Estimating transitions using the Potential-Theoretic approach and a data-based method

A MASTER THESIS IN MATHEMATICAL SCIENCE
UTRECHT UNIVERSITY

Francesca Cavallini

SUPERVISOR
prof. dr. ir. J.E. Frank

SECOND READER
dr. W.M. Ruszel

Utrecht University
October 2021

Abstract

A large variety of random dynamical systems describing real world phenomena stay in a neighbourhood of a stable equilibrium for a long period of time. When zooming out to longer time scales, one can observe the system spontaneously transitioning to an alternate stable state. The transition is rapid and occurs after a random time. Is it possible to estimate quantities related to transitions only using a relatively small amount of data from a time series of the system state? Can one do it without explicitly knowing the underlying equations for the system?

We address the problem for a specific class of stochastic processes, namely continuous-time Markov processes which represent solutions to Stochastic Differential Equations (SDEs). We propose a method based on the Potential-Theoretic approach to metastability developed by Bovier and Den Hollander. The Potential Theory provides the theoretical framework of the present work and, importantly, it defines the Partial Differential Equation (PDE) with Dirichlet boundary conditions associated with the stochastic process that we aim to solve. We combine the theory with numerical methods in order to estimate the unknown stochastic generator of the process, by means of a series of approximations. The recent Extended Dynamic Mode Decomposition method inspires the data-based approach and a Galerkin finite element method is used to solve the PDE.

As a result, a Dirichlet problem is successfully solved for a paradigmatic model for metastability, namely the stochastic diffusion in a double-well potential in one dimension. Our algorithm finds a data-based approximation of the equilibrium potential for this example.

Contents

Introduction

Metastability is a widespread phenomenon that arises in a large variety of complex systems - physical, chemical, biological or economic. It describes the property of a dynamical system to stay long in a stable state before spontaneously transitioning to an alternate stable state [1]. These systems are characterized by the presence of two time scales: on short time scales, one can observe the system staying in a neighbourhood of an apparent equilibrium; on long time scales, rapid transitions between equilibria in different regions of the state space are observed.

Many real world phenomena show metastable behaviours. For example, the mix of two chemical reactants may lead to a metastable state (like a hydrogen and oxygen gases mixture) that can persist for a very long time without activating the reaction but, when triggered, the compound instantly transitions to a coexisting stable state (like water). Other examples of complex systems manifesting this behaviour are found in conformational dynamics. Bio-molecules such as proteins can be found in nature in different conformations, the stable states, and they can transition from one to the other under the influence of thermal noise, the latter being called 'conformational transitions' (Figure 1). Studies have been developed to understand the lifetime and the nature of these states [2] . Different models studied in population dynamics are interpreted in terms of transitions too. Selective sweeps in the genes of populations, triggered by mutations that drive the population from one dominant trait to another, can be viewed as transitions between metastable states. And again, viruses moving through a complex network may cause an epidemic, which might be interpreted as a metastable state of the network that lasts until the virus disappears [3] (Figure 2). This metastable state depends sensitively on the size and the architecture of the network and can produce very complex patterns. These are just a few examples that motivate the interest in such systems.

The main challenges in metastability include identifying the metastable states, estimating how long it takes for the system to transition to a new state (*transition times*), how often transitions do happen (*rate of transitions*) and how likely they are (*transition probabilities*). From a mathematical point of view, many complex systems showing metastability are described by non-linear dynamical systems or stochastic dynamical systems. In particular, one popular class of systems that are investigated in the present work are stochastic diffusion processes. Different theoretical approaches to study metastability are developed in the literature. There are several framework and tools to produce estimates of the aforementioned fundamental quantities and that may yield reliable results with different accuracies under stringent hypothesis. Two opposite approaches are presented in this thesis, namely the Pathwise approach and the Potential-Theoretic approach, the latter being fur-

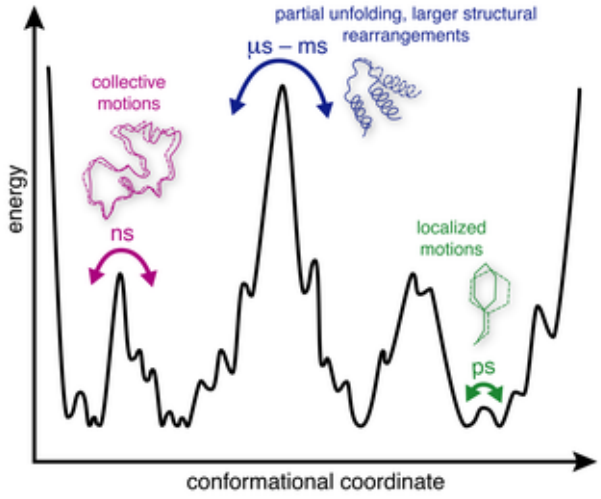

Figure 1: Source: [2]. One-dimensional sketch of the energy landscape which determines the complex dynamics of a protein. Each minimum corresponds to a different protein structure and arrows represent barriers crossovers which correspond to conformational transitions. Smaller barriers are overcome more frequently while larger barriers more rarely.

ther investigated. Subsequently, a new computational approach is proposed which combines the latter with numerical data-driven techniques. Our method is built on the Potential Theory and it relies only on data, making the method effective in different scenarios.

The classical approach to metastability is the Pathwise approach developed by Freidlin and Wentzell in the 70's and is based on the theory of Large Deviations [5]. It gives the most likely path from one metastable state to another one, by means of a variational problem. It describes well the metastable behaviour and produces asymptotic estimates of the transition times. However, most of the information can be obtained from the *rate function* associated with the system which is usually hard to identify. By contrast, the more recent Potential-Theoretic approach presented in the book by Bovier and den Hollander [1] does not consider the system trajectory as a continuous path in the state space. In fact, it rather sees the metastable behaviour as a sequence of visits of a stochastic process to different stable states, focusing on the mean hitting time of these sets. One can intuitively think of the stable states as vertices of a network, and transitions as the connections between vertices. Recent overviews of metastability can be found in in the review papers [6] and [7].

The Potential Theory is proposed as a more effective alternative to the Pathwise approach. It aims for sharper estimates of transition times then the ones achieved by Freidlin and Wentzell's Theory (not only the exponential asymptotics but also the coefficients in the Kramers formula). The goal is to give estimates of transition probabilities, hitting probabilities and expected hitting times of the states. As a first step, they use the classical idea that some fundamental quantities, for instance the *equilibrium potential*, are (or can be inferred from) *solutions of Dirichlet problems*, namely Partial Differential Equations with proper Dirichlet boundary conditions.

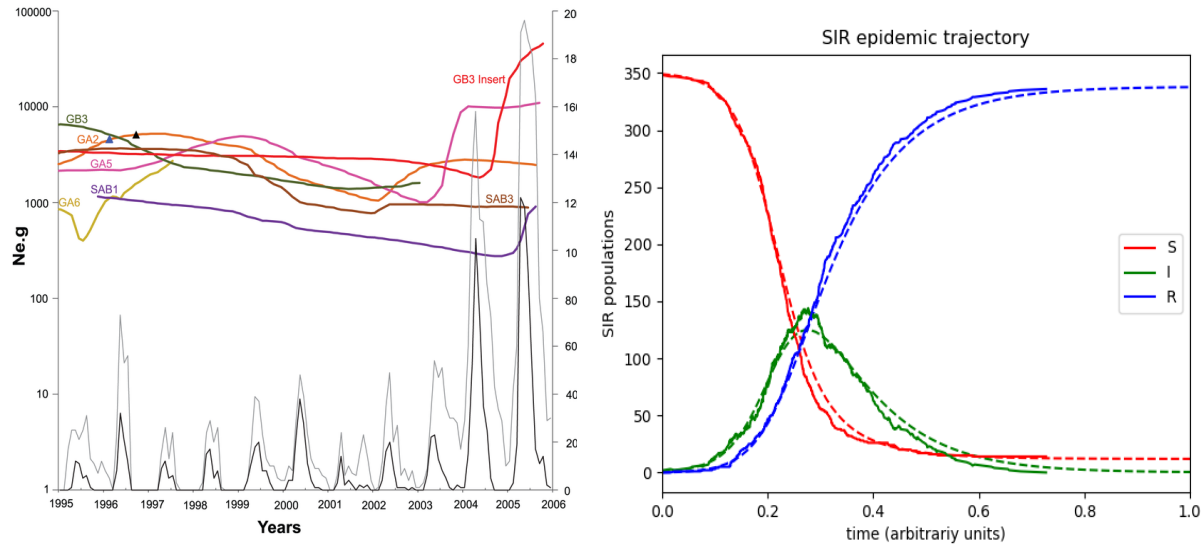

Figure 2: On the left, different genotypes of Human Respiratory Syncytial virus prevalent in São Paulo and seasonal distribution of HRSV cases in São Paulo during 1995–2005. The y-axis (on the left) represents a measure reflecting the change in effective number of infections over time. Genotypes dramatically increase at some random times (source :[3]). On the right, the SIR model. We observe spontaneous transition of population from a majority of susceptibles to a majority of recoveries, corresponding to an outbreak of the epidemic (source: [4]).

When the system comes from a continuous stochastic process, the PDE involves its infinitesimal generator, which is a linear differential operator. However, in most of the cases they can be high-dimensional and very hard to solve, both numerically and analytically. The key observation of the Potential-Theoretic approach is that it is possible bound those quantities in terms of *capacities of sets*. Capacities are quantities associated to pairs of subsets of the domain of the problem and play a crucial role in this approach. Importantly, they can be estimated by means of a variational problem and produce sharp estimates for the average hitting time of sets and for the equilibrium potential. Remarkably, very good estimates can be obtained for some known reversible processes without solving the high-dimensional Dirichlet problems. The so called *Dirichlet quadratic form* defines the variational problem that makes this theory so effective. Nevertheless, it requires to know the explicit equation of the model.

The method developed in this work makes use of the definitions and problem statements of the Potential-Theoretic approach but it is applicable in more diverse and practical cases of study. As an example, suppose we want to study a phenomenon manifesting metastable behaviour and we can observe the system. Furthermore, we are able to collect data by taking measurements. In other words, we have a time series of the system state that shows random 'jumps' between two stable states and we model that with a stochastic process, but we do not know the complete equations for it. In this work, we address the following questions. How can one infer accurate information about transitions from a relatively short time series? Is it possible to do it without explicitly knowing the generator of the stochastic process?

For this purpose, we choose to adopt the framework of Potential Theory. We recall that solving the relatively simple variational problem mentioned before (the one with the Dirichlet form) is not possible in this context, as it requires the complete knowledge of the model. Instead, we propose to solve numerically the Dirichlet problems to retrieve the equilibrium potential and hitting times and probabilities of the process.

If the equations are not available, an approximation of the generator of the process is needed in order to define the boundary value problems. To this purpose, firstly, we investigate the numerical methods in the literature that are used to reconstruct the dynamics of a system from data, i.e. observations of the evolution of the states. One method of particular interest is the Dynamic Mode Decomposition (DMD), which provides approximations of the modes and eigenvalues of the composition operator, also called the *Koopman operator*. The Koopman operator is a linear infinite-dimensional operator acting on the space of *observables* and strictly related to the flow map of the dynamical system. For stochastic data, an alternative version of the method is applied, called Extended Dynamic Mode Decomposition (EDMD), developed by Williams in 2015 [8]. Secondly, a crucial step in this project is to understand the relationships between the underlying stochastic process, the Koopman operator, and the associated PDE. Remarkably, we find that, under some assumptions, and for specific classes of system, the stochastic counterpart of the Koopman Operator corresponds to the infinitesimal generator that defines the PDE of the aforementioned Dirichlet problems. Therefore, this approximation contains precious information that can be used to answer the questions of metastability. Interestingly, similar ideas are developed in [9]. It is shown that, by means of Spectral Theory, the real eigenfunctions can provide a decompostion of the state space into metastable sets. Finally, the Galerkin finite element method provides us with a tool to solve PDE in one and higher dimensions, which combined with other techniques completes our equation-free method.

Compared to the standard approaches, it shows some advantages. On the one hand, the equations of motion of the systems might be very difficult to analyse by hand, or, in some cases, they might be even unknown. The problems arising in the Potential-Theoretic approach are solved in [6] for a one-dimensional diffusion process, and lead to Kramer formulas [10] for diffusion in a double well potential. However, finding the exact solutions is far too complex for stochastic diffusion models in higher dimensions, and no closed formulas of transition times are available. A computational approach might be preferable to avoid this issue.
On the other hand, when observations of the system are available in the form of a time series, data tells us about the evolution with respect to some measure. One could exploit this information and use data to estimate quantities describing the phenomenon. However, because of the random occurrence of transitions, it might be difficult to extract estimates of the quantities from data, as this requires extremely long simulations. In contrast, fewer data snapshots are needed to estimate the generator of the process with Dynamic Mode Decomposition.

To summarize, the goal of this thesis is to motivate, explain and implement a computational method to estimate quantities related to transitions, which does not require running long simulations and is based on the Potential Theory. This implies

the estimation of the stochastic generator from data. In the present work, we combine data-based techniques that are used to retrieve the generator of an unknown process with a theory on metastability, which reveals the information carried by it. The content is organised in three chapters. Chapter 1 is an overview of the theory, includes the Pathwise approach of Large Deviation and the Potential-Theoretic approach, with theory results for some specific examples on Brownian motion and stochastic diffusion. The fundamental quantities of potential theory are defined and the results one can infer from the theory are explained here. In Chapter 2, we delve into the numerical methods such as DMD and EDMD and how they are linked with Koopman theory. We describe our own algorithm used to estimate the generator and list the approximations required together with an estimate of the order of their error contributions. In Chapter 3 we apply the algorithm to a specific problem arising in the theory, namely the Dirichlet problem that defines the equilibrium potential. We solve the problem and explain the interpretation of the solution in terms of stochastic processes.

Chapter 1

Two different approaches to the study of metastability

The object of our study is metastability as the product of a non-linear system dynamics subject to weak random forces or, in statistical physics, also seen as the manifestation of a first order phase transition. By looking at the time series of a metastable system under the effect of a noisy dynamics, one could observe the system moving between different regions of the state space at different time scales: on short time scales the system is in equilibrium within a single region; on long time scales it undergoes rapid transitions to alternate stable states. The transitions are rapid and happen after a long and random time.

