School of Mathematical, Physical and Computational Sciences

Operator Methods and Response in Climate Dynamics

by

Manuel Santos Gutiérrez

Thesis submitted for the degree of Doctor of Philosophy

Department of Mathematics and Statistics

September 2021

University of Reading

Abstract

Understanding how a physical system responds to external stimuli is fundamental in every area of science. To this end, the diverse theories of response in statistical physics come into play in order to predict the change in the average behaviour of a system undergoing perturbations. The general goal of this thesis is to apply the theory of response into the modelling and conceptual understanding of Earth's climate system.

Perturbation theory of operator semigroups shall be employed to derive response formulas in a variety of contexts. Under a stochastic framework, we shall study the effects of adding external forcing in terms of the unperturbed regime, following the spirit of linear response theory. Furthermore, the yielding response formulas are shown to decompose according to the spectral features of the generator of the transfer operator semigroup, allowing to simplify the expressions.

Finite dimensional representations of transfer operators lead to stochastic matrices whose properties give useful information about the system up to finite precision. Thus, it is possible to define a coarse-grained linear response, whose conditions for well-posedness and computability are investigated. This methodology is applied to an Ornstein-Uhlenbeck process and the Lorenz 63 atmospheric convection model, whose linear responses are calculated in agreement with observed simulations of the systems.

Reduced-order equations are derived using operator expansions. The latter provide non-Markovian closures that preserve the statistical properties of the model in question and are proved to posses the structure of multilevel stochastic models. Such structure is also present in the Empirical Model Reduction (EMR), which constructs non-Markovian models out of partially observed data. This analogy is illustrated in a conceptual climate model, suggesting a formal link between the response theoretic methodology and the EMR data-driven protocol.

Declaration: I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Manuel Santos Gutiérrez

Acknowledgements

It is my sincere wish to thank Valerio for his scientific and personal guidance towards elaborating the work presented herein; the way I see mathematics and physics today I largely owe to him.

I must thank my collaborators and mentors, former and current alike, with whom I have had stimulating conversations that have triggered my curiosity for new fields.

I am indebted to the MPE CDT, which has granted me the opportunity to fulfil a desire kindled in 2013 at the Alhóndiga Bilbao where, being a first-year undergraduate student, I first came across the Mathematics of Planet Earth initiative through the words of Christiane Rousseau.

I am grateful to my family and friends for their continuous support throughout this saw-toothed, yet immensely enriching journey.

Table of Contents

List of Tables

List of Figures

Notation

$\sigma(A)$	Spectrum of the operator/matrix A
$\mathfrak{F}[\cdot]$	Fourier transform
$\mathfrak{L}[\cdot]$	Laplace transform
$C_b(\mathcal{X})$	Continuous and bounded functions on $\mathcal{X}$
$[\mathbf{v}]_i$	ith entry of the vector $\mathbf{v}$
LHS	Left-hand side
RHS	Right-hand side
$\mathfrak{Re}$	Real part
$\mathfrak{Im}$	Imaginary part
L^p_μ	p-Integrable functions with respect to μ
$D(A)$	Domain of operator A
$A^\top$	Transpose of matrix A
A^*	Conjugate-transpose of matrix A
$\mathrm{diag}(\mathbf{v})$	Diagonal matrix with entries of vector $\mathbf{v}$
$A_{:,i}$	ith column-vector of matrix A
$\mathbf{e}_i$	Canonical basis vector with ith entry equal to 1
$\mathbf{1}_B$	Characteristic function on the set B
$\mathbb{1}_d$	d-Dimensional vector full of ones
$\mathbb{1}$	Constant function equal to 1

Chapter 1

Introduction

The mathematical framework to understand the evolution of planet Earth's atmosphere and oceans is rapidly growing towards higher levels of precision. The complexity of planetary motion in all its vastness and heterogeneity demands new and almost exclusive theoretical techniques that, in many cases, entail unveiling concepts in mathematics. This thesis precisely investigates mathematical tools useful to understand the physics of our climate's response and modelling. The lines that follow in this chapter will be devoted to introducing motivating elements arising from climate science and defining the mathematical setting for what remains of the dissertation.

The Earth's climate is a forced, dissipative and chaotic complex system. In this sense, the climate is out of thermodynamic equilibrium due to the inhomogeneous absorption of solar radiation, whereby latitudes around the Equator are heated more in comparison with the poles. To a large extent, the resulting global circulation is defined as the system of winds necessary to reduce temperature gradients. Moreover, while the atmosphere is heated from below, the oceans are so from above, thus provoking intrinsically different mechanisms and scales of motion in every component, whose expected state is what we precisely understand as climate. Consequently, the climate system can be seen as an imperfect engine able to transform potential energy arising from temperature differences into kinetic energy in the form of winds and oceanic currents resulting in a wide range of dynamical processes.

The Earth climate system consists of a large network of non-linearly interacting phenomena. This multiscale character is due to a combination of the following factors: the nature of the external forcings, the inhomogeneity of the properties of the system's various components, the complexity of the coupling mechanisms between the latter, and the variety of instabilities, dissipative processes and feedbacks acting at different scales. It is well-nigh impossible, therefore, given our current scientific knowledge and our available or even foreseeable technological capabilities, to create a numerical model able to directly simulate the climate system in all details for all the relevant timescales, which span a range

of over fifteen orders of magnitude [PO92]; [GL20]. In many cases, both the theoretical understanding of such systems and the formulation of numerical models for simulating their properties are based on focusing upon a reduced range of large spatial and long temporal scales of interest, and upon devising an efficient way to capture effectively the impact of the faster dynamical processes acting predominantly in the neglected smaller spatial scales [PS08]; [PW09]. Hence, one has to focus on a specific range of scales through suitably developed, approximate evolution equations that provide the basis for the numerical modelling. Such equations are derived from the fundamental laws of planetary motion through systematic asymptotic expansions that are based on imposing an approximate balance between the forces acting on geophysical flows. These balance relations lead to removing small-scale, fast processes that are assumed to play a minor role at the scales of interest by filtering out the corresponding waves [HH13]; [Val06].

In climate science, parametrisation schemes have been traditionally formulated in such a way that one expresses the net impact on the scales of interest all processes occurring within the unresolved ones via deterministic functions of the resolved variables, as in the pioneering work on the parametrisation of convective activity by Arakawa and Schubert [AS74]. Gravity waves, albedo, vegetation, urban areas and cloud microphysics are some other features that are currently parametrised in weather and climate models [PO92]; [PW09]. At a theoretical level, however, one would still seek for a model that is capable of explaining the nature our climate's variability, upon making physically sound assumptions on the equations of motion. This is the scope taken in [Has76], where climate is assumed to evolve in an infinitely slower manner compared to weather disturbances, thus allowing to parametrise the latter in shape of random, noisy fluctuations. This modelling angle is also justified at a mathematical level where, in the limit of infinite timescale separation, white noise fluctuations appear out of a multiscale system [PS08]. Climate is, thus, regarded as a Brownian particle embedded in a pool of small fluctuating molecules that represent weather, albeit out of thermodynamic equilibrium. This point of view allowed to, for the first time, explain the observed continuous spectrum of climate's variability out of internal processes rather than based on external forcing agents like solar irradiance or other astronomical phenomena.

A clear-cut separation of timescales is, however, defied by Stommel diagrams which do not only reveal the already mentioned variability at a huge range of temporal scales, but also at spatial ones up to the order of approximately 10^4km [Sto63]. This amounts to having a vast collection of internal feedbacks that adjusts climate's response to disturbances, but that can also transfer energy across scales. The hypothesis of infinite timescale separation is, thus, purely academic. The intertwined spectrum of variability makes the presence of delayed effects an inevitable product of neglecting any physical constituent in a dynamic model. Mathematically, this is a well known consequence of projecting variables onto reduced-dimension phase spaces where self-intersecting trajectories no longer lead to

periodic motion, i.e. the semigroup property of the flow is lost. Consequently, the earlier consideration of climate being a "non-equilibrium Brownian particle" is no longer justified, unless the friction term is let to have memory on the past states. In this sense, the generalised Langevin equation (GLE) describes the evolution of Brownian particles embedded in a thermal reservoir, and consists of a stochastic integro-differential equation for the momentum variable containing the integral of a memory kernel and random forcing [Zwa01]; [Pav14].

Efforts should, therefore, be invested in constructing parametrisations with suitable memory kernels in addition to the stochastic forcing. To this end, the theories of statistical physics come to our aid in finding simplified equations that (a), describe the time evolution in reduced phase spaces and (b), incorporate the statistical responses due to coupling physical processes. While item (a) was briefly mentioned in the previous paragraph and stems from the classical theory of Langevin particles [Zwa61], item (b) aligns with more recent advances on the mathematical foundations of modern response theory [WL12]. When a system is forced with another, response theory provides parametrised equations that model the variables of interest without making reference to the coupled steady state, hence providing explicit stochastic and memory corrections. The price to pay is that the accuracy is weighted by the coupling strength between the processes: strong couplings lead to less accurate parametrisations. On the other hand, the derivation does not invoke timescale separations, clearly in line with the climatological phenomenology. On a practical note, one can treat the subgrid processes as the forcing element of the large scale circulation [VL18a]. For instance, one can regard the upper layers of the atmosphere as a subgrid forcing to the lower counterpart and oceans, while still preserving memory effects [DV17].

More recently, it has been recognised also on empirical grounds, that parametrisations should involve stochastic and non-Markovian components [Fra+15]; [PW09]. Indeed, the availability of data brings the opportunity of developing models that, while constructed out of a partially observed system, it allows to emulate its statistical behaviour. Along these lines of thought, one would aim at finding a data-driven approximation of the ultimate theoretical target of the GLE. Successive regressions of empirically obtained tendencies allow to write explicit stochastic equations with general time-correlation properties, albeit without memory in the GLE sense [Wil05]. Extending phase space, on the other hand, is a straightforward manner of incorporating memory, yielding an empirical model reduction (EMR) of the problem's dimensionality [KKG05]. In fact, the EMR procedure resembles the Markovian representation of the GLE and, hence, it is regarded as data-driven approximation of such equation [KCG15]. Thus justifies the success of EMR in capturing the multimodality and variability spectra out of observational input. Further along the data-driven methodologies, machine learning techniques have been proposed as the next frontier of parametrisations, e.g. [Gen+18]; [WDC20], able to deliver a new generation of Earth system models [Sch+17]; see, though, the caveats discussed by [HV18].

Disentangling free from forced variability is at the core of climate sciences. We have so far briefly mentioned how internal processes are likely to explain the spectrum of motion, although forced variability cannot be neglected. Indeed, at an astronomical level, the solar radiation, Earth's eccentricity and translation are outside factors that heavily condition our planet's climate. On the opposite end, endogenous or terrestrial forcings can be largely gathered as those affecting the composition of Earth's atmosphere and can have natural or anthropogenic origin. Volcanic eruptions, for instance, entail a dramatic input of aerosols into the atmosphere in the form of ash together with sudden releases of trapped CO_2. More importantly, variations in the land use and greenhouse gas emissions due to human activity are yielding changes in Earth's climate, which is projected to experience a global increase of up to $+2^{\circ}\mathrm{C}$ average surface temperature by 2050 relative to the end of the nineteenth century [IPC14]; see also the latest IPCC report [IPC21]. Climate sensitivity is, hence, understood as a measure of robustness of climatic variables in the presence of a external stimulus. One way of estimating the latter is by resorting to global circulation models (GCMs) that can be solved parallelly to obtain a projection of climate into the future with certain probability. This is what it is, in essence, done in the recent IPCC reports. However, the coarse resolution of GCMs leads to a lack of internal variability that smudges out non-linear processes that can trigger drastic changes on averaged climatic quantities [Ghi15].

Climate's sensitivity to external forcings has to, therefore, be approached by using raw numerical simulations of comprehensive GCMs, but also tackled at a fundamental level where physical and mathematical tools are useful to explain the nature of smooth, sudden or rough changes in the climate system. Investigations of basic principles for climate's response follow the need of a hierarchy of models which can enhance our understanding of what we mean by *changing climate* [Hel05]. While globally averaged surface temperature suggests our moving to a new climatological regime due to increasing carbon dioxide emissions, other localised phenomena are susceptible of experimenting abrupt changes as a product of external stimuli. For example, it has been noted that the sea surface temperature (SST) anomaly due to the El Niño Southern Oscillation possesses statistics— spread and skewness of the distribution, to be exact— that depend roughly on the travel time of equatorial waves [Che+14]. This means that there are intervals in which the moments of the SST vary smoothly with the tuning of equatorial wave travel time, but others where changes are wildly saw-toothed. Not less interestingly, such a behaviour was associated with a failure of the system's *linear response* as a product of small spectral gaps of empirically learned Markov matrices. This link comes from the modern spectral theory of dynamical systems which will be visited in this thesis.

Exploring parameter space to understand different forcing scenarios is the straightforward way of assessing future climates, although this methodology entails several difficulties. First, long and computationally expensive trajectories need to be integrated (for each

value of the forcing parameter) in order to avoid pathological statistics, like multivalued, regime dependent means [LAH00]. Adjoint methods are also examined in such work, where it is concluded that these do not serve to calculate statistical sensitivities of climate models. The reason being that adjoint techniques heavily depend on initial conditions and long chaotic trajectories provoke diverging estimates of statistical sensitivities, although recent work along these lines has to be mentioned [Wan13]; [NW17]. Hence, one is obliged to find ways of *predicting* the robustness of a system to perturbations. In this regard, *fluctuation-dissipation* theorems of statistical physics grant us a means of relating the response of a system with its natural fluctuations in the shape of correlation functions. Such a theoretical framework gives a recipe where, by examining the present climatology, one can predict its response to a prescribed forcing without examining the forced scenario [Lei75]. The validity of this theorem, though, is restricted to conservative systems and those which possess smooth statistics, possibly due to the presence of a stochastic component that spreads the noise to all model variables.

Recent advances regarding the fluctuation-dissipation theorem clarified that such approach is not so restrictive after all, even for dissipative systems like the climate, if one imposes conditions on the applied forces and the underlying dynamics are sufficiently chaotic [Rue09]; [GC95]. A major consequence is that linear response theory can extensively be exploited with great flexibility since it allows to compute statistical response out of empirically learned Green functions of the unperturbed model [LV07]; [LS11], and it gives a (mathematically) natural definition of climate sensitivity [RLL16]. Moreover, such theory disentangles forcing from time-modulation and, consequently, the methodology can be applied in parallel for a variety of scenarios. More concretely, this approach has been applied to a fully coupled ocean-atmosphere GCM for instantaneous and progressive CO_2 doubling to predict the increase of globally averaged near-surface temperatures but also the intensification of the Antarctic Circumpolar Current, which certainly is a large-scale, yet more localised phenomenon [LLR20].

Response theories of any sort, to which the fluctuation-dissipation theorem belongs, cannot predict statistical changes that result from a *bifurcation point*; where these refer to any substantial and non-smooth change in the dynamical behaviour of a physical system. These are, in fact, attributed to mathematically rigorous changes in a model's trajectories' stability and topological features [ER85]; [Ash+12]. In physical terms, the failure of response theories in this situation is expected since, in the language of the already mentioned work [Lei75], climate's response is determined by the free fluctuations which, by definition, cannot predict radically different unexplored climatological regimes. As a matter of fact, bifurcations in climate science are, to some extent, considered as non-equilibrium phase transitions that can lead to drastic changes in the atmospheric conditions incompatible with life. Far from being academic lucubrations, such critical transitions do occur, for instance, as a result of the well-established Earth's multistability. Indeed, the

climate system supports two stable states, the Warm state, which corresponds to the one we currently live in and the Snowball state, referring to a glaciated planet with an average surface temperatures of 200 to 220K [Ghi76]; [Mar+21]. These stable states alternate with each other as a result of fluctuations in solar radiation intake, mainly affected by the ice-albedo feedback [Bud69]. One can view the Warm and Snowball states as the two competing ends of a tug-of-war where middle-point or *Melancholia state* [LB17] separates the two basins of attraction, similarly to Brownian particles in a double well potential [LB19]. One is, therefore, inclined to device early warning signals for critical transitions. The failure of linear response is an indicator at an abstract level, although this can materialise in other computable quantities, like the observed slowing down of the decay of correlations in the Warm-to-Snowball transitions [Tan+18]. Along these lines, although in broader generality, the divergence of dynamic susceptibility function [ZLP21] or the enhancement of Lagrangian mixing [NPGR21] are recent features observed in systems undergoing a critical transition.

1.1 Dynamical Systems Perspective

Some fluid mechanical systems display periodic patterns, while many others behave in a chaotic, haphazard way. It is now well known that such an irregular behaviour is not due to inaccurate computations or long transient aperiodicity, but it can be an intrinsic property of the system especially those leading to turbulent fluid flows [Lor84]. A defining characteristic in this respect refers to the sensitive dependence on initial conditions where nearby trajectories on phase space separate substantially from each other in finite time [Lor63]; [ER85]. This is believed to be the agent responsible for preventing accurate weather predictions beyond ten days and the justification of why climate refers to the expected state of Earth's components, regardless of its initial configuration. The dynamical system's perspective focuses on the long-term dynamics of the Earth system and aims at determining the probability of encountering its constituents at a certain state. In this regard, it is less important whether the model under study is deterministic or stochastic since, loosely speaking, they will be indistinguishable from the viewpoint of averages and probabilities. In the paragraphs that follow in this introduction, we shall introduce and motivate the key concepts in the ergodic theory of dynamical systems, that will be crucial to develop the main contents of the present dissertation.

In general, the Navier-Stokes equation for the conservation of momentum constitutes the base for any weather or climate model. While the well-posedness of its solutions is an open problem, the two-dimensional version has been shown to, asymptotically, display its dynamics around a finite dimensional attractor, whose fractal dimension is inversely proportional to the kinematic viscosity [Rob01]. Hence, systems of deterministic ordinary

differential equations may be used to represent forced dissipative hydrodynamic flows not only out of procedural reasons, but also out of a fundamental result. Dynamical systems can be defined as the solutions of ordinary differential equations (ODEs) on smooth manifolds $\mathcal{X} \subseteq \mathbb{R}^d$ of integer dimension d [KH95]; [Tes12]. In our context, $\mathcal{X}$ will be equipped with the Euclidean topology so the usual notions of continuity, opennes or compactness can be invoked. Specifically, let $\mathbf{x} : \mathbb{R} \longrightarrow \mathcal{X} \subseteq \mathbb{R}^d$ satisfy the following ODE:

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}), \tag{1.1}$$

where $\mathbf{F} : \mathcal{X} \longrightarrow \mathcal{X}$ is a continuously differentiable and globally Lipschitz vector field. If an initial condition $\mathbf{x}_0$ in $\mathcal{X}$ is prescribed to Eq. (1.1), so that $\mathbf{x}(0) = \mathbf{x}_0$, then, by virtue of classical existence results, $\mathbf{x}$ is the only function satisfying Eq. (1.1) with the said initial condition. Moreover, such solution $\mathbf{x}$ is well defined for every *time* in $\mathbb{R}$, since we are taking a sufficiently regular vector field $\mathbf{F}$. Then, the *dynamical system* or *flow* $\{\phi^t\}_{t\in\mathbb{R}}$ associated with the ODE (1.1) is defined as follows for each t in $\mathbb{R}$:

$$\begin{aligned} \phi^t : \mathcal{X} \subseteq \mathbb{R}^d &\longrightarrow \mathcal{X} \\ \mathbf{x}_0 &\longmapsto \mathbf{x}(t), \end{aligned} \tag{1.2}$$

where $\mathbf{x}_0$ in $\mathcal{X}$ is the initial condition for which $\mathbf{x}$ is the unique solution to Eq. (1.1). The regularity of ϕ^t in time and space variables is inherited from that of $\mathbf{x}$. The physical interpretation of the flow ϕ^t at time t, is the function that maps initial conditions to their respective location on phase space after t time units have passed. Notice that we are allowing for negative values of time, which refer to the inverse function, whose existence is guaranteed because for each $t > 0$, ϕ^t is a diffeomorphism. Consequently, the flow $\{\phi^t\}_{t\in\mathbb{R}}$ enjoys the group property with respect to time.

Physical experiments with dynamical systems exhibit transient behaviour before entering an asymptotic regime in an *attractor* which determines the dynamics in the long term [ER85]; [Tes12]. In this sense, a global attractor is the maximal compact invariant set, such that points in any bounded set of $\mathcal{X}$ get arbitrarily close to it [Rob01]. The geometry of the attractor is fundamental in the understanding of the nature of the dynamics. For instance, flow trajectories determined by uniformly hyperbolic attractors [KH95] experience an average stretching and folding that automatically confers the already mentioned sensitive dependence to initial conditions. However, in order to tackle climatological questions, trajectory-wise approaches to the investigation of dynamical systems have to be replaced by a statistical interpretation. Along these lines, the initial condition mentioned earlier is substituted by ensembles or distributions that are evolved in time under the action of the dynamical system $\{\phi^t\}_{t\in\mathbb{R}}$ associated to Eq. (1.1). Consequently we require the notion of measurability. Let $\mathcal{A}$ denote the Borel σ-algebra on $\mathcal{X}$. It follows from the properties of the flow $\{\phi^t\}_{t\in\mathbb{R}}$, and ultimately from the regularity of $\mathbf{F}$, that ϕ^t is measurable for every t

in $\mathbb{R}$ and its action on measurable sets is defined element-wise. A probability measure μ is said to be *invariant*, relative to the dynamical system $\{\phi^t\}_{t\in\mathbb{R}}$ if the following condition is satisfied:

$$\mu\left(\phi^{-t}(A)\right) = \mu(A) \text{ for all } A \in \mathcal{A}. \tag{1.3}$$

Notice that the composition of the measure with the flow at time $-t$ is the so-called pushforward formula which, in general, transfers the probability measure μ to $\mu \circ f^{-1}$ for a given measurable, not necessarily invertible function f acting on $\mathcal{X}$.

The invariant measure is an object of crucial importance in the study of dynamical systems, since depending on its nature it can tell us the probability of finding the state of trajectories on phase space when the initial condition is not known. Heuristically, if a typical, long trajectory is able to explore the regions of $\mathcal{X}$ where the invariant measure μ is non-zero, then we expect that it will serve as a gauge for the average behaviour of the system in the long term. In fact, this intuition can formally be gathered in terms of *ergodic* measures. An invariant measure μ is said to be ergodic if for any measurable set A in $\mathcal{A}$ such that $\phi^{-t}(A) = A$ for any t in $\mathbb{R}$, $\mu(A) = 0$ or $\mu(\mathcal{X} \setminus A) = 0$. For ergodic measures, we have the following theorem:

Theorem 1.1.1 (Birkhoff)**.** *Let $\{\phi^t\}_{t\in\mathbb{R}}$ be a dynamical system acting on the measure space $(\mathcal{X}, \mathcal{A}, \mu)$ and let μ be an ergodic invariant measure relative to $\{\phi^t\}_{t\in\mathbb{R}}$. Then, the following limit holds*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \Psi\left(\phi^t(\mathbf{x})\right) \mathrm{d}t = \int_{\mathcal{X}} \Psi(\mathbf{x}) \mu(\mathrm{d}\mathbf{x}), \tag{1.4}$$

for any integrable function Ψ and μ-almost every $\mathbf{x}$ in $\mathcal{X}$.

This equality states the conditions for time averages to be interchangeable with phase integrals, and the assumption of its applicability to physical systems constitutes one of the cornerstones of statistical mechanics, namely, the ergodic hypothesis [KG14]. The function Ψ in Theorem 1.1.1 represents a physically measurable function of state or *observable*. Thus, evaluating Ψ after ϕ^t, has to be seen as successive snapshots of the dynamics filtered through a suitable observable [KG14]. The almost-everywhere formulation of Birkhoff's theorem is restrictive if the invariant measure of the system is singular with respect to Lebesgue. In this case, it is operationally impossible to sample a dynamical system from a set of Lebesgue-volume equal to zero. However, some systems do seem do posses agreeing statistics for randomly initialised initial conditions. This suggests the possibility of extending Birkhoff-type averages to initial conditions not necessarily sampled from the invariant measure which, in most cases, is not known. The sort of measures where Theorem 1.1.1 can be concluded for Lebesgue-almost every $\mathbf{x}$ in $\mathcal{X}$ are called *physical measures* [ER85]; [GC95].

In general, ergodicity allows to establish Césaro-convergence results, where limits are taken in the mean sense [Hal17]. When such limits hold also for non-averaged quantities, the dynamical system in question possesses the property of *mixing* [LM94]; [Hal17]. Formally a dynamical system $\{\phi_t\}_{t\in\mathbb{R}}$ is mixing if the limit

$$\lim_{t\to\infty} \mu\left(\phi^t(A)\cap B\right) = \mu(A)\mu(B), \tag{1.5}$$

holds for any pair of $\mathcal{A}$-measurable sets A and B. It can be shown that mixing implies ergodicity and, in this respect, the former is stronger than the latter. Moreover, one of the characteristic properties of mixing systems is that correlation functions decay in time [LM94]; [Rue86]. Correlations are defined between two observed quantities and measure their statistical relationship over time, as well as indicating the loss of memory of a dynamical system with respect to its initial state. Formally, the correlation function between two observable functions Ψ and Φ at time t is defined as:

$$C_{\Psi,\Phi}(t) = \int_{\mathcal{X}} \Psi(\mathbf{x})\Phi(\mathbf{x}(t))\mu(\mathrm{d}\mathbf{x}) - \int_{\mathcal{X}} \Psi(\mathbf{x})\mu(\mathrm{d}\mathbf{x}) \int_{\mathcal{X}} \Phi(\mathbf{x})\mu(\mathrm{d}\mathbf{x}). \tag{1.6}$$

Even if our setting is that of chaotic dynamical systems, the presence of periodic behaviour can be found when correlation functions do not decay with time. If they decay sufficiently fast so that the Wiener-Khinchin theorem can be invoked, regarding Eq. (1.6) in Fourier domain with $\Psi = \Phi$ gives the power spectra of the state-function Ψ, capable of indicating the periodic or quasiperiodic behaviour of the system [ER85]. More importantly, one is able to extract the peaked dominant frequencies which might stand out in a broadband spectrum typical of noisy systems.

1.1.1 Stochastic Systems

By and large, the trajectory-based study of dynamical systems should be replaced by the statistical point of view as a result of not having full knowledge of the initial conditions. If one furthermore adds uncertainty into the driving fields, one is lead to stochastic differential equations (SDEs) whereby noise acts as an intrinsic part of the motion on phase space. This procedure has now become fundamental in the Earth science community where, as noted at the beginning of this Introduction, despite there not being a clear-cut scales separation, climate variability can be framed as the overall response to short-term, weather-type disturbances or arising from subgrid processes [PW09].

Noise is not only added on the basis of model uncertainty or parametrisation purposes, but it is also understood from physical principles. Indeed, Brownian particles embedded in a fluid (or heat bath) experience frictional and fluctuating forces arising from the interaction with the environment that give rise to the Langevin stochastic equations. Those forces are, furthermore, related to each other by the first form of the fluctuation-dissipation theorem

we mentioned earlier in the Introduction, which relates the fluctuating and frictional terms [Ris89]; [Zwa01]. In this sense, fluctuations of Brownian particles are understood as sporadic collisions which, in spite of being deterministic, on average they are *exactly* represented by SDEs.

We consider a d-dimensional dynamical system induced by an Itô SDE of the following form:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t))\mathrm{d}t + \Sigma(\mathbf{x}(t))\mathrm{d}W_t, \tag{1.7}$$

where the drift component is determined by the the vector field $\mathbf{F} : \mathcal{X} \subseteq \mathbb{R}^d \longrightarrow \mathbb{R}^d$ and the stochastic part of Eq. (1.7) consists of the $d \times d$ covariance matrix $\Sigma\Sigma^\top$ and the d-dimensional Wiener process W_t. We shall assume therefore that $\mathbf{x}(t)$ lives in $\mathcal{X}$ for every non-negative value of t. Since Eq. (1.7) allows for non-linear drift vector fields and covariance matrices, we shall require a global Lipschitz and growth condition that ensures the existence of a solution globally in time that continuously depends on the initial condition $\mathbf{x}(0) = \mathbf{x}_0$ in $\mathcal{X}$[1]. We highlight that the unbounded character of the noise can force $\mathcal{X} = \mathbb{R}^d$, although for generality we shall stick to $\mathcal{X}$ for the phase space notation. Notice that while the flow generated by Eq. (1.1) consists of diffeomorphisms, the stochastic flow emanating from Eq. (1.7) is, in general, not reversible and, hence, not univocally defined for negative times. In this sense, the addition of noise naturally introduces a form of diffusion [Pav14].

Invariant measures in stochastic systems have to be understood in the sense of Birkhoff averages like that in Eq. (1.1.1), where, now the system is stochastic [DZ96]. Related concepts like ergodicity and mixing are analogous extensions from deterministic flows [Tan+20]. Additionally, the introduction of noise, has a smoothing effect in the sense that it forces the trajectories to explore wider regions of phase space and long term evolution of volumes are likely to posses non-vanishing Lebesgue measure. However, since we have not made any assumptions on the covariance matrix $\Sigma\Sigma^\top$, we cannot guarantee smoothness of the transition probabilities between states. The idea is that if noise spreads out through all the variables of the system, chances are that the transition probabilities between states are determined by smooth probability densities. To this end, Hörmander's condition is invoked so that at any point of phase space $\mathcal{X}$, the tangent linear can be recovered by the the directions generated by second order differential operator in Eq. (1.7) stemming from the noise. Hence, the smoothness and boundedness of the coefficients of Eq. (1.7) together with Hörmander's condition [Hör67] ensures the existence of smooth transition probabilities solving Eq. (2.16) and the invariant measures μ possesses a density representation with respect to Lebesgue measure [Pav14, Chapter 6].

[1]For well posedness and stability results for SDEs, the reader is referred to [Gar09] for a quick introduction and [DZ96] for more general results.

1.1.2 Response Theory

When forcing is introduced in a system, its behaviour will automatically change and evolve towards a new regime or statistical steady state. This amounts to noting that applying extraneous fields into Eq. (1.1) or Eq. (1.7), will result in a change of the system's climatology or invariant measure; see Eq. (1.3). When one cannot afford to explore all the possible outputs of a system due to applied forces, one would benefit from having a formula capable of describing the induced statistical changes in terms of the unperturbed regime; hence, gaining predictive skill.

If the vector field on the RHS of Eqs. (1.1) and (1.7), is heuristically perturbed in the form of $\mathbf{F} \mapsto \mathbf{F} + \varepsilon\mathbf{G}$, for some vector field $\mathbf{G} : \mathcal{X} \longrightarrow \mathcal{X}$ and a small real number ε, response theory in a technical sense aims at determining the mathematical properties of the following map:

$$\varepsilon \longmapsto \int_{\mathcal{X}} \Psi(\mathbf{x})\mu_\varepsilon(\mathrm{d}\mathbf{x}), \tag{1.8}$$

for some generic observable Ψ and invariant measure μ_ε. In particular, one is interested in understanding in what sense the response map (1.8) could be differentiable so that statistical averages would change smoothly with respect to the parameter ε. Contrarily, the lack of smoothness implies a rough dependence of model statistics and it could, furthermore, be attributed to the presence of a bifurcation point provoking drastic changes in the systems attractor [Bal14].

The first result on response theory was already pointed out earlier by making reference to the fluctuation-dissipation theorem [Kub66], by which the derivative of the response map (1.8) can be recast as the integral of a correlation function in the unperturbed regime; see [Kub57] for the non-rigorous, yet physically meaningful derivation in the context of equilibrium systems subject to time-dependent external fields. Ever since, linear response theory became a pillar or modern statistical physics; see [Mar+08] for a review. With the extension to non-equilibrium systems, correlation formulas no longer explained the system's response, in view that invariant measure μ_ε lacks the regularity that would make the response map differentiable in the usual sense. It is, nevertheless, useful to find a formal derivative of the response map with respect to epsilon [Rue98]; [Rue09]:

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\int_{\mathcal{X}} \Psi(\mathbf{x})\mu_\varepsilon(\mathrm{d}\mathbf{x})|_{\varepsilon=0} = \int_0^\infty \int_{\mathcal{X}} \mathbf{G}(\mathbf{x}) \cdot \nabla \left(\Psi(\mathbf{x}(t))\right) \mu(\mathrm{d}\mathbf{x})\mathrm{d}t, \tag{1.9}$$

although this formula is not rigorously proven to converge in general, unless we impose further stability conditions on attractor that would make its topology and the supported invariant measure more robust to changes in ϵ.

In order to achieve convergence, of Eq. (1.9) the idea is that the external field within the integral on the RHS of Eq. (1.9) is decomposed in into elements that carry different dynamical information, and this can only be attained by imposing a finer geometric

structure to phase space. Indeed, if the underlying attractor is uniformly hyperbolic [KH95], the tangent map $T_{\mathbf{x}}$ at each point $\mathbf{x}$ has a continuous invariant splitting $T_{\mathbf{x}} = E^s(\mathbf{x}) \oplus E^c(\mathbf{x}) \oplus E^u(\mathbf{x})$ with constants $\chi, \lambda > 0$ such that

$$\left\| T_{\mathbf{x}}\phi^t \mathbf{u} \right\| \leq \chi e^{-\lambda t} \|\mathbf{u}\|; \tag{1.10a}$$

$$\left\| T_{\mathbf{x}}\phi^{-t} \mathbf{v} \right\| \leq \chi e^{-\lambda t} \|\mathbf{v}\|, \tag{1.10b}$$

for every $\mathbf{u}$ in E^s and $\mathbf{v}$ in E^u, and, additionally, $E^c(\mathbf{x})$ is one-dimensional. The norm $\|\cdot\|$ comes from a suitably defined Riemannian metric on $\mathcal{X}$ [Rue97]; [Bal14]. While uniformly hyperbolic systems posses sensitive dependence with respect to initial conditions, they are *structurally stable* conferring robustness to perturbations in a topological sense [KH95]. Whether such structural stability translated into a statistical sense was not ascertained until the 90s when D. Ruelle, found that the smoothness of the invariant measure along the unstable manifold ensured the convergence of the linear response formula (1.9) [Rue97]. Indeed, having the splitting (1.10) at hand, one can decompose the perturbation vector field so that $\mathbf{G} = \mathbf{G}^s + \mathbf{G}^c + \mathbf{G}^u$, where, in fact, $\mathbf{G}^c = f\mathbf{F}$ for some scalar function of $\mathcal{X}$, f. Then, the contributions from the stable and unstable directions are disentangled [Rue98]; [Rue09]:

$$\frac{\mathrm{d}}{\mathrm{d}\varepsilon} \int_{\mathcal{X}} \Psi(\mathbf{x}) \mu_\varepsilon(\mathrm{d}\mathbf{x})|_{\varepsilon=0} = \int_0^\infty \int_{\mathcal{X}} \left[\left(T_{\mathbf{x}}\phi^t \right) \mathbf{G}^s(\mathbf{x}) \right] \cdot \nabla \Psi(\mathbf{x}) \mu(\mathrm{d}\mathbf{x}) \mathrm{d}t \tag{1.11a}$$

$$+ \int_0^\infty \int_{\mathcal{X}} \left[-\mathbf{F} \cdot \nabla f(\mathbf{x}) \right] \Psi\left(\phi^t(\mathbf{x}) \right) \mu(\mathrm{d}\mathbf{x}) \mathrm{d}t \tag{1.11b}$$

$$+ \int_0^\infty \int_{\mathcal{X}} \left[-\nabla_u \cdot \mathbf{G}^u(\mathbf{x}) \right] \Psi\left(\phi^t(\mathbf{x}) \right) \mu(\mathrm{d}\mathbf{x}) \mathrm{d}t, \tag{1.11c}$$

where $\nabla_u \cdot$ indicates the divergence on the unstable directions [Rue97]. The convergence of the LHS of Eq. (1.11a) now depends on how quickly $t \mapsto (T_{\mathbf{x}}\phi^t)\, \mathbf{G}^s(\mathbf{x})$ vanishes and the correlation functions of Eqs. (1.11b) and (1.11c) decay at infinity. Consequently, the fluctuation-dissipation theorem is recovered and correlation functions survive non-equilibrium backgrounds, albeit in a different form. It will be discussed later in this thesis that the poles of their Fourier transforms constitute a building block for the establishment of linear response— in the sense of Eq. (1.11)— but, more importantly, they are related to the physically relevant measurable quantities of correlation spectra [Rue86]; [BL07].

Positive results in achieving linear response beyond uniform hyperbolicity exist [Dol04], although it seems that the physical applicability of formulas like Eq. (1.9) can reach as far as analytically intractable flows, as with the applications to climate sensitivity [RLL16] and change [LLR20], both cases in coupled atmosphere-ocean models. In order to reconcile theory and observations, one is obliged to adopt the *chaotic hypothesis* [GC95], by which dynamics filtered through generic observables behave, at a practical level, like uniformly hyperbolic systems. This way, Boltzmann's ergodic hypothesis is refined by further

requiring a degree of structural stability both in a topological and metric sense. Under this framework, numerous works have managed to provide numerical evidence of linear response for non-uniformly hyperbolic systems [Rei02]; [LS11]; [WG19], to name a few.

1.2 This Thesis

In this thesis we shall apply the response theory for general dynamical systems to the analysis of climate response in a variety of contexts. Perturbations are here studied first, as an externally applied field and secondly, as the product of coupling two physical processes together. These two formats are in line with Earth's climate phenomenology depicted earlier in this Introduction, where it was clarified that any model should account for forced variability due to exogenous agents as well as internal variability due to non-linear interactions between multiple feedbacks. A probabilistic and statistical treatment of dynamical systems will be taken, delving into the analytical derivations but also aiming at identifying when the elementary results of response theory can be applied to conceptual climate models.

In Section 1.1, we explained that climatologies of dynamical models boil down to the determination of the invariant measure which is defined equally for deterministic and stochastic systems. In this sense, the evolution of probability measures in deterministic systems is the departing point, for which the Liouville equation for conservation of probability provides a consistent framework of study. Such equation can be understood in terms of operator semigroup theory [EN00]; [Paz12], as will be detailed in Chapter 2. We, furthermore, claim alongside many works, that the perturbation theory of operators is of great use to derive linear response relations. The modern spectral theory of dynamical systems suggest that the functional properties of the Liouville equation and its stochastic Fokker-Planck analogue provide valuable statistical information of the underlying system [Bal00]; [Tan+20]. This, together with the classical perturbation theory for linear operators— see, e.g., [Kat66]— will serve to find analytical formulas of response which translate to physical properties of the underlying system. It will be seen how the linear response formula Eq. (1.9) is recast into an operator equation which, in addition, decomposes into eigenpairs when the associated Fokker-Planck equation generates a quasi-compact semigroup [EN00]. The operator formulation will allow to treat time-dependent forcings allowing to recover the fluctuation-dissipation theorem, as well as the Green function which provide the leading order changes in the statistics. Furthermore, we shall systematically calculate the response due to stochastic perturbations in a seemingly equal way, establishing that the associated Green function formulation— see [Luc12]— can only be employed when noise is interpreted in the Stratonovich sense. The general aim of Chapter 2 is, therefore, twofold. First, we establish the functional analytical basis for what will follow in this

thesis and, second, we hope to convince the reader that operator semigroup point of view is valuable for the study of response in generic dynamical systems.

Since the Liouville equation provides a consistent framework for the treatment of probabilities in dynamical systems, it makes the associated transfer operator theoretically instrumental when dealing with the uncertainty in meteorological [Ehr94] and climatological forecasts [Has76][2]. However, the high number of degrees of freedom of a truncated atmospheric model translates into, practically, intractable dimensionalities under the transfer operator scope. Hence, we require efficient ways of approximating the transfer operator that go beyond discretising the Liouville equation. This problem is tackled in Chapter 3, where we project the transfer operator onto an amenable basis of functions, following the so-called Ulam's method [Ula64]; for comprehensive surveys, see [Fro98]; [Fro01]. The resulting Markov chain approximations recasts the problem of response into the perturbation theory of stochastic matrices, as initiated in [Luc16] and extended in [ADF18]; [SGL20]. At such level, response formulas will be derived, providing finite-dimensional analogues of the linear response of Eq. (1.8) and other dynamical quantities of interest. In this setting, we shall also derive the numerical algebraic conditions for constructing the linear response, giving stability and conditioning results for general stochastic matrices. An application of this methodology is done to the celebrated Lorenz 63 atmospheric convection conceptual model [Lor63], which will be perturbed by external fields. By working on phase space and sampling the unforced statistics, the perturbation theory for finite Markov chains will allow us to compute the linear response and sensitivities of such system, despite it possessing a singularly hyperbolic attractor [Tuc02], hence being ill-posed in the sense of Ruelle's response theory.

In Chapter 4 we address the question of how response theory allows to treat weakly coupled systems following [SG+21]. When two physical processes interact in a weak way, one can regard each of them as a perturbation of the other. Along these lines, response theory gives the model that captures the leading order statistical behaviour of the model of interest, hence parametrising the remaining one [WL12]. More concretely, the yielding parametrised Wouters-Lucarini (WL) equation provides stochastic and memory corrections that account for the neglected process. This point of view is a natural device to identify the non-Markovian effects that one gets when no strict separation of timescales is present [VL18a]. The WL equation thus provides a method to construct a model that captures the coupled statistics without necessarily sampling the full system. Such response theoretic model will be here derived from expansions of the Liouville equation, in lines with the previous chapters and extending the departing work of [WL13]. It will be noted, that having a spectral decomposition of the Liouville equation— in the spirit of Chaper 2— allows to recast the WL equation into an extended Markovian model with the structure of

[2]In [Has76], the author takes, rather, the stochastic extension of the Liouville equation, namely, the Fokker-Planck equation.

an Multilevel Stochastic Model (MSM) [KCG15], where the memory effects are explained by a hierarchy of linearly driven variables which arise from the spectral decomposition of the response due to the coupling. Interestingly, in the theory of the already cited empirical model reduction (EMR) [KKG05], multilevel regressions from partially observed data provide reduced-order equations with the MSM structure [KCG15]. In views of this analogy between the WL equation and the EMR protocol, we shall draw a conceptual link between them, further claiming that the convergence of the data-driven methodology is strongly linked to the spectral properties of the underlying Liouville operators.

Finally, the results of chapters 2, 3 and 4 will be broadly summarised in Chapter 5, along with the main conclusions. Although the theory of response is the backbone of the present dissertation, its investigation in the various contexts here exposed required the diversion onto some general branching problems that have been partially addressed. Some instances are: the spectral theory of stochastic matrices, the loss of semigroup property of projected transfer operators and the data-driven construction of non-Markovian models. Thus, in Chapter 5 we hope to give an outlook of the applicability, limitations and future research along the lines of this thesis.

Appendices are included in order to avoid the reader diverting from the main topics. They contain supporting calculations, but also supplementary content. In Appendix A an homogeneous equation for the linear response is calculated, revisiting the work of [Ken71]. As an extension of Chapter 3, Appendix B is devoted to calculating the leading order changes in the stationary vector of a stochastic matrix subject to atomic perturbations that only affect one state. For completeness, Appendix C applies Itô convolutions to solve a system of equations coupled to a linear stochastic differential equation. Finally, the aforementioned EMR data-driven protocol is applied in Appendix D to the coupled Lorenz 84 [Lor84] and Lorenz 63 [Lor63] systems to asses the problem of convergence under changes in (i), coupling strength and, (ii) the timescale separation.

1.2.1 Contributions

The contributions of this thesis are concisely listed below together with their relative publications:

Chapter 2. A decomposition of the Green function associated with a stochastic differential equation is presented in terms of the spectrum of quasi-compact Fokker-Planck semigroups. Such is the main content of Section 2.3. Applied fields are later generalised to be stochastically modulated in Section 2.3.1, and it is shown that the Green function formalism is equally employed although it implicitly assumes a Stratonovich interpretation of the noise. This is proved in Proposition 2.3.1. In this context, the linear response of correlation functions and power spectra is deduced

using the resolvent perturbation expansion, thus extending the study of [LW17]. These results are yet to be reported in an article.

Chapter 3. The suitability of stochastic matrices for calculating the linear response is investigated at a linear-algebraic level. Results regarding well-posedness, conditioning and stability are given in the propositions of Section 3.2.1. As an application of this theory, two low-dimensional non-equilibrium systems are investigated. Finite differences are then employed as a model for the probability fluxes entailed by prescribed perturbations and shown to be valid to predict the linear response of a given state function, even for the dissipative Lorenz 63 system. The main results in this chapter are gathered in [SGL20].

Chapter 4. In this chapter, systems of coupled equations are treated in a perturbative way. The operator relations of Chapter 2 are used to derive the WL equation—see [WL12]—, which captures the response of a system forced by another, hence extending the preliminary formulas of [WL13] to more general coupling laws. Furthermore, the memory kernel of the WL equation is spectrally decomposed (cf. Section 2.3), allowing to Markovianise said equation in a hierarchy of successively fast variables arranged in a multilevel stochastic model; see Theorem 4.2.1. An analogy is presented with the EMR methodology is drawn and suggested that its convergence depends on the spectral structure of the underlying Fokker-Planck operator, as indicated by the theory. Both techniques, the WL equation and the EMR are then applied to a conceptual climate model found in, e.g., [KCG15] to analyse their numerical performance. These results can be found in [SG+21].

Chapter 2

Operator Semigroups and Response

The realisation that a general system of ODEs can be regarded as a linear operator acting on function spaces dates back to the early 1930s when B. O. Koopman proved that change in state functions (observables) obeyed a unitary transformation in Hamiltonian systems [Koo31]. More importantly, it was then found that the spectral properties of such transformations could be linked to mechanical features of the system in question. The work with J. von Neumann extended the study of such transformations and linked their spectrum to the presence of chaotic motion [KN32]. Specifically, the existence of a continuous spectrum revealed that "the motion of any set M of Ω becomes more and more spread out into an amorphous, everywhere-dense chaos". Which heavily reminds of the concepts of ergodicity and mixing introduced in Section 1.1.

A probabilistic interpretation of chaotic flows, on the other hand, focuses on the evolution of density functions on phase space and, to this regard, the primary result is due to Liouville whose theorem states that conservative systems preserve the Lebesgue measure indefinitely; volumes are not contracted over time [KG14]. Such theorem is based on the continuity equation for the conservation of probability, whose solution generates the so-called *transfer operator* semigroup, to be defined precisely later. The preservation of probability and its non-negativity makes the transfer operator enjoy the properties of positive operators which inherit the consequences of the operator version of the Perron-Frobenius theorem [LM94]. Most important of the latter being that the invariant measure of the system— see Eq. (1.3)— is characterised by the spectral properties of the leading (unit) eigenvalue of the transfer operator.

The extension to general non-Hamiltonian systems far away from equilibrium makes the treatment of the Liouville operator substantially more difficult since usual function spaces do not contain the measures describing the statistical properties of the systems. This is due to the fact that an average contracting of phase space volumes provokes the invariant states to be supported in zero-volume sets making them singular with respect to the Lebesgue measure [Rue09] and hence not having a Radon-Nikodym derivative.

It seemed, thus, necessary to extend the functional setting so that the transfer operator semigroup could be used in such systems to extract dynamical information. It was not until the 2000s where more general Banach spaces of distributions where considered so that singular distributions were also taken into account in the transfer operator formalism [GL06]. This way, the initially conjectured power of the spectral theory of chaotic flows became mathematically available for a wider range of dynamical systems.

When the system under study experiences external perturbations, it is of interest to study and predict the smooth or abrupt effects on its average behaviour. To this end, the transfer and Koopman operator semigroups served to formulate the first results in the context of weakly forced equilibrium systems [Kub57]. Such a discovery allowed to link the response and robustness of the system to its natural fluctuations in the limit of infinitesimal, yet time-dependent forcings. This link is no longer available for non-equilibrium systems, at least, not in the usual way due to the lack of density representations of the invariant measure. It was not until the 1990s again when the smoothness of statistical steady states was established for structurally stable systems possessing uniformly hyperbolic attractors [Rue97]; [Rue98]. These results were, furthermore, achieved in the modern transfer operator framework giving evidence of the power of the methodology. Particularly, the subunitary eigenvalues (point spectra) reveal the dominating rates of decay of correlations [Rue86] as well as giving a tool to prove linear response rigorously [HM10]; [Bal14]. In this regard, the cornerstone result is due to G. Keller and C. Liverani, who demonstrate that the spectral features of the transfer operator are preserved under a wide range of perturbations which include stochastic forces and numerical discretisations [KL99].

The effectiveness of the transfer operator to provide dynamical and statistical information about a system motivates the rest of the chapter. More particularly, we want to apply the perturbation theory for strongly continuous semigroups to the transfer and Koopman operators to derive formulas that describe the statistical properties of a dynamical system subject to static, time-dependent and randomly applied fields. With the adequate functional setting in hand, the matter of investigation here is to determine the role of the spectral structure of such operators in quantifying the sensitivity of the system and finding its relaxation rates. Although the main results here concern stochastic systems, analogies with deterministic flows will be drawn since they are virtually indistinguishable from the operator point of view.

The rest of this chapter is structured as follows. In Section 2.1 the transfer and Koopman operators are introduced and their main properties are given for deterministic flows. This approach is suitably extended for stochastic systems in Section 2.1.1. Response to static forcings is investigated in Section 2.2 and linked to the spectrum of the Fokker-Planck semigroup in Section 2.2.1. An extension to time-dependent perturbations is done in Section 2.3 where the particular case of stochastically forced deterministic systems is included in Section 2.3.1. In Section 2.4, we revisit V. M. Kenkre's homogeneous

response formula [Ken71] under a stochastic a stochastic framework. Finally, a summary and overview is given in Section 2.5.

2.1 Transfer Operators: definition and properties

Departing from the dynamical system $\{\phi^t\}_{t\in\mathbb{R}}$ induced by the ODE in Eq. (1.1), we are interested in how probability density functions evolve on phase space $\mathcal{X}$ as opposed single trajectories. One heuristic way of visualising this would be to initialise system (1.1) with a distribution of initial conditions and ask what such a distribution is going to look like after some time under the action of the flow. To this end, we introduce the Borel σ-algebra $\mathcal{A}$ of open sets in $\mathcal{X}$ and a probability measure μ, as previously done in Section 1.1. For every A in $\mathcal{A}$, we define the transfer operator $\mathcal{P}_t : L^1_\mu(\mathcal{X}) \longrightarrow L^1_\mu(\mathcal{X})$ as:

$$\int_A \mathcal{P}_t\rho(\mathbf{x})\mu(\mathrm{d}\mathbf{x}) = \int_{\phi^{-t}(A)} \rho(\mathbf{x})\mu(\mathrm{d}\mathbf{x}), \tag{2.1}$$

for every t in $\mathbb{R}$, A in $\mathcal{A}$ and ρ in $L^1_\mu(\mathcal{X})$. The equation above defines an operator uniquely since the functions ϕ^t are non-singular [Hal17]; [LM94]. Linearity, positivity and conservation of probability of the transfer operator follow immediately from the definition above. More interestingly, the family of operators $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ satisfies the semigroup property in the same way $\{\phi^t\}_{t\in\mathbb{R}}$ does. Indeed, for every t, s in $\mathbb{R}$, and A in $\mathcal{A}$, we have:

$$\int_A \mathcal{P}_{t+s}\rho(\mathbf{x})\mu(\mathrm{d}\mathbf{x}) = \int_{\phi^{-(t+s)}A} \rho(\mathbf{x})\mu\left(\mathrm{d}\mathbf{x}\right) = \int_{\phi^{-s}(\phi^{-t}(A))} \rho(\mathbf{x})\mu(\mathrm{d}\mathbf{x}) \tag{2.2a}$$

$$= \int_{\phi^{-t}(A)} \mathcal{P}_s\rho(\mathbf{x})\mu(\mathrm{d}\mathbf{x}) = \int_A \mathcal{P}_t\mathcal{P}_s\rho(\mathbf{x})\mu(\mathrm{d}\mathbf{x}). \tag{2.2b}$$

Furthermore, since the transformation ϕ^t is invertible, $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ is, in fact, a group, although we shall nevertheless call it a semigroup for the sake of uniformity. The transfer operator can be defined for a wide range of dynamical systems that need not arise from an ODE, although in the latter case it can be written as a change-of-variable formula involving the Jacobian of the flow [LM94]. Hence, if ρ is continuous with compact support on $\mathcal{X}$, limits of $t \to 0$ can be taken uniformly on $\mathbf{x}$ to conclude that:

$$\lim_{t\to 0} \|\mathcal{P}_t\rho - \rho\| = \lim_{t\to 0} \int_{\mathcal{X}} |\mathcal{P}_t\rho(\mathbf{x}) - \rho(\mathbf{x})|\, \mu\left(\mathrm{d}\mathbf{x}\right) = 0. \tag{2.3}$$

A density argument extends this result to $L^1_\mu(\mathcal{X})$. Equation (2.3) not only exploits the explicit form of the transfer operator for ODEs, but it also shows that the semigroup of transfer operators $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ is a *strongly continuous* semigroup, for which there exists an extensive literature of results [Paz12]; [EN00]. As we shall see later, strong continuity is not a mathematical nicety, but it will allow us to regard the transfer operator semigroup as

a solution to an abstract initial value problem which will greatly facilitate its investigation.

While $L^1_\mu(\mathcal{X})$ is the natural space for density functions (with respect to μ), observable functions are their dual counterpart and are taken in $L^\infty_\mu(\mathcal{X})$ on which it is possible to define the dual operator semigroup $\{\mathcal{U}_t\}_{t\in\mathbb{R}}$, known as the Koopman semigroup:

$$\mathcal{U}_t\Psi(\mathbf{x}) = \Psi\left(\phi^t(\mathbf{x})\right), \tag{2.4}$$

for every t in $\mathbb{R}$ and Ψ in $L^\infty_\mu(\mathcal{X})$. It is not trivial to show that this operator is the dual of $\mathcal{P}_t$, but the proof involves the pairing by which duality is defined for ρ in $L^1_\mu(\mathcal{X})$ and Ψ in $L^\infty_\mu(\mathcal{X})$. This way, the transfer operator and the Koopman operator are related via

$$\langle \mathcal{P}_t\rho, \Psi\rangle = \int_{\mathcal{X}} \mathcal{P}_t\rho(\mathbf{x})\Psi(\mathbf{x})\mu(\mathrm{d}\mathbf{x}) = \int_{\mathcal{X}} \rho(\mathbf{x})\mathcal{U}_t\Psi(\mathbf{x})\mu\left(\mathrm{d}\mathbf{x}\right) = \langle \rho, \mathcal{U}_t\Psi\rangle\,. \tag{2.5}$$

The Koopman operator family $\{\mathcal{U}_t\}_{t\in\mathbb{R}}$ also constitutes a strongly continuous semigroup.

One of the primary results in the theory of strongly continuous semigroups allows to regard them as solutions to abstract initial value problems in suitable Banach spaces, by means of generalising the concept of exponential function. This procedure is achieved by identifying a generating operator densely defined on the set where all strong derivatives of the semigroup exist. In general, if $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ is a strongly continuous semigroup on the Banach space $\mathcal{B}$, its generator $\mathcal{L} : D(\mathcal{L}) \subseteq \mathcal{B} \longrightarrow \mathcal{B}$ is defined as:

$$\mathcal{L}\rho = \lim_{t\to 0} \frac{\mathcal{P}_t\rho - \rho}{t}, \tag{2.6}$$

where $D(\mathcal{L})$ is the set of all functions in $\mathcal{B}$ where the limit in Eq. (2.6) exists in the strong sense [EN00]. In the case of ODE-induced dynamical systems is specially handy, since it can be written in terms of the vector field $\mathbf{F}$, which is usually prescribed to the problem. To see this, the idea is to take continuously differentiable observables with compact support in $C^1_0(\mathcal{X})$ and apply Eq. (2.6) to the Koopman operator $\mathcal{U}_t$ defined in Eq. (2.4). Finally, one obtains a the generating operator $\mathcal{L}^* : C^1_0(\mathcal{X}) \longrightarrow C^1_0(\mathcal{X})$ defined as:

$$\mathcal{L}^*\Psi = \sum_{i=1}^{d} \left[\mathbf{F}\right]_i \partial_{x_i}\Psi = \mathbf{F}\cdot\nabla\Psi, \tag{2.7}$$

for every Ψ in $C^1_0(\mathcal{X})$ and $\mathbf{x}$ in $\mathcal{X}$. The generator formulation allows to view the non-linear ODE (1.1) as a linear partial differential equation (PDE) in an equivalent way. More precisely, if we— making a slight abuse of notation— define $\Psi(\mathbf{x}, t) = \Psi(\mathbf{x}(t))$, it solves the following equation:

$$\partial_t\Psi(\mathbf{x}, t) = \mathcal{L}^*\Psi(\mathbf{x}, t), \tag{2.8}$$

for every $\mathbf{x}$ in $\mathcal{X}$ and t in $\mathbb{R}$. The solution to Eq. (2.8) is precisely the Koopman semigroup

$\{\mathcal{U}_t\}_{t\in\mathbb{R}}$ and due to its analogy to finite dimensional linear ODEs, such semigroup can be denoted as $\{e^{t\mathcal{L}^*}\}_{t\in\mathbb{R}}$, where the usual exponential multiplication relations hold. The adjoint relation of the Koopman and transfer operator semigroups can be further exploited to show that $\{\mathcal{P}_t\}_{t\in\mathbb{R}}$ also satisfies an evolution equation like Eq. (2.8), where in this case the generator is denoted by $\mathcal{L}$ and defined as:

$$\mathcal{L}\rho = -\sum_{i=1}^{d} \partial_{x_i} \left([\mathbf{F}]_i \, \rho \right) = -\nabla \cdot (\rho \mathbf{F}) \,, \tag{2.9}$$

for functions ρ in $C_0^1(\mathcal{X})$. The evolution equation associated with the transfer operator is the Liouville equation [KG14]:

$$\partial_t \rho(\mathbf{x}, t) = \mathcal{L}\rho(\mathbf{x}, t), \tag{2.10}$$

where $\rho(\mathbf{x}, t)$ is the probability of encountering the system in $\mathbf{x}$ at time t. Notice that Eq. (2.10) is a conservation law that models the fact that probability is not created or destroyed. The generators for the Koopman and transfer operator semigroups were deduced for functions in $C_0^1(\mathcal{X})$ which is dense in $L_\mu^1(\mathcal{X})$ although not so in $L_\mu^\infty(\mathcal{X})$ which are the natural spaces where these semigroups are defined. Therefore, if the semigroups act on integrable or bounded functions, any reference to the generator will assume that the functions involve have the necessary regularity or compactness of their support.

An important consequence of the strong continuity of the semigroup generated by Eq. (2.10) is that it provides a way of finding the invariant measure of the system by means of the Hille-Yoshida theorem [LM94]; [EN00]. Indeed, if ρ is a density with respect to μ for the invariant measure of the dynamical system generated by Eq. (1.1), then we have $\mathcal{L}\rho \equiv 0$ and consequently, $e^{t\mathcal{L}}\rho = \rho$. It has already been noted at the starting paragraphs of the present chapter that such a density representation is not common when μ is the Lebesgue measure because in many systems out of equilibrium, non-conservative forces act on phase space so that volumes are slowly contracted making the resulting invariant measure singular with respect to Lebesgue [ER85].

If the dynamical system is measure preserving, for instance, if μ is an invariant measure of the system, the transfer operator $\mathcal{P}_t$ is an isometry in $L_\mu^1(\mathcal{X})$ [Hal17]. To see this, it is enough to prove it for finite combinations of characteristic functions on measurable sets to then extend it to non-negative functions by means of the monotone convergence theorem. This way, since $L_\mu^2 \subset L_\mu^1$ whenever μ is finite, $\mathcal{P}_t$ is also an isometry in L_μ^2, which enjoys a Hilbert structure. Moreover, if the flow ϕ^t is invertible, which in fact we are assuming, then $\mathcal{U}_t$ is also invertible. Hence, we conclude that $\mathcal{U}_t$ is an invertible isometry, for every t, making it a unitary operator. This result was first obtained by Koopman [Koo31], although a clear application into the dynamical systems theory was made by Von Neumann, where he proved that the limiting behaviour of such unitary operator converged, in the mean and

strong operator sense, to projections onto its fixed points [Neu32]. Such result is stated for generic unitary operators, family to which the transfer and Koopman operators belong, provided that they are defined in a suitable functional space [LM94].

Another consequence of the unitarity of the transfer operator semigroup is that its spectrum lies exactly on the complex unit ball. In particular, the invariant elements under $\mathcal{P}_t$ and $\mathcal{U}_t$ can be expressed as the set of eigenfunctions associated to the eigenvalue 1. In fact, the nature of such eigenvalue carries information about the dynamical and statistical features of the underlying system as illustrated in the following theorem [Hal17]:

Theorem 2.1.1. *A measure-preserving dynamical system $\{\phi^t\}_{t\in\mathbb{R}}$ is ergodic if and only if 1 is a simple eigenvalue of the associated Koopman operator $\mathcal{U}_t$ with constant associated eigenfunctions, for any t in $\mathbb{R}$. Furthermore, if $\{\phi^t\}_{t\in\mathbb{R}}$ is ergodic, every eigenvalue of $\mathcal{U}_t$ is simple and the set of eigenvalues forms a group on the complex unit circle with respect to multiplication.*

By the same token, mixing is characterised by the eigenvalue 1 being alone on the complex unit ball with simple algebraic multiplicity, thus, giving another justification of why mixing implies ergodicity [Hal17]; [LM94]. In practical terms, mixing is characterised by correlation functions (see Eq. (1.6)) decaying at infinity, and this, furthermore, indicates that the system approaches an invariant distribution [Liv95]. In fact, correlation functions of two generic observables in $L^2_\mu(\mathcal{X})$ can be expressed in terms of the transfer/Koopman operator:

$$C_{\Psi,\Phi}(t) = \int_{\mathcal{X}} \mathcal{P}_t\Psi(\mathbf{x})\Phi(\mathbf{x})\mu(\mathrm{d}\mathbf{x}) = \int_{\mathcal{X}} \Psi(\mathbf{x})\mathcal{U}_t\Phi(\mathbf{x})\mu(\mathrm{d}\mathbf{x}), \tag{2.11}$$

assuming that $\int_{\mathcal{X}} \Psi \mathrm{d}\mu = \int_{\mathcal{X}} \Phi \mathrm{d}\mu = 0$. If the system is chaotic, typical trajectories will decorrelate in time and observables behave, to a large extent, as random variables. However, since the spectrum of $\mathcal{P}_t$ and $\mathcal{U}_t$ do not display non-unitary eigenvalues by virtue of Theorem 2.1.1, one cannot explain the decay rate of $C_{\Psi,\Phi}(t)$ at infinity. On this note, for the already mentioned uniformly hyperbolic systems— see Section 1.1.2— correlation spectra are rigorously shown to posses poles on a strip in the negative complex plane associated with the leading decay rates and oscillations of $C_{\Psi,\Phi}$ [Rue86]; [Bal00]. In the transfer operator language, the usual L^p domains should be extended to anisotropic Banach spaces that capture the geometry of the attractor so that a non-trivial sub-unitary point spectrum is present [GL06]. More importantly, the latter eigenvalues correspond to the aforementioned poles in correlation spectra [BL07].

2.1.1 Stochastic Systems

In the previous section, we have seen that transfer and Koopman operator semigroups provide a method for studying deterministic flows in which their spectral properties encode

statistical information of the system. In stochastic processes, it is possible to extend the definition of such operator semigroups so that the effects of the noise is taken into account.

The fundamental relation between stochastic flows, like that generated by Eq. (1.7), and linear operators arises from the conservation probability in the same way the Liouville equation stands for deterministic systems. In this sense, if $\rho(\cdot, t)$ denotes a probability function on phase space $\mathcal{X}$, its evolution under the stochastic system (1.7) is determined by the Fokker-Planck equation [Ris89]; [Pav14]:

$$\partial_t \rho(\cdot, t) = \mathcal{L}\rho(\cdot, t) := -\nabla \cdot (\mathbf{F}\rho(\cdot, t)) + \frac{1}{2}\nabla^2 : \left(\Sigma\Sigma^\top \rho(\cdot, t)\right), \tag{2.12}$$

where $\mathcal{L} : D(\mathcal{L}) \subseteq \mathcal{B} \longrightarrow \mathcal{B}$ is a linear differential operator densely defined in a Banach space $\mathcal{B}$. Given an initial density ρ_0, the solution at time t is provided by the exponential operator $\rho(\cdot, t) = e^{t\mathcal{L}}\rho_0$. Furthermore, since the process Eq. (1.7) is Markovian, the family of operators $\{e^{t\mathcal{L}}\}_{t\geq 0}$ satisfies the semigroup property and in this sense, the operator $\mathcal{L}$ is said to be the generator of the semigroup. Notice that, as opposed to deterministic flows, noisy systems are not invertible and the semigroup $\{e^{t\mathcal{L}}\}_{t\geq 0}$ cannot be extended to negative times. This is also present in the Fokker-Planck formulation whereby the second-order differential operator acts as a diffusing agent, preventing reversibility. Furthermore, the existence of an invariant measure μ ensures that the semigroup generated by Eq. (2.12) is strongly continuous in $L^p_\mu(\mathcal{X})$ for every $p \geq 1$ and, hence, the domain of definition $D(\mathcal{L})$ is the set of functions ρ in $\mathcal{B} = L^p_\mu(\mathcal{X})$ such that the limit in Eq. (2.6) exists in the strong sense [Tan+20, Appendix A]. To see this, it is enough to consider Eq. (2.12) in a space of sufficiently regular functions which allows to extend the definition of $\{e^{t(\mathcal{L}_0+\varepsilon\mathcal{L}_1)}\}_{t\geq 0}$ onto $L^p_\mu(\mathcal{X})$ using density arguments.

