

PROBABILISTIC TOOLS AND MONTE-CARLO APPROXIMATIONS FOR SOME BOLTZMANN EQUATIONS

CARL GRAHAM¹ AND SYLVIE MÉLÉARD²

Abstract. This paper aims to introduce to non-specialist mathematicians some probabilistic methods which can be used for the analysis of kinetic equations. We provide the necessary probabilistic background, starting from the basics. We interpret the equations as the Kolmogorov equations for Markov processes, which we define using martingale problem formulations. This interpretation is natural, given the statistical mechanics derivation of many of these equations, and gives information on quantities out of the reach of the kinetic equation formulation. We show how sample path techniques yield new quantitative and qualitative results, and provide Monte-Carlo approximations with rates of convergence. We notably consider some kinetic equations without cutoff, with non-integrable cross-sections.

AMS Subject Classification. 35B65, 60H07, 60K35, 65C05, 65C35, 82C40, 82C80.

The dates will be set by the publisher.

INTRODUCTION

The aim of the course is to present a probabilistic interpretation of some Boltzmann equations in terms of Markov processes. Then, probabilistic tools at our disposal will allow us to study some structural properties of the solutions, as well as some effective Monte-Carlo methods for their approximation and simulation.

The main idea is that the Boltzmann equation is a nonlinear integro-differential equation, which can be interpreted as the equation for the time-marginal laws of a Markov process. Then, instead of working at the level of the PDE, we consider the associated Markov process.

In the first section of these notes, we explain or recall the principal tools we use, and in particular we describe in simple cases the relations between transport or Fokker-Planck equations and Markov processes. In this situation, the Markov processes are diffusion processes with jumps.

In Section 2, we introduce the Boltzmann equation and explain its probabilistic interpretation. The Boltzmann equation describes the evolution of the limit density $f(t, x, v)$ of a rarefied gas at time t in position x and velocity v and is given by

$$\partial_t f + v \cdot \nabla_x f = Q(f, f)$$

where Q is a quadratic integral operator (cf. [6]). This equation can be seen as the equation satisfied by the flow of marginals at time t of a probability measure P defined on a path space. The law P is defined as solution of a nonlinear martingale problem, in which appear two mechanisms: a free transport without interaction and

Keywords and phrases: kinetic equations, Markov processes, nonlinear martingale problems, Monte-Carlo methods.

¹ CMAP, École Polytechnique, 91128 Palaiseau France (UMR CNRS 7641). e-mail: carl@cmapp.polytechnique.fr

² Université Paris 10, MODAL'X, 200 avenue de la République, 92000 Nanterre France. e-mail: sylm@ccr.jussieu.fr

a jump mechanism containing the interaction. This collision terms depends on the law of the process itself, but locally through the density at x of the law. This leads to a ill-defined process, because of a possible unbounded jump measure and a delocalized interaction. This difficulty is of the same nature than the one encountered by the analysts. Existence and uniqueness results, subjects of a large literature, are very restrictive (cf. [9]).

Then we consider in the third section of these lectures a simplified model in which the interaction is delocalized in space (and becomes a mean field interaction) and the jump kernel is bounded (cutoff hypothesis). In this case, existence and uniqueness of a solution of this mollified Boltzmann equation are proved and one can define a corresponding nonlinear Markov process (called Boltzmann process) whose jump measure has a linear dependence on the law of the process (cf. [15]). This linear dependence allows us to define some simple or binary mean-field interacting n -particle systems which give a unified treatment of the Bird and Nanbu models (cf. [1] and [25]). Using a pathwise representation of the particle systems by interacting random graphs and a probabilistic argument called coupling, we prove a strong propagation of chaos result: when n goes to infinity, the law of each subsystem of fixed size k of the n -particle system converges in variation norm to the k -fold product of the law P of a Boltzmann process, with a precise $\mathcal{O}(1/n)$ rate of convergence. This result implies in particular the convergence of the random empirical measures (μ^n) of the system to P .

Since the flow of time-marginals of P solves the mollified Boltzmann equation, the pathwise representation and convergence result yield an effective simulation of the solution of this equation: the n -particle system is simulated using an acceptance-rejection collision method, at collision times given by a Poisson process.

Section 4 is devoted to a case with cutoff and small initial data for which one has existence and uniqueness of the solution of the full Boltzmann equation. (cf. [24], [22]). In this case, our approach allows us to obtain an approximating result for the solution of the full Boltzmann equation by stochastic interacting particle systems.

Next, we are interested in the suppression of the cutoff hypothesis. We consider a very simple model of non cutoff Boltzmann equation called the Kac equation. This is a spatially homogeneous equation in dimension one (cf. [7]). In this model, $f(t, x, v)$ does not depend on x and so the interaction is mean-field. The jump measure defined on \mathbb{R} has an explicit form, explodes near zero but has a second order moment. In this case, we can construct some simulable approximating particle systems. We also prove regularity results on the solution using the stochastic calculus of variations. These results have been generalized by Fournier [11] for a Boltzmann equation without cutoff in dimension 2.

TABLE OF CONTENTS

- 1 - Probabilistic prerequisites
 - 1.1 Random variables, their laws, and expectations
 - 1.2 Conditional expectations
 - 1.3 Random processes, the Skorohod space, and the canonical probability space
 - 1.4 Markov processes and martingale problems
 - 1.5 Laws of large numbers and confidence intervals
 - 1.6 Simulation of random variables and processes
- 2 - The Boltzmann equation and its probabilistic interpretation
 - 2.1 The Boltzmann equation
 - 2.2 The nonlinear martingale problem
- 3 - The mollified problem
 - 3.1 The mollified nonlinear martingale problem
 - 3.2 The approximating stochastic interacting particle systems
 - 3.3 Propagation of chaos in variation norm and rate of convergence

3.4 Algorithms of simulation for the mollified Boltzmann equation

- 4 - The case of a Boltzmann equation with cutoff and small initial data
 - 4.1 The existence and uniqueness result and the nonlinear martingale problem
 - 4.2 Stochastic approximations
- 5 - A generalized Kac equation without cutoff
 - 5.1 The set up
 - 5.2 Probabilistic interpretation.
 - 5.3 Stochastic approximations of the Kac equation without cutoff.
 - 5.4 Numerical results

1. PROBABILISTIC PREREQUISITES

We are going to introduce quickly the main probabilistic notions and tools we shall use. A difficulty resides in the probabilistic terminology, which has been historically developed separately from mathematical analysis; standard mathematical objects may then appear mysterious. We also refer for this to the books of Bouleau [4] and of Lapeyre, Pardoux and Sentis [20].

1.1. Random variables, their laws, and expectations

1.1.1. Probability spaces and random variables

A **probability space** is a measurable space furnished with a positive measure of total mass 1. From a probabilistic point of view, a probability space (Ω, \mathcal{F}, P) models the effect of chance:

- The set Ω models the **universe** of possible outcomes of chance. A point ω in Ω represents some scenario, which may occur due to randomness.
- The σ -algebra \mathcal{F} on Ω represents subsets of Ω , called **events**, which can be quantified probabilistically.
- The **probability measure** P on \mathcal{F} quantifies the probabilistic knowledge we have about any event A in \mathcal{F} by attributing to it its probability of occurrence $P(A) \in [0, 1]$. In particular, we have $P(\Omega) = 1$.

Intuitively, chance furnishes us with an outcome ω chosen or drawn according to the probability measure P .

A subset N of Ω is **negligible** if there exists A in \mathcal{F} such that $N \subset A$ and $P(A) = 0$. A property holds **almost surely** (abbreviated **a.s.**), or with probability 1, if it holds except for ω in a negligible set. Most quantities or properties we consider are defined or hold only a.s., and we do not always bother to mention it.

A **random variable** X is a measurable mapping $X : \Omega \rightarrow E$, where E is a measurable space, called the state or sample space. For some ω in Ω , the value $X(\omega)$ is a sample of X . A difficulty is that probabilists often understate the ω in their formulæ. For instance we use the notation

$$\{X \in B\} = \{\omega : X(\omega) \in B\} = X^{-1}(B), \quad B \in \mathcal{B}(E).$$

We denote the σ -algebra generated by X by

$$\sigma(X) = \{\{X \in B\} : B \in \mathcal{B}(E)\}$$

and the measurability of X is equivalent to the fact that $\sigma(X) \subset \mathcal{F}$.

In general, the space of values E is a **Polish space** (a topological space corresponding to a separable and complete metric space) with the Borel σ -algebra $\mathcal{B}(E)$. The set $\mathcal{P}(E)$ of probability measures on a Polish space E is then furnished with the **weak topology** corresponding to the duality with **continuous bounded** functions, which makes it also a Polish space. The interest of this topology is to preserve the total mass of the measures. If ρ is a measure on E , the weak convergence on $\mathcal{P}(E)$ can be metrized by the metric D defined by

$$D(P, Q) = \sup\{|\langle P, f \rangle - \langle Q, f \rangle| : f \in BL_1(E)\}, \quad \forall P, Q \in \mathcal{P}(E),$$

where $BL_1(E)$ is the space of bounded Lipschitz functions with Lipschitz constant and uniform bound lesser than or equal to 1.

1.1.2. Laws of random variables. Independence of random variables

A part of the difficulty in interpreting probabilistic texts is that the same statement may often be read in two different ways:

- As a statement about random variables. The statement is considered globally in a functional way, and leads to look for properties satisfied almost surely, that means for all ω except for a negligible set.
- As a statement concerning a sample, i.e., concerning an outcome ω which is understated. We consider that chance has provided us with an unknown ω , and we witness the effects of this choice. That leads to the notion of law of a random variable.

Let X be a random variable. For B in $\mathcal{B}(E)$, $\{X \in B\}$ is measurable, and we may consider $P(\{X \in B\})$, which we denote for shorthand by $P(X \in B)$. We hence define a probability measure on $(E, \mathcal{B}(E))$, denoted by $\mathcal{L}(X)$ or \mathcal{L}_X and called the **law** of X under P , by setting

$$\mathcal{L}_X(B) = P(X \in B) = P \circ X^{-1}(B), \quad B \in \mathcal{B}(E).$$

Then $(E, \mathcal{B}(E), \mathcal{L}_X)$ is a probability space, called the **image** of (Ω, \mathcal{F}, P) by X .

Given a random variable X , we are often led to consider the image probability space $(E, \mathcal{B}(E), \mathcal{L}_X)$ rather than (Ω, \mathcal{F}, P) , since E usually has relevant topological properties, while (Ω, \mathcal{F}, P) is abstract and may be huge with respect to notions related to X . Note that a **canonical** way to represent directly a random variable on $(E, \mathcal{B}(E), \mathcal{L}_X)$ with same law as X is to consider the identity mapping from $(E, \mathcal{B}(E), \mathcal{L}_X)$ to $(E, \mathcal{B}(E))$

The simplest random variable is constant: for some a , for any ω in Ω , we have $X(\omega) = a$. Then $\sigma(X) = \{\emptyset, \Omega\}$, and $\mathcal{L}_X = \delta_a$. Such a random variable is said to be **deterministic**. The simplest example of interesting random variable is the **indicator function** \mathbb{I}_A of an event A in \mathcal{F} , defined by

$$\mathbb{I}_A(\omega) = 1 \text{ for } \omega \in A, \quad \mathbb{I}_A(\omega) = 0 \text{ for } \omega \in A^C,$$

and then $\sigma(\mathbb{I}_A) = \{\emptyset, A, A^C, \Omega\}$ and $\mathcal{L}(\mathbb{I}_A) = P(A)\delta_1 + (1 - P(A))\delta_0$.

Remark 1.1. The law \mathcal{L}_X is characterized by the collection of the $\langle \mathcal{L}_X, \phi \rangle$ for a sufficiently large set of ϕ . On a Polish space, we may for example consider the set of bounded Lipschitz functions $BL_1(E)$ we have defined above, and there also exists countable such sets. If X takes values in \mathbb{R}^d , its law is characterized by its Fourier transform, called in the probabilistic language **the characteristic function** of X . The Polish spaces we shall consider are functional spaces, on which there is no Fourier analysis, and which necessitate more intricate tools.

We are now interested in modeling a collection of random variables such that the observation of some of the random variables does not modify the probabilistic knowledge we have about the others. That leads to the notion of independent random variables, the independence being at the level of the statistical behavior.

Definition 1.2. The random variables X_1, \dots, X_k are **independent** if and only if

$$\mathcal{L}(X_1, \dots, X_k) = \mathcal{L}(X_1) \otimes \dots \otimes \mathcal{L}(X_k).$$

An infinite collection of random variables is independent if and only if any finite sub-collection is.

1.1.3. Expectations of random variables

Definition 1.3. For $\phi \in L^\infty(E)$ with values in \mathbb{R} or in a finite-dimensional vector space such as \mathbb{R}^d , the quantity

$$E(\phi(X)) = \int_{\Omega} \phi(X(\omega)) P(d\omega) = \int_E \phi(x) \mathcal{L}_X(dx) = \langle \mathcal{L}_X, \phi \rangle$$

is well defined. It is called the **expectation**, **expected value**, or **mean** (value) of $\phi(X)$, belongs to the same set as $\phi(X)$, and can be considered component-wise.

We extend this notion to arbitrary non-negative measurable functions ϕ by allowing $E(\phi(X)) = \infty$. For arbitrary functions ϕ such that $E(|\phi(X)|) < \infty$, we set $E(\phi(X)) = E(\phi(X)^+) - E(\phi(X)^-)$.

For p in $[1, \infty[$, we set $\|X\|_p = E(|X|^p)^{1/p}$ and say that X belongs to $L^p(\Omega, \mathcal{F}, P)$ if $\|X\|_p < \infty$. We set $\|X\|_\infty = \inf\{m \geq 0 : P(|X| \leq m) = 1\}$ and say that X belongs to $L^\infty(\Omega, \mathcal{F}, P)$ if $\|X\|_\infty < \infty$. For $1 \leq p \leq q \leq \infty$, we have $\|X\|_p \leq \|X\|_q$.

If X is in L^2 then $E(X)$ is the best approximation of X by a constant, which corresponds to the orthogonal projection on the subspace of deterministic random variables.

Note that if X and Y are independent, then for any bounded functions ϕ and ψ

$$E(\phi(X)\psi(Y)) = E(\phi(X))E(\psi(Y)).$$

1.2. Conditional expectations

The scope of probability theory is to assign probabilities to certain events. Restricted probabilistic information or knowledge on (Ω, \mathcal{F}, P) can be often naturally coded by the sub- σ -algebra of events which we can observe and evaluate in that prospect. For instance, $\sigma(X)$ represents the information available by observing X . The probabilistic knowledge in $(E, \mathcal{B}(E), \mathcal{L}_X)$ is equivalent to that in $(\Omega, \sigma(X), P)$, while \mathcal{F} may contain inaccessible or irrelevant information. We state a simple result in this direction.

Lemma 1.4. (*Doob Lemma.*) *On (Ω, \mathcal{F}, P) , let X and Y be random variables with values respectively in E and in F . Assume that X is $\sigma(Y)$ -measurable. Then there is a \mathcal{L}_Y -a.s. unique measurable function $h : F \rightarrow E$ such that $X(\omega) = h(Y(\omega))$ for any $\omega \in \Omega$.*

This is often written $X = h(Y)$, but we explicited the ω to emphasize the fact that h depends on the (functions) X and Y , but not on ω .

We consider a random variable X , and some sub- σ -algebra \mathcal{R} of \mathcal{F} representing the restricted information which is available to us. We want to express the “best” prediction (or approximation) of X “given” this knowledge. This approximation will be \mathcal{R} -measurable, which explains “given”. We provide a definition of “best”, and more details.

Theorem 1.5. *Let a probability space (Ω, \mathcal{F}, P) and a sub- σ -algebra \mathcal{R} of \mathcal{F} be given. For any X in $L^1(\Omega, \mathcal{F}, P)$ there exists an a.s. unique random variable, denoted by $E(X | \mathcal{R})$ and called the **conditional expectation of X given \mathcal{R}** , such that:*

1. $E(X | \mathcal{R})$ belongs to $L^1(\Omega, \mathcal{R}, P)$
2. For any Z in $L^\infty(\Omega, \mathcal{R}, P)$, we have $E(ZX) = E(ZE(X | \mathcal{R}))$.

*If X is in $L^2(\Omega, \mathcal{F}, P)$ then $E(X | \mathcal{R})$ is the **orthogonal projection** of X onto $L^2(\Omega, \mathcal{R}, P)$, and Property 2 is valid for all Z in $L^2(\Omega, \mathcal{R}, P)$ and corresponds to the characteristic property of the orthogonal projection.*

If $\mathcal{R} = \sigma(Y)$ for a F -valued random variable Y we denote $E(X | \mathcal{R})$ by $E(X | Y)$ and call it the **conditional expectation of X given Y** . The Doob Lemma (Lemma 1.4) yields that for some function h ,

$$E(X | Y)(\omega) = h(Y(\omega)), \quad \omega \in \Omega,$$

where h is \mathcal{L}_Y -a.s. unique, which is coherent with the fact that $E(X | Y)$ is P -a.s. defined. We then define the **conditional expectation of X given that Y is equal to a** by

$$E(X | Y = a) = h(a), \quad a \in F.$$

Note that $a \mapsto E(X | Y = a)$ is well defined only \mathcal{L}_Y -a.s. on F , or actually on $Y(\Omega)$.

For an event A we define $P(A | \mathcal{R}) = E(\mathbb{I}_A | \mathcal{R})$, the conditional probability of A given \mathcal{R} , and similarly define other notions of conditional probabilities. If B is an event such that $P(B) > 0$, it is simple to check that the elementary definition $P(A | B) = P(A \cap B)/P(B)$ of the conditional probability of A given B coincides with $P(A | \mathbb{I}_B = 1)$. If $P(B) = 0$ then $P(A | \mathbb{I}_B = 1)$ is arbitrary since $\{1\}$ is negligible for $\mathcal{L}(\mathbb{I}_B) = \delta_0$.

We are now going to list the main properties of the conditional expectation. These are essentially those of the expectation and those of an orthogonal projection.

Theorem 1.6. *Let (Ω, \mathcal{F}, P) , a sub- σ -algebra \mathcal{R} of \mathcal{F} , and X in $L^1(\Omega, \mathcal{F}, P)$ be given.*

1. *If $X \geq 0$ a.s., then $E(X | \mathcal{R}) \geq 0$. (Positivity)*
2. *For $a, b \in \mathbb{R}$ and $X, Y \in L^1(\Omega, \mathcal{F}, P)$, we have $E(aX + bY | \mathcal{R}) = aE(X | \mathcal{R}) + bE(Y | \mathcal{R})$. (Linearity)*
3. *$\|E(X | \mathcal{R})\|_p \leq \|X\|_p$ for $1 \leq p \leq \infty$. (Contraction in the L^p spaces)*
4. *If X is \mathcal{R} -measurable then $E(X | \mathcal{R}) = X$.*
5. *For any \mathcal{R} -measurable Z such that ZX is in L^1 , we have $E(ZX | \mathcal{R}) = ZE(X | \mathcal{R})$.*
6. *If \mathcal{S} is a sub-algebra of \mathcal{R} , then $E(E(X | \mathcal{R}) | \mathcal{S}) = E(X | \mathcal{S})$.*
7. *We have $E(X | \{\emptyset, \Omega\}) = E(X)$. In particular, $E(E(X | \mathcal{R})) = E(X)$.*
8. *If X and Y are independent and integrable, then $E(X/Y) = E(X)$.*

1.3. Random processes, the Skorohod space, and the canonical probability space

We model random evolutions in time, using a sample set which is a set of paths (functions of time) on \mathbb{R}^d , or more generally on a Polish space E . We write $X = (X_t)_{t \geq 0}$ and say that X is a (random) **process**. Actually, we consider simultaneously X as a random path and as a time-indexed collection of random variables, and call $X(\omega) = (X_t)_{t \geq 0}(\omega) = (X_t(\omega))_{t \geq 0}$ a sample path.

In 1933, Kolmogorov gives a rigorous probability model to define properly the law of a process, and this step is considered as the basis of the modern probability theory. If the process $X = (X_t)_{t \geq 0}$ is such that for some measurable space (E, \mathcal{E}) , $X_t : \Omega \rightarrow E$ for $t \geq 0$, then we consider the probability space (called the product space) $(E^{\mathbb{R}^+}, \mathcal{E}^{\mathbb{R}^+})$, where $\mathcal{E}^{\mathbb{R}^+}$ (called the product σ -field) is the smallest σ -field for which the canonical projections $x \in E^{\mathbb{R}^+} \mapsto x_t \in E$ are measurable for all $t \geq 0$. The important result is the following.

Theorem 1.7. *1) The law P of a process X with sample paths in $E^{\mathbb{R}^+}$ is characterized by the family of its **finite-dimensional marginals** $P_{t_1, \dots, t_k} = \mathcal{L}(X_{t_1}, \dots, X_{t_k})$ on E^k , for $k \geq 1$ and $0 \leq t_1 < \dots < t_k$.*

*2) A family of probability measures P_{t_1, \dots, t_k} on E^k , for $k \geq 1$ and $0 \leq t_1 < \dots < t_k$, is **consistent** if whenever $0 \leq s_1 < \dots < s_i$ is a subset of $0 \leq u_1 < \dots < u_j$, then P_{s_1, \dots, s_i} is the corresponding marginal of P_{u_1, \dots, u_j} (the family of finite-dimensional marginals of a process is necessarily consistent). Let E be a Polish space and \mathcal{E} its Borel σ -field. If we are given such a consistent family, then there exists a unique law P on the product space $E^{\mathbb{R}^+}$ with product σ -field $\mathcal{E}^{\mathbb{R}^+}$ having this family as set of finite-dimensional marginals.*

This theorem justifies the existence of the Brownian motion, the Poisson process, etc.

We assume now that there is some regularity in the motions and thus in the sample paths. For this, we define the **Skorohod space** $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ as the set of all right-continuous and left-handed limited functions from \mathbb{R}_+ to \mathbb{R}^d , and consider processes $X = (X_t)_{t \geq 0}$ such that $X : \Omega \rightarrow \mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$. The **Skorohod metric** on $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ is a complicated and technical metric, constructed in order to generalize appropriately the metric of uniform convergence on compact sets on $C(\mathbb{R}_+, \mathbb{R}^d)$, and make $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ a Polish space. The Borel σ -algebra is then the product σ -algebra, and if $X = (X_t)_{t \geq 0}$ is a process with sample paths in $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ then $X_t : \Omega \mapsto \mathbb{R}^d$ are random variables for $t \geq 0$. See Ethier and Kurtz [10], pp. 116 seq., for precisions.

If we consider a restricted time interval $[0, T]$ (finite horizon), we assume that the sample paths are continuous at T , and define the Skorohod space $\mathbb{D}([0, T], \mathbb{R}^d)$ and its metric accordingly.

We define the law $\mathcal{L}(X)$ of X and for any $t \geq 0$ the law $\mathcal{L}(X_t)$ of X_t , and

$$\mathcal{L}(X) \in \mathcal{P}(\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)), \quad \mathcal{L}(X_t) \in \mathcal{P}(\mathbb{R}^d), \quad (\mathcal{L}_{X_t})_{t \geq 0} \in \mathbb{D}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}^d)),$$

and $\mathcal{L}(X_t)$ is called the (one-dimensional) **marginal** of $\mathcal{L}(X)$ on \mathbb{R}^d at time t .

