We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter

Stochastic Theory of Coarse-Grained Deterministic Systems: Martingales and Markov Approximations

Michel Moreau and Bernard Gaveau

Abstract

Many works have been devoted to show that Thermodynamics and Statistical Physics can be rigorously deduced from an exact underlying classical Hamiltonian dynamics, and to resolve the related paradoxes. In particular, the concept of equilibrium state and the derivation of Master Equations should result from purely Hamiltonian considerations. In this chapter, we reexamine this problem, following the point of view developed by Kolmogorov more than 60 years ago, in great part known from the work published by Arnold and Avez in 1967. Our setting is a discrete time dynamical system, namely the successive iterations of a measure-preserving mapping on a measure space, generalizing Hamiltonian dynamics in phase space. Using the notion of Kolmogorov entropy and martingale theory, we prove that a coarse-grained description both in space and in time leads to an approximate Master Equation satisfied by the probability distribution of partial histories of the coarse-grained state.

Keywords: stochastic theory, coarse-grained deterministic systems, Markov processes, martingales

1. Introduction

It is generally admitted that Thermodynamics and Statistical Physics could be deduced from an exact classical or quantum Hamiltonian dynamics, so that the various paradoxes related to irreversibility could also be explained, and nonequilibrium situations could be rigorously studied as well. These questions have been and still are discussed by many authors (see, for instance Refs. [1–4] and many classical textbooks, for instance [5–10]), who have introduced various plausible hypotheses [7–14], related to the ergodic principle [8–11], to solve them. It seems that there are two major kinds of problems. First, to justify that physical systems can reach an equilibrium state when they are isolated, or in contact with a thermal bath (which remains to be defined). Secondly, to justify various types of reduced stochastic dynamics, depending on the phenomena to be described: Boltzmann equations, Brownian motions, fluid dynamics, Bogoliubov-Born-Green-Kirkwood-Yvon (BBGKY) hierarchies, etc..: see for instance Refs [1–11, 15, 16]. Concerning the first type of problems (reaching an equilibrium, if any) very rough estimations

show [17] that the time scales to reach equilibrium, using only Hamiltonian dynamics and measure inaccuracies, are extremely large, contrarily to everyday experience, and quantum estimations are even worse [17]. Essentially, these times scale as Poincaré recurrence times and they increase as an exponential of the number of degrees of freedom (see Section 3 of this chapter for a brief discussion and references).

Here we concentrate on the second type of problems: is it possible to derive a stochastic Markovian process from an "exact" deterministic dynamics, just by coarse graining the microscopic state space? We generalize and complete the formalism recently presented [18] for Hamiltonian systems. Our framework is now more general and applies to all deterministic systems with a measure preserving dynamics, which, by Liouville theorem, include Hamiltonian dynamics.

Following Kolmogorov, we start with a measure space with a discrete time dynamics given by the successive iterations of a measure preserving mapping. The Kolmogorov entropy, or trajectory entropy, has been defined by Kolmogorov as an invariant of stationary dynamical systems (see Arnold and Avez book [19] for a pedagogical presentation). We follow his work and generalize part of his results. We also use martingale theory [20–23] to show that the stationary coarse-grained process almost surely tends to a Markov process on partial histories including n successive times, when n tends to infinity. From this result, we show that in the nonstationary situation, the probability distribution of such partial histories approximately satisfies a Master equation. Its probability transitions can be computed from the stationary distribution, expressed in terms of the invariant measure. It follows that, with relevant hypotheses, the mesoscopic distribution indeed tends to the stationary distribution, as expected.

Our next step is to coarse grain time also. The new, coarse-grained time step is now $n\,\tau$, τ being the elementary time step of the microscopic description, and n being the number of elementary steps necessary to approximately "erase" the memory with a given accuracy. The microscopic dynamics induces new dynamics on partial histories of length n. We show that it is approximately Markovian if n is large enough. This idea is a generalization of the Brownian concept: a particle in a fluid is submitted to a white noise force which is the result of the coarse-graining of many collisions, and the time step is thus the coarse-graining of many microscopic time steps [8, 24]. The Brownian motion emerges as a time coarse-grained dynamics.

In Section 2, we recall various mathematical concepts (Kolmogorov entropy, martingale theory) and use them to derive the approximate Markov property of the partial histories, and eventually to obtain an approximate Master Equation for the time coarse-grained mesoscopic distribution [18].

In Section 3, we briefly consider the problem of relaxation times and recall very rough estimations showing that an exact Hamiltonian dynamics predicts unrealistic, excessively large relaxation times [17], unless the description is completed by introducing other sources of randomness than the measure inaccuracies leading to space coarse-graining. Note that, following Kolmogorov [19], we do not address the Quantum Mechanics formalism.

2. Microscopic and mesoscopic processes in deterministic dynamics

2.1 Microscopic dynamics: Definitions and notations

It has been shown recently [18] that coarse-grained Hamiltonian systems can be approximated by Markov processes provided that they satisfy reasonable

properties, covering many realistic cases. These conclusions can be extended to a large class of deterministic systems generalizing classical Hamiltonian systems, which we now describe. We first specify our hypotheses and notations.

2.1.1 Deterministic microdynamics

Consider a deterministic system $\mathcal{S}$. Its states x, belonging to a state space X, will be called "microstates", in agreement with the usual vocabulary of Statistical Physics. The deterministic trajectory due to the microscopic dynamics transfers the microstate x_0 at time 0 to the microstate $x_t = \varphi_t(x_0)$ at time t. The evolution function φ_t satisfies the current properties of dynamic systems: $\varphi_t\,\varphi_s = \varphi_{t+s}$, $\varphi_0 = I$, t and s being real numbers and I being the identical function.

The dynamics is often invariant by time reversion, as assumed in many works on Statistical Physics: we refer to classical textbooks on the subject for details [5–8], but we will not use such properties in this chapter.

2.1.2 Microscopic distribution

Assume that the exact microstate x_0 is unknown at time 0, but is distributed according to the probability measure μ on the phase space X. The microscopic probability distribution μ_t at time t is given by

$$\mu_t\,(A) = \mu(\varphi_{-t}\,A). \tag{1}$$

for any measurable subset A of X. If μ is stationary, it is preserved by the dynamics: $\mu_t\,(A) = \mu\,(A)$. This condition, however, it not necessarily satisfied, in particular for physical systems during their evolution.

We will focus on two important cases:

a. the finite case: X is finite and consists in N microstates.

b. the absolutely continuous case: $X \subset R^n$, where (*i*) R is the set of real numbers and n is an integer (usually very large, and even in the case of Hamiltonian dynamics), and (*ii*) the measure μ is absolutely continuous with respect to the Lebesgue measure ω on R^n: these exists an integrable probability density $p(x)$ such that for any measurable subset A of X

$$\mu(A) = \int_A p(x)d\omega(x). \tag{2}$$

Furthermore, we assume that (*iii*) the Lebesgue measure of X (or volume of X) $V = \mathrm{vol}(X) \equiv \int_X d\omega(x)$ is finite, and (*iv*) the Lebesgue measure ω is preserved by the dynamics for any t and any measurable subset A of X.:

$$\mathrm{vol}A \ = \mathrm{vol}\,(\varphi_{-t}A). \tag{3}$$

The last two assumptions obviously generalize basic properties of Hamiltonian dynamics in a finite volume of phase space. Thus, by (1)–(3), the probability density is conserved along any trajectory: at time t the probability density is

$$p(x,t) \ = p\,(0, \varphi_{-t}x) \equiv p\,(\varphi_{-t}x). \tag{4}$$

2.1.3 Initial microscopic distribution: The stationary situation

Suppose that $\mathcal{S}$ is an isolated physical system and no observation was made on $\mathcal{S}$ at time 0 nor before 0. Then, in the absence of any knowledge on $\mathcal{S}$, we admit that at the initial time $\mathcal{S}$ is distributed according to the only unbiased probability law, which is the uniform law. This is clearly justified in the finite case, according to the physical meaning traditionally given to probability: in fact, attributing different probabilities for two distinct microstates of X would imply that some measurement would allow one to distinguish them objectively, which is not the case at time 0.