We recall that the smoothness of transition probabilities $\rho(\cdot, t)$ is guaranteed by Hörmander's condition mentioned at the end of Section 1.1.1. In the language of generators, the second-order diffusion operator is sufficiently strong so that its regularising properties affect the solutions of Eq. (2.12); we refer the reader to [Pav14, Chapter 6]. Furthermore, the invariant measure of the system will posses a density representation with respect to the Lebesgue measure ρ_0, which solves and is characterised by $\mathcal{L}\rho_0 \equiv 0$, in allusion to the Hille-Yosida theorem.

The duality between densities and observable functions introduced in the previous section is maintained in the context of stochastic systems. This dual representation is taken in the L^2 sense and results in the *backward-Kolmogorov* equation describing the evolution of the expectation values of observable functions [Pav14]. Indeed, let Ψ be in $C_b(\mathcal{X})$— continuous and bounded functions— and let its expectation value with respect to $\rho(\cdot, t)$ be denoted by $\Psi(\cdot, t)$, then:

$$\partial_t \Psi(\cdot, t) = \mathcal{L}^*\Psi(\cdot, t) := \mathbf{F} \cdot \nabla\Psi(\cdot, t) + \frac{1}{2}\Sigma\Sigma^\top : \nabla^2\Psi(\cdot, t), \tag{2.13}$$

for every $t > 0$. In the same way as the Fokker-Planck equation generates a strongly continuous semigroup in $L^p_\mu(\mathcal{X})$, so does the backward-Kolmogorov equation in $L^q(\mathcal{X})$, where p and q are conjugate real numbers. In particular, if Ψ and Φ are zero μ-mean observables in $L^2_\mu(\mathcal{X})$, we have that correlation functions can be expressed in terms of the semigroup [Tan+20]:

$$C_{\Psi,\Phi}(t) = \int_{\mathcal{X}} e^{t\mathcal{L}}\Psi(\mathbf{x})\Phi(\mathbf{x})\mu\left(\mathrm{d}\mathbf{x}\right) = \int_{\mathcal{X}} \Psi(\mathbf{x})e^{t\mathcal{L}^*}\Phi(\mathbf{x})\mu\left(\mathrm{d}\mathbf{x}\right). \tag{2.14}$$

Because of the irreversibility induced by noise, correlations are in principle more likely to decay in stochastic systems. In fact, under mild conditions the operator $e^{t\mathcal{L}}$ reveals eigenvalues strictly within the unit ball which would imply that correlations decay at infinity at an exponential rate [Tan+20]. This considerations will become fundamental in the following sections and will be treated later with due detail.

Remark 2.1.1 (Notation)**.** *In what follows the subscript $\mathcal{X}$ for phase integrals $\int_{\mathcal{X}}$ shall be omitted when clear.*

Summarising Section 2.1, semigroups arise naturally in deterministic and stochastic flows out of the evolution of observables and probability density functions. While the Fokker-Planck equation generates the stochastic version of the transfer operator, the backward-Kolmogorov equation extends the Koopman semigroup to stochastic systems. It has been observed that, once within the theory of semigroups, the ideas and algebraic manipulations follow equally for deterministic and stochastic systems, albeit with functional analytical differences. Indeed, the singularity of invariant measures typical in non-conservative deterministic flows is replaced by smooth densities as results from applying noise. In addition, while stochastic systems reveal non-trivial eigenvalues in usual L^p spaces and are capable of explaining the decay of correlations, finding their deterministic analogues is much more complicated, see Theorem 2.1.1. In what follows, we shall take the stochastic approach for the very last reasons, although reference to deterministic systems will be made throughout the text.

2.2 Static Response

Static response refers to the leading order changes in the statistical properties of a system as a result of time-independent forcing and its approach to a new steady state. The goal of static response is, hence, twofold: (i) calculate the sensitivity to a prescribed field and (ii), estimate the relaxation timescales. In static response the dynamical system of interest is forced with a field that pushes the system to a new stationary state as time tends to infinity. The key technical advantage is that the resulting system is autonomous and the perturbation theory for semigroups can be conveniently applied; as will be detailed below.

We consider the d-dimensional Itô SDE (1.7) where an external field $\mathbf{G} : \mathcal{X} \longrightarrow \mathcal{X}$ is applied in the following form:

$$\dot{\mathbf{x}}(t) = [\mathbf{F}(\mathbf{x}(t)) + \varepsilon \mathbf{G}(\mathbf{x}(t))]\,\mathrm{d}t + \Sigma(\mathbf{x}(t))\mathrm{d}W_t, \tag{2.15}$$

where ε is a real parameter. The vector field $\mathbf{G}$ should be understood as the perturbation applied to the process and ε as the parameter regulating its strength. We, thus, have that when $\varepsilon = 0$ the process is unperturbed. The size of ε is not discussed just yet, although a perturbative approach to studying Eq. (2.15) would require infinitesimal values of ε in the spirit of linear response theory for equilibrium and non-equilibrium systems, see [Kub57]; [Kub66] and [Rue09] respectively. However, efforts in relating statistical response for finite perturbations and equilibrium fluctuations have to be noted, see [Bof+03]; [Luc08].

The Fokker-Planck equation associated with Eq. (2.15) is now a perturbed version of Eq. (2.12):

$$\partial_t \rho(\cdot, t) = \mathcal{L}\rho(\cdot, t) := -\nabla \cdot (\mathbf{F}\rho(\cdot, t) + \varepsilon \mathbf{G}\rho(\cdot, t)) + \frac{1}{2}\nabla^2 : \left(\Sigma\Sigma^\top \rho(\cdot, t)\right). \tag{2.16}$$

Additionally, the operator $\mathcal{L}$ is divided into two operators $\mathcal{L}_0 : D(\mathcal{L}_0) \subseteq \mathcal{B} \longrightarrow \mathcal{B}$ and $\mathcal{L}_1 : D(\mathcal{L}_1) \subseteq \mathcal{B} \longrightarrow \mathcal{B}$ so that $\mathcal{L} = \mathcal{L}_0 + \varepsilon\mathcal{L}_1$ and their domains are assumed to satisfy $D(\mathcal{L}) = D(\mathcal{L}_0) = D(\mathcal{L}_1)$. These operators are defined as:

$$\mathcal{L}_0 \rho = -\nabla \cdot (\mathbf{F}\rho) + \frac{1}{2}\nabla^2 : \left(\Sigma\Sigma^\top \rho\right); \tag{2.17a}$$

$$\mathcal{L}_1 \rho = -\nabla \cdot (\mathbf{G}\rho), \tag{2.17b}$$

for every ρ in $D(\mathcal{L})$. We immediately identify that $\mathcal{L}_1$ is the operator corresponding to the perturbation introduced in Eq. (2.15) and hence, we refer to it as the perturbation operator. The unperturbed counterpart, $\mathcal{L}_0$, is constituted by an advection due to the drift term and a diffusion component arising from the noise.

Because the perturbed generating operator is time-independent, we can make use of a wide range semigroup theoretic results to derive response formulas; we refer the reader to [EN00, Chapter 9.c] for a detailed and technical discussion on the topic. The main tool to be used is the Dyson expansion which expresses a perturbed semigroup as a power series of the unperturbed generator. Such perturbative expansions were originally introduced by Freeman J. Dyson in the context of quantum electrodynamics [Dys49] and later formulated rigorously in mathematical terms in [Gil17]. The Dyson expansion of the perturbed evolution operator is:

$$e^{t(\mathcal{L}_0 + \varepsilon\mathcal{L}_1)} = \sum_{k=0}^{\infty} P_k(t), \tag{2.18}$$

where $P_0(t) = e^{t\mathcal{L}_0}$ and

$$P_{k+1}(t) = \varepsilon \int_0^t e^{(t-s)\mathcal{L}_0} \mathcal{L}_1 P_k(s) \mathrm{d}s = \varepsilon \int_0^t e^{s\mathcal{L}_0} \mathcal{L}_1 P_k(t-s) \mathrm{d}s, \tag{2.19}$$

for $k \geq 0$. The series in Eq. (2.18) is formal since the operator $\mathcal{L}_1$ is unbounded and a direct norm-estimate cannot be taken to bound P_k. In case $\mathcal{L}_1$ is bounded, convergence of Eq. (2.18) is attained in the operator norm and operator integrals are understood in the strong operator topology; see [EN06] for more convergence and generation criteria. Here we shall, nevertheless, use Eq. (2.18) after evaluation of an initial probability density function $\rho(\cdot, 0)$ in $D(\mathcal{L})$ to find its analogous at time t, namely, $\rho(\cdot, t)$:

$$\rho(\cdot, t) = \sum_{k=0}^{\infty} P_k(t)\rho(\cdot, 0) = e^{t\mathcal{L}_0}\rho(\cdot, 0) + \varepsilon \int_0^t e^{(t-s)\mathcal{L}_0} \mathcal{L}_1 e^{s\mathcal{L}_0} \rho(\cdot, 0) \mathrm{d}s + \mathcal{O}\left(\varepsilon^2\right). \tag{2.20}$$

In the context of equilibrium classical systems governed by a perturbed Hamiltonian, expansions can directly be applied to the Gibbs measure and partition function of the system to obtain the leading order corrections to the statistics. Quantum Hamiltonians can be treated almost equally taking into account that the perturbed distribution function has to be expanded in the operator sense using Eq. (2.18) or resolvent expansions as done in [Zwa01].

Remark 2.2.1. *By virtue of the stability and regularity criteria imposed to the SDE in Eq.* (2.15)*, $\mathcal{L} = \mathcal{L}_0 + \varepsilon\mathcal{L}_1$ automatically generates a strongly continuous semigroup. One could, instead, ask whether $\mathcal{L}_0 + \varepsilon\mathcal{L}_1$ generates a strongly continuous semigroup regardless of the underlying dynamic process. Since $\mathcal{L}_1$ is unbounded, Dyson operator expansions like that in Eq.* (2.18) *cannot be rigorously applied, as already noted. However, it can be shown that the perturbation operator $\mathcal{L}_1$ is $\mathcal{L}_0$-bounded, i.e., there exist non-negative real numbers γ_1 and γ_2 so that $\|\mathcal{L}_1 f\| \leq \gamma_1 \|\mathcal{L}_0 f\| + \gamma_2 \|f\|$ for every f in $D(\mathcal{L})$. This condition is enough to ensure that the sum $\mathcal{L}_0 + \varepsilon\mathcal{L}_1$ generates a strongly continuous semigroup for small values of ε. We refer the reader to [EN06, Chapter III.2].*

Notice that each of the elements P_k of the perturbative series is written in terms of the semigroup generated by $\mathcal{L}_0$ uniquely. In particular, the leading order term involves a sole time integral, in the spirit of seminal works on linear response theory [Kub57]. Higher order terms involve multiple time integrals that result from the Dyson expansion and they agree with formulas derived elsewhere; see [Luc08]. The drawback is that such perturbative series works for infinitesimal values of ε or within a convergence interval. This shall be commented on later.

If one initialises Eq. (2.16) with the invariant density of the unperturbed system ρ_0, one would obtain its evolution towards the new stationary statistical state. Then, since

$\mathcal{L}_0\rho_0 \equiv 0$, we have:

$$\rho(\cdot,t) = e^{t(\mathcal{L}_0+\varepsilon\mathcal{L}_1)}\rho_0 = \sum_{k=0}^{\infty}\varepsilon^k\rho_k(\cdot,t) = \rho_0 + \varepsilon\int_0^t e^{s\mathcal{L}_0}\mathcal{L}_1\rho_0 \mathrm{d}s + \mathcal{O}\left(\varepsilon^2\right), \tag{2.21}$$

where we have introduced the successive corrections of the unperturbed invariant measure $\rho_k(\cdot,t)$ which give the solution at time t and are weighted by ε^k. In particular, $\rho_1(\cdot,t)$ denotes the leading order term and, hence, the formal derivative of $\rho(\cdot,t)$ with respect to ε. Notice that even if the perturbation operator is time-independent, treating it as a inhomogeneous component of the Fokker-Planck equation naturally introduces time-dependence via the integral convolutions. This dependence is only transient since the perturbation in Eq. (2.15) induces a new autonomous system and, if it is let to evolve for enough time, $\rho(\cdot,t)$ would arrive at a new stationary state ρ_ε which solves $(\mathcal{L}_0+\varepsilon\mathcal{L}_1)\rho_\varepsilon \equiv 0$. This way, letting t tend to infinity in Eq. (2.21) we get a perturbative formula for the perturbed invariant measure ρ_ε:

$$\rho_\varepsilon = \sum_{k=0}^{\infty}\varepsilon^k\rho_k = \rho_0 + \varepsilon\int_0^\infty e^{s\mathcal{L}_0}\mathcal{L}_1\rho_0 \mathrm{d}s + \mathcal{O}\left(\varepsilon^2\right), \tag{2.22}$$

where we have dropped the time-dependence of ρ_k. The leading order term $\int_0^\infty e^{s\mathcal{L}_0}\mathcal{L}_1\rho_0\mathrm{d}s$ is formally associated with the derivative of ρ_ε with respect to ε and measures the sensitivity of the unperturbed steady state.

With the application of the perturbation, expectation values of observables change accordingly becoming time-dependent before they reach a new steady value. Let Ψ denote an observable function. Then, its expectation value at time t satisfies the following chain of equalities:

$$\langle\Psi,\rho(\cdot,t)\rangle := \int\Psi(\mathbf{x})\rho(\mathbf{x},t)\mathrm{d}\mathbf{x} = \sum_{k=0}^{\infty}\varepsilon^k\delta^{(k)}\left[\Psi\right](t) \tag{2.23a}$$

$$= \langle\Psi,\rho_0\rangle + \varepsilon\langle\Psi,\rho_1(\cdot,t)\rangle + \mathcal{O}\left(\varepsilon^2\right) \tag{2.23b}$$

$$= \int\Psi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} + \varepsilon\int\Psi(\mathbf{x})\int_0^t e^{s\mathcal{L}_0}\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}s\mathrm{d}\mathbf{x} + \mathcal{O}\left(\varepsilon^2\right) \tag{2.23c}$$

$$= \int\Psi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} + \varepsilon\int_0^t\int e^{s\mathcal{L}_0^*}\Psi(\mathbf{x})\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}\mathrm{d}s + \mathcal{O}\left(\varepsilon^2\right), \tag{2.23d}$$

where the successive correction terms $\delta^{(k)}[\Psi](t)$ denote the kth order response and, particularly, $\delta^{(1)}[\Psi](t)$ is the linear response. In Eq. (2.23d) we have used the generator of the backward-Kolmogorov equation $\mathcal{L}^*$ in Eq. (2.8), so that the expectation value of an observable function Ψ evolves according to:

$$\partial_t\Psi(\cdot,t) = \mathcal{L}^*\Psi(\cdot,t) = \left(\mathcal{L}_0^* + \varepsilon\mathcal{L}_1^*\right)\Psi(\cdot,t), \tag{2.24}$$

and the operators $\mathcal{L}_0^*$ and $\mathcal{L}_1^*$ are

$$\mathcal{L}_0^*\Psi = \mathbf{F}\cdot\nabla\Psi + \frac{1}{2}\Sigma\Sigma^\top : \nabla^2\Psi; \tag{2.25a}$$

$$\mathcal{L}_1^*\Psi = \mathbf{G}\cdot\nabla\Psi, \tag{2.25b}$$

in accordance with the discussion in Section 2.1.1. Moreover, if the perturbed statistical steady state ρ_ε is attained, the functions $\delta^{(k)}[\Psi]$ become independent for every $k \geq 0$. More concretely, we can deduce the expectation value of Ψ in the perturbed regime $\langle\Psi, \rho_\varepsilon\rangle$ in a perturbative fashion by letting t go to infinity in Eq. (2.23c):

$$\langle\Psi, \rho_\varepsilon\rangle = \sum_{k=0}^{\infty}\delta^{(k)}[\Psi] = \langle\Psi, \rho_0\rangle + \varepsilon\int_0^\infty\int\Psi(\mathbf{x})e^{s\mathcal{L}_0}\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}\mathrm{d}s + \mathcal{O}\left(\varepsilon^2\right), \tag{2.26}$$

where the kth order response $\delta^{(k)}[\Psi]$ is now time-independent. On a practical note, since the invariant measure is often inaccessible, changes in a system has to be monitored through suitable observables. In this sense, the scalar quantity $\int_0^\infty\int\Psi(\mathbf{x})e^{s\mathcal{L}_0}\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}\mathrm{d}s$ would be the formal derivative of the response map (1.8) and would equate to Eq. (1.9).

Remark 2.2.2. *The existence of a density function ρ_0 that accounts for the statistical steady state allows to characterise the linear response $\delta^{(1)}[\Psi](t)$ in Eq.* (2.23) *in terms of the unforced fluctuations of the system which take the shape of correlation functions. This is the content of the fluctuation dissipation theorem [Kub66]. Indeed, one has:*

$$\delta^{(1)}[\Psi](t) = \int_0^t\int e^{s\mathcal{L}_0}\Psi(\mathbf{x})\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}\mathrm{d}s \tag{2.27a}$$

$$= \int_0^t\int e^{s\mathcal{L}_0}\Psi(\mathbf{x})\frac{\mathcal{L}_1\rho_0(\mathbf{x})}{\rho_0(\mathbf{x})}\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}\mathrm{d}s = \int_0^t C_{\Psi,\Phi}(s)\mathrm{d}s, \tag{2.27b}$$

where $\Phi = \mathcal{L}_1\rho_0/\rho_0$ and $C_{\Psi,\Phi}(t)$ denotes the correlation function between Ψ and Φ at time t, see Eq. (2.14)*. This result also holds for deterministic systems with an absolutely continuous invariant measure.*

As mentioned earlier, the time-dependence of the linear response $\delta^{(1)}[\Psi](t)$ comes from the Dyson formula applied to the perturbation operator $\varepsilon\mathcal{L}_1$ before the perturbed system attains a steady state. Therefore, one can ask what the response looks like in the spectral domain. To this end, linear response function is characterised spectrally in terms of the generator $\mathcal{L}_0$, which is related by to the evolution operator $e^{t\mathcal{L}_0}$ via the resolvent operator and the Laplace transform [EN00]:

$$R(z, \mathcal{L}_0)f := (z - \mathcal{L}_0)^{-1}f = \mathfrak{L}[\mathcal{L}_0 f](z) = \int_0^\infty e^{-zt}e^{t\mathcal{L}_0}f\mathrm{d}t, \tag{2.28}$$

for every z in $\mathbb{C}$ such that $\mathfrak{Re}z > 0$. Recall that the resolvent operator is bounded

and bijective for every z in the resolvent set $\mathbb{C} \setminus \sigma(\mathcal{L}_0)$. When a perturbation like that in Eq. (2.16) is introduced, the resolvent operator of the perturbed generator can be expressed in terms of the unperturbed one [EN06]:

$$R(z, \mathcal{L}_0 + \varepsilon\mathcal{L}_1) = R(z, \mathcal{L}_0)\left(1 - \varepsilon\mathcal{L}_1 R(z, \mathcal{L}_0)\right)^{-1} = R(z, \mathcal{L}_0)\sum_{k=0}^{\infty} \varepsilon^k \left(\mathcal{L}_1 R(z, \mathcal{L}_0)\right)^k, \tag{2.29}$$

for every z in $\mathbb{C}$ such that $\mathfrak{Re}z > 0$. Notice that convergence is here attained in the operator norm since $R(z, \mathcal{L}_0)$ and $\mathcal{L}_1 R(z, \mathcal{L}_0)$ are bounded. This way one would have to ensure that $\varepsilon\|\mathcal{L}_1 R(z, \mathcal{L}_0)\| < 1$ in order for the expansion Eq. (2.28) to converge, introducing a first explicit constraint to the the size of ε. Taking the Laplace transform of the full response based on Eq. (2.23) we obtain:

$$\mathfrak{L}\left[\langle \Psi, \rho(\cdot, t)\rangle - \langle \Psi, \rho_0\rangle\right](z) = \varepsilon \int R(z, \mathcal{L}_0^*)\mathcal{L}_1^* R(z, \mathcal{L}_0^*)\Psi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} + \mathcal{O}\left(\varepsilon^2\right), \tag{2.30}$$

where we have used that $R(z, \mathcal{L}^*) = R(z, \mathcal{L})^*$ for any operator $\mathcal{L}$, because we are operating on an observable rather than a density function. Truncating at second order in powers of ε, the formula for linear response in complex frequency space is derived:

$$\mathfrak{L}\left[\delta^{(1)}\left[\Psi\right](t)\right](z) = \int R(z, \mathcal{L}_0^*)\mathcal{L}_1^* R(z, \mathcal{L}_0^*)\Psi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} \tag{2.31a}$$

$$= \int \mathcal{L}_1^* R(z, \mathcal{L}_0^*)\Psi(\mathbf{x}) R(z, \mathcal{L}_0)\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} \tag{2.31b}$$

$$= \frac{1}{z}\int \mathcal{L}_1^* R(z, \mathcal{L}_0^*)\Psi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{2.31c}$$

where we have used the fact that ρ_0 is an eigenfunction of $\mathcal{L}_0$ relative to 0 to conclude that $R(z, \mathcal{L}_0)\rho_0 = \rho_0/z$. The introduction of a pole at $z = 0$ is a result of employing the Laplace transform and applying a static external field. As we shall see later, such pole at $z = 0$ will disappear when time-dependence is specified and physically restricted, in accordance with general results in linear response theory [Rue09].

If Ψ and Φ denote two smooth observables with zero ρ_0-mean, its correlation function at time t was defined in Eq. (2.14). Taking the Laplace transform of Eq. (2.14) we obtain:

$$\mathfrak{L}[C_{\Psi,\Phi}](z) = \int_0^\infty e^{-zt}\int \Psi(\mathbf{x})e^{t\mathcal{L}_0}\Phi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}\mathrm{d}t \tag{2.32a}$$

$$= \int \Psi(\mathbf{x})\int_0^\infty e^{-zt}e^{t\mathcal{L}_0}\Phi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}t\mathrm{d}\mathbf{x} \tag{2.32b}$$

$$= \int \Psi(\mathbf{x})R(z, \mathcal{L}_0)\Phi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{2.32c}$$

which is valid for $\mathfrak{Re}z > 0$. Although the resolvent is well defined and bounded for every

$\mathfrak{Re}z > 0$, peaks in $\mathfrak{L}[C_{\Psi,\Phi}]$ obtained from observed data would result from a meromorphic extension onto the complex plane where the poles of $\mathfrak{L}[C_{\Psi,\Phi}]$ correspond to the eigenvalues (point spectrum) of $\mathcal{L}_0$, also known as the Ruelle-Pollicot resonances [Rue86]; [Rue09]; [Tan+20]. This topic will be developed later in the text where we explicitly invoke a spectral decomposition of $\mathcal{L}_0$.

The effects of perturbations on correlation functions are deduced by examining $C_{\Psi,\Phi}$ in Eq. (2.14) where now, the exponential operator involves the perturbation operator $\mathcal{L}_1$ and the invariant measure ρ_0 is substituted by ρ_ε, which is obtained from Eq. (2.22). The leading order response of correlation functions involves two terms, where only one of them can be recast in correlation form [LW17]. Indeed, denoting the perturbed correlation function as $C^\varepsilon_{\Psi,\Phi}$ and applying the Dyson expansion one gets:

$$C^\varepsilon_{\Psi,\Phi}(t) = \int_0^\infty \int e^{(t+s)\mathcal{L}_0^*}\Psi(\mathbf{x})e^{s\mathcal{L}_0^*}\Phi(\mathbf{x})\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}\mathrm{d}s \tag{2.33a}$$

$$+ \int\int_0^t e^{(t-s)\mathcal{L}_0^*}\mathcal{L}_1^* e^{s\mathcal{L}_0^*}\Psi(\mathbf{x})\Phi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}s\mathrm{d}\mathbf{x} + \mathcal{O}\left(\varepsilon^2\right), \tag{2.33b}$$

in agreement with [LW17, Eqs. (11)-(14)]. While the integrand in the RHS of Eq. (2.33a) takes the form of a correlation function with respect to $\mathcal{L}_1\rho_0$, Eq. (2.33b) cannot be rewritten in correlation form. Applying the Fourier transform to $C^\varepsilon_{\Psi,\Phi}(t)$ would yield the response in spectral domain, again following [LW17]. We can, however, exploit the resolvent expansion Eq. (2.28) to obtain the perturbed correlation function in terms of the unperturbed resolvent operator. By definition we have:

$$\mathfrak{L}[C^\varepsilon_{\Psi,\Phi}](z) = \int \Psi(\mathbf{x})R(z,\mathcal{L}_0+\varepsilon\mathcal{L}_1)\Phi(\mathbf{x})\rho_\varepsilon(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{2.34}$$

for $\mathfrak{Re}z > 0$. Expanding $\mathfrak{L}[C^\varepsilon_{\Psi,\Phi}](z)$ in powers of ε we get:

$$\begin{aligned}\mathfrak{L}[C^\varepsilon_{\Psi,\Phi}](z) = &\int \Psi(\mathbf{x})R(z,\mathcal{L}_0)\Phi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}\\ &+\varepsilon\int \Psi(\mathbf{x})R(z,\mathcal{L}_0)\Phi(\mathbf{x})\rho_1(\mathbf{x})\mathrm{d}\mathbf{x}\\ &+\varepsilon\int \Psi(\mathbf{x})R(z,\mathcal{L}_0)\mathcal{L}_1R(z,\mathcal{L}_0)\Phi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} + \mathcal{O}\left(\varepsilon^2\right).\end{aligned} \tag{2.35}$$

Hence, we can formally write

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\varepsilon}\mathfrak{L}[C^\varepsilon_{\Psi,\Phi}](z)|_{\varepsilon=0} = &\int \Psi(\mathbf{x})R(z,\mathcal{L}_0)\Phi(\mathbf{x})\rho_1(\mathbf{x})\mathrm{d}\mathbf{x}\\ &+\int \Psi(\mathbf{x})R(z,\mathcal{L}_0)\mathcal{L}_1R(z,\mathcal{L}_0)\Phi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}.\end{aligned} \tag{2.36}$$

This formula can be obtained by taking the Laplace transform of Eq. (2.33), although the resolvent formalism makes the calculations substantially easier. We, thus, provide an

explicit link between the response in frequency domain and the spectral properties of the unperturbed generator $\mathcal{L}_0$ via the resolvent.

Remark 2.2.3. *The use of complex frequencies by means of the Laplace transform instead of the Fourier domain comes out of convenience and accordance with the semigroup theory literature. The connection is, nevertheless, straightforward by (formally) noticing that, if ω denotes a real frequency, the one-sided Fourier transform of the correlation function reads as:*

$$\mathfrak{F}[C_{\Psi,\Phi}](\omega) = \int_0^\infty e^{-i\omega t} C_{\Psi,\Phi}(t)\mathrm{d}t = \int \Psi(\mathbf{x}) R(i\omega, \mathcal{L}_0)\Phi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{2.37}$$

The discussion in Fourier space will expanded in the next section when time dependent perturbations are investigated. We refer the reader also to [BL07].

2.2.1 Spectral Decomposition of the Response Function

The decomposition of correlation functions and power spectral densities in terms of the functional features of Liouville operators, like $\mathcal{L}_0$ in Eq. (2.16), has been crucial to understand why they decay in deterministic chaotic systems and to identify the conditions under which such decay is exponential or subexponential [Liv95]. The eigenvalues responsible for governing the system's decay of correlations are known as Ruelle-Pollicot resonances [Pol85]; [Rue86], and the barrier that prevents readily calculating such spectral features stands in the choice of Banach space where the Liouville operators are defined [GL06], which need to take into account the complex and possibly fractal geometries of phase space that support singular invariant measures. In fact, the natural L^p spaces cannot capture eigenvalues of $\mathcal{L}_0$ that explain the decay of correlations; we refer the reader back to Theorem 2.1.1. Contrarily, Fokker-Planck equations do reveal non-trivial resonances in L^p domains so that the natural assumptions on Eq. (2.15) allow to impose finer, yet useful structures to $\mathcal{L}_0$, as done below and in previous works [Tan+20]. This, together with the fluctuation-dissipation dissipation theorem of Remark 2.2.2 is suggestive that the Ruelle-Pollicot resonances can also be used to decompose the linear response formulas derived in Eq. (2.26). In this section, we shall investigate the linear response $\delta^1[\Psi](t)$ in terms of the spectrum of the generator of the Fokker-Planck semigroup.

In what follows, the setting of Eq. (2.15) stays practically the same, where it is assumed that its associated strongly continuous semigroup $\{e^{t(\mathcal{L}_0+\varepsilon\mathcal{L}_1)}\}_{t\geq 0}$ generated by the Fokker-Planck equation satisfies the conditions that guarantee the existence of a unique smooth invariant measure [Pav14, Chapter 6]. To write the response in terms of the unperturbed evolution, we shall further require the semigroup $\{e^{t\mathcal{L}_0}\}_{t\geq 0}$ to be *quasi-compact*, this is, it approaches the space of compact operators as t tends to infinity in the operator norm [EN00]. Moreover, quasi-compact semigroups are constituted by operators having an

essential spectral radius strictly less than one, implying that spectra with larger modulus are eigenvalues of finite algebraic multiplicity, which are related to those of the generator by means of the Spectral Mapping Theorem [EN00, Chapter IV]. Namely, if λ_j is an eigenvalue of $\mathcal{L}_0$ with eigenfunction ψ_j, we have that:

$$\mathcal{L}_0\psi_j = \lambda_j\psi_j \iff e^{t\mathcal{L}_0}\psi_j = e^{\lambda_j t}\psi_j, \tag{2.38}$$

making $e^{\lambda_j t}$ and eigenvalue of $e^{t\mathcal{L}_0}$ for every t. This way, the spectrum of $\mathcal{L}_0$ is divided into a dominating point spectrum (eigenvalues) and an essential background with a boundary determined by the essential growth bound ω_{ess}:

$$\omega_{ess} = \inf_{t>0} \frac{1}{t} \log \left\| e^{t\mathcal{L}_0} \right\|_{ess}, \tag{2.39}$$

where $\| \cdot \|_{ess}$ is the essential norm[1]. Therefore, if $\{\lambda_j\}_{j=1}^{M}$ are eigenvalues of finite algebraic multiplicity such that $\lambda_j > \omega_{ess}$ the operator $e^{t\mathcal{L}_0}$ is decomposed as:

$$e^{t\mathcal{L}_0} = \sum_{j=1}^{M} T_j(t) + \mathcal{R}(t), \tag{2.40}$$

where $T_j(t)$ denotes the contribution relative to the eigenvalue λ_j and $\mathcal{R}$ is the operator accounting for the essential spectrum. The operator $T_j(t)$ is defined as:

$$T_j(t) = \sum_{k=0}^{a_j-1} e^{\lambda_j t} \frac{t^k}{k!} \left(\mathcal{L}_0 - \lambda_j\right)^k \Pi_j, \tag{2.41}$$

where Π_j denotes the spectral projector around the eigenvalue λ_j whose (finite) multiplicity is counted by a_j. For notational convenience, we shall further assume that $\lambda_0 = 0 > \mathfrak{Re}\lambda_1 \geq \mathfrak{Re}\lambda_2 \geq \ldots$. The hope is that, as t tends to infinity the contributions from the essential spectrum decay. In fact, if $\omega > \sup\{\omega_{ess}\} \cup \left\{\mathfrak{Re}\lambda : \lambda \in \sigma(\mathcal{L}_0) \setminus \{\lambda_0, \ldots, \lambda_M\}\right\}$ there exists a constant $C > 0$ such that:

$$\|\mathcal{R}(t)\| \leq Ce^{-\omega t}, \tag{2.42}$$

for all positive values of t. This ensures that, as time goes to infinity, the norm of the residual operator decays to zero. The finiteness of the eigenvalues assumed here implies that all of them, but for $\lambda_0 = 0$, are contained in a strip of the complex plane that is at a distance $\gamma = -\mathfrak{Re}\lambda_1$ from the imaginary axis. Such eigenvalues translate as meromorphic poles

[1]The essential norm of an operator T on a Banach space $\mathcal{B}$ is:

$$\|T\|_{ess} = \inf \{\|T - K\| : K \text{ is compact in } \mathcal{B}\}.$$

of the resolvent operator $R(\cdot, \mathcal{L}_0)$ and correspond to the aforementioned Ruelle-Pollicot resonances, albeit in a stochastic context [Tan+20]. The number γ is called the *spectral gap* of $\mathcal{L}_0$ and it guarantees that correlations decay exponentially fast [Tan+20]; [Bal00, §1.3]. More importantly, this quantity gauges the sensitivity of the system with respect to external perturbations [KL99]. Namely, if $0 < \gamma \ll 1$ the linear response function $\delta^{(1)}[\Psi](t)$ will decay to a steady state very slowly and, consequently, its static version (2.26) will be larger. It must be noted, though, that the point spectrum need not be finite, so that it is possible that it accumulates around the leading zero eigenvalue. This pathological case is responsible for subexponential decay rates [Rue86] and a signal of the system approaching a critical point, as is the case of the pitchfork bifurcation [Gas+95].

Remark 2.2.4. *If the spectrum of $\mathcal{L}_0$ consists only of simple eigenvalues, this is $a_j = 1$ for all values of j, $T_j(t)$ has a more explicit expression. Let us suppose that f is in $L^2_\mu(\mathcal{X})$ and projects entirely onto the point spectrum, then*

$$e^{t\mathcal{L}_0} f = \sum_{j=0}^{M} T_j(t) f = \sum_{j=0}^{M} e^{\lambda_j t} \Pi_j f = \sum_{j=0}^{M} e^{\lambda_j t} \langle f, \psi_j^* \rangle \psi_j, \tag{2.43}$$

where $\langle \cdot, \cdot \rangle$ is the L^2-inner product with respect to ρ_0. Here, we have exploited the simplicity of the spectrum and the L^2-dual representation of the eigenfunctions ψ_j^*.

When the observables of interest decompose suitably according to the eigenfunctions of $\mathcal{L}_0^*$, the dominant modes of variability encoded in its leading eigenvalues are susceptible of explaining the decay of the response function at infinity. In order to see this, we shall decompose the linear response function $\delta^{(1)}[\Psi](t)$ in terms of the point spectrum of the generator $\mathcal{L}_0^*$. For that, let us suppose that Ψ projects entirely onto the span of all eigenfunctions so that:

$$e^{t\mathcal{L}_0^*} \Psi = \sum_{j=1}^{M} T_j^*(t) \Psi + \mathcal{R}^*(t) \Psi = \sum_{j=1}^{M} e^{\lambda_j t} \Pi_j^* \Psi, \tag{2.44}$$

where the action of the residual operator $\mathcal{R}^*$ vanishes by assumption. The same equality would hold if densities are considered instead, bearing in mind to take the dual represen-

tation. Now, we expand the linear response of the observable Ψ:

$$\delta^{(1)}\left[\Psi\right](t) = \int_0^t \int \Psi(\mathbf{x}) e^{s\mathcal{L}_0} \mathcal{L}_1 \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x}\mathrm{d}s = \int_0^t \int e^{s\mathcal{L}_0^*} \Psi(\mathbf{x}) \mathcal{L}_1 \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x}\mathrm{d}s \tag{2.45a}$$

$$= \int_0^t \int \left(\sum_{j=0}^{M} T_j^*(s) \Psi(\mathbf{x}) \right) \mathcal{L}_1 \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x}\mathrm{d}s \tag{2.45b}$$

$$= \int_0^t \int \left(\sum_{j=0}^{M} \sum_{k=0}^{a_j-1} e^{\lambda_j s} \frac{s^k}{k!} \left(\mathcal{L}_0^* - \lambda_j\right)^k \Pi_j^* \Psi(\mathbf{x}) \right) \mathcal{L}_1 \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x}\mathrm{d}s \tag{2.45c}$$

$$= \sum_{j=0}^{M} \int_0^t e^{\lambda_j s} \sum_{k=0}^{a_j-1} \frac{s^k}{k!} \int \left(\mathcal{L}_0^* - \lambda_j\right)^k \Pi_j^* \Psi(\mathbf{x}) \mathcal{L}_1 \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x}\mathrm{d}s \tag{2.45d}$$

$$= \sum_{j=0}^{M} \int_0^t e^{\lambda_j s} \sum_{k=0}^{a_j-1} \alpha_j^{(k)} \frac{s^k}{k!} \mathrm{d}s \tag{2.45e}$$

$$= \sum_{j=0}^{M} \sum_{k=0}^{a_j-1} \sum_{l=0}^{k} c_{j,k,l} e^{\lambda_j t} t^l - \sum_{j=0}^{M} \sum_{k=1}^{a_j-1} c_{j,k,0}, \tag{2.45f}$$

where $\alpha_j^{(k)} = \int (\mathcal{L}_0^* - \lambda_j)^k \Pi_j^* \Psi(\mathbf{x}) \mathcal{L}_1 \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x}$ and $c_{j,k,l} = (-1)^{k-l} (l!)^{-1} \alpha_j^{(k)} \lambda_j^{-(k-l+1)}$. Since the semigroup $\{e^{t\mathcal{L}_0}\}_{t\geq 0}$ is quasi-compact, it follows that the eigenvalues on the imaginary axis are simple [EN06, Chapter V. §4] and one can exploit Remark 2.2.4 to express the spectral projector around $\lambda_0 = 0$ in easier terms to find:

$$\alpha_0 = \int \Pi_j^* \Psi(\mathbf{x}) \mathcal{L}_1 \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x} = \int \langle \psi_0, \Psi \rangle \, \psi_0^*(\mathbf{x}) \mathcal{L}_1 \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x} \tag{2.46a}$$

$$= \int \langle \rho_0, \Psi \rangle \, \mathbb{1}(\mathbf{x}) \mathcal{L}_1 \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x} = \langle \rho_0, \Psi \rangle \int \mathcal{L}_1^* \mathbb{1}(\mathbf{x}) \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x} = 0, \tag{2.46b}$$

where we have used the fact that ρ_0 and the constant function $\mathbb{1}$ are the eigenfunctions of $\mathcal{L}_0$ and $\mathcal{L}_0^*$ relative to $\lambda_0 = 0$, respectively. This implies that the indexes j in Eq. (2.45) range from 1 to M. Taking Laplace transforms of Eq. (2.45) for complex frequencies z such that $\mathfrak{Re} z > 0$ we have:

$$\mathfrak{L}\left[\delta^{(1)}\left[\Psi\right]\right](z) = \sum_{j=1}^{M} \sum_{k=0}^{a_j-1} \sum_{l=0}^{k} \int_0^\infty c_{j,k,l} e^{-zt} e^{\lambda_j t} t^l \mathrm{d}t - \sum_{j=1}^{M} \sum_{k=0}^{a_j-1} c_{j,k,0} \int_0^\infty e^{-zt} \mathrm{d}t \tag{2.47a}$$

$$= \sum_{j=1}^{M} \sum_{k=0}^{a_j-1} \sum_{l=0}^{k} c_{j,k,l} \frac{l!}{(z-\lambda_j)^{l+1}} - \frac{1}{z} \sum_{j=1}^{M} \sum_{k=0}^{a_j-1} c_{j,k,0}. \tag{2.47b}$$

Notice that a meromorphic extension onto the complex plane would reveal poles located at the eigenvalues λ_j plus an extra resonance at $z = 0$ that results from the application of the Laplace transform of an integral, which can be expressed as a convolution with the Heaviside distribution. The same formula can be obtained by using the resolvent approach and realising that $\mathcal{L}_0$ and $R(z, \mathcal{L}_0)$ share the same eigenfunctions. In this case, for the

fully simple point spectrum we have:

$$R(z,\mathcal{L}_0)\psi_j = (z-\lambda_j)^{-1}\psi_j, \tag{2.48}$$

for all $j = 1,\dots,M$. Hence, taking the dual and substituting into Eq. (2.31c) we get:

$$\mathfrak{L}\left[\delta^{(1)}\left[\Psi\right]\right](z) = \frac{1}{z}\int \mathcal{L}_1^* R(z,\mathcal{L}_0^*)\Psi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} \tag{2.49a}$$

$$= \frac{1}{z}\int \mathcal{L}_1^*\left(\sum_{j=1}^{M}(z-\lambda_j)^{-1}\Pi_j^*\Psi(\mathbf{x})\right)\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} \tag{2.49b}$$

$$= \sum_{j=1}^{M}\frac{\alpha_j}{z(z-\lambda_j)} = \sum_{j=1}^{M}\frac{\alpha_j}{\lambda_j(z-\lambda_j)} - \sum_{j=1}^{M}\frac{\alpha_j}{\lambda_j z}, \tag{2.49c}$$

which is the same as Eq. (2.47b) for $k = l = 0$.

Remark 2.2.5. *Following Remark 2.2.2, the spectral decomposition of the transfer/Koopman operator can be applied to the correlation function $C_{\Psi,\Phi}(t)$ to get [Tan+20]:*

$$C_{\Psi,\Phi}(t) = \sum_{j=0}^{M}\sum_{k=0}^{a_j-1}\frac{t^k}{k!}e^{\lambda_j t}\beta_j^{(k)}\psi_j + \int \mathcal{R}(t)\Psi(\mathbf{x})\Phi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{2.50}$$

where the coefficients $\beta_j^{(k)}$ as opposed to $\alpha_j^{(k)}$ are now dependent on both observable functions Ψ and Φ. Notice that the second term in the RHS of Eq. (2.50) *is the contribution coming from the essential spectrum which, in the case Ψ and Φ project entirely onto the point spectrum, would vanish.*

2.3 Time-dependent Forcing

The SDE (2.15) includes a perturbation vector field that does not vary in time, making the forced system also autonomous with an associated invariant measure. Based on this setting, classical results in the perturbation theory of semigroups were used to derive the basic linear response formulas. This setting is no longer valid in case the external forcing is modulated in time, since Dyson-like operator expansions (2.18) cannot be invoked and the perturbative expansion of the resolvent operator (2.29) is not available. The aim of this section is to show that one can, nevertheless, derive linear response formulas for external perturbations with physically constrained time-dependence using the operator theoretic approach. Hence, dynamic response, as opposed to static, aims at (i) determining the leading order time-dependent corrections to the statistics as results of a perturbation and (ii) disentangling the time-modulation from the calculation of the response via the Green function.

In this section, we consider the following time-dependent version of Eq. (2.15):

$$\mathrm{d}\mathbf{x}(t) = [\mathbf{F}(\mathbf{x}) + \varepsilon g(t)\mathbf{G}(\mathbf{x})]\,\mathrm{d}t + \Sigma(\mathbf{x})\mathrm{d}W_t, \tag{2.51}$$

which generates a process $\mathbf{x}(t)$ in $\mathbb{R}^d$ for an initial condition at a certain value of time, which is autonomous for $\varepsilon = 0$. When $\varepsilon \neq 0$, a time-modulated forcing vector field $\mathbf{G}$ is activated through the bounded function g. The associated Fokker-Planck equation is:

$$\partial_t \rho = (\mathcal{L}_0 + \varepsilon g(t)\mathcal{L}_1)\,\rho = -\nabla\cdot(\mathbf{F}\rho) - \varepsilon g(t)\nabla\cdot(\mathbf{G}\rho) + \frac{1}{2}\nabla^2 : \left(\Sigma\Sigma^\top\rho\right), \tag{2.52}$$

where the generating operators $\mathcal{L}_0$ and $\mathcal{L}_1$ are defined as done previously and when $\varepsilon = 0$ we assume, as in the previous section, that there is an invariant density function ρ_0 that solves $\mathcal{L}_0\rho_0 \equiv 0$. Because the SDE is non-autonomous, the density functions ρ are dependent on the phase variable $\mathbf{x}$ at time t and the initial condition $\mathbf{x}_0$ at time s. If one, instead, assumes that the initial condition $\mathbf{x}_0$ at time $s = 0$ is distributed according to the unperturbed state ρ_0 ($\varepsilon = 0$), we can understand the densities in Eq. (2.52) as

$$\rho(\mathbf{x}, t) := \int \rho(\mathbf{x}, \mathbf{y}, t, 0)\rho_0(\mathbf{y})\mathrm{d}\mathbf{y}, \tag{2.53}$$

so that $\rho(\mathbf{x}, t)$ determines the probability of encountering the process in $\mathbf{x}$ at time t for a ρ_0-distributed initial condition; see also [Pav14, Chapter 2] for more general considerations. The goal now is to calculate the distribution function for the perturbed system i.e., when $\varepsilon \neq 0$. To this end, we shall use solve Eq. (2.52) for successive orders of ε. Let us assume that a solution of Eq. (2.52) ρ can be written as:

$$\rho = \rho_0 + \varepsilon\rho_1 + \varepsilon^2\rho_2 + \ldots \tag{2.54}$$

While ρ_0 is provided, ρ_1 is obtained by plugging Eq. (2.54) into Eq. (2.52), gathering the ε-order terms and applying the variation-of-parameters formula:

$$\rho_1(\cdot, t) = \int_0^t e^{(t-s)\mathcal{L}_0} g(s)\mathcal{L}_1\rho_0\mathrm{d}s, \tag{2.55}$$

which equals the static version obtained in Eq. (2.21) where now the time-modulation g appears in the integrand. The kth order element of Eq. (2.54) is obtained analogously by gathering ε^k-terms. Let Ψ be an observable, then the change of its expectation value with respect to time is:

$$\langle\Psi, \rho(\cdot, t)\rangle = \sum_{k=0}^{\infty} \delta^{(k)}\left[\Psi\right](t) = \int \Psi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} + \int \Psi(\mathbf{x})\rho_1(\mathbf{x}, t)\mathrm{d}\mathbf{x} + \mathcal{O}\left(\varepsilon^2\right). \tag{2.56}$$

Hence, the linear response accounting for the first order corrections is:

$$\delta^{(1)}[\Psi](t) = \int_0^t g(s) \int \Psi(\mathbf{x}) e^{(t-s)\mathcal{L}_0} \mathcal{L}_1 \rho_0 \mathrm{d}\mathbf{x}\mathrm{d}s. \tag{2.57}$$

The procedure for deriving Eq. (2.57) can be applied to the static case to deduce Eq. (2.23). However, in the dynamic scenario, we shall impose the physical constraint of *causality* [Rue09]; [Luc18], by which at negative times the response is zero. This amounts to having that the function $t \mapsto \int \Psi(\mathbf{x}) e^{t\mathcal{L}_0} \mathcal{L}_1 \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x}$ is zero for $t < 0$. Furthermore, if such function decays sufficiently quickly for large times, we can extend the integration limits to all the real line so that the response formula is written in convolution form without compromising convergence. In particular, the linear response operator would take the role of a *Green function* $\mathcal{G}$ in the following form [Luc08]:

$$\delta^{(1)}[\Psi](t) = (g * \mathcal{G})(t) = \int_{-\infty}^{\infty} g(s) \int \Psi(\mathbf{x}) e^{(t-s)\mathcal{L}_0} \mathcal{L}_1 \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x}\mathrm{d}s. \tag{2.58}$$

This implies that once the Green function is known, the linear response can be readily calculated regardless of the time-modulation g. Higher order terms in Eq. (2.56) correspond to the non-linear response and are obtained following [Kub57]; [Luc08] by gathering the contributions at a given power of ε and can be formulated in terms of kth order Green functions $\mathcal{G}^{(k)}$:

$$\mathcal{G}^{(k)}(\tau_1, \ldots, \tau_k) = \int \Theta(\tau_1) \ldots \Theta(\tau_k - \tau_{k-1}) \mathcal{L}_1 e^{(\tau_k - \tau_{k-1})\mathcal{L}_0^*} \ldots \mathcal{L}_1 e^{\tau_1 \mathcal{L}_0^*} \Psi(\mathbf{x}) \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{2.59}$$

Then, the kth order response of the observable Ψ at time t is given by:

$$\delta^{(k)}[\Psi](t) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \mathcal{G}^{(k)}(\tau_1, \ldots, \tau_k) g(t - \tau_1) \ldots g(t - \tau_k) \mathrm{d}\tau_1 \ldots \mathrm{d}\tau_k. \tag{2.60}$$

Furthermore, the convolution structure allows to extend g to general distributions [Rud91]. Indeed, one can consider an impulse at initial time by taking $g(t) = \delta(t)$ to get

$$\delta^{(1)}[\Psi](t) = \int_{-\infty}^{\infty} \delta(s) \mathcal{G}(t - s) \mathrm{d}s = \mathcal{G}(t) = \int \Psi(\mathbf{x}) e^{t\mathcal{L}_0} \mathcal{L}_1 \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{2.61}$$

Which says that, in the limit of small forcing strength ε, the observed response is exactly the Green function. Gathering higher order terms according to Eq. (2.60) we inductively find:

$$\langle \Psi, \rho(\cdot, t) \rangle = \sum_{k=0}^{\infty} \mathcal{G}^{(k)}(t, \ldots, t) = \sum_{k=0}^{\infty} \varepsilon^k \int \Psi(\mathbf{x}) e^{s\mathcal{L}_0} \mathcal{L}_1^k \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x}. \tag{2.62}$$

If a constant forcing is applied at $t = 0$, then g takes form of the Heaviside distribution

$\Theta(t)$, for which the response reads as:

$$\delta^{(1)}[\Psi](t) = \int_{-\infty}^{\infty} \Theta(s)\mathcal{G}(t-s)\mathrm{d}s = \int_0^t \int \Psi(\mathbf{x})e^{(t-s)\mathcal{L}_0}\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}\mathrm{d}s. \tag{2.63}$$

Notice that this formula is equal to that obtained using expansions in the static version; see Eq. (2.23d). Higher order terms are more convoluted that the impulse perturbations, so they shall be omitted.

The spectral decomposition introduced earlier in Section 2.2.1 for quasi-compact operators allows us to naturally link the theory of the Green function with the eigenvalues of the generator of the unperturbed Fokker-Planck semigroup. Indeed, we assume that our observable of interest Ψ projects entirely onto the point spectrum of $\mathcal{L}_0$ in the same way as we assumed in Eq. (2.45). This gives the following series of equalities:

$$\begin{aligned}
\delta^{(1)}\left[\Psi\right](t) &= \int_{-\infty}^{\infty} g(t-s)\mathcal{G}(s)\mathrm{d}s && \text{(2.64a)}\\
&= \int_{-\infty}^{\infty} g(t-s)\int \Theta(s)e^{s\mathcal{L}_0^*}\Psi(\mathbf{x})\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}\mathrm{d}s && \text{(2.64b)}\\
&= \int_{-\infty}^{\infty} g(t-s)\int \left(\Theta(s)\sum_{j=1}^{M} e^{\lambda_j s}\Pi_j^*\Psi(\mathbf{x})\right)\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}s\mathrm{d}\mathbf{x} && \text{(2.64c)}\\
&= \sum_{j=1}^{M}\sum_{k=1}^{a_j-1}\alpha_j^{(k)}\frac{1}{k!}\int_{-\infty}^{\infty} g(t-s)\Theta(s)e^{\lambda_j s}s^k\mathrm{d}s && \text{(2.64d)}\\
&= \sum_{j=1}^{M}\sum_{k=1}^{a_j-1}\alpha_j^{(k)}\frac{1}{k!}\left[g * \left(\Theta e^{\lambda_j \circ}\circ^k\right)\right](t), && \text{(2.64e)}
\end{aligned}$$

where the coefficients $\alpha_j^{(k)}$ are defined immediately below Eq. (2.45). At a practical level, when a system is weakly forced so that its response is at the linear regime, $\delta^{(1)}[\Psi](t)$ can be observed from numerical simulations. In particular, if $g(t) = \delta(t)$, the perturbation would be equivalent to an ε-sized shift in the initial condition of the unperturbed system. This comment raises the question of using empirical measurements of the response $\delta^{(1)}\left[\Psi\right](t)$ to find the decomposition of the Green function in terms of the point spectrum, which can be thence convoluted against different time-modulations. In effect, Eq. (2.64) provides a basis for the empirically observed response $\delta^{(1)}\left[\Psi\right](t)$ to project onto.

In Fourier domain, the convolution structure of $\delta^{(1)}\left[\Psi\right](t)$ is transformed into a product:

$$\mathfrak{F}\left[\delta^{(1)}\left[\Psi\right]\right](\omega) = \mathfrak{F}\left[g * \mathcal{G}\right](\omega) = \mathfrak{F}\left[g\right](\omega)\mathfrak{F}\left[\mathcal{G}\right](\omega), \tag{2.65}$$

where ω is a real frequency and $\mathfrak{F}\left[\mathcal{G}\right]$ is the susceptibility function which it is analytic in the upper complex plane, by virtue of the causality constraint. Furthermore, if one assumes polynomial decay of $\mathfrak{F}\left[\mathcal{G}\right](\omega)$, Kramers-Kronig relations can be derived, which link the imaginary and real parts of the susceptibility function via Cauchy integrals [Luc09]. Under

suitable integrability assumptions, one can, moreover, link the logarithm of the modulus of the susceptibility with its real part, and then use Kramers-Kronig relations to find the imaginary counterpart [Luc12].

Going back to the impulse force where $g(t) = \delta(t)$, it follows from Eq. (2.65) that the linear response in frequency domain is exactly given by the susceptibility. If $g(t) = \Theta(t)$,

$$\mathfrak{F}\left[\delta^{(1)}\left[\Psi\right]\right](\omega) = \mathfrak{F}\left[\Theta * \mathcal{G}\right](\omega) = \left(\pi\delta(\omega) - i\mathbf{P}\left(\frac{1}{\omega}\right)\right)\mathfrak{F}\left[\mathcal{G}\right](\omega), \tag{2.66}$$

for the real frequency ω and where $\mathbf{P}$ denotes the Cauchy principal value of a singular function.

Higher order terms in the frequency domain are obtained by computing the Fourier transform of Eq. (2.59):

$$\begin{aligned}&\mathfrak{F}\left[\delta^{(k)}\left[\Psi\right]\right](\omega) = \\ &\int_{-\infty}^{\infty}\dots\int_{-\infty}^{\infty}\mathfrak{F}\left[\mathcal{G}^{(k)}\right](\omega_1,\dots,\omega_k)g(\omega_1)\dots g(\omega_k)\delta\left(\omega - \sum_{i=1}^{k}\omega_k\right)\mathrm{d}\omega_1\dots\mathrm{d}\omega_k,\end{aligned} \tag{2.67}$$

where the delta-function forces the integrand to be non-zero only when the sum of the Fourier frequencies sum up to the input ω.

The Fourier transform of the linear response has a more convenient representation in terms of the information coming from a simple point spectrum of $\mathcal{L}_0$:

$$\mathfrak{F}\left[\delta^{(1)}\Psi(t)\right](\omega) = \sum_{j=1}^{M}\alpha_j\mathfrak{F}\left[g * \mathcal{G}\right](\omega) = \mathfrak{F}\left[g\right](\omega)\sum_{j=1}^{M}\frac{\alpha_j}{i\omega - \lambda_j}. \tag{2.68}$$

The susceptibility function $\mathfrak{F}\left[\mathcal{G}\right]$ can be meromorphically extended to all values of ω in $\mathbb{C}$ such that $\mathfrak{Im}\omega > \omega_{ess}$, where the poles correspond to those of the resolvent $R(\cdot, \mathcal{L}_0)$ which are precisely given by the eigenvalues of $\mathcal{L}_0$ with the exception of that at $\omega = 0$ by the observation made in Eq. (2.46). It is then clear that the location of the poles in the susceptibility function depend, uniquely, on the underlying system and not on the time-modulation function g or the observable in question. On the other hand, the spectral coefficients α_j also known as residues [Rue86], do depend on the observable. Comparing Eq. (2.68) to Eq. (2.49c), we notice, once again, that the pole at $z = 0$ in Eq. (2.49c) is solely due to the Laplace transforms taken there.

2.3.1 Stochastically Perturbed Deterministic Systems

The theory of random dynamical systems regards stochastic flows as transformations of phase space parametrised in time but also in the noise realisation [Arn98]. In this sense, one can view noise as the inhomogeneous component of an equation that makes it, loosely

speaking, non-autonomous. Interestingly, when the noise component is relatively weak, it can be regarded as a perturbation to the otherwise unperturbed deterministic system. Consequently, the leading order modification to the system's statistics is expected to be captured by the linear response function, in the spirit of the previous sections where here, instead, the perturbation vector field will be encoded in the covariance matrix. Indeed, such is the case, although two main issues arise. First, the reference deterministic system might not posses an absolutely continuous invariant measure with respect to Lebesgue which prevents the use of density functions. Second, the Green function can be seamlessly applied to derive realisation-dependent response formulas although these do not agree with a perturbation expansion of the Fokker-Planck equation. This section, therefore, aims at (i), determining the leading order correction to the statistics using operator relations and (ii), proving that Green function and operator based response formulas are equal only when the former is taken in the Stratonovich setting.

Stochastic perturbations to deterministic systems have been widely studied in the literature although such topic will not be reviewed here. Instead, the reader is referred to [Gar09, Chapter 7]. We shall, nonetheless, follow previous work based on response theory whereby, taking advantage of Ruelle's formalism, [Luc12] deduces the effects of adding noise to a deterministic system showing, additionally, how to understand Green's function in the stochastic sense. The same question is tackled in [Abr17], although the author resorts to expanding the perturbed flow in terms of the tangent map leading to an apparently different formula to that found in [Luc12]. These two formulas will be compared below.

From the operator point of view, the major background result concerns the stability of the spectrum of the transfer and Koopman semigroups [BK98]; [KL99]. These results are formulated for general transfer operators with a non-trivial point spectrum and conclude that under small and generic stochastic perturbations, the point spectrum will experience relatively small changes. In particular, if a system is exponentially mixing, small applied fields wont change that character, cf. Proposition 3.2.3 in Chapter 3.

In this section we shall start by considering a stochastically perturbed system that obeys the following Itô SDE:

$$\mathrm{d}\mathbf{x}(t) = \mathbf{F}(\mathbf{x})\mathrm{d}t + \varepsilon\mathbf{G}(\mathbf{x})\mathrm{C}\mathrm{d}W_t, \tag{2.69}$$

where $\mathbf{F} : \mathcal{X} \longrightarrow \mathbb{R}^d$ is the drift, $\mathbf{G} : \mathcal{X} \longrightarrow \mathbb{R}^{d\times p}$ is a perturbation matrix, and W_t denotes an independent p-dimensional Wiener process with $p \times p$ covariance matrix $\mathrm{C}^2 = \mathrm{C}\mathrm{C}^\top$. The associated Fokker-Planck equation reads as:

$$\partial_t\rho(\cdot, t) = \left(\mathcal{L}_0 + \varepsilon^2\mathcal{L}_2\right)\rho(\cdot, t), \tag{2.70}$$

where the operator $\mathcal{L}_0$ is the usual Liouvillian and the perturbation operator $\mathcal{L}_2$ above is

defined for ρ in $D(\mathcal{L}_0)$ as:

$$\mathcal{L}_2\rho = \frac{1}{2}\nabla^2 : \left(\rho\mathbf{G}\mathrm{C}^2\mathbf{G}^\top\right). \tag{2.71}$$

This shows why ε is squared in Eq. (2.70) and justifies the subscript in $\mathcal{L}_2$. Equation (2.70) describes the evolution of densities under the action of the stochastic flow, whereas expectation values of observables Ψ evolve according to the backward-Kolmogorov equation:

$$\partial_t\Psi(\cdot,t) = \left(\mathcal{L}_0^* + \varepsilon^2\mathcal{L}_2^*\right)\Psi(\cdot,t). \tag{2.72}$$

While the operator $\mathcal{L}_0^*$ was already defined in Eq. (2.25a), $\mathcal{L}_2^*$ is:

$$\mathcal{L}_2^*\Psi = \frac{1}{2}\left(\mathbf{G}\mathrm{C}^2\mathbf{G}^\top\right) : \nabla^2\Psi, \tag{2.73}$$

for Ψ in $D(\mathcal{L}_2^*)$. Here we immediately observe that stochastic forcing enters at second order in powers of ε when considering the operator representation of the stochastic process (2.69). This is due to the timescale carried by the Wiener process, by which characteristic times are $\sqrt{\mathrm{d}t}$ in Eq. (2.69) producing the squared ε that accompanies $\mathcal{L}_2$ and $\mathcal{L}_2^*$. Using Eq. (2.56) we can derive the leading order response of the steady state to the introduction of noise:

$$\delta^{(2)}\left[\Psi\right](t) = \int\int_0^t \mathcal{L}_2^*\Psi(\mathbf{x}(t))\rho_0(\mathbf{x})\mathrm{d}s\mathrm{d}\mathbf{x} \tag{2.74a}$$

$$= \frac{1}{2}\int\int_0^t \left(\mathbf{G}(\mathbf{x})\mathrm{C}^2\mathbf{G}^\top(\mathbf{x})\right) : \nabla^2\Psi(\mathbf{x}(s))\rho_0(\mathbf{x})\mathrm{d}s\mathrm{d}\mathbf{x}. \tag{2.74b}$$

The right hand side of Eq. (2.74) does not require absolute continuity of the measure ρ_0, making it applicable in dissipative systems with singular measures: $\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} \mapsto \rho_0(\mathrm{d}\mathbf{x})$. Note that we have evaluated the deterministic Koopman operator $e^{t\mathcal{L}_0^*}$ to the observable Ψ to give $e^{t\mathcal{L}_0^*}\Psi(\mathbf{x}) = \Psi(\mathbf{x}(t))$, where $\mathbf{x}(t)$ solves Eq. (2.69) for $\varepsilon = 0$; this is in accordance with Eq. (2.4).

