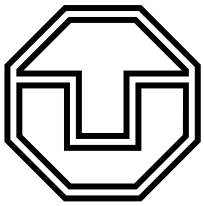

TECHNISCHE UNIVERSITÄT DRESDEN

Importance Sampling of Rare Events in Chaotic Systems

Dissertation
zur Erlangung des wissenschaftlichen Grades
Doctor rerum naturalium

vorgelegt von

Jorge C. Leitão
geboren am 14.02.1989 in Estoi (Portugal)

Institut für Theoretische Physik
Fachrichtung Physik
Fakultät für Mathematik und Naturwissenschaften
Technische Universität Dresden
2016

Eingereicht am 31. Mai 2016

1. Gutachter: Prof. Dr. Holger Kantz
2. Gutachter: Prof. Dr. Tamás Tél
3. Gutachter: Prof. Dr. Roland Ketzmerick

Abstract

Rare events play a crucial role in our society and a great effort has been dedicated to numerically study them in different contexts. This thesis proposes a numerical methodology based on Monte Carlo Metropolis-Hastings algorithm to efficiently sample rare events in chaotic systems. It starts by reviewing the relevance of rare events in chaotic systems, focusing in two types of rare events: states in closed systems with rare chaoticities, characterised by a finite-time Lyapunov exponent on a tail of its distribution, and states in transiently chaotic systems, characterised by a escape time on the tail of its distribution. This thesis argues that these two problems can be interpreted as a traditional problem of statistical physics: sampling exponentially rare states in the phase-space – states in the tail of the density of states – with an increasing parameter – the system size. This is used as the starting point to review Metropolis-Hastings algorithm, a traditional and flexible methodology of importance sampling in statistical physics. By an analytical argument, it is shown that the chaoticity of the system hinders direct application of Metropolis-Hastings techniques to efficiently sample these states because the *acceptance is low*. It is argued that a crucial step to overcome low acceptance rate is to construct a proposal distribution that *uses information about the system* to bound the acceptance rate. Using generic properties of chaotic systems, such as exponential divergence of initial conditions and fractals embedded in their phase-spaces, a proposal distribution that guarantees a bounded acceptance rate is derived for each type of rare events. This proposal is numerically tested in simple chaotic systems, and the efficiency of the resulting algorithm is measured in numerous examples in both types of rare events. The results confirm the dramatic improvement of using Monte Carlo importance sampling with the derived proposals against traditional methodologies: the number of samples required to sample an exponentially rare state increases polynomially, as opposed to an exponential increase observed in uniform sampling. This thesis then analyses the sub-optimal (polynomial) efficiency of this algorithm in a simple system and shows analytically how the correlations induced by the proposal distribution can be detrimental to the efficiency of the algorithm. This thesis also analyses the effect of high-dimensional chaos in the proposal distribution and concludes that an anisotropic proposal that takes advantage of the different rates of expansion along the different unstable directions, is able to efficiently *find rare states*. The applicability of this methodology is also discussed to sample rare states in *non-hyperbolic* systems, with focus on three systems: the logistic map, the Pomeau-Manneville map, and the standard map. Here, it is argued that the different origins of non-hyperbolicity require different proposal distributions. Overall, the results show that by incorporating specific information about the system in the proposal distribution of Metropolis-Hastings algorithm, it is possible to efficiently find and sample rare events of chaotic systems. This improved methodology should be useful to a large class of problems where the numerical characterisation of rare events is important.

To my mother Ana Lopes and my brother Pedro Leitão

Contents

1. Introduction

Extreme events play a crucial role in our society. Landslides, floods, meteorite collisions, solar flares, earthquakes are all events that are rare but often lead to catastrophic consequences to our well being. Science often studies extreme events by recreating the process that generates them sufficiently many times. One way of recreating the process is by preparing it in an experimental setup. For example, the experimental observation of the Higgs boson was made possible by repeating the collision of high-energetic particles in a particle accelerator. Because most of the events in the collision were not relevant to the observation of the Higgs boson, it required millions of repetitions until sufficient events were gathered.

While some events, such as the creation of the Higgs boson close to a detector, can be recreated by technology, ingenuity, and enough public investment, others, such as earthquakes or floods, can not be easily achieved in a controlled setup. This poses a limitation on the type of events that can be recreated in an experimental setup, which hinders our ability to study and understand them. In these situations, we instead resolve to physical models – a simplifying description of the process that reproduces such events.

An ubiquitous feature of extreme events is that often they can be modelled only as an emergent behaviour of a *non-linear* dynamical process. For example, an earthquake happens due to the interaction of different effects that, when combined, store elastic energy. This elastic energy is then suddenly released in the form of seismic waves. The process of storage and subsequence explosive release of energy can be modelled by a non-linear process with memory [1]. Other examples include the modelling of the explosive volcano eruptions [2] and atypical water level of Nile River [1].

A fascinating consequence of the non-linearity of dynamical systems is that it often leads to an extremely sensitive dependence on initial conditions, i.e. chaos. Chaos has a long and fascinating history dating back to the seminal work of Poincaré in 1890 [3] and is today a well established field of research with applications in Biology, Geology, Economy, Chemistry, and Physics [4, 5]. When a system is chaotic, its evolution in time cannot be described by a closed formula. This is a major limitation to the study of dynamical systems, and is a major reason why chaos is often studied numerically, a paradigm that was popularised by Lorenz in the 60s and 70s [4]. The chaotic nature of some non-linear processes is fascinating, but it also hinders our ability to study extreme events on them, as it forbids describing extreme events in a closed formula. In this situation, the best we can do is to numerically simulate the model in a computer and search extremes events, just like in a particle accelerator. The high sensitivity to initial conditions makes the study of individual trajectories difficult as it requires extremely high accuracy about the initial state of the system. For this reason, chaotic systems are typically characterised by how trajectories behave *on average*. For example, one

of the most important quantities in the characterisation of chaotic systems is the Lyapunov exponent of the system. The Lyapunov exponent quantifies the time-scale that two nearby trajectories remain, on average, close to each other in time, and is defined as a time-average over a trajectory, or as an average over an ensemble of trajectories. The calculation of an individual trajectory requires it to be numerically evolved in time, and therefore an average over an ensemble of trajectories requires each one of them to be evolved individually, which makes the study of chaotic systems computationally demanding.

Another important aspect of the study of extreme events in non-linear processes is that some mechanisms responsible for their appearance can be described and understood in simple chaotic systems. For example, bifurcations, where a small change in one parameter leads to a dramatic change in the system, is a phenomena that can explain extreme events [1], and bifurcations are already present in simple chaotic systems. The importance of extreme events to human society, the chaotic nature of non-linear processes, the understanding that simple chaotic systems already possess basic mechanisms responsible for extreme events, and the computational challenges of studying rare events in chaotic systems numerically together constitute the motivation for this thesis.

This thesis is about performing numerical simulations in chaotic systems in such a way that rare events are generated more likely than if they would be generated by chance. It is focused on two classes of problems where rare events in chaotic systems are important:

- finite-time trajectories with high or low Lyapunov exponents;
- long living trajectories in open systems.

Finding and sampling rare states has been approached in the literature of chaotic systems before, with promising results [6–11]. For example, the method *stagger and dagger*, used to find long-living trajectories of transiently chaotic systems, has been an important tool to characterise chaotic saddles [6, 8], and the *Lyapunov weighted dynamics* has been used successfully to find atypically low or high chaotic trajectories [9, 12]. A well known methodology within "trajectory sampling" is transition path sampling [13], that has been used to sample rare trajectories (e.g. molecules movement), typically influenced by thermal noise [13]. These are typically trajectories that transit from one stable configuration to another stable configuration [13, 14]. The different algorithms in the literature achieve their goal through different, often ingenious, solutions. This is often helpful to solve a particular problem, but it may not necessarily be helpful to approach new problems.

A well established numerical technique that has been used in statistical physics to study rare events since the 50s is Metropolis-Hastings (MH) importance sampling [15, 16]. Essentially, MH is a random walk $\boldsymbol{x} \to \boldsymbol{x}'$ in the phase-space of the system that generates rare states more often than they would be found by chance, consequently reducing the computational cost associated with their rareness. The flexibility of MH is confirmed by its success in numerous fields in Physics, Chemistry, Finance, amongst many others [15, 16]. While transition path sampling or Lyapunov weighted dynamics already uses some kind of importance sampling technique, three questions remains largely open: 1. can Metropolis-Hastings be used to systematically sample rare trajectories of (deterministic) chaotic

systems? If yes, 2. how and 3. at what (computational) cost?

To answer these questions, this thesis develops a systematic approach to sample rare events in chaotic systems and shows how Metropolis-Hastings can be used to efficiently sample them in different situations. The crucial concept of this thesis is the notion of proposal distribution of the Metropolis-Hastings algorithm. Essentially, the proposal distribution is the conditional probability of "trying" a state $\boldsymbol{x}'$ in the phase-space of the system, given another state $\boldsymbol{x}$. The crucial question for MH to work can be posed as follows: what proposal distribution guarantees that a property of the trajectory starting at $\boldsymbol{x}'$, e.g. its Lyapunov exponent, is similar to the same property of the trajectory starting from $\boldsymbol{x}$? The main contribution of this thesis is a methodology to answer this question for a broad class of chaotic systems and properties is a methodology (Chapter 4) to incorporate properties of trajectories of chaotic systems (Chapter 2) in this conditional probability such that it leads to an efficient MH algorithm (Chapter 3). This methodology allows to construct efficient MH algorithms to sample rare events in different problems and classes of chaotic systems (Chapter 5-7). We expect the ideas and formalism presented here to find applications in other problems of chaotic systems and in the study of extreme events more generally. Therefore, we expect this thesis to be useful to both people studying extreme events in non-linear systems and to people using numerical techniques.

This thesis is organised in 8 chapters as follows:

- Chapter 2 introduces chaotic systems and shows that there is a set of relevant numerical problems in studying rare events that are suitable to the application of Metropolis-Hastings;
- Chapter 3 introduces the Metropolis-Hastings (MH) algorithm as a flexible and well established approach to these problems. Chapter 3 also argues that, in order to be efficient, the Metropolis-Hastings algorithm requires a proposal distribution that uses information about the system;
- Chapter 4 shows how to incorporate general features of chaotic systems, such as exponential divergence or self-similarity of some of its properties, in the proposal distribution;
- Chapters 5 and 6 illustrate the efficiency of Monte Carlo algorithms to sample rare events in different classes of chaotic systems;
- Chapters 7 discusses the effect of non-hyperbolicity of the system in the construction of an efficient proposal distribution;
- Chapter 8 summarises the results and discusses its implications.

2. Chaotic Systems

2.1. Time evolution

A dynamical system is defined by a state $\boldsymbol{x}$ in a phase-space $\boldsymbol{x} \in \Omega \subset \mathbb{R}^D$ and by a specific rule that governs its time evolution. Most systems in Physics are time continuous and their evolution is described by a differential equation. However, when such differential equation has no analytical solution, the system is studied using numerical solutions. Numerical solutions are obtained in computers by means of a time-discretisation of the differential equation and therefore the time-evolution system is mapped to a discrete-time evolution. This, and the study of continuous dynamics using Poincaré surface of sections [5], motivates the formulation of the discrete-time evolution of dynamical systems: starting at a state $\boldsymbol{x}_0 \equiv \boldsymbol{x}$, the system evolves according to

$$\boldsymbol{x}_{t+1} = \boldsymbol{F}(\boldsymbol{x}_t) \tag{2.1}$$

where $\boldsymbol{F}(\boldsymbol{x}) \in \Omega$. In other terms, $\boldsymbol{x}_t = \boldsymbol{F}^t(\boldsymbol{x}_0)$ where $\boldsymbol{F}^t$ is $\boldsymbol{F}$ composed t times, $\boldsymbol{F}^t(\boldsymbol{x}) = \boldsymbol{F}(\boldsymbol{F}(...\boldsymbol{F}(\boldsymbol{x})...))$.

2.2. Dynamical stability

A basic property of a dynamical system is dynamical stability: what happens in time when a trajectory is perturbed by a small amount $\boldsymbol{h}$? The distance of a state displaced from $\boldsymbol{x}$ by $\boldsymbol{h}$, $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{h}$, to the original state $\boldsymbol{x}$ evolves in time according to

$$\boldsymbol{x}'_t - \boldsymbol{x}_t = \boldsymbol{F}^t(\boldsymbol{x}') - \boldsymbol{F}^t(\boldsymbol{x}) \tag{2.2}$$

Expanding $\boldsymbol{F}^t(\boldsymbol{x}')$ around $\boldsymbol{F}^t(\boldsymbol{x})$ allows $\boldsymbol{x}'_t - \boldsymbol{x}_t$ to be written as

$$\boldsymbol{x}'_t - \boldsymbol{x}_t = J_t(\boldsymbol{x}) \cdot \boldsymbol{h} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 \boldsymbol{F}^t(\boldsymbol{x})}{\partial x_i \partial x_j} h_i h_j + O(|\boldsymbol{h}|^3) \tag{2.3}$$

where $J_t(\boldsymbol{x}) \equiv d\boldsymbol{F}^t(\boldsymbol{x})/d\boldsymbol{x}$ is the Jacobian matrix of $\boldsymbol{F}^t$, and $\partial^2 \boldsymbol{F}^t(\boldsymbol{x})/(\partial x_i \partial x_j)$ is the (i, j) entry of the Hessian matrix of $\boldsymbol{F}$. The first term of Eq. 2.3 can be expanded using the derivative of the

composition and be written as

$$D(\boldsymbol{x}, \boldsymbol{h}, t) \equiv J_t(\boldsymbol{x}) \cdot \boldsymbol{h} = \left(\prod_{i=t-1}^{0} J\left(\boldsymbol{x}_i\right) \right) \cdot \boldsymbol{h} = \boldsymbol{h}_t \tag{2.4}$$

where $J \equiv J_1$ and $\boldsymbol{h}_t$ evolves in the tangent space according to

$$\boldsymbol{h}_0 = \boldsymbol{h} \;\; ; \boldsymbol{h}_{i+1} = J\left(\boldsymbol{x}_i\right) \cdot \boldsymbol{h}_i \;\; . \tag{2.5}$$

The stability of $\boldsymbol{x}'_t - \boldsymbol{x}_t$ is characterised, under a linear approximation of Eq. 2.3, (that is, for small $|\boldsymbol{h}|$), by the stability of $D(\boldsymbol{x}, \boldsymbol{h}, t)$, which depends on the initial direction of $\boldsymbol{h}$. Different directions may lead to different stability, and this can be characterised by the eigenvalues and eigenvectors of J^t. The eigenvectors associated with a eigenvalue whose real part is smaller than 1 correspond to stable directions, as $\boldsymbol{x}'_t - \boldsymbol{x}_t$ along that direction converges, and those associated with eigenvalue larger than 1 diverge. The largest finite-time Lyapunov exponent (FTLE) of a point $\boldsymbol{x}$ can be defined[1] by

$$\lambda_t(\boldsymbol{x}) = \frac{\log(\mu_1)}{t} \;\; , \tag{2.6}$$

where μ_1 is the real part of the largest eigenvalue of J^t. Thus, when at least one direction is unstable, $\boldsymbol{x}'_t - \boldsymbol{x}_t$ increases exponentially with time, and at most by

$$\boldsymbol{x}'_t - \boldsymbol{x}_t = \delta_0 e^{\lambda_t(\boldsymbol{x})t} \;\; . \tag{2.7}$$

When the system is one dimensional, the "Jacobian matrix" is a single number, the product in Eq. 2.4 is a product of numbers, and the only "eigenvalue" is the result of this product. Thus, in this case Eq. 2.6 can be written as

$$\lambda_t(\boldsymbol{x}) = \frac{1}{t} \sum_{i=0}^{t-1} \log \left| \frac{dF(\boldsymbol{x}_i)}{dx} \right| \;\; . \tag{2.8}$$

In higher dimensions, this decomposition is not possible, but it is still useful because it indicates how the FTLE can be interpreted as an average of individual contributions of each step.

A positive $\lambda_t(\boldsymbol{x})$ implies that an infinitesimal perturbation is exponentially amplified, which is the landmark of chaotic behaviour. Still, the existence of a point with a positive FTLE is not a condition for the system to be considered chaotic: an unstable point of a pendulum has exponential sensitivity to a small perturbation, even though the pendulum is not a chaotic system.

Calculation of the FTLE One way to compute the finite-time Lyapunov exponent is to compute the Jacobian matrix J_t directly from the definition, by multiplying the Jacobians of each time step,

[1] There is subtle difference between this definition and the one from the Cauchy-Green strain tensor. This is sometimes denoted as the stability exponent instead of Lyapunov exponent. See Chapter 4 and 6 of Ref. [17] for a discussion on this distinction.

$J(\boldsymbol{x}_i)$, and diagonalize it. [17] In some situations, the computer may not have precision to do so as the entries of the Jacobian matrix grow exponentially with t when the system is chaotic. In this situation, one alternative is to consider v random unitary vectors $\boldsymbol{h}_i$, and, on each time step along the evolution of the trajectory $\boldsymbol{x}_i$, multiply $\boldsymbol{h}_i$ by $J(\boldsymbol{x}_i)$ and renormalise them (normalization $v_i^{(t)}$), while storing the individual sums $V_i = \sum_t \log(v_i^{(t)})$. As $v \to \infty$, the largest V_i/t approaches λ_t.

2.3. Natural measure

A central aspect in dynamical systems is the description of the dynamics in terms of the evolution of a set of trajectories. Informally, this is motivated by the following question: how will a set of initial conditions be distributed as a function of time? This question dates back to the ergodic hypothesis, which has a central role in statistical physics and can be framed by considering an initial set of states distributed according to $P_0(\boldsymbol{x})$ and questioning how will this density be at a later time t, and, in particular, at $t \to \infty$. Formally, the evolution of densities is described by the Frobenius-Perron operator, PF$[P]$, [5]

$$P_{t+1}(\boldsymbol{x}) = \mathrm{PF}[P_t] \equiv \int_\Omega \delta(\boldsymbol{x} - \boldsymbol{F}(\boldsymbol{x}'))P_t(\boldsymbol{x}')d\boldsymbol{x}' \quad . \tag{2.9}$$

that maps a density at time t to a density at time $t+1$. It often happens that a density cannot be properly defined. [5] One example is when there is the need to represent the density of a single state $\{\boldsymbol{x}\} \subset \Omega$, which can only be defined in terms of the measure dirac-δ, that assigns 1 to the set $\{\boldsymbol{x}\}$ and 0 elsewhere. Due to that, it is convenient to use the measure associated with the density when the density can be properly defined, and use a measure when a density cannot be defined [5].

An important notion in dynamical systems is the *invariant density* P^*, that fulfils PF$[P^*] = P^*$. Associated with it is the invariant measure μ^*, that fulfils $\mu^*(S) = \mu^*(\boldsymbol{F}^{-1}(S))$ for every subset $S \subset \Omega$ [2]. Because this measure is invariant, any average performed over it is also invariant.

The relevant[3] invariant measure here is the natural invariant measure, $\mu(\boldsymbol{x})$, which corresponds to the probability that a typical (randomly chosen) long living trajectory is on a given infinitesimal volume of the phase-space centered at $\boldsymbol{x}$. The importance of this measure is that, when it is ergodic [4], the ergodic theorem guarantees that an average over an infinitely long trajectory, such as the Lyapunov exponent in Eq. 2.8, converges to an average over the natural invariant density $\mu(\boldsymbol{x})$:

$$\lim_{t\to\infty} \frac{1}{t} \sum_{i=0}^{t} A(\boldsymbol{x}_i) = \int_\Omega A(\boldsymbol{x}) d\mu(\boldsymbol{x}) \quad . \tag{2.10}$$

Intuitively, this is because successive applications of the Frobenius-Perron operator maintain the

[2] This is equivalent to $\mu^*(S) = \mu^*(\boldsymbol{F}(S))$ when the map is invertible.

[3] There is not a single invariant measure, for example, the measure $\frac{1}{P}\sum_p^P \delta(\boldsymbol{x} - \boldsymbol{x}_p)$ where $\boldsymbol{x}_p$ is the iteration of a periodic orbit of period P is invariant.

[4] A measure μ is ergodic in respect to $\boldsymbol{F}$ when, for every subset G of its support such that $\boldsymbol{F}^{-1}(G) = G$, either $\mu(G) = 0$ or $\mu(G) = 1$.

invariant density unchanged, and therefore the states $\boldsymbol{x}_i$ are as if they were drawn according to the measure μ.

2.4. Chaos

The two definitions above, the finite-time Lyapunov exponent and natural invariant measure, are used here to define a chaotic system: it is a system which contains at least one ergodic natural invariant measure in the phase-space, whose average maximal Lyapunov exponent on that measure,

$$\lambda_L \equiv \lim_{t\to\infty} \lambda_t(\boldsymbol{x}) \tag{2.11}$$

is positive. Informally, chaos is not a property of a single trajectory, but of an ensemble of trajectories. The ergodicity of the measure is important here because it implies that the positive Lyapunov exponent is not a property of a particular state, but of an ergodic set of states.

2.4.1. Strong chaos

It is often useful to distinguish different types of chaos. The strongest notion of chaos is hyperbolic chaos, which requires that the Jacobian of the map at every point in the phase-space is decomposable in a set of eigenvalues only either larger than one (unstable directions) or smaller than one (stable directions), and that the number of unstable and stable directions is the same for all $\boldsymbol{x}$. A paradigmatic example of an hyperbolic system is the tent map, whose state $\boldsymbol{x} = x \in \Omega = [0, 1]$ evolves according to

$$F(x) = \begin{cases} ax & \text{for } x \leq 1/a \\ b(1-x) & \text{for } x > 1/a \end{cases} \tag{2.12}$$

with $a > 1$ and $b \equiv a/(a-1)$. This map is hyperbolic because the "Jacobian matrix" (a number in 1D) is larger than 1 for every x. This map is often used as a test system because most of its properties can be computed analytically. For example, its natural Markov partition [5] can be computed, which is represented in Fig. 2.1. This Markov partition induces a symbolic sequence for each trajectory. A symbolic sequence is the sequence of numbers (symbols) for which the consecutive iterations of $\boldsymbol{x}$ can be represented by, see Fig. 2.1. One example of its usefulness is the calculation of the FTLE. When the state of the trajectory is at the partition "0", $|dF/dx| = a$; in partition "1" $|dF/dx| = b$. Therefore, from Eq. 2.8, $\lambda_t(\boldsymbol{x})$ corresponds to the number of times $i(x)$ the state was in partition "0" times a i.e. $x_i \in [0, 1/a]$, plus the number $t - i(x)$ of times it was in partition "1" times b,

$$t\lambda_t(x) = t\lambda_t(i(x)) \equiv i(x)\log a + (t - i(x))\log b \quad . \tag{2.13}$$

It is also possible to compute analytically the distribution of the FTLE over the natural invariant measure of the system. The natural invariant measure of this system is the constant measure [5], and

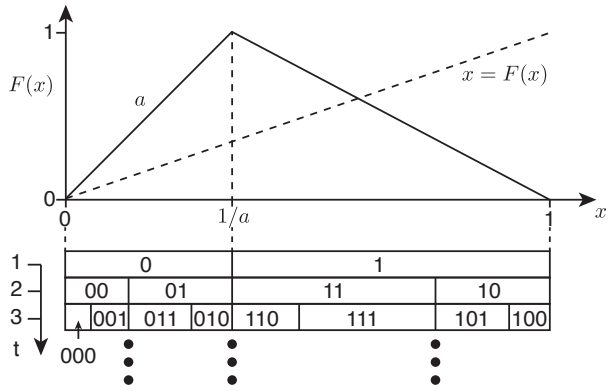

Figure 2.1.: The two scale tent map, Eq. 2.12, its respective Markov partition, and the symbolic sequences of the trajectories. A trajectory with symbolic sequence "010" is a trajectory that starts at partition "0" (left side), its first iteration is at partition "1" and its second iteration is at partition "0". The symbolic sequence of the trajectory can be used to characterise it. For example, the FTLE of the trajectory is related to the number of "0" and "1" of the trajectory.

therefore the average number of times the partition "0" is visited is proportional to its length, $1/a$. Thus, the distribution of $E = t\lambda_t$ is given by

$$P(E) \equiv \int \delta(E - t\lambda_t(\boldsymbol{x}))P(\boldsymbol{x})d\boldsymbol{x} = \sum_{i=0}^{t-1} \delta\left(E - \lambda_t(i)t\right) \binom{t}{i} \frac{1}{a^i} \left(1 - \frac{1}{a}\right)^{t-i} \quad , \tag{2.14}$$

where i is the number of "0"s in the symbolic sequence.

There are very general mathematical results that are applicable to hyperbolic systems [5]. In particular, the dynamics of an hyperbolic system can always be described in terms of a symbolic sequence of a Markov partition of its phase-space, they are structurally stable, and they fulfil the conditions of the shadowing theorem.

2.4.2. Weak chaos

Most chaotic systems in physics are not hyperbolic, and often not ergodic [5]. The paradigmatic example of a non-hyperbolic system is the kicked rotor: consider a massless rod with a mass attached to its end that can frictionless rotate around a pivotal point on the other end. Consider now that, at a fixed period, the mass is kicked with a momentaneous force perpendicular to the rod with amplitude K. This system is fully described by the momentum p of the mass and its angle θ in respect to an arbitrary axis. The evolution of this system is described in terms of a continuous-time differential equation, and, since between kicks the equations of motion are integrable, it can also be described by the standard map [5], uniquely defined by the momentum and angle of the mass immediately after each kick, $\boldsymbol{x} = (p, \theta) \in \Omega = [0, 1] \times [0, 1]$, that evolves in time according to

$$\boldsymbol{F}(p, \theta) = \begin{cases} p + K/(2\pi)\sin(2\pi\theta) \mod 1 \\ \theta + p + K/(2\pi)\sin(2\pi\theta) \mod 1 \end{cases} . \tag{2.15}$$

This system has the main properties of a generic low-dimensional Hamiltonian system: it is non-hyperbolic because the Jacobian of the system has eigenvalue 1 for $\theta = 1/4 + \mathbb{N}$. Fig. 2.2 represents the phase-space of this system. This map is a paradigmatic example of the KAM scenario of mixed phase-space Hamiltonian systems, on which a component of the phase-space, the chaotic sea, is believed to be ergodic, but there are other non-zero-measure components, the KAM islands, that are integrable and non-ergodic [5]. The kicked rotor above is an example, but the KAM scenario appears very generally when a non-linear integrable Hamiltonian system is perturbed, and often has dramatic consequences to the system, as what was before an integrable system, becomes chaotic.

Another important class of non-hyperbolic systems is found in chaotic billiards. A chaotic billiard is a dynamical system whose Hamiltonian corresponds to an infinite potential outside a boundary and a zero potential inside. [5] In other words, a particle moves freely inside the billiard and reflects specularly in its boundary. These systems are essential elements in the theoretical description of experiments in atomic, acoustic, microwave, and optical cavities [18–21]. A paradigmatic example of such system is the stadium billiard, represented in Fig. 2.3. Even though this system is ergodic, it contains a set of points where the Jacobian of the system has eigenvalue 1. One example of such points are bouncing balls: trajectories that bounce perpendicular to the two parallel walls of the billiard, that are marginally unstable (the eigenvalues of the Jacobian matrix are equal to 1). [21] Even though the set of such points is 0-measure, they are fundamental to explain why the distribution of recurrence times does not decay exponentially in time [22].

2.4.3. Shadowing and numerical studies

As introduced in Sec. 2.2, a positive Lyapunov exponent implies that any two arbitrary close trajectories diverge exponentially in time. Thus, the description of the trajectory starting at $\boldsymbol{x}$ requires a very precise knowledge of $\boldsymbol{x}$. This makes the characterisation of a single trajectory un-useful because the working precision of any experimental setup will most likely be smaller than the precision required to represent the exact trajectory under study. For this reason, chaotic systems are often characterised by statistical properties that are averages over a single trajectory or over an ensemble of trajectories [5] and particular examples of such averages are discussed in the next section.

If the system is chaotic, how can a finite-precision numerical solution of the system describe a true trajectory? Specifically, numerical solutions are always bounded to a numerical precision ε_c. Therefore the true value of $\boldsymbol{x}$ can only be represented in a computer up to ε_c, i.e. $\boldsymbol{x}_c = \boldsymbol{x} \pm \varepsilon_c \neq \boldsymbol{x}$. More dramatically, $\boldsymbol{F}(\boldsymbol{x}_c)$ is only computable up to ε_c, and thus its representation in a computer is $\boldsymbol{F}_c(\boldsymbol{x}_c) = \boldsymbol{F}(\boldsymbol{x}_c \pm \varepsilon_c) \pm \varepsilon_c \neq \boldsymbol{F}(\boldsymbol{x}_c) \neq \boldsymbol{F}(\boldsymbol{x})$. How can a numerical solution, or an ensemble of numerical solutions, still represent anything meaningful about the system? [23]

A crucial argument in the numerical study of chaotic systems is the shadowing lema [5]. It states that when the system is hyperbolic, there is a true trajectory $\boldsymbol{x}'$ close to $\boldsymbol{x}$ such that the numerical solution $\boldsymbol{F}_c^t(\boldsymbol{x}_c)$ is close to $\boldsymbol{F}^t(\boldsymbol{x}')$. That is, even though $\boldsymbol{F}_c^t(\boldsymbol{x}_c)$ does not represent $\boldsymbol{F}^t(\boldsymbol{x})$, it is still a good representation of a(nother) true trajectory of the system.

The conditions of shadowing lemma do not hold in non-hyperbolic chaotic systems and, in particu-

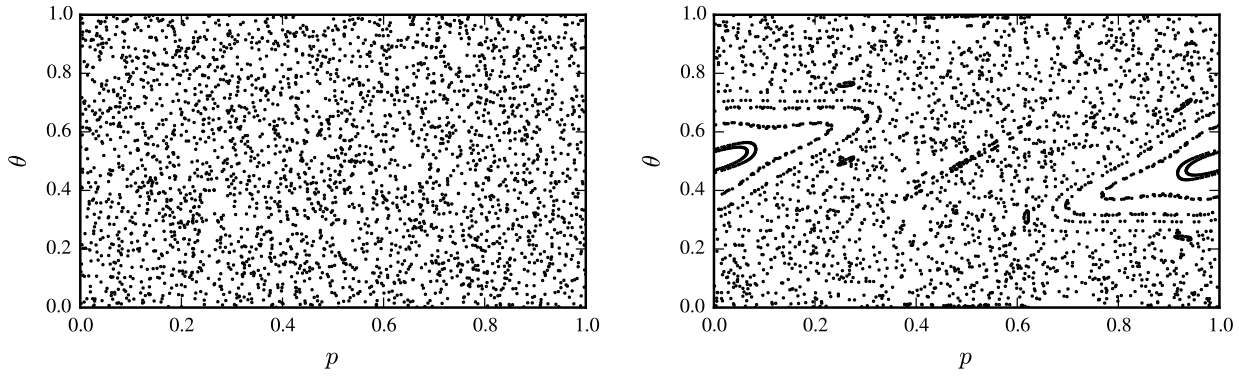

Figure 2.2.: The phase-space of the standard map with $K = 6$ (left) and $K = 2.1$ (right). Each point represents a state, obtained by starting 30 random initial conditions each evolved each of them for 100 steps. For $K = 6$, an individual trajectory wanders deterministic in the full phase-space in a seemly randomly fashion, and a single infinite trajectory is believed to come arbitrarily close to any other point of the phase-space. For $K = 2.1$, some regions of the phase-space (e.g. the island around $(p, \theta) = (0, 1/2)$) are not visitable by a trajectory outside that region, a clear sign that the measure is not ergodic.

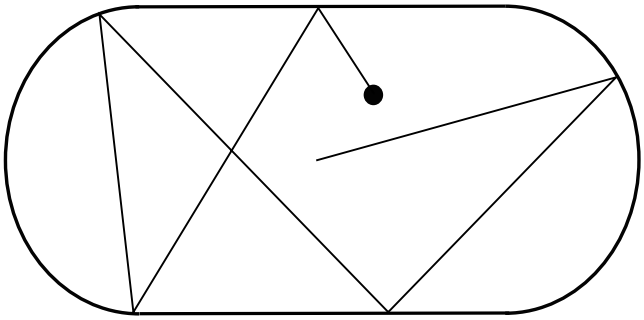

Figure 2.3.: The Bunimovich stadium is as an example of a chaotic billiard with only concave boundaries but that still is chaotic. Trajectories that bounce up and down between the parallel walls are 0-measure and the largest eigenvalue of the corresponding Jacobian is 1. Therefore a small perturbation to them leads to a polynomial divergence, even though a typical trajectory has on average exponential divergence.

lar, it was shown [24] that shadowing breaks in non-hyperbolic systems, i.e. $\boldsymbol{F}_c^t(\boldsymbol{x}_c)$ may not represent any true trajectory $\boldsymbol{F}^t(\boldsymbol{x}')$. Surprisingly, even though individual numerical trajectories do not represent any true trajectory, there is a strong numerical evidence that statistical averages over numerical trajectories are still unbiased, in the sense that they are independent of the precision ε_c used. Shadowing and shadowing breaking, as well as an important number of physical phenomena in chaotic systems, can be understood in terms of the variability of the chaoticity of different trajectories. [24]

2.5. Variability of trajectories' chaoticity

Chaotic systems are characterised by an average exponential divergence of trajectories in time, quantified by the Lyapunov exponent of the system λ_L. For a given finite time t_o, an individual state has a divergence quantified by the FTLE, Eq. 2.6 [25, 26]. It is possible to quantify how different trajectories

are differently chaotic by the distribution of the FTLE, $P(\lambda_{t_o})$, or, equivalently, the distribution of $E \equiv t_o \lambda_{t_o}$:

$$P(E) = \int \delta(E - t_o \lambda_{t_o}(\boldsymbol{x}))P(\boldsymbol{x})d\boldsymbol{x} \quad . \tag{2.16}$$

The FTLE and its distribution was introduced in the 80s [27–29] and has been used to study turbulent flows [30], Hamiltonian dynamics [31], chimera states [32], characterize dynamical trapping [27, 33–36], amongst others [25, 26]. The distribution of FTLE is related to the generalised dimensions [37] and often follows a large deviation principle, where t_o is the extensive parameter [11, 37]. In strongly chaotic systems, the distribution of FTLE is Gaussian [37], whereas for intermittent chaos and other weakly chaotic systems the distribution is typically non-Gaussian [38].

Figure 2.4 represents the typical phase-space dependency of the FTLE and respective distribution: it is a quantity that strongly depends on the state $\boldsymbol{x}$, and the respective distribution, averaged over all $\boldsymbol{x}$ for a finite t_o, is peaked and decays exponentially to zero on both sides for a typical strongly chaotic system. In the limit $t_o \to \infty$, the distribution of FTLE converges to a dirac-delta distribution at the Lyapunov exponent of the system, $P(\lambda_{t_o}) \to \delta(\lambda_{t_o} - \lambda_L)$.

The extremes of the distribution play a significant role in the characterisation of chaotic systems, as, for example, the higher moments of the distribution are related to higher qs in the generalised dimensions D_q of the attractor [37]. The regions of the phase-space with the finite-time Lyapunov exponent about 0 (large) are associated with slow (fast) decay of correlations [29], and their characterisation has been used to get insight on whether the system is ergodic or not [31]. Moreover, trajectories characterised by a low or high finite-time Lyapunov exponent can play a significant role

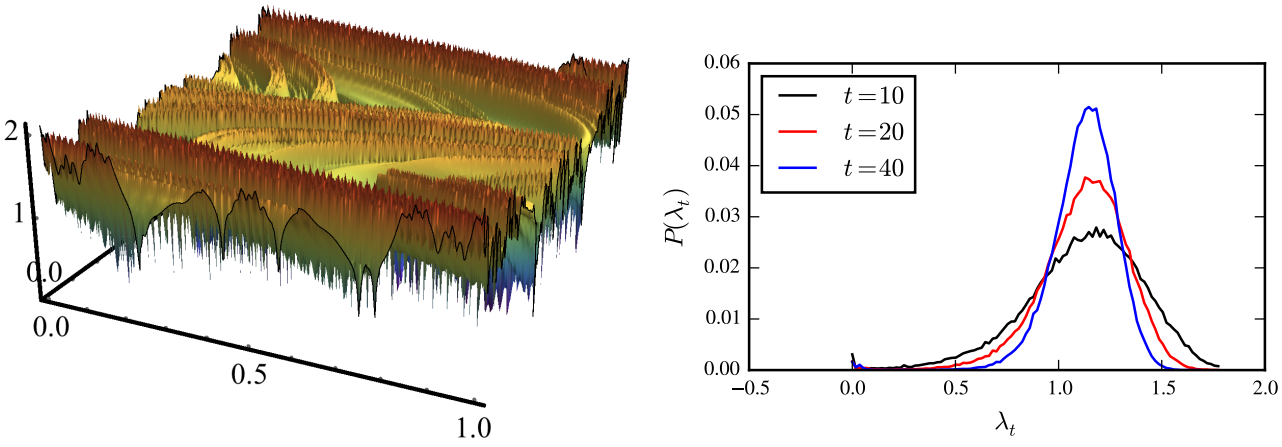

Figure 2.4.: Two main characteristics of the FTLE. Left panel: the intricate dependency of $\lambda_{t_o}(\boldsymbol{x})$ (z-axis) with the state $\boldsymbol{x}$ (2D, x and y axis). Right panel: the distribution of FTLE for different finite times t shows exponential decaying tails with λ_t, and decay with t. The system used was the Standard Map (Eq. 2.15) with $K = 6$, over the full phase-space, $\Gamma = \Omega = [0, 1]^2$. The left figure is $\lambda_4(\boldsymbol{x})$ of the standard map with $K = 6$. The $P(\lambda_t)$ was computed from 10^5 uniformly distributed initial conditions on Γ, $\lambda_t \in [0, \log(6)]$ was divided in 100 equal bins and $P(\lambda_t)$ is the relative count on each.

in the dynamics of interfaces in chaotic flows [30] and others [9].

It will now be argued, based on examples in the literature, that analysis based on the tails of the distribution FTLE, such as those in Refs. [29–31], often requires estimating an integral of the form of

$$I(\lambda_{t_o}) \equiv \int_\Gamma A(\boldsymbol{x})\delta(\lambda_{t_o} - \lambda_{t_o}(\boldsymbol{x}))P(\boldsymbol{x})d\boldsymbol{x} \tag{2.17}$$

where $A(\boldsymbol{x})$ is a pre-selected quantity, Γ is a pre-selected sampling region (typically $\Gamma = \Omega$), and $P(\boldsymbol{x})$ is the measure attributed to $\boldsymbol{x}$ (often constant, $P(\boldsymbol{x}) = 1/|\Gamma|$). First, consider the problem of estimating the distribution $P(\lambda_{t_o})$, e.g. Refs. [9, 29, 31, 38]. The traditional technique to estimate such distribution is to sample M states $\boldsymbol{x}$ according to $P(\boldsymbol{x})$ and estimate $P(\lambda_{t_o})$ using the estimator M_λ/M, where M_λ is the number of samples with $\lambda_{t_o}(\boldsymbol{x}) \in [\lambda_{t_o}, \lambda_{t_o} + \Delta\lambda_{t_o}]$. Formally, this procedure corresponds to compute an integral of the form of $I(\lambda_{t_o})$ with $A(\boldsymbol{x}) = 1$, and with small bin size $\Delta\lambda_{t_o}$. The second example is retrieved from Ref. [29]. There, in order to evaluate the contribution of the algebraic region of the phase-space to the power spectra, the authors decomposed the power-spectra in two terms corresponding to the power spectra of trajectories with low FTLE and trajectories with high FTLE. This required estimating the power spectra from a set of trajectories conditioned to an interval of λs on the tails of the distribution. Associating the power spectrum ($S(f)$ in the ref.) with $A(\boldsymbol{x}) = A_f(\boldsymbol{x})$ where f is the frequency, the power spectra represented in Fig. 4 of Ref. [29] corresponds to compute integrals (for different frequencies f) of the form of

$$\mathbb{E}\left[A|\lambda_{t_o}\right] = \frac{1}{P(\lambda_{t_o})}\int_\Gamma A(\boldsymbol{x})\delta(\lambda_{t_o} - \lambda_{t_o}(\boldsymbol{x}))P(\boldsymbol{x})d\boldsymbol{x} \quad . \tag{2.18}$$

Numerically, such integral can be performed in the same way as done for the first example, by drawing samples from $P(\boldsymbol{x})$ and estimating the average of A (one for each frequency) over the samples with a $\lambda_t(\boldsymbol{x})$ within a range of λ, which corresponds to compute an integral of the form of Eq. 2.17.

Numerically estimating the integral in Eq. 2.17 is challenging for two reasons: First, $P(\lambda_{t_o})$ decays with t_o (right panel of Fig. 2.4): the distribution of FTLE often follows a large deviation principle, $P(\lambda_{t_o}(\boldsymbol{x})) \propto \exp\left(t_o s(\lambda_{t_o})\right)$, where s is intensive in respect to t_o and is often concave. [37] Consequently, find or sample states with increasing λ_{t_o} using e.g. uniform sampling, requires an exponentially high number of samples. The second challenge is that the dependency of $\lambda_{t_o}(\boldsymbol{x})$ with $\boldsymbol{x}$ has a very rough landscape with multiple minima and maxima (left panel of Fig. 2.4). Rough landscapes with multiple minima are known to pose challenges to traditional numerical techniques as they become trapped in a local minima. [39]

Find and sample states with high or low-λ has been addressed in the literature with numerical techniques that go beyond traditional uniform sampling [9–11]. Such techniques have been successfully applied to find [9] and sample [10, 11] states with extremal λs in different chaotic systems. Refs. [9, 11] use a population Monte Carlo canonical ensemble where λ_t plays the role of the energy E to find or sample high/low chaotic states. The method computes stochastic trajectories that, from the numerical tests performed, are indistinguishable from (deterministic) trajectories of the system.

