UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia "Galileo Galilei"

Master Degree in Physics

Final Dissertation

Using Observables as Predictors through Response Theory: From Linear Systems to Nonlinear Climate Models

Thesis supervisor

Prof. Amos Maritan

Thesis co-supervisor

Prof. Valerio Lucarini

Candidate

Umberto Maria Tomasini

Academic Year 2019/2020

Abstract

The goal of response theory, in each of its many statistical mechanical formulations, is to predict the perturbed response of a system from the knowledge of the unperturbed state and of the applied perturbation. A new recent angle of the problem focuses on performing predictions of the change in one observable of the system by using the change in a second observable as a surrogate for the actual forcing. From the angle of control theory, this means that one can reconstruct the full state of a perturbed process from the observation of just a few variables. It turns out that it is not always possible to use perturbed observables as surrogates of the forcing, but no general method has been provided to practically discriminate a priori the unsuccessful cases. In this work, we make the first step towards the filling of this gap in the simple yet relevant case of forced and dissipative linear systems, including the case of systems with memory. We introduce a rigorous test (which we refer to as unpredictability criterion) that can rule out a variable as a predictor of other variables. This applies when the linear feedbacks acting on such a variable are stronger than dissipation. At this regard, we provide some simple yet informative examples and we propose some applications of our results in control theory and rational approximation theory. We also show that the presence of non-Markovian components in the dynamics enhances the predictive power of all variables. Moreover, we extend our results to random matrices, obtaining an average form of that test. One of the reasons for the effectiveness of response theory lies in its extreme flexibility in the definitions of response and perturbation. In spatially extended systems it is possible to look at the response of some local observables, defined in different zones of the system. In this work we apply response theory to the spatially extended system Lorenz 96, which has been proposed in the context of climate science. We perturb the system in a given location and then we study to what extent a given local observable can predict the behaviour of another local observable. We show this kind of study allows to unveil potential information flows and causal links within the system. Given this study of the dynamical system Lorenz 96, we devise a rather general method to study spatially extended systems and their emergent properties through response theory.

Acknowledgements

A lot happened in the last five years. When I first came in Padua my favourite sentence was from Paperon de' Paperoni: "Talent does not exist, just inspiration and ambition exist". I was pursuing a solitary dream, to go out of my house and to discover the world through the lens of Physics. The Galilean School of Higher Education and the University of Padova welcomed me and my dreams, giving me even more, and I will always be thankful for that. Because of them, a flow of people entered in my life, enriching my dream and changing me. The main character of my favourite manga, Mob from Mob Psycho 100 by One, experienced something similar. He says: "people can change through their interactions with others. I was able to change because of everyone. I was able to become a little stronger than if I had been alone". I want now to express my gratitude to the people and institutions which have been in my life in these years.

Let's start from the roots, my family: my mum Maria Pia B., my dad Eugenio T., my sister Lucia T., my aunt Francesca B., my uncle Orville C., my cousin Enrico C. and Gloria P., my grandparents Mirella B. and Franco B. and all the other members of my family (which is very large and impossible to fully thank here). I thank them for their patience, their understanding, their approval and support. I cannot even begin to imagine their emotions when I moved out of my house at 18 years old, when I lived away for weeks, when I spent my Covid quarantine in Padova. They inspired me in many ways; they taught me the importance to go on, to take care of the moments spent together and to empathize with people. They shaped me with their love in many ways, building who I am.

Let's go on with the person with whom I have the deepest and full of emotions relationship: my girlfriend Anna. We met (almost) by accident three years and a half ago, close to the Physics department and the other places which have seen our love grow and develop through the years. My years in Padova would have been entirely different without her. I have learnt together with her how to build relationships: brick by brick, step by step, without fretting. Observing myself with her eyes I rounded my corners and I knew myself better. Observing her with my eyes I learnt that you never stop to discover how much beautiful is a person. Observing us with our eyes I learnt the value of the things of life and the value of the things of life together, and how much we complete each other.

There are some people from your hometown, that you keep with yourself wherever you go. I would like to thank a group of wonderful people with whom I have incredible memories and bonds: Matteo V., Giacomo D., Stefano R., Leonardo L., Andrea Z., Tommaso M., Francesco S. With them I have shared travels and the happiness within it, the curiosity of the world, a common view of life, intense discussions, the good music, a true friendship. I would like to deeply thank Francesco C.. He has been my early companion of hopes for the future when we were children, and still now we confess to each other doubts and thoughts, trying to make sense of the world. I want to thank him, together with Caterina M. Whenever and wherever we meet, I feel that we are progressing in our life and that we reciprocally understand each other. I want to thank Andrea Z., Michel M., Marco M., Matteo C., Simone M. We have been together for a long time, and we still do.

I want now to thank a group of amazing people who entered in my life in these last five years: "I ragazzi di Via Paolotti", or "Paolo Barigelli e la sua squadra". Whenever I have trouble or I need advice, or I want to share a good laugh or to start a discussion, they are the first people who come up in my mind. Let' start from the people with whom I have shared and lived the most, under the Galilean's rooftop: Giovanni F., Enrico M., Antonio C., Andrea B., Matteo Z. They entered in

my daily life so strongly and deeply that I cannot imagine a life without them around. A special thanks goes to Giovanni and Enrico, the members of the laboratory group: "I tre porcellini", which is something fundamental in my life. I want to deeply thank them for their patience and work in the laboratory, in front of my deficiencies. I want to thank them for the endless list of amazing memories I share with them, where we got to deeply know ourselves. Another special thanks goes to Antonio. He is the glue of our group, the glue of my Galilean years. When he was the first of us to go abroad, I missed him a lot. Then we spent together the quarantine months and several other events, and I was truly happy to be with him during them. Ruviano will always be a special place for me. I want to thank Giovanni C., Guglielmo F., Luna C., Paolo B., Nicola B., Giacomo V. and Beatrice C. You have made my years in Padova special and memorable, and I want to keep this bond in the years to come. Moreover, I want to thank the last people who entered in this group: Alessandro L. and Alice G. A special thanks goes to Giovanni. He helped me in may ways, earnestly listening to me and giving me advice that encouraged me. The discussions with you happened in turning points of my life, and your point of view helped me a lot, without failing. Another special thanks goes to Guglielmo. I lost count of the discussions, the crazy things and the laughs we had together. Another special gratitude goes to Luna. She always found time to listen to me, ordering my thoughts and finding the answer to my problems.

I want now to thank all the people I met under the Galilean's rooftop. My life has been enriched a lot by the experience in the Galilean School of Higher education, one of the best experiences in my life so far. Living under the same roof with such a large bunch of clever, friendly, sharp and incredible people inspired and stimulated me in many ways, in both the personal and academic aspects. In such a unique environment you can share very different experiences, from memorable adventures to intense and fulfilling studying. The list of people I would like to thank is very long since it stretches along various cohorts of the School, and each person would need a few lines in these acknowledgements. For this reason I will simply list them, in order of their cohort: Davide R., Alessandro S., Carlo Maria S., Daniele S., Davide G., Gianluca M., Niccolò R., Giordano A., Enrico D.G., Filippo S., Carina N., Omar D., Lorenzo C., Tommaso B., Roberto A., Enrico A., Gabriele, Tommaso, Luca B., Alessandro F., Paolo P., Nicholas T., Federica C., Luca T., Giacomo M. Federico S., Simone P., Daniele N., Vittoria R., Camilla C.T., Davide D., Daniele R., Leonardo Z., Carlo B., Giorgio M., Francesco G., Luca N., Lorenzo B., Michele C., Antonio C., Giulia P., Marta C., Roberta D.P., Francesco V., Nicola C., Pier Federico P., Filippo F., Alberto A., Leonardo S., Daniele T., Alessandro R., Andrea G., Marco G., Andrea M., Enrico A., Teresa G., Asia F., Giacomo P., Dario D.D., Laura O., Luca I.. A special thanks goes to Pier Federico, since he made me discover the volunteering reality during the Covid-19 pandemy. Another special thanks goes to a scientific subgroup of the Galilean people, which is called "Giovani Naturalisti", who discussed with me a few points of the thesis, helping me in a decisive way and proving the collaborative reality of the Galilean School.

I want to express my uttermost gratitude to the institutions which supported me in these years. I want to thank the Galilean School of Higher education, which gave me the unique opportunity to grow in an amazing and interdisciplinary environment. It has been an incredibly valuable added value in my life and in my academic career. I entered in a network of propositive, helpful and open people, who are spread in all the fields and all over the world. The life in the Galilean School prompted my curiosity towards both sciences and other disciplines, giving me the taste of what interdisciplinarity is really like. This point of view greatly affected me in my research interests. I want to thank my Galilean tutors for their support: Professor Livia Conti and Professor Marco Zanetti. A special thanks goes to Professor Paolo Ciatti: his friendly, unconditioned support and his humanity are a plus to the Galilean School. I want to express my gratitude to the directors who directed the Galilean School during my years Professor Michele Cortelazzo and Giancarlo Benettin and to the professors who were responsible to choose the Galilean courses Professor Michele Maggini and Professor Andrea Lucchini. Moreover, I would like to thank the Residence Nord Piovego and its staff, for his warm welcoming and staying. Another thanks goes to the canteens which granted us food and a place to gather together, becoming important places of my university years: Piovego, Pio X, Forcellini, Murialdo and Belzoni.

My gratitude goes to the Department of Physics, which has granted me in these years a higher-level education in a large variety of Physics fields, firmly building my background and my skills. Many professors helped me in this process. My first, big thanks goes to Professor Amos Maritan. In all the key moments and decisions of my Master, I always looked for his suggestions and guidance. In all of these occasions, he gave me the right perspective and the right suggestion, greatly helping me. His approach to Physics, in particular to Theoretical Physics deeply inspired me. His supervising skills, his patience and his sharp glances over the scientific problems really guided me, first in the Galilean research project we did together and then in the Master Thesis. I would like to thank Professor Samir Simon Suweis. In both of his courses, he has taught me the importance of Physics as a quantitative way to better understand the scientific problems of other scientific fields. He always has been open to questions to further the lectures and the project we did together was to me a first touch of the procedure of applying Theoretical physics to other fields. I would like to thank Professor Sandro Azaele, for the discussions we did for the Professor Suweis' research project and for his important corrections to the Master thesis, which contributed to making it more readable and understandable, expanding its background. Another special thanks goes to Professor Luca Martucci who, in addition to being a Galilean tutor, he has also been my supervisor of the Bachelor thesis. Under his guidance, I first learned how to go through the literature, how to write a scientific report and how to push beyond the already existing literature. He also allowed me to be a tutor for his course in Mathematical Methods for Physics, greatly boosting my teaching skills.

I want now to deeply thank the man who is behind all this thesis, my supervisor Professor Valerio Lucarini. He friendly encouraged and helped me in all the process, starting from our first call last summer, then during the months of preparation before my arrival to Reading, during my month in Reading, during the return in Italy due to the pandemic and in all the successive months, even in a smart-working mode. He immediately put me on the boundaries of the state-of-the-art of the theory, and he was able to extract the best out of me in order to push beyond them, with a great maieutics. I learned so much from him: to passionately wonder about your research interests, how to approach a new scientific problem, the flexibility that research needs and about the importance to establish friendly relationships. He has my deepest gratitude. Moreover, I want to thank all the people whom I met during my Reading stay for their warm welcoming and friendship, which have sweetened the difficulties of my first move abroad: Elena S., George M., Manuel S.G., Ieva D., Sebastiano R., Thomas S., Jennifer I., Carlo C., Ben A., Georgios S., Giulia C., Leonardo R., Chiara V. A special thanks goes to Elena S.: she is the one who suggested to me to contact Professor Lucarini, who gave me a lot of suggestions for my Reading stay and who opened her house for me.

Contents

Introduction

Elements of Response Theory

Response Theory (RT) is a Statistical Physics theory that provides a general method for predicting the changes in the statistical properties of observables of interest once the system under investigation is perturbed. A cornerstone in the development of RT has been the work by Kubo [1,2], who considered the case of weakly perturbed systems near *thermodynamic equilibrium*. When a system is in this kind of steady state, there are no steady state currents in the system or, equivalently, the time-reversal property applies on the trajectories of the system in its steady state. Although Kubo's theory has been criticized at an early stage (by the Van Kampen's argument [3], also discussed in [4]), it has been successful in a great variety of disciplines, such as materials science [5]. The main practical tools of RT that are key to many applications are response formulae that allow one to compute the response of the system from the knowledge of the applied perturbation and of statistical properties of the unperturbed system. In particular, the response of the system is defined as the change of the expectation value of an observable of interest Ψ resulting from the applied perturbation. The expectation value of Ψ in the perturbed system is expanded in a perturbative series where the zeroth-order term is the expectation value of Ψ in the unperturbed system. The higher-order terms are expressed in terms of *response functions* which contain information about the statistics of the unperturbed system and the applied forcing. These response functions themselves can be seen as expectation values in the unperturbed system of some quantities that depend on the applied forcing and Ψ. Another key feature of these response functions is that of, once we compute them, they allow the prediction of the response of the system to a continuum of different forcings of the system. We can observe that knowing the statistics of the unperturbed system, in particular its spontaneous fluctuations, it is possible to compute the response of the system to a given weak perturbation. This is the key idea behind the fluctuation-dissipation theorem (FDT) by Kubo [2], which establish a link between the forced and free fluctuations in the perturbative regime. We will go through the Kubo's RT and the FDT in **Chapter 1**.

A crucial issue not treated in Kubo's theory is related to how systems out of (thermodynamic) equilibrium respond after being forced. These intrinsically dissipative systems are extremely important, and their steady states are called non-equilibrium steady states (NESS). Viscous hydrodynamics, granular materials, and the climate are a few relevant examples [4,6]. A rigorous and crucial development in the context of deterministic dynamics was provided by Ruelle [7–10], who derived a RT for Axiom A dynamical systems both at the equilibrium and out of equilibrium. Axiom A systems are a par-

ticular class of dynamical system, but the *chaotic hypothesis* proposed by Gallavotti and Cohen [11, 12] states that they are practically indistinguishable from the effective properties of macroscopic observables in high-dimensional systems which display a sufficient degree of chaotic behaviour. As a consequence, RT should work in this large class of dynamical systems. Equilibrium dynamical systems possess an invariant measure absolutely continuous with respect to Lebesgue. Instead, general non-equilibrium dynamical systems obeying Axiom A dynamics possess an invariant measure that is not smooth along the stable manifolds over which it is defined and smooth along the unstable manifold. On average, the phase space experiences a contraction along those directions. Ruelle proved that the response relations are divided into two contributions: the one which comes from the unstable and central manifolds, which can be framed as a FDT result, and the one from the stable (or contracting) manifold, which cannot be framed as a FDT result [10]. In other words, the natural fluctuations are not equivalent to the forced perturbations along the stable directions [6,13,14]. The two contributions are hard to compute directly in practical applications [15–17].

RT can also be approached by the stochastic dynamics point of view [17–19]. Adding a noise term in the equation of the dynamical system, the invariant measure becomes smooth even along the stable directions and the FDT fully holds [20]. In this setting, the obtained formula is called the Kubo-Agarwal formula, which reduces to the Kubo formula for equilibrium systems. The addition of a noise term has to be justified by the nature of the considered problem. This stochastic perspective becomes relevant in many complex systems, where the focus is on coarse-grained dynamics, which is effectively stochastic as a result of the presence of microscopic degrees of freedom. Note that the coarse-grained dynamics is in general non-markovian, with memory effect becoming negligible in the limit of infinite time-scale separation between the fast and slow variables [21–25]. A generalised version of the FDT valid for higher moments has been proposed by [26] in the spirit of Zubarev's generalization of Kubo's results [27]. In **Chapter 2** we will review and compare these two approaches to non-equilibrium systems.

Predictors and Predictands

A different angle on the problem of defining the response of a system to perturbations has been proposed in [28], where it is discussed in fairly general terms how to relate the response of different observables of a system undergoing a perturbation. The goal is to understand to what extent we can use perturbed observables as surrogates of the perturbation to predict the future state of other observables. This viewpoint is relevant since it allows one to predict the state of the desired observables even if we lack some information on the actual perturbation acting on the systems. It turns out that *not* all choices of predictor and predictand are equally successful: some pairs may work, some other not, and there is an asymmetry in the predictive power of observables.

A surrogate response function that is able to predict the change in the observable Ψ_1 from the knowledge of the change in the observable Ψ_2 is truly predictive if it requires only information on the past values of Ψ_2. This requires the Fourier transform of the surrogate response function to have only poles in the lower complex $\omega-$ plane. The response functions of the usual RT satisfy this requirement by construction, and from this fact it is possible to derive a set of Kramers-Kronig (KK) relations related to these

response functions [6, 13, 14, 26], as shown in Chapters 1 and 2. Due to how surrogate response functions are built, one can have an asymmetry between the predictive power of Ψ_1 on Ψ_2 and Ψ_2 on Ψ_1: some observables cannot predict other observables (see discussion in Chapter 2). In [28] it is possible to find an example of a dynamical system, the climatic model Lorenz 96 (L96), where the state of an observable Φ_1 can be predicted by another observable Φ_2 but it cannot be predicted by a third one Φ_3.

In Section 2.6.2 we present an innovative method which quantifies how much a surrogate response function is non-predictive. The method is rather efficient from the computational point of view and its theoretical justification is based on the fact that the surrogate response functions do not satisfy anymore the KK relations. The problem to actually quantify the predictive ability of the surrogate response functions can emerge in a variety of situations where we have many non-predictive surrogate response functions and we want to choose among them the one providing the best prediction. Moreover, applying the method in such problems can unveil potential flows of information or causal links present in the system, as we will see in the L96 system in Chapter 4.