Formula Eq. (2.74) coincides with what was obtained in [Abr17], where instead, the author resorted to the tangent space in order to evaluate $\left(\mathbf{G}(\mathbf{x})\mathrm{C}^2\mathbf{G}^\top(\mathbf{x})\right) : \nabla^2\Psi(\mathbf{x}(t))$ avoiding at all moments the density representation of the invariant measure ρ_0. An approach based on high-order Green functions was taken in [Luc12], although it yielded a different formula. In such work, the Wiener increments $\mathrm{d}W_t$ were treated as the time-modulation of a sequence of p perturbations encoded in the matrix $\mathbf{G}$. To illustrate this idea, it is useful to rewrite Eq. (2.69) as a series of applied vector fields modulated by

correlated Wiener increments:

$$\mathrm{d}\mathbf{x} = \mathbf{F}(\mathbf{x})\mathrm{d}t + \varepsilon \sum_{i=1}^{p} g_i(t)\mathbf{G}_{:,i}(\mathbf{x}), \tag{2.75}$$

where $\mathbf{G}_{:,i}$ is the ith column of $\mathbf{G}$ and $g_i(t)$ are assumed to satisfy $\mathbb{E}\left[g_i(t)\right] = 0$ and $\mathbb{E}\left[g_k(t)g_l(t)\right] = \mathrm{C}^2_{k,l}$ for every $i, k, l = 1, \ldots, p$. For a given realisation of the noise indexed by σ, one can apply Eq. (2.58) to obtain:

$$\delta^{(1)}_{\sigma}\left[\Psi\right](t) = \sum_{i=1}^{p} \int_{-\infty}^{\infty} \mathcal{G}_i(s) g_i(t-s)\mathrm{d}s, \tag{2.76}$$

where $\mathcal{G}_i$ denotes the Green function associated with the vector field $\mathbf{G}_{:,i}$ for every i. Since, g_i corresponds to a Wiener increment, taking averages over all possible realisations σ makes the linear responses vanish: $\mathbb{E}\left[\delta^{(1)}_{\sigma}\left[\Psi\right]\right] = 0$. Contrarily, the second order response $\delta^{(2)}_{\sigma}\left[\Psi\right]$ survives the averaging and is simplified to a single-time-variable integral of the form:

$$\tilde{\delta}^{(2)}\left[\Psi\right](t) := \mathbb{E}\left[\delta^{(2)}_{\sigma}\left[\Psi\right]\right] = \frac{1}{2}\sum_{k,l=1}^{p} \int \rho_0(\mathrm{d}\mathbf{x}) \int_{-\infty}^{\infty} \Theta(s)\mathcal{L}^*_{1,k}\mathcal{L}^*_{1,l} e^{s\mathcal{L}^*_0}\Psi(\mathbf{x})\mathrm{d}s \tag{2.77a}$$

$$= \frac{1}{2}\sum_{k,l=1}^{p} \int \rho_0(\mathrm{d}\mathbf{x}) \int_{-\infty}^{\infty} \Theta(s)\mathbf{G}_{:,k}\cdot\nabla\left(\mathbf{G}_{:,l}(\mathbf{x})\cdot\nabla e^{s\mathcal{L}^*_0}\Psi(\mathbf{x})\right)\mathrm{d}s, \tag{2.77b}$$

where $\mathcal{L}^*_{1,i} = \mathbf{G}_{:,i}\cdot\nabla$. To compare Eq. (2.77b) with Eq. (2.74), it is handy to disentangle the column vector fields $\mathbf{G}_{:,i}$ from the correlation matrix $\mathbf{G}$ so that, after some vector-matrix manipulations, one deduces that:

$$\delta^{(2)}\left[\Psi\right](t) = \frac{1}{2}\sum_{k=1}^{p}\sum_{l=1}^{p} \mathrm{C}^2_{k,l} \int \rho_0(\mathrm{d}\mathbf{x}) \int_0^t \left(\mathbf{G}_{:,k}\mathbf{G}^{\top}_{:,l}\right) : \nabla^2\Psi(s)\mathrm{d}s. \tag{2.78}$$

It is distinct, then, that the approach described in the present document using operator expansions yields a different formula to that of [Luc12] for the seemingly same question. However, the tackling of such problem is, naturally, different, since the Green function formalism regards the Wiener increments as an explicit time function whereas the operator approach, by default, encodes the all averages in configuration space. To understand the connection between these two approaches, we shall investigate whether the approach taken in [Luc12] can be expressed more rigorously in terms of operators and compare it with the perturbation operator in Eq. (2.73). To this end and for clarity, let us consider Eq. (2.75) where a single perturbation $\mathbf{G}_{:,1}$ is applied, namely, $p = 1$. This means that $\mathrm{C}^2 \equiv 1$ and we can rewrite Eq. (2.75) as:

$$\mathrm{d}\mathbf{x}(t) = \mathbf{F}(\mathbf{x})\mathrm{d}t + \varepsilon\tilde{\mathbf{G}}H\mathrm{d}W_t, \tag{2.79}$$

where $\mathbf{x}(t)$ is in $\mathbb{R}^d$, $\tilde{\mathbf{G}} = \mathrm{diag}(\mathbf{G}_{:,1})$, W_t is a d-dimensional Wiener process and H is a $d \times d$ matrix defined as:

$$H = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} [1, 0, \ldots, 0] = \mathbb{1}_d \mathbf{e}_1^\top. \tag{2.80}$$

Using this setting, we claim that Lucarini's approach coincides with Eq. (2.74) if Eq. (2.79) is interpreted in the Stratonovich sense (or $\mathbf{G}_{:,1}$ is a constant vector). We proceed by expanding the operator product $\mathcal{L}^*_{1,1}\mathcal{L}^*_{1,1}$, which is instrumental in Eq. (2.77a). Let f denote a generic twice differentiable function:

$$\begin{aligned}
\mathcal{L}^*_{1,1}\mathcal{L}^*_{1,1} f &= \mathbf{G}_{:,1} \cdot \nabla \left(\mathbf{G}_{:,1} \cdot \nabla f \right) && \text{(2.81a)} \\
&= \mathbf{G}_{:,1} \cdot \left[\left(\mathbf{G}_{:,1} \cdot \nabla \right) \nabla f + \left(\nabla f \cdot \nabla \right) \mathbf{G}_{:,1} \right] \\
&\qquad + \mathbf{G}_{:,1} \cdot \left[\mathbf{G}_{:,1} \times \left(\nabla \times \nabla f \right) + \nabla f \times \left(\nabla \times \mathbf{G}_{:,1} \right) \right] && \text{(2.81b)} \\
&= \mathbf{G}_{:,1} \cdot \left[\left(\mathbf{G}_{:,1} \cdot \nabla \right) \nabla f + \left(\nabla f \cdot \nabla \right) \mathbf{G}_{:,1} + \nabla f \times \left(\nabla \times \mathbf{G}_{:,1} \right) \right] && \text{(2.81c)} \\
&= \mathbf{G}_{:,1} \cdot \left(\mathbf{G}_{:,1} \cdot \nabla \right) \nabla f + \mathbf{G}_{:,1} \cdot \left(\nabla f \cdot \nabla \right) \mathbf{G}_{:,1} + \mathbf{G}_{:,1} \cdot \left(\nabla f \times \nabla \times \mathbf{G}_{:,1} \right) && \text{(2.81d)} \\
&= \mathbf{G}_{:,1} \cdot \left(\mathbf{G}_{:,1} \cdot \nabla \right) \nabla f + \mathbf{G}_{:,1} \cdot \left(\nabla f \cdot \nabla \right) \mathbf{G}_{:,1} - \nabla f \cdot \left(\mathbf{G}_{:,1} \times \nabla \times \mathbf{G}_{:,1} \right) && \text{(2.81e)} \\
&= \left(\mathbf{G}_{:,1} \mathbf{G}_{:,1}^\top \right) : \nabla^2 f + \mathbf{G}_{:,1} \cdot \left(\nabla f \cdot \nabla \right) \mathbf{G}_{:,1} - \nabla f \cdot \left(\mathbf{G}_{:,1} \times \nabla \times \mathbf{G}_{:,1} \right). && \text{(2.81f)}
\end{aligned}$$

We immediately observe that if $\nabla \mathbf{G}_{:,1} \equiv 0$, the response formula derived in Eqs. (2.74) and (2.78) coincide. Consequently, agreement is found if the forcing is additive white noise. The second and third terms on the RHS of Eq. (2.81f) are extra terms that can be associated with the Itô-to-Stratonovich correction [Pav14]. Indeed, we see this by regarding Eq. (2.79) in the Stratonovich sense:

$$\mathrm{d}\mathbf{x}(t) = \mathbf{F}(\mathbf{x})\mathrm{d}t + \varepsilon \tilde{\mathbf{G}} H \circ \mathrm{d}W_t. \tag{2.82}$$

The corresponding Itô conversion reads and expands as

$$\mathrm{d}\mathbf{x}(t) = \left[\mathbf{F}(\mathbf{x}) + \frac{\varepsilon^2}{2} \left[\nabla \cdot \left((\tilde{\mathbf{G}} H)(\tilde{\mathbf{G}} H)^\top \right) - \left(\tilde{\mathbf{G}} H \right) \nabla \cdot \left(\tilde{\mathbf{G}} H \right) \right] \right] \mathrm{d}t + \varepsilon \tilde{\mathbf{G}} H \mathrm{d}W_t \tag{2.83a}$$

$$= \left[\mathbf{F}(\mathbf{x}) + \frac{\varepsilon^2}{2} \left[\nabla \cdot \left(\mathbf{G}_{:,1} \mathbf{G}_{:,1}^\top \right) - \left(\nabla \cdot \mathbf{G}_{:,1} \right) \mathbf{G}_{:,1} \right] \right] \mathrm{d}t + \varepsilon \tilde{\mathbf{G}} H \mathrm{d}W_t. \tag{2.83b}$$

Note that an ε^2 term has appeared in the drift component. Consequently, the backward-Kolmogorov equation associated with Eq. (2.83) will only have perturbation operators of order ε^2. Following the Dyson expansions that lead to Eq. (2.74), the perturbation operator

$\mathcal{L}_2^*$ for Eq. (2.83) writes as:

$$\mathcal{L}_2^* f = \frac{1}{2}\left[\nabla\cdot\left(\mathbf{G}_{:,1}\mathbf{G}_{:,1}^\top\right) - \left(\nabla\cdot\mathbf{G}_{:,1}\right)\mathbf{G}_{:,1}\right]\cdot\nabla f + \frac{1}{2}\mathbf{G}_{:,1}\mathbf{G}_{:,1}^\top : \nabla^2 f \tag{2.84a}$$

$$= \frac{1}{2}\left[\left(\mathbf{G}_{:,1}\cdot\nabla\right)\mathbf{G}_{:,1}\right]\cdot\nabla f + \frac{1}{2}\mathbf{G}_{:,1}\mathbf{G}_{:,1}^\top : \nabla^2 f \tag{2.84b}$$

$$= \frac{1}{2}\left[\frac{1}{2}\nabla\left(\mathbf{G}_{:,1}\cdot\mathbf{G}_{:,1}\right) - \mathbf{G}_{:,1}\times\nabla\times\mathbf{G}_{:,1}\right]\cdot\nabla f + \frac{1}{2}\mathbf{G}_{:,1}\mathbf{G}_{:,1}^\top : \nabla^2 f. \tag{2.84c}$$

Hence, comparing Eq. (2.84c) and Eq. (2.81f), we are left with showing that $\nabla\left(\mathbf{G}_{:,1}\cdot\mathbf{G}_{:,1}\right)\cdot\nabla f = 2\mathbf{G}_{:,1}\cdot\left(\nabla f\cdot\nabla\right)\mathbf{G}_{:,1}$, which can be seen by direct evaluation:

$$\frac{1}{4}\nabla\left(\mathbf{G}_{:,1}\cdot\mathbf{G}_{:,1}\right)\cdot\nabla f = \frac{1}{4}\begin{bmatrix}2\mathbf{G}_{1,1}\partial_{x_1}\mathbf{G}_{1,1} + \ldots + 2\mathbf{G}_{d,1}\partial_{x_1}\mathbf{G}_{d,1}\\ \vdots \\ 2\mathbf{G}_{1,1}\partial_{x_d}\mathbf{G}_{1,1} + \ldots + 2\mathbf{G}_{d,1}\partial_{x_d}\mathbf{G}_{d,1}\end{bmatrix}\cdot\nabla f \tag{2.85a}$$

$$= \frac{1}{2}\left[\mathbf{G}_{:,1}\cdot\partial_{x_1}f\partial_{x_1}\mathbf{G}_{:,1} + \ldots + \mathbf{G}_{:,1}\cdot\partial_{x_d}f\partial_{x_d}\mathbf{G}_{:,1}\right] \tag{2.85b}$$

$$= \frac{1}{2}\mathbf{G}_{:,1}\cdot\left(\nabla f\cdot\nabla\right)\mathbf{G}_{:,1}. \tag{2.85c}$$

Which proves the claim. This can be formalised in the following proposition:

Proposition 2.3.1. *Consider the SDE* (2.69) *in Stratonovich form and let* $\mathbf{H}:\mathbb{R}^d\longrightarrow\mathbb{R}^d$ *be the Stratonovich-to-Itô correction:*

$$\mathbf{H} = \frac{\varepsilon^2}{2}\left[\nabla\cdot\left(\mathbf{G}\mathrm{C}^2\mathbf{G}^\top\right) - \mathbf{G}\mathrm{C}^2\left(\nabla\cdot\mathbf{G}^\top\right)\right]. \tag{2.86}$$

Then,

$$\tilde{\delta}^{(2)}\left[\Psi\right] - \delta^{(2)}\left[\Psi\right] = \frac{1}{\varepsilon^2}\int\rho_0(\mathrm{d}\mathbf{x})\int_{-\infty}^{\infty}\Theta(s)\mathbf{H}(\mathbf{x})\cdot\nabla\Psi(\mathbf{x}(s))\mathrm{d}s, \tag{2.87}$$

where the first and second terms of the LHS are defined in Eq. (2.77a) *and Eq.* (2.78) *respectively.*

Proof. We note the following chain of equalities:

$$\mathrm{d}\mathbf{x} = \mathbf{F}\mathrm{d}t + \varepsilon\mathbf{G}\mathrm{C}\circ\mathrm{d}W_t = \left[\mathbf{F}+\mathbf{H}\right]\mathrm{d}t + \varepsilon\mathbf{G}\mathrm{C}\mathrm{d}W_t \tag{2.88a}$$

$$= \left[\mathbf{F} + \sum_{k=1}^{p}\sum_{l=1}^{p}\mathbf{H}_{k,l}\right]\mathrm{d}t + \varepsilon\sum_{i=1}^{p}\tilde{\mathbf{G}}_i H_i\mathrm{C}\mathrm{d}W_t, \tag{2.88b}$$

where $\tilde{\mathbf{G}}_i = \mathrm{diag}\left(\mathbf{G}_{:,i}\right)$, $H_i = \mathbb{1}\mathbf{e}_i^\top$ and $\mathbf{H}_{k,l}$ is defined as:

$$\mathbf{H}_{k,l} = \frac{\varepsilon^2}{2}\mathrm{C}_{k,l}\left[\nabla\cdot\left(\mathbf{G}_{:,k}\mathbf{G}_{:,l}^\top\right) - \tilde{\mathbf{G}}_k\nabla\cdot\tilde{\mathbf{G}}_l\right]. \tag{2.89}$$

Now, it is enough to repeat the argument started in Eq. (2.79) for each $\tilde{\mathbf{G}}_i$ and then use the

linearity of the leading order response to extend it for $i = 1, \dots, p$. □

Remark 2.3.1. *In case the applied fields* $\mathbf{G}_{:,i}$ *in Eq.* (2.75) *are independent, we have that* $\mathrm{C} \equiv 1$ *in the SDE* (2.83) *making the cross-terms disappear in views that:*

$$\tilde{\mathbf{G}}_{:,k} H_k \mathrm{C} \tilde{H}_l^\top \mathbf{G}_{:,l} = \tilde{\mathbf{G}}_{:,k} H_k \tilde{H}_l^\top \tilde{\mathbf{G}}_{:,l} = \tilde{\mathbf{G}}_{:,k} \mathbb{1} \mathbf{e}_k^\top \mathbf{e}_l \mathbb{1}^\top \tilde{\mathbf{G}}_{:,l} \equiv 0. \tag{2.90}$$

for every $k \neq l$*. Consequently, the* $\mathcal{L}_2^*$ *operator associated with Eq.* (2.83) *reads as:*

$$\mathcal{L}_2^* = \left[\sum_{i=1}^{p} \nabla \cdot \left(\mathbf{G}_{:,i} \mathbf{G}_{:,i}^\top \right) - \sum_{i=1}^{p} \tilde{\mathbf{G}}_i \nabla \cdot \tilde{\mathbf{G}}_i \right] \cdot \nabla + \frac{1}{2} \sum_{i=1}^{p} \left(\mathbf{G}_{:,i} \mathbf{G}_i^\top \right) : \nabla^2, \tag{2.91}$$

which would equate to applying the linear response formula to each field $\mathbf{G}_{:,i}$ *independently and, then, summing up over* $i = 1, \dots, p$*.*

In [Luc12] linear response of power spectral densities is related to the modulus of the susceptibility function in case of white-noise modulated forcings. Indeed, there it is argued that if $\mathfrak{F}\left[C^{\epsilon}_{\Psi,\Psi}\right](\omega)$ is the perturbed power spectrum of the observable Ψ then we have:

$$\mathbb{E}\left[\mathfrak{F}\left[C^{\varepsilon}_{\Psi,\Psi}\right](\omega)\right] - \mathfrak{F}\left[C_{\Psi,\Psi}\right](\omega) \approx \mathbb{E}\left[\left|\delta^{(2)}_{\sigma}\left[\Psi\right]\right|^2\right] \approx \varepsilon^2 \left|\mathfrak{F}\left[\mathcal{G}\right](\omega)\right|^2. \tag{2.92}$$

In order to prove this, the author of [Luc12] invoked the Wiener-Khinchin theorem and applied it to the autocorrelation function of $\delta^{(1)}_{\sigma}[\Psi](t)$ before taking averages over σ. Equation (2.92) appears to have a term missing if one compares it to Eq. (2.36). This is due to the fact that, before averaging over realisations, the leading order response of the flow is proportional to ε whereas the perturbed invariant measure responds proportionally to ε^2. Therefore, the RHS of Eq. (2.36) yields a single term from which Eq. (2.92) follows.

Notice that the left hand side of the equation above is especially easy to calculate if one is let to sample the perturbed and unperturbed dynamics. Thus, it is possible to obtain estimates of the modulus of the susceptibility function in a fully empirical way. One is left with finding the actual locations of the poles, which the modulus itself cannot reveal. This would be attained by noticing that the logarithm of the susceptibility function can be written as:

$$\log\left(\mathfrak{F}\left[\mathcal{G}\right](\omega)\right) = \log\left(\left|\mathfrak{F}\left[\mathcal{G}\right](\omega)\right|\right) + i\varphi(\omega), \tag{2.93}$$

for every ω and some phase function φ. Assuming, further, that $\log\left(\mathfrak{F}\left[\mathcal{G}\right](\omega)\right)$ is analytic, Kramers-Kronig relations are invoked to link its real and imaginary parts shown in Eq. (2.93) and, thus, that of $\mathfrak{F}\left[\mathcal{G}\right](\omega)$. This way, Eq. (2.92) constitutes a practical way of estimating the susceptibility function of a deterministic flow by using time series of the stochastically forced vector field.

2.4 Homogeneous Equation for the Response

The limit of weak forcing strength is instrumental in deriving meaningful linear response formulas that capture statistical changes due to an applied external field. When small, yet finite-sized perturbations are introduced, power series are resorted to and their convergence depends on a radius of expansion. When the perturbations are not infinitesimal or small, linear response formulas cease to be useful and the power series of high-order responses might diverge. One seems inevitably obliged to solve the perturbed Fokker-Planck equation (2.52) to find the probability state at a certain time to evaluate the phase average or full response of a given observable. In the present section, we aim at deriving a scalar equation for the response function $\langle \Psi, \rho(\cdot, t) \rangle$ in Eq. (2.56) that does not require a power expansion nor the full knowledge of perturbed probability density function. The price to pay, as will be detailed later, is the introduction of projected operators and memory in the resulting equation.

The set of techniques explained here are based on the calculations presented in [Ken71]; [Ken73] in the context of quantum mechanical systems, at a time when linear response and the subsequent fluctuation-dissipation theorem were just being introduced. Letting alone the mathematically rigorous formulation of such results that eventually came with Ruelle, the physical community was far from being unanimous with regards to Kubo's discoveries. The source of criticism for linear response was that it was not clear whether infinitesimal changes in the microscopic nature of the system would provoke relatively equal disturbances in the average behaviour. This was the cornerstone of the objections to the perturbative approaches and it is summarised in [Kam71]. In such work, it is argued that the chaotic nature of statistical mechanical systems can provoke exponentially diverging microscopic trajectories on phase space with a rate given by the first Lyapunov exponent resulting in statistical responses well beyond linearity. Hence, alternative ways of simplifying the study of response were needed, one of them being that originated in [Ken71]; see also [VV78] for a review on attempts to justify linear response theory.

Now, well after Ruelle's mathematical proof of linear response, it is our desire to review the formulas presented of [Ken71]; [Ken73] but in a stochastic framework and with references to recent advances in the treatment of operator equations. To do this, the equation of motion— regardless it being stochastic or deterministic— will suitably be projected to derive, with the aid of the Mori-Zwanzig formalism [Zwa01], a scalar equation for the response $\langle \Psi, \rho(\cdot, t) \rangle$ where the only independent variable is $\langle \Psi, \rho(\cdot, t) \rangle$ itself; this is, an *homogeneous* equation. It will be shown, moreover, that such relations can be exploited to find explicit solutions for the response in frequency domain in a way that they are not expressed in power series as opposed to previous sections.

To achieve this goal, we shall start by considering the forced SDE (2.15). Since the introduction of perturbations makes the equations of motion non-autonomous, the

exponential operator no longer serves as a notation for the solution of Eq. (2.16) unless we introduce some assumptions on $\mathcal{L}(t)$. First, we shall assume that $\mathcal{L}(t)$ generates a strongly continuous unitary semigroup for every $t \geq 0$. Second, we assume that the family $\mathcal{L}(t)$ is time-ordered (see [Gil17] for precise definition) so that the operator $\int_s^t \mathcal{L}(\tau)\mathrm{d}\tau$ is a strong Riemann integral and the generator of a strongly continuous semigroup on t so that if $\rho(\cdot,t) = e^{\int_s^t \mathcal{L}(\tau)\mathrm{d}\tau}\rho_s$ then

$$\partial_t \rho(\cdot,t) = \mathcal{L}(t)\rho(\cdot,t), \text{ with } \rho(\cdot,s) = \rho_s. \tag{2.94}$$

The exponential operator in this sense is, therefore, called a time-ordered exponential [Gil17]. To finish the notation for this section we will define the full response function $R_\Psi(t)$ as:

$$R_\Psi(t) := \langle \Psi, \rho(\cdot,t)\rangle = \int \Psi(\mathbf{x})\rho(\mathbf{x},t)\mathrm{d}\mathbf{x}, \tag{2.95}$$

where $\rho(\cdot,t)$ is a solution of Eq. (2.16). Traditionally, the response function is defined as the difference between the perturbed and unperturbed means of the observable Ψ, but since the latter is stationary, this definition will not introduce differences with other formulations.

The derivation of an homogeneous response equation starts by differentiating Eq. (2.95) with respect to time:

$$\partial_t R_\Psi(t) = \partial_t \int \Psi(\mathbf{x})\rho(\mathbf{x},t)\mathrm{d}\mathbf{x} = \int \Psi(\mathbf{x})\partial_t\rho(\mathbf{x},t)\mathrm{d}\mathbf{x} = \int \Psi(\mathbf{x})\mathcal{L}(t)\rho(\mathbf{x},t)\mathrm{d}\mathbf{x}. \tag{2.96}$$

This equation is exact and describes the evolution of the response to any order in ε for an unlimited amount of time. However, we want to avoid solving the Fokker-Planck equation (2.16) exactly to calculate $\rho(\cdot,t)$, on the contrary, the target is to extract from $\rho(\cdot,t)$ the information necessary in order to describe $R_\Psi(t)$ alone. For such purpose, we introduce the projection $\mathbb{P}$ defined by:

$$\mathbb{P}f = \xi^{-1}\left(\int \Psi(\mathbf{x})f(\mathbf{x})\mathrm{d}\mathbf{x}\right)\rho_0, \tag{2.97}$$

where,

$$\xi := \int \Psi(\mathbf{x})\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{2.98}$$

Consequently with Eq. (2.97), our main assumption here is that ξ is non-zero. To check that, indeed, $\mathbb{P}$ is a projection, it suffices to take a function f and evaluate it twice against $\mathbb{P}$:

$$\mathbb{P}\mathbb{P}f = \xi^{-1}\left(\int \Psi(\mathbf{x})\mathbb{P}f(\mathbf{x})\mathrm{d}\mathbf{x}\right)\rho_0 \tag{2.99a}$$

$$= \xi^{-1}\left(\int \Psi(\mathbf{x})\xi^{-1}\left(\int \Psi(\mathbf{x})f(\mathbf{x})\mathrm{d}\mathbf{x}\right)\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}\right)\rho_0 = \mathbb{P}f \tag{2.99b}$$

as desired. With these definitions, we immediately find that:

$$\mathbb{P}\rho(\cdot,t) = \xi^{-1}\left(\int \Psi(\mathbf{x})\rho(\mathbf{x},t)\mathrm{d}\mathbf{x}\right)\rho_0 = \xi^{-1}R_\Psi(t)\rho_0 \tag{2.100}$$

and that $\mathbb{P}\rho_0 = \rho_0$.

Next step is to derive an equation of motion of the projected function $\mathbb{P}\rho(\mathbf{x},t)$ by making use of the generalised Mori-Zwanzig formalism formulated originally in [MD69]. Applying the projections $\mathbb{P}$ and $1-\mathbb{P}$ to the Fokker-Planck equation Eq. (2.16) gives the following:

$$\begin{aligned}
\partial_t\mathbb{P}\rho(\cdot,t) =& \mathbb{P}\mathcal{L}(t)\rho(\cdot,t) = \mathbb{P}\mathcal{L}(t)\left(\mathbb{P}+1-\mathbb{P}\right)\rho(\cdot,t) && \text{(2.101a)}\\
=& \mathbb{P}\mathcal{L}(t)\mathbb{P}\rho(\cdot,t) + \mathbb{P}\mathcal{L}(t)\left(1-\mathbb{P}\right)\rho(\cdot,t), && \text{(2.101b)}\\
\partial_t\left(1-\mathbb{P}\right)\rho(\cdot,t) =& \left(1-\mathbb{P}\right)\mathcal{L}(t)\rho(\cdot,t) = \left(1-\mathbb{P}\right)\mathcal{L}(t)\left(\mathbb{P}+1-\mathbb{P}\right)\rho(\cdot,t) && \text{(2.101c)}\\
=& \left(1-\mathbb{P}\right)\mathcal{L}(t)\mathbb{P}\rho(\cdot,t) + \left(1-\mathbb{P}\right)\mathcal{L}(t)\left(1-\mathbb{P}\right)\rho(\cdot,t). && \text{(2.101d)}
\end{aligned}$$

Integrating Eq. (2.101d) we find the value of $\left(1-\mathbb{P}\right)\rho(\cdot,t)$:

$$\left(1-\mathbb{P}\right)\rho(\cdot,t) = G(t,0)\left(1-\mathbb{P}\right)\rho_0 + \int_0^t G(t,s)(1-\mathbb{P})\mathcal{L}(s)\mathbb{P}\rho(\cdot,s)\mathrm{d}s, \tag{2.102}$$

where $G(t,s) = e^{\int_s^t(1-\mathbb{P})\mathcal{L}(\tau)\mathrm{d}\tau}$. The final expression for the projected dynamics $\mathbb{P}\rho(\cdot,t)$ is split in four parts:

$$\begin{aligned}
\partial_t\mathbb{P}\rho(\cdot,t) =& && \text{(2.103a)}\\
& \mathbb{P}\mathcal{L}(t)\mathbb{P}\rho(\cdot,t) && \text{(2.103b)}\\
&+ \mathbb{P}\mathcal{L}(t)G(t,0)\left(1-\mathbb{P}\right)\rho_0 && \text{(2.103c)}\\
&+ \mathbb{P}\mathcal{L}(t)\int_0^t G(t,s)\left(1-\mathbb{P}\right)\mathcal{L}(s)\mathbb{P}\rho(\cdot,s)\mathrm{d}s, && \text{(2.103d)}
\end{aligned}$$

where the relevant terms of the equation have been separated into rows. The LHS of Eq. (2.103a) contains the time-derivative of the response function and, applying the definition of the projection operator $\mathbb{P}$, reads as:

$$\partial_t\mathbb{P}\rho(\cdot,t) = \partial_t\xi^{-1}\left(\int \Psi(\mathbf{x})\rho(\mathbf{x},t)\mathrm{d}\mathbf{x}\right)\rho_0 = \xi^{-1}\rho_0\partial_t R_\Psi(t). \tag{2.104}$$

The *streaming* term [MD69] refers to the projected evolution laws for the projected

variables and is present in Eq. (2.103b) and expanded below:

$$\mathbb{P}\mathcal{L}(t)\mathbb{P}\rho(\cdot,t) = \xi^{-2}\rho_0 \int \Psi(\mathbf{x})\mathcal{L}(t)\left(\int \Psi(\mathbf{x})\rho(\mathbf{x},t)\mathrm{d}\mathbf{x}\right)\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} \tag{2.105a}$$

$$= \xi^{-2}\rho_0 R_\Psi(t)\int \Psi(\mathbf{x})\mathcal{L}(t)\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{2.105b}$$

Third, we notice that $(1-\mathbb{P})\,\rho_0 \equiv 0$ so that the initial value term in Eq. (2.103c) vanishes; this results from Eq. (2.100). The last step is to expand the memory term by means of analysing the kernel inside the time integral in Eq. (2.103d):

$$\mathbb{P}\mathcal{L}(t)\int_0^t G(t,s)\,(1-\mathbb{P})\,\mathcal{L}(s)\mathbb{P}\rho(\cdot,s)\mathrm{d}s \tag{2.106a}$$

$$= \xi^{-2}R_\Psi(t)\rho_0\int_0^t\int \mathcal{L}(t)G(t,s)(1-\mathbb{P})\mathcal{L}(s)\rho_0(\mathbf{x})\mathrm{d}s. \tag{2.106b}$$

Upon substitution and simplification we find that the evolution equation for the response function R_Ψ can be written in more compact terms as:

$$\partial_t R_\Psi(t) = B(t)R_\Psi(t) + \int_0^t K(t,s)R_\Psi(s)\mathrm{d}s, \tag{2.107}$$

where the newly introduced function $B(t)$ and memory kernel $K(t,s)$ are defined as:

$$B(t) = \xi^{-1}\int \Psi(\mathbf{x})\mathcal{L}(t)\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{2.108a}$$

$$K(t,s) = \xi^{-1}\int \Psi(\mathbf{x})\mathcal{L}(t)e^{\int_s^t(1-\mathbb{P})\mathcal{L}(\tau)\mathrm{d}\tau}\,(1-\mathbb{P})\,\mathcal{L}(s)\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{2.108b}$$

Equation (2.107) becomes a scalar and homogeneous equation on R_Ψ, where there is not explicit reference to the time-dependent measure $\rho(\mathbf{x},t)$ that would be obtained by solving the full Fokker-Planck equation (2.16). Solving the non-Markovian equation (2.107) would yield the response function and, in principle, should be equal to the perturbative approach in Eq. (2.56), provided that the parameter ε is within the radius of convergence. A first check for the validity of Eq. (2.107) would involve showing that the leading order terms with respect ε would give the Kubo formula (2.58). Indeed, it is calculated in Appendix A, following [Ken71], how the memory equation (2.107) is solved to find that the linear response can be written in terms of the Green function as done earlier.

Remark 2.4.1. *The assumption of $\xi \neq 0$ is central in the derivation of Eq.* (2.107)*, as also noted in the first work [Ken71]. A generalisation for $\xi = 0$ is found in [Ken73], although the resulting equation differs from that of Eq.* (2.107) *because a time-dependent term is added on the RHS; see also [VV78].*

Step-function perturbation. If a constant forcing is activated at $t = 0$, time is modulated by a Heaviside distribution on positive values of time. In particular we are looking at the

Fokker-Planck equation expressed as:

$$\partial_t \rho(\cdot, t) = \mathcal{L}(t)\rho(\cdot, t) = \mathcal{L}_0 \rho\,(\cdot, t) + \varepsilon\Theta(t)\mathcal{L}_1\rho(\cdot, t), \tag{2.109}$$

where Θ is the Heaviside distribution with amplitude controlled by the perturbation parameter ε. The linear response and susceptibility functions were already calculated in Eqs. (2.63) and (2.66), respectively, using the Green function formulation. The same results can be obtained to leading order using the present approach; see Appendix A for the derivation.

We then look at Eq. (2.107) to study the full response function $R_\Psi(t)$. Upon substitution, the function $B(t)$ present in Eq. (2.108a) becomes:

$$B(t) = \xi^{-1} \int \Psi(\mathbf{x})\mathcal{L}(t)\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} = \xi^{-1} \int \Psi(\mathbf{x})\left(\mathcal{L}_0 + \varepsilon\Theta(t)\mathcal{L}_1\right)\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} \tag{2.110a}$$

$$= \varepsilon\Theta(t)\xi^{-1} \int \Psi(\mathbf{x})\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} = \varepsilon q\Theta(t), \tag{2.110b}$$

where we have introduced the scalar quantity $q := \int \Psi(\mathbf{x})\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}$. Furthermore, the integral kernel can be shown to become a convolution kernel by observing that for $t, s \geq 0$ we have:

$$G(t, s) = e^{\int_s^t (1-\mathbb{P})\mathcal{L}_(s)\mathrm{d}s} = e^{\int_s^t (1-\mathbb{P})(\mathcal{L}_0+\varepsilon\Theta(s)\mathcal{L}_1)\mathrm{d}s} = e^{(t-s)(1-\mathbb{P})(\mathcal{L}_0+\varepsilon\mathcal{L}_1)}. \tag{2.111}$$

Consequently, the function $K(t, s)$ depends on $t - s$ uniquely and Eq. (2.107) becomes:

$$\partial_t R_\Psi(t) = \varepsilon q R_\Psi(t) + \int_0^t K(t - s)R_\Psi(s)\mathrm{d}s, \tag{2.112}$$

which has the structure of a convolution. Finally, imposing causality as in Eq. (2.58), we take Fourier transforms:

$$\omega\mathfrak{F}\left[R_\Psi\right](\omega) - R_\Psi(0) = \varepsilon q\mathfrak{F}\left[R_\Psi\right](\omega) + \mathfrak{F}\left[K\right](\omega)\mathfrak{F}\left[R_\Psi\right](\omega), \tag{2.113}$$

for ω in $\mathbb{R}$ and whereby rearranging we get the explicit solution:

$$\mathfrak{F}\left[R_\Psi\right](\omega) = \frac{\xi}{\omega - \varepsilon q + \mathfrak{F}\left[K\right](\omega)}, \tag{2.114}$$

where we have assumed that $R_\Psi(0) = \xi$; for the justification, see Eq. (A.4) in Appendix A. The inverse transform would return the response function to all orders in the parameter ε.

In Section 2.3 the response formulas were given in terms of the kth order Green functions and, provided that the forcing parameter ε is small enough, yielded the full response in time and frequency domains; see Eqs. (2.60) and (2.67), respectively. Here, on the other hand, the response in frequency domain is given in terms of the memory

integral in Eq. (2.107). Such integral component is determined by the memory kernel K which is abstract and analytically intractable since it involves the exponential of a projected (unbounded) differential operator, namely, $(1-\mathbb{P})\mathcal{L}$. However, rearranging Eq. (2.114) and using the perturbative expansion in terms of the Green function based on Eq. (2.62) and its transform Eq. (2.67) we find:

$$\mathfrak{F}\left[K\right](\omega) = \omega - \varepsilon q - \left(\sum_{k=0}^{\infty} \varepsilon^k \delta^{(k)}[\Psi](\omega)\right)^{-1}, \tag{2.115}$$

which gives a formula for relating the memory term in Eq. (2.107) and the higher order responses obtained through the generalised Green functions defined in Eq. (2.59).

Impulse perturbation. For a Delta-modulated forcing, substituting $g(t)$ by $\delta(t)$ in Eq. (2.56) readily gives the linear response function Eq. (2.61), although gathering the leading order terms of Eq. (2.107) would also yield the same formula; see, again, Appendix A. For the full response, the function $B(t)$ in Eq. (2.108a) reads in this case as

$$B(t) = \xi^{-1} \int \Psi(\mathbf{x}) \mathcal{L}(t) \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x} = \xi^{-1} \int \Psi(\mathbf{x}) \varepsilon \delta(t) \mathcal{L}_1 \rho_0(\mathbf{x}) \mathrm{d}\mathbf{x} = \varepsilon q \delta(t). \tag{2.116}$$

The memory kernel $K(t,s)$ in Eq. (2.108b) can also be written in convolution form as for the Heaviside distribution. Indeed,

$$G(t-s) = e^{\int_s^t (1-\mathbb{P})\mathcal{L}(\tau)\mathrm{d}\tau} = e^{\int_s^t (1-\mathbb{P})(\mathcal{L}_0 + \varepsilon\delta(\tau)\mathcal{L}_1)\mathrm{d}\tau} = e^{(t-s)(1-\mathbb{P})\mathcal{L}_0 + \varepsilon(1-\mathbb{P})\mathcal{L}_1}. \tag{2.117}$$

This way, Eq. (2.107) becomes:

$$\partial_t R_\Psi(t) = B(t) R_\Psi(t) + \int_0^t K(t-s) R_\Psi(s) \mathrm{d}s. \tag{2.118}$$

As before, we take Fourier transforms on the previous equation to get:

$$\omega \mathfrak{F}\left[R_\Psi\right](\omega) - R_\Psi(0) = \varepsilon q R_\Psi(0) + \mathfrak{F}\left[K\right](\omega) \mathfrak{F}\left[R_\Psi\right](\omega). \tag{2.119}$$

Taking $R_\Psi(0) = \xi$ and rearranging,

$$\mathfrak{F}\left[R_\Psi\right](\omega) = \frac{\xi\left(1+\varepsilon q\right)}{\omega - \mathfrak{F}\left[K\right](\omega)}. \tag{2.120}$$

As with Eq. (2.115), it is possible to relate the kernel K with the kth order responses provided that ε lies in the interval of the power series convergence.

2.5 Summary and Discussion

In this chapter we have derived response formulas to describe a system undergoing external perturbations. The theory of the transfer and Koopman operators together with their stochastic analogues via the Fokker-Planck and backward-Kolmogorov equation, has been shown to provide a methodology which yields the linear and non-linear responses in agreement with previous works in this area.

Equation (2.23) contains the main perturbative formula for the expectation values of observables in a system subject to static forcing. Time-dependence is here uniquely due to applying variation-of-parameters in Eq. (2.16) and, thus, such response formula has to be understood as transient. After convergence to a steady state, the stationary response is given in Eq. (2.26) and the rate of convergence is determined by the point spectrum of the generator; see Eq. (2.45). This is in line with recent advances in decomposing correlation functions in terms of the point-spectrum of quasi-compact operators [Tan+20]. Taking the advantage of the static character of the perturbation, Laplace transforms allowed to link the theory of perturbed resolvent operators with the frequency representation of the response in Eq. (2.49a).

When the applied fields are time-modulated, physical constraints in the response functions are imposed. Indeed, the enforcement of causality allows to express the linear response as a convolution of the associated Green function and the time-modulation as shown in Eq. (2.58). More importantly, here we observed that the point spectrum of the underlying quasi-compact operator serves to decompose both the Green function and dynamic susceptibility in terms of eigenfunctions: see Eq. (2.64) and Eq. (2.68), respectively.

A particular case of inhomogeneous forcing is considered in Section 2.3.1, where noise is used to perturb a deterministic system. The operator expansion in Eq. (2.26) was seamlessly employed in the backward-Kolmogorov equation associated with Eq. (2.69) to find Eq. (2.74), which contains the leading order response and happens to be proportional to ε^2. If, on the other hand, one treats the Wiener increments in Eq. (2.69) as functions obeying the usual rules of calculus, one obtains a response formula which has to be averaged over all possible realisations of the noise and gives a different result (2.77a), first obtained in [Luc12]. Because of treating the noise as a usual function, one is implicitly led to the Stratonovich stochastic calculus [Pav14, Chapter 3]. In Proposition 2.3.1, we precisely show that the difference between Eq. (2.74) and Eq. (2.77a) is exactly given by the Itô-to-Stratonovich correction.

The second-to-last section takes an alternative angle to the problem of response. In Section 2.4 we projected the Fokker-Planck equation to write a scalar and homogenous equation for the response, first deduced in [Ken71] for quantum systems, which did not resort to finding the perturbed density function at the price of introducing memory; see

Eq. (2.107). The existence of solutions of Eq. (2.107) are tied to the decay properties of the memory kernel $K(t)$, which possesses an exponential of a projected operator. In views of the spectral decomposition of the response function presented in Section 2.2.1, one is inclined to think that convergence of the integral term in Eq. (2.107) as t tends to infinity is determined by the spectral gap in the projected generators, which are not investigated here. It was also observed that viewing Eq. (2.107) in frequency domain allows to derive a non-perturbative formula for the response in the case of step-function (Eq. (2.114)) and impulse perturbations (Eq. (2.120)). The Fourier transform of the memory kernel can, furthermore, be related in the latter cases with the high-order responses and, thus, with the generalised Green functions (see Eq. (2.115)). The latter is rather surprising since the memory kernel on the LHS of Eq. (2.115) involves convoluted projected differential operators whereas the generalised Green functions do not.

Chapter 3

Response and Sensitivity Using Markov Chains

In the Introduction, we mentioned the need of elaborating methodologies that allowed to predict statistical changes of the climate system as a result of applying an external field, without resorting long numerical integrations of the model in question. Towards this goal, the linear response theory for stochastic systems surveyed in Chapter 2 come to our aid since the resulting changes in the statistics are expressed in terms of the unforced evolution operators. Constructing the response operator of a physical system, however, is a difficult and costly task, which heavily depends on the nature of its dynamics and the number of degrees of freedom. As a consequence of such computational expense, developing numerically tractable methods are required. In this chapter we demonstrate that it is possible to calculate, by using finite representations of the transfer operator, the response of a system by sampling its unperturbed dynamics and having prior knowledge of the forcings applied to it. The overall goal is to provide practically usable tools for studying the response of complex non-equilibrium climate-like systems to perturbations.

Obtaining discrete representations of dynamical invariants is a pillar towards a better understanding of a physical system. As for the invariant measure and other steady state statistics, Ulam's conjecture claims that a finer spatial resolution yields more accurate computations [Ula64]. Unfortunately, this conjecture has only been proved positively for very few deterministic dynamical systems; see e.g. [Li76]; [Fro98], although more details and references will be given in the main body of this chapter. Along these lines, Ulam-type finite projections of the transfer operator or Fokker-plank semigroup are, nevertheless, susceptible of providing satisfactory estimates of invariant measures, correlations and responses up to finite precision.

Projected transfer operators are realised to have Markov-matrix structures whose spectral features are linked to dynamical properties of the underlying system. Indeed, the already mentioned rough-parameter dependence of an intermediate complexity ENSO

model is characterised by small spectral gaps in these matrices[1]. Moreover, almost invariant sets on phase space arise from the eigenvectors of such matrices and can be associated with Lagrangian coherent structures in the ocean [Fro+07] as well as midlatitude atmospheric blocking events [TBD15]. Furthermore, when a system is undergoing perturbations, Markov matrices serve to give estimates of the statistical sensitivity of the model [Sen93]; [Luc16] and can reveal its approach to a bifurcation point [TLD18]; [Tan+18]. It remains to show, though, whether information of the unforced attractor can be used to entirely calculate the linear response of the system given prior knowledge of the applied field.

The plan for this chapter is to delve on the finite representation of the operator semigroup techniques exposed in the previous chapter and report the results found in [SGL20]. We shall depart from the deterministic case of the transfer operator introduced in Section 2.1 and observe that an extension to stochastic systems is straightforward. Projecting the transfer operator onto a finite dimensional basis functions naturally leads to matrix representations that inherit a Markovian structure, as is shown in Section 3.1. By working at the level of abstract yet finite Markov chains, we are able, in Section 3.2, to develop the perturbation theory of stochastic matrices that allows for defining the concept of linear response for this type of finite phase space processes and thus, for coarse-grained transfer operators. This is accompanied by technical counterpart in Section 3.2.1, which addresses the algebraic well-posedness of the problem. The numerical conditioning of stochastic matrices towards calculating the linear response is investigated in Section 3.2.1.1. Numerical results are gathered in Section 3.3 which involve the study of two simple yet meaningful dynamical processes. In Section 3.3.1 a two-dimensional Ornstein-Uhlenbeck (O-U) process [UO30] is analysed, whereas Section 3.3.2 deals with the celebrated Lorenz 63 system [Lor63] in the hope of providing the transfer operator point of view to prior investigations in determining its statistical sensitivity.

3.1 Projected Transfer Operators and Markov Chains

We consider a finite subdivision of the d-dimensional phase space $\mathcal{X} \subseteq \mathbb{R}^d$ into N Lebesgue-measurable and non-intersecting subsets or *boxes* $\{B_i\}_{i=1}^N$ and define $\mathbf{1}_{B_i}$ as the characteristic function on box $B_i \subset \mathcal{X}$. The vector space spanned by this set of functions we denote as U_N and it is, by construction, N-dimensional. In case that $\mathcal{X}$ is unbounded, some of the boxes must also be unbounded in order to cover $\mathcal{X}$, but this is fine as long as the chosen collection is finite. Thus, we define the projection

[1]The meaning of spectral gap here is to a large extent the same as defined in the paragraph below Eq. (2.42)

$P_N : L^1_\eta(\mathcal{X}) \longrightarrow U_N = \mathrm{Span}\left(\{\mathbf{1}_{B_i}\}_{i=1}^N\right) \subset L^1_\eta(\mathcal{X})$ as

$$P_N\rho = \sum_{i=1}^{N} \frac{\mathbf{1}_{B_i}}{\eta(B_i)} \int_{B_i} \rho(\mathbf{x})\mathbf{1}_{B_i}(\mathbf{x})\eta(\mathrm{d}\mathbf{x}), \tag{3.1}$$

where η indicates some probability measure that has to be chosen depending on the problem at hand; this shall be discussed in more detail later in this chapter. The choice of functional space follows from the domain of the transfer operator $\mathcal{P}_t$ defined in Eq. (2.1), although it is susceptible of being changed accordingly with the nature of the flow. It follows that any projected function $P_N\rho$ can be represented in vector form $\mathbf{u} = [u_1, \ldots, u_N]^\top$ in $\mathbb{R}^N$ with respect to the basis of U_N, where:

$$u_i = \frac{1}{\eta(B_i)} \int_{B_i} \rho(\mathbf{x})\mathbf{1}_{B_i}(\mathbf{x})\eta(\mathrm{d}\mathbf{x}), \tag{3.2}$$

for every $i = 1, \ldots, N$. In an analogous manner, the transfer operator $\mathcal{P}_t$ introduced in Eq. (2.1) can be projected to obtain matrix approximants. It was shown in Section 2.1 that the operator $\mathcal{P}_t$ is the exponential of a generator $\mathcal{L}$ defined as a strong limit in Eq. (2.9), and such notation will be used from now on. The projected exponential $P_N e^{t\mathcal{L}} : U_N \longrightarrow U_N$, hence, admits a matrix representation $\mathcal{M}^t$, where each element is given by:

$$\mathcal{M}^t_{i,j} := \frac{1}{\eta(B_i)} \int_{B_i} e^{t\mathcal{L}}\mathbf{1}_{B_j}(\mathbf{x})\eta(\mathrm{d}\mathbf{x}), \tag{3.3}$$

with $i, j = 1, \ldots, N$. This matrix describes the proportion of box B_i that will end up in box B_j after t time units. To a large extent this matrix determines the probability of jumping onto B_j conditional on being somewhere distributed in B_i according to η. In general, matrices that describe transition probabilities between states are called *Markov* or *stochastic* [Sen73]:

Definition 3.1.1. *Let N be a positive integer. The matrix $\mathcal{M}$ in $\mathbb{R}^{N\times N}$ is* Markov *or* stochastic *if the following two conditions are met:*

(i) $\mathcal{M}_{i,j} \geq 0$ *and,*

(ii) $\sum_{i=1}^N \mathcal{M}_{i,j} = 1$

for every $i, j = 1, \ldots, N$.

With this definition, it is claimed that $\mathcal{M}^t$ is a stochastic matrix for every $t \geq 0$. We, therefore, have to check conditions *(i)* and *(ii)* on the entries of $\mathcal{M}^t$, stated in Definition 3.1.1. Firstly, since the characteristic functions are positive and η is a probability measure, it is immediate that $\mathcal{M}^t_{i,j} \geq 0$, for every $i, j = 1, \ldots, N$. Secondly, we apply the

definition of the transfer operator (2.1) to obtain the following chain of equalities:

$$\sum_{i=1}^{N} \mathcal{M}_{i,j}^{t} = \sum_{i=1}^{N} \frac{1}{\eta(B_i)} \int_{B_i} e^{t\mathcal{L}} \mathbf{1}_{B_j}(\mathbf{x}) \eta(\mathrm{d}\mathbf{x}) \tag{3.4a}$$

$$= \sum_{i=1}^{N} \frac{1}{\eta(B_i)} \int_{\phi^{-t}(B_i)} \mathbf{1}_{B_j}(\mathbf{x}) \eta(\mathrm{d}\mathbf{x}) \tag{3.4b}$$

$$= \sum_{i=1}^{N} \frac{1}{\eta(B_i)} \int_{\phi^{-t}(B_i)\cap B_j} \eta(\mathrm{d}\mathbf{x}) \tag{3.4c}$$

$$= \sum_{i=1}^{N} \frac{\eta\left(\phi^{-t}\left(B_i\right) \cap B_j\right)}{\eta\left(B_i\right)} = 1, \tag{3.4d}$$

which is true for every $j = 1, \ldots, N$ and every value of t. Consequently, the matrix defined in Eq. (3.3) is stochastic.

It is worth highlighting that the properties of the flow ϕ^t are reflected through $e^{t\mathcal{L}}$ and it raises the question of whether such properties are preserved when projecting $e^{t\mathcal{L}}$ onto U_N. First of all, notice that because of the projection, the semigroup property is lost; see Eq. (2.2). While the transfer operator satisfies $e^{t\mathcal{L}} e^{s\mathcal{L}} = e^{(t+s)\mathcal{L}}$— which results from the underlying autonomous system—, the matrix representation $\mathcal{M}^t$ does not satisfy the semigroup relation in general, leading to a inhomogeneous family of Markov chains $\{\mathcal{M}^t\}_{t\geq 0}$ whose the spectral elements (eigenvalues and eigenvectors) might be time-dependent. This can also be seen as a consequence of projecting the evolution of probability density functions which entail the appearance of memory terms, by virtue of the Mori-Zwanzig formalism; cf. Eq. (2.101) or [Zwa01]. In practise, the projected transfer operator $\mathcal{M}^\tau$ is calculated once for a value of $\tau > 0$ so that the semigroup property is almost held for $n\tau$ time units for every integer $n > 1$:

$$\mathcal{M}^{n\tau} \approx \overbrace{\mathcal{M}^\tau \cdot \ldots \cdot \mathcal{M}^\tau}^{n} = \left(\mathcal{M}^\tau\right)^n . \tag{3.5}$$

Hence, a suitable choice of τ allows to minimise the loss of semigroup property and define the homogeneous discrete time Markov chain $\{(\mathcal{M}^\tau)^n\}_{n=1}^{\infty}$, where the projection of the transfer operator is performed only once. From now on, $\mathcal{M}^{n\tau}$ shall equally denote the projection of $e^{n\tau\mathcal{L}}$ and nth power of $\mathcal{M}^\tau$. In general, a good value for τ depends on the size of the boxes used in the partition. Indeed, coarse partitions are examined with larger values of τ, so that points of phase space have more chances of jumping to the next box;[FR02]; [TLD18]. Consequently, the choice for τ becomes crucial and will be studied in Section 3.3.2 with a worked example.

Alongside the semigroup property, the concepts of ergodicity and mixing of a dynamical system $\{\phi^t\}_{t\geq 0}$ defined in the Introduction can also be extended to Markov chains by means of analysing the structure of the stochastic matrices defining them. Positive elements in

the stochastic matrix $\mathcal{M}^\tau$ indicate that after τ time units, a fraction of B_j will jump into B_i and, in this sense, the two states B_i and B_j are connected. Furthermore, $\mathcal{M}^{2\tau}$ will also be a stochastic matrix and its positive elements indicate connected states after 2τ time units. Hence, one expects that, if the underlying system is ergodic or mixing, the entries of the stochastic matrix $\mathcal{M}^t$ will eventually become positive. Indeed, in the theory of Markov matrices, and more generally of positive matrices, ergodicity and mixing can be suitably defined in terms of their algebraic structure and non-negativity of matrix powers. Indeed, a stochastic matrix $\mathcal{M}$ is *primitive* if there exists a positive integer n such that $\mathcal{M}^n$ has strictly positive elements [Sen73]. Such matrices will, hence, define Markov processes where states are visited with uniformly non-vanishing probability after enough iterations of the chain. These matrices are also known as *irreducible and aperiodic* [Bal00]; [Sen73, §1]. For this class of matrices, the celebrated Perron-Frobenius theorem holds in the form that follows:

Theorem 3.1.1 (Perron-Frobenius)**.** *Let N be a positive integer and $\mathcal{M}$ an $N \times N$ primitive stochastic matrix. Then,*

(i) 1 *is in* $\sigma(\mathcal{M})$

(ii) 1 *is simple*

(iii) if $\mathcal{M}\mathbf{u} = \mathbf{u}$, $\mathbf{u}$ *is strictly positive or negative*

(iv) $1 > |\lambda_j|$, *for all* λ_j *in* $\sigma(\mathcal{M}) \setminus \{1\}$

This theorem is central in the theory of Markov chains since it guarantees the existence of an invariant probability vector, this is, an eigenvector $\mathbf{u} = [u_1, \ldots, u_N]^\top$ associated with the eigenvalue 1 that is strictly positive and such that $\sum_{i=1}^N u_i = 1$. Moreover, the simplicity of such an eigenvalue implies that there are no subsets of states that are invariant under the action of $\mathcal{M}$, reminiscing of ergodicity. Finally, the Perron-Frobenius theorem also implies that the spectrum of a primitive stochastic matrix is within the complex unit ball, in views of item *(iv)* above.

With regards to the projected transfer operator defined in Eq. (3.3), one would like to obtain finite dimensional vector approximations of the invariant measure that can be used to calculate the expectation values of the coarse-grained observables. To this end, and assuming that the resulting matrix is primitive, we exploit Perron-Frobenius theorem to find the positive eigenvector $\mathbf{u}$ associated with the largest eigenvalue 1:

$$\mathcal{M}^\tau \mathbf{u}(\tau) = \mathbf{u}(\tau), \tag{3.6}$$

for some value of τ in $\mathbb{R}$ that minimises the loss of semigroup property in the sense of Eq. (3.5). Notice that because of this assumption, although the family $\{\mathcal{M}^\tau\}_{\tau \in \mathbb{R}}$ is

technically inhomogeneous, $\mathbf{u}$ has a weak τ-dependence to the extent that $\mathbf{u}(\tau) \approx \mathbf{u}(n\tau)$ for every natural n. Moreover, by construction, it follows that $\mathbf{u}$ is exactly invariant for every stochastic matrix in $\{\mathcal{M}^{n\tau}\}_{n=1}^{\infty}$. Hence, from now on we shall drop the τ-dependence notation in Eq. (3.6).

With the homogeneous chain $\{\mathcal{M}^{n\tau}\}_{n=1}^{\infty}$ at hand, convergence to the steady state $\mathbf{u}$ is measured by the rate at which $\|\mathcal{M}^{n\tau}\mathbf{u}_0 - \mathbf{u}\|$ vanishes for any initial ensemble $\mathbf{u}_0$ in $\mathbb{R}^N$. Since we are assuming that $\mathcal{M}^{\tau}$ is mixing, the second largest eigenvalue in modulus λ_2 of $\mathcal{M}^{\tau}$ lies strictly inside the unit ball and regulates the speed at which ensembles relax towards a steady state [Sen73]. In this sense, λ_2 is related to the spectral gap associated with a point spectrum in the generator of the transfer operator or Fokker-Planck semigroup discussed in the paragraph below Eq. (2.42).

Remark 3.1.1. *Projected transfer operator are, effectively, a way of visualising the coarse-grained dynamics of the flow, although it is not here proved whether they constitute a rigorous approximation in a suitable norm. More concretely, a projection P_N is said to be a rigorous approximation of $e^{t\mathcal{L}}$ in the operator norm if*

$$\lim_{N\to\infty} \left\| P_N e^{t\mathcal{L}} - e^{t\mathcal{L}} \right\| = 0. \tag{3.7}$$

This is the content of Ulam's conjecture [Ula64], which states that finer partitions lead to more accurate approximations of dynamical invariants such as the invariant measure or the rate of mixing. Regarding the invariant measure estimated from the eigenvalue problem (3.6)*, such conjecture has turned out to be positive for twice differentiable expanding one-dimensional maps in the L^1 norm [Li76]. Immediate extensions to higher dimensions require the existence of a physical and absolutely continuous invariant measure where partitions of phase space can be taken rather arbitrarily in order to converge [Fro98]. It has been proved, however, that Ulam-type estimations do work as well for uniformly hyperbolic dynamical systems which posses a singular invariant measure, although partitions are required to be Markov-partitions — see [Fro99] for definition— that allow to regard the dynamical system as subshifts of finite type [Bal00].*

In the same way that the eigenvector associated with the largest eigenvalue 1 gives us a first guess for the discretised invariant measure, we can consider approximating the expectation value of an observable Ψ. To this end, we can introduce an inner product $\langle \cdot, \cdot \rangle_{P_N}$, so that if ρ and Ψ are square integrable, we have

$$\langle \rho, \Psi \rangle_{P_N} := \langle P_N\rho, P_N\Psi \rangle = \sum_{i=1}^{N} [P_N\rho]_i [P_N\Psi]_i, \tag{3.8}$$

where $\langle \cdot, \cdot \rangle$ refers to the L^2-inner product and $[P_N\rho]_i$ and $[P_N\Psi]_i$ are the coordinates with respect to the basis of U_N of ρ and Ψ respectively, as determined by Eq. (3.2). Notice

that if η is the invariant measure and possesses a density ρ_0 with respect to Lebesgue, then $\langle \rho_0, \Psi \rangle_{P_N}$ provides a coarse-grained estimate of the expectation value of Ψ at the statistical steady state. Furthermore, since correlation functions can be expressed in terms of the transfer operator, see Eq. (2.11), its discretised counterparts via Ulam-projections should be the candidates for approximating them in finite precision. Letting Ψ and Φ be L^2-integrable, their projected correlation function is given by:

$$C^{P_N}_{\Psi,\Phi}(t) = \sum_{i=1}^{N}\sum_{j=1}^{N} \left[P_N\Psi\right]_i \mathcal{M}^t_{j,i} \left[P_N\rho_0\right]_j \left[P_N\Phi\right]_j - \langle \rho_0, \Psi \rangle_{P_N} \langle \rho_0, \Phi \rangle_{P_N} . \tag{3.9}$$

As before, projected correlations provide estimates of the actual correlation functions, although rigorous results on the limit of fine resolution are, to this date, very limited; see Remark 3.1.1.

In previous chapters we have observed that the invertibility of the flow together with the invariance of the measure μ makes the transfer operator semigroup unitary in $L^2_\mu(\mathcal{X})$; see Theorem 2.1.1 and the comments below. Consequently, the lack of non-unital eigenvalues in such operator semigroup cannot explain the decay of correlations, which is characteristic of chaotic and mixing systems. It was observed that more general Banach spaces are needed to identify the resonances within the unit ball responsible of making correlations decay [BL07]. In this chapter, though, the transfer operator semigroup is projected by means of Eq. (3.3). Such finite truncation introduces artificial irreversibility to the dynamics, and will reflect on the non-unitarity of the resulting approximation $\mathcal{M}^\tau$, but also seen in the Perron-Frobenius theorem 3.1.1. One way of viewing this is to regard Ulam-partitions as an upwind finite-difference scheme that artificially introduces diffusion and, consequently, the approximated transfer operator and other invariant features correspond to that of a stochastically forced system [FJK13]. On physical grounds, the geometric complexity and self-similarity of deterministic attractors that support the invariant measure prevents that an increase of resolution results in unitary transition matrices $\mathcal{M}^\tau$ that, in turn, yield eigenvalues that converge strictly within the complex unit ball [FR02].

3.2 Perturbations of Finite Markov Chains

Applied fields lead to perturbations on the transfer operator $e^{\tau\mathcal{L}}$ and, consequently on the matrix $\mathcal{M}^\tau$. This fact motivates the problem of reverting the question i.e., by considering perturbations of $\mathcal{M}^\tau$, can we estimate the response. Because rigorous and general results regarding the approximation of transfer operators do not exist— see Remark 3.1.1—, we cannot guarantee whether the qualitative features like the existence of a spectral gap, the regularity of the invariant measure or the correlation functions will be preserved under projections. In this section we shall, nevertheless, develop the theory of response for

finite Markov chains at an abstract level following [Sch68]; [Luc16]; [ADF18], so that the matrix $\mathcal{M}^\tau$ can be employed to estimate the response and sensitivity of the underlying system.

Consider a mixing Markov chain with a finite number of states N in $\mathbb{N}$. The transition probability from state to state is determined by a matrix $\mathcal{M}$ in $\mathbb{R}^{N\times N}$ which, without loss of generality, determines the process itself. In other words, a positive vector $\mathbf{u}_0$ in $\mathbb{R}^N$ with $\sum_{i=1}^N u_{0,i} = 1$ would indicate an initial ensemble of states and the sequence $\{\mathcal{M}^n\mathbf{u}_0\}_{n=0}^\infty$ would be a realisation of the Markov process and the probability of transitioning from ith state conditioned on being in the jth is given by $\mathcal{M}_{i,j}$. Positivity and conservation of probability ensure that $\mathcal{M}$ is a stochastic matrix, satisfying conditions *(i)* and *(ii)* of Definition 3.1.1. Furthermore, the process is assumed to be mixing, implying that the matrix $\mathcal{M}$ is primitive, for which the Perron-Frobenius theorem 3.1.1 holds.

In order to obtain a response formula for Markov chains, it is necessary to introduce external perturbations that do not compromise the Markovian structure of the defining matrices. This way, working at the level of stochastic matrices, we consider a perturbation of the matrix $\mathcal{M}$ of the form:

$$\mathcal{M} \longrightarrow \mathcal{M} + \sum_{k=1}^{n} \varepsilon_k m_k, \tag{3.10}$$

where $m_1, \ldots, m_n$ are $N \times N$ matrices and $\varepsilon_1, \ldots \varepsilon_n$ are real parameters. The matrices m_k are what we will call the *perturbation matrices* which model different physical processes involved in changing the nature of the matrix $\mathcal{M}$. This way of introducing perturbations allows to consider general patterns of forcing with, if time-dependence was allowed, different time modulations; see, for instance, Eq. (2.75) or a more general discussion in [Zwa01]. Another reason for this is to investigate whether responses to different perturbations can be seamlessly added so that a superposition principle holds; we shall observe that by the nature of the problem taken here, $n = 1$ in Eq. (3.10) will suffice to determine the well-posedness of the problem; but this is going to be treated in Section 3.2.1.

The matrix on the RHS of Eq. (3.10) must be stochastic in order to describe Markov chain, this amounts to enforcing certain structure on the perturbation matrices m_k so that Definition 3.1.1 is satisfied. This requirement corresponds to having:

$$\sum_{i=1}^{N} [m_k]_{i,j} = 0, \tag{3.11}$$

for any k in $\{1, \ldots, n\}$ and j in $\{1, \ldots, N\}$. This assures that the columns of $\mathcal{M} + \sum_{k=1}^n \varepsilon_k m_k$ add up to one so that total probability adds up to 1. Moreover, non-negativity

must be preserved, so not all choices of ε_k are valid. For this, we define

$$\varepsilon_k^- = \min\left\{\varepsilon \in \mathbb{R} : \forall i,j \in \{1,\dots,N\}, \mathcal{M}_{i,j} + \varepsilon\,[m_k]_{i,j} \geq 0\right\}, \tag{3.12}$$

and

$$\varepsilon_k^+ = \max\left\{\varepsilon \in \mathbb{R} : \forall i,j \in \{1,\dots,N\}, \mathcal{M}_{i,j} + \varepsilon\,[m_k]_{i,j} \geq 0\right\}. \tag{3.13}$$

Hence, to ensure non-negativity of the perturbed Markov process, we must have that $\max_k \varepsilon_k^- \leq \varepsilon_k \leq \min_k \varepsilon_k^+$. To guarantee that the latter interval is non-empty, we need to be certain that $\varepsilon_k^- < 0$ and $\varepsilon_k^+ > 0$, for all k. This would not happen supposing that for i_1, j_1, i_2 and j_2 in $\{1,\dots,N\}$ we have that $\mathcal{M}_{i_1,j_1} = \mathcal{M}_{i_2,j_2} = 0$ and $[m_k]_{i_1,j_1}, [m_k]_{i_2,j_2} < 0$. This would imply that $\varepsilon_k^- = \varepsilon_k^+ = 0$. In such case, we say that the matrices m_k are non-admissible perturbations.

