a direct computation which is detailed in Lemma II-A.1 (a similar computation is also used in the proof of Theorem II-3.10) gives:

$$\begin{aligned}
\mathcal{L}_N [\Phi_2 \circ \boldsymbol{\mu}_N] (\mathbf{x}^N) &= \langle \mu_{\mathbf{x}^N}, L_{\mu_{\mathbf{x}^N}} \varphi^1 \rangle \langle \mu_{\mathbf{x}^N}, \varphi^2 \rangle + \langle \mu_{\mathbf{x}^N}, \varphi^1 \rangle \langle \mu_{\mathbf{x}^N}, L_{\mu_{\mathbf{x}^N}} \varphi^2 \rangle \\
&\quad + \frac{1}{N} \langle \mu_{\mathbf{x}^N}, \Gamma_{L_{\mu_{\mathbf{x}^N}}} (\varphi^1, \varphi^2) \rangle,
\end{aligned}$$

where Γ is the carré du champ operator. In order to have

$$\| (\hat{\mathcal{L}}_N [\Phi_2] - \mathcal{L}_\infty [\Phi_2]) \circ \boldsymbol{\mu}_N \|_\infty \xrightarrow{N \rightarrow +\infty} 0,$$

the operator $\mathcal{L}_\infty$ should necessarily satisfy:

$$\mathcal{L}_\infty \Phi_2(\nu) = \langle \nu, L_\nu \varphi^1 \rangle \langle \nu, \varphi^2 \rangle + \langle \nu, \varphi^1 \rangle \langle \nu, L_\nu \varphi^2 \rangle.$$

This computation is generalised (see Lemma II-A.2) to any k -fold tensorized test function

$$\varphi_k = \varphi^1 \otimes \dots \otimes \varphi^k,$$

and the operator $\mathcal{L}_\infty$ is defined against monomial functions by:

$$\mathcal{L}_\infty \Phi_k(\nu) = \sum_{i=1}^k \langle Q(\nu), \varphi^i \rangle \prod_{j \neq i} \langle \nu, \varphi^j \rangle, \quad (96)$$

where $\Phi_k(\nu) = \langle \nu^{\otimes k}, \varphi_k \rangle$ and with a slight abuse of notation, we write

$$\langle Q(\nu), \varphi \rangle \equiv \langle \nu, L_\nu \varphi \rangle,$$

for the integral of a test function φ against $Q(\nu)$ (which is not a probability measure). The relation (96) can be extended to any polynomial by linearity. We would also get the same relation (with a different operator Q) for a Boltzmann operator (see Lemma II-A.3).

A natural idea would be to use the relation (96) as a *definition* of an operator acting on polynomials and then extend it to the completion of the space of polynomials, which is a large Banach subset of $C_b(\mathcal{P}(E))$. However, this would not necessarily imply that the right-hand side of (95) goes to zero for any Φ (since for a given polynomial, the convergence implied by Lemma II-A.2 may not be uniform in the degree or number of monomials) and proving this convergence does not seem to be an easy task. We recall that the final goal is to apply $\mathcal{L}_\infty$ to the test functions

$$T_{\infty,t} \Phi_k(\nu) = \langle \bar{S}_t(\nu)^{\otimes k}, \varphi_k \rangle, \quad (97)$$

which are not polynomial in general. In particular, the relation (96) needs to be extended to be able to write

$$\mathcal{L}_\infty T_{\infty,t} \Phi_k(\nu) = \sum_{i=1}^k \langle Q(\bar{S}_t(\nu)), \varphi^i \rangle \prod_{j \neq i} \langle \bar{S}_t(\nu), \varphi^j \rangle,$$

in order to have (94).

Remark 22. The stochastic point of view gives more insight on the form of the nonlinearity in (97). Under the assumption that f_t is the law of a nonlinear Markov process in the sense of McKean (see Appendix A.4), one can write the dual form when $\varphi_k = \varphi^1 \otimes \dots \otimes \varphi^k$:

$$T_{\infty,t} \Phi_k(\nu) = \prod_{i=1}^k \langle \nu, T_t^\nu \varphi^i \rangle, \quad (98)$$

where T_t^ν is the nonlinear transition operator of the process. Up to the dependency of T_t^ν on the measure argument ν , the test function $T_{\infty,t} \Phi_k$ is thus close to be a polynomial.

When T_t^ν does not depend on ν , $f_t = \bar{S}(f_0)$ is the law of a classical time homogeneous Markov process and thus satisfies a linear equation. In this case, everything is much simpler because the operator $T_{\infty,t}$ acts on the space of polynomial and it is not necessary to extend $\mathcal{L}_\infty$ to a larger subspace of $C_b(\mathcal{P}(E))$. One could actually bypass the definition of the limit generator. Note also that the linear case is the framework of our toy example in Section 4.3.2. Unfortunately, in the nonlinear case, the dual backward form (98) does not really seem to be more helpful since T_t^ν has no reason to behave well with respect to its measure argument. We finally point out that all the argument in [131, 130] is more general as it is entirely based on (97) and does not use the fact that f_t is the law of a nonlinear Markov process in the sense of McKean (even though it is the underlying application case).

In his seminal article [91], Grünbaum originally relies on a clever completion of the space of polynomial functions. He then identifies $\mathcal{L}_\infty$ and proves the convergence of the generators on a class C' of “continuously differentiable functions” on $\mathcal{P}(E)$. In order to apply Trotter’s result on $T_{\infty,t}$, Grünbaum uses an unproven smoothness assumption on the nonlinear operator $\bar{S}_t$ which ensures that (97) belongs to the class C' . This assumption has since been proved to be false for some models.

The generator $\mathcal{L}_\infty$ is rigorously defined in [131, 130] using a new notion of differential calculus on the space $\mathcal{P}(E)$ that is very briefly sketched below. The fundamental idea is to consider a distance on $\mathcal{P}(E)$ inherited from a normed vector space $\mathcal{G}$ to be chosen. Let $m_{\mathcal{G}} : E \rightarrow \mathbb{R}_+$ be given together with the weighted subspace of probability measures:

$$\mathcal{P}_{\mathcal{G}}(E) := \{f \in \mathcal{P}(E), \langle f, m_{\mathcal{G}} \rangle < +\infty\}.$$

The weight function $m_{\mathcal{G}}$ may typically be a polynomial function (in which case $\mathcal{P}_{\mathcal{G}}$ is the space of probability measures with a bounded moment) but may also depend on the normed vector space $\mathcal{G}$ which is assumed to contain the space of increments:

$$\mathcal{IP}_{\mathcal{G}}(E) := \left\{ f_1 - f_2, f_1, f_2 \in \mathcal{P}_{\mathcal{G}}(E) \right\} \subset \mathcal{G}. \quad (99)$$

This naturally defines a distance on $\mathcal{P}_{\mathcal{G}}(E)$ by:

$$\forall f_1, f_2 \in \mathcal{P}_{\mathcal{G}}(E), \quad d_{\mathcal{G}}(f_1, f_2) := \|f_1 - f_2\|_{\mathcal{G}}.$$

Example 20. A simple example is obtained when taking for $\mathcal{G}$ the Banach space of finite mass signed measures endowed with the total variation distance. $\mathcal{P}_{\mathcal{G}}(E)$ is then a subspace of $\mathcal{P}(E)$ endowed with the total variation distance, while the space $\mathcal{IP}_{\mathcal{G}}(E)$ of its increments is seen as a subset of the finite mass signed measures.

Other examples and their relations with the distances defined in Section 3.1 are detailed in [131, Section 3.2]. With such notion of distance, it is possible to define a differential structure on the space of probability measures which is inherited from the classical differentiable structure of the Banach space $\mathcal{G}$. Namely, a test function $\Phi : \mathcal{P}_{\mathcal{G}}(E) \rightarrow \mathbb{R}$ is

said to be continuously differentiable at $f \in \mathcal{P}_{\mathcal{G}}(E)$ when there exists a continuous linear application $d\Phi[f] : \mathcal{G} \rightarrow \mathbb{R}$ and a constant $C > 0$ such that:

$$\forall g \in \mathcal{P}_{\mathcal{G}}(E), \quad |\Phi(g) - \Phi(f) - \langle d\Phi[f], g - f \rangle_{\mathcal{G}', \mathcal{G}}| \leq C d_{\mathcal{G}}(f, g). \quad (100)$$

Note that the bracket in the inequality is the duality bracket between $\mathcal{G}'$ and $\mathcal{G}$. The main difference with the usual notion of differentiability in a Banach space is that the space of increments (99) has no vectorial structure. More details on this notion of differential calculus is given in [131, Section 3.3 and Section 3.4] with a special focus on polynomial functions. This definition can be extended to higher order differentiability and to functions with values in $\mathcal{P}_{\tilde{\mathcal{G}}}(E)$ instead of $\mathbb{R}$ (which is the case of the operator $\bar{S}_t$).

The definition of $\mathcal{L}_{\infty}$ then directly comes from the differentiation of the definition of the pullback semigroup (93): let Φ be a continuously differentiable function on $\mathcal{P}_{\mathcal{G}}(E)$, then for all $\nu \in \mathcal{P}_{\mathcal{G}}(E)$, by the composition rule (see [131, Lemma 3.12]), it holds that:

$$\begin{aligned} \mathcal{L}_{\infty}\Phi(\nu) &= \left. \frac{d}{dt} T_{\infty, t} \Phi(\nu) \right|_{t=0} = \left. \frac{d}{dt} \Phi(\bar{S}_t(\nu)) \right|_{t=0} = \left\langle d\Phi[\nu], \left. \frac{d}{dt} \bar{S}_t(\nu) \right|_{t=0} \right\rangle \\ &= \langle d\Phi[\nu], Q(\nu) \rangle. \end{aligned} \quad (101)$$

This computation is almost rigorous up to the assumption that $Q(\nu) \in \mathcal{G}$. The precise assumptions on $(\bar{S}_t)_t$ which make this computation fully rigorous are given by [131, Assumption (A2)] and will be summarised in the next section.

Example 21 (Generators estimate for jump processes). Quite formal computations may also motivate the introduction of a proper notion of differential calculus on $\mathcal{P}(E)$ and lead to generators estimates. Taking the example of jump processes, we have seen that:

$$\hat{\mathcal{L}}_N \Phi(\mu_{\mathbf{x}^N}) = \sum_{i=1}^N \int_E \lambda(x^i, \mu_{\mathbf{x}^N}) \left[\Phi\left(\mu_{\mathbf{x}^N} - \frac{1}{N} \delta_{x^i} + \frac{1}{N} \delta_y\right) - \Phi(\mu_{\mathbf{x}^N}) \right] P_{\mu_{\mathbf{x}^N}}(x^i, dy).$$

The term in the integral is precisely of the form (100) with an increment of size $1/N$. Assuming that it is possible to differentiate Φ , we get:

$$\begin{aligned} &\hat{\mathcal{L}}_N \Phi(\mu_{\mathbf{x}^N}) \\ &= \sum_{i=1}^N \int_E \lambda(x^i, \mu_{\mathbf{x}^N}) \left[\left\langle d\Phi(\mu_{\mathbf{x}^N}), -\frac{1}{N} \delta_{x^i} + \frac{1}{N} \delta_y \right\rangle + o\left(\frac{1}{N}\right) \right] P_{\mu_{\mathbf{x}^N}}(x^i, dy) \\ &= \langle d\Phi(\mu_{\mathbf{x}^N}), Q(\mu_{\mathbf{x}^N}) \rangle + o(1) = \mathcal{L}_{\infty} \Phi(\mu_{\mathbf{x}^N}) + o(1), \end{aligned}$$

where Q is given in the weak form by the left-hand side of (19) (with $a = 0$).

4.3.5 The abstract theorem

The main theorem [131, Theorem 2.1] is based on the following set of assumptions. The first one is the only one which concerns the particle system (it is always implicitly assumed). The second and third ones are motivated by the previous sections. The fourth and fifth ones are stated more informally, more details on their role will be given in the sketch of the proof of the main theorem.

Assumption 2. *The following assumptions are respectively numbered (A1) to (A5) in [131].*

- (1) **(On the particle system).** *The N -particle semigroup $(T_{N,t})_{t \geq 0}$ is a strongly continuous semigroup on $C_b(E^N)$ with generator $\mathcal{L}_N$.*
- (2) **(Existence of the pull-back semigroup).** *There exists a Banach space $\mathcal{G}$ such that the nonlinear semigroup $(\bar{S}_t)_t$ on $\mathcal{P}_{\mathcal{G}}(E)$ is Lipschitz for the distance $d_{\mathcal{G}}$ uniformly in time. The operator $Q : \mathcal{P}_{\mathcal{G}}(E) \rightarrow \mathcal{G}$ is bounded and δ -Hölder for a $\delta \in (0, 1]$. This implies the existence of the limit generator $\mathcal{L}_{\infty}$ defined by (101) (see [131, Lemma 4.1]).*
- (3) **(Generators estimate).** *There exists a sequence $\varepsilon(N)$ such that $\varepsilon(N) \rightarrow 0$ as $N \rightarrow +\infty$ and such that for sufficiently regular test functions Φ on $\mathcal{P}_{\mathcal{G}}(E)$,*

$$\|(\hat{\mathcal{L}}_N \Phi - \mathcal{L}_{\infty} \Phi) \circ \mu_N\|_{\infty} \leq \varepsilon(N) \|\Phi\|_{C^{k,1}(\mathcal{P}_{\mathcal{G}}(E))}, \quad (102)$$

where $\|\cdot\|_{C^{k,1}(\mathcal{P}_{\mathcal{G}}(E))}$ is a norm related to the notion of higher order differentiability.

- (4) **(Differential stability of the nonlinear semigroup).** *The nonlinear semi-group $(\bar{S}_t)_t$ is differentiable (for the generalised version of the notion of differentiability mentioned above), and its derivatives can be bounded uniformly in time using adapted $\mathcal{G}$ -dependent norms.*
- (5) **(Weak stability of the nonlinear semigroup).** *For a Banach space $\tilde{\mathcal{G}}$ possibly different from $\mathcal{G}$, the nonlinear semi-group is Lipschitz for the distance $d_{\tilde{\mathcal{G}}}$. More generally this can be replaced by the existence of a concave modulus of continuity $\Theta_T : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ such that for all $f_0, g_0 \in \mathcal{P}_{\tilde{\mathcal{G}}}(E)$,*

$$\sup_{0 \leq t \leq T} d_{\tilde{\mathcal{G}}}(\bar{S}_t(f_0), \bar{S}_t(g_0)) \leq \Theta_T(d_{\tilde{\mathcal{G}}}(f_0, g_0)).$$

The last assumption (5) quantifies the continuity of the operator $\bar{S}_t$ with respect to its measure argument: together with the assumption (4), it should be understood as a regularity assumption with respect to the initial condition for the limit nonlinear PDE. The fact that it holds for a possibly different Banach space $\tilde{\mathcal{G}}$ is essentially technical: it allows more flexibility in the choice of a well-suited topology to deal with the nonlinear problem.

The following abstract theorem is stated and proved in [131, Theorem 2.1].

Theorem 4.5 (The abstract theorem in [131]). *Let Assumption 2 hold true. Let $T > 0$, $k \in \mathbb{N}$ and $N \geq 2k$. Then there exist a continuously embedded subset $\mathcal{F} \subset C_b(E)$ and some absolute constants $C, C(T) > 0$ such that for any tensorized test function*

$$\varphi_k = \varphi^1 \otimes \dots \otimes \varphi^k \in \mathcal{F}^{\otimes k},$$

it holds that:

$$\sup_{0 \leq t \leq T} |\langle f_t^{k,N} - f_t^{\otimes k}, \varphi_k \rangle| \leq C \frac{k^2 \|\varphi_k\|_\infty}{N} + k^2 C(T) \|\varphi_k\|_{\mathcal{F}_1} \varepsilon(N) + k \|\varphi_k\|_{\mathcal{F}_2} \Theta_T \left(\mathcal{W}_{d_{\tilde{\mathcal{G}}}}(\bar{F}_0^N, \delta_{f_0}) \right), \quad (103)$$

where $\varepsilon(N)$ is defined in Assumption 2(3), $\mathcal{W}_{d_{\tilde{\mathcal{G}}}}$ is a Wassertein distance on the space $\mathcal{P}(\mathcal{P}(E))$ related to $\tilde{\mathcal{G}}$ given by Assumption 2(5) and $\|\cdot\|_{\mathcal{F}_1}$ and $\|\cdot\|_{\mathcal{F}_2}$ are some norms on $C_b(E^k)$ which are defined in the complete version of Assumption 2 (see [131, Section 4]).

Proof (main ideas). The proof in [131] relies on three main steps:

- Approximate $f_t^{k,N}$ by $F_t^{k,N}$ thanks to Lemma 3.11.
- Approximate $f_t^{\otimes k} = \bar{S}_t^{\otimes k}(f_0)$ by $\mathbb{E}_{\mathcal{X}_0^N} \bar{S}_t^{\otimes k}(\mu_{\mathcal{X}_0}^N) = \bar{F}_t^{k,N}$. Since initial chaos $\mu_{\mathcal{X}_0}^N \rightarrow f_0$ is assumed, this should stem from the regularity on the limit equation.
- Compare the time evolution of $F_t^{k,N}$ to the one of $\bar{F}_t^{k,N}$, which motivates the content of Section 4.3.3 and Section 4.3.4.

We recall that $F_t^{k,N}$ and $\bar{F}_t^{k,N}$ are the moment measures (Definition 3.13) associated to the laws F_t^N and $\bar{F}_t^N$ defined by (92). Each of the three terms on the right-hand side of (103) thus comes from the splitting:

$$\langle f_t^{k,N} - f_t^{\otimes k}, \varphi_k \rangle = \langle f_t^{k,N} - F_t^{k,N}, \varphi_k \rangle + \langle F_t^{k,N} - \bar{F}_t^{k,N}, \varphi_k \rangle + \langle \bar{F}_t^{k,N} - f_t^{\otimes k}, \varphi_k \rangle. \quad (104)$$

The first term on the right-hand side is handled with rate $\mathcal{O}(k^2 N^{-1})$ by the approximation Lemma 3.11, using purely combinatorial arguments. The second term is technically the most difficult one. Assumption 2(2) gives a precise meaning of the relation (95) formally derived earlier. The role of Assumption 2(3) is thus self-explanatory and Assumption 2(4) ensures that $\Phi = T_{\infty,t} \Phi_k$ has enough regularity to be taken as a test function in (102). The third term contains two approximations: the first one is how well the initial data is approximated by a (random) empirical measure and then how well this error is propagated in time, which is Assumption 2(5). $\square$

Applications of the abstract Theorem 4.5 to classical models can be found in [131]. The assumptions are rigorously justified for Maxwell molecules with cut-off (see Section 2.3.3), the classical McKean-Vlasov diffusion (with a non-optimal convergence rate) and a mixed jump-diffusion model. The main advantage of this abstract method is its wide range of applicability, although each model requires a careful and dedicated verification of the five assumptions. The choice of the different spaces $\mathcal{G}$ indeed strongly depends on the structure of the model. In the companion paper [130], the abstract method is developed in a more general framework: the five assumptions are modified to include conservation relations in order to treat the case of Boltzmann models with unbounded collision rates, possibly uniformly in time. The results will be summarised in Section II-4.5.

Remark 23 (BBGKY hierarchy, statistical solution and limit generator). We previously made the remark that taking the limit $N \rightarrow +\infty$ in the BBGKY hierarchy (51) or (52) leads to an infinite hierarchy of equations called the Boltzmann hierarchy (Remark 15). By the Hewitt-Savage theorem, at every time $t > 0$, the Boltzmann hierarchy is associated to a unique $\pi_t \in \mathcal{P}(\mathcal{P}(E))$ which is sometimes called a *statistical solution* of the limit problem. In [130, section 8], the authors show that for cutoff Boltzmann models, given an initial $\pi_0 \in \mathcal{P}(\mathcal{P}(E))$, the statistical solution π_t is the unique solution to the evolution equation

$$\partial_t \pi = \mathcal{L}^\infty \pi_t,$$

where $\mathcal{L}^\infty$ is the formal adjoint of the limit generator $\mathcal{L}_\infty$ on $C_b(\mathcal{P}(E))$. It means that π_t satisfies the weak equation

$$\forall \Phi \in C_b(\mathcal{P}(E)), \quad \frac{d}{dt} \langle \pi_t, \Phi \rangle = \langle \pi_t, \mathcal{L}_\infty \Phi \rangle.$$

When $\pi_0 = \delta_{f_0}$, there exists a unique statistical solution which is *chaotic* in the sense that this solution is equal to δ_{f_t} where f_t solves the nonlinear PDE. As already explained many times, this fact is equivalent to the propagation of chaos. Moreover it is proven in [130, section 8] that the operators which generate (in a sense which is made rigorous) the BBGKY-hierarchy converge towards the generators of the processes related to moment measures of π_t . However it should be noted that in general there exists many other statistical solutions. The notion of statistical solution is an important notion in fluid mechanics, where it originates.

4.4 Large Deviation Related Methods

Various approaches related to large deviations theory are investigated here. It is possible to motivate them by looking back at Remark 14, which suggests that chaos can be seen as a kind of weak law of large numbers, as it implies the weak convergence:

$$\langle \mu_{\mathcal{X}^N}, \varphi \rangle - \mathbb{E}_{\mathcal{X}^N} [\varphi(X^{1,N})] \xrightarrow{N \rightarrow +\infty} 0.$$

When a strong law of large numbers holds, it is natural to look at the fluctuations of $\langle \mu_{\mathcal{X}^N}, \varphi \rangle$ by establishing some weak central limit theorem. Nonetheless one can look at this issue the other way round, trying to deduce some weak law of large numbers from a fluctuation result. Indeed, the usual central limit theorem implies a weak version of the law of large numbers, although the latter is classically proven using quite different tools. Actually, it should be noted that, if a typical large deviation estimate of the form $\mathbb{P}(D(\mu_{\mathcal{X}_t^N}, f_t) > \varepsilon) \leq e^{-NI(\varepsilon)}$ holds for a distance D on $\mathcal{P}(E)$ and I a rate function which has a unique zero, then propagation of chaos immediately follows from the Borel-Cantelli lemma. Note however that if one is only interested in the law of large numbers, a large deviation result may be quite overworked. As we shall see, the large deviation theory provides new tools and useful insights on propagation of chaos. In Section 4.4.1 we give a mostly historical description of large deviation results which imply as a byproduct a weak form propagation of chaos in some specific cases. These results are related to Laplace's

theory of fluctuations, which has been widely used in statistical physics to study out of equilibrium systems. The relative entropy functional (Definition 3.10) plays a crucial role in this analysis: in Section 4.4.2, we gather classical results which link propagation of chaos and entropy bounds. Section 4.4.3 is devoted to the study of (quantitative) concentration inequalities which will be useful in the following sections to strengthen propagation of chaos results. We will later give a brief overview of “pure” large deviation results which go beyond propagation of chaos in Section II-5.4.1. Some classical material on large deviation theory can be found in Appendix A.5.

4.4.1 Chaos through Large Deviation Principles

In the seminal article [10], the authors improve results from [107] and [23] on Large Deviation Principles (LDP) for Gibbs measure and obtain as a byproduct a pathwise propagation of chaos result for the McKean-Vlasov diffusion. Firstly, [10, Theorem A] below states a large deviation principle for Gibbs measures with a polynomial potential.