Given a probability space (Ω, \mathcal{F}, P) and a process Y with an appropriately defined evolution, we wish to study the law of Y . Since $(\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d), \mathcal{B}(\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)), \mathcal{L}_Y)$ is then a probability space, we consider $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ and its Borel σ -field as a **canonical probability space**, and the **canonical process** $X = (X_t)_{t \geq 0}$ is given by

$X_t(\omega) = \omega_t$ for $\omega = (\omega_t)_{t \geq 0}$ in $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$. We often study directly a law P on $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ and its Borel σ -field by describing the random evolution of X , and then P is automatically the law of X . We denote \mathcal{L}_{X_t} by P_t .

It is clear that $P \in \mathcal{P}(\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)) \mapsto (P_t)_{t \geq 0} \in \mathbb{D}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}^d))$ corresponds to a projection, but $(P_t)_{t \geq 0}$ does not suffice to specify P . Events determined by the state of the process at more than one time, such as first hitting times, sojourn times, or maximal values attained over an interval of time, cannot be given a probability using only $(P_t)_{t \geq 0}$. There are many processes sharing the same time marginals but with very different laws.

1.4. Markov processes and martingale problems

Let (Ω, \mathcal{F}, P) be given. A **filtration** is an increasing family $(\mathcal{F}_t)_{t \geq 0}$ of sub- σ -fields:

$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}, \quad \forall 0 \leq s \leq t,$$

where \mathcal{F}_t models the information available to an observer up to time t . If X is a process on (Ω, \mathcal{F}) , then its **proper filtration** $(\mathcal{F}_t^X)_{t \geq 0}$ is given by

$$\mathcal{F}_t^X = \sigma((X_s)_{0 \leq s \leq t})$$

which corresponds to the information obtained by observing the process up to time t .

We consider a **filtered probability space** $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$. A process X is **adapted to the filtration** $(\mathcal{F}_t)_{t \geq 0}$ if and only if X_t is \mathcal{F}_t -measurable, or equivalently if and only if $\mathcal{F}_t^X \subset \mathcal{F}_t$ for all $t \geq 0$. Processes are **adapted** if not stated otherwise, and their behavior up to time t can be described by events in \mathcal{F}_t .

The **canonical filtered probability space** is the canonical space $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ with its Borel σ -field and the canonical filtration, which is the proper filtration of the canonical process.

1.4.1. Markov processes, their transition kernels and semi-groups

We consider a family of processes used for modeling some irregular phenomena, such as those appearing in statistical physics, which have no memory, in the following sense: *the probabilistic prediction of the future evolution after a time $s \geq 0$, given knowledge of the past before s , depends only on the present state at time s .*

This prediction may or may not depend on s : when it does not, we speak of **Markov processes**, or of **homogeneous** Markov processes to stress this fact; when it does, of **inhomogeneous** Markov processes.

Definition 1.8. $(Y_t)_{t \geq 0}$ is an inhomogeneous Markov process if and only if

$$E(\phi(Y_{s+t}) | \mathcal{F}_s) = E(\phi(Y_{s+t}) | Y_s), \quad \forall t, s \geq 0, \phi \in L^\infty.$$

Equivalently, for arbitrary integer $k \geq 1$, times $u_1, \dots, u_k \geq s$, and bounded function ϕ of k variables, we have $E(\phi(Y_{u_1}, \dots, Y_{u_k}) | \mathcal{F}_s) = E(\phi(Y_{u_1}, \dots, Y_{u_k}) | Y_s)$.

Markov processes generalize ordinary differential equation flows, with a given physical evolution and many possible initial conditions.

According to the Doob Lemma (Lemma 1.4), for a fixed t , there is a function h_s such that we have $E(\phi(Y_{t+s}) | Y_s)(\omega) = h_s(Y_s(\omega))$. For homogeneous Markov processes, h_s will not depend on s and we have

$$E(\phi(Y_{t+s}) | Y_s)(\omega) = h(Y_s(\omega)).$$

In general $h : y \mapsto E(\phi(Y_{t+s}) | Y_s = y)$ is defined only \mathcal{L}_{Y_s} -a.s., but we assume it is well-defined on \mathbb{R}^d .

Note that if $(Y_t)_{t \geq 0}$ is an inhomogeneous Markov process, then $(t, Y_t)_{t \geq 0}$ is a homogeneous Markov process.

In the following, we shall avoid unnecessary technicalities, and consider homogeneous Markov processes with sample paths in the Skorohod space and arbitrary starting points. We have defined a Markov process on an arbitrary probability space. Let us now define a family of Markov laws.

Definition 1.9. 1) On the Skorohod space $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ with canonical filtration and canonical process $(X_t)_{t \geq 0}$, the family of laws $P_x \in \mathcal{P}(\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d))$, for x in \mathbb{R}^d , such that $P_x(X_0 = x) = 1$, is a **family of homogeneous Markov laws** if and only if $(X_t)_{t \geq 0}$ is a Markov process under P_x (see Definition 1.8) for each $x \in \mathbb{R}^d$ and

$$E_x(\phi(X_{s+t}) | X_s = y) = E_y(\phi(X_t)), \quad \forall t, s \geq 0, \quad x, y \in \mathbb{R}^d, \quad \phi \in L^\infty. \quad (1.1)$$

This can be summarized as follows:

$$E_x(\phi(X_{s+t}) | \mathcal{F}_s)(\omega) = E_{X_s(\omega)}(\phi(X_t)), \quad \forall t, s \geq 0, \quad x \in \mathbb{R}^d, \quad \omega \in \mathbb{D}(\mathbb{R}_+, \mathbb{R}^d), \quad \phi \in L^\infty. \quad (1.2)$$

For ν in $\mathcal{P}(\mathbb{R}^d)$, we set $P_\nu = \int P_x \nu(dx)$. Note that $P_{\delta_x} = P_x$. Under P_ν , the canonical process $(X_t)_{t \geq 0}$ is a Markov process such that the law of X_0 is ν , and (1.1) and (1.2) hold with ν instead of x .

2) For $x \in \mathbb{R}^d$ and $t \geq 0$, the marginals $P_{x,t}$ of P_x on \mathbb{R}^d at time t define the **transition kernel** for the Markov process and satisfy the **Chapman-Kolmogorov equations**

$$P_{x,s+t}(\cdot) = \int P_{y,t}(\cdot) P_{x,s}(dy), \quad \forall t, s \geq 0, \quad x \in \mathbb{R}^d. \quad (1.3)$$

3) The family $(T_t)_{t \geq 0}$ of linear positive contraction operators

$$T_t : \phi \in L^\infty \mapsto T_t \phi \in L^\infty(\mathbb{R}^d), \quad T_t \phi(x) = E_x(\phi(X_t)) = \int \phi(y) P_{x,t}(dy), \quad (1.4)$$

satisfies $T_t 1 = 1$ and the **semi-group property**

$$T_0 = Id, \quad T_{s+t} = T_s \circ T_t, \quad \forall t, s \geq 0, \quad (1.5)$$

and is called the **semi-group** of the Markov process.

The Chapman-Kolmogorov equations (1.3) are obtained by integrating (1.1) with respect to $P_{x,s}(dy)$ and are equivalent to the semi-group property (1.5). The notions and properties in 2) and 3) are clearly equivalent.

Assume we are given $P_{x,t}$, x in \mathbb{R}^d , $t \geq 0$. We wish to construct the laws P_x of the Markov process starting at x in \mathbb{R}^d . The Markov property implies that for any $k \geq 1$ and $0 \leq t_1 < \dots < t_k$, the finite-dimensional marginals P_{x,t_1, \dots, t_k} of P_x , which are the law under P_x of $(X_{t_1}, \dots, X_{t_k})$, are given by

$$P_{x,t_1, \dots, t_k}(dx_1, dx_2, \dots, dx_k) = P_{x,t_1}(dx_1) P_{x_1, t_2 - t_1}(dx_2) \cdots P_{x_{k-1}, t_k - t_{k-1}}(dx_k), \quad (1.6)$$

and the Chapman-Kolmogorov property (1.3) is a necessary and sufficient condition for consistency.

Hence, by Theorem 1.7, 1) is an equivalent notion to 2) and 3) **provided** we introduce the proper extension (1.6) for the finite-dimensional marginals, **obtained** from the one-dimensional marginals **using** the Markov property. Indeed, there exists a unique law P_x on the product space $(\mathbb{R}^d)^{\mathbb{R}_+}$ such that (1.6) holds.

We again stress that P_x contains much more probabilistic information on quantities related to sample paths, such as extremal values or hitting times, than its marginal process $(P_{x,t})_{t \geq 0}$, and that this information very often is physically relevant.

1.4.2. Generators of Markov processes

Let $(T_t)_{t \geq 0}$ be a Markov semi-group. Its **generator** \mathcal{A} with domain $\mathcal{D}(\mathcal{A}) \subset L^\infty$ is given by

$$\mathcal{A}\phi = \lim_{t \rightarrow 0} \frac{T_t \phi - \phi}{t}, \quad \mathcal{D}(\mathcal{A}) \text{ is the set of } \phi \text{ in } L^\infty \text{ for which the limit exists.} \quad (1.7)$$

The Hille-Yosida theory gives necessary and sufficient conditions for an operator to be the generator of a Markov semi-group, see Ethier and Kurtz [10] pp. 6 seq., 165 seq. Essentially this defines in a proper way T_t as $e^{t\mathcal{A}}$. A simple consequence of the definition is that

$$T_t\phi - \phi = \int_0^t T_s\mathcal{A}\phi ds, \quad \forall \phi \in \mathcal{D}(\mathcal{A}), t \geq 0.$$

From the Markov process point of view, we have

$$\mathcal{A}\phi(x) = \lim_{t \rightarrow 0} \frac{E_x(\phi(X_t)) - \phi(x)}{t}, \quad \forall \phi \in \mathcal{D}(\mathcal{A}), x \in \mathbb{R}^d, \quad (1.8)$$

$$E_x(\phi(X_t)) - \phi(x) = \int_0^t E_x(\mathcal{A}\phi(X_s)) ds, \quad \forall \phi \in \mathcal{D}(\mathcal{A}), t \geq 0, x \in \mathbb{R}^d, \quad (1.9)$$

which is a linear evolution equation for the marginal processes $(P_{x,t})_{t \geq 0}$. In the sequel we often denote $P_{x,t}$ or $P_{\nu,t}$ simply by P_t , specifying the initial condition $P_0 = \delta_x$ or $P_0 = \nu$ when necessary. The equation (1.9) writes

$$\langle P_t, \phi \rangle - \langle P_0, \phi \rangle = \int_0^t \langle P_s, \mathcal{A}\phi \rangle ds, \quad \forall \phi \in \mathcal{D}(\mathcal{A}), t \geq 0, \quad (1.10)$$

and corresponds to a weak formulation for the evolution equation

$$\partial_t P_t = \mathcal{A}^* P_t \quad (1.11)$$

where \mathcal{A}^* is the formal **dual** or **adjoint** operator of \mathcal{A} . This equation is often called the (forward) **Kolmogorov** equation or the **Fokker-Planck** equation (especially when \mathcal{A}^* is a second-order differential operator) for the Markov process. In the terminology of Kac, it is called the **master** equation.

Remark 1.10. We have seen that if we consider a d -dimensional Markov process $(Y_t)_{t \geq 0}$, possibly inhomogeneous, then $(t, Y_t)_{t \geq 0}$ is a homogeneous $(d+1)$ -dimensional Markov process. Its generator is then of the form $\hat{\mathcal{A}} = \partial_t + \mathcal{A}_t$ acting on functions $\phi(t, x)$ on $\mathbb{R}_+ \times \mathbb{R}^d$ where \mathcal{A}_t acts only on the $x \in \mathbb{R}^d$ variable, and does not depend on $t \geq 0$ if and only if the Markov process $(Y_t)_{t \geq 0}$ is homogeneous.

Remark 1.11. The equation (1.10), even with the addition of the Chapman-Kolmogorov equations (1.3), does not suffice to define the Markov process of generator \mathcal{A} ; this definition requires the introduction of a sample-path property such as the Markov property. The Markov process corresponds to a microscopic description of the evolution, and the equation (1.10) to a macroscopic description in which random fluctuations are averaged out.

1.4.3. Martingale problems

In order to model these random fluctuations, we introduce the following notion.

Definition 1.12. A process M on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ is a martingale if and only if

1. It is adapted and integrable: $M_t \in L^1(\Omega, \mathcal{F}_t, P)$, for all $t \geq 0$.
2. It is conditionally constant: $E(M_t | \mathcal{F}_s) = M_s$, for all $0 \leq s \leq t$.

Then the best prediction of M_t knowing the past before $s \leq t$ is M_s itself, corresponding to a fair game situation. There are many consequences of this definition, obtained in particular using Theorem 1.6. Let us just note that

$$E(M_t | \mathcal{F}_s) = M_s \text{ is equivalent to } E(M_t - M_s | \mathcal{F}_s) = 0 \text{ and implies } E(M_t) = E(M_s), \quad 0 \leq s \leq t.$$

The connection between the Markov process, its generator, and martingales, is that the generator expresses the trend of the Markov process, with a remainder which is a martingale, as seen in the following result.

Theorem 1.13. (*Dynkin formula.*) Let \mathcal{A} be a generator with domain $\mathcal{D}(\mathcal{A})$, and P_x , for x in \mathbb{R}^d , be the corresponding Markov laws on $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$. Then for any x in \mathbb{R}^d and ϕ in $\mathcal{D}(\mathcal{A})$,

$$\phi(X_t) - \phi(X_0) - \int_0^t \mathcal{A}\phi(X_s) ds = M_t^\phi, \quad t \geq 0,$$

defines a martingale under P_x for the canonical filtration.

Proof. For ϕ in $\mathcal{D}(\mathcal{A})$ and $0 \leq s \leq t$, M_t^ϕ is in $L^1(\Omega, \mathcal{F}_t, P)$, and

$$E_x(M_t^\phi - M_s^\phi | \mathcal{F}_s) = E_x\left(\phi(X_t) - \phi(X_s) - \int_s^t \mathcal{A}\phi(X_u) du \mid \mathcal{F}_s\right)$$

and using the Markov property and the Doob Lemma (Lemma 1.4), this is equal to

$$E_x\left(\phi(X_t) - \phi(X_s) - \int_s^t \mathcal{A}\phi(X_u) du \mid X_s\right) = h(X_s)$$

for some function h . Using (1.1) and the Kolmogorov equation (1.9), for $y \in \mathbb{R}^d$, P_s -a.s.,

$$h(y) = E_x\left(\phi(X_t) - \phi(X_s) - \int_s^t \mathcal{A}\phi(X_u) du \mid X_s = y\right) = E_y\left(\phi(X_{t-s}) - \phi(X_0) - \int_0^{t-s} \mathcal{A}\phi(X_u) du\right) = 0.$$

Hence $E_x(M_t^\phi - M_s^\phi | \mathcal{F}_s) = 0$. □

Note that $M_0^\phi = 0$ and thus $E_x(M_t^\phi) = 0$ for $t \geq 0$, yielding the Kolmogorov equation (1.9). There is much more information about sample path behavior in the definition of a martingale than this mean-zero property. We shall see how this enables us to define the Markov process in a way inspired by the Kolmogorov equation.

Definition 1.14. (Martingale problem). Let $\mathcal{C} \subset L^\infty$, and \mathcal{A} be an operator defined on \mathcal{C} . We say that P_x solves the **martingale problem** for \mathcal{A} starting at x in \mathbb{R}^d if $P_x(X_0 = x) = 1$ and for any ϕ in \mathcal{C} ,

$$\phi(X_t) - \phi(X_0) - \int_0^t \mathcal{A}\phi(X_s) ds = M_t^\phi, \quad t \geq 0,$$

defines a martingale under P_x for the canonical filtration.

We do not seek for the maximum generality, and assume that we are in a situation in which we know that the sample paths of solutions to the martingale problem on product space are a.s. in the Skorohod space. The following results are extracted from Ethier and Kurtz [10], pp. 182–188.

Theorem 1.15. *Considering Definition 1.14, assume that for any $x \in \mathbb{R}^d$ there exists a unique solution P_x on $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ for the martingale problem for \mathcal{A} starting at x . Then (P_x) , for x in \mathbb{R}^d , is a family of Markov laws with generator extending \mathcal{A} . Conversely, if (P_x) , for x in \mathbb{R}^d , is a family of Markov laws with generator \mathcal{A} , and $\mathcal{C} \subset \mathcal{D}(\mathcal{A})$ is a separating set, then it is the unique solution for the martingale problem in Definition 1.14.*

We take for \mathcal{C} a convenient set of sufficiently regular functions, sufficiently large for uniqueness to hold. Choices often made contain the set of smooth functions with compact support C_K^∞ . If \mathcal{A} is obviously bounded on L^∞ , then this is also a good choice.

1.4.4. *Probabilistic interpretation of linear evolution equations*

Consider the linear evolution equation for functions $f(t, x)$ defined for $t \geq 0$ and $x \in \mathbb{R}^d$

$$\partial_t f = \mathcal{A}^* f \tag{1.12}$$

with an integrable initial condition $f_0 = f(0, \cdot)$, in which the operator \mathcal{A}^* appears to be the adjoint of some Markov generator \mathcal{A} . We interpret it in weak distributional sense, with measure-valued solutions $f(t, x)dx$. We can separate $f_0(x)dx$ into its positive and negative parts and normalize each to a probability density, and the properties of a Markov generator imply that the set of probability measures is preserved by the evolution corresponding to (1.12). Thus, considering the linearity of the equation, it is no limitation to restrict our attention to probability density initial values f_0 and probability density solutions. A natural extension is to consider arbitrary probability measure initial conditions ν and probability measure solutions. Since $P_\nu = \int P_x \nu(dx)$, it is enough to consider Dirac masses as initial values.

We have thus interpreted (1.12) as the Kolmogorov equation (1.10), (1.11) for the Markov process with generator \mathcal{A} , and obtain a solution by considering the marginals $(P_t)_{t \geq 0}$ of the Markov process law. If moreover P_t has a density f_t on \mathbb{R}^d , for $t > 0$, then this gives a function-valued solution for (1.12).

An important part of the study is to check that \mathcal{A} is indeed a Markov generator. For this task, we shall develop probabilistic tools such as sample path representations and computations and stochastic calculus. We never use the Hille-Yosida theory, but directly exhibit a probabilistic construction for the Markov process itself.

If \mathcal{A} is a generator the Markov process laws P_x are uniquely defined, but there may be probability measure solutions of (1.10) or (1.11) distinct from the marginals $(P_{x,t})_{t \geq 0}$, not satisfying the Chapman-Kolmogorov equations (1.3). Selecting the Markovian solution is usually a physical choice, and introduces a notion of Markovian uniqueness.

Specification of a Markov process using a martingale problem

Depending on the form of \mathcal{A} , we decide on an appropriate set of test functions \mathcal{C} and consider the martingale problem for \mathcal{A} as in Definition 1.14. If we succeed in proving existence and uniqueness of the solution $P_x \in \mathcal{P}(\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d))$ for the martingale problem starting at any x in \mathbb{R}^d , these solutions $(P_x)_{x \in \mathbb{R}^d}$ correspond to a Markov process with generator \mathcal{A} , and their marginal flows $(P_{x,t})_{t \geq 0}$ solve the Kolmogorov equation (1.10), with test functions in $\mathcal{D}(\mathcal{A}) \supset \mathcal{C}$.

1.4.5. *Examples of generators, and sample path constructions*

We shall now, at last, see how the generators of the Markov processes in which we are interested look like.

Flows of differential equations

We assume that the vector field $b : x \in \mathbb{R}^d \mapsto b(x) \in \mathbb{R}^d$ is such that there exists a unique solution to the differential equation $\dot{x}_t = b(x_t)$. The corresponding flow X_t^x , for $x \in \mathbb{R}^d$ and $t \geq 0$, is the unique solution for

$$X_t^x = x + \int_0^t b(X_s^x) ds, \quad t \geq 0. \tag{1.13}$$

The paths are deterministic, and it is obvious that we have thus defined a Markov process such that P_x is the Dirac mass at $(X_t^x)_{t \geq 0}$. Considering (1.8)

$$\mathcal{A}\phi(x) = \left. \frac{\partial}{\partial t} \phi(X_t^x) \right|_{t=0} = \sum_{i=1}^d b_i(x) \partial_i \phi(x) = b(x) \cdot \nabla \phi(x), \quad \forall \phi \in C_b^1, \tag{1.14}$$

and we can write the generator in the concise form $\mathcal{A} = \sum_{i=1}^d b_i \partial_i = b \cdot \nabla$.

Pure-jump processes

We model the random Markov behavior of particles with piecewise constant sample paths, with non-accumulating jump-times, defined by the sequence $x_0 = x, x_1, \dots, x_n, \dots$, of points visited in succession,

and $t_0, t_1, \dots, t_n, \dots$, of respective sojourn times in these states. The processes jump from point to point in \mathbb{R}^d , making successive independent random choices for the sojourn times and the jump destinations.

It is obvious from the Markov property that the sojourn times at any x in \mathbb{R} must have a law depending only on x which is **memoryless**, in the following sense: a random variable T is memoryless if and only if

$$P(T > t + s | T > s) = P(T > t).$$

There are two degenerate cases $T = 0$ a.s. and $T = \infty$ a.s., else there is some $\lambda \in]0, +\infty[$ such that

$$P(T > t) = e^{-\lambda t}, \quad t \geq 0; \quad \mathcal{L}_T(dt) = \lambda \mathbb{1}_{t>0} e^{-\lambda t} dt,$$

which is a law called the **exponential law of parameter λ** and denoted by $\mathcal{E}(\lambda)$. The expected value $E(T)$ is $1/\lambda$, and the degenerate cases correspond to the weak limits $\lambda = \infty$ and $\lambda = 0$ respectively.

The most famous pure-jump process is the Poisson process: the inter-arrival jump times are independent exponential laws with parameter λ and the amplitude of jumps are all equal to 1. Let us define a Poisson process of rate λ : given a sequence of independent $\mathcal{E}(\lambda)$ variables $(S_n)_{n \geq 1}$, setting

$$T_0 = 0, \quad T_n = S_1 + \dots + S_n, \quad n \geq 1; \quad N_t = \sup\{n \geq 0 : T_n \leq t\} = \sum_n \mathbb{1}_{T_n \leq t}, \quad t \geq 0,$$

defines the **Poisson process $(N_t)_{t \geq 0}$ of rate λ** . Giving the counting process $(N_t)_{t \geq 0}$, its sequence of arrival times $(T_n)_{n \geq 1}$, or its sequence of inter-arrival times $(S_n)_{n \geq 1}$, are alternative descriptions of the same random model. The Poisson process $(N_t)_{t \geq 0}$ has independent stationary increments, hence it is a Markov process, and its law is determined by the law of its increments

$$P(N_{s+t} - N_s = k | \mathcal{F}_s^N) = P(N_t = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad t, s \geq 0. \quad (1.15)$$

This law is called the **Poisson law of parameter λt** .

Now, we consider more general pure-jump processes, for which the rate of jumps and the amplitude of jumps depend on the position of the particles. For each $x \in \mathbb{R}^d$, in the state x , after waiting an exponential time of parameter $\lambda(x)$ the process must choose its next state y according to a probability measure $\pi(x, dy)$. The collection of the transition rates $\lambda(x)$ and transition kernels $\pi(x, dy)$, for x in \mathbb{R}^d , is unique if we further ask that $\pi(x, \{x\}) = 0$ (up to indeterminacy of $\pi(x, dy)$ for $\lambda(x) = 0$), in which case y is different from x , a.s. This collection characterizes the Markov process.

We describe how to actually construct samples of the Markov process, i.e., the paths $(Y_t(\omega))_{t \geq 0}$ followed at random. Each step is performed independently of the past.

- When the process is at an x in \mathbb{R}^d , it stays there for a duration distributed according to $\mathcal{E}(\lambda(x))$.
- When that time has elapsed, we choose a location y for the process according to $\pi(x, dy)$.
- We start over the procedure.