In this chapter, two mathematical approaches to study metastability and, in particular, the occurrence of transitions are explained. We compare the techniques used and the results obtained for processes described by stochastic differential equations. Firstly, we summarize the traditional Pathwise approach, developed in the late 1960's by M.I. Freidlin and A.D. Wentzell. We then present the foundations of an alternative and more recent way to tackle metastability, namely the Potential-Theoretic approach, which provides the theoretical setting of the remainder chapters. The exposition in this chapter is based on the book by A. Bovier and F. den Hollander [1], and as such, we use the notation found in their work.

1.1 The Pathwise approach

Freidlin and Wentzell propose the Theory of Large Deviations on path space and apply it to study the behaviour of dynamical systems subject to the effect of small random perturbation on large time scales. The theory is developed in their book *Random Perturbations of Dynamical Systems* [5]. Models of small random perturbations depending on a parameter ϵ are considered, such as the Markov process solution to the stochastic equation

$$dX_t^\epsilon = b(X_t^\epsilon)dt + \epsilon\sigma(X_t^\epsilon)dW_t \tag{1.1}$$

where X_t^ϵ is continuous random variable on $\mathbb{R}^d$, W_t is an m-dimensional Brownian motion, $b : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ is the drift, $\sigma : \mathbb{R}^d \longrightarrow \mathbb{R}^{d\times m}$ is the diffusion coefficient and $\epsilon \in \mathbb{R}$.

They rely on large deviation theory to study the asymptotic behaviour of probabilities of rare events (i.e. *large deviations* from the unperturbed process) as $\epsilon \to 0$. This approach describes well the metastable dynamics and produces estimates of the crossover time.

1.1.1 Large deviations on path space for stochastic process

One central object in the theory of large deviation is the *rate function.*

Definition 1 *[1] A family of probability measures $(\mu_\epsilon)_{\epsilon>0}$ on a Polish space $\mathcal{H}$ is said to satisfy the large deviation principle (LDP) with rate function $I : \mathcal{H} \to [0, \infty]$ if:*
(i) I has compact level sets and is not identically infinite,
(ii) $\liminf_{\epsilon\to 0} \epsilon \ln \mu_\epsilon(O) \geq -I(O)$ for all $O \subset \mathcal{H}$ open,
(iii) $\limsup_{\epsilon\to 0} \epsilon \ln \mu_\epsilon(C) \leq -I(C)$ for all $C \subset \mathcal{H}$ closed,
where, for every $S \subset \mathcal{H}$, $I(S)$ is defined as $I(S) = \inf_{x\in S} I(x)$, .

A more intuitive definition of Large Deviation Principle is given in [1]. It states that if $B_\delta(x)$ is a ball of radius $\delta > 0$ centered at $x \in \mathcal{H}$, then

$$\mu_\epsilon(B_\delta(x)) = e^{-[1+o(1)]I(x)/\epsilon} \tag{1.2}$$

when $\epsilon \to 0$ followed by $\delta \to 0$.

In other words, the measure of any ball centered in x w.r.t. the law μ_ϵ tends to zero exponentially fast as $\epsilon \to 0$, whenever $I(x) > 0$. Thus, if $I(x) > 0$, x is a (exponentially) rare event w.r.t. the probability measure μ_ϵ for small ϵ. Note that, in general, the most probable state is the one at which I takes its smallest value. According to this probabilistic interpretation, the Large Deviation Principle is a distributional property that gives a rough description of the asymptotic behavior of the probabilities of rare events. The rate function I is used to quantify these probabilities.

In Freidlin and Wenzell's Pathwise approach, $\mathcal{H}$ is a function space and rate functions take the form of action functionals.

We include three models which satisfy the Large Deviation properties and for which an action functional can be derived. The results for the Brownian motion and Brownian motion with drift are known as Schilder's theorems, while the Freidlin–Wentzell theorem represents a generalization for diffusion processes.

Brownian motion $B = (B_t)_{t\in\mathbb{R}^+}$

A sample path B of the Brownian motion started at 0, is typically at a distance of order $\sqrt{t}$ from the origin at time t ($B_t/t \to 0$ as $t \to \infty$). We want to compute the probability that B strays from the most likely trajectory and follows a path such that B_t lives on space scale t. We will see how to find the most likely path later. Formally, we fix a time-horizon $T > 0$, and consider any smooth path $\gamma : [0, T] \to \mathbb{R}^d$ starting at the origin and ending somewhere else in the state space. We want to estimate the probability

$$\mathbb{P}(\sup_{s\in[0,T]} ||\epsilon B_{s/\epsilon} - \gamma(s)|| \leq \delta) \tag{1.3}$$

for $\epsilon \to 0$. Here we can see that $t = s/\epsilon$, so in the limit we are considering the infinite time scale.

Let $\mathcal{H} = C_0([0,T])$ be the space of continuous functions $f : [0,T] \to \mathbb{R}^d$ such that $f(0) = 0$ equipped with the supremum norm $||f||_\infty$, and set $B^\epsilon = (\epsilon B_{s/\epsilon})_{s\in[0,T]}$ in order to have a scaled process on a compact interval $[0,T]$. Then, for $\gamma(t) \in \mathcal{H}$, 1.3 is equivalent to the probability

$$\mathbb{P}(B^\epsilon \in B_\delta(\gamma)) \tag{1.4}$$

where $B_\delta(\gamma)$ is the ball centered in γ of radius δ.

We state without proof the Shilder's theorem for the Brownian motion. The proof is presented in [1].

Theorem 1.1 *The family of stochastic processes B^ϵ satisfies the LDP on $\mathcal{H} = C_0([0,T])$ and the rate function $I : \mathcal{H} \longrightarrow \mathbb{R}^+$ is*

$$I(\gamma) = \frac{1}{2}\int_0^T |\dot\gamma(s)|^2 ds \tag{1.5}$$

if γ is absolutely continuous with square-integrable derivative. $I(\gamma) = \infty$ otherwise.

Hence, by applying the informal definition of the Principle 1.2, we find the asymptotic behaviour of probability in 1.4 as $\epsilon \to 0$.

Brownian motion with drift $X = (X_t)_{t\in\mathbb{R}^+}$

We now consider a stochastic process satisfying a stochastic differential equation of the form

$$X_t = B_t + \int_0^t b(X_t)ds, \quad t \in \mathbb{R}^+ \tag{1.6}$$

where $b : \mathbb{R} \longrightarrow \mathbb{R}$ is a globally Lipschitz function. Again we fix a time interval $[0,T]$ and we consider the scaled processes B^ϵ as before so that $X^\epsilon = (X^\epsilon_s)_{s\in[0,T]}$. The following theorem gives an expression for the rate function.

Theorem 1.2 *The family of processes X^ϵ satisfies the LDP on $\mathcal{X} = C_0([0,T])$ with rate function*

$$I(\gamma) = \frac{1}{2}\int_0^T |\dot\gamma(s) - b(\gamma(s))|^2 ds. \tag{1.7}$$

It is possible to extend the argument to more general diffusion and Markov processes. We do not include that in this work and we rather continue with the application to metastability.

The theorems above guarantee the Large Deviation Principle for the given stochastic processes and give an expression for the probability of a solution connecting two points u and v in the phase space in a time T. We have that

$$\lim_{\delta\to 0}\lim_{\epsilon\to 0} \epsilon \ln \mathbb{P}(X^\epsilon_T \in B_\delta(v) | X^\epsilon_0 \in B_\delta(u)) = -\inf_{\gamma:\gamma(0)=u,\gamma(T)=v} I(\gamma). \tag{1.8}$$

Here the arrival time at v is fixed, but one can also choose to allow arbitrary long paths. Freidlin and Wentzell define the *quasi potential* between u and v

$$V(u,v) = \inf_{T<\infty} \inf_{\gamma:\gamma(0)=u,\gamma(T)=v} I(\gamma) \tag{1.9}$$

If u is an asymptotically stable equilibrium and v an arbitrary point in the phase space, one can interpret the quasi-potential $V(u,v)$ with respect to u as the cost to go against the flow from u to v. Moreover, V plays a role in the computation of two central quantities: the escape probability from u to v and the *first hitting time* of the ball $B_\delta(v)$. In fact they proved that $-V(u,v)$ is the exponential asymptotics of the escape probability from u to v.

Finding the escaping probabilities and exit times from a stable solution reduces to a variational problem. In practice, one has to minimize the rate function I, which, in the examples above has the form of a classical action functional.

We show the results for the special case of diffusion with drift generated by a potential F, such that $b(x) = -\nabla F$ for $x \in \mathbb{R}^d$.

Example: Brownian motion in a double-well

Let consider a double-well potential F as in Figure 1.1. We observe that the rate functions I in Section 1.1.1 have the properties of the classical action functionals in Newton mechanics, where, for a motion with drift, the Lagrangian takes the form

$$\mathcal{L}(\gamma(s), \dot\gamma(s), s) = \frac{1}{2}\int_0^T |\dot\gamma(s) + \nabla F(\gamma(s))|^2 ds. \tag{1.10}$$

According to the principle of least action, the most likely paths connecting two points in the state space are the stationary points of the functional I. One can find the solutions to this variational problem, by solving the Euler-Lagrange equations, and find which is the most likely trajectory that the process follows starting from u to v.

It is shown in [5] that the solution γ of the variational problem in 1.9, satisfies

$$\dot\gamma(t) = -\nabla F(\gamma(t)).$$

$I(\gamma) = 0$ and its time reversal $\bar\gamma$ defines the most likely path going uphill from v to u. Importantly, the cost of leaving the potential is $V(v,u) = 2[F(z^*) - F(v)]$. In conclusion, the results on the transitions from one potential well to the other one are:

- the transition probability verifies

$$\mathbb{P}(X_t^\epsilon \text{ starting from } u \text{ escapes to } v) \sim e^{-2[F(z^*)-F(v)]/\epsilon}, \quad \text{as } \epsilon \to 0$$

- and, uniformly in $w_\delta \in B_{\delta/2}(u)$, for arbitrary small $\rho > 0$, the first hitting time of the ball $B_\delta(v)$ is controlled by the relation

$$\lim_{\delta\to 0}\lim_{\epsilon\to 0} \mathbb{P}(e^{2[F(z^*)-F(u)-\rho]/\epsilon} \le \tau_{B_\delta(v)} \le e^{2[F(z^*)-F(u)+\rho]/\epsilon} | X_0^\epsilon = w_\delta) = 1. \tag{1.11}$$

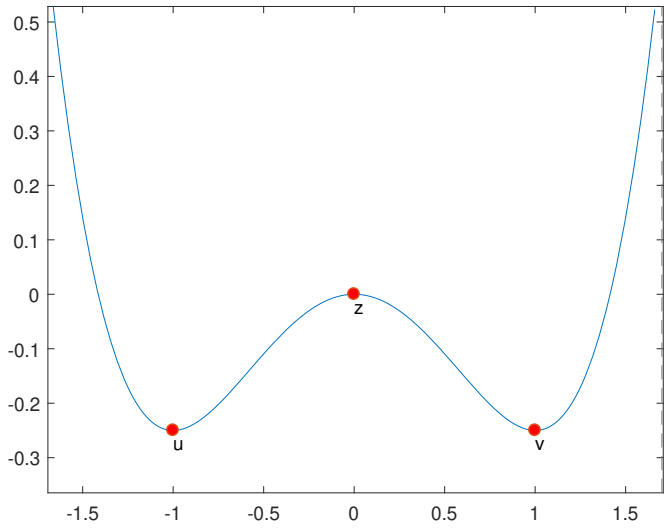

Figure 1.1: Double well potential $F(x) = \frac{1}{4}x^4 - \frac{1}{2}x^2$

.

1.2 The Potential Theory

In contrast to the Theory of Large Deviations for metastability, which considers the most likely continuous path between metastable states, the Potential-Theoretic approach models the metastable behaviour as a sequence of visits of the process to different sets. A first issue that emerges is how to identify the so called *metastable sets* that act as vertices of a network where each transition is a jump from one vertex to another. A second topic entails the estimate of the mean hitting times of those sets and the transition probabilities.

1.2.1 From Markov processes to the Dirichlet problem

In this work we limit the discussion to a special class of stochastic processes. We make a significant simplification and we consider Markov processes which admit a stochastic generator $\mathcal{L}$. However, one should be aware of a number of regularity conditions required, for which we refer to [1]. Continuous Markov processes satisfying such properties are called Feller processes. The associated Markov transition functions $(P_t)_{t\in\mathbb{R}}$ form a semi-group of bounded linear operator with generator $\mathcal{L}$, such that $P_t = e^{t\mathcal{L}}$ for each $t \in \mathbb{R}$, where e^X is the exponential map defined by the ordinary series expansion.

The formal relationship between a Markov process and the associated infinitesimal generator is given by the so called martingale problem. We recall here the definition of martingale.

Definition 2 *Let $(\Omega, \mathcal{G}, \mathbb{P}, (\mathcal{G}_t)_{t\in\mathbb{R}^+})$ a filtered probability space with continuous-time filtration $(\mathcal{G}_t)_{t\in\mathbb{R}^+}$. A stochastic process X on a filtered space is called martingale if: i) X is adapted to $\mathcal{G}_t$, namely X_t is $\mathcal{G}_t$-measurable for every $t \in \mathbb{R}^+$*

ii) $\mathbb{E}[|X_t|] < \infty$ *for all* $t \in \mathbb{R}^+$
iii) $\mathbb{E}[X_t|\mathcal{G}_s] = X_s$ *a.s. for all* $0 \leq s \leq t$.

The martingale problem is stated through the following

Definition 3 *[1] Given a regular domain* $S \subset \mathbb{R}^d$*, a linear operator* $\mathcal{L}$ *with domain* $D(\mathcal{L})$ *and range*$(\mathcal{L}) \subset C_0(S)$*, a* S*-value adapted process* X *on the filtered space* $(\Omega, \mathcal{G}, \mathbb{P}, (F_t)_{t\in\mathbb{R}^+})$ *is called solution of the martingale problem associated with the operator* $\mathcal{L}$ *if, for any* $f \in D(\mathcal{L})$*,*

$$M_t = f(X_t) - \int_0^t \mathcal{L}f(X_s)ds \tag{1.12}$$

is an $(\mathcal{F}_t)_{t\in\mathbb{R}^+}$*-martingale.*

The existence and uniqueness of solution X are proved under certain assumptions, providing a useful characterization of Markov processes. We skip technicalities of the involved proof of probability theory that can be found in [1] and [11].

