January 9, 2005

Ergodicity and the Numerical Simulation of Hamiltonian Systems

P. F. Tupper

Department of Mathematics and Statistics, McGill University
Montréal, Québec, CANADA, H3A 2K6.

Abstract

We discuss the long-time numerical simulation of Hamiltonian systems of ordinary differential equations. Our goal is to explain the ability of symplectic integration schemes such as Störmer-Verlet to compute accurate long-time averages for these systems in the context of molecular dynamics. This paper introduces a weakened version of ergodicity that allows us to study this problem. First, we demonstrate the utility of the weakened ergodicity definition by showing that it is a property of Hamiltonian systems robust to perturbations. Second, we study what the weakened ergodicity of a Hamiltonian system implies about numerical simulations of the system. In the case where a numerical method is volume-conserving and approximately energy-conserving, we show that long-time averages are approximated well for sufficiently small step lengths.

1   Introduction

In the field of molecular dynamics researchers simulate physical systems over long time intervals using numerical integration. Given the length of the time intervals and the size of time steps used, trajectories of the systems are not accurately computed [1]. Despite this, it is believed in many circumstance that the simulations provide reliable information about long-time averages of quantities, as well as qualitative behaviour [17]. Standard arguments from the numerical analysis of ordinary differential equations do not explain this performance [7].

In many cases of interest, the physical systems being simulated are believed to be *ergodic.* To introduce this term, we define two sorts of averages.

1

For any segment of a trajectory, we can consider the average of a function of the state variables along that segment. The limit of this average as the length of the segment goes to infinity is called an *ergodic* or *long-time average*. Alternatively, we can take the average of the function over the entire state-space of the system, which is known as the *ensemble* or *state-space average*. In general, long-time averages will depend on the initial condition chosen to generate the trajectory, and will not necessarily equal the state-space average. However, we say that a system is ergodic if, for almost all initial conditions, the long-time average is equal to the state-space average [18].

Here we will focus on Hamiltonian systems of ordinary differential equations. Given a Hamiltonian function, we select an initial condition and numerically approximate the solution to the resulting differential equations with a one-step method. Since energy is conserved, we take the state-space of our system to be all states with a particular energy. This set has a natural invariant measure induced by the Liouville measure [9].

If the Hamiltonian system is ergodic on the energy level set, then we can make the following non-rigorous argument to explain the accurate computation of long-time averages. Under the postulate of ergodicity, long-time averages for the original system are equal to the corresponding state-space averages. The numerical method applied to the system is another dynamical system with (approximately) the same state-space. If the numerical method is also ergodic (with a nearby invariant measure) then we expect its long-term averages to be equal to the corresponding state-space average as well. By this argument, the long-term averages of the original system and of its numerical approximation should be (approximately) the same for almost all initial conditions.

We could proceed by trying to make the steps of the above argument rigorous for at least some set of ergodic Hamiltonian systems and some numerical methods. Unfortunately, there is a serious problem with this approach. The difficulty stems from the large mismatch between the concept of ergodicity as it is used in the physics community and as it is used in the mathematics community [16]. For the physicist, ergodicity is a property that is postulated of a system in order to make many analytical and computational tasks tractable. The postulate is seldom justified from first principles. Instead, calculations for a system are performed assuming ergodicity. Then if theoretical predictions agree with experiments (actual or numerical), this is taken as evidence that the assumption is valid. Over the years, physicists have accumulated

extensive experience about which systems at which energies are well modeled by the assumption of ergodicity [19, 9].

For the mathematician, the justification of this assumption has posed a difficult task. Starting with the work of Sinai [15], there has been a long tradition of rigorously establishing ergodicity for specific Hamiltonian systems [18]. However, the majority of the systems for which ergodicity has been proved are billiard systems, which form a small subset of of the systems of interest to physicists. To the best of our knowledge, the only examples of physically realizable Hamiltonian systems of ODEs that are ergodic on some energy level set are given in [5] and [8] and are both quite specific in their construction. Ergodicity has not been proved for any system that has been seriously investigated with molecular dynamics simulations.

More troubling than the paucity of examples in this area is the number of negative results. The papers [4, 20] show that some of the ergodic billiard systems mentioned above are not ergodic for a range of energies when the billiard collisions are replaced by a soft collision of arbitrary stiffness. In [11], it is shown that generic Hamiltonian systems are not ergodic for any dense set of energies. Liverani [10] uses KAM theory to argue that interacting particle systems are not ergodic for sufficiently high energies. Although these results are not conclusive, it seems possible that the vast majority of Hamiltonian systems of physical interest are not ergodic in the strict mathematical sense of the word.

If we wish to perform a mathematically rigorous analysis of molecular dynamics based on the concept of ergodicity, we face a difficulty. The physical concept of ergodicity does not yet have a rigorous formulation. However, a theory based on the mathematical definition of ergodicity faces two problems. The first is that it is in danger of being irrelevant, since likely none of the systems one would want to apply the theory to are, in fact, ergodic. The second is that even if we restricted ourselves to mathematically ergodic Hamiltonian systems, we probably could not carry through the program: it is very unlikely that the perturbations induced in these systems by numerical approximation would result in ergodic systems according to the mathematical definition.

The first goal of this paper is to break this impasse by reviving a weakened definition of ergodicity which appears implicitly in Khinchin [9]. We believe that this mathematically precise definition is weak enough to hold true for many systems of physical interest, whilst still being strong enough to be of practical use. In particular, we will show that this definition is stable under perturbations of the dynamics. The second goal of this paper is to

3

justify long-term simulation in molecular dynamics by using this definition to prove theorems on the ability of certain numerical methods to accurately approximate the long-time averages of the original system.

First we formally state the traditional mathematical definition of ergodicity. Let Σ be a measure space with probability measure μ and let $S(t)$ be a measure-preserving flow on it. We denote the integral of a function f over Σ with respect to measure μ by

$$\mathbb{E}f(u) = \bar{f} = \int_{u \in \Sigma} f(u) \, d\mu(u).$$

We say that $S(t)$ is ergodic if for all integrable functions $f : \Sigma \to \mathbb{R}$

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(S(t)u) \, dt = \bar{f} \tag{1.1}$$

for μ-almost all $u \in \Sigma$. There is an analogous definition for maps rather than flows: Let S be a measure-preserving map from Σ to itself. We say that S is ergodic if, for all integrable $f : \Sigma \to \mathbb{R}$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(S^i u) = \bar{f}. \tag{1.2}$$

See, for example, [23].

A consequence of the above definitions of ergodicity is that for all bounded measurable f

$$\mathbb{E} \left| \lim_{T \to \infty} \frac{1}{T} \int_0^T f(S(t)u) \, dt - \bar{f} \right| = 0 \tag{1.3}$$

in the case of ergodic flows and

$$\mathbb{E} \left| \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(S^i u) - \bar{f} \right| = 0 \tag{1.4}$$

in the case of ergodic maps.

Building on (1.3) and (1.4) we can weaken the definition of ergodicity. The new definition depends on a parameter δ and a particular function $f : \Sigma \to \mathbb{R}$. For flows, we say that a μ-preserving flow $S(t)$ on Σ is δ-ergodic with respect to f if

$$\mathbb{E} \left| \lim_{T \to \infty} \frac{1}{T} \int_0^T f(S(t)u) \, dt - \bar{f} \right| \leq \delta. \tag{1.5}$$

4

For maps, we say that a μ-preserving map S on Σ is δ-ergodic with respect to f if

$$\mathbb{E}\left|\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f(S^i u)-\bar{f}\right|\leq\delta. \tag{1.6}$$

We can now state the main result of our paper.

> **Main Theorem** (Theorem 3.2) If a measure-preserving flow on a state-space is δ-ergodic with respect to a function f, then given any $\epsilon>0$, any sufficiently small perturbation of the flow with the same invariant measure and state-space will be $(\delta+\epsilon)$-ergodic with respect to f.

See Section 3 for a formal statement and proof. We will extend this theorem to the case of Hamiltonian flows with perturbed state-spaces and invariant measures later in Section 3, and to perturbations due to numerical discretization in Section 4.

The motivation for the above definitions of δ-ergodicity is a discussion of ergodicity by Khinchin in [9, p. 62]. After considering the difficulty of justifying the assumption of ergodicity for Hamiltonian systems, he discusses the possibility of requiring ergodicity to hold for only some special functions of interest. As an example of this, he goes on to prove that if a function f is nearly constant on the state space Σ then all measure-preserving flows are δ-ergodic with respect to f in the sense above for some small δ. For completeness, we will provide a proof of this statement in Subsection 3.1.

An example of the situation of a nearly constant f can be obtained by considering a gas of many particles in a rectangular container. Here we let Σ be the set of all states with a particular energy, and μ be the invariant measure on Σ induced by the Liouville measure (see below). Let f be the fraction of the total number of particles that are in the left half of the container. This quantity will be nearly equal to $1/2$ for almost all accessible states in Σ, and so Khinchin's theorem shows that (1.5) holds with this f for some small δ. Indeed, in any system of many components with short-range interactions, we expect that functions that are expressible as an average over all components in the system will have this property; see, for example, [13, 9].

However, these considerations are not sufficiently general to explain the utility of the ergodic hypothesis in all cases of interest. Continuing with the above example of the gas, consider the function of state that is 1 when a distinguished particle is on the left half of the box, and 0 otherwise. For

approximately half of all states of the system, the function has value 1, for the other half, 0. The average of this function over the state-space is $1/2$. Since the function is not nearly constant, the above discussion does not apply. However, we still expect that for any trajectory we examine, the particle will spend about half its time on the left-hand side of the box. We expect this because a gas in a box is one of many physical systems for which experience has shown the physical hypothesis of ergodicity to be justified. (As we have discussed, it is probably not ergodic in the formal, mathematical sense.) In the next section we will bolster these claims with some numerical experiments on a simple model of such a gas.