A sufficient condition ensuring that the jump times do not accumulate is

$$\sup_{x \in \mathbb{R}^d} \lambda(x) \leq \lambda^* < \infty$$

in which case we may give a somewhat simpler construction. Each step is performed independently of the past.

- Regardless of the location x of the process, it stays there for a duration distributed according to $\mathcal{E}(\lambda^*)$.
- When that time has elapsed, with probability $1 - \lambda(x)/\lambda^*$, the process stays at x , and with probability $\lambda(x)/\lambda^*$, it takes a value chosen according to $\pi(x, dy)$.
- We start over the procedure.

The succession of times at which we must make a decision about jumps defines a Poisson process of rate λ^* , and at these times the probability kernel ruling the actual choice of the arrival point y starting from point x is

$$\pi^*(x, dy) = \frac{\lambda(x)}{\lambda^*} \pi(x, dy) + \left(1 - \frac{\lambda(x)}{\lambda^*}\right) \delta_x(dy). \quad (1.16)$$

Let us compute the generator for the Markov process in terms of $\lambda(x)$ and $\pi(x)$, x in \mathbb{R}^d , according to (1.8). Starting at a point x in \mathbb{R}^d , let T denote the first time of jump, with law $\mathcal{E}(\lambda(x))$. Let us assume $\sup_x \lambda(x) \leq \lambda^* < \infty$. For ϕ in L^∞ ,

$$E_x(\phi(X_t)) = E_x(\phi(X_t) \mathbb{1}_{T>t}) + E_x(\phi(X_t) \mathbb{1}_{T \leq t}) = e^{-\lambda(x)t} \phi(x) + E_x(\phi(X_t) \mathbb{1}_{T \leq t}),$$

and the probability that there be two jumps or more is

$$P(T + T' \leq t) \leq P(T \leq t)P(T' \leq t) \leq (1 - e^{-\lambda(x)t})(1 - e^{-\lambda^*t}) = O(t^2)$$

where T' is a random variable independent from T which represents the sojourn time in the state reached after the first jump, thus

$$\begin{aligned} E_x(\phi(X_t) \mathbb{1}_{T \leq t}) &= (1 - e^{-\lambda(x)t}) \int \phi(y) \pi(x, dy) + O(t^2), \\ \frac{E_x(\phi(X_t)) - \phi(x)}{t} &= \frac{1 - e^{-\lambda(x)t}}{t} \left(\int \phi(y) \pi(x, dy) - \phi(x) \right) + O(t), \end{aligned}$$

from which we deduce that for $\sup_x \lambda(x) \leq \lambda^* < \infty$ and π^* defined in (1.16), we have

$$\mathcal{A}\phi(x) = \lambda(x) \int \{\phi(y) - \phi(x)\} \pi(x, dy) = \lambda^* \int \{\phi(y) - \phi(x)\} \pi^*(x, dy), \quad \phi \in L^\infty. \quad (1.17)$$

We define the **jump measure** $\mu(x, dy) = \lambda(x)\pi(x, dy)$, which is well-defined up to arbitrary mass at x . We have $\lambda(x) = |\mu(x)|$ and $\pi(x, dy) = \mu(x, dy)/|\mu(x)|$, and

$$\mathcal{A}\phi(x) = \int \{\phi(y) - \phi(x)\} \mu(x, dy) = \int \phi(y) \mu(x, dy) - |\mu(x)|\phi(x), \quad \phi \in L^\infty. \quad (1.18)$$

Mixture of differential equation flows and jumps

Given a vector field $b : x \in \mathbb{R}^d \mapsto b(x) \in \mathbb{R}^d$ such that there exists a unique solution to the differential equation $\dot{y}_t = b(y_t)$, and a collection $\lambda(x) \geq 0$ and $\pi(x) \in \mathcal{P}(\mathbb{R}^d)$, for x in \mathbb{R}^d , such that $\sup_x \lambda(x) \leq \lambda^* < \infty$, we construct a process $(Y_t)_{t \geq 0}$ as follows (all choices are independent):

- We have constructed a sample $(Y_t(\omega))_{0 \leq t \leq s}$ up to time $s \geq 0$. Let $x = Y_s(\omega)$.
- We choose a duration h according to $\mathcal{E}(\lambda^*)$. We compute $(X_u^x)_{0 \leq u \leq h}$ according to the deterministic flow (1.13), and set $(Y_t(\omega))_{s \leq t < s+h} = (X_u^x)_{0 \leq u < h}$. We have $Y_{s+h-}(\omega) = \lim_{t \uparrow s+h} Y_t(\omega) = X_h^x$, denoted by y .
- With probability $1 - \lambda(y)/\lambda^*$ we set $Y_{s+h}(\omega) = Y_{s+h-}(\omega) = y$, and with probability $\lambda(y)/\lambda^*$, the process jumps and we set $Y_{s+h}(\omega) = z$ for z chosen according to $\pi(y, dz)$. Thus, the value of $Y_{s+h}(\omega)$ is chosen according to $\pi^*(y)$ in (1.16).
- We start over the procedure at time $s + h$ with the sample $(Y_t(\omega))_{0 \leq t \leq s+h}$.

We check as above that this defines a Markov process with generator

$$\mathcal{A}\phi(x) = b(x) \cdot \nabla \phi(x) + \int \{\phi(y) - \phi(x)\} \mu(x, dy), \quad \phi \in C_b^1. \quad (1.19)$$

1.4.6. A class of integro-differential operators and sample path evolutions

We wish to extend the class of operators we consider. Let us consider more general jump measure $\mu(x, dy)$ for which $|\mu(x)| = \int \mu(x, dy) = \infty$. Then (1.19) is well defined as soon as $\int |y - x| \mu(x, dy) < \infty$.

Assuming then that $\int |y - x| \mu(x, dy) < \infty$, for x in \mathbb{R}^d , we define the mean jump amplitude

$$d(x) = \int (y - x) \mu(x, dy) \in \mathbb{R}^d, \quad x \in \mathbb{R}^d.$$

Then we may rewrite (1.19) as

$$\mathcal{A}\phi(x) = (b(x) + d(x)) \cdot \nabla\phi(x) + \int \{\phi(y) - \phi(x) - (y - x) \cdot \nabla\phi(x)\} \mu(x, dy), \quad \phi \in C_b^1. \quad (1.20)$$

We have isolated a **drift** term $\tilde{b} = b + d$ due to the differential equation flow and the mean jump amplitude, and written the integral operator in a **centered** form.

Moreover, if we assume now that μ is such that $\int |y - x|^2 \mu(x, dy) < \infty$ but $\int |y - x| \mu(x, dy) = \infty$, it becomes natural to consider directly an operator of the form (for simplicity of notations, we set b instead of \tilde{b})

$$\mathcal{A}\phi(x) = b(x) \cdot \nabla\phi(x) + \int \{\phi(y) - \phi(x) - (y - x) \cdot \nabla\phi(x)\} \mu(x, dy), \quad \phi \in C_b^2. \quad (1.21)$$

Diffusion Approximations

Assume now that we have a sequence of such operators

$$\mathcal{A}^n\phi(x) = b(x) \cdot \nabla\phi(x) + \int \{\phi(y) - \phi(x) - (y - x) \cdot \nabla\phi(x)\} \mu^n(x, dy), \quad \phi \in C_b^2, \quad (1.22)$$

such that

$$a_{ij}^n(x) = \int (y - x)_i (y - x)_j \mu^n(x, dy), \quad \forall n \geq 1, x \in \mathbb{R}^d,$$

is well-defined and converges to a non-negative symmetric matrix $(a_{ij})_{1 \leq i, j \leq d}$. Assume moreover that $(\mu^n)_{n \geq 1}$ concentrates on small jumps in a suitable way. It is tempting to pass to the limit formally in (1.22), yielding

$$\mathcal{A}^\infty\phi(x) = b(x) \cdot \nabla\phi(x) + \frac{1}{2} \sum_{i, j=1}^d a_{ij}(x) \partial_{ij}^2 \phi(x), \quad \phi \in C_b^2. \quad (1.23)$$

Under reasonable assumptions, Markov processes with such operators can be constructed and this limit can be justified. This is the so-called **diffusion approximation**, leading to what physicists refer to as **Fokker-Planck equations**, namely

$$\partial_t P_t = (\mathcal{A}^\infty)^* P_t.$$

These Markov processes have **continuous** paths, corresponding to the **local** nature of their generators.

The Brownian motion

If $b = 0$ and a is the identity matrix, we obtain the famous d -dimensional **Brownian motion** process, with generator $\frac{1}{2}\Delta$. Let us give a definition of the one-dimensional Brownian motion.

Definition 1.16. A one-dimensional Brownian motion is a continuous process $(B_t)_{t \geq 0}$ with independent stationary increments, which is a Markov process with generator $\frac{1}{2} \frac{d^2}{dx^2}$. It is such that

$$\mathcal{L}(B_{s+t} - B_s | \mathcal{F}_s^B) = \mathcal{L}(B_t) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx, \quad t, s \geq 0. \quad (1.24)$$

A d -dimensional Brownian motion $B = (B_t^1, \dots, B_t^d)$ is a \mathbb{R}^d -valued process whose coordinates are independent one-dimensional Brownian motions.

Let us now assume that $b(x)$ and the non-negative symmetric matrix $(a_{ij}(x))_{1 \leq i, j \leq d}$ do not depend on x . Let σ be a matrix such that $\sigma \sigma^* = (a_{ij})_{1 \leq i, j \leq d}$, and $B = (B_t^1, \dots, B_t^d)$ be a d dimensional Brownian motion. It is easy to check that the processes

$$Y_t^i = x_i + b_i t + \sum_{j=1}^d \sigma_{ij} B_t^j, \quad 1 \leq i \leq d,$$

defines a Markov process with generator (1.23) starting at x . In vector notation $Y_t = x + bt + \sigma B_t$.

When $b(x)$ and $(a_{ij}(x))_{1 \leq i, j \leq d}$ depend on x , such an intuitive construction is formally true locally, and can often be made rigorous. The notion of **stochastic differential equations** makes sense of the expression $dY_t = b(Y_t)dt + \sigma(Y_t)dB_t$, and can be used to prove existence of the corresponding Markov process under appropriate regularity assumptions.

A general class of integro-differential operators

We now consider a type of operator susceptible to be the generator of a Markov process, inspired by all of the above. Let b be a vector field, $a = \sigma \sigma^*$ a symmetric positive matrix field, μ a positive measure kernel, and

$$\mathcal{A}\phi(x) = b(x) \cdot \nabla \phi(x) + \frac{1}{2} \sum_{i, j=1}^d a_{ij}(x) \partial_{ij} \phi(x) + \int \{\phi(y) - \phi(x)\} \mu(x, dy), \quad \phi \in C_b^2. \quad (1.25)$$

This expression makes sense only if $\int \min(|y - x|, 1) \mu(x, dy) < \infty$.

If we wish to weaken this assumption to $\int \min(|y - x|^2, 1) \mu(x, dy) < \infty$, we further generalize the form of the operator by centering the part of the jump term corresponding to jumps of amplitude less than 1 as in (1.21).

It can be proven that all Markov processes with sample paths in $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ are of this form.

From a probabilistic perspective, the operator (1.25) prescribes the “instantaneous local random evolution” expected from sample paths of the Markov process, which should at time t and location x :

- Drift continuously according to the vector $b(x)$
- Diffuse around continuously according to a Brownian motion weighted by the matrix $\sigma(x)$
- Jump around according to the jump measure $\mu(x)$.

We may prove by probabilistic means that there is indeed a well-defined Markov process satisfying globally this evolution, under appropriate assumptions on a , b , and μ . Theorem 1.15 gives us a methodology for this purpose, but there exist many other constructive methods. This constructive point of view is in many ways intuitive, and is very close to the physical evolution from which the probabilistic model was drawn.

Except in very special cases such as (1.15) and (1.24) there are no explicit formulæ, and little quantitative and even qualitative information is available. We shall later see that approximations of P_x , $x \in \mathbb{R}^d$, can be derived from this sample path construction.

Let us see the relation between the Markov generator (1.25) and partial differential equations.

We compute the formal adjoint of \mathcal{A} in (1.25) using integration by parts and the Fubini formula. We have

$$A^*p(dx) = - \sum_{i=1}^d \partial_i(b_i(x)p(dx)) + \frac{1}{2} \sum_{i,j=1}^d \partial_{ij}(a_{ij}(x)p(dx)) + \int \mu(z, dx) p(dz) - |\mu(x)|p(dx), \quad p \in \mathcal{P}(\mathbb{R}^d). \quad (1.26)$$

This gives us a large class of linear evolution equations $\partial_t f = A^* f$ which we may try to solve and study, see (1.10), (1.11), and the following section. These equations may correspond to so-called **transport equations** when $a = 0$, or **Fokker-Planck equations** when $\mu = 0$ and a is sufficiently non-degenerate.

A main aspect to remark is that in the integro-differential evolution equation, the integral terms comes from an underlying jump process.

1.5. Laws of large numbers and confidence intervals

We are interested in P_x and $(P_{x,t})_{t \geq 0}$, for x in \mathbb{R}^d . Given a Markov generator such as (1.25), we can, at least theoretically, construct a sample path $(Y_t)_{t \geq 0}(\omega)$ for the process, corresponding to the result of a random draw of ω according to P_x . We can reproduce this procedure independently to obtain sample paths of independent copies of this process. The following results tell us how to use this to obtain a random approximation of P_x and evaluate the precision attained.

1.5.1. Laws of large numbers

The Law of Large Numbers theorem is the first most famous theorem in probability theory and is the basis for all statistics, or for all Monte-Carlo approaches.

Theorem 1.17. (*(Strong) Law of large numbers.*) *Let $(X_i)_{i \geq 1}$ be a sequence of independent identically distributed random variables. Assume that these random variables are \mathbb{R}^d -valued and L^1 ; in particular, $E(X_i)$ exists and does not depend on $i \geq 1$. Then,*

$$\lim_{n \rightarrow \infty} \frac{X_1 + \cdots + X_n}{n}(\omega) = E(X_1), \quad \text{for any } \omega \in \Omega, \text{ a.s.}$$

Let $(X_i)_{i \geq 1}$ be a sequence of independent identically distributed random variables on a Polish space E , with common law Q . If $\phi : E \rightarrow \mathbb{R}$ is any L^∞ function, then $(\phi(X_i))_{i \geq 1}$ is a sequence of independent identically distributed random variables, which are real-valued and bounded, in particular L^1 , and Theorem 1.17 yields

$$\lim_{n \rightarrow \infty} \frac{\phi(X_1) + \cdots + \phi(X_n)}{n}(\omega) = E(\phi(X_1)) = \langle Q, \phi \rangle, \quad \text{for any } \omega \in \Omega, \text{ a.s.}$$

The negligible set depends on ϕ . We define the **empirical measure** of (X_1, \dots, X_n) ,

$$\frac{1}{n} \sum_{i=1}^n \delta_{X_i} \quad \text{with samples in } \mathcal{P}(E), \quad \left\langle \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \phi \right\rangle = \frac{\phi(X_1) + \cdots + \phi(X_n)}{n}.$$

Weak convergence of probability measures on a Polish space has countable convergence-determining sets of test functions, and we obtain a very important result, which we could call the “fundamental theorem in statistics”.

Theorem 1.18. *Let $(X_i)_{i \geq 1}$ be a sequence of independent identically distributed random variables on a Polish space E , with common law Q . Then, for weak convergence on $\mathcal{P}(E)$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \delta_{X_i}(\omega) = Q, \quad \text{for any } \omega \in \Omega, \text{ a.s.}$$

This result gives us the key to the approximation of the Markov laws P_x , x in \mathbb{R}^d , given some generator. We construct sample paths $X_1(\omega), \dots, X_n(\omega)$ in $E = \mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ of independent copies of the process starting at x , and approximate P_x by the sample of the empirical measure

$$\frac{1}{n} \sum_{i=1}^n \delta_{X_i}(\omega) = \frac{1}{n} \sum_{i=1}^n \delta_{X_i(\omega)}.$$

If we are interested in $(P_{x,t})_{t \geq 0}$, general results imply generally that

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n \delta_{X_i(t)(\omega)} \right)_{t \geq 0} = (P_{x,t})_{t \geq 0}$$

in $\mathbb{D}(\mathbb{R}_+, \mathcal{P}(\mathbb{R}^d))$ for the weak convergence on $\mathcal{P}(\mathbb{R}^d)$ uniformly for compact sets of \mathbb{R}_+ .

1.5.2. Confidence intervals

We now evaluate the precision attained, as a function of the number of samples taken.

Theorem 1.19. (*Weak law of large numbers.*) *Let $(X_i)_{i \geq 1}$ be a sequence of independent identically distributed random variables. Assume that these random variables are real-valued and L^2 , and set $\sigma^2 = E((X_i - E(X_i))^2) = E(X_i^2) - E(X_i)^2$. Then, we have L^2 -norm convergence with precise rate*

$$\left\| \frac{X_1 + \dots + X_n}{n} - E(X_1) \right\|_2 = E \left(\left(\frac{X_1 + \dots + X_n}{n} - E(X_1) \right)^2 \right)^{1/2} = \frac{\sigma}{\sqrt{n}}, \quad \forall n \geq 1,$$

from which we deduce the Bienaymé-Chebichev bound

$$P \left(\left| \frac{X_1 + \dots + X_n}{n} - E(X_1) \right| \geq \varepsilon \right) \leq \frac{\sigma^2}{\varepsilon^2 n}, \quad \forall \varepsilon > 0, n \geq 1. \quad (1.27)$$

The L^2 -norm and Bienaymé-Chebichev estimates point to a $1/\sqrt{n}$ rate of convergence, as does the following precise asymptotic rate. These results can be easily extended to \mathbb{R}^d , and the $1/\sqrt{n}$ rate does not depend on the dimension of space.

The following result is the second most famous theorem in probability theory. It gives the rate of convergence in all standard statistics or Monte-Carlo approximations, whatever the dimension of the space of values.

Theorem 1.20. (*Central limit theorem.*) *Let $(X_i)_{i \geq 1}$ be a sequence of independent identically distributed random variables. Assume that these random variables are real-valued and L^2 , and let $\sigma^2 = E((X_i - E(X_i))^2) = E(X_i^2) - E(X_i)^2$. Then, for weak convergence for probability measures,*

$$\lim_{n \rightarrow \infty} \mathcal{L} \left(\frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \dots + X_n}{n} - E(X_1) \right) \right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx,$$

and equivalently

$$\lim_{n \rightarrow \infty} P \left(a \leq \frac{\sqrt{n}}{\sigma} \left(\frac{X_1 + \dots + X_n}{n} - E(X_1) \right) \leq b \right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx, \quad \forall a \leq b.$$

One says that $\frac{\sqrt{n}}{\sigma} (X_1 + \dots + X_n - E(X_1))$, $n \geq 0$, converges in law to the normal law $\mathcal{N}(0, 1)$.

The relation between this tight asymptotic estimate of convergence and the looser exact Bienaymé-Chebichev bound (1.27) is simple, and best explained using the notion of **confidence interval**.

Assume we use the law of large numbers in order to estimate the unknown mean $m = E(X_1)$. We set $\hat{m}_n = (X_1 + \dots + X_n)/n$, which is the (random) estimator for m , and $\lim_{n \rightarrow \infty} \hat{m}_n = m$, a.s. Unfortunately we cannot go on for ever, and we seek a value n^* such that \hat{m}_{n^*} is a “good” estimate of m , and wish to quantify this fact. Setting a confidence interval means stating that “with probability of being mistaken less than p , the estimate \hat{m}_{n^*} is in the range $]m - \varepsilon, m + \varepsilon[$ ”.

There is an interplay between the probabilistic **confidence level** p , the **precision** ε , and the **size** n^* :

- The rigorous Bienaymé-Chebichev bound (1.27) yields

$$p = \frac{\sigma^2}{\varepsilon^2 n^*} \iff \varepsilon = \frac{\sigma}{\sqrt{p} \sqrt{n^*}} \iff n^* = \frac{\sigma^2}{\varepsilon^2 p}.$$

- The precise central-limit theorem yields

$$\lim_{n \rightarrow \infty} P \left(\frac{\sqrt{n}}{\sigma} \left| \frac{X_1 + \dots + X_n}{n} - E(X_1) \right| \geq a \right) = \frac{2}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{x^2}{2}} dx, \quad \forall a \geq 0,$$

and we can decide to use this asymptotic formula as if it were actually true for finite n (there are rates of convergence for the Central Limit Theorem, and usually the sequence is fairly close to its limit for n greater than 50 or so). Then

$$p = \frac{2}{\sqrt{2\pi}} \int_{a(p)}^\infty e^{-\frac{x^2}{2}} dx$$

gives a bijective relation between p and $a(p)$, and

$$a(p) = \frac{\varepsilon \sqrt{n^*}}{\sigma} \iff \varepsilon = \frac{\sigma a(p)}{\sqrt{n^*}} \iff n^* = \frac{\sigma^2 a(p)^2}{\varepsilon^2}.$$

The Bienaymé-Chebichev confidence intervals are pessimistic but hold unconditionally. The asymptotic confidence intervals obtained using the central limit theorem are much tighter than the Bienaymé-Chebichev ones: if p is close to 0, then $a(p)$ is much smaller than $1/\sqrt{p}$, and ε and n^* are respectively proportional to these terms and to the squares of these terms. As examples, for $p = 5\%$ we have approximately $a(p) = 1.96$ and $1/\sqrt{p} = 4.47$, for $p = 1\%$, $a(p) = 2.58$ and $1/\sqrt{p} = 10$. The central limit theorem confidence intervals are often used in practice, and are exact for Gaussian variables.

Ideally, we determine first the level of probability of being mistaken $p \in]0, 1[$ which we are ready to tolerate. It is small, seldom more than 0.05, and often in the few percentage or fraction of percentage level. We then set the precision ε , which should again be small in terms of the order of magnitude of $E(X_1)$. We then obtain the required number of samples n^* from this formula. In practice we are often limited by n^* , and compute the other quantities in terms of this. Note that we must estimate or at least bound σ , which is likely to be unknown.

1.6. Simulation of random variables and processes

The remaining practical problem is how to compute actual sample paths. We use a computer, and we cannot actually introduce randomness in this setting. In practice, pseudo-random number generators are used.

1.6.1. Pseudo-random number generators

The basic pseudo-random number generator is a function such as `drand48` that can be invoked by a program. Successive calls to this function furnish us with a sequence x_1, x_2, \dots , of real numbers between 0 and 1. This sequence should satisfy the probabilistic laws (laws of large numbers, central limit theorem ...) with sufficient precision, so that we can pretend that x_1, x_2, \dots , is a sample $U_1(\omega), U_2(\omega), \dots$, where $(U_i)_{i \geq 1}$ is a sequence of independent uniform random variables (with law the Lebesgue measure on $[0, 1]$).

Number theoretical results and statistical tests are used to obtain and validate such functions. In general, these computer functions use two mathematical functions g and h . A number y_0 , called the **seed**, is furnished to the function and stored in a memory. After the $(n - 1)$ -th call, the number y_{n-1} is stored in the memory, the n -th call to the function computes $y_n = g(y_{n-1})$, gives us $x_n = h(y_n)$, and stores y_n in place of y_{n-1} .