In the absolutely continuous case, initial uniformity is less obvious: it amounts to assuming that the system should be found with equal probability in two regions of the state space with equal volumes if no information allows one to give preference to any of these regions. This is of course a subjective assertion, but for Hamiltonian systems it agrees with the semi-quantum principle which asserts that, in canonical coordinates, equal volumes of the phase space correspond to equal numbers of quantum states.

Another way for choosing the initial probability distribution is to make use of Jaynes' principle [25], which is to maximize the Shannon entropy of the distribution under the known constraints over this distribution: in the present case of an isolated system which has not been previously observed, this principle also leads to the uniform law. It is not really better founded than the previous, elementary reasoning, but it may be more satisfying and it can be safely used in more complex situations. We refer to most textbooks on statistical mechanics for discussing these well-known, basic questions.

The uniform distribution in a finite space, either discrete or absolutely continuous, is clearly stationary. In addition to the previous hypotheses, we will assume that the space X is indecomposable [26]: the only subsets of X which are preserved by the evolution function φ_t are the empty set $\varnothing$ and X itself. Then, the stationary probability distribution is unique [18].

For simplicity, we will henceforth assume that the phase space X is finite.

Initial, nonstationary situation. In certain situations, the system can be prepared by submitting it to specific constraints before the initial time 0. Then it may not be distributed uniformly in X at $t = 0$. We will consider this case in the next paragraph.

2.2 Mesoscopic distributions

2.2.1 Mesoscopic states

Because of the imprecision of the physical observations, it is impossible to determine exactly the microstate of the system, but it is currently admitted that the available measure instruments allow one to define a finite partition of X into subsets $i \in \mathcal{M} \equiv (i^k)$, $k = 1, 2, \dots M$, such that it is impossible to distinguish two microstates belonging to the same subset i. So, in practice the best possible description of the system consists in specifying the subset i where its microstate x lies: i can be called the *mesostate* of the system. The probability for the system to be in the mesostate i at time t will be denoted $p(i,t)$. It is not sure, however, that two microstates belonging two different mesostates can always be distinguished: this point will be considered in Section 3.2.2.

Remark: for convenience, we use the same letter p to denote the probability in a countable state space, as well as the probability density in the continuous case. This creates no confusion when the variable type is explicitly mentioned. This is the case now since, as mentioned previously, we assume that the space X is discrete. The

transposition to the continuous case is generally obvious, although the complete derivations may be more difficult.

2.2.2 The stationary situation

If time 0 is the beginning of all observations and actions, we assume that the initial microscopic distribution μ is uniform and stationary, as discussed previously, and the probability to find system $\mathcal{S}$ in the mesostate i_0 at time 0 is $p(i_0, 0) = \mu(i_0,)$. The probability to be in i at time t is $p^0(i, t) = \mu_t(i) = \mu(\varphi_{-t}(i))$. The stationary joint probability to find $\mathcal{S}$ in i_0 at time 0 and in i at time t is

$$p^0(i_0, 0; i, t)) = \mu(\varphi_{-t} i \cap i_0) = \mu(i \cap \varphi_t i_0) \tag{5}$$

and the conditional probability of finding $\mathcal{S}$ in the i at time t, knowing that it was in i_0 at time 0 is

$$p^0(i, t | i_0, 0) = \frac{p^0(i, t; i_0, 0)}{p^0(i_0, 0)} = \frac{\mu(\varphi_{-t} i \cap i_0)}{\mu(i_0)} = \frac{\mu(i \cap \varphi_t i_0)}{\mu(i_0)} \tag{6}$$

Similarly, the stationary n-times joint probability and related conditional probabilities are readily obtained from

$$p^0(i_0, 0; i_1, t_1; \ldots i_{n-1}, t_{n-1}) = \mu\left(\varphi_{-t_{n-1}} i_{t_{n-1}} \cap \ldots \cap i_0\right). \tag{7}$$

with, for any t: $p^0(i_0, t; i_1, t_1 + t; \ldots i_{n-1}, t_{n-1} + t) = p^0(i_0, 0; i_1, t_1; \ldots i_{n-1}, t_{n-1})$.

For the sake of simplicity, we will discretize the times $0 < t_1 < t_2 \ldots$, and write $t_i = k_i \tau$, k_i being a nonnegative integer and τ a constant time step, which will be taken as time unit.

2.2.3 Non stationary situation

If $\mathcal{S}$ is a physical system, interactions may exist before or at time 0, so that the $\mathcal{S}$ can be constrained to lie in a certain subset A of X at time 0. However, since it is not possible to distinguish two microstates corresponding to the same mesostate, A should be a union of mesostates, or at least one mesostate. If it is known that at time 0 the microstate x of the system belongs to the mesostate i, we should assume that the initial microscopic distribution is uniform over i, since no available observation can give further information on x: so, in the discrete case, if $n(i)$ is the number of microstates included in i and $\chi_i(x)$ the characteristic function of i

$$p(x, 0 | x \in i) = \frac{1}{n(i)} \chi_i(x) \tag{8}$$

In the absolutely continuous case, the similar conditional density is obtained in the same way, replacing the number of microscopic states contained in the mesostate i by its volume $v(i)$. For simplicity, we follow considering the discrete case, with obvious adaptations to the continuous case.

If one only knows the mesoscopic initial distribution $p(i, 0)$ that at time 0 the system belongs to i, for each mesostate i of $\mathcal{M}$, the initial microscopic distribution becomes

$$p(x, 0) = \sum_i \frac{1}{n(i)} p(i, 0) \chi_i(x) = \sum_i \frac{p(i, 0)}{\mu(i)} \frac{\chi_i(x)}{N} \tag{9}$$

N being the total number of microstates in X.

The n-times nonstationary mesoscopic probabilities are obtained from (9)

$$p_n(i_0, 0; i_1, 1, ; i_{n-1}, n-1) = p(i_0 \cap \varphi_{-1} i_1 \cap \ldots \cap \varphi_{-n+1} i_{n-1}, 0)$$
$$= \frac{p(i_0)}{\mu(i_0)} \frac{n(i_0 \cap \varphi_{-1} i_1 \cap \ldots \cap \varphi_{-n+1} i_{n-1})}{N}. \tag{10}$$

where $n(A)$ is the number of microstates belonging to some subset A of X. So

$$p_n(i_0, 0; i_1, 1, ; i_{n-1}, n-1) = \mu(i_0 \cap \varphi_{-1} i_1 \cap \ldots \cap \varphi_{-n+1} i_{n-1}) \frac{p(i_0)}{\mu(i_0)}. \tag{11}$$

and all multiple probabilities follow, for instance

$$p_{n-1}(i_1, 1; \ldots; i_n, n) = \sum_{i_0} \mu(i_0 \cap \varphi_{-1} i_1 \cap \ldots \cap \varphi_{-n} i_n) \frac{p(i_0)}{\mu(i_0)}. \tag{12}$$

The corresponding process is generally not Markovian. For instance, if $i_0 \cap \phi_{-1} i_1 \neq \emptyset$, $i_1 \cap \phi_{-1} i_2 \neq \emptyset$ and $i_0 \cap \phi_{-2} i_2 = \emptyset$, it is easily seen that $p(.i_2, 2 | i_1, 1; i_0, 0) = 0$ but $p(i_2, 2 | i_1, 1) \neq 0$.

From the definition of the relative probabilities, one can formally write

$$p(i_2, t_2) = \sum_{i_1} p(i_2, t_2 | i_1, t_1) p(i_1, t_1). \tag{13}$$

but in general this equation is useless, since the conditional probability $p(i2, t2 | i1, t1)$ cannot be computed independently of $p(i1, t1)$.