Ref. [10] proposes a flat-histogram simulation to find high or low chaotic states by developing an observable to quantify the chaoticity of the state. The development of numerical methods that go beyond traditional uniform sampling, such as those in Refs. [9–11], is a strong motivation to this thesis to study the general applicability of Monte Carlo to sample rare events in chaotic systems.

2.6. Transient chaos

Chaotic behaviour does not always present itself in the form of a dynamics on an attractor [8, 40]. One example is chaotic scattering: the situation where trajectories, parametrised e.g. by an impact parameter, enter in a scatterer, perform a complicated movement on it, and them leave the scatterer. When the scatterer's potential leads to non-linear equations of motion, there is a high chance that the trajectories exhibit chaos, that is, until they leave the scatterer, they exhibit a positive Lyapunov exponent. Transient chaos is found in different phenomena such as chaotic scattering [41], leaked chaotic systems [40, 42], spatially extended systems [43], bifurcations [44] and chaotic billiards [21, 22, 45]. Many of such phenomena appear in different fields and applications, such as acoustics [46, 47], chemical reactions [48], micro-laser cavities [49–53], hydrodynamical flows [54], nanosystems [55], and optical metamaterials [56].

The main characteristic of transient chaos, in opposition to permanent chaos, is that trajectories leave after some time t_e, which implies that the natural invariant measure is always a zero-measure set. Transient chaos is formally characterised on a set of points, called the chaotic saddle, that do not escape neither forward nor backward in time [8]. The importance of this set is that it is possible to define an invariant natural measure on it. This makes the chaotic saddle an extremely useful set to quantify the invariant chaotic properties of the system because, e.g. it is possible to define the Lyapunov exponent of the system as an average over trajectories on it [8].

The escape time of a state $\boldsymbol{x}$ in a pre-selected region $\boldsymbol{x} \in \Gamma \subset \Omega$ is defined as the first passage time of its trajectory to a pre-selected exit set $\Lambda \subset \Omega$:

$$t_e(\boldsymbol{x}) \equiv \min\{t : \boldsymbol{F}^t(\boldsymbol{x}) \in \Lambda\} \quad . \tag{2.19}$$

Almost all trajectories leave the system; still, some leave faster than others. Such variability is quantified by the distribution of escape time: the probability that a random initial condition $\boldsymbol{x}$ leaves at a given time t_e,

$$P(t_e) = \int_\Gamma \delta(t_e - t_e(\boldsymbol{x}))P(\boldsymbol{x})d\boldsymbol{x} \quad , \tag{2.20}$$

where $P(\boldsymbol{x})$ is the probability assigned to each state in Γ, which is often constant, $P(\boldsymbol{x}) = 1/\int_\Gamma d\boldsymbol{x} \equiv 1/|\Gamma|$. Figure 2.5 represents the typical features of the escape time and escape time distribution: the function $t_e(\boldsymbol{x})$ strongly depends on the initial state $\boldsymbol{x}$, and the escape time distribution decays exponentially with $t_e(\boldsymbol{x})$.

The second reason why the chaotic saddle is so important is because trajectories that do *not* belong

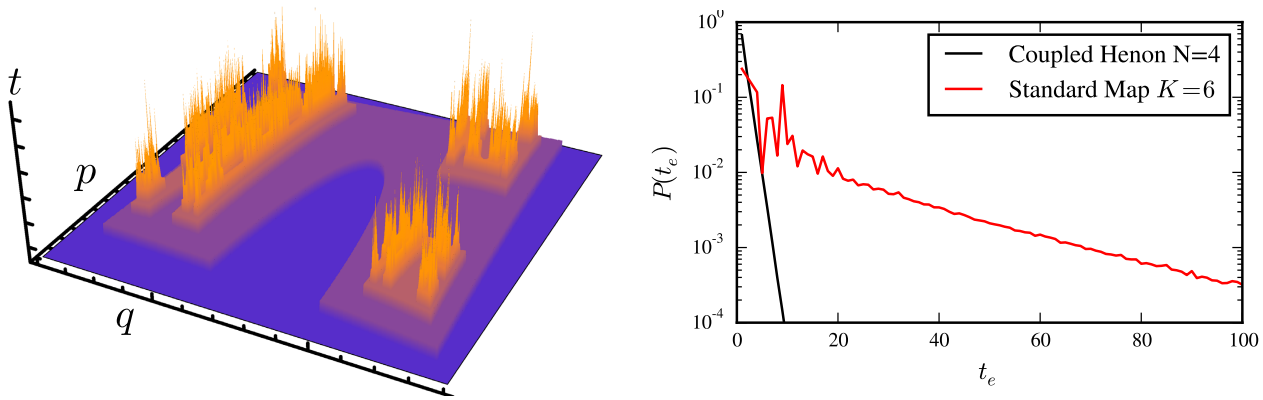

Figure 2.5.: Main characteristics of the escape time of an open chaotic system. Left panel: the dependency of $t_e(\boldsymbol{x})$ with the state $\boldsymbol{x}$, showing an intricate landscape with multiple local and global maxima. The map is the Coupled Hénon map (defined in Eq. 2.27) with $N = 2$, and the two dimensions are a surface of section on the plane $x_2 = 0, y_2 = 0$. Right panel: the exponential decay of the distribution of escape time of (1) the Coupled Hénon map with $N = 4$ and (2) the Standard Map with a leak at $\Lambda = [0, 1] \times [0, 0.1]$. $P(t_e = t)$ was computed by uniformly drawing 10^6 states $\boldsymbol{x} \in \Gamma$, and measure the relative number of times that $t_e(\boldsymbol{x}) = t_e$.

to it but live for a large finite time approach it. Consider a uniformly distributed set of points in the phase-space that is evolved in time. After the first step, some states leave the system as their forward evolution enters the exit set Λ, and only a fraction of initial points survive. Furthermore, at each time step, some of the surviving points leave the system as their forward evolution enters the exit set Λ. The set of points that has not escaped up to time t is approximating the set that never escapes in forward iterations. Moreover, looking at the trajectory of the initial set of points that survived up to time t, the states corresponding to the trajectory at time $t/2$, i.e. $\boldsymbol{F}^{t/2}(\boldsymbol{x})$, will survive up to time $t/2$, *and* will by definition survive by backward iterating them by $t/2$. That is, such states survive for $t/2$ for both backward and forward iterations. Therefore, as $t \to \infty$, such states do not leave the system neither in forward or backward iterations, and thus converge to the chaotic saddle of the system. Any expectation (e.g. the Lyapunov exponent) computed from an ensemble of such states converges to the value computed by an average over states in the chaotic saddle, drawn according to the natural invariant measure on the chaotic saddle, and can thus be used to characterise transient chaos in such systems.

A common numerical technique to characterise transiently chaotic systems is to *find* a long living state $\boldsymbol{x}$ with $t_e(\boldsymbol{x}) \gg 1$ and compute an average over states of its trajectory [6, 8]. For large $t_e(\boldsymbol{x})$, the trajectory between times $[t_s - t_e(\boldsymbol{x})/2, t_e(\boldsymbol{x})/2 + t_s]$ where $t_s \ll t_e(\boldsymbol{x})/2$ is close to a trajectory on the chaotic saddle. When the system is ergodic, an average over this long trajectory corresponds to an average over the natural invariant density and therefore an average over these states characterises invariant properties of the system.

Another common numerical technique to characterise transiently chaotic systems is to compute averages over an ensemble $P(\boldsymbol{x})$ of initial conditions $\boldsymbol{x}$ that leave the system at time $t_e(\boldsymbol{x}) = t$ [8,

22, 57, 58]. There are two motivations for using this technique. One motivation is to characterize invariant properties of the system. Like in the previous technique, the states $\boldsymbol{F}^{t/2}(\boldsymbol{x})$ with $t \to \infty$ are independent samples of the natural invariant density, and thus averages over then characterise invariant properties of the system. Another motivation is to compute non-invariant properties of the system. For example, even though asymptotically $P(t_e)$ becomes independent of $P(\boldsymbol{x})$, for finite times it depends on the particular initial density $P(\boldsymbol{x})$. Formally, numerically computing an average over an ensemble can be interpreted as numerically approximating the integral

$$I(t_e) \equiv \int_\Gamma A(\boldsymbol{x})\delta(t_e - t_e(\boldsymbol{x}))P(\boldsymbol{x})d\boldsymbol{x} \quad , \tag{2.21}$$

where $A(\boldsymbol{x})$ is here a generic function of the phase-space. Let us now enumerate examples in the literature where the numerical problem to be solved can be interpreted as numerically estimating an integral of the form of $I(t_e)$.

Compute the escape time distribution: The calculation of $P(t_e)$ is often done by drawing states from $P(\boldsymbol{x})$ (e.g. uniform density), and counting the relative number of states that exited at escape time t_e [8]. Formally, this corresponds to compute an integral of the form of $I(t_e)$ with $A(\boldsymbol{x}) = 1$.

Compute generalised dimensions of the chaotic saddle: The generalised dimensions are an important property of the chaotic saddle [8], and its calculation is often made by box counting [8, 59–62]. Essentially, the phase-space is divided in boxes $i = 1, ..., B(\varepsilon)$, each of size ε^N, and the generalised dimension D_q is proportional to $\log \sum_i \mu_i^q$, where μ_i is the fraction of points of the saddle that belong to the box i. [60] Numerically, μ_i is computed from a set of points $\boldsymbol{x}_j$ in the saddle by counting how many are in the box i. Such an estimate can be interpreted as the expectation of an indicator function that tells whether the state in the chaotic saddle, $\boldsymbol{F}^{t_e/2}(\boldsymbol{x})$ for $t_e(\boldsymbol{x}) \gg 0$, is inside the box i, $A(\boldsymbol{x}) = \delta(\boldsymbol{F}^{t_e/2}(\boldsymbol{x}) \in i)$. This expectation can be written as a conditional expectation of $A(\boldsymbol{x})$ over states that leave at time t_e,

$$\mathbb{E}\left[A|t_e\right] \equiv \frac{1}{P(t_e)} \int_\Gamma A(\boldsymbol{x})\delta(t_e - t_e(\boldsymbol{x}))P(\boldsymbol{x})d\boldsymbol{x} \quad , \tag{2.22}$$

and taking into account that the interest here is in the limit $t_e \to \infty$. The calculation of this expectation is essentially computing an integral of the form of $I(t_e)$.

Compute the distribution of FTLE on the chaotic saddle: The distribution of FTLE λ_t is another important property of the chaotic saddle [8, 60], and is another example where its calculation requires sampling states close to the saddle. As discussed in section 2.5, the distribution of FTLE is the probability that a state $\boldsymbol{x}$ has a FTLE $\lambda_t(\boldsymbol{x}) \in I_\lambda \equiv [\lambda_t, \lambda_t + d\lambda_t]$, for a fixed t. Because a state on the saddle can be interpreted as states $\boldsymbol{F}^{t_e(\boldsymbol{x})/2}(\boldsymbol{x})$ where $\boldsymbol{x}$ is drawn from $P(\boldsymbol{x})$ and $t_e(\boldsymbol{x}) \gg 1$, the estimation of the distribution of FTLE can be done by generating states $\boldsymbol{x}$ according to $P(\boldsymbol{x})$ (e.g. uniformly distributed in Γ), store the ones with large t_e, compute $\lambda_t(\boldsymbol{F}^{t_e/2}(\boldsymbol{x}))$, and construct an

histogram for the different values of λ. Formally, this can be done by computing an expectation given by Eq. 2.22 for the different bins of the histogram centered at λ, $A(\boldsymbol{x}) = A_\lambda(\boldsymbol{x}) \equiv \delta_{\lambda_t(\boldsymbol{F}^{t_e/2}(\boldsymbol{x}))\in I_\lambda}$.

From the perspective of transiently chaotic systems, the interpretation of the different numerical problems as integrals of the form $I(t_e)$ in Eq. 2.21 may not be insightful, but this interpretation will be crucial to approach these different numerical challenges in a common setting.

2.6.1. Strongly chaotic open systems

Strongly chaotic open systems occur in nature in numerous contexts. [8, 40] One example is the description of the transient preceding the thermal equilibrium of a closed chemical reaction: consider a closed chemical reaction with two or more reactants that mix to form a new compound. One way to describe such dynamics is through a mean field on which the concentration of reactants evolves until the reaction reaches a thermal equilibrium. Ref. [63] hypothesised that such transients of the concentrations can be chaotic, and this was experimentally confirmed in Ref. [44]. Another interesting example is the occurrence of turbulent fluctuations in laminar flows. It was shown that turbulent fluctuations ("puff"s) occur for finite times under a laminar flow with high Reynolds number [64]. These finite-time events are associated with a chaotic transient towards laminar flow, and it was shown that they can be explained in terms of a high-dimensional chaotic transient. A signature of strong chaos in open systems is an exponential decay of the escape time distribution with the escape time, quantified by the asymptotic escape rate κ, [8]

$$\lim_{t_e\to\infty} t_e^{-1} \log P(t_e) = -\kappa \quad . \tag{2.23}$$

The paradigmatic example of a strongly chaotic open system is the open tent map [8], defined on $\Omega = [0, 1]$ by

$$F(x) = \begin{cases} ax & \text{for } x \le b/(a+b) \\ b(1-x) & \text{for } x > b/(a+b) \end{cases} \tag{2.24}$$

where $a > 1$ and $b > a/(a-1)$. The state exits the system when it leaves the unit interval, i.e. $\Lambda = \mathbb{R} - \Omega$. This map contains the main features of an open chaotic system: it has a positive Lyapunov exponent ($\lambda_L = \frac{a\log(b)+b\log(a)}{a+b}$), an exponential decay of the escape time distribution,

$$P(t_e) = \kappa e^{-\kappa t_e} \quad , \tag{2.25}$$

with $\kappa = -\log(1/a + 1/b)$, and a conditionally invariant measure that is fractal with a (non-integer) fractal dimension D_0 given implicitly by

$$a^{-D_0} + b^{-D_0} = 1 \quad . \tag{2.26}$$

To understand how fractals appear in chaotic systems, let us analyse the surviving set, the set of points that have not escaped up to time t_e, in the tent map with $a = b = 3$. After the first iteration, the points in the interval $[1/3, 2/3]$ leave the system, as they are mapped to out of the unit interval. This implies that the surviving set at the first iteration is the interval $[0, 1] - [1/3, 2/3]$. After the second iteration, each of the sets ($[0, 1/3]$ and $[2/3, 1]$) composing the surviving set are now further chopped, as the middle third interval of each of them exits the system. This construction, represented in figure 2.6, corresponds to the construction of the famous middle third Cantor set, which, in the limit as $t_e \to \infty$, has a non-integer dimension of $D_0 = \log 2/\log 3$, which is also given by Eq. 2.26 for $a = b = 3$. It is in this sense that the escape time function is here described as a fractal-like function.

High-dimensional transient chaos

High-dimensional transient chaos corresponds to transient chaos occurring in systems with more than one unstable direction. In the notation of Sec. 2.2, this corresponds to the Jacobian matrix J^t to have more than one eigenvalue with real part larger than one, which implies that states $\boldsymbol{x}'$ around $\boldsymbol{x}$ diverge with different rates depending on the direction they are from $\boldsymbol{x}$. There is a fascinating phenomenology associated with high-dimensional transients [8, 65] that can be observed in high-dimensional chaotic scattering [8, 41] and transient chaos in spatially extended systems [8]. A generic example of a high-dimensional strongly chaotic open system is a set of d coupled Hénon maps on a ring, defined by a state $\boldsymbol{x} = (x_1, y_1, ..., x_d, y_d) \in \Omega = \mathbb{R}^{2d}$ where each individual map (x_i, y_i) evolves according to

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} A_i - x_i^2 + By_i + k(x_i - x_{i+1}) \\ x_i \end{pmatrix}, \tag{2.27}$$

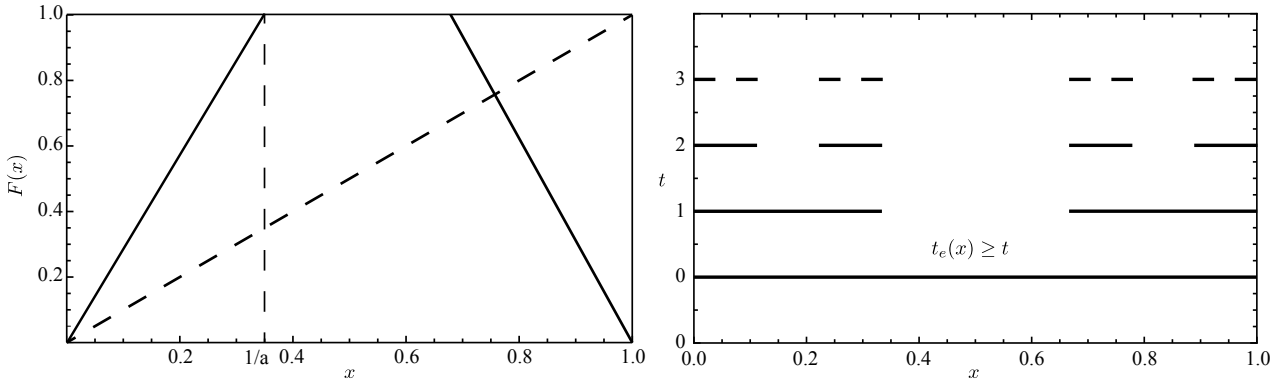

Figure 2.6.: The open tent map, and its corresponding surviving set, equal to the construction of the cantor set. (Left) The open tent map, where the escape correspond to states inside the interval in the middle, that maps to outside the unit interval. (Right) An iteration of the open tent map with $a = b = 3$ corresponds to remove the middle third of each of the plateaus of the surviving set at time t and the set that survives this removal is the surviving set at $t + 1$. The third middle Cantor set is the surviving set at $t \to \infty$.

for $i = 1, ..., d$, $d + 1 \equiv 1$, and with parameters $k = 0.4$, $B = 0.3$, $A_1 = 3$ (if $d > 1$), $A_d = 5$, and $A_i = A_1 + (A_d - A_1)(i - 1)/(d - 1)$. This choice of parameters ensures that a chaotic map is obtained in the $d = 1$ case and corresponds to the map studied in Ref. [6] for $d = 2$. The constraining region is $\Gamma = [-4, 4]^{2d}$ because it covers the chaotic saddle of the system, and $\Lambda = \Omega - \Gamma$, i.e. the trajectory leaves the system if the absolute value of any of the coordinates is higher than 4. The escape function $E(\boldsymbol{x}) = t_e(\boldsymbol{x})$ for $d = 2$ is represented in figure 2.5. Increased dimensionality of the phase-space is known to pose challenges in representation [66], characterisation [8], and is known to be a traditional challenge in minimisation problems [67], and numerical integration [16], and this has also been observed in transient chaos [6–8, 68].

2.6.2. weakly chaotic open systems

Non-hyperbolicity often has dramatic consequences to the statistical properties of transient chaos. One important consequence of non-hyperbolicity is that often $P(t_e)$ no longer asymptotically decays exponentially with t_e, but rather polynomially [8],

$$P(t_e) \sim t_e^{-\alpha} \quad , t_e \gg 1 \quad . \tag{2.28}$$

A great effort has been devoted to understand the mechanism responsible for the power-law behaviour of $P(t_e)$ [21, 22, 69–71]. In particular, an important question that has been addressed is the effect of KAM islands in Hamiltonian systems to the statistics of the escape time distribution, and how it affects e.g. transport [69–71]. It is well known that long-living trajectories "stick" to the borders of the KAM islands and that this phenomena is responsible for the power-law decay [21, 22, 69–71]. This stickiness, in turn, can be understood in terms of the Markov tree model [66, 71, 72]. However, this model is applicable to 2 dimensional systems only, even though stickiness is also observed in higher dimensional Hamiltonian systems (strongly coupled Standard maps) [66, 73].

In practice, understanding stickiness is often achieved by studying statistical properties of long-living trajectories [69, 70]. Due to the power-law decay of $P(t_e)$, typically "long-living" is taken to be logarithmically high in t_e, which is interpreted as states with a high $E(\boldsymbol{x}) = \log t_e(\boldsymbol{x})$ (base 10 in Figs. 1 and 3 of Ref. [70]). For example, the clarification of whether the exponent α in Hamiltonian systems with hierarchical KAM islands is universal or not requires estimating α for a large escape time [69, 70]. The state of the art method to estimate α is to estimate $P(E)$ and fit $\alpha_{\approx} = E^{-1} \log P(E)$ for a high E. This requires estimating $P(E)$, which is traditionally done by sampling N trajectories, count how many have an escape time $t_e(\boldsymbol{x})$ between $[e^E, e^{E+1}]$ (or another base), and estimate $P(E)$ with N_E/N (e.g. in Fig. 3 of Ref. [69, 70], each orange thin line is such an estimate). Such calculation can be interpreted as the numerical estimation of $P(E)$ given by

$$P(E) = \int_{\Gamma} \delta(E - \log t_e(\boldsymbol{x})) P(\boldsymbol{x}) d\boldsymbol{x} \quad , \tag{2.29}$$

for large E. The crucial challenge is that sampling states with large E is often prohibitive because it

requires sampling the tails of $P(E)$: "While this [the universality of the exponent α] may indeed be the case given the hierarchical island structure in the phase space, to attain such a universal exponent one needs a prohibitively long time to perform the needed numerical calculations". [8] Like in the examples given above where the generalised dimensions and distribution of the FTLE with increasing t_e, the same type of numerical estimates can be performed in long living states in weakly chaotic open systems. Thus, more generally, it is possible to interpret that statistical properties of long living trajectories in weakly chaotic systems are obtained by

$$\mathbb{E}\left[A|E\right] = \frac{1}{P(E)} \int_{\Gamma} A(\boldsymbol{x}) \delta(E - \log t_e(\boldsymbol{x})) P(\boldsymbol{x}) d\boldsymbol{x} \quad , \tag{2.30}$$

for large $E = \log t_e$ and where $A(\boldsymbol{x})$ is the statistic of interest (e.g. one of the quantities described after Eq. 2.21). Like Eq. 2.29, this requires computing an integral of the form of

$$I(E) = \int_{\Gamma} A(\boldsymbol{x}) \delta(E - \log t_e(\boldsymbol{x})) P(\boldsymbol{x}) d\boldsymbol{x} \quad . \tag{2.31}$$

A paradigmatic example of a low-dimensional non-hyperbolic open chaotic system used to study stickiness in Hamiltonian systems is the standard map, Eq. 2.15, with a leak [8, 22, 40]. For a parameter where KAM islands exist in the phase-space, such as $K = 2.1$, the map corresponds to a generic mixed-phase-space hamiltonian system (see Fig. 2.7). A leak that does not intersect a KAM island, e.g. $\Lambda = [0, 1] \times [0, 0.1]$, can be used to characterise trajectories that stick close to the KAM islands, as they correspond to the long living trajectories of the leaked system [22].

When the constraining region Γ is chosen to be the image of Λ, $\Gamma = \boldsymbol{F}(\Lambda)$, the escape time distribution corresponds to the distribution of recurrence times of the corresponding closed system [74].

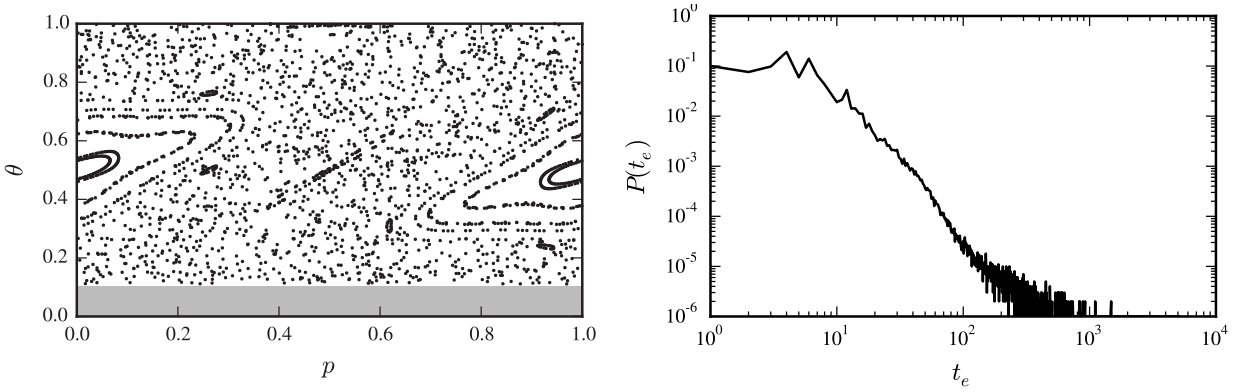

Figure 2.7.: The typical characteristics of a mixed phase-space Hamiltonian system with a leak. (Left) The phase-space with a chaotic component and KAM islands and the leak in the chaotic component. (Right) The escape time distribution $P(t_e)$, computed by starting 10^6 initial conditions on the first image of the leak ($\Gamma = \boldsymbol{F}(\Lambda)$) with density as the first iteration of the uniform distribution on the leak, $P(\boldsymbol{x}) = FP[U(\Lambda)]$. The map is the leaked standard map with $K = 2.1$ and $\Lambda = [0, 1] \times [0, 0.1]$.

2.6.3. Numerical challenges in open chaotic systems

Apart from a small number of analytically trackable examples such as the open tent map, characterising an open system, such as characterising the chaotic saddle of the coupled Hénon maps defined above or sticky trajectories in the open standard map, is challenging. As summarised by Eq. 2.21 and Eq. 2.31, this requires sampling states with high escape times. However, such states are exponentially rare with increasing t_e, as per Eq. 2.23 and Eq. 2.28. Not only they are exponentially rare, they are also fractally distributed in the phase-space, as figure 2.5 illustrates. Rough landscapes are known to pose challenges to numerical maximisation [39] and numerical integration [75–80], and the fractal landscapes present in chaotic systems are no exception [6, 7, 68]. But not only these states are rare and fractally distributed, they may also be embedded in a high-dimensional phase-space.

Several methods have been proposed to *find* long-living states in open systems [6, 7, 68]. One of the most successful is the *stagger and dagger* [6]. This method starts in a random state $\boldsymbol{x}_0$ with an escape time $t_e(\boldsymbol{x}_0) \geq T^*$ and constructs a pseudo-trajectory given by

$$\boldsymbol{x}_{t+1} = \begin{cases} \boldsymbol{F}(\boldsymbol{x}_t) & \text{if } t_e(\boldsymbol{x}_t) > T^* \\ \boldsymbol{F}(\boldsymbol{x}_t + \boldsymbol{h}) & \text{if } t_e(\boldsymbol{x}_t) \leq T^* \end{cases} \tag{2.32}$$

where $\boldsymbol{h}$ is such that it fulfils the condition $t_e(\boldsymbol{x}_t + \boldsymbol{h}) > T^*$. The search of such $\boldsymbol{h}$ is done by proposing $\boldsymbol{h}$ isotropically and with $|\boldsymbol{h}| = 10^{-U(s_{\min}, s_{\max})}$, where U is the uniform distribution with $s_{\min}, s_{\max}$ free parameters, until the condition is fulfilled (a stagger in the original notation). As long as $|\boldsymbol{h}|$ is small (i.e. large $s_{\min}$), and the system fulfils the conditions of shadowing, the pseudo-trajectory shadows a true trajectory of the system with an escape time given by the length of the trajectory, which can be as long as computationally possible. This method was successfully applied to different systems, and it is robust to increased dimensionality: it was applied to the coupled Hénon map, Eq. 2.27, with $d = 2$ and to the double rotor introduced in Ref. [81]. The method in Ref. [7] generalizes this procedure by considering an directional proposal that maximizes $t_e(\boldsymbol{x}_t + \boldsymbol{h})$ at each step. Even though Ref. [7] argues that it becomes computationally expensive to compute the right direction, it was applied to systems up to eight dimensions. The ideas and applicability of these methods to *find* long living states with high escape times is a strong motivation to adapt these ideas to *sample* such states to compute integrals of the form of Eq. 2.21. The crucial distinction between finding and sampling rare states is that a state found by a search method is not necessarily drawn from a priori known distribution, and, in particular, from $P(\boldsymbol{x})$. Therefore, phase-space averages, such as Eq. 2.22 and Eq. 2.30 cannot be performed from such found states, because there is no guarantee that the average over such samples is an unbiased estimate of $I(E)$.

There have been proposals to sample long-living trajectories more efficiently in open weakly chaotic systems [82], and methodologies that are able to characterize such long-living trajectories are of great need to numerically study rare events in open chaotic systems. This motivates the search for methods that are able to sample long living states in such a way that the unbiased estimates of phase-space ensembles can be performed.

2.7. Summary of the numerical problems to study rare events in chaotic systems

The different problems discussed in the previous sections share common characteristics and numerical challenges that can be summarised as follows:

- the analysis is conditioned to a projection of $\boldsymbol{x}$ into a one-dimensional observable, $E(\boldsymbol{x})$.
- the observable $E(\boldsymbol{x})$ increases linearly with a control parameter, N, and the study is interested in large but finite N.
- the probability distribution of E, $P(E)$, depends on E and N as $P(E) \propto \exp Ns(E)$ where s does not depend on N.
- the focus of the analysis is on rare states in respect to $P(E)$: states whose observable E are in one of the tails of $P(E)$.

To map between the problems introduced in the previous sections to the notation introduced above can be made by interpreting $E(\boldsymbol{x})$ as:

- $E(\boldsymbol{x}) = t_o \lambda_{t_o}(\boldsymbol{x})$ in FTLE of closed systems
- $E(\boldsymbol{x}) = t_e(\boldsymbol{x})$ in strongly chaotic open systems
- $E(\boldsymbol{x}) = \log t_e(\boldsymbol{x})$ in weakly chaotic open systems

and the parameter N as:

- t_o in FTLE of closed systems
- a pre-selected maximum escape time $t_{\max}$ in strongly chaotic open systems
- a pre-selected maximum $E_{\max} = \log t_{\max}$ in weakly chaotic open systems

These 3 observables share two different numerical problems: find rare states [6–9] and sample rare states [10, 11, 83, 84] that can be formalised as follows:

- Find rare states: minimize or maximize $E(\boldsymbol{x})$ over a constraining region Γ, $\boldsymbol{x} \in \Gamma \subset \Omega$, for increasing N.
- Sample rare states: compute the integral of a function $A(\boldsymbol{x})$ conditioned to a particular value of E and for increasing N, over an ensemble of states $\boldsymbol{x}$ distributed according to $P(\boldsymbol{x})$ in a constraining region of the phase-space $\boldsymbol{x} \in \Gamma \subset \Omega$:

$$I(E) = \int_{\Gamma} A(\boldsymbol{x}) \delta(E - E(\boldsymbol{x})) P(\boldsymbol{x}) d\boldsymbol{x} \quad . \tag{2.33}$$

The numerical challenges of these two numerical problems are common: the states $\boldsymbol{x}$ are exponentially difficult to find with increasing N (or E), the function $E(\boldsymbol{x})$ contains multiple local and global minima embedded in fractal-like structures, and a potential high dimensionality D of the phase-space. This calls for a systematic approach to tackle the two problems and these four challenges, which is the main goal of this thesis.

3. Monte Carlo Importance Sampling

So the method we employ is actually a modified Monte Carlo scheme, where, instead of choosing configurations randomly, then weighting them with $\exp(-E/kT)$, we choose configurations with a probability $\exp(-E/kT)$ and weight them evenly.

Metropolis et al. [85]

3.1. Importance Sampling

As introduced in the previous chapter, the traditional methodology to compute an integral of the form of Eq. 2.33 in chaotic systems is to draw m samples $\{\boldsymbol{x}_i\}$ distributed according to $P(\boldsymbol{x})$ from the relevant region Γ and approximate the integral $I(E)$ by the estimator

$$\overline{I(E)} \equiv \frac{1}{m} \sum_{i=1}^{m(\{\boldsymbol{x}\},E)} A(\boldsymbol{x}_i) \tag{3.1}$$

where $m(\{\boldsymbol{x}\}, E)$ denotes the number of samples with $E(\boldsymbol{x}_i) \in [E, E + \Delta E]$ (when E is discrete, $\Delta E = 1$).

The relative distance of the estimator $\overline{I(E)}$ to $I(E)$ is quantified by the ratio $\epsilon(E)$ of the standard deviation $\sigma\left[\overline{I(E)}\right] \equiv \sqrt{\mathbb{E}\left[\overline{I(E)}^2\right] - \mathbb{E}\left[\overline{I(E)}\right]^2}$ with $I(E)$, which is given by

$$\epsilon(E) \equiv \frac{\sigma\left[\overline{I(E)}\right]}{I(E)} = \frac{\sigma[I(E)]}{\sqrt{mP(E)}} \quad . \tag{3.2}$$

Therefore, the number of samples m_* required to achieve a given precision σ_* on a given E_* is $m_*(E) = (\sigma[I(E_*)]/\sigma_*)^2/P(E_*)$. The critical problem in sampling rare states is that, because $P(E)$ decays exponentially with E, $m_*(E)$ increases exponentially with E.

Importance sampling techniques aim to improve this scaling by drawing samples from a distribution $\pi(\boldsymbol{x}) \neq P(\boldsymbol{x})$ on the phase-space (e.g. non-uniform in the phase-space) [16]. Specifically, consider m independent samples $\{\boldsymbol{x}_i\}$ drawn from $\pi(\boldsymbol{x})$, and $\pi(\boldsymbol{x})$ to depend only on E, $\pi(\boldsymbol{x}) = \pi(E(\boldsymbol{x})) = \pi(E)$.

The expected number of samples with a given E, $m(E)$, is given by

$$m(E) \equiv m \int_\Gamma \delta(E - E(\boldsymbol{x}))\pi(E(\boldsymbol{x}))P(\boldsymbol{x})d\boldsymbol{x} = mP(E)\pi(E) \quad . \tag{3.3}$$

Because the samples are no longer drawn uniformly, the estimator in Eq. 3.1 would be biased. Importance sampling uses an unbiased estimator for $I(E)$ given by [16]

$$\overline{I(E)} \equiv \frac{1}{m} \sum_{i=1}^{m(\{\boldsymbol{x}\},E)} A(\boldsymbol{x}_i)\frac{P(\boldsymbol{x}_i)}{\pi(\boldsymbol{x}_i)} \quad . \tag{3.4}$$

Intuitively, when a state $\boldsymbol{x}_1$ is sampled twice as much as another state $\boldsymbol{x}_2$, $\pi(\boldsymbol{x}_1) = 2\pi(\boldsymbol{x}_2)$, it only contributes half of the weight of $\boldsymbol{x}_2$ to the sum in Eq. 3.4. When $\pi(\boldsymbol{x}) = P(\boldsymbol{x})$, this estimator reduces to the one in Eq. 3.1. The advantage of importance sampling is that the relative error of this estimator is

$$\epsilon(E) = \frac{\sigma[I(E)/\pi(E)]}{\sqrt{mP(E)\pi(E)}} \quad . \tag{3.5}$$

That is, $\pi(E)$ can be chosen to favor states $\boldsymbol{x}$ with observable E on the tails of $P(E)$ and therefore improve the precision of the estimator on these tails.

The errors in Eqs. 3.2,3.5 were obtained assuming that the m samples were independent. In traditional methodologies such as uniform sampling, this is the case. However, in the algorithms discussed below, it is not. Therefore, it is necessary to modify Eq. 3.5 for the case where the samples $\{\boldsymbol{x}_i\}$ are drawn from $\pi(\boldsymbol{x})$ and are also correlated. This modification is given by (derivation in appendix A, Eq. A.12)

$$\epsilon(E) = \sigma[I(E)/\pi]\sqrt{\frac{1+2T}{mP(E)\pi(E)}} \quad , \tag{3.6}$$

where T is given by (Eq. A.11)

$$T \equiv \frac{1}{M}\sum_{i=1}^{M}\sum_{\tau=1}^{M-i} R(\tau) \tag{3.7}$$

where $R(\tau)$ is the autocorrelation function of $A(\boldsymbol{x})/\pi(\boldsymbol{x})$. For example, when the sample are independent, $R(\tau) = 0$, $\forall\tau$, $T = 0$ and Eq. 3.6 reduces to Eq. 3.5. Designing an efficient importance sampling technique can be summarised by three steps:

1. choose a suitable $\pi(\boldsymbol{x})$
2. have a method to generate samples from $\pi(\boldsymbol{x})$
3. minimise the integrated autocorrelation time T of the method

3.2. Metropolis-Hastings algorithm

Importance sampling requires samples $\boldsymbol{x}$ drawn according to $\pi(\boldsymbol{x})$. Except for a very small set of $\pi(\boldsymbol{x})$ where the inverse method can be used (e.g. uniform distribution, normal distribution), this is not easily achieved. One of the most general (and oldest) algorithms to accomplish this is the Metropolis-Hastings algorithm [16].

The Metropolis-Hastings algorithm is based on a Markovian, ergodic and detailed balance random walk in the sampling region Γ that asymptotically draws states $\boldsymbol{x}$ according to $\pi(\boldsymbol{x})$. This random walk is initialised from a random state $\boldsymbol{x} \in \Gamma$ and evolves to a new state $\boldsymbol{x'} \in \Gamma$ with a transition probability $P(\boldsymbol{x'}|\boldsymbol{x})$ chosen such that the random walk is ergodic and fulfils detailed balance,

$$P(\boldsymbol{x'}|\boldsymbol{x})\pi(\boldsymbol{x}) = P(\boldsymbol{x}|\boldsymbol{x'})\pi(\boldsymbol{x'}) \quad . \tag{3.8}$$

The $P(\boldsymbol{x'}|\boldsymbol{x})$ is composed by two terms: the proposal distribution $g(\boldsymbol{x'}|\boldsymbol{x})$ and an acceptance distribution $a(\boldsymbol{x'}|\boldsymbol{x})$, $P(\boldsymbol{x'}|\boldsymbol{x}) = g(\boldsymbol{x'}|\boldsymbol{x})a(\boldsymbol{x'}|\boldsymbol{x})$. The random walk fulfils detailed balance, Eq. 3.8, because in Metropolis-Hastings the acceptance probability is chosen as [16]

$$a(\boldsymbol{x'}|\boldsymbol{x}) = \min\left(1, \frac{g(\boldsymbol{x}|\boldsymbol{x'})}{g(\boldsymbol{x'}|\boldsymbol{x})}\frac{\pi(\boldsymbol{x'})}{\pi(\boldsymbol{x})}\right) \quad . \tag{3.9}$$

Metropolis-Hastings algorithm Algorithmically, the procedure is implemented as follows: choose a region Γ and a random initial condition $\boldsymbol{x} \in \Gamma$. Evolve the random walk in time according to:

1. Propose a state $\boldsymbol{x'}$ drawn from $g(\boldsymbol{x'}|\boldsymbol{x})$;
2. Compute $a(\boldsymbol{x'}|\boldsymbol{x})$ replacing $\boldsymbol{x}$ and $\boldsymbol{x'}$ in Eq. 3.9;
3. Generate a random number r in $[0, 1]$. If $r < a(\boldsymbol{x'}|\boldsymbol{x})$, make $\boldsymbol{x'}$ to be the new $\boldsymbol{x}$;
4. Store $\boldsymbol{x}$ and go to 1.

The set of sub-steps 1-4 brings the random walk from its current state $\boldsymbol{x}$ and it is here called a Markov step. After a transient number of steps where the algorithm converges to the asymptotic distribution, the stored states $\boldsymbol{x}$ are (correlated) samples drawn from $\pi(\boldsymbol{x})$.

3.3. Sampling distribution

3.3.1. Canonical Ensemble

One of the most common sampling distributions is the canonical distribution, given by [15]

$$\pi(\boldsymbol{x}) = \pi(E(\boldsymbol{x})) \propto e^{-\beta E(\boldsymbol{x})} \quad . \tag{3.10}$$

Historically, this choice was motivated by the problem of computing a sum of the form of Eq. 2.33 where E was the Hamiltonian of the system, and whose $P(\boldsymbol{x}) \propto \exp(-\beta E(\boldsymbol{x}))$ was a consequence of the interaction of the system with a thermal bath at temperature $1/\beta$.

The canonical ensemble, Eq. 3.10, was motivated by the intrinsic $P(\boldsymbol{x})$ of systems in contact with a thermal bath, but it was soon realised that the canonical sampling distribution also allows to reduce the variance of the estimator in Eq. 3.6 on specific regions of the energy, by tuning β to favor such states, even when $P(\boldsymbol{x})$ is not the canonical ensemble [15]. Specifically, in the canonical ensemble, the number of sampled states with $E(\boldsymbol{x}_i) = E$ is

$$m(E) \propto mP(E)e^{-\beta E} \propto me^{-\beta E+S(E)} \tag{3.11}$$

and the maximum of m is at E^* solution of $\beta = dS/dE(E^*)$. Therefore, β tunes which energy E is on average sampled the most.

Another important use of the Metropolis-Hastings algorithm is to minimize functions: as $1/\beta \to 0$, the only states that the system visits are those whose energy is minimum. This led to the development of one of the most important minimisation techniques in high dimensional phase-spaces – stimulated annealing – on which $1/\beta$ is slowly decreased in time for the algorithm to slowly approach the global minimum [86, 87].