Most pseudo-random number generators, such as **drand48**, are linear congruential: the y_n are integer,

$$y_n = ay_{n-1} + b \text{ mod } c$$

for well-chosen integers a , b , and c , and $x_n = y_n/c$. This sequence is perfectly deterministic. It is cyclical, and the actual sequence observed depends only on the initial value y_0 . We should introduce y_0 , and this initialization must be done **once and only once before** the first call to the function. For **drand48**, the seed y_0 is specified using **srand48**, and in order to obtain a seemingly “random” initialization, the seed is often given by the time in seconds in the internal clock of the computer. For instance, the short **c++** program

```
#include <cstdlib>
#include <cmath>
#include <time.h>
main
{
    srand48(time(0));
    for(int i = 0; i <$ 100; i++)
        drand48();
}
```

sets the seed at the time of the internal clock, and then furnishes successively 100 numbers, which could be used as a sample $U_1(\omega)$, $U_2(\omega)$, \dots , $U_{100}(\omega)$.

1.6.2. Simulation of random variables on \mathbb{R} and \mathbb{R}^d

Given some law Q on \mathbb{R} , \mathbb{R}^d , or more generally on a Polish space E , we wish to construct a sequence $(X_i)_{i \geq 1}$ of independent random variables with law Q , starting from a sequence $(U_i)_{i \geq 1}$ of independent uniform random variables. The following result states that you can obtain any law on \mathbb{R} as the image law of the uniform law on $[0, 1]$, and gives us a theoretical way to compute the function. This implies in theory that any law on a Polish space, which is (if infinite) Borel isomorphic to $[0, 1]$ with the Borel σ -field, is the image law of the uniform law, but this is useless in practice. Even on \mathbb{R} , we should try to use as constructive an approach as possible.

Theorem 1.21. *Let Q be a probability measure on \mathbb{R} , and*

$$F : x \in \mathbb{R} \mapsto F(x) = Q(]-\infty, x]) \in [0, 1]$$

*be its **cumulative distribution function**. Then F is non-decreasing and right-continuous, and $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$. We denote by G its left inverse, defined by*

$$G : u \in [0, 1] \mapsto G(u) = \inf \{x : F(x) \geq u\} \in \mathbb{R},$$

which is non-decreasing and left-continuous. Then if U is a uniform random variable, setting $X(\omega) = G(U(\omega))$ for ω in Ω defines a random variable X of law Q .

Proof. If $x \leq y$ then $F(x) = Q(]-\infty, x]) \leq Q(]-\infty, y]) = F(y)$. If $(x_n)_{n \geq 0}$ is a sequence decreasing to x , then $]-\infty, x] = \bigcap_{n \geq 0}]-\infty, x_n]$ and from general measure-theoretic principles $F(x_n) = Q(]-\infty, x_n])$ decreases to $F(x) = Q(]-\infty, x])$. Similarly, $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow +\infty} F(x) = 1$.

Clearly if $F(x) \geq u$ then $G(u) \leq x$. If $G(u) \leq x$ then $F(y) \geq u$ for all $y > x$ and since F is right-continuous then $F(x) \geq u$. Thus

$$G(u) \leq x \iff u \leq F(x)$$

and hence $P(X \leq x) = P(G(U) \leq x) = P(U \leq F(x)) = F(x)$. We conclude, since a law is characterized by its cumulative distribution function. \square

If F is bijective, then G is the inverse F^{-1} . In general G is not easy to compute. A simple case in which we may do so is of great importance for pure-jump Markov processes.

Lemma 1.22. *If U is uniform on $[0, 1]$ and $\lambda > 0$, then $-\ln(U)/\lambda$ is an exponential variable of parameter λ .*

Proof. For the exponential law $\mathcal{E}(\lambda)$, the cumulative distribution function F is given by

$$F(x) = 0, \quad \forall x \leq 0; \quad F(x) = 1 - e^{-\lambda x}, \quad \forall x \geq 0.$$

A simple computation yields $G(u) = -\ln(1-u)/\lambda$. Since $1-U$ and U have the same law, the result follows. \square

A simple result concerns random variables with a countable set of values. In theory we can approximate arbitrarily well any random variable on a Polish space by such a variable.

Lemma 1.23. *Let $p_n \geq 0$, for n in \mathbb{N} , sum up to unity, and let values x_n , for n in \mathbb{N} , be given. Let U be uniform on $[0, 1]$. Then we obtain a random variable X such that $P(X = x_n) = p_n$ for n in \mathbb{N} by setting*

$$X(\omega) = x_n \iff \sum_{i=0}^{n-1} p_i \leq U(\omega) < \sum_{i=0}^n p_i, \quad \forall \omega \in \Omega, \quad n \in \mathbb{N}.$$

Simulating random variables efficiently is almost an art. Many laws are simulated on a case-to-case basis. As an example, we give a method for obtaining a pair of independent Gaussian random variables, centered and of variance 1, from a pair of independent uniform variables (caution about the independence properties of the pseudo-random number generator may lead us to abandon one of these Gaussian variables).

Lemma 1.24. *If U and V are independent uniformly distributed random variables on $[0, 1]$, then*

$$X = \sqrt{-2 \ln U} \cos 2\pi V, \quad Y = \sqrt{-2 \ln U} \sin 2\pi V,$$

form a pair of independent random variables with law $\mathcal{N}(0, 1)$.

Proof. Let (X, Y) be a pair of independent random variables with law $\mathcal{N}(0, 1)$. Then for any bounded function ϕ , using the polar coordinates,

$$\begin{aligned} E(\phi(X, Y)) &= \frac{1}{2\pi} \int \phi(x, y) e^{-\frac{x^2+y^2}{2}} dx dy = \frac{1}{2\pi} \int_{\substack{0 < r < \infty \\ 0 < \theta < 2\pi}} \phi(r \cos \theta, r \sin \theta) e^{-\frac{r^2}{2}} r dr d\theta \\ &= \int \phi(\sqrt{a} \cos \theta, \sqrt{a} \sin \theta) \frac{1}{2} \mathbb{I}_{a>0} e^{-\frac{a}{2}} da \frac{1}{2\pi} \mathbb{I}_{0 < \theta < 2\pi} d\theta \end{aligned}$$

and we recognize that if A is a $\mathcal{E}(1/2)$ random variable and if Θ is independent and uniform on $[0, 2\pi]$ then $(\sqrt{A} \cos \Theta, \sqrt{A} \sin \Theta)$ has same law as (X, Y) . \square

We finish with the **rejection method**. It is a rather general way of simulating \mathbb{R}^d -valued random variables. We assume these random variables have a density, but the method can be extended.

We wish to simulate a sequence of random variables with law $g(x) dx$. We find an easily simulated law $h(x) dx$ such that $g \leq \alpha h$ for some not too large $\alpha > 0$. We draw an independent couple of values (y, u) according to $h(x) dx$ and the uniform law. The drawing is rejected, and we repeat the random draw independently, except when (y, u) is such that $\alpha h(y)u \leq g(y)$, in which case y is used as a drawing according to $g(x) dx$. This whole procedure can be repeated independently to obtain an i.i.d. sequence of law $g(x) dx$. The mathematical result can be stated as follows.

Theorem 1.25. *Let g and h be two probability densities on \mathbb{R}^d . Assume that there is some real number α (necessarily greater than 1) such that $g(x) \leq \alpha h(x)$, for all x in \mathbb{R}^d . Let $(Y_i)_{i \geq 1}$ be a sequence of independent random variables of law $h(y) dy$, $(U_i)_{i \geq 1}$ be an independent sequence of independent random variables uniformly distributed on $[0, 1]$. Let*

$$N = \inf\{i \geq 1 : \alpha h(Y_i) U_i \leq g(Y_i)\}, \quad X = Y_N.$$

Then X and N are independent, X has law $g(x) dx$, and N has the geometric law of mean α given by

$$P(N = k) = \frac{1}{\alpha} \left(1 - \frac{1}{\alpha}\right)^{k-1}, \quad k \geq 1.$$

Proof. For ϕ in L^∞ ,

$$\begin{aligned} E(\phi(N, Y_N)) &= \sum_{k \geq 1} E\left(\phi(k, Y_k) \mathbb{I}_{\alpha h(Y_k) U_k \leq g(Y_k)} \prod_{i=1}^{k-1} \mathbb{I}_{\alpha h(Y_i) U_i > g(Y_i)}\right) \\ &= \sum_{k \geq 1} \int_0^1 \int \phi(k, y) \mathbb{I}_{\alpha h(y) u \leq g(y)} h(y) dy du \left(\int_0^1 \int \mathbb{I}_{\alpha h(y) u > g(y)} h(y) dy du\right)^{k-1} \\ &= \sum_{k \geq 1} \int \phi(k, y) \frac{g(y)}{\alpha h(y)} h(y) dy \left(\int \left(1 - \frac{g(y)}{\alpha h(y)}\right) h(y) dy\right)^{k-1} \\ &= \sum_{k \geq 1} \int \phi(k, y) g(y) dy \frac{1}{\alpha} \left(1 - \frac{1}{\alpha}\right)^{k-1} \end{aligned}$$

and we easily conclude. □

1.6.3. Simulation of Markov processes

We are interested in the effective simulation of Markov processes with sample paths in $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ and integro-differential generators of the form (1.25). This is done by applying the “constructive” description of sample paths given in Sections 1.4.5 and 1.4.6. We describe a way to simulate the sample paths on $[0, T]$. We set $\lambda(x) = |\mu(x)|$ for x in \mathbb{R}^d and assume $\sup_x \lambda(x) \leq \lambda^*$.

- We have simulated a sample path on $[0, s]$, with $s < T$, and the value at time s is x in \mathbb{R}^d . We set $h = -\ln(u)/\lambda^*$, where u is given by the pseudo-random number generator, and simulates a sample $U(\omega)$ of an uniform random variable. We set $t = \min(s + h, T)$.
- We simulate a (continuous) sample path on $[0, t - s]$ corresponding to the differential part of (1.25), which we concatenate with the previous sample path to obtain a sample path on $[0, t]$. This is deterministic if and only if $a = 0$, in which case it requires solving the differential equation (1.13) in order to obtain $(X_u^x)_{0 \leq u \leq t-s}$. If $a \neq 0$, this is random, and stochastic differential equations driven by Brownian motion should be solved. This step may require approximations and time discretization (like for example Euler’s schemes of approximations).
- We must attribute a value at time t for the sample path. If $s + h > T$, we use the value y reached in the previous step. If $s + h \leq T$, we obtain v from the pseudo-random number generator, simulating a sample $U(\omega)$ of an uniform random variable. If $v > \lambda(y)/\lambda^*$, we use the value y reached in the previous step; if $v \leq \lambda(y)/\lambda^*$, we draw a value z according to $\pi(y, dz)$, usually using the rejection method.
- We now have simulated a sample path on $[0, t]$. If $t < T$, we start over the procedure.

2. THE BOLTZMANN EQUATION AND ITS PROBABILISTIC INTERPRETATION

Using the probabilistic tools we have introduced, we develop a probabilistic approach for the study of the Boltzmann equation and for the approximation of its solutions.

2.1. The Boltzmann equation

Let us consider a gas with a unique kind of particles. The mean free path λ of a particle is the average of the distance covered without collisions. A simple computation gives that at one atmosphere of pressure and temperature 0°C , this length λ is very small with respect to the size L of some objects of study (planes, space shuttles...). But at high levels of atmosphere, the gas is dilute and the mean free path increases. So one has to take differently into account the collisions. There are approximatively three cases:

1. high altitude: $\lambda \gg L$, one can neglect the collisions and one obtains a free transport equation
2. low altitude: then λ is so small before L that it is considered to tend to 0. The equation tends to be a fluid equation
3. middle altitude: $\lambda \approx L$.

These three zones give thus very different equations for the quantities related to the gas. We shall restrict our attention to the kinetic description of the middle situation. More precisely, we study the quantity $f(t, x, v)$ which is the density of this rarefied gas at time t , position x and velocity v . It will always satisfy

$$f(t, x, v) \geq 0, \quad \int f(t, x, v) dx dv = 1.$$

The equation satisfied by f in this middle stage is called the Boltzmann equation. It is the kinetic equation

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = Q(f, f); \quad f(0, x, v) = f_0(x, v), \quad (2.1)$$

understood in a weak sense. The function f_0 is a density function on \mathbb{R}^6 . The kernel Q is a quadratic collision kernel acting only on the variable v , preserving momentum and kinetic energy, of the form

$$Q(f, f)(t, x, v) = \int_{S^2} \int_{\mathbb{R}^3} \left(f(t, x, v^*) f(t, x, w^*) - f(t, x, v) f(t, x, w) \right) B(v - w, \nu) dw d\nu \quad (2.2)$$

where v^* and w^* represent the post-collisional velocities of two particles of velocities before collision v and w having collided in a position x in which their centers are on a line of direction given by the unit vector ν . There is a statistical knowledge of the geometry of the impact given by ν . The cross-section $B(v - w, \nu)$ quantifies the likelihood of interactions at point x of particles of velocities v and w with impact parameter ν and depends only on $|v - w|$ and $(v - w) \cdot \nu$.

Conservation of kinetic energy and momentum for binary collisions implies that v^* and w^* are obtained by exchanging the projections of the speeds on ν :

$$v^* = v + ((w - v) \cdot \nu)\nu, \quad w^* = w + ((v - w) \cdot \nu)\nu.$$

There can be various boundary conditions on x , which can be included in the model by changing the free transport operator. All these models and kinds of collision (following the cross section B) are discussed in Cercignani [6], Neunzert and al. [26], Babovsky and Illner [1] and Wagner [39], Graham and Méléard [15].

2.1.1. Probabilistic interpretation

The nonlinear integral term Q in (2.1) comes from the randomness in the geometric configuration of collisions, and it is natural to study the probabilistic interpretation of this equation. We adapt the ideas in Section 1.4.4 to this nonlinear setting; in particular, we need to take into account more general initial conditions.

A solution f of the equation (2.1) is understood in the following weak sense: for all sufficiently smooth test functions ϕ ,

$$\begin{aligned} & \int \phi(x, v) (\partial_t f(t, x, v) + v \cdot \nabla_x f(t, x, v)) dx dv \\ &= \int \phi(x, v) \int_{S^2} d\nu \int_{\mathbb{R}^3} dw B(v - w, \nu) (f(t, x, v^*) f(t, x, w^*) - f(t, x, v) f(t, x, w)) dx dv. \end{aligned}$$

Then we transpose the operators so to have them operating on ϕ . Using integration by parts and the change of variables $(v, w) \rightarrow (v^*, w^*)$ setting $P_t(dx, dv) = f(t, x, v) dx dv$, (2.1) is then rewritten as

$$\partial_t \langle P_t, \phi \rangle - \langle P_t, v \cdot \nabla_x \phi(x, v) \rangle = \left\langle P_t(dx, dv), \int (\phi(x, v^*) - \phi(x, v)) B(v - w, \nu) f(t, x, w) dw d\nu \right\rangle. \quad (2.3)$$

We see easily on the form of this equation that the particle position evolves according to free flow and its velocity changes from v to v^* at point x and time t according to the jump measure $B(v - w, \nu) f(t, x, w) dw d\nu$.

We interpret then (2.3) as the Kolmogorov equation for the marginal laws $(P_t)_{t \geq 0}$ of a stochastic process $(X, V) = (X_t, V_t)_{t \geq 0}$. We recall that (X, V) is a random function from \mathbb{R}_+ to \mathbb{R}^6 , for each t , P_t is the law of (X_t, V_t) , and for each continuous function ϕ on $\mathbb{R}^3 \times \mathbb{R}^3$, $E(\phi(X_t, V_t)) = \langle P_t, \phi \rangle$. We can consider every probability measure initial data P_0 , even degenerated as a Dirac measure.

2.2. The nonlinear martingale problem

We adapt the notions in Section 1.4.3 and 1.4.4 in this nonlinear framework. We are looking for a Markov process (X, V) with sample paths belonging to the Skorohod space $\mathbb{D}([0, T], \mathbb{R}^6)$, such that the marginal laws are solution of (2.3). The law P of (X, V) will satisfy the two following properties:

1) The law P belongs to the space of probability measures on $\mathbb{D}([0, T], \mathbb{R}^6)$ such that for every $t \in]0, T]$, the probability measure P_t on \mathbb{R}^6 has a density function $p(t, x, v)$. ($P_t(dx, dv) = p(t, x, v) dx dv$.) This space is denoted by $\tilde{\mathcal{P}}(\mathbb{D}([0, T], \mathbb{R}^6))$.

2) For every function $\varphi \in C_b^1(\mathbb{R}^6)$,

$$\begin{aligned} \varphi(X_t, V_t) - \varphi(X_0, V_0) - \int_0^t V_s \cdot \nabla_x \varphi(X_s, V_s) ds \\ - \int_0^t \int_{S^2} \int_{\mathbb{R}^3} (\varphi(X_s, V_s + ((w - V_s) \cdot \nu) \nu) - \varphi(X_s, V_s)) B(V_s - w, \nu) p(s, X_s, w) dw d\nu ds \end{aligned} \quad (2.4)$$

is a P -martingale, where $p(t, \cdot)$ is the density of P_t and $P_0(dx, dv) \in \mathcal{P}(\mathbb{R}^6)$ is given.

Definition 2.1. Such a probability measure P is called a solution of the nonlinear martingale problem (\mathcal{M}) characterized by (2.4).

Remark 2.2. Taking expectations in (2.4), we obtain the weak form (2.3) of the Boltzmann equation. Therefore, the martingale problem gives much more information than the evolution equation, since it enables us to consider multidimensional time marginals and quantities based on the whole process as extrema or hitting times which are of interest for gas dynamics.

The expression (2.4) is very unpleasant, mainly due to the local interaction; the density $p(t, x, w) dw$ is quite misbehaved in terms of the law $P_t(dx, dv)$ from both a continuity and a boundedness points of view. This renders any probabilistic existence and uniqueness result difficult to establish. From an analytical point of view, it is also very difficult, and existence and uniqueness for the Boltzmann equation have thus been very elusive, and are usually valid under very restrictive assumptions. In this general context, the best results of existence seem to be those of DiPerna-Lions [9]. We will see in Section 4 a particular case in which we have both existence and uniqueness of the solutions of the Boltzmann equation and of the associated nonlinear martingale problem.

To bypass these problems, let us first consider simpler equations, called mollified Boltzmann equations or Povzner equations, in which both problems of unboundedness and localized interaction disappear.

3. THE MOLLIFIED PROBLEM

3.1. The mollified nonlinear martingale problem

Mollifying consists in delocalizing in space the interaction appearing in the Boltzmann equation in order to obtain a mean field model. We use a regularizing kernel $I(x, y)$ and replace $f(t, x, w)$ by $\int f(t, y, w)dy$. This approximation is close to the statistical physics model and is suited for a probabilistic study.

The main example is given by the grid method. The space \mathbb{R}^3 is partitioned in cubic cells Δ of volume $|\Delta| = \delta^3$ and particles in each cell interact as if in the same location. We introduce the regularizing kernel

$$I^\delta(x, y) = \frac{1}{\delta^3} \sum_{\Delta} 1_{x \in \Delta} 1_{y \in \Delta}. \quad (3.1)$$

The collision operator Q is replaced by Q^δ , defined by

$$Q^\delta(f, f)(t, x, v) = \int (f(t, x, v^*)f(t, y, w^*) - f(t, x, v)f(t, y, w))B(v - w, \nu)I^\delta(x, y)dydw d\nu,$$

which leads to the mollified equation with delocalized cross-section $I^\delta(x, y)B(v - w, \nu)$

$$\partial_t f + v \cdot \nabla_x f = Q^\delta(f, f). \quad (3.2)$$

The unboundedness of the cross-section can also lead to serious difficulties. This may be taken care of by physical considerations (angular cutoff, speeds bounded by the speed of light). Here we assume that $\sup_{v, w} \int B(v - w, \nu)d\nu = \|B\|_\infty < \infty$, and then

$$\sup_{x, y, v, w} \int I^\delta(x, y)B(v - w, \nu)d\nu \leq \frac{\|B\|_\infty}{\delta^3}. \quad (3.3)$$

One easily proves that

Proposition 3.1. *There exists a constant $C(B, \delta) > 0$ depending only on $\|B\|_\infty$ and δ such that for each functions f and g in $L^1(dx dv)$,*

$$\|Q^\delta(f, f) - Q^\delta(g, g)\|_{L^1(dx dv)} \leq C(B, \delta)\|f + g\|_{L^1(dx dv)}\|f - g\|_{L^1(dx dv)}.$$

Given f_0 in $L^1(dx dv)$, the equation (3.2) has a unique density solution f^δ on $[0, T]$.

Exactly as in Section 2 we consider weak solutions of the equation (3.2), and a flow $(P_t)_t$ of probability measures on \mathbb{R}^6 will be a solution of (3.2) if for each $\phi \in C_b^{1,0}(\mathbb{R}^3 \times \mathbb{R}^3)$,

$$\partial_t \langle P_t, \phi \rangle - \langle P_t, v \cdot \nabla_x \phi(x, v) \rangle = \left\langle P_t(dx, dv), \int (\phi(x, v^*) - \phi(x, v))I^\delta(x, y)B(v - w, \nu)d\nu P_t(dy, dw) \right\rangle. \quad (3.4)$$

This equation has a simple mean field interpretation: a particle evolves in an ocean of infinitely many independent particles distributed according to the measure P_t , chooses at random another particle according to the above law, performs a collision with it and then forgets everything about it. **This does not say what happens to the collision partner. Depending on what happens to this collision partner, different**

algorithms have been developed by Bird, Nanbu and others. A physical binary symmetric collision derivation for example leads to write

$$\begin{aligned} & \left\langle P_t(dx, dv), \int (\phi(x, v^*) - \phi(x, v)) I^\delta(x, y) B(v - w, \nu) d\nu P_t(dy, dw) \right\rangle \\ &= \left\langle P_t(dx, dv) P_t(dy, dw), \int \frac{1}{2} (\phi(x, v^*) - \phi(x, v) + \phi(y, w^*) - \phi(y, w)) I^\delta(x, y) B(v - w, \nu) d\nu \right\rangle. \end{aligned} \quad (3.5)$$

We are now looking for a Markov process (X, V) taking values in $\mathbb{D}([0, T], \mathbb{R}^6)$ and whose law P^δ has time-marginals satisfying (3.4). It will be a solution of the so-called nonlinear mollified martingale problem (\mathcal{M}^δ) in the sense that for any function $\phi \in C_b^1(\mathbb{R}^6)$,

$$\begin{aligned} & \varphi(X_t, V_t) - \varphi(X_0, V_0) - \int_0^t V_s \cdot \nabla_x \varphi(X_s, V_s) ds \\ & - \int_0^t \int_{S^2} \int_{\mathbb{R}^3} (\varphi(X_s, V_s + ((w - V_s) \cdot \nu) \nu) - \varphi(X_s, V_s)) I^\delta(X_s, y) B(V_s - w, \nu) P_s^\delta(dy, dw) d\nu ds \end{aligned} \quad (3.6)$$

is a P^δ -martingale and the nonlinearity lies in the term $P_s^\delta = P^\delta \circ (X_s, V_s)^{-1} = \mathcal{L}(X_s, V_s)$ and $P_0^\delta = P_0$ is given.

Remark 3.2. The existence of a solution of this martingale problem gives, by taking expectations, the existence of a solution of (3.4). Conversely, the uniqueness of a flow of probability measures solutions of (3.4) implies the uniqueness in the martingale problem.