It results from (11) that the nonstationary conditional probabilities, *conditioned by the whole past up to time* 0, are identical to the corresponding stationary probabilities: as an example

$$p(i_n, n | i_{n-1}, n-1; \ldots; i_0, 0) = p^0(i_n, n | i_{n-1}, n-1; \ldots; i_0, 0)$$
$$= \frac{\mu(i_0 \cap \varphi_{-1} i_1 \cap \ldots \cap \varphi_{-n} i_n)}{\mu(i_0 \cap \varphi_{-1} i_1 \cap \ldots \cap \varphi_{-n+1} i_{n-1})}. \tag{14}$$

We will make use of this simple but important property later.

2.3 Entropy of the mesoscopic process resulting from deterministic, microscopic system

Kolmogorov and other authors [19] studied the entropy and ergodic properties of the stationary mesoscopic process defined previously, following methods introduced by Shannon in the framework of signal theory [27–30]. These methods, and part of Kolmogorov's results, can be extended to the nonstationary process (11).

2.3.1 The n-times entropy and the instantaneous entropy of the mesoscopic system

Following Kolmogorov, we consider y the Shannon entropy [27–30] of the trajectory $(i)_n = (i_0, \ldots, i_{n-1})$ in the phase space

$$S(p_n) = -\sum_{i_0, \ldots i_{n-1}} p_n(i_0, 0; \ldots; i_{n-1}, n-1) \ln p_n(i_0, 0; \ldots; i_{n-1}, n-1). \tag{15}$$

On the other hand, the new information obtained by observing the system in the mesoscopic state i_n at time t_n, knowing that it was in the respective states $i_0, \ldots i_{n-1}$ at the prior times $0, \ldots n-1$, will be called the instantaneous entropy

$$s_n(p) = S_{n+1}(p) - S_n(p) = -\sum_{i_0,;;;i_n} p(i_0, 0; \ldots i_n, n) \ln p(i_n, n | i_{n-1}, n-1; \ldots; i_0, 0) \geq 0$$

$$= \sum_{i_0,;;;i_{n-1}} p(i_0, 0; \ldots i_{n-1}, n-1) \, S\big(p(\bullet, n | i_{n-1}, n-1; \ldots; i_0, 0)\big). \quad (16)$$

where p denotes the infinite process. The properties of $S(p_n)$ and $s_n(p)$ have been extensively studied by Kolmogorov and other authors in the case of the stationary process (6) [19]: they are summarily mentioned in 2.5. They are not necessarily valid for the nonstationary process.

2.3.2 Maximizing the n-times entropy of the mesoscopic system: The "Markov scheme"

If one knows the first two distributions p_1 and p_2, one can mimics the exact mesoscopic distributions p_n by using the Jaynes' principle, maximizing the entropy $S(q_n)$ of a ditribution q_n under the constraints $q_1 = p_1$ and $q_2 = p_2$. Then it is found that optimal distribution q_n is the Markov distribution $\bar{q}_n$ satifying these constraints [18].

It is shown in Ref. [18] that for n > 2, both the n-times entropy $S_n(\bar{q})$ and the instantaneous entropy $s_n(\bar{q})$ are larger than the correponding entropies $S_n(p)$ and $s_n(p)$ of the exact process p, except if p is Markov: $p = \bar{q}$.

The Markov process $\bar{q}_n$ is not really an approximation of the mesoscopic process p, because $\bar{q}_n$ does not tend to p_n when $n \to \infty$. Approximating the exact mesoscopic process by a Markov process will be the main purpose of the next section.

2.4 Entropy and memory in the stationary situation

2.4.1 Kolmogorov entropy of the stationary process

Here we consider the stationary process arising from the initial uniform microscopic distribution $\mu(x)$, when the n-times stationary probability is p_n^0 given by (7). For the sake of simplicity we omit the index ^{0} in the present Section, unless otherwise specified. It can be shown [19] that the entropy $S_n(p)$ is an increasing, concave function of n

$$s_n \equiv S_{n+1}(p) - S_n(p) \geq 0. \quad (17)$$

$$s_{n+1} - s_n = S_{n+1}(p) - 2S_n(p) + S_{n-1}(p) \leq 0. \quad (18)$$

It results from (17) and (18), and also from 2.5.2, that the limits

$$\lim_{n\to\infty} \frac{1}{n} S_n(p) = \lim_{n\to\infty} s_n(p) = s(p). \quad (19)$$

exist: $s(p)$ is the Kolmogorov entropy of the evolution function f with respect to the partition (i) of the mesoscopic states [19]. More simply, we can call it entropy of the mesoscopic process.

2.4.2 Memory decrease in the stationary mesoscopic process

It has been proved recently [18] that, although it is infinite, the memory of the mesoscopic process fades out with time: for n large enough, if $N > n$ the probability

of i_N at time N conditioned by the n last events, is practically equal to the probability at time N, conditioned by the *whole past* down to time 0.

$$p(i_N, N | i_{N-1}, N-1; \ldots i_n, N-n) \approx p(i_N, N | i_{N-1}, N-1; \ldots i_0, 0) \text{ when } n \to \infty.$$

(20)

More precisely, for any $\varepsilon > 0$, there exists a positive integer n such that for any $N > n$

$$0 < s_N < s_n < \varepsilon.$$

(21)

where s_n is the instantaneous entropy given by (14). In fact, let us write

$$\Pi_N(i_N) = p(i_N, N | i_{N-1}, N-1; \ldots . i_0, 0) = \mu\left(f_{-N} i_N \middle| f_{-N+1} i_{N-1} \cap \ldots \cap i_0\right); \quad (22)$$

$$\Pi_N^{(n)}(i_N) = p(i_N, N | i_{N-1}, N-1; \ldots . ; i_{N-n}, N-n) = \mu\left(f_{-N} i_N \middle| f_{-N+1} i_{N-1} \cap \ldots \cap f_{-N+n} i_{N-n}\right).$$

(23)

For a given n, formula (23) allows one to define a new process $p^{(n)}$ from the original process p, which can be called "the approximate process of order n" of p (see Section 2.6). It results from (21) and from the stationarity of p that for any $\varepsilon > 0$, there is an integer $n(\varepsilon)$ depending only on ε, such that for any integers N, $n > n(\varepsilon)$

$$0 < s_n(p) - s_N(p) = \sum_{i_0, \ldots i_{N-1}} \mu\left(i_0 \cap f_{-1}(i_1) \cap \ldots \cap f_{-N+1}(i_{N-1}) \, S_{0, \ldots N-1}\left(\Pi_N \middle| \Pi_N^{(n)}\right) < \varepsilon.$$

(24)

where $S_{0, \ldots N-1}\left(\Pi_N | \Pi_N^{(n)}\right)$ is the relative entropy of Π_N with respect to $\Pi_N^{(n)}$: the last right hand member of Eq. (22) is the average of this relative entropy on the past of N. Because $s_N(p)$ decreases to a limit $\bar{s}(p)$ when $N \to \infty$, it results that

$$0 < \delta s_n(p) \equiv s_n(p) - \bar{s}(p) \leq \varepsilon \text{ if } n > n(\varepsilon).$$

(25)

The total variation distance $d(P, Q)$ between two distributions P_j and Q_j over the states j of a finite set (j) is

$$d(P, Q) = \frac{1}{2} \sum_j \left| P_j - Q_j \right|.$$

(26)

Then, the total variation distance $d_{0, \ldots N-1}\left(\Pi_N, \Pi_N^{(n)}\right)$ between Π_N and $\Pi_N^{(n)}$ (for a given past trajectory between times 0 and N-1) is related to the relative entropy [18, 31] and it can be concluded that

$$\left\langle d_{0, \ldots N-1}\left(\Pi_N, \Pi_N^{(n)}\right)\right\rangle^2 \leq \left\langle \left[d_{0, \ldots N-1}\left(\Pi_N, \Pi_N^{(n)}\right)\right]^2\right\rangle < \varepsilon/2 \text{ if } n(\varepsilon) < n < N. \quad (27)$$

2.4.3 Convergence properties of the approximate process

Let us write $m = N-n > 0$. It follows [18] from (25) that for any fixed m, the total variation distance between the exact and the approximate probabilities $d_{0, \ldots m+n-1}\left(\Pi_{m+n}, \Pi_{m+n}^{(n)}\right)$ *tends to 0 in probability when $n \to \infty$*

$$d_{0,\ldots m+n-1}\left(\Pi_{m+n}, \Pi_{m+n}^{(n)}\right) \xrightarrow{p} 0 \text{ if } n \to \infty. \tag{28}$$

So, the probability that this distance exceeds a given accuracy $a > 0$ can be made as small as desired by choosing n large enough.