We now formally describe the Hamiltonian framework for our study. Let $H : \mathbb{R}^{2n} \to \mathbb{R}$ be a smooth function. Allow it to define the flow of a differential equation in $\mathbb{R}^{2n}$ via

$$\frac{du}{dt} = -J\nabla H(u),$$

where

$$J := \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}. \tag{1.7}$$

We use $S(t)$ to define the flow map induced by these equations. A property of Hamiltonian systems is that $H(S(t)u) = H(u)$ for all $t \in \mathbb{R}$. So we let our state-space be $\Sigma = \{u \in \mathbb{R}^{2n} | H(u) = E\}$ for some energy E.

If $S(t)$ is viewed as a flow on all of $\mathbb{R}^{2n}$, then by Liouville's theorem the standard Lebesgue measure (Euclidean volume) is invariant under it. However, (for sufficiently smooth H) the flow restricted to Σ has an invariant measure given by

$$d\mu(u) := C\frac{dS(u)}{\|\nabla H(u)\|}, \tag{1.8}$$

where C is some constant and dS is surface area on Σ [9]. Here we will always consider compact Σ and choose C so that $\mu(\Sigma) = 1$. The measure μ has the important feature that for any function $f : \mathbb{R}^{2n} \to \mathbb{R}$ the expectation of f on Σ with respect to the measure μ is given by [9]

$$\mathbb{E}f(u) = \lim_{\delta E \downarrow 0} \frac{\int_{H(v)\in[E,E+\delta E]} f(v)\, dv}{\int_{H(v)\in[E,E+\delta E]} dv}. \tag{1.9}$$

In addition to energy, Hamiltonian systems often have other trivial conserved quantities such as momentum. Σ may be further restricted to factor

6

out these. So when we talk about a Hamiltonian system being ergodic, we usually mean on a particular energy surface, with particular choices of conserved quantities.

We begin in Section 2 with some numerical experiments on a simple two-dimensional model of a gas. We will demonstrate some of the phenomena we hope to explain, examining numerically computed ergodic averages for a variety of functions, with different initial conditions and simulation step-lengths. In particular, we will consider two functions that are nearly constant in Khinchin's sense, and two that are not.

The remainder of the paper is devoted to analytical results about dynamical systems and algorithms. In Section 3 we will consider situations where δ-ergodicity can be shown to hold with respect to some function f. In Subsection 3.1 we provide a proof of Khinchin's result of [9] for the case of a nearly constant function f. In Subsections 3.2 and 3.3 we prove that if we make a small perturbation of an ergodic system, we will obtain a δ-ergodic system with a small δ. This provides us with a wide variety of provably δ-ergodic systems, which can be obtained by perturbing ergodic system such as billiard systems.

In Section 4 we will consider the simulation of δ-ergodic systems with numerical integrators. First we will consider the (unrealistic) situation of an integrator that conserves both phase-space volume and energy. We will show that the weakened ergodicity property is conserved in this situation. Then we consider the more realistic case of an integrator that conserves phase-space volume exactly, and energy approximately, over long periods of time. A shortcoming of the results is that the stepsize required in the proof of the theorem must be small enough that trajectories are accurately estimated over long, though finite, periods of time.

In a recent work, Reich [14] also considers the long-term simulation of Hamiltonian systems. There results are obtained under the strong assumption of uniformly hyperbolic dynamics [22], but with a more moderate restriction on stepsize. The present work relaxes the assumptions to just those of δ-ergodicity, though at the cost of much smaller stepsizes.

2 Ergodic Averages for a Lenard-Jones System: Numerical Experiments

In order to introduce some of the phenomena we will explore in this paper, we present the results of some simple, yet fairly typical, molecular dynamics calculations. We will perform several long-term simulations of a Hamiltonian system, with different initial conditions, and demonstrate the apparent convergence of ergodic averages to ensemble averages for four different functions f. Two of these functions will be what we refer to as nearly constant.

The system we will consider is a model of $N = 400$ monatomic atoms interacting on a two-dimensional periodic domain. Each atom has position $q_i \in [0, R]^2$ and momentum $p_i \in \mathbb{R}^2$. The Hamiltonian is given by

$$
\begin{aligned}
H(q, p) &= \sum_{i=1}^{N} V(\|q_i - q_j\|) + \sum_{i=1}^{N} \frac{\|p_i\|^2}{2m}, \\
V(r) &= \left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}.
\end{aligned}
$$

We fix $m = 1$, $\sigma = 0.2$, $R = 6$. The dynamics are Hamiltonian with periodic boundary conditions in the configuration space. The inter-particle forces are truncated so that particles do not interact if their centres are more than distance 5σ away from each other, as is standard practice [1].

The equations are integrated using the Störmer-Verlet scheme, which for separable Hamiltonians is [6]

$$
\begin{aligned}
p_{i+1/2} &= p_i - \Delta t \frac{\partial H}{\partial q}(q_i), \\
q_{i+1} &= q_i + \Delta t \frac{\partial H}{\partial p}(p_{i+1/2}), \\
p_{i+1} &= p_{i+1/2} - \Delta t \frac{\partial H}{\partial q}(q_{i+1}).
\end{aligned}
$$

This is a symplectic and time-reversible method method [7]. As a consequence of the former property, it conserves state-space volume: the method has the same invariant measure on $[0, 6]^2 \times \mathbb{R}^2$ as the original flow. However, unlike the flow, it does not conserve the energy function H.

We will consider four different functions of state:

$$F_1(q,p) := \frac{\|p_1\|^2}{2m},$$

$$F_2(q,p) := \frac{1}{N}\sum_{i=1}^{N}\frac{\|p_i\|^2}{2m},$$

$$F_3(q,p) := \mathbf{1}_A(q_1),$$

$$F_4(q,p) := \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_A(q_i).$$

Here we use $\mathbf{1}_A$ to denote the indicator function of A on $[0,R]^2$, where A denotes the left half of the periodic domain. The function F_1 gives the kinetic energy of particle 1. Function F_2 gives the average kinetic energy of all the particles, which corresponds to temperature. Functions F_3 is one if particle 1 is located in the set A and is zero otherwise. Function F_4 is the fraction of all the particles that are in A at a particular time. For this many ($N = 400$) particles, functions F_2 and F_4 can be considered nearly constant functions in Khinchin's terms, as we shall see.

We perform simulations with four different initial conditions, denoted by IC_i, $i = 1, \ldots, 4$. With IC_1 and IC_2 the particles are initially placed on a square periodic lattice with lattice spacing 0.3 units, which means that the particles are uniformly spaced over the entire domain. With IC_3 and IC_4 the particles are placed on a square periodic lattice with spacing $2^{1/6}\sigma$ units. (At this distance apart, two particles at rest are in equilbrium.) With regards to initial momenta, for IC_1 and IC_3 both x and y velocities are selected from a standard normal distribution. For IC_2 and IC_4 the 200 particles with the largest y values are given velocities $[1,0]^T$, and the remainder $[-1,0]^T$. In each case the velocities are first translated together so that total momentum is zero (if necessary), and then velocities are scaled together so that the total energy H is 100. We summarize these initial conditions in Table 1.

We denote the flow of the Hamiltonian system by $S(t)$, and the map given by the numerical method by $\hat{S}_{\Delta t}$. We estimate the ergodic averages of the original system

$$\frac{1}{T}\int_0^T F_i(S(t)(\text{IC}_j))\,dt,$$

	Initial q	Initial p_i
IC_1	uniformly spaced lattice	Gaussian
IC_2	uniformly spaced lattice	± 1
IC_3	tight packed lattice	Gaussian
IC_4	tight packed lattice	± 1

Table 1: Summary of the four different sets of initial conditions.

with the ergodic averages of the numerical method

$$\frac{1}{n} \sum_{k=0}^{n-1} F_i(\hat{S}_{\Delta t}^k(IC_j)).$$

Figure 2.1 shows the computed ergodic averages of the four functions versus $T = n\Delta t$ with the four different initial conditions, using a step length of $\Delta t = 0.001$. For an ergodic system, we would expect that the ergodic average of F_i should converge to the ensemble average of F_i with increasing T for each initial condition. For functions F_1, F_2, F_4 we observe the expected convergence. For F_3 it looks as though convergence will occur, though the plot is not conclusive. For all initial conditions the slowest convergence is for F_3 and the fastest for F_4.

Figure 2.2 shows the same data, over the same time interval, but with a steplength $\Delta t = 0.0005$. The plots are not qualitatively different from the previous ones. Thus, increasing the accuracy of the trajectories does not appear to change the ergodic behaviour of the simulation.

These results are consistent with the following picture of the situation: the flow of the underlying Hamiltonian system is δ-ergodic with respect to the four functions F_i for a small value of δ. The numerical method, despite not having the same invariant state-space as the flow, yields approximately the same ergodic averages over long periods of simulation. The subsequent sections of this paper will help establish the plausibility of these claims.