Before deriving the response formulas, we need to make sure that the perturbed Markov chain preserves the mixing property, so that its leading unit eigenvalue is simple and there are no other eigenvalues on the unit ball. To tackle this issue, we refer to the continuity of the matrix-to-eigenvalue mapping, which allows to take sufficiently small non-zero values of ε_k so that the subdominant eigenvalues of $\mathcal{M}$ remain inside the complex unit ball after the perturbation is applied [Wil65]; [TE05]; [ADF18]. In this setting, by virtue of the Perron-Frobenius theorem, the perturbed stochastic matrix will have a dominant simple eigenvalue, whose value is 1 and its associated eigenvector $\mathbf{v} = \mathbf{v}(\varepsilon_1,\dots,\varepsilon_n)$ can be normalised and strictly positive. In other words, $\mathbf{v}$ solves:

$$\left(\mathcal{M} + \sum_{k=1}^{n} \varepsilon_k m_k\right)\mathbf{v} = \mathbf{v}. \tag{3.14}$$

The goal is to express the perturbed invariant measure $\mathbf{v}$ in terms of $\mathcal{M}, m_1, \dots, m_n, \varepsilon_1, \dots$ and ε_n. Not only do we want to calculate $\mathbf{v}$ but also describe how the unperturbed measure $\mathbf{u}$ responds at a given power of ε_k. In particular, we shall extract the linear response, which is specially relevant in the physical literature. Using multiindex notation, we suppose a formal expansion in powers of $\varepsilon_1,\dots,\varepsilon_n$:

$$\mathbf{v} = \mathbf{u} + \sum_{|\alpha|=1}^{\infty} (\varepsilon_1,\dots,\varepsilon_n)^\alpha \mathbf{w}_\alpha, \tag{3.15}$$

where $\mathbf{w}_\alpha = \frac{1}{\alpha!}\left(\partial_{\varepsilon_1},\dots,\partial_{\varepsilon_n}\right)^\alpha \mathbf{v}$. The series in Eq. (3.15) contains the response at all orders of ε_k, in particular, for $|\alpha| = 1$, we get the linear response. Moreover, this is saying that the perturbations can be treated independently and added to give the response at leading order of the parameter ε_k, for every value of k. Substituting the expression

(3.15) in Eq. (3.14) we obtain:

$$\left(\mathcal{M}+\sum_{k=1}^{n}\varepsilon_k m_k\right)\left(\mathbf{u}+\sum_{|\alpha|=1}^{\infty}(\varepsilon_1,\ldots,\varepsilon_n)^{\alpha}\mathbf{w}_{\alpha}\right)=\mathbf{u}+\sum_{|\alpha|=1}^{\infty}(\varepsilon_1,\ldots,\varepsilon_n)^{\alpha}\mathbf{w}_{\alpha}. \quad (3.16)$$

Gathering the terms for $|\alpha|=1$ we get:

$$\mathcal{O}(\varepsilon_k):(1-\mathcal{M})\,\partial_{\varepsilon_k}\mathbf{v}=m_k\mathbf{u}, \quad (3.17)$$

for $k=1,\ldots,n$. Notice that the matrix $1-\mathcal{M}$ cannot be inverted since 1 is an eigenvalue of $\mathcal{M}$; we shall discuss this issue in the next section. At the moment, we directly apply the inverted matrix to find that

$$\partial_{\varepsilon_k}\mathbf{v}=(1-\mathcal{M})^{-1}\,m_k\mathbf{u}. \quad (3.18)$$

We define $\mathcal{G}_k=(1-\mathcal{M})^{-1}\,m_k$ as *linear response matrix* which has also been named a *differential* matrix [Sch68]. The choice of notation for the linear response matrix is not out of coincidence since it plays the role of the Green function discussed in Section 2.3. If we repeat the process for the second order terms ($|\alpha|=2$) we get:

$$\mathcal{O}\left(\varepsilon_k^2\right):\frac{1}{2!}\partial^2_{\varepsilon_k}\mathbf{v}=(1-\mathcal{M})^{-1}\,m_k\partial_{\varepsilon_k}\mathbf{v}=\left(\mathcal{G}_k\right)^2\mathbf{u} \quad (3.19a)$$

$$\mathcal{O}\left(\varepsilon_k\varepsilon_l\right):\partial^2_{\varepsilon_k,\varepsilon_l}\mathbf{v}=(1-\mathcal{M})^{-1}\,m_l\partial_{\varepsilon_k}\mathbf{v}+(1-\mathcal{M})^{-1}\,m_k\partial_{\varepsilon_l}\mathbf{v} \quad (3.19b)$$

$$=\mathcal{G}_k\mathcal{G}_l\mathbf{u}+\mathcal{G}_l\mathcal{G}_k\mathbf{u}. \quad (3.19c)$$

Thus, we inductively construct the whole expansion as:

$$\mathbf{v}=\mathbf{u}+\sum_{i=1}^{\infty}\left(\sum_{k=1}^{n}\varepsilon_k\mathcal{G}_k\right)^{i}\mathbf{u}=\left(1-\sum_{k=1}^{n}\varepsilon_k\mathcal{G}_k\right)^{-1}\mathbf{u}. \quad (3.20)$$

This formula provides a tool to calculate the perturbed invariant measure as long as it converges, in which case it should equal the vector found by solving the eigenvalue problem Eq. (3.14). The problem of convergence will be addressed in the next section.

Moreover, by simply considering the transpose matrices, it is possible to develop a response theory for observables, which are understood to be in duality with probability density functions. Indeed, letting Ψ in $\mathbb{R}^N$ represent a generic coarse-grained observable,

its expectation value with respect to the perturbed stationary vector $\mathbf{v}$ satisfies:

$$\langle \Psi, \mathbf{v} \rangle = \left\langle \Psi, \mathbf{u} + \sum_{i=1}^{\infty} \left(\sum_{k=1}^{n} \varepsilon_k \mathcal{G}_k \right)^i \mathbf{u} \right\rangle = \left\langle \left(1 + \sum_{i=1}^{\infty} \left(\sum_{k=1}^{n} \varepsilon_k \mathcal{G}_k \right)^i \right)^\top \Psi, \mathbf{u} \right\rangle \tag{3.21a}$$

$$= \left\langle 1 + \sum_{i=1}^{\infty} \left(\sum_{k=1}^{n} \varepsilon_k \mathcal{G}_k^\top \right)^i \Psi, \mathbf{u} \right\rangle, \tag{3.21b}$$

where $\langle \cdot, \cdot \rangle$ denotes the usual inner product in $\mathbb{R}^N$, although it is, in fact, a finite dimensional interpretation of the standard pairing between observables and probability density functions. The advantage of formulas (3.20) and (3.21) is that they allow us to identify the response to perturbations at an arbitrary order of non-linearity including the linear case which encodes the sensitivity of the system to forcing, in line of the linear response formulas derived in Chapter 2. In fact, Eq. (3.21) is the finite dimensional counterpart of Eq. (2.23). These matrix relations are, however, still purely formal. We need a deeper understanding of what we mean with $(1 - \mathcal{M})^{-1}$ and for what values of ε_k they are useful.

3.2.1 Well-posedness and Invertibility of $1 - \mathcal{M}$

In this section we will study the necessary conditions for the response formulas presented above to be useful. Generally, we will revisit the problem of the non-invertibility of the linear response matrix and specifically, address the items listed below:

(i) Identify when the matrix $1 - \mathcal{M}$ can be inverted

(ii) Find the radius of convergence of Eq. (3.20)

(iii) Assess the numerical conditioning of $1 - \mathcal{M}$

Item (i) was indicated in the previous section when constructing the linear response matrix. Item (ii) will boil down to calculating the interval of real values of ε_k for which the series in Eq. (3.20) converges. Finally, item (iii) concerns the numerical inversion of the matrix $1 - \mathcal{M}$ and, ultimately, the computation of the linear response matrix: we shall estimate the numerical conditioning (in a sense explained later) in terms of the eigenvalues of the matrix $\mathcal{M}$. To tackle these questions we will reduce the number of perturbation matrices on the RHS of Eq. (3.10) to only one (i.e., with $n = 1$), without compromising the generality of the results.

As mentioned earlier, the linear response matrix is not well defined *a priori* because 1 is an eigenvalue of the matrix $\mathcal{M}$ making $1 - \mathcal{M}$ is not invertible. However, we can define a more suitable normed space for which the norm of $\mathcal{M}$ is less than one, making $1 - \mathcal{M}$ invertible. The idea relies on the fact that $\mathbb{R}^N$, on which $\mathcal{M}$ is defined, admits a splitting of the form:

$$\mathbb{R}^N = \operatorname{span}\{\mathbf{u}\} \oplus \mathcal{V}, \tag{3.22}$$

where $\mathcal{V}$ is the invariant subspace generated by the generalised eigenvectors of $\mathcal{M}$ associated with the eigenvalues distinct to 1 and $\mathrm{span}\{\mathbf{u}\}$ is the vector subspace spanned by $\mathbf{u}$. This space can also be regarded as the kernel of the functional ι defined as

$$\begin{aligned} \iota : \mathbb{R}^N &\longrightarrow \mathbb{R}, \\ \mathbf{x} &\longmapsto \iota(\mathbf{x}) = \sum_{i=1}^{N} x_i. \end{aligned} \tag{3.23}$$

Indeed, one observes that if an N-vector $\mathbf{u}_j$ is a generalised eigenvector of $\mathcal{M}$ associated with the eigenvalue λ_j, the identity $(\mathcal{M} - \lambda_j)^{a_j} \mathbf{u}_j = \mathbf{0}$ holds for some positive integer a_j, namely, the algebraic multiplicity. Hence,

$$\iota\left((\mathcal{M} - \lambda_j)^{a_j} \mathbf{u}_j\right) = \sum_{k=0}^{a_j} \binom{a_j}{k} \lambda_j^k \iota\left(\mathcal{M}^{a_j - k} \mathbf{u}_j\right) = (1 - \lambda_j)^{a_j} \iota(\mathbf{u}_j) = 0. \tag{3.24}$$

Consequently, since the eigenvalue 1 is simple and the rest of eigenvalues have norm strictly less than one, we have that $\iota(\mathbf{u}_j) = 0$. Now, turning to the perturbation problem Eq. (3.10), if $\mathcal{M}$ and $\mathcal{M} + \varepsilon m$ is are stochastic matrices, it follows that $\iota(m\mathbf{x}) = 0$ for any $\mathbf{x}$ in $\mathbb{R}^N$. Moreover, 1 is no longer an eigenvalue of $\mathcal{M}$ if this matrix' action is restricted to $\mathcal{V}$, making $1 - \mathcal{M}$ invertible. The discussion in the paragraphs above can be summarised in the following statement:

Proposition 3.2.1. *Let $\mathcal{M}$ and m be $N \times N$ matrices where $\mathcal{M}$ and $\mathcal{M} + m$ are both mixing stochastic matrices. In addition, let $\mathbf{u}$ denote the N dimensional invariant vector associated with $\mathcal{M}$. Then, using the splitting of $\mathbb{R}^N$ in Eq.* (3.22)*, $1 - \mathcal{M}$ is invertible on $\mathcal{V}$ and $m\mathbf{x}$ is in $\mathcal{V}$, for any $\mathbf{x}$ in $\mathbb{R}^N$.*

Recall that the linear response matrix in Eq. (3.18) only requires the evaluation of $(1 - \mathcal{M})^{-1}$ after having calculated $m_k\mathbf{u}$, so writing the inverse explicitly is not that big an abuse, by virtue of the previous proposition. We must underline, however, that it is not possible directly numerically invert $1 - \mathcal{M}$. To overcome this problem, we must deflate the matrix by removing the dependence on the dominant eigenspace. In other words, we have to find an $N \times N$ matrix $\mathcal{Q}$ such that $1 - \mathcal{M} + \mathcal{Q}$ is invertible. Such matrix is given by the spectral projector of $\mathcal{M}$ around the eigenvalue 1 [Kat66]; [TE05]:

$$\mathcal{Q} = \frac{1}{2\pi i} \int_{\Gamma(1,r)} (z - \mathcal{M})^{-1} \, \mathrm{d}z, \tag{3.25}$$

where $\Gamma(1, r)$ denotes a closed circle of radius r and centred at 1 in the complex plane. The radius r is sufficiently small so that the interior of $\Gamma(1, r)$ contains only the eigenvalue 1. Consequently, the resolvent operator that constitutes the integrand in Eq. (3.25) is well defined for all values of z on $\Gamma(1, r)$; the properties of the resolvent where mentioned in

Section 2.2 although the reader is referred, again, to [EN06]. Spectral projectors of general (possibly unbounded) operators can be difficult to evaluate, although matrices admit more handy expressions in terms of eigenvectors. Indeed, let $\mathbf{z}$ be a constant N dimensional real vector. Then, it follows that $\mathbf{z}$ is a left eigenvector relative to $\mathcal{M}$:

$$\mathbf{z}^\top \mathcal{M} = \mathbf{z}^\top. \tag{3.26}$$

With this definition and recalling that the vector $\mathbf{u}$ denotes the right eigenvector relative to the eigenvalue 1, spectral projector $\mathcal{Q}$ is now written as [TE05]:

$$\mathcal{Q} = \frac{\mathbf{u}\mathbf{z}^\top}{|\mathbf{u}^\top \mathbf{z}|}. \tag{3.27}$$

Furthermore, if the eigenvectors $\mathbf{u}$ and $\mathbf{z}$ are normalised so that $|\mathbf{u}^\top \mathbf{z}| = 1$, the denominator in Eq. (3.27) equals one. It is immediate to check that $\mathcal{Q}$ in Eq. (3.27) is a (rank-one) projector, i.e. $\mathcal{Q}\mathcal{Q} = \mathcal{Q}$. Notice, in fact, that the matrix $\mathcal{Q}$ is also obtained by $\mathcal{Q} = \lim_{\tau\to\infty} \mathcal{M}^\tau$, in the matrix norm [Sen73].

We now have all the components to define a generalised inverse matrix of $1 - \mathcal{M}$ necessary to calculate the linear response matrix $\mathcal{G}$. To this end, we perform Neumann series to express the inverse of $1 - \mathcal{M}$:

$$\mathcal{G} = (1 - \mathcal{M})^{-1} m = \sum_{k=0}^{\infty} \mathcal{M}^k m = \sum_{k=0}^{\infty} (\mathcal{M} - \mathcal{Q})^k m = (1 - \mathcal{M} + \mathcal{Q})^{-1} m, \tag{3.28}$$

where we have exploited the fact that $\mathcal{Q}m = 0$ in the third equality. The second equality in Eq. (3.28) is justified when the matrix $\mathcal{M}$ acts on $\mathcal{V}$ (see Eq. (3.22)), where we would have that the spectral radius of $\mathcal{M}$ is strictly less than one. On the other hand, the fourth equality holds in general, since the spectral radius of the matrix $\mathcal{M} - \mathcal{Q}$ is less than one on $\mathbb{R}^N$. This comes specially convenient in computer calculations, although the numerical conditioning has to be assessed; this will be investigated in Section 3.2.1.1 below.

To tackle the convergence of the series in Eq. (3.20) (item (ii) on the list at the start of this section) we want to be certain that the induced L^1-norm of the series $\sum_{k=1}^{\infty} \varepsilon^k \mathcal{G}^k \mathbf{u}$ does not blow up. For this problem we introduce the matrix norm $\|\cdot\|_{1^*}$ which we define as the norm $\|\cdot\|_1$ restricted to $\mathcal{V}$:

$$\|A\|_{1^*} = \max_{\mathbf{x}\in\mathcal{V}} \frac{\|A\mathbf{x}\|_1}{\|\mathbf{x}\|_1} = \max_{\substack{\mathbf{x}\in\mathbb{R}^N \\ \iota(\mathbf{x})=0}} \frac{\|A\mathbf{x}\|_1}{\|\mathbf{x}\|_1}. \tag{3.29}$$

This matrix norm will allow us to estimate the linear response matrix $\mathcal{G}$ whose definition involves the inverse of $1 - \mathcal{M}$. We are now in conditions of applying the ratio test in

Eq. (3.20):

$$\frac{\left\|\varepsilon^{k+1}\mathcal{G}^{k+1}\mathbf{u}\right\|_1}{\left\|\varepsilon^k\mathcal{G}^k\mathbf{u}\right\|_1} = \frac{\left\|\varepsilon\mathcal{G}\varepsilon^k\mathcal{G}^k\mathbf{u}\right\|_1}{\left\|\varepsilon^k\mathcal{G}^k\mathbf{u}\right\|_1} = \frac{\left\|\varepsilon\left(1-\mathcal{M}\right)^{-1}m\varepsilon^k\mathcal{G}^k\mathbf{u}\right\|_1}{\left\|\varepsilon^k\mathcal{G}^k\mathbf{u}\right\|_1} \tag{3.30a}$$

$$= \frac{\left\|\varepsilon\left(1-\mathcal{M}\right)^{-1}m\varepsilon^k\mathcal{G}^k\mathbf{u}\right\|_{1^*}}{\left\|\varepsilon^k\mathcal{G}^k\mathbf{u}\right\|_1} \tag{3.30b}$$

$$\leq \frac{\varepsilon\left\|(1-\mathcal{M})^{-1}\right\|_{1^*}\left\|m\varepsilon^k\mathcal{G}^k\mathbf{u}\right\|_{1^*}}{\left\|\varepsilon^k\mathcal{G}^k\mathbf{u}\right\|_1} \tag{3.30c}$$

$$\leq \varepsilon\left\|(1-\mathcal{M})^{-1}\right\|_{1^*}\|m\|_1 \leq \varepsilon\left(1-\|\mathcal{M}\|_{1^*}\right)^{-1}\|m\|_1. \tag{3.30d}$$

Since we want that the ratio remains less than 1 to ensure absolute convergence, we choose ε so that

$$|\varepsilon| < \varepsilon_{max} := \frac{1-\|\mathcal{M}\|_{1^*}}{\|m\|_1}. \tag{3.31}$$

The real parameter ε_{max} determines a radius of expansion of the response formulas and, hence, a tolerance on the size of perturbations. We are still left, with showing that ε_{max} is non-zero. For that we shall analyse the number $\|\mathcal{M}\|_{1^*}$ which is commonly known as the *ergodicity coefficient* of the Markov chain $\mathcal{M}$ [Sen84]; [IS14]. Roughly speaking, $\|\mathcal{M}\|_{1^*}$ is an indicator of the mixing strength of $\mathcal{M}$, similarly to what the second largest eigenvalue λ_2 of $\mathcal{M}$ does. Contrary to λ_2, the ergodicity coefficient possesses an explicit formula in terms of the elements of the matrix $\mathcal{M}$ [Dob56]; [IS14]:

$$\|\mathcal{M}\|_{1^*} = \frac{1}{2}\max_{1\leq i,j\leq N}\|\mathcal{M}(\mathbf{e}_i-\mathbf{e}_j)\|_1 = \frac{1}{2}\max_{1\leq i,j\leq N}\sum_{k=1}^{N}|\mathcal{M}_{i,k}-\mathcal{M}_{j,k}|, \tag{3.32}$$

which does not involve solving eigenvalue problems. From Eq. (3.32) we observe that $\|\mathcal{M}\|_{1^*} < 1$ if any two columns of $\mathcal{M}$ share a positive element. For the class of mixing Markov chains considered here we have that $\|\mathcal{M}\|_{1^*} < 1$. Consequently, Eq. (3.31) implies that $\varepsilon_{max} > 0$. More importantly, the ergodicity coefficient constitutes a bound for the size of the non-unit eigenvalues of a stochastic matrix, giving an *a priori* bound on the rate of convergence to equilibrium: $|\lambda_2| \leq \|\mathcal{M}\|_{1^*}$. Moreover, the ergodicity coefficient enjoys the submultiplicative property and the second eigenvalue does not. This is critical in the study of convergence of inhomogeneous Markov chains determined by collection of stochastic matrices $\{\mathcal{M}^k\}_{k=1}^{\infty}$, where

$$\left\|\mathcal{M}^1\cdot\ldots\cdot\mathcal{M}^{k_0}\right\|_{1^*} \leq \left\|\mathcal{M}^1\right\|_{1^*}\cdot\ldots\cdot\left\|\mathcal{M}^{k_0}\right\|_{1^*}, \tag{3.33}$$

for any k_0 in $\mathbb{N}$. Hence, the ergodicity coefficient gauges the rate of convergence to equilibrium of a family of stochastic matrices.

The ergodicity coefficient can be regarded as a condition number for the invariant vector

$\mathbf{u}$ in the sense that it estimates the effects of applying a perturbation on general Markov matrices [Sen88]. Indeed, using similar arguments to the ones in Eqs. (3.30) we have:

$$\|\mathbf{v}-\mathbf{u}\|_1 \leq \frac{\varepsilon\|m\|_1}{1-\|\mathcal{M}\|_{1^*}}, \tag{3.34}$$

where $\mathbf{v}$ is the perturbed invariant vector given in Eq. (3.14). Taking the point of view of observables (see Eq. (3.21)), bounds on the expected value of a coarse-grained observable Ψ, can be immediately obtained. Along these lines, we refer the reader to [Inu19] for a practical use of the ergodicity coefficient, in the context of optical chaos.

The ergodicity coefficient is clear to provide a measure of the sensitivity of a Markov chain with respect to perturbations but it does not tell us how they affect the localisation of the eigenvalues of the matrix $\mathcal{M}$, something crucial if one wants to control the spectral gap between the leading eigenvalues. This spectral stability problem can be understood in terms of the algebraic conditioning of the matrix $\mathcal{M}$. For this purpose we introduce the condition number $\kappa_p(A)$ of a matrix A in $\mathbb{R}^{N\times N}$ with respect to the p-norm [Wil65, §4]:

$$\kappa_p\left(A\right) = \|A\|_p\|A^{-1}\|_p, \tag{3.35}$$

where in case A is singular, $\kappa_p(A)=\infty$. With this definition, the following stability result holds [Mit03]:

Proposition 3.2.2. *Let $\mathcal{M}$ be an $N\times N$ diagonalisable, primitive and stochastic matrix with eigenvalues $\{\lambda_j\}_{j=1}^N$ satisfying: $\lambda_1=1>|\lambda_2|\geq|\lambda_3|\geq\ldots\geq|\lambda_N|$. Let X be an $N\times N$ non-singular matrix such that $\mathcal{M}=X^{-1}\mathrm{diag}(\lambda_1,\ldots,\lambda_N)X$. Suppose that the $N\times N$ matrix m is such that $\mathcal{M}+m$ is stochastic and that*

$$\kappa_2(X)\|m\|_1 < \frac{1-|\lambda_2|}{2}. \tag{3.36}$$

Then, $\mathcal{M}+m$ has a unique invariant vector.

The proof of this result relies on classical perturbation theory, in particular on the Bauer-Fike theorem [BF60]. With a bit more work, one can deduce a bound on the rate of convergence to equilibrium of the perturbed chain, also presented in the work cited above. The condition shown in Eq. (3.36) can very restrictive if the process is governed by a highly non-normal Markov matrix, as in this case $\kappa_2\left(X\right)$ can be very large [TE05]. This non-normality reflects on pairs of eigenvectors being parallel to each other making the algebraic condition number exceedingly large. However, in order to preserve the spectral gap, the only thing needed is a good conditioning of the eigenvalues closest to the unit circle. Hence, to find a sharper stability bound like in Eq. (3.36) the *eigenvalue condition number* might be the object to look at [Wil65]; see e.g. Appendix B for its definition and [Tan+19] for an application of this concept.

Proposition 3.2.2 only holds for diagonalisable matrices, so that $\kappa_p(X)$ stays finite. In practise, an especially for projected transfer operators, Markov matrices are typically singular and posses clusters of eigenvalues close to the origin. This makes the stability bound in Eq. (3.36) of no use, since the eigenvectors can become almost parallel making the condition number arbitrarily large. The ergodicity coefficient, on the other hand, is more convenient to this end since it does not depend on the algebraic structure of $\mathcal{M}$, but on its mixing character. In particular, we can formulate the stability result of Proposition 3.2.2 for general Markov matrices in terms of the ergodicity coefficient:

Proposition 3.2.3. *Let $\mathcal{M}, m$ be $N \times N$ matrices and ε a real number such that $\mathcal{M}$ and $\mathcal{M}+\varepsilon m$ are stochastic matrices with sets eigenvalues $\{\lambda_i\}_{i=1}^N$ and $\{\lambda_i(\varepsilon)\}_{i=1}^N$, respectively, satisfying: $\lambda_1 = 1 > \|\mathcal{M}\|_{1^*} \geq |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_N|$ and $\lambda_1(\varepsilon) = 1 \geq |\lambda_2(\varepsilon)| \geq |\lambda_3(\varepsilon)| \geq \ldots \geq |\lambda_N(\varepsilon)|$. If $\varepsilon\|m\|_1 < 1 - \|\mathcal{M}\|_{1^*}$, then $\lambda_2(\varepsilon) < 1$ and $\mathcal{M} + \varepsilon m$ has a unique invariant vector.*

Proof. The result follows from the subadditive property of the norm $\|\cdot\|_{1^*}$ on stochastic matrices [IS14]. This implies that:

$$|\|\mathcal{M} + \varepsilon m\|_{1^*} - \|\mathcal{M}\|_{1^*}| \leq \varepsilon\|m\|_{1^*} = \varepsilon\|m\|_1. \tag{3.37}$$

where the last equality follows from Eq. (3.29). Hence, using the assumption of $\|m\|_1 < 1 - \|\mathcal{M}\|_{1^*}$,

$$\|\mathcal{M} + \varepsilon m\|_{1^*} \leq \|\mathcal{M}\|_{1^*} + \varepsilon\|m\|_1 < 1, \tag{3.38}$$

or,

$$\|\mathcal{M} + \varepsilon m\|_{1^*} \leq \|\mathcal{M}\|_{1^*} < 1. \tag{3.39}$$

Secondly, since the ergodicity norm bounds the magnitude of every non-unit eigenvalue, we have that $\lambda_2(\varepsilon) \leq \|\mathcal{M} + \varepsilon m\|_{1^*} < 1$. This last inequality implies that the simplicity of the unit eigenvalue of $\mathcal{M}$ is preserved under the perturbation m. Therefore, the perturbed stochastic matrix $\mathcal{M} + \varepsilon m$ has a unique invariant vector. □

As a consequence, we have that the more mixing the Markov chain is, the more stable the second eigenvalue will be with respect to perturbations. However, Proposition 3.2.3 does not impose conditions on m beyond its size. So it remains to investigate whether structured perturbation matrices can amplify the changes in spectral elements of general Markov chains. In [ADF18], the authors consider the problem of finding a perturbation matrix that maximises the rate of mixing of a Markov chain, by optimising the aforementioned eigenvalue condition number. In case the perturbation matrix only affects the transition probability of a single state, explicit formulas can be derived for the sensitivity of the second eigenvalue, the rate of mixing and linear response; the reader is referred to Appendix B for details.

3.2.1.1 Conditioning of the Linear Response Matrix

The linear response matrix has been shown to be well defined by virtue of the fact that the matrix $1-\mathcal{M}$ is invertible on the range of the perturbation matrix m. The numerical computation of the linear response matrix goes through the calculation of the generalised inverse of Eq. (3.28). In general, the generalised inverse matrix can be inverted directly or by implicitly solving a linear system of equations using an adequate stable method. In any case, the computational well-posedness of finding the inverse of the matrix $1-\mathcal{M}+\mathcal{Q}$ is given by the condition number $\kappa_p\left(1-\mathcal{M}+\mathcal{Q}\right)$, whereby the larger this scalar is, the more numerically difficult it is to invert the matrix. This is because κ_p determines a relative distance of a matrix to its closest non-invertible matrix of the same dimensions. Hence, it is of computational interest to determine the factors that make $\kappa_p\left(1-\mathcal{M}+\mathcal{Q}\right)$ large. In particular, we seek lower bounds of $\kappa_p\left(1-\mathcal{M}+\mathcal{Q}\right)$ in terms of the spectrum of $\mathcal{M}$, thus, avoiding having to calculate inverses explicitly. In particular, the following result holds:

Proposition 3.2.4. *Let N be a positive integer and $\mathcal{M}$ in $\mathbb{R}^{N\times N}$ be a mixing stochastic matrix with eigenvalues $\{\lambda_j\}_{j=1}^N$ satisfying: $\lambda_1=1>|\lambda_2|\geq|\lambda_3|\geq\ldots|\lambda_N|$. Let $\mathcal{Q}$ in $\mathbb{R}^{N\times N}$ denote the spectral projector around the eigenvalue $\lambda_1=1$, defined in Eq.* (3.27)*. Then, the condition number with respect to the spectral norm $\kappa_2\left(1-\mathcal{M}+\mathcal{Q}\right)$ satisfies:*

$$\kappa_2\left(1-\mathcal{M}+\mathcal{Q}\right)\geq\frac{1}{\operatorname{dist}\left(\{1\},\sigma\left(\mathcal{M}\right)\setminus\{1\}\right)}. \tag{3.40}$$

Proof. First, let us find a lower bound for $\|\left(1-\mathcal{M}+\mathcal{Q}\right)^{-1}\|_2$ in Eq. (3.35). Notice that since $\mathcal{Q}$ is a spectral projector around 1, the spectrum of $\mathcal{M}-\mathcal{Q}$ is given by $\sigma\left(\mathcal{M}-\mathcal{Q}\right)=\{\lambda_2,\ldots,\lambda_N\}\cup\{0\}$. Furthermore, the resolvent norm of $\mathcal{M}-\mathcal{Q}$ at the complex number z is bounded from below by the inverse of the distance to the spectrum [EN00]:

$$\|\left(z-\left(\mathcal{M}-\mathcal{Q}\right)\right)^{-1}\|_2\geq\frac{1}{\operatorname{dist}\left(\{z\},\sigma\left(\mathcal{M}-\mathcal{Q}\right)\right)}, \tag{3.41}$$

where $\operatorname{dist}(\cdot,\cdot)$ indicates the usual distance between two compact sets in the complex plane. In fact, this inequality holds for every induced matrix norm.

Secondly, we shall bound the term $\|1-\mathcal{M}+\mathcal{Q}\|_2$ in Eq. (3.35). By the definition of the spectral norm, we have

$$\|1-\mathcal{M}+\mathcal{Q}\|_2=\sigma_1\left(1-\mathcal{M}+\mathcal{Q}\right)=\sqrt{\lambda_{max}\left(\left(1-\mathcal{M}^*+\mathcal{Q}^*\right)\left(1-\mathcal{M}+\mathcal{Q}\right)\right)}, \tag{3.42}$$

where $\sigma_1(A)$ denotes the largest singular value of a matrix A and $\lambda_{max}(A)$ denotes its largest eigenvalue in modulus. Next, by virtue of spectral radius theorem, $\|A\|_2=$

$\sigma_1(A) \geq |\lambda_{max}(A)|$, for any matrix A. In our case, we observe that:

$$(1 - \mathcal{M} + \mathcal{Q})\,\mathbf{u} = \mathbf{u}, \tag{3.43}$$

where $\mathbf{u}$ is the invariant vector associated with $\mathcal{M}$. This shows that 1 is an eigenvalue of $1 - \mathcal{M} + \mathcal{Q}$ and therefore, we get the following inequality:

$$\|1 - \mathcal{M} + \mathcal{Q}\|_2 = \sigma_1\,(1 - \mathcal{M} + \mathcal{Q}) \geq |\lambda_{max}\,(1 - \mathcal{M} + \mathcal{Q})\,| \geq 1. \tag{3.44}$$

Using this inequality and Eq. (3.41), we get the desired result. □

Notice that equality in Eq. (3.40) is attained when $\mathcal{M} = \mathcal{Q}$. Proposition 3.2.4 states the requirements on the spectrum of $\mathcal{M}$ for the associated linear response matrix to be well conditioned for numerical computation: it demonstrates that the presence of an eigenvalue close to unity makes the condition number arbitrarily large. In many applications, as will be apparent later, each iteration of the matrix $\mathcal{M}$ represents the transition probabilities in phase space relative to short physical timescales. This implies that stochastic matrices will be diagonally dominant and the eigenvalues are clustered around 1. Consequently, the second largest eigenvalue λ_2 (in the notation of Proposition 3.2.4) not only does it set the mixing rate of the Markov chain, but will also determine the numerical conditioning for estimating the linear response.

3.3 Numerical Results

The present section aims at applying the discretised transfer operator and perturbation theory of Markov chains in the study two specific dynamical systems subject to external stimuli. The link between Markov chains and dynamical systems was explained in Section 3.1 where the projection of the transfer operator describing the evolution of probability density functions naturally gives matrix approximants with the structure of Markov chains. In Section 3.2, we exploited the theory of stochastic matrices to define, at a finite dimensional and abstract level, the notions of linear response and sensitivity of a Markov chain, which will be here used to calculate and investigate a simple O-U process [UO30] and the Lorenz 63 system [Lor63]. To this end, time series will be used to sample the statistics of the systems in question, to then elaborate detailed climatologies in its unperturbed regime to then predict their response without resorting to observations of the forced system.

The dynamical systems chosen here are different in nature and yield different statistical properties reflected on their associated operator semigroups and invariant measures. Broadly, while the O-U process is stochastic and possesses a stationary probability density with respect to the Lebesgue measure, the deterministic Lorenz 63 system will display an average contraction of volumes on phase space provoking that the associated invariant

measure is singular. The regularity of the invariant measure allows for employing the response formulas derived in Chapter 2, where a density function was explicitly used. For the Lorenz 63 system, however, such linear response formulas will be purely formal, although we shall show that a seamless application of the theory in Section 3.2 onto a coarse-grained phase space can yield correct sensitivities in agreement with experimental data and previous works.

It was shown on Section 2.2 and 2.3 that an external forcing could be associated with perturbation operators affecting the Fokker-Planck equation of a stochastic process; see Eq. (2.16). If one considers a perturbation of the kind $\mathbf{F} \mapsto \mathbf{F} + \sum_k \varepsilon_k \mathbf{G}_k$ in Eq. (2.15), we obtain a perturbed Fokker-Planck equation:

$$\partial_t \rho(\cdot, t) = \mathcal{L}_0 \rho(\cdot, t) + \sum_{k=1}^{n} \varepsilon_k \mathcal{L}_k \rho(\cdot, t), \tag{3.45}$$

where $\mathcal{L}_k = -\nabla \cdot (\mathbf{G}_k \circ)$, in accordance with Eq. (2.17b). We immediately identify the unperturbed and perturbed components of the equation. Assuming sufficient regularity of the measure ρ with respect to t, we can Taylor expand Eq. (3.45) around t. Indeed, let $\tau > 0$ be a small increment in the time variable. Then,

$$\rho(\mathbf{x}, t + \tau) \simeq \rho(\mathbf{x}, t) + \tau \mathcal{L} \rho(\mathbf{x}, t) + \tau \sum_{k=1}^{n} \varepsilon_k \mathcal{L}_k \rho(\mathbf{x}, t), \tag{3.46}$$

where $\mathcal{O}(\tau^2)$ terms have been dropped, following the strategy outlined in [Luc16]. The unperturbed component of Eq. (3.46) represents the pushforward operator to first order in τ. Projecting Eq. (3.45) onto a finite basis of characteristic functions in the spirit of Eq. (3.1), translates the problem into vector and matrix notation:

$$\mathbf{u}_{t+\tau} = \mathcal{M}^\tau \mathbf{u}_t + \sum_{k=1}^{n} \varepsilon_k m_k, \tag{3.47}$$

where the perturbation matrices m_k are differentiation matrices that will be constructed suitably for each of the systems analysed later. We have, therefore, arrived at a perturbation problem of stochastic matrices analogous to Eq. (3.10), making the formulas derived in Section 3.2 susceptible of being applied in continuous time dynamical systems.

As mentioned in the introductory paragraph of this section, we wish to calculate the sensitivity of a system by only sampling the statistics in its unperturbed state. Hence, in order to approximate $\mathcal{M}^\tau$ we shall estimate the climatology of the system using long time series that capture trajectories evolving on a bounded region of phase space $\mathcal{X}$. After transients have died, sample points will populate $\mathcal{X}$, possibly *shadowing* [Rob01] the dynamics on an attractor in case of deterministic systems. In stochastic systems, even though phases spaces are typically unbounded, $\mathcal{X}$ can be chosen so that it capture most of

the probability of the system. Such region can be subdivided into N equally sized boxes $\{B_i\}_{i=1}^N$ with Lebesgue-zero measure intersections. If $\{\mathbf{x}_1, \dots \mathbf{x}_T\}$ denotes the T sample points living in $\mathcal{X}$, the matrix $\mathcal{M}^\tau$ in Eq. (3.46) is constructed as:

$$\mathcal{M}^\tau_{i,j} = \frac{\#\Big\{\{\mathbf{x}_i\}_{i=1}^T \cap B_i \cap \phi^{-\tau} B_j\Big\}}{\#\Big\{\{\mathbf{x}_i\}_{i=1}^T \cap B_j\Big\}}, \tag{3.48}$$

for $i, j = 1, \dots, N$ and where $\#$ is the counting measure. A matrix whose entries are defined following Eq. (3.48) is called a *transition matrix* with transition time τ. In this case, since the sample points $\{\mathbf{x}_i\}_{i=1}^T$ populate the attractor of the system, the probability of transitioning from box B_j to box B_i is relative to the invariant measure of the system. In other words, the transition matrix $\mathcal{M}^\tau$ is an approximation of the projected transfer operator in Eq. (3.3) where η is the invariant measure of the system.

Remark 3.3.1. *The matrix $\mathcal{M}^\tau$ estimated from time series is the procedure for applying Eq.* (3.3) *when the finite measure η corresponds to the invariant measure, in accordance with the usual definition of the transfer and Koopman operator semigroups; see Section 2.1. If the system possesses an absolutely continuous invariant measure, η can be chosen as the Lebesgue measure. In this case, the construction of $\mathcal{M}^\tau$ consists of populating each box B_i in Eq.* (3.3) *with sample points that will be evolved τ time units forward in time and transitions will be counted in the spirit of Eq.* (3.48)*. One can, nevertheless, seamlessly apply the Lebesgue-measure approximation to study the rates of convergence to the steady state regardless of this being a singular measure. The reader is referred to [Fro98] for seminal rigorous results along these lines and to [Tan+18] for a comparative approach in the context of dissipative dynamical systems.*

Remark 3.3.2. *Transition matrices learnt from discretised data like that in Eq.* (3.48)*, correspond to an* empirical frequency estimator *that maximises the log-likelihood function of transition probabilities from state to state [AG57]. The derivation, though, is strongly based on the independence assumption that allows to write the log-likelihood function as a sum. Lifting such assumption entails extending the physical state space to a collection of latent (unobserved) Markovian states[2] which carry the memory of the process. This approach to estimate the invariant measure of the system is, firstly, of greater generality that of Eq.* (3.48) *since it can reveal hidden dependencies that were initially neglected by the coarse-graining and secondly, it appears to be more resilient to dimensionality and lack of data points [Ger+18]. The drawback, though, is that the resulting latent Markov model cannot be related so easily to a projected transfer operator, whose algebraic and spectral properties can be of great use, as is clear from Section 3.1.*

[2]Not to confuse with Hidden Markov Models [Ger+18].

The perturbation matrices m_k in Eq. (3.47) represent projected differential operators and they should be such that they constitute admissible perturbations to the stochastic matrix $\mathcal{M}^\tau$, namely, their columns need to add up to zero (see Eq. (3.11)) and ε_+ and ε_- in Eqs. (3.12) and (3.13), respectively, should be larger than zero. We now examine how two implement them so that they are compatible with the domain discretisation carried out for the transition matrix $\mathcal{M}^\tau$. For simplicity, we shall describe the procedure in two dimensions which still allows for cross terms in higher order derivatives. The idea is to assign a value to the derivatives at each box B_i based on its value at the centre. The centre of box B_i is denoted by $[x_i, y_i]$ in $\mathbb{R}^2$, for every $i = 1, \ldots, N$. Since the boxes are equally sized, we let the real numbers δ_x and δ_y denote the length of the sides along the x and y-axis, respectively. Since the box-discretisation need not have a Cartesian product structure, there are three kinds of boxes. The box B_i is said to be

(i) *interior* if $[x_i, y_i] \pm [\delta_x, 0]$ and $[x_i, y_i] \pm [0, \delta_y]$ correspond to a center of some box in $\{B_i\}_{i=1}^N$.

(ii) *isolated* if $[x_i, y_i] \pm [\delta_x, 0]$ and $[x_i, y_i] \pm [0, \delta_y]$ do not correspond to a center of any box in $\{B_i\}_{i=1}^N$.

(iii) a *boundary* box if it is neither interior or isolated.

Performing several Taylor expansions we find the partial derivative of ρ with respect to x at the interior box $[x_i, y_i]$ to be:

$$\partial_x \rho\left([x_i, y_i]\right) = \frac{\rho([x_i + \delta_x, y_i]) - \rho([x_i - \delta_x, y_i])}{2\delta_x} + \mathcal{O}\left(\delta_x^2\right). \tag{3.49}$$

which corresponds to a centred stencil. The same scheme is used for the y-direction. For the second and cross derivatives, we implement the usual second order centred discretisation:

$$\partial_x^2 \rho([x_i, y_i]) = \frac{\rho([x_i + \delta_x, y_i]) - 2\rho([x_i, y_i]) + \rho([x_i - \delta_x, y_i])}{2\delta_x^2} + \mathcal{O}(\delta_x^2), \tag{3.50a}$$

$$\begin{aligned}\partial_{x,y}^2 \rho([x_i, y_i]) =& \frac{\rho([x_i + \delta_x, y_i + \delta_y]) - \rho([x_i + \delta_x, y_i - \delta_y])}{4\delta_x \delta_y} \\ &- \frac{\rho([x_i - \delta_x, y_i + \delta_y]) - \rho([x_i - \delta_x, y_i - \delta_y])}{4\delta_x \delta_y} + \mathcal{O}(\delta_x \delta_y).\end{aligned} \tag{3.50b}$$

This stencil is completed with the analogous schemes in the y-direction. Since Eqs. (3.49)-(3.50) involve the centres of every box surrounding B_i, we define the derivatives with respect to x and y as the finite difference increments obtained by dropping the second order terms in δ_x and δ_y. Derivatives at the boundary boxes are defined similarly, although the centred scheme employed in Eqs. (3.49)-(3.50) has to be substituted by a first order

increment. Indeed, without loss of generality, suppose B_i is a boundary box so that $[x_i - \delta_x, y_i]$ is not a centre of any box in $\{B_i\}_{i=1}^N$. Then, the partial derivative of ρ with respect to x at $[x_i, y_i]$ is:

$$\partial_x \rho\left([x_i, y_i]\right) = \frac{\rho([x_i + \delta_x, y_i]) - \rho([x_i, y_i])}{\delta_x} + \mathcal{O}\left(\delta_x\right). \tag{3.51}$$

Derivatives at the rest of the boundary boxes are defined following this strategy. Finally, we enforce that isolated boxes have derivatives equal to zero. Naturally, these numerical derivatives can be arranged into matrices so that their multiplication with vectors approximate their respective differentiation. Thus we obtain matrix representations m_k of the operators $\mathcal{L}_k$.

3.3.1 An O-U Process

As a preliminary application, we shall consider the O-U process, whose statistical properties are well known in analytical terms; see e.g., [MPP02]; [Pav14, Chapter 3]. This process $\{\mathbf{x}(t)\}_{t\geq 0}$ is generated by the following linear SDE:

$$\mathrm{d}\mathbf{x}(t) = A\mathbf{x}(t)\mathrm{d}t + \Sigma \mathrm{d}W_t, \tag{3.52}$$

where W_t is a d-dimensional Wiener process, A is a $d \times d$ matrix which represents the drift of the process and the $d \times d$ matrix Σ is such that $\Sigma\Sigma^\top$ is positive-definite indicating the covariance of the noisy component. Performing an integration of Eq. (3.52) in the Itô sense, one obtains strong solutions driven by the exponential of the matrix A:

$$\mathbf{x}(t) = e^{tA}\mathbf{x}_0 + \int_0^t e^{(t-s)A}\mathrm{d}W_s, \tag{3.53}$$

for some d-dimensional initial condition $\mathbf{x}_0$ which can be prescribed exactly or assumed to be suitably distributed. The stability of Eq. (3.53) is determined by the spectrum of the matrix A which determines the norm-growth of its matrix exponential. Hence, we shall require that spectral abscissa (the maximum of the real parts of the eigenvalues) of A is strictly less that zero. In fact, such condition is necessary for this process to posses an Gaussian invariant measure [DZ96]:

$$\rho(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det \mathcal{S}}} e^{-\frac{1}{2}\mathbf{x}^\top \mathcal{S}^{-1}\mathbf{x}}, \tag{3.54}$$

where $\mathcal{S}$ is the $d \times d$ positive-definite stationary covariance matrix, determined by the Lyapunov equation:

$$A\mathcal{S} + \mathcal{S}A^\top = \Sigma\Sigma^\top, \tag{3.55}$$

which uniquely determines the matrix $\mathcal{S}$ since $\Sigma\Sigma^\top$ is strictly positive-definite [Gar09]. As $\mathbf{x}(t)$ is let to evolve in time, ensembles of trajectories on phase space are exponentially quickly expected to oscillate around zero. If such ensemble is regarded as a probability density function ρ, it obeys the Fokker-Planck equation (3.45). Again, the spectral stability of the matrix A translates on the operator $\mathcal{L}_0$ having a zero spectral abscissa, with a simple eigenvalue located exactly at 0 and relative eigenfunction given by a Gaussian invariant measure [DZ96]; [Ris89]. Moreover, the entirety of the spectrum of $\mathcal{L}_0$ consists of simple eigenvalues given by linear combinations of those of A with integer coefficients [MPP02]. Consequently, the semigroup generated by the Fokker-Planck equation is, trivially, quasi-compact with zero residual for which spectral decompositions of the response function are available; see Section 2.2.1. Furthermore, the strongly continuous semigroup generated by $\mathcal{L}_0$ is a contraction which converges (in the operator norm sense) to the spectral projector relative to the eigenvalue 0 exponentially fast, with a rate determined by a finite spectral gap equal to the maximum real part of the eigenvalues of A.

Numerical Setting and Results. As a numerical example, we shall consider a two dimensional O-U processes, where the constituent matrices A and Σ in Eq. (3.52) are given by:

$$A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \text{ and } \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \tag{3.56}$$

It is obvious that the variables in this process are uncorrelated and we shall investigate the response of the system when its mean is shifted and correlations are introduced in the noise. This way, the process in question is:

$$\mathrm{d}\mathbf{x}^{\varepsilon_1,\varepsilon_2}(t) = A\left(\mathbf{x}^{\varepsilon_1,\varepsilon_2}(t) - \varepsilon_1\mu\right)\mathrm{d}t + \sqrt{\Sigma\Sigma^* + \varepsilon_2 H}dW_t, \tag{3.57}$$

where

$$\mu = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \text{ and } H = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{3.58}$$

and $\varepsilon_1, \varepsilon_2 \in \mathbb{R}$ have to satisfy the conditions for the perturbed process to be well defined, namely, $\Sigma\Sigma^* + \varepsilon_2 H$ has to be symmetric positive definite. The resulting Fokker-Planck equation takes the form of Eq. (3.45) where $\mathcal{L}_1$ and $\mathcal{L}_2$ are a first and a second order differential operators, respectively, and are defined component-wise as:

$$\mathcal{L}_1\rho(\mathbf{x}) = -\varepsilon_1 \sum_{k=1}^{2} \partial_{x_k}\left(\sum_{l=1}^{2}, \left[A\mu\right]_l \rho(\mathbf{x})\right); \tag{3.59a}$$

$$\mathcal{L}_2\rho(\mathbf{x}) = \frac{\varepsilon_2}{2}\sum_{k=1}^{2}\sum_{l=1}^{2}\partial^2_{x_k,x_l}\left(H_{k,l}\rho(\mathbf{x})\right), \tag{3.59b}$$

for every sufficiently differentiable ρ and $\mathbf{x}$ in $\mathcal{X}$. The projection onto a finite dimensional basis would yield the discretised equation Eq. (3.47), falling into the perturbation theory developed in Section 3.2.

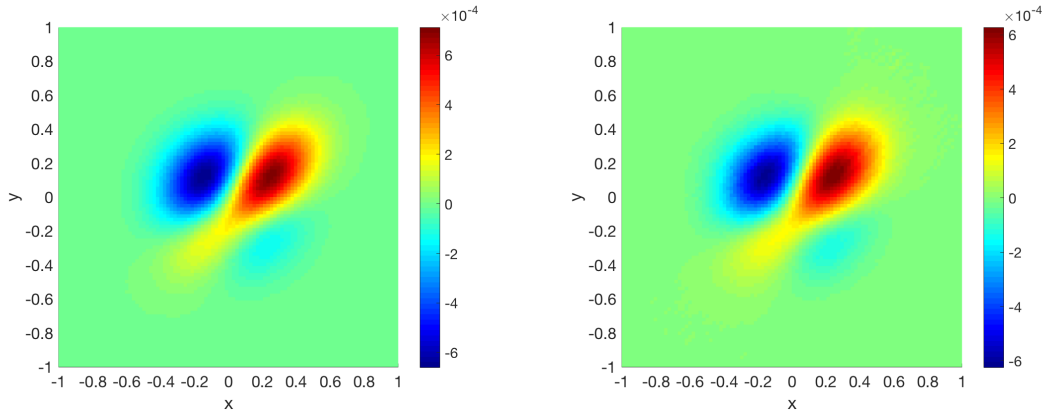
In order to construct a transition matrix $\mathcal{M}^\tau$, we let $\{B_i\}_{i=1}^{2^N}$ be a collection of boxes covering four standard deviations of the process on each direction. In the experiments performed, we have considered 2^N boxes by means of discretising each axis into $2^{N/2}$ equally sized segments. We have examined the values $N = 10, 12$ and 14 trying to keep a balance between numerical tractability and precision. The box subdivision of phase space is done using the MATLAB package GAIO [DFJ01] which creates a hierarchy of boxes based on a binary tree structure which greatly accelerates the process of associating a point on phase space with the unique box containing it. The unperturbed process is sampled by integrating Eq. (3.52) forward using an Euler-Maruyama scheme for 10^6 time units with a time step of $\mathrm{d}t = 10^{-2}$, thus obtaining a time series of points on phase space. We then construct the transition matrix $\mathcal{M}^\tau$ according to Eq. (3.48), where ϕ^t is now replaced by the process $\mathbf{x}(t)$ and the transition time is chosen to be $\tau = \mathrm{d}t$.

The perturbation matrices m_1 and m_2 that result from discretising the operators $\mathcal{L}_1$ and $\mathcal{L}_2$ are calculated using Eqs. (3.49)-(3.50). In the present O-U case, since the invariant measure has an absolutely continuous density with respect to the Lebesgue measure which is smoothly defined on $\mathbb{R}^2$, there will not be any complexity on its support, therefore we shall not expect to find many isolated boxes or boundary boxes having relatively big probability of occurrence. Moreover, provided that the time series at hand is long enough relative to the domain discretisation, the boundary boxes will have almost zero probability, therefore, there is no practical need of implementing explicit boundary derivatives as done in Eq. (3.51).

A good compromise was found provided that the resolution was high enough. Indeed, on Fig. 3.1, we see that the predicted response of the observable x was well approximated. This is checked on Table 3.1 where we show that the error of approximating the perturbed invariant measure using Eq. (3.8) and the response formulas of Eq. (3.21) presented earlier is small. The linear response (see Fig. 3.2) is precisely doing what one expects: mass is pumped to the right as a consequence of the mean being shifted and the introduction of correlations inflicts a rotation. Higher order correction terms (see Fig. 3.2) can also give an insight on how the measure is gradually modified. As a safety check, the sum of the components of the response are checked to add up to (almost) zero, meaning that probability is not introduced or depleted.

3.3.2 The Lorenz 63 System

The existence of an invariant density with respect to the Lebesgue measure in the O-U process has greatly facilitated the computation and characterisation of its response to



(a) (b)

Figure 3.1: Response of the O-U system to the perturbations considered in Eq. (3.58). The figure (a) shows the true response calculated by subtracting the unperturbed invariant measure from the perturbed one. Figure (b) is the predicted response calculated from Eq. (3.20).

Table 3.1: Mean value (first column) and linear responses (second and third columns) of the observable x with respect to the perturbations in Eq. (3.58). The first row refers to the values obtained from integrating the O-U process. The rest are values calculated via discretisation of the transfer operator where N is the resolution. We defined Error1 as the L^2 norm of the difference between the coarse-grained perturbed invariant measure and the first order correction. Error2 is the same but with higher order correction terms.

	$\langle x \rangle$	$\delta^{(1)}_{\varepsilon_1}[x]_1$	$\delta^{(1)}_{\varepsilon_2}[x]_1$	Error1	Error2
O-U	0	1	0	0	0
$N = 10$	10^{-4}	0.47	3×10^{-3}	0.02	0.02
$N = 12$	10^{-4}	0.77	2×10^{-3}	$7{\times}10^{-3}$	5×10^{-3}
$N = 14$	10^{-4}	0.93	8×10^{-4}	$3{\times}10^{-3}$	8×10^{-4}

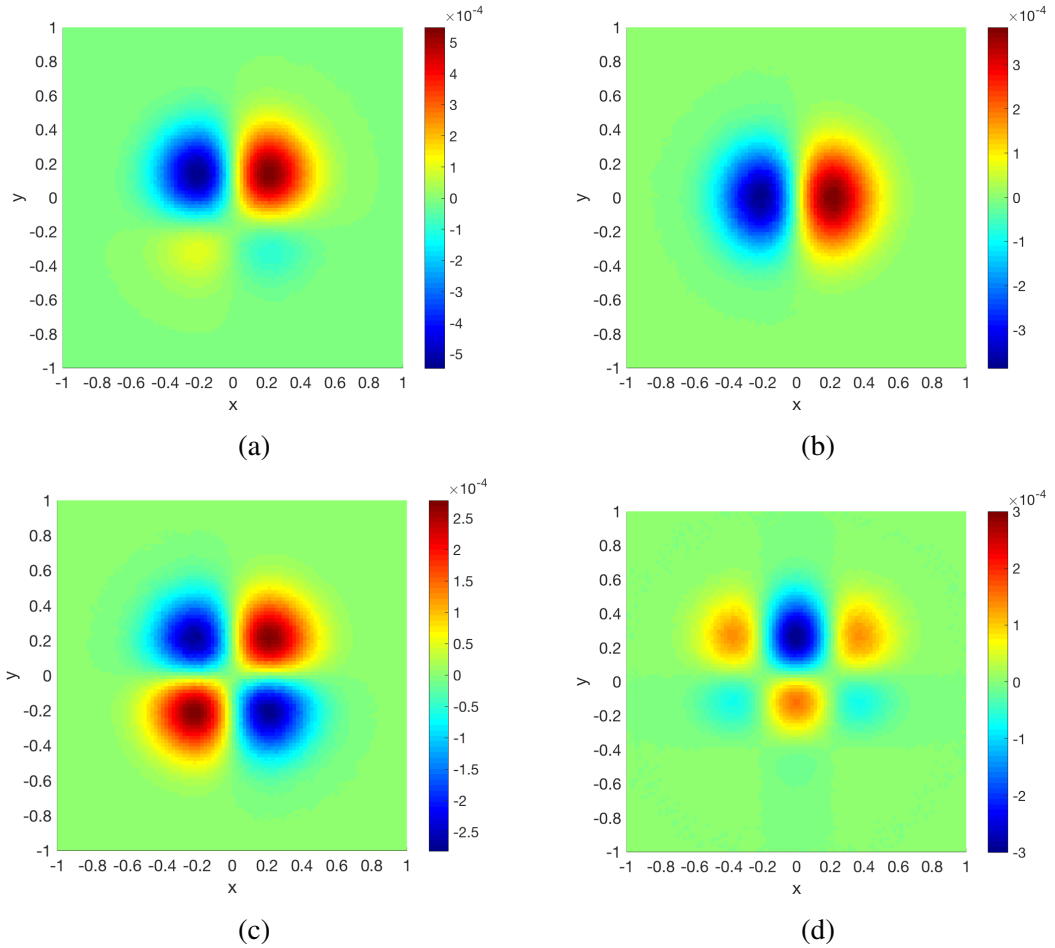


Figure 3.2: Panel (a) shows the linear response of the O-U process with respect to the perturbations considered in Eq. (3.58) calculated by truncating Eq. (3.21) at the first order. Figures (b) and (c) show the linear response to changes in ε_1 and ε_2, respectively. Figure (d) contains the second order correction obtained via Eq. (3.21).

external forcing. The Lorenz 63 system, instead, generates a non-conservative and non-uniformly hyperbolic flow making it ill-posed for calculating response in the sense of Ruelle (1.9). The Lorenz 63 system is a deterministic three-dimensional flow $\{\phi^t\}_{t\in\mathbb{R}}$ that is obtained by truncating Saltzman's equations for convection, and results in the set of ODEs governed by the vector field $\mathbf{F} : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}) = \begin{cases} s(y-x) \\ x(r-z) - y \\ xy - bz \end{cases}, \tag{3.60}$$

where the Prandtl number, aspect ratio and Rayleigh factor are $s = 10$, $b = 8/3$ and $r > 0$, respectively. The Rayleigh factor is not determined just yet in view that it is the parameter that, ultimately, regulates the temperature difference between the convecting layers and it is responsible for the appearance of well-studied bifurcations [Spa82]. The Lorenz 63 equation (3.60) is quadratic and, hence, non-linear which, together with having dimension greater than two, makes it susceptible of displaying complex chaotic dynamics on a compact region of phase space.

The presence of a global attractor boils down to the existence a quadratic Lyapunov function whose Lie derivative has negative values on an ellipsoid [Spa82]. The axis and centre of this ellipsoid can be written in terms of the model parameters r, s and b. One considers, for instance, the Lyapunov function $L(\mathbf{x}) = rx^2 + sy^2 + s(z-2r)^2$, with Lie derivative satisfying the following relation:

$$\dot{L}(\mathbf{x}) = \nabla L(\mathbf{x}) \cdot \mathbf{F}(\mathbf{x}) = -2s(rx^2 + y^2 + bz^2 - 2rbz), \tag{3.61}$$

for every $\mathbf{x}$ in $\mathbb{R}^3$. Consequently, the set $\mathcal{E} := \{\mathbf{x} \in \mathbb{R}^3 : rx^2 + y^2 + bz^2 = 2rbz\}$ defines an ellipsoid and the points at which the Lie derivative of L vanishes. Furthermore, the points in the interior of $\mathcal{E}$ are the only ones having positive Lie derivative, implying that all points on $\mathbb{R}^3$ eventually fall into $\mathcal{E}$. Thus is found a compact trapping region on phase space where entering trajectories can never leave. This fact, together with the regularity of the vector field $\mathbf{F}$, implies that solutions are well defined for any time.

The non-conservative forces present in the Lorenz system result from the dissipative processes in the Saltzman's equation. More concretely, phase pace experiences an average exponential contraction of volumes proportional to the divergence of the governing vector field. Indeed, let $V(t) : \mathbb{R} \longrightarrow \mathbb{R}$ denote the volume of a compact subset $S \subset \mathbb{R}^3$ with smooth boundary and normal vector $\mathbf{n}$ in $\mathbb{R}^3$. Then, Liouville's formula and the divergence theorem give:

$$\dot{V}(t) = \int_{\phi(\partial S,t)} \mathbf{F} \cdot \mathbf{n} \mathrm{d}s = \int_{\phi(S,t)} \nabla \cdot \mathbf{F} \mathrm{d}v = -(1+\sigma+b)V(t). \tag{3.62}$$

From where one deduces that:

$$V(t) = e^{-(1+s+b)t}V(0). \tag{3.63}$$

Hence, if the trapping volume determined by the ellipsoid $\mathcal{E}$ is let to evolve forward in time, it will eventually have zero volume. Notice that this exponential contraction of volumes does not imply stability in any way, it just means that the limiting set or attractor, if it exists, will have zero Lebesgue volume. Moreover, the invariant measure of the system, regardless of the nature of the attractor, will be singular with respect to the Lebesgue measure. Let us note, furthermore, that Eq. (3.63) implies that the Lorenz flow cannot have repelling equilibria. If the latter were true, there would exist an open neighbourhood of a source in $\mathbb{R}^3$ whose volume is exponentially divergent with rate equal to the sum of the eigenvalues of the linearised equation, in contradiction with Eq. (3.63).

Solving $\mathbf{F} = 0$ gives three equilibrium points on phase space, namely, $\mathbf{O} = (0, 0, 0)$ and $\mathbf{p}_\pm = (\pm\sqrt{b(r-1)}, \pm\sqrt{b(r-1)}, r-1)$. For the values of $0 < r < 1$ the origin is the only real equilibrium value, whose local stability is determined by the eigendecomposition of the linearised equation at $\mathbf{O}$. Global stability, on the other hand, follows from the existence of a Lyapunov function whose Lie derivative is negative for all values outside a compact ellipsoid[3]. At $r = 1$, a degeneracy is introduced in the linearised equation at the origin and two new equilibria $\mathbf{p}_\pm$ appear. This way a supercritical pitchfork is traversed at $r = 1$, since the origin loses stability and the non-trivial equilibria are locally stable. The stability of $\mathbf{p}_\pm$ is broken when the real parts of two of the eigenvalues determining the linear stability trespass the imaginary axis, giving rise to two unstable and one stable local directions. Such loss of stability corresponds to a subcritical Hopf bifurcation attained at [Spa82]:

$$r_H = \frac{s(s+b+3)}{s-b-1} \approx 24.7368... \tag{3.64}$$

corresponding to the "critical value of r for the instability of steady convection" [Lor63]. Beyond the Hopf bifurcation $r > r_H$, the lack of sinks or sources and the numerical evidence of a positive Lyapunov exponent, revealing sensitive dependence with respect to initial conditions, suggest chaotic behaviour of the flow supported on a global attractor, albeit non-rigorously.

The quest for a mathematical proof of chaos in the Lorenz system started since the publication of E. Lorenz' seminal paper, and was gathered as one of S. Smale's unresolved problems [Sma98]. Attempts on the proof of such fact kicked off with Lorenz' local maxima map, which by naked eye resembled the (chaotic) tent map to the extent of being mutually homeomorphic. The geometric Lorenz model was later introduced by J. Guckenheimer [GW79], where the Poincaré map was studied and argued to be faithful

[3]The Lyapunov function L can be taken to be in this case $L(\mathbf{x}) = rx^2 + sy^2 + sz^2$.

with the original model. Such one-to-one correspondence was proved in 2000 concluding the result below [Tuc02]:

Theorem 3.3.1 (W. Tucker)**.** *For the classical parameter values ($s = 10$, $b = 8/3$ and $r = 28$), the Lorenz equations support a robust strange attractor, which supports a unique SRB measure.*

While strangeness in this theorem refers to the sensitive dependence on initial conditions, robustness alludes to the persistence of the attractor's topology under small changes in the governing parameters. In this sense, the Lorenz attractor is shown to posses a hyperbolic structure, although in a singular way. Such non-uniform hyperbolicity results from the intersection of the stable manifold with the return plane, provoking a discontinuity in the Poicaré map. Consequently, robustness in Theorem 3.3.1 cannot be in the linear response sense, since the splitting of the tangent space is not continuous and, thus, not suitable for having a smooth density along the unstable directions. There exists, nevertheless, attempts in providing numerical evidence of linear response in the Lorenz 63 system point to the existence of linear response in the Lorenz system. In [Rei02], linear response is analysed in the frequency domain providing a methodology to extract from particular time series the value of the susceptibility function at a given frequency. It is, in fact, established that the modulus of the susceptibility should be inversely proportional to the forcing strength when the latter is small, up to approximately 0.001 in the Lorenz 63 model. An extension to high-order responses was done in [Luc09], where generalised Kramer-Kronig relations were numerically calculated, this time departing from the numerical calculation of the Green function. Other approaches focus on the numerical shadowing of trajectories and, in fact, constitute the dual approach to the operator methods here presented [LAH00]; [NW17]

The perturbation problem we tackle here is that of changing the value of the control parameter $r \to r + \varepsilon_1$ for ε_1 in $\mathbb{R}$, which leads to bifurcations numerically surveyed in [Spa82]. We also study the additive perturbation on the z-variable by adding a real parameter ε_2 on the third component of the vector field $\mathbf{F}$. These perturbations incur a modification on the vector field of the form:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}) + \varepsilon_1 \mathbf{G}_1(\mathbf{x}) + \varepsilon_2 \mathbf{G}_2(\mathbf{x}), \tag{3.65}$$

where $\mathbf{G}_1(\mathbf{x}) = [0, x, 0]^\top$ and $\mathbf{G}_2(\mathbf{x}) = [0, 0, 1]^\top$. Consequently, the Liouville equation reads as Eq. (3.45), with perturbation operators $\mathcal{L}_1$ and $\mathcal{L}_2$ defined according to Eq. (2.17b). The resulting changes in the statistics of the system were monitored through the observable z whose mean value was calculated for values of r ranging in $[23.5, 30]$ and ε_2 in $[-1, 1]$. When $r \approx 24.7 + 0.47\varepsilon_2$, a sharp gradient in the mean value of the observable z is identified which results from the Hopf bifurcation. This unbounded gradient located at r_H is tracked

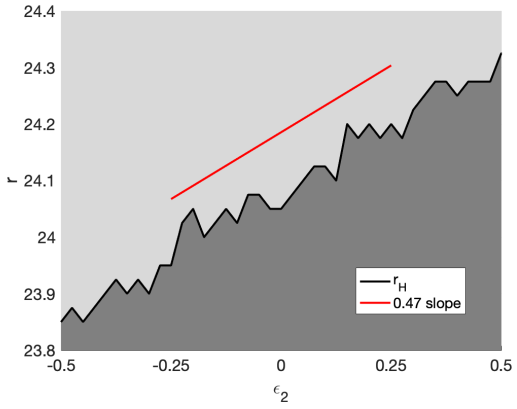


Figure 3.3: Location of the statistically inferred bifurcation point r_H as a function of the parameters r and ε_2. The solid black line r_H indicates sharp gradients of $\langle z \rangle$. Darker and lighter areas roughly correspond to values for $\langle z \rangle > 22.5$ and $\langle z \rangle < 20.5$, respectively. The expectation values were computed by integrating the equations for $5 \cdot 10^2$ time units with twenty ensemble members for each value parameter choice.

and shown in Fig. 3.3 (a) as a function of r and ε_2. Far away from this bifurcation point, the statistics change smoothly with respect to ε_1 and ε_2. For the calculation of these means, ensembles of twenty integrations of $5 \cdot 10^2$ time units and a time step 10^{-2} time units were analysed after having removed 10% of each time series to discard transients. This number of ensemble members are length of integrations resulted in negligible standard deviations provided r is far away from the bifurcation point.