Further results can be obtained by newly using the sationnarity of process p. In fact, it can be shown [18] that p is a martingale [20–22]. Then, general results from martingales theory (see below) show that when $n \to \propto$ the distance between the stationary conditional probability Π_{m+n} and its approximation $\Pi_{m+n}^{(n)}$ tends to 0 *almost surely* [18], as well as and in probability

$$d_{0,\ldots m+n-1}\left(\Pi_{m+n}, \Pi_{m+n}^{(n)}\right) \xrightarrow{a.s.} 0 \text{ if } n \to \infty. \tag{29}$$

So, the approximation $\Pi_{m+n}^{(n)}$ converges to Π_{m+n} for almost all trajectories [18]. We now sketch the derivation of this conclusion from martingale theory.

2.5 Martingale theory and almost sure convergence

For convenience, we first summarize some definitions and results of martingale theory [20–22], before applying them to the mesoscopic laws of deterministic systems. We refer to [20] for adressing more general cases.

2.5.1 Definitions

  i. *simplified definition*: a (discrete time) sequence of stochastic variables X_n is a martingale if for all n:

$$\langle |X_n| \rangle < \infty \text{ and } \langle X_{n+1}|X_n, \ldots X_1 \rangle = X_n. \tag{30}$$

where $\langle X \rangle$ denotes the average (mathematical expectation) of the stochastic variable X.

  ii. *more generally* (see the general definition, for instance, in [20])

If • $(\Omega, \mathcal{F}, P)$ is a probability space (where Ω is the state space, P is the probability law, and $\mathcal{F}$ is the set of all subspaces (σ-algebra) for which P is defined),

 • $\mathcal{F}_n$ is an increasing sequence of σ-algebras extracted from $\mathcal{F}$ ($\mathcal{F}_n \subset \mathcal{F}_{n+1} \subset \ldots \subset \mathcal{F}$), and.

 • for all $n \geq 0$, X_n is a stochastic variable defined on $(\Omega, \mathcal{F}_n, P)$,

the sequence X_n is a martingale if $\langle |X_n| \rangle < \infty$ and $\langle X_{n+1}|\mathcal{F}_n \rangle = X_n$.

2.5.2 Convergence theorem for martingales

Among the remarkable properties of martingales, the following convergence theorem holds [20, 21]:

If (X_n) is a positive martingale, the sequence X_n converges almost surely to a stochastic variable X.

So, for almost all trajectories ω, $X_n(\omega) \to X(\omega)$ with probability 1 when $n \to \infty$. Stronger and more general results can be found in the references.

2.5.3 Application to the nth approximation of the stationary mesoscopic process

The stochastic variable $Y_N = p^0(i_N, N | i_{N-1}, N-1; \ldots ; i_0, 0)$ is a martingale. In fact, because of the stationarity of p^0 we have, renumbering the states

$$p^0(i, N | i_{-1}, N-1; \ldots ; i_{-N}, 0) = p^0(i, 0 | i_{-1}, -1; \ldots ; i_{-N}, -N) \equiv p^0(i, 0 | \mathcal{F}_N). \quad (31)$$

where $\mathcal{F}_N$ is the σ-algebra generated by $i_{-1}, \ldots i_{-N}$. Let us write

$$\pi_N = p^0(i, 0 | i_{-1}, -1; \ldots ; i_{-N}, -N) = p^0(i, 0 | \mathcal{F}_N). \quad (32)$$

We have, because $\mathcal{F}_{N-1} \subset \mathcal{F}_N$

$$\langle \pi_N | \mathcal{F}_{N-1} \rangle = \langle p^0(i, 0 | \mathcal{F}_N) | \mathcal{F}_{N-1} \rangle = p^0(i, 0 | \mathcal{F}_{N-1}) = \pi_{N-1}. \quad (33)$$

So, π_N is a martingale on the σ-algebra $\mathcal{F}_N$, and by the convergence theorem, it converges almost surely to a.
limit π when $N \to \infty$.
Now if $N > n$, let us write $m = N-n > 0$. Because of the stationarity of p^0

$$p^0(i, n+m | i_{-1}, n+m-1; \ldots ; i_{-m}, m) = p^0(i, 0 | i_{-1}, -1; \ldots ; i_{-n}, -n) = \pi_n(i). \quad (34)$$

Thus, for any fixed, positive m

$$\pi_{n+m} - \pi_n \overset{a.s.}{\to} 0. \quad (35)$$

The absolute value distance between π_{n+m} and π_n is obtained by summing $|\pi_{n+m}(i) - \pi_n(i)|$ over the M possible states i, So

$$d_{0, \ldots n+m-1}\left(q_{n+m}, q_{n+m}^{(n)}\right) = d(\pi_{m+n}, \pi_n) \overset{a.s.}{\to} 0 \text{ if } n \to \infty. \quad (36)$$

which is (29), one of our main, formal results.

2.6 *n*-times Markov approximation of the mesoscopic stationary process

Returning to inequalities (19), when the value ε is fixed for obtaining a required precision, the value $n = n(\varepsilon)$ is determined and a satisfying approximation of the exact mesoscopic process is obtained by neglecting the memory effects at time differences larger than n [18] Thus, one replaces $p(i_N, N | i_{N-1}, N-1; \ldots \ldots i_0, 0)$ by

$$p^{(n)}(i_N, N | i_{N-1}, N-1; \ldots \ldots i_0, 0) = p(i_N, N | i_{N-1}, N-1; \ldots \ldots i_{N-n}, N-n) \quad \text{if } N > n \quad (37)$$

With the convention

$$p^{(n)}(i_0, 0; \ldots i_N, N) = p(i_0, 0; \ldots i_N, N) \quad \text{if } N \leq n. \quad (38)$$

all the probabilities related to the approximate process $p^{(n)}$ are defined from the probabilities of p: this defines $p^{(n)}$, the approximate process of order n of p. So, $p^{(n)}$ has a finite memory of size n, whereas p has in general an infinite memory.
The process $p^{(n)}$ is a Markov process on the partial trajectories I_K consisting of groups of n successive mesoscopic states

$$I_K = (i_{Kn}, i_{Kn+1}, \ldots i_{(K+1)n-1}) \in \mathcal{M}^n \tag{39}$$

Its probability distributions can be written in abbreviated notations

$$P_K^{(n)}(I_0, T_0; I_1, T_1; \ldots I_{K-1}, T_{K-1}) \equiv p^{(n)}(I_0, 0, 1, \ldots n-1; I_1, n, n+1, \ldots 2n-1; \ldots; I_{K-1},$$
$$(K-1)n, \ldots, Kn-1). \tag{40}$$

T_K being the group of n successive times: $T_k = kn, kn+1, \ldots, (k+1)n-1$. From the approximation (18) it follows (see Appendix A) that

$$P^0(I_K, T_K \mid I_{K-1}, T_{K-1}; \ldots I_0, T_0) \approx P^{0(n)}(I_K, T_K \mid I_{K-1}, T_{K-1}) \ . \tag{41}$$

where we now use the upper index^0 in P^0 and $P^{0(n)}$ to recall that, in the present section, p is the stationary distribution. Note that, because of this stationarity

$$P^{0(n)}(I_K, T_K \mid I_{K-1}, T_{K-1}) = P^{0(n)}(I_K, T_1 \mid I_{K-1}, T_0) = P^0(I_K, T_1 \mid I_{K-1}, T_0)$$
$$\equiv W(I_K \mid I_{K-1}). \tag{42}$$

So, the transition matrix W is well defined from the known stationary distribution p^0.