As we have already mentioned in the introduction, the basic idea behind the Potential Theory is to express probabilistic quantities of Markov processes as solutions of Dirichlet problems for suitable boundary conditions. Let $X = (X_t)_{t\in\mathbb{R}^+}$ be a continuous-time Markov process on state space S with generator $\mathcal{L}$. Consider an open subset $D \subset S$ and the continuous functions $g, k : D \longrightarrow \mathbb{R}$ and $\bar{g} : D^C \longrightarrow \mathbb{R}$.

Definition 4 *The Dirichlet problem for* $f : D \longrightarrow \mathbb{R}$ *with source* k *and boundary conditions* $\bar{g}$ *is*

$$(-\mathcal{L}f)(x) + k(x)f(x) = g(x) \qquad x \in D \tag{1.13}$$
$$f(x) = \bar{g}(x) \qquad x \in D^C. \tag{1.14}$$

Hence, we set up the Dirichlet problem associated with a martingale problem. As a consequence, the solution f can be directly computed from the PDE but it also has a 'probabilistic representation' which involves the stochastic process X. Indeed, f does also depend on the domain and boundary condition in 1.14. This is a key observation. In fact, provided one sets up a suitable problem with the correct boundary conditions and solves it numerically, one could infer some information about the process using consequences of the following theorem. We refer to [1] for the proof in the discrete-time case.

Theorem 1.3 *Let* $\mathcal{L}$ *be the generator of a continuous-time Markov process* X *so that* X *is the solution of the martingale problem associated with* $\mathcal{L}$*. Assume* $f \in D(\mathcal{L})$ *solves the Dirichlet problem in 1.13 and 1.14. Let* $\tau_{D^C} = \{\inf t > 0 : X_t \in D^C\}$*. If*

$$\mathbb{E}_x\left[\tau_{D^c} \exp\left(\inf_{x\in D} k(x)\tau_{D^c}\right)\right] < \infty, \quad x \in D \tag{1.15}$$

then

$$f(x) = \mathbb{E}_x\left[\bar{g}(X_{\tau_{D^c}}) \exp\left(-\int_0^{\tau_{D^c}} k(X_s)ds\right) + \right.$$
$$\left. + \int_0^{\tau_{D^c}} g(X_t) \exp\left(-\int_0^t k(X_s)ds\right) dt, \quad x \in D. \right. \tag{1.16}$$

Whenever at the boundary $\bar{g} \equiv 0$ and $k(x) \equiv \lambda$ is such that the problem has a unique solution, then it can be expressed in terms of the Dirichlet Green's function $G^{\lambda}_{D^C}(x, dz)$, i.e. for any $g \in C_0(D)$

$$f(x) = \int_D G^{\lambda}_{D^C}(x, dz) g(z) \tag{1.17}$$

where

$$G^{\lambda}_{D^C}(x, dz) = \mathbb{E}_x \left[\int_0^{\tau_{D^c}} e^{-\lambda t} \mathbb{1}_{X_t \in dz} dt \right] \quad x, z \in D. \tag{1.18}$$

We define two central objects in the Potential Theory: the *equilibrium potential* and the *equilibrium measure*.

Definition 5 *Let $A, B \subset D$. The equilibrium potential $h_{A,B}$ is defined as the solution of the Dirichlet problem*

$$\begin{array}{ll} (-\mathcal{L}h)(x) = 0 & x \in (A \cup B)^C \\ h(x) = 1 & x \in A \\ h(x) = 0 & x \in B \end{array} . \tag{1.19}$$

The representation of the solution in equation 1.16 gives a natural probabilistic interpretation of $h_{A,B}$ in terms of hitting probabilities of the process. In this problem, $g = 0$ and $\bar{g}(x) \neq 0$ only on $x \in A$, hence, the Representation Theorem implies

$$h_{A,B}(x) = \mathbb{E}_x \left[\mathbb{I}_A(X_{\tau_{A \cup B}}) \right] = \mathbb{P}_x(\tau_A < \tau_B), \qquad x \in (A \cup B)^C. \tag{1.20}$$

From a different prospective, this result gives an analytical expression of the probability on the right-hand side of 1.20. If $x \in (A \cup B)^C$, we find that

$$(\mathcal{L}h_{A,B})(x) = \mathcal{P}_x(\tau_A < \tau_B), \quad x \in B \tag{1.21}$$

$$(-\mathcal{L}h_{A,B})(x) = \mathcal{P}_x(\tau_B < \tau_A), \quad x \in A. \tag{1.22}$$

The equilibrium measure is the second important quantity in our study.

Definition 6 *We define for $x \in D$ the equilibrium measure*

$$e_{A,B}(x) = (-\mathcal{L}h_{A,B})(x) \tag{1.23}$$

We observe that $e_{A,B}(x) = 0$ for all $x \notin A$ and it represents the probability that the process escapes from A to B.
In conclusion, we state a connection between the objects we just introduced. Suppose we know the equilibrium measure on A, then the equilibrium potential solves the nonhomogeneous Dirichlet problem

$$\begin{array}{ll} (-\mathcal{L}h)(x) = e_{A,B}(x) & x \in S \setminus B \\ h(x) = 0 & x \in B \end{array} \tag{1.24}$$

And since solution of 1.24 can be written in terms of Green's function, according to 1.17, the following holds:

Theorem 1.4 *With the notation introduced above and S, A and B as before,*

$$h_{A,B}(x) = \int_A G^0_B(x, y) e_{A,B}(dy) \quad x \in S. \tag{1.25}$$

1.2.2 The Dirichlet principle

The success of the potential approach has been proved in the case of reversible Markov processes, while its non-reversible counterpart shows some limitations and is still wide open for future research. Indeed, this is one of the drawbacks of this approach compared to the Pathwise approach which, in contrast, does not rely on reversibility assumptions. Aware of the lack of generality, we limit the theory in the remainder of the chapter to the reversible continuous-time setting.

First, we recall some basic facts about reversibility. Consider a Markov continuous-time process X on the state space S with Markov semigroup $(P_t)_{t\in\mathbb{R}^+}$ and a probability measure μ on S. We now assume μ is invariant with respect to the semigroup $(P_t)_{t\in}$. This allows to define the adjoint operators $(P_t^*)_{t\in\mathbb{R}^+}$ and the adjoint infinitesimal generator $\mathcal{L}^*$, via the relation

$$\int_S \mu(dx)(\mathcal{L}^* g)(x) f(x) = \int_S \mu(dx)(\mathcal{L} f)(x) g(x) \qquad \forall f, g \in D(\mathcal{L}). \tag{1.26}$$

A Markov process is called *reversible* if the infinitesimal generator is self-adjoint with respect to some invariant measure μ.

Let μ be an invariant measure for X, one can define two more central objects of the theory.

Definition 7 *The generator $\mathcal{L}$ of a reversible process X defines a non-negative quadratic form*

$$\mathcal{E}(f, g) = \int_S \mu(dx) g(x)(-\mathcal{L} f)(x) \qquad \forall f, g \in D(\mathcal{L}) \tag{1.27}$$

called the Dirichlet form.

Definition 8 *The capacity of the capacitor (A, B) is defined as*

$$\mathrm{cap}(A, B) = \int_A \mu(x) e_{A,B}(dx) \tag{1.28}$$

where $e_{A,B}(dx)$ is the equilibrium measure defined in 1.23.

The definitions of equilibrium measure and equilibrium potential imply an alternative representation of capacities in term of Dirichlet functions. If $A, B \subset S$ are two disjoint non-empty sets, $\mathrm{cap}(A, B)$ can be expressed as

$$\mathrm{cap}(A, B) = \mathcal{E}(h_{A,B}, h_{A,B}). \tag{1.29}$$

Remarkably, capacities play a crucial role in the computation of the mean hitting times of sets, as will be evident in the next section about reversible diffusion processes. However, it is often very challenging to find them as they imply to solve the Dirichlet problems above, which are often high-dimensional. The Potential-Theoretic approach exploits important variational principles that produce lower and upper bounds for capacities without the explicit solution $h_{A,B}$. The most relevant is the *Dirichlet Principle*, proved by Bovier and den Hollander in [1].

Theorem 1.5 *(Dirichlet Principle) Let A, B and S as in the Dirichlet problem in 1.19, and $\mathcal{H}_{A,B}$ the space of continuous functions f on $\bar{D}$ such that:*
i) $\mathcal{E}(f,f) < \infty$ and
ii) $f \geq 1$ on A and $f \leq 0$ on B.
If 1.19 admits a unique solution $h_{A,B}$, the equilibrium potential, then

$$\mathrm{cap}(A,B) = \inf_{f \in \mathcal{H}_{A,B}} \mathcal{E}(f,f) \tag{1.30}$$

Moreover, if $\mathcal{H}_{A,B} \neq \emptyset$, the infimum is attained by the equilibrium potential.

Trivially, this principle provides us with an elementary upper bound for any $f \in \mathcal{H}_{A,B}$. Compared to the Large Deviation approach, exploiting variational principles leads to sharper estimates on capacities and, therefore, on hitting probabilities. The lower bounds are less obvious and are investigated in [6].

We stress the fact that the computational approach we want to develop in this work is equation-free. Hence, the variational principles and the Dirichlet form are not useful tools in this case. In fact, we recall Definition 7 which involves the unknown generator $\mathcal{L}$ and the explicit formula for the invariant measure.

1.2.3 Example: Diffusions with noise

We present the main result of Potential Theory on a specific class of stochastic processes, namely elliptic diffusion processes in $\mathbb{R}^d$. Let consider an open regular domain $D \subset \mathbb{R}^d$. The SDE that defines the process is

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \tag{1.31}$$

where B is the d-dimensional Brownian motion, and b, σ measurable and bounded vector functions of the domain. In particular, b is a time independent drift vector and σ a time independent diffusion matrix. The infinitesimal generator restricted to $C^2(D)$ is

$$(\mathcal{L}f)(x) = \sum_{i=1}^{d} b_i(x)\frac{\partial f(x)}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d} \sigma^T(x)\sigma(x)\frac{\partial^2 f(x)}{\partial x_i \partial x_i} \tag{1.32}$$

We consider the simplified case with $\sigma^T(x)\sigma(x) = I_d$. Assume there exists a probability measure μ reversible with respect to $\mathcal{L}$. For reversible diffusion, invariant measures of the form $\mu(dx) = e^{-F(x)}dx$ are often considered so that the drift is the gradient of the potential up to a factor $-1/2$, i.e. $b(x) = -\frac{1}{2}\nabla F(x)$. Our goal is to find an expression of the mean hitting time in terms of some of the aforementioned quantities.

Let $A, B \subset D$ two disjoint and non-empty sets such that $\partial D = A \cup B$. The solution of the Dirichlet problem in 1.19 is the equilibrium potential $h_{A,B}(x)$. It represents the probability of transition to the set A. According to Theorem 1.5, and applying the first Green's identity, the quantity

$$\mathrm{cap}(A,B) = \mathcal{E}(h_{A,B}, h_{A,B}) = \int_D \mu(x)(-\mathcal{L}h_{A,B})(dx) =$$

$$= \int_D e^{-F(x)} h_{A,B}(x)(-\mathcal{L}h_{A,B}(x))dx = \int_{\partial A} e^{-F(x)}\partial_{n(x)} h_{A,B}(x) d\sigma_A(x)$$

Hence, the equilibrium measure $e_{A,B}$ in this case reads

$$e_{A,B}(dx) = \partial_{n(x)} h_{A,B}(x) d\sigma_A(x)$$

To conclude, we state the theorem for reversible continuous-time diffusion proved in [1], which gives the link between the expectation of the first hitting time τ_B and capacity. The first statement is the analogue of Theorem 1.4 for reversible diffusion.

Theorem 1.6 *For D and A, B as before,*

$$h_{A,B}(x) = \int_A \frac{\mu(y)}{\mu(x)} G_B(y,x) e_{A,B}(dy), \quad x \in A \cup B \cup D. \tag{1.33}$$

In particular, if f is the solution of the Dirichlet problem

$$(-Lf)(x) = g(x) \qquad x \in A \cup D \tag{1.34}$$

$$f(x) = 0 \qquad x \in B \tag{1.35}$$

introducing the probability measure $\nu_{A,B}$ on A

$$\nu_{A,B}(dy) = \frac{e^{-F(y)} e_{A,B}(dy)}{\mathrm{cap}(A,B)}, \tag{1.36}$$

we have

$$\int_A \nu_{A,B}(dy) f(x) = \frac{1}{\mathrm{cap}(A,B)} \int_{D \cup A} e^{-F(x)} h_{A,B}(x) g(x) dx. \tag{1.37}$$

Picking $g \equiv 1$ and writing the probabilistic representation of the solution f given in Theorem 1.3, we observe that $f(x) = \mathbb{E}_x[\tau_B]$. Finally, we get the link between mean hitting time in B and capacity:

Corollary 1.6.1 *For A, B, D as before*

$$\int_A \nu_{A,B}(dy) \mathbb{E}_y[\tau_B] = \frac{1}{\mathrm{cap}(A,B)} \int_{D \cup A} e^{-F(x)} h_{A,B}(x) dx. \tag{1.38}$$

To sum up, we tried to compare the approaches in order to derive the average transition time to a metastable set in stochastic diffusion processes. As an example, we pick F being a double well potential and choose sets A and B containing small balls centered at the locations of the minima of F. We end up studying the transitions between attractors located at the minima of F.
In one dimension, this problem was solved by Kramer in 1940 [10], who rigorously derived the classical Eyring–Kramers formula.
On the one hand, the Pathwise approach computes the exponential asymptotic term of the formula, without relying on reversibility assumptions, even though it does not provide the prefactor computed by Kramer.
On the other hand, Potential Theory provides a method based on PDEs which are hard to solve in dimension higher than one. However, the Potential-Theoretic approach reduces the problem to that of the computation of Newtonian capacities, which are estimated using variational principles and monotonicity properties. We tried to give a general idea of this approach and we refer to [12] for a complete description of the method.

Chapter 2

Data-based methods to approximate the generator

Before numerically solving the Dirichlet problem arising in the Potential Theory, it is important to understand the data-based methods in the literature that describe the evolution of a dynamical system. An emerging method is the Dynamic Mode Decomposition (DMD), a data-based technique which approximates the best linear least square fit dynamical system for the data measurements [13]. In their seminal work, Mezic [14] and Schmid [15] establish a connection between DMD and the Koopman theory, and approximate from data the Koopman operator of an unknown dynamical system. Our method has its roots in one of the multiple extensions of the DMD, the Extended Dynamic Mode Decomposition (EDMD), suitable for data coming from stochastic systems.