Surrogate RT and Control in Linear Systems

The requirement for the success of the prediction turns out to be that the Fourier transform of the surrogate response function has only poles in the lower complex $\omega-$ plane. This is equivalent to the stability requirement found in the control theory literature [29]. This is usually observed after the full computation of the response functions, which can be quite cumbersome. We would like to have a criterion which allows us to discriminate a priori, before obtaining the response function, whether an observable is actually predictive or not. We present in **Chapter 3** a first original step towards the solution of this problem. We focus on simple linear systems (Markovian and non-Markovian ones), and we find a criterion which clarifies when an observable cannot be used for prediction purposes. We refer to this result as the *Unpredictability Criterion* (UC). We show that adding memory effects the UC changes, improving the predictive ability of the variables. Furthermore, we randomize the matrix which defines the linear system, deriving a form of the UC which is valid on average over the dynamical variables. This approach is meaningful in cases where we have just partial information on the structure of the system, maybe due to measurement errors or noise, and hence we have to add some randomness in the matrix of the dynamical system.

Even though this work is meant to be a first step towards the design of general methods, focusing just on linearized systems could seem to be purely of academic interest. On the other hand, linearized systems are extensively studied and used in the broad context of *control theory* [29, 30]. Control theory studies how to implement feedbacks to control the value of a variable of interest of a dynamical system [31], using the measured value of that variable. In such a problem, the perturbation of the system is seen as an input, and the variable of interest depending on the state of the system is seen as an output. Controlling the system can be loosely divided into three tasks: (i) sensing of the state or the output of the system, (ii) comparison of the measured values with the desired ones and (iii) implementations of corrective actions on the input.

In the context of control theory, response functions are usually referred to as transfer

functions, connecting input and outputs signals. A rich wealth of tools concerning linearized systems have been developed in control theory, where they find extensive applications. Remarkably, a large variety of non-linear systems can be reduced to a linear version of them, employing a careful choice of the input of the system or by expanding around an equilibrium point of interest [32]. Moreover, it is possible to implement feedbacks which can keep a system in a linear regime [29, 30]. The main motive of the success of linear systems is that explicit results can be derived much more easily than non-linear systems and they can be studied and controlled more transparently.

We propose here the idea that surrogate RT could provide some useful insights for control theory. One of the problems that emerge when designing a controller is the partial information of the state of the dynamical system. To sense the output of the system, we have to use an observer, which can consist of a few sensors. In many applications, it is not practical for the observer to directly measure the whole system state, so typically it focuses on a few observables of interest, which are the most meaningful to reconstruct the full state. Surrogate RT could identify the most meaningful observables in terms of their ability to predict other observables. In this sense, the predictive power of the dynamical variables is the ability of the dynamical variables to reconstruct the state of the other dynamical variables. In particular, our work in this thesis could provide an efficient criterion to rule out a priori the useless observables, which are the ones without predictive power. Another relevant application of our work in control theory deals with the identification of non-minimum phase systems [29]. This class of dynamical systems features controllability problems which can have detrimental effects on industrial controllers [29, 33]. The UC gives an effective method to discover whether a dynamical system is non-minimum or not, preventing eventual damages of the controller.

RT and Climate

Response Theory is a rather general and flexible theory, which can be applied to a large variety of contexts in very different scientific disciplines. One of these contexts is the climate, where it provides new insights and alternative methods to very long and time-consuming numerical simulations. The climate is a NESS since it is a forced and dissipative complex system. The forcing acting on the climate is given by variations in time and space of the net energy flux at the top of the atmosphere. The system tends to re-equilibrate to a steady state thanks to the transport of mass and energy provided by moving geophysical fluids and to the exchange of infrared radiation [34–36].

One of the greatest issues which are studied within climate science is climate change. The goal is to predict the change of the statistical properties after some internal (e.g the atmospheric composition) or external parameter (e.g. the solar irradiance) is perturbed. This is a very complicated task since the climate is a system extended in space and very complex, with a lot of processes and feedbacks acting within it in a wide range of spatial and temporal scales. A first and simplistic -but heavily used- start for tackling this gargantuan problem can be found in the equilibrium climate sensitivity (ECS) [34], which measures the globally and annually averaged surface air temperature increase which would result from the doubling of the concentration of a given greenhouse gas

(GHG) versus that of the reference state with temperature T_0:

$$\text{ECS} \equiv \Delta T = \Lambda(T_0)\Delta R_{2\times\text{GHG}}, \tag{1}$$

where the factor $\Lambda(T_0)$ is called linear gain factor and $\Delta R_{2\times\text{GHG}}$ is the extra net radiative forcing which will be present in the atmosphere as a result of the doubling of the concentration of the GHG. The underlying hypothesis of the definition of the ECS is that ΔT is measured between two steady states. We can see that the ECS has some serious shortcomings: it provides no temporal information since it addresses just long term changes and no spatial information too.

To overcome this problem, recently it has been proposed to use the Ruelle's RT presented in Chapter 2 to effectively predict the response of observables of interest, such as the global average temperature, to a given perturbation, in a well-defined context, where the climate is seen as a chaotic dissipative deterministic system. The FDT cannot be applied since we are dealing with a NESS. As a consequence, due to the presence of the stable manifold contribution, particular events called climate surprises can emerge [37]. Remarkably, the climate surprises cannot be predicted by the natural fluctuations present in the unperturbed system. The Ruelle's approach has been successfully applied to climatic models with increasing degree of complexity, going from simple models [14], to intermediate-complexity ones [6,15] and even to Global Climate Models (GCM) [38,39], spanning over a large range of temporal scales [40]. the usefulness of RT lies in the fact that once we obtain the response functions for a given observable from a known perturbation with a simple temporal pattern, we can use it to make predictions for a continuum of temporal patterns, improving the climate change predictions [34].

In **Chapter 4** we apply the surrogate RT on a climate model: the Lorenz 96 (L96) model [41–43]. The L96 model is a rather simple climatic model, but it contains all the basic physical processes that are generally present in a non-linear dynamical system: advection, mechanical damping and forcing. Moreover, it displays chaotic behaviour and the presence of travelling waves within it. The L96 has already been a subject of RT investigations in [6], showing successfully the applicability of the Ruelle's RT. The first application of the surrogate RT on the L96 model has been proposed in the seminal paper [28], where surrogate response functions were built for global perturbations, looking at global observables. In Chapter 4 we apply the surrogate RT again on the L96 system, but focusing on local perturbations and local observables. This is possible since the L96 system is an extended model in space. In other words, perturbing the system in a given location, we analyze when a local dynamical variable is able to predict the time behaviour of another local dynamical variable. A crucial role in solving this problem is played by the form and the time scale of the propagation of the perturbation. In particular, it is possible to define an *hierarchy* of dynamical variables in terms of their predictive power, which is closely related to how fast they react to the perturbation. The sooner a dynamical variable feels the perturbation and the better it could predict other variables. In addition to that, we use the L96 system as a benchmark for our ratio method derived in Chapter 2, showing that it could discover causal links or flows of information with a definite direction. We conclude showing that perturbing the system in two different ways and considering a combination of global and local observables, possible bad predictions performed by local observables can be drastically improved.

The experimental setting related to that observation is that of a local observer who tries to predict what is happening to an observable in other parts of the globe. First, he/she perturbs the system just with one forcing. If he/she discovers that the information provided by that forcing is not enough, he/she considers additional forcings in order to extract more information from the system, improving his predictions. Our application of the surrogate RT on the spatially extended and chaotic L96 dynamical system shows that it is a theory which can be successfully applied to non-trivial systems, providing new insights on them. Using the L96 system as a toy model, we derive a methodology which can be generally applied to a generic dynamical system to understand which local observables can predict other local observables, using a well-defined hierarchy in terms of their predictive power.

Chapter 1

Kubo's Response Theory

In this section, we resume the Kubos's theory, which can be considered as the first cornerstone of RT [1, 2]. In **Section 1.1** we present the Langevin equation and the Fokker-Planck (FP) equation, and the connection between them [44]. Using this setting, in **Section 1.2** we remark the difference between thermodynamic equilibrium and non-equilibrium steady-state (NESS). Next, in **Section 1.3** we derive the Kubo's response formula in the Hamiltonian formulation, following [4], which holds for systems at the thermodynamic equilibrium. In **Section 1.4** we derive the Krames-Kronig (KK) relations related to the Kubo's response formula, whose only ingredient is the causality of the response function. In **Section 1.5** we study the FDT, which focuses on relating the fluctuations of unperturbed systems at the thermodynamic equilibrium with their properties after being slightly perturbed. In **Section 1.6** we present the Green-Kubo relations, which compute transport coefficients by means of equilibrium averages. Lastly, in **Section 1.7** we explain in nutshell the famous van Kampen's objection to Kubo's response theory despite its effectiveness, and how his argument can be objected supporting the effectiveness of RT.

1.1 Langevin equation and FP equation

1.1.1 Brownian Motion

Let's first present the Langevin equation for the motion of a mesoscopic particle in d dimensions over a fluid medium [44].This kind of motion is called Brownian Motion (BM), and it features a separation of time scales between the collision time τ_c (the typical time-scale between two collisions of the mesoscopic particle with the fast-moving particles of the fluid) and the dissipation time τ_d (the typical time-scale over which the mesoscopic particle reduces its speed due to viscosity):

$$\tau_c << \tau_d. \tag{1.1}$$

We take the point of view of the mesoscopic particle, whose motion has characteristic time τ_d. We have that the position $\vec{x}$ of the mesoscopic particle follows the following equation, which is called Langevin equation:

$$m\ddot{x}_i = -\tilde{\gamma}\dot{x}_i + \eta_i, \qquad i = 1, ..., d, \tag{1.2}$$

where m is the mass of the particle and $\tilde{\gamma}$ is the viscosity coefficient. We have introduced the white noise η_i, which is such that:

$$\langle \eta_i \rangle = 0, \qquad \langle \eta_i(t)\eta_j(t') \rangle = \tilde{\Gamma}\delta_{ij}\delta(t-t'), \tag{1.3}$$

for each $i, j = 1...d$ and t, t'. The average is over many realizations of the system, i.e. over many trajectories of the mesoscopic particle obtained by different realizations of the white noise. The white noise expresses the influence of the collisions of the microscopic degrees of freedom of the medium with the mesoscopic particle: their average influence is zero and the collisions are uncorrelated over different dimensions and different time instants.

From the Langevin equation we can extract a first FDT. In order to do that, we use the following relation [44] for the velocity of the particle $v = \dot{x}$:

$$\langle v_i^2(t) \rangle = \frac{\tilde{\Gamma}}{2m\tilde{\gamma}}(1 - e^{-2\frac{\tilde{\gamma}}{m}t}) + v_i^2(0)e^{-2\frac{\tilde{\gamma}}{m}t}, \tag{1.4}$$

where $v_i^2(0)$ is the initial condition. In the limit of large times $t \to \infty$ we have:

$$\langle v_i^2(t) \rangle \to \frac{\tilde{\Gamma}}{2m\tilde{\gamma}}. \tag{1.5}$$

On the other hand, the equipartition theorem gives the following relation at the equilibrium:

$$\frac{1}{2}m\langle v_i^2(t) \rangle = \frac{k_B T}{2}, \tag{1.6}$$

where T is the temperature and k_B is the Boltzmann constant. Comparing (1.5) and (1.6) we obtain our first dissipation-fluctuation relation [4, 44]:

$$\tilde{\Gamma} = 2\tilde{\gamma}k_B T, \tag{1.7}$$

It is called fluctuation-dissipation relation because it links dissipation coefficients ($\tilde{\gamma} \sim m/\tau_d$) with fluctuation elements (T and $\tilde{\Gamma}$). The first one measures how fast the system reacts to a perturbation, relaxing to equilibrium, whilst the second ones are property of the unperturbed system.

In the Langevin equation the stochastic component of the system is described by the noise η_i, which makes the trajectories of the dynamical variable x_i random. We can alternatively describe the system by means of a probability distribution function (PDF) $p(x, t)$ over all the possible positions at a given instant t, following [44]. Note that we denote by x the vector $\{x_i\}_{i=\{1,...,d\}}$. We will use this notation for each vector in our work. The quantity $p(x, t)d^d x$ is the probability to find the mesoscopic particle at time t in the infinitesimal volume $x + d^d x$ around the position x. We want to write an equation which describes the evolution in time of this PDF. In order to do that, we introduce the transition rate $W(x', x)$, which is the probability per unit time that the particle jumps from x to x'. By means of this quantity, we can write the evolution in time of the PDF through the following relation [44]:

$$\frac{\partial p(x,t)}{\partial t} = \int_{-\infty}^{\infty} d^d x' [W(x, x')p(x', t) - W(x', x)p(x, t)] \tag{1.8}$$

which resembles a master equation, where the first term is a gain term (particles jump in x) and the second one a loss term (particles jump away from x). In this formalism the characteristic time is way larger than τ_c, since the instantaneous influence of the collisions is lost. At this point we can introduce the vectorial displacement $\chi = x' - x$ and rewrite the jump probability as follows:

$$W(x', x) = W(x; \chi). \tag{1.9}$$

Then we take (1.8) and we change the integration variable from x' to χ:

$$\frac{\partial p(x,t)}{\partial t} = \int_{-\infty}^{\infty} d^d\chi [W(x+\chi; -\chi)p(x+\chi, t) - W(x;\chi)p(x,t)], \tag{1.10}$$

and we change the integration variable from χ to $-\chi$ in the first integral:

$$\frac{\partial p(x,t)}{\partial t} = \int_{-\infty}^{\infty} d^d\chi [W(x-\chi; \chi)p(x-\chi, t) - W(x;\chi)p(x,t)]. \tag{1.11}$$

Now we Taylor expand the first term within in the integral in (1.11) around x for small χ (this expansion is called Kramers-Moyal expansion [45]) and we stop at the second order. This is justified by the fact that in the BM large displacements are very unlikely to happen. Note that we cannot hope to obtain a better equation truncating the expansion after the second order, thanks to the Pawula Theorem [45]. This theorem states that for a positive transition probability W, the Kramers-Moyal expansion may stop either after the first term or after the second term. If the expansion does not stop after the second term, all the infinite number of terms have to be kept. Truncating up to the second order and defining the following averages:

$$a_i(x) = \int d^d\chi W(x;\chi)\chi_i \tag{1.12}$$

$$b_{ij}(x) = \int d^d\chi W(x;\chi)\chi_i\chi_j, \tag{1.13}$$

we find the FP equation:

$$\partial_t p = -\nabla(ap) + \frac{1}{2}\sum_{i,j}\partial_i\partial_j(b_{ij}p), \tag{1.14}$$

which describes the evolution in time of the PDF $p(x,t)$ of a system described by the set of N dynamical variables x and subject to the drift coefficients $\{a_i(x,t)\}_i$ and diffusion coefficients $\{b_{ij}(x,t)\}_{ij}$.

1.1.2 General stochastic differential equation

In this section we generalize the Langevin equation and the FP equation taking the point of view of stochastic differential equations [44, 46]. The main problem in the Langevin equation derived in the BM case (1.2) is the fact that the deterministic part is differentiable with respect to time, whilst the white noise is discontinuous at each time t. To overcome this problem, we introduce the notion of Wiener process, which is the stochastic process defined by the integration in time of white noise:

$$W_i(t) \equiv \int_0^t dt' \eta_i(t'). \tag{1.15}$$

The statistical properties of the white noise (1.3) are transferred in the following statistical properties of the Wiener process:

$$\langle W_i \rangle = 0, \qquad \langle W_i(t) W_j(t') \rangle = \tilde{\Gamma} \delta_{ij} \min(t, t'), \tag{1.16}$$

in particular:

$$\langle W_i(t)^2 \rangle = \tilde{\Gamma} t, \tag{1.17}$$

hence the variance of the Wiener process diffuses linearly in time. The Wiener process has the following properties:

- $W_i(t)$ is continuous in time and it has zero average at any instant t.
- For any $t_1 < t_2 < t_3$, the non-overlapping increments $(W(t_2) - W(t_1)$ and $(W(t_3) - W(t_2)$ are independent and identically distributed (iid).
- For any $t_1 < t_2$, the increment $(W(t_2) - W(t_1)$ follows a Gaussian PDF with zero average and variance $\tilde{\Gamma}(t_2 - t_1)$, as a consequence of the central limit theorem (since the Wiener process is built summing up iid realizations of the white noise).

It is possible to define an infinitesimal increment of a Wiener process as follows:

$$dW_i(t) = W_i(t + dt) - W_i(t) \sim \mathcal{N}(0, \tilde{\Gamma} dt). \tag{1.18}$$

At this point we can define the general form of the overdamped Langevin equation (we do not consider the inertial term in (1.2)) as a well-defined stochastic differential equation, for a generic stochastic process X_i:

$$dX_i(t) = a_i(X(t), t)dt + c_{ij}(X(t), t)dW_j(t), \tag{1.19}$$

where the a_i are the generalized drift coefficient and the c_{ij} are coefficients which generalize the previous Langevin equation (1.2) (they are linked to the generalized diffusion coefficients, as it follows).