Stimulated annealing and Metropolis-Hastings algorithm with a canonical sampling distribution can both fail to achieve their goal: when the phase-space has a rough landscape, for a fixed $\beta \gg 1$, the algorithm can become trapped in a local minimum for an arbitrary long time as it may take an arbitrary amount of time to cross the energy barrier in the phase-space that separates one minimum from the other, see figure 3.1. To reach a new minimum, one of the two situations must occur: (i) the proposal distribution proposes a new minimum, effectively tunnelling the energy barrier; (ii) the random walk crosses the energy barrier by crawling it. The difficulty in (i) is that there are a very small number of states with low E (P decays with E), and therefore these states are also difficult to be proposed by the sampling distribution without knowing them a priori (trivial situation where the algorithm is not needed). The difficulty in (ii) is that it requires crossing an energy barrier of height ΔE, which, in the canonical ensemble, has a probability proportional to $\exp(-\beta\Delta E)$. Therefore, as β is increased to better approximate the global minimum, the algorithm more likely becomes trapped on a local minimum. This challenge is also present in sampling rare states: the sampling distribution is maximal at states whose energy is minimum, but the algorithm may be sampling from a different distribution for an arbitrarily long time because in practice it is not able to cross energy barriers. The existence of multiple local minima in observables of chaotic systems, described in Chapter 1, hints that, in applications to chaotic systems, the canonical ensemble may become trapped in local minima.

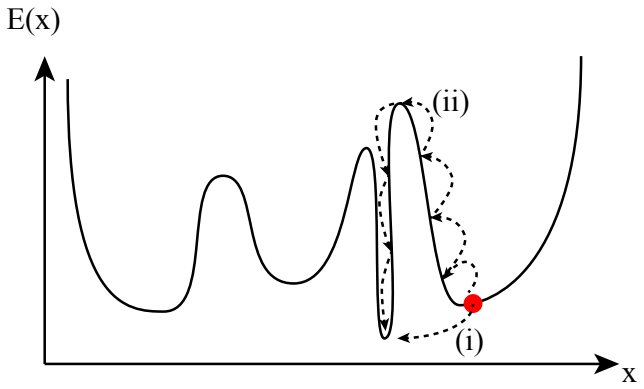

Figure 3.1.: Metropolis-Hastings with a canonical ensemble typically fails to find the global minimum, or to sample from the correct distribution, when the energy landscape is rough. This is because it may not be able to leave a local minimum to a better minimum in any of the two possible paths: (i) tunnel the energy barrier separating two minima or (ii) climb the energy barrier with multiple small steps. Case (i) is unlikely because the "basin of attraction" of the global minimum is small and is therefore unlikely to find; case ii) is unlikely because the algorithm has to climb an energy barrier.

3.3.2. Flat-histogram ensemble

The limitations of the canonic ensemble described above were extensively discussed in the literature [15, 16, 75–80, 88], and one of the most successful candidates to address the problem of multiple minima (and other problem important in statistical physics, that $dS/dE(E)$ may not be monotonic and therefore $\beta = dS/dE$ has multiple solutions) is to use the sampling distribution [80, 88]

$$\pi(\boldsymbol{x}) = \pi(E(\boldsymbol{x})) \propto \frac{1}{P(E(\boldsymbol{x}))}, \quad E \in [E_{\min}, E_{\max}], \tag{3.12}$$

for a given choice of $E_{\min}, E_{\max}$ that defines the region of interest on the observable E. This is known as flat-histogram (or multicanonical) because, replacing Eq. 3.12 in Eq. 3.3 leads to a constant average number of samples on each E,

$$m(E) = \frac{m}{E_{\max} - E_{\min}} \quad . \tag{3.13}$$

Consequently, the dependence of the variance in Eq. 3.6 is only due to the autocorrelation T, which implies that the computational cost to draw a state on the tail of $P(E)$ is no longer driven by the exponential decrease of $P(E)$, but by the computational cost to draw uncorrelated samples from $\pi(\boldsymbol{x})$.

3.3.3. Wang-Landau algorithm

The main limitation of the flat-histogram is that it requires knowing $P(E)$ in advance, which is very often unknown. A major development in the usage of flat-histogram was the Wang-Landau algorithm [88], that modifies the Metropolis-Hastings algorithm to a non-markovian chain that asymptotically converges to a flat-histogram Metropolis-Hastings. The Wang-Landau algorithm starts with

an approximation of $P(E)$, $P_{WL}(E) = 1$, and, on step 4 of the MH (see algorithm 3.2), it multiplies $P_{WL}(E(\boldsymbol{x}))$ ($\boldsymbol{x}$ is the current state of the random walk) by a constant $f > 1$. After a given number of steps, f is reduced by a factor 2 and this procedure is repeated until a final $f_{\min} \simeq 1$ is reached. [88] The value $f_{\min}$ and how often each refinement of f is made dictates how close $P_{WL}(E)$ will be from $P(E)$ [88–90].

This algorithm is non-markovian because the acceptance probability $a(\boldsymbol{x}'|\boldsymbol{x})$, that depends on $P(E)$ via $\pi(\boldsymbol{x}) \propto 1/P(E(\boldsymbol{x}))$, now depends not only on $\boldsymbol{x}$, but also on which E the previous state was, and the previous one, and so on. While there is still no formal proof that Wang-Landau converges to $P(E)$, there is strong numerical evidence that it does for specific update rules of how f is decreased [89, 90]. More importantly, the convergence of the Wang-Landau algorithm can always be confirmed a posteriori, by analysing the histogram of E, and comparing it to the expected flatness of $m(E)$.

3.4. Efficiency

The quantifier of precision in Monte Carlo simulations is the standard deviation $\sigma(\overline{I(E)})$, Eq. 3.6: the lower it is, the smaller the error bars around the mean, and therefore the better the numerical approximation it is from the true value, for a fixed number of samples m. The interest here is how the required number of samples m for a fixed error σ scales with N, the parameter discussed in sec. 2.7 that tends to infinity. As discussed in Sec. 3.1, uniform sampling requires m to exponentially increase with N to maintain a constant standard deviation. In the Metropolis-Hastings, the samples are correlated and therefore the variance of the estimator, Eq. 3.6, also depends on the integrated autocorrelation time T. In principle, the autocorrelation time depends on the particular value of E, which makes the analysis of the efficiency, in terms of the integrated autocorrelation time, dependent on the particular value of E considered. To avoid this complication, the efficiency is often quantified in terms of the average round-trip [91–93], which is an upper bound for the number of samples m (Markov steps) required to obtain an uncorrelated sample from $\pi(\boldsymbol{x})$ [93, 94]. The round-trip, τ, is the average number of Markov steps (samples) required for the algorithm to go from a state $\boldsymbol{x}$ with $E(\boldsymbol{x}) = E_{\min}$ to a state $\boldsymbol{x}$ with $E(\boldsymbol{x}) = E_{\max}$ and return back. The idea is that when $E_{\min}$ and $E_{\max}$ are chosen such that a complete round-trip crosses E that maximizes $P(E)$, at that maximum the algorithm de-correlates from any previous state. Numerically, the round-trip time is computed by having a boolean value tracking whether the random walk is moving in the direction $E_{\min} \to E_{\max}$ or $E_{\max} \to E_{\min}$, and use it to measure the total number of steps required to make the full round-trip.

Importance sampling Monte Carlo is widely used in statistical physics because the computational cost often scales scales polynomially with N [16, 93], which is a dramatic improvement over uniform sampling. To see how Monte Carlo can achieve a polynomial scaling of the round-trip with N, consider an hypothetical implementation of the Metropolis-Hastings with flat-histogram ensemble, over a region of interest $[E_{\min}, E_{\max}]$, and let us compute the round-trip time under simplifying assumptions [93]. To that, let us analyse how the random walk in $\boldsymbol{x}$ is mapped to a random walk in E. Formally, the Metropolis-Hastings algorithm is a Markovian random walk in the phase-space $\boldsymbol{x}$

defined by $P(\boldsymbol{x}'|\boldsymbol{x})$. Given a function $E(\boldsymbol{x})$, the random walk on $\boldsymbol{x}$ can be mapped to a random walk on E by marginalising over all $\boldsymbol{x}$. When this is done, the random walk on E is no longer (1st order) Markovian because the dynamics over $\boldsymbol{x}$ induce correlations on the random walk on E. Still, let us assume that these correlations are negligible, such that the random walk on E is a Markovian process, uniquely determined by $P(E'|E)$. By construction, flat-histogram guarantees that the asymptotic distribution in E is $m(E) = \text{const}$. Therefore, this random walk is a simple diffusion process on a one-dimensional box of size $[E_{\min}, E_{\max}]$. Let us further assume that the random walk only moves to neighbourhood Es of the current E it is, or, in other terms, $P(E'|E)$ can be represented by a tridiagonal matrix with a probabilities to go to $E+\Delta E$, $E-\Delta E$, and stay in the same E. In general, this is not easily achieved because it requires proposing a state $\boldsymbol{x}'$ close to $\boldsymbol{x}$ in such a way that $\Delta E \equiv E(\boldsymbol{x}') - E(\boldsymbol{x}) \approx 1$. Under the above assumptions, an initial density evolves in time by diffusing to neighbourhood states and the mean displacement is $\sigma_E(t) = \Delta E\sqrt{t}$. For an initial state to move from $E_{\min}$ to $E_{\max}$ (half round-trip), it has to diffuse and the time τ required to make half round-trip is therefore obtained by solving $\sigma_E(\tau) = E_{\max} - E_{\min}$, or

$$\tau = \left(\frac{E_{\max} - E_{\min}}{\Delta E}\right)^2 \quad . \tag{3.14}$$

The observable increases linearly with N, and therefore, when ΔE does not depend on N,

$$\tau(N) \sim N^2 \quad . \tag{3.15}$$

For example, consider the problem of sampling states of an open chaotic system with different escape times t_e from $t_e = 1$ up to $t_e = t_e^{\max}$. In this case, $E_{\min} = 1$ and $E_{\max} = N = t_e^{\max}$. In this problem, the round-trip is expected to scale as

$$\tau(t_e^{\max}) \sim (t_e^{\max})^2 \quad . \tag{3.16}$$

There are known deviations of the scaling in Eq. 3.15 leading to a higher exponent, N^{2+z} with $z > 0$ [91–93]. These can happen, for example, when the acceptance rate decreases with N, which makes $\Delta E \to 0$ as $N \to \infty$, or due to correlations on the Markov chain on $\boldsymbol{x}$ such that the random walk on E is not Markovian [93]. Nevertheless, these do not qualitatively change the argument: Monte Carlo with flat-histogram has the potential of generating rare states polynomially with increasing N, while uniform sampling generates rare states exponentially with increasing N.

3.5. The challenge of Monte Carlo in Chaotic Systems

A key ingredient for an efficient Monte Carlo simulation is the proposal distribution $g(\boldsymbol{x}'|\boldsymbol{x})$ [95]. The ideal proposal distribution of a Metropolis-Hastings algorithm draws $\boldsymbol{x}'$ independently of $\boldsymbol{x}$ according to $\pi(\boldsymbol{x})$, $g(\boldsymbol{x}'|\boldsymbol{x}) = \pi(\boldsymbol{x})$. This is because i) the acceptance in Eq. 3.9 is always one and ii) each step of the random walk generates an independent sample $\boldsymbol{x}$, which implies that the variance in Eq. 3.6

is minimal. The difficulty of sampling rare events in chaotic systems is that a useful $\pi(\boldsymbol{x})$ to sample them is a difficult function to sample from. For concreteness, consider the particular problem of sampling high-escape times in the open tent map, Eq. 2.24, whose $t_e(\boldsymbol{x})$ is represented in Fig. 2.6, and consider the canonical sampling distribution, $\pi(\boldsymbol{x}) \propto \exp(-\beta t_e(\boldsymbol{x}))$. In this example, $\pi(\boldsymbol{x}) \propto \exp(-\beta)$ between $[1/a, 1-1/a]$, the large interval in Fig. 2.6, and so forth $(\exp(-\beta t_e))$ in subsequent intervals. The number of intervals increases as 2^{t_e}. Therefore, sampling from $\pi(\boldsymbol{x})$ would require enumerating every interval, sample one at random according to a correct distribution that depends on β, and then sample a uniform point within that interval. While this could be achieved in simple maps such as the open tent map, this is unfeasible in a general chaotic system where the $t_e(\boldsymbol{x})$ dependency is unknown.

One alternative would be to consider $g(\boldsymbol{x}'|\boldsymbol{x})$ to be the uniform distribution over Γ. One could imagine that changing the sampling distribution $\pi(\boldsymbol{x})$ alone could decrease the variance of the estimator, since this gives more preference to rarer states. However, changing the sampling distribution alone has no gain in decreasing the variance of the estimator. The number of samples on a given high t_e increases, but at the expense of an equal increase of the autocorrelation of the samples. This is seen by calculating the average acceptance rate of a sample at a given escape time t_e in the canonical ensemble. The acceptance in this case is given by $A(\boldsymbol{x}'|\boldsymbol{x}) = \min\{1, \exp(-\beta(t_e(\boldsymbol{x}') - t_e(\boldsymbol{x}))\}$ where $\beta < 0$, since the aim is to reach higher $E(\boldsymbol{x}) = t_e(\boldsymbol{x})$. The acceptance of a state $\boldsymbol{x}$ is given by

$$a(\boldsymbol{x}) = \int_\Gamma d\boldsymbol{x}' A(\boldsymbol{x}'|\boldsymbol{x}) g(\boldsymbol{x}'|\boldsymbol{x}) \quad . \tag{3.17}$$

Because $g(\boldsymbol{x}'|\boldsymbol{x}) = 1/|\Gamma|$ and π only depends on $t_e(\boldsymbol{x})$, $a(\boldsymbol{x})$ does not depend on $\boldsymbol{x}$, only on t_e: $a(\boldsymbol{x}) = a(t_e(\boldsymbol{x}))$. The average acceptance rate at a given t_e, $A(t_e) \equiv \mathbb{E}\left[a(\boldsymbol{x})|t_e\right]$, is given by

$$A(t_e) = \frac{1}{m(t_e)} \int_\Gamma d\boldsymbol{x} \delta(t_e - t_e(\boldsymbol{x})) e^{-\beta t_e(\boldsymbol{x})} a(\boldsymbol{x}) \quad . \tag{3.18}$$

Because $a(\boldsymbol{x})$ only depends on t_e, it can be pulled out of the integral, and thus $A(t_e) = a(t_e)$. Taking into account that $P(t_e') = \kappa \exp(-t_e' \kappa)$, the acceptance rate can be computed analytically by integrating Eq. 3.17 and leads to

$$A(t_e) = \frac{e^{-\kappa t_e}\beta + e^{t_e \beta}\kappa}{\beta + \kappa} \quad . \tag{3.19}$$

This shows that the acceptance rate decays exponentially with increasing t_e (recall that $\beta < 0$ to reach states with high t_e). This same argument applies to a flat-histogram simulation, where $\pi(\boldsymbol{x}) \propto \exp(\kappa t_e(\boldsymbol{x}))$. In the notation of the derivation of Eq. 3.14, this implies that the there is an exponentially increasing probability for the state to *stay* in the same $E = t_e$. This implies that ΔE in Eq. 3.14 depends on N, and, in particular, it decreases exponentially with N, leading to an exponential increase of the round-trip time.

In summary, this chapter showed that Metropolis-Hastings is an excellent candidate to approach the numerical challenges found in the study of rare events in chaotic systems. Firstly, because it

is grounded in strong mathematical results such as importance sampling theorem and asymptotic convergence of Markov processes. Secondly, because it is formulated with very little assumptions about the system, the observable of interest or the dynamics of the system, which gives enough freedom to adapt it to the specific aim (sampling or finding), observable, and system. Thirdly, because there seems to be no theoretical reason for the sampling to be exponential; Monte Carlo is used to sample rare states in polynomial time in other problems of statistical physics. Finally, because the numerical problems found in chaotic systems can be re-written as problems where Monte Carlo is suitable for, as done in chapter 2. On the other hand, this chapter also showed that the optimal proposal of Metropolis-Hastings is unfeasible in chaotic systems, and that, without any extra information about the system, Metropolis-Hastings is as efficient as the traditional uniform sampling. It is on this duality that this thesis is laid on. The core of this thesis, the next chapter, introduces a theoretical methodology to consistently add information about the system (e.g. it is chaotic, the landscape is fractal) to the proposal distribution in order to efficiently use Metropolis-Hastings to study rare events in chaotic systems.

4. Efficient Proposals for Chaotic Systems

If σ is too low, the Metropolis steps are too short and move too slowly throught the target distribution; if σ is too high, the algorithm almost always rejects and stays in the same place.

Gelman, Roberts, and Gilks [16]

This chapter constructs a proposal distribution for the Metropolis-Hastings algorithm in chapter 3 suitable for problems of chaotic systems outlined in chapter 2. This construction is divided in three steps that correspond to the three next sections. Section 1 answers to the question: what should the proposal distribution aim for? The main conclusion of this section is that the proposal distribution should propose the state $\boldsymbol{x}'$ from $\boldsymbol{x}$ such that the average distance $\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right]$ is constrained to a given value, given by Eq. 4.6 below. This equation is a condition to the proposal distribution that the remaining sections aim to obtain. Section 2 answers to the question: how to generate a state $\boldsymbol{x}'$ that is somehow correlated with $\boldsymbol{x}$, and how to quantify such correlation (in order to later correlate their observables)? The main result of this section is that the correlation can be quantified by the time $t_\star$ the trajectories are close by, and that there are two types of proposal distributions that guarantee a correlation of $t_\star$ of $\boldsymbol{x}'$ with $\boldsymbol{x}$, given by Eq. 4.11 and Eqs 4.13, 4.15 below. Section 3 answers to the question: how is the correlation time $t_\star$ related to the average distance $\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right]$? The main result of this section are explicit formulas that relate $t_\star$ with $\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right]$ for two observables introduced in Chapter 2, the escape time in open systems $E(\boldsymbol{x}) = t_e(\boldsymbol{x})$ with $t_\star$ given by Eq. 4.34 below, and the FTLE, $E(\boldsymbol{x}) = t_o\lambda_{t_o}(\boldsymbol{x})$, with $t_\star$ given by Eq. 4.24 below. These specific results together compose the main result of this chapter because they construct a proposal distribution to be used in Metropolis-Hastings that, in principle, makes it an efficient method to study rare events in chaotic systems.

4.1. Aim of the proposal distribution

The impossibility to construct an optimal proposal distribution $g(\boldsymbol{x}'|\boldsymbol{x})$ (sec. 3.5) motivates the use of an heuristic that is still efficient. Because efficiency means minimising the round-trip time, the major goal here is to construct a proposal distribution that makes the round-trip time to scale as N^2. As discussed in section 3.4, one crucial aspect of the round-trip time to scale as N^2 is that the acceptance rate must not depend on N. A low acceptance makes the proposals $\boldsymbol{x}'$ to be more likely rejected and

because a rejection makes the Markov process to stay on the same state $\boldsymbol{x}$, it increases the round-trip time. When $\boldsymbol{x}'$ is independent of $\boldsymbol{x}$, the acceptance rate decreases exponentially with increasing N, as discussed in Sec. 3.5. This implies that $\boldsymbol{x}'$ has to be somehow similar to $\boldsymbol{x}$. However, similarity between states slows down the diffusion on E, and therefore also decreases the round-trip time.

On the one hand, $\boldsymbol{x}'$ cannot be too similar from $\boldsymbol{x}$, but on the other hand, $\boldsymbol{x}'$ cannot be too different from $\boldsymbol{x}$. These two extremes can be understood in terms of the acceptance alone. When the two states are independent, the acceptance is too low (0), when the states are too similar, the acceptance is high (1). The above considerations motivate a proposal distribution constructed to avoid these the two extremes. One condition that avoids these extremes is to impose a proposal distribution that bounds the acceptance rate away from 0 and 1. One way to achieve this is to impose a constance acceptance,

$$a(\boldsymbol{x}'|\boldsymbol{x}) = a_\star \quad , \tag{4.1}$$

away from 0 and 1.[1] This avoids low acceptance rates because $a_\star > 0$ is fixed, and also avoids high correlations because the more $\boldsymbol{x}'$ and $\boldsymbol{x}$ are correlated, the closer the acceptance is to $1 \neq a_\star$. Still, the crucial goal is not that the acceptance is exactly $a_\star$, but rather that it does not depend on N.

To achieve Eq. 4.1, see Eq. 3.9, $g(\boldsymbol{x}'|\boldsymbol{x})$ must generate $\boldsymbol{x}'$ from $\boldsymbol{x}$ in such a way that it exactly compensates the difference that $g(\boldsymbol{x}'|\boldsymbol{x})$ has from $\pi(\boldsymbol{x})$, which, in general, is non-trivial to achieve. The next step is thus to approximate Eq. 4.1 by a simpler condition. Lets first notice that $\pi(\boldsymbol{x})$ only depends on $\boldsymbol{x}$ through $E(\boldsymbol{x}) = E_{\boldsymbol{x}}$, $\pi(\boldsymbol{x}) = \pi(E_{\boldsymbol{x}})$. Therefore, at the very least, the proposal $g(\boldsymbol{x}'|\boldsymbol{x})$ should guarantee that $\boldsymbol{x}'$ is generated from $\boldsymbol{x}$ in such a way that $\pi(E_{\boldsymbol{x}'})$ is neither too close (high acceptance) nor too far (low acceptance) from $\pi(E_{\boldsymbol{x}})$. Quantitatively, this can be written as

$$\frac{\pi(E_{\boldsymbol{x}'})}{\pi(E_{\boldsymbol{x}})} = a \tag{4.2}$$

where $0 < a < 1$ is a constant. When the proposal is able to achieve a small variation of E, $\pi(E_{\boldsymbol{x}'})$ can be expanded in Taylor series around $E_{\boldsymbol{x}'} = E_{\boldsymbol{x}}$, which allows to write

$$\frac{\pi(E_{\boldsymbol{x}'})}{\pi(E_{\boldsymbol{x}})} = 1 + \frac{d\log\pi(E_{\boldsymbol{x}})}{dE}(E_{\boldsymbol{x}'} - E_{\boldsymbol{x}}) \quad . \tag{4.3}$$

In statistical physics an heuristics often used in Metropolis-Hastings is $\Delta E \equiv E_{\boldsymbol{x}'} - E_{\boldsymbol{x}} \approx 1$. E.g. it was used in the derivation of Eq. 3.15, and has been used for example in spin systems (single spin flip) [93], ensemble of complex networks (single link exchange) [96], and proteins [97].

The condition in Eq. 4.2 is a condition on $\boldsymbol{x}'$ and $\boldsymbol{x}$. However, $\boldsymbol{x}'$ is obtained from $\boldsymbol{x}$ through $g(\boldsymbol{x}'|\boldsymbol{x})$. Therefore, Eq. 4.2 as to be written in the form of an expectation over all possible $\boldsymbol{x}'$. Formally, this

[1] For example, $a_\star$ could be 0.234, proposed in Ref. [95].

expectation is computed as

$$\mathbb{E}\left[\frac{\pi(E_{\boldsymbol{x}'})}{\pi(E_{\boldsymbol{x}})}|\boldsymbol{x}\right] = \int_{\Gamma} \frac{\pi(E_{\boldsymbol{x}'})}{\pi(E_{\boldsymbol{x}})} g(\boldsymbol{x}'|\boldsymbol{x}) d\boldsymbol{x}' \quad , \tag{4.4}$$

and thus, formally, the proposal distribution should be such that

$$\mathbb{E}\left[\frac{\pi(E_{\boldsymbol{x}'})}{\pi(E_{\boldsymbol{x}})}|\boldsymbol{x}\right] = a \quad . \tag{4.5}$$

Using Eq. 4.3, an average constant acceptance is thus achieved when

$$\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right] = \frac{a-1}{d\log\pi(E_{\boldsymbol{x}})/dE} \quad . \tag{4.6}$$

This equation, the main result of this section, is a condition that an efficient Metropolis-Hastings imposes to the proposal distribution in terms of the average difference in the observable E. This condition is non-trivial because it depends on the particular π, E, $\boldsymbol{F}$, and $\boldsymbol{x}$. Lets first specify it for the two sampling distributions discussed in the previous chapter:

Canonical ensemble When $\pi(\boldsymbol{x}) = \exp(-\beta E(\boldsymbol{x}))$, the condition in Eq. 4.6 is given by

$$\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right] = \frac{1-a}{\beta} \quad . \tag{4.7}$$

That is, the higher the β, the closer the proposed $E(\boldsymbol{x}')$ has to be from $E(\boldsymbol{x})$.

Flat-histogram When $\pi(\boldsymbol{x}) \propto 1/P(E(\boldsymbol{x}))$, the condition in Eq. 4.6 is given by

$$\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right] = \frac{1-a}{\frac{d\log P}{dE}(E(\boldsymbol{x}))} \quad . \tag{4.8}$$

which indicates that, in particular, when $E(\boldsymbol{x})$ is close to the maximum of $P(E)$, the derivative of $\log P$ approaches 0 and $E(\boldsymbol{x}')$ can be arbitrary distant from $E(\boldsymbol{x})$. On the other hand, as E deviates from the maximum of $P(E)$, smaller and smaller changes are necessary to guarantee a constant acceptance.

The condition in Eq. 4.6 was derived here in the context of chaotic systems, but it is valid for a Metropolis-Hastings algorithm on *any* problem. In particular, it has an interesting connection to traditional Monte Carlo used in spin systems. Eq. 4.8 implies that the proposed E' can be arbitrarily distant from E (arbitrary large number of spins can be flipped) when the configuration has an energy close to the maximum of the density of states.

4.2. Propose correlated trajectories

The proposal distribution requires correlating the trajectory starting at $\boldsymbol{x}'$ with a trajectory starting at $\boldsymbol{x}$ such that Eq. 4.6 holds. Fulfilling this requirement requires the ability to control $\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right]$,

which requires a procedure to correlate the state $\boldsymbol{x}'$ with $\boldsymbol{x}$. The aim of this section is to introduce a quantification of the correlation of two trajectories of finite-time t_o that can be related with $\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right]$, and present two different proposal distributions that propose a state $\boldsymbol{x}'$ on which this correlation is controlled.

The observables E introduced in chapter 3 are all dependent not only on $\boldsymbol{x}$, but also on the dynamics of $\boldsymbol{x}$.[2] Therefore, it is insightful to quantify the correlation of $\boldsymbol{x}$ and $\boldsymbol{x}'$ as a similarity of their respective trajectories. Trajectories that start at different but close positions $\boldsymbol{x}$ and $\boldsymbol{x}'$ are not independent of each other. For example, in Fig. 2.1, trajectories that are in the same box have the same symbolic sequence and the same FTLE. One natural way to quantify the similarity of two trajectories of finite-time t_o is to quantify for how long the two trajectories are close within a distance $\Delta < |\Gamma|$. Formally, this can be quantified by $t_\star(\boldsymbol{x}, \boldsymbol{x}') = \max\{t_\dagger(\boldsymbol{x}, \boldsymbol{x}')\}$, where $t_\dagger(\boldsymbol{x}, \boldsymbol{x}')$ is such that $|\boldsymbol{F}^t(\boldsymbol{x}) - \boldsymbol{F}^t(\boldsymbol{x}')| \leq \Delta \ , \ \forall t < t_\dagger \leq t_o$. The $t_\star = t_\star(\boldsymbol{x}, \boldsymbol{x}')$ fulfils $0 < t_\star(\boldsymbol{x}, \boldsymbol{x}') < t_o$, as $\boldsymbol{x}' = \boldsymbol{x}$ implies that $t_\star = t_o$, and $t_\star = 0$ when $\boldsymbol{x}'$ is far from $\boldsymbol{x}$. Yet, a trajectory starting from $\boldsymbol{x}' = \boldsymbol{F}(\boldsymbol{x})$ may not not fulfil $|\boldsymbol{x} - \boldsymbol{F}(\boldsymbol{x})| > \Delta$, but it is still very similar to the one starting from $\boldsymbol{x}$, see figure 4.1. Thus, two trajectories can be similar for two reasons: because one is a shift of the other, and because the starting states are close to each other. One way to include both two cases is to define $t_\star(\boldsymbol{x}, \boldsymbol{x}')$ as

$$t_\star(\boldsymbol{x}, \boldsymbol{x}') \equiv \max\left\{t_\dagger : |\boldsymbol{F}^t(\boldsymbol{x}') - \boldsymbol{F}^{t_{\text{shift}}+t}(\boldsymbol{x})| \leq \Delta \ , \ -t_o < t_{\text{shift}} < t_o, t + t_{\text{shift}} < t_\dagger \leq t_o \ , \ \forall t_{\text{shift}}, t\right\} \ . \tag{4.9}$$

For $t_{\text{shift}} = 0$, this recovers the case of two trajectories starting close to each other; $t_{\text{shift}} \neq 0$ includes situations where a trajectory starts at $\boldsymbol{x}'$ close to $\boldsymbol{F}^{t_{\text{shift}}}(\boldsymbol{x}) \neq \boldsymbol{x}$. This definition is also motivated by the concept of symbolic sequences introduced in sec. 2.4.1. The similarity of the trajectory starting from $\boldsymbol{x}'$ with the one starting from $\boldsymbol{x}$ can be quantified by the number $t_\star = t_\star(\boldsymbol{x}, \boldsymbol{x}')$ of symbols that both trajectories share, which corresponds to the $t_\star(\boldsymbol{x}, \boldsymbol{x}')$ in Eq. 4.9. The definition in Eq. 4.9 avoids the necessity of the existence of a phase-space partition, but, for the purposes of the argument below, "the two trajectories share a sequence of $t_\star$ states that are close within Δ" and "the two trajectories share a sequence of $t_\star$ symbols" are equivalent for a given unknown Δ.

The expected correlation between two states whose one is drawn according to a proposal distribution is here defined by

$$t_\star(\boldsymbol{x}) \equiv \mathbb{E}\left[t_\star(\boldsymbol{x}, \boldsymbol{x}')|\boldsymbol{x}\right] = \int_\Gamma g(\boldsymbol{x}'|\boldsymbol{x}) t_\star(\boldsymbol{x}, \boldsymbol{x}') d\boldsymbol{x}' \ . \tag{4.10}$$

This naturally depends on the observable and on the proposal distribution used. The aim of the next sections is to introduce two parametric proposals distributions on which $t_\star(\boldsymbol{x})$ can be computed explicitly.

[2] The escape time t_e requires that $\boldsymbol{F}^t(\boldsymbol{x}) \notin \Lambda \ , \forall t$, the FTLE λ_{t_o} is a sum of terms computed over the trajectory.

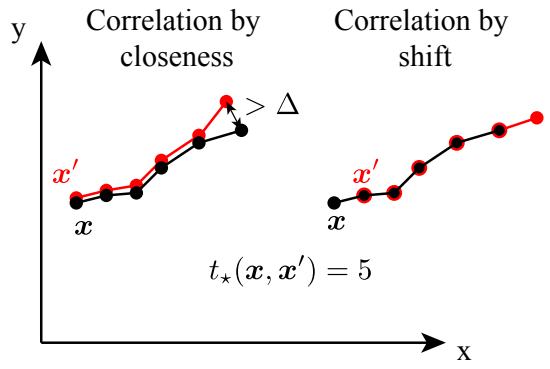

Figure 4.1.: Two trajectories starting at $\boldsymbol{x}$ and $\boldsymbol{x}'$ can be correlated up to $t_\star$ either by being close to each other or by one being a shift of the other. In both cases represented here, 5 of their states are close within Δ.

4.2.1. Shift proposals

One proposal that guarantees that trajectories are correlated by $t_\star$ is the shift proposal, originally introduced in Ref. [98] in the context of sampling paths of chemical reactions. It consists in proposing a state $\boldsymbol{x}'$ that is a forward or backward iteration of $\boldsymbol{x}$, $\boldsymbol{x}' = \boldsymbol{F}^{t_{\text{shift}}}(\boldsymbol{x})$, where t_{shift} is a free parameter. The relation between t_{shift} and $t_\star$ is that a shift of $\pm t_{\text{shift}}$ guarantees that $t_o - t_{\text{shift}}$ elements of the original trajectory are preserved. Therefore, this proposal guarantees that $t_\star$ elements are preserved when $|t_{\text{shift}}| = t_o - t_\star$, see Fig. 4.1. Because detailed balance has to be guaranteed in the sampling problem, backward and forward shifts must be equally likely. A proposal that fulfils detailed balance and guarantees that $t_\star$ symbols are preserved is therefore

$$g(\boldsymbol{x}'|\boldsymbol{x}) = \frac{1}{2}\delta\left(\boldsymbol{x}' - \boldsymbol{F}^{t_{\text{shift}}}(\boldsymbol{x})\right) + \frac{1}{2}\delta\left(\boldsymbol{x}' - \boldsymbol{F}^{-t_{\text{shift}}}(\boldsymbol{x})\right) \tag{4.11}$$

with $t_{\text{shift}} = t_o - t_\star(\boldsymbol{x})$.

This proposal unfortunately has different disadvantages: i) a priori there is no guarantee that $\boldsymbol{F}^{t_{\text{shift}}}(\boldsymbol{x}) \in \Gamma$. It is applicable when $\Gamma = \Omega$, which e.g. is not the case in open systems; ii) it requires the map to be invertible as per Eq. 4.11. iii) the proposal can only propose states that are forward or backward iterations of $\boldsymbol{x}$. Consequently, for the random walk to be ergodic in the phase-space, the map itself must be ergodic. iv) this proposal diffuses without drift on a trajectory passing through $\boldsymbol{x}$, by shifting the starting point forward or backward. Thus, it will always sample fewer states than a time average of a trajectory starting at $\boldsymbol{x}$. Nevertheless, as shown below, this proposal can be useful, when the conditions i)-iv) are fulfilled, to reduce correlations, as it performs non-local moves in the phase-space.

4.2.2. Neighbourhood proposals

Another strategy to construct a proposal on which on average the states are correlated by $t_\star(\boldsymbol{x})$ is to perturb $\boldsymbol{x}$ by a finite amount $\boldsymbol{\delta}$, $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{\delta}$, characterised by a direction $\hat{\boldsymbol{\delta}}$ and a norm δ, $\boldsymbol{\delta} \equiv \hat{\boldsymbol{\delta}}\delta$,

$g(\boldsymbol{x}'|\boldsymbol{x}) = \boldsymbol{x} + P(\boldsymbol{\delta}|\boldsymbol{x})$. A common case is when the probability distribution is separated in two independent terms [16]:

$$P(\boldsymbol{\delta}|\boldsymbol{x}) = P(\hat{\boldsymbol{\delta}}|\boldsymbol{x})P(\delta|\boldsymbol{x}) \tag{4.12}$$

and that $P(\hat{\boldsymbol{\delta}}|\boldsymbol{x})$ is uniformly distributed in the D directions and $P(\delta|\boldsymbol{x})$ has zero mean (i.e. an isotropic proposal). Consider this case here. Consider also that $P(\delta|\boldsymbol{x})$ is characterised by a well defined scale, e.g. it is an half-normal distribution[3] with mean $\delta_x(\boldsymbol{x})$:

$$P(\delta|\boldsymbol{x}) = \frac{\sqrt{2}}{\sqrt{\pi\delta_x(\boldsymbol{x})^2}} e^{-\frac{\pi\delta^2}{4\delta_x(\boldsymbol{x})^2}} \text{ for } \delta > 0 \quad . \tag{4.13}$$

This choice makes the ratio $g(\boldsymbol{x}|\boldsymbol{x}')/g(\boldsymbol{x}'|\boldsymbol{x})$ in Eq. 3.9 to be

$$\frac{g(\boldsymbol{x}|\boldsymbol{x}')}{g(\boldsymbol{x}'|\boldsymbol{x})} = \frac{\delta_x(\boldsymbol{x})}{\delta_x(\boldsymbol{x}')} \exp\left[-\frac{\pi|\boldsymbol{x}'-\boldsymbol{x}|^2}{4\delta_x(\boldsymbol{x})^2}\left(1-\frac{\delta_x(\boldsymbol{x})^2}{\delta_x(\boldsymbol{x}')^2}\right)\right] \tag{4.14}$$

The main motivation for this choice is that the proposal distribution is described by a single function, $\delta_x(\boldsymbol{x})$, that quantifies the distance $\boldsymbol{x}' - \boldsymbol{x}$, $\mathbb{E}\left[|\boldsymbol{x}' - \boldsymbol{x}||\boldsymbol{x}\right] = \delta_x(\boldsymbol{x})$.

The goal is now to relate $\delta_x(\boldsymbol{x})$ with $t_\star(\boldsymbol{x})$. In the limit $\delta_x(\boldsymbol{x}) \to 0$, the states are the same and therefore $\lim_{\delta_x(\boldsymbol{x})\to 0} t_\star(\boldsymbol{x}) = t_o$. In the limit $\delta_x(\boldsymbol{x}) \to |\Gamma|$, the proposal is equivalent to draw $\boldsymbol{x}'$ uniformly from Γ, and $\boldsymbol{x}'$ is independent of $\boldsymbol{x}$ and $t_\star = 0$. To preserve a correlation of $t_\star(\boldsymbol{x})$, it is necessary that $\delta_x(\boldsymbol{x})$ is such that the two trajectories starting at $\boldsymbol{x}$ and $\boldsymbol{x}'$ are close together up a time $t_\star(\boldsymbol{x})$. For small $\delta_x(\boldsymbol{x})$, two trajectories diverge exponentially in time according to Eq. 2.4, and, in particular, their maximal distance is given by Eq. 2.7. Therefore, to guarantee that two trajectories are distanced at most by Δ after a time $t_\star(\boldsymbol{x})$, $\delta_x(\boldsymbol{x})$ must be given by

$$\delta_x(\boldsymbol{x}) = \Delta e^{-\lambda_{t_\star}(\boldsymbol{x})t_\star(\boldsymbol{x})} \quad . \tag{4.15}$$

This equation is the main result of this section because it relates the parameter of the proposal distribution, $\delta_x(\boldsymbol{x})$, with the average correlation $t_\star(\boldsymbol{x})$ that the two states $\boldsymbol{x}$ and $\boldsymbol{x}'$ will have.

In 2008, ref. [14] proposed a proposal called "precision shooting" in the context of transition path sampling, that constructs a trajectory $\{\boldsymbol{x}'_i\}$ with $\boldsymbol{x}'_0 \equiv \boldsymbol{x}' = \boldsymbol{x} + \delta\hat{\boldsymbol{\delta}}$ (where δ is a free parameter) that is not a true trajectory but, within the numerical precision of a computer, is indistinguishable (in the system considered) from a trajectory starting at $\boldsymbol{x}'$ with δ small. In light of the discussion on shadowing in sec. 2.4.3, precision shooting constructs a trajectory $\{\boldsymbol{x}'_i\}$ that shadows the true trajectory starting at $\boldsymbol{x}'$, in the same spirit as the algorithm develop in Ref. [99] to construct a pseudo-trajectory. Assuming that shadowing holds, which seems to be the case in the system of the ref. [14], precision shooting proposal can therefore be interpreted as the neighbourhood proposal, Eq. 4.15, with $t_\star$ a free parameter (related to δ via Eq. 4.15), that takes advantage of shadowing theorem to simplify the construction $\boldsymbol{x}'$.

[3] An exponential distribution would be equally acceptable and would not change the main conclusions.

Ref. [14] presents also an insightful discussion on how, in the system and observable considered, the acceptance rate depends on δ, suggesting that the acceptance rate increases with decreasing δ (Fig. 9 of the ref.). In light of the discussion in section 4.2, this result is interpreted as follows: as δ decreases, $\boldsymbol{x}'$ becomes more correlated with $\boldsymbol{x}$ (since $t_\star$ is related with δ by Eq. 4.15), and therefore the acceptance is expected to increase, as indicated in Fig. 9 of Ref. [14]. This is however insufficient to the goal of correlating the states according to the condition in Eq. 4.6 because it does not allow to derive δ (or $t_\star$) that fulfils the condition.

The crucial advantage of Eq. 4.15 and Eq. 4.6 is that they allow to relate $\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right]$ (in Eq. 4.6) with $\boldsymbol{x}' - \boldsymbol{x}$ (in Eq. 4.15) through $t_\star$ (in Eq. 4.9), which is the aim of the next section.