Numerical Setting and Results. Unlike the O-U system considered in Section 3.3.1, the dynamics of the Lorenz 63 model is supported on a compact attractor. Negative Lie derivatives on the ellipsoid $\mathcal{E}$ (see Eq. (3.61)) provide an analytical bound of the location of the attractor although numerical investigations, on the other hand, guarantee that, once transients have died away, the attractor is tightly enclosed by the Cartesian product $[-20, 20] \times [-30, 30] \times [0, 50]$. Then, to obtain the box partition, we divide each axis into two and repeat such procedure in the resulting segments. This way, a total of 2^N boxes $\{B_i\}_{i=1}^{2^N}$ were constructed for the values of $N = 12, 15$ and 18. As mentioned earlier, the box discretisation and searching algorithms were executed using GAIO [DFJ01]. We note at this stage that only a few of the 2^N boxes will support a non-zero value of the invariant measure and hence, only these will be kept for computations reducing the sizes of the later constructed transition matrices. On this note, the box-counting dimension of the estimated compact set can be readily calculated, giving here an approximation of 1.95, slightly lower in comparison with previous works [GP83]. It is not intended here to provide rigorous computations on this matter.

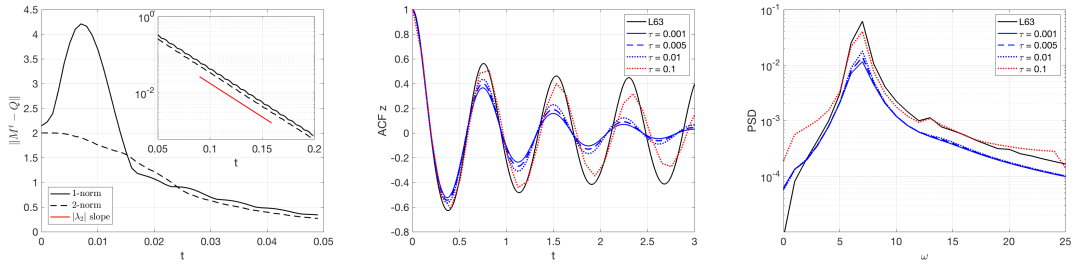
Looking back at Eq. (3.47), we want to sample the unperturbed system (3.60) in order to approximate the transfer operator via the stochastic matrix $\mathcal{M}^\tau$ for some suitable choice of

transition time τ. To obtain the sample points, we integrated the model for 10^5 time units with a time step of $\mathrm{d}t = 10^{-3}$ time units using an adaptive Runge-Kutta scheme, where 10^3 time units had been previously discarded to make sure the system has reached the steady state. Such length of integration translates into approximately 10^5 Lyapunov times and guarantees that each box is populated by a sufficient amount of points; a failure of the latter would create numerical artifacts like having multiple irreducible states— of positive Lebesgue measure—, in disagreement with the presence of physical invariant measure [Tuc02]. Once the time series is at hand, the transitions between boxes are recorded following Eq. (3.48) where four transition times τ have been studied: $\tau = \mathrm{d}t, 5 \cdot \mathrm{d}t, 10 \cdot \mathrm{d}t$ and $100 \cdot \mathrm{d}t$. The choice of these transition times aims at keeping a balance between the accuracy of the time-discretised Liouville equation (3.46) and the coarse-graining due to the box subdivision.

The dissiptiveness of the flow together with the irreversibility of the coarse-grained transfer operator $\mathcal{M}^\tau$, reveals resonances within the unit circle with non-zero imaginary parts, explaining the oscillations in the correlation functions. This results in having a non-selfadjoint generator in the Liouville equation which further implies the non-normality of the resulting transition matrix $\mathcal{M}^\tau$. Transient growths and delays in convergence to steady state are, hence, expected when taking successive powers of $\mathcal{M}^\tau$[4]. This is illustrated in Fig. 3.4 (a) where the convergence to steady state is measured in terms of the powers of $\mathcal{M}^\tau - \mathcal{Q}$, where $\mathcal{Q}$ is the projection matrix to which $\mathcal{M}^{n\tau}$ converges as n tends to infinity. The Euclidean and 1-norms are considered and in both cases there is a delay before they enter the asymptotic regime where the convergence rate is measured by the modulus second largest eigenvalue λ_2 of $\mathcal{M}^\tau$. More concretely, we observe that at $t \approx 0.025$ the exponentially decaying regime starts. While the 1-norm shows an subexponential initial decay, the Euclidean or 2-norm exhibits transient growth. In either case, these phenomena are explained by the numerical range of $\mathcal{M}^\tau$ exceeding the complex unit ball; see [TE05, Chapter IV] for an overview on norm-growth of matrix powers.

It is well known that the artificial diffusion due to a coarse discretisation is lessened by a large transition time τ, especially when approximating the eigenvalues of the Liouville/Fokker-Planck equation [TLD18]. Unlike conservative systems, the Lorenz attractor will be composed of infinitely self-replicating structures that will be smudged by any spatial discretisation, regardless of the resolution. As commented earlier, artificial diffusion equates to the introduction of irreversibility having two relevant implications in the present case study: (a), the sampled invariant measure is smoothed [FJK13] and (b), eigenvalues within the unit circle appear, although they are susceptible of revealing the location of the Ruelle-Pollicot resonances [FR02]. Such eigenvalues, however, do not have a formula in terms of the vector field $\mathbf{F}$ in Eq. (3.60) although it is known that they explain

[4]This is mathematically attributed to non-normal matrices possessing a numerical range that strictly contains the convex hull of its eigenvalues; see [TE05].



(a) (b) (c)

Figure 3.4: (a) 1- and 2-norms of the decay matrix $\mathcal{M}^t - \mathcal{Q}$ as a function of time and its logarithmically scaled version against the slope given by the modulus of the second largest eigenvalue λ_2 of $\mathcal{M}^{5\cdot \mathrm{d}t}$ in the inset figure. (b) Autocorrelation function of the variable z using Eq. (3.9) for transition times τ indicated in the legend vs. full integration (black curve). (c) Power spectral densities of the variable z obtained from the Fourier transform of Eq. (3.9) for different values of transition time τ indicated in the legend vs. full integration (black curve).

the evolution and decay of correlation functions [Rue86]; [BL07]. Hence, an immediate sanity check is the evaluation of correlation functions as done in Fig. 3.4, where in panel (b) we computed the autocorrelation function of the variable z. The black curve indicates the autocorrelation function obtained from direct numerical simulations of the Lorenz 63 system, whereas the coloured lines are approximated autocorrelation functions using a transition matrices $\mathcal{M}^\tau$ and formula (3.9) with a resolution of $N = 15$ and transition times τ indicated in the legend. We highlight that for each curve, an independent run of the model has been performed accordingly with the integration parameters explained earlier in this subsection, so that dependences on initial conditions are ruled out. It is distinct that for larger values of transition times, a better approximation of the correlation function is obtained, as results of a minimisation of the artificial diffusion entailed by the discretisation. The case of $\tau = 0.1$ is plotted in red since, although being a valid transition time, it will not be considered for the study of the response in favour of the accuracy of the time-discretised Louville equation (3.46). Fig. 3.4 (c) shows the power spectral density of the variable z for different transition times. It is remarkable that the location of the frequency peak is indistinguishable from the reference integration (black curve), located at around $\omega = 8$. This suggests that, while the rate of correlation decay is overestimated due to artificial diffusion, the dominating frequency variability is well captured. In other words, the real parts of the eigenvalues of $\mathcal{M}^\tau$ are more sensitive to transition times, whereas the imaginary parts are more robust.

A coarse-grained estimation of the invariant measure can be readily obtained by solving the eigenvalue problem Eq. (3.6) and it is plotted on Fig. 3.5 (a) against the corresponding box-covering for $N = 15$. We immediately identify the complex "butterfly"-like set of boxes that supports the coarse-grained measure, which concentrates probability around

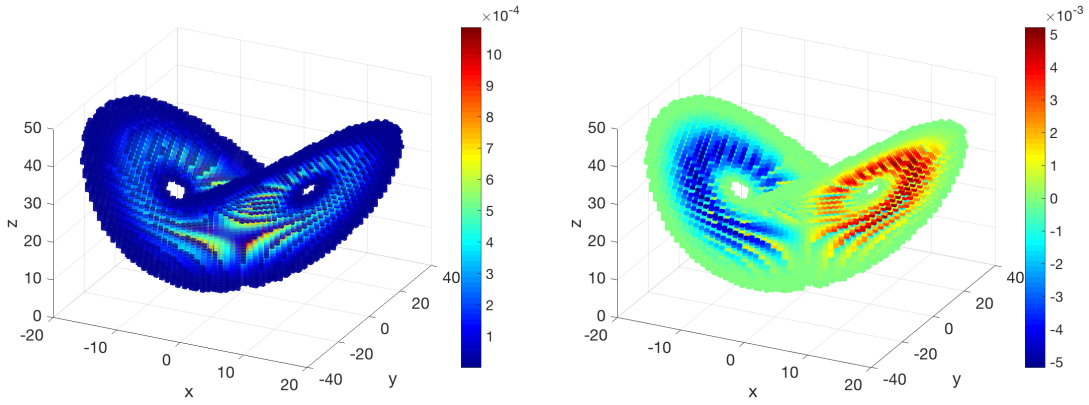
the centre of the butterfly compared to the edges of the lobes where the dark blue colours indicated a low probability of occurrence. Regarding the perturbation operators, the same techniques as with the O-U process where employed. The difference here that one needs to take special care due to the complex geometry of the attractor, as illustrated by Fig. 3.5. Hence, one needs to take special care of the boundary and isolated boxes. As noted at the end of Section 3.3, isolated boxes will be assigned zero-valued derivatives, whereas boundary box derivatives will be estimated with forward increments (see Eq. (3.51)). This technical detail is of crucial importance in the treatment of the edges of the lobes of the attractor depicted in Fig. 3.5 (a), where the values of probability, although close to zero, sharply vanish once outside the box-covering. Note that this is not a significant problem in the O-U process, since the Gaussian density smoothly vanishes at the outer boxes. Bearing all these details in mind, the operators $\mathcal{L}_1$ and $\mathcal{L}_2$ were discretised into m_1 and m_2. In particular, Fig. 3.5 (b) shows the evaluation of m_1 onto $\mathbf{u}$, revealing the probability increase and depletion flows resulting from the applied perturbation; see red and blue values respectively. Ulam's method is prone to errors when estimating the invariant measure, and such errors are larger (in relative terms) where the empirical occupation rate is smaller. This is made worse when one considers finite differences. Nonetheless, we expect that the errors we introduce are small in absolute terms and localised in the phase space. Therefore, if one considers smooth observables, the overall contribution coming from those regions will end up being small. The relevance of using smooth observables should not come as a surprise: Ruelle's [Rue09] response theory only works for C^3 observables, while the extension to less regular functions entails modifications that will not be treated here [BKL17].

At a qualitative level, the results of applying this methodology are presented in Fig. 3.6, where we show that Eq. (3.21) can indeed approximate the expectation values of the chosen observables for a wide range of values of ε_1. In this case, we are showing the calculations obtained with a transition time of $\tau = \mathrm{d}t$. We underline that these plots demonstrate the validity of the formulae not only to compute the linear response but to predict the perturbed statistics of the system. As is natural, the formulas cannot be expected to work for large values of the perturbation parameter, letting alone beyond the bifurcation point at r_H.

Quantitatively, tables 3.2-3.4 contain a series of statistics computed using stochastic matrices $\mathcal{M}^\tau$ for transition times $\tau = \mathrm{d}t, 5 \cdot \mathrm{d}t$ and $10 \cdot \mathrm{d}t$, respectively, with perturbation parameters of $\varepsilon_1 = 0.1$ and $\varepsilon_2 = 0.1$. In these tables, the perturbative expansions and linear response formulas obtained in Section 3.2 are applied to calculate the expectation values of the observables x^2, y^2, z^2 and z using Eq. (3.8), together with their linear responses using Eq. (3.21). The first row of these tables is labelled with Lorenz 63 and contains the reference values calculated from direct simulations of Eq. (3.60). Differently to the calculations for Fig. 3.3, an ensemble of twenty independent members was taken in

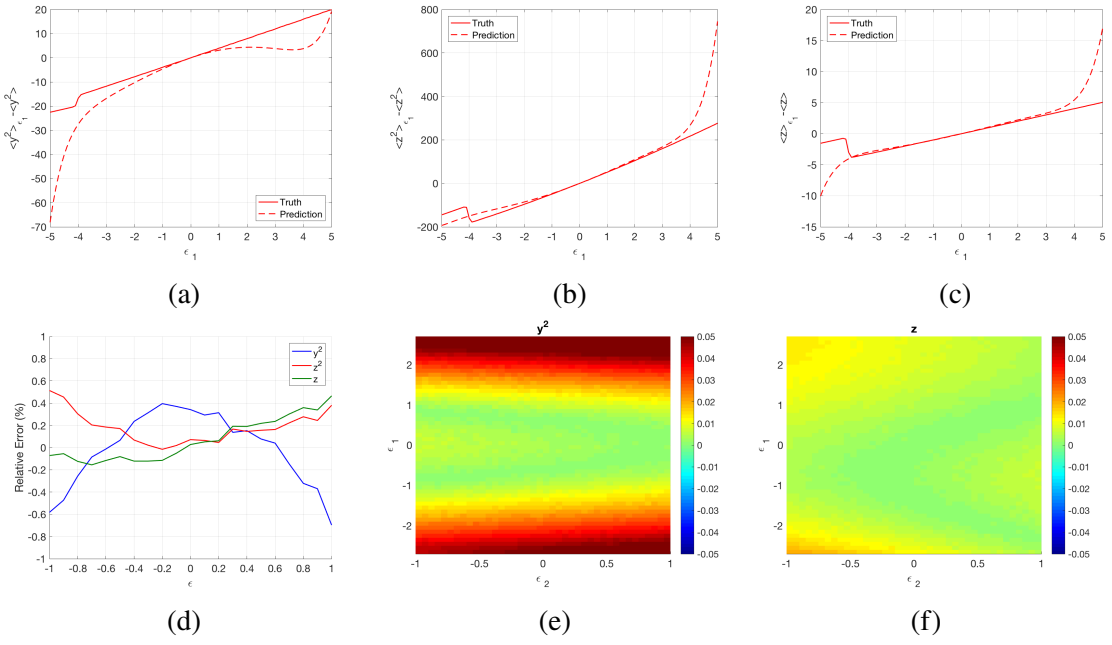
each case and integrated for 10^5 time units and with a time step of 10^{-2} time units, after the removal of transient trajectories of 10^3 time units. These type of ensemble integrations served both to calculate the unperturbed and perturbed expectation values and the linear responses, whose standard errors are not shown here since they are below 1%. The values of $\varepsilon_1 = \pm 0.05$ and $\varepsilon_2 = \pm 0.05$ were also considered in the empirical estimation of the linear response. When only $\mathcal{L}_1$ is applied, the predicted perturbed expectation values using Eq. (3.21) are shown in columns five to eight of tables 3.2-3.4. The perturbed expectation values as results of applying simultaneous forcings $\varepsilon_1 \mathbf{G}_1$ and $\varepsilon_2 \mathbf{G}_2$ are shown in columns nine to twelve of the same tables, demonstrating the predictive skill of the formulas. More importantly, the linear responses with respect to $\mathcal{L}_1$ and $\mathcal{L}_2$ are shown in the last eight columns of tables 3.2-3.4, while the combined linear response is given by the sum, by virtue of the additivity. We highlight that the computations shown here agree with those of prior works where, instead, the authors resort to trajectory-wise algorithms for the computation of the linear response [Wan13]; [NW17].

Increasing resolution yielded more precise results in the estimation of expectation values, although this was not the case when computing the linear responses for $N = 15$ and $N = 18$ in tables 3.2-3.4. The reason is that a high box-resolution needs of exponentially more sample points on phase space, in detriment of invariant measure approximation. We observe, furthermore, that the accuracy of such approximations heavily depend on the choice of observable and perturbation. While the linear and total responses of z present a good agreement with the full model, the rest of observable's linear responses become dependent on the perturbation chosen. Not so when calculating the full responses where, in relative terms, the performance is better. We highlight that the Markov matrix $\mathcal{M}$ calculated from a time series will not capture the effects of the stable directions, necessary for the computation of Ruelle's formula (1.9), and (1.11) in particular. Indeed, a stochastic matrix estimated with respect to the Lebesgue measure (see Remark 3.1.1) encodes also the effects of both stable and unstable directions and would yield a set of eigenvalues that contain the those of $\mathcal{M}$ [TLD18]. We therefore, conclude that tables 3.2-3.4 contain the information relative to the unstable directions observed in long integrations of the system. More work is, hence, needed to discern the way applied fields project along the stable and unstable directions of the flow.



(a) (b)

Figure 3.5: The coarse-grained Lorenz 63 invariant measure is shown on figure (a) for a box-resolution of $N = 15$. Relative occupancy is colour-coded. Using the same level of coarse-graining, figure (b) shows the discretised operator $\mathcal{L}_1$ applied onto the invariant measure. Positive (negative) values show the direction of increase (depletion) of probability due to the perturbation.



(a) (b) (c)

(d) (e) (f)

Figure 3.6: Panels (a), (b) and (c) are the expected values of the observables y^2, z^2 and z, respectively, computed using Eq. (3.21) (dashed line) vs. empirically obtained means (solid line) as a function of ε_1. The relative error is shown on figure (d). Relative error $(\%)$ of the prediction of the expected values of the observables indicated in the plots. Figures (e) and (f) are 2-dimensional errors for y^2 and z as a function of ε_1 and ε_2.

Table 3.2: Expectation values (first twelve columns) and linear responses (last eight columns) of the observables x^2, y^2, z^2 and z with respect to the unperturbed and perturbed ($\varepsilon_1 = 0.1$ and $\varepsilon_2 = 0.1$) invariant measure. In the first row empirical time averages were used whereas the rest indicate the expectation values obtained by means of evaluating Eq. (3.21) with a transition matrix of size $N \times N$, and transition time $\tau = \mathrm{d}t$.

	$\langle x^2 \rangle$	$\langle y^2 \rangle$	$\langle z^2 \rangle$	$\langle z \rangle$	$\langle x^2 \rangle_{\varepsilon_1}$	$\langle y^2 \rangle_{\varepsilon_1}$	$\langle z^2 \rangle_{\varepsilon_1}$	$\langle z \rangle_{\varepsilon_1}$	$\langle x^2 \rangle_{\varepsilon_1,\varepsilon_2}$	$\langle y^2 \rangle_{\varepsilon_1,\varepsilon_2}$
L-63	62.7975	81.2079	628.9203	23.5490	63.0685	81.5965	633.9853	23.6507	62.9642	81.4554
$N = 12$	63.3087	82.3652	629.7495	23.5497	63.9594	82.6604	634.0191	23.6397	63.8450	82.4960
$N = 15$	62.9322	81.4755	629.2876	23.5527	63.6321	81.8329	634.3466	23.6618	63.5166	81.6583
$N = 18$	62.8354	81.2423	629.2227	23.5542	63.4534	81.4851	634.3259	23.6622	63.3296	81.2907
	$\langle z^2 \rangle_{\varepsilon_1,\varepsilon_2}$	$\langle z \rangle_{\varepsilon_1,\varepsilon_2}$	$\delta^{(1)}_{\varepsilon_1}[x^2]$	$\delta^{(1)}_{\varepsilon_1}[y^2]$	$\delta^{(1)}_{\varepsilon_1}[z^2]$	$\delta^{(1)}_{\varepsilon_1}[z]$	$\delta^{(1)}_{\varepsilon_2}[x^2]$	$\delta^{(1)}_{\varepsilon_2}[y^2]$	$\delta^{(1)}_{\varepsilon_2}[z^2]$	$\delta^{(1)}_{\varepsilon_2}[z]$
L-63	633.8258	23.6491	2.6060	3.886	50.50	1.008	-1.006	-1.478	-1.172	-0.002
$N = 12$	633.4111	23.6260	6.488	3.022	42.386	0.894	-1.143	-1.651	-6.103	-0.138
$N = 15$	633.8201	23.6505	6.985	3.656	50.272	1.087	-1.147	-1.743	-5.257	-0.113
$N = 18$	633.6526	23.6526	6.215	2.608	50.802	1.080	-1.216	-1.922	-4.737	-0.094

90

Table 3.3: Expectation values (first twelve columns) and linear responses (last eight columns) of the observables x^2, y^2, z^2 and z with respect to the unperturbed and perturbed ($\varepsilon_1 = 0.1$ and $\varepsilon_2 = 0.1$) invariant measure. In the first row empirical time averages were used whereas the rest indicate the expectation values obtained by means of evaluating Eq. (3.21) with a transition matrix of size $N \times N$, and transition time $\tau = 5 \cdot \mathrm{d}t$.

	$\langle x^2 \rangle$	$\langle y^2 \rangle$	$\langle z^2 \rangle$	$\langle z \rangle$	$\langle x^2 \rangle_{\varepsilon_1}$	$\langle y^2 \rangle_{\varepsilon_1}$	$\langle z^2 \rangle_{\varepsilon_1}$	$\langle z \rangle_{\varepsilon_1}$	$\langle x^2 \rangle_{\varepsilon_1,\varepsilon_2}$	$\langle y^2 \rangle_{\varepsilon_1,\varepsilon_2}$
L-63	62.7975	81.2079	628.9203	23.5490	63.0685	81.5965	633.9853	23.6507	62.9642	81.4554
$N = 12$	63.3154	82.3735	629.7792	23.5502	63.9725	82.6588	634.0774	23.6409	63.8575	82.4928
$N = 15$	62.9284	81.4895	629.1764	23.5505	63.6315	81.8222	634.3041	23.6609	63.5154	81.6443
$N = 18$	62.8305	81.2722	629.0002	23.5500	63.4426	81.4444	634.1984	23.6596	63.3210	81.2436
	$\langle z^2 \rangle_{\varepsilon_1,\varepsilon_2}$	$\langle z \rangle_{\varepsilon_1,\varepsilon_2}$	$\delta^{(1)}_{\varepsilon_1}[x^2]$	$\delta^{(1)}_{\varepsilon_1}[y^2]$	$\delta^{(1)}_{\varepsilon_1}[z^2]$	$\delta^{(1)}_{\varepsilon_1}[z]$	$\delta^{(1)}_{\varepsilon_2}[x^2]$	$\delta^{(1)}_{\varepsilon_2}[y^2]$	$\delta^{(1)}_{\varepsilon_2}[z^2]$	$\delta^{(1)}_{\varepsilon_2}[z]$
L-63	633.8258	23.6491	2.6060	3.886	50.50	1.008	-1.006	-1.478	-1.172	-0.002
$N = 12$	633.6270	23.6270	6.556	2.936	42.675	0.902	-1.148	-1.666	-6.190	-0.140
$N = 15$	633.7810	23.6478	7.029	3.453	50.928	1.100	-1.149	-1.768	-5.223	-0.110
$N = 18$	633.7781	23.6521	6.182	2.124	50.001	1.081	-1.179	-1.939	-4.202	-0.075

91

Table 3.4: Expectation values (first twelve columns) and linear responses (last eight columns) of the observables x^2, y^2, z^2 and z with respect to the unperturbed and perturbed ($\varepsilon_1 = 0.1$ and $\varepsilon_2 = 0.1$) invariant measure. In the first row empirical time averages were used whereas the rest indicate the expectation values obtained by means of evaluating Eq. (3.21) with a transition matrix of size $N \times N$, and transition time $\tau = 10 \cdot \mathrm{d}t$.

	$\langle x^2 \rangle$	$\langle y^2 \rangle$	$\langle z^2 \rangle$	$\langle z \rangle$	$\langle x^2 \rangle_{\varepsilon_1}$	$\langle y^2 \rangle_{\varepsilon_1}$	$\langle z^2 \rangle_{\varepsilon_1}$	$\langle z \rangle_{\varepsilon_1}$	$\langle x^2 \rangle_{\varepsilon_1,\varepsilon_2}$	$\langle y^2 \rangle_{\varepsilon_1,\varepsilon_2}$
L-63	62.7975	81.2079	628.9203	23.5490	63.0685	81.5965	633.9853	23.6507	62.9642	81.4554
$N = 12$	63.3167	82.3746	629.7816	23.5503	63.9828	82.6286	634.1533	23.6425	63.8667	82.4596
$N = 15$	62.9281	81.4919	629.1595	23.5501	63.6447	81.7800	634.4634	23.6643	63.5294	81.5974
$N = 18$	62.8317	81.2747	629.0003	23.5500	63.4431	81.3503	634.2905	23.6600	63.3220	81.1361
	$\langle z^2 \rangle_{\varepsilon_1,\varepsilon_2}$	$\langle z \rangle_{\varepsilon_1,\varepsilon_2}$	$\delta^{(1)}_{\varepsilon_1}[x^2]$	$\delta^{(1)}_{\varepsilon_1}[y^2]$	$\delta^{(1)}_{\varepsilon_1}[z^2]$	$\delta^{(1)}_{\varepsilon_1}[z]$	$\delta^{(1)}_{\varepsilon_2}[x^2]$	$\delta^{(1)}_{\varepsilon_2}[y^2]$	$\delta^{(1)}_{\varepsilon_2}[z^2]$	$\delta^{(1)}_{\varepsilon_2}[z]$
L-63	633.8258	23.6491	2.606	3.886	50.50	1.008	-1.006	-1.478	-1.172	-0.002
$N = 12$	633.5229	23.6284	6.652	2.648	43.403	0.917	-1.157	-1.694	-6.325	-0.142
$N = 15$	633.9628	23.6542	7.188	3.101	52.579	1.137	-1.134	-1.797	-5.008	-0.101
$N = 18$	633.9375	23.6547	6.207	1.523	50.59	1.063	-1.129	-1.979	-3.499	-0.050

92

3.4 Summary and Discussion

Projecting the transfer operator or Fokker-Planck equation onto a finite dimensional vector space using Ulam's method naturally gives stochastic or Markovian matrices that describe the dynamical system of interest up to finite precision. Although rigorous approximation results are scarce, we have provided numerical evidence of its usefulness in calculating certain dynamical invariants.

By considering perturbations of a generic finite phase space Markov chain, one can express the resulting invariant vector as a power series of differential matrices that provide the response at all orders of non-linearity, including the leading order term that is denominated as the linear response; see Eq. (3.21). The well-posedness of linear response in this sense boils down to determining the singularity of said differential matrix; see Proposition 3.2.1. A stability result is presented in Proposition 3.2.3, where the problem of having non-simple eigenvalues is overcome in comparison to the results of [Mit03]. Furthermore, the numerical conditioning of the linear response is described in terms of the eigenvalues of the chain; see Proposition 3.2.4.

Alongside previous works [Luc16]; [ADF18], it is possible to link the perturbation theory of stochastic matrices to dynamical systems via the projected transfer operator. The linear component in the perturbative expansion (3.21) indicates the coarse-grained sensitivity of a dynamical system to prescribed perturbations, in the same way Eq. (2.23) does in general. We examined the Fokker-Planck/Liouville equation to identify the operators that incur the perturbations on the evolution of densities. Then, by considering simple finite difference methods, we were able to model (to finite-precision) the flow of probability that results from the perturbation. This gives matrix perturbations that allowed to exploit the perturbative formulas giving us access to the linear and non-linear response of the systems. Notice that using this method, we only need one integration of the (unforced) model to determine the response and sensitivity.

The two numerical experiments performed in sections 3.3.1 and 3.3.2 were intrinsically different. While the O-U process possesses a smooth invariant measure, the Lorenz 63 model is a dissipative deterministic system with an singularly hyperbolic attractor of zero Lebesgue-measure. Along these lines, we highlight that the partitioning of phase space naturally induces artificial diffusion [FJK13], which can smoothen the sampled distribution and, hence, making the calculations better posed [FR02]. In either example, the unperturbed models were integrated in order to sample their respective climatologies and to construct the projected transfer operator. Finite differences were, then, constructed taking into account the geometry of the partition— see bulleted list above Eq. (3.49)— and the estimated linear responses agreed with the empirically calculated values, as seen in tables 3.1-3.4. We observed dependence on the chosen observable and resolution in the Lorenz 63 case study, and could be attributable to the fact that the stable directions of the

flow are not sampled using this methodology; see also [TLD18].

One is led to think that, at a coarse-grained level, the more mixing the stochastic matrix is, the more robust it is to forcing in views of Eq. (3.31). However, we noted that the mixing rate given by the modulus of the second eigenvalue is deceiving in determining the convergence rates as shown in Figure 3.4 (a), where transient growth is observed for small values of time. This because the Lorenz 63 model is dissipative and, thus, the projected semigroup is prone to be non-normal and likely to reveal sub-unitary complex eigenvalues with numerical ranges exceeding the unit ball. This suggests that the mixing rate of the Markov chain may not be enough to quantify the sensitivity of the chains, but one should account for the non-normality of the underlying semigroup generator or associated stochastic matrix, the very least.

Chapter 4

Reduced-Order Dynamical Models

The projection of transfer operator semigroups via Markov chains is a costly procedure that is restricted to low-dimensional phase spaces. It, therefore, remains to answer what happens when physical domains are projected onto a set of relevant variables. Recent works have been devoted to this question [Che+14]; [Tan+20], although we shall here focus on the similar problem of *parametrising* unresolved, subgrid variables that are neglected by coarse-graining.

Many theoretically rigorous parametrisations have been devised, that are broadly divided into *top-down* and *data-driven* approaches: the former aim at deriving the parametrisations by applying suitable approximations to the equations describing the dynamics of the whole system, see e.g. [GM13]; [MTV01]; [WL12], while the latter are built by constructing a statistical-dynamical model of the impact of unresolved scales on the ones of interest. In this chapter, we will discuss and compare the properties of the Wouters-Lucarini (WL) top-down equation [WL12] and of the Empirical Model Reduction (EMR) data-driven parametrisation [KKG05]. We will also see when and how the integro-differential equation occurring in the WL parametrisation can be recast into a set of Markovian stochastic differential equations (SDEs). In other words, we investigate the *quasi-Markovianity* of the latter parametrisation [Pav14].

The two aforementioned methodologies are conceptually and practically different, even though the ultimate goal is to provide a computationally practical approximation for the Mori-Zwanzig or GLE integro-differential equation mentioned in the Introduction and to be revisited more rigorously in Section 4.1. In other words, both approaches— top-down and data-driven— provide fluctuations in the form of stochastic noise and memory effects determined by an integral kernel. On the one hand, the WL approach assumes prior knowledge about the decoupled hidden dynamics but no information about the statistical properties of the coupled system, in the lines of linear response theory. The empirical approach, on the other hand, samples the observed variables evolving according to the latter. The structure of Multilevel Stochastic Models (MSMs) that generalise EMRs

[KCG15] allows one, moreover, to derive explicit formulas for the fluctuating and memory correction terms that parametrise the influence of hidden processes.

The overall goal of this chapter is to provide a conceptual and analytical link between these two approaches aiming, first, to buttress the practical relevance of the WL perturbative approach and, on the other, to provide further insight into the well-documented robustness of the EMR method. Moreover, we will clarify how multilevel systems arise from both the top-down (WL) and the bottom-up (EMR) approaches. The present chapter explores the complete set of boxes and explains all the arrows in Fig. 4.1. The diagram in the figure shows that, starting from the top box, one can arrive at a memory equation, via top-down or bottom-up methods, as indicated by the left and right sequence of arrows, respectively.

This chapter is based on a recent publication [SG+21] and is structured as follows. The Mori-Zwanzig formalism for projected equations of motion is recalled in Section 4.1, highlighting the need for non-Markovian modelling. Section 4.2 revisits the derivation of the WL parametrisation method by applying the Dyson expansion— already defined in Section 2.2— to the Koopman operator associated with two weakly coupled dynamical systems. We show in Section 4.2.1.1 that such an expansion need not be truncated for additively coupled models and consider more general coupling laws than those in [WL13]. Furthermore, we study the problem of finding Markovian representations of the memory equation in the WL approach based on the spectral decomposition of the Koopman semigroup. Specifically, Theorem 4.2.1 shows, in the case of a scalar observable, how to recast the stochastic integro-differential equation arising in the WL parametrisation as a multilevel Markovian stochastic system involving explicitly the spectral elements of the (uncoupled) Koopman operator, and we point out in Remark 4.2.5 how such a Markovianisation extends to the multidimensional setting. Section 4.4 provides new insights into the Markovian representation adopted in the MSM framework; these insights help one to determine, in particular, the number of levels required for EMR to converge. Finally, Section 4.5 presents a comparison of the data-driven and top-down parametrisation approaches using a simple conceptual stochastic climate model.

The present chapter has supporting information included in a number of appendices. Appendix C discusses the stochastic Itô integration of the elementary form of an MSM. Appendix D shows further details on how EMR approaches can capture the dynamical properties of partially observed systems; in it, we consider a simple climate model, obtained by coupling the Lorenz atmospheric model [Lor84] and the Lorenz 63 convection model [Lor63], already studied in Section 3.3.2.

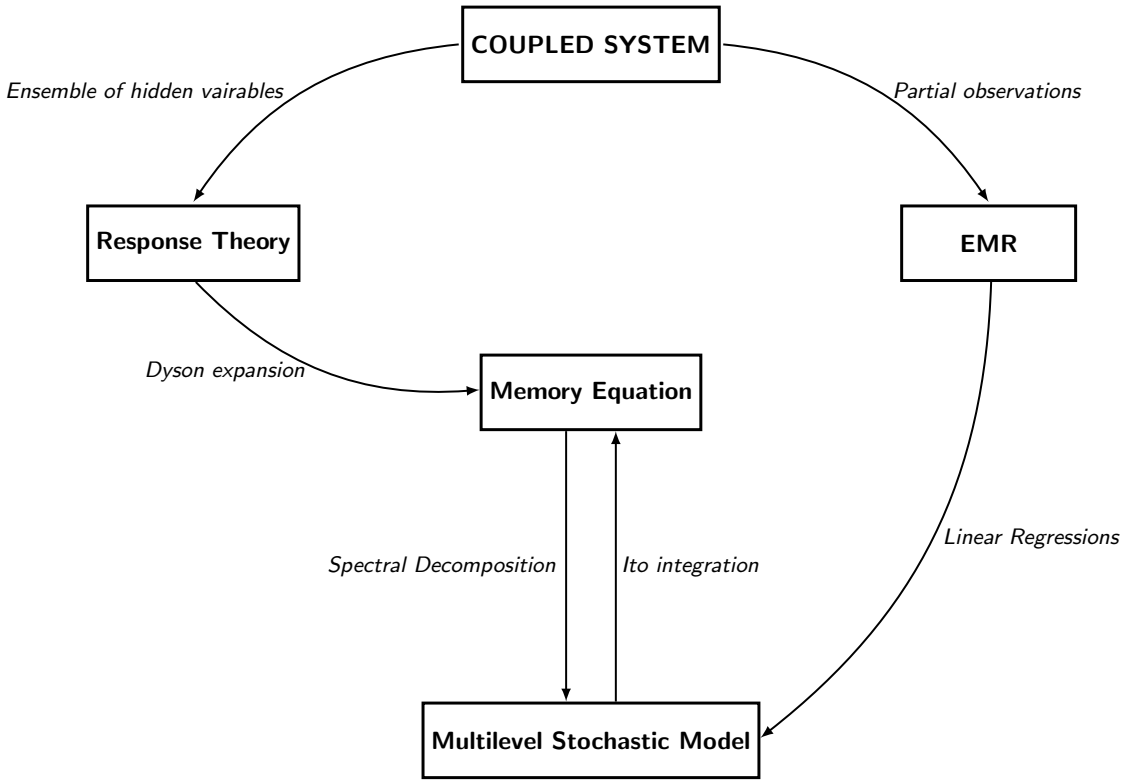


Figure 4.1: Schematic view of the two complementary approaches studied in this chapter. The arrows on the left-hand side indicate top-down, perturbative parametrisations; on the right, they refer to bottom-up, empirical parametrisations.

4.1 The Mori-Zwanzig Formalism

Let us reformulate the problem of constructing parametrisations as a projection of the dynamics onto the set of resolved variables. By working at the level of observables, Mori [Mor65] and Zwanzig [Zwa61] showed that the evolution laws for the projected dynamics incorporate a deterministic term that would be obtained by neglecting altogether the impact of the unresolved variables, to which a stochastic and a non-Markovian correction had to be added. A. J. Chorin and co-authors [CHK02]; [CL15] played an important role in developing further these ideas and applying them to several important problems. We briefly recapitulate below the Mori-Zwanzig projection operator approach.

Formally, let Ψ denote a generic observable— possibly in C_b— defined on a state space viewed as the product of two finite-dimensional spaces $\mathcal{X} \times \mathcal{Y}$, with variables $\mathbf{x}$ in $\mathcal{X}$ and $\mathbf{y}$ in $\mathcal{Y}$ being the resolved and unresolved variables, respectively. Next, let us define $\mathbb{P}$ to be a projector onto functions depending only on the target variables in $\mathcal{X}$, with the complementary projector on the unresolved variables being defined by $\mathbb{Q} = 1 - \mathbb{P}$.

Given a smooth flow ϕ_t on $\mathcal{X} \times \mathcal{Y}$ arising as solutions of Eq. (1.1), its action on smooth observables $\Psi = \Psi(\mathbf{x}, \mathbf{y})$ was introduced in Section 2.1 via the Koopman operator semigroup $\{\mathcal{U}_t\}_{t\geq 0}$; see also Section 4.2 below for further details on context of coupled

systems. We wish to recall that $\mathcal{U}_t$ is represented formally as the exponential of $\mathcal{L}^*$, $\mathcal{U}_t = e^{t\mathcal{L}^*}$, although it is rigorously justified by operator semigroup theory [EN00]; [Paz12]. With this representation, $\mathcal{L}^*$ satisfies the following identities

$$\partial_t\big(\mathcal{U}_t\Psi\big) = \mathcal{L}^* e^{t\mathcal{L}^*}\Psi = e^{t\mathcal{L}^*}\mathcal{L}^*\Psi = e^{t\mathcal{L}^*}\left(\mathbb{P}+\mathbb{Q}\right)\mathcal{L}^*\Psi \tag{4.1a}$$

$$= e^{t\mathcal{L}^*}\mathbb{P}\mathcal{L}^*\Psi + e^{t\mathbb{Q}\mathcal{L}^*}\mathbb{Q}\mathcal{L}^*\Psi + \int_0^t e^{(t-s)\mathcal{L}^*}\mathbb{P}\mathcal{L}^* e^{s\mathbb{Q}\mathcal{L}^*}\mathbb{Q}\mathcal{L}^*\Psi \mathrm{d}s, \tag{4.1b}$$

where we have employed the Dyson identity to obtain Eq. (4.1b), as was discussed earlier in Chapter 2, around Eq. (2.18). The first term in Eq. (4.1b) is the contribution of the resolved variables $\mathbf{x}$ alone to the instantaneous rate of change of Ψ. The second term models the fluctuating effects of the unresolved $\mathbf{y}$-variable by itself, while the third and last term represents, via an integral, the time-delayed influence upon $\mathbf{x}$ of its interactions with $\mathbf{y}$.

This formal calculation suggests that any closed model for the $\mathbf{x}$-variables should incorporate a fluctuating term to account for the $\mathbf{y}$ contributions and a memory or integral term for the $\mathbf{y}$–$\mathbf{x}$ interactions. Unfortunately, the Mori-Zwanzig equation (4.1b)— also known as the Generalised Langevin Equation (GLE) in the context of equilibrium systems [Pav14]— does not provide explicit analytic formulas to determine each of the three summands in the RHS of Eq. (4.1b). Hence, we need efficient ways to approximate such an equation.

In the limit of perfect timescale separation between the $\mathbf{x}$- and $\mathbf{y}$-variables, the non-Markovian term drops out and the fluctuating term can be represented as a— possibly multiplicative— white-noise term, thus recovering the basic results obtained via homogenisation theory [PS08]. When no such separation exists, however, one has to resort to finding an integral kernel beyond the abstract formulation of Eq. (4.1b); see, for instance, the theoretical ansatz based on perturbation expansions presented in [WL12]; [WL13] and discussed later in the chapter, and [KCG15]; [VL18a] for concrete applications.

In parallel with the theoretical approaches to approximate the Mori-Zwanzig equation (4.1b), data-driven methods have been proposed to model fluctuations and memory effects arising as a result of projecting a large state space onto (much) smaller subspaces. To this end, the work in [KCG15] provides a rigorous connection between the Mori-Zwanzig equation (4.1b) and multilevel regression models that were initially introduced for climatological purposes in [KKG05]. More recently, [LL21] proposed Nonlinear Auto-Regressive Moving Average with exogenous input models as a data-driven methodology that is comparable with the Mori-Zwanzig formalism and applied such models to the deterministic Kuramoto-Sivashinsky equation and to a stochastic Burgers equation. The complementarity of theory-based and data-driven model reduction methods in the absence of timescale separation is very well documented in [LL21] as well.

Efforts at using ingeniously selected basis functions as a stepping stone in data-driven methods for model reduction, effective simulation with partial data, and even prediction are multiple. Thus, the eigenvalues and eigenfunctions of the Koopman and transfer operators have been used to capture the modes of variability of the underlying flow, regardless of the latter being deterministic or stochastic; see, respectively, [Bal00] or [Tan+20]. Dynamic Mode Decomposition [Sch10, DMD] allows one to reconstruct from observations the eigenvectors and eigenvalues of the Koopman operator for observables of interest even in high-dimensional dynamical systems [Mez05]; [Row+09]. The latter approach is complementary to the one presented herein because we shall use the eigenvectors of the Koopman operator to build the projected dynamics for the observables of interest, which can then be rewritten as a multilevel Markovian stochastic model. Examples of other types of selection of dynamically interesting and effective bases are multichannel singular-spectrum analysis analysis [Ghi+02, MSSA] and data-adaptive harmonic decomposition [CK17]; [KCG18, DAH-MSLM]. A more thorough discussion of the complementary approaches involved is beyond the scope of this chapter.

4.2 Revisiting the Weak-Coupling Parametrisation

To study dynamical systems in which one can separate the variables into two groups with weak coupling between the two, one often resorts to parametrisations of the effects of one group on the other. In the limit of weak coupling, the coupling itself can be treated as a perturbation of the main dynamics [WL12]. Granted such an assumption and some degree of structural stability of the system, one can apply response theory to derive explicit stochastic and memory terms to describe the impact of the variables we want to neglect on those of interest, in the Mori-Zwanzig spirit. Note that, to do so, no assumption on timescale separation between the two groups of variables is necessary. This point is particularly relevant in the context of climate, where no clear timescale separation is observed as pointed out in the Introduction. This implies that asymptotic expansions of the kind used in homogenisation theory— see [PS08]— are of limited utility.

4.2.1 Responses Due to Coupling

Here, using a perturbative approach, we review the derivation of the parametrisation presented in [WL12]; [WL13]. Formally, we want to couple two dynamical systems generated independently by two vector fields $\mathbf{F} : \mathcal{X} \subseteq \mathbb{R}^{d_1} \longrightarrow \mathcal{X}$ and $\mathbf{G} : \mathcal{Y} \subseteq \mathbb{R}^{d_2} \longrightarrow \mathcal{Y}$ with possibly $d_1 \neq d_2$ and typically $d_1 \ll d_2$. We study a broad class of systems of the

form:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)) + \varepsilon \mathbf{C}^{\mathbf{x}}_{\mathbf{x}}(\mathbf{x}(t)) : \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}(t)), \tag{4.2a}$$

$$\dot{\mathbf{y}}(t) = \mathbf{G}(\mathbf{y}(t)) + \varepsilon \mathbf{C}^{\mathbf{y}}_{\mathbf{x}}(\mathbf{x}(t)) : \mathbf{C}^{\mathbf{y}}_{\mathbf{y}}(\mathbf{y}(t)). \tag{4.2b}$$

The operation indicated by the colon $\mathbf{x} : \mathbf{y}$ denotes the Hadamard product that multiplies vectors or matrices component-wise. Here, four new vector fields have been introduced to model the coupling law, namely $\mathbf{C}^{\mathbf{x}}_{\mathbf{x}} : \mathcal{X} \longrightarrow \mathcal{X}$, $\mathbf{C}^{\mathbf{y}}_{\mathbf{x}} : \mathcal{X} \longrightarrow \mathcal{Y}$, $\mathbf{C}^{\mathbf{x}}_{\mathbf{y}} : \mathcal{Y} \longrightarrow \mathcal{X}$ and $\mathbf{C}^{\mathbf{y}}_{\mathbf{y}} : \mathcal{Y} \longrightarrow \mathcal{Y}$. The real parameter ε controls the strength of the coupling between the two groups of variables, $\mathbf{x}(t)$ and $\mathbf{y}(t)$, so that the $\mathbf{x}$- and $\mathbf{y}$-variables are uncoupled for $\varepsilon = 0$. We assume that the vector fields $\mathbf{F}$ and $\mathbf{G}$, as well as the coupling laws in Eq. (4.2) are such that the system possesses a global attractor. Furthermore, we assume throughout this chapter that this global attractor supports an physical invariant probability measure μ that describes the distribution of trajectories onto the global attractor, so that Birkhoff averages hold for sets with positive Lebesgue measure; see Theorem 1.1.1 and comments below.

The WL parametrisation views the coupling as an ε-perturbation of the otherwise independent $\mathbf{x}$- and $\mathbf{y}$-processes, with $\mathbf{x}$ the observed and $\mathbf{y}$ the hidden variables. One next assumes that the impacts of perturbations applied to these processes can be addressed using Ruelle's response theory [Rue97]; [Rue09], so that response formulas of Chapter 2 can be used to derive an effective equation for the $\mathbf{x}$-variables.

Taking the Mori-Zwanzig point of view, we wish to calculate the evolution of observables that depend on the observed variables $\mathbf{x}$ alone, $\Psi = \Psi(\mathbf{x}(t))$. The idea, following [WL13], is to perform a perturbative expansion of the differential operator $\mathcal{L}^*$ governing the evolution of Ψ under the action of the flow associated with Eq. (4.2). Denoting by $\Psi(\mathbf{x}, \mathbf{y}, t)$ the time evolution of a smooth observable Ψ in $\mathcal{C}_b(\mathcal{X} \times \mathcal{Y})$, the first step of this Dyson-like operator expansion reads as follows:

$$\partial_t \Psi = \mathcal{L}^* \Psi = (\mathcal{L}^*_0 + \varepsilon \mathcal{L}^*_1)\Psi, \tag{4.3}$$

exactly as Eq. (2.24) where, instead, the second order operators vanish because of the absence of noise. Here $\mathcal{L}^*_0$ and $\mathcal{L}^*_1$ account for the advective effects of the uncoupled and coupling terms, respectively, that compose the RHS of Eq. (4.2). Namely, looking back at Eq. (2.25) we deduce

$$\mathcal{L}^*_0 = \begin{bmatrix} \mathbf{F}(\mathbf{x}) \\ \mathbf{G}(\mathbf{y}) \end{bmatrix} \cdot \begin{bmatrix} \nabla_{\mathbf{x}} \\ \nabla_{\mathbf{y}} \end{bmatrix}; \tag{4.4a}$$

$$\mathcal{L}^*_1 = \begin{bmatrix} \mathbf{C}^{\mathbf{x}}_{\mathbf{x}}(\mathbf{x}) : \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}) \\ \mathbf{C}^{\mathbf{x}}_{\mathbf{x}}(\mathbf{x}) : \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}) \end{bmatrix} \cdot \begin{bmatrix} \nabla_{\mathbf{x}} \\ \nabla_{\mathbf{y}} \end{bmatrix}, \tag{4.4b}$$

in Eq. (4.4), $\nabla_{\mathbf{x}}$ and $\nabla_{\mathbf{y}}$ denote the vector differential operators with respect to the variables $\mathbf{x}$ and $\mathbf{y}$. Recalling Section 2.1, the solution operator of Eq. (4.3) is the Koopman operator. Formally, its dual counterpart acts on densities and it is the transfer operator [Bal00]; see, also, Section 2.1. Equation (4.3) is thus a transport law, where the physical quantity or observable is advected by the vector field on the RHS of Eq. (4.2).

Note that the operator formalism presented here in the deterministic dynamical systems setting —and the associated semigroup theory— extends to Markov diffusion processes driven by a stochastic forcing, as clarified in Chapter 2; see also [Tan+20]. In the latter case, we remind the reader that the transport equation (4.3) becomes the so-called backward-Kolmogorov equation (2.13) that describes the evolution of the expected value of observables. Loosely speaking, the corresponding extension amounts to adding a Laplacian-like operator to the advection operator $\mathcal{L}^*$; more details are given in Section 2.1.1.

More precisely, one associates to the solution $\Psi(\mathbf{x}, \mathbf{y}, t)$ of Eq. (4.3) unfolding from an "initial" observable $\Psi = \Psi(\mathbf{x}, \mathbf{y})$ at time $t = 0$, a family of linear Koopman operators indexed by time $\{\mathcal{U}_t\}_{t\geq 0}$ such that $\Psi(\mathbf{x}, \mathbf{y}, t) = \mathcal{U}_t\Psi(\mathbf{x}, \mathbf{y})$, for any $t \geq 0$ and $(\mathbf{x}, \mathbf{y})$ in $\mathcal{X} \times \mathcal{Y}$. These operators are defined— as mentioned already in connection with introducing the GLE (4.1b) and Sections 2.1 and 2.1.1— as exponentials of the operator $\mathcal{L}^*$, i.e., $\mathcal{U}_t = e^{t\mathcal{L}^*}$. This notation is formal as the operator $\mathcal{L}^*$ is unbounded; it is, however, useable as $\{\mathcal{U}_t\}_{t\geq 0}$ satisfies the semigroup property, i.e. $\mathcal{U}_{t+s} = \mathcal{U}_t\mathcal{U}_s$, $t, s \geq 0$, as for a standard exponential. Over the appropriate function space of observables Ψ, this family actually forms a strongly continuous contracting semigroup [EN06]; we refer to the functional setting employed in the definition of the semigroup generators around Eqs. (2.6) and (2.13).

The action of the flow on an observable Ψ becomes thus more transparent thanks to the operator $\mathcal{U}_t$, according to the equation

$$\mathcal{U}_t\Psi(\mathbf{x}_0, \mathbf{y}_0) = e^{t\mathcal{L}^*}\Psi(\mathbf{x}_0, \mathbf{y}_0) = \Psi(\mathbf{x}(t, \mathbf{x}_0), \mathbf{y}(t, \mathbf{y}_0)), \tag{4.5}$$

where $(\mathbf{x}(t, \mathbf{x}_0), \mathbf{y}(t, \mathbf{y}_0))$ denotes the system's solution at time t emanating from the initial state $(\mathbf{x}_0, \mathbf{y}_0)$ at time $t = 0$. In what follows, we omit the subscript 0 in $(\mathbf{x}_0, \mathbf{y}_0)$ but still take it as an initial state.

When the coupling parameter ε in system (4.2) is small, one can use formal perturbation expansions of the Koopman semigroup to better isolate and assess the coupling effects at the level of observables, in the same way it was done for the static response of the Fokker-Planck equation in Section 2.2. To do so, we follow here the Dyson expansion

(2.18) which in the present case reads as follows:

$$e^{t\mathcal{L}^*}\Psi(\mathbf{x},\mathbf{y}) = e^{t\mathcal{L}_0^*+t\varepsilon\mathcal{L}_1^*}\Psi(\mathbf{x},\mathbf{y}) \tag{4.6a}$$

$$= e^{t\mathcal{L}_0^*}\Psi(\mathbf{x},\mathbf{y}) + \varepsilon\int_0^t e^{s\mathcal{L}^*}\mathcal{L}_1^* e^{(t-s)\mathcal{L}_0^*}\Psi(\mathbf{x},\mathbf{y})\mathrm{d}s \tag{4.6b}$$

$$= e^{t\mathcal{L}_0^*}\Psi(\mathbf{x},\mathbf{y}) + \varepsilon\int_0^t e^{(t-s)\mathcal{L}_0^*}\mathcal{L}_1^* e^{s\mathcal{L}^*}\Psi(\mathbf{x},\mathbf{y})\mathrm{d}s, \tag{4.6c}$$

and it yields the following expansion of the Koopman operator in ε:

$$e^{t\mathcal{L}^*}\Psi(\mathbf{x},\mathbf{y}) = e^{t\mathcal{L}_0^*}\Psi(\mathbf{x},\mathbf{y}) + \varepsilon\int_0^t e^{(t-s)\mathcal{L}_0^*}\mathcal{L}_1^* e^{s\mathcal{L}_0^*}\Psi(\mathbf{x},\mathbf{y})\mathrm{d}s + \mathcal{O}\left(\varepsilon^2\right). \tag{4.7}$$

This identity shows that the evolution of a generic observable can be described as an ε-perturbation of its decoupled evolution according to $\mathcal{L}_0^*$. We note, again, that these expansions are purely formal and in particular it is not clear in which sense this expansion might converge. For a bounded perturbation operator $\mathcal{L}_1^*$, it would be straightforward to prove boundedness of the resulting perturbed semigroup. However, $\mathcal{L}_1^*$ here is a differential linear operator, for which direct estimates are more laborious; see comments immediately below Eq. (2.18) and Remark 2.2.2. Leaving alone the functional analysis framework that would make such an expansion rigorously convergent, we shall use nevertheless the expansion (4.7) throughout this rest of the present chapter.

The objective now is, using this operator expansion, to derive an effective reduced-order model for the evolution of the $\mathbf{x}$-variable without having to resolve the $\mathbf{y}$-process. We start observing the system at $t = 0$, but assume that it has already attained a steady state. Since we are only concerned with observables depending solely on the $\mathbf{x}$-variables, we formulate now an evolution equation for such observables. To do so, we consider first the adjoint Liouville equation (4.3) for a generic $\mathbf{y}$-independent observable Ψ, this is, $\Psi(\mathbf{x},\mathbf{y}) = \Psi(\mathbf{x})$, for every $\mathbf{x}$ and $\mathbf{y}$. For such an observable, at the time we start observing the coupled system, Eq. (4.3) reduces to

$$\partial_t\left(e^{t\mathcal{L}^*}\Psi\right)|_{t=0} = \left[\mathbf{F}(\mathbf{x}) + \varepsilon\mathbf{C}_{\mathbf{x}}^{\mathbf{x}}(\mathbf{x}) : \mathbf{C}_{\mathbf{y}}^{\mathbf{x}}(\mathbf{y})\right]\cdot\nabla_{\mathbf{x}}\Psi. \tag{4.8}$$

Equation (4.8) illustrates the trivial fact that the time evolution in Eq. (4.2) of an $\mathbf{x}$-dependent physical quantity is also affected by the $\mathbf{y}$-variables.

Following [WL12]; [WL13], the decoupled equations are assumed to have been evolving for some time prior to the coupling. Hence, we have to formally parametrise the evolution of the $\mathbf{C}_{\mathbf{y}}^{\mathbf{x}}(\mathbf{y})$-contribution to the vector field which is, ultimately, a vector-valued observable. We do so by introducing an extended version of the Koopman operators that act on vectors component-wise, rather than just on real-valued observables. Consider $\mathbf{v} : \mathcal{X}\times\mathcal{Y} \longrightarrow \mathbb{R}^d$, for some positive integer d, and define the action of the Koopman

operator $e^{t\mathcal{L}^*}$ on $\mathbf{v}$ as:

$$\left[e^{t\mathcal{L}^*}\mathbf{v}(\mathbf{x},\mathbf{y})\right]_i = e^{t\mathcal{L}^*}\left[\mathbf{v}(\mathbf{x},\mathbf{y})\right]_i \tag{4.9}$$

for every $i = 1,\ldots,d$. The definition (4.9) will allow us to use the semigroup notation for observables of possibly different dimensions, all of which take their inputs in the phase space $\mathcal{X}\times\mathcal{Y}$. Ultimately, this is a component-wise evaluation of our extended Koopman operator family, and its generator can be obtained analogously by taking strong limits in each entry. As mentioned above, we have to model the effects of the coupling vector field $\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y})$, whose state at time $t = 0$ is the product of the evolution from time $-t$ to 0. We then have, with the dynamics starting at time $-t$ and initial state $(\mathbf{x}_0,\mathbf{y}_0)$,

$$\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}) = e^{t\mathcal{L}^*}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0,-t) = e^{t\mathcal{L}^*}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0)\,. \tag{4.10}$$

Now, by using the perturbative expansions in Eqs. (4.6b)-(4.7), we obtain:

$$\begin{aligned}
\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}) &= e^{t\mathcal{L}^*}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0) = e^{t\mathcal{L}^*_0+t\varepsilon\mathcal{L}^*_1}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0) && \text{(4.11a)}\\
&= e^{t\mathcal{L}^*_0}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0) + \varepsilon\int_0^t e^{s\mathcal{L}^*}\mathcal{L}^*_1 e^{(t-s)\mathcal{L}^*_0}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0)\,\mathrm{d}s && \text{(4.11b)}\\
&= e^{t\mathcal{L}^*_0}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0) + \varepsilon\int_0^t e^{(t-s)\mathcal{L}^*_0}\mathcal{L}^*_1 e^{s\mathcal{L}^*}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0)\,\mathrm{d}s && \text{(4.11c)}\\
&= e^{t\mathcal{L}^*_0}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0) + \varepsilon\int_0^t e^{(t-s)\mathcal{L}^*_0}\mathcal{L}^*_1 e^{s\mathcal{L}^*_0}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0)\,\mathrm{d}s + \mathcal{O}\left(\varepsilon^2\right). && \text{(4.11d)}
\end{aligned}$$

Plugging the identity in Eq. (4.11c) into Eq. (4.8), we find the following expression:

$$\begin{aligned}
\partial_t\left(e^{t\mathcal{L}^*}\Psi\right)|_{t=0} = &\left[\mathbf{F}(\mathbf{x}) + \varepsilon\mathbf{C}^{\mathbf{x}}_{\mathbf{x}}(\mathbf{x}) : \left\{e^{t\mathcal{L}^*_0}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0)\right\}\right]\cdot\nabla_{\mathbf{x}}\Psi\\
&+\left[\varepsilon\mathbf{C}^{\mathbf{x}}_{\mathbf{x}}(\mathbf{x}) : \left\{\int_0^t e^{s\mathcal{L}^*}\mathcal{L}^*_1 e^{(t-s)\mathcal{L}^*_0}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0)\,\mathrm{d}s\right\}\right]\cdot\nabla_{\mathbf{x}}\Psi.
\end{aligned} \tag{4.12}$$

This equation is an exact reformulation of the problem induced by Eq. (4.8). This reformulation demonstrates that memory effects enter at second order in powers of the coupling parameter. Notice, though, that even if Eq. (4.12) reduces the dimensionality of the problem from $d_1 + d_2$ to d_1, it does not constitute an approximation for the evolution of Ψ as an observable of $\mathbf{x}$ alone, since it depends on the evolution of the $\mathbf{y}$-variables in the coupled regime by means of the action of $e^{s\mathcal{L}^*}$ onto $\mathcal{L}^*_1$. Therefore, we need to perform a further approximation by considering Eq. (4.11d) instead, which leads to a second order response formula:

$$\begin{aligned}
\partial_t\left(e^{t\mathcal{L}^*}\Psi\right)|_{t=0} \simeq &\left[\mathbf{F}(\mathbf{x}) + \varepsilon\mathbf{C}^{\mathbf{x}}_{\mathbf{x}}(\mathbf{x}) : \left\{e^{t\mathcal{L}^*_0}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0)\right\}\right]\cdot\nabla_{\mathbf{x}}\Psi\\
&+\left[\varepsilon\mathbf{C}^{\mathbf{x}}_{\mathbf{x}}(\mathbf{x}) : \left\{\int_0^t e^{(t-s)\mathcal{L}^*_0}\mathcal{L}^*_1 e^{s\mathcal{L}^*_0}\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0)\,\mathrm{d}s\right\}\right]\cdot\nabla_{\mathbf{x}}\Psi,
\end{aligned} \tag{4.13}$$

where the terms of order ε^3 have been dropped. In some sense, Eq. (4.13) is our equivalent of the Dyson approximation for interactions quantum electrodynamics, where such interactions take place between the generic $\mathbf{x}$ and $\mathbf{y}$ variables. Furthermore, Eq. (4.13) approximates the evolution of the $\mathbf{x}$-variables with no need for the evolution of the $\mathbf{y}$-variables in the coupled regime, in line with the theory of linear response. This result amounts to saying that— by observing only the statistical properties of the decoupled dynamics of the $\mathbf{y}$-process, obtained with $\varepsilon = 0$— one can construct a Markovian contribution

$$\mathbf{C}^{\mathbf{x}}_{\mathbf{x}}(\mathbf{x}) : \Big\{ e^{t\mathcal{L}_0^*} \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0) \Big\}, \tag{4.14}$$

and a non-Markovian contribution

$$\mathbf{C}^{\mathbf{x}}_{\mathbf{x}}(\mathbf{x}) : \Big\{ \int_0^t e^{(t-s)\mathcal{L}_0^*} \mathcal{L}_1^* e^{s\mathcal{L}_0^*} \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0)\, \mathrm{d}s \Big\}, \tag{4.15}$$

to the dynamics of the $\mathbf{x}$-variables that is able to describe the delayed impact of the coupling and involves integrating a memory kernel. Expanding the kernel $\tilde{\mathcal{K}}$ of the integral correction, we get:

$$\begin{aligned}
\tilde{\mathcal{K}}(t, s, \mathbf{x}_0, \mathbf{y}_0) :=& e^{(t-s)\mathcal{L}_0^*} \mathcal{L}_1^* e^{s\mathcal{L}_0^*} \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0) && \text{(4.16a)}\\
=& e^{(t-s)\mathcal{L}_0^*} \Big[\mathbf{C}^{\mathbf{x}}_{\mathbf{x}}(\mathbf{x}_0) : \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0) \Big] \cdot \nabla_{\mathbf{x}} e^{s\mathcal{L}_0^*} \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0) && \text{(4.16b)}\\
&+ e^{(t-s)\mathcal{L}_0^*} \Big[\mathbf{C}^{\mathbf{y}}_{\mathbf{x}}(\mathbf{x}_0) : \mathbf{C}^{\mathbf{y}}_{\mathbf{y}}(\mathbf{y}_0) \Big] \cdot \nabla_{\mathbf{y}} e^{s\mathcal{L}_0^*} \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0) && \text{(4.16c)}\\
=& \Big[e^{(t-s)\mathcal{L}_0^*} \Big(\mathbf{C}^{\mathbf{y}}_{\mathbf{x}}(\mathbf{x}_0) : \mathbf{C}^{\mathbf{y}}_{\mathbf{y}}(\mathbf{y}_0) \Big) \Big] \cdot \nabla_{\mathbf{y}} e^{s\mathcal{L}_0^*} \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}_0)\,. && \text{(4.16d)}
\end{aligned}$$

Note that the leading-order Koopman operator $e^{s\mathcal{L}_0^*}$ models the evolution of the observables in the uncoupled regime. Since there is no prior knowledge on initialising the coupled system at time $-t$, the initial state $\mathbf{y}_0$ in the hidden variables should be drawn from an ensemble, according to a probability density function. At this stage, there is freedom in the choice of such a prior. However, since we are assuming that the coupled system was initialised at time $-t$, it is natural to draw $\mathbf{y}_0$ according to the invariant measure ν associated with the dynamical system generated by the vector field $\mathbf{G}$ from Eq. (4.2). In effect, we wish to sample initial conditions from the coupled steady state, but do not assume any prior knowledge of the coupled statistics. As discussed in [WL12]; [WL13], we can take advantage of response theory to address this situation. Indeed, for any sufficiently smooth observable Ψ, we have the following perturbation expansion at hand (see also Eq. (2.23)):

$$\langle \Psi \rangle_\varepsilon := \int \Psi(\mathbf{x}, \mathbf{y}) \mu(\mathrm{d}\mathbf{x}, \mathrm{d}\mathbf{y}) = \langle \Psi \rangle_{\varepsilon=0} + \sum_{k=1}^{\infty} \varepsilon^k \delta^{(k)}[\Psi], \tag{4.17}$$

where $\langle \Psi \rangle_\varepsilon$ is the expectation value of Ψ in the coupled system (4.2), $\langle \Psi \rangle_{\varepsilon=0}$ is the

expectation value of Ψ according to the statistics generated by the uncoupled process obtained by setting $\varepsilon = 0$ in Eq. (4.2), and $\varepsilon^k \delta^{(k)}[\Psi]$ is the kth-order response. This formula is analogous to that in Eq. (2.23). In what follows, we remove the subscripts for the averages when $\varepsilon = 0$. Therefore, we have that the expected value of the coupling function reads as:

$$\left\langle \mathbf{C}_{\mathbf{y}}^{\mathbf{x}} \right\rangle_{\varepsilon} = \left\langle \mathbf{C}_{\mathbf{y}}^{\mathbf{x}} \right\rangle + \sum_{k=1}^{\infty} \varepsilon^k \delta^{(k)}[\mathbf{C}_{\mathbf{y}}^{\mathbf{x}}]. \tag{4.18}$$

Likewise, we can calculate the average of such function at time t:

$$\left\langle e^{t\mathcal{L}_0^*} \mathbf{C}_{\mathbf{y}}^{\mathbf{x}} \right\rangle_{\varepsilon} = \left\langle e^{t\mathcal{L}_0^*} \mathbf{C}_{\mathbf{y}}^{\mathbf{x}} \right\rangle + \sum_{k=1}^{\infty} \varepsilon^k \delta^{(k)}[e^{t\mathcal{L}_0^*} \mathbf{C}_{\mathbf{y}}^{\mathbf{x}}]. \tag{4.19}$$

Now, by letting $\tilde{\eta}(t, \mathbf{y}_0) = e^{t\mathcal{L}_0^*} \mathbf{C}_{\mathbf{y}}^{\mathbf{x}}(\mathbf{y}_0)$, we find that in order for the approximate statistics to agree up to second order in ε with the exact one, we only need to impose the following conditions upon the first two moments of the parametrised noisy fluctuations (see also [WL12] for the derivation):

$$\left\langle \tilde{\eta}(t, \mathbf{y}_0) \right\rangle = \int \nu\left(\mathrm{d}\mathbf{y}_0\right) \mathbf{C}_{\mathbf{y}}^{\mathbf{x}}\left(\mathbf{y}_0\right), \tag{4.20a}$$

$$\left\langle \tilde{\eta}(t, \mathbf{y}_0) \tilde{\eta}^{\top}(0, \mathbf{y}_0) \right\rangle = \int \nu(\mathrm{d}\mathbf{y}_0) e^{t\mathcal{L}_0^*} \mathbf{C}_{\mathbf{y}}^{\mathbf{x}}\left(\mathbf{y}_0\right) \left(\mathbf{C}_{\mathbf{y}}^{\mathbf{x}}\left(\mathbf{y}_0\right)\right)^{\top}. \tag{4.20b}$$

It follows that any stochastic noise $\eta(t)$ that satisfies the two conditions above will be suitable for parametrising the fluctuations in the $\mathbf{y}$-dynamics tied to the lack of knowledge in the initial state. Each of the entries in the correlation matrix given by Eq. (4.20b) is the correlation function between the components of the vector field $\mathbf{C}_{\mathbf{y}}^{\mathbf{x}}$ and these will become explicit provided a suitable spectral decomposition is at hand. Such a decomposition will be provided later in Section 4.2.2.