If the initial condition has a density, we have moreover

Theorem 3.3. *Under the cutoff assumption (3.3), and if we assume that $P_0(dx, dv) = f_0(x, v) dx dv$, then there exists a unique probability measure $P^\delta \in \tilde{\mathcal{P}}(\mathbb{D}([0, T], \mathbb{R}^6))$, solution of the mollified nonlinear martingale problem. Moreover, for each $t \in [0, T]$, P_t^δ has a density with respect to the Lebesgue measure, which is the unique solution f^δ of the mollified Boltzmann equation defined in (3.1).*

Proof. 1) Uniqueness: if P^δ is a solution of (\mathcal{M}^δ) , then by taking expectations, we obtain that the flow of marginals $(P_t^\delta)_t$ is solution of (3.4). Moreover, for each $t \in [0, T]$ and by some probabilistic arguments related to the existence of a density for P_0 and to the form of the jump measure, we can prove that the probability measure P_t^δ has a density function $p^\delta(t, \cdot)$. This density p^δ is then a weak solution of the mollified Boltzmann equation and Proposition 3.1 implies that $p^\delta = f^\delta$ almost surely.

Hence two solutions Q^1 and Q^2 of (\mathcal{M}^δ) have the same flow of time marginals with density f^δ . So they both solve a standard martingale problem (\mathcal{M}^{f^δ}) with bounded jump measure $f^\delta(t, y, w) B(v - w, \nu) I^\delta(x, y) d\nu dy dw$, for which existence and uniqueness of a solution is well known. Thus $Q^1 = Q^2$.

2) Existence: Let us consider a solution Q of (\mathcal{M}^{f^δ}) , which, as before, has a density q w.r.t. the Lebesgue measure. Then by taking expectations, we obtain that q is solution of the linear equation

$$\partial_t q + v \cdot \nabla_x q = Q^\delta(q, f^\delta); \quad q(0, x, v) = f_0(x, v), \quad (3.7)$$

with

$$Q^\delta(q, f^\delta)(t, x, v) = \int \left(q(t, x, v^*) f^\delta(t, y, w^*) - q(t, x, v) f^\delta(t, y, w) \right) I^\delta(x, y) B(v - w, \nu) d\nu dy dw.$$

One easily proves that for two density functions q_1 and q_2 ,

$$\|Q^\delta(q_1, f^\delta) - Q^\delta(q_2, f^\delta)\|_{L^1(dx, dv)} \leq \frac{\|B\|_\infty}{\delta^3} \|q_1 - q_2\|_{L^1(dx, dv)},$$

and then we obtain the uniqueness of the density function q solution of (3.7). Since f^δ is a solution of this equation, then $q = f^\delta$ and Q is a solution of the nonlinear martingale problem. \square

If P_0 has no density, then the above proof does not work. One can show an existence and uniqueness result if we assume Lipschitz regularity of the cross section and of the regularizing kernel I^δ , see [13] Proposition 4.4 and [15] Theorem 4.3, but this is out of our context. In the next subsections, we will obtain the existence of a solution of (\mathcal{M}^δ) for any law P_0 as the limit of the laws of some interacting particle systems.

3.2. The approximating stochastic interacting particle systems

One defines mean field interacting particle systems which approximate a solution of the nonlinear martingale problem (3.6). These systems correspond to different physical models: first a simple mean field model (Nambu's system), second a binary mean field model (Bird's system).

These particle systems are homogeneous Markov processes on a product space, which we define through their generators. We refer to Section 1.4 for the definitions, and Section 1.6 for their simulation on a computer (which we further discuss later).

Let $(x^n, v^n) = ((x_1, v_1), (x_2, v_2), \dots, (x_n, v_n))$ be the generic point in $(\mathbb{R}^6)^n$. We introduce the mapping $\mathbf{e}_i : h \rightarrow \mathbf{e}_i \cdot h = (0, \dots, 0, h, 0, \dots, 0) \in (\mathbb{R}^3)^n$ with h at the i -th place. We consider $\phi \in C_b^1((\mathbb{R}^6)^n)$ and define two systems of particles:

The Nambu system is a Markov process in $\mathbb{D}(\mathbb{R}_+, (\mathbb{R}^6)^n)$ with generator

$$\sum_{i=1}^n v_i \cdot \nabla_{x_i} \phi(x^n, v^n) + \frac{1}{n-1} \sum_{1 \leq i \neq j \leq n} \int \left(\phi(x^n, v^n + \mathbf{e}_i \cdot ((v_j - v_i) \cdot \nu) \nu) - \phi(x^n, v^n) \right) I^\delta(x_i, x_j) B(v_i - v_j, \nu) d\nu. \quad (3.8)$$

The Bird system is a Markov process on the same space with generator

$$\sum_{i=1}^n v_i \cdot \nabla_{x_i} \phi(x^n, v^n) + \frac{1}{n-1} \sum_{1 \leq i \neq j \leq n} \int \frac{1}{2} \left(\phi(x^n, v^n + \mathbf{e}_i \cdot ((v_j - v_i) \cdot \nu) \nu) + \phi(x^n, v^n + \mathbf{e}_j \cdot ((v_i - v_j) \cdot \nu) \nu) - \phi(x^n, v^n) \right) \times I^\delta(x_i, x_j) B(v_i - v_j, \nu) d\nu. \quad (3.9)$$

In both cases, $Z^{\delta, n} = (X^{\delta, n}, V^{\delta, n}) = (Z^{\delta, 1n}, Z^{\delta, 2n}, \dots, Z^{\delta, nn})$ denotes the Markov process.

An important point which helps to understand the choice of the particle systems is the following. The nonlinearity appears in the jump term of the mollified martingale problem through the marginal laws of the process and is a usual mean field nonlinearity. That gives an obvious choice of approximating particle systems based on the law of large numbers, as in Theorem 1.18 and leads to the Nambu system. (Nambu in [25] presented an algorithm for the Boltzmann equation using this approach). But this choice is very far from the physical interpretation of the Boltzmann equation. It can be more convenient to construct approximating particle systems corresponding to the underlying statistical physics model by considering (3.5), which leads to the Bird system, in reference to the historical Bird algorithm for the Boltzmann equation (cf. [39]).

In the Markov process generated by the Nambu generator, there are no simultaneous jumps of particles, while in the Bird model, two interacting particles jump simultaneously. One advantage of the Bird model is the exact conservation of momentum and kinetic energy for the particle system.

3.3. Propagation of chaos in variation norm and rate of convergence

The propagation of chaos is a probabilistic limit behavior in which the law of a fixed number k of particles in an interacting system with independent initial data converges to the k -fold product of a limiting probability measure as the size of the system goes to infinity. This implies weak convergence in probability of the empirical measures to the limiting probability measure.

This is to be compared to the law of large numbers for a system of independent particles, see Section 1.5.

Let us give now our main theorem in which we adapt the interaction graph techniques in Graham-Méléard [14]. The convergence results are on the path space and imply results on the flow of time marginals.

Let $|\cdot|_T$ denote the variation norm in the space of signed measures on $\mathbb{D}([0, T], (\mathbb{R}^6)^k)$, for any $k \in \mathbb{N}$.

Theorem 3.4. *Let us consider i.i.d. variables $(Z_0^{\delta, in})_{1 \leq i \leq n}$ with law P_0 and assume the cutoff hypothesis*

$$\sup_{v, w \in \mathbb{R}^3} \int B(v - w, \nu) d\nu = \|B\|_\infty < +\infty.$$

1) *We have propagation of chaos: setting $\Lambda_\delta = \frac{\|B\|_\infty}{\delta^3}$, for any fixed k ,*

$$|\mathcal{L}(Z^{\delta, 1n}, \dots, Z^{\delta, kn}) - \hat{P}^{\delta \otimes k}|_T \leq k(k-1) \frac{e^{\Lambda_\delta T}}{n-1} \tag{3.10}$$

where \hat{P}^δ solves the nonlinear martingale problem (\mathcal{M}^δ) .

2) *Let us denote by $\mu^{\delta, n}$ the empirical measures*

$$\mu^{\delta, n} = \frac{1}{n} \sum_{i=1}^n \delta_{Z^{\delta, in}}.$$

Then the sequence $(\mu^{\delta, n})_n$ converges to \hat{P}^δ with $O(1/\sqrt{n})$ estimates.

The asymptotic behavior (3.10) is called *Propagation of chaos* in a terminology introduced by Kac and then by McKean, and it means the propagation of the independence, in reference to the hypothesis on the initial data. It describes the behavior of a **fixed** subsystem when the size of the particle system tends to infinity.

The convergence in variation norm for probability measures on the path space is a very strong convergence. It implies the convergence in variation for each time marginal, uniformly in time, and if the probability measures have densities, it implies the L^1 -convergence for the densities.

Proof. In order to prove Theorem 3.4, we prove two results: first

$$|\mathcal{L}(Z^{\delta, 1n}, \dots, Z^{\delta, kn}) - \mathcal{L}(Z^{\delta, 1n}) \otimes \dots \otimes \mathcal{L}(Z^{\delta, kn})|_T \leq k(k-1) \frac{\Lambda_\delta T + (\Lambda_\delta)^2 T^2}{n-1}$$

and second

$$|\mathcal{L}(Z^{\delta, 1n}) - \hat{P}^{\delta \otimes k}|_T \leq k(k-1) \frac{e^{\Lambda_\delta T} - 1}{n-1}.$$

The first part consists in computing the probability that there exists an interaction between two particles i and j during the time T by using some interaction graphs representing the past of the particles, and the second part consists in defining the limit probability \hat{P}^δ on a limiting Boltzmann tree.

1) The interaction graphs.

The first step consists in giving for any $k \leq n$ a pathwise representation of the process $(Z^{\delta, 1n}, \dots, Z^{\delta, kn})$ on the time interval $[0, T]$ using interaction graphs.

At time T , the evolution of the state of a particle has been directly affected by other particles with which it has collided, and recursively these particles have been affected by still other particles which thus influence indirectly our first particle, and so on. We now describe this construction on $[0, T]$ in reverse time so that we can build the path of our particle with the least amount of superfluous knowledge: this defines its past history, containing all particles that have influenced it through the chains of interactions described above. If we prove

that the histories of two particles become disjoint in the limit, it will be possible by a coupling argument to show that they become independent.

These chains of interactions are described by interaction graphs, which are random subsets of $[0, T] \times \{1, \dots, n\}$. Let us first introduce independent Poisson processes $(N_{ij})_{1 \leq i < j \leq n}$ of rate $\Lambda^\delta / (n-1)$. For $i < j$, N_{ij} are random clocks giving the epochs at which $Z^{\delta, in}$ and $Z^{\delta, jn}$ are authorized to jump simultaneously and possibly interact, and $N_{ji} = N_{ij}$. We imagine then time as being vertical and directed upwards, and the indices of particles as being on a horizontal level. We work our way backward in time starting from time T to build a graph rooted on the given subset $\{1, \dots, k\}$. Every time we encounter a jump of a Poisson process N_{ij} for an i (or j) already in the graph we include the index j (or i) in the graph at that time. This branching is deterministic given the Poisson processes and generates two branches, see Figure 1.

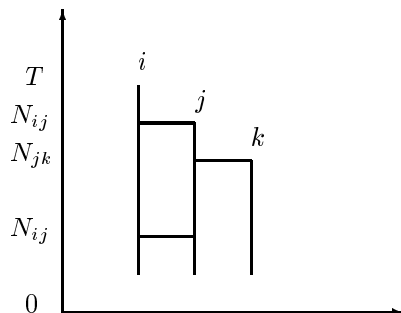


FIGURE 1. Interaction graph rooted at i .

Once an index is selected, the whole vertical line from the point of its first selection down to zero belongs to the graph, and we proceed recursively from the selected indices to define the graph. We do not have a tree, since a particle may influence another one several times. We can thus build down from time T an interaction graph $G_{1, \dots, k}^n$ rooted on $(1, \dots, k)$.

Once we reach time 0, we construct a pathwise representation of $(Z^{\delta, 1n}, \dots, Z^{\delta, kn})$ in direct time on $[0, T]$. The interaction graph represents all the information necessary to construct the process, and we only need to consider at time t the indices appearing then in the graph. We use independent variables of law P_0 at time 0. Then we follow the indices in the graph. Between jumps of the N_{ij} we follow the free transport. At the jump times of N_{ij} we compute the total mass $|\hat{B}|$ of the jump measure \hat{B} at the position reached by $(Z^{\delta, in}, Z^{\delta, jn})$. With probability $1 - |\hat{B}|/\Lambda_\delta$ we do nothing and with probability $|\hat{B}|/\Lambda_\delta$ we choose the joint amplitude of jumps according to the law $\frac{\hat{B}}{|\hat{B}|}$. All this must be done independently. This yields the correct Markovian evolution.

2) The coupling, the interaction chains and the estimates of convergence.

For simplicity we restrict ourselves to compare two particles with indices i and j .

Let us roughly describe the coupling. We build on the same probability space the subgraph G_i^n stemming from i describing the past history of the particle i , and the same one for j . The two particles will interact during $[0, T]$ if in their past history before time T , one can find a pair (k, l) of indices such that for example k has had an influence on i and l has had an influence on j . We call A_{ij}^n the event describing the random set for which such a pair (k, l) exists.

The definition of the variation norm implies clearly that

$$|\mathcal{L}(Z^{\delta, in}, Z^{\delta, jn}) - \mathcal{L}(Z^{\delta, in}) \otimes \mathcal{L}(Z^{\delta, jn})|_T \leq 2P(A_{ij}^n).$$

Similarly for k indices we have :

$$|\mathcal{L}(Z^{\delta,1n}, \dots, Z^{\delta,kn}) - \mathcal{L}(Z^{\delta,1n}) \otimes \dots \otimes \mathcal{L}(Z^{\delta,kn})|_T \leq 2P\left(\cup_{1 \leq p < q \leq k} A_{pq}^n\right).$$

We now have to compute $P(A_{ij}^n)$.

A_{ij}^n happens only if we encounter a jump of a pair of conflicting Poisson processes before reaching 0. This event happens if there exists an **interaction chain** between i and j , described by Figure 2.

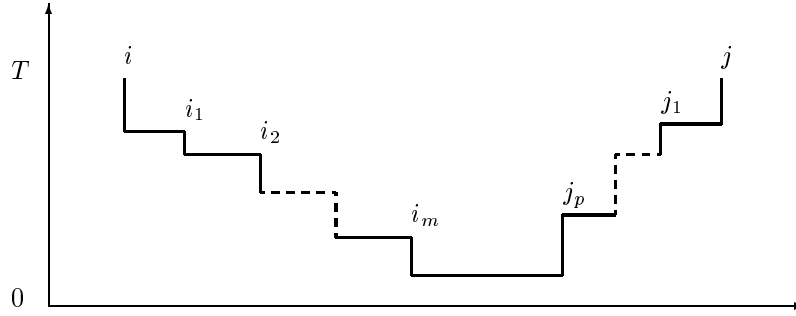


FIGURE 2. Chain interaction.

We define the interaction chain length as $m + p + 1$. We now compute the probability of existence of an interaction chain during the time interval $[0, T]$. Let us denote by Q_T^n this probability. We have $Q_T^n \leq \sum_{q \geq 1} Q_T^n(q)$, where $Q_T^n(q)$ is a bound of the probability of occurrence of an interaction chain of length q during $[0, T]$. We presently compute $Q_T^n(q)$ by induction on q .

We start with $q = 1$. For a direct interaction to happen between 0 and T , there is a route containing both i and j such that N_{ij} jumps in $[0, T]$. The probability of this event is less than

$$Q_T^n(1) \leq 1 - e^{-2\frac{\Lambda_\delta}{n-1}T} \leq 2\frac{\Lambda_\delta}{n-1}T.$$

We now assume we have a bound $Q_T^n(q - 1)$ of the probability of occurrence of an interaction chain of length $q - 1$, for $q \geq 2$. Then for an interaction chain of length q to happen, there must be first (backward from T) the birth of a new branch at either i or j after a waiting time $t \in [0, T]$, and this new branch must then be joined in time $T - t$ by an interaction chain of length $q - 1$ to the one of i or j which did not branch first. The maximal rate of this new branch is $2\Lambda_\delta$ and thus :

$$Q_T^n(q) \leq \int_0^T Q_{T-t}^n(q - 1) 2\Lambda_\delta e^{-2\Lambda_\delta t} dt.$$

If we denote the exponential density of parameter $2\Lambda_\delta$ by $e_{2\Lambda_\delta}$ and convolution by $*$, for $q \geq 2$,

$$Q_T^n(q) \leq Q_T^n(q - 1) * e_{2\Lambda_\delta}(T) \leq Q_T^n(1) * e_{2\Lambda_\delta}^{*(q-1)}(T).$$

Since $e_{2\Lambda_\delta}^{*k}(t) = (2\Lambda_\delta)^k \frac{t^{k-1}}{(k-1)!} e^{-2\Lambda_\delta t}$ we obtain

$$\begin{aligned} Q_T^n &\leq Q_T^n(1) + 2\Lambda_\delta \int_0^T Q_{T-t}^n(1) \sum_{q \geq 2} \frac{(2\Lambda_\delta t)^{q-2}}{(q-2)!} e^{-2\Lambda_\delta t} dt \leq Q_T^n(1) + 2\Lambda_\delta \int_0^T (1 - e^{-2\frac{\Lambda_\delta}{n-1}(T-t)}) dt \\ &\leq 2 \frac{\Lambda_\delta T + (\Lambda_\delta)^2 T^2}{n-1}. \end{aligned}$$

Considering the $k(k-1)/2$ pairs of indices we obtain the first bound (3.10) in Theorem 3.4.

3) The limiting Boltzmann tree.

A similar coupling argument between the interaction graph issued from one index for a given n and a limit Boltzmann tree, where the links are taken amongst an infinite supply of independent similar links, shows that the laws of the processes $Z^{\delta, in}$ converge to the law \hat{P}^δ of a process constructed similarly on the tree. The construction is inspired from step 2 of the proof and will not be developed here.

4) The empirical measure and the nonlinear martingale problem.

For a bounded function ϕ defined on $\mathbb{D}([0, T], \mathbb{R}^6)$, we have

$$\begin{aligned} E(\langle \mu^{\delta, n} - \hat{P}^\delta, \phi \rangle^2) &= E\left(\left(\frac{1}{n} \sum_{i=1}^n \phi(Z^{\delta, in}) - \langle \hat{P}^\delta, \phi \rangle\right)^2\right) \\ &= \frac{1}{n} E(\langle \phi(Z^{\delta, 1n}) - \langle \hat{P}^\delta, \phi \rangle \rangle^2) + \frac{n-1}{n} E(\langle \phi(Z^{\delta, 1n}) - \langle \hat{P}^\delta, \phi \rangle \rangle \langle \phi(Z^{\delta, 2n}) - \langle \hat{P}^\delta, \phi \rangle \rangle) \\ &= E(\phi(Z^{\delta, 1n}) \phi(Z^{\delta, 2n})) - 2\langle \hat{P}^\delta, \phi \rangle E(\phi(Z^{\delta, 1n})) + \langle \hat{P}^\delta, \phi \rangle^2 + O(1/n) \end{aligned}$$

which is $O(1/n)$ uniformly in T and $\|\phi\|_\infty$. This implies the convergence of $\mu^{\delta, n}$ to \hat{P}^δ and gives an $O(1/\sqrt{n})$ rate for appropriate metrics.

The convergence of the law of each process $Z^{\delta, in}$ being in variation norm, one can prove easily that the limiting law \hat{P}^δ solves the nonlinear martingale problem (\mathcal{M}^δ) . The proof of Theorem 3.4 is now finished. \square

Theorem 3.4 gives upper-bounds for the rate of convergence. As it can be seen in Méléard [23], one can moreover obtain a central-limit theorem, which assures that $\frac{1}{\sqrt{n}}$ is the exact rate of convergence and justifies the interest of considering Monte-Carlo algorithms.

3.4. Algorithms of simulation for the mollified Boltzmann equation

As seen previously, the empirical measures $\mu^{\delta, n}$ approximate the law of the mollified Boltzmann process. We now quickly show how to simulate the particle systems defined in Section 3.2, and refer to Section 1.6 for details.

The ideas of the graph construction in Section 3.3 can be used. The jump operator can be simulated exactly, by using an acceptance-rejection or fictitious-collisions method as in the proof of Theorem 3.4. The core idea is that adding mass at zero to the jump measure does not change the jump operator, and this involves no approximation whatsoever. Since the total mass $|\hat{B}|$ of the jump measure \hat{B} is bounded by Λ_δ , one can always add $\Lambda_\delta(1 - |\hat{B}|/\Lambda_\delta)\delta_0$ to \hat{B} and obtain a positive measure of constant total mass Λ_δ .

There are n particles and the total rate for the $n(n-1)/2$ pairs of possible interactions is $n\Lambda_\delta/2$ as seen in (3.9). A Poisson process of rate $n\Lambda_\delta/2$ gives the sequence of collision times. At each of these we choose uniformly the pair of particles which interact, update the states of these particles following the free transport, compute $|\hat{B}|$ at these states, discard the jump with probability $1 - |\hat{B}|/\Lambda_\delta$ and with probability $|\hat{B}|/\Lambda_\delta$ choose the joint jump amplitude according to $\hat{B}/|\hat{B}|$. All this is done independently. This method only necessitates at each step the evaluation of the cross-section of the interacting pair and not those of the $n(n-1)/2$ pairs.

This simulation is exact if we simulate exactly the exponential variables related to the Poisson process, instead of discretizing in time.

4. THE CASE OF A BOLTZMANN EQUATION WITH CUTOFF AND SMALL INITIAL DATA.

We have seen in Section 2 that the Boltzmann equation presents many important difficulties, due to the unboundedness of B and to the localization in space in the quadratic collision term. Existence and uniqueness have been studied by many authors under restrictive assumptions on the cross-section B (principally a cutoff assumption) and results have been obtained in particular in small time or under small initial data, as it can be found in [19], [2], [17], [36], [3] and more recently in Mischler-Perthame [24]. In this Section, we prove the existence and uniqueness of the solution of the Boltzmann equation in a case of cutoff collision operator and small initial data, by a fixed point argument in a well chosen functional space \mathcal{B}_α . We follow [24]. This existence and uniqueness result (Theorem 4.1) is very close to *a priori* assumptions in Babovsky and Illner [1].

One also obtains in this case the existence and uniqueness of the solution of the associated martingale problem in the space of probability measures having a measurable version of densities in \mathcal{B}_α .

Then, as in the previous section, one delocalizes the interaction by considering the mollified Boltzmann equation with the regularizing kernel I^δ . Theorem 3.4 shows the convergence of some interacting particle systems to the solution of the mollified Boltzmann equation and gives a precise rate of convergence in $\mathcal{O}\left(\frac{1}{n} \exp\left(\frac{K}{\delta^3}\right)\right)$, where n is the size of the particle system.

Next, we consider such systems in which the size of the cells of the grid depends on the size of the system. More precisely, we assume that δ depends on n in such a way that $\delta(n) \rightarrow 0$ and $\frac{1}{n} \exp\left(\frac{K}{\delta(n)^3}\right) \rightarrow 0$ as $n \rightarrow \infty$. Then we prove that the empirical measures of the associated interacting particle systems converge in law to a deterministic probability measure whose density flow is the solution of the full Boltzmann equation.

At our knowledge, this result seems to be the first pathwise approximation result in a non mollified case and in dimension 3. Let us quote Caprino and Pulvirenti [5] and Rezakhanlou [30], who obtain the convergence of stochastic particle systems to a one-dimensional Boltzmann equation at fixed times. Our approach allows to understand the similarity between Bird's and Nanbu's algorithms. Moreover it gives a precise asymptotic relation between δ and n ; even if it is not optimal, this gives a partial answer to an open question in [1].

4.1. The existence and uniqueness result and the nonlinear martingale problem

Let us now prove the existence and uniqueness result obtained for the Boltzmann equation in a case of bounded collision operator and small initial data with finite energy.