4.3. Guarantee local proposals

Now that the tools to enforce the preservation of a given correlation $t_\star(\boldsymbol{x})$ are available, the only missing piece is to relate $t_\star(\boldsymbol{x})$ with $\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right]$ in Eq. 4.6. Because the computation of $t_\star(\boldsymbol{x})$ depends on the particular observable E, a different derivation is presented for two observables discussed in Chapter 2, t_e and λ_{t_o}. The observable $\log t_e$ used in weakly chaotic systems is discussed in chapter 7. Given the limitations of the shift proposal described above, the argumentation below is limited to neighbourhood proposals. [4]

4.3.1. FTLE in closed systems

As introduced in section 2.5, the observable in this case is given by $E(\boldsymbol{x}) = t_o\lambda_{t_o}(\boldsymbol{x})$, where t_o is the observation time. The aim in this case is thus to write $\mathbb{E}\left[t_o\lambda_{t_o}(\boldsymbol{x}') - t_o\lambda_{t_o}(\boldsymbol{x})|\boldsymbol{x}\right]$ as a function of $t_\star(\boldsymbol{x})$. The finite-time Lyapunov exponent considered in Eq. 2.8 is a sum of t_o terms and, thus, it can be written as

$$t_o\lambda_{t_o}(\boldsymbol{x}) = t_\star\lambda_{t_\star}(\boldsymbol{x}) + (t_o - t_\star)\lambda_{t_o-t_\star}(\boldsymbol{x}_{t_\star}) \quad , \tag{4.16}$$

where $\boldsymbol{x}_{t_\star} \equiv \boldsymbol{F}^{t_\star}(\boldsymbol{x})$ ($\boldsymbol{x}'_{t_\star} \equiv \boldsymbol{F}^{t_\star}(\boldsymbol{x}')$), and the first term represents the divergence up to time $t_\star$, and the second term divergence from $t_\star$ to t_o. Likewise, the trajectory $\boldsymbol{x}'$ has a finite-time Lyapunov exponent given by

$$t_o\lambda_{t_o}(\boldsymbol{x}') = t_\star\lambda_{t_\star}(\boldsymbol{x}') + (t_o - t_\star)\lambda_{t_o-t_\star}(\boldsymbol{x}'_{t_\star}) \quad . \tag{4.17}$$

Because $\boldsymbol{x}'$ is proposed according to Eq. 4.15, by construction, the first $t_\star$ states of the trajectory starting at $\boldsymbol{x}'$ are close (within Δ) to the states of the trajectory starting at $\boldsymbol{x}$ up to $t_\star(\boldsymbol{x})$. Therefore, one can approximate that the respective Lyapunovs up to time $t_\star$ are equal,

$$\mathbb{E}\left[\lambda_{t_\star}(\boldsymbol{x}')|\boldsymbol{x}\right] \approx \lambda_{t_\star}(\boldsymbol{x}) \quad . \tag{4.18}$$

[4] In principle an equivalent argumentation can be made to the shift proposal, and it is possible that the assumptions are equivalent to the ones used here. However, this was not analysed in detailed here.

Subtracting Eq. 4.16 from Eq. 4.17 and using Eq. 4.18 gives

$$\mathbb{E}\left[t_o\lambda_{t_o}(\boldsymbol{x}') - t_o\lambda_{t_o}(\boldsymbol{x})|\boldsymbol{x}\right] = (t_o - t_\star)\left(\mathbb{E}\left[\lambda_{t_o-t_\star}(\boldsymbol{x}'_{t_\star})|\boldsymbol{x}\right] - \lambda_{t_o-t_\star}(\boldsymbol{x}_{t_\star})\right) \tag{4.19}$$

The left side of this equation is the same as in Eq. 4.6 and thus the aim now is to write the right side as a function of properties of the system. Let us focus on the calculation of $\mathbb{E}\left[\lambda_{t_o-t_\star}(\boldsymbol{x}'_{t_\star})|\boldsymbol{x}\right]$ first. By construction, $\boldsymbol{x}'$ is generated such that $|\boldsymbol{x}'_{t_\star} - \boldsymbol{x}_{t_\star}| \approx \Delta$. Because the system is chaotic, one can approximate that $\boldsymbol{x}'_{t_\star}$ is sufficiently separated from $\boldsymbol{x}_{t_\star}$ such that $\lambda_{t_o-t_\star}(\boldsymbol{x}'_{t_\star})$ is independent of $\boldsymbol{x}$ and given by the mean of the distribution of FTLE at time $t_o - t_\star$, i.e.

$$\mathbb{E}\left[\lambda_{t_o-t_\star}(\boldsymbol{x}'_{t_\star})|\boldsymbol{x}\right] = \lambda_{t_o-t_\star} \approx \lambda_L \tag{4.20}$$

where the second approximation is valid for large $t_o - t_\star$. Let us now focus on $\lambda_{t_o-t_\star}(\boldsymbol{x}_{t_\star})$ in Eq. 4.19. It can be simplified under the approximation that, on average, the two terms of Eq. 4.16 contribute equally to $\lambda_{t_o}(\boldsymbol{x})$, such that each of them is, on average, equal to $\lambda_{t_o}(\boldsymbol{x})$, i.e.

$$\lambda_{t_o-t_\star}(\boldsymbol{x}_{t_\star}) \approx \lambda_{t_o}(\boldsymbol{x}) \quad . \tag{4.21}$$

Replacing Eq. 4.21 and Eq. 4.20 in Eq. 4.19 gives

$$\mathbb{E}\left[t_o\lambda_{t_o}(\boldsymbol{x}') - t_o\lambda_{t_o}(\boldsymbol{x})|\boldsymbol{x}\right] = (t_o - t_\star(\boldsymbol{x}))\left(\lambda_L - \lambda_{t_o}(\boldsymbol{x})\right) \quad . \tag{4.22}$$

Before proceeding, let us confirm that this equation is intuitively consistent: when $t_\star(\boldsymbol{x}) = t_o$, the two trajectories are indistinguishable and therefore the respective FTLE should be the same, which is indeed obtained from the equation. When $t_\star(\boldsymbol{x}) = 0$, the two trajectories are independent and therefore the proposed state should have a FTLE given approximately by λ_L, which is also obtained from the equation. This equation is now ready to be plugged in the condition of constant acceptance. Replacing the left side of Eq. 4.6 by the expectation in Eq. 4.22 and solving to $t_\star(\boldsymbol{x})$ gives

$$t_\star(\boldsymbol{x}) = t_o - \frac{a-1}{d\log\pi(\lambda_t(\boldsymbol{x}))/d\lambda_t}\frac{1}{\lambda_L - \lambda_{t_o}(\boldsymbol{x})} \quad , \tag{4.23}$$

valid for $t_\star \in [0, t_o]$. It does not make sense for $t_\star(\boldsymbol{x})$ to be smaller than 0 as it would imply arbitrarily large proposals, that would have to be re-mapped to the constraining region Γ. Likewise, $t_\star$ should also never be larger than t_o as it physically does not make sense. For these reasons, in practice $t_\star$ should be given by

$$t_\star(\boldsymbol{x}) = \max\left\{0, t_o - \left|\frac{a-1}{d\log\pi(\lambda_t(\boldsymbol{x}))/d\lambda_t}\frac{1}{\lambda_L - \lambda_{t_o}(\boldsymbol{x})}\right|\right\} \quad , \tag{4.24}$$

which is the main result of this section. This equation provides an expression to $t_\star(\boldsymbol{x})$ that can be inserted in the parameter of the proposal distribution, Eq. 4.15, that under the approximations used above achieves a constant acceptance rate.

One of the most interesting aspects of Eq. 4.24 is that when $\lambda_{t_o}(\boldsymbol{x})$ is close to maximum of $\pi(E)$, λ_L, $t_\star(\boldsymbol{x}) \approx 0$ and $|\boldsymbol{x}' - \boldsymbol{x}| \gg 1$, i.e. the state $\boldsymbol{x}'$ is approximately drawn uniformly on Γ. Intuitively, this is the proposal that minimises the correlations of $\boldsymbol{x}'$ with $\boldsymbol{x}$ because when the algorithm is close to the maximum of the $\pi(E)$, there is no need to correlate the two states as any random proposal is accepted. Another interesting aspect of Eq. 4.24 is that it reduces the proposal distribution to a uniform distribution when the sampling distribution is the uniform distribution: $d\log\pi(E)/dE(E) = 0$ implies $t_\star(\boldsymbol{x}) = 0$.

In Ref. [84] (from the same author of this thesis), the $t_\star$ used was different from the one in Eq. 4.24, namely, it was $t_\star = t_o - 1$. There are two differences between the derivation above and the derivation on Ref. [84]. The first difference is that Ref. [84] used the assumption that Monte Carlo required $\delta E \approx 1$. The approximation $d\log\pi/dE \approx 1$ is common in Monte Carlo because $d\log\pi/dE$ does not depend on N (both $\log\pi$ and E increase linearly with N). Eq. 4.6 does not use this approximation and instead presents the un-approximated value of $d\log\pi/dE$, which reveals what exactly the proposal should aim for. The second difference is that the argumentation leading to Eq. 4.24 had not been formalised in Ref. [84]; instead, a simpler approximation based on changing the last term of the sum of λ_{t_o} was used (see argumentation after Eq. (14)).

This generalisation solves an important issue originally present in Refs. [14, 84]. There, it was argued that the shift proposal was required in order to guarantee ergodicity of the Monte Carlo process. This was because, in both references, Eq. 4.15 was used with $t_\star$ replaced by t_o and a fixed Δ. This implies that the size of the proposal $\delta_x(\boldsymbol{x})$ is always much smaller (e.g. 10^{-10} smaller) than the size of the constraint region $|\Gamma|$. Consequently, the random walk would only move on a very small region of Γ. In both references, the shift proposal was used to avoid such situation. However, as discussed in section 4.2.1, the applicability of shift is limited. An advantage of Eq. 4.24 is that it allows proposals that guarantee ergodicity without using the shift proposal: close to the maximum of the density of states, the proposal naturally covers all states $\boldsymbol{x} \in \Gamma$.

Finally, the derivation of Eq. 4.24, in particular Eqs. 4.16 and 4.19, only required λ_{t_o} to be an average over a trajectory. Therefore, equation Ref. 4.24 can be generalised as follows: consider

$$e_{t_o}(\boldsymbol{x}) \equiv \frac{1}{t_o}\sum_{i=1}^{t_o} f(\boldsymbol{x}_i) = E_{t_o}(\boldsymbol{x})/t_o \tag{4.25}$$

where $f(\boldsymbol{x}_i)$ is an arbitrary function of the phase-space (the logarithm of the derivative of the map corresponds to $e_{t_o}(\boldsymbol{x}) = \lambda_{t_o}(\boldsymbol{x})$, $E_{t_o}(\boldsymbol{x}) = \lambda_{t_o}(\boldsymbol{x})t_o$). Replacing this quantity in the derivation of Eq. 4.24 mutatis mutandis and without using the approximation in Eq. 4.21, one obtains

$$t_\star(\boldsymbol{x}) = \max\left\{0, t_o - \left|\frac{a-1}{d\log\pi(E_{t_o}(\boldsymbol{x}))/dE}\frac{1}{E^* - E_{t_o - t_\star}(\boldsymbol{x})}\right|\right\} \quad , \tag{4.26}$$

where E^* is approximately the maximum of the distribution $P(E_{t_o})$ (using the approximation in Eq. 4.20). This generalizes Eq. 4.24 for an arbitrary average over trajectories of size t_o, $e_{t_o}(\boldsymbol{x})$, and it

should be useful to sample rare states in respect to this observable.[5]

4.3.2. Escape time in strongly chaotic open systems

As introduced in Chapter 2, in strongly chaotic open systems $E(\boldsymbol{x}) = t_e(\boldsymbol{x})$ and $P(t_e) \sim \exp(-\kappa t_e)$. The aim here is to compute $t_\star(\boldsymbol{x})$ that fulfils Eq. 4.6 by taking into account that $\boldsymbol{x}$ is given and $\boldsymbol{x}' = \boldsymbol{x} + \hat{\boldsymbol{\delta}}\delta_x(\boldsymbol{x})$ with $\delta_x(\boldsymbol{x})$ given by Eq. 4.15. Note that in this case the trajectory's length, t_o, is not a constant because the system is open. Instead, the trajectory's length is given by $t_e(\boldsymbol{x})$.

A trajectory starting at $\boldsymbol{x}'$ proposed according to Eq. 4.15 fulfils $|\boldsymbol{x}'_{t_\star} - \boldsymbol{x}_{t_\star}| \leq \Delta$. Therefore, up to $t_\star$, the two trajectories are indistinguishable. Under this assumption, and because from the definition of $t_\star$ in Eq. 4.9, $t_\star(\boldsymbol{x}) \leq t_e(\boldsymbol{x})$, $t_e(\boldsymbol{x}') = t_\star + t_e(\boldsymbol{x}'_{t_\star})$ and therefore

$$\mathbb{E}\left[t_e(\boldsymbol{x}')|\boldsymbol{x}\right] = t_\star(\boldsymbol{x}) + \mathbb{E}\left[t_e(\boldsymbol{x}'_{t_\star})|\boldsymbol{x}\right] \quad . \tag{4.27}$$

Before proceeding, let us take a moment to describe $t_e(\boldsymbol{x}'_{t_\star})$ and $\mathbb{E}\left[t_e(\boldsymbol{x}'_{t_\star})|\boldsymbol{x}\right]$. The $t_e(\boldsymbol{x}'_{t_\star})$ is the escape time of the state $\boldsymbol{x}'$ iterated $t_\star$ times and that has not escaped up to time $t_\star$. The $\mathbb{E}\left[t_e(\boldsymbol{x}'_{t_\star})|\boldsymbol{x}\right]$ is, from the definition (Eq. 4.4),

$$\mathbb{E}\left[t_e(\boldsymbol{x}'_{t_\star})|\boldsymbol{x}\right] = \int_\Omega g(\boldsymbol{x}'|\boldsymbol{x}) t_e(\boldsymbol{x}'_{t_\star}) d\boldsymbol{x}' \quad . \tag{4.28}$$

That is, $\mathbb{E}\left[t_e(\boldsymbol{x}'_{t_\star})|\boldsymbol{x}\right]$ is an expectation over a neighbourhood of $\boldsymbol{x}$ of size $\delta_x(\boldsymbol{x})$ given by Eq. 4.15 of $t_e(\boldsymbol{x}'_{t_\star})$. In other words, $\mathbb{E}\left[t_e(\boldsymbol{x}'_{t_\star})|\boldsymbol{x}\right]$ is an integration over a small effective circle that isotropically "zooms" around $\boldsymbol{x}$ with a length $\delta_x(\boldsymbol{x})$. Let us now analyse in detail the properties of the function $t_e(\boldsymbol{x})$ in the open tent map defined by Eq. 2.24, and see how $\mathbb{E}\left[t_e(\boldsymbol{x}'_{t_\star})|\boldsymbol{x}\right]$ can be approximated in this case. The escape time $t_e(\boldsymbol{x})$ of this map is shown in Fig. 4.2, on which every interval can be identified and its size, that can be computed analytically from a and b. Specifically, a given interval has size $\varepsilon(x) = (1 - 1/a - 1/b)a^{-i(x)}b^{-t_e(x)+i(x)}$, where $i(x)$ is the number of times $0 < F^t(x) < 1/a$, for $t = 1, ..., t_e(x)$. The crucial observation is that $\lambda_{t_e(x)}(x)$ is proportional to $\log(\varepsilon(x))/t_e(x)$. Thus, proposing according to Eq. 4.15 is equivalent to "zoom" the landscape of $t_e(\boldsymbol{x})$ around $\boldsymbol{x}$ with a scale correspondent to $t_\star$'th iteration of the construction of the landscape. Under this zoom, the landscape of $t_e(\boldsymbol{x}_{t_\star})$ is equal to the landscape of $t_e(\boldsymbol{x})$. and therefore $\mathbb{E}\left[t_e(\boldsymbol{x}'_{t_\star})|\boldsymbol{x}\right]$ should be a constant independent of t_e. The value of this constant can be computed by assuming that the escape time of trajectories $\boldsymbol{x}'$ nearby a state $\boldsymbol{x}$ does not depend on the state $\boldsymbol{x}$ but only on the condition that $\boldsymbol{x}'$ did not exit up to time $t_\star$, denoted as $\star$. This approximation allows to write

$$\mathbb{E}\left[t_e(\boldsymbol{x}'_{t_\star})|\boldsymbol{x}\right] = \mathbb{E}\left[t_e(\boldsymbol{x}'_{t_\star})|\star\right] \tag{4.29}$$

[5] One example of such an observable is the average magnetisation of a trajectory of the Hamiltonian mean-field (HMF) model [100].

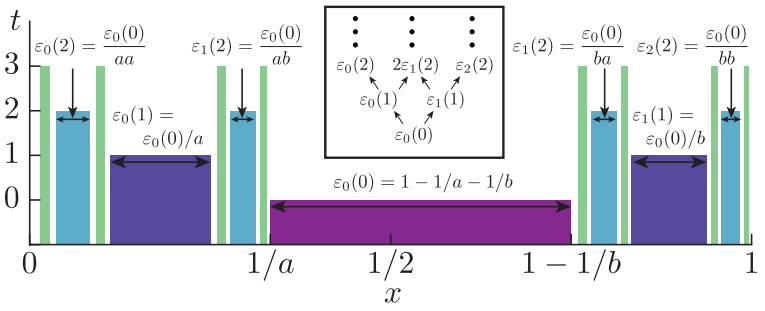

Figure 4.2.: The escape time function, $t_e(x)$, of the tent map for a generic a and b. There are 2^t intervals, and the size of each plateau can be analytically computed from the one at a previous time, and therefore be written analytically.

and inserting Eq. 4.29 in Eq. 4.27 gives

$$\mathbb{E}\left[t_e(\boldsymbol{x}')|\boldsymbol{x}\right] = t_\star(\boldsymbol{x}) + \mathbb{E}\left[t_e(\boldsymbol{x}'_{t_\star})|\star\right] \quad . \tag{4.30}$$

The idea behind this equation is represented in Fig. 4.3. A small ball of size $\delta_x(0)$ around $\boldsymbol{x}$ (the proposal density $g(\boldsymbol{x}'|\boldsymbol{x})$) evolves in time according to the dynamics of the system, being stretched along the unstable directions and compressed along the stable directions. The most unstable direction eventually reaches the size Δ by the time $t_\star(\boldsymbol{x})$ ($t_\star(\boldsymbol{x}) = t_e(\boldsymbol{x}) - 1$ is exemplified in the figure), after which the trajectory starting at $\boldsymbol{x}'$ differentiates from the one starting from $\boldsymbol{x}$. In particular, $\boldsymbol{x}'_t$ will be approximately independent of $\boldsymbol{x}$. Therefore, $\mathbb{E}\left[t_e(\boldsymbol{x}'_{t_\star})|\star\right]$ is the average escape time of an independent state $\boldsymbol{x}'_{t_\star}$, which is the average of $P(t_e)$ and is given by $1/\kappa$:

$$\mathbb{E}\left[t_e(\boldsymbol{x}'_{t_\star})|\star\right] = 1/\kappa \quad . \tag{4.31}$$

Replacing Eq. 4.31 in Eq. 4.30 gives

$$\mathbb{E}\left[t_e(\boldsymbol{x}')|\boldsymbol{x}\right] = t_\star(\boldsymbol{x}) + 1/\kappa \quad . \tag{4.32}$$

Subtracting $t_e(\boldsymbol{x})$ on both sides of Eq. 4.32 gives

$$\mathbb{E}\left[t_e(\boldsymbol{x}') - t_e(\boldsymbol{x})|\boldsymbol{x}\right] = t_\star(\boldsymbol{x}) + \frac{1}{\kappa} - t_e(\boldsymbol{x}) \quad . \tag{4.33}$$

The left side is equation is the term appearing in the condition of constant acceptance rate, Eq. 4.6. Equating both left sides and solving for $t_\star(\boldsymbol{x})$ gives

$$t_\star(\boldsymbol{x}) = t_e(\boldsymbol{x}) - \frac{1}{\kappa} - \frac{a-1}{d\log\pi(t_e)/dt_e} \quad , \tag{4.34}$$

which is the central result of this section.

Lets take a moment to re-visit the full argument of this section: (A) the Metropolis-Hastings requires a proposal that guarantees a specific variation of t_e given by Eq. 4.6. (B) the proposal with

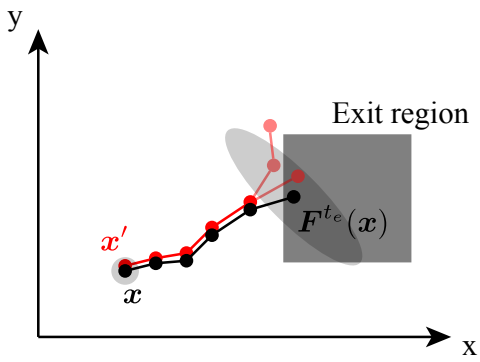

Figure 4.3.: A trajectory starting close to $\boldsymbol{x}$ and correlated for a time $t_\star = 4$ can leave at a time $t_\star > 4$. An initial ball around $\boldsymbol{x}$ (black) expands along the unstable direction according proportionally to $\exp(t\lambda_t(\boldsymbol{x}))$. The trajectory $\boldsymbol{x}$ leaves at time $t_e(\boldsymbol{x})$, but it is possible to find a trajectory starting at $\boldsymbol{x}'$ (red) to leave at a higher (or lower) time by controlling for how long ($t_\star$) it stays close to the trajectory $\boldsymbol{x}$. Once the trajectories are sufficiently apart, the escape time of the state $\boldsymbol{F}^{t_\star}(\boldsymbol{x}')$ is independent of $\boldsymbol{x}$.

a scale given by Eq. 4.15 guarantees that on average the trajectory starting at $\boldsymbol{x}'$ stays close to the trajectory starting at $\boldsymbol{x}$ up to $t_\star(\boldsymbol{x})$. (C) The variation of t_e is related to $t_\star(\boldsymbol{x})$ via Eq. 4.33. (A), (B) and (C) leads to Eq. 4.34. This equation is one of the main results of this thesis because, along with Eq. 4.15, it defines the proposal distribution required to have an average constant acceptance in open fully chaotic systems.

When $d\log\pi(t_e)/dt_e$ is a constant χ, which is the case in the canonical ensemble ($\chi = \beta$) and asymptotically in the flat-histogram ($\chi = \kappa$), $\delta_x(\boldsymbol{x})$ can be re-written as

$$\delta_x(\boldsymbol{x}) = \delta_0 e^{-\lambda_{t_e}(\boldsymbol{x})t_e(\boldsymbol{x})} \tag{4.35}$$

where $\delta_0 \equiv \Delta\exp(-\lambda_t(1/\kappa + (a-1)/\chi)$ is a constant and $\lambda_{t_\star}(\boldsymbol{x})$ was approximated by $\lambda_{t_e}(\boldsymbol{x})$ for the same argument that led to Eq. 4.21. The expression in Eq. 4.35 generalizes the result in Ref. [83]. There, it is argued that an isotropic proposal as $\delta_x(\boldsymbol{x}) = \delta_0\exp(-\lambda_L t_e(\boldsymbol{x}))$ where δ_0 is a free parameter guarantees a constant acceptance. In light of Eq. 4.35, $\delta_x(\boldsymbol{x})$ used in Ref. [83] uses the approximation $\lambda_{t_e}(\boldsymbol{x}) \approx \lambda_L$, which is a valid when $\lambda_t(\boldsymbol{x})$ is close to the maximum of the density of the FTLE. Because the constant δ_0 is independent of t_e, it does not influence how the acceptance depends on t_e, which is consistent to why in Ref. [83] a constant acceptance ratio is observed.

4.4. Summary

At this moment it is worth taking a moment to summarise the main results of this chapter, which constitute the core theoretical part of this thesis. This chapter contains a set of explicit formulas to perform efficient Monte Carlo simulations in chaotic systems. These formulas do not make assumptions about the sampling distribution $\pi(\boldsymbol{x})$, reduce to the uniform proposal distribution when the sampling distribution is itself uniform, and only depend on properties of the system such as the Lyapunov

exponent. They generalize previous results in the literature, Refs. [14, 83, 84].

The analysis that led to the explicit formulas for the correlation time $t_\star(\boldsymbol{x})$ are also valid for the problem of *finding rare states* introduced in Sec. 2.7. Specifically, the formulas were derived for the condition in Eq. 4.6, which dictates how different $\mathbb{E}\left[E(\boldsymbol{x}')|\boldsymbol{x}\right]$ has to be from $E(\boldsymbol{x})$ to guarantee a constant acceptance. This condition is stronger than the condition required for an algorithm to find minima or maxima of E, which requires only proposing states $\boldsymbol{x}'$ such that $\mathbb{E}\left[E(\boldsymbol{x}')|\boldsymbol{x}\right] > E(\boldsymbol{x})$ (or vice-versa for minimising E). This result is independent of the particular minimisation algorithm (e.g. stimulated annealing, step descent, stagger and dagger in open systems) because it only discusses which new state $\boldsymbol{x}'$ should be *tried*, given the current state $\boldsymbol{x}$. Therefore, the results above should also be valid to efficiently find rare states, such as the ones found in Refs. [6, 9, 10, 101].

It is useful to summarise the different proposals in an algorithmic form so they can be easily referred to. In all cases, the proposal requires the current state of the random walk, denoted by $\boldsymbol{x}$.

FTLE in closed systems

1. Generate a unitary vector $\hat{\boldsymbol{\delta}}$ in D dimensions
2. Compute $t_\star = t_o - \frac{a-1}{d\log\pi(E)/dE(E)}\frac{1}{|\lambda_L - \lambda_{t_o}(\boldsymbol{x})|}$, Eq. 4.24.
3. Compute $\delta_x = \Delta\exp(-t_\star\lambda_{t_\star}(\boldsymbol{x}))$, Eq. 4.15
4. Generate a random number δ from a normal distribution with mean 0 and variance δ_x^2
5. Make $\boldsymbol{x}' = \boldsymbol{x} + \hat{\boldsymbol{\delta}}|\delta|$

where $\Delta \in \mathbb{R}$ is free parameter (e.g. $\Delta = 0.1$) and a is the chosen average acceptance (e.g. 0.5). For example, in a canonic ensemble with parameter β, $d\log\pi(E)/dE(E) = \beta$ and therefore $t_\star = t_e(\boldsymbol{x}) - \frac{a-1}{\beta}\frac{1}{|\lambda_L - \lambda_{t_o}(\boldsymbol{x})|}$. The value of λ_L can be estimated using e.g. the first samples of the random walk. In the flat-histogram ensemble, λ_L is the maximum of $P(\lambda_{t_o})$ and $\log\pi(E)/dE(E)$ is given by the state density, or by an approximation of it, e.g. $\log P_{WL}$ of the Wang-Landau algorithm.

Fully chaotic open systems

1. Generate a unitary vector $\hat{\boldsymbol{\delta}}$ in D dimensions
2. Compute $t_\star = t_e(\boldsymbol{x}) - 1/\kappa - \frac{a-1}{d\log\pi/dt_e}$, Eq. 4.34
3. Compute $\delta_x = \delta_0\exp(-t_e(\boldsymbol{x})\lambda_{t_e(\boldsymbol{x})}(\boldsymbol{x}))$, Eq. 4.35
4. Generate a random number δ from a normal distribution with mean 0 and variance δ_x^2
5. Make $\boldsymbol{x}' = \boldsymbol{x} + \hat{\boldsymbol{\delta}}|\delta|$

where δ_0 is a free parameter (e.g. $\delta_0 = 0.1$). Both $\lambda_{t_e}(\boldsymbol{x})$ and $t_e(\boldsymbol{x})$ are required by the proposal and both can be computed during the same evolution of the system: $t_e(\boldsymbol{x})$ is the time until the trajectory enters the exit region Λ, $\lambda_{t_e}(\boldsymbol{x})$ is the FTLE of this trajectory.

4.5. Simplifications

This section presents approximations that can be used to simplify both the implementation time and the computational cost of the proposals derived in the previous section.

4.5.1. Propose with the Lyapunov exponent in fully chaotic open systems

For a fixed sampling distribution $\pi(\boldsymbol{x})$, the proposal in Eq. 4.35 requires computing two quantities from $\boldsymbol{x}$ and $\boldsymbol{x}'$: $t_e(\boldsymbol{x})$ and $\lambda_{t_e(\boldsymbol{x})}(\boldsymbol{x})$. The calculation of $\lambda_{t_e(\boldsymbol{x})}(\boldsymbol{x})$ requires either using a numerical algorithm or multiplying a product of matrixes [17], both of which have an associated computational cost. A simplification to this proposal is to approximate $\lambda_{t_e(\boldsymbol{x})}(\boldsymbol{x})$ by the maximum of the distribution of FTLE with finite-time t_e, $\lambda_L(t_e)$,

$$\lambda_{t_e(\boldsymbol{x})}(\boldsymbol{x}) \approx \lambda_L(t_e(\boldsymbol{x})) \quad . \tag{4.36}$$

This approximation is valid as long as $\lambda_{t_e}(\boldsymbol{x})$ is not on the tails of the distribution of FTLE with finite-time t_e, $P(\lambda_{t_e})$. A sampling distribution that only depends on t_e, $\pi(\boldsymbol{x}) = \pi(t_e(\boldsymbol{x}))$, guarantees that states with the same t_e are equally likely, $P(\boldsymbol{x}|t_e) = P(\boldsymbol{x}) = \pi(\boldsymbol{x})$, and therefore $P(\lambda_{t_e}(\boldsymbol{x})|t_e) = P(\lambda_{t_e})$. The approximation of using the maximum of the distribution holds because the tails of $P(\lambda_{t_e})$ decay exponentially with increasing t_e (see sec. 2.5). Under these approximations, Eq. 4.35 can be simplified to

$$\delta_x(t_e(\boldsymbol{x})) = \delta_0 e^{-\lambda_L(t_e(\boldsymbol{x}))t_e(\boldsymbol{x})} \quad . \tag{4.37}$$

Furthermore, $\lambda_L(t_e)$ converges to the Lyapunov exponent of the system λ_L with increasing t_e. Therefore, a further simplification is to use the Lyapunov exponent of the system instead of $\lambda_L(t_e)$ in Eq. 4.37,

$$\delta_x(t_e(\boldsymbol{x})) = \delta_0 e^{-\lambda_L t_e(\boldsymbol{x})} \quad . \tag{4.38}$$

These simplifications, first derived in Ref. [83], avoid computing $\lambda_{t_e}(\boldsymbol{x})$ on every proposal step, and instead require just an estimation of either $\lambda_L(t_e)$ or even just λ_L.

Adaptively estimate the Lyapunov exponent

Using the proposal distribution with λ_L requires a priori knowledge of it, which typically is not available. This difficulty resembles the same problem that flat-histogram simulations have: $P(E)$ is required, but it is typically unknown a priori. This analogy motivates a Monte Carlo procedure that on the fly computes $\delta_x(t)$ that scales with λ_L.

Consider an hypothetical simulation with an isotropic proposal distribution (Eq. 4.13) with

$$\delta_x(\boldsymbol{x}) = \sigma(t_e(\boldsymbol{x})) \quad , \tag{4.39}$$

where $\sigma(t)$ is initially set to be $\sigma(t) = 1$ for every t. Consider also that the simulation reached a state $\boldsymbol{x}$ with a high escape time (e.g. $t_e = t_e(\boldsymbol{x}) = 10/\kappa$). A proposed state, $\boldsymbol{x}' = \boldsymbol{x} + \hat{\boldsymbol{h}}\sigma(t_e)$, will most likely have a much lower escape time (e.g. $t_e(\boldsymbol{x}') = 1/\kappa$). From Eq. 4.15 and Eq. 4.34, this indicates that $\sigma(t_e)$ is much higher than the "correct" proposal, $\delta_x(\boldsymbol{x})$, and therefore it should be reduced in the next proposal. The opposite is also true: when $\sigma(t_e)$ is much smaller than $\delta_x(\boldsymbol{x})$, $t_e(\boldsymbol{x}') = t_e(\boldsymbol{x})$ and it should be increased. This hypothetical simulation suggests that, in the same spirit as the Wang-Landau algorithm to approximate the density $P(t_e)$, there is the possibility to approximate $\delta_x(\boldsymbol{x})$ using an update scheme that can be inserted in the Metropolis-Hastings algorithm, and that is given by the same algorithm as the Wang-Landau (see sec. 3.3.3), but instead of updating $P_{WL}(t)$, it updates also $\sigma(t)$ [83]:

$$\sigma(t_e) = \begin{cases} \sigma(t_e)f & \text{for } t_e(\boldsymbol{x}') = t_e \\ \sigma(t_e)/f & \text{for } t_e(\boldsymbol{x}') < t_e \end{cases} . \tag{4.40}$$

This update scheme, which generalises the Wang-Landau procedure to the proposal distribution, was here extensively tested in different systems (tent map, full chaotic standard map with a leak, Coupled Hénon map with different Ds). It converged to a function $\sigma(t)$ that decays exponentially with the Lyapunov exponent of the system, and a proposal distribution with a constant acceptance rate in a flat-histogram simulation. The result of one of such simulations is shown in Fig. 4.4.

4.5.2. Power-law proposal distribution

The proposal distributions derived in the previous sections requires some knowledge about the state and the system: $\lambda_{t_o}(\boldsymbol{x})$, $t_e(\boldsymbol{x})$ (in open systems), λ_L of the system, and, in some situations, $P(E(\boldsymbol{x}))$. Such knowledge, required to compute $\delta_x(\boldsymbol{x})$ (which requires $t_\star$), may not be available or it may be computationally expensive to obtain. One alternative to this lack of knowledge is to consider a proposal on which the correlation time $t_\star$ is not imposed by $t_\star(\boldsymbol{x})$, but is a uniformly random variable between $[0, t_o]$ (FTLE) or $[0, t_e(\boldsymbol{x})]$ (open systems) that is generated on each proposal. Some values of $t_\star$ will be far from the optimal $t_\star(\boldsymbol{x})$ and the corresponding $\boldsymbol{x}'$ will be rejected or it will be too close from $\boldsymbol{x}$, but others $t_\star$ will still be close from the optimal $t_\star(\boldsymbol{x})$ and therefore useful.

Having a uniformly distributed correlation $t_\star$ still requires computing $\delta_x(\boldsymbol{x})$ in Eq. 4.15, which requires $\lambda_{t_o}(\boldsymbol{x})$. In the case $\lambda_t(\boldsymbol{x})$ is unknown (e.g. in open systems one could be only interested in the escape time and therefore not compute $\lambda_t(\boldsymbol{x})$), one may further approximate it by an uniform distribution between two extremes. Because the product of two uniformly distributed random variables is also a uniform random variable, this leads to the proposal distribution where $\delta_x(\boldsymbol{x})$ is given by $\exp(-U(s_{\min}, s_{\max}))$ where $s_{\min}$ and $s_{\max}$ are two free parameters. This leads to a scale $\delta_x(\boldsymbol{x})$ that is power-law distributed and given by

$$P(\delta_x|\boldsymbol{x}) = P(\delta_x) = \frac{1}{\delta_x}\frac{1}{s_{\max} - s_{\min}}, \quad \delta_x \in [\delta_{\min}, \delta_{\max}] \ , \tag{4.41}$$

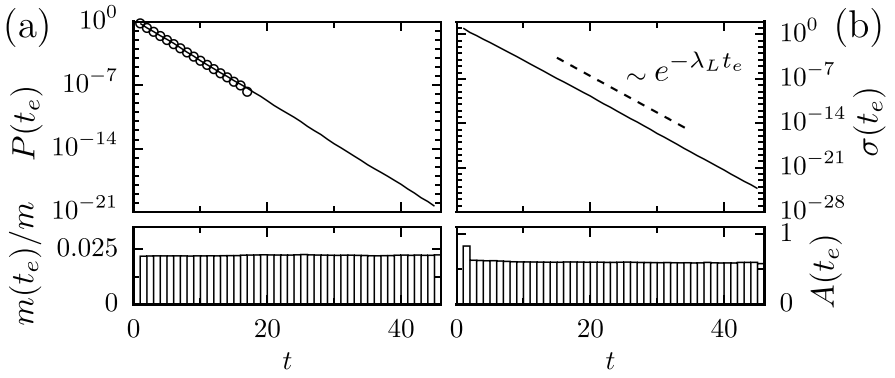

Figure 4.4.: The escape time distribution and $\sigma(t_e)$ of the $D = 4$ coupled Hénon map, Eq. 2.27 computed by the Wang-Landau algorithm introduced in Sec. 3.3.3, with the adaptive proposal derived in Eq. 4.40. The upper left panel represents the escape time distribution computed using Uniform sampling (circles) and Wang-Landau algorithm (line). The lower left panel is the relative number of samples obtained for each t_e, confirming the achieved flat-histogram. The upper right panel is the $\sigma(t_e)$ computed from the adaptive proposal, with a decay coinciding with the largest Lyapunov exponent of the system. The lower right panel is the acceptance ratio, confirming that the acceptance is constant. Uniform sampling used the same number of samples as the Wang-Landau algorithm. Adapted from Ref. [83].

where $\delta_{\min} = \exp(-s_{\max})$ and $\delta_{\max} = \exp(-s_{\min})$. The δ in $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{h}\delta$ is an half-normal distribution with a scale δ_x, but since this scale is now power-law distributed, it no longer makes sense to use the half-normal distribution altogether; instead, it is possible to just use the power-law proposal distribution where δ is drawn from $P(\delta_x)$ in Eq. 4.41, i.e. $|\boldsymbol{x}' - \boldsymbol{x}|$ is power-law distributed according to Eq. 4.41. Without the half-normal distribution the proposal distribution no longer depends on $\boldsymbol{x}$ and therefore $g(\boldsymbol{x}'|\boldsymbol{x})/g(\boldsymbol{x}|\boldsymbol{x}') = 1$.

The stagger part of the algorithm of Ref. [6] is exactly Eq. 4.41 and the argumentation above is an explanation to why the proposal distribution used in Ref. [6] to find states with high-escape time t_e is reported to work well: it is a proposal distribution that proposes $\boldsymbol{x}'$ correlated with $\boldsymbol{x}$ with a correlation $t_\star$ that is uniformly distributed, which eventually proposes $\boldsymbol{x}'$ with the optimal correlation $t_\star(\boldsymbol{x})$. To confirm this explanation, let us consider a flat-histogram simulation with a power-law proposal distribution on the open tent map and consider the measurement of $\mathbb{E}\left[\log \delta_x | A^*, t_e\right]$, where A^* is the condition $\varepsilon < a(\boldsymbol{x}'|\boldsymbol{x}) < 1 - \varepsilon$ (of bounded acceptance). Under the above argumentation, the scale $\log \delta_x$ that contributes to a bounded acceptance is given by $-\lambda_{t_e} t_e$, per Eq. 4.35. This is indeed confirmed by the results of such numerical simulation, shown in figure 4.5. This result, combined with the derivation of $t_\star(\boldsymbol{x})$, explains the success of the proposal (the stagger part) used in Ref. [6] from basic notions of chaotic systems and numerical methods.

4.6. Discussion

This chapter introduced a theoretical methodology to construct an efficient proposal distribution to the Metropolis-Hastings algorithm in chaotic systems. The first section of this chapter proposed an heuristics, constant acceptance rate, as the starting point to construct the proposal distribution. This

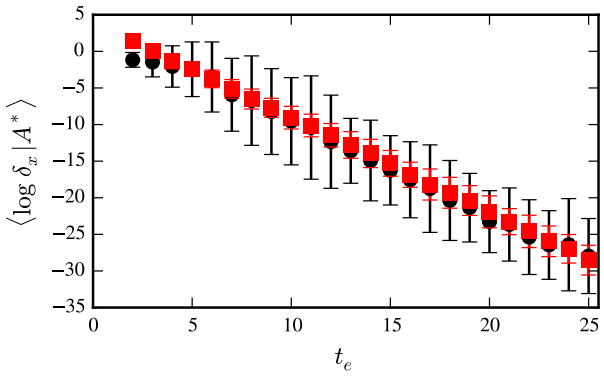

Figure 4.5.: The relevant proposals of the the power-law proposal are those on which the scale is given by $\delta_x(\boldsymbol{x})$ in Eq. 4.35. The x axis is the escape time; the y axis is the best estimator of $\mathbb{E}\left[\log \delta_x | A^*, t_e\right]$ (black dots, 2σ) over 4×10^5 samples obtained from a flat-histogram simulation with the power-law proposal given by Eq. 4.41, for each escape time t_e and conditional to on acceptance $A^* = \{\varepsilon < a(\boldsymbol{x}'|\boldsymbol{x}) < 1 - \varepsilon\}$ with $\varepsilon = 0.1$. The power-law proposal distribution samples all scales, but the scales suitable for Metropolis-Hastings depend on t_e as $\exp(-\lambda t_e)$, as expected from the results of Sec. 4.5.1. This simulation was made on the open tent map, Eq. 2.24, with $a = 3$ and $b = 5$. The best estimate of $-\lambda_{t_e}(\boldsymbol{x})t_e + \log(\Delta)$ with $\Delta = 50$ corresponds to the red line (2σ). The correspondance of the two curves indicates that the scale of the power-law proposal whose acceptance rate is bounded corresponds to the scale given by $\delta_x(\boldsymbol{x})$. The parameters used in the power-law proposal were $\delta_{\max} = 1$, $\delta_{\min} = 2^{-40}$.

generalizes the traditional heuristic [93] used in Metropolis-Hastings, $E(\boldsymbol{x}') - E(\boldsymbol{x}) \sim 1$ and can therefore be used very generally in any Metropolis-Hastings simulation.[6]

The second section of this chapter introduced an auxiliar quantity, the correlation time $t_\star$, to quantify the similarity between any two states of a dynamical system. This correlation is motivated by the notion that the observables considered in chapter 2 are computed over trajectories, and quantifies the similarity of these trajectories. This section then introduced two proposal distributions, shift proposal and neighbourhood proposal, which guarantee that the correlation time $t_\star(\boldsymbol{x})$ between $\boldsymbol{x}'$ and $\boldsymbol{x}$ is controlled.