We can obtain the generalized FP equation from the generalized Langevin equation (1.19). To do that, we consider a test function $f(X, t)$ at least twice differentiable with respect to X and we take its Taylor expansion up to the second order (without taking care of the indices):

$$df = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial X} dX + \frac{1}{2} \frac{\partial^2 f}{\partial X^2} dX^2 + O(dt^2, dX^3). \tag{1.20}$$

At this point we introduce the Langevin equation (1.19) in the increments and we use the fact that $dW^2 = dt$:

$$df = (\frac{\partial f}{\partial t} + a \frac{\partial f}{\partial X} + \frac{1}{2} c^2 \frac{\partial^2 f}{\partial X^2}) dt + c \frac{\partial f}{\partial X} dW, \tag{1.21}$$

which is the celebrated Ito's formula [44]. At this point we assume that $\partial_t f = 0$ (i.e. the test function f does not depend explicitly on time) and we take the average of (1.21), recalling that $\langle dW \rangle = 0$:

$$\frac{d}{dt} \langle f \rangle = \langle a \frac{\partial f}{\partial X} + \frac{1}{2} c^2 \frac{\partial^2 f}{dX^2} \rangle. \tag{1.22}$$

We explicit the PDF $p(X,t)$ of the stochastic variable X at time t as follows:

$$\langle f \rangle = \int_S dX p(X,t) f(X),$$

where S is the state space of X. Doing some integration by parts in (1.22) and assuming that the $p(X,t)$ vanishes at the boundaries of S, we obtain the generalized FP equation [44]:

$$\partial_t p(X,t) = \sum_i \Big\{ -\frac{\partial}{\partial X_i}(a_i(X,t)p(X,t)) + \frac{1}{2}\sum_j \frac{\partial}{\partial X_i}\frac{\partial}{\partial X_j}[b_{ij}(X,t)p(X,t)] \Big\}, \quad (1.23)$$

where the generalized diffusion coefficients b_{ij} are linked to the coefficients c_{ij} introduced in (1.19) as follows:

$$b_{ij} = \sum_k c_{ik}c_{kj}. \quad (1.24)$$

1.2 Thermodynamic equilibrium and NESS

We want now to define the possible types of steady state which can occur in system [44]. We do that considering the Fokker-Planck equation (1.14), recasting it in an alternative form by means of the definition of a density current $\vec{j}$:

$$\partial_t p = -\nabla j, \quad (1.25)$$

where the vector j has components $j_i = a_i - \frac{1}{2}\sum_j \partial_j[b_{ij}p]$. At the equilibrium, the PDF $p(x,t)$ does not change, so we impose the LHS to be vanishing. This translates to the condition $\nabla j = 0$. This is the point where the difference between thermodynamic equilibrium and NESS comes up:

- If we impose the stationarity condition $\nabla j = 0$ together with $j = 0$, we are in a thermodynamic equilibrium, or equilibrium.
- If we impose the stationarity condition $\nabla j = 0$ with a non-vanishing current $j \neq 0$, we are in a NESS.

1.2.1 DBE and equilibrium

We show now that the equilibrium condition can be linked to the time reversal condition, which refers to a situation where the probability of a forward path is equal to the probability of the same path, but reversed [44]. To show this equivalence, we consider the Langevin equation of a Brownian Motion in one dimension within a conservative force with potential $U(x)$ [44]:

$$m\ddot{x} = -\tilde{\gamma}\dot{x} - \nabla U + \eta, \quad (1.26)$$

which can be recasted in the following first-order differential system:

$$\begin{cases} \dot{x} = p/m \\ \dot{p} = -\gamma p - \nabla U + \eta \end{cases} \quad (1.27)$$

where we have defined $\gamma = \tilde{\gamma}/m$. In this case, we refer to a couple of values of space coordinates and momentum (x, p) as a state α, and we consider a path as a sequence of states in time $\{\alpha_t\}_t$. The time inversal operation $\mathcal{T}$ applied to a state α works as follows:

$$\alpha^* = \mathcal{T}(\alpha) = \mathcal{T}(x, p) = (x, -p) \tag{1.28}$$

The condition of time reversal is the following (for simplicity we consider just a discrete set of time instants $\{t_i\}_{i=1...N}$ and we say that a path is a sequence of states α_i in this discrete set):

$$p(\alpha_{t_1} = \alpha_1, ..., \alpha_{t_N} = \alpha_N) = p(\alpha_{t_1} = \alpha_N^*, ..., \alpha_{t_N} = \alpha_1^*) \tag{1.29}$$

Before proving the equivalence between the time reversal property and the equilibrium condition, we remark that we can have the condition (1.29) if and only if the following set of relations is satisfied by the probabilities to be in a state p_α and the jump probabilities $W_{\alpha\beta}$ [44]:

$$W_{\beta\alpha} p_\alpha = W_{\alpha^*\beta^*} p_{\beta^*}. \tag{1.30}$$

This set of equations is called Detailed Balance Equations (DBE). This equivalence can be proven using the Kolgomorov's criterion in the context of ergodic Markov chains and it is presented in Appendix A.

In order to prove that the DBE are present at the equilibrium condition, we consider as state α a couple (x, p) and for state β the couple (x', p') such that:

$$\begin{cases} x'(t) = x(t + dt) = x(t) + \frac{p}{m} dt \\ p'(t) = p(t + dt) = p(t) + (-U'(x(t)) - \gamma p)dt + dW, \end{cases} \tag{1.31}$$

where dW is the infinitesimal increment of a Wiener process (1.18), hence it follows a Gaussian PDF with zero mean and variance $\tilde{\Gamma} dt$. At this point the jump probabilities become:

$$W_{\beta\alpha} = \delta(x' - x - \frac{p}{m} dt) \frac{1}{\sqrt{2\pi\tilde{\Gamma} dt}} \exp\left[-\frac{1}{2\tilde{\Gamma} dt} (p' - p + (\frac{\tilde{\gamma} p}{m} + U'(x))dt)^2 \right] \tag{1.32}$$

$$W_{\alpha^*\beta^*} = \delta(x - x' + \frac{p'}{m} dt) \frac{1}{\sqrt{2\pi\tilde{\Gamma} dt}} \exp\left[-\frac{1}{2\tilde{\Gamma} dt} (p' - p + (\frac{-\tilde{\gamma} p'}{m} + U'(x'))dt)^2 \right], \tag{1.33}$$

where the Dirac deltas take care of the deterministic part while the Gaussians of the stochastic part of the transition. Taking the ratio of the jump probabilities we obtain [44]:

$$\frac{W_{\beta\alpha}}{W_{\alpha^*\beta^*}} = \exp\left[-\frac{2\tilde{\gamma}}{\tilde{\Gamma}} (\frac{(p')^2}{2m} - \frac{(p)^2}{2m} + U(x') - U(x)) \right] = \tag{1.34}$$

$$= \exp\left[-\frac{2\tilde{\gamma}}{\tilde{\Gamma}} (E' - E) \right] = \exp\left[-\frac{1}{k_B T} (E' - E) \right], \tag{1.35}$$

where we have defined the energy:

$$E(\alpha) = \frac{p^2}{2m} + U(x). \tag{1.36}$$

Note that the delta functions cancel each other in the ratio (1.34), since their arguments are equal up to terms of order greater than dt, as it can be verified inserting the relation (1.31) into (1.33) ($p'\,dt$ can be substituted by $p\,dt$). In the last equality we have used the FDT (1.7). The last equality can be rewritten as the ratio $\frac{p_{\beta^*}}{p_\alpha}$ since at the equilibrium we have that the steady-state distribution is given by [44]:

$$p_\alpha = \frac{1}{\mathcal{Z}} e^{-\frac{E(\alpha)}{k_B T}} = p_{\alpha^*}, \tag{1.37}$$

where $\mathcal{Z}$ is a normalization factor. All in all, we have showed that at the equilibrium the DBE are valid, and hence the time reversal property for the trajectories holds. This fact characterizes the equilibrium condition since in a NESS the time reversal property is no more guaranteed.

1.3 Linear Response Theory

Kubo's theory has been a milestone in Response Theory. In his works [1, 2] he has addressed the problem of studying the properties of slightly perturbed systems, relating them to the properties of the unperturbed system [4, 44]. He focused on Hamiltonian systems firstly taken at the equilibrium, with Hamiltonian $\mathcal{H}_0(p, q)$ and equation of the dynamics:

$$\dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i}, \qquad \dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i}, \tag{1.38}$$

which are invariant under time reversal. At this point we perturb the system, acting directly on the Hamiltonian, which becomes time-dependent: $\mathcal{H}(p, q, t) = \mathcal{H}_0(p, q) + \mathcal{F}(t)V(p, q)$, where $\mathcal{F}$ is considered small with respect to the other terms in the Hamiltonian. Note that $V(p, q)$ does not depend in an explicit way on time t. The Hamilton equations become:

$$\dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i} - \mathcal{F}(t)K_i^q, \qquad \dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i} - \mathcal{F}(t)K_i^p, \tag{1.39}$$

where we have defined the generalized forces:

$$K_i^q = -\frac{\partial V}{\partial p_i}, \qquad K_i^p = \frac{\partial V}{\partial q_i}. \tag{1.40}$$

The phase space PDF in the unperturbed case $f_{eq}(p, q)$ is given by the Boltzmann-Gibbs formula:

$$f_{eq}(p, q) = \frac{1}{\mathcal{Z}} e^{-\beta \mathcal{H}_0(p,q)}, \qquad \beta \equiv \frac{1}{k_B T} \tag{1.41}$$

which satisfies the Liouville equation in the unperturbed case [4]:

$$\frac{\partial}{\partial t} f_{eq}(p, q) + i\mathcal{L}_0 f_{eq}(p, q) = i\mathcal{L}_0 f_{eq}(p, q) = 0, \tag{1.42}$$

where $\mathcal{L}_0$ is the Liouville operator whose action on a function $g(p, q, t)$ is given by the following Poisson brackets:

$$i\mathcal{L}_0[g] = \{g, \mathcal{H}_0\} = \sum_j \left(\frac{\partial \mathcal{H}_0}{\partial p_j} \frac{\partial}{\partial q_j} - \frac{\partial \mathcal{H}_0}{\partial q_j} \frac{\partial}{\partial p_j} \right) g. \tag{1.43}$$

In the perturbed case the phase space PDF $f(p,q,t)$ depends on time and it satisfies the perturbed Liouville equation:

$$\frac{\partial}{\partial t} f(p,q,t) + i[\mathcal{L}_0 + \mathcal{L}_{\text{ext}}] f(p,q,t) = 0, \tag{1.44}$$

where the operator $\mathcal{L}_{\text{ext}}$ is such that:

$$i\mathcal{L}_{\text{ext}}[g] = \mathcal{F}(t)\{g, V\} = \mathcal{F}(t) \sum_j \left(\frac{\partial V}{\partial p_j} \frac{\partial}{\partial q_j} - \frac{\partial V}{\partial q_j} \frac{\partial}{\partial p_j} \right) g. \tag{1.45}$$

An approximated solution of (1.44), valid up to first order in $i\mathcal{L}_{\text{ext}}$ and with initial condition $f(p,q,0) = f_{eq}(p,q)$ at $t = 0$, is the following [4]:

$$f(p,q,t) = f_{eq}(p,q) - i \int_0^t dt' e^{-i(t-t')\mathcal{L}_0} \mathcal{L}_{\text{ext}} f_{eq}(p,q). \tag{1.46}$$

We remark that we are stopping at the first order since we are developing a linear response theory, but we can consider even higher-order terms without much effort. Now we consider a generic observable $\Psi(p,q)$ and we compute its average over the phase space at time t:

$$\langle \Psi(t) \rangle = \int dpdq \, \Psi(p,q) f(p,q,t). \tag{1.47}$$

We want to compute the linear perturbation in the average value of Ψ with respect to equilibrium [4]:

$$\langle \Delta\Psi(t) \rangle = \langle \Psi(t) \rangle - \langle \Psi \rangle_{eq}. \tag{1.48}$$

To do that we substitute (1.46) into (1.48), obtaining:

$$\langle \Delta\Psi(t) \rangle = \int dpdq \, \Psi(p,q) \int_0^t dt' e^{-i(t-t')\mathcal{L}_0} \mathcal{F}(t') \{f_{eq}, V\} \tag{1.49}$$

Since f_{eq} depends on p, q just through $\mathcal{H}_0$, we have that:

$$\{f_{eq}, V\} = \{\mathcal{H}_0, V\} \frac{\partial f_{eq}}{\partial \mathcal{H}_0}, \tag{1.50}$$

where:

$$\{\mathcal{H}_0, V\} = -\{V, \mathcal{H}_0\} = -\left(\frac{dV}{dt} \right)_{t=0}, \qquad \frac{\partial f_{eq}}{\partial \mathcal{H}_0} = -\beta f_{eq}. \tag{1.51}$$

Note that in the first relation we have used the fact that $\frac{dV}{dt} = \{V, H_0\} + \frac{\partial V}{\partial t}$, in addition to the fact that V does not depend in an explicit way on t (hence $\partial_t V = 0$). As a consequence, we have that (1.49) becomes:

$$\langle \Delta\Psi(t) \rangle = \beta \int_0^t dt' \Big\{ \int dpdq \, \Psi(p,q) e^{-i(t-t')\mathcal{L}_0} \left(\frac{dV}{dt} \right)_{t=0} f_{eq}(q,p) \Big\} \mathcal{F}(t'). \tag{1.52}$$

As a last step we use the unitarity of the Liouvillian operator, obtaining:

$$\langle \Delta\Psi(t) \rangle = \beta \int_0^t dt' \mathcal{F}(t') \int dpdq \, \left(\frac{dV}{dt} \right)_{t=0} f_{eq} \, e^{i(t-t')\mathcal{L}_0} \Psi(p,q), \tag{1.53}$$

which can assume a simplified form defining a response function $\Gamma_{\Psi,V}$:

$$\langle \Delta\Psi(t)\rangle = \int_0^t dt' \mathcal{F}(t')\Gamma_{\Psi,V}(t-t'), \tag{1.54}$$

where [4]:

$$\Gamma_{\Psi,V}(t) = \Theta(t)\beta \int dpdq \left(\frac{dV}{dt}\right)_{t=0} f_{eq}\, e^{it\mathcal{L}_0}\Psi(p,q). \tag{1.55}$$

We can see in (1.54) that the effect of the time pattern of the perturbation $\mathcal{F}$ at time $\tau < t$ on the observable Ψ at time t is mediated by the response functions $\Gamma_{\Psi,V}$. We added a theta in the response function (1.55) to highlight the fact that the response function is predictive: we can predict the state of the perturbed observable Ψ at time t just using just the perturbation $\mathcal{F}(\tau)$ up to time t, as it is shown in (1.54). It is remarkable that (2.16), as (1.55), does not depend on the time pattern of the perturbation, while it depends on the observable and the space pattern of the forcing. The practical use of this fact is the possibility to compute the response function for a given time pattern, and then use it to predict the response of the given observable Ψ to another time pattern (performing the convolution of the response function with the new time pattern). We can obtain a more compact form of the response function as follows. First, we define the time-correlation functions of two observables $A(q,p)$ and $B(q,p)$ at two times $t_1 < t_2$ and at the equilibrium:

$$\begin{aligned}\langle A(t_2)B(t_1)\rangle_0 =& \int dpdq\, A(q,p))e^{-i\mathcal{L}_0(t_2-t_1)}B(q,p)f_{eq}(q,p)\\ =&\langle A(t_2-t_1)B(0)\rangle_0\\ =&\langle A(0)B(t_1-t_2)\rangle_0,\end{aligned} \tag{1.56}$$

where we have used the invariance under time translations at the equilibrium of the correlators. We can thus rewrite the response function $\Gamma_{\Psi,V}$ as follows:

$$\begin{aligned}\Gamma_{\Psi,V}(t) =&\Theta(t)\beta\langle \dot{V}(0)\Psi(t)\rangle_0\\ =&-\Theta(t)\beta\langle V(0)\dot{\Psi}(t)\rangle_0,\\ =&-\Theta(t)\beta\frac{d}{dt}\langle V(0)\Psi(t)\rangle_0\end{aligned} \tag{1.57}$$

where $\dot{V}(0) = \left(\frac{dV}{dt}\right)_{t=0}$ and $V(0) = V(q(0),p(0))$. The last formula has been obtained from (1.55), taking the adjoint of the operator $i\mathcal{L}_0$:

$$\begin{aligned}\Gamma_{\Psi,V}(t) =&\Theta(t)\beta \int dpdq\ (i\mathcal{L}_0 V)\, f_{eq}\, e^{it\mathcal{L}_0}\Psi(p,q)\\ =&\Theta(t)\beta \int dpdq\, V f_{eq}\, (i\mathcal{L}_0)^\dagger e^{it\mathcal{L}_0}\Psi(p,q).\end{aligned} \tag{1.58}$$

It is important to observe that the response function $\Gamma_{\Psi,V}(t)$, which describes the perturbed system, is obtained from an average over the unperturbed system. Lastly, we notice that we can rewrite (1.55) in another meaningful form defining the dissipative flux $j(p,q)$ as follows [4]:

$$j(p,q)f_{eq}(p,q) = -\{\mathcal{H}_0, V\}f_{eq}(p,q) = \sum_j \left(K_j^q F_j^{(0)} - K_j^p \frac{p_j}{m}\right) f_{eq}(p,q), \tag{1.59}$$

where we have used the generalized forces (1.40) and we have defined the force acting on the j-th particle in the system defined by $\mathcal{H}_0$ as $F_j^{(0)} = -\frac{\partial \mathcal{H}_0}{\partial q_j}$. We can see that due to the presence of the generalized forces the dissipative flux (1.59) is different from zero in general. Using (1.59), the response function (1.55) becomes:

$$\Gamma_{\Psi,V}(t) = \Theta(t)\beta \int dpdq\, j(p,q) f_{eq}\, e^{it\mathcal{L}_0}\Psi(p,q) = \beta\langle j(0)\Psi(t)\rangle_0. \tag{1.60}$$