In this chapter, we show why the Koopman operator is a powerful tool for the analysis of a dynamical system. Then, we summarize the aforementioned data-driven procedures explained in [13] and [8] to build an estimate of it. We identify the link between the Stochastic Koopman Operator and the generator $\mathcal{L}$ associated with the SDE, as it appears in the Dirichlet Problem of Potental Theory. In the remainder of the chapter, we detail a method used to estimate $\mathcal{L}$ and the solution to the PDE. Remarkably, it differs from other algorithms in the literature as it does not rely on any piece of information on the governing equations of the system, representing a step forward to an equation-free estimation of transitions.

2.1 Dynamic Mode Decomposition and the Koopman operator

When we want to study a phenomenon of interest by recording the different states of the system, we first have to make the choice of which quantity we want to observe at each state. For instance one can describe a two-dimensional fluid flow by recording velocity and pressure fields, or stream function and vorticity. Let us consider a time interval Δt and let $\mathbf{x}(n) := \mathbf{x}(t_n) = \mathbf{x}(n\Delta t)$ be the state of the system at discrete time n. We consider a dynamical system defined by the map

$$\mathbf{x}(n+1) = N(\mathbf{x}(n)) \tag{2.1}$$

where N is a non-linear function of the state. The function N prescribes the dynamics and it might be yet undiscovered, as we might not know the process that produced the data but only have the emerging data from the system. One could also consider to look at the state by measuring some non-linear function of the state $\mathbf{z}(n) = \mathbf{T}(\mathbf{x}(n))$ and rewrite the system in terms of the new variable $\mathbf{z}$,

$$\mathbf{T}(\mathbf{x}(n+1) = \mathbf{T}(N(\mathbf{x}(n)). \tag{2.2}$$

Both observables (i.e., $\mathbf{x}$ or $\mathbf{z}$) describe the same fundamental phenomenon, yet one may be preferable to others.

As we mentioned before, the state of the system can be described by measuring different *observables*, that are functions defined on the state space. The evolution of the observables is governed by the Koopman operator, which describes how measurements of a dynamical system evolve through the nonlinear dynamics. Interestingly, the advantage of lifting the dynamics from the space of the states to the space of observables is that the Koopman Operator is a linear operator. However, because these measurements are functions and form a functions space, the Koopman operator is infinite dimensional.

The DMD, if combined with the Koopman theory, provides a finite-dimensional approximation to eigenvalues, eigenvectors and modes (vectors that enable us to reconstruct the state of the system as a linear combination of the Koopman eigenfunctions) of the infinite-dimensional Koopman operator. The infinite nature of Koopman operator is of course problematic, but if it could be truncated without great loss of accuracy, then we would obtain a linear and finite-dimensional approximation of a nonlinear system, without directly linearizing around a particular fixed point.

2.1.1 Koopman theory

We start from the broad prospective of discrete-time systems and we give the definition of the Koopman Operator, following the notation used in [8]. Consider a dynamics governed by an autonomous, discrete-time dynamical system $(\mathcal{M}, n, F)$, where $\mathcal{M} \subset \mathbb{R}^N$ is the state space, $n \in Z$ is (discrete) time and $F : \mathcal{M} \to \mathcal{M}$ is the evolution operator such that

$$x_{n+1} = F(x_n), \qquad \forall x_n \in \mathcal{M}.$$

Let us denote with $\mathcal{H}$ the space of observables, namely a Hilbert space containing all the scalar-valued measurement functions $g : \mathcal{M} \to \mathbb{C}$ of the state $x \in \mathcal{M}$.

Definition 9 *The Koopman operator $\mathcal{K}$ is an infinite-dimensional linear operator that acts on $\mathcal{H}$. The action of the Koopman operator on a measurement function g equates to composition of the function g with the flow map F, which advances the state of a Δt interval:*

$$\mathcal{K}(g) = g \circ F \Longrightarrow \mathcal{K}(g)(x_n) = (g \circ F)(x_n) = g(F(x_n)) = g(x_{n+1}). \tag{2.3}$$

In other words, the Koopman operator, also known as the *composition operator*, advances measurements along with the flow F [13]. We stress the fact that $\mathcal{K}$ maps observables to observables. More rigorously, it defines a new dynamical system, $(\mathcal{H}, n, \mathcal{K})$, that governs the evolution of observables $g \in \mathcal{H}$. These are two different parametrization of the same fundamental behaviour, linked by the full-state observable $\bar{g}(x) = x$.

Starting from an arbitrary state $x_i \in \mathcal{M}$ at time i, we can obtain the next state at time $i+1$ by applying the definition of the flow map F or, alternatively, considering the action of $\mathcal{K}$ on the full-state observable and apply that to x_i, namely:

$$\mathcal{K}(\bar{g})(x_i) = (\bar{g} \circ F)(x_i) = \bar{g}(F(x_i)) = F(x_i) = x_{i+1}.$$

The Koopman operator may also be defined for continuous-time dynamical system of ODEs, such as

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, t). \tag{2.4}$$

For each time interval Δt, we denote $\Phi_{\Delta t} : \mathcal{M} \longrightarrow \mathcal{M}$ the flow map which maps an initial point x to the solution passing from x after a time Δt. There exists a semigroup of Koopman operators $\{\mathcal{K}_{\Delta t}\}$ associated with the flow maps $\{\Phi_{\Delta t}\}$, for which usually there is no formula. The action of the semigroup on the observables is defined by

$$\mathcal{K}_{\Delta t} g(\mathbf{x}) = g(\Phi_{\Delta t}(\mathbf{x})). \tag{2.5}$$

Definition 10 *The continuous-time Koopman operator [8] is the infinitesimal generator $\mathcal{L}$ of the semigroup $\{\mathcal{K}_{\Delta t}\}_{\Delta t \in \mathbb{R}^+}$ and, similarly to the discrete case, acts on the space of the observables $\mathcal{H}$ so that, if $g \in \mathcal{H}$,*

$$\mathcal{L} g = \mathbf{f} \cdot \nabla g. \tag{2.6}$$

Formally, this relationship comes from the limit

$$\mathcal{L} g = \lim_{\Delta t \longrightarrow 0} \frac{\mathcal{K}_{\Delta t} g - I(g)}{\Delta t} = \lim_{\Delta t \longrightarrow 0} \frac{g \circ \Phi_{\Delta t} - g}{\Delta t}.$$

For each time interval τ and under continuity assumptions, the following formal relation between the Koopman operators and the generator of the semigroup holds:

$$\mathcal{K}_\tau = \exp(\tau \mathcal{L}). \tag{2.7}$$

Spectral decomposition of the continuous Koopman operator is fundamental in representing the evolution of the observables. A detailed study of spectral analysis on function spaces is presented in [13]. Here, we assume a finite spectrum of the generator $\mathcal{K}$. However, it is possible to extend the results to dynamical systems with continuous spectra.
Let us denote by $\phi_k(\mathbf{X})$ the k-th eigenfunction and λ_k the k-th eigenvalue of $\mathcal{K}$.

Assume $\mathbf{g} = [g_1 g_2 ... g_n]$ a set of observables in the span of the first N eigenfunctions with coefficients $\{\mathbf{v}_k \in \mathbb{C}^n\}_{k=1}^N$, called Koopman modes. Then it is possible to express the evolution of $\mathbf{g}$ using an eigenfunction expansion so that

$$\mathbf{g}(\mathbf{x}_{t+\tau}) = \mathcal{K}_\tau(\mathbf{g})(\mathbf{x}_t) = \mathcal{K}_\tau(\sum_{k=1}^{N} \mathbf{v}_k \phi_k(\mathbf{x}_t)) = \sum_{k=1}^{N} \mathbf{v}_k \mathcal{K}_\tau(\phi_k(\mathbf{x}_t)) = \sum_{k=1}^{N} e^{\lambda_k \tau} \mathbf{v}_k \phi_k(\mathbf{x}_t)$$

The equation above shows that, under some assumptions, the future observebles are represented by means of Koopman eigenfunctions, eigenvectors and modes (the coefficients $\mathbf{v}_k$). This motivates the application of numerical data-based methods such as the DMD to this problem, methods that are proven to be extremely effective when we do not have access to the function $\mathbf{f}$.

2.1.2 DMD to approximate Koopman operator

Let us consider a generic system of ODEs as in 2.4 and suppose we do not know the governing functions on the right hand side. Similarly to the discrete time map in 2.2, the DMD method constructs an approximate linear dynamical system such that $\tilde{\mathbf{x}}(t)$ is a certain approximation of the original state variable $\mathbf{x}(t)$ such that

$$\frac{d\tilde{\mathbf{x}}(t)}{dt} = \mathcal{A}\tilde{\mathbf{x}}(t), \tag{2.8}$$

where $\mathcal{A}$ is a matrix. The advantage of the linear approximation is of course that we have explicit solutions in terms of eigenvectors and eigenvalues of the matrix $\mathcal{A}$, (ϕ_k, λ_k) for $k = 1, 2, ..., N$. In particular, for given initial conditions,

$$\tilde{\mathbf{x}}(t) = \sum_{k=1}^{N} b_k \phi_k e^{\lambda_k t} \tag{2.9}$$

and we seek to minimize the error $||\tilde{\mathbf{x}} - \mathbf{x}||$ for all times. One way of doing that using data emerging from the system is by applying the Dynamic Mode Decomposition, a data-driven algorithm that finds the least square fit linear dynamical system.

Given the continuous time system in 2.8 it is always possible to define the discrete flow map $\tilde{\Phi}_{\Delta t}$ corresponding to the evolution of 2.8 for a fixed time Δt and describe the discrete-time system as:

$$\tilde{\mathbf{x}}_{k+1} = \tilde{\Phi}_{\Delta t}(\tilde{\mathbf{x}}_k) \approx \mathbf{A}\tilde{\mathbf{x}}_k \tag{2.10}$$

where

$$\mathbf{A} = \exp(\mathcal{A}\Delta t). \tag{2.11}$$

However, we usually do not have an explicit formula for the flow map $\Phi_{\Delta t}$ of the original (nonlinear) system in 2.4 for $\mathbf{x}$.

The DMD method requires collecting m pairs of snapshots of the system state $(\mathbf{x}_k, \mathbf{y}_k)_{k=1}^m$ such that $\mathbf{y}_k = \Phi_{\Delta t}(\mathbf{x}_k)$ and order them into two data matrices $\mathbf{X}$ and $\mathbf{Y}$ such that each column contains one state measurement, namely

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & ... & \mathbf{x}_m \end{pmatrix} \tag{2.12}$$

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \mathbf{y}_3 & ... & \mathbf{y}_m \end{pmatrix}. \tag{2.13}$$

It computes the least-square fit linear operator $\mathbf{A}$ relating the data $\mathbf{Y} \approx \mathbf{AX}$, namely

$$\mathbf{A} = \mathbf{Y}\mathbf{X}^{\dagger} \tag{2.14}$$

where $\mathbf{X}^{\dagger}$ denote the Moore-Penrose pseudoinverse.

The matrix A may be very large in dimensions and not directly tractable. A detailed description of the DMD algorithm and the model reduction techniques used for the eigendecomposition of the matrix A can be found in [13].

We are now ready to link the DMD and the Koopman theory. We recall that the Koopman Operator is a linear infinite-dimensional operator acting on measurement functions g of the state space. It advances the measurements of a fixed time Δt so that

$$\mathcal{K}_{\Delta t} g(\mathbf{x}_t) = g(\Phi_{\Delta t}(\mathbf{x}_k)) = g(\mathbf{x}_{t+\Delta t}). \tag{2.15}$$

We consider a collection of n nonlinear observables ψ_i, $i = 1, 2, 3..., n$ and build a vector-value function $\mathbf{\Psi} : \mathcal{M} \longrightarrow \mathbb{C}^n$ such that

$$\mathbf{\Psi}(\mathbf{x}) = \begin{pmatrix} \psi_1(\mathbf{x}) & \psi_2(\mathbf{x}) & \psi_3(\mathbf{x}) & ... & \psi_n(\mathbf{x}) \end{pmatrix}^T. \tag{2.16}$$

In order to build the two matrices required by the method we apply the observables to the data matrices defined in 2.12. In formula, we define

$$\mathbf{\Psi}(\mathbf{X}) := \begin{pmatrix} \mathbf{\Psi}(\mathbf{x}_1) & \mathbf{\Psi}(\mathbf{x}_2) & \mathbf{\Psi}(\mathbf{x}_3) & ... & \mathbf{\Psi}(\mathbf{x}_m) \end{pmatrix} \tag{2.17}$$

$$\mathbf{\Psi}(\mathbf{Y}) := \begin{pmatrix} \mathbf{\Psi}(\mathbf{y}_1) & \mathbf{\Psi}(\mathbf{y}_2) & \mathbf{\Psi}(\mathbf{y}_3) & ... & \mathbf{\Psi}(\mathbf{y}_m) \end{pmatrix} \tag{2.18}$$

and apply DMD on the space of observable. The algorithm finds the best-fit for K such that $\mathbf{\Psi}(\mathbf{Y}) \approx K\mathbf{\Psi}(\mathbf{X})$, namely

$$K = \mathbf{\Psi}(\mathbf{Y})\mathbf{\Psi}(\mathbf{X})^{\dagger}. \tag{2.19}$$

K is a finite dimensional approximation of the infinite linear operator $\mathcal{K}_{\Delta t}$ in 2.15. It is worth noting that the dimension and the quality of the approximation strongly depend on the choice of the observables and the quality of data. A complete discussion on the set of observables follows in the next section about the Extended Dynamic Mode Decomposition.

In conclusion, one can observe that the DMD method is just the execution of the Koopman idea where $\mathbf{\Psi}$ is chosen to be the full state observable $\mathbf{\Psi}(\mathbf{x}) = \mathbf{x}$, (i.e. the identity operator). As previously stated, the DMD finds the best approximation in a least-square sense of the operator A, projecting the whole non-linear dynamics into a linear dynamical system. The Koopman theory does the same but on the space of observables. We take some collections of non linear observable and the model tells us how to project the observables into the future. By choosing the correct observables the system is completely determined by a linear dynamics.