The four initial conditions we have considered are highly unrepresentative of the state-space of the system, which is why the initial values of F_2 and F_4 are not at all close to their ensemble average of $1/2$. To show that these functions are indeed nearly constant, we need to look at the distribution of each function applied to points randomly drawn from the state space according to the invariant measure. It is difficult to randomly generate such
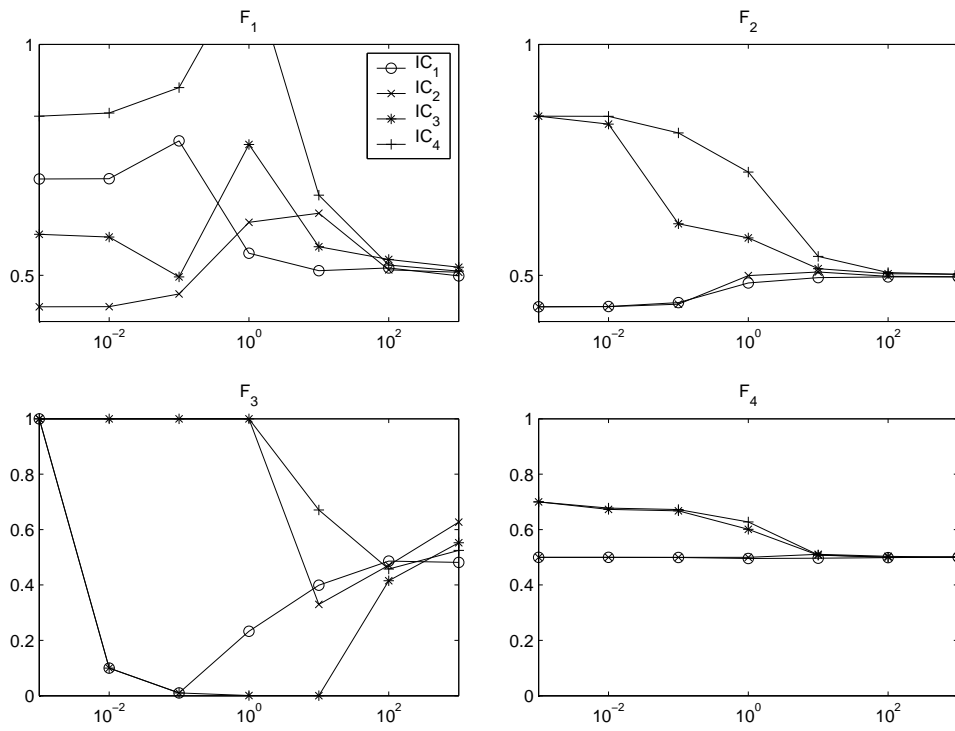
10

Figure 2.1: Ergodic averages of the functions versus T for four different initial conditions. $\Delta t = 0.001$.
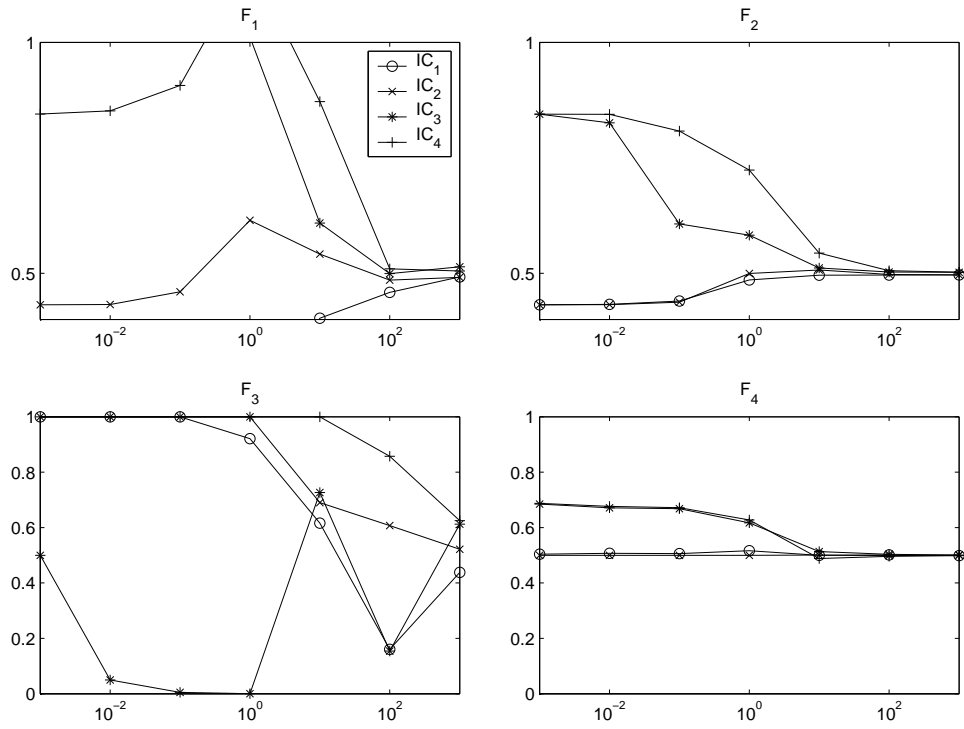
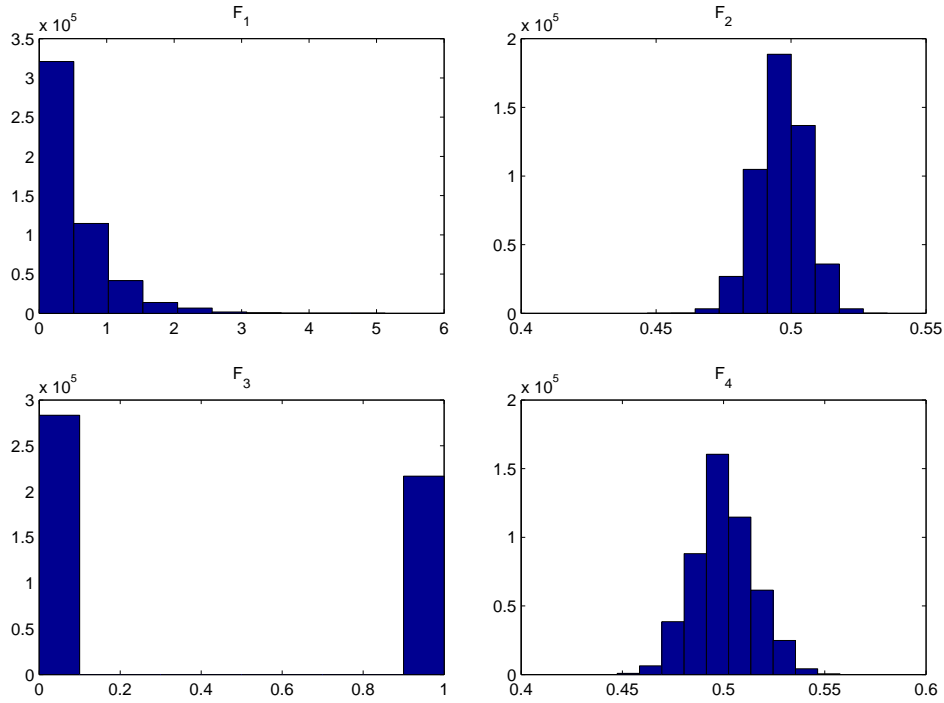Figure 2.2: Same as previous figure, but with $\Delta t = 0.001$.

Figure 2.3: Histograms of the four functions sampled over the state-space.

points so instead we will assume that numerical trajectories run long enough yield points with the correct statistics. Accordingly, we begin by running the simulation with $\Delta t = 0.001$ for 5×10^5 steps starting from IC_1. After this, we run for an additional 5×10^5 steps with the same steplength, but storing the function values of the state of the system after each step. In Figure 2.3 we show histograms of the functions values for each of the four functions. Assuming that after the initial run the computed trajectory samples the state-space with the correct density reasonably accurately, these plots show how the functions are distributed. The functions F_1 and F_3 show great variations, but the distribution of F_2 and F_4 have standard deviations less than 0.03 of their means, and thus could be considered nearly constant with $\delta \sim 0.03$. If N were increased with density and average energy kept constant, these standard deviations would converge to 0 like $N^{-1/2}$ [9].

Later in this paper we will consider the time T it takes for an ergodic average to converge to its limit within a tolerance of some $\delta > 0$, for most initial conditions. This is *not* what one observes in Figures 2.1 and 2.2, as
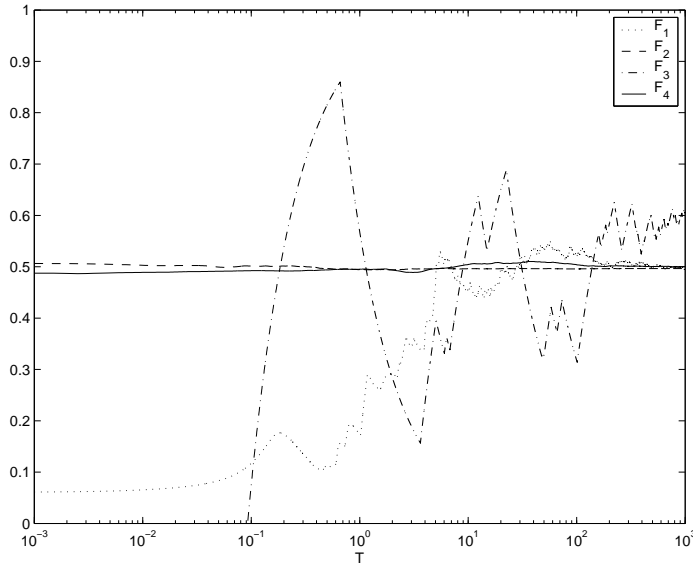
13

Figure 2.4: Computed ergodic averages versus time for the four functions, with initial conditions taken from end of MD simulation of length $T = 10^2$ started with IC_1. $\Delta t = 0.0005$.

the initial conditions are quite atypical of the entire state-space. To get more representative trajectories, we run a simulation from IC_1 for time 100 with $\Delta t = 0.0005$, and then take that as our initial condition for the computation of ergodic averages. Figure 2.4 shows the ergodic averages for the four functions with these initial conditions for increasing times.

We observe that for the two nearly constant functions, F_2 and F_4, the ergodic average is never far from the ensemble average, and would not be even with $T = \Delta t$. For the other two functions it still takes a considerable duration of time for the ergodic averages to converge, even with these initial conditions. This sharp difference in rate of convergence will have a large effect on the strength of our theoretical results in Section 4.

The number of papers featuring molecular dynamics simulations of Lenard-Jones systems is immense. However, the seminal references are [12] and [21], both of which are contained in [3]. An interesting numerical investigation of a Lenard-Jones system with ergodicity as its primary focus is [16]. Both [16] and [1] contain the results of molecular dynamics simulations showing the rapid divergence of trajectories, and thus establishing the unreliabilty of individually computed trajectories in these situations.

14

3 δ-Ergodic Systems

There are two approaches to using the weakened ergodicity definition (1.5) of the Introduction. We may simply postulate that the system we are working with is δ-ergodic with respect to some f of interest. Working from this assumption, we may infer what we can about the system and numerical approximations of it. Alternatively, we may try and prove that the system is indeed δ-ergodic for some δ and f.