1.4 KK relations

In this section we derive the KK equations [44] which are peculiar of the response function (1.57). These equations link the real and the imaginary part of the Fourier transform of the response function, requiring just the causality of the response function (1.55) (guaranteed from the presence of the Theta function). The KK equations can be regarded as a set of constraints that any response function of the type (1.55) must satisfy. First, we perform the Fourier transform of the response function $\Gamma_{\Psi,V}$, using the third form of (1.57) and denoting as $C(\omega)$ the Fourier transform of the correlation function $C(t) = \langle V(0)\Psi(t)\rangle$:

$$\Gamma_{\Psi,V}(\omega) = -\beta \int \frac{d\omega'}{2\pi}\Theta(\omega-\omega')(i\omega')C(\omega'), \tag{1.61}$$

where we used the fact that the transform of a product in the time domain is a convolution in the frequency domain. We recall now that the Fourier transform of the Heaviside theta is the following [44]:

$$\begin{aligned}\Theta(\omega) =& \lim_{\varepsilon\to 0^+}\frac{1}{\varepsilon - i\omega}\\ =&\pi\delta(\omega) + iP\left[\frac{1}{\omega}\right],\end{aligned} \tag{1.62}$$

where $P[.]$ is the principal part. As a consequence, we can express $\Gamma_{\Psi,V}$ as follows:

$$\Gamma_{\Psi,V}(\omega) = -i\frac{\beta}{2}\omega C(\omega) + \beta P\left[\int \frac{d\omega'}{2\pi}\frac{\omega'}{\omega-\omega'}C(\omega')\right]. \tag{1.63}$$

In particular, we have that:

$$\Gamma^I_{\Psi,V}(\omega) = -\frac{\beta}{2}\omega C(\omega). \tag{1.64}$$

We can deduce from (1.63) the following relation, which is the KK relation [44]:

$$\Gamma^R_{\Psi,V}(\omega) = -P\left[\int \frac{d\omega'}{\pi}\frac{\Gamma^I_{\Psi,V}(\omega')}{\omega-\omega'}\right], \tag{1.65}$$

where $\Gamma^R_{\Psi,V}$ and $\Gamma^I_{\Psi,V}$ stands for the real and imaginary part of the response function. An equivalent form of (1.65) is the following:

$$\Gamma^I_{\Psi,V}(\omega) = P\left[\int \frac{d\omega'}{\pi}\frac{\Gamma^R_{\Psi,V}(\omega')}{\omega-\omega'}\right], \tag{1.66}$$

In other words, the real and the imaginary part of the response function are linked by an Hilbert transform [13]. An alternative form for the KK relation can be found plugging (1.62) and (1.64) into (1.61):

$$\Gamma_{\Psi,V}(\omega) = -\lim_{\varepsilon\to 0^+}\int \frac{d\omega'}{\pi}\frac{\Gamma^I_{\Psi,V}(\omega')}{(\omega-\omega')+i\varepsilon}. \tag{1.67}$$

Another interesting observation about the real and imaginary part of the response function is the following. From the definition of Fourier transform of $\Gamma_{\Psi,V}$, we have that $\Gamma_{\Psi,V}(\omega) = \Gamma^\dagger_{\Psi,V}(-\omega)$. As a consequence, we have that $\Gamma^I_{\Psi,V}(\omega)$ is an odd function in ω while $\Gamma^R_{\Psi,V}(\omega)$ is an even function in ω.

It could be of interest to expand more on the case $V = \Psi = X$, since in that case the correlation function is $C(t) = \langle X(t)X(0)\rangle_0$ and the Wiener.-Khinchin theorem holds for its Fourier transform [44]. This theorem states that the following relation holds:

$$C(\omega) = \lim_{T\to\infty}\langle |X_T(\omega)|^2\rangle_0, \tag{1.68}$$

where $X_{T(\omega)} = \frac{1}{T}\int_{-T/2}^{T/2} dt e^{i\omega t}X(t)$. As a consequence we have that $C(\omega)$ is greater or equal to zero: $C(\omega) \geq 0$.

1.5 Fluctuation dissipation theorem

We want now to obtain some relations, which form the FDT, which connect the spontaneous fluctuations present in an unperturbed system at the equilibrium with the properties of the same system slightly perturbed [44]. We address this problem considering a system prepared very far in the past in the equilibrium state of the perturbed Hamiltonian $\mathcal{H} = \mathcal{H}_0 - h(t)A$, with $h(t) = h$ (assumed to be small) for $t < 0$ and $h(t) = 0$ for $t \geq 0$. What happens is that we switch off the perturbation at $t = 0$ and the system relaxes to the equilibrium state of the unperturbed Hamiltonian $\mathcal{H}_0$.

Now we consider a generic observable $X(p, q)$. Its average value up to $t = 0$ is given by [4, 44]:

$$\langle X\rangle_{eq} = \int dpdq\, f(p,q)X(p,q), \tag{1.69}$$

where $f(p, q)$ is the canonical distribution of the perturbed system:

$$f(p,q) = \frac{e^{-\beta(\mathcal{H}_0 - hA)}}{\int dpdq e^{-\beta(\mathcal{H}_0-hA)}} \tag{1.70}$$

We want to study the average value $\langle X(t)\rangle$ of the observable X during the relaxation process. During the relaxation process (for $t \geq 0$) the evolution of $X(t)$ is governed by the Hamilton equation given by the Hamiltonian $\mathcal{H}_0$, with initial condition $\{p_0, q_0\}$. This initial condition must be weighted by the canonical distribution $f(p, q)$:

$$\langle X(t)\rangle = \frac{\int dpdq\, e^{-\beta(\mathcal{H}_0 - hA(0))}X(t)}{\int dpdq\, e^{-\beta(\mathcal{H}_0-hA(0))}}, \tag{1.71}$$

where we recall that the evolution of $X(t)$ is governed by $\mathcal{H}_0$. At this point we expand the canonical weight in h up to first order:

$$\langle X(t)\rangle \approx \frac{\int dpdq\, e^{-\beta\mathcal{H}_0}(1+h\beta A(0))X(t)}{\int dpdq\, e^{-\beta\mathcal{H}_0}(1+h\beta A(0))}, \tag{1.72}$$

We divide numerator and denominator by the unperturbed partition function $\mathcal{Z}_0 = \int dpdq\, e^{-\beta\mathcal{H}_0}$ and we expand again up to first order in h, obtaining the following relation [44]:

$$\begin{aligned}\langle \Delta X(t)\rangle =& \langle X(t)\rangle - \langle X\rangle_0 \\ =& \beta h(\langle X(t)A(0)\rangle_0 - \langle X(0)\rangle_0\langle A(0)\rangle_0),\end{aligned} \tag{1.73}$$

where $\langle . \rangle_0$ denotes the average using the canonical distribution given by $\mathcal{H}_0$ and where we have replaced $\langle X(t)\rangle_0$ with $\langle X(0)\rangle_0$, since averaging over the equilibrium distribution we obtain quantities independent on time. The relation (1.73) is crucial in linear response theory. We can see from that relation that the evolution in time (related to dissipation) of a perturbed macroscopic observable $X(p,q)$ cannot be discriminated by the thermal fluctuations around the equilibrium of the same observable. In this linear regime, the time-scale of the out-of-equilibrium behavior of the observable X is proportional to the time auto-correlation function of X in the unperturbed system.

This line of thinking can be applied directly on the response function $\Gamma_{\Psi,V}$. The non-equilibrium response of the considered system to a forcing can be computed by means of a time auto-correlation function (1.57), which we recall now:

$$\Gamma_{\Psi,V}(t) = -\Theta(t)\beta\frac{d}{dt}\langle V(0)\Psi(t)\rangle_0. \tag{1.74}$$

The meaning is the same as (1.73): the response of a system slightly perturbed out of equilibrium can be fully described by a correlation function in the unperturbed system, i.e. from its fluctuations. This fact can be observed also from the KK relation (1.64), performing the inverse Fourier transform of that equation. Thanks to the Heaviside theta present in (1.57), the response function has support just for positive times, hence all its poles are located in the lower complex $\omega-$plane, i.e. they all have a negative imaginary part. These poles will be the same as the Fourier transform of the correlation function $C(\omega)$ for positive times. This connection between $C(\omega)$ and $\Gamma_{\Psi,V}(\omega)$ is basically the FDT.

Static susceptibility from the FDT

Sending $h \to 0$ in (1.73) we can obtain the static susceptibility $\chi = \frac{\partial\langle X\rangle}{\partial h}_{|h\to 0}$ of a macroscopic extensive observable X, which gives the variation of the average value of X with respect to variations of the conjugated intensive variable h in the limit $h \to 0$ [44]:

$$\begin{aligned}\chi =& \frac{\partial\langle X\rangle}{\partial h}_{|h\to 0} \\ =& \beta[\langle X^2(0)\rangle - \langle X(0)\rangle_0^2]\end{aligned} \tag{1.75}$$

If we impose $\langle X(0)\rangle = 0$ without loss of generality, we have:

$$\chi = \beta\langle X^2(0)\rangle \tag{1.76}$$

Moreover, it is remarkable that the static susceptibility can be obtained from the following limit:

$$\chi = \lim_{\omega \to 0} \Gamma_{\Psi,V}(\omega), \tag{1.77}$$

where $\Gamma_{\Psi,V}(\omega)$ is the Fourier transform of the linear response function (1.61)m with $V = \Psi = X$. The relation (1.77) is important, because it shows that the static susceptibility can be obtained considering the long time limit of the response function $\Gamma_{\Psi,V}$. Let's prove (1.77), proceeding as follows from (1.57):

$$\begin{aligned} \lim_{\omega \to 0} \Gamma_{\Psi,V}(\omega) =& -\beta \lim_{\omega \to 0} \int e^{i\omega t} \Theta(t) \frac{d}{dt} \langle X(t) X(0) \rangle_0 dt \\ =& -\beta \lim_{\omega \to 0} \lim_{\varepsilon \to 0^+} \int_0^\infty e^{i\omega t} e^{-\varepsilon t} \frac{d}{dt} \langle X(t) X(0) \rangle_0 dt, \end{aligned} \tag{1.78}$$

where we used $\Theta(t)$ to get rid of half of the integral and we regularized the integral adding a convergence factor $e^{-\varepsilon t}$. At this point we perform a partial integration, obtaining [44]:

$$\begin{aligned} \lim_{\omega \to 0} \Gamma_{\Psi,V}(\omega) =& -\beta \lim_{\varepsilon \to 0^+} \left[e^{-\varepsilon t} \langle X(t) X(0) \rangle_0 \right]_0^\infty \\ =& \beta \langle X^2(0) \rangle \\ =& \chi, \end{aligned} \tag{1.79}$$

where we have used (1.76), hence proving (1.77). Another unveiling formula for χ can be derived by means of (1.77), (1.67) and (1.64):

$$\begin{aligned} \chi =& \lim_{\omega \to 0} \lim_{\varepsilon \to 0^+} \int \frac{d\omega'}{\pi} \frac{\Gamma^I_{\Psi,V}(\omega')}{(\omega - \omega') - i\varepsilon} \\ \chi =& \beta \int \frac{d\omega}{2\pi} C(\omega), \end{aligned} \tag{1.80}$$

which shows that the susceptibility χ is related to the sum of the correlation at all frequencies ω. This relation is also called sum rule.

A particular instance of (1.75) could be determined by the choice $X = \mathcal{H}_0$, with conjugated variable $\beta = \frac{1}{k_B T}$. With that choice we have [44]:

$$-k_B T^2 C_V = \left(\langle \mathcal{H}_0^2 \rangle_0 - \langle , \mathcal{H}_0 \rangle_0^2 \right). \tag{1.81}$$

we have defined the heat capacity at constant volume C_V. This FDT connects the response to an energy perturbation (the heat capacity) to the thermal energy fluctuations of the system at the equilibrium.

Another relevant example comes up choosing as extensive variable the magnetization M, with conjugated intensive variable the magnetic field H:

$$\chi = \frac{\partial M}{\partial H}_{|h \to 0} = \beta \left(\langle M^2 \rangle_0 - \langle M \rangle_0^2 \right). \tag{1.82}$$

The FDT (1.82) can be put in local form considering an Hamiltonian $\mathcal{H}_0(m(r))$ which depends on the magnetization density $m(r)$ and perturbing it in the following way:

$$\mathcal{H}(m(r)) = \mathcal{H}_0(m(r)) - \int dr h(r) m(r), \tag{1.83}$$

Note that we have chosen to present also the constant second term for the sake of showing the meaning of (1.81) and (1.82) in a clear way. Focusing on the observable $m(r)$, it is possible to show that the following local FDT holds between two spatial points in r and r' [4]:

$$\begin{aligned}\chi(r,r') =&\beta\left(\langle m(r)m(r')\rangle_0 - \langle m(r)\rangle_0\langle m(r')\rangle_0\right)\\ =&\beta\Gamma(r,r'),\end{aligned} \tag{1.84}$$

where $\Gamma(r,r')$ is the correlation function.

1.6 Green-Kubo relations

In this section we derive the Green-Kubo relations, a set of equations within the context of response theory which allows to compute transport coefficients (non-equilibrium properties) by means of some correlations computed at the equilibrium [44]. A first relation of this kind can be derived in the contest of the FP equation (1.14). We consider just one variable, a vanishing drift coefficient and an homogeneous and constant diffusion coefficient D. Given these assumptions the FP equation becomes:

$$\partial_t p = D\partial_x^2 p. \tag{1.85}$$

We multiply (1.85) by x^2 and then we integrate over x. We obtain:

$$\partial_t\langle x^2\rangle = 2D, \tag{1.86}$$

whose result is the following, given an initial condition $x(0)$:

$$\langle x^2(t)\rangle = \langle x^2(0)\rangle + 2Dt. \tag{1.87}$$

We have that $x(t) = \int_0^t v(\tau)d\tau$, where v is the velocity. We insert this relation in (1.87), obtaining:

$$D = \langle \int_0^t v(t)v(\tau)\rangle_0. \tag{1.88}$$

Since at the equilibrium the time translation invariance holds, we obtain the following relation [44]:

$$D = \langle \int_0^t v(\tau)v(0)\rangle_0, \tag{1.89}$$

which gives the diffusion coefficient D in terms of the integral in time of the time autocorrelation function of the velocity. This is reasonable: the more correlated is the velocity in time and the more the system diffuses. It is important to remark that in practice the formula (1.89) holds in the limite of large t:

$$D = \lim_{t\to\infty}\langle \int_0^t v(\tau)v(0)\rangle_0. \tag{1.90}$$

1.6.1 Hydrodynamic approach

A more general set of Green-Kubo relations can be derived by means of an hydrodynamic approach [44]. Let's consider a space density ρ_a of some extensive conserved quantity a. We will start from two base equations. The first one is the continuity equation for a:

$$\partial_t \rho_a = -\nabla j_a, \tag{1.91}$$

where $j_a(x,t)$ is the vector density current associated to a, with components j_i. The second equation we will consider is a phenomelogical one:

$$\langle j_a(x,t)\rangle = -D_a \nabla \langle \rho_a(x,t)\rangle, \tag{1.92}$$

where the averages are non-equilibrium ones since we are considering fluctuating quantities in a situation where a density gradient is applied. We condensate (1.91) and (1.92) in the following relation, dropping the index a:

$$\partial_t \langle \rho(x,t)\rangle = D\nabla^2 \langle \rho(x,t)\rangle. \tag{1.93}$$

At this point we use the FDT (1.73), which tells us that fluctuations in the local density cannot be discriminated from the relaxation in time of the same density if the perturbation of the system is reasonably small. Using this idea, we have the following proportionality, where the average in the RHS is at the equilibrium:

$$\langle \rho(x,t)\rangle \propto \langle \rho(x,t)\rho(y,t')\rangle_0. \tag{1.94}$$

We remark that we have assumed without loss of generality $\langle \rho(x,t)\rangle_0 = 0$. Then, we define the following correlation function C and we assume space translation invariance (time space translation invariance is given by the fact that we are doing an equilibrium average):

$$C(x,t;t,t') \equiv \langle \rho(x,t)\rho(y,t')\rangle_0 = C(x-y;t-t'). \tag{1.95}$$

We insert (1.94) into (1.93), obtaining:

$$\partial_t C(x-y;t-t') = D\nabla^2 C(x-y;t-t'). \tag{1.96}$$

We solve this equation performing the Fourier transform of the correlation function:

$$C(k;t-t') = \int dx e^{ik(x-y)} C(x-y;t-t'). \tag{1.97}$$

Then we use the fact that C is space translation invariant:

$$\begin{aligned} C(k;t-t') =& \frac{1}{V}\int dx \int dy e^{ik(x-y)} C(x-y;t-t') \\ =& \frac{1}{V}\langle \rho(k,t)\rho(-k,t')\rangle_0. \end{aligned} \tag{1.98}$$

Now we go back to (1.96) in the Fourier domain, and we impose $t > 0$ and $t' = 0$:

$$\partial_t C(k;t) = -Dk^2 C(k;t), \tag{1.99}$$

which gives as solution:

$$C(k;t) = e^{-Dk^2 t} C(k;0), \tag{1.100}$$

where $k^2 = |k|^2$. Using the time translation invariance we can extend this solution to $t < 0$:

$$\begin{aligned} C(k,-t) =& \frac{1}{V}\langle \rho(k,-t)\rho(-k,0)\rangle_0 \\ =& \frac{1}{V}\langle \rho(k,)\rho(-k,T)\rangle_0 \\ =& \frac{1}{V}\langle \rho(k,)\rho(-k,T)\rangle_0 \\ =& C(k,t), \end{aligned} \tag{1.101}$$

where we have exploited the fact that $C(r,t) = C(-r,t)$ due to spatial parity. As a consequence, for any t we have that:

$$C(k;t) = e^{-Dk^2|t|}C(k;0). \tag{1.102}$$

At this point we perform the Fourier transform in time of $C(k,t)$, obtaining:

$$\begin{aligned} C(k,\omega) =& \int_{-\infty}^{\infty} dt\, e^{i\omega t}C(k,t) \\ =& C(k,0)\int_{-\infty}^{\infty} dt\, e^{i\omega t}e^{-Dk^2|t|}, \end{aligned} \tag{1.103}$$

which gives the following:

$$C(k,\omega) = C(k,0)\frac{2Dk^2}{\omega^2 + (Dk^2)^2}. \tag{1.104}$$

Now we can extrapolate by (1.104) a useful formula for the generalized diffusion coefficient, multiplying both sides for ω^2/k^2 and the taking the limit $k \to 0$ and then $\omega \to 0$ (hence infinite time and infinite space) [44]:

$$D = \frac{1}{2\phi}\lim_{\omega\to 0}\lim_{k\to 0}\frac{\omega^2}{k^2}C(k,\omega), \tag{1.105}$$

where $\phi = C(k=0;t=0)$. We can see that the transport coefficent D is obtained by the equilibrium correlation function C, by means of a double limit.