Theorem 4.6 (Polynomial Potential). *Let $\mathcal{E}$ be a Polish measurable space. Let $\mu_0 \in \mathcal{P}(\mathcal{E})$. Let us consider a random vector $\mathcal{X}^N$ in $\mathcal{E}^N$, distributed according to the Gibbs measure:*

$$\mu^N(d\mathbf{x}^N) = \frac{1}{Z_N} \exp[NG(\mu_{\mathbf{x}^N})] \mu_0^{\otimes N}(d\mathbf{x}^N), \quad (105)$$

where Z_N is a normalization constant and G is a polynomial function on $\mathcal{P}(\mathcal{E})$ (called the energy functional) of the form:

$$G(\mu) = \sum_{k=2}^r \langle \mu^{\otimes k}, V_k \rangle,$$

for some symmetric continuous bounded functions V_k on $\mathcal{E}^k$. Then the laws of $\mu_{\mathcal{X}^N}$ satisfy a large deviation principle in $\mathcal{P}(\mathcal{P}(\mathcal{E}))$ with speed N^{-1} and rate function $\mu \mapsto H(\mu|\mu_0) - G(\mu) - \inf_{\mathcal{P}(\mathcal{E})}(H(\cdot|\mu_0) - G)$.

Denote by m_0 the infimum of $H(\cdot|\mu_0) - G$ in $\mathcal{P}(\mathcal{E})$ and $\mathcal{P}^{m_0}$ the set of probability measures which achieve it. The study of $\mathcal{P}^{m_0}$ is related to the study of the quadratic form Θ_ν on $L_0^2(\mathcal{E}, d\nu)$ (the space of centered square ν -integrable functions on E) defined for any ν in $\mathcal{P}^{m_0}$ by:

$$\forall f, g \in L_0^2(\mathcal{E}, d\nu), \quad \langle \Theta_\nu f, g \rangle := \sum_{k=2}^r k(k-1) \langle \nu^k, f \otimes g \otimes 1^{k-2} V_k \rangle.$$

The following [10, Theorem B] quantifies the fluctuations of $\mu_{\mathcal{X}^N}$ in the *non-degenerate case*. Analogous results for the degenerate case are given in [10, Theorem C].

Theorem 4.7 (Chaos and Fluctuations). *Assume that $\mathcal{P}^{m_0}$ is non degenerate in the sense that for all $\nu \in \mathcal{P}^{m_0}$,*

$$\text{Ker}(\text{Id} - \Theta_\nu) = \{0\}.$$

In this case, let us consider the quantities:

$$d(\nu) := [\det(\text{Id} - \Theta_\nu)]^{-1/2}, \quad \bar{d}(\nu) := \frac{d(\nu)}{\sum_{\nu' \in \mathcal{P}^{m_0}} d(\nu')}.$$

Then the following properties hold.

1. The set $\mathcal{P}^{m_0}$ is finite.
2. $\lim_{N \rightarrow \infty} e^{Nm_0} Z_N = \sum_{\nu \in \mathcal{P}^{m_0}} d(\nu)$.
3. For any integer $k \geq 1$ and any φ in $C_b(\mathcal{E}^k)$,

$$\langle \mu^N, \varphi \otimes 1^{\otimes N-k} \rangle \xrightarrow{N \rightarrow \infty} \sum_{\nu \in \mathcal{P}^{m_0}} \bar{d}(\nu) \langle \nu^{\otimes k}, \varphi \rangle.$$

4. The random measures $\mu_{\mathcal{X}^N} \in \mathcal{P}(\mathcal{E})$ satisfy local and global Central Limit Theorems.

When $\mathcal{P}^{m_0}$ reduces to a single non-degenerate minimizer f , then the third assertion exactly tells that the sequence $(\mu^N)_N$ is f -chaotic.

Going back to an interacting particle system, let $f_0^N \in \mathcal{P}(E^N)$ be an initial law and $f_{[0,T]}^N \in \mathcal{P}(C([0,T], E^N)) \simeq \mathcal{P}(C([0,T], E)^N)$ be the pathwise law of the particle system with initial law f_0^N . In the same way, let $f_{[0,T]} \in \mathcal{P}(C([0,T], E))$ be the law of the targeted limit nonlinear process with initial law $f_0 \in \mathcal{P}(E)$. Following [10] and [11], pathwise chaos on $[0, T]$ can be recovered from the above theorem, essentially by taking $\mathcal{E} = C([0, T], E)$.

Corollary 3 (Pathwise chaos from LDP). *Assume that the following properties hold.*

1. f_0^N is a Gibbs measure of the form (105) with respect to $f_0^{\otimes N}$ for a polynomial energy functional $G \in C_b(\mathcal{P}(E))$.
2. The functional $\mu \in \mathcal{P}(E) \mapsto H(\mu|f_0) - G(\mu)$ admits a unique minimizer $\mu_\star$ which is non-degenerate.
3. $f_{[0,T]}^N$ is a Gibbs measure of the form (105) with respect to $f_{[0,T]}^{\otimes N}$ for a polynomial energy functional $\mathcal{G} \in C_b(\mathcal{P}(C([0,T], E)))$.
4. The functional $\nu \in \mathcal{P}(C([0,T], E)) \mapsto H(\nu|f_{[0,T]}) - \mathcal{G}(\nu)$ has a unique minimizer $f_{[0,T]}^\star$ which is non-degenerate. Moreover, $f_{[0,T]}^\star$ is the pathwise law of the nonlinear process with initial condition $\mu_\star$.

Then the sequence $(f_{[0,T]}^N)_N$ is $f_{[0,T]}^\star$ -chaotic.

The two first two assumptions are related to the initial data. The *propagated* property is more the LDP than the chaoticity since the initial measure is assumed to be Gibbsian and no more chaotic as usual. To recover the usual setting, the first assumption has to be replaced by the f_0 -chaoticity of f_0^N , that is to say $G = 0$. The unique minimizer of $H(\cdot|\mu_0)$ is thus in this case $\mu_\star = f_0$ and $f_{[0,T]}^\star = f_{[0,T]}$ is the desired law for the limit nonlinear process. The third assumption tells that the Gibbs form of the density is also valid at

the pathwise level. For a McKean-Vlasov diffusion with regular coefficients (typically Lipschitz [52],[121]), the Gibbs density $\frac{df_{[0,T]}^N}{df_{[0,T]}^{\otimes N}}$ can typically be computed using Girsanov's formula (see Appendix A.6). Thus the remaining difficulty often lies in the fourth point.

Example 22 (Application to several models). Checking that the above assumptions hold can be very technical. To give a flavour of the possible applications, we mention here a few examples.

- (*McKean-Vlasov system with regular gradient forces and constant diffusion*). The assumptions are exhaustively checked in the original paper [10], leading to the desired pathwise chaos on finite time intervals.
- (*McKean-Vlasov system with only continuous drift and Hölder position dependent diffusion*). Pathwise chaos is proved in the seminal work [55] by establishing a LDP principle and by showing that the limit law is the only minimizer of the related rate function. The method is close to the one which is described above, but it is driven in some abstract dual spaces in order to weaken the regularity conditions: the diffusion can depend on the position with Hölder-regularity (but does not depend on its law), but no such regularity is needed for the non-linear drift.
- (*Hamiltonian systems with random medium interactions*). Pathwise chaos is proved in [52] by extending the above method. The main difficulty in this system comes from the control of the random jumps and from the random medium.
- (*Curie-Weiss and Kuramoto models*). Once again, it is an application of the above method. The Curie-Weiss model is obtained as a corollary in [55] and [52], while the Kuramoto model is the last part of [52]. The method is applied to analogous jump processes with random interactions in [115]. See also Section II-5.2.1.

4.4.2 Chaos from entropy bounds

The theorems stated in the previous Section 4.4.1 strongly suggest that the relative entropy (between the N -particle distribution and the tensorised limit law) is an important quantity to look at. In fact, Pinsker inequality (50) implies that

$$\|f_t^N - f_t^{\otimes N}\|_{\text{TV}}^2 \leq 2H(f^N | f^{\otimes N}),$$

so if the right-hand side goes to zero as $N \rightarrow +\infty$, it implies propagation of chaos in Total Variation norm. But as it can be expected, it is very demanding to prove that the relative entropy vanishes. The following lemma shows that a simple bound may be sufficient for a slightly weaker result.

Lemma 4.8 (Dimensional bounds on entropy, [50]). *Let $\mathcal{E}$ be a measurable space. For every symmetric probability measure f^N on $\mathcal{E}^N$, and every nonnegative integer $k(N) \leq N$, it holds that*

$$H(f^{k(N),N} | f^{\otimes k(N)}) \leq \frac{k(N)}{N} H(f^N | f^{\otimes N}). \quad (106)$$

A bound on $H(f^N|f^{\otimes N})$ thus implies propagation of chaos in Total Variation norm for blocks of size $k(N) = o(N)$. This technique is by now classical and various applications will be presented in the following sections.

Remark 24. Note that a bound on $H(f^N|f^{\otimes N})$ implies that the normalised entropy goes to zero as $N \rightarrow +\infty$:

$$\tilde{H}(f^N|f^{\otimes N}) := \frac{1}{N} H(f^N|f^{\otimes N}) \xrightarrow{N \rightarrow +\infty} 0.$$

A first historical example of entropy bound for Gibbs measures (with a continuous bounded but non necessarily polynomial potential) can be found in the article [11] subsequent to [10].

Theorem 4.9 (Entropy bound for Gibbs measures, [11]). *Let μ^N be a non degenerate Gibbs measure of the form (105). Then, with the same notations as in Theorem 4.7, the measure $\mu_\star^N = \sum_{\nu \in \mathcal{P}^{m_0}} \bar{d}(\nu) \nu^{\otimes N}$ satisfies the entropy bound*

$$\limsup_{N \rightarrow +\infty} H(\mu^N | \mu_\star^N) < +\infty.$$

Note that this result strengthens [11, Theorem B] (Theorem 4.7). As before, this readily implies pathwise propagation of chaos.

Corollary 4 (Pathwise McKean-Vlasov, C_b^2 potentials [11]). *If the rate function has a unique zero [11, Assumption (A1)], the pathwise entropy bound on $[0, T]$ holds for McKean-Vlasov gradient systems with C_b^2 coefficients.*

Theorem 4.9 is very strong and quite general. It is mainly intended for true Gibbs measures and in our case, it may look too powerful (besides, the hypothesis may not be easily checked). In the literature, there are more direct approaches to bound the entropy.

The two following lemmas are valid respectively in the pathwise and the pointwise cases for the McKean-Vlasov diffusion. They do not directly prove a propagation of chaos result but suggest a way to bound the relative entropy under very weak assumptions on the drift. They are a direct consequence of Girsanov theorem (Appendix A.6) which provides an explicit expression of the relative entropy as an observable of the particle system. The classical application is a strengthening result: if this observable can be controlled by a weak form of propagation of chaos, then the entropy bound strengthens the weak propagation of chaos result into strong (pathwise) propagation of chaos in TV norm, see for instance Corollary II-1. Moreover, in some cases where the drift is singular and when no easy propagation of chaos result is available, computing a bound on only one observable is often a quite sensible strategy. This will be reviewed in particular in Section II-3.4.

Lemma 4.10 (Pathwise entropy bound). *Let $T > 0$ and $I = [0, T]$. Assume that the nonlinear martingale problem associated to the McKean-Vlasov diffusion (Definition 2.3 and (10)) with*

$$b : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d, \quad \sigma = I_d,$$

is wellposed and let $f_I \in \mathcal{P}(C([0, T], \mathbb{R}^d))$ be its solution. For $N \in \mathbb{N}$, let $f_I^N \in \mathcal{P}(C([0, T], (\mathbb{R}^d)^N))$ be the law of the associated particle system $(\mathcal{X}_t^N)_t$. Then, for any $k \leq N$ it holds that

$$H(f_I^{k,N} | f_I^{\otimes k}) \leq \frac{k}{2} \mathbb{E} \left[\int_0^T |b(X_t^1, \mu_{\mathcal{X}_t^N}) - b(X_t^1, f_t)|^2 dt \right]. \quad (107)$$

This lemma is a mere application of Girsanov's theorem; for simplicity, the result is stated in the case of a constant diffusion matrix but it is also valid in the case of a diffusion matrix which depends on the positional argument but not on the measure argument (see Remark 25).

Proof. Since the nonlinear martingale is well-posed, it is well-known (see [106, Chapter 5, Proposition 4.6] or [75, Chapter 5, Proposition 3.1]) that we can construct a filtration and N independent adapted f_I -Brownian motions $(\overline{B}_t^i)_t$ on the path space such that

$$d\mathbf{X}_t^i = b(\mathbf{X}_t^i, f_t)dt + d\overline{B}_t^i, \quad (108)$$

where we recall that $\mathbf{X}_t^N = (\mathbf{X}_t^1, \dots, \mathbf{X}_t^N)$ is the canonical process on $C([0, T], \mathbb{R}^{dN}) \simeq C([0, T], \mathbb{R}^d)^N$. In other word, the canonical process is a weak solution of the nonlinear McKean-Vlasov SDE on the path space $(C([0, T], \mathbb{R}^d), \mathcal{F}, f_I)$. For $i \in \{1, \dots, N\}$ let us define the processes:

$$\Delta_t^i = b(\mathbf{X}_t^i, \mu_{\mathbf{X}_t^N}) - b(\mathbf{X}_t^i, f_t).$$

and

$$H_t^N := \sum_{i=1}^N \left[\int_0^t \Delta_s^i \cdot d\overline{B}_s^i - \frac{1}{2} \int_0^t |\Delta_s^i|^2 ds \right].$$

It is classical to check that $\exp(H^N)$ defines a f_I -martingale (see [106, Chapter 3, Corollary 5.16]). Then by Girsanov theorem (see Appendix A.6) on the product space $(C([0, T], \mathbb{R}^d)^N, \mathcal{F}^{\otimes N}, f_I^{\otimes N})$, it is possible to define a probability measure f_I^N on $C([0, T], \mathbb{R}^d)^N$ such that for $i \in \{1, \dots, N\}$, the processes

$$B_t^i := \overline{B}_t^i - \int_0^t \mathbf{X}_s^i ds \quad (109)$$

are N independent f_I^N -Brownian motions. Reporting (109) into (108) we see that

$$d\mathbf{X}_t^i = b(\mathbf{X}_t^i, \mu_{\mathbf{X}_t^N})dt + dB_t^i.$$

In other words, $(\mathbf{X}_t^N)_t$ is a weak solution of the McKean-Vlasov particle system on the path space $(C([0, T], \mathbb{R}^d)^N, \mathcal{F}, f_I^N)$ and, as the notation implies, f_I^N is the N -particle distribution. Moreover, the Girsanov theorem gives a formula for the Radon-Nikodym derivative:

$$\frac{df_I^N}{df_I^{\otimes N}} = \exp(H_T^N) = \exp \left(\sum_{i=1}^N \left[\int_0^T \Delta_s^i \cdot d\overline{B}_s^i - \frac{1}{2} \int_0^T |\Delta_s^i|^2 ds \right] \right).$$

As an immediate consequence, it is possible to compute the relative entropy as follows:

$$\begin{aligned}
H(f_I^N | f_I^{\otimes N}) &:= \mathbb{E}_{f_I^N} \left[\log \frac{df_I^N}{df_I^{\otimes N}} \right] \\
&= \mathbb{E}_{f_I^N} [H_T^N] \\
&= \mathbb{E}_{f_I^N} \left[\sum_{i=1}^N \left[\int_0^t \Delta_t^i \cdot d\bar{B}_t^i - \frac{1}{2} \int_0^T |\Delta_t^i|^2 dt \right] \right] \\
&= \mathbb{E}_{f_I^N} \left[\sum_{i=1}^N \left[\int_0^T \Delta_t^i \cdot dB_t^i + \frac{1}{2} \int_0^T |\Delta_t^i|^2 dt \right] \right] \\
&= \frac{1}{2} \mathbb{E}_{f_I^N} \left[\sum_{i=1}^N \int_0^T |\Delta_t^i|^2 dt \right],
\end{aligned}$$

which, by exchangeability, eventually gives:

$$H(f_I^N | f_I^{\otimes N}) = \frac{N}{2} \mathbb{E}_{f_I^N} \left[\int_0^T |b(\mathbf{X}_t^1, \mu_{\mathbf{X}_t^N}) - b(\mathbf{X}_t^1, f_t)|^2 dt \right].$$

Coming back to our usual notations on the abstract probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which a particle system $(\mathcal{X}_t^N)_t \sim f_I^N$ is defined, it simply means that

$$H(f_I^N | f_I^{\otimes N}) = \frac{N}{2} \mathbb{E} \left[\int_0^T |b(X_t^1, \mu_{\mathcal{X}_t^N}) - b(X_t^1, f_t)|^2 dt \right].$$

The conclusion follows from Lemma 4.8. $\square$

Remark 25. The inequality (107) is actually an equality (see the proof of [108, Theorem 2.6(3)]). This relatively direct computation can be seen as a very special case of [116, Theorem 2.4]. The result readily extends to the case of time-dependent parameters b, σ and to the case of a non constant diffusion matrix $\sigma \equiv \sigma(t, x)$ which does not depend on the measure argument and which is assumed to be invertible everywhere. The only difference in (107) is that b should be replaced by $\sigma^{-1}b$. An even more general setting is the one given in [108] where

$$b : [0, T] \times C([0, T], \mathbb{R}^d) \times \mathcal{P}(C([0, T], \mathbb{R}^d)) \rightarrow \mathbb{R}^d, \quad \sigma : [0, T] \times C([0, T], \mathbb{R}^d) \rightarrow \mathcal{M}_d(\mathbb{R}),$$

are assumed to be jointly measurable. This does not affect the final result (107) nor the argument.

Remark 26. It is worth noticing that this approach does not seem to be restricted to diffusion processes. On the one hand, the full Girsanov theory can be applied to jump processes as well (see [116] and the references therein) and it is actually a very powerful and general result in the theory of stochastic integration [111, Section 5.5]. On the other hand, any model presented in this review can be written as the solution of a very general martingale problem. To the best of our knowledge, an analogous generalised result does not seem to exist in the literature yet. For the Nanbu system, it may be contained in [113, Theorem 2.11].

In a pointwise setting, the time derivative of the relative entropy can be directly computed using the generator of the particle system.

Lemma 4.11 (General bound on the time-derivative entropy). *Let $f_t \in \mathcal{P}(\mathbb{R}^d)$ be the solution of (12) at time t with*

$$b : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d, \quad \sigma = I_d,$$

and let $f_t^N \in \mathcal{P}((\mathbb{R}^d)^N)$ be the law of the associated particle system. Then for every $\alpha > 0$ it holds that

$$\frac{d}{dt} H(f_t^N | f_t^{\otimes N}) \leq \frac{\alpha - 1}{2} I(f_t^N | f_t^{\otimes N}) + \frac{N}{2\alpha} \mathbb{E} \left[|b(X_t^1, \mu_{\mathcal{X}_t^N}) - b(X_t^1, f_t)|^2 \right]. \quad (110)$$

The following proof is mostly formal as we assume that the limit f_t and $\log f_t$ are regular enough to be taken as test functions. The computations can be fully justified in the cases where the lemma will be applied.

Proof. Let us recall that the generator of the N -particle system is defined by

$$\mathcal{L}_N \varphi_N(\mathbf{x}^N) = \sum_{i=1}^N L_{\mu_{\mathbf{x}^N}} \diamond_i \varphi_N(\mathbf{x}^N),$$

where, given $\mu \in \mathcal{P}(E)$,

$$L_\mu \varphi(x) := \langle b(x, \mu), \nabla \varphi \rangle + \frac{1}{2} \Delta \varphi.$$

The Kolmogorov equation for the N -particle system reads:

$$\frac{d}{dt} \langle f_t^N, \varphi_N \rangle = \langle f_t^N, \mathcal{L}_N \varphi_N \rangle.$$

The Kolmogorov equation associated to a system of N independent f_t -distributed particles reads:

$$\frac{d}{dt} \langle f_t^{\otimes N}, \varphi_N \rangle = \langle f_t^{\otimes N}, L_{f_t}^{\otimes N} \varphi_N \rangle,$$

where we define the generator

$$L_{f_t}^{\otimes N} \varphi_N := \sum_{i=1}^N L_{f_t} \diamond_i \varphi_N.$$

The relative entropy is defined by:

$$H(f_t^N | f_t^{\otimes N}) = \mathbb{E}_{\mathcal{X}_t^N \sim f_t^N} \left[\log \frac{df_t^N}{df_t^{\otimes N}}(\mathcal{X}_t^N) \right] \equiv \left\langle f_t^N, \log \frac{f_t^N}{f_t^{\otimes N}} \right\rangle.$$

In the last term, $\frac{df_t^N}{df_t^{\otimes N}}$ has been replaced by $\frac{f_t^N}{f_t^{\otimes N}}$, which makes sense since f_t^N and f_t are probability density functions. Using the product derivation rule:

$$\frac{d}{dt} H(f_t^N | f_t^{\otimes N}) = \left\langle f_t^N, \mathcal{L}_N \left(\log \frac{f_t^N}{f_t^{\otimes N}} \right) \right\rangle + \left\langle f_t^N, \frac{d}{dt} \log \frac{f_t^N}{f_t^{\otimes N}} \right\rangle$$

The last term can be written

$$\left\langle f_t^N, \frac{d}{dt} \log \frac{f_t^N}{f_t^{\otimes N}} \right\rangle = \left\langle f_t^N, \frac{f_t^{\otimes N}}{f_t^N} \frac{d}{dt} \frac{f_t^N}{f_t^{\otimes N}} \right\rangle = \left\langle f_t^{\otimes N}, \frac{d}{dt} \frac{f_t^N}{f_t^{\otimes N}} \right\rangle.$$

The mass conservation for f_t^N gives

$$\left\langle f_t^{\otimes N}, \frac{f_t^N}{f_t^{\otimes N}} \right\rangle = \langle f_t^N, 1 \rangle = 1,$$

and therefore

$$\left\langle f_t^{\otimes N}, \frac{d}{dt} \frac{f_t^N}{f_t^{\otimes N}} \right\rangle = - \left\langle f_t^{\otimes N}, L_{f_t}^{\otimes N} \frac{f_t^N}{f_t^{\otimes N}} \right\rangle.$$

Using the definition of the generator, we get:

$$L_{f_t}^{\otimes N} \frac{f_t^N}{f_t^{\otimes N}}(\mathbf{x}^N) = \sum_{i=1}^N \left\langle b(x^i, f_t), \frac{f_t^N}{f_t^{\otimes N}}(\mathbf{x}^N) \nabla_{x^i} \log \frac{f_t^N}{f_t^{\otimes N}}(\mathbf{x}^N) \right\rangle + \frac{1}{2} \Delta \left(\frac{f_t^N}{f_t^{\otimes N}} \right).$$

And thus it holds that

$$\begin{aligned} \frac{d}{dt} H(f_t^N | f_t^{\otimes N}) &= \mathbb{E}_{f_t^N} \left[\sum_{i=1}^N \left\langle b(X_t^i, \mu_{\mathcal{X}_t^N}) - b(X_t^i, f_t), \nabla_{x^i} \log \frac{f_t^N}{f_t^{\otimes N}}(\mathcal{X}_t^N) \right\rangle \right] \\ &\quad + \frac{1}{2} \left\langle f_t^N, \Delta \left(\log \frac{f_t^N}{f_t^{\otimes N}} \right) \right\rangle - \frac{1}{2} \left\langle f_t^{\otimes N}, \Delta \left(\frac{f_t^N}{f_t^{\otimes N}} \right) \right\rangle \\ &= \mathbb{E}_{f_t^N} \left[\sum_{i=1}^N \left\langle b(X_t^i, \mu_{\mathcal{X}_t^N}) - b(X_t^i, f_t), \nabla_{x^i} \log \frac{f_t^N}{f_t^{\otimes N}}(\mathcal{X}_t^N) \right\rangle \right] \\ &\quad - \frac{1}{2} \left\langle f_t^{\otimes N}, \left| \nabla \frac{f_t^N}{f_t^{\otimes N}} \right|^2 \right\rangle. \end{aligned} \tag{111}$$

and the last term involves

$$\left\langle f_t^{\otimes N}, \left| \nabla \frac{f_t^N}{f_t^{\otimes N}} \right|^2 \right\rangle = \left\langle f_t^N, \left| \nabla \log \frac{f_t^N}{f_t^{\otimes N}} \right|^2 \right\rangle =: I(f_t^N | f_t^{\otimes N}).$$

Therefore Cauchy-Schwarz inequality and Young inequality give for any $\alpha > 0$

$$\frac{d}{dt} H(f_t^N | f_t^{\otimes N}) \leq \frac{\alpha - 1}{2} I(f_t^N | f_t^{\otimes N}) + \frac{1}{2\alpha} \sum_{i=1}^N \mathbb{E} \left[|b(X_t^i, \mu_{\mathcal{X}_t^N}) - b(X_t^i, f_t)|^2 \right].$$

The conclusion follows since particles are exchangeable. $\square$

Remark 27. Several points should be noticed :

- It is possible to take $\alpha = 1$ in order to get rid of the Fisher information as in [92], but a further control on $W_2(\mu_{\mathcal{X}_t^N}, f_t)$ is then needed, see [92].