In this section we consider a few situations where the latter is possible. In Subsection 3.1 we consider the situation where the function f of interest is nearly constant over most of the state-space. In Subsection 3.2 we will consider the case in which the state-space and invariant measure of the original system and the perturbed system are identical. In these two cases our results are stated for very general flows, not necessarily arising from differential equations. In Subsection 3.3 we will consider the situation when the state-space as well as the flow is perturbed. Here we will state results only for the situations when the flow is induced by a Hamiltonian function and the perturbed flow arises from a perturbation of the Hamiltonian.

3.1 Nearly Constant Function f

Frequently in statistical mechanical applications we are not interested in functions of state that depend upon only one particle, but functions which are averages taken over all the particles in the system. A classic example of this is temperature, which is related to the average kinetic energy of all particles. We showed two examples of such functions in Section 2: F_2 and F_4. Typically, these functions are nearly constant for most accessible states [13, 9]. We will show that under these circumstances any measure-preserving flow is weakly ergodic.

Let Σ be a compact manifold in $\mathbb{R}^m$ with the Borel σ-algebra $\mathcal{M}$. We let μ be a probability measure on $(\Sigma, \mathcal{M})$. We let $\mathbb{E}$ denote the expectation of a function on Σ with respect to μ. One way to formalize the hypothesis of a nearly constant function f is to assume $\mathbb{E}|f(u) - \bar{f}|$ is small, where $\bar{f} := \mathbb{E}f$. This covers the possibility of f being close to its average on most of state-space, as well as it being far from its average on a small portion of state-space.

Theorem 3.1 *Suppose*

1. $f : \Sigma \to \mathbb{R}$ is bounded and measurable with $\mathbb{E}|f(u) - \bar{f}| \leq \delta$.

2. $S(t)$ is a μ-preserving flow on Σ.

Then

$$\mathbb{E}\left|\lim_{T \to \infty} \frac{1}{T} \int_0^T f(S(t)u)\, dt - \bar{f}\right| \leq \delta. \qquad (3.10)$$

That is, $S(t)$ is δ-ergodic with respect to f.

Proof By Birkhoff's Ergodic Theorem [23, 9] the limit

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(S(t)u)\, dt$$

is defined for almost all u. So the quantity on the left of (3.10) is defined. Accordingly,

$$\mathbb{E}\left|\lim_{T \to \infty} \frac{1}{T} \int_0^T f(S(t)u)\, dt - \bar{f}\right| \leq \mathbb{E}\lim_{T \to \infty} \frac{1}{T} \int_0^T |f(S(t)u) - \bar{f}|\, dt$$

$$= \lim_{T \to \infty} \frac{1}{T} \int_0^T \mathbb{E}|f(S(t)u) - \bar{f}|\, dt$$

where we have used the Bounded Convergence Theorem. Since $S(t)$ is μ invariant we can remove the $S(t)$ from the last expression. So the quantity of interest is less than or equal to

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \mathbb{E}|f(u) - \bar{f}|\, dt = \mathbb{E}|f(u) - \bar{f}| \leq \delta$$

as required. $\square$

An application of the Markov inequality gives

$$\mu\left\{u : \left|\lim_{T \to \infty} \frac{1}{T} \int_0^T f(S(t)u)\, dt - \bar{f}\right| \geq \alpha\right\} \leq \frac{\delta}{\alpha}$$

which is precisely Khinchin's result in [9] but with a better constant.

3.2   Perturbations of Ergodic Flows: Same State-Space

As before, Σ is a compact manifold in $\mathbb{R}^m$, with induced Borel σ-algebra $\mathcal{M}$. We let μ be an arbitrary probability measure on $(\Sigma, \mathcal{M})$. We will use $\mathbb{E}$ to denote expectation with respect to the measure μ on Σ. We let Σ have the metric it inherits from the norm on $\mathbb{R}^m$, which is denoted by $|\cdot|$.

Suppose we have an ergodic measure-preserving flow $S(t)$ on Σ that leaves μ invariant. Suppose $S_\gamma(t)$, $\gamma > 0$ are perturbations of the flow that also act on Σ and have the same invariant measure. Without assuming that the $S_\gamma(t)$ are ergodic, what can we say about long-time averages of the perturbed flow? In the following, we will show that if $S(t)$ is a δ-ergodic flow with respect to a function f and $\epsilon > 0$, then $S_\gamma(t)$ is $(\delta + \epsilon)$-ergodic with respect to f for sufficiently small γ.

Theorem 3.2 *Let Σ be a compact manifold in $\mathbb{R}^n$. Suppose*

1. *$f : \Sigma \to \mathbb{R}$ is Lipschitz-continuous.*

2. *$S(t)$ is a μ-preserving flow on Σ that is δ-ergodic with respect to f.*

3. *For each $\gamma > 0$, $S_\gamma(t)$ is a μ-preserving flow on Σ such that for all $T \in [0, \infty)$*
$$\lim_{\gamma \to 0} \sup_{t \in [0,T]} \mathbb{E} |S_\gamma(t)u - S(t)u| = 0.$$

Then for all $\epsilon > 0$ there is a $\gamma_0 > 0$ such that if $\gamma \leq \gamma_0$ then

$$\mathbb{E} \left| \lim_{T \to \infty} \frac{1}{T} \int_0^T f(S_\gamma(t)u) \, dt - \bar{f} \right| \leq \delta + \epsilon, \tag{3.11}$$

where $\bar{f} = \mathbb{E} f(u)$. That is, for any $\epsilon > 0$, $S_\gamma(t)$ is $(\delta + \epsilon)$-ergodic with respect to f for sufficiently small γ.

Remark: The third hypothesis above is implied by the assumption that for all $T \in [0, \infty)$

$$\lim_{\gamma \to 0} \left(\sup_{u \in \Sigma} \sup_{t \in [0,T]} |S_\gamma(t)u - S(t)u| \right) = 0. \tag{3.12}$$

(Recall that Σ is compact.) Though stronger than the stated hypothesis, (3.12) is perhaps simpler.

Proof For each $\gamma > 0$ Birkhoff's Ergodic Theorem implies that

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T f(S_\gamma(t)u)\, dt$$

exists for μ-almost all u. Accordingly the quantity on the left of (3.11) is defined. Moreover, since f is bounded on Σ we have

$$\mathbb{E} \left| \lim_{T \to \infty} \frac{1}{T} \int_0^T f(S_\gamma(t)u)\, dt - \bar{f} \right| = \lim_{T \to \infty} \mathbb{E} \left| \frac{1}{T} \int_0^T f(S_\gamma(t)u)\, dt - \bar{f} \right|,$$

so that the latter limit also exists. Thus if we can show that

$$\mathbb{E} \left| \frac{1}{T} \int_0^T f(S_\gamma(t)u)\, dt - \bar{f} \right| \leq \delta + \epsilon$$

for a sequence of times T that go to infinity, we will be done.

Since $S(t)$ is δ-ergodic with respect to f

$$\lim_{T \to \infty} \mathbb{E} \left| \frac{1}{T} \int_0^T f(S(t)u)\, dt - \bar{f} \right| \leq \delta.$$

Choose a T such that

$$\mathbb{E} \left| \frac{1}{T} \int_0^T f(S(t)u)\, dt - \bar{f} \right| < \delta + \epsilon/2.$$

Now given this T, for any γ

$$\mathbb{E} \left| \frac{1}{T} \int_0^T f(S_\gamma(t)u)\, dt - \frac{1}{T} \int_0^T f(S(t)u)\, dt \right| \leq \frac{1}{T} \int_0^T \mathbb{E}|f(S_\gamma(t)u) - f(S(t)u)|\, dt$$
$$\leq F \sup_{t \in [0,T]} \mathbb{E}|S_\gamma(t)u - S(t)u|,$$

where F is the Lipschitz constant of f on Σ. Thus by hypothesis, we can choose a γ_0 such that for $\gamma \leq \gamma_0$

$$\mathbb{E} \left| \frac{1}{T} \int_0^T f(S_\gamma(t)u)\, dt - \frac{1}{T} \int_0^T f(S(t)u)\, dt \right| \leq \epsilon/2.$$

So for $\gamma \leq \gamma_0$

$$\mathbb{E} \left| \frac{1}{T} \int_0^T f(S_\gamma(t)u)\, dt - \bar{f} \right| \leq \delta + \epsilon.$$

18

This gives us a result for a particular time interval $[0, T]$. For integer multiples mT of the time T, we obtain

$$\mathbb{E}\left|\frac{1}{mT}\int_0^{mT} f(S_\gamma(t)u)\,dt - \bar{f}\right| = \mathbb{E}\left|\frac{1}{m}\sum_{i=1}^m \frac{1}{T}\int_0^T f(S_\gamma((i-1)T+t)u)\,dt - \bar{f}\right|$$

$$\leq \frac{1}{m}\sum_{i=1}^m \mathbb{E}\left|\frac{1}{T}\int_0^T f(S_\gamma((i-1)T+t)u)\,dt - \bar{f}\right|.$$

Since the measure μ is invariant under the flow of $S_\gamma(t)$, we can replace $S_\gamma((i-1)T+t)$ with $S_\gamma(t)$ in the last expression. So

$$\mathbb{E}\left|\frac{1}{mT}\int_0^{mT} f(S_\gamma(t)u)\,dt - \bar{f}\right| \leq \frac{1}{m}\sum_{i=1}^m \mathbb{E}\left|\frac{1}{T}\int_0^T f(S_\gamma(t)u)\,dt - \bar{f}\right|$$

$$= \mathbb{E}\left|\frac{1}{T}\int_0^T f(S_\gamma(t)u)\,dt - \bar{f}\right|$$

$$\leq \delta + \epsilon.$$

Since $mT \to \infty$ as $m \to \infty$ we have established our result. $\square$

Taking the case where $\delta = 0$ gives the following result:

Theorem 3.3 *Let the assumptions 1 and 3 of Theorem 3.2 hold. Let $S(t)$ be an ergodic μ-preserving flow on Σ. Then for any $\epsilon > 0$, $S_\gamma(t)$ is ϵ-ergodic with respect to f for sufficiently small γ. Alternatively, we may write*

$$\lim_{T \to \infty} \frac{1}{T}\int_0^T f(S_\gamma(t)u)\,dt \to \bar{f}.$$

in L_1 as $\gamma \to 0$. $\square$

Remarks: The hypotheses make quite minimal assumptions about the ergodic properties of the flow S. In particular, no mixing or hyperbolicity assumptions are used. The cost of this is that truly minuscule values of γ are needed to obtain a reasonably small ϵ. For example, consider a case where S is ergodic ($\delta = 0$) and is generated by a vector field on the manifold Σ. Suppose the S_γ are generated by perturbations of size $\mathcal{O}(\gamma)$ in the vector field. In general, $S(t)u$ and $S_\gamma(t)u$ will differ on the order of $\gamma \exp(Ct)$ for some constant C. By analogy with the Central Limit Theorem, we might assume that

$$\mathbb{E}\left|\frac{1}{T}\int_0^T f(S(t)u)\,dt - \bar{f}\right| = \mathcal{O}(T^{-1/2}).$$

19

So in the proof of Theorem 3.2 we need a T of order ϵ^{-2}. The second part of the proof requires that we choose a γ small enough so that perturbed trajectories closely agree within ϵ of the original trajectories over time T. This would require a γ on the order of $\epsilon \exp(-CT)$, Putting this together we find that γ depends on ϵ like

$$\gamma \sim \epsilon \exp[-C\epsilon^{-2}]$$

for some C. For most practical situations this would lead to a γ far too small to explain the prevalence of apparently ergodic behaviour. We discuss this point further with regards to numerical integrators in Subsection 4.2. We will attempt to address this point in future work by making stronger assumptions on the flow of the unperturbed Hamiltonian system.

3.3 Perturbations of Ergodic Flows: Perturbed State-Space

A very strong assumption in Theorem 3.2 is that the perturbed flows have the same state-space Σ and invariant measure μ as the original flow. In general we do not expect this to happen. Here we will consider the situation where the original system is the flow of an autonomous Hamiltonian system on an energy level-set, and the perturbed system is obtained by perturbing the Hamiltonian function. The arguments here can be applied to more general situations as well.

In particular, let our flow $S(t)$ be generated by a Hamiltonian function $H : \mathbb{R}^{2n} \to \mathbb{R}$ via the system of differential equations $du/dt = -J\nabla H(u)$, where J is the $2n \times 2n$ constant matrix defined in (1.7). We take the state-space to be

$$\Sigma := \{u \in \mathbb{R}^{2n} | H(u) = E\}$$

for some energy E. Thus, Σ is invariant under the induced flow $S(t)$ and has an invariant measure μ given by (1.8).

Suppose the perturbed flows $S_\gamma(t)$ are generated by perturbed Hamiltonians $H_\gamma : \mathbb{R}^{2n} \to \mathbb{R}$, such that $H_\gamma \to H$ as γ goes to zero. For each γ, the state-space of the perturbed flow is

$$\Sigma_\gamma := \{u \in \mathbb{R}^{2n} | H_\gamma(u) = E\}.$$

For each sufficiently small γ, Σ_γ will have a probability measure μ_γ invariant under $S_\gamma(t)$, corresponding to the Liouville measure.

More formally, we define our Hamiltonian function $H(u)$ and perturbed Hamiltonian functions $H_\gamma(u)$ through a function $\mathcal{H}(u, \gamma)$, by letting $H_\gamma(u) := \mathcal{H}(u, \gamma)$ and $H(u) := \mathcal{H}(u, 0)$. We make the following assumptions on $\mathcal{H}(u, \gamma)$ and a corresponding energy level E:

Assumptions 3.4 *The function $\mathcal{H}(u, \gamma) : \mathbb{R}^{2n} \times [0, \infty) \to \mathbb{R}$ and the value $E \in \mathbb{R}$ satisfy*
(i) $\mathcal{H}$ is C^2 (twice continuously differentiable) and $\nabla_u^2 \mathcal{H}$ is locally Lipschitz
(ii) $\Sigma := \{u : \mathcal{H}(u, 0) = E\}$ is bounded
(iii) $\nabla_u \mathcal{H}(u, 0) \neq 0$ on Σ

These assumptions guarantee that Σ is a compact manifold, as is Σ_γ for sufficiently small γ. Moreover, as is shown in the appendix in Lemma A.1, Σ and Σ_γ are isomorphic for small enough γ.

In the following theorem the symbol $\mathbb{E}$ ($\mathbb{E}_\gamma$) will denote integration over Σ (Σ_γ) with measure μ (μ_γ). Also note that though $S(t)$ leaves Σ invariant, and $S_\gamma(t)$ leaves Σ_γ invariant, we view both as flows on all of $\mathbb{R}^{2n}$.

Theorem 3.5 *Suppose*

1. *$\mathcal{H}$ and E satisfy Assumptions 3.4*

2. *$f : \mathbb{R}^{2n} \to \mathbb{R}$ is locally Lipschitz-continuous.*

3. *The flow $S(t)$ generated by H is δ-ergodic with respect to f on (Σ, μ).*

Then for all $\epsilon > 0$ there is a $\gamma_0 > 0$ such that if $\gamma \leq \gamma_0$ then

$$\mathbb{E}_\gamma \left| \lim_{T \to \infty} \frac{1}{T} \int_0^T f(S_\gamma(t)u) \, dt - \bar{f}_\gamma \right| \leq \delta + \epsilon$$

where $\bar{f}_\gamma := \mathbb{E}_\gamma f$. That is, $S_\gamma(t)$ is $(\delta + \epsilon)$-ergodic with respect to f on Σ_γ for sufficiently small γ.

Proof First note that the vector field that generates $S(t)$ is $J\nabla_u \mathcal{H}(u, 0)$ and the vector field that generates $S\gamma(t)$ is $J\nabla_u \mathcal{H}(u, \gamma)$. By Assumptions 3.4 the latter vector field converge to the former for every $u \in \mathbb{R}^{2n}$ as $\gamma \to 0$. Thus, for each $u \in \mathbb{R}^{2n}$ and $T > 0$,

$$\lim_{\gamma \downarrow 0} \sup_{t \in [0, T]} \|S_\gamma(t)u - S(t)u\|, \tag{3.13}$$

and this will occur uniformly on bounded sets in $\mathbb{R}^{2n}$.

From here, the proof is similar to that of Theorem 3.2. Choose a T such that

$$\mathbb{E}\left|\frac{1}{T}\int_0^T f(S(t)u)\,dt - \bar{f}\right| < \delta + \epsilon/3. \qquad (3.14)$$

Define

$$g(u) := \left|\frac{1}{T}\int_0^T f(S(t)u)\,dt - \bar{f}\right|$$

and for any γ

$$g_\gamma(u) := \left|\frac{1}{T}\int_0^T f(S_\gamma(t)u)\,dt - \bar{f}_\gamma\right|,$$

for $u \in \mathbb{R}^{2n}$ in a neighbourhood of Σ. Expression (3.14) implies that $\mathbb{E}g(u) < \delta + \epsilon/3$. We will show that for small enough γ, $\mathbb{E}_\gamma g_\gamma(u) < \delta + \epsilon$. To do this observe

$$|\mathbb{E}g(u) - \mathbb{E}_\gamma g_\gamma(u)| \le |\mathbb{E}g(u) - \mathbb{E}_\gamma g(u)| + |\mathbb{E}_\gamma[g(u) - g_\gamma(u)]|.$$

By Lemma A.2 the first term on the right can be made smaller than $\epsilon/3$ with a small enough γ. To show that the second term can be made smaller than $\epsilon/3$ as well, note

$$
\begin{aligned}
|\mathbb{E}_\gamma[g(u) - g_\gamma(u)]| \;&\le\; \mathbb{E}_\gamma|g(u) - g_\gamma(u)| \\
&\le\; \sup_{u \in \Sigma_\gamma}\left|\frac{1}{T}\int_0^T f(S_\gamma(t)u) - f(S(t)u)\,dt\right| + |\bar{f} - \bar{f}_\gamma| \\
&\le\; F\sup_{u \in N}\sup_{t \in [0,T]}\|S_\gamma(t)u - S(t)u\| + |\mathbb{E}f(u) - \mathbb{E}_\gamma f(u)|
\end{aligned}
$$

where N is a neighbourhood of Σ, and F is the Lipschitz constant of f on N. The first term can be made small by 3.13 and the second term can be made small by Lemma A.2.

We have now shown that

$$\mathbb{E}_\gamma\left|\frac{1}{T}\int_0^T f(S_\gamma(t)u)\,dt - \bar{f}_\gamma\right| \le \delta + \epsilon$$

for small enough γ. As before we can go from this result on $[0,T]$ to the result on $[0, mT]$, and then to the limit as $T \to \infty$. $\quad\square$

4   Numerical Approximations

We now consider the numerical integration of Hamiltonian flows that are δ-ergodic with respect to some f. The first subsection considers the situation of integrating a system with a volume- and energy-conserving integrator. Here the results are particularly clear and easy to prove. We then consider the more important case of integrators that conserve phase-space volume exactly, and energy approximately, over long time intervals. The Störmer-Verlet method, used in Section 2, falls into this class when applied to many Hamiltonian systems of interest.

4.1   Volume- and Energy- Conserving Integrators

In this section we will consider the situation in which a numerical integrator is applied which conserves both the Hamiltonian function (energy) and the phase-space volume. Since no general purpose integrator for Hamiltonian systems is known that preserves both these quantities, the purpose of this subsection is primarily to prepare us for the next subsection in which we consider symplectic methods.