We now want to rewrite (1.105) in such a way that D can be computed starting from the density current j which appears in the continuity equation (1.91), which form in the Fourier space is the following [44]:

$$\partial_t \rho(k,t) + ikj(k,t) = 0, \tag{1.106}$$

where there is a product between the two vectors k and j. We now consider the following relation:

$$\begin{aligned} \partial_t\partial_{t'}C(k,t-t') =& \frac{1}{V}\langle \partial_t\rho(k,t)\partial_{t'}\rho(-k,t')\rangle_0 \\ =& \frac{1}{V}\sum_{i,j} k_i k_j \langle j_i(k,t)j_j(-k,t')\rangle_0. \end{aligned} \tag{1.107}$$

We now perform the Fourier transform of (1.107), then we divide by k^2 and we take the limits for vanishing k and ω:

$$\begin{aligned}\lim_{\omega\to 0}\lim_{k\to 0}\frac{\omega^2}{k^2}C(k,\omega) &= \lim_{\omega\to 0}\int_{-\infty}^{\infty} d(t-t')e^{-i\omega(t-t')}\lim_{k\to 0}\sum_{i,j}\frac{k_i k_j}{k^2}\langle j_i(k,t)j_j(-k,t')\rangle_0 \\ &= \lim_{\omega\to 0}\int_{-\infty}^{\infty} d(t-t')e^{-i\omega(t-t')}\sum_{i,j}\frac{k_i k_j}{k^2}\langle j_i^T(t)j_j^T(t')\rangle_0,\end{aligned} \tag{1.108}$$

where we have defined the total current at time t, since we are doing the limit $k \to 0$:

$$j_i^T(t) = \int dx j_i(x,t). \tag{1.109}$$

Now, assuming isotropy among the d dimensions at the equilibrium, we have:

$$\langle j_i^T(t)j_i^T(t')\rangle_0 = \frac{1}{d}\delta_{ij}\langle j^T(t)j^T(t')\rangle_0. \tag{1.110}$$

Then we plug (1.105) and (1.110) into (1.108) and we use the fact that $\sum_{i,j}\frac{k_i k_j}{k^2}\delta_{ij} = 1$, obtaining [44]:

$$D = \frac{1}{dV\phi}\int_0^{\infty} dt\langle j^T(t)j^T(0)\rangle_0, \tag{1.111}$$

where we have used the invariance under time reversal $t \leftrightarrow -t$ of the equilibrium autocorrelation function. The equation (1.111) is the known Green-Kubo relation, which connects the generalized transport coefficient D with an integral over time of the time autocorrelation function of the density current at the equilibrium. The diffusion coefficient can refer to any conserved quantity such as the mass or the energy. This is of practical interest since we can compute non-equilibrium coefficients through equilibrium averages, which can be obtained by means of numerical simulations.

1.7 Van Kampen's objection

Van Kampen moved an objection towards the Kubo's derivation of the Response formula (1.55), claiming that the effectiveness of such a formula is very limited [3]. His argument can be summarized as follows [4]. We consider a dynamical system with state $x(0)$ at time 0 and we perturb it with an instantaneous perturbation $\delta x(0)$. Then we take the difference $\delta x(t)$ between the perturbed trajectory and the unperturbed trajectory at time t, whose components are:

$$\delta x_i(t) = \sum_j \frac{\partial x_i(t)}{\partial x_j(0)}\delta x_j(0) + O(|\delta x(0)^2). \tag{1.112}$$

At this point we average over an ensemble of initial conditions:

$$\langle\frac{\partial x_i(t)}{\partial x_j(0)}\rangle = \int \frac{\partial x_i(t)}{\partial x_j(0)}\rho(x(0))dx(0), \tag{1.113}$$

The relation (1.113) consists in the linear response function of the observable x_i with respect to a perturbation of x_j:

$$\Gamma_{ij}(t) = \langle\frac{\partial x_i(t)}{\partial x_j(0)}\rangle. \tag{1.114}$$

To connect with Kubo's theory, we consider Hamiltonian systems whose stationary distribution at $t = 0$ is the canonical distribution: $\rho(x) = \frac{1}{Z}\exp(-\beta H(x))$. We can now integrate by parts within (1.114), obtaining:

$$\Gamma_{ij}(t) = \langle x_i(t)\frac{\partial H(x(0))}{\partial x_j(0)}\rangle, \tag{1.115}$$

which is the Kubo's response function (1.55). If $x_i = q_i$ and $x_j = q_j$ the perturbation potential is $V = p_j$ and the perturbed observable is x_i. To see this, it is sufficient to use the Hamilton equations (1.38) (identifying q with x), since the factor $\frac{\partial H(x(0))}{\partial x_j(0)}$ which appears in (1.115) is $-\dot{p}_j$. As a consequence we have that $V = p_j$, looking at the formula (1.55). This is consistent with the fact that the variable x_i is perturbed by an instantaneous kick at $t = 0$, since the perturbed Hamiltonian equations (1.39) for q_i with the perturbation potential $V = p_j$ and time pattern of the perturbation $\mathcal{F}(t) = \delta(t)\delta x(0)$ is the following:

$$\dot{x}_i = \frac{\partial H}{\partial p_i} + \delta x(0)\delta(t)\frac{p_j}{p_j}$$

$$\Rightarrow x_i(t) = \int^t d\tau \frac{\partial H}{\partial p_i} + \delta x(0).$$

If x_i and x_j are instead momenta coordinates, the line of thinking is the same but with different perturbation potential $V = q_j$.

The van Kampen's argument is related to(1.114) and stems from the following consideration about Lyapunov exponents. Given a dynamical system of the following form [47, 48]:

$$\dot{x} = F(x), \tag{1.116}$$

we perturb it with a small perturbation $\delta x(t)$, and the linearised dynamics become:

$$\frac{d\delta x(t)}{dt} = \frac{\partial f}{\partial x}_{|x(t)}\delta x(t), \tag{1.117}$$

with formal solution:

$$\delta x(t) = M(t, 0)\delta x(0). \tag{1.118}$$

where the matrix M is called propagator. This matrix contracts/expands the errors from time 0 to time t given by the perturbation with respect to the initial trajectory. It has been shown [48] that if the system is ergodic[1], then the following limit exists for almost all initial conditions $x(0)$:

$$\lim_{t\to\infty}(MM^\dagger)^{t/2} = \lambda_{x_0}. \tag{1.119}$$

[1]Intuitively, a system is ergodic if the evolution in time (for a sufficiently long period) of a zone of its phase space encompasses thoroughly the entire phase space. More mathematically, a transformation $T : X \to X$ which preserves the measure (with $\mu(X) = 1$) is ergodic if for every E in the phase space such that $T^{-1}(E) \subset E$, either $\mu(E) = 0$ or $\mu(E) = 1$. From the physical point of view, this is translated that, for long periods of time, the time spent by a system over a region of the phase space is equivalent to the measure of that region. As a consequence, averages over long periods of time are equal to averages over the phase space.

The logarithm of the eigenvalues of $\lambda_{x_0}$are the Lyapunov exponents, which give the rates of exponential separation between close orbits. In a chaotic system one or more of them are positive, leading to an exponential increase over time of the errors $\delta x(0)$. As a consequence the expansion (1.112) is not valid for time larger than $(1/\lambda)\ln(L/|\delta x(0)|)$, where L is the typical fluctuation size of x. Van Kampen used that fact to claim that the response formula (1.115) is valid just for extremely small perturbations $\delta x(0)$ or for very short times, making the use of response formula (1.115) limited to a very narrow range of cases.

Despite Van Kampen's argument, response theory has been successfully applied to various disciplines, such as materials science [5]. To explain that apparent contradiction, it has to be noted that van Kampen's based his argument of single trajectories, whilst response theory is founded on averages over ensembles of trajectories. While various physical observables can behave chaotically along their evolution in time. their spatial or temporal averages will be much smoother [17], dampening the exponential growth of the errors predicted by van Kampen, who observed that the "instability of the trajectories favours the stability of distribution functions". This is the key to the effectiveness of response theory.

Chapter 2

Beyond Kubo's Theory

Kubo's key results were derived in the context of weakly perturbed Hamiltonian dynamical systems which were at their thermodynamic equilibrium. Non-Hamiltonian systems, such as dissipative systems, i.e. open systems which evolution in time features a variation in energy, are out of his scope even though they can be found in a large variety of scientific fields: viscous hydrodynamics, granular materials, and the climate are a few relevant examples [4, 6]. These are systems whose steady state is a NESS, differently from the systems considered by Kubo. The extension of Kubo's response theory to such systems has been performed successfully in the next decades [4,7,10,17]. In **Section 2.1** we derive a generalized response formula which can be approached by two main approaches, as explained in the next sections. In **Section 2.2** we present a first pathway, which works on the invariant measure of the considered system, requiring its smoothness. This can be achieved by adding a stochastic term or studying a reduced phase space [20]. In this setting, the obtained formula is called the Kubo-Agarwal formula, which reduces to the Kubo formula for equilibrium systems. In **Section 2.3** we present the second approach, which focuses on chaotic dissipative deterministic system, where the measure is singular. In this context, the pioneering work of Ruelle [7, 10] introduced a response formula divided into two contributions: one which can be framed as a FDT result, and another which cannot. In other words, the natural fluctuations are not equivalent to the forced perturbations [6]. Then, we present a simplified algorithm to compute the linear response, proposed in [17], which slightly generalizes the one described in [16]. In **Section 2.4** we derive the generalized KK relations [6, 13, 14, 26], starting from the causality property of the response functions. These relations provide the possibility to compute the whole response function starting just from the knowledge of either the real or imaginary part of the Fourier transform of the response function. In **Section 2.5** we show the higher-order response theory, which allows the computation of the response of the system beyond the first order [13, 26, 49]. In **Section 2.6** we present the surrogate response theory [28]: a new recent perspective in response theory which allows us to understand to what extent we can use perturbed observables as surrogates of the perturbation to predict the future state of other observables. This viewpoint is relevant since it allows one to predict the state of the desired observables even if we lack some information on the actual perturbation acting on the systems.

2.1 General derivation of the response formula

In this section we derive a general response formula [4,17], of much more general scope than the Kubo's one (1.55). Kubo's theory is restricted to perturbations of Hamiltonian systems which were at the equilibrium, while the following response formula is valid also for perturbed systems starting in a NESS situation. We consider the setting of Markovian systems defined over regions of $\mathcal{R}^d$. This landscape of systems includes jump processes described by master equations, diffusion processes and deterministic dynamical systems. We will focus on diffusion processes, since they consider also deterministic dynamical systems in the limit of vanishing diffusion coefficient. As a consequence, our derivation works for both stochastic and deterministic systems. For simplicity we consider overdamped diffusion, with overdamped Langevin equation:

$$\dot{x}_i(t) = F_i(x(t)) + c_{ij}(x(t))\eta_j(t), \tag{2.1}$$

where $c = \{c_{ij}\}$ are coefficients linked to the generalized diffusion coefficient, $\eta = \{\eta_j\}$ is a white noise and $F(x) = \{F_i(x)\}$ describes a general dynamical system in the limit $c \to 0$. Now we define the backward generator L, which works on the observables [50,51]. This generator defines the evolution backwards in time of a generic observable Ψ, as follows. We consider the expectation value of Ψ at time t, averaging over the possible initial conditions at time 0 with stationary measure $\rho(dx)$ (which can refer to an equilibrium state or to a NESS):

$$\langle \Psi(x(t)) \rangle = \int \rho(dx)(e^{tL}\Psi), \tag{2.2}$$

where $\rho(dx)$ can also be not differentiable with respect to x, or it can be singular with respect to the volume element dx. We can restate (2.2) as follows, differentiating with respect t:

$$\frac{\partial}{\partial t}\langle \Psi(t) \rangle = \langle L\Psi(t) \rangle. \tag{2.3}$$

For the overdamped diffusion the backward generator is the following (as it can be seen in (1.22)):

$$L\Psi = F\nabla(\Psi) + \frac{1}{2}c_{ik}c_{kj}\partial_i\partial_j(\Psi). \tag{2.4}$$

Now we perturb the equation of dynamics adding a small perturbation in the drift term for $t \geq 0$. The added perturbation has as time pattern $e(t)$ and space pattern $G(x) = \{G_i(x)\}$, and we denote with $F^e = \{F_i^e\}$ the perturbed dynamical system:

$$\begin{aligned} \dot{x}_i(t) =& F_i^e(x(t)) + c_{ij}(x(t))\eta_j(t) \\ =& F_i(x(t)) + e(t)G_i(x(t)) + c_{ij}(x(t))\eta_j(t). \end{aligned} \tag{2.5}$$

We consider now the change in the expectation value of the observable Ψ due to the perturbation [17]. First, we introduce the perturbed backward generator:

$$L^h = L + e(t)L_1, \tag{2.6}$$

where:

$$L_1\Psi = G\nabla(\Psi). \tag{2.7}$$

The change in the expectation value of Ψ is the following [17]:

$$\langle \Psi(t) \rangle^e - \langle \Psi(t) \rangle = \int \rho(dx)(e^{tL^h} - e^{tL})\Psi(x). \tag{2.8}$$

We use now the variation of parameters formula [52, 53] [1]:

$$e^{tL^h} - e^{tL} = \int \rho(dx) \int_0^t ds\, e^{sL^h}(e(s)G(x)\nabla)e^{(t-s)L}\Psi(x). \tag{2.9}$$

We can substitute the factor e^{sL^h} with the unperturbed evolution operator e^{sL}, introducing correction terms of order e^2 and beyond:

$$e^{tL^h} - e^{tL} = \int \rho(dx) \int_0^t ds\, e^{sL}(e(s)G(x)\nabla)e^{(t-s)L}\Psi(x) + \mathcal{O}(e^2). \tag{2.10}$$

Substituting (2.10) into (2.8) and truncating up terms of the first order, we obtain:

$$\langle \Psi(x(t)) \rangle^e - \langle \Psi(x(t)) \rangle = \int \rho(dx) \int_0^t ds\, e^{sL}(e(s)G(x)\nabla)e^{(t-s)L}\Psi(x). \tag{2.11}$$

Then we have that the linear change in the expectation value of the observable Ψ is:

$$\delta\langle \Psi(x(t)) \rangle = \int_0^t ds\, e(s) \int \rho(dx) e^{sL}(G(x)\nabla)e^{(t-s)L}\Psi(x) \tag{2.12}$$

Looking at (2.12) we can notice that we can define a response function (where the observable Ψ evolves for a time $(t-s)$ and then the operator $G\nabla$ is applied to that evolved observable) [17]:

$$\Gamma_{\Psi,G}(t,s) = \int \rho(dx) e^{sL}(G(x)\nabla)e^{(t-s)L}\Psi(x), \tag{2.13}$$

which is such that:

$$\delta\langle \Psi(x(t)) \rangle = \int_0^t ds\, e(s)\Gamma_{\Psi,G}(t,s). \tag{2.14}$$

[1] The formula can be quickly proven as follows [53] in a more general context (we will not enter into the mathematical details). Let's consider two continuous semigroups $\{T(t)\}_{t\geq 0}$ with generator A and $\{S(t)\}_{t\geq 0}$ with generator C on the Banach space X such that $C = A + B$ for some bounded operator B. Then we have that $S(t)x = T(t)x + \int_0^t ds\, T(t-s)BS(s)x$. To prove this statement let's define the functions $\chi_x(s) \equiv T(t-s)S(s)x \in X$, for $s \in [0,t]$. We have that:

$$\frac{d}{ds}\chi_x(s) = T(t-s)CS(s)x - T(t-s)AS(s)x = T(t-s)BS(s)x.$$

As a consequence we have that:

$$\begin{aligned} S(t)x - T(t)x = \chi_x(t) - \chi_x(0) = \int_0^t \frac{d}{ds}\chi_x(s)\, ds &= \int_0^t ds T(t-s)BS(s)x, \\ &= \int_0^t ds S(s)BT(t-s)x, \end{aligned}$$

proving the desidered statement for each $x \in X$. Choosing $T(t) = e^{tL}$ and $S(t) = e^{tL^h}$ we obtain (2.9).