- Before the splitting which introduces α , Cauchy-Schwarz's inequality would have lead to a bound close to the *HWI* inequality in our special case.

To end this section, we would like to emphasize the fact that these results do not require any particular regularity on the drift (this is a well-known but remarkable property of Girsanov's transform). As a general rule, entropy related methods are well suited to handle cases with singular interactions. An example with exceptionally weak regularity assumptions will be given in Section II-3.4. Another example with a complex abstract interaction mechanism will be presented in Section II-3.6.2.

4.4.3 Tools for concentration inequalities

Large Deviation principles imply propagation of chaos, but they do not always give a way to quantify it since their result is often purely asymptotic (for instance, Sanov theorem is non-quantitative). In this section, we gather some results which quantify the deviation of an empirical measure of N samples around its mean. These results are valid for any fixed (sufficiently large) N . We first state a classical concentration inequality, obtained as the consequence of a Log-Sobolev inequality. This inequality and other related functional inequalities are deep structural properties of the system which will also be used to study ergodic properties and long-time propagation of chaos (Section II-3.1.3). Then we state quantitative versions of Sanov theorem which strengthen the above concentration inequality. In this section, E is assumed to be $\mathbb{R}^d$ (or at least a smooth finite dimensional manifold).

The following Log-Sobolev inequality is another kind of entropy bound.

Definition 4.12 (Log-Sobolev Inequality). For $\lambda > 0$, a probability measure μ with finite second moment satisfies a Logarithmic-Sobolev Inequality $LSI(\lambda)$ when for all ν in $\mathcal{P}(E)$

$$H(\nu|\mu) \leq \frac{1}{2\lambda} I(\nu|\mu),$$

where I is the Fisher information (Definition 3.10).

An important consequence is the following lemma.

Lemma 4.13 (Concentration, see Ledoux [112]). *If a probability measure μ satisfies a $LSI(\lambda)$ then for any Lipschitz test function φ with Lipschitz constant bounded by 1 and for any $\varepsilon > 0$, it holds that*

$$\mathbb{P}_{X \sim \mu}(|\varphi(X) - \mathbb{E}[\varphi(X)]| \geq \varepsilon) \leq 2e^{-\frac{\lambda \varepsilon^2}{2}}. \quad (112)$$

This lemma is typically applied in E^N for $\mathcal{X}_t^N \sim f_t^N$ with the function

$$\bar{\varphi}(\mathcal{X}_t^N) = \frac{1}{N} \sum_{i=1}^N \varphi(X_t^i),$$

where φ is 1-Lipschitz on E . The Lipschitz norm of the function $\bar{\varphi}$ is bounded by $1/\sqrt{N}$.

A classical result [137, Theorem 1] shows that under mild assumptions, the Log-Sobolev inequality also implies the following Talagrand inequality.

Definition 4.14 (Talagrand Inequalities). For any real $p \geq 1$ and $\lambda > 0$, a probability measure μ with finite p -th moment satisfies a Talagrand inequality $T_p(\lambda)$ when for all $\nu \in \mathcal{P}(E)$,

$$W_p(\nu, \mu) \leq \sqrt{\frac{2H(\nu|\mu)}{\lambda}}. \quad (113)$$

This inequality is all the more strong as λ and p are big (it is a consequence of Jensen's inequality). It is known that T_2 implies some Poincaré inequality and a handful characterization is available for T_1 inequalities, see [22]. Talagrand inequalities are also useful to quantify ergodicity with respect to the Wassertein distance. Here is another characterization.

Lemma 4.15 (Square-exponential moment). *A probability measure μ with finite expectation satisfies a T_1 inequality if and only if there exist $\alpha > 0$ and $x \in E$ such that $\int_E e^{\alpha|x-y|^2} \mu(dy) < +\infty$.*

Note that Talagrand inequalities allow to recover usual Wassertein convergence (and then convergence in law) from entropic convergence. Concentration inequalities can also stem from Talagrand inequalities, although the stronger Logarithmic Sobolev inequality is more often used in this context. The Logarithmic Sobolev inequality can be established thanks to the following criterion.

Proposition 13 (Bakry, Emery [6, 5]). *Consider a diffusion process with semi-group $(P_t)_{t \geq 0}$, generator $\mathcal{L}$ and carré du champ operator $\Gamma : (\varphi, \psi) \mapsto \frac{1}{2}[\mathcal{L}(\varphi\psi) - \varphi\mathcal{L}(\psi) - \psi\mathcal{L}(\varphi)]$. Assume that there exists a real λ such that for every regular function φ*

$$\Gamma_2(\varphi) \geq \lambda \Gamma(\varphi)$$

where $\Gamma_2(\varphi) = \frac{1}{2}[\mathcal{L}(\Gamma(\varphi)) - 2\Gamma(\varphi\mathcal{L}(\varphi))]$. Then the following properties hold.

- For all $x \in \mathbb{R}^d$ and all $t > 0$, if $P_0 = \delta_x$, then P_t satisfies $LSI\left(\frac{\lambda}{1-e^{-\lambda t}}\right)$ (where the semi-group P_t is identified with the transition probability measure that it generates).
- If $\lambda > 0$, the semi-group is ergodic and P_t converges towards the invariant measure μ with rate

$$H(P_t|\mu) \leq Ce^{-2\lambda t}.$$

In the previous concentration result (112), the test function is fixed before computing the probability. One could take the supremum over all 1-Lipschitz test functions, this would correspond to a weak chaos for the D_1 distance. To get stronger estimates on the Wasserstein distance between $\mu_{\mathcal{X}_t^N}$ and f_t , the supremum needs to come inside the probability. This is not an easy task, it requires a quantitative version of Sanov theorem, which is proved in [21, Theorem 2.1].

Theorem 4.16 (Pointwise Quantitative Sanov [21]). *Consider a probability measure μ on $\mathbb{R}^d$ which satisfies $T_p(\lambda)$ for $p \in [1, 2]$ and $\lambda > 0$ and which has a bounded square-exponential moment. Let $\mathcal{X}^N \sim \mu^{\otimes N}$ be a system of N i.i.d μ -distributed random variables.*

Then for any $\lambda' < \lambda$ and $\varepsilon > 0$, there exists a constant N_ε (which depends also on d and the square-exponential moment of μ) such that for all $N \geq N_\varepsilon$,

$$\mathbb{P}(W_p(\mu_{\mathcal{X}^N}, \mu) > \varepsilon) \leq e^{-\gamma_p \frac{\lambda'}{2} N \varepsilon^2},$$

where $\gamma_p > 0$ is an explicit constant which depends only on p .

The following pathwise generalization is proved in [18, Theorem 1].

Theorem 4.17 (Pathwise Quantitative Sanov and Pathwise chaos [18]). *Under the same assumptions, the above theorem holds for a measure μ on the Hölder space $C^{0,\alpha}([0, T], \mathbb{R}^d)$ with $\alpha \in (0, 1]$.*

Except the last one, the results in this section are stated in a static framework. Examples of time dependent systems and applications to propagation of chaos are detailed in Section II-3.5.

4.5 Tools for Boltzmann interactions

4.5.1 Series expansions

Let us consider first the homogeneous Boltzmann system with $L^{(1)} = 0$. When the collision rate λ satisfies the uniform bound (31), then it can be directly checked that the operator $\mathcal{L}_N$ is continuous for the L^∞ norm:

$$\forall \varphi_N \in C_b(E^N), \quad \|\mathcal{L}_N \varphi_N\|_\infty \leq \Lambda(N-1) \|\varphi_N\|_\infty.$$

Without loss of generality (see Proposition 3), we will assume here that $\lambda \equiv \Lambda$ is constant. As a consequence, the exponential series $e^{t\mathcal{L}_N}$ is absolutely convergent for $t < 1/(\Lambda(N-1))$ and there is a semi explicit formula for an observable $\varphi_N \in C_b(E^N)$ at any time $t \geq 0$:

$$\mathbb{E}[\varphi_N(\mathcal{Z}_t^N)] = \sum_{k=0}^{+\infty} \frac{t^k}{k!} \langle f_0^N, \mathcal{L}_N^k \varphi_N \rangle,$$

where $(\mathcal{Z}_t^N)_t$ is the particle process with initial law $f_0^N \in \mathcal{P}(E^N)$. Then, considering a test function $\varphi_N \equiv \varphi_s \otimes 1^{\otimes(N-s)}$ which depends only on s variables for a fixed $s \in \mathbb{N}$, the term on the right-hand side depends only on N through known quantities, namely the initial law f_0^N and the operator $\mathcal{L}_N$. The initial law f_0^N is assumed to be f_0 -chaotic so there is an asymptotic control of all its marginals when $N \rightarrow +\infty$. In its seminal article [104], Kac managed to pass to the limit directly in the series on the right-hand side using a dominated convergence theorem argument. This necessitates in particular to prove the absolute convergence of the series on a time interval independent of N when s is fixed. The argument has been generalised in [34] and will be thoroughly discussed in Section II-4.1. As a byproduct it will show the existence of a solution of the Boltzmann equation in the form of an explicit series expansion. The final formula (II-115) will be a direct extension of the “exponential formula” obtained by McKean in [124] for the solution of a simpler Boltzmann model in $E = \{-1, 1\}$ (the famous 2-state Maxwellian gas). In a

famous work, Wild [163] showed that the solution of the Boltzmann equation for cut-off Maxwellian molecules has a semi-explicit representation in the form of an infinite sum (see also [159, Chapter 4, Section 1], [31] and the references therein). McKean showed that for the 2-state Maxwellian gas, the Formula (II-115) obtained by propagation of chaos can be interpreted as the dual version of a Wild sum.

This approach is only focused on the evolution of observables of the form $\langle f_t^N, \varphi_N \rangle$ for a fixed φ_N and does not study directly the evolution of the N -particle law f_t^N . The evolution of f_t^N is given by the forward Kolmogorov equation. This dual point of view on Kac's theorem is studied in [140] and will also be reviewed in Section II-4.1. The starting point is the BBGKY hierarchy:

$$\partial_t f_t^{s,N} = \frac{s}{N} \mathcal{L}^s f_t^{s,N} + \frac{N-s}{N} \mathcal{C}_{s,s+1} f_t^{s+1,N},$$

where $\mathcal{C}_{s,s+1} : \mathcal{P}(E^{s+1}) \rightarrow \mathcal{P}(E^s)$ is defined for $f^{(s+1)} \in \mathcal{P}(E^{s+1})$ and $\varphi_s \in C_b(E^s)$ by

$$\langle \mathcal{C}_{s,s+1} f^{(s+1)}, \varphi_s \rangle := \sum_{i=1}^s \int_{E^{s+1}} L^{(2)} \diamond_{i,s+1} [\varphi_s \otimes 1](\mathbf{z}^{s+1}) f^{(s+1)}(d\mathbf{z}^{s+1}).$$

Let $\mathbf{T}_N^{(s)}(t) = \exp(t \frac{s}{N} \mathcal{L}^s)$ denote the semi-group acting on $\mathcal{P}(E^s)$ generated by $\frac{s}{N} \mathcal{L}^s$. Then, interpreting the last term on the right-hand side as a perturbation of a linear differential equation, Duhamel's formula reads:

$$f_t^{s,N} = \mathbf{T}_N^{(s)}(t) f_0^{s,N} + \frac{N-s}{N} \int_0^t \mathbf{T}_N^{(s)}(t-\tau) \mathcal{C}_{s,s+1} f_\tau^{s+1,N} d\tau.$$

Iterating this formula gives a semi-explicit series expansion in terms of the initial condition:

$$f_t^{s,N} = \sum_{k=0}^{+\infty} \alpha_N^{(s,k)} \int_0^t \int_0^{t_1} \dots \int_0^{t_{k-1}} \mathbf{T}_N^{(s)}(t-t_1) \mathcal{C}_{s,s+1} \mathbf{T}_N^{(s+1)}(t_1-t_2) \mathcal{C}_{s+1,s+2} \dots \mathcal{C}_{s+k-1,s+k} \mathbf{T}_N^{(s+k)}(t_k) f_0^{s+k,N} dt_1 \dots dt_k,$$

where $\alpha_N^{(s,k)} = (N-s) \dots (N-s-k+1)/N^k$ if $s+k \leq N$ and $\alpha_N^{(s,k)} = 0$ otherwise. Taking the limit $N \rightarrow +\infty$ in this series is the dual viewpoint of the previous approach. If the limit exists, propagation of chaos holds whenever the result can be identified as the infinite hierarchy of tensorised laws $f_t^{\otimes s}$ where f_t solves the Boltzmann equation. Note that this approach can be easily extended to inhomogeneous Boltzmann systems where $L^{(1)} \neq 0$, in which case the semi-group $\mathbf{T}_N^{(s)}$ should be replaced by the semi-group generated by $\sum_{i=1}^s L^{(1)*} \diamond_i + \frac{s}{N} \mathcal{L}^s$. A famous example is given by Lanford's theorem (see Section II-4.6). The study of the evolution of observables is however more natural for abstract systems when there is no known explicit formula for the dual operators $\mathcal{L}^s$ and $\mathcal{C}_{s,s+1}$ acting on the particle probability distributions.

4.5.2 Interaction graph

In an abstract Boltzmann model given by the generator (24) in Section 2.3, the binary interactions can be represented by graph structures. Given a trajectorial realisation of the particle system, the *interaction graph* of a particle (or a group of particles) is built backward in time and retain the genealogical interactions which determine the particle at the current time (i.e. the history of the collisions). Before building graphs from particle realisations, the minimal structure of such a possible graph is detailed in the following definition.

Definition 4.18 (Interaction graph). Consider an index $i \in \{1, \dots, N\}$ (it will stand later for the index of a particle). An interaction graph for i at time $t > 0$ is the data of

1. a k -tuple $\mathcal{T}_k = (t_1, \dots, t_k)$ of *interaction times* $t > t_1 > t_2 > \dots > t_k > 0$,
2. a k -tuple $\mathcal{R}_k = (r_1, \dots, r_k)$ of pairs of indexes, where for $\ell \in \{1, \dots, k\}$, the pair denoted by $r_\ell = (i_\ell, j_\ell)$ is such that $j_\ell \in \{i_0, i_1, \dots, i_{\ell-1}\}$ with the convention $i_0 = i$ and $i_\ell \in \{1, \dots, N\}$.

Such an interaction graph is denoted by $\mathcal{G}_i(\mathcal{T}_k, \mathcal{R}_k)$.

Given a trajectorial realisation of a Boltzmann particle system, the interaction graph of the particle i retains the minimal information needed to compute the state of particle i at time $t > 0$. It is constructed as follows.

- The set $(i_1, \dots, i_k)$ is the set of indexes of the particles which interacted directly or indirectly with particle i during the time interval $(0, t)$ (an indirect interaction means that the particle has interacted with another particle which interacted directly or indirectly with particle i) – note that the i_ℓ 's may not be all distinct.
- The times $(t_1, \dots, t_k)$ are the times at which an interaction occurred.
- For $\ell \in \{1, \dots, k\}$, the indexes (i_ℓ, j_ℓ) are the indexes of the two particles which interacted together at time t_ℓ .

Following the terminology of [89], a *route* of size q between i and j is the union of q elements $r_{\ell_k} = (i_{\ell_k}, j_{\ell_k})$, $k = 1, \dots, q$ such that $i_{\ell_1} = i$, $i_{\ell_{k+1}} = j_{\ell_k}$ and $j_{\ell_q} = j$. A route of size 1 (i.e a single element r_ℓ) is simply called a route. A route which involves two indexes which were already in the graph before the interaction time (backward in time) is called a *recollision*. This construction is more easily understood with the graphical representation of an interaction graph shown on Figure 1.

The definition of interaction graphs can be extended straightforwardly starting from a group of particles instead of only one particle. This representation does not take into account the physical trajectories of the particles, it only retains the history of the interactions among a group of particles. Note that the graph is not a tree in general since the i_ℓ 's are not necessarily distinct. It is a tree when no recollision occurs.

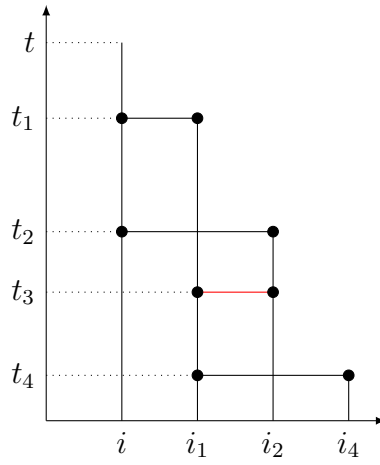


Figure 1: An interaction graph. The vertical axis represents time. Each particle is represented by a vertical line parallel to the time axis. The index of a given particle is written on the horizontal axis. The construction is done backward in time starting from time t where only particle i is present. At each time t_ℓ , if i_ℓ does not already belong to the graph, it is added on the right (with a vertical line which starts at t_ℓ). The couple $r_\ell = (i_\ell, j_\ell)$ of interacting particles at time t_ℓ is depicted by an horizontal line joining two big black dots on the vertical line representing the particles i_ℓ and j_ℓ . for instance, on the depicted graph, $r_2 = (i_2, i)$. Note that at time t_3 , $r_3 = (i_1, i_2)$ (or indifferently $r_3 = (i_2, i_1)$) where i_1 and i_2 were already in the system. Index i_3 is skipped and at time t_4 , the route is $r_4 = (i_4, i_1)$. The *recollision* occurring at time t_3 is depicted in red.

Interaction graphs are a classical tool in the study of Boltzmann particle systems. As we shall see in Section II-4.6 they are particularly useful to give a physical interpretation of the series expansions discussed in the previous section. The connections between interaction graphs and series expansions are more thoroughly discussed in [124] and [31].

In a more probabilistic setting, the following definition extends the construction of an interaction graph to the case of random parameters.

Definition 4.19 (Random interaction graph). Let $\Lambda > 0$, $N \in \mathbb{N}$, $i \in \{1, \dots, N\}$ and $t > 0$. Let $(T^{m,\ell})_{1 \leq k < \ell \leq N}$ be $N(N-1)/2$ independent Poisson processes with rate Λ/N . For each Poisson process $T^{m,\ell}$ we denote by $(T_n^{m,\ell})_n$ its associated increasing sequence of jump times. The sets of times $\mathcal{T}_k = (t_1, \dots, t_k)$ and routes $\mathcal{R}_k = (r_1, \dots, r_k)$ are defined recursively as follows. Initially, $t_0 = t$ and $i_0 = i$ and for $k \geq 0$,

$$t_{k+1} = \max_{\ell,p,n} \{T_n^{i_\ell,p} \mid T_n^{i_\ell,p} < t_k, \ell \leq k\}. \quad (114)$$

Then, given (ℓ, p, n) such that $t_{k+1} = T_n^{i_\ell,p}$, $i_{k+1} = p$ and $j_{k+1} = i_\ell$ so that $r_{k+1} = (i_{k+1}, j_{k+1})$. The procedure is stopped once the set on the right-hand side of (114) is empty (it happens almost surely after a finite number of iterations). The resulting interaction graph $\mathcal{G}_i(\mathcal{T}_k, \mathcal{R}_k)$ is called the random interaction graph with rate Λ rooted on i at time t . The definition is extended similarly starting from a finite number of indexes $(i_0, i_1, \dots, i_k)$ instead of just i .

As explained before, a realisation of a Boltzmann particle system immediately gives an interaction graph for each particle. More importantly, given an interaction graph, it is possible to construct a forward realisation of a Boltzmann particle. More precisely, when the interaction graph is sampled as a random interaction graph following, then the following straightforward lemma constructs a forward realisation of a stochastic process whose pathwise law is equal to $f_{[0,t]}^{1,N}$, the first marginal of the law $f_{[0,t]}^N$ of a Boltzmann particle system given by the generator (24) on the time interval $[0, t]$.

Lemma 4.20. *Let us consider the Boltzmann setting given by Assumption 1 together with the uniform bound (31) on λ . Given a realisation of a random interaction graph sampled beforehand as in the previous definition, apply the following procedure:*

1. At time $t = 0$, let the particles $Z_0^{i_\ell}$ be distributed according to the initial law.
2. Between two collision times, the particles evolve according to $L^{(1)}$.
3. At a collision time t_ℓ , with probability $\lambda(Z_{t_\ell}^{i_\ell}, Z_{t_\ell}^{j_\ell})/\Lambda$, the new states of particles i_ℓ and j_ℓ are sampled according to

$$(Z_{t_\ell}^{i_\ell}, Z_{t_\ell}^{j_\ell}) \sim \Gamma^{(2)}(Z_{t_\ell}^{i_\ell}, Z_{t_\ell}^{j_\ell}, dz_1, dz_2).$$

Then the process $(Z_s^i)_{s \leq t}$ is distributed according to the one-particle marginal $f_{[0,t]}^{1,N}$ of the law $f_{[0,t]}^N$ of a Boltzmann particle system given by the generator (24) on the time interval $[0, t]$.

This trajectorial construction of a particle (sub-)system is used in [89, 127]. Using purely combinatorial arguments the authors prove a pathwise version of Kac's theorem with an explicit optimal convergence rate in total variation norm. The main idea is that it is possible to compute the probability of sampling a *bad* interaction graph, that is a graph which would give a system of particles with too much correlation. When $N \rightarrow +\infty$, this probability goes to zero. As a consequence, when N is large, with high probability, the particle system is close to a system of independent particles which are shown to be distributed according to the solution of the Boltzmann equation. This will be reviewed in Section II-4.2. As in the proof of Lanford's theorem (Section II-4.6), a fundamental idea is to reduce the problem to the estimation of the number of recollisions in a sampled graph or to the number of graphs in which two given particles are linked by a route of arbitrary size. Indeed, in the probabilistic setting, if a random binary *tree* with branching rate Λ is sampled first and then a particle system is constructed as above but starting from independent particles, then this gives a trajectorial representation of a process whose law is the solution of the Boltzmann equation (29).

Acknowledgments

The authors wish to thank Pierre Degond for his precious advice and careful proofreading of this manuscript. The authors also thank Paul Thevenin for fruitful comments and discussions.

A Probability reminders

For the convenience reader, classical elements of stochastic analysis and probability theory which are used throughout this review are gathered in this section. The Sections A.1, A.2 and A.3 are devoted to the classical construction of stochastic processes on the Skorokhod space. Notions related to the theory of Markov processes and their links with linear and nonlinear PDEs can be found in Section A.4. Section A.5, Section A.6 and Section A.7 summarize probability results regarding respectively large deviations, the Girsanov theorem and Poisson random measures.

A.1 Convergence of probability measures

The two following classical theorems complete the results of Section 3.1.1 and Section 3.1.2 to study the limit of sequences of probability measures.