From the approximate relation (41) if follows that the exact stationnary process P^0 on the partial history I_K during the time interval T_K approximately obeys the n-times Markov Equation (see Section 2.7)

$$P^0(I_K, T_K) \approx \sum_{I_{K-1}} W(I_K \mid I_{K-1}) \ P^0(I_{K-1}, T_{K-1}). \tag{43}$$

while the nth approximation $P^{0(n)}$ satisfies (33) exactly.

2.7 Markov approximations of the nonstationary mesoscopic process

We return to the nonstationary process p generated by the deterministic microscopic process from an arbitrary initial distribution of the mesoscopic states, given by (11). As in paragraph 2.6, it is now necessary to distinguish the stationary process p° by the upper index^0.

One can write the trivial equality

$$p(i_N, N; \ldots i_{N+n-1}, N+n-1) = \sum_{i_{N-1}, \ldots, i_0} p(i_{N+n-1}, N+n-1; \ldots; i_N, N$$
$$\mid i_{N-1}, N-1; \ldots; i_0, 0) p(i_0, 0; \ldots; i_{N-1}, N-1). \tag{44}$$

We now use remark (14): the conditional probabilities, conditioned by the whole past up to time 0, are identical in the stationary and nonstationary situations. The stationary distributions p^0 can be approximated by its nth approximation $p^{0(n)}$ introduced in Section 2.6. Thus we can write

$$p^0(i_{N+n-1}, N+n-1; \ldots; i_N, N \mid i_{N-1}, n-1; \ldots; i_0, 0) =$$

$$p^0(i_{N+n-1}, N+n-1 \mid i_{N+n-2}, N+n-2; \ldots; i_0, 0) p^0(i_{N+n-2}, N+n-2 \mid i_{N+n-3}, N+n-3; \ldots; i_0, 0) \ldots$$

$$\ldots p^0(i_N, N \mid i_{N-1}, N-1; \ldots; i_0, 0) \approx p^{0(n)}(i_{N+n-1}, N+n-1; \ldots; i_N, N \mid i_{N-1}, n-1; \ldots; i_{N-n}, N-n). \tag{45}$$

With (35), Eq. (34) yields the approximate n-times Markov Equation

$$p(i_N, N; \ldots; i_{N+n-1}, N+n-1) \approx \sum_{i_{N-1}, \ldots, i_0} p^0(i_{N+n-1}, N+n-1; \ldots; i_N, N$$

$$|i_{N-1}, n-1; \ldots; i_{N-n}, N-n) p(i_{N-n}, N-n; \ldots; i_{N-1}, N-1). \tag{46}$$

Taking $N = Kn$ for an integrer $K \geq 0$, using the condensed notations of § 2.6 and definition (42), Eq. (46) yields an approximate Master Equation for the probability $P(I_K, T)$ of the partial history I_K during the time interval T_K

$$P(I_K, T_K) \approx \sum_{I_K} W(I_K | I_{K-1}) \ P(I_{K-1}, T_{K-1}) \tag{47}$$

which is the Eq. (43) obtained for the stationary probability $P^0(I_K, T)$. Let $P^{(n)}(I_K, T)$ be the exact solution of Eq. (47) that coincides with the exact P at the n first elementary times 0, 1, ... n-1 of the system history: $P^{(n)}(I_0, T) = P(I_0, T)$. Then, $P^{(n)}(I_K, T)$ defines the nth approximation of $P(I_K, T)$: in principle, it can be computed from Eq. (47) since the probability transitions W are known by (41).

The stationary probabilities approximation $P^{0(n)}$ deduced from p^0 provide the stationary solution of (47)

$$P^{0(n)}(I_K, T_K) = p^0(i_{Kn}, Kn; \ldots; i_{K(n+1)-1}, K(n+1)-1)$$
$$= p^0(i_{Kn}, 0; \ldots; i_{K(n+1)-1}, n-1). \tag{48}$$

So, when $K \to \infty$,

$$P^{(n)}(I_K, T_K) \to P^{0(n)}(I_K, T_K). \tag{49}$$

and consequently, for any integer $k \in [0, n\text{-}1]$, the nth approximation of the mesoscopic distribution p satisfies

$$p^{(n)}(i, Kn+k) \to \mu(i, k) = \mu(i) \ \text{ if } K \to \infty. \tag{50}$$

for any initial mesocopic distribution, which is the basic assumption of statistical thermodynamics. Supplementary assumptions allow one to conclude that, in realistic situations, the mesoscopic distribution p itself satisfies this property (see Appendix B).

2.8 Time averages and simple Markov approximation

Up to now, we took as time unit some time step τ which gives the time scale of microscopic phenomena. By considering some finite partition (i) of the phase space X and replacing the microscopic states $x \in X$ by the mesocopic states $i \in (i^k)$, we have performed a space coarse graining, as necessary for taking practical observations into account. For the same purpose, one should also introduce [18] a space coarse graining, since the time scale $\theta = n \tau$ of current observations is much larger than τ: $n >> 1$.

All mesoscopic functions remaining practically constant on the time scale θ, their averages can be computed from the time averages $\overline{P}_K$ of the probabilities p_k over θ

$$\overline{p}_K = \frac{1}{n} \sum_{k \in T_K} p_k. \tag{51}$$

where K is an integer ≥ 1 and T_K is the time interval (τ being the time unit) $T_K = (K-1)n, K+1, \dots Kn\text{-}1$.

Suppose (a) that the mesoscopic probabilities p are slowly variating functions of the mesoscopic states, (*i.e.* for any positive α, $|p(i) - p(j)| < \alpha$ if the distance between the mesostates i and j is small enough, with an appropriate metric in the spase of mesostates), and (b) that discontinuous trajectories have low probabilities and can be neglected. Of course, these assumptions are not verified for some important, well known processes such as Brownian processes, but they seem to be reasonable for modeling physical processes where the inertial effects are strong enough. Then, a simple approximation is to consider that

$$p^0\left(i_{Kn-1}, Kn-1; \dots; i_{(K-1)n}, (K-1)n \,\middle|\, i_{(K-1)n-1}, (K-1)n-1; \dots; i_{(K-2)n}, (K-2)n\right)$$

$$\approx p^0\left(i_{\overline{K}}, Kn-1; \dots; i_{\overline{K}}, (K-1)n \,\middle|\, i_{\overline{K}-1}, (K-1)n-1; \dots; i_{\overline{K}-1}, (K-2)n\right)$$

$$\equiv \overline{W}\left(i_{\overline{K}} \,\middle|\, i_{\overline{K}-1}\right). \tag{52}$$

where

$$\overline{K} = \frac{1}{n}\sum_{k \in T_K} k \;=\; \frac{1}{n}\sum_{k=(K-1)n}^{Kn-1}. \tag{53}$$

Consider the time-averaged probability

$$\overline{P}\left(i_{\overline{K}}, K\right) \equiv \frac{1}{n}\sum_{k \in T_K} p\left(i_k, k\right) \approx \frac{1}{n}\sum_{k \in T_K} p\left(i_{\overline{K}}, k\right). \tag{54}$$

Using the Markov Eq. (47) and the complementary approximations (42), we obtain the new Master Equation

$$\overline{P}(i, K) \approx \sum_j \overline{W}(i \,|\, j)\; \overline{P}(j, K-1). \tag{55}$$

This equation is much simpler than Eq. (47), since it applies in the space $\mathcal{M}$ of the M mesostates (i), whereas (47) is valid in the space $\mathcal{M}^n$ of n successive mesostates. However, Eq. (45) relies on several approximations that are difficult to control. In spite of these difficulties, which can only be precisely discussed for specific examples, Master Equations like (55), resulting from deterministic micro-scopic systems by coarse-graining both their states and time, are a practical way to study their evolution of a mesoscopic scale, used in innumerable works.