As before, we consider a Hamiltonian system with Hamiltonian function H and let

$$\Sigma = \{u \in \mathbb{R}^{2n} | H(u) = E\} \tag{4.15}$$

for some energy E. We denote the flow induced by the Hamiltonian by $S(t)$. We let $\hat{S}_{\Delta t}$ denote a one-step integrator with constant step length, and $\hat{S}_{\Delta t}^i$ the composition of this map with itself i times. Given an initial value $u \in \mathbb{R}^{2n}$, $\hat{S}_{\Delta t}u$ is an approximation to $S(\Delta t)u$. Moreover, $\hat{S}_{\Delta t}^i u$ is an approximation to $S(i\Delta t)u$, for $i \geq 0$.

We remark that if we assume that the approximate map $\hat{S}_{\Delta t}$ is ergodic on Σ in addition to the assumptions below, then immediately for μ-almost all initial conditions numerically computed ergodic averages are equal to $\bar{f}$. However, we will not make this assumption about the numerical method.

Theorem 4.1 *Let $H : \mathbb{R}^{2n} \to \mathbb{R}$, and $E \in \mathbb{R}$. Suppose*

  1. *∇H is locally Lipschitz and the set Σ defined by (4.15) is a compact manifold*

  2. *$f : \mathbb{R}^{2n} \to \mathbb{R}$ is locally Lipschitz,*

23

3. $S(t)$ is δ-ergodic with respect to f on (Σ, μ) for some $\delta \geq 0$,

4. $\hat{S}_{\Delta t}$, $\Delta t > 0$ is a family of volume-preserving isomorphisms of $\mathbb{R}^{2n}$ that leave Σ invariant and such that for any $T > 0$

$$\lim_{\Delta t \to 0} \mathbb{E} \sup_{0 \leq i \leq \lfloor T/\Delta t \rfloor} \|\hat{S}_{\Delta t}^i u - S(i\Delta t)u\| = 0.$$

Then for all $\epsilon > 0$, there is a Δt_0 such that for all $\Delta t \leq \Delta t_0$,

$$\mathbb{E} \left| \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(\hat{S}_{\Delta t}^i u) - \bar{f} \right| \leq \delta + \epsilon.$$

In other words, $\hat{S}_{\Delta t}$ is $(\delta + \epsilon)$-ergodic with respect to f, for sufficiently small Δt.

Remark: The fourth hypothesis merely states that $\hat{S}_{\Delta t}$ is a convergent numerical method, though with some uniformity in u required.

Proof Since $\hat{S}_{\Delta t}$ is μ-preserving, as in Theorem 3.2 it suffices to show that for sufficiently small Δt

$$\mathbb{E} \left| \frac{1}{n} \sum_{i=0}^{n-1} f(\hat{S}_{\Delta t}^i u) - \bar{f} \right| \leq \delta + \epsilon$$

for some n which may depend on Δt.

Choose a T_0 large enough so that

$$\mathbb{E} \left| \frac{1}{T} \int_0^T f(S(t)u) \, dt - \bar{f} \right| \leq \delta + \epsilon/2, \tag{4.16}$$

for all $T \geq T_0$. Now choose a $\Delta t_0 \leq T_0/2$ such that whenever $\Delta t \leq \Delta t_0$, and n is an integer such that $n\Delta t \leq 2T_0$ then

$$\mathbb{E} \left| \frac{1}{n\Delta t} \int_0^{n\Delta t} f(S(t)u) \, dt - \frac{1}{n} \sum_{i=0}^{n-1} f(\hat{S}_{\Delta t}^i u) \right| \leq \epsilon/2. \tag{4.17}$$

24

This is possible for the following reasons. For any such n we can write

$$\left| \frac{1}{n\Delta t} \int_0^{n\Delta t} f(S(t)u)\,dt - \frac{1}{n} \sum_{i=0}^{n-1} f(\hat{S}_{\Delta t}^i u) \right|$$

$$\leq \left| \frac{1}{n\Delta t} \int_0^{n\Delta t} f(S(t)u)\,dt - \frac{1}{n} \sum_{i=0}^{n-1} f(S(i\Delta t)u) \right|$$

$$+ \left| \frac{1}{n} \sum_{i=0}^{n-1} f(S(i\Delta t)u) - \frac{1}{n} \sum_{i=0}^{n-1} f(\hat{S}_{\Delta t}^i u) \right|$$

The expectation of the first term yields

$$\mathbb{E} \left| \frac{1}{n\Delta t} \int_0^{n\Delta t} f(S(t)u)\,dt - \frac{1}{n} \sum_{i=0}^{n-1} f(S(i\Delta t)u) \right|$$

$$\leq \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E} \left| \frac{1}{\Delta t} \int_0^{\Delta t} f(S(t + i\Delta t)u)\,dt - f(S(i\Delta t)u) \right|$$

$$\leq \mathbb{E} \left| \frac{1}{\Delta t} \int_0^{\Delta t} f(S(t)u)\,dt - f(u) \right|$$

$$\leq F\mathbb{E} \sup_{t \in [0,\Delta t]} \|S(t)u - u\| \leq FC\Delta t$$

for some constant C, which converges to zero as Δt goes to zero. (F is the Lipschitz constant of f on Σ). The expectation of the second term is

$$\mathbb{E} \left| \frac{1}{n} \sum_{i=0}^{n-1} f(S(i\Delta t)u) - \frac{1}{n} \sum_{i=0}^{n-1} f(\hat{S}_{\Delta t}^i u) \right| \leq F\mathbb{E} \sup_{0 \leq i \leq n} \|\hat{S}_{\Delta t}^i u - S(i\Delta t)u\|$$

$$\leq F\mathbb{E} \sup_{0 \leq i \leq \lfloor 2T_0/\Delta t \rfloor} \|\hat{S}_{\Delta t}^i u - S(i\Delta t)u\|$$

which converges to zero as Δt goes to zero by the fourth hypothesis. Thus there is some Δt_0 such that the bound (4.17) holds for all $\Delta t \leq \Delta t_0$ and for all $n\Delta t \leq 2T_0$.

Now for any $\Delta t \leq \Delta t_0$ there is an integer n such that $T := n\Delta t \in [T_0, 2T_0]$. So by (4.16) and (4.17)

$$\mathbb{E} \left| \frac{1}{n} \sum_{i=1}^n f(\hat{S}_{\Delta t}^i u) - \bar{f} \right| \leq \delta + \epsilon$$

as required. □

25

4.2 Volume-Conserving and Approximately Energy-Conserving Integrators

We now relax the assumption of energy conservation which we used in the previous subsection. We replace it by an assumption that on long time-intervals the integrator keeps the energy of the system within a narrow range of the initial energy.

The most important class of volume-preserving integrators are the symplectic integrators, such as the Störmer-Verlet method described in Section 2. In many cases they can be shown, using backward error analysis, to approximately conserve energy over long time intervals [2]. Suppose we are integrating a Hamiltonian system of differential equations with Hamiltonian function H and we are using a symplectic integrator $\hat{S}_{\Delta t}$ with steplength Δt. In many situations it is possible to prove that for each Δt there is a perturbed Hamiltonian $H_{\Delta t} \approx H$ with induced flow $S_{\Delta t}(t)$ such that

$$\|\hat{S}_{\Delta t} u - S_{\Delta t}(\Delta t) u\| \leq C \Delta t \exp[-D/\Delta t],$$

for constants C, D, [2, 14, 6]. This in turn implies that

$$H_{\Delta t}(\hat{S}_{\Delta t}^n u) = H_{\Delta t}(u) + \mathcal{O}(e^{-h_0/2\Delta t})$$

and

$$H(\hat{S}_{\Delta t}^n u) = H(u) + \mathcal{O}(\Delta t^p) \tag{4.18}$$

for $n\Delta t \leq e^{-h_0/2\Delta t}$.

Here, we will not explicitly use any backward error analysis. Rather, we will assume that the energy error satisfies uniform bounds over long time intervals, as in (4.18). Our reasons for this are two-fold. Firstly, energy conservation of this kind has been shown to exist for symplectic integrators in situations where backward error analysis is not valid, [6, Ch. XIII]. Secondly, our theorem applies to volume-conserving integrators that are not symplectic as long as they show a similar energy-conservation property.

We note that there is another class of integrators for which a backward error analysis has been developed: time-reversible integrators [6]. These integrators can also be shown to nearly conserve energy over long periods of time, as well as perform well at computing long-time averages. However, our present theory does not apply to them, since they are not, in general, volume-conserving.

26

We face a new technical difficulty in this subsection. In the previous subsection, the exact energy conservation of the integrator implied that it had an energy level set as an invariant domain. In the present case, it may be that even over short times there will be no non-trivial invariant sets. So we must resort to an approximate invariance. We want to define sets Σ such that $\Sigma \backslash \hat{S}^n_{\Delta t} \Sigma$ has small measure relative to Σ for large n. In general this is impossible to do if Σ has co-dimension one, as it does for energy level-sets. Accordingly, our theorem will apply to sets Σ of the following form. We let

$$\Sigma := \{u \in \mathbb{R}^{2n} | H(u) \in [E_1, E_2]\}. \tag{4.19}$$

We let μ be the volume measure restricted to Σ and normalized so that $\mu\Sigma = 1$. Thus for functions g defined on Σ we define

$$\mathbb{E}g := \int_{\Sigma} g(u) du / \text{vol} \, (\Sigma).$$

For other sets A of non-zero volume in $\mathbb{R}^{2n}$ we define

$$\mathbb{E}_A g := \int_A g(u) du / \text{vol} \, (A).$$

The flow $S(t)$ leaves Σ and μ invariant. Although $S(t)$ cannot be ergodic on Σ due to energy conservation, it can be δ-ergodic with respect to some f— especially if E_1 and E_2 are close. In particular, it is not difficult to show that if $S(t)$ is δ-ergodic with respect to a continuous f on the set of all states of energy E_1 then for all $\epsilon > 0$ it is $(\delta + \epsilon)$-ergodic on Σ for small enough $E_2 - E_1$.

Theorem 4.2 *Let $H : \mathbb{R}^{2n} \to \mathbb{R}$, and $E_1 < E_2 \in \mathbb{R}$. Suppose*

1. *∇H is locally Lipschitz and*

$$\{u \in \mathbb{R}^{2n} | H(u) = E\}$$

   is a compact manifold for all E in an open neighbourhood of $[E_1, E_2]$.