The first theorem links weak convergence and almost sure convergence of random variables.

Theorem A.1 (Skorokhod's representation theorem). *Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of probability measures on a Polish space E which converges weakly towards $f \in \mathcal{P}(E)$ as $n \rightarrow +\infty$. Then there exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and some E -valued random*

variables X, X_n defined on this space for all $n \in \mathbb{N}$, such that

$$\text{Law}(X_n) = f_n, \quad \text{Law}(X) = f, \quad X_n(\omega) \xrightarrow[n \rightarrow +\infty]{} X(\omega), \quad \mathbb{P}\text{-a.s.}$$

The second theorem is the widely used Prokhorov's theorem which gives a helpful characterization of compactness for the weak convergence topology. The following results can be found for instance in [16, Section 5]. Compactness is linked to the notion of tightness defined below.

Definition A.2 (Tightness). A family $(f_i)_{i \in I}$ of probability measures on a separable metric space E (endowed with its Borel σ -field) is said to be *tight* when for every $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset E$ such that

$$\forall i \in I, \quad f_i(K_\varepsilon) > 1 - \varepsilon.$$

A sequence of random variables is said to be tight when the sequence of their laws is tight.

Theorem A.3 (Prokhorov's Theorem). *A tight sequence $(f_n)_{n \in \mathbb{N}}$ of probability measures on E is weakly relatively compact. Conversely, if E is also complete, any weakly relatively compact family $(f_n)_{n \in \mathbb{N}}$ is tight.*

A.2 Skorokhod's topology and tightness on the Skorokhod space

A stochastic process is a random function from a time interval to a state space (E, ρ) assumed to be Polish. Throughout this article, the stochastic process are assumed to belong (at least) to the Skorokhod space of càdlàg functions.

Definition A.4 (càdlàg). Let T in $(0, +\infty]$. A function $x : [0, T] \rightarrow E$ is said to belongs to the Skorokhod space $D([0, T], E)$ of càdlàg functions when x is right-continuous and has a left-limit at any time $t \in [0, T]$:

$$x(t^-) := \lim_{\substack{s \rightarrow t \\ s < t}} x(s) \text{ exists, } \quad x(t) = x(t^+).$$

We recall that a càdlàg function admits an at most countable number of discontinuities.

The law of a stochastic process is therefore an element of $\mathcal{P}(D([0, T], E))$. In order to characterize the compact sets of this space, it is first necessary to precise the topology on $D([0, T], E)$. For a much more detailed study of the Skorokhod space, we refer to [16, Section 12].

Definition A.5 (Skorokhod $J1$ topology). Let Λ denote the set of strictly increasing homeomorphisms from $[0, T]$ onto itself. The Skorokhod $J1$ metrics on $D([0, T], E)$ is defined by

$$d(x, y) := \inf_{\lambda \in \Lambda} \left\{ \sup_{0 \leq t \leq T} \rho(x(t), y(\lambda(t))) + \sup_{s < t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right| \right\}.$$

Endowed with this metric, the Skorokhod space $D([0, T], E)$ is complete and separable.

This topology is weaker than the one of continuous functions, which does not make $D([0, T], E)$ a complete space. However, estimates in the strong $\|\cdot\|_\infty$ metrics implies estimates a complete metric, up to a fixed multiplicative constant. In practise, working with the $\|\cdot\|_\infty$ metrics is thus often sufficient.

Let us recall that for a continuous function $x \in C([0, T], E)$, the continuity modulus is defined for $0 < \delta < T$ by

$$w_x(\delta) := \sup_{0 \leq t \leq T-\delta} w_x[t, t+\delta], \quad w_x I = \sup_{s, t \in I} \rho(x(t), x(s)).$$

A function x belongs to $C([0, T], E)$ if and only if $\lim_{\delta \rightarrow 0} w_x(\delta) = 0$. For càdlàg functions, another notion of modulus is defined.

Definition A.6 (càdlàg modulus). The càdlàg-modulus on $D([0, T], E)$ is defined by

$$w'_x(\delta) := \inf_{\{t_i\}} \max_i w_x[t_{i-1}, t_i],$$

where the infimum is taken over sub-divisions $\{t_i\}$ of $[0, T]$ such that $\min_i t_{i+1} - t_i > \delta$. A function x belongs to $D([0, T], E)$ if and only if $\lim_{\delta \rightarrow 0} w'_x(\delta) = 0$.

An analog of the Ascoli-Arzelà theorem in the space of càdlàg functions is given by [16, Theorem 12.3]. It states that a subset $A \subset D([0, T], E)$ is relatively compact if and only if

- (1) $\sup_{x \in A} \|x\|_\infty < \infty$;
- (2) $\lim_{\delta \rightarrow 0} \sup_{x \in A} w'_x(\delta) = 0$.

In some cases, it is easier to use the modulus

$$w''_x(\delta) := \sup_{\substack{t_1 \leq t \leq t_2 \\ t_2 - t_1 \leq \delta}} |x(t_1) - x(t)| \wedge |x(t) - x(t_2)|,$$

see [16, Theorem 12.4]. Using these results the following tightness criterion for probability measures on $D([0, T], E)$ is proved in [75, Chapter 3, Corollary 7.4].

Theorem A.7 (Basic tightness criterion in D). *For each $n \in \mathbb{N}$, let $(X_t^n)_t$ be an adapted E -valued càdlàg process on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$. The sequence $(X_t^n)_t$ is tight if and only if the following two conditions hold.*

- (1) *For every $\varepsilon > 0$ and every rational number $t \geq 0$, there exists a compact set $K_{\varepsilon, t} \subset E$ such that*

$$\liminf_{n \rightarrow +\infty} \mathbb{P}(X_t^n \in K_{\varepsilon, t}) \geq 1 - \varepsilon.$$

- (2) *For every $\varepsilon > 0$ and $T > 0$, there exists $\delta > 0$ such that*

$$\limsup_{n \rightarrow +\infty} \mathbb{P}(w'_{X^n|_{[0, T]}}(\delta) \geq \varepsilon) \leq \varepsilon.$$

Still, this criterion remains abstract and requires to find a suitable partition of the time interval to evaluate the càdlàg modulus. The tightness criteria which are used to prove propagation of chaos will be detailed in Appendix II-C.

A.3 Stochastic processes and martingales

In the following, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. This section reminds the reader of the basic properties related to stochastic processes and martingales. An exhaustive rigorous study can be found in the classical books [75, Chapter 2], [100], [141] or [111]. The present section also summarizes elements of [101].

A.3.1 Martingales

Definition A.8 (Stochastic process). A *stochastic process* with state space E (a measurable space) and indexed by a set I (often $I = [0, T]$) is a function $X : I \times \Omega \rightarrow E$ such that $X_t = X(t, \cdot)$ is a E -valued random variable (that is is to say $\omega \mapsto X_t(\omega)$ is measurable) for every $t \in I$.

- (1) The (pathwise) *law* of $(X_t)_t$ is the push-forward measure $f_I = X_{\#}\mathbb{P}$ on a space $\mathcal{F}(I, E)$ of functions $I \rightarrow E$.
- (2) The *time marginal laws* are the push-forward measures on E defined for any $t \in I$ by $f_t = X_t^E_{\#}f_I$, provided that the evaluation maps $X_t^E : \mathcal{F}(I, E) \rightarrow E, \omega \mapsto \omega(t)$ are measurable.

The stochastic process $X = (X_t)_{t \in I}$ can be seen as a $\mathcal{F}(I, E)$ -valued random variable. For a given $\omega \in \Omega$, $(X_t(\omega))_{t \geq 0}$ is called a *sample path* (or trajectory) of X .

From now on let $I = [0, T]$ with $T \in (0, +\infty]$.

Example 23 (Canonical stochastic process). Given a probability distribution on the space of functions $I \rightarrow E$, $f_I \in \mathcal{P}(\mathcal{F}(I, E))$, the *canonical stochastic process with law f_I* is $X = (X_t^E)_{t \in I}$ defined on the probability space $(\mathcal{F}(I, E), \mathcal{F}, f_I)$: the law of X is indeed $X_{\#}f_I = f_I$.

Definition A.9 (Filtration). A filtration is an increasing family of σ -algebras $(\mathcal{F}_t)_{t \geq 0}$, i.e. such that $\mathcal{F}_s \subset \mathcal{F}_t$ for $s \leq t$. A filtration is said to be:

- (1) *complete* when

$$\forall t \geq 0, \quad \{\mathcal{A} \subset \Omega, \mathcal{A} \subset \mathcal{B}, \mathbb{P}(\mathcal{B}) = 0\} \subset \mathcal{F}_t;$$

- (2) *right-continuous* when

$$\forall t \geq 0, \quad \mathcal{F}_t = \mathcal{F}_{t+}, \quad \mathcal{F}_{t+} := \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}.$$

In the following, let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ be a filtered probability space, whose filtration is assumed to be complete and right-continuous, together with $\mathcal{F} = \bigcup_{t \geq 0} \mathcal{F}_t$.

Definition A.10 (Regularity for stochastic processes). A stochastic process $X = (X_t)_{t \geq 0}$ is said to be

- (1) *adapted* to the filtration $(\mathcal{F}_t)_{t \geq 0}$ when the random variable X_t is $\mathcal{F}_t$ -measurable;

- (2) *predictable* when X is measurable for the σ -algebra generated by the sets $(s, t] \times F$ for $0 \leq s \leq t$ and $F \in \mathcal{F}_s$;
- (3) *with finite variation paths* when a.s. $(X_t(\omega))_{0 \leq t \leq T}$ has bounded variation on any finite time-interval $[0, T]$;
- (4) *continuous* (resp. càdlàg, right-continuous...) when for $\mathbb{P}$ -almost every $\omega \in \Omega$, the sample path $(X_t(\omega))_{t \geq 0}$ is a continuous (resp. càdlàg, right-continuous...) function of t .

Definition A.11 (Martingale and sub-martingale). An adapted real-valued process $X = (X_t)_{t \geq 0}$ such that $\mathbb{E}|X_t| < +\infty$ for every $t \geq 0$ is a $(\mathcal{F}_t)_t$ -martingale when

$$\forall t, s \geq 0, \quad \mathbb{E}[X_{t+s} | \mathcal{F}_t] = X_t.$$

It is a *sub-martingale* when

$$\forall t, s \geq 0, \quad \mathbb{E}[X_{t+s} | \mathcal{F}_t] \geq X_t.$$

These definitions are extended componentwise to $\mathbb{R}^N$ -valued processes.

Sub-martingales enjoy many useful properties, starting with the following proposition [75, Chapter 2, Proposition 2.9].

Proposition 14 (Càdlàg modification). *If $X = (X_t)_{t \geq 0}$ is a sub-martingale, then there exists an adapted real-valued process $Y = (Y_t)_{t \geq 0}$ such that*

$$\forall t \geq 0, \quad P(X_t = Y_t) = 1, \tag{115}$$

and Y is càdlàg outside a countable set of times t . In the following this set is assumed to be $\emptyset$, so (sub-)martingales will be considered as càdlàg processes.

Equation 115 means that Y is a *modification* of X . The following inequality controls the growth of sub-martingales.

Proposition 15 (Doob's maximal inequality). *Given a sub-martingale $(X_t)_{t \geq 0}$, $T > 0$ and any $p > 1$, it holds that*

$$\mathbb{E} \left| \sup_{0 \leq t \leq T} X_t \right|^p \leq \left(\frac{p}{p-1} \right)^p \mathbb{E}|X_T|^p.$$

A.3.2 Local martingales and quadratic variation

In order to define a stochastic integral, the notion of martingale needs to be weakened to the notion of local martingale.

Definition A.12 (Local martingale). A real-valued process $(M_t)_{t \geq 0}$ is a *local martingale* when there exists a sequence $(\tau_n)_{n \in \mathbb{N}}$ of stopping times such that $\tau_n \rightarrow +\infty$ and $(M_{t \wedge \tau_n})_{t \geq 0}$ is a martingale for every $n \geq 0$. This definition is extended componentwise to $\mathbb{R}^N$ -valued processes.

Definition A.13 (Quadratic variation). The quadratic variation $[M] = ([M]_t)_{t \geq 0}$ of a (local) square integrable martingale $(M_t)_{t \geq 0}$ is defined as the limit in probability

$$[M]_t := \lim_{\delta(\pi) \rightarrow 0} \sum_{t_n \in \pi} (M_{t_{n+1} \wedge t} - M_{t_n \wedge t})^2,$$

where π is a subdivision $0 = t_0 < t_1 < \dots$ of $\mathbb{R}_+$ with mesh $\delta(\pi) = \sup_n (t_{n+1} - t_n)$. It is the only (up to modification) adapted increasing process with jumps $\Delta[M]_t = \Delta M_t^2 = M_t^2 - M_{t-}^2$ such that $(M_t^2 - [M]_t)$ is a (local) martingale.

Example 24 (The case of finite variation paths). When $(M_t)_{t \geq 0}$ has finite variation paths, it is straightforward to check that

$$[M]_t = \sum_{s \leq t} (\Delta M_s)^2.$$

Definition A.14 (Cross-variation). The cross-variation $[M, N] = ([M, N]_t)_{t \geq 0}$ of two (local) square integrable martingales $(M_t)_{t \geq 0}$ and $(N_t)_{t \geq 0}$ is defined by

$$[M, N]_t := \frac{1}{2}([M + N]_t - [M]_t - [N]_t).$$

It is the only (up to modification) adapted increasing process with jump at time $t > 0$, $\Delta[M, N]_t = \Delta(MN)_t = M_t N_t - M_{t-} N_{t-}$ such that $(M_t N_t - [M, N]_t)$ is a (local) martingale.

Proposition 16 (BDG inequality). *For every $p \geq 1$, there exist two constants $c_p, C_p > 0$ such that for any local martingale $(M_t)_{t \geq 0}$, the supremum $M_t^* := \sup_{0 \leq s \leq t} |M_s|$ can be controlled by the quadratic variation in L^p -norm:*

$$c_p \mathbb{E}[M]_t^{p/2} \leq \mathbb{E}(M_t^*)^p \leq C_p \mathbb{E}[M]_t^{p/2}.$$

Definition A.15 (Predictable quadratic and cross-variation). The predictable quadratic variation $\langle M \rangle = (\langle M \rangle_t)_{t \geq 0}$ of a (local) square integrable martingale $(M_t)_{t \geq 0}$ is the unique *predictable* càdlàg increasing process (with finite variation paths) such that

$$(M_t^2 - \langle M \rangle_t)_{t \geq 0}$$

is a (local) martingale. The existence of the predictable quadratic variation stems from the Doob-Meyer decomposition theorem for DL -supermartingales (see [75, Chapter 2, Proposition 5] for more details). The predictable cross-variation of two (local) square integrable martingales M, N is defined the same way, setting

$$\langle M, N \rangle_t := \frac{1}{2}(\langle M + N \rangle_t - \langle M \rangle_t - \langle N \rangle_t).$$

Note that by subtracting the martingale characterizations of the quadratic variation and of the predictable quadratic variation, we get that $([M]_t - \langle M \rangle_t)_{t \geq 0}$ is a (local) square integrable martingale. The predictable quadratic variation is said to be the *compensator* of the quadratic variation. Moreover, the equality $[M] = \langle M \rangle$ holds as soon as $(M_t)_{t \geq 0}$ is a continuous process.

A.3.3 Semimartingales

The notion of local martingales is then extended to the notion of semimartingale, which forms a large class of processes against which a stochastic integral can be defined (a notion that will not be detailed here).

Definition A.16 (Semimartingale). A real-valued process $(X_t)_{t \geq 0}$ is a semimartingale when it can be decomposed as

$$X_t = M_t + A_t,$$

where $(M_t)_{t \geq 0}$ is a local martingale and $(A_t)_{t \geq 0}$ is an adapted process with finite variation paths. This definition is extended component-wise to $\mathbb{R}^N$ -valued processes.

The cross-variation $[X, Y]$ and predictable cross-variation $\langle X, Y \rangle$ can be extended to semimartingales $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$ as the limit in probability

$$[X, Y]_t = \lim_{\delta(\pi) \rightarrow 0} \sum_{t_n \in \pi} (X_{t_{n+1} \wedge t} - X_{t_n \wedge t})(Y_{t_{n+1} \wedge t} - Y_{t_n \wedge t}).$$

Writing $X_t = M_t + A_t$, note that $[X] = [M]$ and $\langle X \rangle = \langle M \rangle$ because A has finite variation paths.

For vector-valued semimartingales, the cross-variation is defined componentwise as follows.

Definition A.17 (Vectorial cross-variation). Let $(\mathbf{X}_t)_{t \geq 0}$ be a $\mathbb{R}^d$ -valued semimartingale with the notation $\mathbf{X}_t = (X_t^1, \dots, X_t^d)$, the matrix-valued cross-variations are defined by

$$[\mathbf{X}] = ([X^i, X^j])_{1 \leq i, j \leq d}, \quad \langle \mathbf{X} \rangle = (\langle X^i, X^j \rangle)_{1 \leq i, j \leq d},$$

and the related scalar quantities are the traces of these matrices (defined as the sum of diagonal elements)

$$[\mathbf{X}] = \text{Tr}([\mathbf{X}]), \quad \langle \mathbf{X} \rangle = \text{Tr}(\langle \mathbf{X} \rangle).$$

The integration of locally bounded predictable processes $(H_t)_{t \geq 0}$ against finite variation processes is well-defined using Stieltjes formalism. The theory of stochastic integration extends this to integrate $(H_t)_{t \geq 0}$ against any square integrable semimartingale $(X_t)_{t \geq 0}$. Notable identities are the Itô isometry

$$\mathbb{E} \left[\left(\int_0^t H_{s-} dX_s \right)^2 \right] = \mathbb{E} \left[\int_0^t H_{s-}^2 d[X]_s \right] = \mathbb{E} \left[\int_0^t H_{s-}^2 d\langle X \rangle_s \right],$$

(the last equality holds because $[X] - \langle X \rangle$ is a local martingale) and the integration by parts formula

$$X_t Y_t = X_0 Y_0 + \int_0^t X_{s-} dY_s + \int_0^t Y_{s-} dX_s + [X, Y]_t,$$

for general semimartingales $(X_t)_{t \geq 0}$ and $(Y_t)_{t \geq 0}$. Given a (local) square integrable martingale $(M_t)_{t \geq 0}$, the process

$$\left(\int_0^t H_{s-} dM_s \right)_{t \geq 0}$$

is built in such a way that it is a (local) square integrable martingale.

Example 25 (Dynkin's representation formula). Anticipating on the next section, a Markov process $(X_t)_{t \geq 0}$ with generator L is such that for any φ in $\mathcal{D}(L)$ (see Theorem A.29)

$$M_t^\varphi := \varphi(X_t) - \varphi(X_0) - \int_0^t L\varphi(X_s)ds,$$

is a martingale. In this case, the predictable quadratic variation $\langle M^\varphi \rangle$ can be computed as

$$\langle M^\varphi \rangle_t = \int_0^t \Gamma_L(\varphi, \varphi)(X_s)ds,$$

involving the *carré du champ operator*

$$\Gamma_L(\varphi, \psi) := L[\varphi\psi] - \varphi L\psi - \psi L\varphi,$$

for any $\varphi, \psi \in \mathcal{D}(L)$. These properties can be generalized to obtain a wider class of semimartingales, which are non-necessarily Markov processes.

A.3.4 D -semimartingales

Dynkin's formula can be turned into a definition to obtain a wider class of (non-necessarily Markov) $\mathbb{R}^d$ -valued semimartingales, named class D (in the honour of Dynkin). The following definition can be found in [101]. The notation π_i denotes the coordinate function $\mathbf{x}^d \mapsto x^i$.

Definition A.18 (Semimartingale of class D). A $\mathbb{R}^d$ -valued semimartingale $\mathbf{X}_t$ belongs to the class D when there exist an increasing càdlàg function $t \mapsto A(t)$, a vector space $\mathcal{C}$ of $\mathbb{R}$ -valued continuous functions on $\mathbb{R}^d$ and a mapping $L : \mathcal{C} \times \mathbb{R}^d \times \mathbb{R}_+ \times \Omega \rightarrow \mathbb{R}$ such that the following properties hold.

- (1) For every $1 \leq i, j \leq d$, π_i and $\pi_i\pi_j$ belong to $\mathcal{C}$.
- (2) For every $(\mathbf{x}^d, t, \omega) \in \mathbb{R}^d \times \mathbb{R}_+ \times \Omega$, the map $\varphi \mapsto L(\varphi, \mathbf{x}^d, t, \omega)$ is linear and maps $\mathcal{C}$ to itself. Moreover for every $\varphi \in \mathcal{C}$, the map $(\mathbf{x}^d, t, \omega) \mapsto L(\varphi, \mathbf{x}^d, t, \omega)$ is measurable for the σ -algebra on $\mathbb{R}^d$ of Borel sets, and the one of predictable events on Ω .
- (3) For every $\varphi \in \mathcal{C}$,

$$M_t^\varphi := \varphi(\mathbf{X}_t) - \varphi(\mathbf{X}_0) - \int_0^t L(\varphi, \mathbf{X}_{s-}, s, \cdot) dA(s),$$

is a local square integrable martingale.

To this process are associated the *local coefficients*

$$b_i(\mathbf{x}^d, t, \omega) = L(\pi_i, \mathbf{x}^d, t, \omega), \quad a_{ij}(\mathbf{x}^d, t, \omega) = \Gamma_{L(\cdot, \mathbf{x}^d, t, \omega)}(\pi_i, \pi_j)(\mathbf{x}^d),$$

and the *drift vector* $\mathbf{b} = (b_i)_{1 \leq i \leq d}$ and the *diffusion matrix* $\mathbf{a} = (a_{ij})_{1 \leq i, j \leq d}$. In the following we often omit to write the dependency in ω .

The following lemma is proved in [101, Lemma 3.1.3].

Lemma A.19 (Predictable variation). *Given a D -semimartingale $\mathbf{X}$, define*

$$\mathbf{M}_t := \mathbf{X}_t - \mathbf{X}_0 - \int_0^t \mathbf{b}(\mathbf{X}_{s-}, s) dA(s).$$

Then $(\mathbf{M}_t)_{t \geq 0}$ is a $\mathbb{R}^d$ -valued local square integrable martingale, whose (scalar) predictable quadratic variation reads

$$\langle \mathbf{X} \rangle_t = \int_0^t \text{Tr}(\mathbf{a}(\mathbf{X}_{s-}, s)) dA(s) - \sum_{s \leq t} \|\mathbf{b}(\mathbf{X}_{s-}, s)\|^2 |\Delta A(s)|^2.$$

Remark 28 (Generalized SDE and intrinsic randomness). The previous notions naturally extend the notion of diffusion SDE, since at least formally

$$d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_{t-}, t) dA(t) + \mathbf{a}(\mathbf{X}_{t-}, t) d\mathbf{M}_t.$$

This parallel could be used to extend coupling and completeness methods to irregular processes or even non-Markov ones. However, remember that in this case, the generator L depends on ω and has therefore *its own source of randomness*, together with the local coefficients $\mathbf{b}$ and $\mathbf{a}$.

Similarly to Markov processes, a convenient way to build a D -semimartingale on $\Omega = D([0, T], \mathbb{R}^d)$ (for $T \in (0, +\infty]$) is to solve a martingale problem.

Definition A.20 (Martingale problem). A probability distribution on the path space $f_I \in \mathcal{P}(D([0, T], \mathbb{R}^d))$ is a solution to the martingale problem issued from $f_0 \in \mathcal{P}(\mathbb{R}^d)$ whenever for all $\varphi \in \mathcal{C}$,

$$M_t^\varphi := \varphi(\mathbf{X}_t) - \varphi(\mathbf{X}_0) - \int_0^t L(\varphi, \mathbf{X}_{s-}, s, \cdot) dA(s),$$

is a f_I -martingale, where $\mathbf{X}_t$ is the canonical process $D([0, T], \mathbb{R}^d) \rightarrow \mathbb{R}^d$.