In the memory term, though, we neglect ε-corrections to its statistics since memory effects are of order ε^2 already. Thus, we have, by averaging the kernel $\tilde{\mathcal{K}}(t, s, \mathbf{x}_0, \mathbf{y}_0)$ in Eq. (4.16b) with respect to the ensemble of $\mathbf{y}$-variables, ν,

$$\mathcal{K}(t, s, \mathbf{x}) := \int \nu(\mathrm{d}\mathbf{y}_0) \left[e^{(t-s)\mathcal{L}_0^*} \left(\mathbf{C}_{\mathbf{x}}^{\mathbf{y}}(\mathbf{x}_0) : \mathbf{C}_{\mathbf{y}}^{\mathbf{y}}(\mathbf{y}_0) \right) \right] \cdot \nabla_{\mathbf{y}} e^{s\mathcal{L}_0^*} \mathbf{C}_{\mathbf{y}}^{\mathbf{x}}\left(\mathbf{y}_0\right) \tag{4.21a}$$

$$= \int \nu(\mathrm{d}\mathbf{y}_0) \left[e^{(t-s)\mathcal{L}_0^*} \mathbf{C}_{\mathbf{x}}^{\mathbf{y}}(\mathbf{x}_0) : e^{(t-s)\mathcal{L}_0^*} \mathbf{C}_{\mathbf{y}}^{\mathbf{y}}(\mathbf{y}_0) \right] \cdot \nabla_{\mathbf{y}} e^{s\mathcal{L}_0^*} \mathbf{C}_{\mathbf{y}}^{\mathbf{x}}\left(\mathbf{y}_0\right) \tag{4.21b}$$

$$= \int \nu(\mathrm{d}\mathbf{y}_0) \left[\mathbf{C}_{\mathbf{x}}^{\mathbf{y}}(\mathbf{x}(t-s)) : e^{(t-s)\mathcal{L}_0^*} \mathbf{C}_{\mathbf{y}}^{\mathbf{y}}(\mathbf{y}_0) \right] \cdot \nabla_{\mathbf{y}} e^{s\mathcal{L}_0^*} \mathbf{C}_{\mathbf{y}}^{\mathbf{x}}\left(\mathbf{y}_0\right). \tag{4.21c}$$

This way, the memory kernel only depends on the $\mathbf{x}$-variables and no information about the coupled invariant measure μ. Hence, we find a self-consistent evolution of the $\mathbf{x}$-variables, subject to the influence of unobserved variables $\mathbf{y}$, in the form of a stochastic

integro-differential equation (SIDE) resembling the GLE (4.1b):

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}) + \varepsilon \mathbf{C}^{\mathbf{x}}_{\mathbf{x}}(\mathbf{x}) : \eta(t) + \varepsilon^2 \mathbf{C}^{\mathbf{x}}_{\mathbf{x}}(\mathbf{x}) \cdot \int_0^t \mathcal{K}(t, s, \mathbf{x}) \mathrm{d}s, \tag{4.22}$$

where $\eta(t)$ is a stochastic forcing that agrees with the mean and correlation properties stated in Eq. (4.20). We emphasise that the solution $\mathbf{x}(t)$ of the original system of ordinary differential equations (4.2) does not satisfy Eq. (4.22): it is just the proposed reduced-order model for the $\mathbf{x}$-variables. The closure provided by expressing the corrections in the second and third term on the RHS of Eq. (4.22) as functions of $\mathbf{x}$ alone constitutes the WL parametrisation. There are two sources of error in the model proposed in Eq. (4.22). First, the truncation performed in the Dyson expansion neglects higher-order statistical responses which are weighted by the third power of the small coupling parameter. Secondly, averaging over the statistics of the uncoupled dynamics can also introduce errors. In fact, the nature of the stochastic correction is not fully determined except for its lagged correlation.

Note that there is considerable freedom in the choosing the noise, since we only require agreement up to the second moment. However, a direct consequence of this weak-coupling parametrisation is that realisations of the noise can be produced by directly integrating the decoupled hidden variables, or by representing it using simple autoregressive models [VL18b]. We are assuming here that the uncoupled dynamics leads to a noisy signal; this can be achieved either by the presence of stochastic forcing in the hidden variables [VL18b]; [Wou+16] or by their uncoupled dynamics being chaotic [VL18a].

The perturbation operator approach taken here is analogous to that of [WL13], who only considered the independent or additive-coupling cases; the latter is expanded below in Section 4.2.1.1. Here, though, we generalise further the parametrisation formulas that can be obtained via perturbative expansion of linear operators. In fact, the present approach can also be extended to weakly coupled systems of the form:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)) + \varepsilon \mathbf{C}^{\mathbf{x}}\left(\mathbf{y}(t)\right), \tag{4.23a}$$

$$\dot{\mathbf{y}}(t) = \mathbf{G}(\mathbf{y}(t)) + \varepsilon \mathbf{C}^{\mathbf{y}}\left(\mathbf{x}(t), \mathbf{y}(t)\right), \tag{4.23b}$$

where $\mathbf{C}^{\mathbf{y}}$ encodes interactions that need not be separable between the $\mathbf{x}$- and $\mathbf{y}$-variables in the hidden layer of the model. Note that the full parametrisation of arbitrary couplings was discussed by the two authors of [WL13] in a previous work [WL12], in which they directly employed the first and second order Green functions, although they had to resort to Schauder basis to express arbitrary couplings as a series of separable product on $\mathbf{x}$ and $\mathbf{y}$.

4.2.1.1 The Additive-Coupling Case

The approximate Dyson expansion given in Eq. (4.13) is exact in the case of additive coupling. Such systems take the form:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)) + \varepsilon \mathbf{C}^{\mathbf{x}}(\mathbf{y}(t)), \tag{4.24a}$$

$$\dot{\mathbf{y}}(t) = \mathbf{G}(\mathbf{y}(t)) + \varepsilon \mathbf{C}^{\mathbf{y}}(\mathbf{x}(t)). \tag{4.24b}$$

Indeed, letting $\mathbf{C}^{\mathbf{y}}(\mathbf{x}, \mathbf{y}) = \mathbf{C}^{\mathbf{y}}(\mathbf{x})$ in Eq. (4.23b) and using Eq. (4.11b) allows us to avoid the truncation of the Dyson expansion and yields the following expression for the memory term:

$$\tilde{\mathcal{K}}(t, s, \mathbf{x}, \mathbf{y}_0) = e^{s\mathcal{L}^*} \mathcal{L}_1^* e^{(t-s)\mathcal{L}_0^*} \mathbf{C}^{\mathbf{x}}(\mathbf{y}_0) \tag{4.25a}$$

$$= e^{s\mathcal{L}^*} \left(\mathbf{C}^{\mathbf{x}}(\mathbf{y}_0) \cdot \nabla_{\mathbf{x}} + \mathbf{C}^{\mathbf{y}}(\mathbf{x}) \cdot \nabla_{\mathbf{y}}\right) e^{(t-s)\mathcal{L}_0^*} \mathbf{C}^{\mathbf{x}}(\mathbf{y}_0) \tag{4.25b}$$

$$= e^{s\mathcal{L}^*} \left(\mathbf{C}^{\mathbf{y}}(\mathbf{x})\right) \cdot \nabla_{\mathbf{y}} e^{(t-s)\mathcal{L}_0^*} \mathbf{C}^{\mathbf{x}}(\mathbf{y}_0), \tag{4.25c}$$

which is exact. Next, taking averages with respect to ν, we obtain:

$$\mathcal{K}(t, s, \mathbf{x}) = \int \nu(\mathrm{d}\mathbf{y}_0) e^{s\mathcal{L}^*} \mathbf{C}^{\mathbf{y}}(\mathbf{x}) \cdot \nabla_{\mathbf{y}} e^{(t-s)\mathcal{L}_0^*} \mathbf{C}^{\mathbf{x}}(\mathbf{y}_0). \tag{4.26}$$

Hence, the parametrisation in this additive-coupling case is exact, as no terms proportional to ε^k, $k \geq 3$ are present. The only assumption made is that the statistics in the $\mathbf{y}$-variables have reached a steady state according to the unperturbed system. Finally, the full SIDE in this case takes the form:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}) + \varepsilon \eta(t) + \varepsilon^2 \int_0^t \mathcal{K}(t, s, \mathbf{x}) \mathrm{d}s, \tag{4.27}$$

where the stochastic process η has the mean and correlation properties given by Eq. (4.20). This equation is, thus, exactly the one obtained in [WL13].

4.2.2 Markovian Representation through Koopman Eigenfunctions

In the context of Langevin dynamics, there are known conditions on memory kernels that allow one to recast certain SIDEs into Markovian SDEs by means of extended variables; see [Pav14, Section 8.2]. Those stochastic processes are called quasi-Markovian [Pav14], although such a Markovianisation procedure is actually not limited to stochastic processes and it relies on the same type of ideas in other contexts; see [Che+11]; [CGH12]; [Daf70] and [KCG15, Section 1.3]. This Markovianisation theory can be formulated in the setting of near-equilibrium statistical mechanics, where one uses fluctuation-dissipation-like relations that link the decay properties of the memory kernel and the decorrelation rates of the

fluctuations. Here, we follow the approach in [Pav14] but without making any assumptions on the Hamiltonian behaviour of the $\mathbf{y}$-variables. We need, though, to impose conditions on the spectral properties of the generator of the $\mathbf{y}$-dynamics, as explained below.

We define the generator $\mathcal{L}^*_{0,\mathbf{y}}$ of the Koopman semigroup associated with the $\mathbf{y}$-dynamics by:

$$\mathcal{L}^*_{0,\mathbf{y}}\Psi(\mathbf{y}) = \mathbf{G}(\mathbf{y}) \cdot \nabla\Psi(\mathbf{y}), \tag{4.28}$$

for every real-valued observable Ψ in $D\left(\mathcal{L}^*_{0,\mathbf{y}}\right) \subset C_b(\mathcal{Y})$ and we denote the associated Koopman operator at time t by $e^{t\mathcal{L}^*_{0,\mathbf{y}}}$; the subscript "$\mathbf{y}$" has been dropped from the ∇ operator herewith, for notational clarity. Recall that the spectrum of such operators provides useful insights into the statistical properties of the system; see Section 2.2.1.

It suffices to show below that the spectrum of $e^{t\mathcal{L}^*_{0,\mathbf{y}}}$ allows one to characterise the constitutive ingredients of the WL parametrisation (4.22) and (4.27), subject to natural assumptions. Even though we have clarified in Eq. (4.9) how the Koopman operator acts on vector-valued observables of any dimension, we restrict now its action for simplicity to scalar real-valued observables, as in Eq. (4.28). In this case, along the lines of Section 2.2.1, we can (formally) decompose the operator as:

$$e^{t\mathcal{L}^*_{0,\mathbf{y}}} = \sum_{j=1}^{M} e^{\lambda_j t}\Pi^*_j + \mathcal{R}^*(t), \tag{4.29}$$

where $\{\lambda_j\}_{j=1}^M$ are the simple eigenvalues that form the point spectrum of $\mathcal{L}^*_{0,\mathbf{y}}$ and Π^*_j is the spectral projector onto the eigenspace spanned by the eigenfunction ψ^*_j. Recall that dual eigenfunctions are here denoted with the superscript $(\cdot)^*$, in line with Remark (2.2.4) and Eq. (2.44). Here, $\mathcal{R}^*(t)$ is the residual operator associated with the essential spectrum of $\mathcal{L}^*_{0,\mathbf{y}}$ and its norm is controlled by a decaying exponential; see Eq. (2.42). We assume furthermore that the spectrum of $\mathcal{L}^*_{0,\mathbf{y}}$ lies in the complex left half-plane, and that in particular $\mathfrak{Re}\lambda_j \leq 0$ for any j. Recall that such a spectral decomposition and its properties can be rigorously justified for a broad class of differential equations perturbed by small noise disturbances; see [Tan+20, Theorem 1 and Appendix A.5]. Based on these rigorous results, we assume, roughly speaking, that these properties survive in a certain small-noise limit, and concentrate here on the vector field given by $\mathbf{G}$ in (4.28) for which a decomposition such as (4.29) holds and a spectral gap does exist in the appropriate functional space; we refer the reader to the paragraphs immediately below Eq. (2.42).

In the following lines, we heuristically examine the expression of the memory kernel $\mathcal{K}$ appearing in Eq. (4.27) using the eigendecomposition proposed in Eq. (4.29). In particular,

we study such an integral kernel $\mathcal{K}$ component-wise:

$$[\mathcal{K}(t,s,\mathbf{x})]_i = \mathbf{C}^{\mathbf{y}}(\mathbf{x}(s)) \cdot \left\langle \nabla \sum_{j=1}^{M} e^{\lambda_j (t-s)} \alpha_{i,j} \psi_j^*(\mathbf{y}) \right\rangle + \mathcal{R}^*(t-s)[\mathbf{C}^{\mathbf{x}}]_i \tag{4.30a}$$

$$\approx \mathbf{C}^{\mathbf{y}}(\mathbf{x}(s)) \cdot \left\langle \nabla \sum_{j=1}^{M} e^{\lambda_j (t-s)} \alpha_{i,j} \psi_j^*(\mathbf{y}) \right\rangle \tag{4.30b}$$

$$= \mathbf{C}^{\mathbf{y}}(\mathbf{x}(s)) \cdot \sum_{j=1}^{M} e^{\lambda_j (t-s)} \alpha_{i,j} \left\langle \nabla \psi_j^*(\mathbf{y}) \right\rangle, \tag{4.30c}$$

for $i = 1, \ldots, d_1$, where

$$\alpha_{i,j} = \langle \psi_j, [\mathbf{C}^{\mathbf{x}}]_i \rangle = \int \nu(\mathrm{d}\mathbf{y}) \overline{\psi_j(\mathbf{y})} [\mathbf{C}^{\mathbf{x}}(\mathbf{y})]_i, \tag{4.31}$$

and we have neglected the contribution coming from the essential spectrum. The coefficients $\alpha_{i,j}$ are the analogues to those defined earlier in Section 2.2.1 immediately below Eq. 2.45.

This decomposition highlights the fact that the leading eigenvalues of the operator governing the evolution of observables in the uncoupled $\mathbf{y}$-dynamics set the timescale for the memory kernel. Furthermore, this spectral approximation implies that the correlation functions of the noise have the same decay properties, as will become apparent later in the proof of Theorem 4.2.1. It follows that the correspondence between the noise and integral timescales allows us to recast the SIDE in the WL equation (4.27) into a fully Markovian version with linearly driven hidden variables that are forced by the observed variables, through a functional dependence that can be non-linear. More exactly, we have the following theorem.

Theorem 4.2.1. *Consider the system* (4.24) *where Eq.* (4.24a) *is, instead, a scalar equation for a real-valued variable $x(t)$. Let ν be the physical invariant measure associated with the equation*

$$\dot{\mathbf{y}} = \mathbf{G}(\mathbf{y}), \tag{4.32}$$

i.e., with the flow determined by the vector field $\mathbf{G}$ in system (4.24)*, for $\varepsilon = 0$. Moreover, let $\mathcal{L}^*_{0,\mathbf{y}}$ be the (uncoupled) Koopman operator associated with* (4.32) *as defined in Eq.* (4.28)*.*

The point spectrum of $\mathcal{L}^*_{0,\mathbf{y}}$ is assumed to be constituted of M simple eigenvalues, whose corresponding eigenpairs $\left\{(\lambda_j, \psi_j^*),\ j = 1, \ldots, M\right\}$ are ordered as follows: $0 \geq \mathfrak{Re}\lambda_j \geq \mathfrak{Re}\lambda_{j+1}$ and $\lambda_j = \overline{\lambda_{j+1}}$ when $\mathfrak{Im}\lambda_j > 0$, for j in $\{1, \ldots, M\}$.

We assume that $\mathbf{C}^{\mathbf{x}}$ in (4.24) *lies in the $\mathrm{span}\left\{\psi_j^*,\ j = 1, \ldots, M\right\}$ and has ν-mean zero.*

Then, the WL equation (4.27) *associated with system* (4.24) *admits a Markovianisation*

of the form:

$$\dot{x}(t) = \mathbf{F}(x(t)) + \varepsilon \Lambda \cdot \mathbf{Z}(t), \tag{4.33a}$$

$$\mathrm{d}\mathbf{Z}(t) = [\varepsilon \mathbf{R}(x(t)) + \mathrm{D}\mathbf{Z}(t)]\,\mathrm{d}t + \Sigma \mathrm{d}W_t, \tag{4.33b}$$

where Λ *and* $\mathbf{Z}(t)$ *lie in* $\mathbb{C}^M$ *for every* t*, while the inner product* $\Lambda \cdot \mathbf{Z}(t)$ *is real. Here,* $\mathbf{R}$ *is mapping* $\mathbb{R}$ *into* $\mathbb{C}^M$*,* W_t *is a (real-valued)* M*-dimensional Wiener process with covariance matrix* Σ*, and* D *is an* $M \times M$ *matrix with complex entries, as specified below.*

More precisely,

$$\Lambda = \left[\alpha_1^{1/2}\beta_1^{1/2}, \ldots, \alpha_M^{1/2}\beta_M^{1/2}\right]^\top, \tag{4.34}$$

where

$$\alpha_j = \langle \psi_j, \mathbf{C}^{\mathbf{x}} \rangle = \int \nu(\mathrm{d}\mathbf{y}) \overline{\psi_j(\mathbf{y})} \mathbf{C}^{\mathbf{x}}(\mathbf{y}), \tag{4.35a}$$

$$\beta_j = \left\langle \mathbf{C}^{\mathbf{x}}, \psi_j^* \right\rangle = \int \nu(\mathrm{d}\mathbf{y}) \mathbf{C}^{\mathbf{x}}(\mathbf{y}) \psi_j^*(\mathbf{y}). \tag{4.35b}$$

The $\mathbb{C}^M$*-valued mapping* $\mathbf{R}$ *is defined as*

$$\mathbf{R}\,(x) = \left(\mathbf{C}^{\mathbf{y}}(x) \cdot \frac{\alpha_1^{1/2}}{\beta_1^{1/2}} \left\langle \nabla \psi_1^*(\mathbf{y}) \right\rangle, \ldots, \mathbf{C}^{\mathbf{y}}(x) \cdot \frac{\alpha_M^{1/2}}{\beta_M^{1/2}} \left\langle \nabla \psi_M^*(\mathbf{y}) \right\rangle \right), \tag{4.36}$$

where $\mathbf{C}^{\mathbf{y}}$ *is defined in* (4.24b)*, and* $\langle \cdot \rangle$ *denotes averaging with respect to the invariant measure* ν*.*

Finally, $\mathrm{D} = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_M)$ *and the covariance matrix* Σ *is given by*

$$\Sigma = (-(\mathrm{D} + \mathrm{D}^*))^{1/2}\,\mathrm{H}, \tag{4.37}$$

where H *is an* $M \times M$ *matrix whose entries are defined as follows: If* λ_j *is real, then* $\mathrm{H}_{j,j} = 1$*, and if* $\lambda_j = \overline{\lambda_{j+1}}$*, then*

$$\begin{aligned} \mathrm{H}_{j,j} &= 1, \\ \mathrm{H}_{j+1,j+1} &= 0, \\ \mathrm{H}_{j+1,j} &= 1, \end{aligned} \tag{4.38}$$

while all other entries are zero.

Proof. The aim is to show that under the assumptions of this Theorem — which require that the coupling function $\mathbf{C}^{\mathbf{x}} : \mathbb{R} \longrightarrow \mathbb{R}$ projects entirely onto $\mathrm{span}\left\{\psi_j^*,\ j = 1, \ldots, M\right\}$ — the memory and noise terms of the WL equation (4.27) are obtained from the term $\varepsilon\Lambda \cdot \mathbf{Z}(t)$ in Eq. (4.33a), after integration of Eq. (4.33b).

Step 1. In this step, we expand the memory term and the lag-correlations of the noise in the WL equation (4.27) in terms of the leading eigenelements of the uncoupled Koopman

operator $\mathcal{L}^*_{0,\mathbf{y}}$. These expansions will serve us in Step 2 below to compare the noise and memory terms of the WL-equation with those produced after integration of Eq. (4.33b).

Let the λ_j be sorted as in the hypotheses of the theorem. Hence, to distinguish real and complex conjugate eigenvalues and for notational convenience, we introduce the set of indices I_r and I_+ defined as:

$$I_r = \{j \in \{1, \dots, M\} : \lambda_j \text{ is real}\}, \tag{4.39a}$$

$$I_+ = \{j \in \{1, \dots, M\} : \mathfrak{Im}\lambda_j > 0\}. \tag{4.39b}$$

An immediate consequence that is used several times throughout the proof is that the sum of the eigenvalues is real, and may be split as follows:

$$\sum_{j=1}^{M} \lambda_j = \sum_{j \in I_r} \lambda_j + \sum_{j \in I_+} \lambda_j + \sum_{j \in I_+} \overline{\lambda_j} = \sum_{j=1}^{M} \mathfrak{Re}\lambda_j \in \mathbb{R}. \tag{4.40}$$

As previously stated, we expand the mean and correlation functions of the scalar noise term η in the WL equation (4.27) in terms of the eigenpairs. The mean is zero by assumption, but the autocorrelation function can be expanded as follows, based on the spectral decomposition of correlation functions [Tan+20] later reviewed in Section 4.3:

$$\left\langle \eta(t)\eta(0) \right\rangle = C_{\mathbf{C}^{\mathbf{x}},\mathbf{C}^{\mathbf{x}}}(t) = \sum_{j=1}^{M} e^{\lambda_j t} \alpha_j \beta_j; \tag{4.41}$$

herein, α_j and β_j are as defined in (4.35a) and (4.35b), respectively. The expansion of the correlation function in Eq. (4.41) is a finite sum by virtue of the assumption that $\mathbf{C}^{\mathbf{x}}$ lies in $\operatorname{span}\left\{\psi_j^*,\ j = 1, \dots, M\right\}$ and therefore there is no contribution from the essential spectrum.

Regarding the complex scalars α_j and β_j defined (4.35a) and (4.35b), it follows that for each j such that $\lambda_j = \overline{\lambda_{j+1}}$, we get $\alpha_j = \overline{\alpha_{j+1}}$ and $\beta_j = \overline{\beta_{j+1}}$. Indeed,

$$\alpha_j = \int \nu(\mathrm{d}\mathbf{y}) \overline{\psi_j(\mathbf{y})} \mathbf{C}^{\mathbf{x}}(\mathbf{y}) = \overline{\int \nu(\mathrm{d}\mathbf{y}) \psi_j(\mathbf{y}) \mathbf{C}^{\mathbf{x}}(\mathbf{y})} \tag{4.42a}$$

$$= \overline{\int \nu(\mathrm{d}\mathbf{y}) \overline{\psi_{j+1}(\mathbf{y})} \mathbf{C}^{\mathbf{x}}(\mathbf{y})} = \overline{\alpha_{j+1}}, \tag{4.42b}$$

in which we have exploited the fact that $\psi_j = \overline{\psi_{j+1}}$ when λ_j is complex and j in $\{1, \dots, M\}$.

The same proof can be repeated for β_j. Such a conjugacy relation also holds for the gradients of the eigenfunctions $\nabla\psi_j$, for those j in $\{1 \dots, M\}$ such that $\lambda_j = \overline{\lambda_{j+1}}$. This is observed by the following equality:

$$\nabla\psi_j(\mathbf{y}) = \nabla\overline{\psi_{j+1}}(\mathbf{y}) = \overline{\nabla\psi_{j+1}(\mathbf{y})}, \tag{4.43}$$

since ∇ is a differential operator that only involves here differentiation with respect to a real variable.

Exploiting (4.43), the memory kernel $\mathcal{K}$ then expands as (recalling that $\mathcal{R}^*(t)\mathbf{C}^{\mathbf{x}} \equiv 0$):

$$\begin{aligned}\mathcal{K}(t,s,x) =& \mathbf{C}^{\mathbf{y}}(x(s)) \cdot \left\langle \nabla \sum_{j=1}^{M} e^{\lambda_j (t-s)} \alpha_j \psi_j^*(\mathbf{y}) \right\rangle \\ =& \mathbf{C}^{\mathbf{y}}(x(s)) \cdot \sum_{j \in I_r} e^{\lambda_j (t-s)} \alpha_j \left\langle \nabla \psi_j^*(\mathbf{y}) \right\rangle \\ &+ \mathbf{C}^{\mathbf{y}}(x(s)) \cdot \sum_{j \in I_+} e^{\lambda_j (t-s)} \alpha_j \left\langle \nabla \psi_j^*(\mathbf{y}) \right\rangle \\ &+ \mathbf{C}^{\mathbf{y}}(x(s)) \cdot \sum_{j \in I_+} \overline{e^{\lambda_j (t-s)} \alpha_j \left\langle \nabla \psi_j^*(\mathbf{y}) \right\rangle}\end{aligned} \tag{4.44}$$

which leads to

$$\mathcal{K}(t,s,x) = \mathbf{C}^{\mathbf{y}}(x(s)) \cdot \sum_{j=1}^{M} \mathfrak{Re}\left(e^{\lambda_j (t-s)} \alpha_j \left\langle \nabla \psi_j^*(\mathbf{y}) \right\rangle \right). \tag{4.45}$$

Note that to go from (4.44) to (4.45), we have made use of the aforementioned conjugacy relations, that led to a real-valued memory kernel at the end. With (4.41) and (4.45) at hand, the noise and memory terms in the WL equation (4.27) are thus characterised in terms of the leading eigenelements of the uncoupled Koopman operator $\mathcal{L}_{0,\mathbf{y}}^*$.

Step2. The second step consists of analysing the noise and memory terms produced by integration of Eq. (4.33b) and to compare these terms with those of the WL equation.

Performing an Itô integration of Eq. (4.33b)— cf. Appendix C— leads to:

$$\mathbf{Z}(t) = e^{t\mathrm{D}}\mathbf{Z}(0) + \underbrace{\int_0^t e^{(t-s)\mathrm{D}} \Sigma \mathrm{d}W_s}_{\text{noise term}} + \varepsilon \underbrace{\int_0^t e^{(t-s)\mathrm{D}} \mathbf{R}(x(s)) \mathrm{d}s}_{\text{memory term}}, \tag{4.46}$$

where (for simplicity) we have assumed that the initial condition distributes normally with mean zero and variance equal to the identity matrix, and the function $\mathbf{R} : \mathbb{R} \longrightarrow \mathbb{C}^M$ is as defined in (4.36). The noise and memory contributions of $\mathbf{Z}(t)$ are as indicated by the brackets in (4.46).

Let us denote the noisy component of Eq. (4.46) as $\mathbf{q}(t)$ in $\mathbb{R}^M$. Then, it is clear that $\mathbf{q}$ has zero mean and the lag cross-correlations read as:

$$\mathbb{E}\left[\mathbf{q}(t)\mathbf{q}^\top(0)\right] = e^{t\mathrm{D}} = \begin{bmatrix} e^{\lambda_1 t} & & \\ & \ddots & \\ & & e^{\lambda_M t} \end{bmatrix}. \tag{4.47}$$

Let Λ be defined as in (4.34) and let us calculate the mean and lag-correlations of the one-dimensional stochastic process $\Lambda \cdot \mathbf{q}$, aimed at approximating η in the WL equation. First, note that $\Lambda \cdot \mathbf{q}$ is a zero-mean Gaussian process, with lag correlations given by:

$$\mathbb{E}\left[(\Lambda \cdot \mathbf{q}(t))\,(\Lambda \cdot \mathbf{q}(0))\right] = \left(e^{t\mathrm{D}}\Lambda\right) \cdot \Lambda. \tag{4.48}$$

Now, expanding $\left(e^{t\mathrm{D}}\Lambda\right) \cdot \Lambda$ in (4.48) shows that we recover the RHS of (4.41).

However, the noise term η in the WL equation is a real-valued stochastic process, and we are dealing with complex scalars, so therefore we still have to show that $\Lambda \cdot \mathbf{q}(t)$ is real for every t in $\mathbb{R}$. To do so, let us denote by $\mathbf{w}^\top = [w_1, \ldots, w_M]$ any arbitrary row vector in $\mathbb{R}^M$. Consider the following inner product:

$$\Lambda \cdot e^{t\mathrm{D}}\Sigma\mathbf{w} = \Lambda \cdot e^{t\mathrm{D}} \begin{bmatrix} \sqrt{-2\mathfrak{Re}\lambda_1} & & \\ & \ddots & \\ & & \sqrt{-\mathfrak{Re}\lambda_M} \end{bmatrix} \mathrm{H}\mathbf{w} \tag{4.49a}$$

$$= \Lambda \cdot \begin{bmatrix} e^{\lambda_1 t}\sqrt{-2\mathfrak{Re}\lambda_1} & & \\ & \ddots & \\ & & e^{\lambda_M t}\sqrt{-2\mathfrak{Re}\lambda_M} \end{bmatrix} \mathrm{H}\mathbf{w}. \tag{4.49b}$$

By construction of the matrix H in Eqs. (4.37)-(4.38), the product $\mathrm{H}\mathbf{w}$ is given componentwise, for $j = 2, \ldots, M$, as:

$$[\mathrm{H}\mathbf{w}]_j = \begin{cases} w_j, \text{ if } j \in I_r \text{ or, } j \in I_+, \\ w_{j-1}, \text{ if } \lambda_j = \overline{\lambda_{j-1}}, \end{cases} \tag{4.50}$$

while $[\mathrm{H}\mathbf{w}]_1 = w_1$. This implies in particular that $[\mathrm{H}\mathbf{w}]_j = [\mathrm{H}\mathbf{w}]_{j+1}$ whenever $\lambda_j = \overline{\lambda_{j+1}}$. As a consequence, we get

$$\begin{aligned} \Lambda \cdot e^{t\mathrm{D}}\Sigma\mathbf{w} = &\sum_{j\in I_r} \alpha_j^{1/2}\beta_j^{1/2}e^{\lambda_j t}\sqrt{-2\lambda_j}w_j + \sum_{j\in I_+} \alpha_j^{1/2}\beta_j^{1/2}e^{\lambda_j t}\sqrt{-2\mathfrak{Re}\lambda_j}w_j \\ &+ \sum_{j\in I_+} \overline{\alpha_j^{1/2}\beta_j^{1/2}e^{\lambda_j t}}\sqrt{-2\mathfrak{Re}\lambda_j}w_j, \end{aligned} \tag{4.51}$$

which shows that $\Lambda \cdot e^{t\mathrm{D}}\Sigma$ is a real-valued quantity, and thus for any realisation of the M-dimensional Wiener process W_t, the product $\Lambda \cdot e^{(t-s)\mathrm{D}}\Sigma \mathrm{d}W_s$ is real and hence, $\Lambda \cdot \mathbf{q}(t)$ is also real for every t.

Finally, we are left with showing that the memory kernel $\mathcal{K}$ of the WL equation coincides with that of the memory term in Eq. (4.46), when multiplied by the vector Λ. To do so, we exploit the expansion (4.45) of $\mathcal{K}$ for this comparison, namely using the

expression of $\mathbf{R}$ in (4.36), we observe that

$$\begin{aligned}\Lambda \cdot \int_0^t e^{(t-s)\mathrm{D}}\mathbf{R}(x(s))\mathrm{d}s =&\mathbf{C}^{\mathbf{y}}(x(s)) \cdot \sum_{j\in I_r} e^{\lambda_j(t-s)}\alpha_j\left\langle \nabla\psi_j^*(\mathbf{y})\right\rangle \\ &+ \mathbf{C}^{\mathbf{y}}(x(s)) \cdot \sum_{j\in I_+} e^{\lambda_j(t-s)}\alpha_j\left\langle \nabla\psi_j^*(\mathbf{y})\right\rangle \\ &+ \mathbf{C}^{\mathbf{y}}(x(s)) \cdot \sum_{j\in I_+} e^{\lambda_j(t-s)}\alpha_j\left\langle \nabla\psi_j^*(\mathbf{y})\right\rangle \\ =&\mathbf{C}^{\mathbf{y}}(x(s)) \cdot \sum_{j=1}^{M} \mathfrak{Re}\left(e^{\lambda_j(t-s)}\alpha_j\left\langle \nabla\psi_j^*(\mathbf{y})\right\rangle\right),\end{aligned} \tag{4.52}$$

which indeed coincides with the expression of $\mathcal{K}$ in Eq. (4.45), as desired. The proof is complete. □

Remark 4.2.1. *Note that the Koopman operator of interest here is the one associated with the* $\mathbf{y}$*-subspace* $\mathcal{Y} \subseteq \mathbb{R}^{d_2}$ *and not with the entire* $(\mathbf{x},\mathbf{y})$*-space* $\mathcal{X}\times\mathcal{Y} \subseteq \mathbb{R}^{d_1+d_2}$*. Other techniques, like the DMD mentioned in Section 4.1, aim at extracting the modes of variability of the full system by means of studying the Koopman operator in the entire phase space through suitably defined observables. To this end, the latter methods employ projections of observables onto the eigenfunctions of the Koopman operator to obtain the so-called Koopman modes, which are susceptible of capturing the underlying dynamics. Notice that in Theorem 4.2.1, instead, we are using the Koopman eigenfunctions to identify the closure model, while projections only come into play in the definition of the coefficients* α_j *and* β_j*; see Eqs.* (4.35a) *and* (4.35b)*, respectively.*

Remark 4.2.2.

(i) Assumptions on $\mathbf{F}$*,* $\mathbf{R}$ *and* Λ *that ensure that* (4.33) *possesses a global random attractor— and thus a stable asymptotic behaviour in the pullback sense— appear in [KCG15, Theorem 3.1 and Corollary 3.2].*

(ii) Note that Theorem 4.2.1 can be viewed as a generalisation of other Markovianisation results for GLEs that appeared in the literature; see [Pav14]. For instance, the scalar GLE in $\mathbb{R}$ *reduces to*

$$\dot{x} = F(x(t)) - \int_0^t K(t-s)x(s)\mathrm{d}s + \eta(t), \tag{4.53}$$

where λ *is in* $\mathbb{R}^n$*,* M *is a positive definite* $n\times n$ *matrix and* $K(t-s) = \left(e^{(t-s)M}\lambda\right)\cdot\lambda$ *determines the autocorrelation of the process* $\eta(t)$*. In this setting, Eq.* (4.53) *is*

equivalent to the following SDE:

$$\begin{aligned} \dot{x} &= F(x(t)) + \lambda \cdot z \\ \mathrm{d}z &= [x\lambda - Mz]\,\mathrm{d}t + \Sigma \mathrm{d}W_t, \end{aligned} \tag{4.54}$$

with $\Sigma\Sigma^* = M + M^*$. Theorem 4.2.1 allows for non-linear dependence on x in the z-equation, and thus for memory kernels that are more complicated than in (4.53)*. Such a generalisation is of practical importance since the process z can then have a more complex correlation dependence on the observed variable x than the one afforded by linear memory terms.*

Remark 4.2.3. *When $\mathcal{L}^*_{0,\mathbf{y}}$ is self-adjoint — in a suitable Hilbert space, as outlined in Section 4.3 but also in Chapter 2—, i.e., when $\mathcal{L}^*_{0,\mathbf{y}} = \mathcal{L}_{0,\mathbf{y}}$ the eigenvalues are real and the eigenvectors are mutually orthogonal. Self-adjointness thus implies that there are no oscillations in the correlation functions of the noise or, equivalently, peaks in their power spectrum. With respect to Theorem 4.2.1, the matrix* H *in this case would be the identity, since the eigenvalues and eigenfunctions are real and the Itô solutions of Eq.* (4.33b) *are, hence, real as well.*

Remark 4.2.4. *The resulting system given by Eq.* (4.33) *is now fully Markovian and the only sources of error with respect to the original SIDE* (4.22) *lie (i), in the effects of the essential spectrum, which are neglected herein and (ii), the assumptions about the coupling terms. Neglecting the essential spectrum is only valid for Koopman operators with a point spectrum capable of capturing the correlations in the decoupled $\mathbf{y}$-system; the latter might only hold in the case of Markovian diffusion processes and not for deterministic ones. Also, the assumption that the coupling functions project solely on the point spectrum might not hold in general.*

From a practical perspective, though, a suitable choice of dominant eigendirections can reduce the number of extra dimensions needed to integrate the system. Such a suitable choice boils down to neglecting particular eigendirections and this can be done according to two handy criteria:

(i) The weight determined by the α_j and β_j coefficients defined in Eqs. (4.35a) *and* (4.35b) *is small; and*

(ii) The eigenvalues λ_j of $\mathcal{L}^*_{0,\mathbf{y}}$ satisfy $\mathfrak{Re}\lambda_{j_0} \ll \lambda^\dagger$, for some j_0 in $\{1,\ldots,M\}$, in which case $e^{\lambda_{j_0} t}$ decays rapidly as t grows; here $\lambda^\dagger < 0$ and $|\lambda^\dagger|$ is some characteristic inverse time for the deterministic system $\mathbf{F}$. In addition, if $\alpha_{j_0} > \alpha_j$ for every j in $\{1,\ldots,M\}$ and $j \neq j_0$, both the memory and the noise correlations die out fast. Hence, one can neglect the integral terms and perform a fully Markovian parametrisation, which is possible in the presence of white noise.

Remark 4.2.5. *Theorem 4.2.1 is stated for x being a scalar for the sake of simplicity, but this result extends to the d_1-dimensional Eq.* (4.24) *for the (observed) variables* $\mathbf{x}$*. In this remark, we sketch the main elements that permit such a generalisation.*

Aside from the obvious generalisation of the assumptions in Theorem 4.2.1 to a multidimensional setting, the main hypothesis consists of assuming that the now vector-valued coupling function $\mathbf{C}^{\mathbf{x}}$ in Eq. (4.24) *has components $\{[\mathbf{C}^{\mathbf{x}}]_i : i = 1, \dots, d_1\}$ that project onto the M simple eigenspaces of the decoupled Koopman operator introduced in Eq.* (4.28)*. In this case, the construction of a multilevel Markovianisation like Eq.* (4.33) *can be done in the following fashion:*

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)) + \varepsilon \Lambda \mathbf{Z}(t), \tag{4.55a}$$

$$\mathrm{d}\mathbf{Z}(t) = [\varepsilon \mathbf{R}(\mathbf{x}(t)) + \mathcal{D}\mathbf{Z}(t)]\,\mathrm{d}t + \mathcal{S}\mathrm{d}W_t. \tag{4.55b}$$

Here W_t is a $d_1 M$-dimensional Wiener process, $\mathbf{Z}(t)$ is a $d_1 M$-dimensional vector, Λ is a matrix of size $d_1 \times d_1 M$, $\mathbf{R} : \mathbb{R}^{d_1} \longrightarrow \mathbb{C}^{d_1 M}$ and $\mathcal{D}$ and $\mathcal{S}$ are $d_1 M \times d_1 M$ block-diagonal matrices given by:

$$\mathcal{D} = \begin{bmatrix} \mathrm{D}_1 & & \\ & \ddots & \\ & & \mathrm{D}_M \end{bmatrix} \text{ and } \mathcal{S} = \begin{bmatrix} \Sigma_1 & & \\ & \ddots & \\ & & \Sigma_M \end{bmatrix}, \tag{4.56}$$

where D_j and Σ_j are $d_1 \times d_1$ diagonal matrices with every (non-zero) element being equal to λ_j or $\sqrt{-2\mathfrak{Re}\lambda_j}$, respectively. More importantly, the vectors $\mathbf{Z}(t)$ and $\mathbf{R}(\mathbf{x})$ are split into M column vectors $\mathbf{z}_j(t)$ and $\mathbf{r}_j(\mathbf{x})$ of length d_1 with j in $\{1, \dots, M\}$. This way, $\mathbf{Z}(t) = \left[\mathbf{z}_1^\top(t), \dots, \mathbf{z}_M^\top(t)\right]^\top$ and $\mathbf{R}(\mathbf{x}) = \left[\mathbf{r}_1^\top(t), \dots, \mathbf{r}_M^\top(t)\right]^\top$.

Therefore Eq. (4.55) *can be written as:*

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)) + \varepsilon \Lambda \mathbf{Z}(t), \tag{4.57a}$$

$$\mathrm{d}\mathbf{z}_1(t) = (\varepsilon \mathbf{r}_1(\mathbf{x}(t)) + \mathrm{D}_1 \mathbf{z}_1(t))\,\mathrm{d}t + \Sigma_1 \mathrm{d}W_t^{(1)}, \tag{4.57b}$$

$$\vdots$$

$$\mathrm{d}\mathbf{z}_M(t) = (\varepsilon \mathbf{r}_M(\mathbf{x}(t)) + \mathrm{D}_M \mathbf{z}_M(t))\,\mathrm{d}t + \Sigma_M \mathrm{d}W_t^{(M)}, \tag{4.57c}$$

where $W_t^{(j)}$ is a d_1-dimensional Wiener process. The vectors $\mathbf{r}_j(\mathbf{x})$ are given by:

$$\mathbf{r}_j(\mathbf{x}) = \left[\mathbf{C}^{\mathbf{y}}(\mathbf{x}) \cdot \gamma_{1,j} \left\langle \nabla \psi_j^*(\mathbf{y}) \right\rangle, \dots, \mathbf{C}^{\mathbf{y}}(\mathbf{x}) \cdot \gamma_{d_1,j} \left\langle \nabla \psi_j^*(\mathbf{y}) \right\rangle\right]^\top, \tag{4.58}$$

where $\gamma_{i,j}$ are defined in terms of the parameters (α_j, β_j) introduced in Eqs. (4.35a) *and* (4.35b)*, respectively. Here we don't give the explicit expression of Λ, but its role is to*

provide suitable weights, in the spirit of Eq. (4.34)*, to the levels in Eq.* (4.57) *so that (a), the correlation functions match those of the coupling function* $\mathbf{C}^{\mathbf{x}}$ *in the uncoupled regime; and (b), the resulting term* $\Lambda\mathbf{Z}(t)$ *is real.*

The system Eq. (4.57) *has the general structure one would obtain if the coupling function* $\mathbf{C}^{\mathbf{x}}$ *projected along all the eigendirections in the point spectrum. This might not be true in general, but the drift matrix* $\mathcal{D}$ *can be rearranged so that only the relevant modes of variability be modelled — following criteria (i) and (ii), as formulated in Remark 4.2.4 — and still afford a reduction of the number of levels* M*. Notice that the* M*th level variables* $\mathbf{z}_M$ *described by Eq.* (4.57c) *decorrelate the fastest compared to the rest, since their exponential decorrelation rate is given by* $|\mathfrak{Re}\lambda_M| \geq |\mathfrak{Re}\lambda_j|$*, for all* $j = 1, \ldots, M-1$*.*

The advantages of the Markovian system (4.33) and (4.57) over the original WL equation (4.27) are twofold. First, we identify situations in which the WL equation can be Markovianised by introducing extended, hidden variables. This idea was already introduced in a preliminary application of the WL parametrisation [Wou+16], in which the authors resorted to a Markovian system to perform their simulations. In fact, one of their examples is studied in the present framework; see Section 4.2.3.

Second, memory equations contain non-local terms that are cumbersome and computationally expensive to integrate, as well as requiring much larger storage for the full history of the system's variables. The efficient Markovianisation of evolution equations with memory terms is an active field of research in diverse areas of mathematics and the applied sciences; these areas include the study of bifurcations of delay differential equations [Che+16]; [CKL20], the reduction of stochastic partial differential equations to stochastic invariant manifolds [CLW15a]; [CLW15b], and material sciences [Daf70], among many others.

4.2.3 Preliminary Example

As seen earlier in Theorem 4.2.1, if the coupling function is resonant with the Koopman operator associated with the $\mathbf{y}$-dynamics, one can identify the dominant exponential rates of decay of the memory term and the characteristic decorrelation time of the noise. As a consequence, one can Markovianise the parametrisation and greatly facilitate the numerical integrations involved.

To illustrate the above statement, we revisit here the preliminary application of the WL parametrisation in the context of multiscale triads [Wou+16]. In that work, the authors implemented the parametrisation for a collection of three-dimensional models that do exhibit timescale separation and compare the corresponding outputs to those obtained via homogenisation. The results are encouraging, since the parametrisations in [Wou+16] were obtained only from the decoupled hidden dynamics, in the lines of the present chapter as well; see derivation of Eq. (4.22).

One of the first multiscale triads studied in [Wou+16] is the following:

$$\dot{x}(t) = \varepsilon B^{(0)} y_1 y_2, \tag{4.59a}$$
$$\dot{y_1}(t) = \varepsilon B^{(1)} x y_2 - \gamma_1 y_1 + \sigma_1 \mathrm{d}W_t^{(1)}, \tag{4.59b}$$
$$\dot{y_2}(t) = \varepsilon B^{(2)} x y_1 - \gamma_2 y_2 + \sigma_2 \mathrm{d}W_t^{(2)}. \tag{4.59c}$$

Here we require that $\sum_j B^{(j)} = 0$, $\mathrm{d}W_t^{(1)}$ and $\mathrm{d}W_t^{(2)}$ are scalar Wiener increments and the parameter ε indicates both the timescale separation and the coupling strength. Notice that when the system is decoupled, i.e. when $\varepsilon = 0$, the fast dynamics evolve according to an O-U process whose steady state statistics are governed by Gaussian distributions with explicit mean and variance; see, e.g. [Pav14]. Hence, by virtue of the previous formulas or by following [Wou+16], the WL parametrisation yields the following scalar SIDE:

$$\dot{x}(t) = \varepsilon \eta(t) + \varepsilon^2 \int_0^t \mathcal{K}(s, x(t-s)) \mathrm{d}s; \tag{4.60}$$

here,

$$\langle \eta(t) \rangle = 0, \tag{4.61a}$$
$$\langle \eta(t+s)\eta(s) \rangle = \left(B^{(0)}\right)^2 e^{-(\gamma_1+\gamma_2)t} \frac{\sigma_1^2}{2\gamma_1} \frac{\sigma_2^2}{2\gamma_2}, \tag{4.61b}$$
$$\mathcal{K}(s,x) = \begin{bmatrix} x \\ x \end{bmatrix} \cdot \left\langle \begin{bmatrix} B^{(1)} y_2 \\ B^{(2)} y_1 \end{bmatrix} : \nabla_{\mathbf{y}} y_1(s) y_2(s) \right\rangle, \tag{4.61c}$$

where the angular brackets refer to the ensemble averages according to the already mentioned Gaussian distributions arising from the decoupled model. Expanding these averages Eq. (4.61c) leads to:

$$\mathcal{K}(s,x) = x e^{-(\gamma_1+\gamma_2)} \left\langle B^{(1)} y_2^2 + B^{(2)} y_1^2 \right\rangle \tag{4.62a}$$
$$= x B^{(0)} e^{-(\gamma_1+\gamma_2)s} \left(B^{(1)} \frac{\sigma_2^2}{2\gamma_2} + B^{(2)} \frac{\sigma_1^2}{2\gamma_1} \right). \tag{4.62b}$$

The timescales are indicated by the exponents in the formulas above and they are the same for the noise and the memory kernel. This equality suggests the possibility of

Markovianising the memory equation into the following two-dimensional system:

$$\dot{z_1}(t) = \varepsilon B^{(0)} z_2, \tag{4.63a}$$

$$\dot{z_2}(t) = -(\gamma_1+\gamma_2)z_2 + \frac{\sigma_1\sigma_2}{2\gamma_1\gamma_2}\{2(\gamma_1+\gamma_2)\}^{1/2}\mathrm{d}W_t + \varepsilon\left(B^{(1)}\frac{\sigma_2^2}{2\gamma_2} + B^{(2)}\frac{\sigma_1^2}{2\gamma_1}\right)z_1. \tag{4.63b}$$

Clearly, performing a numerical integration of this system is easier than for a memory equation like Eq. (4.60).

The results of Section 4.2.2 allow us to carry out the dimension reduction of the multiscale triad by analysing the spectral properties of the Koopman operator associated with the decoupled $\mathbf{y}$-dynamics. Since the $\mathbf{y}$-variables evolve stochastically, the Koopman operator becomes the backward-Kolmogorov equation (2.13), which governs the evolution of the expectation values of the observables. Thus, for a generic observable Ψ in the $\mathbf{y}$-phase space, the evolution of its expectation value in the uncoupled regime is given by:

$$\begin{aligned}\partial_t\Psi(y_1,y_2) &= \mathcal{L}^*_{0,\mathbf{y}}\Psi(y_1,y_2)\\ &= \begin{bmatrix}-\gamma_1 y_1\\ -\gamma_2 y_2\end{bmatrix}\cdot\nabla\Psi(y_1,y_2) + \sigma_1^2\partial^2_{y_1}\Psi(y_1,y_2) + \sigma_2^2\partial^2_{y_2}\Psi(y_1,y_2).\end{aligned} \tag{4.64}$$

Now, letting $\Psi(y_1,y_2) = y_1y_2$ be the coupling function of the triad system (4.59), we find that:

$$\mathcal{L}^*_{0,\mathbf{y}}\Psi(y_1,y_2) = -(\gamma_1+\gamma_2)\Psi(y_1,y_2). \tag{4.65}$$

The above equation is an eigenvalue problem, showing that this particular observable Ψ is an eigenfunction of the generating operator $\mathcal{L}^*_{0,\mathbf{y}}$ operator associated with the eigenvalue $-(\gamma_1+\gamma_2)$. This is no surprise, since y_1 and y_2 are respectively the Hermite polynomial eigenfunctions of the backward-Kolmogorov equation of the scalar O-U process [UO30]; [Pav14, Section 4.4]. Hence, the product y_1y_2 is also an eigenfunction of the same equation for the joint process. Therefore, we can immediately re-Markovianise the parametrisation according to Eq. (4.33), where $\mathrm{D} = \gamma_1+\gamma_2$ and

$$\mathbf{R}(x(t)) = \left(B^{(1)}\frac{\sigma_2^2}{2\gamma_2} + B^{(2)}\frac{\sigma_1^2}{2\gamma_1}\right)x(t), \tag{4.66}$$

in agreement with [Wou+16].

4.3 Semigroup Property of the Projected Koopman Operator Family

Memory effects represented by integral terms seem unavoidable unless the memory kernels vanish quickly with respect to time. Infinite timescale separation between the two sets of variables, for instance, leads to the vanishing of the associated integral expressions yielding a Markovian process [PS08]. Here, we are not assuming no such property in the coupled dynamical system under study; see Eqs. (4.2) and (4.24). On the other hand, reduced phase spaces can help explain the statistics of the dynamical system without resorting to delayed effects that entail the integrals in Eqs. (4.22) and (4.27). Following [Tan+20], we briefly review here a criterion based on Koopman operators— and, more generally, Markov operators— that enables one to decide whether memory effects can help explain the dynamics and statistics in reduced phase spaces.

It was shown in [Che+14, Theorem A] that projection onto a reduced state space is closely related with a coarse graining of the (full) probability transitions on the original system's attractor, while [Tan+20, Theorem 2] dealt recently with the impact of such a projection in terms of reduction of the Koopman semigroup. In [Tan+20], the authors proposed a criterion based on the spectral theory of Markov semigroups to ascertain whether the reduced state space associated with a given projection can fully explain the statistics of the desired variables. In fact, it follows that the analysis of correlation functions is not only of physical interest but also of methodological utility in determining the need for modelling non-Markovian effects by inspecting the loss of the semigroup property, as explained below.

Let μ denote an ergodic invariant measure of the system and take two observables Ψ, Φ in the space L^2_μ of zero-mean functions that are square-integrable with respect to μ. Assume furthermore that the spectrum of the operator $\mathcal{L}$ in Eq. (2.12) is densely defined in L^2_μ and it is a pure point spectrum, given by the eigenvalues $\{\lambda_j\}_{j=1}^\infty$ and their associated eigenfunctions $\{\psi_j\}_{j=1}^\infty$, where the eigenvalues are ordered by their decreasing real parts. Then, the correlation function $C_{\Psi,\Phi}(t)$ between the functions Ψ and Φ is given by:

$$C_{\Psi,\Phi}(t) = \int \Psi \cdot e^{t\mathcal{L}^*} \Phi \mathrm{d}\mu = \int e^{t\mathcal{L}} \Psi \cdot \Phi \mathrm{d}\mu \tag{4.67}$$

and it can be expanded, formally, as

$$C_{\Psi,\Phi}(t) = \sum_{j=1}^{\infty} e^{\lambda_j t} \left\langle \psi_j, \Psi \right\rangle_\mu \left\langle \Phi, \psi_j^* \right\rangle_\mu . \tag{4.68}$$

It is recalled that the dual operators in (4.67) and the adjoint eigenvectors in (4.68) are indicated by the superscript $(\cdot)^*$, while $\langle \cdot, \cdot \rangle_\mu$ denotes the inner product with respect to μ.

We refer to [Tan+20, Corollary 1] for a proof of (4.68) in the context of Markov semigroups. The proof actually applies to the case of the Koopman semigroups considered here as long as the Koopman semigroup $\mathcal{U}_t$ defined by (4.5) is a strongly continuous semigroup in L^2_μ with a simple spectrum. The RHS of Eq. (4.68) consists of a linear combination of exponential terms whose coefficients are calculated by projecting Ψ and Φ onto the corresponding eigenspaces. These coefficients weight each exponential function and they can become exceedingly large if the Koopman operator deviates very much from normality [TE05, Chapter X].

The interactions between the resolved and hidden variables that are modelled by the Dyson expansion of the Koopman operator in Section 4.1 may introduce memory effects into the closed, reduced model for the $\mathbf{x}$-variables, as given by Eqs. (4.22)-(4.26). In certain situations, such memory effects can be neglected, even in the absence of exact slaving relationships between the resolved and hidden variables [CLM19]. But the loss of slaving relationships requires, in general, an explicit representation of memory effects [CLM17] to achieve an efficient model reduction.

Furthermore it was shown in [Che+14]; [Tan+20] that the reduction of the Koopman semigroup to observables that act only on the reduced state space leads, in most circumstances, to a family of operators that, while Markovian, no longer satisfy the semigroup property. One might then ask to which extent this loss of the semigroup property arising from the reduction, and the related emergence of memory effects, are crucial for providing a faithful reduced model of the observed variables.

When considering reduced state spaces obtained by projection, along with observables Ψ and Φ defined on them, Theorem 2 in [Tan+20] shows the existence of a family of Markov operators $\{\mathcal{T}_t\}_{t\geq 0}$ that satisfies:

$$\int \Psi \cdot \mathcal{T}_t \Phi \mathrm{d}\mu_{\mathbf{x}} = \int [\Psi \circ \pi_{\mathbf{x}}] \cdot e^{t\mathcal{L}^*} [\Phi \circ \pi_{\mathbf{x}}] \mathrm{d}\mu = C_{\Psi \circ \pi_{\mathbf{x}},\, \Phi \circ \pi_{\mathbf{x}}}(t), \tag{4.69}$$

for every $t \geq 0$, where $\pi_{\mathbf{x}}$ is the canonical projection onto the reduced subspace and $\mu_{\mathbf{x}}$ is the *disintegrated* or *sample measure* associated with $\pi_{\mathbf{x}}$; see [Tan+20, Remark 3]. However, due to the projection, the semigroup property is lost, namely, $\mathcal{T}_s\mathcal{T}_t \neq \mathcal{T}_{t+s}$ for some t, s.

Following the reasoning above, one can establish a criterion for the need to model a memory contribution when performing the dimension projection. Formally, if there exist $\tau > 0$ and T in $\mathbb{N}$ such that, for every t in $\{k\tau \in \mathbb{R} : 0 \leq k \leq T\}$, we have

$$C_{\Psi \circ \pi_{\mathbf{x}},\, \Phi \circ \pi_{\mathbf{x}}}(t) = \int \Psi \cdot \mathcal{T}_t \Phi \mathrm{d}\mu_{\mathbf{x}} = \int \Psi \cdot (\mathcal{T}_\tau)^k \, \Phi \mathrm{d}\mu_{\mathbf{x}}, \tag{4.70}$$

for some natural k, one can, then, say that the semigroup is preserved, to some extent, depending on how large T and small τ can be taken in Eq. (4.70) above. Other such criteria

are available in the context of mutually dual Koopman and transfer operators. Indeed, [TBD15] had already considered empirical ways of quantifying the loss of the semigroup property in reduced dimensions.

The interpretation of τ in Eq. (4.70) can come from the practical implementation of the semigroups following Ulam's method described in Section 3.1. In particular, if we look at the construction of transition matrices in Eq. (3.48), τ would represent the time-lag between transitions. Indeed, going back to Fig. 3.4 in Section 3.3.2, Eq. (4.70) was implemented and shown that there existed a value for τ (namely, $\tau = 100 \cdot \mathrm{d}t$) for which correlations were better approximated. According to the present criterion, one is led to think that memory effects were not so relevant in that application, probably, because there is no dimension reduction; this was implicitly assumed in Eq. (3.5). On the other hand, if for all reasonable values of τ and T, Eq. (4.70) is not satisfied, memory effects can be assumed significant.

4.4 Multilevel Stochastic Models and Empirical Model Reduction

4.4.1 Multilevel Stochastic Models (MSMs)

MSMs are a general class of SDEs that were introduced in [KCG15] and are, by their layered structure, susceptible to provide a good approximation of the GLE (4.1b) formulated by Mori and Zwanzig when a high-dimensional system is partially observed; see [KCG15, Proposition 3.3 and Sec. 5]. The MSM framework allows one to provide such approximations that are accompanied by useful dynamical properties, such as the existence of random attractors [KCG15, Theorem 3.1]. As discussed in [KCG15], MSMs arise in a variety of data-driven protocols for model reduction that typically use successive regressions from partial observations; see Section 4.4.2 below. The general form of an MSM is given by [KCG15, Eq. (**MSM**)]; we only use herein its most basic version, which has the following structure:

$$\mathrm{d}\mathbf{x}(t) = [\mathbf{F}(\mathbf{x}(t)) + \varepsilon \Pi \mathbf{r}(t)]\,\mathrm{d}t, \tag{4.71a}$$

$$\mathrm{d}\mathbf{r}(t) = [\varepsilon \mathrm{C}\mathbf{x}(t) - \mathrm{D}\mathbf{r}(t)]\,\mathrm{d}t + \Sigma \mathrm{d}W_t. \tag{4.71b}$$

Here the observed vector variable $\mathbf{x}(t)$ lies in $\mathbb{R}^{d_1}$ and, for $\varepsilon = 0$, the hidden variables $\mathbf{r}(t)$ in $\mathbb{R}^{d_2}$ evolve in time independently. Otherwise, the dynamics of the $\mathbf{x}$-variables is linearly coupled to that of the $\mathbf{r}$-variables, which act upon (4.71a) as a stochastic forcing, via the canonical projection $\Pi : \mathbb{R}^{d_2} \longrightarrow \mathbb{R}^{d_1}$, while W_t in (4.71b) is a d_2-dimensional Wiener process. The matrix C in $\mathbb{R}^{d_2 \times d_1}$ models the feedback of the $\mathbf{x}$-process onto the

$\mathbf{r}$-variables. In the case of $\mathrm{C} \equiv 0$, $\mathbf{r}$ would evolve according to an O-U process with drift matrix D and covariance matrix $\Sigma\Sigma^\top$. For the sake of simplicity, in the calculations that follow, we restrict ourselves to the case $d_1 = d_2$ so that the projection Π reduces to the identity.

The more general MSM with non-linear coupling considered in [KCG15] was shown to be equivalent to a SIDE with explicit expressions for the memory kernels and stochastic forcing being obtained; see [KCG15, Proposition 3.3]. The noise term there results from successive convolutions of the homogeneous solutions of the lower levels of the system with an O-U process. In particular, using the Itô stochastic calculus, one readily obtains a SIDE that is equivalent to an MSM; see [KCG15, Section 3.2] and Appendix C below.