We observe that since we have assumed that ρ is stationary at time zero, we can further simplify the expression of the response function (2.13). We define the adjoint operator $\mathcal{L}$ of the backward generator L. Then we rewrite the expression (2.13) using $\mathcal{L}$:

$$\Gamma_{\Psi,G}(t,s) = \int \left(e^{s\mathcal{L}}\rho(dx)\right)(G(x)\nabla)e^{(t-s)L}\Psi(x). \tag{2.15}$$

Since ρ is the steady-state distribution, we have that it does not evolve under the operator $e^{s\mathcal{L}}$ (also called transfer operator [55]). In other words, ρ is the eigenvector with unitary eigenvalue related As a consequence, the response function becomes [17]:

$$\Gamma_{\Psi,G}(t) = \Theta(t)\int \rho(dx)(G(x)\nabla)e^{tL}\Psi(x), \tag{2.16}$$

hence the response function $\Gamma_{\Psi,G}(t,s)$ depends only on the difference $(t-s) > 0$. Similarly to (1.55), the response function $\Gamma_{\Psi,G}(t,s)$ connects the perturbation at time s with the perturbed observable Ψ at time $t > s$. Performing a Fourier transform, the relation (2.16) becomes:

$$\delta\langle\Psi\rangle(\omega) = e(\omega)\Gamma_{\Psi,G}(\omega). \tag{2.17}$$

00since the Fourier transform of a convolution is a product. The obtained formula (2.16) provides a method to compute the non-equilibrium response of a system to a perturbation by means of an equilibrium average. The response function defined in (2.16) has support just for positive times. This is equivalent to the fact that the poles of its Fourier transform are all located in the lower complex $\omega-$plane, i.e. they all have negative imaginary part. This fact is guaranteed by the stability of the considered system [28]. To see this, we remark that the backward operator L must have all eigenvalues σ_k with real part negative to have the system stable. In fact, if there is an eigenvalue with positive real part, the evolution in time of the generic observable $\langle\Psi(t)\rangle$ would explode. The same applies to the response function (2.16): given a perturbation, we would have an increasing in time response. Given that, we proceed taking the Fourier transform of (2.16), obtaining [28, 54]:

$$\Gamma_{\Psi,G}(\omega) = \sum_{k=1}^{\infty}\frac{\alpha_k\{\Psi,G\}}{i\omega+\sigma_k}, \tag{2.18}$$

where $\alpha_k\{\Psi,G\}$ are factors which depend on Ψ and G. The derivation of such a formula entails the decomposition of the evolution operator $\exp(tL)$ along its eigenvectors, with eigenvalues $\exp(t\sigma_k)$, and the Fourier transform of such eigenvalues $\exp(t\sigma_k)$. Remarkably, since $\mathrm{Re}[\sigma_k] < 0$, we have that the poles $\omega_k = i\sigma_k$ all have imaginary part negative, as claimed before. This equivalence between stability of the system and causality of the response function can be see more clearly in the context of linearized system in Chapter 3, in the relation (3.9). We can thus say that the response function is predictive: the state of the perturbed observable Ψ at time t can be predicted using just the perturbation $e(\tau)$ up to time t, as it is shown in (2.14). Moreover, note that the response function (2.16) does not depend on the time pattern of the perturbation, while it depends on the observable and the space pattern of the forcing. In the following two sections we provide a variety of different approaches to the key formula (2.16).

2.2 Working on the measure: the Kubo-Agarwal formula

If we look at the response function $\Gamma_{\Psi,G}$ we can observe that we can make further manipulations to it if the measure ρ is smooth with respect the reference volume element: $\rho(dx) = \rho(x)dx$. This happens adding any small amount of noise in the equation of the dynamical system [56–58]. The addition of the beneficial noise term has to be justified by the nature of the considered problem. This stochastic perspective becomes relevant in many complex systems, where the focus is on a coarse grained dynamics (mesoscopic or macroscopic scale), which is effectively stochastic as a result of the presence of microscopic degrees of freedom [20]. The basic idea can be showed as follows. Let's consider a linear system of two ODEs [34, 59]:

$$\begin{aligned} \dot{x} &= a_{11}x + a_{12}y \\ \dot{y} &= a_{21}x + a_{22}y \end{aligned} \tag{2.19}$$

Now we want to focus on the detailed behavior of $x(t)$, considering the dynamics of $y(t)$ just in its statistical impact on the dynamics of $x(t)$. This is relevant in cases where y is a variable which fluctuates faster than x, which is the case when y is a degree of freedom in a smaller scale than x. We can proceed solving the second formula in (2.19) for y:

$$y(t) = e^{a_{22}t}y(0) + \int_0^t e^{a_{22}(t-s)}a_{21}x(s)ds, \tag{2.20}$$

and then plugging (2.20) into the first equation in (2.19), obtaining:

$$\dot{x}(t) = a_{11}x + \int_0^t K(t-s)x(s)ds + f(t), \tag{2.21}$$

where $f(t) = y(0)a_{12}\exp(a_{22}t)$ and $K(t) = a_{12}a_{21}\exp(a_{22}t)$. The relation (2.21) is a generalized Langevin equation (GLE) where the second term is a memory term (and the kernel K is a memory kernel) and the third term is the noise term, since the factor $y(0)$ is thought to be sampled from the fast variable $y(t)$. The Mori-Zwanzig formalism extends this idea, considering a general set of Markovian equation of the dynamics with a large number of degrees of freedom (even an infinite number of them) and focusing on a single variable, projecting the dynamics of the other variables in a GLE which describes its evolution in time. Note that the coarse-grained dynamics is in general non-markovian, with memory effect becoming negligible in the limit of infinite time scale separation between the fast and slow variables [21–25]. In [20] the authors argue that the projection of the singular measure of chaotic non-Hamiltonian systems along a certain direction is generally smooth (after a coarse-graining procedure), apart for a few particular cases in very simple low-dimensional systems. This procedure can be summarized as follows. Given an invariant measure $\rho(dx)$, defined over a $d-$ dimensional phase space, we $\varepsilon-$ partition it in a finite set of $d-$ dimensional hypercubes $\Lambda_k(\varepsilon)$ centered in x_k. At this point the coarse grained invariant density $\rho(x)_\varepsilon$ becomes the following:

$$\rho(x)_\varepsilon = \sum_k \rho(x)_{\varepsilon,k}, \tag{2.22}$$

where:

$$\rho(x)_{\varepsilon,k} = \begin{cases} \frac{1}{\varepsilon^d}\int_{\Lambda_k(\varepsilon)} d\mu(x), & x \in \Lambda_k(\varepsilon) \\ 0, & \text{else} \end{cases} \tag{2.23}$$

We can now project this coarse grained invariant density along a certain direction i marginalizing over the other degrees of freedom. We call Z_i the number of bins of form $[x_i^{(q)} - \varepsilon/2, x_i^{(q)} + \varepsilon/2)$, for $q \in \{1, ..., Z_i\}$. We compute now the probability density to be in $x_i^{(q)}$, by means of the probability to be in one of these Z_i bins:

$$\rho(x_i^{(q)})_\varepsilon = \frac{1}{\varepsilon} \int_{x_i^{(q)}-\varepsilon/2}^{x_i^{(q)}+\varepsilon/2} \left\{ \int \rho(x)_\varepsilon \prod_{j \neq i} dx_j \right\} dx_i, \tag{2.24}$$

which is expected to be smoother than the previous distribution $\rho(x)_\varepsilon$. This procedure can be repeated for the perturbed distribution at any instant of its evolution in time. As the number of degrees of freedom increases, this projection procedure is expected to give smoother results. This is the case of macroscopic systems in non-equilibrium steady states, since the dynamics of the observables of interest takes place in reduced phase space, whose dimensions is extremely smaller than the one of the entire phase space. In this case, the computation of the response can be safely carried out considering the distribution ρ to be smooth.

Now that we have reviewed were the smoothness assumption of the invariant measure ρ is valid, let's see the consequences of such an assumption. If ρ is smooth we can do a partial integration step within (2.16). Let's start rewriting (2.16) introducing the operator $L_1 = G\nabla$, i.e. the perturbation of the backward generator:

$$\Gamma_{\Psi,G}(t) = \Theta(t) \int \rho(dx) L_1 e^{tL} \Psi(x). \tag{2.25}$$

At this point we define the adjoint operator $\mathcal{L}_1$ of the operator L_1: $\int dx(\mathcal{L}_1\phi_1)\phi_2 = \int dx\phi_1(L_1\phi_2)$, for two smooth observables ϕ_1 and ϕ_2. The adjoint operator of the backward generator is related to the forward evolution in time. Moreover, the adjoint $\mathcal{L}$ of the backward generator L of the diffusion (2.4) is the operator which defines the FP equation related to the diffusion process [45]:

$$\partial_t \rho = \mathcal{L}\rho, \tag{2.26}$$

where:

$$\mathcal{L}\rho = -\nabla(F\rho) + \partial_i\partial_j(\frac{1}{2}c_{ik}c_{kj}\rho). \tag{2.27}$$

As a consequence the form of the adjoint operator $\mathcal{L}_1$, for the perturbation of the deterministic part in (2.5) is the following:

$$\mathcal{L}_1\rho = -\nabla(G\rho). \tag{2.28}$$

Given the form (2.28) of $\mathcal{L}_1$, we perform a partial integration within (2.16):

$$\begin{aligned} \Gamma_{\Psi,G}(t) =& \Theta(t) \int dx(\mathcal{L}_1\rho(x))e^{tL}\Psi(x) \\ =& -\Theta(t) \int dx\nabla(G\rho)e^{tL}\Psi(x). \end{aligned} \tag{2.29}$$

which can be written in a more compact form:

$$\Gamma_{\Psi,G}(t) = \Theta(t)\langle\frac{\mathcal{L}_1\rho}{\rho}(0)\Psi(t)\rangle. \tag{2.30}$$

The relation (2.30) is called the Kubo-Agarwal formula, and its derivation does not require the unperturbed system to be in a NESS or at the thermodynamic equilibrium; the only hypothesis is the smoothness of the measure ρ. For completely deterministic systems, (2.30) is associated with equilibrium [10] since the smoothness of ρ is a feature of equilibrium states in that kind of systems. We can see that the FDT holds, as in the Kubo's formula (1.74): the time correlation function of two suited observables fully describes the forced response of a system to a weak perturbation. The fact that the FDT holds for both equilibrium systems and NESS (with the main assumption of the smoothness of ρ) can be fundamentally explained by the fact that a thermodynamic theory can be constructed for both systems. In particular, the thermodynamics theory for the NESS is called Steady State Thermodynamics (SST) and it has been introduced in the seminal work [60]. In that paper, the notion of entropy has been extended to include also a contribution, called excess entropy, whose existence discriminate the NESS from the equilibrium systems.

We can observe that the Kubo-Agarwal formula (2.30) is more general than the Kubo's formula thanks to the fact that no specific form is assumed for the distribution ρ and hence is a generalization of the Kubo's formula (2.29) to general non-Hamiltonian systems at the equilibrium. To recover the Kubo's formula, we can consider a smooth canonical distribution $\rho = \frac{1}{Z}\exp(-\beta U)$ with $F = -\nabla U$ and coefficient $c(x) = \sqrt{2/(k_B\beta)}$ (related to an equilibrium state) and with perturbation $G = -\nabla V$, we have that:

$$\frac{\mathcal{L}_1\rho}{\rho} = -\beta\nabla V\cdot\nabla U + \nabla^2 V = \beta LV, \tag{2.31}$$

where we have used the backward generator (2.4) of the diffusion process. As a consequence (2.30) can be rewritten as follows:

$$\begin{aligned}\Gamma_{\Psi,G}(t) =& \Theta(t)\beta\langle(LV)(0)\Psi(t)\rangle\\ =& \Theta(t)\beta\langle\dot{V}(0)\Psi(t)\rangle,\end{aligned} \tag{2.32}$$

which is the Kubo's formula (1.55) for the response function.

We remark that Kubo-Agarwal formula (2.30) can be derived in the same way as we derived the more general (2.16), working with the adjoint operator $\mathcal{L}$ in the FP formalism [4, 61]. The idea is the same: we consider the unperturbed FP equation (2.26) (concerning ρ) and the FP perturbed with (2.28) (concerning ρ^e). The solution of the latter equation is the following (following a very similar approach to the one which leads to the equation (2.11)):

$$\rho^e(x,t) = \rho(x) - \int_0^t ds\, e^{(t-s)\mathcal{L}} e(s)\nabla(G(x)\rho(x)) + O(e^2). \tag{2.33}$$

Considering a general observable Ψ we can compute the difference $\delta\langle\Psi(t)\rangle$ between the expectation value using the perturbed measure $\rho^e(x,t)$ and the expectation value using the unperturbed measure $\rho(x)$, obtaining the response function (2.30). We remark that this derivation of the Kubo-Agarwal formula, which is the adjoint version to the one related to (2.16), is only possible because we have assumed that the steady state distribution ρ is smooth, which allows the Taylor expansion (2.33) and the its truncation at first order.

2.2.1 Link with Information Theory

It is possible to relate the Kubo-Agarwal formula (2.30) to information theory quantities, in particular to the information potential $\mathfrak{I}$:

$$\mathfrak{I} \equiv -\log \rho. \tag{2.34}$$

To see how to explicit that relation, let's proceed as follows [18, 62]. Without loss of generality, we consider a perturbation which does not depend on time: $e(t) = e$. Then we consider the consider the perturbed system with stationary density ρ^e and perturbed generator L^e. We rewrite the linear perturbation of the stationary density simply as:

$$\rho^e(x) = \rho(x) + h\rho_1(x) + ..., \tag{2.35}$$

where ρ_1 expresses the linear perturbation of the invariant measure. Since ρ^e is stationary, we have that $\mathcal{L}^e\rho^e = 0$, where $\mathcal{L}^e$ is the adjoint of the perturbed backward generator L^e. As a consequence, we have that:

$$\mathcal{L}_1\rho + \mathcal{L}\rho_1 = 0. \tag{2.36}$$

We can use the relation (2.36) to re-express the Kubo-Agarwal formula as follows:

$$\Gamma_{\Psi,G}(t) = -\Theta(t)\langle\left(\frac{\mathcal{L}\rho_1}{\rho}\right)(0)\Psi(t)\rangle. \tag{2.37}$$

Now we take into consideration the information potential $\mathfrak{I}^e$ (2.34) for the invariant measure of the perturbed system ρ^e and we derive it with respect to e (we denote with ∂_e the partial derivation with respect to e, taken for $e = 0$):

$$\begin{aligned}\partial_e\mathfrak{I}^e =& -\partial_e \log \rho^e \\ =& -\lim_{e\to 0}\frac{\rho^e - \rho}{e\rho^e} \\ =& \frac{\rho_1}{\rho}.\end{aligned} \tag{2.38}$$

We now plug (2.38) into (2.37), obtaining the following relation:

$$\begin{aligned}\Gamma_{\Psi,G}(t) =& \Theta(t)\langle\left(\frac{\mathcal{L}(\rho\partial_e\mathfrak{I}^e)}{\rho}\right)(0)\Psi(t)\rangle \\ =& \int dx\mathcal{L}(\rho(x)\partial_e\mathfrak{I}^e)(x)\Psi(x(t)) \\ =& \int dx\rho(x)\partial_e\mathfrak{I}^e(x)L\,e^{tL}\Psi(x) \\ =& \frac{d}{dt}\langle\partial_e\mathfrak{I}^e(0)\Psi(t)\rangle,\end{aligned} \tag{2.39}$$

which makes explicit the relation between the information potential and response theory. Note that the formula (2.39) is valid both for NESS and for equilibrium systems, since it is a rewriting of the Kubo-Agarwal formula (2.30). An advantage of (2.39) is that detailed information on the dynamics is not needed (on the contrary of (2.30), where the operator $\mathcal{L}_1$ explicitly appears). A very practical use of (2.39) could be to

use parametrized versions of the stationary density ρ (for example, quasi-Gaussian approximations) in order to get a formula for the information potential $\mathcal{I}$. Note that the information potential (2.34) can be linked with the Shannon entropy as follows [63]:

$$\begin{aligned} S &= -\int dx \rho(x) \log \rho(x) \\ &= \int dx \rho(x) \mathcal{I}(x) \\ &= \langle \mathcal{I} \rangle. \end{aligned} \tag{2.40}$$

As a consequence, it is possible to directly relate the Kubo-Agarwal response formula (2.30) with a notion of entropy by means of the information potential $\mathcal{I}$.

2.2.2 Breaking time-reversal symmetry (NESS)

In an equilibrium state, we have seen that the distribution of trajectories is time-symmetric: the probability of a path is equal to the probability of the reversed path (1.29). In a NESS there is generally a breaking of the time-reversal property of the trajectories. We observe now that the Kubo's formula (1.55) is related equilibrium showing that the Kubo-Agarwal formula (2.16) differs from the Kubo's one by a term which is not time-reversal symmetric.