A.4 Markov processes and Markov representation for PDEs

The purpose of this section is to briefly review the probabilistic framework for linear and nonlinear Markov processes. Classical references and review articles on the subject include [75, 141, 27, 25]. The prototypical nonlinear Markov process is the solution of the McKean-Vlasov SDE:

$$dX_t = b(X_t, f_t) dt + \sigma(X_t, f_t) dB_t, \quad X_t \sim f_t,$$

where $(B_t)_{t \geq 0}$ is a standard Brownian motion and f_t the law of the random variable X_t at time t . Such a process is said to be nonlinear (in the sense of McKean) since its definition depends on its own law. This type of nonlinearity has been introduced in the seminal

[125]. Unlike, classical (linear) Markov process, the law f_t of the nonlinear Markov process satisfies the nonlinear the Fokker-Planck equation

$$\partial_t f_t(x) = -\nabla_x \cdot \{b(x, f_t)f_t\} + \frac{1}{2} \sum_{i,j=1}^d \partial_{x_i} \partial_{x_j} \{a_{ij}(x, f_t)f_t\},$$

where the matrix $a = (a_{ij})$ is defined by $a = \sigma\sigma^T$. This kind of equations is derived in the limit $N \rightarrow +\infty$ in the Kolmogorov equations associated to large interacting particle systems defined by N linear Markov processes. While the theory of linear Markov processes is well-established, nonlinear Markov processes are not so classical in the literature and require specific tools to be built (for instance, the well-posedness result proved in Proposition 1 for the McKean-Vlasov SDE).

The classical theory of time homogeneous linear Markov processes is presented in Section A.4.1. Elements of the theory of time-inhomogeneous Markov processes can be found in Section A.4.2. Using these concepts and following [102] and [123], a theoretical framework is presented in Section A.4.3 for nonlinear Markov processes (in the sense of McKean). Before that, we first recall the basic notions about Markov process (linear or not). Following [75, Chapter 4], the very general definition of Markov processes is the following.

Definition A.21 (Markov process). A stochastic process $(X_t)_{t \geq 0}$ is a Markov process on $(\Omega, \mathcal{F}, \mathbb{P})$ when for every $s, t \geq 0$

$$\forall \mathcal{A} \in \mathcal{B}(E), \quad \mathbb{P}(X_{s+t} \in \mathcal{A} | (X_r)_{0 \leq r \leq t}) = \mathbb{P}(X_{s+t} \in \mathcal{A} | X_t). \quad (116)$$

On the left-hand side the probability is conditioned by the filtration generated by the process, but stronger definitions could involve wider filtrations. The relation (116) will be referred as the *Markov property*. When the time t is replaced by a random *stopping time*, this relation is called the *strong Markov property*.

This means the law of X_{s+t} at time $s+t$ conditionally on the past history up to time t , is the same as the law at time $s+t$ conditionally on the state at time t only. The definition can be equivalently written in terms of bounded measurable test functions $\varphi \in B_b(E)$ as

$$\mathbb{E}[\varphi(X_{s+t}) | (X_r)_{0 \leq r \leq t}] = \mathbb{E}[\varphi(X_{t+s}) | X_t]. \quad (117)$$

This definition remains abstract and does not tell how to *build* a Markov process. If the law f_I is known (built from given processes or e.g. by solving a martingale process), a Markov process with law f_I is given by the canonical process $\mathbf{X}^E = (\mathbf{X}_t^E)_{t \geq 0}$ on the probability space $\Omega = D(I, E)$ endowed with f_I and an adequate filtration (see Example 23). Further constructions which are closer to the PDE point of view are presented in the next subsections.

A.4.1 Time-homogeneous Markov processes and linear PDEs

The content of this section is quite classical and can be found in the classical reference [75].

Transition functions.

A usual way to present Markov processes is to first think about time discrete Markov chains with jumps which are given by a transition kernel (or transition matrix when E is discrete). This definition can be generalized to the time continuous framework using the notion of transition function.

Definition A.22 (Transition functions, homogeneous case). A family of maps $(P_t)_{t \geq 0}$ where $P_t : E \rightarrow \mathcal{P}(E)$ is a family of transition functions when the following properties hold.

- The map $(t, x) \mapsto P_t(x, \cdot)$ is a measurable map $[0, \infty) \times E \rightarrow \mathcal{P}(E)$.
- For all $x \in E$, $P_0(x, \cdot) = \delta_x$.
- For all $s, t \geq 0$ and all $\mathcal{A} \subset \mathcal{B}(E)$, $P_{s+t}(x, \mathcal{A}) = \int_E P_s(y, \mathcal{A}) P_t(x, dy)$.

This last relation is called the *Chapman-Kolmogorov property*.

A family of transition functions $(P_t)_{t \geq 0}$ is said to be *adapted to the Markov process* $(X_t)_{t \geq 0}$ when for all $s, t \geq 0$, and all $\mathcal{A} \subset \mathcal{B}(E)$,

$$\mathbb{P}(X_{s+t} \in \mathcal{A} | (X_r)_{0 \leq r \leq t}) = P_s(X_t, \mathcal{A}),$$

which is equivalent to

$$\mathbb{E}[\varphi(X_{t+s}) | (X_r)_{0 \leq r \leq t}] = \int_E \varphi(y) P_s(X_t, dy),$$

for any bounded Borel measurable test function $\varphi \in B_p(E)$. Note that this relation and the Markov property (116) imply the Chapman-Kolmogorov property with $x = X_r$ for $r \geq 0$. Moreover given the initial distribution $X_0 \sim f_0$, the finite dimensional distributions of $(X_t)_t$ can be computed by

$$\begin{aligned} & \mathbb{P}(X_0 \in \mathcal{A}_0, X_{t_1} \in \mathcal{A}_1, \dots, X_{t_n} \in \mathcal{A}_n) \\ &= \int_{\mathcal{A}_0} \dots \int_{\mathcal{A}_{n-1}} P_{t_n - t_{n-1}}(y_{n-1}, \mathcal{A}_n) P_{t_{n-1} - t_{n-2}}(y_{n-2}, dy_{n-1}) \\ & \quad \dots P_{t_1}(y_0, dy_1) f_0(dy_0), \end{aligned} \tag{118}$$

for any $\mathcal{A}_0, \dots, \mathcal{A}_n \in \mathcal{B}(E)$ and $0 \leq t_1 \leq \dots \leq t_n$. In fact transition functions are sufficient to build a Markov process; this is a consequence of the Kolmogorov extension theorem (or more generally of the Ionescu-Tulcea theorem), see for instance [75, Chapter 4, Theorem 1.1] or [141, Chapter 3, Theorem 1.5].

Theorem A.23 (Markov process built from transition functions). *Let E be a Polish space and $f_0 \in \mathcal{P}(E)$. Given the transition functions $(P_t)_{t \geq 0}$, there exists a Markov process whose finite dimensional distributions are uniquely determined by (118). Its law on the path space is a probability measure $f_I \in \mathcal{P}(D(I, E))$.*

If $f_0 = \delta_x$, then f_I is denoted by f_I^x and [75, Chapter 4, Proposition 1.2] proves that the map $x \mapsto f_I^x(\mathcal{B})$ is measurable for any Borel set $\mathcal{B} \subset D(I, E)$.

Example 26 (Brownian motion). Transition functions are not explicit in general. A notable exception is the d -dimensional Brownian motion for which

$$P_t(x, \mathcal{A}) = (2\pi t)^{-d/2} \int_{\mathcal{A}} \exp\left(-\frac{|x-y|^2}{2t}\right) dy.$$

Note that the map $t \mapsto P_t(x, dy)$ is the measure solution of the 1D heat equation $\partial_t u = \partial_{xx}^2 u$ with initial condition $u(t=0, \cdot) = \delta_x$.

Semigroup representation.

The connection between Markov processes and linear PDEs is given by the semigroup representation as explained below. For $t \geq 0$, let the linear operator T_t acting on $\mathcal{D}$ be defined by:

$$T_t \varphi(x) := \int_E \varphi(y) P_t(x, dy), \quad (119)$$

for any test functions $\varphi \in B_b(E)$. Thanks to the Chapman-Kolmogorov property, this defines a positive measurable semigroup of contractions on $B_b(E)$, where we recall that a family of bounded operators $(T_t)_{t \geq 0}$ on a closed subspace $\mathcal{D} \subset B_b(E)$ is a semigroup when $T_0 = \text{Id}$ and $T_{t+s} = T_t T_s$ for all $s, t \geq 0$. It is said to be a contraction semigroup when the operators are bounded with norm smaller or equal to 1. The semi-group $(T_t)_t$ is said to correspond to a Markov process $(X_t)_{t \geq 0}$ when

$$\forall s, t \geq 0, \forall \varphi \in \mathcal{D}, \quad T_t \varphi(X_{t+s}) = \mathbb{E}[\varphi(X_{s+t}) | (X_r)_{0 \leq r \leq t}]$$

The semi-group representation characterises a Markov process, as stated in [75, Chapter 4, Proposition 1.6].

Theorem A.24. *Let E be a Polish space and let $\mathcal{D} \subset B_b(E)$ be a closed subspace assumed to be separating. Let f_0 in $\mathcal{P}(E)$ and let $(T_t)_{t \geq 0}$ be a semigroup on $\mathcal{D}$ corresponding to a Markov process $(X_t)_t$. Then the finite dimensional distributions of $(X_t)_t$ are determined by $(T_t)_{t \geq 0}$ and f_0 .*

In the following, starting from a semigroup $(T_t)_t$, the goal is to construct a corresponding Markov process. With the notable exception of jump processes, stronger assumptions on E and $(T_t)_t$ are often needed, as the ones given in the following definition.

Definition A.25 (Feller semigroup). Let E be a locally compact Polish space. A semi-group $(T_t)_{t \geq 0}$ is a Feller semigroup when its elements satisfy the following properties.

- (1) **(Feller).** For all $t \geq 0$, T_t maps $C_0(E)$ to $C_0(E)$, where $C_0(E)$ is the space of continuous functions vanishing at infinity. It means that

$$\forall \varphi \in C_0(E), \quad T_t \varphi \in C_0(E).$$

- (2) **(Contraction)**. For all $t \geq 0$ and all $\varphi \in C_0(E)$, $\|T_t \varphi\|_\infty \leq \|\varphi\|_\infty$.
- (3) **(Mass preserving)**. For all $t \geq 0$, $T_t 1 = 1$.
- (4) **(Positivity)**. For all $t \geq 0$ and all $\varphi \in C_0(E)$ such that $\varphi \geq 0$, then $T_t \varphi \geq 0$.
- (5) **(Strongly continuous)**. For all $\varphi \in C_0(E)$, $\|T_t \varphi - \varphi\|_\infty \rightarrow 0$ when $t \rightarrow 0$.

The properties (2) to (4) hold when T_t is defined from (119). Also note that every semigroup defined on $C_0(E)$ can be uniquely extended to the whole space $B_b(E)$ (see [27, Theorem 1.5]). A Markov process corresponding to a Feller semigroup is called a *Feller process*. There is a one-to-one correspondence between Feller semigroups and Feller processes; this is a consequence of the Riesz representation theorem, see for instance [27, Section 1.2], [141, Chapter 3, Proposition 2.2] or [75, Chapter 4, Theorem 2.7].

Theorem A.26 (Càdlàg Markov process from a Feller semigroup). *Let E be a locally compact Polish space and let $f_0 \in \mathcal{P}(E)$. Let $(T_t)_t$ be a Feller semigroup on E . Then there exists a unique transition function $(P_t)_t$ on E such that (119) holds. As a consequence, there exists a Markov process corresponding to $(T_t)_t$ with initial distribution f_0 , whose finite dimensional distributions are uniquely determined by $(T_t)_t$. Moreover this process has a càdlàg modification and satisfies the strong Markov property with respect to the right-continuous filtration $\mathcal{G}_t = \cap_{\varepsilon > 0} \sigma((X_s)_{s \leq t+\varepsilon})$.*

Example 27 (Brownian motion). The d -dimensional Brownian motion is a Feller process. More generally diffusion processes are Feller processes under mild assumptions on the diffusion coefficients.

Example 28 (Markov jump processes and C_b -Feller processes). The definition of Feller process is not universal in the literature. An important variant is the notion of C_b -Feller process for which the space $C_0(E)$ is replaced by $C_b(E)$ in Definition A.25. In this case, the local compactness assumption on E can be dropped. Diffusion processes are *not* C_b -Feller processes because the diffusion semigroup is not strongly continuous on $C_b(E)$ (see [27, Example 1.7d]). The main class of C_b -Feller process that are considered in this review is the class of Markov jump processes, defined by the transition function:

$$\forall x \in E, \quad P_t(x, dy) = \sum_{k=0}^{+\infty} \frac{e^{-t} t^k}{k!} P^k(x, dy),$$

where $P : E \times \mathcal{B}(E) \rightarrow \mathcal{P}(E)$ is a transition probability and for $k \in \mathbb{N}$, $x \in E$ and $\mathcal{A} \in \mathcal{B}(E)$,

$$P^k(x, \mathcal{A}) = \int_E P(x, dy) P^{k-1}(y, \mathcal{A}).$$

An explicit construction of a Markov jump process can be found in [75, Chapter 4, Section 2], see also [141, Chapter 3, Exercise 1.8]. Markov jump processes are more easily understood through their generator as defined below. More generally, the strong continuity property holds on $C_b(E)$ for semigroups which have a bounded generator (for the $\|\cdot\|_\infty$ topology). Further links between Feller semigroups, C_b -Feller semigroups and other notions of Feller semigroups can be found in [27, Section 1.1] and the references therein.

Infinitesimal generator and PDE.

Strongly continuous contraction semigroups are determined by their infinitesimal generator.

Definition A.27 (Infinitesimal generator). The infinitesimal generator L of a strongly continuous contraction semigroup $(T_t)_{t \geq 0}$ on a closed subspace $\mathcal{D} \subset B_p(E)$ is the linear operator defined by

$$L\varphi = \lim_{\substack{t \rightarrow 0 \\ t > 0}} \frac{T_t\varphi - \varphi}{t}$$

on the domain $\mathcal{D}(L) \subset \mathcal{D}$ of functions φ such that the limit exists (for the topology on $\mathcal{D}$).

The first connection between Markov processes and PDE comes from the observation that for all $\varphi \in \mathcal{D}(L)$,

$$\frac{d}{dt}T_t\varphi = T_tL\varphi, \quad \frac{d}{dt}T_t\varphi = LT_t\varphi, \quad (120)$$

which are called the *Kolmogorov equations*. The first equation is called the forward Kolmogorov equation and the second, the backward Kolmogorov equation. The terminology will appear more clearly in the time-inhomogeneous setting below. Since T_t and L commute, the two equations are of course equivalent but each one has its own physical interpretation. The backward equation gives a Markov representation of the solution of the linear PDE $\partial_t u = Lu$ with initial condition φ as the conditional expectation $u(t, x) = T_t\varphi(x) = \mathbb{E}[\varphi(X_t) | X_0 = x]$. The backward equation thus describes the evolution of an observable of the Markov process. A more general version when source terms or boundary conditions are added is the Feynman-Kac formula. More on the forward equation is given in the next paragraph. Note that the generator L can include differential and jump terms in which case the (backward) Kolmogorov equation is an integro-differential PDE.

Example 29. The generator of the d -dimensional Brownian motion is $L\varphi = \frac{1}{2}\Delta\varphi$ and $\mathcal{D}(L) \subset C_0^2(\mathbb{R}^d)$. The generator of a Markov jump process on a space E is

$$L\varphi(x) = \int_E \{\varphi(y) - \varphi(x)\}P(x, dy),$$

and L is bounded on $(C_b(E), \|\cdot\|_\infty)$.

In stochastic analysis, we often use the notion of *full generator* which turns the (forward) Kolmogorov equation into a definition.

Definition A.28 (Full generator). The full generator of a strongly continuous contraction semigroup $(T_t)_t$ on $\mathcal{D}$ is the subset

$$\widehat{L} := \left\{ (\varphi, \psi) \in \mathcal{D} \times \mathcal{D}, T_t\varphi - \varphi = \int_0^t T_s\psi \, ds \right\}.$$

The (forward) Kolmogorov equation says that $\{(\varphi, L\varphi), \varphi \in \mathcal{D}(L)\} \subset \widehat{L}$. The full generator is often used in connection with the martingale characterisation of a Markov process due to Stroock and Varadhan, see [27, Corollary 1.37] and [75, Chapter 4, Section 3].

Theorem A.29. *Let $(X_t)_t$ be a strong Markov process with full generator $\widehat{L}$. For $\varphi, \psi \in B_p(E)$, let us define for $t \geq 0$,*

$$M_t^{\varphi, \psi} := \varphi(X_t) - \varphi(X_0) - \int_0^t \psi(X_s) ds.$$

Then the full generator is characterised by

$$\widehat{L} = \left\{ (\varphi, \psi) \in B_p(E) \times B_p(E), (M_t^{\varphi, \psi})_t \text{ is a } \sigma(X)\text{-martingale} \right\}.$$

Taking $\psi = L\varphi$ shows that

$$M_t^\varphi = \varphi(X_t) - \varphi(X_0) - \int_0^t L\varphi(X_s) ds$$

is a martingale. The (forward) Kolmogorov equation is retrieved by simply taking the expectation.

Conversely, a linear operator (bounded or unbounded) L with domain $\mathcal{D} \subset B_p(E)$ is the generator of a strongly continuous contraction semigroup (and thus is the generator of a Markov process) if and only if it satisfies the Hille-Yosida theorem. In the context of Feller processes, the result is stated in [27, Theorem 3.1] or [75, Chapter 4, Theorem 2.2]. The following example is a classical and important application. Further examples using various points of view (SDE, martingale problem...) can be found in [27, Chapter 3].

Example 30. The second order differential operator on $\mathbb{R}^d$,

$$L\varphi(x) = c(x)\varphi(x) + \sum_{i=1}^d b_i(x) \frac{\partial \varphi}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^d a_{ij}(x) \frac{\partial^2 \varphi}{\partial x_i \partial x_j},$$

is the generator of a Feller semigroup when a_{ij} , b_i and c are respectively $C_b^3(\mathbb{R}^d)$, $C_b^2(\mathbb{R}^d)$ and $C_b^1(\mathbb{R}^d)$ with $c(x) \leq 0$ and the matrix $a = (a_{ij})$ is uniformly elliptic in the sense that there exists $\lambda > 0$ such that

$$\forall x, \xi \in \mathbb{R}^d, \quad \langle a(x)\xi, \xi \rangle \geq \lambda |x|^2.$$

The corresponding Feller process is said to be a diffusion process.

The dual semi-group.

Let E be locally compact and let $(T_t)_t$ be a Feller semigroup. To obtain a forward (or strong) version of the PDE representation (120), let the dual version of (119) be given by:

$$\forall \nu \in \mathcal{P}(E), \quad \varphi \in C_0(E) \mapsto \langle S_t \nu, \varphi \rangle := \langle \nu, T_t \varphi \rangle \in \mathbb{R}.$$

By the Riesz representation theorem, this defines a family of operator $(S_t)_t$ on $\mathcal{P}(E)$. For all $\nu \in \mathcal{P}(E)$, $S_t\nu \in \mathcal{P}(E)$ and by the Chapman-Kolmogorov property, the family of operators $(S_t)_t$ also forms a semigroup. When transition functions $(P_t)_{t \geq 0}$ are available, this duality relation implies

$$S_t\nu(dy) = \int_E P_t(x, dy)\nu(dx).$$

Given the generator L of the semigroup $(T_t)_t$, the dual semigroup $(S_t)_t$ satisfies the dual Kolmogorov equations:

$$\partial_t S_t = L^* S_t, \quad \partial_t S_t = S_t L^*,$$

where L^* is the dual operator of L . This time, the forward Kolmogorov equation (the first one) can be interpreted as the initial value problem $\partial_t f_t = L^* f_t$. According to (118), the solution of the forward equation $f_t = S_t f_0$ is the law at time $t \geq 0$ of the Feller process with initial distribution f_0 .

Example 31. The law f_t of a diffusion process (see Example 30) satisfies the forward Kolmogorov equation, also called the Fokker-Planck equation in this context:

$$\partial_t f_t(x) = c(x)f_t(x) - \sum_{i=1}^d \partial_{x_i} \{b_i(x)f_t(x)\} + \frac{1}{2} \sum_{i,j=1}^d \partial_{x_i} \partial_{x_j} \{a_{ij}(x)f_t(x)\}.$$

A.4.2 Time-inhomogeneous Markov processes

One of the goal of this subsection is to extend this formalism to cases where the generator L_t has a time-dependence. The analog of transition functions is the following time-inhomogeneous version.

Definition A.30 (Transition functions, inhomogeneous case). A family of maps $(P_{s,t})_{0 \leq s \leq t}$ where $P_{s,t} : E \rightarrow \mathcal{P}(E)$ is a family of time-inhomogeneous transition functions when the following properties hold.

- The map $(s, t, x) \mapsto P_{s,t}(x, \cdot)$ is a measurable map $[0, \infty) \times [0, \infty) \times E \rightarrow \mathcal{P}(E)$.
- For all $t \in \mathbb{R}_+$ and for all $x \in E$, $P_{t,t}(x, \cdot) = \delta_x$.
- For all $0 \leq r \leq s \leq t$, for all $x \in E$ and for all $\mathcal{A} \in \mathcal{B}(E)$,

$$P_{r,t}(x, \mathcal{A}) = \int_E P_{s,t}(y, \mathcal{A}) P_{r,s}(x, dy).$$

This last relation is the equivalent of the Chapman-Kolmogorov property.

Similarly to the time-homogeneous case, the transition functions $(P_{s,t})_{0 \leq s \leq t}$ are said to be adapted to the time-inhomogeneous Markov process $(X_t)_{t \geq 0}$ when

$$\mathbb{P}(X_t \in \mathcal{A} | (X_r)_{0 \leq r \leq s}) = P_{s,t}(X_s \in \mathcal{A}),$$

or equivalently,

$$\mathbb{E}[\varphi(X_t)|(X_r)_{0 \leq r \leq s}] = \int_E \varphi(y) P_{s,t}(X_s, dy),$$

for every $0 \leq s \leq t$ and any $\varphi \in B_p(E)$. The time-homogeneous setting is recovered when $P_{s,t} \equiv P_{t-s}$ depends on $t - s$ only. Moreover given the initial distribution $f_0 \in \mathcal{P}(E)$, the finite dimensional distributions can be computed as

$$\begin{aligned} & \mathbb{P}(X_0 \in \mathcal{A}_0, X_{t_1} \in \mathcal{A}_1, X_{t_2} \in \mathcal{A}_2, \dots, X_{t_n} \in \mathcal{A}_n) \\ &= \int_{\mathcal{A}_0} \dots \int_{\mathcal{A}_{n-1}} P_{t_{n-1}, t_n}(y_{n-1}, \mathcal{A}_n) P_{t_{n-2}, t_{n-1}}(y_{n-2}, dy_{n-1}) \\ & \dots P_{0, t_1}(y_0, dy_1) f_0(dy_0), \end{aligned}$$

so that given f_0 , a time inhomogeneous Markov process is fully characterised by the time-inhomogeneous transition functions. The link with PDEs is retrieved similarly by introducing the operators

$$\forall \varphi \in B_p(E), \quad T_{s,t}\varphi(x) := \int_E \varphi(y) P_{s,t}(x, dy).$$

The family $(T_{s,t})_{s \leq t}$ is an evolution system in the following sense, see [26], [144] and the references therein.