2. *$f : \mathbb{R}^{2n} \to \mathbb{R}$ is a locally Lipschitz*

3. *the flow $S(t)$ generated by H is δ-ergodic on (Σ, μ) with respect to f.*

4. $\hat{S}_{\Delta t}$, $\Delta t > 0$ is a family of volume-preserving maps on $\mathbb{R}^{2n}$ such that for any $T > 0$

$$\lim_{\Delta t \to 0} \sup_{0 \le i \le \lfloor T/\Delta t \rfloor} \|\hat{S}^i_{\Delta t} u - S(i\Delta t)u\| = 0.$$

uniformly over bounded sets in $\mathbb{R}^{2n}$.

5. there exists constant C, exponent $p > 0$, and times $\mathcal{T}_{\Delta t}$, with $\mathcal{T}_{\Delta t} \to \infty$ as $\Delta t \to 0$ such that

$$|H(\hat{S}^n_{\Delta t} u) - H(u)| \le C\Delta t^p$$

when $n\Delta t \in [0, \mathcal{T}_{\Delta t}]$.

It follows that, for all $\epsilon > 0$, there is a Δt_0 and a $\mathcal{T}_{\min}$ such that if $\Delta t \le \Delta t_0$ and $n\Delta t \in [\mathcal{T}_{\min}, \mathcal{T}_{\Delta t}]$ then

$$\mathbb{E} \left| \frac{1}{n} \sum_{i=0}^{n-1} f(\hat{S}^i_{\Delta t} u) - \bar{f} \right| \le \delta + \epsilon.$$

Proof We begin by assuming that $|f|$ is bounded and has supremum $f_{\max}$. The conclusion for arbitrary f follows by the compactness of the energy level sets.

Choose a T such that

$$\mathbb{E} \left| \frac{1}{T} \int_0^T f(S(t)u)\, dt - \bar{f} \right| < \delta + \epsilon/6.$$

Now choose Δt_1 such that if $\Delta t \le \Delta t_1$ and $\tilde{n} = \lfloor T/\Delta t \rfloor$ then

$$\mathbb{E} \left| \frac{1}{\tilde{n}} \sum_{i=0}^{\tilde{n}-1} f(\hat{S}^i_{\Delta t} u) - \bar{f} \right| < \delta + \epsilon/3. \tag{4.20}$$

This is possible by an argument similar to that used in the proof of Theorem 4.1.

Fixing Δt, for any n let $n = \tilde{n}K + r$ where K and r are nonnegative integers and $r < \tilde{n}$. Then

$$\left| \frac{1}{n} \sum_{i=0}^{n-1} f(\hat{S}_{\Delta t}^i u) - \bar{f} \right| = \left| \frac{1}{n} \sum_{i=0}^{\tilde{n}K-1} f(\hat{S}_{\Delta t}^i u) + \frac{1}{n} \sum_{i=\tilde{n}K}^{\tilde{n}K+r-1} f(\hat{S}_{\Delta t}^i u) - \bar{f} \right|$$

$$\leq \left| \frac{1}{\tilde{n}K} \sum_{i=0}^{\tilde{n}K-1} f(\hat{S}_{\Delta t}^i u) - \bar{f} \right| + \left| \left(\frac{1}{\tilde{n}K} - \frac{1}{n} \right) \sum_{i=0}^{\tilde{n}K-1} f(\hat{S}_{\Delta t}^i u) \right|$$

$$+ \left| \frac{1}{n} \sum_{i=\tilde{n}K}^{n-1} f(\hat{S}_{\Delta t}^i u) \right|$$

Some straightforward manipulation yields that

$$\mathbb{E} \left| \frac{1}{n} \sum_{i=0}^{n-1} f(\hat{S}_{\Delta t}^i u) - \bar{f} \right| \leq \frac{1}{K} \sum_{J=0}^{K-1} \mathbb{E} \left| \frac{1}{\tilde{n}} \sum_{i=0}^{\tilde{n}-1} f(\hat{S}_{\Delta t}^{i+J\tilde{n}} u) - \bar{f} \right| + 2 f_{\max} \frac{\tilde{n}}{n}. \quad (4.21)$$

Now suppose $n\Delta t \geq \mathcal{T}_{\min}$ (which we have not yet determined). Working with the second term we obtain:

$$2 f_{\max} \frac{\tilde{n}}{n} \leq 2 f_{\max} \frac{T}{n\Delta t} \leq 2 f_{\max} \frac{T}{\mathcal{T}_{\min}}.$$

So if we let $\mathcal{T}_{\min} = 6 f_{\max} T / \epsilon$, the second term on the right of (4.21) will be less than $\epsilon/3$ whenever $n\Delta t \geq \mathcal{T}_{\min}$.

It remains to bound the first term on the right of (4.21). For each J

$$\mathbb{E} \left| \frac{1}{\tilde{n}} \sum_{i=0}^{\tilde{n}-1} f(\hat{S}_{\Delta t}^{i+J\tilde{n}} u) - \bar{f} \right| = \mathbb{E}_{\Sigma_J} \left| \frac{1}{\tilde{n}} \sum_{i=0}^{\tilde{n}-1} f(\hat{S}_{\Delta t}^i u) - \bar{f} \right|,$$

where $\Sigma_J = \hat{S}_{\Delta t}^{J\tilde{n}} \Sigma$. We need to show that the expression on the right is close to that in (4.20). Lemma 4.3 below allows us to do this. We let

$$g(u) := \left| \frac{1}{\tilde{n}} \sum_{i=0}^{\tilde{n}-1} f(\hat{S}_{\Delta t}^i u) - \bar{f} \right|$$

which is less than $f_{\max}$ for all u. Let $E_{\max}$ be the maximum value of $H(\hat{S}_{\Delta t}^n u)$ over all $u \in \Sigma$ and all $n\Delta t \leq \mathcal{T}_{\Delta t}$, and $E_{\min}$ be the minimum. Let $B = \{u \in \mathbb{R}^{2n} | H(u) \in [E_{\min}, E_{\max}]\}$. Now choose a Δt_2 such that when $\Delta t < \Delta t_2$,

29

$\mathrm{vol}(B) \leq (1 + \epsilon/6f_{\max})\mathrm{vol}(\Sigma)$. This is possible by the fifth hypothesis, since both Σ and Σ_J are subsets of B. The idea is that since B converges to Σ as Δt goes to 0, Σ_J must converge to Σ. Then Lemma 4.3 shows that

$$|\mathbb{E}_{\Sigma_J} g(u) - \mathbb{E}_\Sigma g(u)| \leq \epsilon/3. \tag{4.22}$$

Putting this together with (4.20), we have that if $\Delta t \leq \Delta t_0 := \min(\Delta t_1, \Delta t_2)$ and $n\Delta t \leq \mathcal{T}_{\Delta t}$

$$
\begin{aligned}
\frac{1}{K} \sum_{J=0}^{K-1} \mathbb{E} \left| \frac{1}{\tilde{n}} \sum_{i=0}^{\tilde{n}-1} f(\hat{S}_{\Delta t}^{i+J\tilde{n}} u) - \bar{f} \right| &\leq \frac{1}{K} \sum_{J=0}^{K-1} \left[\mathbb{E}_\Sigma \left| \frac{1}{\tilde{n}} \sum_{i=0}^{\tilde{n}-1} f(\hat{S}_{\Delta t}^{i} u) - \bar{f} \right| + \epsilon/3 \right] \\
&= \mathbb{E}_\Sigma \left| \frac{1}{\tilde{n}} \sum_{i=0}^{\tilde{n}-1} f(\hat{S}_{\Delta t}^{i} u) - \bar{f} \right| + \epsilon/3 \\
&\leq \delta + 2\epsilon/3.
\end{aligned}
$$

Combining this with (4.21) gives that if $\Delta t \leq \Delta t_0$ and $n\Delta t \in [\mathcal{T}_{\min}, \mathcal{T}_{\Delta t}]$ then

$$\mathbb{E}_\Sigma \left| \frac{1}{n} \sum_{i=0}^{n-1} f(\hat{S}_{\Delta t}^{i} u) - \bar{f} \right| \leq \delta + \epsilon,$$

as required. $\square$

Lemma 4.3 *Let B be a subset of $\mathbb{R}^{2n}$ with finite volume. Let $g : \mathbb{R}^{2n} \to \mathbb{R}$ satisfy $\sup_{u \in B} |g(u)| \leq G$. Let $A, A^* \subset B$, $\mathrm{vol}(A) = \mathrm{vol}(A^*)$, and $\mathrm{vol}(B) \leq (1 + \gamma)\mathrm{vol}(A)$. Then*
$$|\mathbb{E}_A g - \mathbb{E}_{A^*} g| \leq 2G\gamma.$$

Proof

$$
\begin{aligned}
|\mathbb{E}_A g - \mathbb{E}_{A^*} g| &= \frac{1}{\mathrm{vol}\,(A)} \left| \int_A g - \int_{A^*} g \right| \\
&\leq \frac{1}{\mathrm{vol}\,(A)} \left\{ \left| \int_A g - \int_B g \right| + \left| \int_{A^*} g - \int_B g \right| \right\} \\
&\leq \frac{1}{\mathrm{vol}\,(A)} 2G(\mathrm{vol}\,(B) - \mathrm{vol}\,(A)) \\
&\leq 2G\gamma,
\end{aligned}
$$

as required. $\square$

Remarks: The remarks after Theorem 3.3 apply here as well. The proof works by choosing a time T over which the values of f on Σ are well sampled by trajectories of the system. Then a step-size Δt is chosen such that the numerical trajectory approximates the true trajectory over this time. Thus, accurate computation of long-time averages is only guaranteed when the trajectories are accurately computed over long periods of time. This is not the regime in which most MD calculations are performed. Typically, computed and actual solutions to the equations diverge rapidly.