First, we introduce the state-space velocity u, defining it from the stationary density current j_ρ which solves the continuity equation in the steady state:

$$\partial_t \rho = -\nabla j_\rho = 0. \tag{2.41}$$

The state-space velocity u is defined as follows:

$$u \equiv \frac{j_\rho}{\rho}, \tag{2.42}$$

which becomes for the diffusion process with generator (2.27):

$$u_i = F_i - \frac{\partial_j (\frac{1}{2} c_{ik} c_{kj} \rho)}{\rho}. \tag{2.43}$$

We consider now two general smooth observables Ψ_1 and Ψ_2 and we consider their time correlations $\langle \Psi_1(0)\Psi_2(t) \rangle$ and $\langle \Psi_1(0)\Psi_2(-t) \rangle$, where the averages are over the steady state distribution ρ . These steady-state time-correlations are invariant under time translations, but they are equal just in an equilibrium state. We have already defined the operator L which generates the time-forward motion:

$$\langle \Psi_1(0)\Psi_2(t) \rangle = \int \rho(dx) \Psi_1(0) e^{tL} \Psi_2(0). \tag{2.44}$$

In addition to that, we define the operator L^* which generates the time-reversed motion:

$$\langle \Psi_1(0)\Psi_2(-t) \rangle = \int \rho(dx) \Psi_1(0) e^{tL^*} \Psi_2(0). \tag{2.45}$$

We observe that detailed balance condition, i.e. the time-reversal property, holds for $L = L^*$. An explicit form for L^* can be found in the following way. Since time-translation invariance holds for such correlations, we have that:

$$\int \rho(dx)(L\Psi_1)(x)\Psi_2(x) = \int \rho(dx)\Psi_1(x)(L^*\Psi_2)(x), \tag{2.46}$$

from which we have that: $L^*(\Psi) = \mathcal{L}(\rho\Psi)/\Psi$ for any smooth Ψ. In the case of diffusion processes (2.1), we have that (using (2.27) and (2.43)):

$$\begin{aligned} L^* =& L - 2u\nabla \\ =& -F\nabla + \partial_i(b_{ij}\partial_j) + 2\,b\partial_i(\log\rho)\partial_j, \end{aligned} \tag{2.47}$$

where we have shortened the notation defining the matrix $b = \{b_{ij}\} = \frac{1}{2}c_{ik}c_{kj}$. Now we observe that if we choose in the Kubo-Agarwal formula the following identification:

$$\frac{\mathcal{L}_1\rho}{\rho} = -\beta L^* V, \tag{2.48}$$

the Kubo-Agarwal formula (2.30) reduces to the Kubo's formula (1.57):

$$\begin{aligned} \Gamma_{\Psi,G}(t) =& -\Theta(t)\beta\langle(L^*V)(0)\Psi(t)\rangle \\ =& -\Theta(t)\beta\langle V(0)L\Psi(t)\rangle. \end{aligned} \tag{2.49}$$

As observed before, we have that $L = L^*$ at equilibrium. As a consequence, we can identify corrections to the Kubo's formula making a different substitution:

$$\begin{aligned} \frac{\mathcal{L}_1\rho}{\rho} =& -\beta\frac{(L+L^*)}{2}V \\ =& -\beta L^*V - \frac{(L-L^*)}{2}V, \end{aligned} \tag{2.50}$$

which reduces to the (2.48) in an equilibrium state. The second term is the non-equilibrium part in the Kubo-Agarwal formula (2.30) since it is non-invariant under time reversal. We can introduce the state velocity u in the Kubo-Agarwal formula for diffusion processes using the relation (2.47), obtaining the following response function:

$$\Gamma_{\Psi,G}(t) = -\Theta(t)\beta\langle\dot{V}(0)L\Psi(t)\rangle - \beta\langle[u\nabla V](0)A(t)\rangle. \tag{2.51}$$

We can see from (2.51) that the Kubo's relation form is restored moving in the Lagrangian frame with relative drift velocity u with respect to the observer:

$$\frac{d}{ds} \to \frac{d}{ds} - u\nabla.$$

This transformation, making the observer move with a velocity proportional to the stationary density current j_ρ defined in (2.41), removes the non-equilibrium part obtaining an equilibrium-like response function.

2.3 Working on the observable: Ruelle's approach

2.3.1 The setting: chaotic dissipative systems

In this section we examine how to approach the response formula (2.16) in the case the measure ρ is not smooth. This is the case of chaotic dissipative deterministic systems. The steady state of this class of system is generally a NESS, since the time-reversal symmetry of the trajectories is broken by the presence of dissipation. Let's see how the dissipation enters in our description. We consider the following deterministic system:

$$\dot{x}(t) = F(x). \tag{2.52}$$

where the deterministic system F is t-independent. The phase space is called Γ. The related continuity equation, which describes the time evolution of PDF ρ, is the following:

$$\begin{aligned} \partial_t \rho(x,t) &= -\nabla(F\rho) \\ &= -(\nabla F)\rho - F(\nabla\rho). \end{aligned} \tag{2.53}$$

Now there are two possibilities [64]:

- The velocity field F is incompressible: $\nabla F = 0$. As a consequence the phase space volume is conserved along trajectories, since [65]:

  $$\frac{D\rho}{Dt} = \frac{\partial\rho}{\partial t} + (\nabla\rho)F = 0,$$

  where we have inserted the continuity equation (2.53). This class of systems is called conservative. The Hamiltonian systems considered by Kubo are conservative, as it can be shown using the Hamltonian equations[2].

- If the velocity field F is compressible, hence if $\nabla F < 0$, we are referring to dissipative dynamical systems[3]. In this kind of systems, phase space volumes are contracted along the trajectories, since:

  $$\frac{D\rho}{Dt} < 0.$$

  As a consequence, the set of points reached at the steady state by the trajectories forms a space of dimension smaller than the dimension of the whole phase space Γ. This set of points is called an attractor. There could be more than one attractor in a system and we call basin of attraction the set of point whose evolution goes into one of these attractors. The geometry of the attractors can be quite peculiar and not smooth. Usually, the geometry of the attractor is fractal [64] and we refer to these attractors as "strange". On the contrary, conservative systems evolve occupying the whole Γ, hence they do not possess an attractor. We remark that the contraction of phase space is related to a non-zero entropy production in time (which is linked to the asymmetry between the trajectories and their reversals), but we will not further discuss this link [10, 64, 66].

[2] For an Hamiltonian system with Hamiltonian $\mathcal{H}$ we have that the dynamical system related to the coordinates (q,p) is given by the Hamiltonian equations: $(\dot{q},\dot{p}) = (\partial_p\mathcal{H}, -\partial_q\mathcal{H})$. If we compute the divergence of the dynamical system, we have $\nabla F = \partial_q\partial_p\mathcal{H} - \partial_p\partial_q\mathcal{H} = 0$. As a consequence, Hamiltonian systems are conservative.

[3] The case $\nabla F > 0$ is not that interesting because it describes a motion without boundaries in the phase space.

We will focus on dissipative dynamical systems, taken in their chaotic regime. In this setting, there is a very high sensitivity to the initial conditions. This is described by the fact that at least one Lyapunov exponent (1.119) is greater than zero, hence small variations $\delta x(0)$ at time 0 increase exponentially in time. In these chaotic dissipative regimes, there is no invariant measure ρ absolutely continuous with respect to the Lebesgue measure, since their support are the strange attractors and hence they are usually singular with respect to Lebesgue. Note that instead in conservative dynamical systems the invariant measure is ergodic and absolutely continuous with respect to the Lebesgue measure: $\rho(dx) = \rho(x)dx$, since there is no fractal attractor within the phase space.

SRB and Axiom A systems

In order to well define an ergodic measure for chaotic dissipative dynamical systems, among all the possible ones, it is interesting to introduce the notion of Sinai-Bowen-Ruelle (SRB) measure [7,64]. A point x in the phase space Γ is called hyperbolic if its associated tangent space T_x can be decomposed in the following direct sum:

$$T_x = E_x^u \oplus E_x^s \oplus E_x^0, \tag{2.54}$$

where E_x^u is the unstable subspace, E_x^s the stable subspace and E_x^0 the neutral subspace, defined as follows. If a tangent vector $z(0)$ belongs to:

- E_x^u, there exists $K > 0$ and $\alpha \in (0,1)$ such that $z(-t) \leq K\alpha z(0)$, i.e. the backward time-evoluted vector $z(t)$ expands exponentially.
- E_x^s, there exists $K > 0$ and $\alpha \in (0,1)$ such that $z(t) \leq K\alpha z(0)$, i.e. the forward time-evoluted vector $z(t)$ expands exponentially.
- E_x^0, we have that $z(\pm t)$ is bounded and finite at any time t. For dynamical systems $F(x)$ which are not dependent on t the direction tangent to the trajectory does not shrink or expand along the trajectory, hence it belongs to the neutral subspace. For ODE it is possible to show that E_x^c is one dimensional [64].

If the points of the strange attractor of a chaotic dissipative dynamical system are hyperbolic, we can expect that the unstable stretching directions are traversed by rough discontinuous stable directions. At this point, we say that the ergodic invariant measure ρ of a dynamical system is an SRB measure if and only if it smooth along the unstable directions, while it could be singular along the stable and neutral directions [8,9,66]. In a sense, the SRB measure generalizes the invariant measure in conservative systems, which was smooth along all the directions. We can argue that the SRB measure identifies mathematically a NESS in chaotic dissipative dynamical systems.

Now we introduce two classes of systems [10,64], based on the notion of hyperbolicity. First we define the hyperbolic set K as a set of points within the phase space such that all points in K are hyperbolic and the splitting into the stable, unstable and neutral directions is continuous over K (uniform hyperbolicity). At this point we define the Asonov systems as systems defined by a smooth F over a smooth manifold which is a hyperbolic set. Next, we define a generalization of the Asonov systems: if a system defined by a dissipative and smooth F has an attractor K which is hyperbolic and its periodic orbits are dense in K, we call it an Axiom A system. It is possible to

show in a rigorous way that Axiom A systems have exactly one SRB measure, and hence they are of great interest. On the other hand, Axiom A systems are a particular class of dynamical system and they are far to represent all the dynamical systems (a not Axiom A system can be found for example in [10]). However, the chaotic hypothesis proposed by Gallavotti and Cohen [11, 12] states that Axiom A systems are practically indistinguishable from the effective properties of macroscopic observables in high-dimensional systems which display a sufficient degree of chaotic behaviour. This hypothesis fundamentally explains the effectiveness of Ruelle's approach (presented below), which is based on SRB measures.

2.3.2 RT derivation

We will present now the derivation of the general formula (2.16) under the perspective of Ruelle [67]. We consider a chaotic dissipative deterministic system (2.52) with invariant ergodic measure ρ given by a SRB measure. We have seen before the characterization of the SRB measures in terms of their smoothness along the unstable directions. Other equivalent characterizations of SRB measures are the following [66]:

- ρ is the zero-noise limit of small random perturbation of the dynamical system F. In other words, let's take a family of Markov chains P^ε, with $\varepsilon > 0$, such that $P^\varepsilon \to \delta_F$, i.e. the Dirac delta which dictates the deterministic time evolution given by F. If the invariant measure for P^ε tends to ρ, we have that ρ is a SRB measure.

- ρ is the long-time limit of an absolutely continuous measure m under the time evolution given by F, For every smooth observable Ψ this condition is given by:

$$\lim_{s\to-\infty} \int dx\, m(x)\Psi(f^{-s}x) = \int \rho(dx)\Psi(x), \tag{2.55}$$

  where f is the evolution operator such that:

$$x(s,t) = x(t-s) = f^{t-s}x(0). \tag{2.56}$$

  where we used the fact that the dynamical system F is time independent. In other words, the absolutely continuous measure m evolves and shrinks in a singular SRB measure ρ.S

Now we perturb the dynamical system (2.52) with a perturbation with time pattern $e(t)$ and space pattern $G(x)$, as done previously in (2.5). The consequence on the dynamical variables x is the following:

$$\begin{aligned} \delta x(t,s) =& \delta \int_s^t d\tau F(x(\tau,s)) \\ =& \int_s^t d\tau (T_{x(\tau,s)} f^{t-\tau}) e(\tau) G(x(\tau,s)). \end{aligned} \tag{2.57}$$

where $T_x f^\sigma$ is the tangent map at x to f^σ (in canonical coordinates it is the matrix of partial derivatives). We consider now a generic smooth observable Ψ. Its expectation

value at time t changes as follows:

$$\begin{aligned}\delta\langle\Psi\rangle(t) =& \delta \lim_{s\to-\infty} \int dx\, m(x)\Psi(x(t,s)) \\ =& \lim_{s\to-\infty} \int dx\, m(x)\delta x(t,s)\nabla_{x(t,s)}\Psi.\end{aligned} \tag{2.58}$$

Now we plug (2.57) inside (2.58) and we use the notation $(f^*m)(\phi) = m(\phi \circ f)$ (for a generic smooth ϕ):

$$\delta\langle\Psi\rangle(t) = \lim_{s\to-\infty} \int_s^t d\tau \int ((f^{\tau-s})^*m)(dy)((T_y f^{t-\tau})e(\tau)G(y))\nabla_{y(t,\tau)}\Psi, \tag{2.59}$$

where we substituted $x(\tau, s) = y$. Using the SRB measure ρ property (2.55) and the fact that F is time independent (2.56) we have that:

$$\delta\langle\Psi\rangle(t) = \int_{-\infty}^t d\tau \int \rho(dy)((T_y f^{t-\tau})e(\tau)G(y))\nabla_{y(t-\tau)}\Psi, \tag{2.60}$$

which can be rewritten in the following equivalent form:

$$\delta\langle\Psi\rangle(t) = \int_{-\infty}^t d\tau \int \rho(dx)((T_{x(\tau-t)} f^{t-\tau})e(\tau)G(x(\tau-t)))\nabla_x\Psi, \tag{2.61}$$

or even:

$$\delta\langle\Psi\rangle(t) = \int_{-\infty}^t d\tau \int \rho(dy)e(\tau)G(y)\nabla_y(\Psi \circ f^{t-\tau}). \tag{2.62}$$

The last formula (2.62) can be rewritten as the response formula (2.14):

$$\begin{aligned}\delta\langle\Psi\rangle(t) =& \int_{-\infty}^t d\tau e(\tau) \int \rho(dy)G(y)\nabla_y(\Psi \circ f^{t-\tau}) \\ =& \int_{\infty}^t d\tau e(\tau)\Gamma_{\Psi,G}(t-\tau).\end{aligned} \tag{2.63}$$

where we have defined the response function (2.16):

$$\Gamma_{\Psi,G}(t) = \Theta(t) \int \rho(dy)G(y)\nabla_y(\Psi \circ f^t). \tag{2.64}$$

adding a $\Theta(t)$ to enforce causality. It is noticeable that causality enters in our derivation thanks to the time asymmetry present in the SRB measure property (2.55), where the time evolution goes in the forward direction. These response formulas can be proved rigorously considering Axiom A systems [8]. Moreover, these formulas can be extended to time-dependent F and ρ, and to random forces [8].

2.3.3 A partial FDT

Let's consider briefly the equilibrium case. In this case the measure is absolutely continuous $\rho(dx) = \rho(x)dx$. In this case we can make a partial integration step, obtaining the following response formula:

$$\Gamma_{\Psi,G}(t) = -\Theta(t) \int \rho(dy)\frac{\nabla_y(G\rho)}{\rho}(0)\Psi \circ f^t. \tag{2.65}$$

Performing the Fourier transform of (2.65) we obtain the following relation:

$$\begin{aligned}\Gamma_{\Psi,G}(\omega) = &- \int_{-\infty}^{\infty} dt\, e^{i\omega t}\Theta(t)\langle\frac{\nabla_y(G\rho)}{\rho}(0)\Psi\circ f^t\rangle \\ =&S_{\Psi,G}(\omega),\end{aligned} \tag{2.66}$$

where $S(\omega)$ is the Fourier transform of the correlation function multiplied with the Heaviside theta within the response function (2.65), also called spectral density. As discussed for the formula (1.64), the poles of the spectral density are the same of the response function. This is basically the FDT: the non-equilibrium response of the forced system (which enters in the dissipative regime) is fully described by fluctuations of the unperturbed system at the equilibrium.

On the other hand, if we consider a general SRB measure, or equivalently a NESS, the FDT is just partially preserved. To see this, we use the characterization of the SRB measure in terms of its smoothness along the unstable directions. First, we decompose the spatial pattern of the perturbation along the unstable, stable and neutral directions:

$$G(x) = G^s(x) + G^{uc}(x), \tag{2.67}$$

where we have grouped together the neutral direction (tangent to the trajectory) and the unstable direction. We can now decompose the response function (2.64) in two contributes:

$$\Gamma_{\Psi,G}(t) = \Gamma^s_{\Psi,G}(t) + \Gamma^{uc}_{\Psi,G}(t), \tag{2.68}$$

where:

$$\Gamma^s_{\Psi,G}(t) = \Theta(t)\int \rho(dy)G^s(y)\nabla_y(\Psi\circ f^t) \tag{2.69}$$

and

$$\begin{aligned}\Gamma^{uc}_{\Psi,G}(t) = &- \Theta(t)\int \rho(dy)\frac{\nabla^{uc}_y(G^{uc}\rho)}{\rho}(0)\Psi\circ f^t \\ = &- \Theta(t)\langle\frac{\nabla^{uc}_y(G^{uc}\rho)}{\rho}(0)\Psi\circ f^t\rangle.\end{aligned} \tag{2.70}$$

Notice that along the unstable and neutral directions we have performed the partial integration step since the SRB is smooth along these directions. This step is not allowed in general along the stable direction, where the SRB measure ρ is singular. We can see that just a contribute (the unstable-neutral one) can be rewritten in a time auto-correlation function form (which expresses the natural fluctuations in the system). As a consequence, we can infer that the non-equilibrium response of these chaotic dissipative deterministic system to a weak forcing can not be fully described by spontaneous fluctuations. The FDT can be recovered in this context considering just perturbations tangent everywhere to the unstable-neutral directions in each point of the phase space. To understand why the full FDT is lost, we have to recall that the support of the SRB measure ρ is a strange attractor, i.e. a fractal geometrical object. As a consequence, a general forcing will have a component along the unstable manifold, which can be described by the natural fluctuations (since we are on the support of the invariant SRB distribution ρ), and another component along the stable direction which induces a motion (which will dissipate exponentially in time) outside the support of ρ [6, 20] and hence can not be described by spontaneous fluctuations. If

the ρ is smooth this problem does not exist anymore, since there is no attractor and the forced motion evolves within the support of the measure. Nevertheless, we remark that still the response function $\Gamma_{\Psi,G}$ can be computed by means of the steady-state average of a suitable quantity.