Definition A.31 (Evolution system). An evolution system $(T_{s,t})_{0 \leq s \leq t}$ is a family of bounded linear operators on a closed subspace $\mathcal{D} \subset B_p(E)$ such that

- for all $t \geq 0$, $T_{t,t} = \text{Id}$;
- for all $0 \leq s \leq t$, $T_{s,t}1 = 1$;
- for all $0 \leq r \leq s \leq t$, $T_{r,t} = T_{r,s}T_{s,t}$.

These properties can be directly checked in the case of a system defined by time inhomogeneous transition functions.

An evolution system is said to correspond to a Markov process $(X_t)_{t \geq 0}$ when

$$\forall 0 \leq s \leq t, \quad \forall \varphi \in \mathcal{D}, \quad T_{s,t}\varphi(X_s) = \mathbb{E}[\varphi(X_t)|(X_r)_{0 \leq r \leq s}].$$

The previous notion of Feller semigroup readily extends to evolution systems in a time-inhomogeneous setting with the strong continuity property being replaced by

$$\forall \varphi \in C_0(E), \quad \|T_{s,t}\varphi - \varphi\|_\infty \xrightarrow{(s,t) \rightarrow (0,0)} 0.$$

In the time inhomogeneous case, there are two notions of infinitesimal generators, depending on if the derivative is taken from the left or from the right. The left and right generators are defined respectively by

$$L_t^- \varphi := \lim_{\varepsilon \rightarrow 0^+} \frac{T_{t-\varepsilon, t}\varphi - \varphi}{\varepsilon}, \quad L_t^+ \varphi = \lim_{\varepsilon \rightarrow 0^+} \frac{T_{t, t+\varepsilon}\varphi - \varphi}{\varepsilon}.$$

They are respectively defined on the domains denoted by $\mathcal{D}(L_t^-)$ and $\mathcal{D}(L_t^+)$. Note that in both cases, the generators depend on a time variable. In general, the left and right generators do not coincide but they do under stronger uniform continuity assumptions with respect to the time variable, see [26, Lemma 2.2]. They also coincide for time homogeneous systems.

If $(T_{s,t})_{s \leq t}$ is strongly continuous then the forward and backward Kolmogorov equations reads, respectively:

$$\frac{d^\pm}{dt} T_{s,t} = T_{s,t} L_t^\pm, \quad \frac{d^\pm}{ds} T_{s,t} = -L_s^\pm T_{s,t},$$

where $\frac{d^+}{dt}$ (resp. $\frac{d^-}{dt}$) denotes the right (resp.) left derivative. Given $\varphi \in \mathcal{D}(L_s^-)$ and a fixed $t > 0$, the solution of the backward Kolmogorov equation is

$$u(s, x) = T_{s,t} \varphi(x) = \mathbb{E}[\varphi(X_t) | X_s = x].$$

The terminology *backward* refers to the fact that u satisfies a final value problem with terminal condition φ at $s = t$. In the time homogeneous setting,

$$T_{s,t} = T_{0,t-s} \equiv T_{t-s},$$

and consequently for $\tau \leq t$, the quantity

$$U(\tau, x) := u(t - \tau, x) = T_\tau \varphi(x),$$

satisfies

$$\partial_\tau U = LU, \quad U(\tau = 0) = \varphi,$$

which is the backward Kolmogorov equation previously obtained (since in this case $L_s^+ = L_s^- = L$ does not depend on s).

As before, the forward equation is better understood with the dual formulation. Given $s \leq t$, the operator $S_{s,t}$ acting on $\mathcal{P}(E)$ is defined by

$$\forall \nu \in \mathcal{P}(E), \forall \varphi \in B_p(E), \quad \langle S_{s,t} \nu, \varphi \rangle = \langle \nu, T_{s,t} \varphi \rangle.$$

This duality relation implies

$$S_{s,t} \nu(dy) = \int_E P_{s,t}(x, dy) \nu(dx)$$

Let $f_0(xdx)$ be the initial distribution. Then

$$f_t(dx) := S_{0,t} f_0(dx) = \int_E P_{0,t}(y, dx) f_0(dy),$$

is the law of the associated Markov process at time t . In this last equality, note the change of variables which exchanges the roles of the y and x variables. Since the family of operators $(S_{s,t})_{s \leq t}$ satisfies the dual Chapman-Kolmogorov property $S_{s,t} S_{r,s} = S_{r,t}$ for $r \leq s \leq t$, then $S_{s,t} f_s = f_t$ for all $0 \leq s \leq t$. Let us assume that the left and right generators coincide and let L_t^* be the formal adjoint of $L_t = L_t^+$. Then, the forward Kolmogorov equation becomes:

$$\partial_t S_{s,t} = L_t^* S_{s,t},$$

which is an initial value problem for the density f_t with initial condition f_s at s .

Example 32. The time inhomogeneous setting allows to consider diffusion processes with time variable coefficients (see Example 30). The backward Kolmogorov equation reads:

$$-\partial_s u(s, x) = \sum_{i=1}^d b_i(s, x) \partial_{x_i} u(s, x) + \frac{1}{2} \sum_{i,j=1}^d a_{ij}(s, x) \partial_{x_i} \partial_{x_j} u(s, x), \quad u(t, x) = \varphi(x),$$

and the forward Kolmogorov equation (or Fokker-Planck equation) is:

$$\partial_t f_t(x) = - \sum_{i=1}^d \partial_{x_i} \{b_i(t, x) f_t(x)\} + \frac{1}{2} \sum_{i,j=1}^d \partial_{x_i} \partial_{x_j} \{a_{ij}(t, x) f_t(x)\}, \quad f_{t=0}(x) = f_0(x).$$

A.4.3 Non-linear Markov processes

The previous steps have given Markov representations for linear PDEs (with and without time varying coefficients). The class of *nonlinear* PDEs studied in this review can also be seen as the forward Kolmogorov equation associated to a particular class of time-inhomogeneous Markov processes. This point of view was introduced by the seminal work of McKean [123, 125] and Johnson [102].

Let us first introduce an extension of the notion of transition functions.

Definition A.32 (Transition functions, nonlinear case). A family of maps $(P_t^\nu)_{t \geq 0}$ from $E \rightarrow \mathcal{P}(E)$ defined for any $\nu \in \mathcal{P}(E)$ is a family of non-linear transition functions when it satisfies the following properties.

- The map $(t, x, \nu) \mapsto P_t^\nu(x, \cdot)$ is a measurable map $[0, \infty) \times E \times \mathcal{P}(E) \rightarrow \mathcal{P}(E)$.
- For all $x \in E$ and for all $\nu \in \mathcal{P}(E)$, $P_0^\nu(x, \cdot) = \delta_x$.
- For all $s, t \geq 0$, for all $x \in E$, for all $\nu \in \mathcal{P}(E)$ and for all $\mathcal{A} \in \mathcal{B}(E)$, it holds that

$$P_{s+t}^\nu(x, \mathcal{A}) = \int_E P_s^\nu(x, dy) P_t^\nu(y, \mathcal{A}) P_t^\nu(x, dy).$$

This last relation is a nonlinear version of the Chapman-Kolmogorov property; the linear case is recovered when P_t^ν does not depend on ν .

At this point, it is natural to introduce the nonlinear operator $\bar{S}_t : \mathcal{P}(E) \rightarrow \mathcal{P}(E)$ defined for $\nu \in \mathcal{P}(E)$ by

$$\bar{S}_t(\nu)(dy) = \int_E P_t^\nu(x, dy) \nu(dx),$$

so that the non-linear Chapman-Kolmogorov relation reads

$$P_{s+t}^\nu(x, \cdot) = \int_E P_s^{\bar{S}_t(\nu)}(y, \cdot) P_t^\nu(x, dy). \quad (121)$$

Integrating against ν , this gives

$$\bar{S}_{t+s}(\nu) = \int_E P_s^{\bar{S}_t(\nu)}(y, \cdot) \left(\int_E P_t^\nu(x, dy) \nu(dx) \right) = \bar{S}_s(\bar{S}_t(\nu)). \quad (122)$$

In particular $\bar{S}_0 = \text{Id}$, and $(\bar{S}_t)_{t \geq 0}$ appears to be a *non-linear semi-group*: the bar notation reminds of the non-linearity when writing $\bar{S}_t$ alone. The semigroup $(\bar{S}_t)_{t \geq 0}$ is the analog of the previous dual semi-group.

Finally a nonlinear Markov process in the sense of McKean with initial distribution $f_0 \in \mathcal{P}(E)$ is a time-inhomogeneous Markov process with transition functions of the form

$$\bar{P}_{s,t} = P_{t-s}^{\bar{S}_s(f_0)},$$

for $s \leq t$. This corresponds to the construction of Johnson [102] established when the state space E is the 2-state space $E = \{-1, +1\}$.

The nonlinear evolution system is defined for any $s \leq t$ by

$$\forall \varphi \in B_p(E), \quad \bar{T}_{s,t} \varphi(x) = \int_E \varphi(y) \bar{P}_{s,t}(x, dy) = \int_E \varphi(y) P_{t-s}^{f_s}(x, dy).$$

In the above expression, $f_s := \bar{S}_s(f_0)$ is the law at time s of the associated nonlinear Markov process in the sense of McKean. Its right generator is given by:

$$L_{f_t} \varphi(x) := \lim_{\varepsilon \rightarrow 0} \frac{\bar{T}_{t,t+\varepsilon} \varphi(x) - \varphi(x)}{\varepsilon} = \lim_{\varepsilon \rightarrow 0} \frac{\int_E \varphi(y) P_\varepsilon^{f_t}(x, dy) - \varphi(x)}{\varepsilon},$$

from which it can be seen that it depends on f_t only. With this particular form for the dependence in time, the forward Kolmogorov equation given in Example 32 thus appears to be the nonlinear Fokker-Planck equation satisfied by the law of the solution of the McKean-Vlasov diffusion SDE

$$d\bar{X}_t = b(\bar{X}_t, f_t)dt + \sigma(\bar{X}_t, f_t)dB_t, \quad \bar{X}_t \sim f_t,$$

the wellposedness of which has been studied in Proposition 1.

A.5 Large Deviation Principles and Sanov theorem

Definition A.33 (Large Deviation Principle). Given a sequence sequence $(a_N)_N$ of positive numbers $a_N \rightarrow 0$ and a non-negative lower-semicontinuous function I on E , a sequence $(\mu_N)_N$ in $\mathcal{P}(E)$ satisfies a Large Deviation Principle (LDP) with speed a_N and rate function I when for any Borel set $A \subset E$, it holds that

$$-\inf_{\overset{\circ}{A}} I \leq \liminf_{N \rightarrow \infty} a_N \log \mu_N(A) \leq \limsup_{N \rightarrow \infty} a_N \log \mu_N(A) \leq -\inf_{\bar{A}} I,$$

where $\overset{\circ}{A}$ and $\bar{A}$ denote respectively the interior and closure of A .

Given a sequence $(X^i)_{i \in \mathbb{N}}$ of i.i.d. real-valued random variables, Cramér's theorem states that the sequence of the laws $\text{Law}(\frac{1}{N} \sum_{i=1}^N X^i)$, $N \in \mathbb{N}$, satisfies a LDP with rate function $\Lambda^*(x) = \sup_{t \in \mathbb{R}} (tx - \log \mathbb{E}[\exp(tX^1)])$ and speed $a_N = 1/N$. By taking the image random variables $\varphi(X^1), \varphi(X^2), \dots$ for a fixed test function φ , a LDP can be obtained for the sequence of laws $\text{Law}(\langle \mu_{\mathcal{X}^N}, \varphi \rangle)$, where $\mathcal{X}^N = (X^1, \dots, X^N)$. However, in this case, the rate function depends on the choice of the test function φ . A more precise theorem which gives a LDP for the laws of the sequence of empirical measures is Sanov theorem.

Theorem A.34 (Sanov). *Let μ be a probability measure on a Polish space E , and let $(X^i)_{i \in \mathbb{N}}$ be a sequence of independent μ -distributed random variables. For $N \in \mathbb{N}$, we recall the notation $\mathcal{X}^N = (X^1, \dots, X^N) \in E^N$. Then the laws in $\mathcal{P}(\mathcal{P}(E))$ of the measure-valued random variables $\mu_{\mathcal{X}^N}$ satisfies a large deviation principle with speed N^{-1} and rate function the relative entropy $\nu \mapsto H(\nu|\mu)$.*

A.6 Girsanov transform

There are many versions of Girsanov theorem. We do not give the most general result (which can be found for instance in [111, Theorem 5.22]) but only the one which will be used in this review and which can be found in [106, Chapter 3, Theorem 5.1].

Theorem A.35 (Girsanov). *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$ be a filtered probability space and let $(B_t)_t$ be a d -dimensional Brownian motion on this space with $\mathbb{P}(B_0 = 0) = 1$. Let $(X_t)_t$ be a $\mathbb{R}^d$ -valued adapted measurable process and let the process be defined (whenever it exists) for $t < +\infty$ by:*

$$H_t := \int_0^t X_s \cdot dB_s - \frac{1}{2} \int_0^t |X_s|^2 ds.$$

Let $\mathbb{Q}$ the probability measure on $(\Omega, \mathcal{F})$ defined by its Radon-Nikodym derivative on each $\mathcal{F}_T$, $T < +\infty$:

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_T} = \exp(H_T).$$

Assume that $\exp(H)$ is a martingale and let us define the process

$$\tilde{B}_t = B_t - \int_0^t X_s ds.$$

Then for each fixed $T \in [0, +\infty)$, $(\tilde{B}_t)_{t \leq T}$ is a Brownian motion on $(\Omega, \mathcal{F}_T, \mathbb{Q}|_{\mathcal{F}_T})$.

A.7 Poisson random measures

This section briefly explains how to model jump processes using Poisson random measures. The theory of random measures is explained in great details in [100]. Another classical reference on stochastic integration with respect to random measures is [96]. The following presentation is also inspired by [7, Appendix A] and [133, Section 3].

Let us fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$.

Definition A.36 (Poisson random measure). Let $(\mathcal{E}, \mu)$ be a measurable Polish space endowed with a σ -finite measure μ . Let $\widehat{\mathcal{M}}(\mathcal{E})$ be the set of all measures λ on $\mathcal{E}$ which are expressed as a countable sum of Dirac measures on $\mathcal{E}$ and such that $\lambda(\mathcal{A}) < +\infty$ for any μ -finite set $\mathcal{A}$. A Poisson random measure with intensity μ is a mapping $\mathcal{N} : \Omega \rightarrow \widehat{\mathcal{M}}(\mathcal{E})$ with the following properties.

- (i) The mapping $\omega \in \Omega \mapsto \mathcal{N}(\omega, \mathcal{A})$ is measurable for any μ -finite set $\mathcal{A}$.
- (ii) For every disjoint μ -finite sets $\mathcal{A}_1, \dots, \mathcal{A}_k$, the random variables $\mathcal{N}(\mathcal{A}_j)$, $j \in \{1, \dots, k\}$, are independent and $\mathcal{N}(\mathcal{A}_j)$ follows a Poisson distribution on $\mathbb{N}$ with parameter $\mu(\mathcal{A}_j)$.

We will only consider the case $\mathcal{E} = \mathbb{R}_+ \times \Theta$ where Θ is a Polish space and μ is of the form $\mu(dt, d\theta) = dt \otimes \nu(d\theta)$. The results below also extend to the case where ν is replaced by a family of σ -finite measures $(\nu_t)_t$ on Θ which depend on the time parameter. The Poisson random measure $\mathcal{N}$ is assumed to be *adapted* which means that it satisfies the following properties.

- (i) $\mathcal{N}(\mathcal{A})$ is $\mathcal{F}_t$ -measurable for each Borel measurable set $\mathcal{A} \in \mathcal{B}([0, t] \times \Theta)$ with $t > 0$
- (ii) The σ -field generated by $\{\mathcal{N}(\mathcal{A}), \mathcal{A} \in \mathcal{B}((t, +\infty) \times \Theta)\}$ is independent of $\mathcal{F}_t$.

Some ideas on the construction of the stochastic integral against $\mathcal{N}$ are gathered below.

To better understand what a Poisson random measure does, it is useful to consider first the case $\nu(\Theta) < +\infty$. In this case, on every finite time interval $[0, T]$, $\mathcal{N}_t(\Theta) := \mathcal{N}((0, t] \times \Theta)$ defines a classical Poisson process. The Poisson random measure $\mathcal{N}$ can be shown to admit the representation:

$$\mathcal{N}(dt, d\theta) = \sum_{n=1}^{\gamma} \delta_{(T_n, \theta_n)}(dt, d\theta),$$

where $T_1, \dots, T_\gamma$ are the jump times of $\mathcal{N}_t(\Theta)$ and θ_n are i.i.d. random variables with distribution $\nu(d\theta)/\nu(\Theta)$. For any measurable function $a \equiv a(\omega, t, \theta)$ on $\Omega \times \mathbb{R}_+ \times \Theta$ with values in $\mathbb{R}$, the integral with respect to $\mathcal{N}$ is defined by:

$$\int_0^T \int_{\Theta} a(\omega, s, \theta) \mathcal{N}(ds, d\theta) := \sum_{n=1}^{\gamma} a(\omega, T_n, \theta_n).$$

That is to say, it is the sum of the random amplitudes $a(\omega, T_n, \theta_n)$ added at each jumping time T_n .

To extend the previous construction to the case $\nu(\Theta) = +\infty$, let us consider a predictable real-valued function $a \equiv a(\omega, t, \theta)$ on $\Omega \times \mathbb{R}_+ \times \Theta$. We do not write the dependence in ω in the following. We recall that $(\Theta, \nu(d\theta))$ is σ -finite, so there exists an increasing sequence of subsets $(\Theta_p)_p$ such that $\nu(\Theta_p) < +\infty$ and $\Theta = \cup_p \Theta_p$. By the previous construction, the integral of a against $\mathcal{N}$ is well-defined on each subset $[0, T] \times \Theta_p$. There are two cases to distinguish.

1. When a satisfies the L^1 condition

$$\mathbb{E} \left[\int_0^t \int_{\Theta} |a(s, \theta)| \nu(d\theta) ds \right] < +\infty,$$

it is possible to show that the sequence $\left(\int_0^T \int_{\Theta_p} a(s, \theta) \mathcal{N}(ds, d\theta) \right)_p$ is Cauchy in L^1 .

Its limit is denoted by $\int_0^T \int_{\Theta} a(s, \theta) \mathcal{N}(ds, d\theta)$. In this case, the process

$$M_t = \int_0^t \int_{\Theta} a(s, \theta) \mathcal{N}(ds, d\theta) - \int_0^t \int_{\Theta} a(s, \theta) \nu(d\theta) ds, \quad (123)$$

is a $\mathcal{F}_t$ -martingale (in fact, M_t characterises $\mathcal{N}$).

2. When a satisfies the L^2 condition

$$\mathbb{E} \left[\int_0^T \int_{\Theta} |a(s, \theta)|^2 \nu(d\theta) ds \right] < +\infty,$$

then it is possible to prove that (123) still defines a square integrable martingale with quadratic variation

$$\langle M \rangle_t = \int_0^t \int_{\Theta} |a(s, \theta)|^2 \tilde{\mathcal{N}}(ds, d\theta),$$

where

$$\tilde{\mathcal{N}}(dt, d\theta) := \mathcal{N}(dt, d\theta) - \nu(d\theta)dt,$$

is called the compensated measure of $\mathcal{N}$. Note however that the quantity $\int_0^T \int_{\Theta} a(s, \theta) \mathcal{N}(ds, d\theta)$ may not be defined.

The next step is to make sense of the jump-diffusion SDE:

$$\begin{aligned} X_t = X_0 + \int_0^t b(X_s) ds + \int_0^t \sigma(X_s) dB_s + \int_0^t \int_{\Theta} \alpha(X_{s-}, \theta) \mathcal{N}(ds, d\theta) \\ + \int_0^t \int_{\Theta} \tilde{\alpha}(X_{s-}, \theta) \tilde{\mathcal{N}}(ds, d\theta), \end{aligned} \quad (124)$$

where this time $\alpha, \tilde{\alpha} : E \times \Theta \rightarrow E$ for $E = \mathbb{R}^d$. We always assume the following Lipschitz integrability conditions:

- (i) For all $x \in E$ and $T > 0$, it holds that $\int_0^T \int_{\Theta} |\alpha(x, \theta)| \nu(d\theta) < +\infty$ and $\int_0^T \int_{\Theta} |\tilde{\alpha}(x, \theta)|^2 \nu(d\theta) < +\infty$.
- (ii) There exists $C > 0$ such that for any $x, y \in E$,

$$|\sigma(x) - \sigma(y)|^2 + |b(x) - b(y)|^2 + \int_{\Theta} |\tilde{\alpha}(x, \theta) - \tilde{\alpha}(y, \theta)|^2 \nu(d\theta) \leq C|x - y|^2.$$

In the classical theory of SDE (see [96, Chapter IV, Section 9]) the Lipschitz integrability condition

$$\int_{\Theta} |\alpha(x, \theta) - \alpha(y, \theta)|^p \nu(d\theta) \leq C|x - y|^p$$

with $p = 2$ is also assumed. However, from a modelling point of view, it makes more sense in the context of this review to assume this condition with $p = 1$ (see [4, Remark 2.1] or the introduction of [87]). In this L^1 setting, strong existence and uniqueness for the SDE (124) is proved in [87, Theorem 1.2]. Moreover, the generator of the process is the sum of the three generators

$$\begin{aligned} \mathcal{L}\varphi(x) &= \sum_{i=1}^d b(x) \cdot \nabla \varphi(x) + \frac{1}{2} \sum_{i,j=1}^d (\sigma \sigma^T)_{ij}(x) \partial_{x_i} \partial_{x_j} \varphi(x), \\ \mathcal{J}\varphi(x) &= \int_{\Theta} \{\varphi(x + \alpha(x, \theta)) - \varphi(x)\} \nu(d\theta), \\ \tilde{\mathcal{J}}\varphi(x) &= \int_{\Theta} \{\varphi(x + \tilde{\alpha}(x, \theta)) - \varphi(x) - \tilde{\alpha}(x, \theta) \cdot \nabla \varphi(x)\} \nu(d\theta). \end{aligned}$$

B A strengthened version of Hewitt-Savage theorem and its partial system version

The following computation gives the Cauchy-estimate of Theorem 3.17 using stronger metrics.

Corollary 5 (Extending Cauchy-estimates to other metrics). *The Cauchy-estimate of Theorem 3.17 can be obtained in $W_{\delta,p}$ -distance for $M \leq N$ with rate $\varepsilon(M)$, where δ is any metric on $\mathcal{P}(E)$ such that a uniform δ^p -law of large numbers holds, i.e. for every $N \geq 1$, there exists $\varepsilon(N) > 0$ with $\varepsilon(N) \rightarrow 0$ as $N \rightarrow \infty$, such that for every f in $\mathcal{P}_p(E)$ and every $f^{\otimes N}$ -distributed vector $\bar{\mathcal{X}}^N$,*

$$\mathbb{E}[\delta^p(\mu_{\bar{\mathcal{X}}^N}, f)]^{1/p} \leq \varepsilon(N).$$

Thanks to [78], this includes all the Wassertein- p metrics on a compact space E or on any E endowed with a bounded distance.