The third section of this chapter introduced a methodology to derive an expression for $t_\star(\boldsymbol{x})$ that guarantees a constant acceptance in the Metropolis-Hastings algorithm. It focused on two basic observables of chaotic systems introduced in chapter 2: the escape time $E(\boldsymbol{x}) = t_e(\boldsymbol{x})$ and $E(\boldsymbol{x}) = \lambda_{t_o}(\boldsymbol{x})$ for closed systems. This methodology in principle can be extended to any observable and its greatest strength lies in the notion of adding information about the system to the proposal distribution. For example, the derivation of $t_\star(\boldsymbol{x})$ added the understanding that the landscape of t_e is fractal, which is known from the theory of transient chaos. This approach contrasts with other approaches, such as Refs. [6, 14] and many others in other usages of Monte Carlo, because it connects the proposal distribution with the acceptance rate. In this connection, a number of hypothesis and properties of

[6] For example, using this heuristics, in a Metropolis-Hastings flat-histogram in the Ising model (in this case $\boldsymbol{x}$ corresponds to a list of all spins, $E(\boldsymbol{x})$ corresponds to the energy of the configuration), when the configuration of the system is close to the maximum of the density of states (maximal energy, no magnetisation), there is no need to flip just one spin at the time: one can flip all spins at once because, there, correlating such states brings no advantage to increase the acceptance rate (it is going to be accepted anyway), but it increases r.

the system have been used. This is extremely powerful because, when the hypothesis are satisfied, the methodology constructs an algorithm that is efficient in the sense that the acceptance rate is bounded. But, more importantly, when the algorithm is not efficient, this implies that one of the hypothesis is violated. Such violations should then be understood, and this understanding can then be inserted back in this methodology to generate a new algorithm adapted for that situation. More generally, the first and second section introduce two tools that are used in the third section to add hypothesis to the proposal distribution of the Metropolis-Hastings algorithm.

Overall, this chapter described how to add information of chaotic systems, in this case the self-similar properties of the landscape, the exponential diverge of trajectories, and the exponential decay of correlations, to construct an efficient Metropolis-Hastings algorithm to sample them. The next chapters are devoted to test the assumptions used here on each of the problems, escape time and FTLE, and confirm the practical usefulness of the methodology.

5. Application to the finite-time Lyapunov exponent

Section 4.3.1 concluded with a set of formulas, Eq. 4.15 with Eq. 4.24 for proposing states $\boldsymbol{x}'$ that guarantee a constant acceptance when the observable is the finite-time Lyapunov exponent. The first section of this chapter tests the different approximations made to derive the formula of $t_\star(\boldsymbol{x})$ on a system which can be studied analytically, the tent map. This section confirms that the approximations hold in this system, but that the neighbourhood proposal alone still leads to an exponential increase with the cost, and considers a proposal that overcomes this in different chaotic systems, confirming the dramatic improvement of Metropolis-Hastings over uniform sampling. The second section analyses the efficiency of such algorithm, and analytically shows how the limited ways to propose a state $\boldsymbol{x}'$ leads to a sub-optimal (still polynomial) efficiency of the algorithm.

5.1. Tests in the Tent map

5.1.1. $\delta_x(\boldsymbol{x})$ controls $\lambda_{t_\star}(\boldsymbol{x}') - \lambda_{t_\star}(\boldsymbol{x})$

Recall that the first approximation made in the derivation of Eq. 4.24 was that when $\delta_x \equiv |\boldsymbol{x}' - \boldsymbol{x}|$ is drawn from a half-normal distribution with scale parameter $\delta_x(\boldsymbol{x})$ given by Eq. 4.15, $\boldsymbol{F}^{t_\star}(\boldsymbol{x}')$ is sufficiently close from $\boldsymbol{F}^{t_\star}(\boldsymbol{x})$ such that $\mathbb{E}\left[\lambda_{t_\star}(\boldsymbol{x}')|\boldsymbol{x}\right] = \lambda_{t_\star}(\boldsymbol{x})$, Eq. 4.18, holds. This approximation can be tested numerically by randomly drawing states $\boldsymbol{x}_i$, and, for each, propose a state $\boldsymbol{x}'_i$ according to Eq. 4.13 with $\delta_x(\boldsymbol{x})$ given by Eq. 4.15. The result of such test is shown in figure 5.1 and confirms that states $\boldsymbol{x}'$ proposed with a scale $\delta_x(\boldsymbol{x})$ given by Eq. 4.15 with $t_\star = t$ have the same finite-time Lyapunov exponent up to time t, as expected from Eq. 4.18.

5.1.2. $\lambda_{t_o - t_\star}(\boldsymbol{F}^{t_\star}(\boldsymbol{x}))$ is independent of $\lambda_{t_\star}(\boldsymbol{x})$

The second approximation made in the derivation of Eq. 4.24 is that there is no dependence between $\lambda_{t_\star}(\boldsymbol{x})$ and $\lambda_{t_o - t_\star}(\boldsymbol{F}^{t_\star}(\boldsymbol{x}))$, Eq. 4.20. In other words, that the finite-time Lyapunov exponent of the trajectory starting at $\boldsymbol{F}^{t_\star}(\boldsymbol{x})$ and ending at $t_o - t_\star$ is indistinguishable from a finite-time Lyapunov exponent drawning from the distribution of finite-time Lyapunov exponents with finite-time $t_o - t_\star$. This approximation was numerically tested by drawing points $\boldsymbol{x}_i$, computing the pairs $\lambda_t(\boldsymbol{x}_i), \lambda_t(\boldsymbol{F}^t(\boldsymbol{x}_i))$ (i.e. $2t_\star = t_o = 2t$), and test whether the conditional probability of $\lambda_t(\boldsymbol{F}^t(\boldsymbol{x}_i))$ equals the (unconditional) probability of $\lambda_t(\boldsymbol{x}_i)$, Fig. 5.2. The results indicate that the probability of $\lambda_{t_o - t_\star}(\boldsymbol{F}^{t_\star}(\boldsymbol{x}))$

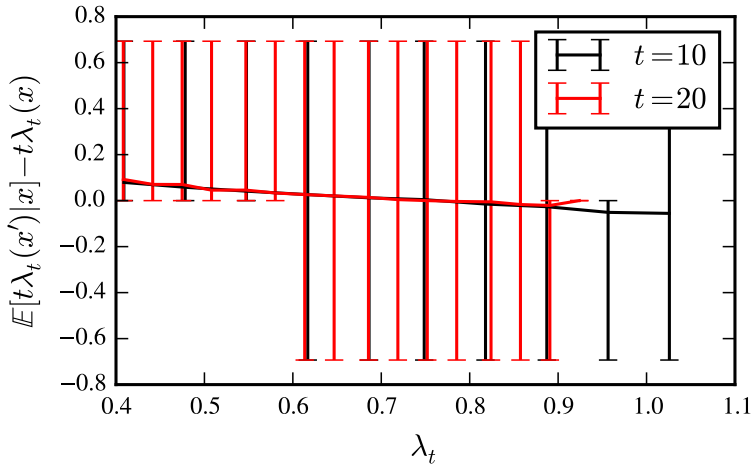

Figure 5.1.: Proposing with $\delta_x(\boldsymbol{x})$ given by Eq. 4.15 with $t_\star = t$ guarantees that $\mathbb{E}\left[t\lambda_t(\boldsymbol{x}')|\boldsymbol{x}\right] = t\lambda_t(\boldsymbol{x})$, i.e. that the trajectories have the same $E = t\lambda_t$ up to time t. The graph was obtained by uniformly sampling 10^5 states $\boldsymbol{x}_i$ and compute $(\lambda_t(\boldsymbol{x}_i), \lambda_t(\boldsymbol{x}_i + \boldsymbol{\delta}(\boldsymbol{x})))$, in the tent map with $a = 3$, with $\Delta = 0.1$. The x-axis represent equally spaced bins with $\lambda_t(\boldsymbol{x}_i)$ and the y axis represents the best estimator, 5% and 95% quantiles of $\mathbb{E}\left[t\lambda_t(\boldsymbol{x}_i + \boldsymbol{\delta}(\boldsymbol{x}))|\boldsymbol{x}\right] - t\lambda_t(\boldsymbol{x}_i)$.

is independent of the particular $\lambda_{t_\star}(\boldsymbol{x})$ obtained, as predicted by the argumentation that lead to Eq. 4.20.

5.1.3. Proposing with $t_\star(\boldsymbol{x})$ guarantees a bounded acceptance

The above tests indicate that the hypothesis used to derive Eq. 4.24 is valid in the tent map. Therefore, a proposal with $t_\star(\boldsymbol{x})$ given by Eq. 4.24 should guarantee a constant acceptance rate, which this section aims to test. The test consisted in sampling 10^6 states according to the following procedure: 1) uniformly draw a state $\boldsymbol{x}_i$ and compute $\lambda_i \equiv \lambda_{t_o}(\boldsymbol{x}_i)$; 2) generate a state $\boldsymbol{x}_i'$ according to the proposal distribution Eq. 4.13 with $\delta_x(\boldsymbol{x})$ given by Eq. 4.15 and $t_\star$ given by Eq. 4.24 ($\delta_0 = 1$), and compute $\lambda_i' \equiv \lambda_{t_o}(\boldsymbol{x}_i')$; 3) store $t_i \equiv g(\boldsymbol{x}|\boldsymbol{x}')/g(\boldsymbol{x}'|\boldsymbol{x})$ computed from Eq. 4.14: δ_x is given by Eq. 4.15, and $|\boldsymbol{x}' - \boldsymbol{x}|$ is given by storing $\delta_i = |\boldsymbol{x}_i' - \boldsymbol{x}_i|$, and $r_i \equiv \pi(E')/\pi(E) = \exp(-\beta t_o(\lambda_i' - \lambda_i))$ for the canonical ensemble and $r_i = P(E)/P(E')$, where $P(E) = P(t_o\lambda_i)$, is given by Eq. 2.14 for the flat-histogram.

In figure 5.3 is plotted the mean of $\Pi \equiv \min(1, \pi(\boldsymbol{x}')/\pi(\boldsymbol{x}))$. It shows that, contrary to the expected from the derivation, the ratio is not independent of λ_{t_o} and therefore $\boldsymbol{x}$. In the canonical ensemble, there is linear dependency of Π with λ_{t_o}, and in the flat-histogram ensemble, the ratio is 1 in the maximum of $P(\lambda_{t_o})$, and decays to around 0.2 on the tails. There are possible explanations for these observations. In the flat-histogram, close to the maximum of $P(\lambda_{t_o})$, $t_\star(\boldsymbol{x}) = 0$ because Eq. 4.8 has a singularity. Being at a maximum acceptance is thus just a consequence of proposing any state $\boldsymbol{x}'$, and naturally accepting it because $\pi(\boldsymbol{x}') = \pi(\boldsymbol{x})$. This is a consequence of Eq. 4.8 not being valid when $\lambda_{t_o}(\boldsymbol{x}) = \lambda_L$, which is specially pronunced in this system because $\lambda_{t_o}(\boldsymbol{x})$ can only take a discrete set of values. Another possibility is that the approximation in Eq. 4.18 is not valid, specially when $t_o - t_\star \approx 1$, where the maximum of $P(\lambda_{t_o - t_\star})$ can be very different from λ_L. Nevertheless, the crucial point here is that the effect of increasing $N = t_o$ does not dramatically changes the ratio of $\pi(E')/\pi(E)$. The

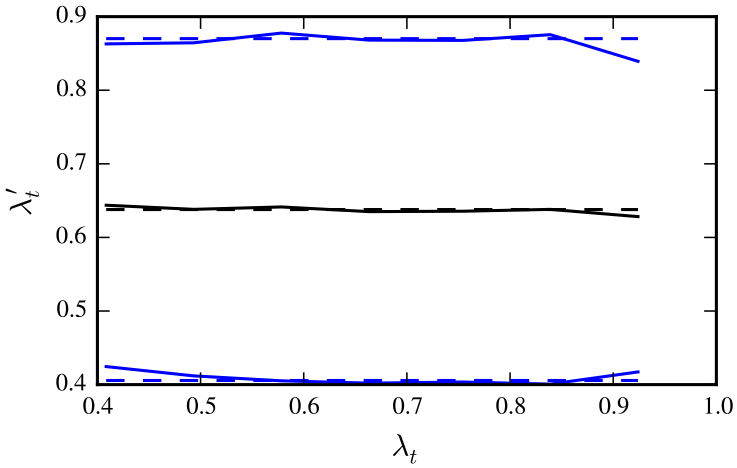

Figure 5.2.: The finite-time Lyapunov exponent of the first half of the trajectory is independent from the one of the second half of the trajectory. The graph was obtained by sampling 10^5 random initial conditions $\boldsymbol{x}_i$ and compute $\big(\lambda_t(\boldsymbol{x}_i), \lambda_t(\boldsymbol{F}^t(\boldsymbol{x}_i))\big)$ = (first half, second half) of a $2t = 16$ steps trajectory. The y axis represents the mean (full black) $\pm$ 2 standard deviations (full blue) of $\lambda_t(\boldsymbol{F}^t(\boldsymbol{x}_i))$ conditioned to a given λ_t. The dashed lines represent the same mean and standard deviation, but over all points (without conditioning). If $\lambda_t(\boldsymbol{F}^t(\boldsymbol{x}_i))$ is independent of $\lambda_t(\boldsymbol{x}_i)$, the dashed and full lines are the same within fluctuations, as observed. A Kolmogorov-Smirnov test comparing the un-conditioned and conditioned distributions is higher than 0.001 (hypothesis that they are independent is not rejected).

acceptance ratio, given by the average of $\min(1, g(\boldsymbol{x}|\boldsymbol{x}')/g(\boldsymbol{x}'|\boldsymbol{x})\pi(\boldsymbol{x}')/\pi(\boldsymbol{x}))$ (Eq. 3.9) is shown in Fig. 5.4 and confirms that the proposal guarantees a bounded acceptance, therefore achieving the main goal set for the proposal distribution.

Overall, these results suggest that the proposal distribution obtained in Section 4.3.1 allows Metropolis-Hastings to be used to compute averages conditioned to the maximal finite-time Lyapunov exponent.

5.2. Analysis of the efficiency of the flat-histogram ensemble

The previous sections showed that, on the tent map, the approximations used in the derivation of the $t_\star$ are valid. Therefore, one would expect a Metropolis-Hastings algorithm using the proposal derived in Sec. 4.3.1 to also be efficient. However, as discussed in section 3.4, a constant acceptance is not sufficient for Metropolis-Hastings to be efficient, and, surprisingly, the tent map is an example where this is the case. To see why, let us analyse the landscape of λ_t, illustrated in Figure 5.5, and imagine a flat-histogram simulation on this system, for $t_o = 4$ (black curve in the figure), and analyse what happens to it in terms of a round-trip. Lets suppose that the simulation was recently at the minimum λ_t (0.41) and that the next round-trip is made by going to the maximum λ_t (1.09) and return back. Lets further suppose that the simulation eventually got to a state with $\lambda_t \approx 0.92$. Because $\pi(\boldsymbol{x}) = \pi(\lambda_t(\boldsymbol{x}))$, every state at that λ_t is equiprobable. Therefore, the state can be at any plateau (of the 4, see fig.), proportionally to their plateau-size. However, not every plateau contains, on its neighbourhood, a neighbour plateau with higher λ_t, for example, the plateau around 0.3. Therefore,

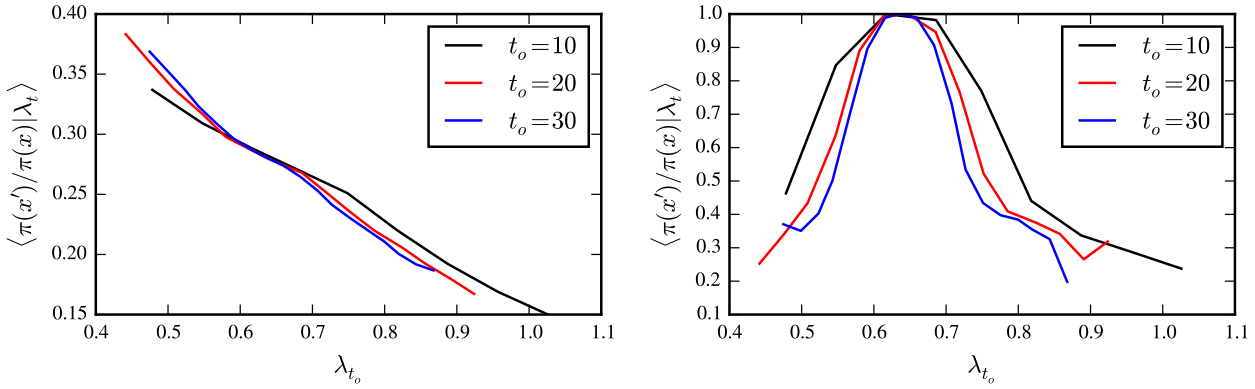

Figure 5.3.: The $t_\star(\boldsymbol{x})$ given by Eq. 4.24 guarantees a bounded ratio $\pi(\boldsymbol{x}')/\pi(\boldsymbol{x})$ independently of t_o. The graph represents the average ratio $\pi(\boldsymbol{x}')/\pi(\boldsymbol{x})$ as a function of $\lambda_{t_o}(\boldsymbol{x})$ obtained from uniformly sample 10^6 pairs of states $(\boldsymbol{x}, \boldsymbol{x}'(\boldsymbol{x}))$ according to Eq. 4.24 (see text for details) in the $a = 3$ tent map, for different finite-times t_o and sampling distributions. Left panel: $\pi(\boldsymbol{x})$ is the canonical ensemble, Eq. 3.10, with $\beta = 1$. Right panel: $\pi(\boldsymbol{x})$ is the flat-histogram ensemble, Eq. 3.12.

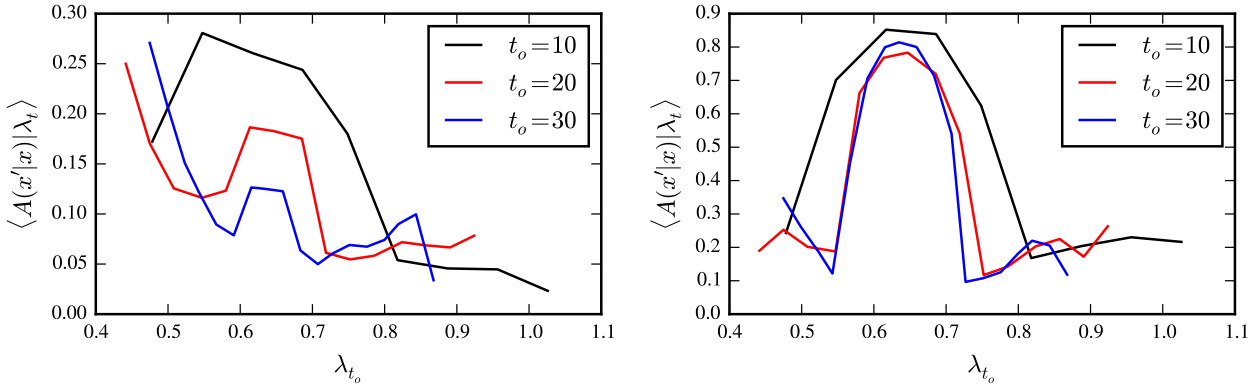

Figure 5.4.: The $t_\star(\boldsymbol{x})$ given by Eq. 4.24 guarantees a bounded acceptance ratio independently of t_o. The graph represents the average acceptance as a function of $\lambda_{t_o}(\boldsymbol{x})$ obtained from uniformly sample of 10^6 pairs of states $(\boldsymbol{x}, \boldsymbol{x}'(\boldsymbol{x}))$ (see text for details) in the tent map with $a = 3$, for different finite-times t_o and sampling distributions. Left panel: $\pi(\boldsymbol{x})$ is the canonical ensemble, Eq. 3.10, with $\beta = 1$. Right panel: $\pi(\boldsymbol{x})$ is the flat-histogram ensemble, Eq. 3.12.

a local proposal would never be able to reach a higher plateau from a state on such a plateau. First, it would need to go backward, reach the maximum of $P(\lambda_t)$ (around 0.69), where the proposal proposes any other state, and then try to find another path towards a higher λ_t. This would already be the problem if the simulation would be a canonical ensemble with a β favouring higher λ_t's, since it would require decreasing λ_t by an amount $\Delta\lambda$, and that only happens with a probability $\exp(-\beta\Delta\lambda t)$, as discussed in Sec. 3.3.1 and exemplified by Fig. 3.1. This is solved by using a flat-histogram ensemble. However, the crucial challenge here is that as t increases, the number of local maxima also increases, but the number maxima connected with the global maximum is constant: the red curve, with $t = 6$, contains now 11 plateaus for $\lambda_t \approx 0.87$, but only 1 is locally connected to the maximum λ_t, the one around 0. This means that, as t increases, it becomes more difficult to perform a round-trip: not only because the expected time to diffuse increases (see Eq. 3.15), but also because there are

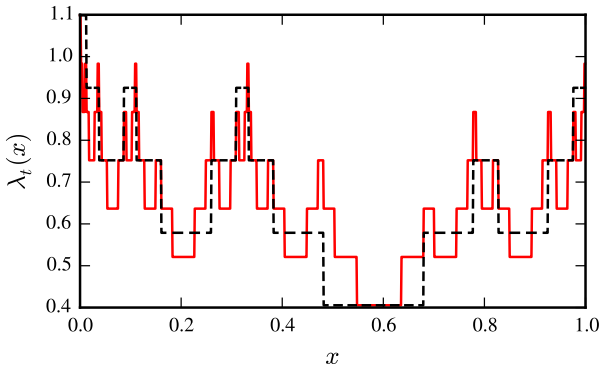

Figure 5.5.: The plot represents $\lambda_t(x)$ of the tent map with $a = 3$, with $t = 4$ (dashed black) and $t = 6$ (full red), computed using Eq. 2.13. The maximum $\lambda_t(x)$ is $\lambda_{\max} \approx 1.09$, and the minimum is $\lambda_{\min} \approx 0.4$.

more times where the simulation diffuses forward and backward until it reaches the global maxima. The existence of multiple local maxima is a well known challenge in Monte Carlo simulations and optimisation algorithms in glassy systems [86, 93, 102–104], and the aim here is to see how this can be solved here, and what consequences it has to the efficiency.

In this simple system, this can also be understood from the perspective of symbolic sequences. The symbolic dynamics of the tent map was introduced in Fig. 2.1, which is here reproduced in Fig. 5.6 to avoid the reader to go back and forth. As described in section 2.4.1, a state x in the tent map can be represented by a symbolic sequence of t_o symbols. Moreover, as described in section 4.2, the neighbourhood proposal is expected to guarantee that $t_\star$ symbols of the sequence of $\boldsymbol{x}'$ are equal to the one from $\boldsymbol{x}$. Thus, the neighbourhood proposal is essentially maintains $t_\star$ symbols, and changes the remaining $t_o - t_\star$. To complete half a round-trip, the neighbourhood proposal needs to, in multiple steps, diffuse from $E_{\min}$ to $E_{\max}$ (recall efficiency discussion in section 3.4). Lets suppose that, in the process of doing half a round-trip, by multiple steps, the random walk starting at a state x where $s(x) = [1...1]$ (with $E(x) = E_{\min}$) is now at a state with a symbolic sequence with $t_o - 1$ zeros, and 1 one (e.g. [010...0]). In terms of the random walk in the observable E, the random walk on E would only require one step to reach $E_{\max}$ to complete half round-trip, by changing the remaining symbol. The crucial point here is that, the neighbourhood proposal can only change specific $t_o - t_\star$ symbols of the sequence. For example, when one symbol is aimed to be changed, by using $t_\star \approx t_o - 1$, not every symbol of the sequence can be changed, which correspond to the neighbourhood symbolic sequences on the phase-space (for example, from sequence [010], neighbourhood proposal can propose the sequence [010], [011] or [110]). Thus, even though on E the random walk is close to $E_{\max}$, in practice it may need to go back in E until it is able to change the particular symbol 1 that would move to $E_{\max}$, and then try again to reach to the sequence [0...0]. Relating to the derivation of Eq. 3.14, the assumption that is violated here is that the random walk in E is not a Markovian process. In this case, being at a given $E = E_{\max} - \Delta E$ does not necessarily allow the state to transit to the state $E_{\max}$, which is to say that the transition probability $P(E'|E)$ alone does not fully describe the process because there

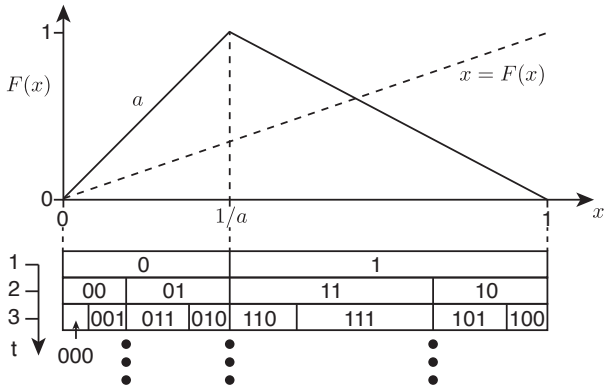

Figure 5.6.: The tent map, Eq. 2.12, and its symbolic sequence. A state with symbolic sequence "010" is a state that starts at partition "0" (left side), its first iteration is at partition "1" and its second iteration is at partition "0". The observable $t_o\lambda_{t_o}$ corresponds to $E(x) = i(x)\log a + (t_o - i(x))\log a/(a-1)$ where $i(x)$ is the number of times the trajectory starting from x was in the interval $[0, 1/a]$, see section 2.4.1. $i(x)$ is thus the number of zeros in the symbolic sequence $s(x)$ of x. $E_{\max}$ corresponds to the sequence [0...0] and $E_{\min}$ to the sequence [1...1].

are correlations in x that are not accounted for by a Markovian process in E.

The above considerations suggest that neighbourhood proposals alone may not be sufficient to efficiently perform round-trips in this system. For this reason, the shift proposal is here used in combination with the neighbourhood proposal to propose a state $\boldsymbol{x}'$ in a non-neighbour plateau of the plateau of $\boldsymbol{x}$, and therefore allow to move from one plateau to the other without having to make the long path of crossing the barrier, using the mechanism (i) in Fig. 3.1. Since, as discussed in sec. 4.2.1, the shift proposal is also able to make small changes in λ, depending on t_{shift}, it implies that such addition will only de-correlate the proposed state $\boldsymbol{x}'$ from $\boldsymbol{x}$ by a finite and small amount, and, at the same time, facilitates the jump from the plateau at 0.8 to the plateau at 0 in one step. To confirm the usefulness of the shift proposal in this case, consider two flat-histogram simulations on the tent map, one with a neighbourhood proposal only, and another with a proposal distribution that is the superposition of a neighbourhood and a shift proposal with $t_{\text{shift}} = 1$ (i.e. at each step, choose one of the proposals with probability 1/2), and lets measure the round-trip time τ with increasing t. Figure 5.7 shows how a flat-histogram with neighbourhood proposal scales exponentially with t, and how including the shift proposal makes the round-trip to scale polynomially with t.

Ref. [84] argues that the shift proposal is required in order to achieve ergodicity, due to the extremely small neighbour proposals that were considered there (see discussion in Sec. 4.3.1). The $t_\star$ derived in Sec. 4.3.1 naturally fixed this issue by allowing a proposal step that is of the size of the phase-space when λ_t is close to the maximum of $P(\lambda_t)$. However, the results of figure 5.7 show that such proposal is not sufficient to obtain an efficient Metropolis-Hastings, not because of ergodicity of the chain, but because of the high correlations induced by the proposal distribution. Therefore, the discussion above complements the work of Ref. [84] because it provides an extra argument of why the shift proposal is required in this situation.

The results in figure 5.7 shows that the number of samples required to obtain an independent state

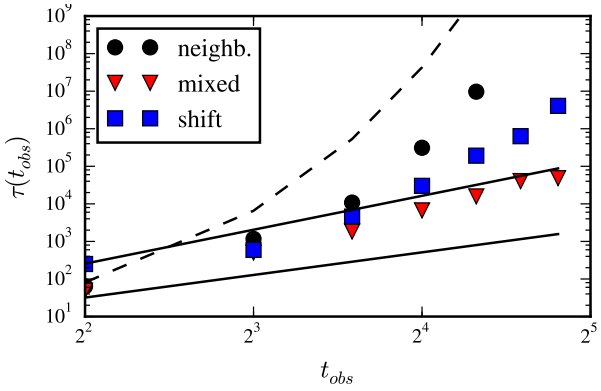

Figure 5.7.: Average round-trip of a flat-histogram in the tent map with a mixed proposal is polynomial, as opposed to uniform sampling, or using only shift or only neighbourhood proposals. The simulation was made using a flat-histogram simulation on the tent map with $a = 3$, where $P(\lambda_t)$ is given from Eq. 2.14, and where the round-trip was defined as going from $\lambda_{\min}$ to $\lambda_{\max}$ and return. The dashed black line represents $1/P(\lambda_{\max})$, the expected number of samples required in uniform sampling; the bottom full line is proportional to t_o^2, the upper line is proportional to t_o^3. Shift proposal requires proposing states using a backward iteration. This is problematic here because the tent map is not invertible; this was solved here by proposing one of the two pre-images of the state.

with the maximum or minimum λ_t in a Metropolis-Hastings with the proposal derived in the previous chapter increases polynomially, as opposed to the exponential increase in uniform sampling in the tent map. This confirms the dramatic improvement of using Metropolis-Hastings algorithm to efficiently sample states with low or high FTLE, on which the proposal distributions derived in the previous chapter are fundamental. To confirm the generality of this result, flat-histogram simulations, with the Wang-Landau algorithm to estimate $P(\lambda_t)$, were performed on different chaotic systems: the tent map considered above, the logistic map, defined by $F(x) = 4x(1-x)$, and the standard map in Eq. 2.15. The results are shown in Fig. 5.8 and confirm the dramatic improvement and generality of using Metropolis-Hastings to sample rare states. The simulations in Fig. 5.8 use $t_\star(\boldsymbol{x}) = t_o - 1$ instead of the one given by Eq. 4.24. This is computationally always more expensive because the correlations due to the neighbourhood proposal are maximal (see the discussion after Eq. 4.24), but on the other hand this proposal is simpler because it does not require estimating $d\log P/dE$ and λ_L.

5.2.1. Analytical analysis of the scaling of the round-trip using symbolic sequences

The results in Fig. 5.8 indicate that the scaling of the round-trip is larger than the expected scaling of the round-trip derived in Sec. 3.4: the expected is $\tau \sim N^2$, Eq. 3.15, and the observed is between N^2 and N^3, and in the tent Map is of N^3 ($N = t_o$). The aim of this section is to explain this scaling. This analysis is published in Ref. [84], and the following discussion is retrieved from it.

Analytical calculations of how the computational effort of a Monte Carlo simulation scales with the system size are typically impossible because of the complexity of the underlying Markov chain [16]. In applications to dynamical systems, this problem is even more difficult because the Markov chain is defined on a continuous phase-space. The advantage of the tent map is that the Monte Carlo

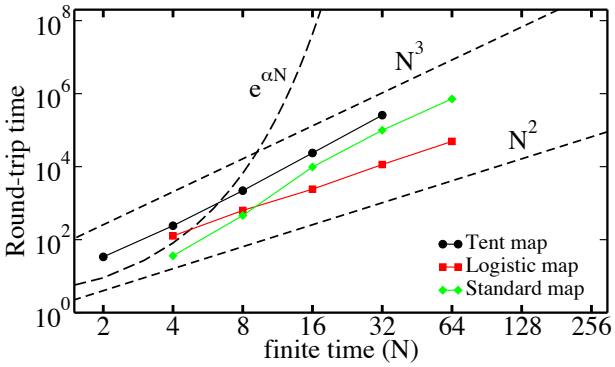

Figure 5.8.: The number of samples required to sample a rare state, proportional to the round-trip time, scales polynomially with $N = t_o$ in Metropolis-Hastings with the proposal distribution derived in Sec. 3.4, as opposed to the exponential increase in uniform sampling. Adapted from Ref. [84]. Tent map: Eq. (2.12) with $a = 3$; Logistic map: $F(x) = 4x(1-x)$; Standard map [5]: $K = 8$ and $P(\boldsymbol{x}) = \text{const.}$ was used in every case. The Wang-Landau algorithm introduced in Sec. 3.3.3 was used to estimate the distribution prior to perform the flat-histogram and the distribution agrees with the analytical one when available [37, 38]. The proposal distribution used was a mixed proposal composed by 50% chance of being the neighbourhood proposal with $t_\star = t_o - 1$ and $\Delta = 1$, and 50% chance of being the shift proposal with $t_{\text{shift}} = 1$.

simulation in its phase-space can be mapped to a Monte Carlo simulation in the (discrete) space of symbolic sequences. Specifically, consider a new Monte Carlo process defined on the set of all possible 2^N symbolic sequences of the tent map that moves through a standard discrete-space Metropolis-Hastings process: from a symbolic sequence s propose a new sequence s' with a given proposal distribution and accept or reject it according to an acceptance distribution. There are two distinct simulations:

1. in the phase-space, using $x \in \Omega = [0, 1]$ with the procedure previously outlined;

2. in the symbolic sequences, with a state space given by the set of all 2^N binary sequences.

The crucial step is to construct the simulation 2. in such a way that it is equivalent to simulation 1.

First, lets map the two proposals, shift and neighbourhood, and for simplicity consider here $t_\star(\boldsymbol{x}) = t_o - 1$, as used in Fig. 5.6, which means that the calculations here are a lower bound to the scaling:

1. shift corresponds to have the whole sequence $s(x)$ shifted by one symbol, where the last symbol is dropped and a new symbol s is added in the beginning (or the opposite to the backward shift), see Fig. 5.9(a). The new symbol s_i (0 or 1) appears with probability μ_{s_i} to correspond to the respective measure of the phase-space [1]

2. neighbourhood corresponds to propose either the same symbolic sequence s or a neighbour sequence *in the phase-space*. E.g. from $s = [011]$ in Fig. 5.6, it proposes $s' = [011]$, $s' = [001]$,

[1] In the phase-space, a forward shift followed by a backward shift sends the state x exactly to the same state it was two steps before because the system is deterministic. The approximation here is that the randomisation due to neighbourhood proposal in simulation 1. leads to the new symbol s_i to appear with probability μ_{s_i} as if the initial condition x with symbolic sequence $s(x)$ would be in any position y such that $s(y) = s(x)$.

or $s' = [010]$. For the tent map, this can be written as a simple rule, see Fig. 5.9(b).

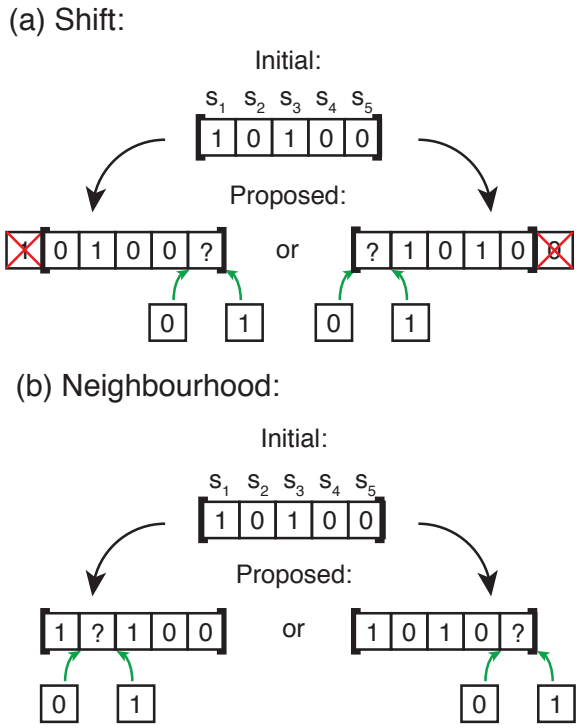

Figure 5.9.: Local proposals on the symbolic sequences of the tent map. (a) *shift* proposes to shift the sequence, where the first symbol is dropped and a new symbol is added in the end (left), or vice-versa (right). (b) *neighbourhood* proposes a change in the symbol after the first "1" when counting from the right (left), or in the last symbol (right).

The acceptance (Eq. (3.9) with $\boldsymbol{x}$ replaced by s) is mapped by taking into account that the proposal is symmetric (i.e. $g(s|s') = g(s'|s)$) and that λ of s is computed using Eq. (2.13) by counting the number of 0's in s.

Fig. 5.10 compares simulations 1. and 2. and indicates no quantitative difference in the scaling of the round-trip time, indicating that both simulations have equivalent scalings. Furthermore, the comparison of the scaling of simulation 2. with and without precision shooting allows to conclude that the effect of the neighbourhood proposal is to move τ vertically (making simulations less efficient[2]) with no effect on its scaling. Therefore, the analysis that follows focus on the shift proposal only.

To explain the sub-optimal scaling of the computational effort, lets consider an ensemble of independent simulations and compute how the average round-trip time τ scales with increasing N. The derivation of the round-trip given by Eq. 3.15 was made under the hypothesis that the ensemble diffuses in $\lambda \in [\lambda_{min}, \lambda_{max}]$ with a variance $\sigma_\lambda \sim \sqrt{t}$. The scaling $\tau \sim N^3$ evident in Fig. 5.10 rejects this hypothesis. Lets then drop this hypothesis and compute explicitly how σ_λ evolves with t. A simulation on the symbolic sequences with only shifts is equivalent to a window of size N moving on a

[2]Notice that the neighbourhood proposal is required in the simulation in phase-space for the simulation to de-correlate from the initial state

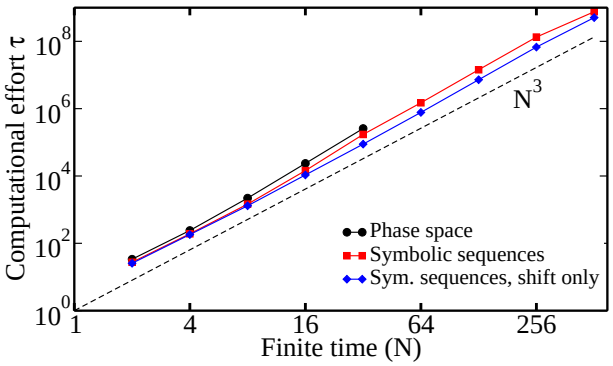

Figure 5.10.: Equivalence between a Monte Carlo in the phase-space of the tent map and in its symbolic sequences. Scaling of the average round-trip as function of the finite-time N of a flat-histogram simulation: in its phase-space (black circles); in its symbolic sequences (red rectangles); in its symbolic sequences using only shift proposals (blue diamonds). The dashed line represents the scaling N^3. The tent map with $a = 3$ was used here, the simulation in the phase-space corresponds to a flat-histogram with the exact distribution Eq. (2.14), $\Delta_0 = 0.1$ was used in Eq. (4.15), and the average round-trip was computed over 100 round-trips. In simulations on the symbolic sequences, the neighbourhood proposal in the symbolic sequence used probability 1/2 of proposing the same sequence, and probability 1/4 to propose to one of the two neighbour sequences, but there is no qualitative difference with other values.

tape of 0's and 1's that, at each step, moves to the left or to the right and randomises the symbol that enters the window (see Fig. 5.9a, where the boundaries of the window are the right brackets in bold) λ is proportional to the sum of 1's in the window (from Eq. (2.13)) and, in particular, λ_{min} occurs in the sequence [00...0] and λ_{max} in the sequence [11...1]. The challenge is then to compute how the round-trip time τ — the number of window moves required to complete the path [00...0] $\rightarrow$ [11...1] $\rightarrow$ [00...0] — scales with N. First, notice that this window performs a simple random walk and thus the number of symbols $\Delta(t)$ that change after t Monte Carlo steps scales as

$$\Delta(t) \sim \sqrt{t} \quad . \tag{5.1}$$

Since, on average, at time t only $\Delta(t)$ of the N symbols changed, the variance $\sigma^2_\lambda(t)$ only depends on the symbols that changed. Since λ is proportional to the sum of symbols by Eq. (2.13), its variance is proportional to the number of symbols that changed $\Delta(t)$, and thus

$$\sigma^2_\lambda(t) \sim \Delta(t) \sim \sqrt{t} \quad , \tag{5.2}$$

which confirms the existence of a subdiffusion in of the random walk in λ. The average time to obtain an independent sequence s is the time τ_s such that all symbols have changed, or $\Delta(\tau_s) \approx N$. Replacing this in Eq. (5.2) leads to

$$\tau_s \sim N^2 \quad . \tag{5.3}$$

To appreciate the relevance of this result, compare it to how would τ_s scale if any symbol of the

symbolic sequence s was allowed to change at each time. Because s has N symbols, this would require $t = N$ steps and thus $\tau_s \sim N$, different from τ_s given by Eq. 5.3. Eq. (5.3) thus indicates that the proposals derived to correspond to the proposals in phase-space dramatically limit the allowed moves, and this changes the scaling of τ_s.