3. Discussion of the Markov representation derived from Hamiltonian dynamics, and estimation of the uniformization time

The previous results show that the coarse grained mesoscopic dynamics can eventually be represented by a Master Equation, because the memory of this dynamics is gradually lost over time. However, they do not provide the time scale of this fading. In order to estimate its order of magnitude simply, we make an intuitive remark: the conditional probability to jump from some mesostate i to another one can be evaluated without knowing the past history of the system if one knows the initial microscopic distribution over i. The only unbiaised initial distribution is the uniform one. Thus, one can consider that the system has a memory limited to one time step if uniformity is approximately realized in each mesoscopic cell: this is the basis of the elementary Markov models of mesoscopic evolution. Let T be the average time

needed to reach uniformity in a mesoscopic scale, starting from strong inhomogeneity. In a first approximation it is reasonable to use this uniformization time T to characterize the time scale over which a Markov evolution can describe the system.

3.1 Uniformization time in a mesoscopic cell: An elementary estimation for Hamiltonian systems

Using oversimplified, but reasonable arguments [17], we now coarsely estimate the uniformization time T in a mesoscopic cell. As an example, we consider n identical particles initially located in this cell, among N identical particles in an isolated vessel. The complete system obeys Hamilton mechanics.

Assume that the particles constitute a gas under normal conditions, with density $\rho \approx 3. \, 10^{25}$ molecules.m^{-3}. A mesocopic state can be reasonably represented by a cube of size $l \approx 10^{-6}$ m (as an order of magnitude), which contains $n \approx 3. \, 10^{7}$ molecules. We now divide the mesoscopic cell into m "microscopic" cells whose size λ is comparable to the size of a molecule: each of these microscopic cells, however, should contain a sufficient number particles for allowing them to interact from time to time. We can take $\lambda \approx 10^{-8}$ m, so each microscopic cell approximately contains 30 molecules, and there are $m \approx 10^{6}$ microscopic cells in a mesoscopic cell. The particles have an average absolute value $v \approx 500$ m.s^{-1} in typical conditions. They can jump between the various microcells of the same mesoscopic cell. They can also jump out of their initial mesoscopic cell, but they are replaced by molecules proceeding from other cells, and we assume that these contrary effects coarsely compensate themselves, except in the first stage of the evolution if the initial mesocopic distribution is strongly inhomogeneous.

Because all particles are identical, an almost microscopic configuration of a mesoscopic cell can be defined by specifying the number of particles in each of its microscopic cells. Focusing on a given mesoscopic cell, we compute the number of its possible configurations, and we estimate the average time θ necessary for the system to visit all these configurations. Note that the uniformization time T is obviously much larger than θ: $T >> \theta$. So, θ is a lower bound of T.

The number of ways of partitioning the n identical particles into the m microscopic cells is

$$C = \frac{(m+n-1)!}{n!(m-1)!} \approx \exp\left[(m+n)\,\varphi(x)\right] \text{ with } \varphi(x)$$
$$= -x \ln x - (1-x) \ln(1-x) \text{ and } x = n/(m+n). \tag{56}$$

The system jumps from one of these configurations to another one each time one the present particles jumps to another microscopic cell. The order of magnitude of the time needed for a particle to cross a micro-cell is λ/v, and the time between two configurations changes is $\tau \approx (1/n)\,\lambda/v.$. In order that all configurations are visited during time θ we should have at least $\theta \approx C\,\tau$ (in fact, θ should be much larger than $C\tau$ because of the multiple visits during θ). So we conclude from (46) and relevant approximations that a lower bound of Θ satisfies

$$n\frac{\varphi(x)}{x} \approx \ln\frac{v\,\Theta}{\lambda} \text{ with } x = n/(m+n) \approx 1. \tag{57}$$

With the previous numerical values

$$\theta \approx \frac{\lambda}{v}\left(\frac{n}{m}\right)^{m} \approx 2.10^{-11}.(30)^{10^{6}} \text{ s.} \tag{58}$$

which is far larger than the age of universe (now estimated to be about 14×10^9 years, or 4.4×10^{17} s)!

Although these calculations are very rudimentary, it is clear that, in the framework of purely Hamiltonian systems, the microscopic distribution within a mesocopic cell remains far from uniformity during any realistic time if it is initially fairly inhomogeneous.

More generally, it is clear that the uniformization time T should be of the order of Poincaré time [32–36] in a mesoscopic cell, which is known to be extraordinarily long [9, 37].

3.2 An elementary, empirical approach of mesoscopic systems

The practical relevance of Markov processes to model a large class of physical systems is supported by a vast literature. We have seen that the progressive erasure of its memory over time allows one to justify the use of a Markov process to represents the evolution of the coarse-grained system. However, such representation can also stem from random disturbances due to the measurements or other sources of stochasticity: then, one has to renounce to a purely deterministic microscopic dynamics, as formerly proposed by many authors, even without adopting the formalism of Quantum Mechanics. It is interesting to compare the time scales of the relaxation to equilibrium in both approaches with an elementary example.

3.2.1 Uniformization induced by randomization

Suppose now that the measure process does not induce any significant change in the average molecules energy - so, their average velocity remains unchanged – but that it causes a random reorientation of their velocity. A rudimentary, one dimensional model of such a randomization could be to assume that each time a molecule is about to pass to a neighboring cell, it will go indifferently to one of the neighboring microscopic cell. In a one dimensional version of the model, a molecule perform a random walk on the $\eta = l/\lambda = 10^2$ points representing the microscopic cells contained in the mesoscopic cell, and we adopt periodic conditions at the boundaries of the mesocopic cell. The $\eta \times \eta$ transition matrix of the process is a circulant matrix which, in its simplest version, has transition probabilities ½ to jump from any state to one of its neighbors, and it is known that its eigenvalues λ_k are $\lambda_k = \cos(2\pi k/\eta)$, $k = 0, 1, \dots [\eta/2]$. The number of jumps necessary for relaxing to the uniform, asymptotic distribution is of the order

$$1/(-\ln \lambda_1) \propto 2(2\pi/\eta)^{-2} \approx 500 \text{ s}.$$

which correspond to a relaxation time of 500. $\lambda/v \approx 10^{-8}$ s, which is very short for current measurements, but comparable with (or even larger than) the time scale of fast modern experiments. Considering a 3-dim model would not change this time scale significantly. It is conceivable that he molecules are not necessarily reoriented each time they leave a microscopic cell. Even if the proportion of reoriented molecules is as low as 10^{-6}, the relaxation time is of order 10^{-2} s, which is insignificant in many simple measures. In this case the Markov representation can be justified.

3.2.2 Semi-classical Hamiltonian systems

In analogy with the previous randomized system, we can introduce a new source of stochasticity in the coarse-grained deterministc systems considered in Sections 2 and 3. This could be done by assuming that a particle cannot be described by a

point, but by a probability density centered on the point that would represent it classically: such a description borrows one, but not all, of the axioms of wave mechanics, and it can be qualified as a "semi-quantical" description. A similar assumption can be introduced without referring to quantum mechanics, by noticing that a particle cannot be localized in a given mesoscopic cell with complete certainty, because of its finite size: if it is mainly attributed to a given cell, there exists a small probability that it also belongs to a neighboring cell. Even without formalizing these possibilities, one can presume that such random effects shorten drastically the memory of the mesoscopic process, and make it short with respect to ordinary measure times: then the Markov approximation described in Section 2 can correctly represent the evolution of the observed coarse-grained process.

4. Conclusion

We have studied the mesoscopic, stochastic process derived from a deterministic dynamics applied to the cells determined by measure inaccuracies. The stationary process, which arises when the microscopic initial state is distributed according to a time invariant measure, was studied by Kolmogorov and further authors: we extended their methods and some of their results, and considered the nonstationary process which stems from a noninvariant initial measure. We have shown that, according to Jaynes' principle, the "exact" mesocopic process can be approximately replaced by the Markov process which, at any time n, reproduces the one-time probability of each mesostate and the transition probabilities from it. This Markov process maximizes the trajectory entropy up to time n, as well as the entropy at time n, conditioned by prior events. The Jaynes' principle, however, does not control the accuracy of this estimate: this was our next concern.