2.2 Extended Dynamic Mode Decomposition and the Stochastic Koopman Operator

This section gives an overview of the Extended Dynamic Mode Decomposition (EDMD), an entirely data-driven method presented by Williams et al. in [8] which

efficiently produces numerical finite approximations of Koopman Operator and its eigenfunction for an unknown nonlinar dynamical system.

The main advantange of the Extended Dynamic Mode Decomposition is that, while the approaches described so far assume the data provided to the algorithm coming from a deterministic system, the EDMD method consider the case when we assume a stochastic process underlying the dynamics and provides an estimate of the so called Stochastic Koopman Operator leading tuples.

This method is particularly relevant for the purpose of this work because it allows to extend the previous discussion to dynamical systems with stochastic noise, such as the stochastic diffusion problem which is addressed and solved in the following chapter.

2.2.1 EDMD definition

The EDMD method requires

- a set of N_K observables $\mathcal{D} = \{\psi_1, \psi_2, \psi_3 ... \psi_{N_K}\} \subset \mathcal{H}$, $\psi_i : \mathcal{M} \longrightarrow \mathbb{C}$, called *dictionary functions*. For brevity, we collect them in a vector-value function $\mathbf{\Psi} : \mathcal{M} \longrightarrow \mathbb{C}^{N_K}$,

$$\mathbf{\Psi}(x) := \begin{pmatrix} \psi_1(x) & \psi_2(x) & \psi_3(x) & ... & \psi_{N_k}(x) \end{pmatrix} \tag{2.20}$$

- and a dataset organized in m pairs of snapshots of the system $\{(x_i, y_i)\}_{i=1}^m$ such that $y_i = \Phi_{\Delta t}(x_i)$.

It projects the dynamics of the observables onto the subspace $\mathcal{H}_\mathcal{D}$ generated by the dictionary $\mathcal{D}$ and provides a finite dimensional approximation of $\mathcal{K}$ and eigenfunctions with respect to the basis.

As the subspace $\mathcal{H}_\mathcal{D}$ is generally not invariant under the action of $\mathcal{K}$, the algorithm computes $\mathbf{K} \in \mathcal{C}^{N_k} \times \mathcal{C}^{N_k}$ by minimizing a residual term that emerges when applying $\mathcal{K}$ to an arbitrary observable $\psi \in \mathcal{H}_\mathcal{D}$.

Definition 11 *[8] Given the dataset $\{(x_i, y_i)\}_{i=1}^m$ and N_K dictionary functions $\mathcal{D} = \{\psi_1, \psi_2, \psi_3 ... \psi_{N_K}\}$, define $\mathbf{\Psi}$ as in 2.20. The EDMD finite-dimensional approximation of $\mathcal{K}$ is*

$$\mathbf{K} = \mathbf{G}^\dagger \mathbf{A}, \tag{2.21}$$

where

$$\mathbf{G} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{\Psi}(x_i)^* \mathbf{\Psi}(x_i), \tag{2.22}$$

$$\mathbf{A} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{\Psi}(x_i)^* \mathbf{\Psi}(y_i), \tag{2.23}$$

and * indicates the conjugate transpose vector.

Let $\mathbf{v}_j$ be the j-th eigenvector and μ_j the j-th eigenvalue of K, then the EDMD approximation of the j-th eigenfunction and j-th eigenvector of the continuous-time $\mathcal{K}$ are, respectively,

$$\phi_j(x) = \mathbf{\Psi}(x)\mathbf{v}_j, \qquad \lambda_j = \ln(\mu_j)/\Delta t. \tag{2.24}$$

We observe that $\mathbf{K}$, $\mathbf{A}$ and $\mathbf{G}$ are $N_K \times N_K$-matrices. The rank of K depends on the data and on the definition of basis functions, but with sufficient (and generic) data and the commonly used dictionaries discussed below, it is full rank. Moreover, $\mathbf{K}$ represents a projection into the subspace $\mathcal{H}_\mathcal{D}$. Thus, its dimensions corresponds to the size of the dictionary $\mathcal{D}$ and also how well it approximates $\mathcal{K}$ indeed depends on how rich is $\mathcal{D}$ and the generated subspace of $\mathcal{H}$. In particular, we would like $\mathcal{H}_\mathcal{D}$ to include the dominant Koopman eigenfunctions or at least an accurate approximation of those.

2.2.2 Relationship with other methods

We now explain how the Extended Dynamic Mode Decomposition relates to the Galerkin finite element method.
As is argued in [8], the approximation given by (2.21) converges, in the data limit, to the approximation of the generator that would come from the numerical solution of a PDE by the finite element method. Suppose the m pairs of data are drawn from a probability distribution ρ on the domain. It is observed that the elements of the matrices $\mathbf{A}$ and $\mathbf{G}$ consist of the sample means of the following quantities:

$$\psi_i(x)\psi_i(x), \tag{2.25}$$
$$\psi_i(x)\mathcal{K}_t\psi_i(x). \tag{2.26}$$

The Galerkin approximation of $\mathcal{K}_t$, according to [8], would be $\mathbf{K} = \mathbf{G}^{-1}\mathbf{A}$ with

$$\hat{G}_{i,j} = \langle \psi_i(x), \psi_j(x) \rangle_\rho \tag{2.27}$$
$$\hat{A}_{i,j} = \langle \psi_i(x), \mathcal{K}_t\psi_j(x) \rangle_\rho \tag{2.28}$$

where $\langle a(x), b(x) \rangle_\rho = \int_\mathcal{M} a(s)b(s)d\rho(s)$.

Thus, if the size of data $m \to \infty$, the central limit theorem applies and the sample means $G_{i,j}$ and $A_{i,j}$ converge to the expected value of 2.25 and 2.26 with respect the measure ρ, which are exactly the terms $\hat{G}_{i,j}$ and $\hat{A}_{i,j}$.

2.2.3 The stochastic data

As mentioned at the beginning of this chapter, one of the most interesting characteristics of the EDMD method is that it is completely data-driven, meaning that the algorithm does not require the knowledge of the underlying equations. However, some assumptions on the model are necessary to interpret the approximation given in 2.21. In the DMD approach, we looked at (discrete and continuous) deterministic systems defined by a map or a system of ODEs, where the map and the vector field (the RHS of the equation) are unknown.

Now, we extend the DMD approach to the class of stochastic dynamical systems. In particular, we are interested in the stochastic processes that exhibit metastability and evolve according to a Stochastic Differential Equation of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t \tag{2.29}$$

where W_t is the Wiener process. We now assume that the data are collected by observing a stochastic process X_t.

Firstly, we need to define the Koopman Operator in the stochastic setting. Among the large number of definitions in the literature, we choose as a reference the theoretical framework for random dynamical system developed in [16] and we apply it to SDE.

Let $(\Omega_s, \mathcal{F}, P)$ be a probability space associated with a stochastic dynamics and $\mathbb{T}$ a semigroup representing time. A random dynamical system is represented by the mappings $\phi(t, \omega, x) : \mathbb{T} \times \Omega_s \times \mathcal{M} \longrightarrow \mathcal{M}$, such that, for example, if the time is discrete ($\mathbb{T} = \mathbb{Z}$), $\phi(1, \omega, \cdot) : \mathcal{M} \longrightarrow \mathcal{M}$ is the one step map of the discrete random process. The Koopman Operator takes an observable and returns its expected value over the probability space Ω_s after a time t.

Definition 12 *[16] The stochastic Koopman operator $\mathcal{K}_t$ associated with the random dynamical system ϕ is defined on the space of observables $g : \mathcal{M} \longrightarrow \mathbb{C}$ for which the functional*

$$\mathcal{K}_t g(x) = \mathbb{E}[g(\phi(t, \omega, x))] \qquad \forall x \in \mathcal{M} \tag{2.30}$$

exists. We refer to the family of operators $\{\mathcal{K}_t\}_{t\in\mathbb{T}}$ as the stochastic Koopman operator family.

Under some assumptions, the family of operators satisfies the property of a semigroup. We limit to the (continuous) Markovian processes, for which strong semigroup properties always hold [16]. In this case, the generator of the Koopman semigroup acting on the observables space is defined as

$$\mathcal{L}g(x) = \lim_{t\to 0} \frac{\mathcal{K}_t g(x) - g(x)}{t} \tag{2.31}$$

and is called *continuous-time stochastic Koopman Operator*.

We conclude that, if we apply the EDMD method to stochastic data and choose a good set of dictionary function, it recovers the leading eigenfunctions and eigenvalues of the (continuous) Stochastic Koopman Operator.

We now make a crucial observation that allows to connect the theory presented in Chapter 1 and the numerical methods in this chapter. Going back to equation 2.29, the solution $X_t(x)$ with initial condition $X_{t_0}(\omega) = x$ is formally defined using the Itô stochastic integral

$$X_t(x) = x + \int_0^t b(X_s)ds + \int_0^t \sigma(X_s)dWs. \tag{2.32}$$

It is shown in different books that the stochastic solutions X_t form a semigroup. The generator of the semigroup is called the Kolmogorov operator and appears in the

Kolmogorov backward equation. We highlight that in Section 1.2 we denote it with $\mathcal{L}$. It determines how the expected values of the observables evolve in time. Therefore, under reasonable assumptions about b and σ it coincides with the continuous Stochastic Koopman Operator [8].

For the stochastic Itô diffusion in 2.29, the action of the continuous time stochastic Koopman Operator on the observable g is given by

$$\mathcal{L}g(x) = \sum_i b_i(x)\frac{\partial g}{\partial x_i}(x) + \tfrac{1}{2}\sum_{i,j}\left(\sigma(x)\sigma(x)^\top\right)_{i,j}\frac{\partial^2 g}{\partial x_i\,\partial x_j}(x). \tag{2.33}$$

This observation is central in our work as one can now make use the data-driven approach to Koopman theory in order to find an estimate of the generator $\mathcal{L}$ of stochastic processes underlying metastable dynamical systems.

2.3 Approximation of the generator

With a background on data-based methods for dynamical systems, one can now go back to theory in Chapter 1 and ask if it is possible to solve the Dirichlet Problem in definition 4 without a complete knowledge of the generator $\mathcal{L}$. Although there are several standard methods in the literature that solve elliptic PDEs by discretizing the differential linear operator, this approximation is usually built starting from the model equations. In fact, techniques such as finite difference approximations generally require the explicit formula for $\mathcal{L}$. The method described in this section provides a solution to an unknown PDE by means of a finite-dimensional approximation of $\mathcal{L}$ generated only using time series data from the unknown process. It combines different numerical methods such as the Galerkin finite element method, Monte Carlo method and finite difference approximations.

Further important topics to consider when solving the PDE are the choice of the dictionary functions and how to incorporate the boundary conditions imposed by the problem. Those are discussed in Chapter 3.

2.3.1 Galerkin approximation

As a first step we introduce a finite element basis, project the PDE onto this, and find the combination of basis functions that is the best approximate solution of the problem. This procedure is known as Galerkin approximation and is introduced here for a one-dimensional PDE boundary value problem.

Consider the one-dimensional problem on a domain $\Omega = \{x \in [\alpha, \beta] \subset \mathbf{R}\}$, for some $a, b \in \mathbb{R}$ and $f : \Omega \longrightarrow \mathbb{R}$ a continuous function, of the form

$$\mathcal{L}u = f \tag{2.34}$$
$$u(\alpha) = a \tag{2.35}$$
$$u(\beta) = b, \tag{2.36}$$

where $\mathcal{L}$ is the infinitesimal generator of the diffusion process

$$dx = b(x)dt + \sigma(x)dW, \tag{2.37}$$

for some terms $b(x)$ and $\sigma(x)$, namely,

$$\mathcal{L} = b(x)\frac{\partial}{\partial x} + \frac{\sigma(x)^2}{2}\frac{\partial}{\partial x^2}. \tag{2.38}$$

Applying the finite element method to find the solution u to this problem implies the choice of some set of basis functions.

Let us define the set $\{\psi_i\}_{i=1}^n$, that spans a function space V_ψ. We approximate the solution u of the differential equation as linear combination of basis functions, namely the best approximation to u from the space V_ψ. A function of the space $\tilde{u} \in V_\psi$ may be written as combination of the elements ψ_i so that

$$\tilde{u} = \sum_{i=1}^{n} u_i \psi_i(x). \tag{2.39}$$

Different techniques exist to determine the coefficients u_i, $i \in \{1, 2, ..., n\}$. The Galerkin method is a residual method that requires the following condition, for each element of the basis, given the measure ρ on the domain Ω

$$\int_\Omega (\mathcal{L}u(x) - f(x))\psi_j(x)d\rho = 0, \quad \forall j \in \{1, 2, ..., N\}. \tag{2.40}$$

Assume that f is also an element of V_ψ, then one can rewrite it as a linear combination and let f_i be the coefficients such that the projection of f into V_ψ is $f = \sum_{i=1}^n f_i\psi_i(x)$. Substituting the expansions in 2.40 gives

$$\begin{aligned}
&\int_\alpha^\beta (\mathcal{L}\sum_{i=1}^{n} u_i\psi_i(x) - \sum_{i=1}^{n} f_i\psi_i(x))\psi_j(x)d\rho = \\
&\sum_{i=1}^{n} \int_\alpha^\beta (u_i\mathcal{L}\psi_i(x) - f_i\psi_i(x))\psi_j(x)d\rho = \\
&\sum_{i=1}^{n} u_i \int_\alpha^\beta (\mathcal{L}\psi_i(x))\psi_j(x)d\rho - \sum_{i=1}^{n} f_i \int_\alpha^\beta \psi_i(x)\psi_j(x)d\rho = 0
\end{aligned}$$

By defining two matrices $A, G \in \mathbb{R}^{n\times n}$

$$G_{i,j} = \int_\Omega \psi_i(x)\psi_j(x)d\rho \tag{2.41}$$

$$A_{i,j} = \int_\Omega \mathcal{L}\psi_i(x)\psi_j(x)d\rho, \tag{2.42}$$

the residual condition in matrix formula reads

$$\sum_{i=1}^{n} A_{i,j}u_i - \sum_{i=1}^{n} G_{i,j}f_i = 0 \quad \forall j \in \{1, 2, ..., N\} \implies A^T\mathbf{u} = G^T\mathbf{f}. \tag{2.43}$$

If the problem is homogeneous, namely $f \equiv 0$, then (2.43) reduces to

$$A^T\mathbf{u} = 0. \tag{2.44}$$

The finite-dimensional Galerkin approximation of $\mathcal{L}$ is $L = G^\dagger A^T$ or, alternatively, depending on the definition of the right hand side $\mathbf{f}$, it can also be $L = A^T$. Thus, the vector of coordinates $\mathbf{u}$ is determined by solving a finite dimensional system of linear equations [17].