More quantitatively, for the theorem to work, we expect that $T = \epsilon^{-2}$ is required; see the remarks after Theorem 3.3. The error of an rth order numerical integrator over a time interval T goes like $\Delta t^r \exp(CT)$ for some $C > 0$. So the theorem requires

$$\Delta t \sim \exp[-C\epsilon^{-2}/r],$$

which results in a Δt that must decrease extremely fast with respect to ϵ.

However, the situation is more optimistic in the case of a nearly constant f, as in Theorem 3.1, and as illustrated by the functions F_2 and F_4 in Section 2. Suppose that

$$\mathbb{E}|f(u) - \bar{f}| \leq \delta.$$

In this case the time T in Theorem 4.2 can be taken to be an arbitrarily small, and Δt_1 can be taken to be arbitrarily large. The only restriction on stepsize is provided by the demands of energy conservation; that is, $\Delta t \leq \Delta t_2$ in the theorem. This is in line with computational practice, where the largest stepsize that yields stable energy is used. More quantitatively, suppose the system is being integrated over a long time interval T. The maximum energy error over this period goes like Δt^p for some p. This is required to be smaller than a constant times ϵ. So the theorem requires

$$\Delta t \sim \epsilon^{1/p}.$$

This is similar to the restriction on step length that would result from accuracy considerations.

31

References

[1] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, Oxford. (1987).

[2] G. Benettin and A. Giorgilli, *On the Hamiltonian interpolation of near to the identity symplectic mappings with application to symplectic integration algorithms*, **74** (1994) 1117–1143.

[3] G. Ciccotti, D. Frenkel, I. R. McDonald, editors, Simulation of Liquids and Solids: molecular dynamics and Monte Carlo methods in statistical mechanics, North-Holland, Amsterdam. (1987).

[4] V. J. Donnay, *Non-ergodicity of two particles interacting via a smooth potential,* J. Statist. Phys. **96** (1999), no. 5-6, 1021–1048.

[5] V. Donnay and C. Liverani, *Potentials on the two-torus for which the Hamiltonian flow is ergodic,* Comm. Math. Phys. **135** (1991), no. 2, 267–302.

[6] E. Hairer, C. Lubich, and G. Wanner, *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*, Springer, Berlin. (2002).

[7] E. Hairer, C. Lubich, and G. Wanner, *Geometric numerical integration illustrated by the Störmer/Verlet method,* Acta Numerica 2003, Cambridge Univ. Press, Cambridge, 1–51.

[8] T. J. Hunt and R. S. MacKay, *Anosov parameter values for the triple linkage and a physical system with a uniformly chaotic attractor,* Nonlinearity. **16** (2003), no. 4, 1499–1510.

[9] A. I. Khinchin, *Mathematical Foundations of Statistical Mechanics*, Dover, New York. (1949).

[10] C. Liverani, *Interacting Particles*, in D. Szász, ed., Hard Ball Systems and the Lorentz Gas, Springer (2000) 145–178.

[11] L. Markus and K. R. Meyer, *Generic Hamiltonian dynamical systems are neither integrable nor ergodic*, Memoirs of AMS, No. 114. (1974).

[12] A. Rahman, *Correlation in the Motion of Atoms in Liquid Argon*, Phys. Rev. **136** (1964) A405–A411.

[13] F. Reif, *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill, New York. (1965).

[14] S. Reich, *Backward error analysis for numerical integrators*, SIAM J. Numer. Anal. **36** (1999) 1549–1570.

[15] Ya. G. Sinai, *On the Foundation of the Ergodic Hypothesis for a Dynamical System of Statistical Mechanics*, Dokl. Akad. Nauk SSSR **153**(1963) 1261–1264.

[16] S. D. Stoddard and J. Ford, *Numerical Experiments on the Stochastic Behavior of a Lennard-Jones Gas System*, Phys. Rev A, **8** (1973) 1504–1512.

[17] H. Sigurgeirsson and A. M. Stuart, *Statistics from Computations*, in Foundations of Computational Mathematics (Oxford, 1999), 323–344, Cambridge Univ. Press, Cambridge, 2001.

[18] D. Szász, *Boltzmann's Ergodic Hypothesis, a Conjecture for Centuries?*, in D. Szász, ed., Hard Ball Systems and the Lorentz Gas, Springer (2000) 421–446.

[19] R. C. Tolman, *The principles of statistical mechanics.* Clarendon Press, Oxford. (1938).

[20] D. Turaev and V. Rom-Kedar, *Elliptic islands appearing in near-ergodic flows*, Nonlinearity, **11** (1998), 575–600.

[21] L. Verlet, *Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules*, Phys. Rev. **159** (1967) 98–103.

[22] M. Viana, *Stochastic Dynamics of Deterministic Systems*, Instituto de Matemática Pura e Applicada (IMPA), Rio de Janeiro, (1997).

[23] P. Walters, *An Introduction to Ergodic Theory*, Springer-Verlag, New York. (1982).

A   Appendix

In the following $\mathcal{H}(u, \gamma)$ is a given function by which we define $H(u) := \mathcal{H}(u, 0)$ and $H_\gamma(u) = \mathcal{H}(u, \gamma)$. Given an $E \in \mathbb{R}$, Σ is the set of all u with $H(u) = E$ and Σ_γ is the set of all u with $H_\gamma(u) = E$.

Lemma A.1 *Let $\mathcal{H}(\cdot, \cdot) : \mathbb{R}^{2n} \times \mathbb{R} \to \mathbb{R}$ and $E \in \mathbb{R}$ satisfy Assumptions 3.4. There is a $\gamma_0 > 0$ and a family of continuous bijective $P_\gamma : \mathbb{R}^{2n} \to \mathbb{R}^{2n}$, $\gamma \in [0, \gamma_0)$ defined in a neighbourhood of Σ such that*

1. *$H_\gamma(P_\gamma u) = H(u)$ wherever P_γ is defined.*

2. *P_γ converges to the identity uniformly on Σ as $\gamma \to 0$*

3. *$\det \nabla P_\gamma$ converges to 1 uniformly on Σ as $\gamma \to 0$*

Thus Σ and Σ_γ, as defined above, are isomorphic for $\gamma \in [0, \gamma_0)$.

Proof   We will construct the maps P_γ from the solution to an ordinary differential equation. Let

$$h(v, \gamma) = -\|\nabla_v \mathcal{H}(v, \gamma)\|^{-2} \frac{\partial \mathcal{H}}{\partial \gamma}(v, \gamma) \nabla_v \mathcal{H}(v, \gamma)$$

which is Lipschitz continuous for u in a neighbourhood of Σ (that is, for γ sufficiently small). Let $P_\gamma(u)$ be the solution $v(\gamma)$ to the initial value problem

$$\frac{dv}{d\gamma} = h(v, \gamma), \quad v(0) = u.$$

The flow of this differential equation is defined in a neighbourhood of Σ and for sufficiently small $\gamma > 0$. Observe that

$$\frac{d}{d\gamma} \mathcal{H}(v(\gamma), \gamma) = \nabla_v \mathcal{H}(v(\gamma), \gamma)^T \frac{dv}{d\gamma} + \frac{\partial H}{\partial \gamma}(v(\gamma), \gamma)$$

which is zero due to our choice of $h(v, \gamma)$. So

$$\mathcal{H}(P_\gamma u, \gamma) = \mathcal{H}(u, 0)$$

wherever P_γ is defined. In particular

$$H_\gamma(P_\gamma u) = H(u) = E$$

for all $u \in \Sigma$. Moreover, by standard theorems $P_\gamma u$ depends continuously on u. This establishes items 1 and 2.

The matrix-valued function ∇P_γ satisfies the variational equation

$$\frac{d}{d\gamma} \nabla P_\gamma u = \nabla_u h(P_\gamma u, \gamma), \quad \nabla P_0 u = I.$$

Some checking shows that this right-hand-side is Lipschitz continuous under the stated assumptions. Thus, uniformly in u, $\nabla P_\gamma u$ converges to I as $\gamma \to 0$. This establishes item 3. $\quad\square$

Lemma A.2 *Let g be a continuous function $\mathbb{R}^{2n} \to \mathbb{R}$. Let $H, H_\gamma, \Sigma, \Sigma_\gamma, P_\gamma$ be as defined in the previous theorem. Let $\mathbb{E}$ ($\mathbb{E}_\gamma$) denote taking an average with respect to the Liouville measure for H (H_γ) on Σ (Σ_γ). Then*

$$\lim_{\gamma \downarrow 0} |\mathbb{E}g(u) - \mathbb{E}_\gamma g(u)| = 0.$$

Proof Let $v = P_\gamma u$, $dv = |\det \nabla P_\gamma u| du$. Then

$$
\begin{aligned}
\mathbb{E}_\gamma g(v) &= \lim_{\delta E \downarrow 0} \frac{\int_{H_\gamma(v) \in [E, E+\delta E]} g(v) dv}{\int_{H_\gamma(v) \in [E, E+\delta E]} dv} \\
&= \lim_{\delta E \downarrow 0} \frac{\int_{H(u) \in [E, E+\delta E]} g(P_\gamma u) |\det \nabla P_\gamma u| du}{\int_{H(u) \in [E, E+\delta E]} |\det \nabla P_\gamma u| du} \\
&= \lim_{\delta E \downarrow 0} \frac{\int_{H(u) \in [E, E+\delta E]} g(P_\gamma u) |\det \nabla P_\gamma u| du}{\int_{H_\gamma(v) \in [E, E+\delta E]} du} \\
&\quad \times \lim_{\delta E \downarrow 0} \frac{\int_{H_\gamma(v) \in [E, E+\delta E]} du}{\int_{H_\gamma(v) \in [E, E+\delta E]} |\det \nabla P_\gamma u| du} \\
&= \frac{\mathbb{E}g(P_\gamma u) |\det \nabla P_\gamma u|}{\mathbb{E}|\det \nabla P_\gamma u|}.
\end{aligned}
$$

Since $|\det \nabla P_\gamma|$ converges uniformly to 1 on Σ and $P_\gamma u$ converges uniformly to u, this last quantity converges to $\mathbb{E}g(u)$ as $\gamma \downarrow 0$, as required. $\quad\square$