Proof. In order to couple $\mathcal{X}^M$ to the sub-vector $\mathcal{X}^{M,N}$ of $\mathcal{X}^N$, we choose the transference plane $(\mu_{M,N}, \mu_N)_{\#} \pi^N \otimes \pi^N$: this is well defined thanks to the compatibility property for $\pi^{M,N}$ and π^N . Thus

$$W_{\delta,p}^p(\text{Law}(\mu_{\mathcal{X}^M}), \text{Law}(\mu_{\mathcal{X}^N})) \leq \mathbb{E}_{\mathcal{X}^N \sim \pi^N} [\delta^p(\mu_{\mathcal{X}^{N,M}}, \mu_{\mathcal{X}^N})] = \langle \pi^N, \delta^p(\mu_{M,N}, \mu_N) \rangle.$$

Theorem 3.17 showed that π^N is the N -th moment measure of the limit π in $\mathcal{W}_{H-s-}$

distance, allowing to take advantage of superposing i.i.d. states:

$$\begin{aligned}\langle \pi^N, \delta^p(\boldsymbol{\mu}_{M,N}, \boldsymbol{\mu}_N) \rangle &= \int_{\mathcal{P}(E)} \langle \nu^{\otimes N}, \delta^p(\boldsymbol{\mu}_{M,N}, \boldsymbol{\mu}_N) \rangle \text{Law}(\mu_{\mathcal{X}^N})(d\nu) \\ &\leq C(p) \int_{\mathcal{P}(E)} \langle \nu^{\otimes N}, \delta^p(\boldsymbol{\mu}_{M,N}, \nu) \rangle \text{Law}(\mu_{\mathcal{X}^N})(d\nu) \\ &\quad + C(p) \int_{\mathcal{P}(E)} \langle \nu^{\otimes N}, \delta^p(\nu, \boldsymbol{\mu}_N) \rangle \text{Law}(\mu_{\mathcal{X}^N})(d\nu),\end{aligned}$$

for a constant $C(p)$ which only depends on p . Since $\nu^{\otimes N}$ is the law of a vector of i.i.d. particles

$$\langle \nu^{\otimes N}, \delta^p(\boldsymbol{\mu}_{M,N}, \nu) \rangle = \langle \nu^{\otimes M}, \delta^p(\boldsymbol{\mu}_M, \nu) \rangle = \mathbb{E}_{\bar{\mathcal{X}}^M \sim \nu^{\otimes M}} [\delta^p(\mu_{\bar{\mathcal{X}}^M}, \nu)],$$

and the quantitative assumption on the δ^p -law of large numbers concludes. $\square$

The following proposition gives a useful Cauchy-estimate for the empirical measure of a sub-system of a finite exchangeable particle system. More precisely, let $M < N$ and let $\mathcal{X}^N \sim f^N$ be a finite exchangeable particle system. One wish to compare $\mu_{\mathcal{X}^N}$ and $\mu_{\mathcal{X}^{M,N}}$. This is not possible by a direct coupling argument since the two empirical measures do not have the same size. We thus use an alternative argument based on the special polynomial structure of the H^{-s} norm (see Lemma 3.4). This argument is also used in the proof of the Hewitt-Savage Theorem 3.17.

Proposition 17 (Block empirical measures approximation). *Let $s > d/2$. Let $M < N$ and let $\mathcal{X}^N \sim f^N$ be a finite exchangeable particle system. It holds that*

$$\mathcal{W}_{H^{-s}}(\text{Law}(\mu_{\mathcal{X}^{M,N}}), \text{Law}(\mu_{\mathcal{X}^M})) \leq 2\|\Phi_s\|_{\infty} \left(\frac{1}{M} - \frac{1}{N} \right),$$

where $\Phi_s(z) := \int_{\mathbb{R}^d} e^{-iz \cdot \xi} (1 + |\xi|^2)^{-s} d\xi$.

Proof. Thanks to the identity (46), we get:

$$\begin{aligned}
& \mathbb{E}[\|\mu_{\mathcal{X}^{M,N}} - \mu_{\mathcal{X}^N}\|_{H^{-s}}^2] \\
&= \int_{(\mathbb{R}^d)^N} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi_s(x-y) [\mu_{\mathbf{x}^{M,N}}^{\otimes 2} - \mu_{\mathbf{x}^{M,N}} \otimes \mu_{\mathbf{x}^N}] (dx, dy) \right) f^N(d\mathbf{x}^N) \\
&\quad + \int_{(\mathbb{R}^d)^N} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi_s(x-y) [\mu_{\mathbf{x}^N}^{\otimes 2} - \mu_{\mathbf{x}^N} \otimes \mu_{\mathbf{x}^{M,N}}] (dx, dy) \right) f^N(d\mathbf{x}^N) \\
&= \int_{(\mathbb{R}^d)^N} \left(\frac{1}{M^2} \sum_{k,\ell=1}^M \Phi_s(x^k - x^\ell) - \frac{1}{NM} \sum_{k=1}^M \sum_{j=1}^N \Phi_s(x^k - x^j) \right) f^N(d\mathbf{x}^N) \\
&\quad + \int_{(\mathbb{R}^d)^N} \left(\frac{1}{N^2} \sum_{i,j=1}^N \Phi_s(x^i - x^j) - \frac{1}{NM} \sum_{i=1}^N \sum_{\ell=1}^M \Phi_s(x^i - x^\ell) \right) f^N(d\mathbf{x}^N) \\
&= \left(\frac{M}{M^2} - \frac{M}{NM} \right) \Phi_s(0) \\
&\quad + \left(\frac{M^2 - M}{M^2} - \frac{NM - M}{NM} \right) \int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi_s(x-y) f^{2,N}(dx, dy) \\
&\quad + \left(\frac{N}{N^2} - \frac{M}{NM} \right) \Phi_s(0) \\
&\quad + \left(\frac{N^2 - N}{N^2} - \frac{NM - M}{NM} \right) \int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi_s(x-y) f^{2,N}(dx, dy) \\
&= \left(\frac{1}{M} - \frac{1}{N} \right) \left(\Phi_s(0) - \int_{\mathbb{R}^d \times \mathbb{R}^d} \Phi_s(x-y) f^{2,N}(dx, dy) \right).
\end{aligned}$$

Since Φ_s is bounded, this gives the desired estimate. $\square$

In order to compare the empirical measures of two sub-systems with different sizes but which come from the same exchangeable finite particle system, one can use the same method as in Corollary 5. However, we need to replace the marginals $f^{k,N}$ by their moment measures approximation. Thanks to Lemma 3.11, this error term is quantitative which gives an explicit maximal size for the subsystems.

Proposition 18 (Strong Cauchy-estimates for block empirical measures). *Under the assumptions of Corollary 5, let $M < N$ and let $k < \ell$ such that $\varepsilon(\ell) \geq \ell^2 N^{-1}$. Then it holds that*

$$W_{\delta,p}^p(\text{Law}(\mu_{\mathcal{X}^{k,N}}), \text{Law}(\mu_{\mathcal{X}^{\ell,N}})) \leq \varepsilon(k).$$

Proof. The transference plane is the same as in Corollary 5. Let us consider for simplicity the Wassertein-2 distance (other adaptations are straightforward). Assuming the distance on $E = \mathbb{R}^d$ to be bounded, the condition on ℓ allows to write

$$\begin{aligned}
\mathbb{E}[W_2^2(\mu_{\mathcal{X}^{k,N}}, \mu_{\mathcal{X}^{\ell,N}})] &= \langle f^{\ell,N}, W_2^2(\boldsymbol{\mu}_{k,\ell}, \boldsymbol{\mu}_\ell) \rangle \\
&= \langle \mathbb{E}_{\mathcal{X}^N} \mu_{\mathcal{X}^N}^{\otimes \ell}, W_2^2(\boldsymbol{\mu}_{k,\ell}, \boldsymbol{\mu}_\ell) \rangle + \mathcal{O}(\ell^2 N^{-1}),
\end{aligned}$$

where $\mu_{\mathcal{X}^N}^{\otimes \ell}$ is the ℓ -th moment measure of $\text{Law}(\mu_{\mathcal{X}^N})$. This leads to

$$\begin{aligned} & \langle \mathbb{E}_{\mathcal{X}^N} \mu_{\mathcal{X}^N}^{\otimes \ell}, W_2^2(\boldsymbol{\mu}_{k,\ell}, \boldsymbol{\mu}_\ell) \rangle \\ &= \int_{\mathcal{P}(E)} \langle \nu^{\otimes \ell}, W_2^2(\boldsymbol{\mu}_{k,\ell}, \boldsymbol{\mu}_\ell) \rangle \text{Law}(\mu_{\mathcal{X}^N})(d\nu) \\ &\leq \int_{\mathcal{P}(E)} \langle \nu^{\otimes \ell}, W_2^2(\boldsymbol{\mu}_{k,\ell}, \nu) \rangle \text{Law}(\mu_{\mathcal{X}^N})(d\nu) \\ &\quad + \int_{\mathcal{P}(E)} \langle \nu^{\otimes N}, W_2^2(\nu, \boldsymbol{\mu}_\ell) \rangle \text{Law}(\mu_{\mathcal{X}^N})(d\nu). \end{aligned}$$

Since $\nu^{\otimes \ell}$ is the law of a vector of i.i.d. particles

$$\begin{aligned} \langle \nu^{\otimes \ell}, W_2^2(\boldsymbol{\mu}_{k,\ell}, \nu) \rangle &= \langle \nu^{\otimes k}, W_2^2(\boldsymbol{\mu}_k, \nu) \rangle \\ &= \mathbb{E}_{\bar{\mathcal{X}}^k \sim \nu^{\otimes k}} W_2^2(\mu_{\bar{\mathcal{X}}^k}, \nu), \end{aligned}$$

and the quantitative law of large numbers for the Wassertein-2 distance in [78] concludes. $\square$

Remark 29 (Recovering Cauchy-estimates on finite marginals). If the Wassertein-1 distance is considered instead, this leads to Cauchy estimates on $f^{\ell,N}$ thanks to Proposition 9

$$W_1(f^{k,M}, f^{\ell,N}) = \mathcal{W}_1(\text{Law}(\mu_{\mathcal{X}^{k,M}}), \text{Law}(\mu_{\mathcal{X}^{\ell,N}})) \leq \mathbb{E}[W_1(\mu_{\mathcal{X}^{k,M}}, \mu_{\mathcal{X}^{\ell,N}})].$$

References

- [1] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort and R. Spigler, The Kuramoto model: A simple paradigm for synchronization phenomena, *Rev. Modern Phys.*, **77** (2005), 137–185, <https://link.aps.org/doi/10.1103/RevModPhys.77.137>.
- [2] G. Albi, N. Bellomo, L. Fermo, S.-Y. Ha, J. Kim, L. Pareschi, D. Poyato and J. Soler, Vehicular traffic, crowds, and swarms: From kinetic theory and multi-scale methods to applications and research perspectives, *Math. Models Methods Appl. Sci.*, **29** (2019), 1901–2005, <https://www.worldscientific.com/doi/abs/10.1142/S0218202519500374>.
- [3] L. Ambrosio, N. Gigli and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd edition, Lectures in Mathematics ETH Zürich, Birkhäuser, Basel, 2008.
- [4] L. Andreis, P. Dai Pra and M. Fischer, McKean–Vlasov limit for interacting systems with simultaneous jumps, *Stoch. Anal. Appl.*, **36** (2018), 960–995.
- [5] D. Bakry, L’hypercontractivité et son utilisation en théorie des semigroupes, in *Lectures on Probability Theory: Ecole d’Été de Probabilités de Saint-Flour XXII-1992* (eds. D. Bakry, R. D. Gill, S. A. Molchanov and P. Bernard), Lecture Notes

- in Mathematics, Springer Berlin Heidelberg, 1994, 1–114, <https://doi.org/10.1007/BFb0073872>.
- [6] D. Bakry and M. Émery, Diffusions hypercontractives, in *Séminaire de Probabilités XIX 1983/84* (eds. J. Azéma and M. Yor), no. 1123 in Lecture Notes in Mathematics, Springer Berlin Heidelberg, 1985, 177–206, <http://link.springer.com/10.1007/BFb0075847>, Series Title: Lecture Notes in Mathematics.
 - [7] V. Bansaye and S. Méléard, *Stochastic Models for Structured Populations*, no. 1.4 in Mathematical Biosciences Institute Lecture Series, Springer International Publishing Switzerland, 2015.
 - [8] Nicola Bellomo, Pierre Degond and Eitan Tadmor (eds.), *Active Particles, Volume 1: Advances in Theory, Models, and Applications*, Modeling and Simulation in Science, Engineering and Technology, Springer International Publishing, 2017, <http://link.springer.com/10.1007/978-3-319-49996-3>.
 - [9] Nicola Bellomo, Pierre Degond and Eitan Tadmor (eds.), *Active Particles, Volume 2: Advances in Theory, Models, and Applications*, Modeling and Simulation in Science, Engineering and Technology, Springer International Publishing, 2019, <http://link.springer.com/10.1007/978-3-030-20297-2>.
 - [10] G. Ben Arous and M. Brunaud, Methode de Laplace: étude variationnelle des fluctuations de diffusions de type “champ moyen”, *Stochastics and Stochastics Reports*, **31** (1990), 79–144.
 - [11] G. Ben Arous and O. Zeitouni, Increasing propagation of chaos for mean field models, *Ann. Inst. Henri Poincaré Probab. Stat.*, **35** (1999), 85–102.
 - [12] L. Berlyand, R. Creese, P.-E. Jabin and M. Potomkin, Continuum Approximations to Systems of Correlated Interacting Particles, *J. Stat. Phys.*, **174** (2019), 808–829, <http://link.springer.com/10.1007/s10955-018-2205-8>.
 - [13] L. Bertini, G. Giacomin and K. Pakdaman, Dynamical Aspects of Mean Field Plane Rotators and the Kuramoto Model, *J. Stat. Phys.*, **138** (2009), 270–290.
 - [14] L. Bertini, G. Giacomin and C. Poquet, Synchronization and random long time dynamics for mean-field plane rotators, *Probab. Theory Related Fields*, **160** (2014), 593–653, <http://link.springer.com/10.1007/s00440-013-0536-6>.
 - [15] P. L. Bhatnagar, E. P. Gross and M. Krook, A Model for Collision Processes in Gases. I. Small Amplitude Processes in Charged and Neutral One-Component Systems, *Phys. Rev.*, **94** (1954), 511–525, <https://link.aps.org/doi/10.1103/PhysRev.94.511>.
 - [16] P. Billingsley, *Convergence of Probability Measures*, 2nd edition, Wiley Series in Probability and Statistics., Wiley, New York, 1999.

- [17] T. Bodineau, I. Gallagher and L. Saint-Raymond, The Brownian motion as the limit of a deterministic system of hard-spheres, *Invent. math.*, **203** (2016), 493–553, <http://link.springer.com/10.1007/s00222-015-0593-9>.
- [18] F. Bolley, Quantitative concentration inequalities on sample path space for mean field interaction, *ESAIM Probab. Stat.*, **14** (2010), 192–209.
- [19] F. Bolley, I. Gentil and A. Guillin, Convergence to equilibrium in Wasserstein distance for Fokker–Planck equations, *J. Funct. Anal.*, **263** (2012), 2430–2457, <https://linkinghub.elsevier.com/retrieve/pii/S0022123612002777>.
- [20] F. Bolley, I. Gentil and A. Guillin, Uniform Convergence to Equilibrium for Granular Media, *Arch. Ration. Mech. Anal.*, **208** (2013), 429–445, <http://link.springer.com/10.1007/s00205-012-0599-z>.
- [21] F. Bolley, A. Guillin and C. Villani, Quantitative concentration inequalities for empirical measures on non-compact spaces, *Probab. Theory Related Fields*, **137** (2006), 541–593.
- [22] F. Bolley and C. Villani, Weighted Csiszár–Kullback–Pinsker inequalities and applications to transportation inequalities, *Ann. Fac. Sci. Toulouse Math. (6)*, **14** (2005), 331–352, http://www.numdam.org/item/AFST_2005_6_14_3_331_0/.
- [23] E. Bolthausen, Laplace approximations for sums of independent random vectors, *Probab. Theory Related Fields*, **72** (1986), 305–318, <https://doi.org/10.1007/BF00699109>.
- [24] L. Boltzmann, Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen, *Sitzungsberichte der Akademie der Wissenschaften*, **66** (1872), 275–370, Translation: Further studies on the thermal equilibrium of gas molecules, in *Kinetic Theory 2*, 88–174, Ed. S.G. Brush, Pergamon, Oxford (1966).
- [25] M. Bossy and N. Champagnat, Markov Processes, in *Encyclopedia of Quantitative Finance* (ed. R. Cont), John Wiley & Sons, Ltd, 2010, <http://doi.wiley.com/10.1002/9780470061602.eqf02016>.
- [26] B. Böttcher, Feller evolution systems: Generators and approximation, *Stoch. Dyn.*, **14** (2014), 1350025, <https://www.worldscientific.com/doi/abs/10.1142/S0219493713500251>.
- [27] B. Böttcher, R. Schilling and J. Wang, *Lévy Matters III. Lévy-type processes: construction, approximation and sample path properties*, no. 3 in Lévy Matters, Springer International Publishing, 2013.
- [28] D. Bresch, P.-E. Jabin and Z. Wang, On mean-field limits and quantitative estimates with a large class of singular kernels: Application to the Patlak–Keller–Segel model, *C. R. Math. Acad. Sci. Paris*, **357** (2019), 708–720.

- [29] P. Cardaliaguet, Notes on mean field games (from P.-L. Lions' lectures at Collège de France), in *Lecture given at Tor Vergata*, 2010, 1–59.
- [30] P. Cardaliaguet, F. Delarue, J.-M. Lasry and P.-L. Lions, *The Master Equation and the Convergence Problem in Mean Field Games*, no. 201 in Annals of Mathematics Studies, Princeton University Press, 2019, <http://www.jstor.org/stable/10.2307/j.ctvckq7qf>.
- [31] E. Carlen, M. C. Carvalho and E. Gabetta, Central limit theorem for Maxwellian molecules and truncation of the Wild expansion, *Commun. Pure Appl. Math.*, **53** (2000), 370–397.
- [32] E. Carlen, M. C. Carvalho, J. Le Roux, M. Loss and C. Villani, Entropy and chaos in the Kac model, *Kinet. Relat. Models*, **3** (2008), 85–122.
- [33] E. Carlen, R. Chatelin, P. Degond and B. Wennberg, Kinetic hierarchy and propagation of chaos in biological swarm models, *Phys. D*, **260** (2013), 90–111, <https://linkinghub.elsevier.com/retrieve/pii/S0167278912001492>.
- [34] E. Carlen, P. Degond and B. Wennberg, Kinetic limits for pair-interaction driven master equations and biological swarm models, *Math. Models Methods Appl. Sci.*, **23** (2013), 1339–1376.
- [35] R. Carmona, *Lectures on BSDEs, Stochastic Control, and Stochastic Differential Games with Financial Applications*, SIAM, 2016.
- [36] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I, Mean Field FBSDEs, Control, and Games*, no. 83 in Probability Theory and Stochastic Modelling, Springer International Publishing, 2018.
- [37] R. Carmona and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications II, Mean Field Games with Common Noise and Master Equations*, no. 84 in Probability Theory and Stochastic Modelling, Springer International Publishing, 2018.
- [38] J. A. Carrillo, M. Fornasier, G. Toscani and F. Vecil, Particle, kinetic, and hydrodynamic models of swarming, in *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences* (eds. G. Naldi, L. Pareschi and G. Toscani), Birkhäuser Boston, 2010, 297–336, http://link.springer.com/10.1007/978-0-8176-4946-3_12.
- [39] J. A. Carrillo, S. Jin, L. Li and Y. Zhu, A consensus-based global optimization method for high dimensional machine learning problems, *ESAIM Control Optim. Calc. Var.*, **27** (2021), 1–22, <https://www.esaim-cocv.org/10.1051/cocv/2020046>.
- [40] J. A. Carrillo, M. Delgadino and G. Pavliotis, A λ -convexity based proof for the propagation of chaos for weakly interacting stochastic particles, *J. Funct. Anal.*, **279** (2020), 108734.

- [41] J. A. Carrillo, M. R. D’Orsogna and V. Panferov, Double milling in self-propelled swarms from kinetic theory, *Kinet. Relat. Models*, **2** (2009), 363–378, <http://aims sciences.org/article/doi/10.3934/krm.2009.2.363>.
- [42] P. Cattiaux, F. Delebecque and L. Pédèches, Stochastic Cucker–Smale models: Old and new, *Ann. Appl. Probab.*, **28** (2018), 3239–3286, <https://projecteuclid.org/journals/annals-of-applied-probability/volume-28/issue-5/Stochastic-CuckerSmale-models-Old-and-new/10.1214/18-AAP1400.full>.
- [43] C. Cercignani, *The Boltzmann Equation and Its Applications*, no. 67 in Applied Mathematical Sciences, Springer New York, 1988, <http://link.springer.com/10.1007/978-1-4612-1039-9>.
- [44] C. Cercignani, *Ludwig Boltzmann, the Man Who Trusted Atoms*, Oxford University Press, 2006, <https://oxford.universitypressscholarship.com/view/10.1093/acprof:oso/9780198570646.001.0001/acprof-9780198570646>.
- [45] C. Cercignani, R. Illner and M. Pulvirenti, *The Mathematical Theory of Dilute Gases*, no. 106 in Applied Mathematical Sciences, Springer-Verlag New York, 1994.
- [46] T. Champion and L. De Pascale, The Monge problem in $\mathbb{R}^d$, *Duke Math. J.*, **157** (2011), 551–572.
- [47] L. Chizat and F. Bach, On the Global Convergence of Gradient Descent for Overparameterized Models using Optimal Transport, in *Advances in Neural Information Processing Systems 31 (NeurIPS 2018)* (eds. S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi and R. Garnett), Curran Associates, Inc., Montreal, Canada, 2018, 3040–3050.
- [48] R. Cortez and J. Fontbona, Quantitative propagation of chaos for generalized Kac particle systems, *Ann. Appl. Probab.*, **26** (2016), 892–916, <http://projecteuclid.org/euclid.aoap/1458651823>.
- [49] R. Cortez and J. Fontbona, Quantitative Uniform Propagation of Chaos for Maxwell Molecules, *Commun. Math. Phys.*, **357** (2018), 913–941, <http://link.springer.com/10.1007/s00220-018-3101-4>.
- [50] I. Csiszár, Sanov Property, Generalized I-Projection and a Conditional Limit Theorem, *Ann. Probab.*, **12** (1984), 768–793, <https://projecteuclid.org/euclid.aoap/1176993227>.
- [51] F. Cucker and S. Smale, On the mathematics of emergence, *Jpn. J. Math.*, **2** (2007), 197–227, <http://link.springer.com/10.1007/s11537-007-0647-x>.
- [52] P. Dai Pra and F. den Hollander, McKean-Vlasov limit for interacting random processes in random media, *J. Stat. Phys.*, **84** (1996), 735–772.