2.3.2 Approximation of the PDE solution from data

Importantly, we want to adopt a data-driven approach that is completely independent of the definition of the SDE defined in equation 2.37 and only relies on time series data from the unknown underlying process. However, it is observed that computing A requires to know the action of $\mathcal{L}$ on basis functions (equation 2.42). The definition of A needs to be changed and an approximation of $\mathcal{L}\psi_i$ is required. We accomplish that by looking at the Stochastic Taylor expansion of the functions ψ_i along the solution of the SDE.

For the stochastic differential equation in (2.37), and the autonomous case, i.e. $b(x)$ and $\sigma(x)$ do not depend on time explicitly, the Ito lemma leads to

$$d\psi_i(x(t)) = \left[b(x)\frac{\partial\psi_i(x(t))}{\partial x} + \frac{\sigma^2(x(t))}{2}\frac{\partial^2\psi_i(x(t))}{\partial x^2}\right]dt + \sigma(x(t))\frac{\partial\psi_i(x(t))}{\partial x}dW(t). \quad (2.45)$$

Let us consider a time step Δt small and the stochastic process $x(t)$ defined in 2.37. With the notation introduced in Section 3.1.3 for random dynamical systems in the probability space $(\Omega_S, \mathcal{F}, P)$, we denote $\phi(\Delta t, \omega, \cdot) : \mathcal{M} \longrightarrow \mathcal{M}$ the map that advances the process x by a time step Δt. We have that $x(t + \Delta t) = \phi_{\Delta t}(\omega, x(t))$ and, by definition (see [16]), $\phi_0(\omega, x) = Id$, for any $x \in \mathcal{M}$ and $\omega \in \Omega_S$.

For any Δt small, 2.45 implies

$$\psi_i(\phi(\Delta t, \omega, x)) = \psi_i(x) + \int_0^{\Delta t} \mathcal{L}\psi_i(\phi(s, \omega, x))ds + \int_0^{\Delta t} \sigma(\phi(s, \omega, x))\frac{\partial\psi_i(\phi(s, \omega, x))}{\partial x}dW(s). \quad (2.46)$$

By using the elementary rectangular rule for numerical integration, we have

$$\psi_i(\phi(\Delta t, \omega, x)) = \psi_i(x) + \Delta t\mathcal{L}\psi_i(x) + \Delta W_t\sigma(x)\frac{\partial\psi_i(x)}{\partial x} + \mathcal{O}(\Delta t^2) \implies \quad (2.47)$$

$$\psi_i(\phi(\Delta t, \omega, x)) - \psi_i(x) = \Delta t\mathcal{L}\psi_i(x) + \Delta W_t\sigma(x)\frac{\partial\psi_i(x)}{\partial x} + \mathcal{O}(\Delta t^2). \quad (2.48)$$

The stochastic contribution to this term is given by the stochastic dynamics and it is associated to the space $(\Omega_S, \mathcal{F}, W)$ and the Wiener measure. One can observe that the increments

$$\Delta W_t = W(t + \Delta t) - W(t) \sim N(0, \Delta t), \quad (2.49)$$

so their expectation is zero. If we consider the expected value with respect to the Wiener measure we get the following approximation of the action of the continuous generator upon the dictionary functions

$$\mathbb{E}\left[\psi_i(\phi(\Delta t, \omega, x)) - \psi_i(x)\right] = \mathbb{E}\left[\Delta t\mathcal{L}\psi_i(x)\right] + \mathbb{E}\left[\Delta W_t\sigma(x)\frac{\partial\psi_i(x)}{\partial x}\right] + \mathcal{O}(\Delta t^2) \quad (2.50)$$

$$\implies \mathcal{L}\psi_i(x) = \mathbb{E}\left[\frac{\psi_i(\phi(\Delta t, \omega, x)) - \psi_i(x)}{\Delta t}\right] + \mathcal{O}(\Delta t^2) \quad (2.51)$$

Therefore, we replace the definition of A with

$$A'_{i,j} = \int_\Omega \mathbb{E}\left[\frac{\psi_i(\phi(\Delta t, \omega, x)) - \psi_i(x)}{\Delta t}\right] \psi_j(x) d\rho(x) \tag{2.52}$$

which implies solving the double integrals

$$\int_{\Omega\times\Omega_S} \left(\frac{\psi_i(\phi(\Delta t, \omega, x)) - \psi_i(x)}{\Delta t}\right) \psi_j(x) dW(\omega) d\rho(x), \qquad \forall i, j = 1, 2, ..., N. \tag{2.53}$$

The integrals are solved by numerical approximation using Monte Carlo integration, with the samples taken from the distribution ρ and their evolution according to the map $\phi(\Delta, \omega, \cdot)$. The same technique is applied in the EDMD method in definition 11. The main advantage here is that this method works independently from the dimension of the problem, although other techniques can also be used to obtain better convergence. An alternative to Monte Carlo integration is integrating by quadrature rules, which implies choosing a set of fixed initial points and weights, and then evaluating the function at the mesh points. Instead, Monte Carlo integration requires choosing a probability distribution ρ from which initial conditions are sampled. In such a way, we are allowed to sample data from a stationary time series as drawn from the equilibrium distribution. Hence, it is sufficient to pick initial data and its evolution after an interval Δt.

Let us consider M pairs of samples $\{(x_0^m, x_1^m)\}_{m=1}^M$, where $x_0^m \sim \rho$ is the m-th initial sample, drawn from the distribution ρ, and x_1^m is its realization after a time step Δt. The definition of A used in the algorithm is

$$\hat{A}'_{ij} = \frac{1}{M}\sum_{m=1}^{M} \left(\frac{\psi_i(x_1^m) - \psi_i(x_0^m)}{\Delta t}\right) \psi_j(x_0^m). \tag{2.54}$$

If the data are in the form of a time series $\{x_n\}_{n=1}^M$, then 2.54 takes the form

$$\hat{A_{ij}}' = \frac{1}{M}\sum_{n=1}^{M} \left(\frac{\psi_i(x_{n+1}) - \psi_i(x_n)}{\Delta t}\right) \psi_j(x_n). \tag{2.55}$$

Notice that the choice of the sampling distribution plays a role in the convergence of the method. A uniform distribution on Ω is considered when we have pairs of snapshots $(x_i, y_i)_{i=1}^N$ of the system, at time t_0 and $t_0 + \Delta t$. While in the more realistic case of a time series data, we should consider the equilibrium distribution of the process. In practice, it is not necessary to make explicit use of the measure to compute the integrals.

2.3.3 Convergence of the method to a Galerkin approximation

We prove that the method converges to a Galerkin approximation method in the large-data limit. Using a similar argument to the one used for the EDMD in [8], we see that, by the law of large numbers,

$$\lim_{M\to\infty} \hat{G}_{i,j} = \mathbb{E}[\psi_i(x)\psi_j(x)] = \int_\Omega \psi_i(s)\psi_j(s)\rho(s)ds = G_{i,j}$$

It also holds for $\hat{A}'_{i,j}$ defined in 2.54 that

$$\begin{aligned}\lim_{M\to\infty} \hat{A}'_{i,j} &= \mathbb{E}[\left(\frac{\psi_i(x^m_{n+1}) - \psi_i(x^m_n)}{\Delta t}\right)\psi_j(x^m_n)] = \\ &= \int_{\Omega\times\Omega_s}\left(\frac{\psi_i(\phi(\Delta t, s, \omega)) - \psi_i(s)}{\Delta t}\right)\psi_j(s)\rho(s)dsdW(\omega) = \\ &= \int_{\Omega}\mathbb{E}\left[\frac{\psi_i(\phi(\Delta t, s, \omega)) - \psi_i(s)}{\Delta t}\right]\psi_j(s)\rho(s)ds = \\ &= \int_{\Omega}(\mathcal{L}\psi_i(s) + \mathcal{O}(\Delta t^2))\psi_j(s)\rho(s)ds = A_{i,j} + \mathcal{O}(\Delta t^2),\end{aligned}$$

where ω is an element in Ω_s, the probability space associated with the stochastic dynamics and W is the probability measure.

2.3.4 Error estimate

In conclusion of this section, we add a discussion on the error and the rate of convergence of this method. A complete proof of the convergence rate of the method is not included in this work, however, we summarize the aforementioned approximations and explain under which limits they are consistent. We also try to give an estimate of order of the errors up to an unknown proportionality coefficient.

There are three contributions to take into account when analysing the performance of this method:

- the first approach that was introduced consists of projecting the PDE onto the subspace of some finite element basis (Galerkin method). It is known from the literature that the Galerkin method produces the best approximation to the true solution of the given problem, in the subspace spanned by test functions (dictionary functions). A proof of this fact uses the projection theorem from linear algebra and is discussed in [17]. However, proving the convergence to the true solution as the finite element mesh is refined requires understanding how well a given function can be approximated as a linear combination of the elements of the basis. Indeed, the accuracy grows with the number of test functions N, since the approximating subspace becomes richer, but the order of convergence still depends on the particular set of functions. With the notation used in Section 2.3.1 and $\tilde{u}$ as in 2.39, one should try to understand what is the order of the error in

  $$u = \tilde{u} + err(N)$$

  for the actual choice of basis functions. A further discussion on the order of this approximation is postponed to the first section of Chapter 3.

- A further approximation concerns the term $\mathcal{L}\psi_i(x^m)$ in the computation of $\hat{A}_{i,j}$, which defines the finite dimensional approximation of $\mathcal{L}$. While equation 2.46 is an exact equality given by Itô's formula, an approximation is made

in eq. 2.47. Here, by using the rectangular rule, we assume the integrand function constant between time t and $t + \Delta t$. It is straightforward to show that the error is proportional to Δt^2, namely $err = \mathcal{O}(\Delta t^2)$. In other words, by reducing the time step Δt to $\Delta t/2$, one should also see the error decreasing of $1/4$ factor.

- Finally, the matrices A and G are not computed exactly. Instead Monte Carlo integration is applied to compute the integrals. The error, which depends on the number of samples M, decreases as $1/\sqrt{M}$ as M grows.

  By combining this together, we can see that

$$A_{i,j} = \int_\Omega \mathcal{L}\psi_i(x)\psi_j(x)d\rho = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{\psi_i(x_1^m) - \psi_i(x_0^m)}{\Delta t}\right)\psi_j(x_0^m) + \mathcal{O}(\Delta t^2) + \mathcal{O}\left(\frac{1}{\sqrt{M}}\right) \quad (2.56)$$

Chapter 3

Solution of the Dirichlet problem of Potential Theory

Chapter 1 illustrated the main results of Potential Theory and, in particular, it was highlighted how some of the key quantities that have a probabilistic interpretation concerning transitions are expressed as solutions of Boundary Value Problems. As an example, in the general case of an SDE with uniformly elliptic generator $\mathcal{L}$, the equilibrium potential $h_{A,B}$ on an open domain D, between two disjoint neighbourhoods of two metastable equilibria A, B, such that $A \cup B = \partial D$, represents the probability that a process starting in x first touches the set A, namely

$$h_{A,B}(x) = \mathbb{P}_x(\tau_A < \tau_B) \qquad x \in D. \tag{3.1}$$

We recall that $h_{A,B}$ is defined in [6] as the solution to the homogeneous Dirichlet problem associated to the stochastic PDE, where the value of the unknown function at the boundary is specified by the following Dirichlet conditions:

$$\begin{array}{ll} (-\mathcal{L}h)(x) = 0 & x \in D \setminus (A \cup B) \\ h(x) = 1 & x \in A \\ h(x) = 0 & x \in B, \end{array} \tag{3.2}$$

The inhomogeneous Dirichlet problem

$$\begin{array}{ll} (-\mathcal{L}f)(x) = g(x) & x \in D \\ f(x) = 0 & x \in \partial D, \end{array} \tag{3.3}$$

plays also a role in our study. As argued in Section 1.2, under some assumptions the solution f has a probabilistic representation in terms of expected exit time from a the set D.

The result presented in this chapter is the numerical solution of the Dirichlet Problem in 3.2 for a stochastic diffusion process with a double well potential. This is achieved without knowing the generator $\mathcal{L}$ and using time series data. We apply the equation-free method described in Section 2.3.2 to project the PDE onto a certain subspace spanned by basis functions. To do that, we first need to study how one can impose the Dirichlet boundary conditions (BCs) to the projected solution. In the remainder of the chapter, we present two examples of PDEs with unknown generator and Dirichlet boundary conditions which are solved with our data-driven approach. Finally, we show the plot of the solutions with a brief discussion on the error.

3.1 The choice of the basis functions and error estimate

Choosing a set of basis functions, as for the EDMD method, is a necessary step to apply the algorithm. How do we do that when applying the EDMD to estimate Koopman eigenfunctions? In the EDMD, the choice of the elements in $\mathcal{D}$ is important as they span the subspace $\mathcal{H}_\mathcal{D} \subset \mathcal{H}$ of the space of scalar observables where we seek the approximation of the eigenfunctions. The optimal choice for $\mathcal{D}$ is still an open problem, and indeed it depends on the underlying process and the range and distribution of data. However, as explained by William et al. in [8], standard choices with different advantages are usually Hermite polynomials, radial basis functions or high-degree piecewise polynomials. Hermite polynomials are the easiest to implement and suitable for data distributed according to a normal distribution; the piecewise polynomials have the advantage of producing a matrix G which is block-diagonal but require a mesh on the domain for the definition. Radial basis functions are an optimal choice for more complex and high-dimensional domains as the functions can be easily extended to higher dimension without specifying a mesh.

In this work we build the architecture of the method and apply it to one dimensional problems. As explained in Chapter 2, our method follows the approach of a Galerkin approximation which is a finite element approach. The definitions of matrix $\hat{A}'$ in 2.54 allows to rewrite the Galerkin condition in terms of the basis $\{\psi_i\}_{i=1}^N$ and, consequently, to find the approximation of the solution $\tilde{u}$ with respect to the same basis. The solution projected onto the space V_ψ spanned by $\{\psi_i\}_i$ is given by equation 2.39. It is important to observe that the error we make depends on the subspace V_ψ and hence on our choice of basis functions. A general definition of the finite elements and further discussions on different basis and the error can be found in [18].