We show next that a SIDE with the same linear-memory structure can actually be obtained by using the operator formalism presented in Section 4.2 above. One might object that an MSM is a stochastic system, due to the presence of white noise in the hidden layer, whereas the theory presented above applies to deterministic dynamics. However, it was clarified in Chapter 2 that the algebraic manipulations with operators can be done seamlessly for deterministic or stochastic systems. In fact, given a smooth, C^∞ observable

$$\begin{aligned} \Psi : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} &\longrightarrow \mathbb{R}, \\ (\mathbf{x}, \mathbf{r}) &\longmapsto \Psi(\mathbf{x}, \mathbf{r}), \end{aligned} \tag{4.72}$$

its expected value along a stochastic trajectory $X_t = (\mathbf{x}(t), \mathbf{r}(t))^\top$ solving Eq. (4.71), namely $\mathbb{E}\left[\Psi(X_t)\right]$, defines a Markov semigroup by means of the exponential of the generator $\mathcal{L}^*$ in Eq. (2.13), which solves the backward-Kolmogorov equation associated with Eq. (4.71):

$$\partial_t\left(e^{t\mathcal{L}^*}\Psi\right) = \begin{bmatrix} \mathbf{F}(\mathbf{x}) + \varepsilon\mathbf{r} \\ \varepsilon\mathrm{C}\mathbf{x} - \mathrm{D}\mathbf{r} \end{bmatrix} \cdot \nabla e^{t\mathcal{L}^*}\Psi + \frac{1}{2}\begin{bmatrix} 0 \\ \Sigma\Sigma^* \nabla_{\mathbf{r}}^2 e^{t\mathcal{L}^*}\Psi \end{bmatrix}, \tag{4.73}$$

the only difference with respect to the transport equation (4.3) lies in the presence of a second order differential operator induced by the white noise. We introduce the operators $\mathcal{L}_0^*$ and $\mathcal{L}_1^*$ in the same spirit as those in Eqs. (2.25) and (4.4):

$$\mathcal{L}_0^* = \begin{bmatrix} \mathbf{F}(\mathbf{x}) \\ -\mathrm{D}\mathbf{r} \end{bmatrix} \cdot \nabla + \begin{bmatrix} 0 \\ \Sigma\Sigma^* : \nabla_{\mathbf{r}}^2 \end{bmatrix}; \tag{4.74a}$$

$$\mathcal{L}_1^* = \begin{bmatrix} \mathbf{r} \\ \mathrm{C}\mathbf{x} \end{bmatrix} \cdot \nabla, \tag{4.74b}$$

which play a role that is analogous to their deterministic relatives in Eq. (4.4) of the previous section. Again, the operator $\mathcal{L}_1^*$ is viewed as a perturbation to the operator $\mathcal{L}_0^*$ due to the coupling. If one considers observables $\Psi = \Psi(\mathbf{x})$, Eq. (4.73) becomes at time

$t = 0$:

$$\partial_t\Big(e^{t\mathcal{L}^*}\Psi\Big)|_{t=0} = \Big[\mathbf{F}(\mathbf{x}) + \varepsilon\mathbf{r}\Big] \cdot \nabla_{\mathbf{x}}\Psi, \tag{4.75}$$

and we apply now, as in Section 4.2, the Dyson perturbative expansion. By virtue of the formula (4.22), the parametrisation leads to a memory equation of the form:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)) + \varepsilon\eta(t) + \varepsilon^2 \int_0^t \mathcal{K}(s, \mathbf{x}(t-s))\mathrm{d}s, \tag{4.76}$$

where the hidden variables in the decoupled regime are governed by an O-U process with invariant measure $\mu_{\mathbf{r}}$. The properties of the stochastic noise $\eta(t)$ are given by:

$$\begin{aligned}
\Big\langle \eta(t)\eta^\top(0) \Big\rangle &= \int \mathrm{d}\mu_{\mathbf{r}}(\mathbf{r}_0) e^{t\mathcal{L}_0^*} \mathbf{r}_0\mathbf{r}_0^\top && \text{(4.77a)}\\
&= \int \mathrm{d}\mu_{\mathbf{r}}(\mathbf{r}_0) \mathbb{E}\left[\mathbf{r}(t)|\mathbf{r}_0\right] \mathbb{E}\left[\mathbf{r}(0)|\mathbf{r}_0\right]^\top && \text{(4.77b)}\\
&= \int \mathrm{d}\mu_{\mathbf{r}}(\mathbf{r}_0) e^{-t\mathrm{D}} \mathbf{r}_0\mathbf{r}_0^\top && \text{(4.77c)}\\
&= \int \mathrm{d}\mu_{\mathbf{r}}(\mathbf{r}_0) e^{-t\mathrm{D}} \mathbf{r}_0\mathbf{r}_0^\top = e^{-t\mathrm{D}}\Sigma\Sigma^\top, && \text{(4.77d)}
\end{aligned}$$

where $\mathbf{r}$ is a function analogous to the coupling function $\mathbf{C}_{\mathbf{y}}^{\mathbf{x}}$ in the previous section and the initial condition $\mathbf{r}_0$ is assumed to be normally distributed with zero mean and variance $\Sigma\Sigma^*$. The memory kernel is reads as:

$$\begin{aligned}
\mathcal{K}(s, \mathbf{x}(t-s)) &= \int \mathrm{d}\mu_{\mathbf{r}}(\mathbf{r}_0) \mathrm{C}\mathbf{x}(t-s) \cdot \nabla_{\mathbf{r}_0} \mathbb{E}\left[\mathbf{r}(s)|\mathbf{r}_0\right] && \text{(4.78a)}\\
&= \int \mathrm{d}\mu_{\mathbf{r}}(\mathbf{r}_0) \mathrm{C}\mathbf{x}(t-s) \cdot \nabla_{\mathbf{r}_0} e^{-s\mathrm{D}} \mathbf{r}_0 && \text{(4.78b)}\\
&= \mathrm{C}\mathbf{x}(t-s) \cdot e^{-s\mathrm{D}} && \text{(4.78c)}\\
&= e^{-s\mathrm{D}} \mathrm{C}\mathbf{x}(t-s). && \text{(4.78d)}
\end{aligned}$$

Using the intermediate steps above, the explicit d_1-dimensional equation becomes, finally:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)) + \varepsilon\eta(t) + \varepsilon^2 \int_0^t e^{-s\mathrm{D}} \mathrm{C}\mathbf{x}(t-s)\mathrm{d}s. \tag{4.79}$$

The integro-differential equation above is the same as Eq. (C.2) one obtains using the Itô integration described in Appendix C. This similarity of results occurs because we are considering the case of additive coupling, and the Dyson expansion can be truncated after the memory term proportional to ε^2, cf. Section 4.2.1.1.

4.4.2 Empirical Model Reduction (EMR)

As discussed in Sections 4.1 and 4.2, and illustrated in Fig. 4.1, the evolution of the resolved variables is forced by fluctuating terms and the effects of the previous state of

the system. It is desirable, therefore, to construct a full model of the system even when only capable to partially observe it. The EMR methodology [KKG05] aims at achieving this goal; we discuss it below in the broader context of MSMs. Note that EMR provides a solution for the dynamical closure of partially observed systems and thus it differs from the methodology recently proposed in [BPK16], which requires one to fully observe the system for the data-driven discovery of its underlying equations to work.

Having a set of reduced n, d_1-dimensional observations $\{\mathbf{x}_i : i = 1, \ldots, n\}$ recorded every $\mathrm{d}t$ time units, one seeks to regress the tendencies[1] $\{\mathrm{d}\mathbf{x}_i : i = 1, \ldots, n\}$ of the data onto a quadratic function of the form:

$$\mathbf{F}(\mathbf{x}) = \mathbf{f} + \mathbf{b} \cdot \mathbf{x} + \mathbf{Q}(\mathbf{x}), \tag{4.81}$$

where $\mathbf{b}$ in $\mathbb{R}^{d_1 \times d_1}$ describes dissipative processes and $\mathbf{Q}$ is a quadratic form describing self-interaction between the $\mathbf{x}$-variables. The ith component of the quadratic form is given by:

$$[\mathbf{Q}]_i = \mathbf{x}^\top \mathbf{A}_i \mathbf{x}, \tag{4.82}$$

where $\mathbf{A}_i$ in $\mathbb{R}^{d_1 \times d_1}$. The function $\mathbf{F}$ is expected to approximate the vector field driving the dynamics in the absence of hidden external influences. Of course, performing regressions yields an error called *residual* $\{\mathbf{r}_i : i = 1, \ldots, n\}$. Hence, the evolution of the $\mathbf{x}$-variables satisfies the equation:

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + \mathbf{r}. \tag{4.83}$$

At this point, one can study the properties of the residual time series $\{\mathbf{r}_i\}_{i=1}^n$ and construct a model able to reproduce its main statistical features. However, we know that if it is possible to sample all the variables of the dynamical system of interest, one expects that the residuals are explained by the errors committed exclusively in the regression algorithm. If some sort of subsampling is done, whether spatial or temporal, the residuals are due also to the delayed influence of unresolved processes that are involved in the coupling.

Allowing for the main level variables $\mathbf{x}$ to be linearly coupled with the residual $\mathbf{r}$, we are creating a model that is able to incorporate memory effects as well. Hence, for each

[1] Tendencies can be calculated using a various finite-difference methods of different orders. One can use a $\mathrm{d}t^2$ order forward difference

$$\mathrm{d}\mathbf{x}_i = \frac{-3\mathbf{x}_i + 4\mathbf{x}_{i+1} - \mathbf{x}_{i+2}}{2\mathrm{d}t}, \tag{4.80}$$

as done, for instance, in [KKG05].

component i of $\mathbf{x}$ we have:

$$\frac{\mathrm{d}[\mathbf{x}]_i}{\mathrm{d}t} = [\mathbf{f}]_i + \mathbf{b}_i^{(0)} \cdot \mathbf{x} + \mathbf{x}^\top \mathbf{A}_i \mathbf{x} + [\mathbf{r}^{(0)}]_i, \tag{4.84a}$$

$$\frac{\mathrm{d}[\mathbf{r}^{(0)}]_i}{\mathrm{d}t} = \mathbf{b}_i^{(1)} \cdot [\mathbf{x}, \mathbf{r}^{(0)}] + [\mathbf{r}^{(1)}]_i, \tag{4.84b}$$

$$\frac{\mathrm{d}[\mathbf{r}^{(1)}]_i}{\mathrm{d}t} = \mathbf{b}_i^{(2)} \cdot [\mathbf{x}, \mathbf{r}^{(0)}, \mathbf{r}^{(1)}] + [\mathbf{r}^{(2)}]_i, \tag{4.84c}$$

$$\vdots \tag{4.84d}$$

$$\frac{\mathrm{d}[\mathbf{r}^{(l)}]_i}{\mathrm{d}t} = \mathbf{b}_i^{(l+1)} \cdot [\mathbf{x}, \mathbf{r}^{(0)}, \mathbf{r}^{(1)}, \ldots, , \mathbf{r}^{(l)}] + [\mathbf{r}^{(l+1)}]_i, \tag{4.84e}$$

where we have introduced new matrices $\mathbf{b}^{(j)}$ in $\mathbb{R}^{d_1 \times (j+2)d_1}$ that model the linear coupling. The residual at the last level $[\mathbf{r}^{(l+1)}]_i$ is assumed to obey a Wiener process for which the correlation matrix is obtained from the last residual time series. The choice of stochastic process in the last step can only be done if the decorrelation of $[\mathbf{r}^{(l+1)}]_i$ is sufficiently fast according to the timescale set by $\mathrm{d}t$. This motivates the problem of choosing the optimal number l of levels.

Several criteria have been established to determine the optimal number of levels l. The basic idea is that the resulting $(l+1)$-residual in Eq. (4.84e) should be well approximated by Gaussian white noise with some learnt covariance [KCG15]; [KKG10]. One has, therefore, to test whether the residual variables decorrelate at lag $\mathrm{d}t$ and whether the lag-0 covariance matrix is invariant in the last levels. Consequently, regression on the tendency of the optimal level $\mathbf{r}^{(l)}$ should yield:

$$\mathbf{r}^{(l+1)} - \mathbf{r}^{(l)} \simeq -\mathbf{r}^{(l)} + \gamma^{(l)}, \tag{4.85}$$

where $\gamma^{(l)}$ is the residual of the previous regression and is approximately equal to $\mathbf{r}^{(l+1)}$. Hence, $\gamma^{(l)}$ would become a lagged version of $\mathbf{r}^{(l+1)}$. Subject to this assumption, it is possible to estimate the optimal value of the coefficient of determination R^2:

$$R^2 = 1 - \frac{\sum_k \gamma_l^2}{\sum_k \left(\mathbf{r}^{(l+1)} - \mathbf{r}^{(l)}\right)^2} \simeq 1 - \frac{\sum_l \mathbf{r}^{(l+1)^2}}{\sum_l \mathbf{r}^{(l+1)^2} + \mathbf{r}^{(l)^2}} \simeq 0.5. \tag{4.86}$$

This means that, when the amount of unexplained variance of the last regression is 50%, one has reached the optimal number of levels. It is worth stressing that the empirical model (4.84) has the structure of an MSM (4.71), as discussed in [KCG15], since the memory is fully explained by a set of linearly driven extended variables. It can, therewith, be integrated to transform it into an integro-differential equation with explicit formulas for the fluctuating noise and memory kernel, cf. [KCG15, Proposition 3.3]; see also Section 4.4.1 for such a transformation from another perspective.

Finally, note that the aforementioned stopping criterion for EMR — namely $R^2 \simeq 0.5$, see [KCG15, Appendix A] — is based on decorrelation times and it is also present in the multilevel WL equation (4.57). We noted, in fact, in Remark 4.2.5 of Section 4.2.2 that the last level modelled by Eq. (4.57c) decorrelates the fastest with respect to the rest. Ultimately, making this point amounts to saying that a low number of levels is expected to arise in the EMR method, provided most of the eigenvalues λ_j in Theorem 4.2.1 are located far away from the imaginary axis, except for a very few of them. Conversely, if the Koopman eigenvalues cluster near the imaginary axis or do not exhibit a spectral gap located at a suitably defined, small negative real part, many levels are expected to be needed to capture the hidden dynamics; see again Remark 4.2.5.

4.5 Numerical Experiments

In sections 4.2 and 4.4, we have shown that both the WL top-down approach and the EMR data-driven method yield a set of multilevel equations for the variables of interest in a multiscale system. In particular, both approaches give explicit formulas for the fluctuation term and memory kernel in the GLE (4.1b) of the Mori-Zwanzig formalism. Furthermore, their Markovian representation share that the hidden layers are linearly driven with a white noise background; see Eqs. (4.33) and (4.71). We now compare the two approaches to model reduction in a simple, conceptual stochastic climate model.

Since the modelling of geophysical flows is the primary motivation for this research, we consider a set of SDEs proposed in [FMB07, among others] as a physically consistent climate "toy" model. In such a model, the main $\mathbf{x}$-variables are slow and weakly coupled to the fast $\mathbf{y}$-variables. The latter correspond to weather fluctuations and carry, in fact, most of the system's variance. The model's governing equations are:

$$\mathrm{d}x_1 = \{-x_2\,(L_{12} + a_1 x_1 + a_2 x_2) - d_1 x_1 + F_1 + \varepsilon\,(L_{13} y_1 + c_{134} y_1 y_2)\}\,\mathrm{d}t, \tag{4.87a}$$

$$\mathrm{d}x_2 = \{x_1\,(L_{21} + a_1 x_1 + a_2 x_2) - d_2 x_2 + F_2 + \varepsilon L_{24} y_2\}\,\mathrm{d}t, \tag{4.87b}$$

$$\mathrm{d}y_1 = \left\{\varepsilon\,(-L_{13} x_1 + c_{341} y_2 x_1) + F_3 - \frac{\gamma_1}{h} y_1\right\}\mathrm{d}t + \frac{\sigma_1}{\sqrt{h}}\mathrm{d}W_t^{(1)}, \tag{4.87c}$$

$$\mathrm{d}y_2 = \left\{-\varepsilon\,(L_{24} x_2 + c_{413} y_1 x_2) + F_4 - \frac{\gamma_2}{h} y_2\right\}\mathrm{d}t + \frac{\sigma_2}{\sqrt{h}}\mathrm{d}W_t^{(2)}, \tag{4.87d}$$

where $W_t^{(1)}$ and $W_t^{(2)}$ are two independent Wiener processes. These equations describe the evolution of four real variables $\mathbf{x} = (x_1, x_2)$ and $\mathbf{y} = (y_1, y_2)$; their timescale separation is determined by the parameter h and the coupling strength is controlled by ε. The parameter values used herein are: $c_{134} = c_{341} = 0.25$, $c_{413} = -0.5$, $L_{12} = L_{21} = 1$, $L_{24} = -L_{13} = 1$, $a_1 = -a_2 = 1$, $d_1 = 0.2$, $d_2 = 0.1$, $F_1 = -0.25$, $F_2 = F_3 = F_4 = 0$, $\gamma_1 = 2$, $\gamma_2 = 1$ and $\sigma_1 = \sigma_2 = 1$. The timescale separation and the coupling strength are

set to be $h = 0.1$ and $\varepsilon = 0.5$, respectively.

4.5.1 WL Approximation

Notice that the hidden variables evolve according to a decoupled O-U process. Taking advantage of this fact, we calculate the weak-coupling–limit approximation of the model, according to the formulas presented in Section 4.2 for the separable coupling functions given by:

$$\mathbf{C}^{\mathbf{x}}(\mathbf{x},\mathbf{y}) = \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}) = \begin{bmatrix} L_{13}y_1 + c_{134}y_1y_2, \\ L_{24}y_2 \end{bmatrix}, \tag{4.88a}$$

$$\mathbf{C}^{\mathbf{y}}(\mathbf{x},\mathbf{y}) = \mathbf{C}^{\mathbf{y}}_{\mathbf{x}}(\mathbf{x}) : \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}) = \begin{bmatrix} -L_{13}x_1 + c_{341}y_2x_1 \\ -L_{24}x_2 + c_{413}y_1x_2 \end{bmatrix}. \tag{4.88b}$$

The coupling function $\mathbf{C}^{\mathbf{x}}$ in the slow equation (4.88a) is independent of the $\mathbf{x}$-variables, indicating that the noise correction can be additively incorporated and implemented by examining the decoupled hidden process. Note that the functional form of $\mathbf{C}^{\mathbf{y}}$ implies that the WL parametrisation cannot be exact in ε, as noted in Eqs. (4.12) and (4.13). This indicates that the WL reduced model will not only introduce an error in averaging over the decoupled steady state, but also that the Dyson expansion Eq. (4.6) has to be truncated at ε^3, rather than merely at ε^2 where no memory effects would be included.

According to the WL model (4.22) discussed in Section 4.2, the fluctuation terms correspond to the decoupled evolution of the coupling function $\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}$, concretely as in Eq. (4.20). This allows to directly compute the correlation function:

$$\left\langle \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y})\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y}(t))^{\top} \right\rangle = \begin{bmatrix} L_{13}^2 e^{-(\gamma_1/h)t}\frac{\sigma_1^2}{2\gamma_1} + c_{134}^2 e^{-(\gamma_1+\gamma_2)t/h}\frac{\sigma_1^2\sigma_2^2}{2\gamma_1\gamma_2} & 0 \\ 0 & L_{24}^2 e^{-(\gamma_2/h)t} \end{bmatrix}. \tag{4.89}$$

From Eq. (4.89) we deduce that the noise covariance matrix for the given parameter values is given by:

$$\left\langle \mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y})\mathbf{C}^{\mathbf{x}}_{\mathbf{y}}(\mathbf{y})^{\top} \right\rangle = \begin{bmatrix} 0.2578... & 0 \\ 0 & 0.5 \end{bmatrix}. \tag{4.90}$$

The memory kernel $\mathcal{K}$, which is a vector of two components $[\kappa_1, \kappa_2]^{\top}$, is given by:

$$\mathcal{K}(s,\mathbf{x}) = \begin{bmatrix} \kappa_1(s,\mathbf{x}) \\ \kappa_2(s,\mathbf{x}) \end{bmatrix} = \left\langle \mathbf{C}^{\mathbf{y}}(\mathbf{x},\mathbf{y}) \cdot \nabla_{\mathbf{y}}\mathbf{C}^{\mathbf{x}}(\mathbf{x}(s),\mathbf{y}(s)) \right\rangle; \tag{4.91}$$

here the brackets $\langle\cdot\rangle$ indicate the averages for the uncoupled equilibrium in the $\mathbf{y}$-variables,

which happen to be a set of independent O-U processes. Explicitly,

$$\kappa_1(s,\mathbf{x}) = \left\langle \left(-L_{13}x_1 + c_{341}y_2x_1\right)\partial_{y_1}\left(L_{13}y_1(s) + c_{134}y_1(s)y_2(s)\right)\right\rangle \tag{4.92a}$$

$$- \left\langle \left(L_{24}x_2 + c_{413}y_1x_2\right)\partial_{y_2}\left(L_{13}y_1(s) + c_{134}y_1(s)y_2(s)\right)\right\rangle \tag{4.92b}$$

$$= - L_{13}^2 e^{-(\gamma_1/h)s}x_1 + c_{341}c_{134}e^{-(\gamma_1+\gamma_2)s/h}\frac{\sigma_2^2}{2\gamma_2}x_1 \tag{4.92c}$$

$$+ c_{134}c_{341}e^{-(\gamma_1+\gamma_2)s/h}\frac{\sigma_1^2}{2\gamma_1}x_1, \tag{4.92d}$$

$$\kappa_2(s,\mathbf{x}) = \left\langle \left(-L_{13}x_1 + c_{341}y_2x_1\right)\partial_{y_1}\left(L_{24}y_2(s)\right)\right\rangle - \left\langle L_{24}x_2\partial_{y_2}\left(L_{24}y_2(s)\right)\right\rangle \tag{4.92e}$$

$$= - L_{24}^2 e^{-(\gamma_2/h)s}x_2. \tag{4.92f}$$

The reduced-order model obtained herewith does give explicit formulas for the evaluation of the stochastic noise and the memory kernel, independently of the timescale separation h, although these formulas are rather complicated. Still, the scheme remains the same when changing parameter values, so it is flexible in studying different scenarios.

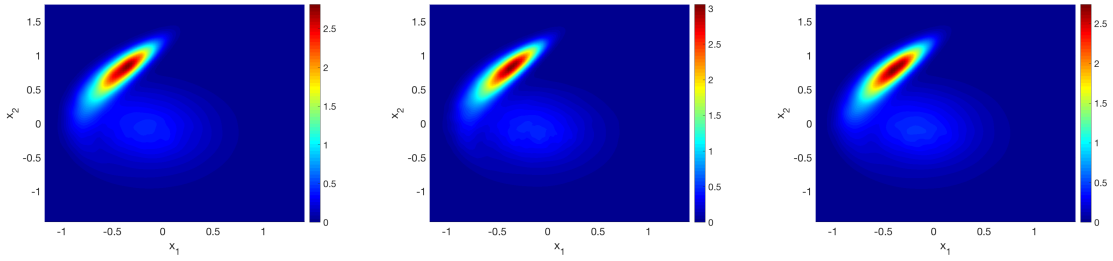
4.5.2 EMR Model and Results

Basic EMR algorithm implementation. Regarding the data-driven EMR protocol, we integrated the full model with a time step of $\mathrm{d}_\ell t = 10^{-3}$ time units for a duration of $T_\ell = 10^4$ time units in order to learn the model parameters. Then, a separate run was performed as a test case. This time, the EMR system was integrated together with the full model using a time step of $\mathrm{d}_\tau t = 10^{-2}$ time units for a total of $T_\tau = 10^5$ time units. The equations were solved using a fourth-order Runge-Kutta and a Euler-Maruyama method for the deterministic and stochastic components, respectively. By sampling every time step, we learned an EMR model whose coefficients were explicitly found. The convergence criterion $R^2 \simeq 0.5$ was attained by adding two extra levels, for a total of three.

The climatologies of the slow $\mathbf{x}$-variables are obtained using data from the full model, the EMR model and the WL parametrisation. The two-dimensional probability density functions (PDFs) of the stochastic model (4.87) in the (x_1, x_2)-plane are shown in Fig. 4.2. These PDFs were calculated by employing the MATLAB R2019a kernel smoothing function *ksdensity*. Their respective marginals are shown in Fig. 4.3. The agreement between the two methodologies when approximating the clearly non-Gaussian density arising from the full model is clearly excellent.

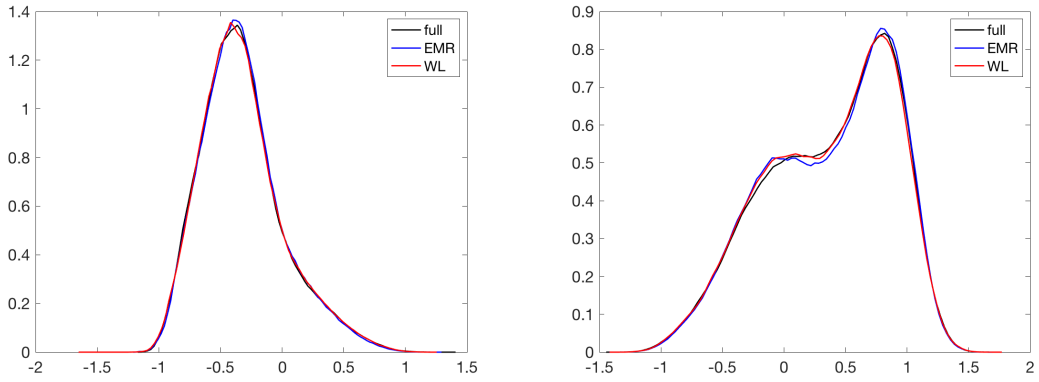
The timescale separation between the $\mathbf{x}$-variables and the $\mathbf{y}$-variables is clearly depicted on panel (a) of Fig. 4.4, where the fast variables decorrelate almost instantly compared to the slow ones. The approximation of these autocorrelation functions is also obtained

using the EMR and WL methods.



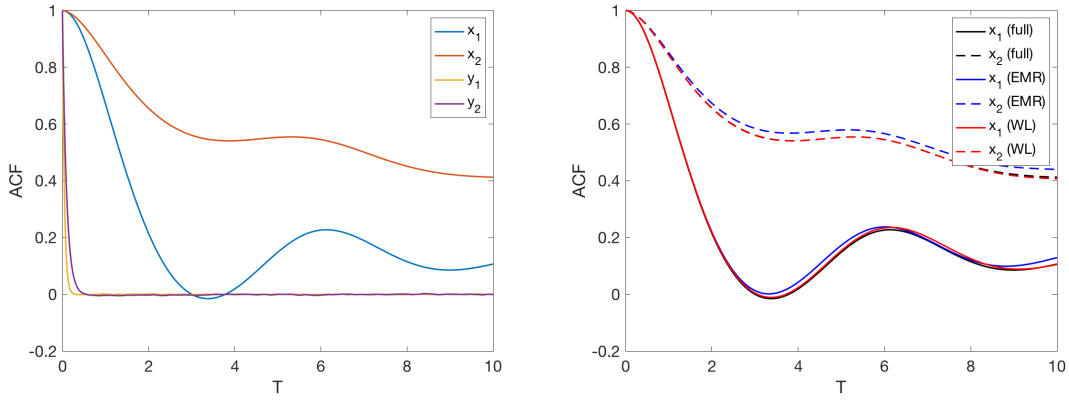
(a) Full model (b) EMR model (c) WL model

Figure 4.2: Two-dimensional probability density functions (PDFs) of the stochastic model (4.87) in the (x_1, x_2)-plane, as obtained with: (a) the full integration; (b) an integration of the EMR model; and (c) the WL parametrisation. The timescale separation parameter used is $h = 0.1$. The PDFs shown here and in Fig. 4.5 were obtained by using the MATLAB R2019a kernel smoothing function *ksdensity*.



(a) x_1-PDF (b) x_2-PDF

Figure 4.3: PDFs of (a) the x_1 variable; and (b) the x_2 variable. The separation parameter is $h = 0.1$ and colors used for each method are indicated by the legends inside the panels.



(a) (b)

Figure 4.4: Autocorrelation functions for the four variables x_1, x_2, y_1, y_2 obtained (a) from the full model; and (b) the comparison of the corresponding results for x_1, x_2 with the full model, the EMR model, and the WL parametrisation. See legend for the choice of lines; $h = 0.1$.

In general, cf. [KKG05], the regressions performed in the main level (4.84a) of the EMR model allow one to effectively reconstruct the coefficients of a weakly coupled model; see another example in Appendix D. The EMR methodology, though, only allows for linear coupling between the slow $\mathbf{x}$'s and the fast $\mathbf{y}$'s. The non-linear coupling between the slow and fast variables in system (4.87) compromises the estimation of the main model parameters in Eq. (4.84a), so that we cannot expect to recover the original, full model's behaviour given by system (4.87). The EMR model coefficients at the first and second levels are as shown on tables 4.1 and 4.2, respectively.

Table 4.1: Empirically estimated EMR model coefficients at the first level, Eq. (4.84a), for $h = 0.1$. First column gives the coefficients for the constant forcing $\mathbf{f}^{(0)}$, the second and third columns indicate the linear component of the vector field $\mathbf{b}^{(0)}$ and the last three columns determine the quadratic form $\mathbf{A}$.

$\mathbf{f}^{(0)}$	x_1	x_2	x_1^2	x_1x_2	x_2^2
-0.31404	-0.50954	-0.065313	0	-1.0092	0.99704
-0.15356	0.12353	0.21979	1.0092	-0.99704	0

Table 4.2: Empirically estimated EMR model coefficients at the second level, Eq. (4.84b), for $h = 0.1$. First column gives the coefficients for the constant forcing $\mathbf{f}^{(1)}$, the next two columns indicate the linear coupling to the main level (i.e. the first two columns of $\mathbf{b}^{(1)}$) and the last two columns determine the linear drift for the second level (i.e. the last two columns of $\mathbf{b}^{(1)}$).

$\mathbf{f}^{(1)}$	x_1	x_2	$r_1^{(1)}$	$r_2^{(1)}$
0	-5.6397e-4	-6.9382e-05	-20.2823	0.016165
0	-1.648e-4	-4.72e-4	0.086621	-10.0426

As discussed in Section 4.4.2, the EMR has the structure of an MSM and it can be recast into an integro-differential equation. If, for simplicity, one only considers the first added level, the EMR can be readily integrated giving the following equation for the evolution of the slow variables $\mathbf{x} = (x_1, x_2)$:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)) + e^{-\mathrm{D}t}\mathbf{y}(0) + \int_0^t e^{-\mathrm{D}(t-s)}\Sigma \mathrm{d}W_s + \int_0^t e^{-\mathrm{D}(t-s)}\mathrm{C}\mathbf{x}(s)\mathrm{d}s. \tag{4.93}$$

Here, W_s is an independent two-dimensional Wiener process and

$$\mathrm{D} = \begin{bmatrix} -19.9982 & 2.1122 \cdot 10^{-3} \\ -0.77528 & -10.116 \end{bmatrix}, \mathrm{C} = \begin{bmatrix} -5 \cdot 10^{-3} & -5 \cdot 10^{-4} \\ -1 \cdot 10^{-3} & -5 \cdot 10^{-3} \end{bmatrix}, \tag{4.94a}$$

$$\Sigma = \begin{bmatrix} 0.2626 & -0.0014 \\ -0.0014 & 0.5013 \end{bmatrix}. \tag{4.94b}$$

First thing to note is that the matrix C has a small norm and, by virtue of Eq. (4.93), it means that memory effects are going to be very small. On the other hand, the eigenvalues of the matrix D are $\lambda_1 \simeq -20, \lambda_2 \simeq -10$, which are approximately the drift coefficients of the uncoupled O-U process driving the $\mathbf{y}$-variables. This indicates that the exponential kernel is damping the effects of the $\mathbf{x}$-variables in past times rather quickly. Moreover, the covariance matrix Σ corresponds to that obtained by integrating the $\mathbf{y}$-dynamics independently, according to Eq. (4.90).

Regarding the WL approximation, we stress that the $\mathbf{y}$-variables are no longer present, after taking the averages in its construction. The memory kernel $\mathcal{K}$ in this case differs from the matrix D above, although its dominant terms correspond to its eigenvalues. Note that, if $L_{13} = 0$, the coupling function $\mathbf{C}_{\mathbf{y}}^{\mathbf{x}}$ would project entirely onto the eigenfunction of the O-U process associated with the generator eigenvalue $-(\gamma_1 + \gamma_2)$. The same statement would hold for $c_{134} = 0$, where in this case $\mathbf{C}_{\mathbf{y}}^{\mathbf{x}}$ projects onto the eigenfunctions associated with the eigenvalue $-\gamma_1$.

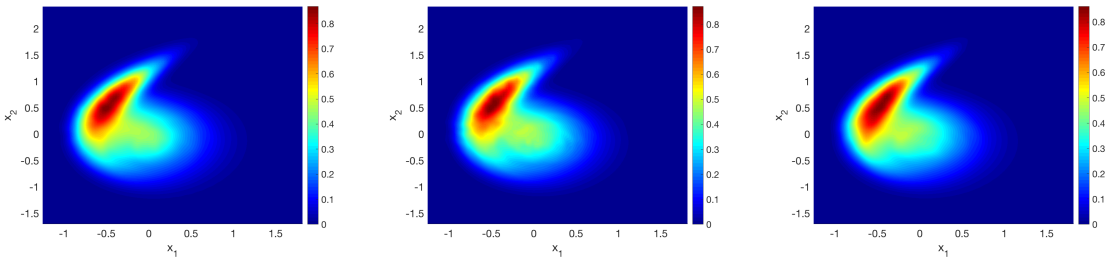
Reduced Time Scale Separation. The parameter h controls the timescale separation in the evolution of the $\mathbf{x}$- and $\mathbf{y}$-variables. Here we set $h = 1$ so that this separation is reduced by an order of magnitude as illustrated by the autocorrelation functions in Fig. 4.7.

In the WL parametrisation, there is no need to sample the dynamics in order to construct it, since the formulas of Section 4.2 are explicit and the construction does not depend on h. In the case of model (4.87), the covariance matrix and time correlations of the WL noise correction are thus given by Eq. (4.89) with no reference to h. The memory term, though, is expected to change as the kernel $\mathcal{K}$ will decay more slowly by a factor of 10. Therefore, memory effects are more important, as expected.

The EMR approach, on the other hand, requires a new learning phase for this value of $h = 1$. We used the same numerical integration parameters $\mathrm{d}_\ell t = 10^{-3}$ and $T_\ell = 10^4$ time units as for the previous case. In the first level regression, one observes that the coefficient values listed in Table 4.3 are essentially the same from those estimated in the previous case, for $h = 0.1$, and listed in Table 4.1. The second level coefficients, for $h = 1$, are shown on Table 4.4. Regarding the convergence levels of EMR, it was not affected by changes in the timescale separation parameter namely, from $h = 0.1$ to $h = 1$. Probably the value of h was not that important here because of the low dimensionality and stochastic nature of the hidden process. However, convergence is likely to be altered in more complicated models, as illustrated in Appendix D.

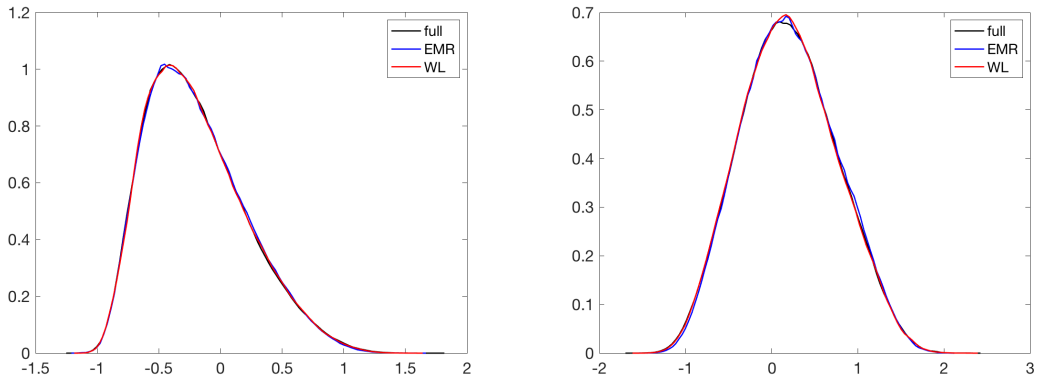
The covariance matrix Σ of the noise correction is indicated in Eq. (4.95b) and it agrees well with the previous values, for $h = 0.1$, as given in Eq. (4.94b). The matrix C that indicates the strength of the memory effects also has a magnitude that is of the same order as that in the previous case of $h = 0.1$, which is rather suprising, given the factor of 10 in timescale separation h; compare Eqs. (4.94a) and (4.95a). This observation tells us that the loss of Markovianity might be intrinsic to the nature of the coupling rather than being due to the timescale separation, even though, in the limit case of infinite scale separation, memory effects will disappear entirely. The memory kernel, as determined by matrix D, scales almost exactly with the timescale separation and it is expected to change depending on how the coupling functions project onto the eigenspaces of the underlying Orstein-Uhlenbeck process, as discussed more generally earlier in Theorem 4.2.1.

The performance of both parametrisation techniques is summarised in Figs. 4.5–4.7. These figures are the exact counterparts of Figs. 4.2–4.4 for the reduced timescale separation $h = 1$. First of all, we note from Fig. 4.7 (a) that, for $h = 1$ there is indeed no strict timescale separation, as indicated by the autocorrelation functions obtained from the full model. Secondly, consideration of Figs. 4.5(a)–(c), 4.6(a,b) and 4.7(b) shows that neither the WL nor the EMR approach seems to be affected by the timescale reduction.



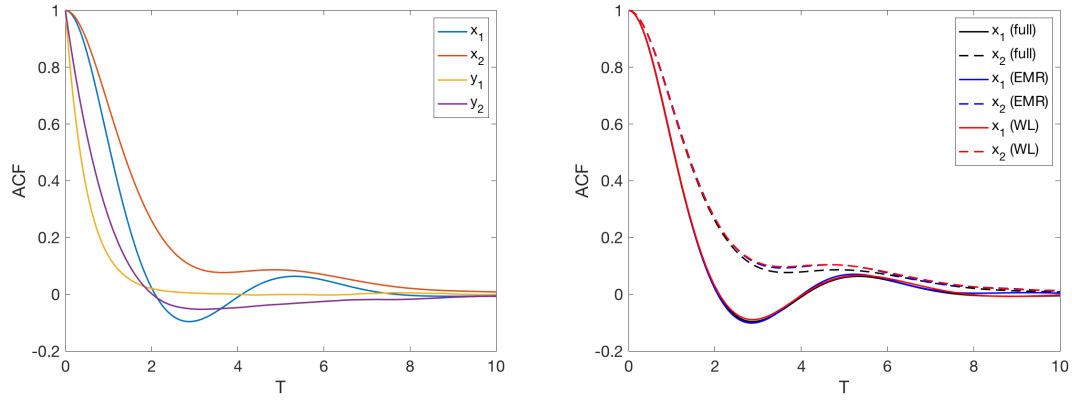
(a) Full model (b) EMR model (c) WL parametrisation

Figure 4.5: Two-dimensional smoothed PDFs of the stochastic model (4.87), but with a timescale separation of $h = 1$. Panels (a), (b) and (c) are calculated by integrating the full model, the EMR model and WL approximation, respectively, as in Fig. 4.2.



(a) x_1-PDF (b) x_2-PDF

Figure 4.6: PDFs of (a) the x_1 variable and (b) the x_2 variable, for a timescale separation of $h = 1$; compare with Fig. 4.3.



(a) (b)

Figure 4.7: Autocorrelation functions for the four variables x_1, x_2, y_1, y_2 obtained (a) from the full model; and (b) the comparison of the corresponding results for x_1, x_2 with the full model, the EMR model, and the WL parametrisation. See legend for the choice of lines; $h = 1$.

Table 4.3: Empirically estimated EMR model coefficients at the first level, for $h = 1$.

f	x_1	x_2	x_1^2	x_1x_2	x_2^2
-0.31181	-0.339	-0.42944	0	-0.93439	0.97958
-0.16925	0.46833	0.17513	0.93439	-0.97958	0

Table 4.4: Empirically estimated EMR model coefficients at the second level, for $h = 1$.

$f^{(1)}$	x_1	x_2	$r_1^{(1)}$	$r_2^{(1)}$
0	-3.9452e-3	-1.5027e-4	-2.1001	0.016509
0	-5.0312e-4	-3.79e-3	0.054861	-1.1214

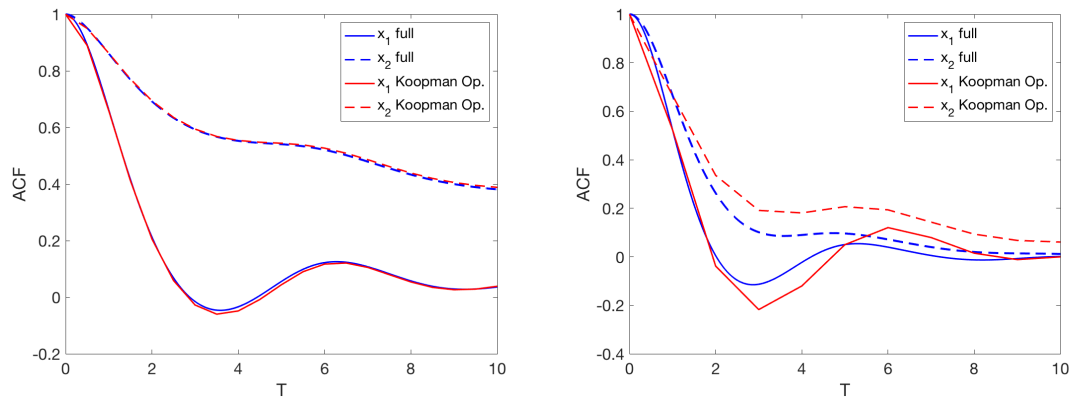
$$\mathrm{D} = \begin{bmatrix} -2.1001 & 0.016509 \\ 0.054861 & -1.1214 \end{bmatrix}, \mathrm{C} = \begin{bmatrix} -4 \cdot 10^{-3} & -1 \cdot 10^{-4} \\ -6 \cdot 10^{-4} & -4 \cdot 10^{-3} \end{bmatrix}, \tag{4.95a}$$

$$\Sigma = \begin{bmatrix} 0.2618 & 0.0011 \\ 0.0011 & 0.4554 \end{bmatrix}. \tag{4.95b}$$

4.5.3 Memory Effects

We would like to end this results section by analysing the role of the memory effects when performing a reduction of the highly idealised model given by Eqs. (4.87). For this purpose, we apply the criterion (4.70) discussed in Section 4.3. We thus spectrally approximate the autocorrelation functions of the variables x_1 and x_2 using Eq. (4.70) with the Koopman/backward-Kolmogorov semigroup $\mathcal{T}_\tau$ estimated using Ulam's method (3.3) with a transition time of $\tau = 0.5$ time units for the case where $h = 0.1$ and $\tau = 1$ time unit for $h = 1$. The choice of different transition times τ depends on how well Eq. (4.70) is approximated. Indeed, for a range of positive but small values of $\tau \leq 1$ we chose those which, for either case, best approximated the the autocorrelation functions of x_1 and x_2 for $T = 10$ time units using transition matrices.

We clearly observe in Fig. 4.8 that the correlation functions can be accurately reconstructed in the case of (a), large timescale separation $h = 0.1$, but not so for (b), $h = 1$. This indicates, naturally, that memory effects are negligible in the first case and relevant in the second, in line with more general considerations of homogenisation theory [PS08]. Furthermore, the impossibility of fulfilling Eq. (4.70) when $h = 1$ indicates that errors due to projection of the Koopman semigroup are attributed exclusively to memory effects and not to artificial diffusion; we refer the reader back to Section 3.3.2 and comments around Fig. 3.4.



(a) (b)

Figure 4.8: Autocorrelation functions for the two $\mathbf{x}$-variables obtained from the full model in the spectral reconstruction using the projected Koopman operator: (a) $h = 0.1$ and (b) $h = 1$.

4.6 Summary and Discussion

Providing efficient and accurate parametrisations for model order-reduction is a key goal in many areas of science. There are two main approaches for constructing parametrisations:

top-down, by deriving the reduced model directly from the evolution equations governing the system through the use of suitable approximations; and data-driven, in which the parametrisations are constructed through suitable optimisation procedures, which are first tuned in a training phase and then actually used in the prediction phase. Both approaches aim to derive the effective dynamics for the variables of interest: formally, this is achieved by applying the Mori-Zwanzig projection operator [Mor65]; [Zwa61] to the full dynamics. The result of doing so is to describe the impact of the hidden variables by formulating a generalised Langevin equation (GLE) (4.1b) for the variables of interest that includes a deterministic, a stochastic, and a non-Markovian component. Top-down and data-driven approaches are conceptually complementary and have different practical advantages and disadvantages. In this chapter, we have shown the fundamental link between a top-down and a data-driven approach that have been formulated and applied in the recent literature. This equivalence was illustrated schematically in Fig. 4.1.

We first revisited in Section 4.2 the WL parametrisation of [WL12]; [WL13], which relies on an assumption of weak coupling between the hidden and observed variables, and have extended the previous results by considering more general coupling classes. We have also shown that the perturbative expansion that yields the WL parametrisation is exact when the coupling between the hidden and resolved variables is additive. In doing so, we noted that he Dyson formalism (4.13) is to essential for computing the effects of the hidden processes on the dynamics of the observed variables, when working at the level of the system's observables. This methodology is explicit, in the sense that no information about the actual coupled process is needed, because the formal computations are performed by considering the limit in which no coupling is present. This is in line with linear response theory and the results of Chapter 2. Other advantages of this approach are that it can be implemented without the need for any hypothesis on the timescale separation between the hidden variables and the observed ones, and that it is also scale adaptive [VL18a].

We addressed systematically the problem of re-Markovianising the WL memory equation, which was first pointed out in [Wou+16] and discussed further in Section 4.2.3 here. This example system (4.59) had one observed and two hidden variables that yielded a scalar WL parametrisation that was re-Markovianised to a extended system with just two scalar differential equations. Throughout Section 4.2, we provided a broader framework for re-Markovianisation later formalised for a scalar equation in Theorem 4.2.1 and described for higher dimensional systems in Remark 4.2.5. The required assumptions for this treatment boil down to having a spectral decomposition like that of Eqs. (2.40) and (4.29). More concretely, it is assumed that the coupling law between the processes projects entirely onto a (simple) point spectrum so that the memory kernel can be written as a sum of exponentially decaying functions; see Eq. (4.30). In addition, if the projection onto eigenfunctions is small, one can afford the reduction of the dimension of the extended model (4.33), as noted in Remark 4.2.4.

The multilevel structure of the re-Markovianisation obtained in Section 4.2 motivated the comparison with multilayer stochastic models (MSMs) in Section 4.4. Such MSMs arise naturally in data-driven reduction methods [KKG05] and they had been shown in [KCG15] to approximate the GLE predicted by Mori [Mor65] and Zwanzig [Zwa01]. We showed in Section 4.4.1 that a seamless application of the WL parametrisation solves the MSM of Eq. (4.71) and coincides with its Itô integration; see Appendix C for clarification. Note that an MSM can be obtained from partial observations of the coupled system, which amounts to the special case of the data-driven empirical model reduction (EMR) methodology [KKG05]. The EMR protocol was revisited here in Section 4.4.2 and it is, in principle, dual to the WL parametrisation, in the sense that only partial observations of the coupled system are required, without the need for knowing the actual equations of motion. Comparing the multilevel structure of Eq. (4.57) with that of Eq. (4.84) suggests that the Koopman eigenvalues λ_j highlighted in Theorem 4.2.1 may help provide insights into the number of levels needed for EMR to converge.

Additionally, we considered in Section 4.5 a conceptual climate model to which we applied both of the methodologies revisited herein. Since both the MSM and the WL parametrisation yield a memory equation that involves integrals and stochastic noise, we were able to compare their structure, as well as their statistical outputs. We found that both methodologies produced equivalent numerical results and that the memory kernel and noise predicted in the WL parametrisation agreed with what was found using the data-driven EMR approach.

Access to the resonances indicated in Theorem 4.2.1 is restricted to analytically tractable processes like that in sections 4.2.3 and 4.5, or low-dimensional systems for which the Markov chain approach of Chapter 3 can be taken to estimate the resonances. A step up in complexity, but with physical relevance, would be to study the coupled two-layer-atmosphere-ocean model of [DV17] where the upper layer of the atmosphere is modeled as an O-U process with a nonlinear coupling with the rest of components. More generally, it is worth noticing, though, that the resonances can be obtained through reduced phase-spaces filtered through adequate observables [Tan+20]. These observables can be seen as the coupling function of Eq. (4.2a), which is more likely to have a tractable dimension and, thence, be suitable for estimating the resonances and modes of Theorem 4.2.1.

Chapter 5

Conclusion

In this work, we have taken an ergodic theoretic scope to understand the effects of perturbations in physical systems. The problems of climate response, sensitivity and variability are motivating of each chapter herein, and have been addressed at a conceptual level. To this end, the theory of operator semigroups has been instrumental and, furthermore, appeared impossible to detach from physical interpretations. In this chapter, we shall reflect on the results presented earlier in this dissertation with vistas to future work.

Estimating climate response and sensitivity is a straightforward task if one has infinite computational power, since one would be able to sample different forcing scenarios and empirically observe the system's reaction to a given perturbation. Unfortunately, this is today impossible, obliging us to utilise a hierarchy of models to achieve a greater understanding of the non-linear processes that can trigger abrupt changes in the climate system [Hel05]; [Ghi15]; [GL20]. We have argued that the linear response theory of statistical physics is relevant tool for understanding climate response, because it provides a framework for calculating non-equilibrium sensitivities which are, in fact, currently not regarded in works of crucial relevance like the latest IPCC reports [IPC14]; [Ghi15]; [IPC21]. It was proposed throughout this work that the differentiability of statistical steady states based on Ruelle's formula (1.9) is an adequate quantitative method for calculating the robustness of a system to a prescribed forcing. In this sense, the stochastic perspective taken in Chapter 2 has facilitated accessing the leading order correction to the system's statistics out of the properties of the Fokker-Planck semigroup generator describing the unforced evolution of probabilities. Indeed, assuming that said semigroup is quasi-compact, relaxation rates in terms of the dominating eigenvalues were identified by means of looking at the response function itself: see Eq. (2.45). The fluctuation-dissipation theorem, equivalently, asserts that response functions should decay as correlation functions, for which spectral decompositions exist [Tan+20]. Furthermore, the Green function formalism was recovered, where the latter is decomposed in terms of the point spectrum regardless of the time-modulation of the forcing: see Eq. (2.64). More interestingly, we proved that the

Green function survives the noise-modulated and correlated forcings as firstly suggested by [Luc12], although the corresponding SDE has to be conceived in the Stratonovich sense. Only in this case can one reconcile the results of [Luc12], [Abr17] and those shown in Sections 2.2 and 2.3.1. This is gathered in Proposition 2.3.1.

Changes in the power spectral densities due to stochastic forcings are given Eq. (2.92) and originally found in [Luc12]. Such formula provides an amenable algorithm for estimating the Ruelle-Pollicot resonances out of time series, as detailed in the paragraph immediately below. While the Ulam projections of the transfer operator employed in Chapter 3 provided such spectral estimates via Markov modelling, they required expensive computations. The method arising from Eq. (2.92), on the other hand, would only require the study of model ensemble trajectories. This idea was partially implemented in [Luc12] to calculate the modulus of the susceptibility function, although the author did not consider the problem of locating its poles. Additionally, here we speculate on the possibility of relating the noise strength employed in Eq. (2.92) with the artificial diffusion entailed by Ulam's projection [FJK13]. An experiment to gauge the scope of these ideas would be most valuable.

The scalar and homogeneous equation (2.107) derived in [Ken71] and revisited here in a classical and stochastic framework, allows to compute the full response of a system without solving the Fokker-Planck equation with the price of introducing memory. The results in this section, however, are abstract, since they involve projected differential operators that do not always have explicit representations, although future investigations should endeavour to find suitable approximations of the different components in Eq. (2.107), especially the memory kernel.

Ulam's method has been used to approximate the transfer operator and Fokker-Planck semigroup in Chapter 3, although we noted that rigorous convergence results are limited to systems possessing absolutely continuous invariant measures or those having (computationally amenable) Markov partitions; we refer the reader back to Remark 3.1.1. It is well known that such method provides stochastic matrix estimates, whose perturbation theory has been tackled in Section 3.2. An Ulam-projected version of Ruelle's linear response formula (1.9) was given in Eq. (3.21) where, in addition, all the non-linear corrections are included. This formula was conceived in the context of stochastic matrices although, similarly to the convergence of Eq. (1.9), the validity of the series expansion (1.9) boiled down to having a sufficiently mixing Markov chain as captured by the ergodicity coefficient: see Eq. (3.31). The ergodicity coefficient can be computed from the matrix entries and provides a bound to the modulus of the second largest eigenvalue of the chain [Dob56]; [IS14]. Furthermore, it is shown in Proposition 3.2.3 that the ergodicity coefficient gives a stability bound for the second eigenvalue. In essence, this result says that the uniqueness of the invariant probability and mixing character of the chain is preserved for perturbations whose size is bounded by the ergodicity coefficient. This result extends

that in Proposition 3.2.2, where the bound is given in terms of the algebraic condition number. Along this lines, it was already suggested immediately below, that the eigenvalue condition number— see Appendix B or [Wil65]; [TE05] for definitions— could improve Proposition 3.2.3 in the limit of small perturbations. We would like to note that a stability result of this sort would be analogous to the full transfer operator version found in [KL99], where the perturbation theory of [Kat66] is resorted to. On a practical note, the eigenvalue condition number has been recently used to check the robustness of transition matrices learned from time series [Tan+19].

Projecting the evolution operators onto a finite basis inherently provokes the loss of semigroup property, which prevents the use of the spectral mapping theorem linking the spectrum of the generator with that of the semigroup [EN00, Chapter IV]. We understand that this is a consequence of the Mori-Zwanzig formalism revisited here in Section 4.1. This question has not been addressed in mathematical terms and the modelling community would greatly benefit from having a priori estimates of the accuracy of Ulam projections. Throughout the numerical approximations of Section 3.3 we have assumed that the semigroup property was satisfied in the sense of Eq. (3.5). Additionally, the introduction of artificial diffusion inherently compromises the accuracy of computations [FR02]; [FJK13]. In the study of the Lorenz 63 system reported in Section 3.3.2, a diagnosis was done in terms of correlation functions— see Fig. 3.4— were it was shown that coarse-grained correlations decay more quickly, although the power spectral peaks were well preserved.

While the O-U process studied in Section 3.3.1 has a purely simple real spectrum with orthogonal eigenfunctions [MPP02], the dissipativeness of the Lorenz 63 model provoked the appearance of complex eigenvalues in the approximate stochastic matrices with non-orthogonal eigenvectors. We observed that the non-normality of such matrices translated in a non-trivial delay in convergence to the steady state, as illustrated in Figure 3.4 (a). Furthermore, the eigenelements of non-normal matrices experience a more sensitive dependence to perturbations which can be better studied using ε-pseudospectra [TE05]. We are, thus, inclined to conjecture that the sensitivity of the system is not only measured by the shrinking of the spectral gap, as calculated for the pitchfork bifurcation [Gas+95] or observed in the transitions between Warm and Snowball climate states [Tan+18], but by large ε-pseudospectral values at the dominating resonances.

The low dimensionality of the problems considered in Section 3.3 allows to perform the box subdivision on the whole phase space. Unfortunately, many physically relevant models posses high-dimensional domains that are numerically intractable. Because of this dimensionality barrier, reduced phase spaces are considered. Results along this line support the applicability of the transfer operator methods in a climatological context, see, e.g., [Che+14]; [TBD15]. However, the inherent loss of Markovianity in the dimensionality reduction requires control on the memory effects introduced by the hidden variables,

again as a consequence of the Mori-Zwanzig formalism. This is something that complicates the study of the response, as pointed out in [LW17] where the robustness of reduced systems with the presence of forcing is assessed. Further research should be oriented on adapting this methodology based on the transfer operator in high-dimensional systems with views to predicting the response of physically relevant models. A first step along these lines is done, in fact, in the already cite work of [Che+14], where narrow spectral gaps in reduced Markov matrices indicate a lack of linear response in an intermediate complexity ENSO model. It was also raised that the linear responses estimated using the Markov chain approach and gathered in tables 3.2-3.4 were heavily dependent on the choice of observable, possibly, because the stable contribution in Ruelle's formula 1.11a is not captured; see [TLD18] for further reference. In this regard, algorithms for calculating the linear response along orbits would be useful to precisely determine the decomposition of the applied fields into the stable and unstable directions of the flow [NW17]. We hope that these techniques provide a way of categorising the suitability of different observables to the methodology of Chapter 3.

To formulate accurate and efficient parametrisations for multiscale processes is a crucial challenge in many areas of science and technology for one of two reasons: either the numerical simulation of all scales active in a given system is computationally unfeasible; or there is a mismatch between model resolution and the granularity and homogeneity of the observations, as in the case in geophysical flows and in the climate system. Moreover, the construction of parametrisations is instrumental to help understand the nature of non-linear fluxes across scales and the physical processes responsible for cascades, instabilities, and feedbacks. In this context, the WL equation first proposed in [WL12] using Ruelle's high-order response formulas and revisited here in Section 4.2 using the operator semigroup expansions, captures the leading order corrections in the statistics of interacting physical processes where there is no need of timescale separation. The Dyson expansion of Eq. (2.18) proved to be instrumental in deriving the WL equation (4.22). It was shown in [Wou+16] through worked examples that the WL equation could be Markovianised to yield reduced sets of equations. We systematically addressed such problem in Theorem 4.2.1, where we assumed that the coupling functions involved in Eq. (4.24) projected entirely onto the point spectrum so that the associated WL equation could be recast into a Markovian system of extended variables, making the latter quasi-Markovian [Pav14]. The resulting system (4.33) appeared to posses the structure of a Multilevel Stochastic Model (MSM), where the hidden variables are linearly driven and explain the integral kernel in the WL equation. More importantly, on Remark 4.2.5 the hidden variables were hierarchised according to their decorrelation time. This note suggests that when the timescales separate, one could potentially remove the linear drift being left with a fully Markovian equation, reminiscing of classical homogenisation results [PS08]. Furthermore, the hierarchised equation can posses a gap in timescales within the hidden variables so that one could

afford a reduction of the dimensions of system (4.33); we refer back to Remark (4.2.4). This practical role of the λ_j's in Theorem 4.2.1 deserves, therewith, a more careful examination in further work.

The response-theory-based WL equation has been employed as top-down parametrisation of conceptual models [Wou+16]; [VL18b], but also as subgrid processes in atmosphere-ocean coupled systems [VL18a]; [DV17]. If the full equations are not known, one is left with data-driven methodologies that provide candidate models to emulate partial observations. To this end, the EMR methodology [KKG05] reviewed here in Section 4.4 is seen as the bottom-up counterpart to the WL parametrisation since they are both written in an MSM manner; this is illustrated in Fig. 4.1. It was suggested at the end of Section 4.4.2 that the number of levels needed for EMR to converge was related with the spectrum underlying Koopman operator. Effectively, if the generator eigenvalues are located close to the imaginary axis, one expects EMR to converge at a high number of levels, whereas if the eigenvalues have very negative real parts, the parametrised signal would be almost white making EMR converge instantly. More broadly, the theoretical results of Chapter 4 establish how the spectral features of semigroups determine in turn the constitutive elements of data-driven non-Markovian closure of partially observed complex systems, when rewritten as an MSM. These results highlights, in particular, new bridges with Koopman modes and the DMD decomposition [Mez05]; [Sch10], as well with other kinds of projections onto spectral bases [CK17]; [Ghi+02]; [Has88]; [Pen96].

Appendix A

Homogeneous Equation for the Linear Response

This appendix aims at checking that, indeed, the Kubo response formula (describing the response function to first order in the perturbation parameter) is recovered when gathering the first order terms in the integro-differential equation (2.107) as done in [Ken71]. Noting that $\mathcal{L}_0\rho_0 \equiv 0$, we can further simplify $B(t)$ and $K(t,s)$:

$$B(t) = \xi^{-1}\int \Psi(\mathbf{x})\mathcal{L}(t)\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} = \xi^{-1}\int \Psi(\mathbf{x})\left(\mathcal{L}_0 + \varepsilon g(t)\mathcal{L}_1\right)\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} \tag{A.1a}$$

$$= \varepsilon g(t)\xi^{-1}\int \Psi(\mathbf{x})\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} = \varepsilon g(t)q, \tag{A.1b}$$

where we have introduced $q := \xi^{-1}\int \Psi(\mathbf{x})\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}$. From this calculation, we immediately observe that $B(t)$ is of order ε. From Eq. (2.108b) and the perturbed Fokker-Planck equation (2.16), the kernel K reads as:

$$K(t,s) = \xi^{-1}\int \Psi(\mathbf{x})\mathcal{L}(t)G(t,s)\left(1-\mathbb{P}\right)\mathcal{L}_0\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} \tag{A.2a}$$

$$+ \varepsilon g(s)\xi^{-1}\int \Psi(\mathbf{x})\mathcal{L}(t)G(t,s)\left(1-\mathbb{P}\right)\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} \tag{A.2b}$$

$$= \varepsilon g(s)\xi^{-1}\int \Psi(\mathbf{x})\mathcal{L}(t)G(t,s)\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} \tag{A.2c}$$

$$- \varepsilon g(s)q\xi^{-1}\int \Psi(\mathbf{x})\mathcal{L}(t)G(t,s)\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{A.2d}$$

where we have used that $\mathcal{L}_0\rho_0 \equiv 0$ and $\mathbb{P}\mathcal{L}_1\rho_0 = q\rho_0$. As with $B(t)$, this calculation shows that the minimum order for $K(t,s)$ is ε.

The next step is to find the equations for the response at the zeroth and first orders of powers of ε corresponding to the expansion:

$$R_\Psi(t) = R_\Psi^{(0)}(t) + \varepsilon R_\Psi^{(1)}(t) + \varepsilon^2 R_\Psi^{(2)}(t) + \dots \tag{A.3}$$

This expansion is the same as Eq. (2.56), although we denote it differently in accordance with Section 2.4. The reason is that the method presented here is different to the approach involving the Green function (see Eq. (2.60)); the target of this subsection is to show that these two approaches coincide in their leading order term. The equation for $R_\Psi^{(0)}(t)$ reads as:

$$\partial_t R_\Psi^{(0)}(t) = 0, \tag{A.4}$$

from where we deduce that the zeroth order response is constant equal to $R_\Psi^{(0)}(0) = \xi$; see Eq. (2.98) for the definition of ξ. We now extract the first order terms in $K(t,s)$, that we denote as $K^{(1)}(t,s)$:

$$K^{(1)}(t,s) = \varepsilon g(s)\xi^{-1} \int \Psi(\mathbf{x})\mathcal{L}_0 G_0(t,s)\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} \tag{A.5a}$$

$$- \varepsilon g(s) q \xi^{-1} \int \Psi(\mathbf{x})\mathcal{L}_0 G_0(t,s)\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{A.5b}$$

which was obtained by replacing the time-dependent perturbed operator $\mathcal{L}(t)$ by the unforced analogue $\mathcal{L}_0$. We have also introduced the operator $G_0(t,s)$ which is defined as:

$$G_0(t,s) = G_0(t-s) = e^{(t-s)(1-\mathbb{P})\mathcal{L}_0}, \tag{A.6}$$

and corresponds to the zeroth order elements in the expansion of $G(t,s)$. Hence, we have that:

$$G_0(t-s)\rho_0 = e^{(t-s)(1-\mathbb{P})\mathcal{L}_0}\rho_0 = \rho_0. \tag{A.7}$$

Consequently, Eq. (A.5b) vanishes.

Thus, the equation for the linear term $R_\Psi^{(1)}(t)$ is:

$$\partial_t R_\Psi^{(1)}(t) = B(t) + \int_0^t K^{(1)}(t,s)\mathrm{d}s. \tag{A.8}$$

This is the non-Markovian equation that describes the time evolution of the linear response as results of extracting the leading order terms in Eq. (2.107). The objective, then, is to prove that Eq. (A.8) is the same as the Kubo formula. This entails proving a few operator identities necessary to manipulate such equation. First we note that for any function f:

$$\mathcal{L}_0 G_0(t,s) f = e^{(t-s)\mathcal{L}_0(1-\mathbb{P})}\mathcal{L}_0 f. \tag{A.9}$$

Furthermore, knowing that $\mathcal{L}_0\mathbb{P}f \equiv 0$ for any function f,

$$\mathcal{L}_0 G_0(t,s) f = e^{(t-s)\mathcal{L}_0(1-\mathbb{P})}\mathcal{L}_0 f = e^{(t-s)\mathcal{L}_0}\mathcal{L}_0 f = \partial_t e^{(t-s)\mathcal{L}_0} f. \tag{A.10}$$

Using these identities, we find a simpler expression for $K^{(1)}(t,s)$:

$$K^{(1)}(t,s) = \varepsilon g(s)\xi^{-1}\int \Psi(\mathbf{x})\mathcal{L}_0 G_0(t,s)\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x} \tag{A.11a}$$

$$= \varepsilon g(s)\xi^{-1}\int \Psi(\mathbf{x})\partial_t e^{(t-s)\mathcal{L}_0}\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{A.11b}$$

The final step is to use Leibniz integration rule in Eq. (A.8) to realise that there exists a function $\mathcal{G}$ such that:

$$R^{(1)}_\Psi(t) = \int_0^t \mathcal{G}(t-s)g(s)\mathrm{d}s, \tag{A.12}$$

and such function is given by:

$$\mathcal{G}(t-s) = \int \Psi(\mathbf{x})e^{(t-s)\mathcal{L}_0}\mathcal{L}_1\rho_0(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{A.13}$$

which is the Green function that gives the linear response (2.58). We note, though, that the procedure for obtaining this linear response was departing from the exact formula of the full response given in Eq. (2.107), which did not resort to linearising the Fokker-Planck equation at the beginning.