The evolution of the random walk can be summarize in the following picture: on time-scales up to τ_s given by Eq. (5.3), the random walk in λ subdiffuses according to Eq. (5.2); on larger time-scales, the random walk diffuses normally as it draws independent sequences. Results shown in Fig. 5.11

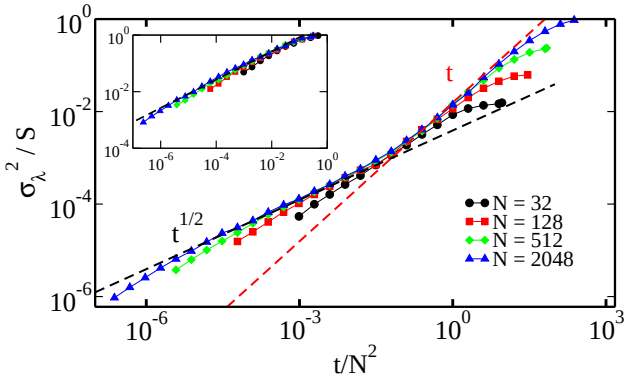

Figure 5.11.: The sub-optimal scaling of Monte Carlo simulations on the symbolic sequences of the tent map. The variance in λ $\sigma_\lambda^2(t)$ is estimated as $1/M \sum_{i=1}^{M} (\lambda_i(t) - \overline{\lambda}(t))^2$, where $\overline{\lambda}(t) = 1/M \sum_{i=1}^{M} \lambda_i(t)$, of $M = 1000$ independent flat-histogram simulations starting from states s with a particular FTLE $\lambda_0 = \lambda_N(N/2)$ [see Eq. (2.13)]. Results are shown as a function of the number of Monte Carlo steps t, for different finite-times N, from $N = 32$ to $N = 2048$. $S = (N+1)^2/12)/N$ is the variance of a flat histogram, divided by N and emphasizes that $\sigma_\lambda^2(\tau_s)$ scales with N, as expected from Eqs. (5.2)-(5.3). Two scalings, $\sqrt{t}$ and t, are shown as dashed lines. Inset shows the same simulations but using the canonic ensemble, Eq. (3.10), with $\beta = 1$. The axis are the same of the main plot, but with σ_λ^2 normalized by the expected variance for the canonical ensemble, derived in Eq. (A.6) of Ref. [84]. The scaling $\sqrt{t}$ is also shown in dashed. Changing λ_0 in the flat-histogram or β in the canonic ensemble does not change the scaling and only shifts all curves.

confirm this picture: a) a transition from $\sigma_\lambda^2 \sim t^{\frac{1}{2}}$ to $\sigma_\lambda^2 \sim t$ is observed; b) the transition occurs at a transition time independent of N when time is rescaled by $1/N^2$, as predicted by Eq. (5.3). Furthermore, the scaling $\sigma_\lambda^2 \sim \sqrt{t}$ does not depend on the sampling distribution, the same scaling is observed in the canonic ensemble, see inset of Fig. 5.11. In fact, it is possible to map a flat-histogram starting at a given initial λ_0 with a canonic simulation with a specific β. [84]

The calculation τ_s now allows to compute the round-trip time τ. Consider the Markov process obtained as τ_s iterations of the original process. The original time t relates to the new time t' by $t' = t/\tau_s$. The new process generates independent symbolic sequences and makes steps in λ of size $\sigma_\lambda(\tau_s) \sim \sqrt{N}$. This implies that the process now diffuses normally in λ with a variance $\sigma_{\tau_s}^2(t')$ given by

$$\sigma_{\tau_s}^2(t') = 2Dt' \sim \sigma_\lambda^2(\tau_s)t' \sim t/N \quad , \tag{5.4}$$

which was obtained by using that $t' = t/\tau_s$, Eqs. (5.2)-(5.3), and that the diffusion coefficient D

of a random walk with step-size normally distributed is proportional to $\sigma^2_\lambda(\tau_s)$. As argued before, performing a round-trip corresponds to $\sigma_{\tau_s}(\tau) \approx N$. Using Eq. (5.4) leads to

$$\tau \sim N^3 \quad , \tag{5.5}$$

in agreement with the scaling observed in Fig. 5.8.

In summary, this analysis described how importance sampling Monte Carlo methods can improve simulations in sampling states on the tails of the distribution of the finite-time Lyapunov exponent. Using a simple system where the round-trip increases faster than N^2, it shows that in this system the Monte Carlo Markov chain decorrelates with $\tau_s \sim N^2$ due to a subdifusion on the finite-time Lyapunov exponent. This allowed to derive the scaling of the round-trip time $\tau \sim N^3$, in excellent agreement with the simulations.

The importance of these results is not limited to flat-histogram simulations. The sub-optimal sampling is a direct consequence of the limited options of proposals that guarantee a bounded acceptance and that can be generated in chaotic systems and therefore should affect importance sampling Monte Carlo simulations using these proposals. For instance, canonic simulations using these proposals – which have been used before for estimating the distribution of finite-time Lyapunov exponents [11, 12] – in the tent map follow the scaling $\tau_s \sim N^2$ derived in Eq. (5.3), as the inset of Fig. 5.11 shows. Extending the validity of the scalings derived in this paper to other methods, such as those in Refs. [11, 12] is not straightforward, but these results show the need of a more careful investigation of the efficiency of modern computational methods applied to dynamical systems.

The efficiency of the simulations is not only a property of the specific Monte Carlo method, it is a result of an interplay between the method (e.g., the proposals) and the phase space structures of the chaotic system (e.g., fractals). Indeed, sub-optimal scaling is also observed in sampling long-living states, as will be analysed in chapter 6.

6. Application to strongly chaotic open systems

This chapter focus on the application of the framework developed in the Chapter 4 to problems in open chaotic systems described in section 2.6. The first section focus on testing the approximations made in the derivations of Eq. 4.34 in section 4.3.2 for strongly chaotic open systems in a simple system, the open tent map. It also confirms the polynomial scaling of the Metropolis-Hastings algorithm with the proposal derived in section 4.3.2 in the Coupled-Hénon maps introduced in section 2.6.1 for different phase-space dimensions. The second section focus on the application of the results of the previous chapter to find long living trajectories in high-dimensional systems and, in particular, how to incorporate the different expanding directions in the proposal distribution to avoid the search algorithm become stuck for arbitrarily high times.

6.1. Test assumptions

Section 4.3.2 introduced a proposal distribution that, under certain assumptions, guarantees a constant acceptance. The aim of this section is to test some of those assumptions on a paradigmatic strongly chaotic system to confirm the usefulness of the proposal distribution.

The derivation of Eq. 4.34 uses the assumption that when $\boldsymbol{x}'$ is proposed with a scale given by Eq. 4.15 with $t_\star = t_e$, $\mathbb{E}\left[t_e(\boldsymbol{x}')|\boldsymbol{x}\right] = t_e(\boldsymbol{x})$, as per Eq. 4.27. In other words, the scale in Eq. 4.15 guarantees that the state $\boldsymbol{x}'$ is correlated with $\boldsymbol{x}$ up to $t_\star(\boldsymbol{x}) = t_e(\boldsymbol{x})$, after which it can exit the system. The test of this assumption is essentially equivalent to the one in section 5.1.1 in Fig. 5.1. The results are summarised in fig. 6.1 and confirm that proposing with a scale $\delta_x(\boldsymbol{x})$ given by Eq. 4.15 with $t_\star(\boldsymbol{x}) = t_e(\boldsymbol{x})$ guarantees that the escape time $t_e(\boldsymbol{x}') \approx t_e(\boldsymbol{x})$.

A second assumption tested here is the self-similarity argument used in deriving Eq. 4.33. This assumption is tested as follows: consider pairs of points $(t_e(\boldsymbol{x}_i), t_e(\boldsymbol{x}_i + \boldsymbol{\delta}(\boldsymbol{x})))$ where $\boldsymbol{x}$ is a random state and $\boldsymbol{\delta}(\boldsymbol{x})$ is given by Eq. 4.35. The self-similarity argument holds if, for a fixed Δ in Eq. 4.33, the distance $(t_e(\boldsymbol{x}_i), t_e(\boldsymbol{x}_i + \boldsymbol{\delta}(\boldsymbol{x})))$ does not depend on t_e. Figure 6.2 presents an estimation of $P(t_e(\boldsymbol{x}') - t_e(\boldsymbol{x})|t_e)$ for different Δ, indicating that the average difference is constant independent of t_e, thus confirming that the proposal distribution achieves its main goal of not depending on t_e.

The previous tests indicate that the proposal distribution should induce a constant acceptance rate when the Lyapunov exponent of the system, Eq. 4.37, is used. To confirm that this is the case, a flat-histogram simulation with an isotropic proposal distribution with width $\delta_x(\boldsymbol{x}) = \delta_x(t_e(\boldsymbol{x}))$ given by Eq. 4.37 with $\lambda_L(t_e) = \lambda_L$ was made. The results, Fig 6.3, reproduced from Ref. [83], confirm that

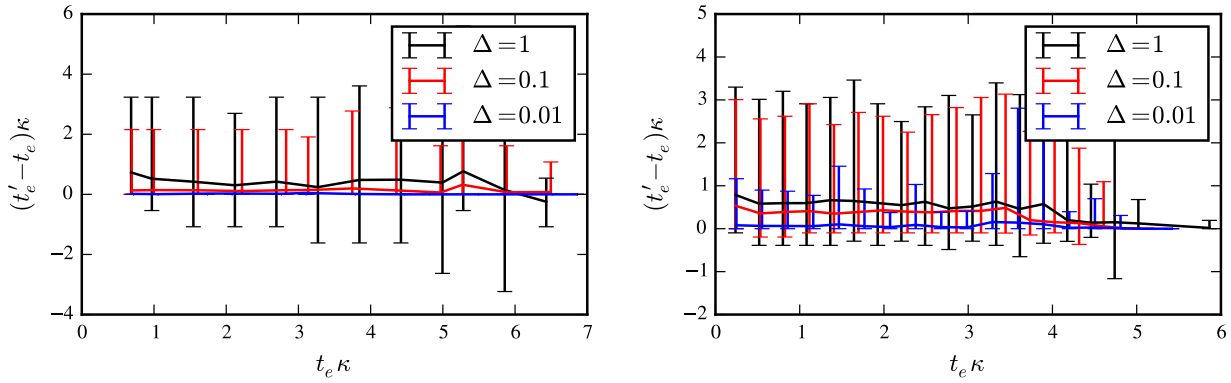

Figure 6.1.: Proposing with the FTLE (Eq. 4.15) with $t_\star = t_e$ guarantees that on average $t_e(\boldsymbol{x} + \boldsymbol{\delta}(\boldsymbol{x})) = t_e(\boldsymbol{x})$. The graph was obtained by uniformly sampling 10^5 $\boldsymbol{x}_i$ and compute $(t_e(\boldsymbol{x}_i), t_e(\boldsymbol{x}_i + \boldsymbol{\delta}(\boldsymbol{x})))$ in the open tent map with $a = 3, b = 5$ (left) and in the standard map with $K = 6$ and $\Lambda = [0, 0.1] \times [0, 1]$ (right); $\boldsymbol{\delta}(\boldsymbol{x})$ was drawn according to Eq. 4.15 (Δ is the pre-factor) with $t_\star = t_e(\boldsymbol{x})$. The x-axis represents the measured $t_e(\boldsymbol{x}_i)\kappa$, where κ was estimated from a simple fit of the escape time distribution. The y axis represents the mean and 0.05'th and 0.95'th quantile of $(t_e(\boldsymbol{x}_i + \boldsymbol{\delta}(\boldsymbol{x})) - t_e(\boldsymbol{x}_i))\kappa$.

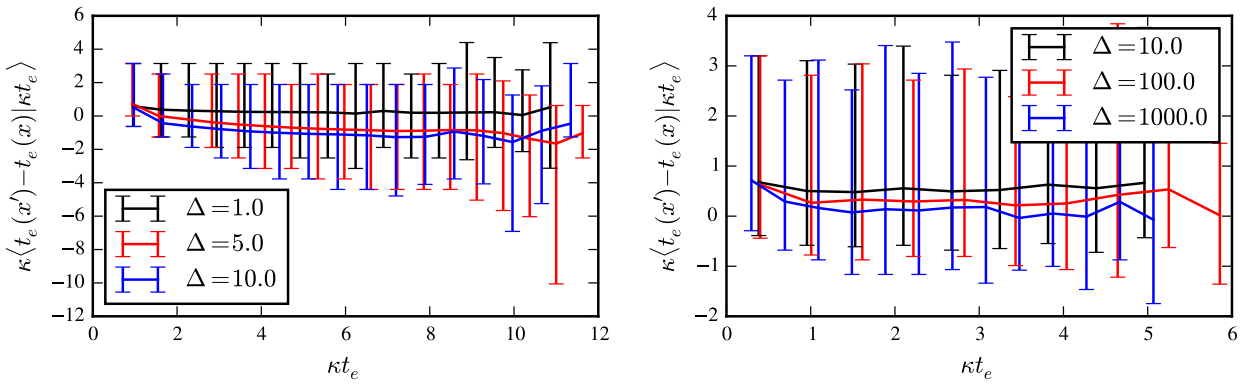

Figure 6.2.: The pre-factor in Eq. 4.33 controls the mean difference $t_e(\boldsymbol{x}') - t_e(\boldsymbol{x})$ for all t_e. The graph was obtained by uniformly sampling 10^6 $\boldsymbol{x}_i$ and compute $(t_e(\boldsymbol{x}_i), t_e(\boldsymbol{x}_i + \boldsymbol{\delta}(\boldsymbol{x})))$ in the open tent map with $a = 3, b = 5$ (left) and in the standard map with $K = 6$ and $\Lambda = [0, 0.1] \times [0, 1]$ (right); $\boldsymbol{\delta}(\boldsymbol{x})$ was drawn according to Eq. 4.15 (Δ is the pre-factor) with $t_\star(\boldsymbol{x}) = t_e(\boldsymbol{x})$. The x-axis represents equally spaced bins with $t_e(\boldsymbol{x}_i)\kappa$ where κ was estimated from a simple fit of the escape time distribution. The y axis represents the mean and 0.05'th and 0.95'th quantile of $P((t_e(\boldsymbol{x}_i + \boldsymbol{\delta}(\boldsymbol{x})) - t_e(\boldsymbol{x}_i))\kappa | t_e\kappa)$.

proposing with λ_L guarantees a constant acceptance, and that any other exponent in Eq. 4.37 fails to achieve so.

The above tests confirm that the derivation made in Sec. 4.3.2 holds for a paradigmatic strongly chaotic open system. These tests were also present the major advantage of using the approach of this thesis: it allows to test the assumptions made on each step, something that other approaches, such as the ones in Refs. [6, 10, 11], do not explicitly allow, where the assumptions are implicit or inexistent.

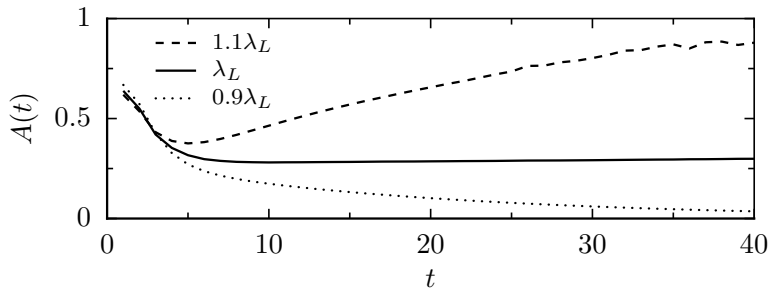

Figure 6.3.: The acceptance rate of a Monte Carlo flat-histogram simulation is constant as a function of the escape time t when the Lyapunov exponent of the system λ_L is used. The simulation was made on the open tent map with $a = 3$ and $b = 5$ with the exact $P(t_e)$ given by Eq. 2.25. Different curves represent using the proposal with $\delta_x(\boldsymbol{x})$ given by Eq. 4.38 with three different exponents. When the exponent is larger than λ_L, $\delta_x(\boldsymbol{x})$ effectively aims for a smaller $t_\star$ and therefore a larger distance $t'_e(\boldsymbol{x}) - t_e(\boldsymbol{x})$, consequently decreasing the acceptance. When the exponent is smaller than λ_L, $\delta_x(\boldsymbol{x})$ aims for a larger $t_\star$ than the one given by Eq. 4.34, and therefore $t'_e(\boldsymbol{x}) - t_e(\boldsymbol{x})$ decreases to 0 as $t_e(\boldsymbol{x}) \to \infty$, and the acceptance converges to 1.

6.2. Efficiency

The tests in the previous section confirm that the proposal distribution derived in section 4.3.2 leads to a constant acceptance rate, suggesting that such proposal leads to an efficient Metropolis-Hastings algorithm. Figure 6.4 presents the average round-trip time as a function of the maximal escape time considered, $t_{\max}$, $\tau(t_{\max})$ in the generic coupled Hénon maps defined by Eq. 2.27, and it confirms the dramatic improvement of Metropolis-Hastings with the proposal derived in section 2.6 over uniform sampling: the scaling is polynomial using importance sampling, and exponential in uniform sampling.

The derivation in Sec. 4.3.2, the tests performed in Sec.6.1, and results in Figure 6.4, show the why and how importance sampling Metropolis-Hastings can efficiently sample long-living trajectories in strongly chaotic open systems. The proposal distribution should be applicable to strongly chaotic open systems more generally, as the approximations made are expected to be valid in other strongly chaotic systems.

6.3. High-dimensional open systems

As described in Sec. 2.6.1, rare events in high dimensional strongly chaotic systems are of major interest, and this section analyses in more detail the application of Monte Carlo to sample extreme events in these systems.

It is not obvious why the assumptions of the derivation of $t_\star$ in section 4.3.2 should hold for high-dimensional systems. The argument used was that since $t_e(\boldsymbol{x})$ is a fractal-like function, a proper scaling of the proposal width $\delta_x(\boldsymbol{x})$ leads to the probability of finding $t_e(\boldsymbol{x}')$ to be independent of $\boldsymbol{x}$, under the assumption that the proposal distribution is isotropic. In high-dimensional systems, where more than one unstable direction exists, different unstable directions are stretched with different rates. In particular, with two unstable directions, the most unstable direction will be stretched faster than

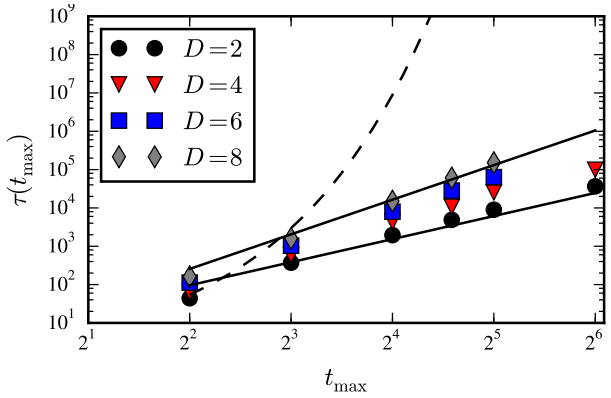

Figure 6.4.: Polynomial scaling of the number of samples required to perform a round-trip ($1 \to t_{\max} \to 1$) as a function of $t_{\max}$ of the Metropolis-Hastings algorithm with the proposal derived in Sec. 4.3.2, as opposed to the exponential scaling expected in uniform sampling. This plot represents the average round-trip time of a flat-histogram simulation with the proposal given in Sec. 4.4 and number of samples required to sample $t_{\max}$ in using uniform sampling (line) in the coupled Hénon map, Eq. 2.27, for different dimensions. The two full lines represent $t_{\max}^2$ (lower) and $t_{\max}^3$ (upper), and the dashed line represents $1/P(t_e)$ for $D = 4$ (e.g. from Fig. 2.5; $D > 4$ have a even higher exponent). The flat-histogram was obtained by first running a Wang-Landau algorithm for 10 refinement steps, each with 100 round-trips. Each point represents the average round-trip time over 100 round-trips after the 10 refinement steps. The proposal distribution used was the isotropic, Eq. 4.13, with $\delta_x(\boldsymbol{x})$ given by Eq. 4.35 with $\delta_0 = 10$.

the other, and therefore an isotropic proposal on $\boldsymbol{x}$ will become a non-isotropic proposal that is not isotropic *along the unstable directions*. This is represented in the top panel of Fig. 6.5. This indicates that the argumentation used in section 4.3.2 is valid only when there is self-similarity along the most unstable direction.

Section 6.1 showed that there is a set of systematic tests that can be performed in order to confirm the assumptions behind the proposal derived in Sec. 4.3.2 on a particular system or classes of systems. To check whether the approximations used are valid in high-dimensional systems, let us perform the same tests on the high-dimensional Coupled Hénon map. Specifically, let us test whether proposing a state isotropically with a scale given Eq. 4.15 achieves an average constant variation of t_e, $t_e(\boldsymbol{x}') - t_e(\boldsymbol{x})$. The results are shown in figure 6.6 and indicate that $t_e(\boldsymbol{x}') - t_e(\boldsymbol{x})$ is largely independent of $t_e(\boldsymbol{x})$ in the high-dimensional Coupled Hénon Map introduced in section 2.6.1: independently of the dimensionality D of the phase-space, $t_e(\boldsymbol{x})' - t_e(\boldsymbol{x})$ is independent of $t_e(\boldsymbol{x})$, confirming that the proposal with a scale given by Eq. 4.15 with $t_\star$ given by Eq. 4.34 achieves a constant acceptance ratio, even in high-dimensional systems.

In figure 6.4, the scaling of the average round-trip with the escape time is not the optimal: t_e^2 is the scaling expected from Eq. 3.15, but the figure indicates t_e^3 instead A qualitative argument that can explain this behaviour is now sketched. As represented in Figure 6.5(a), the proposal distribution at a time t_e becomes narrower and narrower with increase t_e along the most unstable direction (see also Fig. 4.3). A priori, there is no guarantee that there is a state $\boldsymbol{x}_t'$ within that narrow region with $t_e(\boldsymbol{x}_t') > t_e(\boldsymbol{x}_{t-1})$. When no such state exists, the proposal will never be able to reach a higher escape

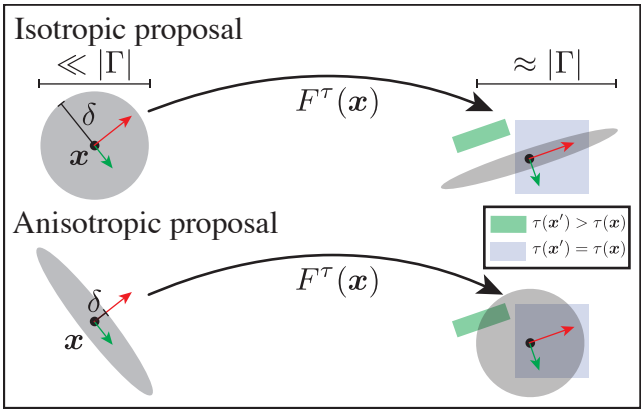

Figure 6.5.: The figure represents a N dimensional system with 2 unstable directions where the plane x-y represent the unstable directions only. The two arrows correspond to the two directions that, after τ iterations, are mapped, via the Jacobian matrix, to the two most unstable directions of $\boldsymbol{x}_\tau$. An isotropic proposal with a scale δ corresponds, after τ iterations, to a proposal along the most unstable direction in $\boldsymbol{x}_\tau$. There is an anisotropic proposal in $\boldsymbol{x}$ that corresponds to an isotropic proposal after τ iterations.

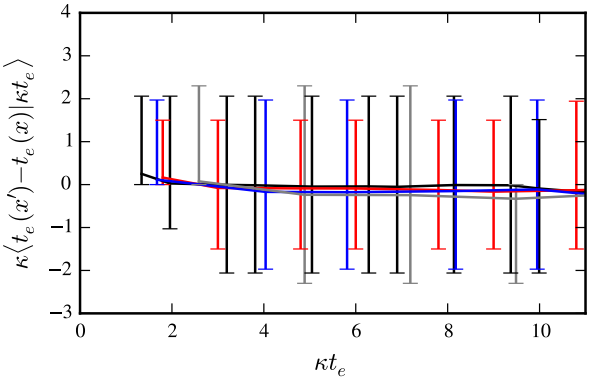

Figure 6.6.: In high-dimensional systems, a constant average difference $t_e(\boldsymbol{x}') - t_e(\boldsymbol{x})$ is achieved by the proposal distribution given by Eq. 4.35. Each point represents the mean, the 0.05 and 0.95 quantile of $P(t_e(\boldsymbol{x}') - t_e(\boldsymbol{x})|t_e)$ (re-scaled by κ), estimated using 10^7 pairs $(\boldsymbol{x}, \boldsymbol{x}')$, where $\boldsymbol{x}$ is uniformly generated in $\Gamma = [-4, 4]^D$ and $\boldsymbol{x}'$ is proposed according to Eq. 4.13 with $\delta_x(\boldsymbol{x})$ given by Eq. 4.35 with $\delta_0 = 8$, for $D = 4, 6, 8, 10$ (black, red, blue, grey). The map is the coupled Hénon map defined in Eq. 2.27.

time, and therefore the completion of a round-trip (the trip $1 \to t_{\max} \to 1$) is hindered when trying to go from $1 \to t_{\max}$. In this situation, the random walk must necessarily go to lower $t_e(\boldsymbol{x})$ to find a region of the phase-space where it can then move to a higher escape time. This corresponds to a similar mechanism described in Sec. 5.2 responsible for an anomalous increase of the round-trip with t_o. As the number D_u of unstable dimensions increases, this behaviour becomes more pronounced, as the proposal distribution at t is proposing in a 1 dimensional line (the most unstable direction), while regions of higher escape time are embedded in a space of D_u dimensions.

The behaviour described above is a consequence of using an isotropic proposal and casts doubts about the generality of the power-law proposal for higher-dimensional systems claimed in Ref. [6],

described in detail in Sec. 2.6.1. Ref. [6] uses an isotropic proposal with the power-law proposal described in Sec. 4.5.2, and, as observed in that section, the scales $\delta = |\boldsymbol{x}' - \boldsymbol{x}|$ on which this proposal is useful corresponds to the scale derived in Sec. 4.3.2. Therefore, it is expected that such proposal will also encounter regions of the phase-space on which the proposal will struggle to find a state $\boldsymbol{x}'$ with a higher escape time than $\boldsymbol{x}$.

One solution to this general problem of proposing in high-dimensional systems, consider an *anisotropic* proposal distribution such that, after t iterations, it corresponds to an isotropic proposal, as represented in Fig. 6.5(b). This ensures that the neighbourhood of $\boldsymbol{x}_t$ is visited equally likely in all unstable directions, thus maximising the chance of $\boldsymbol{x}'_t$ be proposed in to region of higher t_e. This proposal can be constructed by noting that the mapping of a vector $\boldsymbol{h}$ in time is given by Eq. 2.5. Therefore, the evolution of $\boldsymbol{h}$ can be described in the phase-space by the singular value decomposition of J^t. [17] On a single value decomposition, $J_t = U\Sigma V^T$ where U and V are unitary matrixes and Σ is a diagonal matrix, and Σ contains entries higher than 1 (corresponding to unstable directions), and smaller than 1 (corresponding to stable directions). In this decomposition, $U\Sigma V^T\boldsymbol{h}$ represents $\boldsymbol{h}$ being rotated (via V^T), stretched/compressed (via Σ), and then rotated again (via U). The stable directions are not relevant here because it is guaranteed $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{h}$ approximates $\boldsymbol{x}$ along those directions. Let us then consider the matrix Σ^+ to be the matrix Σ with all entries < 1 to be set to zero. The operation $U\Sigma^+V^T$ maps the unstable components of $\boldsymbol{h}$ to their respective forward evolutions. Thus, in order to guarantee an isotropic proposal at time t, the isotropic vector at time t has to be mapped to the initial position as

$$\boldsymbol{h} = \delta_0 V(\Sigma^+)^{-1}U^T\boldsymbol{h}_t \quad , \tag{6.1}$$

where δ_0 is a constant and $\boldsymbol{h}_t$ is a uniformly generated unitary vector ($U^{-1}\boldsymbol{h}_t$ is also uniformly generated).

To analyse the consequences of using the different proposals, let us consider here the problem of finding rare high escape time states in high-dimensional systems, and let us focus on the coupled Hénon map introduced in section 2.6.1. As outlined in Sec. 2.7, the goal is to find a state $\boldsymbol{x}'$ such that $t_e(\boldsymbol{x}') \geq t_e^*$, where $t_e^* = N$ dictates the difficulty of the problem (as $P(t_e)$ decays exponentially). Consider the canonical ensemble outlined in Sec. 3.3.1 to maximize functions, correspondent to $\beta = -\infty$ (minus to maximize), on which the proposed state $\boldsymbol{x}'$ is accepted only when $t_e(\boldsymbol{x}') \geq t_e(\boldsymbol{x})$. To analyse the effect of the different proposal distributions, consider three different proposal distributions to obtain $\boldsymbol{x}'$ from $\boldsymbol{x}$:

- isotropic proposal with scale given by the FTLE, Eq. 4.37;
- isotropic proposal with power-law scale, given by Eq. 4.41;
- anisotropic proposal with vector given by Eq. 6.1.

The parameter $\delta_{\min}$ of the power-law scale is typically unknown a priori. The discussion in section 4.5.2 showed that $\delta_{\min}$ has to be such that $[\delta_{\min}, \delta_{\max}]$ includes the scale $\delta_x(\boldsymbol{x})$ in Eq. 4.15 with

$t_\star = t_e$. Therefore, the $\delta_{\min}$ considered here is $2^{-t_e^*} < \exp(-\lambda_t(\boldsymbol{x})t_e(\boldsymbol{x}))$ for all cases considered here. As discussed before, from $\boldsymbol{x}$ with $t_e(\boldsymbol{x}) = 1$, the proposal distribution is very unlikely to propose a state with $t_e(\boldsymbol{x}') = t_e^*$ due to the decay of $P(t_e)$. Instead, it has to do multiple steps to intermediary escape times, in order to reach the final objective, $t_e(\boldsymbol{x}') = t_e^*$. To quantify the efficiency of the maximisation algorithm, let us derive the optimal efficiency. Consider the hypothetical situation on which the proposal distribution is such that $\Delta E \equiv \Delta t_e = t_e(\boldsymbol{x}') - t_e(\boldsymbol{x}) = \text{const.} > 0$. That is, on each proposal, the proposal is able to reach a higher escape time ($t_e(\boldsymbol{x}') > t_e(\boldsymbol{x})$), and the distance between the two is constant ($\Delta t_e = \text{const.}$). Assuming that the algorithm starts always from a state with $t_e(\boldsymbol{x}) = 1$ (the most likely situation), the number of steps required for the algorithm to reach t_e^* is given by

$$\tau(t_e^*) = \frac{t_e^*}{\Delta t_e} \propto t_e^* \quad . \tag{6.2}$$

Eq. 6.2 is the equivalent of the optimal round-trip derived in Eq. 3.15, for a search algorithm. The scaling is linear with t_e^* (N), and not quadratic as in Eq. 3.15, because this algorithm performs a biased random walk in the observable $E(\boldsymbol{x}) = t_e(\boldsymbol{x})$ towards increasing $t_e(\boldsymbol{x})$. Like in the derivation of the optimal round-trip, Eq. 3.15, there are two important assumptions that lead to Eq. 6.2: the first assumption is that the proposal distribution is able to propose a state $\boldsymbol{x}'$ with $t_e(\boldsymbol{x}') > t_e(\boldsymbol{x})$ independently of the particular $t_e(\boldsymbol{x})$ the algorithm is currently in. The second assumption is that Δt_e does not depend on the history of the algorithm. That is, $P(t_e(\boldsymbol{x}') > t_e(\boldsymbol{x})|\boldsymbol{x})$ is independent of which particular state $\boldsymbol{x}$ the algorithm was previously. Deviations from the scaling in Eq. 6.2 implies that one, or both, of the two assumptions is being violated. Figure 6.7 summarises the costs $\tau(t_e^*)/t_e^*$ of the different proposal distributions considered above and shows how the anisotropic proposal is the only proposal able to achieve the optimal efficiency, i.e. $\tau(t_e^*) \propto t_e^*$ for the different dimensions $D = 2, 4, 6, 8$ considered. Both isotropic proposals, the power-law proposal and FTLE proposal, lead $\tau(t_e^*)$ to increase faster than t_e^*. This indicates that, to reach t_e^*, the algorithm with any of the isotropic proposals requires more steps than the optimal number of steps. This implies that one, or both, of the assumptions used in the derivation of Eq. 6.2 are violated in the isotropic proposals, as anticipated in the discussion motivating the anisotropic proposal. Indeed, $\tau(t_e^*)/t_e^*$ in the anisotropic proposal does not increase with t_e^*, which confirms how it is able to propose states $\boldsymbol{x}'$ that are more likely to guarantee that $t_e(\boldsymbol{x}') > t_e(\boldsymbol{x})$.

This analysis shows how the anisotropic proposal does improve the likelihood that the proposal distribution finds $\boldsymbol{x}'$ such that $t_e(\boldsymbol{x}') > t_e(\boldsymbol{x})$. On the contrary, the cost of the two isotropic proposals discussed here increase exponentially with t_e^* for $D > 4$. While the anisotropic proposal is more efficient to find $t_e(\boldsymbol{x}') > t_e(\boldsymbol{x})$, it does require the calculation of the eigenvalues of the Jacobian matrix. Therefore, the computations necessary to propose $\boldsymbol{x}'$ has a higher cost than the other proposals. This was also the main conclusion of Ref. [7], which also proposes an anisotropic proposal. Still, the analysis here shows how improved information about the state $\boldsymbol{x}$, the different expanding directions and corresponding Lyapunov exponents, allows to better propose states $\boldsymbol{x}'$, and that improves the

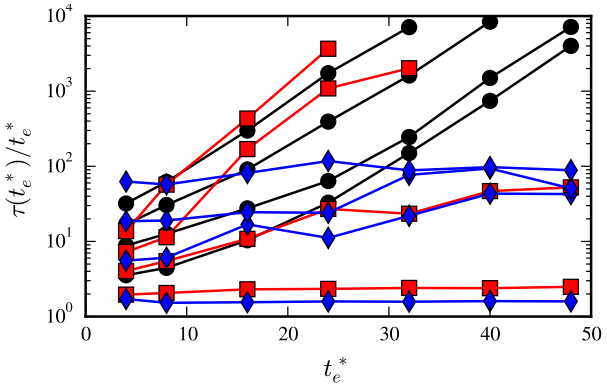

Figure 6.7.: The average cost per escape time of finding a state $\boldsymbol{x}$ with a target escape time t_e^* increases in the two isotropic proposals considered, and is constant in the anisotropic proposal. Each curve represents the average number of steps per escape time, $\tau(t_e^*)/t_e^*$ (y axis), necessary to reach a given t_e^* (x-axis) over 1000 independent simulations starting at a state $\boldsymbol{x}$ with $t_e(\boldsymbol{x}) = 1$ and ending at a state $\boldsymbol{x}^*$ such that $t_e(\boldsymbol{x}^*) \geq t_e^*$, in the Coupled Hénon map. Each curve of the same marker/color corresponds to a different dimension, $D = 2, 4, 6, 8$ (respectively for increasing $\tau(t_e^*)/t_e^*$ in all cases). The black circles represent the isotropic power-law proposal, Eq. 4.41 ($\delta_{\min} = 2^{-t_e^*}$, $\delta_{\max} = 1$). The red squares represent the isotropic proposal, Eq. 4.37 ($\delta_0 = 10$). The blue diamonds represent the anisotropic proposal, Eq. 6.1 ($\delta_0 = 1$). In some cases, the simulation got stuck at a given t_e for more than 10^6 trials to get $t_e(\boldsymbol{x}') > t_e(\boldsymbol{x})$; these cases were discarded in this plot.

efficiency of a search algorithm from an exponential cost with t_e^* to a polynomial and optimal cost.

7. Effects of non-hyperbolicity in the proposal distribution

The results so far indicate that the construction in Chapter 4 is applicable to a general class of problems in chaotic systems and leads to an efficient Metropolis-Hastings algorithm. Still, some of the approximations made in the derivation of explicit formulas for $\delta_x(\boldsymbol{x})$ in Chapter 4 required properties that are not found in weakly chaotic systems. This chapter is devoted to understand the role of non-hyperbolicity of the dynamics in the proposal distribution, and, in particular, how some of the approximations used in Chapter 4 can be violated when the system is non-hyperbolic. In some situations, possible extensions are presented to coup with these violations.

7.1. The FTLE can be smaller than zero

A crucial approximation used in Chapter 4 is that $\delta_x(\boldsymbol{x})$ in Eq. 4.15 guarantees that the trajectory starting from $\boldsymbol{x}'$ is within Δ of the trajectory starting from $\boldsymbol{x}$ up to time $t_\star$. This approximation was made based on the assumption that the first term of the expansion of $\boldsymbol{x}'_t - \boldsymbol{x}_t$ describes $\boldsymbol{x}'_t - \boldsymbol{x}_t$ when the distance $\boldsymbol{x}'_0 - \boldsymbol{x}_0$ is given by Eq. 4.15. To see how this assumption can be violated in the presence of non-hyperbolicity, let us analyse what happens in the logistic map, defined on $\Omega = [0, 1]$ by

$$x_{n+1} = F(x_n) = 4x_n(1 - x_n) \quad . \tag{7.1}$$

This map contains a non-hyperbolic points at $x = 3/8$ and $x = 5/8$, where the absolute of the derivative $|dF/dx| = |4 - 8x|$ is equal to 1. Furthermore, this derivative is 0 at $x = 1/2$. The distribution of FTLE of this map, $P(\lambda_t)$, has negative values for any finite t [38], which implies that there is a non-zero measured set where $\lambda_t(x) < 0$ for all x in that set. For a state x with $\lambda_t(x) < 0$, for a fixed Δ, there is $t_\star$ such that Eq. 4.15 is larger than 1. This implies that, for some $t_\star$, x' is approximately drawn uniformly from $[0, 1]$. In this situation, it is never expected that the two trajectories x' and x are close within Δ up to time $t_\star$, which violates the initial assumption that they would be.

The crucial violation here is that even though the divergence of two trajectories in time is, in the limit that their initial separation goes to 0, given by the first term of the expansion in Eq. 2.3, the divergence of two trajectories separated by a *finite distance* $\delta_x(\boldsymbol{x}) = \exp(-\lambda_t(x)t)$ is not necessarily well described by the first term of the expansion alone. This violation is more evident in states x where $\lambda_t(x) < 0$. To see why, consider a trajectory starting at x on which at time t_i, $x_{t_i} \approx 1/2$, and consider

that all others are not close to 1/2. Since the map is chaotic, up to t_i, trajectories starting from x' distanced from x by $\delta_0 \exp(-\lambda_{t_i}(x)t_i)$ were approximately, at time t_i, within δ_0 of x_{t_i}. However, at that particular point x_{t_i}, because the derivative is zero, the first order approximation dictates that the states x_{t_i+1} and x'_{t_i+1} will be arbitrarily close to each other, as the first order approximation predicts that

$$\left|x'_{t_i+1} - x_{t_i+1}\right| \approx \left|\frac{dF}{dx}(x_{t_i})\right| \delta_0 \approx 0 \quad . \tag{7.2}$$

However, this prediction is far from the expected distance because when the first order term is zero, the second order term dominates. Therefore, the expected distance between these states would need to be approximated by the second order term, that is

$$x'_{t_i+1} - x_{t_i+1} \approx \frac{1}{2}\left|\frac{dF}{dx}(x_{t_i})\right| \delta_0^2 = 4\delta_0^2 \quad . \tag{7.3}$$

To confirm this violation, let us consider two trajectories initially separated by $\Delta \exp(-\lambda_t(x)t)$, and let us compare the distance in time, $\delta_t \equiv |x'_t - x_t|$, with the distance expected from the first term of the Taylor's expansion, $\delta_x(t) = \Delta \exp(\lambda_t(x)t)$. A violation of the assumption happens when the ratio $r(t) \equiv \delta_t/\delta_x(t)$ is different from 1. Figure 7.1 illustrates how the ratio $r(t)$ can be orders of magnitude different from 1, indicating that the distance $|x'_t - x_t|$ is much larger than the expected distance given by $\delta_x(t)$. This ratio becomes 1 by decreasing the initial distance Δ, because there exists a Δ such that the ratio is arbitrarily close to 1 for all t. Still, the crucial point here is that Δ strongly depends on the particular x and can be orders of magnitude different for different x (e.g. one with $\lambda_t > 0$ vs one with $\lambda_t < 0$). In other words, the assumption that is violated in the logistic map is that there is a Δ independent of x that makes $\delta_x(\boldsymbol{x})$ in Eq. 4.15 to guarantee a correlation time $t_\star$ between any two trajectories. This does not necessarily imply that Metropolis-Hastings cannot be applied to systems where $\lambda_t < 0$ for some states; it implies that Eq. 4.15 has to be extended for Δ to depend on x and a relationship for $\Delta(x)$ is required in this situation.

7.2. Non-exponential divergence of trajectories

One important approximation in the derivation of $t_\star(\boldsymbol{x})$ in Eq. 4.34 and Eq. 4.24 is that a trajectory starting $\boldsymbol{x}'_{t_\star}$ rapidly becomes independent of $\boldsymbol{x}'$, such that it can be approximated that $\boldsymbol{x}'_{t_\star}$ is independent of $\boldsymbol{x}'$. This approximation was based on the notion that trajectories diverge exponentially and thus two trajectories e.g. starting from $\boldsymbol{x}_{t_\star}$ and $\boldsymbol{x}'_{t_\star}$ separated by $\Delta \approx 1$, rapidly become independent of each other. This approximation is naturally violated for states x such that $\lambda_t(x) \approx 0$, on which the divergence of nearby trajectories is not exponential.