Theorem 4.1. *Let $\alpha > 0$ and T be a positive time. Let us assume that*

$$(H_1): \sup_{z \in \mathbb{R}^3} \int B(z, \nu) d\nu = \|B\|_\infty < +\infty.$$

$$(H_2): f_0 \text{ is a density function satisfying for a real constant } C_0 \text{ such that } C_0 < \frac{(\sqrt{\alpha})^3}{\|B\|_\infty (\sqrt{\pi})^3 T} = \frac{1}{C_\alpha T}$$

$$0 \leq f_0(x, v) \leq \frac{C_0}{6} \exp(-\alpha|v|^2). \tag{4.1}$$

Then there exists a unique function $f \in L^\infty([0, T] \times \mathbb{R}^3 \times \mathbb{R}^3)$ solution of the Boltzmann equation (2.1) satisfying

$$0 \leq f(t, x, v) \leq \frac{C(t)}{6} \exp(-\alpha|v|^2) \tag{4.2}$$

where $C(t)$ is a positive and bounded function on $[0, T]$ defined by $\frac{1}{C(t)} = \frac{1}{C_0} - C_\alpha t$.

Proof. The proof is completely inspired from the proof of Theorem 2 in [24] given in a more general case of infinite energy. It consists first in introducing an upper solution related to the Boltzmann equation, and second in obtaining a fixed point theorem in a functional space related to this upper solution.

Let us begin with some notations. We set

$$Q^+(f, f)(t, x, v) = \int_{S^2} \int_{\mathbb{R}^3} f(t, x, v^*) f(t, x, w^*) B(v - w, \nu) dw d\nu, \quad Q^-(f, f)(t, x, v) = f(t, x, v) Lf(t, x, v),$$

with

$$Lf(t, x, v) = \int_{S^2} \int_{\mathbb{R}^3} B(v-w, \nu) f(t, x, w) dw d\nu.$$

Let us consider $g(t, v) = C(t)h(v)$, where $h(v) = \exp(-\alpha|v|^2)$. One would like

$$\partial_t g(t, v) = \dot{C}(t)h(v) \geq Q^+(g, g)(t, v) = C^2(t)Q^+(h, h)(v) = C^2(t)h(v)L(h)(v),$$

since $Q(h, h) = 0 = Q^+(h, h) - Q^-(h, h) = Q^+(h, h) - hL(h)$.

Then one is looking for C such that $\dot{C} \geq C^2 \sup_v L(h)(v)$. Therefore let us consider $C \in \mathcal{C}^1([0, T], \mathbb{R})$ such that $C(0) = C_0$ and solving $\dot{C}(t) = \|B\|_\infty \frac{\sqrt{\pi^3}}{\sqrt{\alpha^3}} C^2(t)$. By denoting $C_\alpha = \|B\|_\infty \frac{\sqrt{\pi^3}}{\sqrt{\alpha^3}}$, one finally obtains that $\frac{1}{C(t)} = \frac{1}{C_0} - C_\alpha t$ and the function

$$\hat{h}(t, v) = \frac{C(t)}{6} h(v) = \frac{C(t)}{6} e^{-\alpha|v|^2} \quad (4.3)$$

satisfies

$$\partial_t \hat{h}(t, v) \geq C_\alpha C(t) \hat{h}(t, v) \geq 6Q^+(\hat{h}, \hat{h})(t, v). \quad (4.4)$$

Now one considers the Banach space

$$\mathcal{B}_\alpha = \{\varphi \in L^\infty([0, T] \times \mathbb{R}^3 \times \mathbb{R}^3); 0 \leq \varphi(t, x, v) \leq \hat{h}(t, v)\}, \quad \|\varphi\|_\alpha = \text{esssup}_{t,x,v} \frac{|\varphi(t, x, v)|}{\frac{C(t)}{6} \exp^{-\alpha|v|^2}}.$$

The global existence and uniqueness of the solution of the Boltzmann equation is deduced from a fixed point theorem in this space. As in [24], let us define the mapping $\Lambda : \varphi \in \mathcal{B}_\alpha \rightarrow \psi = \Lambda(\varphi)$, where ψ is the solution of

$$\partial_t \psi + v \cdot \nabla_x \psi + C_\alpha C(t) \psi = Q^+(\varphi, \varphi) + (C_\alpha C(t) - L(\varphi)) \varphi; \quad \psi(0, x, v) = f_0(x, v). \quad (4.5)$$

By a maximum principle one observes that Λ sends \mathcal{B}_α into \mathcal{B}_α and by (4.4) that moreover

$$\forall \varphi_1, \varphi_2 \in \mathcal{B}_\alpha, \|\Lambda \varphi_1 - \Lambda \varphi_2\|_\alpha \leq \frac{5}{6} \|\varphi_1 - \varphi_2\|_\alpha.$$

Hence Λ is a contraction and admits a unique fixed point in \mathcal{B}_α , which is solution of the Boltzmann equation. \square

As in Section 2.2, we associate with the Boltzmann equation the nonlinear martingale problem described in Definition 2.1.

Let us denote by $\tilde{\mathcal{P}}_\alpha(\mathbb{D}([0, T], \mathbb{R}^6))$ the subspace of $\tilde{\mathcal{P}}(\mathbb{D}([0, T], \mathbb{R}^6))$ such that the densities p satisfy $0 \leq p(t, x, v) \leq \hat{h}(t, v)$ for almost every $(t, x, v) \in [0, T] \times \mathbb{R}^6$, \hat{h} being defined in (4.3).

Theorem 4.2. *Under assumptions (H_1) , (H_2) , the nonlinear martingale problem (2.4) has a unique solution P in $\tilde{\mathcal{P}}_\alpha(\mathbb{D}([0, T], \mathbb{R}^6))$ such that $P_0(dx, dv) = f_0(x, v) dx dv$.*

Moreover, the density flow of P is almost surely equal to the solution f of the Boltzmann equation defined in Theorem 4.1.

Let us first observe that if P is solution of (2.4), then by taking the expectations in the martingale problem, the flow of densities p is solution of the Boltzmann equation (2.1). Moreover if we assume that $0 \leq p(t, x, v) \leq \hat{h}(t, v)$, then p is almost surely equal to f by Theorem 4.1. Therefore, we first study the following classical martingale problem associated with the function f , and we prove that

Proposition 4.3. *Under assumptions (H_1) , (H_2) , there exists a unique probability law P^f in $\mathcal{P}(\mathbb{D}([0, T], \mathbb{R}^6))$ absolutely continuous with respect to the Lebesgue measure such that for any function φ in $C_b^1(\mathbb{R}^6)$*

$$\begin{aligned} \varphi(X_t, V_t) - \varphi(X_0, V_0) - \int_0^t V_s \cdot \nabla_x \varphi(X_s, V_s) ds \\ - \int_0^t \int_{S^2} \int_{\mathbb{R}^3} (\varphi(X_s, V_s + ((w - V_s) \cdot \nu) \nu) - \varphi(X_s, V_s)) B(V_s - w, \nu) f(s, X_s, w) dw d\nu ds \end{aligned} \quad (4.6)$$

is a P -martingale and $P_0(dx dv) = f_0(x, v) dx dv$. We say that P^f is a solution of the martingale problem (\mathcal{M}^f) . Moreover, the density flow q of P^f is solution of the mild equation

$$q(t, x, v) = f_0(x - tv, v) + \int_0^t (S_{t-s})^* Q(q, f)(s, x, v) ds, \quad (4.7)$$

where S_t is the semi-group associated with the flow solution of $\partial_t q + v \cdot \nabla_x q = 0$, and $(S_t)^*$ is the dual operator.

Proof. (H_1) and (4.2) imply that the jump kernel $B(V_s - w, \nu) f(s, X_s, w) dw d\nu$ has a finite total mass uniformly in s, X_s, V_s . Moreover, the drift part in (4.6) is Lipschitz. In this case, the existence and uniqueness of a solution P^f of (\mathcal{M}^f) is well known, and one can prove that P^f has a density q with respect to the Lebesgue measure.

By taking expectations in (4.6), we obtain moreover that the flow (q_t) satisfies for φ in $C_b^1(\mathbb{R}^6)$

$$\partial_t \langle q_t, \varphi \rangle - \langle q_t, v \cdot \nabla_x \varphi \rangle = \langle q_t(x, v) dx dv, \int (\varphi(x, v + ((w - v) \cdot \nu) \nu) - \varphi(x, v)) B(v - w, \nu) f(t, x, w) d\nu dw \rangle. \quad (4.8)$$

We extend this formula to functions $\psi(t, x, v)$ in $C_b^1([0, T] \times \mathbb{R}^6)$ by Itô's formula. Let S_t be the semi-group associated with the flow $\partial_t q + v \cdot \nabla_x q = 0$ and $S_t^* = S_{-t}$ the dual semi-group. Of course $S_t \varphi(x, v) = \varphi(x + tv, v)$. For fixed t in $[0, T]$ and φ in $C_b^1(\mathbb{R}^6)$ we set $\psi(s, x, v) = S_{t-s} \varphi(x, v) = \varphi(x + (t-s)v, v)$. Then $\partial_s \psi + v \cdot \nabla_x \psi = 0$ and $\psi(t, \cdot) = \varphi$. The equation (4.8) extended to ψ implies that for any function φ in $C_b^1(\mathbb{R}^6)$

$$\begin{aligned} \int \varphi(x, v) q(t, x, v) dx dv &= \int S_t \varphi(x, v) f_0(x, v) dx dv \\ &+ \int_0^t \int S_{t-s} \varphi(x, v) [f(s, x, w^*) q(s, x, v^*) - f(s, x, w) q(s, x, v)] B(v - w, \nu) d\nu dw dx dv ds \\ &= \int \varphi(x, v) S_t^* f_0(x, v) dx dv \\ &+ \int \varphi(x, v) \int_0^t S_{t-s}^* ([f(s, x, w^*) q(s, x, v^*) - f(s, x, w) q(s, x, v)] B(v - w, \nu) d\nu dw) dx dv ds \end{aligned}$$

and then we deduce that for every $t \in [0, T]$, $dx dv$ almost surely,

$$q(t, x, v) = f_0(x - vt, v) + \int_0^t (S_{t-s})^* Q(q, f)(s, x, v) ds. \quad (4.9)$$

□

Proposition 4.4. *Under assumptions (H_1) and (H_2) , the mild equation (4.9) has a unique solution $q(t, x, v)$ in $L^\infty([0, T], L^1(dx dv))$.*

Proof. Let q' be another solution of the evolution equation. Then

$$\begin{aligned}
\|q(t) - q'(t)\|_{L^1(dx dv)} &= \left\| \int_0^t (S_{t-s})^* [Q(q, f) - Q(q', f)](s, x, v) ds \right\|_{L^1(dx dv)} \\
&\leq \int_0^t \|(S_{t-s})^* [Q(q, f) - Q(q', f)](s, x, v)\|_{L^1(dx dv)} ds \\
&= \int_0^t \|Q(q, f) - Q(q', f)](s, x, v)\|_{L^1(dx dv)} ds \\
&\leq \int_0^t \int \{ |f(s, x, w^*)| |q(s, x, v^*) - q'(s, x, v^*)| + |f(s, x, w)| |q(s, x, v) - q'(s, x, v)| \} \\
&\quad \times B(v - w, \nu) dv dw dx dv ds \\
&\leq C_\alpha \frac{C(T)}{6} \int_0^t \|q(s) - q'(s)\|_{L^1(dx dv)} ds,
\end{aligned}$$

by (H_1) and (4.2). We conclude by the Gronwall Lemma. \square

Proof. (of Theorem 4.2). Let us first prove that the solution f defined in Theorem 4.1 is solution of (4.9). We first consider $\psi \in C_b^1([0, T] \times \mathbb{R}^6)$ with compact support. We can then prove by using Fubini's theorem and the integration by part formula that

$$\begin{aligned}
\partial_t \langle f_t, \psi \rangle - \langle f_t, v \cdot \nabla_x \psi + \partial_t \psi \rangle \\
= \langle f_t(x, v) dx dv, \int (\psi(t, x, v + ((w - v) \cdot \nu) \nu) - \psi(t, x, v)) B(v - w, \nu) f(t, x, w) dw dv \rangle.
\end{aligned}$$

We obtain by approximation the same formula for every function $\psi \in C_b^1([0, T] \times \mathbb{R}^6)$, and considering $\psi(s, x, v) = S_{t-s} \varphi(x, v)$, $\varphi \in C_b^1(\mathbb{R}^6)$, we obtain as before that the solution f of the full Boltzmann equation (1.1) satisfies for each $t \in [0, T]$

$$f(t, x, v) = f_0(x - vt, v) + \int_0^t (S_{t-s})^* Q(f, f)(s, x, v) ds, \quad \text{almost surely in } x, v. \quad (4.10)$$

The uniqueness result in Proposition 4.4 yields $q = f$, hence the solution of (\mathcal{M}^f) is in fact a solution of (2.4).

Let us now consider two solutions P and Q of (2.4) with measurable versions of the densities bounded by \hat{h} . Thus these densities are solutions of (2.1) bounded by \hat{h} and so are almost surely equal, and equal to f , by Theorem 4.1. Then P and Q are solutions of (\mathcal{M}^f) , and the uniqueness in this martingale problem implies that $P = Q$, hence we get Theorem 4.2. \square

Let us now give a regularity result for the function f , useful later in the proof of Proposition 4.8. The property stated below, as well as Theorem 4.1, are very close to properties given as conjectures in Babovsky-Illner [1].

Proposition 4.5. *Let us assume that*

(H_3) : *There exists $K > 0$ such that for every $h \in \mathbb{R}^3$,*

$$\text{esssup}_{x \in \mathbb{R}^3, v \in \mathbb{R}^3} \frac{|f_0(x + h, v) - f_0(x, v)|}{e^{-\alpha|v|^2}} \leq K|h|, \quad (4.11)$$

then the same property holds for f : there exists $K_T > 0$ such that

$$\text{esssup}_{t \in [0, T], x \in \mathbb{R}^3, v \in \mathbb{R}^3} \frac{|f(t, x + h, v) - f(t, x, v)|}{e^{-\alpha|v|^2}} \leq K_T|h|, \quad \forall h \in \mathbb{R}^3. \quad (4.12)$$

Proof. Let f be a solution of the evolution equation (4.10). Then

$$\begin{aligned} f(t, x+h, v) - f(t, x, v) &= f_0(x+h-tv, v) - f_0(x-tv, v) \\ &+ \int_0^t \int \left[\left(f(s, x+h-(t-s)v, w^*) f(s, x+h-(t-s)v, v^*) - f(s, x-(t-s)v, w^*) f(s, x-(t-s)v, v^*) \right) \right. \\ &\quad \left. - \left(f(s, x+h-(t-s)v, w) f(s, x+h-(t-s)v, v) - f(s, x-(t-s)v, w) f(s, x-(t-s)v, v) \right) \right] \\ &\quad \times B(v-w, \nu) d\nu dw ds, \end{aligned}$$

$$\begin{aligned} \frac{|f(t, x+h, v) - f(t, x, v)|}{e^{-\alpha|v|^2}} &\leq \frac{|f_0(x+h-tv, v) - f_0(x-tv, v)|}{e^{-\alpha|v|^2}} \\ &+ \int_0^t \int \left(\frac{|f(s, x+h-(t-s)v, w^*)| |f(s, x+h-(t-s)v, v^*) - f(s, x-(t-s)v, v^*)|}{e^{-\alpha|v|^2}} \right. \\ &\quad \left. + \frac{|f(s, x-(t-s)v, v^*)| |f(s, x+h-(t-s)v, w^*) - f(s, x-(t-s)v, w^*)|}{e^{-\alpha|v|^2}} \right) B(v-w, \nu) d\nu dw ds \\ &+ \int_0^t \int \left(\frac{|f(s, x+h-(t-s)v, w)| |f(s, x+h-(t-s)v, v) - f(s, x-(t-s)v, v)|}{e^{-\alpha|v|^2}} \right. \\ &\quad \left. + \frac{|f(s, x-(t-s)v, v)| |f(s, x+h-(t-s)v, w) - f(s, x-(t-s)v, w)|}{e^{-\alpha|v|^2}} \right) B(v-w, \nu) d\nu dw ds. \end{aligned}$$

Let $\Delta_{h,t}^\alpha(f) = \text{esssup}_{x,v} \frac{|f(t, x+h, v) - f(t, x, v)|}{e^{-\alpha|v|^2}}$. We have

$$\begin{aligned} \frac{|f(t, x+h, v) - f(t, x, v)|}{e^{-\alpha|v|^2}} &\leq K|h| + \frac{C(T)}{6} \int_0^t \int \left(\frac{e^{-\alpha|w^*|^2} \Delta_{h,s}^\alpha(f) e^{-\alpha|v^*|^2}}{e^{-\alpha|v|^2}} + \frac{e^{-\alpha|v^*|^2} \Delta_{h,s}^\alpha(f) e^{-\alpha|w^*|^2}}{e^{-\alpha|v|^2}} \right. \\ &\quad \left. + e^{-\alpha|w|^2} \Delta_{h,s}^\alpha(f) + e^{-\alpha|w|^2} \Delta_{h,s}^\alpha(f) \right) B(v-w, \nu) d\nu dw ds. \end{aligned}$$

We use the conservation of energy: $|v|^2 + |w|^2 = |v^*|^2 + |w^*|^2$. So we get

$$\begin{aligned} \frac{|f(t, x+h, v) - f(t, x, v)|}{e^{-\alpha|v|^2}} &\leq K|h| + 4 \frac{C(T)}{6} \int_0^t \int \Delta_{h,s}^\alpha(f) e^{-\alpha|w|^2} B(v-w, \nu) d\nu dv_* \\ &\leq K|h| + \frac{2C(T)}{3} C_\alpha \int_0^t \Delta_{h,s}^\alpha(f) ds, \end{aligned}$$

and finally

$$\Delta_{h,t}^\alpha(f) \leq K|h| + \frac{2}{3} C(T) C_\alpha \int_0^t \Delta_{h,s}^\alpha(f) ds.$$

We conclude by Gronwall's Lemma. □

4.2. Stochastic approximations

We consider first the mollified equation obtained by the grid method, as in Section 3. Under the assumptions of cutoff and small initial data, one can prove moreover that

Proposition 4.6. *Under assumptions $(H_1), (H_2)$, the unique solution f^δ of (3.2) with initial condition f_0 satisfies for almost $(t, x, v) \in [0, T] \times \mathbb{R}^6$,*

$$0 \leq f^\delta(t, x, v) \leq \hat{h}(t, v).$$

Proof. Since $Q^\delta(\hat{h}, \hat{h}) = Q(\hat{h}, \hat{h})$, the function \hat{h} is also an upper solution for the mollified problem. The same arguments as in the proof of Theorem 4.1 imply that the solution of the mollified equation is also obtained by a fixed point theorem in \mathcal{B}_α associated with the contraction Λ^δ defined as in (4.5) by replacing Q^+ by $Q^{\delta+}$. \square

One deduces from Proposition 4.6 the existence and uniqueness of the solution $P^\delta \in \tilde{\mathcal{P}}_\alpha(\mathbb{D}([0, T], \mathbb{R}^6))$ for the mollified nonlinear martingale problem associated with Q^δ , with initial data $f_0(x, v)dx dv$, as in Theorem 3.3.

In this present situation and if we consider the particle systems defined in Subsection 3.2, Theorem 3.4 yields

Theorem 4.7. *Let us consider i.i.d. variables $(Z_0^{\delta, in})_{1 \leq i \leq n}$ with law $P_0(dx, dv) = f_0(x, v)dx dv$ and assume (H_1) and (H_2) . Let us denote by Λ_δ the real number $\frac{\|B\|_\infty}{\delta^3}$. Then for given T and k ,*

$$|\mathcal{L}(Z^{\delta, 1n}, \dots, Z^{\delta, kn}) - (P^\delta)^{\otimes k}|_T \leq K_{k, T} \frac{e^{\Lambda_\delta T}}{n}. \quad (4.13)$$

Moreover, the empirical measures $\mu^{\delta, n} = \frac{1}{n} \sum_{i=1}^n \delta_{Z^{\delta, in}}$ converge to P^δ in $\mathcal{P}(\mathbb{D}([0, T], \mathbb{R}^6))$ with rate $\sqrt{\frac{e^{\Lambda_\delta T}}{n}}$.

Let us now consider interacting systems in which the size of the cells depends on the size of the system. We will prove that if we assume $(H_1), (H_2)$ and take an asymptotics $(n, \delta(n))$ which tends to $(+\infty, 0)$ such that $\frac{1}{n} e^{\frac{T \|B\|_\infty}{\delta(n)^3}}$ tends to zero, then the empirical measures of the system $(Z^{\delta(n), n})$ converge when n tends to infinity to the unique solution P of the nonlinear martingale problem (2.4) defined in Theorem 4.2.

Consider a sequence $(\delta(l))_{l \in \mathbb{N}}$ which tends to zero when l tends to infinity. Denote $P^l = P^{\delta(l)}$, f^l the flow of its densities, $Q^l = Q^{\delta(l)}$ and $I^l = I^{\delta(l)}$. We will prove the convergence of P^l to P when l tends to infinity.

Proposition 4.8. *Assume $(H_1), (H_2)$, and (H_3) (Cf. Proposition 4.5). If f denotes the unique solution of the Boltzmann equation in \mathcal{B}_α , then*

$$\sup_{t \leq T} \|f^l(t, \cdot) - f(t, \cdot)\|_{L^1} \leq K_T \delta(l) \quad (4.14)$$

and thus tends to zero when l tends to infinity.

Proof. Let us first remark that there exists a constant K_T such that for each (x, v) in \mathbb{R}^6 and t in $[0, T]$,

$$\left| \int f(t, y, v) I^l(x, y) dy - f(t, x, v) \right| \leq K_T \delta(l) e^{-\alpha|v|^2}. \quad (4.15)$$

Indeed, using Proposition 4.5,

$$\begin{aligned} \left| \int f(t, y, v) I^l(x, y) dy - f(t, x, v) \right| &= \left| \int (f(t, y, v) - f(t, x, v)) \frac{1}{\delta(l)^3} 1_{\{y \in \Delta_x^l\}} dy \right| \\ &\leq K_T \int |y - x| e^{-\alpha|v|^2} \frac{1}{\delta(l)^3} 1_{\{y \in \Delta_x^l\}} dy \leq K_T \delta(l) e^{-\alpha|v|^2}. \end{aligned}$$

We now use the mild equations satisfied by f^l and f :

$$\begin{aligned} f(t, x, v) &= f_0(x - vt, v) + \int_0^t (S_{t-s})^* Q(f, f)(s, x, v) ds, \quad \text{almost surely in } x, v, \\ f^l(t, x, v) &= f_0(x - vt, v) + \int_0^t (S_{t-s})^* Q^l(f^l, f^l)(s, x, v) ds, \quad \text{almost surely in } x, v. \end{aligned}$$

So we have

$$f^l(t, x, v) - f(t, x, v) = \int_0^t (S_{t-s})^* \left(Q^l(f^l, f^l) - Q(f, f) \right)(s, x, v) ds,$$

and if we denote $\|f\|_{L^1, t} = \|f(t, \cdot)\|_{L^1(dx dv)}$, we obtain

$$\|f^l - f\|_{L^1, t} \leq \int_0^t \|Q^l(f^l, f^l) - Q(f, f)\|_{L^1, s} ds.$$

We have

$$\begin{aligned} (Q^l(f^l, f^l) - Q(f, f))(s, x, v) &= \int B(v - w, \nu) \left(\int I^l(x, y) dy (f^l(s, x, v^*) f^l(s, y, w^*) - f^l(s, x, v) f^l(s, y, w)) \right. \\ &\quad \left. - [f(s, x, v^*) f(s, y, w^*) - f(s, x, v) f(s, y, w)] \right) dv dw. \end{aligned}$$

Tedious but easy computations using (4.15) lead to

$$\int |(Q^l(f^l, f^l) - Q(f, f))(s, x, v)| dx dv \leq 2K_T C_\alpha \delta(l) + 4C_\alpha \frac{C(T)}{6} \|f^l - f\|_{L^1, s}.$$

We deduce that

$$\|f^l - f\|_{L^1, t} \leq 2K_T C_\alpha \delta(l) T + 4C_\alpha \frac{C(T)}{6} \int_0^t \|f^l - f\|_{L^1, s} ds$$

and by Gronwall's Lemma there exists a constant $K > 0$ such that $\sup_{t \leq T} \|f^l - f\|_{L^1, t} \leq K \delta(l)$.