For the one-dimensional problems, we use piecewise linear basis functions, locally defined as follows: the domain $\Omega = [\alpha, \beta]$ is divided into N equal segments of length $\Delta x = \frac{\beta-\alpha}{N}$ by $N+1$ equally spaced points. These points are called *nodes* and are defined as $x_i = \alpha + i\Delta x$, for $i = 0, 1, 2, ..., N$. We define a set of *basis functions* $\Psi = \{\psi_i(x) : i = 0, 1, ..., N+1\}$ by

$$\psi_i(x) = \begin{cases} \frac{x-x_{i-1}}{\Delta x} & x_{i-1} < x < x_i, \\ \frac{x_{i+1}-x}{\Delta x} & x_i < x < x_{i+1} \\ 0 & otherwise \end{cases} . \tag{3.4}$$

for all $i \in \{1, 2, ..., N-1\}$, as well as

$$\psi_0(x) = \begin{cases} \frac{x-x_0}{\Delta x} & x_0 < x < x_1, \\ 0 & otherwise \end{cases} \qquad \psi_N(x) = \begin{cases} \frac{x-x_{N-1}}{\Delta x} & x_{N-1} < x < x_N, \\ 0 & otherwise \end{cases} \tag{3.5}$$

The basis consists of $N-1$ tent functions centered in the inner nodes, plus two more functions centered at the extreme points of the domain. It is proved in [17] that $\{\psi_i\}_{i=1}^N$ spans the piece-wise linear polynomial space S and that, for constant

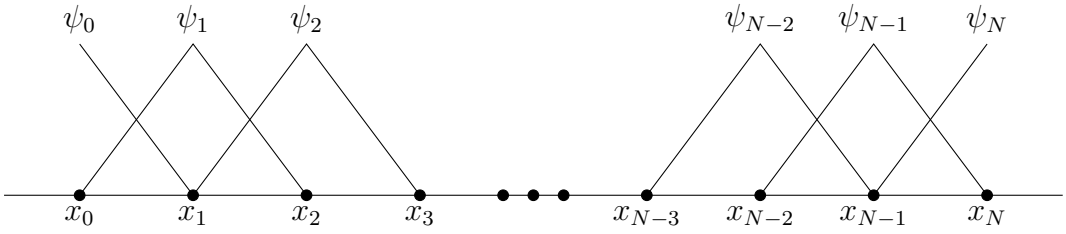

Figure 3.1: Basis functions $\{\psi_i\}_{i=0}^N$

uniform mesh, the error

$$||u - \tilde{u}_S|| \leq C\Delta x^2 ||u''|| \tag{3.6}$$

with C is a constant independent of Δx and u. Therefore, we conclude that it is a second order method in the parameter Δx. However, we recall there are other contributions to the algorithmic error due to the additional approximations.

There are many other options with various properties and higher accuracy in different contexts. For instance, under some assumptions, choosing piecewise polynomials of degree p generates an order $p + 1$ error. For n-dimensional spaces, we might choose radial basis functions, as they do not require a structured grid and are very effective even in high dimensions.

3.2 Implementing Dirichlet Boundary Conditions

The Galerkin method combined with data-based estimates discussed above provide a finite dimensional approximation of the generator $\mathcal{L}$ and the solution to the PDE $\mathcal{L}u = f$. However, nothing has been said about the boundary conditions on the function u. Many questions can arise at this point. Does the approximation given by the data-driven method already imply some conditions at the boundary of the domain? If yes, how can these be modified? How do the data set and the basis functions affect the value of u at α and β? Here, we try to explain how to impose the Dirichlet boundary conditions when we choose a finite element basis as defined in Section 3.1.

Different options may be applicable to the problem we aim to solve. Since the approximate solution is a projection onto the subspace spanned by the dictionary, one could think of choosing Ψ such that all the linear combinations of elements satisfy the BCs. A second approach relies on the fact that, in the finite element approach, each element of the basis is locally defined and associated to a subregion of the domain. For instance, considering the basis Ψ and an arbitrary value $x \in \Omega$, which falls into the segment (x_j, x_{j+1}) for some j, only two basis functions namely ψ_j and ψ_{j+1} determine the value $u(x)$. Thus, in order to impose the value of u at the boundary which satisfies the conditions, one can fix the coefficients u_0 and u_N in the solution in 2.39 equal to a suitable constant value and reduce the number of unknown coefficients by two.

The second way is implemented. Let us denote by L the $(N+1)\times(N+1)$ matrix

representing the Galerkin approximation of the generator $\mathcal{L}$, by

$$\mathbf{U} =: \begin{pmatrix} u_0 & u_1 & u_2 & ... & u_N \end{pmatrix}^T$$

the vector of unknown coefficients and by $\mathbf{f}$ the coefficients of the right hand side, which is given. We consider the Dirichlet boundary conditions $u(\alpha) = 0$ and $u(\beta) = 1$. The linear system that gives the coefficients reads

$$L\mathbf{U} = \mathbf{f}. \tag{3.7}$$

Firstly, $u_0 = 0$ is fixed and removed from the vector of unknowns. The system is reduced by removing the first column of L (the one multiplied by u_0) and the first row of L, $\mathbf{U}$ and $\mathbf{f}$. Hence, the reduced linear system is

$$\left[\begin{array}{c|c} \tilde{L} & \mathbf{l}_N \\ \hline \mathbf{l}^N & l_{N,N} \end{array}\right] \cdot \left[\begin{array}{c} \tilde{\mathbf{u}} \\ \hline u_N \end{array}\right] = \left[\begin{array}{c} \tilde{\mathbf{f}} \\ \hline f_N \end{array}\right]. \tag{3.8}$$

Then also $u_N = 1$ is fixed and the system reduced further by removing the last row of each matrix and vector. Finally, $\tilde{\mathbf{u}}$ is computed by solving the square linear system

$$\tilde{L}\tilde{\mathbf{u}} + \mathbf{l}_N = \mathbf{f} \implies \tilde{L}\tilde{\mathbf{u}} = \mathbf{f} - \mathbf{l}_N \tag{3.9}$$

and the coefficients vector is reconstructed so that $\mathbf{U} = \begin{pmatrix} 0 & \tilde{\mathbf{u}} & 1 \end{pmatrix}^T$.

It is worth noticing one more aspect concerning the data. Even though the domain of the PDE is fixed, the collection of data (regardless if we record a time series or pairs of measurements) may include points outside this domain. One could think of discarding those values or, alternatively, keep them in the data set and define all the basis functions to be zero outside the domain. This is the case for the chosen basis Ψ which allows to keep all the samples.

3.3 Example: the Laplace equation

In this section we include an example of a PDE that we solve using the method described in Section 2.3.2. While it does not give any information about transitions as the ones defined in the Potential Theory, it is still interesting to acknowledge the power of the algorithm which builds a data-driven approximation of the operator from the stochastic data of the corresponding SDE and numerically compute the solution.

Let $\mathcal{L} = \nabla^2$ denote the standard Laplace second-order operator and consider a non-homogeneous Laplace equation on the domain $\Omega = [0, 1]$, with boundary conditions

$$\begin{array}{ll} \mathcal{L}u(x) = x(x-1) & x \in (0,1) \\ u(0) = 0 & \\ u(1) = 0 & \end{array} . \tag{3.10}$$

The operator $\mathcal{L}$ trivially corresponds to the second order derivative in one-dimension. It is known from the theory on mathematical models in [6] that the operator $\frac{\sigma}{2}\mathcal{L}$ is

the infinitesimal generator of the pure diffusion stochastic process described by the equation

$$dX_t = \sigma dW_t, \tag{3.11}$$

where X_t is a continuous-time Markov process in $\mathbb{R}$ describing the position of a particle at time t and W_t denotes a standard one-dimensional Wiener process.

We show that by using data simulated from the process X_t and a correct scaling, our method builds an approximation of $\mathcal{L}$ which is in agreement with the discretization in space computed using the standard second order, central finite difference method. Figure 3.3 shows two approximate solutions to the Dirichlet Problem in 3.10 computed in MATLAB. One is achieved using our data-based approximation of $\mathcal{L}$ and the boundary conditions as described in the previous section. The other one is the solution computed with a finite difference method.

In order to build an approximation of the Laplace operator, we simulate data from the process above and build the snapshot pairs necessary to apply our method. This is done by sampling a set of initial points $(x^m)_{m=1}^M$ from a uniform distribution on Ω and generating their evolutions $(y^m)_{m=1}^M$ after an interval Δt. We use one step of the Euler-Maruyama method with time step $dt = 0.01$ to define $(y^m)_{m=1}^M$.
To summarize, data are collected into two matrices, namely

$M = 100000;$ *%size of data*
$X = rand(1, M);$ *%initial points*
$Y = X + sqrt(dt) * sigma * randn(1, M);$ *%euler-maruyama increment*

Thus, we use the data $(x^m, y^m)_{m=1}^M$ and basis functions ψ_i as in Section 3.1, and the algorithm compute L. We observe that the Laplace operator corresponds to L multiplied by the coefficient $2/\sigma^2$.

Finally, the conditions at 0 and 1 are applied through the coefficients u_0 and u_N following the example in the previous section. To avoid perturbations at the boundary, we decide to enforce the boundary conditions by considering a larger interval than the domain, so additional point outside $[0, 1]$, and impose BCs on the external mesh points too. In this way we enforce BC on neighbourhoods of 0 and 1. The coefficients $\mathbf{U}$ are retrieved solving the reduced linear system. We recall that $\mathbf{U}$ determines the projection of the solution onto that basis.

Given the simplicity of this example, where the generator is known a priori, it is possible to compare the solution with the one computed using a numerical method in the literature. We choose to implement the standard second order, central finite difference method, although other methods can be applied to this problem. A description of the implementation of the finite difference for this problem can be found in [19].

The example shows how to apply this method to numerically solve certain classes of Dirichlet problems. In particular, it is only possible when the differential equation involves a differential operator corresponding to the infinitesimal generator of some stochastic process. Although it does not seem to have advantages compared to other standard methods in this context, it would be interesting to investigate its application to more complex and higher dimensional boundary value problems.

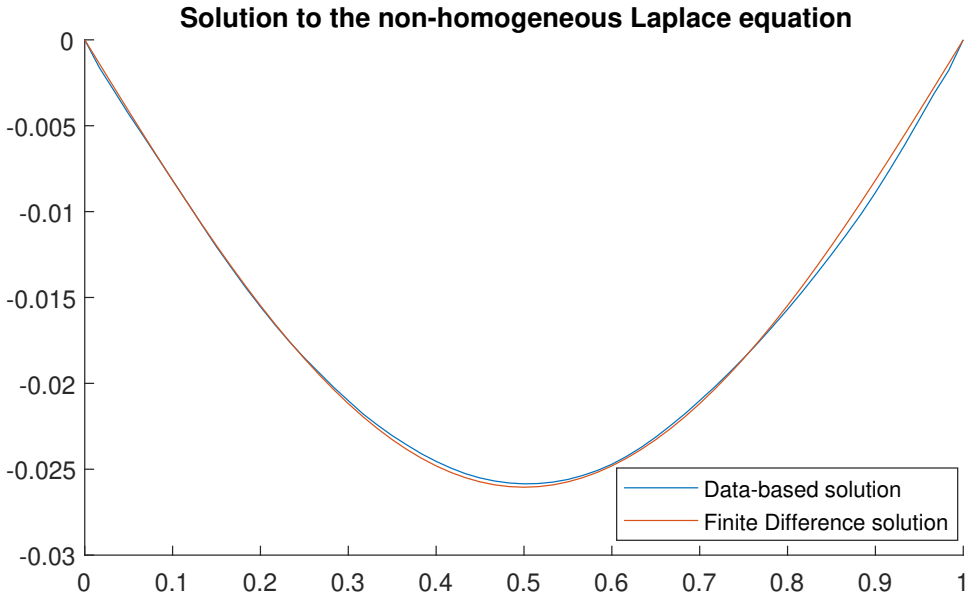

Figure 3.2: A set of $N = 60$ basis functions and $M = 10^6$ data are used. $\sigma = 0.1$ and $dt = 0.01$. The FDM also uses a discretization of the domain in sixty equally-spaced points. The data based solution computed by the algorithm is very close to the Finite Difference one and the two almost overlap.

3.3.1 Analysis of the convergence

Before studying how the error scales with the parameters of the algorithm, it is worth noting that a main difference between our approach and the Finite Difference one in the literature is that our PDE solution is dependent on data. Therefore, it is useful to study the variance of the solution when different time series from the same process are used (Figure 3.3). We perform $r = 10$ runs sampling new data set each time, and consider the mean solution and the mean error for each set of parameters N, Δt and M.

According to our estimate of error $\mathcal{O}(\frac{1}{\sqrt{M}})$, the error scales by a factor $1/10$ when M is increased by a factor 10^2.