- [53] S. Danieri and G. Savaré, Lecture notes on gradient flows and optimal transport, in *Optimal Transportation* (eds. H. Pajot, Y. Ollivier and C. Villani), Cambridge University Press, Cambridge, 2014, 100–144, https://www.cambridge.org/core/product/identifier/CB09781107297296A015/type/book_part.
- [54] D. Dawson, Measure-valued Markov processes, in *École d’Été de Probabilités de Saint-Flour XXI-1991* (ed. P. Hennequin), no. 1541 in Lecture Notes in Mathematics, Springer Berlin Heidelberg, 1993.
- [55] D. Dawson and J. Gärtner, Large deviations from the McKean-Vlasov limit for weakly interacting diffusions, *Stochastics*, **20** (1987), 247–308.
- [56] V. De Bortoli, A. Durmus and X. Fontaine, Quantitative Propagation of Chaos for SGD in Wide Neural Networks, in *Advances in Neural Information Processing Systems 33 (NeurIPS 2020)*, 2020, 278–288, <https://proceedings.neurips.cc/paper/2020/file/02e74f10e0327ad868d138f2b4fdd6f0-Paper.pdf>.
- [57] A. De Masi, A. Galves, E. Löcherbach and E. Presutti, Hydrodynamic Limit for Interacting Neurons, *J. Stat. Phys.*, **158** (2015), 866–902, <http://link.springer.com/10.1007/s10955-014-1145-1>.
- [58] P. Degond, Macroscopic limits of the Boltzmann equation: a review, in *Modeling and Computational Methods for Kinetic Equations* (eds. N. Bellomo, P. Degond, L. Pareschi and G. Russo), Birkhäuser Boston, Boston, MA, 2004, 3–57, http://link.springer.com/10.1007/978-0-8176-8200-2_1, Series Title: Modeling and Simulation in Science, Engineering and Technology.
- [59] P. Degond, Mathematical models of collective dynamics and self-organization, in *Proceedings of the International Congress of Mathematicians ICM 2018*, vol. 4, Rio de Janeiro, Brazil, 2018, 3943–3964.
- [60] P. Degond, A. Frouvelle and J.-G. Liu, Phase Transitions, Hysteresis, and Hyperbolicity for Self-Organized Alignment Dynamics, *Arch. Ration. Mech. Anal.*, **216** (2015), 63–115, <http://link.springer.com/10.1007/s00205-014-0800-7>.
- [61] P. Degond, A. Frouvelle, S. Merino-Aceituno and A. Trescases, Alignment of Self-propelled Rigid Bodies: From Particle Systems to Macroscopic Equations, in *Stochastic Dynamics Out of Equilibrium, Institut Henri Poincaré, Paris, France, 2017* (eds. G. Giacomin, S. Olla, E. Saada, H. Spohn and G. Stoltz), no. 282 in Springer Proceedings in Mathematics & Statistics, Springer, Cham, 2019, 28–66, http://link.springer.com/10.1007/978-3-030-15096-9_2.
- [62] P. Degond and S. Motsch, Continuum limit of self-driven particles with orientation interaction, *Math. Models Methods Appl. Sci.*, **18** (2008), 1193–1215, <https://www.worldscientific.com/doi/abs/10.1142/S0218202508003005>.
- [63] P. Del Moral, Measure-valued processes and interacting particle systems. Application to nonlinear filtering problems, *Ann. Appl. Probab.*, **8** (1998), 438–495.

- [64] P. Del Moral, *Feynman-Kac Formulae, Genealogical and Interacting Particle Systems with Applications*, Probability and Its Applications, Springer-Verlag New York, 2004.
- [65] P. Del Moral, *Mean field simulation for Monte Carlo integration*, no. 126 in Monographs on Statistics and Applied Probability, CRC Press, Taylor & Francis Group, 2013.
- [66] P. Del Moral and S. Penev, *Stochastic Processes: From Applications to Theory*, Texts in Statistical Science, CRC Press : Taylor & Francis Group, 2017.
- [67] P. Del Moral and J. Tugaut, Uniform propagation of chaos and creation of chaos for a class of nonlinear diffusions, *Stoch. Anal. Appl.*, **37** (2019), 909–935.
- [68] P. Diaconis and D. Freedman, Finite Exchangeable Sequences, *Ann. Probab.*, **8** (1980), 745–764, <https://projecteuclid.org/euclid.aop/1176994663>.
- [69] A. Diez, Propagation of chaos and moderate interaction for a piecewise deterministic system of geometrically enriched particles, *Electron. J. Probab.*, **25** (2020), 1–38.
- [70] M. R. D’Orsogna, Y. L. Chuang, A. L. Bertozzi and L. S. Chayes, Self-Propelled Particles with Soft-Core Interactions: Patterns, Stability, and Collapse, *Phys. Rev. Lett.*, **96** (2006), 104302, <https://link.aps.org/doi/10.1103/PhysRevLett.96.104302>.
- [71] A. Durmus, A. Eberle, A. Guillin and R. Zimmer, An elementary approach to uniform in time propagation of chaos, *Proc. Amer. Math. Soc.*, **148** (2020), 5387–5398.
- [72] A. Eberle, Reflection couplings and contraction rates for diffusions, *Probab. Theory Related Fields*, **166** (2016), 851–886.
- [73] A. Eberle, A. Guillin and R. Zimmer, Quantitative Harris-type theorems for diffusions and McKean-Vlasov processes, *Trans. Amer. Math. Soc.*, **371** (2019), 7135–7173.
- [74] N. El Karoui and S. Méléard, Martingale measures and stochastic calculus, *Probab. Theory Related Fields*, **84** (1990), 83–101, <http://link.springer.com/10.1007/BF01288560>.
- [75] S. N. Ethier and T. G. Kurtz, *Markov processes: characterization and convergence*, Wiley series in probability and mathematical statistics, Wiley, New York, 1986.
- [76] M. Feldman and R. J. McCann, Monge’s transport problem on a Riemannian manifold, *Trans. Amer. Math. Soc.*, **354** (2001), 1667–1697.
- [77] J. Fontbona, H. Guérin and S. Méléard, Measurability of optimal transportation and convergence rate for Landau type interacting particle systems, *Probab. Theory Related Fields*, **143** (2009), 329–351, <http://link.springer.com/10.1007/s00440-007-0128-4>.

- [78] N. Fournier and A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure, *Probab. Theory Related Fields*, **162** (2015), 707–738, Publisher: Springer.
- [79] N. Fournier and E. Löcherbach, On a toy model of interacting neurons, *Ann. Inst. Henri Poincaré Probab. Stat.*, **52** (2016), 1844–1876, <https://projecteuclid.org/journals/Annales-de-linstitut-henri-poincare-probabilites-et-statistiques/volume-52/issue-4/On-a-toy-model-of-interacting-neurons/10.1214/15-AIHP701.full>.
- [80] N. Fournier and S. Méléard, A Markov Process Associated with a Boltzmann Equation Without Cutoff and for Non-Maxwell Molecules, *J. Stat. Phys.*, **104** (2001), 359–385, <http://link.springer.com/10.1023/A:1010322130480>.
- [81] N. Fournier and S. Mischler, Rate of convergence of the Nanbu particle system for hard potentials and Maxwell molecules, *Ann. Probab.*, **44** (2016), 589–627, <http://projecteuclid.org/euclid.aop/1454423051>.
- [82] T. Funaki, A Certain Class of Diffusion Processes Associated with Nonlinear Parabolic Equations, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **67** (1984), 331–348.
- [83] I. Gallagher, L. Saint-Raymond and B. Texier, *From Newton to Boltzmann: Hard Spheres and Short-Range Potentials*, no. 18 in Zurich Lectures in Advanced Mathematics, European Mathematical Society, 2014.
- [84] J. Gärtner, On the McKean-Vlasov limit for Interacting Diffusions, *Math. Nachr.*, **137** (1988), 197–248.
- [85] F. Golse, On the Dynamics of Large Particle Systems in the Mean Field Limit, Lecture notes, arXiv:1301.5494.
- [86] H. Grad, Asymptotic theory of Boltzmann equation, *Phys. Fluids*, **6** (1963), 147–181.
- [87] C. Graham, McKean-Vlasov Itô-Skorohod equations, and nonlinear diffusions with discrete jump sets, *Stochastic Process. Appl.*, **40** (1992), 69–82.
- [88] C. Graham, Nonlinear diffusion with jumps, *Ann. Inst. Henri Poincaré Probab. Stat.*, **28** (1992), 393–402.
- [89] C. Graham and S. Méléard, Stochastic particle approximations for generalized Boltzmann models and convergence estimates, *Ann. Probab.*, **25** (1997), 115–132.
- [90] S. Grassi and L. Pareschi, From particle swarm optimization to consensus based optimization: stochastic modeling and mean-field limit, *Math. Models Methods Appl. Sci.*, **31** (2021), 1625–1657.

- [91] F. A. Grünbaum, Propagation of chaos for the Boltzmann equation, *Arch. Ration. Mech. Anal.*, **42** (1971), 323–345.
- [92] A. Guillin and P. Monmarché, Uniform long-time and propagation of chaos estimates for mean field kinetic particles in non-convex landscapes, *J. Stat. Phys.*, **185** (2020), 1–20.
- [93] S.-Y. Ha, K. Lee and D. Levy, Emergence of time-asymptotic flocking in a stochastic Cucker-Smale system, *Commun. Math. Sci.*, **7** (2009), 453–469.
- [94] M. Hauray and S. Mischler, On Kac’s chaos and related problems, *J. Funct. Anal.*, **266** (2014), 6055–6157.
- [95] T. Holding, Propagation of chaos for Hölder continuous interaction kernels via Glivenko-Cantelli, preprint, arXiv:1608.02877.
- [96] N. Ikeda and S. Watanabe, *Stochastic differential equations and diffusion processes*, no. 24 in North-Holland Mathematical Library, Elsevier, 1989, <http://qut.ebiblib.com.au/patron/FullRecord.aspx?p=1888657>.
- [97] P.-E. Jabin, A review of the mean field limits for Vlasov equations, *Kinet. Relat. Models*, **7** (2014), 661–711.
- [98] P.-E. Jabin and Z. Wang, Mean Field Limit for Stochastic Particle Systems, in *Active Particles, Volume 1 : Advances in Theory, Models, and Applications* (eds. N. Bellomo, P. Degond and E. Tadmor), Modeling and Simulation in Science, Engineering and Technology, Birkhäuser Basel, 2017, 379–402, https://doi.org/10.1007/978-3-319-49996-3_10.
- [99] P.-E. Jabin and Z. Wang, Quantitative estimates of propagation of chaos for stochastic systems with $W^{-1,\infty}$ kernels, *Invent. Math.*, **214** (2018), 523–591.
- [100] J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*, Second edition, no. 288 in Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg, 2003, <http://link.springer.com/10.1007/978-3-662-05265-5>.
- [101] A. Joffe and M. Métivier, Weak convergence of sequences of semimartingales with applications to multitype branching processes, *Adv. in Appl. Probab.*, **18** (1986), 20–65, https://www.cambridge.org/core/product/identifier/S0001867800015585/type/journal_article.
- [102] D. P. Johnson, On a class of stochastic processes and its relationship to infinite particle gases, *Trans. Amer. Math. Soc.*, **132** (1968), 275–275, <http://www.ams.org/jourcgi/jour-getitem?pii=S0002-9947-1968-0256452-X>.
- [103] B. Jourdain and S. Méléard, Propagation of chaos and fluctuations for a moderate model with smooth initial data, *Ann. Inst. Henri Poincaré Probab. Stat.*, **34**

- (1998), 727–766, http://www.numdam.org/item/AIHPB_1998__34_6_727_0, Publisher: Gauthier-Villars.
- [104] M. Kac, Foundations of kinetic theory, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 3, University of California Press Berkeley and Los Angeles, California, 1956, 171–197.
 - [105] M. Kac, Some Probabilistic Aspects of the Boltzmann Equation, in *The Boltzmann Equation. Acta Physica Austriaca (Supplementum X Proceedings of the International Symposium “100 Years Boltzmann Equation” in Vienna 4th–8th September 1972)* (eds. E. G. D. Cohen and W. Thirring), Springer Vienna, 1973, 379–400, http://link.springer.com/10.1007/978-3-7091-8336-6_17.
 - [106] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd edition, no. 113 in Graduate Texts in Mathematics, Springer-Verlag New York, 1998.
 - [107] S. Kusuoka and Y. Tamura, Gibbs measures for mean field potentials, *J. Fac. Sci. Univ. Tokyo*, **31** (1984), 223–245.
 - [108] D. Lacker, On a strong form of propagation of chaos for McKean-Vlasov equations, *Electron. Commun. Probab.*, **23** (2018), 1–11.
 - [109] D. Lacker, Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions, preprint, arXiv:2105.02983.
 - [110] O. E. Lanford, Time evolution of large classical systems, in *Dynamical Systems Theory and Application, Battelle Seattle 1974 Rencontres* (ed. J. Moser), Springer-Verlag Berlin Heidelberg, 1975.
 - [111] J.-F. Le Gall, *Brownian Motion, Martingales, and Stochastic Calculus*, no. 274 in Graduate Texts in Mathematics, Springer International Publishing, 2016.
 - [112] M. Ledoux, Concentration of measure and logarithmic Sobolev inequalities, *Séminaire de probabilités de Strasbourg*, **33** (1999), 120–216, http://www.numdam.org/item/SPS_1999__33__120_0/.
 - [113] C. Léonard, Une loi des grands nombres pour des systèmes de diffusions avec interaction et à coefficients non bornés, *Ann. Inst. Henri Poincaré Probab. Stat.*, **22** (1986), 237–262.
 - [114] C. Léonard, Large deviations for long range interacting particle systems with jumps, *Ann. Inst. Henri Poincaré Probab. Stat.*, **31** (1995), 289–323, http://www.numdam.org/item/AIHPB_1995__31_2_289_0/.
 - [115] C. Léonard, On large deviations for particle systems associated with spatially homogeneous Boltzmann type equations, *Probab. Theory Related Fields*, **101** (1995), 1–44, <https://doi.org/10.1007/BF01192194>.

- [116] C. Léonard, Girsanov theory under a finite entropy condition, in *Séminaire de Probabilités XLIV*, no. 2046 in Lecture Notes in Mathematics, Springer, Berlin, Heidelberg, 2012, 429–465, https://doi.org/10.1007/978-3-642-27461-9_20.
- [117] G. Letta, Sur les théorèmes de Hewitt-Savage et de de Finetti, *Séminaire de probabilités de Strasbourg*, **23** (1989), 531–535, http://www.numdam.org/item/SPS_1989__23__531_0/.
- [118] T. Lindvall and L. C. G. Rogers, Coupling of Multidimensional Diffusions by Reflection, *Ann. Probab.*, **14**, <https://projecteuclid.org/journals/annals-of-probability/volume-14/issue-3/Coupling-of-Multidimensional-Diffusions-by-Reflection/10.1214/aop/1176992442.full>.
- [119] W. Liu, L. Wu and C. Zhang, Long-time behaviors of mean-field interacting particle systems related to McKean-Vlasov equations, *Commun. Math. Phys.*, **387** (2021), 179–214.
- [120] E. Luçon, Large Population Asymptotics for Interacting Diffusions in a Quenched Random Environment, in *From Particle Systems to Partial Differential Equations II* (eds. P. Gonçalves and A. J. Soares), no. 129 in Springer Proceedings in Mathematics & Statistics, Springer, Cham, 2015, 231–251, http://link.springer.com/10.1007/978-3-319-16637-7_8.
- [121] F. Malrieu, Logarithmic Sobolev inequalities for some nonlinear PDE's, *Stochastic Process. Appl.*, **95** (2001), 109–132.
- [122] D. Matthes and G. Toscani, On Steady Distributions of Kinetic Models of Conservative Economies, *J. Stat. Phys.*, **130** (2008), 1087–1117, <http://link.springer.com/10.1007/s10955-007-9462-2>.
- [123] H. P. McKean, A class of Markov processes associated with nonlinear parabolic equations, *Proc. Nat. Acad. Sci.*, **56** (1966), 1907, Publisher: National Academy of Sciences.
- [124] H. P. McKean, An exponential formula for solving Boltzmann's equation for a Maxwellian gas, *Journal of Combinatorial Theory*, **2** (1967), 358–382, <http://www.sciencedirect.com/science/article/pii/S0021980067800358>.
- [125] H. P. McKean, Propagation of chaos for a class of non-linear parabolic equations, in *Lecture Series in Differential Equations, Volume 2* (ed. A. K. Aziz), no. 19 in Van Nostrand Mathematical Studies, Van Nostrand Reinhold Company, 1969, 177–194.
- [126] S. Mei, A. Montanari and P.-M. Nguyen, A mean field view of the landscape of two-layer neural networks, *Proc. Natl. Acad. Sci. USA*, **115** (2018), E7665–E7671, <http://www.pnas.org/lookup/doi/10.1073/pnas.1806579115>.

- [127] S. Méléard, Asymptotic Behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models, in *Probabilistic Models for Nonlinear Partial Differential Equations* (eds. D. Talay and L. Tubaro), no. 1627 in Lecture Notes in Mathematics, Springer-Verlag Berlin Heidelberg, 1996.
- [128] S. Méléard and S. Roelly-Coppoletta, Systèmes de particules et mesures-martingales : un théorème de propagation du chaos, *Séminaire de probabilités (Strasbourg)*, **22** (1988), 438–448.
- [129] S. Mischler, Kac’s chaos and Kac’s program, in *Séminaire Laurent Schwartz - EDP et applications*, vol. 2012-2013, Institut des hautes études scientifiques & Centre de mathématiques Laurent Schwartz, 2012, Exposé no XXII, 1–17.
- [130] S. Mischler and C. Mouhot, Kac’s program in kinetic theory, *Invent. Math.*, **193** (2013), 1–147, Publisher: Springer.
- [131] S. Mischler, C. Mouhot and B. Wennberg, A new approach to quantitative propagation of chaos for drift, diffusion and jump processes, *Probab. Theory Related Fields*, **161** (2015), 1–59.
- [132] Adrian Muntean and Federico Toschi (eds.), *Collective Dynamics from Bacteria to Crowds: An Excursion Through Modeling, Analysis and Simulation*, no. 553 in CISM International Centre for Mechanical Sciences, Springer, Vienna, 2014, <http://link.springer.com/10.1007/978-3-7091-1785-9>.
- [133] H. Murata, Propagation of chaos for Boltzmann-like equation of non-cutoff type in the plane, *Hiroshima Math. J.*, **7** (1977), 479–515, <https://projecteuclid.org/euclid.hmj/1206135751>.
- [134] Giovanni Naldi, Lorenzo Pareschi and Giuseppe Toscani (eds.), *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences*, Modeling and Simulation in Science, Engineering and Technology, Birkhäuser Boston, 2010, <http://link.springer.com/10.1007/978-0-8176-4946-3>.
- [135] K. Nanbu, Direct Simulation Scheme Derived from the Boltzmann Equation. I. Monocomponent Gases, *Journal of the Physical Society of Japan*, **49** (1980), 2042–2049, <http://journals.jps.jp/doi/10.1143/JPSJ.49.2042>.
- [136] K. Oelschläger, A Martingale Approach to the Law of Large Numbers for Weakly Interacting Stochastic Processes, *Ann. Probab.*, **12** (1984), 458–479, <https://projecteuclid.org/euclid.aop/1176993301>.
- [137] F. Otto and C. Villani, Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality, *J. Funct. Anal.*, **173** (2000), 361–400.
- [138] K. R. Parthasarathy, *Probability Measures on Metric Spaces*, Academic Press, 1967, <https://linkinghub.elsevier.com/retrieve/pii/C20130081078>.

- [139] R. Pinnau, C. Totzeck, O. Tse and S. Martin, A consensus-based model for global optimization and its mean-field limit, *Math. Models Methods Appl. Sci.*, **27** (2017), 183–204, <https://www.worldscientific.com/doi/abs/10.1142/S0218202517400061>.
- [140] M. Pulvirenti, Kinetic limits for stochastic particle systems, in *Probabilistic Models for Nonlinear Partial Differential Equations* (eds. D. Talay and L. Tubaro), no. 1627 in Lecture Notes in Mathematics, Springer-Verlag Berlin Heidelberg, 1996.
- [141] D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, 3rd edition, no. 293 in Grundlehren der mathematischen Wissenschaften, Springer-Verlag Berlin Heidelberg, 1999, <http://link.springer.com/10.1007/978-3-662-06400-9>.
- [142] G. M. Rotskoff and E. Vanden-Eijnden, Trainability and Accuracy of Neural Networks: An Interacting Particle System Approach, preprint, arXiv:1805.00915.
- [143] N. Rougerie, De Finetti theorems, mean-field limits and Bose-Einstein condensation, *Lectures notes from a course at the LMU, Munich, 2015*, arXiv:1506.05263.
- [144] L. Rueschendorf, A. Schnurr and V. Wolf, Comparison of time-inhomogeneous Markov processes, *Adv. in Appl. Probab.*, **48** (2016), 1015–1044.
- [145] S. Salem, A gradient flow approach to propagation of chaos, *Discrete Contin. Dyn. Syst.*, **40** (2020), 5729–5754.
- [146] S. Serfaty, Systems of points with Coulomb interactions, in *Proceedings of the International Congress of Mathematicians (ICM 2018)*, World Scientific, Rio de Janeiro, Brazil, 2019, 935–977, https://www.worldscientific.com/doi/abs/10.1142/9789813272880_0033.
- [147] J. Sirignano and K. Spiliopoulos, Mean Field Analysis of Neural Networks: A Law of Large Numbers, *SIAM J. Appl. Math.*, **80** (2020), 725–752, <https://epubs.siam.org/doi/10.1137/18M1192184>.
- [148] D. W. Stroock and S. R. S. Varadhan, *Multidimensional Diffusion Processes*, Classics in Mathematics, Springer Berlin Heidelberg, 1997, <http://link.springer.com/10.1007/3-540-28999-2>.
- [149] A.-S. Sznitman, Équations de type de Boltzmann, spatialement homogènes, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **66** (1984), 559–592.
- [150] A.-S. Sznitman, Nonlinear Reflecting Diffusion Process, and the Propagation of Chaos and Fluctuations Associated, *J. Funct. Anal.*, **56** (1984), 311–336.
- [151] A.-S. Sznitman, Topics in propagation of chaos, in *Éc. Été Probab. St.-Flour XIX—1989*, Springer, 1991, 165–251.

- [152] Denis Talay and Luciano Tubaro (eds.), *Probabilistic Models for Nonlinear Partial Differential Equations*, no. 1627 in Lecture Notes in Mathematics, Springer-Verlag Berlin Heidelberg, 1996, <http://link.springer.com/10.1007/BFb0093175>.
- [153] H. Tanaka, Probabilistic Treatment of the Boltzmann Equation of Maxwellian Molecules, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **46** (1978), 67–105.
- [154] H. Tanaka, Some probabilistic problems in the spatially homogeneous Boltzmann equation, in *Theory and Application of Random Fields, Proceedings of the IFIP-WG 7/1 Working Conference, Bangalore 1982* (ed. G. Kallianpur), Lecture Notes in Control and Information Sciences, Springer-Verlag Berlin Heidelberg, 1983, 258–267.
- [155] C. Totzeck, Trends in Consensus-based optimization, preprint, arXiv:2104.01383.
- [156] H. F. Trotter, Approximation of semi-groups of operators, *Pacific J. Math.*, **8** (1958), 887–919.
- [157] T. Vicsek and A. Zafeiris, Collective motion, *Phys. Rep.*, **517** (2012), 71–140, <https://linkinghub.elsevier.com/retrieve/pii/S0370157312000968>.
- [158] C. Villani, Limite de champ moyen, *Cours de DEA*.
- [159] C. Villani, A Review of Mathematical Topics in Collisional Kinetic Theory, in *Handbook of Mathematical Fluid Dynamics* (eds. S. Friedlander and D. Serre), vol. 1, Elsevier Science, 2002, 71–74, <https://linkinghub.elsevier.com/retrieve/pii/S1874579202800040>.
- [160] C. Villani, *Topics in Optimal Transportation*, no. 58 in Graduate Studies in Mathematics, American Mathematical Society, 2003.
- [161] C. Villani, *Optimal Transport, Old and New*, no. 338 in Grundlehren der mathematischen Wissenschaften, Springer-Verlag Berlin Heidelberg, 2009, <http://link.springer.com/10.1007/978-3-540-71050-9>.
- [162] W. Wagner, A functional law of large numbers for Boltzmann type stochastic particle systems, *Stoch. Anal. Appl.*, **14** (1996), 591–636.
- [163] E. Wild, On Boltzmann’s equation in the kinetic theory of gases, *Math. Proc. Camb. Phil. Soc.*, **47** (1951), 602–609, https://www.cambridge.org/core/product/identifier/S0305004100026992/type/journal_article.

Received xxxx 20xx; revised xxxx 20xx.