Let us analyse one simple one-dimensional system where this behaviour is present, the Pomeau-Manneville map with $z = 2$, defined in $\Omega = [0, 1]$ by [105]

$$x_{n+1} = F(x_n) = x_n + x_n^2 \mod 1 \quad . \tag{7.4}$$

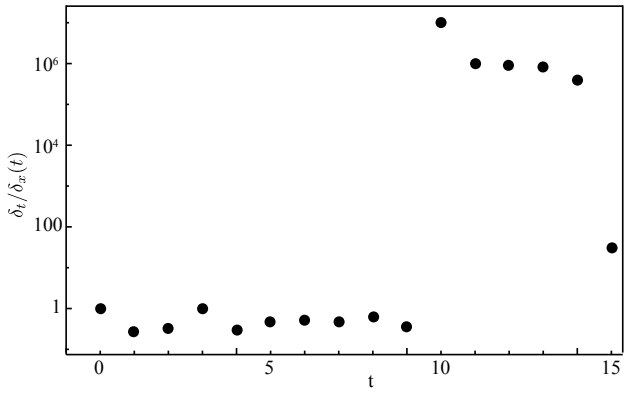

Figure 7.1.: The relative distance $r(t) = |x'_t - x_t|/\delta_x(t)$ of one trajectory x of finite time 15 with $\lambda_{15}(x) \approx -0.15 < 0$ from another trajectory $x' = x + 10^{-5}\exp(-\lambda_{15}(x)15)$. Up to time $t_i = 10$, the distance $x'_t - x_t$ is well described by $\exp(-\lambda_t(x)t)$, but this dramatically changes at that $t_i + 1 = 11$, on which the distance between the two trajectories becomes much larger than the expected from the first order term of the expansion of $x'_t - x_t$ in Taylor's series. This is known in the literature as a glitch [17], on which because $dF/dx \approx 0$ at x_{t_i}, the first term of the Taylor expansion alone does not describe $x' - x$.

This map is a model for intermittence, a phenomena on which trajectories irregularly alternate between apparent integrable and chaotic dynamics [5]. The intermittence in this system appears because the Jacobian matrix (a number in 1D) of this map is $1 + 2x$ and has a non-hyperbolic point, 0, on which the Jacobian matrix is 1, which justifies why it is a non-hyperbolic system.

The interest here is in an open system, which can be achieved by considering a leak in this map [40], here considered by an exit region $\Lambda = [a, 1]$ given by

$$a = \frac{1}{2}\left(-1 + \sqrt{3 + 2\sqrt{5}}\right) \quad . \tag{7.5}$$

This value guarantees that the landscape can be described by a symbolic sequence with forbidden sequences. Its (fractal) escape time function is represented in figure 7.2. The distribution of escape times $P(t_e)$ is known to be power-law distributed with an exponent $\alpha = -2$. Thus, the distribution of $E(\boldsymbol{x}) = \log t_e(\boldsymbol{x})$, can be computed by a change in variables to be an exponential distribution with an exponent $\alpha' = \alpha + 1 = -1$, which is represented in figure 7.3.

Qualitatively, a typical long living trajectory can be pictured by a trajectory that, for a time t_{chaos}, behaves as if it was a chaotic trajectory, and that at some time, denoted here as a time t_i, is injected to very close to the non-hyperbolic point at 0. The trajectory then spends a large amount of time t_{stick} close to 0, until it eventually leaves the region, returning to a chaotic movement. An example of such a trajectory is represented in Fig. 7.4.

As introduced in Sec. 2.6.2, the setting in these systems is to sample states with an increasing logarithm of the escape time, $E(\boldsymbol{x}) = \log t_e(\boldsymbol{x})$, also because $P(t_e) \propto t_e^{-\alpha}$. As in section 4.3.2, the goal here is to obtain an expression for $t_\star(\boldsymbol{x})$, by expressing $\mathbb{E}\left[\log t_e(\boldsymbol{x}')|\boldsymbol{x}\right]$ as a function of $\boldsymbol{x}$ only, and use this in Eq. 4.6.

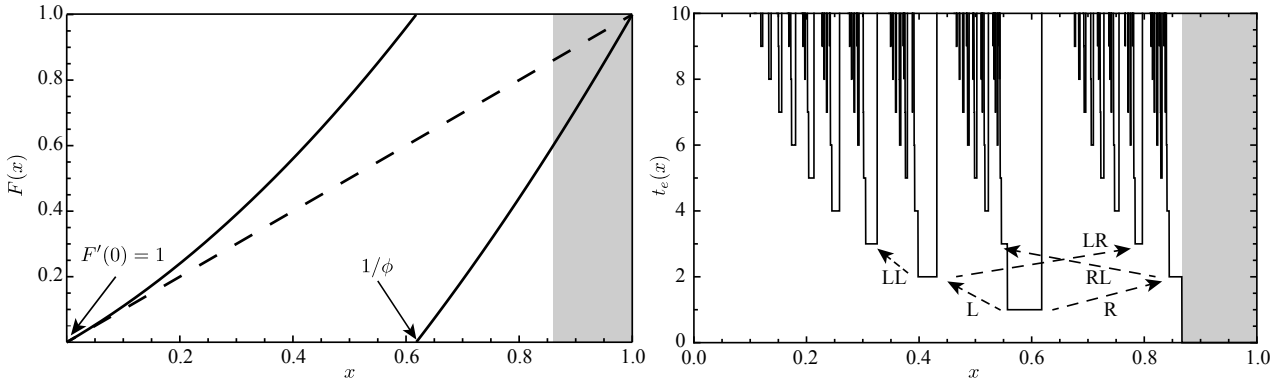

Figure 7.2.: The open Pomeau-Maneville map. Left: the Pomeau-Maneville map and the hole in gray. Right: the escape time function of the map. The landscape is fractal, and has a symbolic sequence associated: each new interval at $t+1$ is constructed to the left (L) or to the right (R) of an interval at t, except when the interval at t was to the right from it's own previous $t-1$. It is thus a restricted symbolic dynamics $s_1 s_2 s_3 ... s_t$ with the forbidden sequence RR.

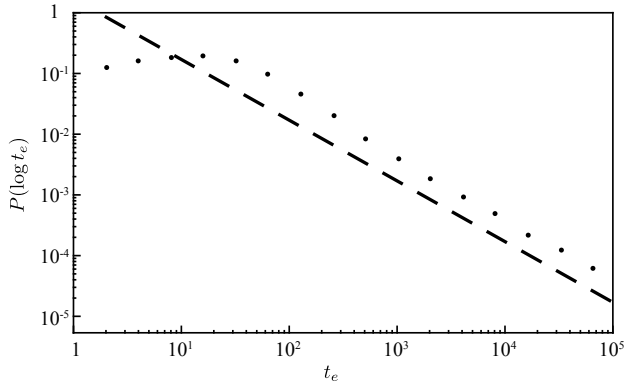

Figure 7.3.: The escape time distribution of the open Pomeau-Maneville map is power-law distributed. The black dots represent $P(E = \log_2 t_e)$ (with bin size $\Delta E = 1$) and the dashed line the expected asymptotic scaling, t_e^{-1}. The x axis is t_e in logarithmic scale. If the x-axis would be $E = \log_2 t_e$ instead, the curve would be an exponential decay in $E = \log_2 t_e$.

The interest here is to understand what distinguishes this map from a strongly chaotic system, and how the proposal distribution can be constructed to this case. The two major modifications here are: a) the observable is $E = \log t_e$, instead of t_e, and b) there is a non-hyperbolic point. Let us start by analysing a typical trajectory x with a high escape time $t_e(x) = 16458$, perturbations x' around it with different distances given by

$$\delta_x(\boldsymbol{x}, t_\star) = \Delta e^{-\lambda_{t_\star} t_\star} \quad . \tag{7.6}$$

This equation is similar to Eq. 4.13, but $t_\star$ here does not dependent on $\boldsymbol{x}$ and is instead a free-parameter that we use to define the scale. Using this proposal, let us measure $P(\log t'_e - \log t_e | \boldsymbol{x})$ for the different $t_\star$. The goal is to test the hypothesis that proposing with Eq. 4.13 guarantees that the two trajectories remain, on average, close together up to time $t_\star$. For example, if the assumption

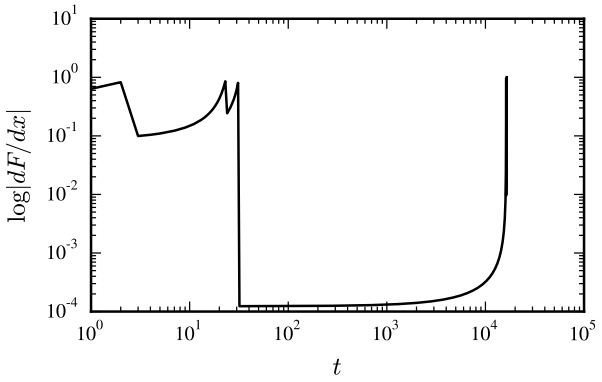

Figure 7.4.: A typical trajectory of a system with intermittence, in this case the Pomeau-Maneville map. The trajectory starts in the chaotic region, where the derivative is high, and it is eventually mapped to very close to $x = 0$, where the derivative is very close to 1 (log close to 0). At this moment, the trajectory spends a long time leaving this region (16458) until it re-enters the chaotic region, and leaves trough the exit region. The FTLE of this trajectory is the average of this curve, and it is already noticable that there are large deviations from the average.

would be true in this system, $t_\star = 0.5t_e$ would imply that on average $t_e(\boldsymbol{x}') \geq 0.5t_e(\boldsymbol{x})$. In terms of the logarithm, this would imply that

$$\log t_e(\boldsymbol{x}') \geq \log t_e(\boldsymbol{x}) - \log 2 \approx \log t_e(\boldsymbol{x}) - 0.7 \quad . \tag{7.7}$$

Figure 7.5 presents numerical estimates of $P(\log t'_e - \log t_e | x, t_\star)$ for different values of $t_\star$ and for a trajectory x with an escape time $t_e(x) = 16458$, or $\log t_e \approx 10$. The first observation is that, independently of $t_\star$, 50% of the trajectories have $\log t'_e - \log t_e \approx 6$, which is unexpected. The second observation is that the average of $\log t'_e - \log t_e$ of the remaining trajectories is much larger than the expected from Eq. 7.7 with $\log t_e(\boldsymbol{x}) = 10$. Instead of the expected $\log t_e(\boldsymbol{x}') - \log t_e(\boldsymbol{x}) \approx -0.7$ for $t_\star = 0.5t_e(\boldsymbol{x})$, the data shows at least $\log t'_e - \log t_e \approx -5$. This shows that the assumption that Eq. 4.13 guarantees that the states are close up to $t_\star$ is here violated. This is expected due to the non-hyperbolic nature of the point $x = 0$. Comparing to the derivations of $t_\star(\boldsymbol{x})$ in chapter 4, what was lost here is the validity of Eq. 4.13 for $t_\star(\boldsymbol{x}) > t_i$. Still, the figure does indicate a dependency of $\log t_e(\boldsymbol{x}') - \log t_e(\boldsymbol{x})$ with $t_\star$, which suggests that it may still be possible to derive a distance between $\boldsymbol{x}'$ and $\boldsymbol{x}$ that leads to a bounded acceptance.

To explore the behaviour of $\log t'_e - \log t_e$ with $t_\star$, let us plot the different trajectories x' generated with $t_\star = 0.9t_e$ that led to the distribution in Fig. 7.5. A useful plot at this moment is to compare the expected divergence given by the first order term of the Taylor expansion with the actual distance between the trajectories, as done in Fig. 7.1. This is shown in figure 7.6 for the Pomeau-Manneville. The first observation is that there are trajectories that largely deviate from x at the time t_i and immediately exit the system. These correspond to points that in Fig. 7.5 have a low escape time $t_e(x')$, and can be explained as follows: the pre-image of $x = 0$ is $1/\phi$, the golden ratio. At $1/\phi$, F is discontinuous. For a trajectory, such as x, to be long living, it must approximate 0, which requires its pre-image to be very close to $1/\phi$ (from the right, i.e. $1/\phi + \epsilon$). The proposal distribution is a

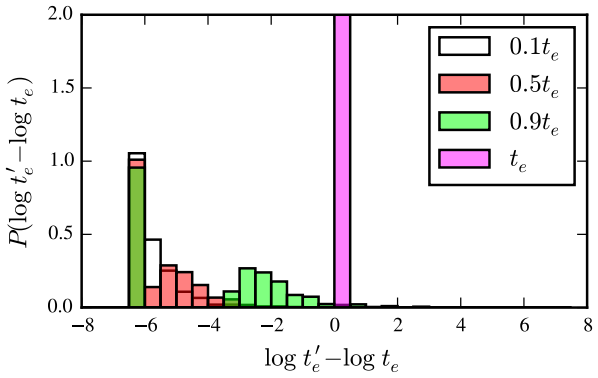

Figure 7.5.: The $P(\log t_e(x') - \log t_e(x)|x)$ for the state x whose trajectory is represented in Fig. 7.4 ($t_e(x) = 16458$), and for x' generate according to $\delta_x(\boldsymbol{x})$ in Eq. 7.6, for $\Delta = 1$ and different values of $t_\star$. Each distribution was estimated using 1000 states x'.

normal distribution around x, and therefore half of the trajectories can be to the left of x and cross $1/\phi$. These will not be mapped close to 0 and quickly leave the system. A crucial observation from this is that the escape time of the trajectory x is mainly driven by how close it passed from 0. This is because, close to 0, the trajectory will not leave that region exponentially fast. Specifically, let us assume that the escape time of a trajectory starting at ε escapes only once it reaches $\varepsilon < b \ll 1$. Let us then compute the escape time of such trajectory. The derivative of the map is $1+2x$, and therefore the divergence of a trajectory starting at $x = \varepsilon$ from one starting at 0 is approximately $\varepsilon + t\varepsilon^2$ (the two first terms of the expansion of $F^t(\varepsilon)$ around $\varepsilon = 0$). The escape time is given by equating this to b, which leads to $t_e(\varepsilon) = \frac{b-\varepsilon}{\varepsilon^2}$. For $b \propto \varepsilon$, this leads to the scaling $t_e(\varepsilon) \sim 1/\varepsilon$. This scaling is confirmed by numerical evolving a trajectory starting at ε, as shown in figure 7.7, which also indicates that the proportionality is really an equality, $\log t_e(\varepsilon) = -\log \varepsilon$.

The above observations can now be used to construct the scale $\delta_x(\boldsymbol{x})$ that guarantees a given variation of $\log t_e(x') - \log t_e(x)$ required by Eq. 4.6. Let us start by denoting $\varepsilon(x)$ as the smallest distance of x_t from 0 for all $t = 1, ..., t_e(x)$, and $t_i(x)$ as the time where the state is in the pre-image of $\varepsilon(x)$. Now, assume that there are no re-injections, such that the time from t_i until the state crosses ϕ, $t_e^*(x)$, is the leading term contributing to the escape time $t_e(x)$, i.e. $t_e^*(x) \approx t_e(x)$. This allows to write that $E(x) = E(\varepsilon(x))$. The aim is then to write $E(\varepsilon(x')) - E(\varepsilon(x))$ as a function of x only, to insert it in Eq. 4.6. This can be computed by noting that

$$E(\varepsilon') - E(\varepsilon) = \log t_e(\varepsilon') - \log t_e(\varepsilon) = \log \varepsilon - \log \varepsilon' \quad . \tag{7.8}$$

Let us denote $\delta_i(x) \equiv \varepsilon - \varepsilon'$ as the distance between two states that are close to 0. This allows to write

$$E(\varepsilon') - E(\varepsilon) = \log \frac{\varepsilon}{\delta_i(x) - \varepsilon} \quad . \tag{7.9}$$

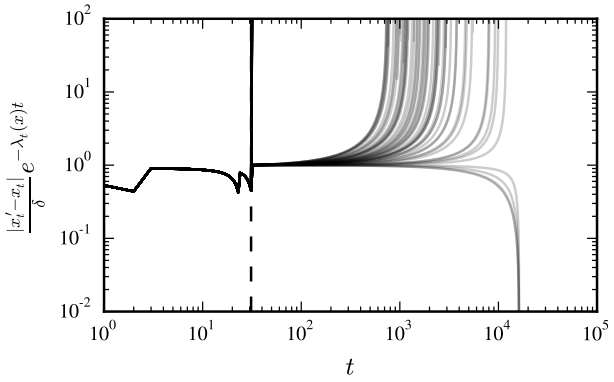

Figure 7.6.: The individual trajectories starting at x' nearby from x diverge from x at the time x is injected to the non-hyperbolic point 0. The y axis is the ratio of the true divergence, $(x'_t - x_t)/\delta$, and the expected divergence of the first order approximation, $\exp(t\lambda_t)$. These are 100 of the 1000 trajectories in Fig. 7.5, with $t_\star = 0.9t_e$. The dashed vertical line corresponds to the injection time t_i (see Fig. 7.4).

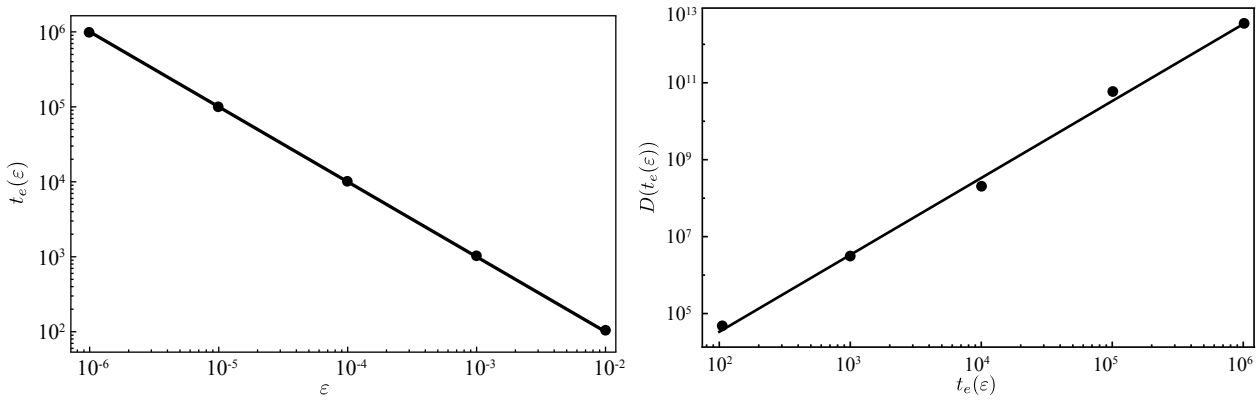

Figure 7.7.: (Left) The escape time t_e depends with distance of the state from 0, ε, as decays as $t_e(\varepsilon) = 1/\varepsilon$. Each point corresponds to an initial state at a different distance from zero (x axis) and its respective escape time (y-axis). The line represents $t_e(\varepsilon) = 1/\varepsilon$. (Right) The total divergence $D(\varepsilon)$ of a state starting at ε is related to its escape time $t_e(\varepsilon)$ according to $D(t_e(\varepsilon)) = 3.3t_e(\varepsilon)^2$ (fit in log scale). Each point corresponds to an initial state at a different distance from zero (ε). The x axis is the escape time of the point, and the y axis is the divergence, computed as $D(\varepsilon) = \sum_{i=0}^{t_e(\varepsilon)} \log(1+2x_i)$ (see Eq. 2.8).

Inserting this relationship in Eq. 4.6 and solving to $\delta_i(x)$ gives

$$\delta_i(x) = \left(e^{-\Delta E(x)} + 1\right) \frac{1}{t_e(x)} \tag{7.10}$$

where $\Delta E(x) \equiv \frac{\alpha - 1}{d \log \pi(E(x))/dE}$. To guarantee that the two states are within $\delta_i(x)$ at time t_i, the initial distance between x' and x should be (according to Eq. 4.15)

$$\delta_x(\boldsymbol{x}) = e^{-\lambda_{t_i}(x)t_i}\delta_i(x) = e^{-\lambda_{t_i}(x)t_i}\left(e^{-\Delta E(x)} + 1\right)\frac{1}{t_e(x)} \quad . \tag{7.11}$$

This can be further simplified by noting that the total divergence $D(\varepsilon) = \exp\left(\lambda_{t_e(\varepsilon)}(\varepsilon)t_e(\varepsilon)\right)$ of a

state starting at ε is $D(\varepsilon) = 3.3t_e(\varepsilon)^2$ (see Fig. 7.7). Because the total divergence $D(x, t_e)$ is the product of the divergence up to t_i and the diverge from t_i to t_e, this allows to write $\delta_x(\boldsymbol{x})$ as

$$\delta_x(\boldsymbol{x}) = 3.3\left(e^{-\Delta E(x)} + 1\right) e^{-\lambda_{t_e}(x)t_e} t_e(x) \quad , \tag{7.12}$$

which relates the scale that two states x and x' should be in order to achieve a given expected variation in the observable E.

To confirm the usefulness of Eq. 7.12, figure 7.8 shows the estimated escape time distribution of the map computed using Metropolis-Hastings flat-histogram with the proposal given by Eq. 7.12 and with Wang-Landau algorithm, confirming that it is able to sample states with high t_e. Figure 7.9 presents the polynomial scaling of the round-trip time of the flat-histogram simulation with increasing E. The round-trip scales polynomially with increasing E, and scales with E^2, which is the optimal situation, as derived in Sec. 3.4, confirming that the proposal distribution derived in Eq. 7.12 achieves its goal of obtaining an efficient Metropolis-Hasting simulation.

The specific derivation presented here is valid for this map only, which is already well studied in the literature. However, the reasoning of deriving a $\delta_x(\boldsymbol{x})$ that allows to change $\log t_e$ can be applied more generally to maps with marginally unstable points. The main conclusion here is that the methodology developed in Chapter 4 is applicable to weakly chaotic open systems with power-law distributions of the escape time, and reinforces the thesis that the strength of the methodology lies in its ability to construct the proposal distribution from the properties of the system.

7.3. Unknown relationship between distance and observable

The previous section considered an example of a weakly chaotic system on which, due to its simplicity, it was possible to derive a relationship between the distance of two trajectories $\boldsymbol{x}' - \boldsymbol{x}$ and the difference in their respective observables, $E(\boldsymbol{x}') - E(\boldsymbol{x})$. Let us now consider a weakly chaotic system where such relationship is not as easily obtainable, and see what can be achieved here. The example considered here is 2D open standard map introduced in Sec. 2.6.2, a paradigmatic example of an Hamiltonian system with mixed phase-space. To recall what was introduced in Sec. 2.6.2, this system is composed by two relevant regions of the phase-space: regions with integrable motion, the KAM islands, and a chaotic see, with a non-integrable motion. This leads to a power-law decay of the escape time distribution, as shown in Fig. 2.7, which motivates the observable $E(\boldsymbol{x}) = \log t_e(\boldsymbol{x})$.

Contrary to the previous cases, and as described in Sec. 2.6.2, in a mixed-phase space system it is less clear what happens to trajectories that approach the boundary of the KAM islands with the chaotic see.

For this reason, the approach taken here is to try to construct a proposal distribution by measuring how trajectories starting close to a state $\boldsymbol{x}$ with high $\log t_e$ (i.e. on the power-law tail of $P(t_e)$) behave. The first step to construct a proposal distribution is to confirm whether $\delta_x(\boldsymbol{x})$ in Eq. 4.15 guarantees that two trajectories stay together up to time $t_\star$. If this is the case, then Eq. 4.15 can be directly used to correlate two states up to $t_\star$, and the remaining question is what $t_\star(\boldsymbol{x})$ should be used in

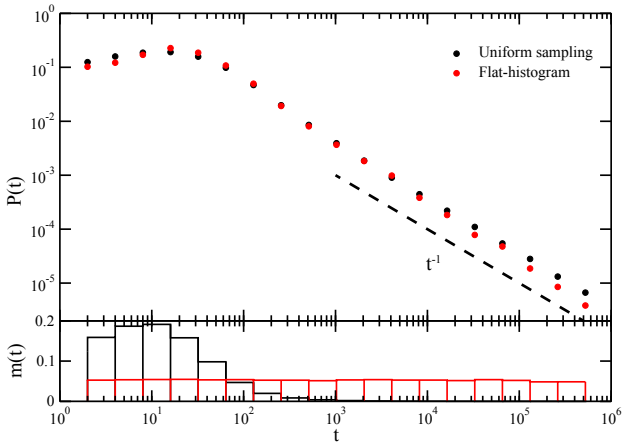

Figure 7.8.: The escape time distribution of the Pomeau-Maneville map using a Metropolis-Hastings with Wang-Landau algorithm, and with the proposal given by Eq. 7.12, and the respective histogram $m(t)$ flat in the variable $E = \log_2(t_e)$. The simulation used 10 WL refinement steps, with 5 round-trips on each refinement step. Since $P(E)$ decays exponentially, the term $3.3e^{-\Delta E(x)} + 1$ in Eq. 7.12 was approximated to a constant in the simulations, equal to 0.1. The expected scaling is shown in dashed, and the uniform sampling in black, confirming the agreement of both simulations.

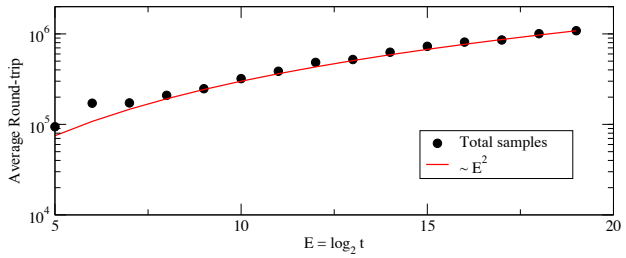

Figure 7.9.: The average round-trip time τ of Metropolis-Hastings with flat-histogram and proposal distribution given by Eq. 7.12 scales polynomially. The simulation used the same parameters described in Fig. 7.8, and 32 round-trips were used for each point.

this case. To test this, consider the same procedure as used in Sec. 7.2, by considering different $t_\star$ in Eq. 7.6 and check whether trajectories remain close to $\boldsymbol{x}_t$ at least up to a time $t_\star$. The results of such test are shown in Fig. 7.10 and confirm that trajectories remain close up to a time $t_\star$. Comparing to the results in Fig. 7.6 for the Pomeau-Manneville map, there seems to be no special time t_i at which trajectories behave fundamentally different. This result indicates that $\delta_x(\boldsymbol{x})$ in Eq. 4.15 can be used to control the time up to which trajectories are close, and the remaining step is to identify which $t_\star(\boldsymbol{x})$ should be used.

Using the same argumentation as in the derivation of Eq. 4.34, the question is what happens in a neighbourhood $\delta_x(\boldsymbol{x})$ of $\boldsymbol{x}$. The relationship between $t_e(\boldsymbol{x}')$ with $t_e(\boldsymbol{x}'_{t_\star})$ is given by

$$\mathbb{E}\left[\log t_e(\boldsymbol{x}')|\boldsymbol{x}\right] = \mathbb{E}\left[\log\left(t_\star(\boldsymbol{x}) + t_e(\boldsymbol{x}'_{t_\star})\right)|\boldsymbol{x}\right] \quad . \tag{7.13}$$

The right side now has to be related to a function that is independent of $\boldsymbol{x}'$. As in the derivation of

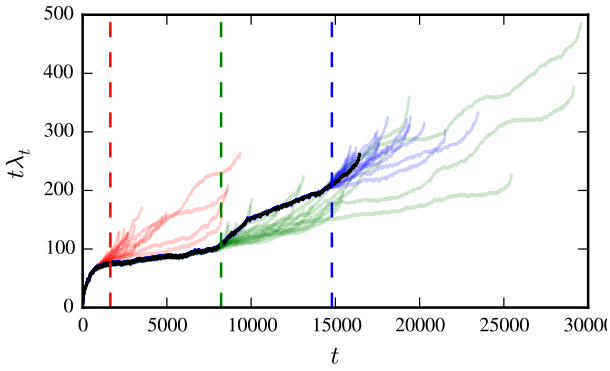

Figure 7.10.: In the open standard map with $K = 2.3$, trajectories starting from a distance of $\boldsymbol{x}$ given by $\exp(-\lambda_{t_\star}(\boldsymbol{x})t_\star)$ (Eq. 7.6) remain close to $\boldsymbol{x}$ up to a time $t_\star$. The black curve represents a trajectory starting at $\boldsymbol{x}$ with $t_e(\boldsymbol{x}) = 16437$ (on the tail of $P(t_e)$, see Fig. 2.7) obtained randomly. Each other curve represents a trajectory starting nearby $\boldsymbol{x}$, with different initial distances given by Eq. 7.6 ($\Delta = 1$) with $t_\star = 0.1t_e(\boldsymbol{x})$ (red), $t_\star = 0.5t_e(\boldsymbol{x})$ (green) and $t_\star = 0.9t_e(\boldsymbol{x})$ (blue). Trajectories are plotted until they leave, i.e. the last time corresponds to $t_e(\boldsymbol{x}')$. The vertical dashed lines represent $0.1t_e(\boldsymbol{x})$, $0.5t_e(\boldsymbol{x})$ and $0.9t_e(\boldsymbol{x})$ respectively. The trajectories $\boldsymbol{x}'$ differentiate from $\boldsymbol{x}$ very close to $t_\star$, indicating that these trajectories are indistinguishable from the one starting from $\boldsymbol{x}$ up to $t_\star$. These simulations were made with a precision of $2^{-2048} \ll e^{-500}$.

Eq. 4.28, by definition,

$$\mathbb{E}\left[\log\left(t_\star(\boldsymbol{x}) + t_e(\boldsymbol{x}'_{t_\star})\right)|\boldsymbol{x}\right] = \int_\Omega g(\boldsymbol{x}'|\boldsymbol{x}) \log\left(t_\star(\boldsymbol{x}) + t_e(\boldsymbol{x}'_{t_\star})\right) d\boldsymbol{x}' \quad . \tag{7.14}$$

Contrary to the derivation of Eq. 4.34, it is unclear which scale $\delta_x(\boldsymbol{x})$ in $g(\boldsymbol{x}'|\boldsymbol{x})$, if any, allows to write the right side as something independent of $\boldsymbol{x}'$. One way to approach this problem is to measure the neighbourhood of a long living state $\boldsymbol{x}$. Consider a state $\boldsymbol{x}$ with a large escape time $t_e(\boldsymbol{x})$ (on the power-law tail of $P(t_e)$), and consider an ensemble of trajectories $\boldsymbol{x}'$ starting around $\boldsymbol{x}$ such that $|\boldsymbol{x}' - \boldsymbol{x}| = \delta$. It was already been established that $t_\star$ controls the correlation time of $\boldsymbol{x}'$ with $\boldsymbol{x}$, and therefore it is possible to write δ as a function of $t_\star$ using Eq. 4.15. Is it possible to obtain a relationship of $t_\star$ with $\boldsymbol{x}$ that controls $\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right]$? The conditional probability $P(\log t_e(\boldsymbol{x}') - \log t_e(\boldsymbol{x})|\boldsymbol{x}, t_\star = t)$ characterises the neighbourhood of $\boldsymbol{x}$, and, in particular, allows to compute the required $\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right]$ for a given proposal distribution with scale $\delta = \delta_x(\boldsymbol{x})$. The interest here is thus to characterise $P(\log t_e(\boldsymbol{x}') - \log t_e(\boldsymbol{x})|\boldsymbol{x}, t_\star = t)$ for a typical point $\boldsymbol{x}$ with a large escape time and use this characterisation to gain insight on what happens on its neighbourhood. Figure 7.11 presents $P(\log t_e(\boldsymbol{x}') - \log t_e(\boldsymbol{x})|\boldsymbol{x}, t_\star = t)$ for different t, and for given point $\boldsymbol{x}$ chosen randomly a priori within different integer values of $\log_2 t_e(\boldsymbol{x})$. These values were chosen because they are on the power-law tail of $P(t_e)$, which is the regime of interest here. The first observation is that, as expected, a small $t_\star$ implies that the proposed $\boldsymbol{x}'$ has a $t_e(\boldsymbol{x}')$ far from $t_e(\boldsymbol{x})$, and $t_\star \approx t_e$ makes $\log t_e(\boldsymbol{x}') \approx \log t_e(\boldsymbol{x})$. The second observation is that, for these particular points $\boldsymbol{x}$, there seems to exist a correlation $t_\star$ that allows proposals with $\log t_e(\boldsymbol{x}') > \log t_e(\boldsymbol{x})$. This indicates that there seems to exist a proposal suitable to search states with higher escape times, which correspond

to a proposal that can be used in a maximisation algorithm such as the Metropolis-Hastings with a canonical ensemble with $\beta = \infty$.

The results in Fig. 7.11 motivate a more detailed analysis to the case $t_\star = 0.5t_e$. Figure 7.12 aggregates the cases $t_\star(\boldsymbol{x}) = 0.5t_e(\boldsymbol{x})$ from the different points $\boldsymbol{x}$ with different $t_e(\boldsymbol{x})$ in a single plot. The results suggest that a proposal with $t_\star = 0.5t_e$ maintains a relatively constant $\log t_e(\boldsymbol{x}') - \log t_e(\boldsymbol{x})$, suggesting that this is a reasonable heuristics to achieve a proposal that, at the very least, maintains $\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right]$ relatively independent of $\log t_e(\boldsymbol{x})$ and thus independent of N. Thus, even with an analytical derivation lacking, these results suggest that $t_\star(\boldsymbol{x}) = 0.5t_e(\boldsymbol{x})$ can be used in flat-histogram Metropolis-Hastings.

Recall that, in the derivation of Eq. 4.6, a crucial assumption used is that a constant acceptance is guaranteed when a constant ratio $\pi(E(\boldsymbol{x}'))/\pi(E(\boldsymbol{x}))$ is guaranteed. This approximation was used to obtain an analytical condition for $\mathbb{E}\left[E(\boldsymbol{x}') - E(\boldsymbol{x})|\boldsymbol{x}\right]$, and was extremely important to derive an expression for $t_\star$ on the different cases considered so far. However, the validity of this approximation was nowhere guaranteed. It was motivated by the qualitative argument that the proposal distribution should, at least, guarantee a constant ratio. It will be argued below that the proposal with $t_\star(\boldsymbol{x}) = 0.5t_e(\boldsymbol{x})$ in the standard map guarantees a bounded $\pi(E(\boldsymbol{x}'))/\pi(E(\boldsymbol{x}))$, but fails to guarantee a constant acceptance ratio. In fact, results in Fig. 7.10 suggest that, even though the escape time for $t_\star = 0.5t_e$ lead to similar $\log t_e$, their respective FTLE varies dramatically. Let us recall that the acceptance is composed by the ratio of two quantities (Eq. 3.9):

$$a(\boldsymbol{x}'|\boldsymbol{x}) = \min\left(1, \frac{g(\boldsymbol{x}|\boldsymbol{x}')}{g(\boldsymbol{x}'|\boldsymbol{x})}\frac{\pi(\boldsymbol{x}')}{\pi(\boldsymbol{x})}\right) \quad . \tag{7.15}$$

For the isotropic half-normal proposal distribution considered so far, the ratio of the proposals g is given by (Eq. 4.14),

$$r_g(\boldsymbol{x}', \boldsymbol{x}) \equiv \frac{g(\boldsymbol{x}|\boldsymbol{x}')}{g(\boldsymbol{x}'|\boldsymbol{x})} = \frac{\delta_x(\boldsymbol{x})}{\delta_x(\boldsymbol{x}')}\exp\left[-\frac{\pi|\boldsymbol{x}' - \boldsymbol{x}|^2}{4\delta_x(\boldsymbol{x})^2}\left(1 - \frac{\delta_x(\boldsymbol{x})^2}{\delta_x(\boldsymbol{x}')^2}\right)\right] \tag{7.16}$$

By definition, $\boldsymbol{x}'$ is constructed to be drawn such that $|\boldsymbol{x}' - \boldsymbol{x}| \approx \delta_x(\boldsymbol{x})$. Thus, the ratio $r_g(\boldsymbol{x}', \boldsymbol{x})$ essentially depends on the ratio $R(\boldsymbol{x}, \boldsymbol{x}') \equiv \delta_x(\boldsymbol{x})/\delta_x(\boldsymbol{x}')$. This allows to write $r_g(\boldsymbol{x}', \boldsymbol{x}) = f(R(\boldsymbol{x}, \boldsymbol{x}'))$ as

$$f(R) = R\exp\left[-\left(1 - R^2\right)\pi/4\right] \quad . \tag{7.17}$$

This function fulfils $f(1) = 1$ and $f(0) = 0$, and decreases to zero for $R \ll 1$ and $R \gg 1$. Thus, the more the distributions differ, the larger/smaller $f(R)$ is (depending on whether $R < 1$ or $R > 1$). To guarantee a constant acceptance, the proposal distribution also needs to guarantee a bounded ratio $r_g(\boldsymbol{x}', \boldsymbol{x})$, which thus equates to guarantee a bounded $R(\boldsymbol{x}', \boldsymbol{x})$. Since there are no more free parameters of the proposal distribution, what remains to be analysed is whether $R(\boldsymbol{x}', \boldsymbol{x})$ is bounded or not. Figure 7.13 illustrates the different values of $R(\boldsymbol{x}', \boldsymbol{x})$ obtained from the same points used to construct the histogram of Fig. 7.12, for the case $t_e = 16437$. As anticipated, it indicates, that the

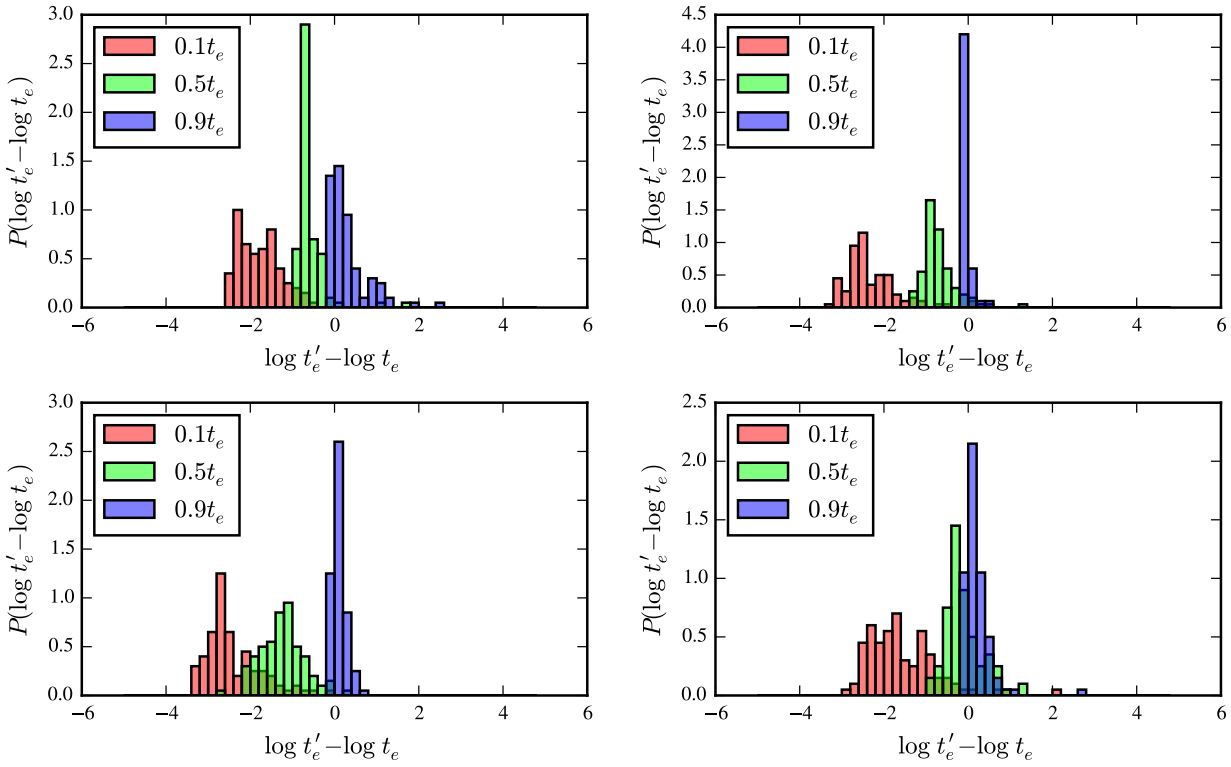

Figure 7.11.: The conditional probability of $t_e(\boldsymbol{x}') = t'_e$ around a particular state $\boldsymbol{x}$ (randomly generated) with an escape time $t_e(\boldsymbol{x})$ given by 1296 (top left), 3594 (top right), 5976 (bottom left) and 14995 (bottom right), corresponding to different values of $E = \log_2 t_e$ ($\approx 10, 11, 12, 13$ respectively). The distribution was estimated by 100 states $\boldsymbol{x}'$ generated by $\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{h}\delta_x(\boldsymbol{x})$ with $\delta_x(\boldsymbol{x})$ given by Eq. 4.15 with $\Delta = 1$ and with different correlation times $t_\star$ given by $0.1t_e$, $0.5t_e$ and $0.9t_e$. The $\lambda_{t_e}(\boldsymbol{x})$ in Eq.$\delta_x(\boldsymbol{x})$ was computed by generating a random unitary vector $\boldsymbol{h}$ (same as the one used to generate $\boldsymbol{x}'$) and evolving it in the tangent space, by multiplying it by the Jacobian matrix J_t. All simulations used a numerical precision of 2^{-2048} and tests were made to guarantee that increasing this precision did not changed $t_e(\boldsymbol{x})$ (decreasing precision did not influenced $t_e(\boldsymbol{x})$), which is an indication that, up to $t_e(\boldsymbol{x})$, the trajectory $\{\boldsymbol{x}_i\}$ is a true trajectory of the system. The map considered was the standard map with $K = 2.3$ and with an exit region given by $\Lambda = [0, 0.1] \times [0, 1]$. These histograms are obtained from an histogram over the total length of the trajectories in Fig. 7.10.

ratio $R(\boldsymbol{x}', \boldsymbol{x})$ is orders of magnitude different from 1 specially with $t_e(\boldsymbol{x}') > t_e(\boldsymbol{x})$, which, from the preceding discussion, leads to an arbitrarily small acceptance rate.