3.4 Solution of the Dirichlet Problem for the diffusion process

In contrast to the previous section, one can start the discussion from a stochastic process and then derive the associated PDE, which can be solved using only time series data from the process. The model considered to analyse transitions is the well-known stochastic diffusion with drift generated by a double well potential. It describes the motion of a particle that tends to remain in one well of the potential, but after some time due to the random perturbation 'jumps' to the second well. Again, we look at the one-dimensional stochastic equation that reads

$$dX_t = -F'(X_t)dt + \sigma dW_t \tag{3.12}$$

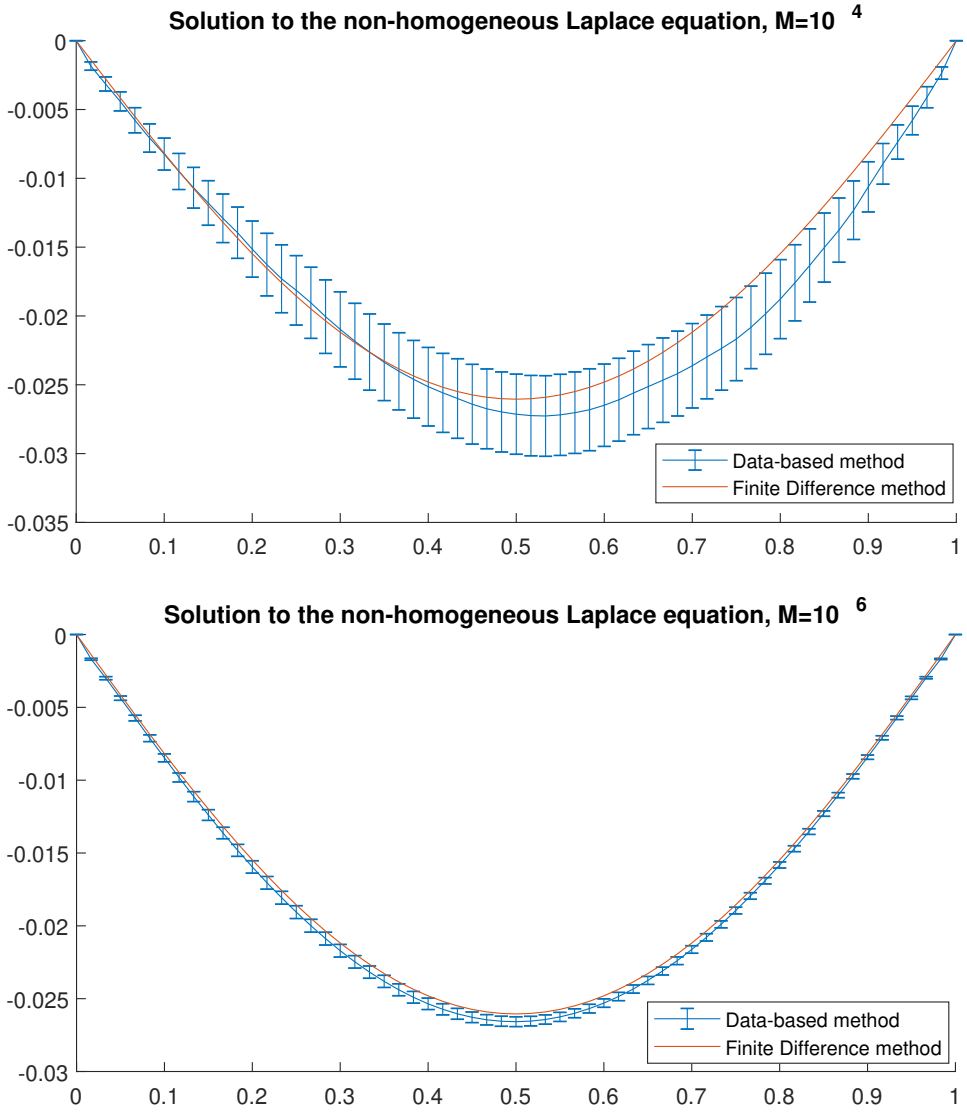

Figure 3.3: A set of $N = 60$ basis functions are used, $\sigma = 0.1$ and $dt = 0.01$. The average solution of $r = 10$ runs is plotted for data size $M_1 = 10^4$ and $M_2 = 10^6$. The error bars represent the standard error of the solutions at each one of the N grid points. The mean errors (computed using the Euclidean norm) are $e_1 = 0.063$ and $e_2 = 0.0070$. The accuracy increases and the solutions are closer to the mean, as the size of data grows.

where $F : \mathbb{R} \longrightarrow \mathbb{R}$ is a continuous and differentiable function with two local minima and $\sigma \in \mathbb{R}$ a coefficient which scales the noise. As an example, we consider $F(x) = \frac{1}{2}x^2 - \frac{1}{4}x^4$. We assume measuring a time series from a diffusion process of this form, showing metastable behaviour between two regions A and B. We solve numerically from data the Dirichlet Problem that defines the equilibrium potential $h_{A,B}$ for the underlying process, defined in the Potential Theory and in equation 3.2.

The two metastable equilibria are located in $x = -1$ and $x = 1$ such that the

associated Dirichlet Problem is

$$\begin{aligned} &(-\mathcal{L}h)(x) = 0 \qquad x \in (-1,1) \\ &h(1) = 1 \\ &h(-1) = 0 \end{aligned} \tag{3.13}$$

The solution $h(x)$ represents the probability that the process starting from a given x reaches the right well before the left one. In formula, let U_{-1} and U_1 denote two neighbourhoods of -1 and 1, then

$$h(x) = \mathbb{P}_x(\tau_{U_1} < \tau_{U_{-1}}). \tag{3.14}$$

In order to check the result, we collect the time series by simulating a specific diffusion process, choosing

- $\sigma = 0.5$ and
- $F(x) = \frac{1}{2}x^2 - \frac{1}{4}x^4$.

We generate the data set by applying M steps of the Euler-Maruyama method with a time interval $dt = 0.01$. The next observation is such that

$$X(t+dt) = X(t) + X(t)(1 - X(t)^2)dt + \sigma\sqrt{dt}\Delta W, \tag{3.15}$$

where $\Delta W = W_{t+1} - W_t$ is the increment of the Wiener processs. It is remarkable that using the data-based approach we manage to solve the problem only from the time series, and without using the information about $\mathcal{L}$ given above. Finally, we compare the result with the PDE solution computed with finite difference method. Figure 3.4 shows the two solutions seem to agree. A further analysis on the error of the method should be conducted to understand convergence with respect the parameter M, N and Δt.

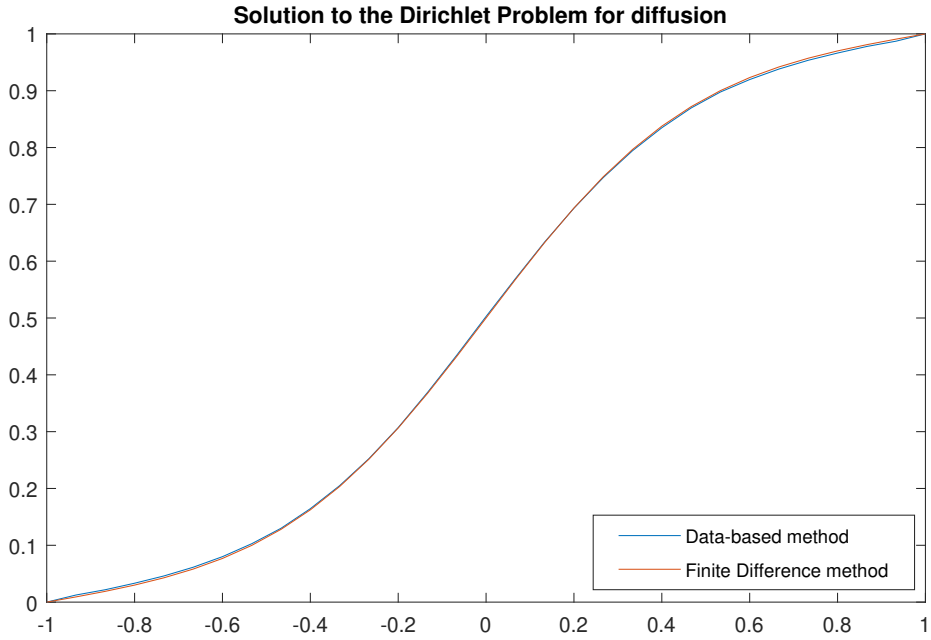

Figure 3.4: $\sigma = 0.5$. The basis consists of $N = 30$ tent functions centered in mesh points inside the domain $(-1, 1)$ plus four additional functions centered in four nodes outside the domain that are used to enforce the boundary conditions in the neighbourhoods of the boundary. The data are is $M = 10^6$ simulated points. The finite difference discretization uses $N = 30$ equally spaced points too.

Conclusion

In this thesis we created a method to study systems subject to random noise that show metastable behaviour, by combining recent data-driven algorithms and the Potential-Theoretic approach to metastability. Metastable random dynamical systems are characterised by the occurrence of rapid and random transitions between stationary states. Estimating fundamental quantities such as average hitting time of metastable sets, transition probabilities and rate of transitions are central challenges in metastability. Two main approaches were studied that are available in the literature. In Chapter 1, we investigated the Pathwise approach and the more recent Potential-Theoretic approach, and we applied them to stochastic differential equations. The first one is based on theory of Large Deviations and computes the probability of the most likely continuous path connecting two stable states. The second one, sets up Dirichlet boundary value problems for the quantities associated to metastable sets which involve the generator of the underlying stochastic process. Instead of solving the high dimensional PDE, it provides sharp bounds on the solutions which are proved to be effective for reversible processes.

However, we observed that both theories require the knowledge of the equations for the system. We overcame this limitation and looked for a method that can be applied when one has just observations from a time series of the system state and wants to investigate the system without knowing the underlying equations. To do that, we studied the Extended Dynamic Mode Decomposition, a method that approximates the Stochastic Koopman Operator of the process from snapshot pairs of the system. We used the same idea and combined different standard numerical methods to achieve the solution of a Dirichlet problem defined in the potential theory for metastability, using an approximation of the generator. As a result, a method which is completely data-driven and equation-free is proposed and implemented for one-dimensional Dirichlet problems.

In Chapter 3 we tested the algorithm on two simple one-dimensional problems. We generated data from a pure diffusion equation and from stochastic diffusion with a double well potential, and we used the data set to run the algorithm applying the boundary conditions imposed by the theory. We showed that, if we choose tent basis functions, the method reconstructs the solution of the problem, with an error that corresponds to a standard second order symmetric finite difference method. Due to the approximation of the integrals that define the linear system of equations, there are two more contributions to the error which are taken into account and analysed. However, due to the different levels at which the approximations apply, it was not straightforward to give a final estimate of the error.

Results in Chapter 3 are in agreement with the probabilistic interpretation of

potential theory and with the analytical results in the literature. They are also not new: in fact, the one-dimensional case is a well-known problem and was firstly solved by Kramers in the paper [10]. What is worth noting here is the approach that combines recent methods that perform well with only few data and do not have big computational costs, to the recent theory, which, in general, would be effective only for systems for which we already know the equations. We conclude that this work shows fertile perspectives and offers good foundations for a possible research whose direction is the generalisation to higher dimensional problems, where the method could be exploited to estimate transitions.

There are several aspects and challenging questions that could be addressed in further studies. The Dirichlet problem we solved imposing Dirichlet boundary conditions gives the equilibrium potential associated to the two minima of the potential. A similar problem is defined in [20] for the mean exit time, namely the average time that the process takes to escape from a potential well, but with different boundary conditions. The framework of the algorithm could be indeed applied to the problem, provide we implement the Neumann condition (the first derivative vanishes) on one extreme of the interval. Even in one dimension, this condition is not trivial to impose and one could try to implement stencils (expanding the domain), and use the method of matched asymptotic expansions as in singular perturbation problems to approximate a uniform solution.

Different choices of basis functions could be more suitable for some problems. Analogously, choosing a good dictionary provides an optimal convergence rate of the EDMD method. The latter is still an open problem in the literature. It is interesting to investigate which other basis are available for our method and, for example, if radial basis functions can be a suitable choice in order to extend the results to higher dimensional processes.

Moreover, another interesting question is how the method extends to more general potentials. One could investigate if the estimation of the SDE operator remain consistent and also how to identify the metastable states and corresponding boundary conditions when the potential has multiple local minima.

Bibliography

[1] A. Bovier and F. den Hollander, *Metastability: A Potential-Theoretic Approach*. Grundlehren der mathematischen Wissenschaften 351, Springer International Publishing, 1 ed., 2015.

[2] H. Grubmüller, "Conformational Dynamics of Proteins." `https://www.mpibpc.mpg.de/grubmueller/research/conformationalFlooding`.

[3] H. Katzov-Eckert, V. Botosso, and al., "Phylodynamics and dispersal of hrsv entails its permanence in the general population in between yearly outbreaks in children," *PloS one*, vol. 7, p. e41953, 10 2012.

[4] Wikipedia, "Compartmental models in epidemiology — Wikipedia, the free encyclopedia." `http://en.wikipedia.org/w/index.php?title=Compartmental%20models%20in%20epidemiology&oldid=1049812645`, 2021.

[5] A. D. W. a. M. I. Freidlin, *Random Perturbations of Dynamical Systems*. Grundlehren der mathematischen Wissenschaften 260, Springer-Verlag New York, 1 ed., 1984.

[6] A. Bovier, *Metastability: A Potential Theoretic Approach*, vol. 3 of *Proceedings of ICM 2006*. EMS Publising House, 2006.

[7] F. den Hollander, "Invited paper," *Stochastic Processes and their Applications*, vol. 114, no. 1, pp. 1–26, 2004.

[8] M. O. Williams, I. G. Kevrekidis, and C. W. Rowley, "A data–driven approximation of the koopman operator: Extending dynamic mode decomposition," *Journal of Nonlinear Science*, vol. 25, p. 1307–1346, Jun 2015.

[9] C. Schütte, W. Huisinga, and S. P. Meyn, "Metastability of diffusion processes," 2003.

[10] H. Kramers, "Brownian motion in a field of force and the diffusion model of chemical reactions," *Physica*, vol. 7, no. 4, pp. 284–304, 1940.

[11] T. G. K. Stewart N. Ethier, *Markov Processes: Characterization and Convergence*. Wiley Series in Probability and Statistics, John Wiley Sons, Inc., 2005.

[12] A. Bovier, M. Eckhoff, V. Gayrard, and M. Klein, "Metastability in reversible diffusion processes i. sharp asymptotics for capacities and exit times," *Journal of the European Mathematical Society*, vol. 6, pp. 399–424, 10 2004.

[13] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic Mode Decomposition*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2016.

[14] H. Arbabi and I. Mezić, "Ergodic theory, dynamic mode decomposition, and computation of spectral properties of the koopman operator," *SIAM Journal on Applied Dynamical Systems*, vol. 16, p. 2096–2126, Jan 2017.

[15] P. Schmid, "Dynamic mode decomposition of numerical and experimental data," *Journal of Fluid Mechanics*, vol. 656, pp. 5 – 28, 2010.

[16] - Nelida, M. Senka, and I. Mezić, "Koopman operator spectrum for random dynamical systems," 09 2019.

[17] M. S. Gockenbach, *Understanding And Implementing the Finite Element Method*. USA: Society for Industrial and Applied Mathematics, 2006.

[18] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*. Texts in Applied Mathematics 15, Springer New York, 3rd ed ed., 1994.

[19] P. Knabner and L. Angermann, *Numerical Methods for Elliptic and Parabolic Partial Differential Equations*. Texts in Applied Mathematics 44, Springer-Verlag New York, 1 ed., 2003.

[20] G. A. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*. Texts in Applied Mathematics 60, Springer-Verlag New York, 1 ed., 2014.