So, a sequence of successive approximations has been defined for the stationary mesoscopic process, based on one of our main results: the probability of any mesostate state conditioned by all past events, can be approximated by its probability conditioned by the n last past events only, the integer n being determined by the maximum distance allowed between these probabilities, as small as it may be. This property entails that the nonstationary mesoscopic process can be approximated by a n-times Markov process or even, after a time coarse-graining, by an ordinary one-time Markov process. These approximations require certain conditions which should be fulfilled by "normal" physical systems, with possible exceptions for slowly relaxing systems. If they are satisfied, the existence of a thermodynamic equilibrium is derived for a coarse-grained system obeying a measure-preserving deterministic dynamics, in particular an Hamiltonian dynamics, without introducing *ad-hoc* external noises. However, very rough estimations of the relaxation time show that for reasonable values of the parameters this time is extraordinarily long and completely unrealistic.

We conclude that, although the basic hypotheses of thermodynamics can be justified from a Hamiltonian or deterministic microscopic dynamics applied to the mesoscopic cells, the observed time scales of the relaxation to equilibrium cannot be explained without going beyond pure Hamilton mechanics, by introducing additional random effects, in particular due to the intrinsic imprecision of the particles localization.

Appendix A: Approximating the n-times conditional probability

With the notations of Section 4.2, we consider approximation (55), which is the basis of the n-times Markov approximation both in the stationary and nonstationary situations. Repeating approximation (51) we can write

$$p\left(i_{N+n-1}, N+n-1; \ldots; i_N, N \mid i_{N-1}, N-1; \ldots; i_{N-n}, N-n; \ldots; i_0, 0\right) =$$

$$= p\left(i_{N+n-1}, N+n-1 \mid i_{N+n-2}, N+n-2; \ldots; i_0, 0\right) \ldots \ldots p\left(i_N, N \mid i_{N-1}, N-1; \ldots; i_0, 0\right)$$

$$\approx p^{(n)}\left(i_{N+n-1}, N+n-1 \mid i_{N+n-2}; \ldots i_{N-1}, N-1\right) \ldots p^{(n)}\left(i_N, N \mid i_{N-1}, N-1; \ldots; i_{N-n}, N-n\right).$$

$$(59)$$

The last line of (49) is $p^{(n)}\left(i_{N+n-1}, N+n-1; \ldots; i_N, N \mid i_{N-1}, N-1; \ldots; i_0, 0\right)$.
We write

$$p\left(i_{N+n-1}, N+n-1; \ldots; i_N, N \mid i_{N-1}, N-1; \ldots; i_{N-n}, N-n; \ldots; i_0, 0\right) \equiv$$

$$p^{(n)}\left(i_{N+n-1}, N+n-1; \ldots; i_N, N \mid i_{N-1}, N-1; \ldots; i_{N-n}, N-n; \ldots; i_0, 0\right) \ Q_N^{(n)}.$$

$$(60)$$

and for $k > n$ we define l_k, using the abbreviations (32)

$$l_k(i_0, \ldots i_{k-1}) \equiv \ln \frac{\Pi_k(i_k)}{\Pi_k^{(n)}(i_k)}.$$

$$(61)$$

We have by (24)

$$s_n(p) - s_k(p) = \sum_{i_0, \ldots i_k} p_N(i_0, 0; \ldots; i_k, k) \ \ln \frac{\Pi_k(i_k)}{\Pi_k^{(n)}(i_k)} = \ \langle l_k(i_0, \ldots i_{k-1}) \rangle \equiv \sigma_k(n).$$

$$(62)$$

(Note that σ_k is positive, although this not necessarily true for l_k). By (24) for any positive ε

$$2 \left\langle d^2\left(p_k, p_k^{(n)}\right) \right\rangle \ \leq \ \langle \sigma_k(i_0, \ldots i_{k-1}) \rangle \ < \ \varepsilon \text{ if } n \text{ is large enough.} \tag{63}$$

Averaging the logarithm of Eq. (60) we have

$$\left\langle L_N^{(n)} \right\rangle \equiv \left\langle \ln Q_N^{(n)} \right\rangle = \sum_{k=0}^{n-1} [s_n(p) - s_{N+k}(p)] \propto n \ [s_n(p) - s_\infty(p)] \ = \ n \, \delta s(n).$$

$$(64)$$

$\delta s(n) \equiv s_n(p) - s_\infty(p)$ can be interpreted as an entropy fluctuation with respect to its equilibrium thermodynamic value. If such a fluctuation relaxes exponentially to 0 with time, as usual, the last term of (54) tends to 0 when $n \rightarrow \infty$. Then, the n-times Markov approximations 4.2 and 5.1 are justified. Although exponential relaxation can be considered as a characteristic of "normal" physical systems, slower relaxations can occur: in this case the Markov approximation may be invalid.

Appendix B: Tendency to the stationary mesoscopic distribution

This tendency can be reasonably expected from the approximation of the exact mesocopic process by Markov processes, but it can only be affirmed by adding additional assumptions to the basic assumptions. We first prove a simple, useful lemma.

B.1. Lemma. Consider a d-dim sequence $u_{n,k}$ with 2 positive, integer indices n, k, satisfying the following properties:

i. it is absolutely bounded: there is a positive real number M such that $|u_{n,k}| < M$ for all integers n, k.

ii. for all n, k, there are positive numbers ε_n (independent of n) and ν (independent of n and k) such that

$$|u_{n,k}| < \varepsilon_n + \nu |u_{n,k-1}| \text{ and } \varepsilon_n \to 0 \text{ if } n \to \infty. \tag{65}$$

Then $u_{n,k} \to 0$ if $n \to \infty$ and $k \to \infty$.

In fact, for any positive ε, there is an integer n_0 such that $\varepsilon_n < \varepsilon$ if $n > n_0$, and

$$|u_{n,k}| < \varepsilon \left(1 + \nu + \dots + \nu^{k-1}\right) + \nu^k u_{n,0} < \frac{\varepsilon}{1-\nu} + \nu^k |u_{n,0}| \text{ if } n > n_0. \tag{66}$$

So, $|u_{n,k}|$ can be made as small as desired by chosing n and k large enough.

B.2. For given integers n and K larger than 1, and states $i_k \in M$, $k = 0, 1, \dots (K + 1)n-1$, we will write.

$p\left(i_0, 0; i_1, 1; \dots; i_{(K+1)n-1}, (K+1)n - 1\right) \equiv p(0, 1, \dots, (K+1)n - 1)$ for the sake of simplicity. In these abbreviated notations, we have

$$p(Kn, Kn + 1; \dots, (K+1)n - 1) = \sum_{i_0, i_1, \dots i_{Kn-1}} p((K+1)n, -1 \dots Kn | Kn - 1, \dots 0)$$

$$p(0; 1; \dots, Kn - 1). \tag{67}$$

We know that

$$p\left((K+1)n, -1 \dots Kn | Kn - 1, \dots 0\right) = p^0((K+1)n, -1 \dots Kn | Kn - 1, \dots 0). \tag{68}$$

p^0 being the stationary probabilities, and that, for large n

$$p^0((K+1)n, -1 \dots Kn | Kn - 1, \dots 0) \approx p^0((K+1)n, -1 \dots Kn | Kn - 1, \dots K(n-1)). \tag{69}$$

More precisely, in the conditions discussed previously, for any given positive ε, there is an integer $n(\varepsilon)$ such that

$$|p^0((K+1)n, -1 \dots Kn | Kn - 1, \dots 0) - p^0((K+1)n, -1 \dots Kn | Kn - 1, \dots K(n-1))| < \varepsilon \text{ if } n \geq n(\varepsilon) \tag{70}$$

So, Eq. (67) becomes

$$p(Kn, Kn + 1; \dots, (K+1)n - 1) =$$

$$\sum_{i_0, \dots i_{Kn-1}} [p((K+1)n, -1 \dots Kn | Kn - 1, \dots 0) - p^0((K+1)n, -1 \dots Kn | Kn - 1, \dots 0)] p(0; 1; \dots, Kn - 1)$$

$$+ \sum_{i_0, \dots i_{Kn-1}} p^0((K+1)n, -1 \dots Kn | Kn - 1, \dots (K-1)n) p(K-1)n; \dots, Kn - 1)$$

$$\equiv a_{n,K} + \sum_{i_0, \dots i_{Kn-1}} p^0((K+1)n, -1 \dots Kn | Kn - 1, \dots (K-1)n) p(K-1)n; \dots, Kn - 1). \tag{71}$$

where the 1st term of the last line satisfies $|a (n, K)| < \varepsilon$ if $n \geq n(e)$.