Appendix B

Atomic Perturbations of Markov Chains

A particularly interesting class of perturbations of Markov chains only affect the transition probabilities of a single state. This can confidently be considered the simplest perturbation of a stochastic matrix and allows to derive perturbative bounds explicitly without resorting to norm estimates.

Let N denote a positive integer and consider an $N \times N$ stochastic matrix $\mathcal{M}$. If the probability of state k is altered, it means that part of the transition probability to state i increases in detriment of that of state j. This class of *atomic* perturbations constitutes a family of matrices $\mathfrak{A}_N$:

$$\mathfrak{A}_N = \left\{ m \in \mathbb{R}^{N\times N} : m = (\mathbf{e}_i - \mathbf{e}_j)\mathbf{e}_k^\top, 1 \leq i, j, k \leq N \right\}. \tag{B.1}$$

Not all elements in $\mathfrak{A}_N$ are admissible perturbations for the stochastic matrix $\mathcal{M}$. Indeed, let $m = (\mathbf{e}_i - \mathbf{e}_j)\mathbf{e}_k^\top$ where i, j and k are so that $\mathcal{M}_{j,k} = 0$ and $\mathcal{M}_{i,k} = 0$. Then, for every real value of ε, $\mathcal{M}_{i,k} + \varepsilon m_{i,k} < 0$ or $\mathcal{M}_{j,k} + \varepsilon m_{j,k} < 0$. This means that $\mathcal{M} + \varepsilon m$ has negative entries for all ε in $\mathbb{R}$ and it cannot constitute a stochastic matrix; see Definition 3.1.1. This motivates the following definition:

$$\mathfrak{A}_N(\mathcal{M}) = \{ m \in \mathfrak{A}_N : \mathcal{M} + m \text{ is stochastic} \}. \tag{B.2}$$

In this setting, let εm be in $\mathfrak{A}_N(\mathcal{M})$ for some ε in $\mathbb{R}$. Then, if $\mathcal{Q}$ is the spectral projector around 1 (see Eq. (3.25)) and $\mathbf{u} = [u_1, \ldots, u_N]^\top$ is the stationary vector of $\mathcal{M}$, there exist i, j and k in $\{1, \ldots, N\}$ such that the linear response is given by:

$$\mathcal{G}\mathbf{u} = (1 - \mathcal{M} + \mathcal{Q})^{-1} m\mathbf{u} = (1 - \mathcal{M} + \mathcal{Q})^{-1} (\mathbf{e}_i - \mathbf{e}_j) u_k \tag{B.3a}$$

$$= u_k \left[(1 - \mathcal{M} + \mathcal{Q})^{-1} \right]_{:,i} - u_k \left[(1 - \mathcal{M} + \mathcal{Q})^{-1} \right]_{:,j}. \tag{B.3b}$$

Equation (B.3b) is in agreement with what was found in [MS88] and this form of the linear response allows to derive explicit bounds on the norm of $\mathcal{G}\mathbf{u}$. This, furthermore,

determines the radius of expansion of the response formula in Eq. (3.16). Indeed, one has

$$\|\mathcal{G}\mathbf{u}\|_1 = u_k \left[(1-\mathcal{M}+\mathcal{Q})^{-1}\right]_{:,i} - u_k \left[(1-\mathcal{M}+\mathcal{Q})^{-1}\right]_{:,j} \leq \frac{u_k}{1-\|\mathcal{M}\|_{1^*}}. \tag{B.4}$$

From where one deduces that the invariant vector $\mathbf{u}$ is more robust to atomic perturbations at state k if the invariant probability there is small. The radius of expansion ε_{max} for the perturbative formulas, on the other hand, becomes independent of u_k by means of the ratio test. Summarising the proof in Eq. (3.30), we take the ratio of two successive elements in the series:

$$\frac{\|\mathcal{G}\mathcal{G}^n\mathbf{u}\|_1}{\|\mathcal{G}^n\mathbf{u}\|_1} \leq \|\mathcal{G}\|_1 \leq \frac{\|m\|_1}{1-\|\mathcal{M}\|_{1^*}} = \frac{2}{1-\|\mathcal{M}\|_{1^*}}, \tag{B.5}$$

which yields the following bound on the radius of expansion

$$\varepsilon_{max} = \frac{1-\|\mathcal{M}\|_{1^*}}{2}. \tag{B.6}$$

While the leading eigenvalue of the perturbed Markov chain remains unchanged, the rest of the spectrum $\{\lambda_l(\varepsilon)\}_{l=1}^N$ is susceptible of being altered as a function of ε. Here, we are assuming that the spectrum is totally simple for $\varepsilon = 0$ and is sorted according to: $\lambda_1 = 1 \geq |\lambda_2| \geq \ldots \geq |\lambda_N|$. To study the sensitivity of the eigenvalues, we calculate the leading order change in λ_j as results of perturbing $\mathcal{M}$ by εm, for $i = 2, \ldots, N$. By invoking the continuity of the eigenvalues with respect to matrix entries and their simplicity [Wil65]; [TE05], we can perform Taylor series around $\varepsilon = 0$ so that

$$\lambda_l(\varepsilon) = \lambda_2 + \varepsilon \frac{\mathrm{d}\lambda_l(\varepsilon)}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} + \mathcal{O}\left(\varepsilon^2\right). \tag{B.7}$$

Let $\mathbf{v}_l = [v_{l,1}, \ldots, v_{l,N}]^\top$ and $\mathbf{u}_l = [u_{l,1}, \ldots, u_{l,N}]^\top$ be the left and right eigenvectors of $\mathcal{M}$ associated with λ_l, respectively, and assume that $\|\mathbf{v}_l\|_2 = \|\mathbf{u}_l\|_2 = 1$, then, if εm is in $\mathfrak{A}_N(\mathcal{M})$ and using standard arguments, there exist i, j and k in $\{1, \ldots, N\}$:

$$\frac{\mathrm{d}\lambda_l(\varepsilon)}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} = \frac{1}{|\mathbf{v}_l^*\mathbf{u}_l|}\mathbf{v}_l^*\left(\mathbf{e}_i - \mathbf{e}_j\right)\mathbf{e}_k^\top\mathbf{u}_l = \frac{1}{|\mathbf{v}_l^*\mathbf{u}_l|}\mathbf{v}_l^*\left(\mathbf{e}_i - \mathbf{e}_j\right)u_{l,k} \tag{B.8a}$$

$$= \frac{1}{|\mathbf{v}_l^*\mathbf{u}_l|}\left(\overline{v_{l,i}} - \overline{v_{l,j}}\right)u_{l,k} = \kappa_2\left(\lambda_l\right)\left(\overline{v_{l,i}} - \overline{v_{l,j}}\right)u_{l,k}, \tag{B.8b}$$

where $k_2(\lambda_l)$ is the eigenvalue condition number [Wil65]; [TE05]. This formula gives a complex number whose absolute value estimates of the rate of change of λ_l as a function of ε. This calculation reveals that the local quantity $\left(\overline{v_{l,i}} - \overline{v_{l,j}}\right)u_{l,k}$ controls the robustness of the eigenvalue λ_l provided that the global quantity $\kappa_2\left(\lambda_l\right)$ stays finite.

The mixing time of the stochastic matrix $\mathcal{M}$ is indicated by the modulus of the second largest eigenvalue, whose sensitivity to atomic perturbations can be estimated by applying

Eq. (B.8b) to λ_2. However, since the modulus can be cumbersome to manipulate, it is hard to get the exact derivative formula for $|\lambda_2|$. It is, therefore, convenient to consider the real part of the logarithm of λ_2, which indicates the rate of convergence. Indeed, we have:

$$\frac{\mathrm{d}\mathfrak{Re}\left(\log(\lambda_2(\varepsilon))\right)}{\mathrm{d}\varepsilon} = \mathfrak{Re}\left(\frac{\mathrm{d}\log(\lambda_2(\varepsilon))}{\mathrm{d}\varepsilon}\right) = \mathfrak{Re}\left(\frac{1}{\lambda_2(\varepsilon)}\frac{\mathrm{d}\lambda_2(\varepsilon)}{\mathrm{d}\varepsilon}\right). \tag{B.9}$$

Now, inserting the value of the sensitivity derivative in Eq. (B.8b) into Eq. (B.9) and evaluating at $\varepsilon = 0$, we get

$$\mathfrak{Re}\left(\frac{1}{\lambda_2(\varepsilon)}\frac{\mathrm{d}\lambda_2(\varepsilon)}{\mathrm{d}\varepsilon}\right)\Big|_{\varepsilon=0} = \frac{\kappa_2(\lambda_2)}{|\lambda_2|^2}\mathfrak{Re}\left(\overline{\lambda_2}\left(\overline{v_{2,i}} - \overline{v_{2,j}}\right)u_{2,k}\right). \tag{B.10}$$

for some i, j and k in $\{1, \ldots, N\}$. We do not expand this expression further, although a generic form in terms of arbitrary perturbation matrices m can be found in [ADF18].

Appendix C

Itô Integration of the MSM

In the main text, we proposed a solution of the MSM given by Eq. (4.71) using the Dyson expansion for the linear operators involved in the backward-Kolmogorov equation. advection acting on functions. Therefore, we substituted non-linear ordinary differential equations for a partial differential equation, for the sake of having linear operators in hand. The same solution can be attained by direct integration of the MSM in the form (4.71). We convolute, in the Itô sense, Eq. (4.71b) to find an explicit solution for $\mathbf{r}(t)$ when $d_1 = d_2$:

$$\mathbf{r}(t) = e^{-\mathrm{D}t}\mathbf{r}(0) + \int_0^t e^{-\mathrm{D}(t-s)}\Sigma \mathrm{d}W_s + \varepsilon \int_0^t e^{-\mathrm{D}(t-s)}\mathrm{C}\mathbf{x}(s)\mathrm{d}s; \tag{C.1}$$

here, $\mathbf{r}(0)$ indicates an initial state that can be assumed to be distributed in a prescribed way. For the more general, non-linear MSMs considered there, see [KCG15, Proposition 3.3].

By substituting the expression (C.1) into Eq. (4.71a), we find an exact expression for the evolution of $\mathbf{x}(t)$:

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x}(t)) + \varepsilon e^{-\mathrm{D}t}\mathbf{r}(0) + \varepsilon \int_0^t e^{-\mathrm{D}(t-s)}\Sigma \mathrm{d}W_s + \varepsilon^2 \int_0^t e^{-\mathrm{D}(t-s)}\mathrm{C}\mathbf{x}(s)\mathrm{d}s, \tag{C.2}$$

in which the memory effects in the fourth term are of second order in ε. Note that the ε-order terms arise from a noise realisation in the decoupled regime, whereas the memory term is exclusively due to the coupling of the main variables with the hidden ones. Hence, the only degree of freedom left is the distribution of the initial states $\mathbf{r}(0)$.

Appendix D

The Coupled Lorenz 84–63 System

The EMR methodology's ability to capture the statistics of low-dimensional dynamical systems was illustrated in [KKG05], where the authors considered the Lorenz 63 system [Lor63] as a test case in which the phase space can be fully sampled. Moreover, provided that the integration time step is short enough, the parameters of the underlying model can be fully captured with a high degree of confidence. Here, we repeat the analysis of [KKG05] to illustrate the effectiveness of EMR in capturing statistical and dynamical properties in an extended, partially observed deterministic system.

The model we consider is the result of coupling the $\mathbf{X} = (X, Y, Z)$ variables of the Lorenz 84 system [Lor84] with the $\mathbf{x} = (x, y, z)$ variables of the Lorenz 63 system [Lor63], namely:

$$\dot{X} = -Y^2 - Z^2 - aX + a(F_0 + hx), \tag{D.1a}$$
$$\dot{Y} = XY - bXZ - Y + G, \tag{D.1b}$$
$$\dot{Z} = XZ + bXY - Z, \tag{D.1c}$$
$$\dot{x} = \tau s(y - x), \tag{D.1d}$$
$$\dot{y} = \tau(rx - y - xz), \tag{D.1e}$$
$$\dot{z} = \tau(xy - \beta z); \tag{D.1f}$$

the parameter values are: $a = 0.25, b = 4$, $F_0 = 8, G = 1$, and $s = 10$, $r = 28, \beta = 8/3$, respectively. The parameter h measures the strength of the coupling, while τ scales the rate of change in the Lorenz 63 system and, therefore, the timescale ratio between the two subsystems.

This system is a skew-product, in the sense of [Sel71], since the coupling is one-way only, with the Lorenz 63 system driving the Lorenz 84 dynamics. Hence, one has— as noted in [VL18b]— a fully Markovian WL parametrisation of the Lorenz 63 variables. Furthermore, the correlation function that defines the stochastic noise $\eta(t)$ can be further

expanded and simplified,with respect to Eq. (4.20). One can, in fact, write explicitly:

$$C(\eta(0), \eta(t)) = \langle (x(0), 0, 0) \cdot (x(t), 0, 0) \rangle \,, \tag{D.2}$$

where the angular brackets $\langle \cdot \rangle$ indicate averages with respect to the physical measure associated with the Lorenz 63 system. Since Lorenz 84 does not feed back into Lorenz 63, the evolution of $\mathbf{x}(t)$ only obeys the dynamics of Lorenz 63, and thus the decorrelation of the noise scales with τ. Notice that this trivially follows from Theorem 4.2.1.

In most of the numerical experiments, the timescale separation between the two systems is $\tau = 5$ unless otherwise stated. The relevance of this timescale parameter was investigated in previous work [VL18b]. Here, we focus as well on the effects of the coupling strength h, and we shall study the cases of $h = 0.25$ and 0.025. Partial observations only will be used in these experiments, by sampling the three-dimensional outputs of the Lorenz 84 system. Then, the observed tendencies are regressed and sequentially layered following the EMR approach, as explained in Section 4.4.

D.1 EMR Outputs

The Lorenz 84–63 model is integrated for 730 time units that correspond in Lorenz 84 to 10 natural years, with a time step of $5 \cdot 10^{-3}$ time units; two separate runs are made for the coupling strengths $h = 0.25$ and $h = 0.025$. These two full-model runs are used to train the corresponding EMR model versions, both of which use only the slow $\mathbf{X}$-variables and eliminate the fast $\mathbf{x}$-variables. Then, two separate full-model simulations are run, for testing purposes, over 7 300 time units, and the EMR model's output is compared with it, for the two parameter values. Below we show the main statistical outputs of the EMR methodology compared to the two reference integrations of the full model. The results for the two separate h-values are shown in Figs. D.1–D.3 and Figs. D.4–D.6, respectively.

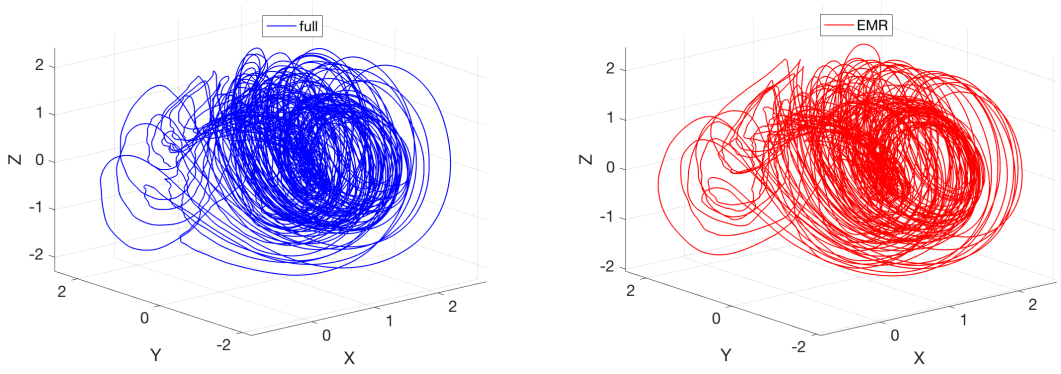
The region of phase space explored by the EMR model clearly coincides with the one visited by the full model, as seen in Figs. D.1 and D.4, and the relative occupancies within this region—as indicated by the smoothed PDFs shown in Figs. D.2 and D.5, respectively—agree very well. The timescales are also well captured, as indicated by the good approximation of the autocorrelation functions, cf. Figs. D.3 and D.6.

Notice that, while the original Lorenz 84–63 system is purely deterministic, the EMR model includes white noise acting on the hidden layers of the learned model. This fact could suggest that a smoothing of the invariant measure is inevitable and that the EMR methodology may not be able to capture fractal geometries in phase space, since the EMR model would not satisfy the Hörmander's hypoellipticity condition [CSG11]; [Hör67]. The numerical evidence in Figs. D.1 and D.4, however, illustrates a strikingly good approximation of the full model's attractor, including its very fine, and presumably

fractal, structure.

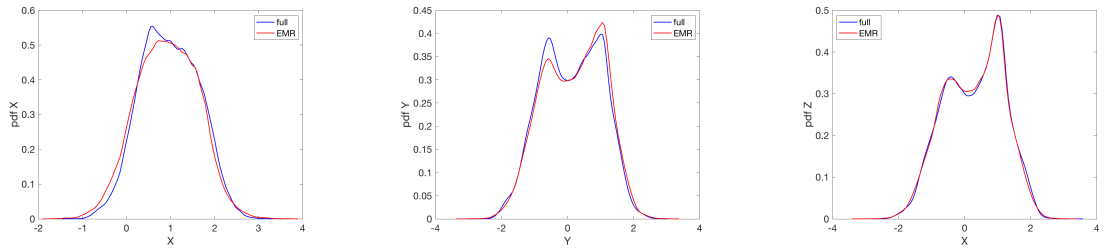
Actually, [KCG15, Theorem 3.1 and Corollary 3.2] provided sufficient conditions for the existence of a random attractor for a broad class of MSMs that are not subject to a non-degeneracy condition of Hörmander type. In other words, one can have an MSM that possesses a random attractor and is thus dynamically quite stable, while exhibiting in a forward sense an invariant measure of the associated Fokker-Planck equation that is singular with respect to the Lebesgue measure. This mathematically rigorous result helps explain what is observed numerically not only in the present paper for the EMR of the Lorenz 84–63 model, but also in the case of the EMR model of the Lotka-Volterra example in [KCG15, Fig. 7].

Ulam's method was used on the projection of the full $(\mathbf{X}, \mathbf{x})$ phase space onto the $\mathbf{X}$ subspace to approximate the spectrum of the Koopman operator, since it can provide further information on the characteristics of the time series, beyond PDFs and correlation functions. In particular, we learned transition matrices $\mathcal{M}^\tau$ according to Eq. (3.48) with transition times of $\tau = \mathrm{d}t$. The observed spectra (red $\times$ symbols) using a coarse partition of phase space into 512 non-intersecting boxes showed good agreement with the spectra based on the full model (blue open circles); see Fig. D.7. This agreement confirms further that, at this level of coarse graining, the EMR model captures well the the characteristics of the full model's solutions.



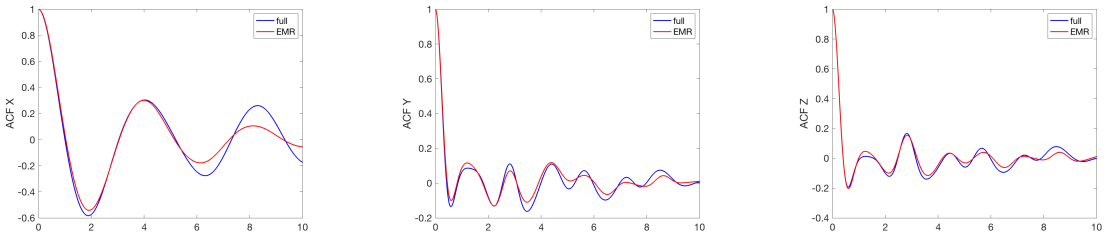
(a) Full model (b) EMR model

Figure D.1: Trajectories of the Lorenz 84–63 model in the three-dimensional (X, Y, Z) phase space of the Lorenz 84 model, for $h = 0.25$ and 200 time units: (a) for the full Lorenz 84–63 model governed by Eqs. (D.1) (blue); (b) for the EMR model (red).



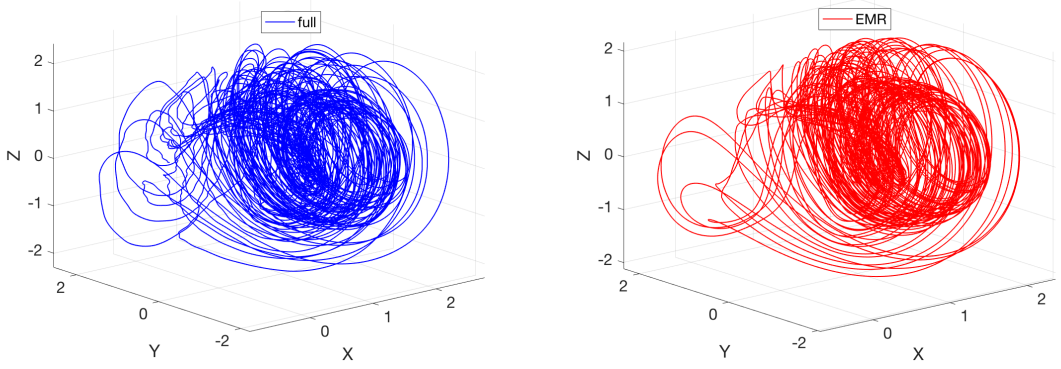
(a) X-PDF (b) Y-PDF (c) Z-PDF

Figure D.2: Smoothed PDFs of the Lorenz 84–63 variables (a) X, (b) Y, and (c) Z with a coupling strength of $h = 0.25$. The blue curve corresponds to the full model; the red curve corresponds to the EMR model. These PDFs and those in Fig. D.5 were obtained by using the MATLAB R2019a kernel smoothing function *ksdensity*.



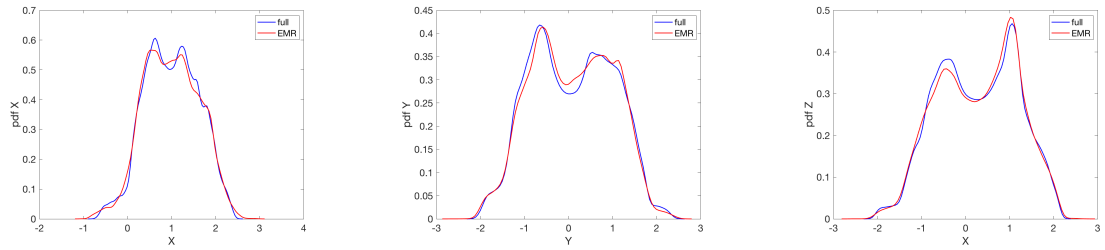
(a) X-ACF (b) Y-ACF (c) Z-ACF

Figure D.3: Autocorrelation functions (ACFs) of the Lorenz 84–63 variables (a) X, (b) Y, and (c) Z for a coupling strength of $h = 0.25$.



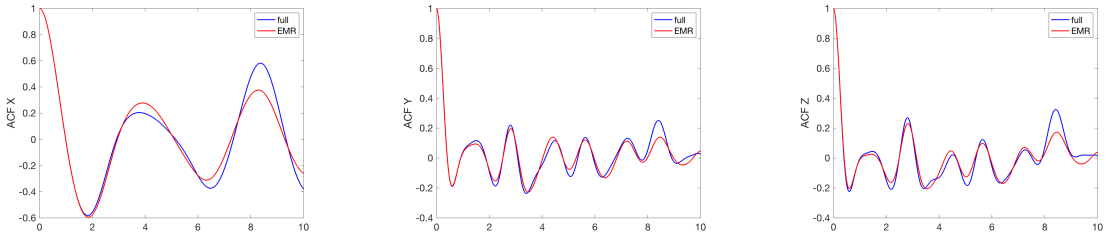
(a) Full model (b) EMR model

Figure D.4: Example trajectories of the Lorenz 84- 63 model on the (X, Y, Z) domain with a coupling strength of $h = 0.025$ integrated for 200 time units. Subfigure (a) corresponds to the full model (D.1) (blue) and subfigure (b) refers to the EMR model (red).



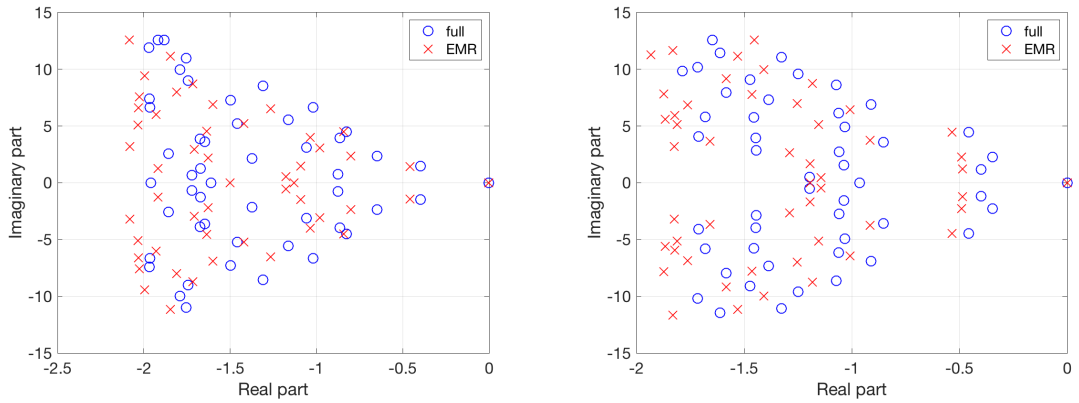
(a) X-PDF (b) Y-PDF (c) Z-PDF

Figure D.5: Smoothed PDFs of the Lorenz 84- 63 variables (a) X, (b) Y, and (c) Z, with a coupling strength of $h = 0.025$. The blue curve corresponds to the full model; the red curve corresponds to the EMR model.



(a) X-ACF (b) Y-ACF (c) Z-ACF

Figure D.6: ACFs of the Lorenz 84- 63 variables (a) X, (b) Y, and (c) Z for $h = 0.025$. The blue curve corresponds to the full model; the red curve corresponds to the EMR model.



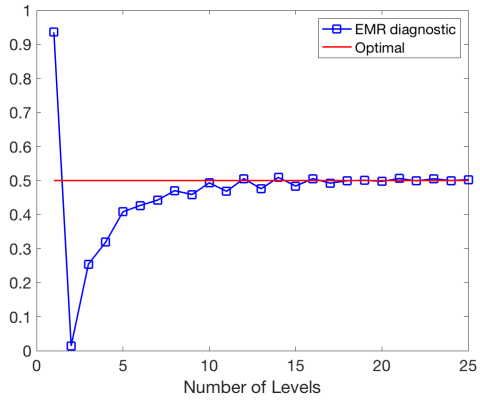
(a) $h = 0.25$ (b) $h = 0.025$

Figure D.7: Leading eigenvalues of the discretized Koopman operator in the L84 model's phase space. The blue open circles correspond to the data obtained by integrating the full model's Eqs. (D.1), the red $\times$ symbols correspond to the EMR model. (a) $h = 0.25$; and (b) $h = 0.025$.

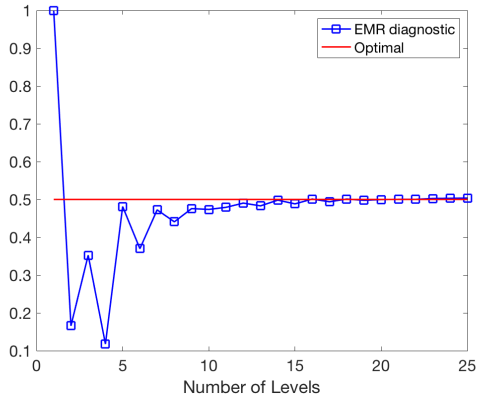
D.2 Convergence

Convergence in the EMR approach is determined by the "whiteness" of the last-level residual, as explained in Section 4.4.2; see Eq. (4.86) and discussion thereof. In Fig. D.8, we plotted the mean of the determination coefficients R^2 for the three $\mathbf{X}$-variables and we show that its convergence in the EMR approach depends only mildly on the coupling parameter h. Indeed, for $h = 0.25$ we observe in panel (a) that around 18 levels are necessary before achieving the optimal level, whereas for weaker coupling with $h = 0.025$ convergence is attained in panel (b) already with 15 levels, as one might expect.

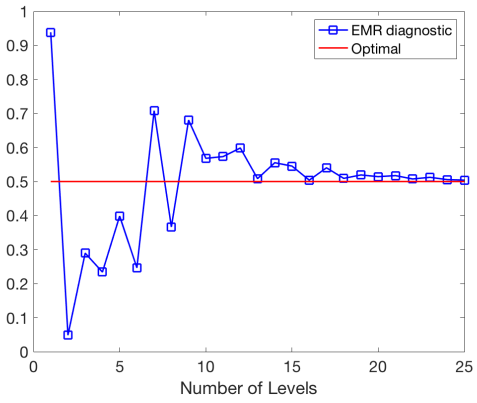
Furthermore, as already pointed out in [KCG15] on a different example, the results in Fig. D.8 (c) illustrate that a smaller timescale separation τ can require a higher number of levels for EMR to attain convergence: in the case at hand, around 25 levels are needed. For completeness, Fig. D.8 (d) shows that including additive white noise in the Lorenz 63 system can, in fact, accelerate the convergence of the method, with convergence achieved at $\ell = 7$.



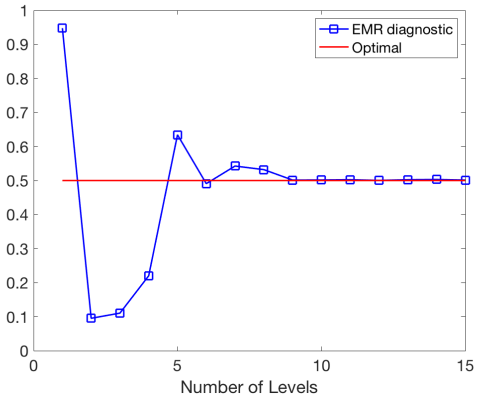
(a) $h = 0.25$ and $\tau = 5$



(b) $h = 0.025$ and $\tau = 5$



(c) $h = 0.25$ and $\tau = 2$

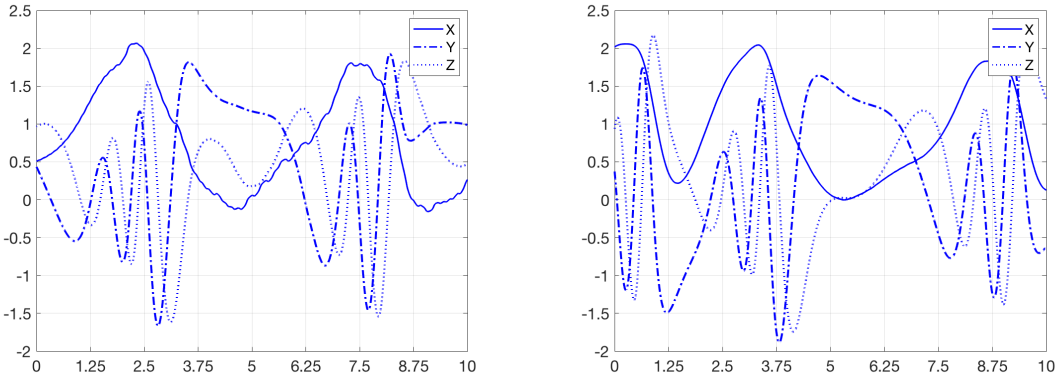


(d) Noisy Lorenz 63 model with $h = 0.25$ and $\tau = 5$

Figure D.8: Determination coefficients R^2 of the EMR method as a function of the number ℓ of levels. (a) $h = 0.25$; (b) $h = 0.025$; and (d) $h = 0.25$ but with the Lorenz 63 model including additive noise. Panels (a,b,d) all have the timescale separation $\tau = 5$, while in panel (c) $h = 0.25$ and $\tau = 2$.

D.3 Model Coefficients

We show here that the EMR model coefficients can be efficiently approximated when phase space subsampling is carried out. Here, regressions are performed over fifty short time series of 10 time units each, with a time step of $5 \cdot 10^{-3}$, as in Section D.1. The reason for taking this sample length here is that 10 time units is visually enough for the slow variable $\mathbf{X}$ to go through a cycle, as illustrated in Fig. D.9, for both $h = 0.25$ and 0.025.



(a) $h = 0.25$ (b) $h = 0.025$

Figure D.9: Time-series of the Lorenz 84 variables (X, Y, Z) over 10 time units: (a) $h = 0.25$, and (b) 0.025.

The estimated coefficients and their standard deviations using the EMR regressions are listed below in Tables D.1 and D.2 for $h = 0.25$ and in Tables D.3 and D.4 for $h = 0.025$. The tables show the coefficients of the linear and quadratic forms at first level in the EMR regressions: see Eq. (4.84a).

As expected, a stronger coupling of $h = 0.25$ leads to greater uncertainty in the estimation, as indicated by the corresponding standard errors. For the fairly complex and chaotic system at hand, we note that no memory effects are artificially introduced in the regressions at the second level. Indeed, we found that the coupling of the main level with the subsequent ones was 0 to the fourth decimal place.

EMR	1	x	y	z	x^2	xy	y^2	xz	yz	z^2
f_X	1.949	-0.352	-0.002	-0.104	-0.015	0.01	0.052	-0.946	-0.001	-0.921
f_Y	0.999	0.001	-1.001	0.002	0	1.001	-4.003	0	0	0
f_Z	0.002	-0.003	-0.002	-1.001	0.001	4.003	1.001	0	0	0

Table D.1: Means of the EMR coefficients of the Lorenz 84–63 model, estimated from an ensemble of 50 runs over 10 time units for $h = 0.25$.

EMR	1	x	y	z	x^2	xy	y^2	xz	yz	z^2
f_X	0.543	1.005	0.246	0.291	0.428	0.152	0.188	0.104	0.076	0.099
f_Y	0.001	0.002	0	0.001	0.001	0	0	0	0	0
f_Z	0.001	0.002	0.001	0.001	0.001	0.001	0	0	0	0

Table D.2: Standard deviations of the EMR coefficients of the Lorenz 84–63 model, estimated from an ensemble of 50 runs over 10 time units for $h = 0.25$.

EMR	1	x	y	z	x^2	xy	y^2	xz	yz	z^2
f_X	2.006	-0.253	-0.001	-0.003	-0.004	0	0.001	-1.002	0.001	-1.003
f_Y	1	0	-1.001	0.002	0	1.001	-4.003	0	0	0
f_Z	0.001	-0.002	-0.002	-1.001	0.001	4.003	1.001	0	0	0

Table D.3: Means of the EMR coefficients of the Lorenz 84–63 model, estimated from an ensemble of 50 runs over 10 time units for $h = 0.025$.

EMR	1	x	y	z	x^2	xy	y^2	xz	yz	z^2
f_X	0.034	0.063	0.019	0.021	0.026	0.01	0.014	0.007	0.008	0.006
f_Y	0.001	0.002	0	0.001	0.001	0	0	0	0	0
f_Z	0.001	0.002	0.001	0.001	0.001	0.001	0	0	0	0

Table D.4: Standard deviations of the EMR coefficients of the Lorenz 84–63 model, estimated from an ensemble of 50 runs over 10 time units for $h = 0.025$.

Bibliography

[Abr17] R. V. Abramov. “Leading Order Response of Statistical Averages of a Dynamical System to Small Stochastic Perturbations”. In: *Journal of Statistical Physics* 166.6 (2017), pp. 1483–1508.

[ADF18] F. Antown et al. “Optimal Linear Responses for Markov Chains and Stochastically Perturbed Dynamical Systems”. In: *Journal of Statistical Physics* 170.6 (2018), pp. 1051–1087.

[AG57] T. W. Anderson and L. A. Goodman. “Statistical Inference about Markov Chains”. In: *The Annals of Mathematical Statistics* 28.1 (1957), pp. 89–110.

[Arn98] L. Arnold. *Random Dynamical Systems*. New York: Springer, 1998.

[AS74] A. Arakawa and W. H. Schubert. “Interaction of a cumulus cloud ensemble with the large-scale environment, Part I”. In: *Journal of the Atmospheric Sciences* 31.3 (1974), pp. 674–701.

[Ash+12] P. Ashwin et al. “Tipping points in open systems: Bifurcation, noise-induced and rate-dependent examples in the climate system”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 370.1962 (2012), pp. 1166–1184.

[Bal00] V. Baladi. *Positive Transfer Operators and Decay of Correlations*. Singapore: World Scientific, 2000.

[Bal14] V. Baladi. “Linear response, or else”. In: *Proceedings of the International Congress of Mathematicians-Seoul* 3 (2014), pp. 525–545.

[BF60] F. L. Bauer and C. T. Fike. “Norms and Exclusion Theorems”. In: *Numerische Mathematik* 2.1 (1960), pp. 137–141.

[BK98] M. Blank and G. Keller. “Random perturbations of chaotic dynamical systems: Stability of the spectrum”. In: *Nonlinearity* 11.5 (1998), pp. 1351–1364.

[BKL17] V. Baladi et al. “Linear and fractional response for the SRB measure of smooth hyperbolic attractors and discontinuous observables”. In: *Nonlinearity* 30.3 (2017), pp. 1204–1220.

[BL07] O. Butterley and C. Liverani. “Smooth anosov flows: Correlation spectra and stability”. In: *Journal of Modern Dynamics* 1.2 (2007), pp. 301–322.

[Bof+03] G. Boffetta et al. “Relaxation of finite perturbations: Beyond the fluctuation-response relation”. In: *Chaos* 13.3 (2003), pp. 806–811.

[BPK16] S. L. Brunton et al. “Discovering governing equations from data by sparse identification of nonlinear dynamical systems”. In: *Proceedings of the National Academy of Sciences* 113.15 (2016), pp. 3932–3937.

[Bud69] M. I. Budyko. “The effect of solar radiation variations on the climate of the Earth”. In: *Tellus* 21 (1969), pp. 611–619.

[CGH12] M. D. Chekroun and N. E. Glatt-Holtz. “Invariant measures for dissipative dynamical systems: abstract results and applications”. In: *Commun. Math. Phys.* 316.3 (2012), pp. 723–761.

[Che+11] M. D. Chekroun et al. “Asymptotics of the Coleman-Gurtin model”. In: *Discrete Contin. Dyn. Syst. Ser. S* 4.2 (2011), pp. 351–369.

[Che+14] M. D. Chekroun et al. “Rough parameter dependence in climate models and the role of ruelle-pollicot resonances”. In: *Proceedings of the National Academy of Sciences of the United States of America* 111.5 (2014), pp. 1684–1690.

[Che+16] M. D. Chekroun et al. “Low-dimensional Galerkin approximations of nonlinear delay differential equations”. In: *Discrete and Continuous Dynamical Systems* 36.8 (2016), pp. 4133–4177.

[CHK02] A. J. Chorin et al. “Optimal prediction with memory”. In: *Physica D: Nonlinear Phenomena* 166.3-4 (2002), pp. 239–257.

[CK17] M. D. Chekroun and D. Kondrashov. “Data-adaptive harmonic spectra and multilayer Stuart-Landau models”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 27.9 (2017), p. 93110.

[CKL20] M. D. Chekroun et al. “Efficient reduction for diagnosing Hopf bifurcation in delay differential systems: Applications to cloud-rain models”. In: *Chaos* 40.8 (2020), p. 53130.

[CL15] A. J. Chorin and F. Lu. “Discrete approach to stochastic parametrization and dimension reduction in nonlinear dynamics”. In: *Proceedings of the National Academy of Sciences* 112.32 (2015), pp. 9804–9809.

[CLM17] M. D. Chekroun et al. “The emergence of fast oscillations in a reduced primitive equation model and its implications for closure theories”. In: *Computers & Fluids* 151 (2017), pp. 3–22.

[CLM19] M. D. Chekroun et al. “Variational Approach to Closure of Nonlinear Dynamical Systems: Autonomous Case”. In: *Journal of Statistical Physics* 179 (2019), pp. 1073–1160.

[CLW15a] M. D. Chekroun et al. *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs I*. Springer Briefs in Mathematics, Springer, 2015.

[CLW15b] M. D. Chekroun et al. *Stochastic Parameterizing Manifolds and Non-Markovian Reduced Equations: Stochastic Manifolds for Nonlinear SPDEs II*. Springer Briefs in Mathematics, Springer, 2015.

[CSG11] M. D. Chekroun et al. “Stochastic climate dynamics: Random attractors and time-dependent invariant measures”. In: *Physica D* 240.21 (2011), pp. 1685–1700.

[Daf70] C. M. Dafermos. “Asymptotic stability in viscoelasticity”. In: *Archive for Rational Mechanics and Analysis* 37.4 (1970), pp. 297–308.

[DFJ01] M. Dellnitz et al. “The algorithms behaind GAIO-set oriented numerical methods for dynamical systems”. In: *Ergodic theory, analysis, and efficient simulation of dynamical systems*. Berlin: Springer, 2001, pp. 145–174.

[Dob56] R. L. Dobrushin. “Central Limit Theorem for Nonstationary Markov Chains”. In: *Theory of Probability and its Applications* 1.1 (1956), pp. 65–80.

[Dol04] D. Dolgopyat. “On differentiability of SRB states for partially hyperbolic systems”. In: *Inventiones Mathematicae* 155.2 (2004), pp. 389–449.

[DV17] J. Demaeyer and S. Vannitsem. “Stochastic parametrization of subgrid-scale processes in coupled ocean-atmosphere systems: benefits and limitations of response theory”. In: *Quarterly Journal of the Royal Meteorological Society* 143.703 (2017), pp. 881–896.

[Dys49] F. J. Dyson. “The radiation theories of Tomonaga, Schwinger, and Feynman”. In: *Physical Review* 75.3 (1949), pp. 486–502.

[DZ96] G. Da Prato and J. Zabczyk. *Ergodicity for Infinite Dimensional Systems*. Cambridge University Press, 1996.

[Ehr94] M. Ehrendorfer. “The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Applications”. In: *Monthly Weather Review* 122.4 (1994), pp. 714–728.

[EN00] K.-J. Engel and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. New York: Springer-Verlag, 2000.

[EN06] K.-J. Engel and R. Nagel. *A Short Course on Operator Semigroups*. Springer, 2006.

[ER85] J. P. Eckmann and D Ruelle. "Ergodic theory of chaos and strange attractors". In: *Reviews of Modern Physics* 57.3 (1985), pp. 617–656.

[FJK13] G. Froyland et al. "Estimating long term behavior of flows without trajectory integration: the infinitesimal generator approach". In: *SIAM J. Numerical Analysis* 51 (2013), pp. 223–247.

[FMB07] C. L. E. Franzke et al. "The origin of nonlinear signatures of planetary wave dynamics: Mean phase space tendencies and contributions from non-Gaussianity". In: *Journal of the Atmospheric Sciences* 64.11 (2007), pp. 3987–4003.

[FR02] S. Fishman and S. Rahav. "Relaxation and Noise in Chaotic Systems". In: *Dynamics of Dissipation*. Berlin: Springer, 2002, pp. 165–192.

[Fra+15] C. L. E. Franzke et al. "Stochastic climate theory and modeling". In: *Wiley Interdisciplinary Reviews: Climate Change* 6.1 (2015), pp. 63–78.

[Fro01] G. Froyland. "Extracting dynamical behaviour via Markov models". In: *Nonlinear Dynamics and Statistics*. Birkhäuser, Boston, MA, 2001.

[Fro+07] G. Froyland et al. "Detection of coherent oceanic structures via transfer operators". In: *Physical Review Letters* 98.22 (2007), pp. 1–4.

[Fro98] G. Froyland. "Approximating physical invariant measures of mixing dynamical systems in higher dimensions". In: *Nonlinear Analysis* 32.7 (1998), pp. 831–860.

[Fro99] G. Froyland. "Using Ulam's method to calculate entropy and other dynamical invariants". In: *Nonlinearity* 12 (1999), pp. 79–101.

[Gar09] C. Gardiner. *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Heildelberg: Springer-Verlag Berlin, 2009.

[Gas+95] P. Gaspard et al. "Spectral signature of the pitchfork bifurcation: Liouville equation approach". In: *Physical Review E* 51.1 (1995), pp. 74–94.

[GC95] G. Gallavotti and E. D. Cohen. "Dynamical ensembles in stationary states". In: *Journal of Statistical Physics* 80.5-6 (1995), pp. 931–970.

[Gen+18] P. Gentine et al. "Could Machine Learning Break the Convection Parameterization Deadlock?" In: *Geophysical Research Letters* 45.11 (2018), pp. 5742–5751.

[Ger+18] S. Gerber et al. "A scalable approach to the computation of invariant measures for high-dimensional Markovian systems". In: *Scientific Reports* 8.1 (2018), pp. 1–9.

[Ghi+02] M. Ghil et al. "Advanced spectral methods for climatic time series". In: *Rev. Geophys.* 40.1 (2002), pp. 1–41.

[Ghi15] M. Ghil. "A mathematical theory of climate sensitivity or, How to deal with both anthropogenic forcing and natural variability?" In: *Climate Change : Multidecadal and Beyond*. Ed. by C. P. Chang et al. World Scientific Publishing Co./Imperial College Press, 2015, pp. 31–51.

[Ghi76] M. Ghil. "Climate stability for a Sellers-type model". In: *J. Atmos. Sci.* 33 (1976), pp. 3–20.

[Gil17] T. L. Gill. "The Feynman-Dyson View". In: *Journal of Physics: Conference Series* 845 (2017), p. 012023.

[GL06] S. Gouëzel and C. Liverani. "Banach spaces adapted to Anosov systems". In: *Ergodic Theory and Dynamical Systems* 26.1 (2006), pp. 189–217.

[GL20] M. Ghil and V. Lucarini. "The physics of climate variability and climate change". In: *Reviews of Modern Physics* (2020).

[GM13] I. Grooms and A. J. Majda. "Efficient stochastic superparameterization for geophysical turbulence". In: *Proceedings of the National Academy of Sciences* 110.12 (2013), pp. 4464–4469.

[GP83] P. Grassberger and I. Procaccia. “Measuring the strangeness of strange attractors”. In: *Physica D: Nonlinear Phenomena* 9.1-2 (1983), pp. 189–208.

[GW79] J. Guckenheimer and R. F. Williams. “Structural Stability of Lorenz Attractors”. In: *Publications Mathématiques d’I.H.É.S* 50 (1979), pp. 59–72.

[Hal17] P. Halmos. *Lectures in Ergodic Theory*. Dover Publications, 2017.

[Has76] K. Hasselmann. “Stochastic climate models Part I. Theory”. In: *Tellus* 28.6 (1976), pp. 473–485.

[Has88] K. Hasselmann. “PIPs and POPs: the reduction of complex dynamical systems using principal interaction and oscillation patterns”. In: *Journal of Geophysical Research* 93.D9 (1988), pp. 11015–11021.

[Hel05] I. M. Held. “The gap between simulation and understanding in climate modeling”. In: *Bull. Am. Meteorol. Soc.* 86 (2005), pp. 1609–1614.

[HH13] J. Holton and G. Hakim. *An Introduction to Dynamic Meteorology, 4th ed.* San Diego, CA: Academic Press, 2013, p. 531.

[HM10] M. Hairer and A. J. Majda. “A simple framework to justify linear response theory”. In: *Nonlinearity* 23.4 (2010), pp. 909–922.

[Hör67] L. Hörmander. “Hypoelliptic second order differential equations”. In: *Acta Mathematica* 119.0 (1967), pp. 147–171.

[HV18] H. Hosni and A. Vulpiani. “Data science and the art of modelling”. In: *Lettera Matematica* 6.2 (2018), pp. 121–129.

[Inu19] M. Inubushi. “Unpredictability and robustness of chaotic dynamics for physical random number generation”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.3 (2019), p. 033133.

[IPC14] IPCC. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by T. Stocker et al. Cambridge: Cambridge University Press, 2014.

[IPC21] IPCC. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by V. Masson-Delmotte et al. In press. Cambridge University Press, 2021.

[IS14] I. C. F. Ipsen and T. Selee. “Ergodicity Coefficients Defined by Vector Norms”. In: *SIAM Journal on Matrix Analysis and Applications* 32.1 (2014), pp. 153–200.

[Kam71] N. G. van Kampen. “The Case Against Linear Response Theory”. In: *Physica Norvegica* 5 (1971), p. 279.

[Kat66] T. Kato. *Perturbation Theory for Linear Operators*. Springer-Verlag, 1966.

[KCG15] D. Kondrashov et al. “Data-driven non-Markovian closure models”. In: *Physica D: Nonlinear Phenomena* 297 (2015), pp. 33–55.

[KCG18] D. Kondrashov et al. “Data-adaptive harmonic decomposition and prediction of Arctic sea ice extent”. In: *Dynamics and Statistics of the Climate System* 3.1 (2018).

[Ken71] V. M. Kenkre. “Integrodifferential Equation for Response Theory”. In: *Physical Review A* 4.6 (1971), pp. 2327–2331.

[Ken73] V. M. Kenkre. “Equations for the Theory of Response and Transport in Statistical Mehcanics”. In: *Physical Review A* 7.2 (1973), pp. 772–781.

[KG14] A. I. Khinchin and G. Gamow. *Mathematical Foundations of Statistical Mechanics*. Martino Publishing, 2014.

[KH95] A. Katok and B. Hasselblatt. *Modern Theory of Dynamical Systems*. Cambridge University Press, 1995.

[KKG05] S. Kravtsov et al. “Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability”. In: *Journal of Climate* 18.21 (2005), pp. 4404–4424.

[KKG10] S. Kravtsov et al. “Empirical model reduction and the modeling hierarchy in climate dynamics and the geosciences”. In: *Stochastic Physics and Climate Modelling*. Ed. by T. N. Palmer and P Williams. Cambridge: Cambridge University Press, 2010, pp. 35–72.

[KL99] G. Keller and C. Liverani. “Stability of the spectrum for transfer operators”. In: *Annali della Scuola Normale Superiore di Pisa* 28.1 (1999), pp. 141–152.

[KN32] B. O. Koopman and J. von Neumann. “Dynamical Systems of Continuous Spectra”. In: *Proceedings of the National Academy of Sciences* 18 (1932), pp. 255–263.

[Koo31] B. O. Koopman. “Hamiltonian Systems and Transformations in Hilbert Space”. In: *Proceedings of the National Academy of Sciences* 17 (1931), pp. 315–318.

[Kub57] R. Kubo. “Statistical-mechanical theory of irreversible processes I”. In: *Journal of the Physical Society of Japan* 12.6 (1957), pp. 570–586.

[Kub66] R. Kubo. “The fluctuation-dissipation theorem”. In: *Reports on Progress in Physics* 29.1 (1966), pp. 255–284.

[LAH00] D. J. Lea et al. “Sensitivity analysis of the climate of a chaotic system”. In: *Tellus* 52.A (2000), pp. 523–532.

[LB17] V. Lucarini and T. Bódai. “Edge states in the climate system: exploring global instabilities and critical transitions”. In: *Nonlinearity* 30.7 (2017), R32.

[LB19] V. Lucarini and T. Bódai. “Transitions across Melancholia States in a Climate Model: Reconciling the Deterministic and Stochastic Points of View”. In: *Phys. Rev. Lett.* 122.15 (2019), p. 158701.

[Lei75] C. E. Leith. “Climate Response and Fluctuation Dissipation”. In: *Journal of the Atmospheric Sciences* 32 (1975), pp. 2022–2026.

[Li76] T. Y. Li. “Finite Approximation for the Perron-Frobenius Operator. A Solution to Ulam’s Conjecture”. In: *Journal of Approximation Theory* 17 (1976), pp. 177–186.

[Liv95] C. Liverani. “Decay of Correlations”. In: *Annals of Mathematics* 142.2 (1995), pp. 239–301.

[LL21] K. K. Lin and F. Lu. “Data-driven model reduction, Wiener projections, and the Koopman-Mori-Zwanzig formalism”. In: *Journal of Computational Physics* 424 (2021), p. 109864.

[LLR20] V. Lembo et al. “Beyond Forcing Scenarios: Predicting Climate Change through Response Operators in a Coupled General Circulation Model”. In: *Scientific Reports* 10.1 (2020), p. 8668.

[LM94] A. Lasota and M. C. Mackey. *Chaos, fractals and noise*. Springer, New York, 1994.

[Lor63] E. N. Lorenz. “Deterministic Nonperiodic Flow”. In: *Journal of the Atmospheric Sciences* 20 (1963), pp. 130–141.

[Lor84] E. N. Lorenz. “Irregularity: a fundamental property of the atmosphere”. In: *Tellus A* 36 A.2 (1984), pp. 98–110.

[LS11] V. Lucarini and S. Sarno. “A statistical mechanical approach for the computation of the climatic response to general forcings”. In: *Nonlinear Processes in Geophysics* 18.1 (2011), pp. 7–28.

[Luc08] V. Lucarini. “Response theory for equilibrium and non-equilibrium statistical mechanics: Causality and generalized kramers-kronig relations”. In: *Journal of Statistical Physics* 131.3 (2008), pp. 543–558.

[Luc09] V. Lucarini. “Evidence of dispersion relations for the nonlinear response of the Lorenz 63 system”. In: *Journal of Statistical Physics* 134 (2009), pp. 381–400.

[Luc12] V. Lucarini. “Stochastic Perturbations to Dynamical Systems: A Response Theory Approach”. In: *Journal of Statistical Physics* 146.4 (2012), pp. 774–786.

[Luc16] V. Lucarini. "Response Operators for Markov Processes in a Finite State Space: Radius of Convergence and Link to the Response Theory for Axiom A Systems". In: *Journal of Statistical Physics* 162.2 (2016), pp. 312–333.

[Luc18] V. Lucarini. "Revising and Extending the Linear Response Theory for Statistical Mechanical Systems: Evaluating Observables as Predictors and Predictands". In: *Journal of Statistical Physics* 173.6 (2018), pp. 1698–1721.

[LV07] G Lacorata and A Vulpiani. "Fluctuation-Response Relation and modeling in systems with fast and slow dynamics". In: *Nonlinear Processes in Geophysics* 14 (2007), pp. 681–694.

[LW17] V. Lucarini and J. Wouters. "Response formulae for n-point correlations in statistical mechanical systems and application to a problem of coarse graining". In: *Journal of Physics A: Mathematical and Theoretical* 50.35 (2017), p. 355003.

[Mar+08] U. M. B. Marconi et al. "Fluctuation-Dissipation: Response Theory in Statistical Physics". In: *Phys. Rep.* 461 (2008), p. 111.

[Mar+21] G. Margazoglou et al. "Dynamical landscape and multistability of a climate model". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 477 (2021), p. 20210019.

[MD69] A. Muriel and M. Dresden. "Projection techniques in non-equilibrium statistical mechanics. I. A new hierarchy of equations". In: *Physica* 43.3 (1969), pp. 424–448.

[Mez05] I. Mezić. "Spectral Properties of Dynamical Systems, Model Reduction and Decompositions". In: *Nonlinear Dynamics* 41.1 (2005), pp. 309–325.

[Mit03] A. Y. Mitrophanov. "Stability and exponential convergence of continuous-time Markov chains". In: *Journal of Applied Probability* 40.4 (2003), pp. 970–979.

[Mor65] H. Mori. "Transport, Collective Motion, and Brownian Motion". In: *Progress of Theoretical Physics* 33.3 (1965), pp. 423–455.

[MPP02] G. Metafune et al. "Spectrum of Ornstein-Uhlenbeck operators in Lp spaces with respect to invariant measures". In: *Journal of Functional Analysis* 196.1 (2002), pp. 40–60.

[MS88] C. D. Meyer and G. W. Stewart. "Derivatives and Perturbations of Eigenvectors". In: *SIAM J. Numerical Analysis* 25.3 (1988), pp. 679–691.

[MTV01] A. J. Majda et al. "A mathematical framework for stochastic climate models". In: *Communications on Pure and Applied Mathematics* 54.8 (2001), pp. 891–974.

[Neu32] J. von Neumann. "Proof of the Quasi-Ergodic Hypothesis". In: *Proceedings of the National Academy of Sciences* 18 (1932), pp. 70–82.

[NPGR21] M. Ndour et al. "Spectral early-warning signals for sudden changes in time-dependent flow patterns". In: *Fluids* 6.2 (2021), pp. 1–24.

[NW17] A. Ni and Q. Wang. "Sensitivity analysis on chaotic dynamical systems by Non-Intrusive Least Squares Shadowing (NILSS)". In: *Journal of Computational Physics* 347 (2017), pp. 56–77.

[Pav14] G. A. Pavliotis. *Stochastic Processes and Applications*. Vol. 60. Springer, New York, 2014.

[Paz12] A. Pazy. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Vol. 44. Springer Science & Business Media, 2012.

[Pen96] C. Penland. "A stochastic model of IndoPacific sea surface temperature anomalies". In: *Physica D: Nonlinear Phenomena* 98.2-4 (1996), pp. 534–558.

[PO92] J. P. Peixoto and A. H. Oort. *Physics of Climate*. New York: AIP Press, 1992.

[Pol85] M Pollicott. "On the rate of mixing of Axiom A flows". In: *Inventiones Mathematicae* 81.3 (1985), pp. 413–426.

[PS08] G. A. Pavliotis and A. M. Stuart. *Multiscale Methods*. New York: Springer, 2008.

[PW09] T. N. Palmer and P Williams, eds. *Stochastic Physics and Climate Modelling*. Cambridge: Cambridge University Press, 2009.

[Rei02] C. H. Reick. "Linear response of the Lorenz system". In: *Physical Review E* 66 (2002), p. 36103.

[Ris89] H. Risken. *The Fokker-Planck Equation*. Second. Springer, 1989.

[RLL16] F. Ragone et al. "A new framework for climate sensitivity and prediction: a modelling perspective". In: *Climate Dynamics* 46.5 (2016), pp. 1459–1471.

[Rob01] J. C. Robinson. *Infinite Dimensional Dynamical Systems*. Cambridge University Press, 2001.

[Row+09] C. W. Rowley et al. "Spectral analysis of nonlinear flows". In: *Journal of Fluid Mechanics* 641 (2009), pp. 115–127.

[Rud91] W. Rudin. *Functional Analysis*. Second. New York: McGraw-Hill, 1991.

[Rue09] D. Ruelle. "A review of linear response theory for general differentiable dynamical systems". In: *Nonlinearity* 22.4 (2009), pp. 855–870.

[Rue86] D. Ruelle. "Resonances of chaotic dynamical systems". In: *Physical Review Letters* 56.5 (1986), pp. 405–407.

[Rue97] D. Ruelle. "Differentiation of SRB states". In: *Communications in Mathematical Physics* 187.1 (1997), pp. 227–241.

[Rue98] D. Ruelle. "General linear response formula in statistical mechanics, and the fluctuation-dissipation theorem far from equilibrium". In: *Physics Letters, Section A: General, Atomic and Solid State Physics* 245.3-4 (1998), pp. 220–224.

[Sch10] P. J. Schmid. "Dynamic mode decomposition of numerical and experimental data". In: *Journal of Fluid Mechanics* 656 (2010), pp. 5–28.

[Sch+17] T. Schneider et al. "Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations". In: *Geophysical Research Letters* 44.24 (2017), pp. 12,312–396,417.

[Sch68] P. J. Schweitzer. "Perturbation theory and finite Markov chains". In: *Journal of Applied Probability* 5.2 (1968), pp. 401–413.

[Sel71] G. R. Sell. *Topological Dynamics and Ordinary Differential Equations*. Van Nostrand Reinhold, 1971.

[Sen73] E. Seneta. *Non-negative matrices*. George Allen and Unwin, 1973.

[Sen84] E. Seneta. "Explicit forms for ergodicity coefficients and spectrum localization". In: *Linear Algebra and Its Applications* 60.C (1984), pp. 187–197.

[Sen88] E. Seneta. "Perturbation of the Stationary Distribution Measured by Ergodicity Coefficients". In: *Advances in Applied Probability* 20 (1988), pp. 228–230.

[Sen93] E. Seneta. "Sensitivity of finite Markov chains under perturbation". In: *Statistics and Probability Letters* 17.2 (1993), pp. 163–168.

[SG+21] M. Santos-Gutiérrez et al. "Reduced-order models for coupled dynamical systems: Data-driven methods and the Koopman operator". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 31.5 (2021), p. 053116.

[SGL20] M. Santos-Gutiérrez and V. Lucarini. "Response and Sensitivity Using Markov Chains". In: *Journal of Statistical Physics* 179 (2020), pp. 1572–1593.

[Sma98] S. Smale. "Mathematical problems for the next century". In: *The Mathematical Intelligencer* 20 (1998), pp. 7–15.

[Spa82] C. Sparrow. *The Lorenz Equations*. Springer, 1982.

[Sto63] H. Stommel. "Varieties of oceanographic experience". In: *Science* 139 (1963), pp. 572–576.

[Tan+18] A. Tantet et al. “Crisis of the chaotic attractor of a climate model: A transfer operator approach”. In: *Nonlinearity* 31.5 (2018), pp. 2221–2251.

[Tan+19] A. Tantet et al. “Ruelle–Pollicott Resonances of Stochastic Systems in Reduced State Space. Part III: Application to the Cane–Zebiak Model of the El Niño–Southern Oscillation”. In: *Journal of Statistical Physics* (2019).

[Tan+20] A. Tantet et al. “Ruelle–Pollicott resonances of stochastic systems in reduced state space. Part I: Theory”. In: *Journal of Statistical Physics* 179 (2020), pp. 1366–1402.

[TBD15] A. Tantet et al. “An early warning indicator for atmospheric blocking events using transfer operators”. In: *Chaos* 25.3 (2015), p. 036406.

[TE05] L. N. Trefethen and M. Embree. *Spectra and Pseudospectra*. Princeton University Press, 2005.

[Tes12] G. Teschl. *Ordinary Differential Equations and Dynamical Systems*. Providence: American Mathematical Society, 2012.

[TLD18] A. Tantet et al. “Resonances in a Chaotic Attractor Crisis of the Lorenz Flow”. In: *Journal of Statistical Physics* 170.3 (2018), pp. 584–616.

[Tuc02] W. Tucker. “A Rigorous ODE Solver and Smale’s 14th Problem”. In: *Foundations of Computational Mathematics* 2.1 (2002), pp. 53–117.

[Ula64] S. M. Ulam. *Problems in Modern Mathematics*. New York: John Wiley and Sons, 1964.

[UO30] G. E. Uhlenbeck and L. S. Ornstein. “On the Theory of the Brownian Motion”. In: *Phys. Rev.* 36.5 (1930), pp. 823–841.

[Val06] G. K. Vallis. *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-scale Circulation*. Cambridge: Cambridge University Press, 2006.

[VL18a] G. Vissio and V. Lucarini. “A proof of concept for scale-adaptive parametrizations: the case of the Lorenz ’96 model”. In: *Quarterly Journal of the Royal Meteorological Society* 144.710 (2018), pp. 63–75.

[VL18b] G. Vissio and V. Lucarini. “Evaluating a stochastic parametrization for a fast-slow system using the Wasserstein distance”. In: *Nonlinear Processes in Geophysics* 25.2 (2018), pp. 413–427.

[VV78] J. F. Van-Velsen. “On linear response theory and area preserving mappings”. In: *Physics Reports* 41.2 (1978), pp. 135–190.

[Wan13] Q. Wang. “Forward and adjoint sensitivity computation of chaotic dynamical systems”. In: *Journal of Computational Physics* 235 (2013), pp. 1–13.

[WDC20] J. A. Weyn et al. “Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere”. In: *Journal of Advances in Modeling Earth Systems* 12.9 (2020), e2020MS002109.

[WG19] C. L. Wormell and G. A. Gottwald. “Linear response for macroscopic observables in high-dimensional systems”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.11 (2019), p. 113127.

[Wil05] D. S. Wilks. “Effects of stochastic parametrizations in the Lorenz’96 system”. In: *Quarterly Journal of the Royal Meteorological Society* 131.606 (2005), pp. 389–407.

[Wil65] J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Oxford University Press, 1965.

[WL12] J. Wouters and V. Lucarini. “Disentangling multi-level systems: Averaging, correlations and memory”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2012.3 (2012).

[WL13] J. Wouters and V. Lucarini. “Multi-level Dynamical Systems: Connecting the Ruelle Response Theory and the Mori-Zwanzig Approach”. In: *Journal of Statistical Physics* 151.5 (2013), pp. 850–860.

[Wou+16] J. Wouters et al. “Parameterization of stochastic multiscale triads”. In: *Nonlinear Processes in Geophysics* 23.6 (2016), pp. 435–445.

[ZLP21] N. Zagli et al. “Spectroscopy of phase transitions for multiagent systems”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 31.6 (2021), pp. 1–8.

[Zwa01] R. Zwanzig. *Nonequilibrium Statistical Mechanics*. Oxford University Press, 2001.

[Zwa61] R. Zwanzig. “Memory effects in irreversible thermodynamics”. In: *Physical Review* 124.4 (1961), pp. 983–992.