In summary, this section showed how the proposal distribution with $t_\star(\boldsymbol{x}) = 0.5t_e(\boldsymbol{x})$ in the open standard map allows to propose states with an increasing $\log t_e(\boldsymbol{x}')$, which is extremely useful in an algorithm that aims to *find* long living trajectories. However, this proposal distribution leads to an low acceptance rate of a flat-histogram Metropolis-Hastings simulation, which implies that it is not suitable to *sample* long living trajectories. Nevertheless, including the ratio of the proposal distributions on the acceptance rate can be used to extend Eq. 4.6, which sets a tighter condition for a bounded acceptance rate. While this was not necessary in the derivations of $t_\star(\boldsymbol{x})$ of Chapter 4, the analysis done here and, in particular, the results in Fig. 7.13, shows that this may be necessary in some observables $E(\boldsymbol{x})$.

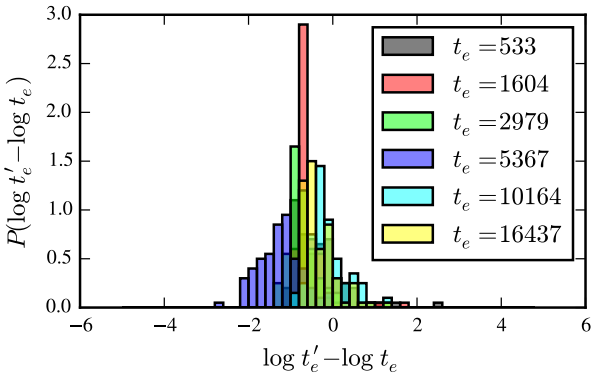

Figure 7.12.: The same data of Fig. 7.11, but where the case $t_\star = 0.5t_e$ is aggregated in a single plot. The distribution of distances $\log t_e(\boldsymbol{x})' - \log t_e(\boldsymbol{x})$ with $\boldsymbol{x}'$ proposed with a correlation time $t_\star(\boldsymbol{x}) = 0.5t_e(\boldsymbol{x})$ from $\boldsymbol{x}$ is relatively robust to an increasing $t_e(\boldsymbol{x})$. In particular, it does not decay to $-\log t_e(\boldsymbol{x})$ with increasing $t_e(\boldsymbol{x})$, which is the case in a uniform proposal, $t_\star(\boldsymbol{x}) = 0$.

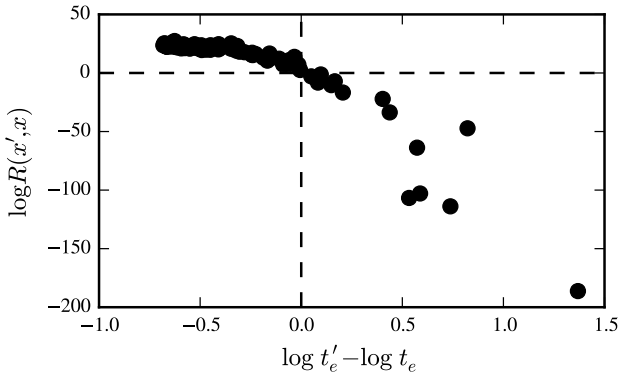

Figure 7.13.: The ratio of the proposal distributions varies by orders of magnitude with the isotropic proposal in the open standard map with $t_\star(\boldsymbol{x}) = t_e(\boldsymbol{x})/2$. The x-axis represents the distance in $\log t_e$ of $t_e(\boldsymbol{x}')$ and $t_e(\boldsymbol{x})$, whose marginal on $\log t'_e - \log t_e$ is exactly the one in Fig. 7.12 for $t_e = 16437 \approx 2^{14}$. The y-axis represents the logarithm of the ratio $\log R(\boldsymbol{x}', \boldsymbol{x})$ of the proposal widths δ at $\boldsymbol{x}$ and $\boldsymbol{x}'$, that is appears in the ratio of the proposal distribution in the acceptance rate. It shows that the ratio is very different from 1, which implies that the ratio of the proposal distributions varies by orders of magnitude.

7.4. Discussion

This chapter discussed the effect of non-hyperbolicity in the proposal distribution constructed in Chapter 4. It was focused on three types of non-hyperbolicity: non-hyperbolicity due to a point with derivative zero (logistic map, section 1), non-hyperbolicity due to a point with derivative 1 (Pommeau-Manneville map, section 2), non-hyperbolicity due to KAM islands (standard map, section 3). This chapter illustrated how some of the approximations used in Chapter 4 are not valid in non-hyperbolic systems. In all situations, this was understood in terms of what exactly fails. In the case of the Pommeau-Manneville map, the analysis of how the distance $\boldsymbol{x}' - \boldsymbol{x}$ affects the distance in the respective observable E allowed to construct a proposal distribution that bounds the acceptance rate. In the case of the logistic map, this seems to be possible to do, but it was not explored here. In

the case of the standard map with $E = \log t_e$, a proposal that guarantees a bounded variation on E seems possible, but it was shown that this is not sufficient for the acceptance rate to be bounded. Thus, it remains an open question whether there is a proposal distribution that guarantees a bounded acceptance rate on this system.

8. Conclusions

This thesis presented a framework to construct an efficient Importance Sampling Monte Carlo algorithm to study rare events in different classes of chaotic systems. The argumentation of this thesis can be summarised as follows:

1. Extreme events are often studied by non-linear models of dynamical processes, some of them chaotic (Chapter 1);

2. There is a set of relevant numerical problems in studying rare events of chaotic systems that can be described in a common setting suitable to apply importance sampling Monte Carlo (Chapter 2);

3. Metropolis-Hastings (MH) is a flexible and well established Monte Carlo method to address these problems; In order to be efficient, it requires a proposal distribution that uses information about the chaotic system (Chapter 3);

4. It is possible to incorporate general features of chaotic systems, such as self-similarity of some of its properties, in the proposal distribution (Chapter 4);

5. Incorporating these features in the proposal distribution leads to an efficient algorithm to generate rare events in chaotic systems (Chapter 5-6);

6. Sampling rare events in non-hyperbolic systems often requires adapting the proposal distribution for the particular characteristics of the system (Chapter 7).

This thesis is the first systematic study on the applicability of importance sampling techniques to chaotic systems, and it shows how an algorithm to study rare events crucially depends on the assumptions about the system incorporated in the proposal distribution. By incorporating critical information about the system, such as "trajectories diverge exponentially", the thesis derived proposal distributions, Eq. 4.24, 4.34, and 7.12 (partially summarised in Table 8.1), that lead to an efficient Monte Carlo algorithm to sample rare events in chaotic systems, as evidenced by Fig. 5.8, 6.4, and 7.9, on which the number of samples to obtain an independent rare sample scales polynomially with the difficulty of the problem, as opposed to the exponential increase observed in traditional uniform sampling.

One aspect of the results in this thesis is that they allow to analyse some of the existing methods in the literature. For example, the stagger part of the *stagger and dagger* (Ref. [6]) corresponds to an isotropic power-law proposal distribution (Eq. 4.41). This proposal distribution was derived here

Target distribution		Proposal correlation time $t_\star(x)$	
		Escape time $E(x) = t_e(x)$	FTLE $E(x) = t\lambda_t(x)$
canonical	$\pi(x) \propto e^{-\beta E(x)}$	$t_e(x) + \frac{a-1}{\beta} - \frac{1}{\kappa}$	$t - \left\lvert \frac{a-1}{\beta} \frac{1}{\lambda_L - \lambda_t(x)} \right\rvert$
flat-histogram	$\pi(x) \propto \frac{1}{P(E(x))}$	$t_e(x) - \frac{a}{\kappa}$	$t - \left\lvert \frac{a-1}{d \log P(\lambda_t)/d\lambda_t} \frac{1}{\lambda_L - \lambda_t(x)} \right\rvert$

Table 8.1.: The four $t_\star$ derived for the two problems (escape time and FTLE as the observable) and two target distributions (canonical and flat-histogram). The four values of $t_\star(\boldsymbol{x})$ reported in the table should be used in $\delta_x(\boldsymbol{x})$ of Eq. 4.15 and specify the time the proposed trajectory $\boldsymbol{x}'$ should stay close to the trajectory $\boldsymbol{x}$ for the acceptance to be bounded.

(Sec. 4.5.2) under a set of approximations about the system ($t\lambda_t$ can be approximated by a uniform random variable) and thus it should also be applicable under these approximations. Moreover, from Fig. 4.5, it is expected that such proposal distribution is less efficient than using a proposal distribution with the Lyapunov exponent of the system, because some of the proposed states $\boldsymbol{x}'$ are too far or too close from $\boldsymbol{x}$. Another example of an algorithm that can be directly analysed by the framework developed here is the *precision shooting* proposed in Ref. [9]. Essentially, this method proposes a state $\boldsymbol{x}'$ isotropically distanced from $\boldsymbol{x}$ by $\delta_x(\boldsymbol{x})$ given by Eq. 4.15, with $t_\star(\boldsymbol{x}) = t_o$ where t_o is the length of the trajectory. The derivations in chapter 4 show why such proposal should guarantee a bounded acceptance rate, and therefore explain why it can be used to sample rare events in chaotic systems. Still, the derivations in chapter 4 show that such proposal always over-correlates the state $\boldsymbol{x}'$ with $\boldsymbol{x}$, because $t_\star(\boldsymbol{x}) = t_o$ corresponds to the highest possible correlation time between $\boldsymbol{x}'$ and $\boldsymbol{x}$. This thesis proposes a modified $t_\star(\boldsymbol{x})$ that improves the efficiency of the algorithm because it allows to reduce the correlation time $\boldsymbol{x}'$ with $\boldsymbol{x}$ depending on the system and sampling distribution π. Another application of the results of chapter 4 is in the algorithm *Lyapunov Weighted Dynamics* of Ref. [11], which uses a population Monte Carlo algorithm. As mentioned in the reference, there is a parameter ε that controls how far new clones $\boldsymbol{x}'$ should be distanced from the existing clone $\boldsymbol{x}$, and that it should be neither too small nor too large. This seems to play an equivalent role as the $\delta_x(\boldsymbol{x})$ in chapter 4 and an optimal ε should therefore be related with it.

For the study of Monte Carlo algorithms, this thesis reinforces the idea that an efficient proposal *requires information about the system.* The importance of the proposal distribution has long been emphasised in the literature [16, 95] and, in the context of sampling chaotic systems, different proposal distributions were discussed before, e.g. in refs. [6, 9]. However, how the efficiency of the algorithm is related to the proposal distribution and how an efficient proposal can be constructed from assumptions about the system has not been discussed before [6–11]. The exercise taken here of constructing a Metropolis-Hastings algorithm from first principles – a bounded acceptance rate – evidences how a proposal distribution can be derived from specific knowledge about the system, such as "trajectories diverge exponentially" or "the escape time function is a fractal-like function".

The proposal distribution is a way of moving in the phase-space stochastically and how to select a new state $\boldsymbol{x}'$ from a given state $\boldsymbol{x}$ is a general problem in different numerical techniques. Some of the most successful numerical algorithms in the literature, such as the golden section search or gradient descent, are essentially generic and efficient ways of selecting a new state. The results in Figs. 5.7, 5.11, and 6.7 show that the proposal distribution strongly influences the computational cost of the different procedures, often irrespectively of the particular sampling procedure (canonical or flat-histogram) and problem (sampling or finding). This reinforces the notion that the selection of the new tentative state from the current state is a crucial factor to take into account when developing numerical techniques for optimisation and numerical integration. The main insight of this thesis is direct connection between fundamental properties of chaotic systems and the optimal proposal distribution.

The development of a numerical algorithm requires compromising between how fast it solves a particular problem, and how it is able to solve different problems. One interesting aspect of the algorithms (i.e. proposal distributions) introduced in this thesis is that even though they can be made very specific (e.g. propose with the FTLE of the trajectory, Eq. 4.15), they can also be made more general (e.g. propose using the Lyapunov exponent of the system, or the power-law proposal distribution, that does not use any specific information about the system). That is, more specificity requires more information (the FTLE of the trajectory) and makes the algorithm more efficient, and less information (only the Lyapunov of the system) makes the algorithm less specific, but also less efficient. This demonstrated capability of this methodology shows how it is not only useful to study a particular system on which some information about it is known, but also useful to situations on which less is known about the system. This does not imply that this framework is useful to any problem: as shown in Chapter 7, the open standard map is an example on which obtaining an efficient proposal distribution remains unclear. Nevertheless, because the framework was outlined in the form of adding known information about the system, it is possible that improved insights about a class of chaotic systems can be translated to a faster algorithm.

One advantage of the methodology presented here is its generality with respect to the observable $E(\boldsymbol{x})$. As shown in chapter 4 and 7, it can been used to construct proposal distributions to sample rare states in different observables E, $E(\boldsymbol{x}) = t\lambda_t$, $E(\boldsymbol{x}) = t_e(\boldsymbol{x})$, and $E(\boldsymbol{x}) = \log t_e(\boldsymbol{x})$. While it remains unclear what is the precise class of observables for which our methodology allows to construct an efficient proposal distribution, the different derivations of the proposal distribution do provide insights on the observables for which a proposal distribution could be constructed from properties of the system. As argued at the end of Sec 4.3.1, observables computed as an average along the trajectories are similar to the FTLE and therefore the proposal distribution, derived in Eq. 4.26 should lead to efficient algorithms in these cases.

Altogether, this thesis reveals a fascinating interplay between the chaotic nature of some non-linear systems and the numerical techniques available to study rare events in these systems. The analysis of this interplay allows to both better understand these systems and better understand these numerical techniques. This understanding opens perspectives to develop better techniques to numerical study

rare trajectories and extreme events in non-linear systems more generally, and therefore this thesis is useful to both people studying extreme events in non-linear systems and to people using numerical techniques.

Future perspectives

One question that this thesis leaves open is whether there is a proposal distribution that allows to sample systems with stickiness, and, in specific, sample long living trajectories in leaked Hamiltonian systems with mixed phase-space. Sampling long living trajectories in Hamiltonian systems with mixed phase-space is an important numerical problem for the understanding of stickiness in both low-dimensional and high-dimensional Hamiltonian systems [69–71]. Sec. 7.3 constructed a proposal distribution that generates states $\boldsymbol{x}'$ suitable to find long living trajectories (in $\log t_e$). However, this proposal leads to a very small acceptance rate, Fig. 7.13. This indicates that, under the hypothesis considered in chapter 4, this proposal is useful to *find* rare states (i.e. it solves the optimisation problem), but it is not useful to *sample* long living states with Metropolis-Hastings algorithm (i.e. it can't efficiently estimate the probabilities of events). There are two paths that can be pursued to tackle this problem. One path is to try to derive an expression for $t_\star(\boldsymbol{x})$ based on theories that explain stickiness, such as the Markov tree model[72, 106, 107]. Another path is to consider a proposal distribution that does not define a specific scale on the distance $\boldsymbol{x}' - \boldsymbol{x}$. The strong variability of the logarithm of the divergence, shown in Fig. 7.13, indicates that, even when two states have a small distance $\log t_e(\boldsymbol{x}') - \log t_e(\boldsymbol{x})$ (that $t_\star(\boldsymbol{x}) = 0.5t_e$ seems to guarantee), this still leads to a large variation of the respective divergences, that is responsible for the low acceptance rate. This suggest that even if these states are close in $E = \log t_e$, they are very different in $\lambda_t(\boldsymbol{x})$, which suggests that a proposal distribution with a specific scale may not be suitable to coup with the properties of the landscape. The challenge is to understand what are the statistical properties of this landscape that can be used in constructing a suitable proposal distribution.

A second question that this thesis does not explore but is worth pursuing is how the procedure described in chapter 4 is related to the theoretical notion of information. Conceptually, the proposal distribution is constructed by using specific pieces of information about the system, such as self-similarity of the escape time function. It is unclear how such pieces of information can be quantified in terms of information content. It would be interesting to investigate whether there is a quantitative connection between the information added to the proposal distribution and the efficiency of the Metropolis-Hastings procedure, and, if yes, whether there is an analytical relationship between these two.

Finally, a question that this thesis leaves open and is worth pursuing is whether the system needs to be fully deterministic for the ideas presented here be applicable. Specifically, a stochastic evolution given by a (high-dimensional) Langevian equation with a non-linear potential is often used as a model of a process because it allows to integrate specific fast degrees of freedom of the process, and focus on the time scale relevant for the events under study. When this model is evolved with a quenched noise

is in practice deterministic and thus, in principle, some of the approximations made in chapter 4 are still valid. This makes the methodology developed here in principle adaptable also to generate extreme events in these models. The noise added to each step t of evolution is equivalent to a random variable χ_t that can be interpreted as a phase-space variable such that the algorithm that now performs a random walk in the extended phase-space $(\boldsymbol{x}, \chi_1, ..., \chi_t)$, and the problem is to understand how the properties of the deterministic dynamics influence a proposal distribution on this extended phase-space. Particular problems where this generalisation can be useful include stochastic dynamics [11], transition paths in molecules [98], and anomalous diffusion [108].

Appendix

A. Variance of correlated samples

Consider $\{x_i\}$, $i = 1, ..., M$ samples from a second order stationary process (i.e. $\mathbb{E}[x]$ and $\mathbb{E}\left[x^2\right]$ are independent of i). One example of such process is the Metropolis-Hastings algorithm, which generates correlated states s_i according to a distribution $\pi(s)$. In this problem, the aim is to estimate an integral of the form

$$\mathbb{E}[f] = \int P(s)f(s)ds \quad , \tag{A.1}$$

which Metropolis-Hastings algorithm estimates by sampling (correlated) states s_i according to a distribution $\pi(s)$, and then use a consistent estimator for $\mathbb{E}[f]$ given by

$$\overline{f} = \frac{1}{M}\sum_{i=1}^{M}\frac{P(s_i)}{\pi(s_i)}f(s_i) \quad . \tag{A.2}$$

The aim here is to compute consistent estimators for the mean $\overline{f}$ and variance $\mathbb{V}\left[\overline{f}\right]$, by taking into account that $x_i = \frac{P(s_i)}{\pi(s_i)}f(s_i)$ are samples from a second order stationary process. In Metropolis-Hastings, the standard deviation $\sigma \equiv \sqrt{\mathbb{V}\left[\overline{f}\right]}$ quantifies the distance of the mean to its true value.

The estimator for the mean is given by

$$\overline{x} \equiv \frac{1}{M}\sum_{i=1}^{M} x_i \tag{A.3}$$

because $\mathbb{E}[\overline{x}] = \frac{1}{M}\sum_{i=1}^{M}\mathbb{E}[x_i] = \mathbb{E}[x]$. To compute the estimator for the variance of mean of x,

$$\mathbb{V}[\overline{x}] \equiv \mathbb{E}\left[\overline{x}^2\right] - \mathbb{E}[\overline{x}]^2 \quad , \tag{A.4}$$

notice that $\mathbb{E}[\overline{x}]^2 = \mathbb{E}[x]^2$. Furthermore, from the definition,

$$\overline{x}^2 = \left(\frac{1}{M}\sum_{i=1}^{M} x_i\right)^2 = \frac{1}{M^2}\sum_{i=1}^{M} x_i^2 + \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j\neq i}^{M} x_i x_j \quad . \tag{A.5}$$

Thus,

$$\mathbb{E}\left[\overline{x}^2\right] = \frac{1}{M}\mathbb{E}\left[x^2\right] + \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j\neq i}^{M}\mathbb{E}\left[x_i x_j\right] \tag{A.6}$$

Notice that the last term depends on the correlation between x_i and x_j.

Let us re-write the double summation in Eq. A.6 as follows:

$$\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j\neq i}^{M}\mathbb{E}\left[x_i x_j\right] = \frac{2}{M^2}\sum_{i=1}^{M}\sum_{\tau=1}^{M-i}\mathbb{E}\left[x_i x_{i+\tau}\right] \quad . \tag{A.7}$$

Recalling the definition of autocorrelation,

$$R\left(\tau\right) \equiv \frac{\mathbb{E}\left[x_t x_{t+\tau}\right] - \mathbb{E}\left[x\right]^2}{\mathbb{V}\left[x\right]} \quad , \tag{A.8}$$

and that the process is stationary (and thus $\mathbb{E}\left[x_i x_{i+\tau}\right]$ only depends on τ), Eq. A.7 can be written as

$$\begin{aligned}\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j\neq i}^{M}\mathbb{E}\left[x_i x_j\right] &= \frac{2}{M^2}\sum_{i=1}^{M}\sum_{\tau=1}^{M-i}\left(\mathbb{E}\left[x\right]^2 + \mathbb{V}\left[x\right]R(\tau)\right) \\ &= \frac{M^2-M}{M^2}\mathbb{E}\left[x\right]^2 + \frac{2}{M^2}\mathbb{V}\left[x\right]\sum_{i=1}^{M}\sum_{\tau=1}^{M-i}R(\tau) \quad .\end{aligned} \tag{A.9}$$

Inserting this in Eq. A.6 and $\mathbb{E}\left[\overline{x}^2\right]$ in Eq. A.4 gives

$$\mathbb{V}\left[\overline{x}\right] = \frac{1}{M}\mathbb{V}\left[x\right] + \frac{2}{M^2}\mathbb{V}\left[x\right]\sum_{i=1}^{M}\sum_{\tau=1}^{M-i}R(\tau) \quad . \tag{A.10}$$

Defining the integrated autocorrelated time as

$$T \equiv \frac{1}{M}\sum_{i=1}^{M}\sum_{\tau=1}^{M-i}R\left(\tau\right) \quad , \tag{A.11}$$

allows to write Eq. A.10 as

$$\mathbb{V}\left[\overline{x}\right] = \frac{\mathbb{V}\left[x\right]}{M}(1+2T) \quad . \tag{A.12}$$

Bibliography

[1] M. Ghil, P. Yiou, S. Hallegatte, B. D. Malamud, P. Naveau, A. Soloviev, P. Friederichs, V. Keilis-Borok, D. Kondrashov, V. Kossobokov, et al., Nonlinear Processes in Geophysics **18**, 295 (2011), ISSN 1607-7946. Cited on pages 1 and 2.

[2] O. Melnik and R. S. J. Sparks, **402**, 37 (1999), ISSN 0028-0836. Cited on page 1.

[3] P. Holmes, Physics Reports **193**, 137 (1990), ISSN 03701573. Cited on page 1.

[4] A. E. Motter and D. K. Campbell, Physics Today **66**, 27 (2013), ISSN 00319228. Cited on page 1.

[5] E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, Cambridge, 1993), 2nd ed., ISBN 0 521 43215 4. Cited on pages 1, 5, 7, 8, 9, 10, 60, and 75.

[6] D. Sweet, H. E. Nusse, and J. A. Yorke, Physical Review Letters **86**, 2261 (2001). Cited on pages 2, 15, 19, 21, 22, 47, 50, 51, 66, 69, 70, 87, and 88.

[7] E. M. Bollt, International Journal of Bifurcation and Chaos **15**, 1615 (2005). Cited on pages 2, 19, 21, 22, 71, and 88.

[8] Y.-C. Lai and T. Tél, *Transient chaos: Complex dynamics in finite time scales*, vol. 173 (Springer, New York, 2010), 1st ed., ISBN 978-1-4419-6986-6 e-ISBN. Cited on pages 2, 14, 15, 16, 17, 18, 19, 20, 22, and 88.

[9] J. Tailleur and J. Kurchan, Nature Physics **3**, 203 (2007), ISSN 1745-2473. Cited on pages 2, 13, 14, 22, 47, and 88.

[10] A. Kitajima and Y. Iba, Computer Physics Communications **182**, 251 (2011), ISSN 00104655. Cited on pages 2, 13, 14, 22, 47, 66, and 88.

[11] T. Laffargue, K.-D. N. T. Lam, J. Kurchan, and J. Tailleur, Journal of Physics A: Mathematical and Theoretical **46**, 254002 (2013), ISSN 1751-8113. Cited on pages 2, 12, 13, 14, 22, 64, 66, 88, and 91.

[12] P. Geiger and C. Dellago, Chemical Physics **375**, 309 (2010). Cited on pages 2 and 64.

[13] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, Annual review of physical chemistry **53**, 291 (2002), ISSN 0066-426X. Cited on page 2.

[14] M. Grünwald, C. Dellago, and P. L. Geissler, The Journal of chemical physics **129**, 194101 (2008), ISSN 1089-7690. Cited on pages 2, 40, 41, 43, 47, and 51.

[15] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Oxford University Press, USA, New York, 2002), ISBN 0198517971. Cited on pages 2, 27, 28, and 29.

[16] C. P. Robert and G. Casella, *Monte Carlo statistical methods*, Springer texts in statistics (Springer, Berlin, 2005), 2nd ed., ISBN 0387212396. Cited on pages 2, 19, 25, 26, 27, 29, 30, 35, 40, 59, and 88.

[17] P. Cvitanovic, R. Artuso, R. Mainieri, G. Tanner, and G. Vattay, *Chaos book* (Niels Bohr Institute, Copenhagen, 2016). Cited on pages 6, 7, 48, 70, and 75.

[18] N. Friedman, A. Kaplan, D. Carasso, and N. Davidson, Physical Review Letters **86**, 1518 (2001). Cited on page 10.

[19] U. Kuhl, H. J. Stockmann, and R. Weaver, Journal of Physics a-Mathematical and General **38**, 10433 (2005). Cited on page 10.

[20] T. Harayama and S. Shinohara, Laser and Photonics Reviews **5**, 247 (2011). Cited on page 10.

[21] E. G. Altmann, J. C. Leitão, and J. Viana Lopes, Chaos **22**, 026114 (2012), ISSN 10541500. Cited on pages 10, 14, and 19.

[22] E. G. Altmann, Ph.D. thesis, University of Wuppertal (2007). Cited on pages 10, 14, 16, 19, and 20.

[23] S. Dawson, C. Grebogi, T. Sauer, and J. Yorke, Physical Review Letters **73**, 1927 (1994), ISSN 0031-9007. Cited on page 10.

[24] T. Sauer, C. Grebogi, and J. Yorke, Physical Review Letters **79**, 59 (1997), ISSN 0031-9007. Cited on page 11.

[25] H. D. I. Abarbanel, R. Brown, and M. B. Kennel, Journal of Nonlinear Science **1**, 175 (1991), ISSN 0938-8974. Cited on pages 11 and 12.

[26] H. D. I. Abarbanel, R. Brown, and M. B. Kennel, Journal of Nonlinear Science **2**, 343 (1992), ISSN 0938-8974. Cited on pages 11 and 12.

[27] P. Grassberger and H. Kantz, Physics Letters A **113**, 167 (1985), ISSN 03759601. Cited on page 12.

[28] P. Grassberger, R. Badii, and A. Politi, Journal of Statistical Physics **51**, 135 (1988), ISSN 0022-4715. Cited on page 12.

[29] M. Sepúlveda, R. Badii, and E. Pollak, Physical review letters **63**, 1226 (1989), ISSN 1079-7114. Cited on pages 12 and 13.

[30] D. Beigie, A. Leonard, and S. Wiggins, Physical Review Letters **70**, 275 (1993), ISSN 0031-9007. Cited on pages 12 and 13.

[31] C. Amitrano and R. Berry, Physical review letters **68**, 729 (1992), ISSN 1079-7114. Cited on pages 12 and 13.

[32] S. Olmi, Chaos (Woodbury, N.Y.) **25**, 123125 (2015), ISSN 1089-7682. Cited on page 12.

[33] G. Zaslavsky, Physica D: Nonlinear Phenomena **168-169**, 292 (2002), ISSN 01672789. Cited on page 12.

[34] J. Szezech, S. Lopes, and R. Viana, Physics Letters A **335**, 394 (2005), ISSN 03759601. Cited on page 12.

[35] R. Artuso and C. Manchein, Physical Review E **80**, 036210 (2009), ISSN 1539-3755. Cited on page 12.

[36] C. Manchein, M. W. Beims, and J. M. Rost, Chaos (Woodbury, N.Y.) **22**, 033137 (2012), ISSN 1089-7682. Cited on page 12.

[37] C. Beck and F. Schögl, *Thermodynamics of chaotic systems: an introduction* (Cambridge University Press, 1995), ISBN 9780521433679. Cited on pages 12, 13, and 60.

[38] A. Prasad and R. Ramaswamy, Physical Review E **60**, 2761 (1999), ISSN 1063-651X. Cited on pages 12, 13, 60, and 73.

[39] U. H. E. Hansmann and Y. Okamoto, Journal of Computational Chemistry **14**, 1333 (1993), ISSN 0192-8651. Cited on pages 13 and 21.

[40] E. G. Altmann, J. S. E. Portela, and T. Tél, Reviews of Modern Physics **85**, 869 (2013), ISSN 0034-6861. Cited on pages 14, 17, 20, and 75.

[41] Q. Chen, M. Ding, and E. Ott, Physics Letters A **145**, 93 (1990), ISSN 03759601. Cited on pages 14 and 18.

[42] L. A. Bunimovich and C. P. Dettmann, Epl **80** (2007). Cited on page 14.

[43] T. Tél and Y.-C. Lai, Physics Reports **460**, 245 (2008), ISSN 03701573. Cited on page 14.

[44] J. Wang, P. G. Soerensen, and F. Hynne, The Journal of Physical Chemistry **98**, 725 (1994), ISSN 0022-3654. Cited on pages 14 and 17.

[45] C. P. Dettmann and O. Georgiou, Journal of Physics A:Mathematical and Theoretical **44**, 195102 (2011). Cited on page 14.

[46] F. Mortessagne, O. Legrand, and D. Sornette, Chaos **3**, 529 (1993). Cited on page 14.

[47] O. Legrand, The Journal of the Acoustical Society of America **88**, 865 (1990), ISSN 00014966. Cited on page 14.

[48] G. S. Ezra, H. Waalkens, and S. Wiggins, The Journal of chemical physics **130**, 164118 (2009), ISSN 1089-7690. Cited on page 14.

[49] J. U. Nöckel and A. D. Stone, Nature **385**, 45 (1997). Cited on page 14.

[50] C. Gmachl, F. Capasso, E. E. Narimanov, J. U. No, A. D. Stone, D. L. Sivco, and A. Y. Cho, Science **280**, 1556 (1998). Cited on page 14.

[51] E. G. Altmann, Physical Review A **79**, 013830 (2009). Cited on page 14.

[52] S. Shinohara, M. Hentschel, J. Wiersig, T. Sasaki, and T. Harayama, Physical Review A **80** (2009), ISSN 1050-2947. Cited on page 14.

[53] S. Shinohara, T. Harayama, T. Fukushima, M. Hentschel, T. Sasaki, and E. E. Narimanov, Physical Review Letters **104**, 163902 (2010). Cited on page 14.

[54] C. Jung, T. Tél, and E. Ziemniak, Chaos **3**, 555 (1993). Cited on page 14.

[55] X. Ni, L. Ying, Y.-C. Lai, Y. Do, and C. Grebogi, Physical Review E **87**, 052911 (2013), ISSN 1539-3755. Cited on page 14.

[56] X. Ni and Y.-C. Lai, Chaos (Woodbury, N.Y.) **21**, 033116 (2011), ISSN 1089-7682. Cited on page 14.

[57] H. Kantz and P. Grassberger, Physica D: Nonlinear Phenomena **17**, 75 (1985), ISSN 01672789. Cited on page 16.

[58] G.-H. Hsu, E. Ott, and C. Grebogi, Physics Letters A **127**, 199 (1988), ISSN 03759601. Cited on page 16.

[59] A. Block, W. von Bloh, and H. J. Schellnhuber, Physical Review A **42**, 1869 (1990), ISSN 1050-2947. Cited on page 16.

[60] M. Dhamala, Y. C. Lai, and E. J. Kostelich, Physical review. E, Statistical, nonlinear, and soft matter physics **64**, 056207 (2001), ISSN 1539-3755. Cited on page 16.

[61] M. J. Körber, M. Michler, A. Bäcker, and R. Ketzmerick, Physical review letters **111**, 114102 (2013), ISSN 1079-7114. Cited on page 16.

[62] M. Schönwetter and E. G. Altmann, Physical review. E, Statistical, nonlinear, and soft matter physics **91**, 012919 (2015), ISSN 1550-2376. Cited on page 16.

[63] S. K. Scott, B. Peng, A. S. Tomlin, and K. Showalter, The Journal of Chemical Physics **94**, 1134 (1991), ISSN 00219606. Cited on page 17.

[64] J. Peixinho and T. Mullin, Physical review letters **96**, 094501 (2006), ISSN 0031-9007. Cited on page 17.

[65] M. Thiel, J. Kurths, M. C. Romano, G. Károlyi, and A. Moura, eds., *Nonlinear Dynamics and Chaos: Advances and Perspectives*, Understanding Complex Systems (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010), ISBN 978-3-642-04628-5. Cited on page 18.

[66] M. Richter, S. Lange, A. Bäcker, and R. Ketzmerick, Physical Review E **89**, 022902 (2014), ISSN 1539-3755. Cited on page 19.

[67] R. Bellman, *Dynamic Programming* (Princeton University Press, 1957), ISBN 069107951X. Cited on page 19.

[68] H. E. Nusse and J. A. Yorke, Physica D: Nonlinear Phenomena **36**, 137 (1989), ISSN 01672789. Cited on pages 19 and 21.

[69] A. E. Motter, A. P. S. de Moura, C. Grebogi, and H. Kantz, Physical Review E **71**, 1 (2005), ISSN 1539-3755. Cited on pages 19 and 90.

[70] G. Cristadoro and R. Ketzmerick, Physical Review Letters **100**, 184101 (2008). Cited on pages 19 and 90.

[71] R. Venegeroles, Physical Review Letters **102**, 64101 (2009). Cited on pages 19 and 90.

[72] J. D. Meiss and E. Ott, Physica D **20**, 387 (1986). Cited on pages 19 and 90.

[73] S. Lange, M. Richter, F. Onken, A. Bäcker, and R. Ketzmerick, Chaos (Woodbury, N.Y.) **24**, 024409 (2014), ISSN 1089-7682. Cited on page 19.

[74] E. G. Altmann and T. Tél, Physical Review E **79**, 16204 (2009). Cited on page 20.

[75] G. Torrie and J. Valleau, Journal of Computational Physics **23**, 187 (1977), ISSN 00219991. Cited on pages 21 and 29.

[76] R. H. Swendsen and J.-S. Wang, Physical Review Letters **57**, 2607 (1986), ISSN 0031-9007. Cited on pages 21 and 29.

[77] E. Marinari and G. Parisi, Europhysics Letters (EPL) **19**, 451 (1992), ISSN 0295-5075. Cited on pages 21 and 29.

[78] Y. Sugita and Y. Okamoto, Chemical Physics Letters **314**, 141 (1999), ISSN 00092614. Cited on pages 21 and 29.

[79] J. Lee, Physical review letters **71**, 211 (1993), ISSN 1079-7114. Cited on pages 21 and 29.

[80] B. A. Berg and T. Neuhaus, Physics Letters B **267**, 249 (1991), ISSN 03702693. Cited on pages 21 and 29.

[81] C. Grebogi, E. Kostelich, E. Ott, and J. A. Yorke, Physica D **25**, 347 (1987), ISSN 01672789. Cited on page 21.

[82] M. Wagner, Ph.D. thesis, Technical University of Dresden (2015). Cited on page 21.

[83] J. C. Leitão, J. M. V. P. Lopes, and E. G. Altmann, Physical Review Letters **110**, 220601 (2013), ISSN 0031-9007. Cited on pages 22, 46, 47, 48, 49, 50, and 65.

[84] J. C. Leitão, J. M. V. P. Lopes, and E. G. Altmann, Physical Review E **90**, 052916 (2014), ISSN 1539-3755. Cited on pages 22, 43, 47, 58, 59, 60, and 63.

[85] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, The Journal of Chemical Physics **21**, 1087 (1953), ISSN 00219606. Cited on page 25.

[86] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science (New York, N.Y.) **220**, 671 (1983), ISSN 0036-8075. Cited on pages 28 and 57.

[87] V. Černý, Journal of Optimization Theory and Applications **45**, 41 (1985), ISSN 0022-3239. Cited on page 28.

[88] F. Wang and D. P. Landau, Physical Review Letters **86**, 2050 (2001), ISSN 0031-9007. Cited on pages 29 and 30.

[89] C. G. Zhou and R. N. Bhatt, Physical Review E **72**, 025701 (2005), ISSN 1539-3755. Cited on page 30.

[90] R. E. Belardinelli, S. Manzi, and V. D. Pereyra, Physical Review E **78**, 067701 (2008). Cited on page 30.

[91] S. Trebst, D. A. Huse, and M. Troyer, Physical Review E **70**, 046701 (2004). Cited on pages 30 and 31.

[92] P. Dayal, S. Trebst, S. Wessel, D. Wurtz, M. Troyer, S. Sabhapandit, and S. N. Coppersmith, Physical Review Letters **92**, 097201 (2004). Cited on pages 30 and 31.

[93] J. Viana Lopes, Ph.D. thesis, Universidade do Porto (2006). Cited on pages 30, 31, 36, 51, and 57.

[94] M. D. Costa, J. V. Lopes, and J. M. B. L. dos Santos, Europhysics Letters (EPL) **72**, 802 (2005), ISSN 0295-5075. Cited on page 30.

[95] G. O. Roberts, A. Gelman, and W. R. Gilks, The Annals of Applied Probability **7**, 110 (1997), ISSN 2168-8737. Cited on pages 31, 36, and 88.

[96] R. Fischer, J. C. Leitão, T. P. Peixoto, and E. G. Altmann, Physical review letters **115**, 188701 (2015), ISSN 1079-7114. Cited on page 36.

[97] P. Grassberger, Physical Review E **56**, 3682 (1997), ISSN 1063-651X. Cited on page 36.

[98] C. Dellago, P. Bolhuis, and P. Geissler, Advances in Chemical Physics (2002). Cited on pages 39 and 91.

[99] C. Grebogi, S. Hammel, J. Yorke, and T. Sauer, Physical Review Letters **65**, 1527 (1990), ISSN 0031-9007. Cited on page 40.

[100] F. Bouchet, S. Gupta, and D. Mukamel, Physica A: Statistical Mechanics and its Applications **389**, 4389 (2010), ISSN 03784371. Cited on page 44.

[101] Y. Iba, N. Saito, and A. Kitajima, Annals of the Institute of Statistical Mathematics **66**, 611 (2014), ISSN 0020-3157. Cited on page 47.

[102] P. Charbonneau, J. Kurchan, G. Parisi, P. Urbani, and F. Zamponi, Nature communications **5**, 3725 (2014), ISSN 2041-1723. Cited on page 57.

[103] B. Kryzhanovsky and M. Malsagov, Optical Memory and Neural Networks **25**, 1 (2016), ISSN 1060-992X. Cited on page 57.

[104] D. J. Wales, *Energy Landscapes Applications to Clusters, Biomolecules and Glasses* (Cambridge University Press, 2003). Cited on page 57.

[105] Y. Pomeau and P. Manneville, Communications in Mathematical Physics **74**, 189 (1980), ISSN 0010-3616. Cited on page 74.

[106] J. D. Hanson, J. R. Cary, and J. D. Meiss, Journal of Statistical Physics **39**, 327 (1985), ISSN 0022-4715. Cited on page 90.

[107] O. Alus, S. Fishman, and J. D. Meiss, Physical Review E **90**, 062923 (2014), ISSN 1539-3755. Cited on page 90.

[108] J. Klafter and I. M. Sokolov, *First Steps in Random Walks: From Tools to Applications* (OUP Oxford, 2011), ISBN 0199234868. Cited on page 91.

Danksagung / Acknowledgements

This thesis would not be possible without the invaluable supervision and guidance of Eduardo G. Altmann, for which I am truly indebtedly grateful. He is an inspiring leader and made his group "social dynamics and dynamical systems" an exceptional environment to be in.
I would also like to thank my great friend and collaborator Dr. João Parente Lopes for the invaluable support and discussions that resulted in several ideas presented in this thesis, and José Miotto for the invaluable and stimulating discussions we had on numerous topics ranging from philosophy and economics to p-value and Korgomorov-Smirnov test, which contributed to the intellectually rewarding environment of the past 3 years. The results in section 6.3 were obtained together with Dr. Matteo Sala, to whom I am grateful for the collaboration. I would also like to thank my friends and colleagues Moritz Schönwetter and Martin Gerlach for the stimulating discussions that inspired different ideas presented in this thesis, as well as other people in the group: Julian, Orestis, Shamik, Rico, and Fakhteh for their invaluable discussions and support.
I would also like to thank my family, in particular Marta Matos, Ana Lopes, and Pedro Leitão, for their unconditional support during the past 3.5 years.
I would also like to acknowledge the Max Planck Institute for the Physics of Complex Systems (PKS), and in particular its administrative and IT staff, for the excellent conditions it provided. Finally, I would like to acknowledge PKS for funding the first 6 months of my stay and Fundação para a Ciência e Tecnologia for funding 3.5 years of my stay in the PKS and in Germany.

Versicherung

Diese Arbeit wurde am Max-Planck-Institut für Physik komplexer Systeme unter der wissenschaftlichen Betreuung Prof. Dr. Holger Kantz durchgeführt.
Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.
Darüber hinaus erkenne ich die Promotionsordnung der Fakultät Mathematik und Naturwissenschaften der Technischen Universität Dresden vom 23 Februar 2011 an.

Jorge C. Leitão

Dresden, 31. Mai 2016