Finally, $\sup_{t \leq T} \|f^l - f\|_{L^1, t}$ tends to zero when l tends to infinity. \square

Theorem 4.9. *Under $(H_1), (H_2), (H_3)$ and if P_0 has a second order moment, then the sequence (P^l) converges to the unique probability measure $P \in \tilde{\mathcal{P}}_\alpha(\mathbb{D}([0, T], \mathbb{R}^6))$ defined in Theorem 4.2.*

Proof. The proof uses standard compactness-uniqueness arguments in a probabilistic point of view and will not be developed here. The main key is Proposition 4.8. \square

We finally conclude by our main result which proves that for a good asymptotics $(n, \delta(n))$, the laws of the stochastic interacting particle system converge when n tends to infinity to the law P associated with the solution of the full Boltzmann equation. More precisely we obtain from Theorems 4.7 and 4.9 the following.

Theorem 4.10. *Let us assume $(H_1), (H_2), (H_3)$ and that P_0 has a second order moment. Let $n \in \mathbb{N}^*$ and consider a sequence of positive real numbers $\delta(n)$ which tends to zero in an asymptotics such that $\frac{1}{n} e^{\frac{\|B\|_\infty T}{\delta(n)^3}}$ tends to zero when n tends to infinity, then*

1) *For every $1 \leq i \leq n$, the sequence of laws of $Z^{\delta(n), in}$ converges in $\mathcal{P}(\mathbb{D}([0, T], \mathbb{R}^6))$ to the probability measure P defined in Theorem 4.2.*

2) *The empirical measures of the interacting particle system $(Z^{\delta(n), in})_{1 \leq i \leq n}$ converge to P in the space $\mathcal{P}(\mathbb{D}([0, T], \mathbb{R}^6))$.*

Two algorithms for the Boltzmann equation

We deduce from the above study two algorithms associated either with the simple mean-field interacting particle system or with the binary mean-field interacting particle system. The description of the algorithms is the same in both cases, since the theoretical justification is unified for the two systems.

As seen previously, the empirical measure $\mu^{\delta(n),n}$ approximates the law of the Boltzmann process whose marginal at time t is equal to the solution $f(t, \cdot)$ of the Boltzmann equation.

As in Subsection 3.2, we simulate the particle system. For a fixed n , let $\Lambda_{\delta(n)} = \frac{\|B\|_\infty}{\delta(n)^3}$. A Poisson process of rate $\frac{n\Lambda_{\delta(n)}}{2}$ gives the sequence of collision times. At each of these times, we choose uniformly the pair of particles which interact, update the particles under the free transport, and simulate the jump operator by an acceptance-rejection procedure, exactly as described above.

The dependence of $\delta(n)$ on n is certainly not optimal (see Section 5.3 of these notes for a numerical study and comments) but that suggests that such an algorithm of simulation can be considered.

5. A GENERALIZED KAC EQUATION WITHOUT CUTOFF

5.1. The set-up

We are now interested in omitting the boundedness assumption on the cross section B . The full Boltzmann equation is then very difficult to study and no existence theorem of global solutions has been proved so far. Therefore, we restrict ourselves here to the study of a spatially homogeneous equations in dimension one.

We consider in fact the simplest case of the non cutoff Kac equation

$$\frac{\partial f}{\partial t} = K_\beta(f, f), \quad (5.1)$$

where $f \equiv f(t, v)$, $t \geq 0$, $v \in \mathbb{R}$, and

$$K_\beta(f, f)(t, v) = \int_{w \in \mathbb{R}} \int_{\theta = -\pi}^{\pi} \left(f(t, v^*) f(t, w^*) - f(t, v) f(t, w) \right) \beta(\theta) d\theta dw \quad (5.2)$$

with

$$v^* = v \cos \theta - w \sin \theta, \quad w^* = v \sin \theta + w \cos \theta.$$

The cross section $\beta : [-\pi, \pi] - \{0\} \mapsto \mathbb{R}_+$ is an even function such that

$$\int_0^\pi \theta^2 \beta(\theta) d\theta < +\infty,$$

but not necessarily $\int_0^\pi \beta(\theta) d\theta < +\infty$. This justifies the terminology of “non cutoff” Kac equation. (When $\int_0^\pi \beta(\theta) d\theta < +\infty$, the equation is said to be “with cutoff”).

The following existence theorem for (5.1) without cutoff can be found in [7].

Theorem 5.1. *Let $f_0 \geq 0$ be such that*

$$\int_{v \in \mathbb{R}} f_0(v) (1 + |v|^2 + |\log f_0(v)|) dv < +\infty$$

and let β be a cross-section satisfying the following property:

$$\exists C_0, C_1 > 0, r \in]1, 3[, \text{ such that } \forall \theta \in]0, \pi[, C_0 |\theta|^{-r} \leq \beta(\theta) \leq C_1 |\theta|^{-r}. \quad (5.3)$$

Then there exists a nonnegative solution $f^\beta(t, v) \in L^\infty([0, +\infty[; L^1(\mathbb{R}_v, (1 + |v|^2)dv)) \cap C([0, +\infty[; \mathcal{D}'(\mathbb{R}_v))$ to (5.1) with initial datum f_0 in the following weak sense: for any $\phi \in C_b^2(\mathbb{R}_v)$,

$$\frac{\partial}{\partial t} \int_{v \in \mathbb{R}} f^\beta(t, v) \phi(v) dv = \int_{v \in \mathbb{R}} \int_{w \in \mathbb{R}} K_\beta^\phi(v, w) f^\beta(t, v) f^\beta(t, w) dw dv, \tag{5.4}$$

where

$$K_\beta^\phi(v, w) = \int_{-\pi}^\pi \left(\phi(v \cos \theta - w \sin \theta) - \phi(v) + w \sin \theta \phi'(v) \right) \beta(\theta) d\theta. \tag{5.5}$$

2) Moreover, if for all $p \in \mathbb{N}$, there exists $C(p)$ such that

$$\int_{v \in \mathbb{R}} f_0(v) \left(1 + |v|^p + |\log f_0(v)| \right) dv \leq C(p), \tag{5.6}$$

every non-negative solution of (5.1) belongs to $L^\infty([0, +\infty[; C^\infty(\mathbb{R}_v))$.

Remark 5.2. 1) There is a compensated term in the operator (5.5), necessary to give a sense to (5.5). If we moreover assume that $\int_0^\pi \theta \beta(\theta) d\theta < +\infty$, then $\int_{-\pi}^\pi \sin \theta \beta(\theta) d\theta$ is well-defined and equal to 0 (since β is even), and eq. (5.5) can be rewritten as $K_\beta^\phi(v, w) = \int_{-\pi}^\pi \{ \phi(v^*) - \phi(v) \} \beta(\theta) d\theta$.

2) In Theorem 5.1, the solution is obtained by a compactness result, and there is no uniqueness result, except in the cutoff case $\beta \in L^1([0, \pi])$, for which one has existence and uniqueness of the solution even if $f_0 |\log f_0|$ is not integrable, see [7], Appendix A.

We will prove the existence and uniqueness of a solution of the associated martingale problem for every initial probability measure with a second order moment. Next, applying the previous results of Section 3, we will exhibit some simulable interacting particle systems whose laws converge to a solution of the Kac equation, and give a precise rate of convergence. Since it is not possible to directly simulate eq. (5.1) when $\int_0^\pi \beta(\theta) d\theta = +\infty$, we introduce a cutoff equation by considering $\beta_\ell = \beta \wedge \ell$ and we simulate its solution using Nanbu's or Bird's systems of n stochastic particles. When $\ell \rightarrow 0$ and $n \rightarrow +\infty$ in a good asymptotics, the laws of the particle system converge to the solution of the non cutoff equation. Estimates of convergence are also given.

Sznitman in [31] studies a spatially homogeneous hard-sphere Boltzmann equation taking in account the large velocities. In that model, there is no angular dependence of the collision kernel. He obtains convergence results without estimates using a compactness-uniqueness method.

Let us denote by $\mathcal{P}_2(\mathbb{R})$ the space of probability measures with a second moment. We again define probability measure solutions of the Kac equation and the associated nonlinear martingale problem.

Definition 5.3. Let β be a cross section such that $\int_0^\pi \theta^2 \beta(\theta) d\theta < +\infty$ and P_0 in $\mathcal{P}_2(\mathbb{R})$, and K_β^ϕ be as in (5.5).

A probability measure flow $(P_t)_{t \geq 0}$ is said to solve eq. (5.1) if for any ϕ in $C_b^2(\mathbb{R})$,

$$\langle P_t, \phi \rangle = \langle P_0, \phi \rangle + \int_0^t \langle P_s \otimes P_s, K_\beta^\phi \rangle ds = \langle P_0, \phi \rangle + \int_0^t \langle P_s(dv) P_s(dw), K_\beta^\phi(v, w) \rangle ds. \tag{5.7}$$

We are now looking for a Markov process $(V_t)_{t \geq 0}$ whose time-marginal laws are solution of (5.7). We have then to study the existence of its law $P \in \mathcal{P}(\mathbb{D}(\mathbb{R}_+, \mathbb{R}))$ starting at P_0 in $\mathcal{P}_2(\mathbb{R})$ such that for any ϕ in $C_b^2(\mathbb{R})$,

$$\phi(V_t) - \phi(V_0) - \int_0^t \langle P_s(dw), K_\beta^\phi(V_s, w) \rangle ds \tag{5.8}$$

is a square-integrable martingale. Here, P_s denotes the marginal of P .

Note that if P solves (5.8), then $(P_t)_{t \geq 0}$ solves (5.7).

Since the function β has a moment of order 2, we can develop some probabilistic arguments based on computations in the space L^2 and prove the existence and uniqueness of a probability measure satisfying (5.8). Let us fix $T > 0$ and define $\mathbb{D}_T = \mathbb{D}(\mathbb{R}_+, \mathbb{R})$.

Theorem 5.4. *Let β be a cross section such that $\int_0^\pi \theta^2 \beta(\theta) d\theta < +\infty$, and suppose that $P_0 \in \mathcal{P}_2(\mathbb{R})$. Then, there exists a unique probability measure $P^\beta \in \mathcal{P}(\mathbb{D}_T)$ with initial datum P_0 satisfying (5.8).*

Moreover, P^β is in $\mathcal{P}_2(\mathbb{D}_T)$, and the flow $(P_t^\beta)_{t \geq 0}$ is a measure solution to eq. (5.1) in the weak sense of Definition 5.3. This flow satisfies the following properties of momentum and energy: for any $t \in \mathbb{R}_+$, $\langle P_t^\beta(dv), v \rangle = \exp(-bt) \langle P_0(dv), v \rangle$ and $\langle P_t^\beta(dv), v^2 \rangle = \langle P_0(dv), v^2 \rangle$. Finally, if $\langle P_0(dv), |v|^p \rangle < +\infty$ for $p \geq 2$, then P^β is in $\mathcal{P}_p(\mathbb{D}_T)$.

If we assume that the function β is sufficiently singular at zero with appropriate controls on the derivatives, for example

$$\beta(\theta) = \frac{k_0}{|\theta|^r} 1_{[-\theta_0, +\theta_0]} + \beta_1(\theta), \quad \theta \neq 0, \tag{5.9}$$

with $r \in]1, 3[$, $k_0 > 0$, $\beta_1 \geq 0$ and $0 < \theta_0 < \frac{\pi}{2}$, (hypothesis inspired from (5.3)), then we obtain the following.

Theorem 5.5. *Let the second order moment initial data P_0 be given and different of δ_0 . Then for every $t > 0$, the law P_t^β has a density $f^\beta(t, \cdot)$ with respect to the Lebesgue measure.*

Since P_t^β solves (5.7), this result has an immediate corollary generalizing the existence result in Theorem 5.1 to non-density initial data:

Corollary 5.6. *Let $P_0 \in \mathcal{P}_2(\mathbb{R})$ and $P_0 \neq \delta_0$. Then there exists a solution $f \in L^\infty(]0, +\infty[_t, L^1((1 + |v|^2)dv))$ to (5.4) with initial data P_0 , and $f(t, \cdot)$ is a probability density function for each $t > 0$.*

Theorem 5.5 is proved by using the stochastic calculus of variations, which is not the topic of this notes. The proof can be found in [16]. Let us remark that the interest of the probabilistic approach lies in particular in the very general initial data we can consider.

Next, we are interested in the regularity of the density of V_t . Given Theorem 5.5 and its Corollary 5.6, the proof of the regularity result in Desvillettes [7] Theorem 3.1 (result recalled in Theorem 5.1) remains valid for our solution, since it only uses the existence of the solution and the moment conditions, but not the integrability of $f_0 |\log f_0|$. We obtain:

Corollary 5.7. *Let all the moments of V_0 be finite and $\mathcal{L}(V_0) \neq \delta_0$. For $t > 0$, let $f^\beta(t, \cdot)$ denote the density of $\mathcal{L}(V_t)$. Then f^β belongs to $L^\infty(]0, +\infty[_t; C^\infty(\mathbb{R}_v))$.*

The proof of Desvillettes shows that the Fourier transform \hat{f} of any solution f of (5.1) satisfies that $|y|^n |\hat{f}(y)|$ is integrable for any n (t is fixed, and not written down), which classically implies the result.

By a probabilistic Malliavin calculus proof we get a slightly better regularity result under more stringent assumptions, which is a variant of Corollary 5.7:

Theorem 5.8. *Assume that all the moments of V_0 are finite, $\mathcal{L}(V_0) \neq \delta_0$, and*

$$\int_{\frac{\pi}{3}}^\pi \left| \frac{\pi}{2} - \theta \right|^{-k} \beta_1(\theta) d\theta < +\infty, \quad \forall k \in \mathbb{N}.$$

For $t > 0$, let $f^\beta(t, \cdot)$ denote the density of $\mathcal{L}(V_t)$. Then f^β belongs to $L^\infty(]0, +\infty[_t; C_{bv}^\infty(\mathbb{R}_v))$ (finite variation functions), and solves (5.1) with initial data $\mathcal{L}(V_0)$.

5.2. Stochastic approximations for the Kac equation without cutoff

We will approximate P^β by a simulable interacting particle system. As an intermediate step, we introduce approximations with cutoff.

Convergence of approximations with cutoff

We consider cross sections $(\beta_\ell)_{\ell \geq 0}$ and β , and set

$$\delta_\ell = \int_{-\pi}^{\pi} (1 - \cos \theta) |\beta - \beta_\ell|(\theta) d\theta, \quad c_\ell = \int_{-\pi}^{\pi} (1 - \cos \theta) (\beta \wedge \beta_\ell)(\theta) d\theta. \quad (5.10)$$

We endow $\mathcal{P}_2(\mathbb{R})$ with a standard metric, called Vaserstein metric or Kantorovich-Rubinstein metric

$$\rho(p, q) = \inf \left\{ \left(\int_{\mathbb{R} \times \mathbb{R}} ((x - y)^2 \wedge 1) r(dx, dy) \right)^{1/2} : r \text{ has marginals } p \text{ and } q \right\}$$

for which $\mathcal{P}_2(\mathbb{R})$ is a complete space. We also define an analogous metric on the path space. If P and Q are probability measures on $\mathbb{D}([0, T], \mathbb{R})$ with second order moments,

$$\rho_T(P, Q) = \inf \left\{ \left(\int_{\mathbb{D}_T \times \mathbb{D}_T} \sup_{0 \leq t \leq T} |X_t - Y_t|^2 R(dX, dY) \right)^{1/2} : R \text{ has marginals } P \text{ and } Q \right\}$$

Theorem 5.9. *Let $P_0 \in \mathcal{P}_2(\mathbb{R})$ be given, and let P^β and P^{β_ℓ} be defined in Theorem 5.4 with cross sections β and β_ℓ respectively. Then*

$$\sup_{0 \leq t \leq T} \rho(P_t^{\beta_\ell}, P_t^\beta)^2 \leq \rho_T(P^{\beta_\ell}, P^\beta)^2 \leq (16\delta_\ell T + 2\delta_\ell^2 T^2) \exp(16c_\ell T + 2c_\ell^2 T^2) \langle P_0(dv), v^2 \rangle.$$

Hence if $\lim_{\ell \rightarrow \infty} \delta_\ell = 0$, then $\lim_{\ell \rightarrow \infty} \sup_{0 \leq t \leq T} \rho(P_t^{\beta_\ell}, P_t^\beta) = \lim_{\ell \rightarrow \infty} \rho_T(P_t^{\beta_\ell}, P_t^\beta) = 0$. This is the case when the β_ℓ are cutoff versions of β , such as $\beta \wedge \ell$ or $\beta(\theta) \mathbf{1}_{|\theta| \geq 1/\ell}$.

Proof. The proof uses a coupling technique and some stochastic pathwise computation. One constructs on the same probability space two processes V^l and V with law respectively P^{β^l} and P^β , such that

$$E \left(\sup_{0 \leq t \leq T} |V_s^l - V_s|^2 \right) \leq (16\delta_\ell T + 2\delta_\ell^2 T^2) \exp(16c_\ell T + 2c_\ell^2 T^2) \langle P_0(dv), v^2 \rangle.$$

We will not develop the proof in more details. □

Convergence estimates for particle systems

1) The cutoff case

We consider first the simpler cutoff Kac equation for which $\beta \in L^1([0, \pi])$. We are able to describe some simulable interacting particle systems whose laws converge to P^β when the size of the system tends to infinity.

Under the cutoff assumption $\int_0^\pi \beta(\theta) d\theta < +\infty$, we define two different mean-field interacting particle systems which approximate the solution of the nonlinear martingale problem (5.8). Let $\mathbf{v}^n = (v_1, v_2, \dots, v_n)$ be the generic point in \mathbb{R}^n , and $\mathbf{e}_i : h \in \mathbb{R} \mapsto \mathbf{e}_i, h = (0, \dots, 0, h, 0, \dots, 0) \in \mathbb{R}^n$ with h at the i -th place. We consider $\phi \in C_b(\mathbb{R}^n)$.

The simple mean-field system is a Markov process in $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^n)$ with generator

$$\frac{1}{n-1} \sum_{1 \leq i \neq j \leq n} \int_{-\pi}^{+\pi} \left(\phi(\mathbf{v}^n + \mathbf{e}_i (v_i (\cos \theta - 1) - v_j \sin \theta)) - \phi(\mathbf{v}^n) \right) \beta(\theta) d\theta. \quad (5.11)$$

The binary mean-field system is a Markov process on the same space with generator

$$\frac{1}{n-1} \sum_{1 \leq i \neq j \leq n} \int_{-\pi}^{+\pi} \frac{1}{2} \left(\phi(\mathbf{v}^n + \mathbf{e}_i \cdot (v_i(\cos \theta - 1) - v_j \sin \theta) + \mathbf{e}_j \cdot (v_j(\cos \theta - 1) - v_i \sin \theta)) - \phi(\mathbf{v}^n) \right) \beta(\theta) d\theta. \quad (5.12)$$

We denote in both cases the Markov process by $V^{\beta,n} = (V^{\beta,1n}, \dots, V^{\beta,nn})$, and by $|\cdot|_T$ the variation norm in the space of signed measures on $\mathbb{D}([0, T], \mathbb{R})$. Theorem 3.4 yields in this context

Theorem 5.10. *1) Let $(V_0^{\beta, in})_{1 \leq i \leq n}$ be i.i.d. with law P_0 . Then we have propagation of chaos in strong sense: for given $T > 0$ and $k \in \mathbb{N}^*$,*

$$|\mathcal{L}(V^{\beta,1n}, \dots, V^{\beta,kn}) - (P^\beta)^{\otimes k}|_T \leq K k^2 \frac{\exp(\|\beta\|_1 T)}{n}, \quad (5.13)$$

where P^β is the unique solution of the nonlinear martingale problem (5.8) with initial law P_0 . Here, K denotes a constant independent of k, T, β, n .

2) The empirical measures defined by $\mu^{\beta,n} = \frac{1}{n} \sum_{i=1}^n \delta_{V^{\beta, in}}$ converge to P^β in $\mathcal{P}(\mathbb{D}([0, T], \mathbb{R}))$ with an estimate of convergence in $\sqrt{K \exp(\|\beta\|_1 T) / \sqrt{n}}$.

2) The case without cutoff

We consider here a cross section β satisfying $\beta(\theta) \leq C_1 |\theta|^{-r}$ for some $C_1 > 0$ and $r \in]1, 3[$, and the cutoff approximations $\beta_\ell(\theta) = \beta(\theta) \mathbf{1}_{\frac{1}{2} \leq |\theta|}$. Then $\beta_\ell \in L^1([0, \pi])$ and

$$\|\beta_\ell\|_1 = \int_{-\pi}^{+\pi} \beta_\ell(\theta) d\theta \leq \frac{2C_1}{r-1} (\ell^{r-1} - \pi^{1-r}).$$

With every function β^ℓ , we can associate a particle system $(V^{\beta_\ell, n})$. Theorems 5.10 and 5.9 yield the following.

Theorem 5.11. *Let β be a cross section such that $\beta(\theta) \leq C_1 |\theta|^{-r}$ for some $C_1 > 0$ and $r \in]1, 3[$, and $\ell(n)$ be a sequence of integers going to $+\infty$ in such a way that $\exp(\frac{2C_1}{r-1} \ell(n)^{r-1} T) = o(n)$. Let $(V_0^{\beta_{\ell(n)}, in})_{1 \leq i \leq n}$ be i.i.d. with a second order law P_0 .*

1) For every $k \in \mathbb{N}^*$, the sequence $\mathcal{L}(V^{\beta_{\ell(n)}, 1n}, \dots, V^{\beta_{\ell(n)}, kn})$ converges to $(P^\beta)^{\otimes k}$, P^β being defined in Theorem 5.4 with initial datum P_0 . Moreover we have the convergence estimate

$$\begin{aligned} \sup_{0 \leq t \leq T} \rho(\mathcal{L}(V_t^{\beta_{\ell(n)}, kn}), P_t^\beta) &\leq \rho_T(\mathcal{L}(V^{\beta_{\ell(n)}, kn}), P^\beta) \\ &\leq K \left(\frac{\exp(\frac{2C_1}{r-1} \ell(n)^{r-1} T)}{n} + (16\delta_{\ell(n)} T + 2\delta_{\ell(n)}^2 T^2) \exp(16bT + 2b^2 T^2) \langle P_0(dv), v^2 \rangle \right) \end{aligned}$$

where $\delta_\ell \leq 2C_1 \int_0^{1/\ell} (1 - \cos \theta) \theta^{-r} d\theta$ tends to zero when ℓ tends to infinity since $r \in]1, 3[$.

2) The empirical measures $\mu^{\beta_{\ell(n)}, n}$ defined in Theorem 5.10 converge to P^β in $\mathcal{P}(\mathbb{D}([0, T], \mathbb{R}))$.

The simulation algorithms

We deduce from the above study two algorithms associated respectively with the simple mean-field interacting particle system and the binary mean-field interacting particle system. The description of the algorithms is the same in both cases, since the theoretical justification is unified for the two systems.