The second term in the last line of (71) is, in other notations, the righth hand side term of the approximate Master Eq. (47) of Section 2.8

$$P(I_K, T_K) \approx \sum_{I_K} W(I_K \mid I_{K-1}) \ P(I_{K-1}, T_{K-1}). \tag{72}$$

This approximate Master Equation can now be written more precisely, in the notations of 2.8

$$P(I_K, T_K) = A_{n,K}(I_K, T_k) + \sum_{I_K} W(I_K, T_K \mid I_{K-1}, T_{K-1}) \ P(I_{K-1}, T_{K-1}) \ . \tag{73}$$

where $A_{n,K}(I_K, T_k)$ is just the $a_{n,K}$ term of (71) expressed in the notations of 2.8 where T_K is the group of n successive times: $kn, kn + 1, .., (k + 1)n - 1$, and $I_K = (i_{Kn}, i_{Kn+1}, ... i_{(K+1)n-1}) \in \mathcal{M}^n$ describes the corresponding partial history of the mesoscopic system.

On the other hand, we know that the stationary distribution P^0 satisfies the Master Eq. (72) exactly. So, writing

$$U_{n,K}(I_K, T_K) \equiv P(I_K, T_K) - P^0(I_K, T_K). \tag{74}$$

we have

$$U_{n,K}(I_K, T_K) = A_{n,K}(I_K, T_k) + \sum_{I_K} W(I_K \mid I_{K-1}) \ U_{n,K-1}(I_{K-1}, T_{K-1}). \tag{75}$$

Note that W depends of n, but is independent of K.

From (72)–(75) it results that $U_{n,K}(I_K, T_K)$ tend to 0 when n and K tend to infinite if, furthermore, the following.

condition (c) holds. In fact, the (M^n- dim) vector $U_{n,K}$ is orthogonal to the left-eigenstate with eigenvalue 1.

of matrix W. All the eigenvalues of the projection of W in the corresponding subspace have an absolute value smaller than 1. Thus the lemma 1 applies if condition (c) is satisfied:

(c) When n increases, the absolute values of the nonstationary eigenvalues of W have an upper bound <1.

This property is likely to hold if the actual stationary mesoscopic process is not too different from an exact Markov process. So, it is reasonable to conjecture that property (c) holds for typical actual systems.

Author details

Michel Moreau* and Bernard Gaveau
Sorbonne Université, Paris, France

*Address all correspondence to: mlmoreau@free.fr

IntechOpen

References

[1] Boltzmann L. Vorlesungen über Gastheorie. In: *part*. Vol. *2*. Leipzig: J. A. Barth; 1898

[2] L. Boltzmann. Reply to Zermelo's Remarks on the theory of heat.In: History ofmodern physical sciences: The kinetic theory of gases, ed. S. Brush, ImperialCollege Press, **57**, 567 (1896).

[3] Ehrenfest P, Ehrenfest T. *The conceptual foundations of the statisticalapproach in Mechanics*. New York: Dover; 1990

[4] Uhlenbeck GE. *Anoutline of Statistical Mechanics, in Fundamental problemsin Statistical Mechanics*. II. ed. Cohen, North Holland, Amsterdam: E. G. D; 1968

[5] Landau LD, Lifshitz EM. *Statistical Physics*. 3rd ed. Oxford: Pergamon Press; 1980

[6] Landau LD, Pitaevskii LP. *Physical Kinetics*. Oxford: Pergamon Press; 1981

[7] H.B. Callen,*Thermodynamics and an Introduction to Thermostatistics*(John Wiley and Sons, New York, (1985).

[8] Reif F. *Fundamentals of Statistical and Thermal Physics*. NY: Mc Graw Hill; 1965

[9] J.P. Eckmann and D. Ruelle, *Ergodic theory of chaos and strange attractors*, Rev Mod Phys**57**, 617 (1985).

[10] J.R. Dorfman, A, Introduction to Chaos I, Nonequilibrium Statistical Mechanics (Cambridge, New York, (1999).

[11] Gallavotti G. Statistical Mechanics: a Short Treatise. Berlin: Springer; 1999

[12] Gallavotti G. Journal of Statistical Physics. 1995;**78**:1571

[13] Evans DJ, Cohen EGD, Morries GP. Phs. Rev. Lett**.** 1993;**71**:15

[14] Gallavotti G, Cohen EGD. Phys. Rev. Lett. 1995;**74**:2694

[15] Yvon J. *La Théorie Statistique des Liquides*. Paris: Herman; 1935

[16] Forster D. *Hydrodynamic Fluctuations*. Benjamen, Reading, Masschusetts: *Broken Symmetry and Correlation Functions*; 1975

[17] B. Gaveau and L.S. Schulman, EPJST **224**, 8, p 891 (2015).

[18] Gaveau B, Moreau M. Chaos. 2020; **30**:083104

[19] Arnold VI, Avez A. *Ergodic probems of Classical Mechanics*. Mathematical Physics Monographs: Benjamin; 1968

[20] Doob J. *Stochastic Processes*. N.Y: Wiley; 1953

[21] Levy P. *Théorie de l'addition des variables aléatoires*. Paris: Gauthier-Villars; 1937

[22] Feller W. *An Introduction to Probability Theory and Its Applications*, vol II. N.Y: Wiley; 1971

[23] McKean HP. *Stochastic Integrals*. London: Academic Press, NY; 1968

[24] Gaveau B, Gaveau MA. *Diffusion of a particle in a very rarefied gas*, Symposium on rarefied gas dynamics, Muntz, Weaver, Campell eds. Progress in Astronomics and Aeronautics. 1988; **118**(61)

[25] Jaynes ET. Phys. Rev. II **108**, 171 (1957). Phys. Rev. II. 1957;**106**:620

[26] Khinchin AI. *Mathemtical Foundations of Statistical Mechanics*. N.Y: Dover; 1949

[27] C. E. Shannon, *A mathematical theory of communication*, Bell System

Technical Journal, vol. 27, p. 449–423
and 623–656 (1948).

[28] Brillouin L. *Science an Information
Theory*. NY: Academic Press; 1956

[29] Khinchin AI. *Mathematical
Foundations of Information Theory*. N.Y:
Dover; 1957

[30] MacMillan R. *The basic theorems of
information theory*. Ann Math Statistics.
1953;**24**:193

[31] T. Cover. and J. Thomas,*Elements of
Information Theory*. Wiley: N.Y; 1991

[32] H. Poincaré. Acta Math. **13**, 1
(10890).

[33] Kac M. Bull. Amer.Math. Soc. 1947;
53:1002

[34] Wolfowitz J. Bull. Amer. Math. Soc.
1949;**55**:394

[35] Blum JR, Rosenblatt JI. J. Math. Sci.
(Delhi). 1967;**2**:1

[36] Schulman LS. Phys. Rev. A. *1978*;*18*
(5):2379

[37] Gallavotti G, Cohen EGD. J. Stat.
Phys. 1995;**80**:931