As seen previously, the empirical measures $\mu^{\beta_{\ell(n)}, n}$ approximate the law of the Kac process whose marginal at time t is equal to the solution $f(t, \cdot)$ of the Kac equation.

We simulate the particle system of size n . The total jump rate is $n \|\beta_{\ell(n)}\|_1$ for (5.11) and $n \|\beta_{\ell(n)}\|_1 / 2$ for (5.12). A Poisson process of same rate gives the sequence of collision times, at each of which we choose uniformly

among the $n(n-1)/2$ possibilities the pair of particles which collide. We then choose the impact parameter θ according to $\beta_{\ell(n)}(\theta) d\theta / \|\beta_{\ell(n)}\|_1$, and in the simple mean-field particle system we only update the velocity of one of the colliding particles, while in the binary one we update both. This simulation is exact if we simulate exactly the exponential variables related to the Poisson process, instead of discretizing time.

We refer to Section 1.6 for further details about the actual simulation.

Remark 5.12. Fournier, in [11] and [12] has generalized the previous approach in the case of a Boltzmann equation without cutoff in dimension 2. He proves in a similar way the existence and uniqueness of the law of the underlying Markov process and the existence and regularity of a flow of densities for this law, then proving the existence of solutions for the equation under very weak initial conditions. The techniques are more intricate, because of the dimension 2, but ideas are similar as in dimension one.

Monte-Carlo approximations by stochastic interacting particle systems are obtained in this context.

5.3. Numerical results

We are going to illustrate the results we have obtained on some one-dimensional models. We stress that the Monte-Carlo methods are essentially independent from the dimension of the space. Indeed, the speed of convergence is always of order $1/\sqrt{n}$, and the actual computer programs can be written in rather a straightforward fashion, whatever the dimension.

5.3.1. The classical Kac equation with cutoff

We give a short computer program, written in c++, which we used in the computer sessions for the ASCII. It is very simple and short, and not optimized in any way. It concerns the classical Kac equation, for which $\beta(\theta) = 1/2\pi$. We thank Christophe Prud'homme for helping us write this program.

```
#include <cstdlib>
#include <cmath>
#include <time.h>
#include <sstream>
#include <iostream>
#include <string>
#include <fstream>
#include <time.h>

const double pi2 = 8*atan(1.0);
const double sqrt3 = sqrt(3.);

void
save(double const* part, int Npart, int k)
{
    stringstream cstr;
    cstr << "hist_" << k << ".mtv";
    ofstream fout(cstr.str());
    fout << "$ DATA=HISTOGRAM

% binwidth=";

    double binwidth = 0.05;
    fout << binwidth << "\n";
    fout << "

% binstart=-5.025";
```

```

    for(int i = 0;i < Npart;i++)
        fout << part[i] << "\n";
    fout << "$END";
}

void
compute(double tt, int Nvis, double* part, int Npart)
{
    double freq = tt/Nvis;

    // elapsed time
    double S = 0;
    double lambda = 2./Npart;

    int k = 0;
    while(k <= Nvis)
    {
        S -= log(drand48()*lambda);

        while (S > k*freq && k <= Nvis)
        {
            save(part,Npart,k);
            k++;
        }
        int index1, index2;

        while ((index1 = drand48()*Npart) == (index2 = drand48()*Npart));
        double w1 = part[index1];
        double w2 = part[index2];
        double theta = pi2*drand48();

        part[index1] = cos(theta)*w1-w2*sin(theta);
        part[index2] = sin(theta)*w1+w2*cos(theta);
    }
}

void
init(int cond, double* part, int Npart)
{
    switch(cond)
    {
        case 1:
        {
            for(int i = 0;i < Npart;i++)
                if (drand48() < 0.5)
                    part[i] = -1;
                else
                    part[i] = 1;
        }
    }
}

```

```

    }
    break;

    case 2:
    {
        for(int i = 0;i < Npart;i++)
            {
                double t = drand48();
                if (t < 0.5)
                    part[i] = 0;
                else if (t < 0.8333333333333333)
                    part[i] = -1;
                else
                    part[i] = 2;
            }
    }

    break;
    case 3:
        for(int i = 0;i < Npart;i++)
            part[i]= sqrt3*(2*drand48()-1);
        break;
    default:
        cout << "wronf initial condition either 1 , 2 or 3\n";
    }
}

int
main(int argc, char **argv)
{
    srand48(time(0));
    ofstream fout("hist.mtv");

    int N,Npart,cond;
    double lambda;
    double total_time;

    cout << "total_time ? "; cin >> total_time; //
    cout << "number of visualization ? ";cin >> N; // number of visualization
    cout << "number of particles ? ";cin >> Npart; //number of particles
    cout << "initial condition [123] ? ";cin >> cond; //choice initial data

    // velocity array of particle
    double *part = new double[Npart];

    init(cond,part,Npart);

    compute(total_time, N, part, Npart);
}

```

5.3.2. The generalized Kac equation without cutoff

This part has been developed by Desvillettes in [8]. We do not include the program because of its length, but it is very similar to the previous one. We focus on the study of the dependency between the number n of particles we take and the numerical cutoff related to the parameter $\ell(n)$.

In Theorem 5.11, a criterion on the function $n \rightarrow \ell(n)$ is established, in order to ensure the convergence of the algorithms described in the previous subsection when $n \rightarrow +\infty$ towards the solution of the non cutoff Kac equation. In this part, we study how to choose, in practice, the dependence of ℓ with respect to n , in order to optimize the computations.

We select a typical solution of the non cutoff Kac equation (5.1), (5.2). We choose

$$\beta(\theta) = |\sin \theta|^{-2} \mathbf{1}_{\{\theta \in [-\pi/2, \pi/2]\}} (2\pi)^{-1}$$

as a typical non cutoff cross section. Note that it is not integrable and does not have a first moment.

We also choose the initial datum with very simple particle discretization

$$f_0(v) = \mathbf{1}_{\{v \in [-1/2, 1/2]\}}.$$

The corresponding solution of Kac equation is denoted by $f(t, v)$.

We also introduce for $\ell > 1$ the cutoff cross section $\beta_\ell(\theta) = \beta(\theta) \mathbf{1}_{\{|\theta| \geq 1/\ell\}}$, and the corresponding solution $f^\ell(t, v)$ of the cutoff Kac equation (with the same initial datum).

The mass and energy of f as well as f^ℓ are independent of t and given by

$$a_0^f(t) = a_0^{f^\ell}(t) = \int_{\mathbb{R}} f(t, v) dv = 1, \quad \frac{a_2^f(t)}{2} = \frac{a_2^{f^\ell}(t)}{2} = \int_{\mathbb{R}} f(t, v) \frac{|v|^2}{2} dv = \frac{1}{24}.$$

Therefore, f and f^ℓ have the same (Gaussian) limit when t tends to infinity, given by

$$\lim_{t \rightarrow +\infty} f(t, v) = \lim_{t \rightarrow +\infty} f^\ell(t, v) = \sqrt{\frac{6}{\pi}} e^{-6|v|^2}.$$

The fact that f and f^ℓ are identical at times 0 and $+\infty$ makes it difficult to choose a time t_0 where it is interesting to compare $f(t_0, \cdot)$ and $f^\ell(t_0, \cdot)$, that is, a time t_0 such that $\|f(t_0, \cdot) - f^\ell(t_0, \cdot)\|$ is of the same order of magnitude as $\sup_{t \in \mathbb{R}} \|f(t, \cdot) - f^\ell(t, \cdot)\|$, for some reasonable norm $\|\cdot\|$. In our case, after an empirical study, we choose $t_0 = 1.8$.

For the initial datum chosen here, the only known explicitly computable quantities (depending on f or f^ℓ) for an arbitrary time t are the moments of order $2N$, where $N \in \mathbb{N}$ (cf. [37]), that is

$$a_{2N}^f(t) = \int_{\mathbb{R}} f(t, v) |v|^{2N} dv, \quad a_{2N}^{f^\ell}(t, v) = \int_{\mathbb{R}} f^\ell(t, v) |v|^{2N} dv.$$

But a_0^f and a_2^f (as well as $a_0^{f^\ell}$ and $a_2^{f^\ell}$) are independent of t , so that the first moment which is explicitly computable and really depending on time is $a_4^f(t)$ (and $a_4^{f^\ell}(t)$). The formulas are the following:

$$a_4^f(t) = \frac{1}{48} (1 - e^{-t/2}) + \frac{1}{80} e^{-t/2}, \quad a_4^{f^\ell}(t) = \frac{1}{48} (1 - e^{-R_\ell t/2}) + \frac{1}{80} e^{-R_\ell t/2}; \quad R_\ell = 1 - \frac{1}{2\pi\ell} - \frac{1}{4\pi} \sin\left(\frac{2}{\ell}\right). \quad (5.14)$$

We shall compare these theoretical values with the values obtained by the Nanbu (simple mean-field) algorithm.

The initial datum is discretized under the form $f_0(v) := \frac{1}{n} \sum_{i=0}^{n-1} \delta_{\frac{i}{n} - \frac{1}{2}}(v)$, and the Poisson process corresponding to the Nanbu algorithm is implemented in the way described in Section 5.2: at each iteration, two

particles are selected randomly (with a uniform law), an exponential time is added to a time counter, and the velocity of only one particle is changed (except if the time counter becomes bigger than t_0), according to the usual rule of collisions (i.e., eq. (5.2)). The angle θ used in this collision is taken randomly according to the cutoff cross section β_ℓ .

We then get a discretized version of f^ℓ , denoted by $\tilde{f}^{\ell,n}(t_0, v) = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{v_i(t_0)}(v)$, and the corresponding fourth moment is computed by the formula $a_4^{\tilde{f}^{\ell,n}}(t_0) = \frac{1}{n} \sum_{i=0}^{n-1} v_i(t_0)^4$. We are now interested in the behavior of the quantity $|a_4^{\tilde{f}^{\ell,n}}(t_0) - a_4^f(t_0)|$ when ℓ and n vary.

More precisely, we choose to estimate how ℓ and n have to be related in order to give an error of discretization and an error due to the cutoff which are the same.

It means that we try to find the quantity $\ell(n)$ (when n varies) such that

$$| \langle a_4^{\tilde{f}^{\ell(n),n}}(t_0) \rangle - a_4^{f^{\ell(n)}}(t_0) | = | a_4^{f^{\ell(n)}}(t_0) - a_4^f(t_0) |. \tag{5.15}$$

In this equality, the right-hand side quantity is explicitly computable thanks to eq. (5.1), (5.2), and the notation $\langle \dots \rangle$ means the mean value “over all possible experiments”.

Of course, in order to estimate the quantity $\langle a_4^{\tilde{f}^{\ell,n}}(t_0) \rangle$ (for a given ℓ, n), we can carry out only a finite number of numerical experiments.

Therefore, for each n , we choose a number $m(n)$ of simulations, made each time with a different set of random numbers. The corresponding mean value is denoted by $\langle a_4^{\tilde{f}^{\ell,n}}(t_0) \rangle_{m(n)}$, and replaces $\langle a_4^{\tilde{f}^{\ell,n}}(t_0) \rangle$ when we try to estimate $\ell(n)$ in such a way that (5.16) holds. The number $m(n)$ is chosen as large as possible. It is limited by the speed of the computer.

In order to find $\ell(n)$, we use a fixed point method, (this is easy since the dependence in ℓ of the values of $|\langle a_4^{\tilde{f}^{\ell,n}}(t_0) \rangle_{m(n)} - a_4^f(t_0)|$ is almost undetectable as soon as ℓ is confined in a “reasonable” interval).

In this process, we can also compute a confidence interval $[\ell^+(n), \ell^-(n)]$, in which $\ell(n)$ lies with a “large” probability.

We now present the numerical results. For each n belonging to a geometric progression, we give $m(n)$, and the computed quantities $\ell^+(n)$, $\ell(n)$, and $\ell^-(n)$.

n	m(n)	$\ell_1^+(n)$	$\ell_1(n)$	$\ell_1^-(n)$
125	5E5	1.6085	1.6135	1.6185
250	5E5	1.954	1.958	1.962
500	2E5	2.420	2.426	2.431
1E3	1E5	3.0530	3.0595	3.0645
2E3	5E4	3.921	3.931	3.940
4E3	5E4	5.125	5.135	5.150
8E3	5E4	6.79	6.81	6.83
16E3	2E4	9.11	9.15	9.19
32E3	2E4	12.380	12.485	12.550
64E3	1E4	17.10	17.15	17.20
128E3	1E3	23.12	23.50	24.00
256E3	600	32.6	33.6	34.4
512E3	400	45.0	45.5	46.0
1024E3	300	64.0	65.0	66.0
2048E3	100	90.0	92.5	95.0

Table 1

We now display curves made with Table 1. In Figure 3, $\ell^+(n)$, $\ell(n)$, and $\ell^-(n)$ are represented as functions of n . In Figure 4, they are represented in a log/log scale. The dashed lines correspond to $\ell^+(n)$ and $\ell^-(n)$, while the continuous lines are related to $\ell(n)$.

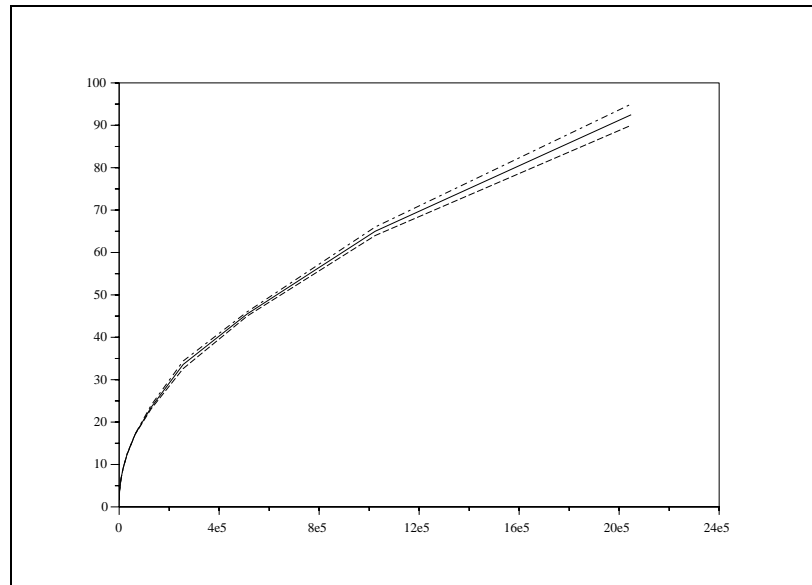


FIGURE 3. $\ell(n)$ as function of n

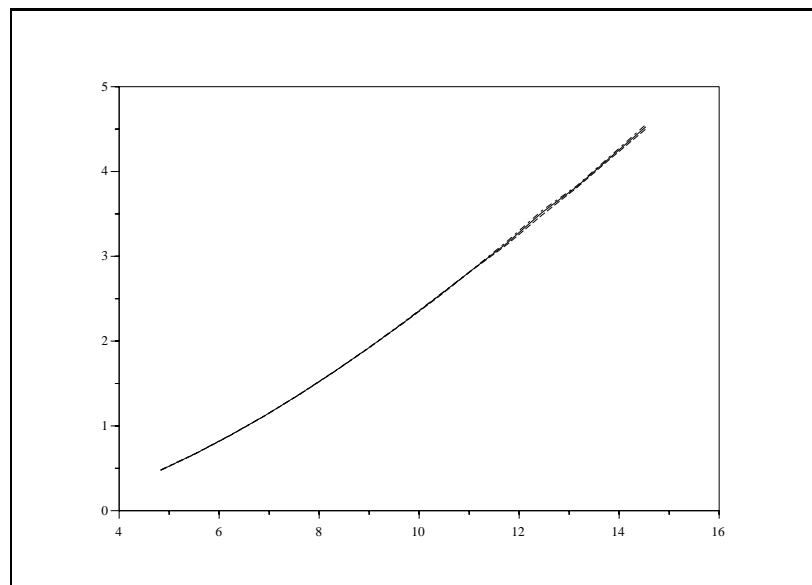


FIGURE 4. $\ell(n)$ as function of n in log/log scale

Fig. 3 clearly shows a concave curve, which is in accord with the guess that $\ell(n)$ should increase less rapidly than n . Remember that in Section 5.2, a sufficient condition of convergence of the method was that (up to different constants) $\exp(\ell(n)) = o(n)$ ($\alpha = 2$ in our example).

However, we can see on Fig. 4 that the curve giving $\ell(n)$ with respect to n is convex when represented on a log/log scale (and in fact almost a straight line). Therefore, a good approximation for $\ell(n)$ seems to be some power n^k , for $k \in]0, 1[$. This means of course that the condition $\exp(\ell(n)) = o(n)$ is not at all fulfilled, and suggests that Theorem 5.11 is far from optimal.

Of course our numerical study is limited and one should not draw hasty conclusions from it. We think however that in practice, a choice of $\ell(n)$ as a power of n might not be so bad.

REFERENCES

- [1] Babovsky, H.; Illner, R.: A convergence proof for Nanbu's simulation method for the full Boltzmann equation, *SIAM J. Numer. Anal.* 26, 45-65 (1989)
- [2] Bellomo, N.; Toscani, G.: on the Cauchy problem for the nonlinear Boltzmann equation. Global existence, uniqueness and asymptotic stability, *J. Math. Phys.* 26 (1985)
- [3] Bellomo, N.; Palczewski, A.; Toscani, G.: *Mathematical topics in nonlinear kinetic theory*, World scientific, Singapore (1988)
- [4] Bouleau, N.: *Probabilités de l'ingénieur, Variables aléatoires et simulation*, Hermann (1986)
- [5] Caprino, S; Pulvirenti, M.: A cluster expansion approach to a one-dimensional Boltzmann equation: a validity result, *Comm. Math. Phys.* 166, 603-621 (1995)
- [6] Cercignani, C.; Illner, R.; Pulvirenti, M.: *The mathematical theory of dilute gases*, Appl. Math. Sci., Springer, Berlin (1994)
- [7] Desvillettes, L: About the regularizing properties of the non cutoffed Kac equation. *Comm. Math. Phys.* 168, 416-440 (1995)
- [8] Desvillettes, L, Graham, C.; Méléard, S.: Probabilistic interpretation and numerical approximation of a Kac equation without cutoff, *Stoch. Proc. Appl.* 84, 115-136 (1999)
- [9] DiPerna, R.J.; Lions P-L.: On the Cauchy problem for Boltzmann equations: global existence and weak stability, *Ann. Math.* 130, 321-366 (1989)
- [10] Ethier, S.N.; Kurtz, T.G.: *Markov processes, Characterization and convergence*, Wiley (1986)
- [11] Fournier, N.: Existence and regularity study for a 2d spatially homogeneous Boltzmann equation without cutoff by a probabilistic approach, To appear in *Annals Appl. Probab.* (1999)
- [12] Fournier, N.: Strict positivity of a solution to a one-dimensional Kac equation without cutoff, *Prépublication 488 du Laboratoire de Probabilités, Paris 6* (1999)
- [13] Graham, C.: Nonlinear diffusion with jumps, *Ann. Inst. H. Poincaré* 28, 393-402 (1992)
- [14] Graham, C.; Méléard, S.: Chaos hypothesis for a system interacting through shared resources, *Probab. Theory Rel. Fields* 100, 157-173 (1984)
- [15] Graham, C.; Méléard, S.: Stochastic particle approximations for generalized Boltzmann models and convergence estimates, *Annals Probab.* 25, 115-132 (1997)
- [16] Graham, C.; Méléard, S.: Existence and regularity of a solution of a Kac equation without cutoff using Malliavin calculus, *Comm. Math. Phys.* 205, 551-569 (1999)
- [17] Hamdache, K.: Existence in the large and asymptotic behaviour for the Boltzmann equation, *Japan J. Appl. Math.* 2, 65-84 (1985)
- [18] Illner, R.; Neunzert, H.: On simulation methods for the Boltzmann equation, *Transport Theory Stat. Phys.* 16, 141-154 (1987)
- [19] Kaniel, S.; Shinbrot, M.: The Boltzmann equation I: uniqueness and global existence, *Comm. Math. Phys.*, 95, 117-126 (1978)
- [20] Lapeyre, B.; Pardoux, E.; Sentis, R.: *Méthodes de Monte-Carlo pour les équations de transport et de diffusion*, Mathématiques et Applications 29, Springer (1998)
- [21] Méléard, S.: Asymptotic behaviour of some interacting particle systems, McKean-Vlasov and Boltzmann models, cours du CIME mai 1995, Probabilistic models for nonlinear pde's, *Lect. Notes Math.* 1627, Springer (1996)
- [22] Méléard, S.: Stochastic approximations of the solution of a full Boltzmann equation for small initial data, *ESAIM, P&S* vol. 2, electronic journal, <http://www.emath.fr/ps/> (1998)
- [23] Méléard, S.: Convergence of the fluctuations associated with generalized mollified Boltzmann equations, *Stochastics and stochastic reports* 63, 195-225 (1998)
- [24] Mischler, S; Perthame, B.: Boltzmann equation with infinite energy, *SIAM J. Math. Analysis* 28 5, 1015-1027 (1997)
- [25] Nanbu, K.: Interrelations between various direct simulation methods for solving the Boltzmann equation, *J. Phys. Soc. Japan* 52, 3382-3388 (1983)
- [26] Neunzert, H.; Gropengeisser, F.; Struckmeier, J. : *Computational methods for the Boltzmann equation*. Applied and Indust. Maths, R. Spigler ed., Dordrecht : Kluwer Acad. Publ., 111-140 (1991)
- [27] Perthame, B.: Introduction to the theory of random particle methods for Boltzmann equation. In: *Progresses on Kinetic Theory*, World Scientific, Singapore (1994)
- [28] Perthame, B.; Pulvirenti, M.: On some large systems of random particles which approximate scalar conservation laws. To appear in *Asympt. Anal.*
- [29] Pulvirenti, M.; Wagner, W., and Zavelani Rossi, M.B.: Convergence of particle schemes for the Boltzmann equation, *Preprint* 49, Institut für Angewandte Analysis und Stochastik, Berlin (1993)

- [30] Rezakhanlou, F.: Kinetic limits for a class of interacting particle systems, *Probab. Theory Rel. Fields*, 104, 97-146 (1996)
- [31] Sznitman, A.S.: Equations de type Boltzmann spatialement homogènes, *Z. Wahrsch. Verw. Geb.* 66, 559-592 (1984)
- [32] Sznitman, A.S.: Topics in propagation of chaos. *Ecole d'été de Probabilités de Saint-Flour XIX - 1989*, Lect. Notes Math. 1464, Springer (1991)
- [33] Tanaka, H.: On the uniqueness of Markov process associated with the Boltzmann equation of Maxwellian molecules, *Proc. Intern. Symp., Kyoto*, 409-42 (1976)
- [34] Tanaka, H.: Probabilistic treatment of the Boltzmann equation of Maxwellian molecules, *Z. Wahrsch. Verw. Geb.* 46, 67-105 (1978)
- [35] Tanaka, H.: Limit theorems for certain diffusion processes with interaction. *Taniguchi Symp. on Stochastic Analysis, Katata*, 469-488 (1982)
- [36] Toscani, G.: On the nonlinear Boltzmann equation in unbounded domains, *Arch. Rat. Mech. Anal.* 95, 37-49 (1986)
- [37] Truesdell, C.: On the pressure and the flux energy in a gas according to Maxwell's kinetic theory II, *J. Rat. Mech. Anal.* 5, 55 (1980)
- [38] Uchiyama, K.: Derivation of the Boltzmann equation from particle dynamics, *Hiroshima Math. J.* 18, 245-297 (1988)
- [39] Wagner, W.: A convergence proof for Bird's direct simulation method for the Boltzmann equation. *J. Stat Phys.* 66, 1011-1044 (1992)