We can observe that the FDT partially holds also from the point of view of the Fourier transform:

$$\Gamma_{\Psi,G}(\omega) = \Gamma^s_{\Psi,G}(\omega) + \Gamma^{uc}_{\Psi,G}(\omega). \tag{2.71}$$

For $\Gamma^{uc}_{\Psi,G}$ we can apply the same line of thinking used in the equilibrium context (2.65): since the poles with negative imaginary part of the spectral density are the same of the response function, the FDT holds. On the other hand, this is no more true for the stable component $\Gamma^s_{\Psi,G}$. To clearly see this let's consider that component using the form of the response function (2.61):

$$\begin{aligned}\Gamma^s_{\Psi,G}(\omega) &= \int_0^\infty dt\, e^{i\omega t} \int \rho(dx)((T_{x(-t)}f^t)G^s(f^{-t}x))\nabla_x\Psi \\ &= \int \rho(dx)\left(\int_0^\infty dt\, e^{i\omega t}(T_{x(-t)}f^t)G^s(f^{-t}x)\right)\nabla_x\Psi.\end{aligned} \tag{2.72}$$

At this point we introduce the contraction semigroup $\{\mathcal{T}^t\}_{t\geq 0}$ along the stable direction:

$$(\mathcal{T}^t X^s)(x) \equiv (T_{x(-t)}f^t)G^s(f^{-t}x). \tag{2.73}$$

Let's call $-H$ the infinitesimal generator of the semigroup. We have, by the Hille-Yosida theorem [68, 69], the following relation:

$$\int_0^\infty dt\, e^{i\omega t}(T_{x(-t)}f^t)G^s(f^{-t}x) = (H - i\omega)^{-1}, \tag{2.74}$$

hence the Fourier transform of the stable component of the response function is:

$$\Gamma^s_{\Psi,G}(\omega) = \int \rho(dx)(H - i\omega)^{-1}\nabla_x\Psi. \tag{2.75}$$

We can see that the poles of $\Gamma^s_{\Psi,G}(\omega)$ are linked to the spectrum of the infinitesimal generator H, which are different from those of a spectral density $S(\omega)$. We can see that due to this stable component the FDT holds only in part [67].

2.3.4 An algorithm to compute the linear response

The two contributions to the response function (2.68) are hard to directly compute in practical applications, especially the one along the stable directions. Nevertheless some algorithms have been developed to numerically evaluate them [15,16]. We present here the simplified version proposed in [17], valid for both deterministic and stochastic system. This algorithm does not need any knowledge of ρ because it focuses entirely on computing the factor related to the observable $\nabla_x\Psi(x(t))$, where the derivation is with respect to the initial condition x. Let's consider first a discrete evolution in time $n = 0, 1, ...$ of the following $d-$ dimensional system:

$$x_{n+1} = g_n(x_n) = x_n + v_n(x_n), \qquad x_0 = x. \tag{2.76}$$

and:

$$v_n(x) = \varepsilon[F(x) + \xi_n], \tag{2.77}$$

where the parameter $\varepsilon > 0$ is such that the continuous evolution in time is recovered for $\varepsilon \to 0^+$ and ξ_n is a stochastic term. We apply now the chain rule recursively:

$$\begin{aligned}\nabla_x \Psi(x_n) =&(\nabla\Psi)(x_n) \cdot \nabla_x (g_{n-1} \circ ... \circ g_0)(x) \\ =&(\nabla\Psi)(x_n) \cdot G_{n-1}(x) \\ =&(\nabla\Psi)(x_n)(\nabla g_{n-1})(x_{n-1}) \cdot \nabla_x (g_{n-2} \circ ... \circ g_0)(x) \\ =&(\nabla\Psi)(x_n)(\nabla g_{n-1})(x_{n-1}) \cdot G_{n-2}(x),\end{aligned} \tag{2.78}$$

where $\nabla\Psi$ is a $1 \times d$ row vector. We have introduced in the above relations the $d \times d$ propagation matrices $G_k(x) = \nabla(g_k \circ ... \circ g_0)$, which satisfy the following recursive relations:

$$\begin{aligned}G_n(x) =&\nabla g_n(x_n) \cdot G_{n-1}(x) \\ G_0(x) =&\nabla g_0(x).\end{aligned} \tag{2.79}$$

Now we explicit the form of the row vector: $\nabla g_k = \mathbb{1} + \varepsilon \nabla F$. We can notice that the form of this vector does not depend nor on the index k and neither on the noise ξ_k. The term $\nabla_x \Psi(x(t))$ reduces to the following expression:

$$\nabla_x \Psi(x_n) = (\nabla\Psi)(x_n) \cdot [\mathbb{1} + \varepsilon \nabla F(x_{n-1})] \cdot ... \cdot [\mathbb{1} + \varepsilon \nabla F(x)]. \tag{2.80}$$

We can obtain the continuous time evolution sending the number of steps n to ∞ or equivalently making the time step $\varepsilon = t/n$ vanish (with a suitable rescaling of the noise). In this limit we have that $[\mathbb{1} + \varepsilon \nabla F(x)] \approx \exp(\varepsilon \nabla F(x))$, hence:

$$\nabla_x \Psi(x(t)) = (\nabla\Psi)(x(t)) \cdot T \exp\left[\int_0^t ds (\nabla F)(x(s))\right]. \tag{2.81}$$

where $x(s)$ depends on the initial condition x through the equation of the dynamics with a frozen realization of the noise ξ (we will later average over these realizations) and T is the time-ordering operator. Once the relation (2.80) or (2.81) is computed recursively. we plug it into (2.16) and then we average first over the possible realizations of the noise and then over an ensemble of possible initial conditions (i.e. over ρ). The computational time required by our algorithm to estimate the propagator matrices G_k for $k \leq n$ is $\mathcal{O}(n)$. Note that this algorithm is suited for systems which are at the steady state before being perturbed, otherwise the propagator matrices G_k are different from the propagator starting from index j and ending in index $j + k$, and the required computational time increases to $\mathcal{O}(n^2)$.

2.4 KK relations

In this section we reconsider the KK relations, introduced in section 1.4, within the much general context of the response function (2.16). We can derive these constraints for the response function using as a main ingredient the fact that the response function $\Gamma_{\Psi,G}$ is causal [26, 70]. For $t > 0$ we have that the following identity holds:

$$\Gamma_{\Psi,G}(t) = \Theta(t)\Gamma_{\Psi,G}(t), \tag{2.82}$$

whose Fourier transform is:

$$\Gamma_{\Psi,G}(\omega) = \int \frac{d\omega'}{2\pi}\Theta(\omega-\omega')\Gamma_{\Psi,G}(\omega'), \tag{2.83}$$

We recall now the form of the Fourier transform of the Heaviside theta (1.62). Using that expression, we can rewrite (2.83) as follows:

$$\Gamma_{\Psi,G}(\omega) = \frac{1}{2\pi}\left(\pi\delta(\omega) + iP\left[\frac{1}{\omega}\right]\right) * \Gamma_{\Psi,G}(\omega), \tag{2.84}$$

where $*$ denotes the convolution operation. We can now derive from (2.84) the following relation:

$$\Gamma_{\Psi,G}(\omega) = iP\left[\int \frac{d\omega'}{\pi}\frac{\Gamma_{\Psi,G}(\omega')}{\omega-\omega'}\right], \tag{2.85}$$

which is an Hilbert transform of the response function in itself. At this point we want to derive the KK relations, which link the real and the imaginary part of the response function through an Hilbert transform. Writing $\Gamma_{\Psi,G} = \Gamma^R_{\Psi,G} + i\Gamma^I_{\Psi,G}$, we obtain them:

$$\Gamma^R_{\Psi,G}(\omega) = -P\left[\int \frac{d\omega'}{\pi}\frac{\Gamma^I_{\Psi,G}(\omega')}{\omega-\omega'}\right] \tag{2.86}$$

and:

$$\Gamma^I_{\Psi,G}(\omega) = P\left[\int \frac{d\omega'}{\pi}\frac{\Gamma^R_{\Psi,G}(\omega')}{\omega-\omega'}\right]. \tag{2.87}$$

These relations are equal to the ones already found within the Kubo's theory. On the other hand, in the Kubo's theory an explicit form for the real and imaginary was derived, due to the particular form of the response function (1.55). To connect with Kubo's theory, we rewrite the response function $\Gamma_{\Psi,G}$ as follows [26]:

$$\Gamma_{\Psi,G}(t) = \Theta(t)R_{\Psi,G}(t). \tag{2.88}$$

where $R_{\Psi,G}(t) \equiv \int \rho(dx)(G(x)\nabla)e^{tL}\Psi(x)$ contains all the details of the physical process we are considering. In the Kubo's theory, R is proportional to the derivative of a time-correlation function $C(t)$:

$$R(t) = -\beta\frac{d}{dt}C(t) \tag{2.89}$$

Now we perform the Fourier transform of (2.88), obtaining:

$$\begin{aligned}\Gamma_{\Psi,G}(\omega) =& \frac{1}{2\pi}\left(\pi\delta(\omega) + iP\left[\frac{1}{\omega}\right]\right) * R_{\Psi,G}(\omega)\\ =& \Gamma^R_{\Psi,G} + i\Gamma^I_{\Psi,G}.\end{aligned} \tag{2.90}$$

Explicit expressions connecting $R_{\Psi,G}$ and the real and imaginary part of the full response function $\Gamma_{\Psi,G}$ can be found if R has some definite symmetry with respect to time reversal [26]. In fact, if $R(t)$ is even or odd under time reversal, we have that the

Fourier transform $R(\omega)$ is respectively real or imaginary. To see this, it is sufficient to notice that (dropping here the indices):

$$\begin{aligned} R(\omega) =& \int dt\, e^{i\omega t} R(t) \\ =& \epsilon_R \int d(-t)\, e^{i\omega(-t)} R(-t) \\ =& \epsilon_R R(-\omega) \\ =& \epsilon_R R^*(\omega), \end{aligned} \tag{2.91}$$

where $\epsilon_R = \pm 1$ if R is even or odd under time reversal. In the case of an even symmetry we have that $R(\omega) = R^*(\omega)$, hence $R(\omega)$ is real. On the other hand, for an odd symmetry $R(\omega) = R^*(\omega)$, hence $R(\omega)$ is imaginary. In the former case, from (2.90), we derive $R(\omega) = 2\Gamma^R_{\Psi,G}(\omega)$, while in the latter $R(\omega) = 2i\Gamma^I_{\Psi,G}(\omega)$. In section 1.4, we have considered the case of an odd-symmetric $R(t)$ (since usually the correlation function $C(t)$ in (2.89) is symmetric under time reversal), obtaining (1.64). We recall that it is (1.64) the main cause of the FDT, since the Fourier transform of the response function can be reconstructed by the spontaneous fluctuations described by the correlation function.

Before presenting other general KK dispersion relations, we want to explicit the link between the short-time behaviour of the response function $\Gamma_{\Psi,G}(t)$ and the asymptotic behaviour of its Fourier transform [6]. In order to do that, we consider the following relation:

$$\begin{aligned} \int dt \Theta(t) t^k e^{i\omega t} =& (-i)^k \frac{d^k}{d\omega^k} \left(\pi\delta(\omega) + iP\left[\frac{1}{\omega}\right] \right) \\ \approx& k! \frac{i^{k+1}}{\omega^{k+1}}. \end{aligned} \tag{2.92}$$

where in (2.92) we have overlooked the distribution nature of the solution and considered $\omega \neq 0$. As a consequence, if the Taylor expansion of the response function $\Gamma_{\Psi,G}$ is:

$$\Gamma_{\Psi,G}(t) \approx \alpha \Theta(t) t^\beta + .., \tag{2.93}$$

the asymptotic behaviour of its Fourier transform, in the limit $\omega \to \infty$ is:

$$\Gamma_{\Psi,G}(\omega) \approx \overline{\alpha} \frac{1}{\omega^{\beta+1}} + ..., \tag{2.94}$$

where $\overline{\alpha} = \alpha(i)^{\beta+1}\beta!$. The specific values of α and β depend on the considered system and on the perturbation. Moreover, since $\Gamma_{\Psi,G}(\omega) = \Gamma^*_{\Psi,G}(-\omega)$ for all real ω (including the asymptotic limit), we have that $\Gamma^R_{\Psi,G}$ is even in ω and $\Gamma^I_{\Psi,G}$ is odd in ω. From that we can deduce that $\overline{\alpha}$ is real for odd β, while it is imaginary for even β.

It is possible to show, considering the asymptotic limit (2.94) and given the causality of the response function, that the following set of generalized KK dispersion relations holds [71]:

$$-\frac{\pi}{2}\omega^{2p-1}\Gamma^I_{\Psi,G}(\omega) = P\int_0^\infty d\omega' \frac{(\omega')^{2p}\Gamma^R_{\Psi,G}(\omega')}{(\omega')^2 - \omega^2}, \tag{2.95}$$

$$\frac{\pi}{2}\omega^{2p}\Gamma^R_{\Psi,G}(\omega) = P\int_0^\infty d\omega' \frac{(\omega')^{2p+1}\Gamma^I_{\Psi,G}(\omega')}{(\omega')^2 - \omega^2}, \tag{2.96}$$

where $p \in [0, ..., (\beta-1)/2]$ if β is odd or $p \in [0, ..., (\beta-1)/2]$ if β is even. The KK relations (2.95) and (2.96) provide a set of self-consistency relations that the response function must satisfy. A practical use of them is to test if an obtained response function is correct. Another possible use of these constraints could be to perform a parametric fit in order to obtain a well-defined response function. The number of these constraints increases with the parameter β, which measures how fast the response function is around $t = 0$. In particular, we can consider certain limits for them, in order to get more handy constraints. For example we can consider (2.96) for $p = 0$ and $\omega \to 0$, obtaining:

$$\Gamma^R_{\Psi,G}(0) = \frac{2}{\pi} P \int_0^\infty d\omega' \frac{\Gamma^R_{\Psi,G}(\omega')}{\omega'}, \tag{2.97}$$

which links the static susceptibility (obtained for long times) with the response of the system at all frequencies. This formula is exactly the sum rule (1.80) derived within the Kubo's theory, We observe that for $\omega \to 0$, the imaginary part of the response function must vanish since $\Gamma^I_{\Psi,G}(0) = 0$, since $\Gamma^I_{\Psi,G}(\omega)$ is asymmetric in ω. On the other hand the analiticity of the Fourier transform of the response function ensures the finitess of $\Gamma^R_{\Psi,G}(0)$. This is consistent with the fact the the integral in the RHS of (2.97) is not divergent, since $\Gamma^I_{\Psi,G}$ goes to zero at least fast as a linear function in ω' (the linear expansion of $\Gamma^I_{\Psi,G}$ around 0 admits only odd terms in ω).

We can derive from (2.95) and (2.96) additional constraints in the limit $\omega \to \infty$, as explained in [72]:

$$\int_0^\infty d\omega' \, (\omega')^{2p+1} \Gamma^I_{\Psi,G}(\omega') = 0, \qquad \begin{array}{ll} p \in [0, \beta/2 - 1], & \beta \text{ even} \\ p \in [0, (\beta-3)/2], & \beta \text{ odd}, \end{array} \tag{2.98}$$

$$\int_0^\infty d\omega' \, (\omega')^{2p} \Gamma^R_{\Psi,G}(\omega') = 0, \qquad \begin{array}{ll} p \in [0, \beta/2 - 1], & \beta \text{ even} \\ p \in [0, (\beta-1)/2], & \beta \text{ odd}, \end{array} \tag{2.99}$$

which are referred to as vanishing sum rules. For $\beta = 0$ no vanishing sum rule can be written, while for $\beta = 1$ only (2.99) gives a rule. Interestingly, in addition to a set of KK relations, another non-vanishing sum rule can be found. If β is odd that rule is:

$$\int_0^\infty d\omega' \, (\omega')^{\beta} \Gamma^I_{\Psi,G}(\omega') = -\frac{\pi}{2} \overline{\alpha}, \tag{2.100}$$

while for even β:

$$\int_0^\infty d\omega' \, (\omega')^{\beta} \Gamma^R_{\Psi,G}(\omega') = i\frac{\pi}{2} \overline{\alpha}, \tag{2.101}$$

where $\overline{\alpha}$ is defined by the asymptotic behaviour of the Fourier transform of the response function (2.94).

2.5 Higher-order response theory

Until now, we have focused on computing the linear correction $\delta\langle\Psi(x(t))\rangle$ (2.14) to the expectation value of an observable Ψ after the considered system is perturbed by a forcing with time pattern $e(t)$ and space pattern $G(x)$. In order to do that, we have

defined a linear response function $\Gamma_{\Psi,G}$ (2.16). We now want to focus on the higher order terms of the following expansion:

$$\langle\Psi\rangle^e(t) = \langle\Psi\rangle + \sum_{n\geq 1} \delta^{(n)}\langle\Psi\rangle, \tag{2.102}$$

where $\langle\Psi\rangle^e$ is the expectation value of Ψ in the perturbed system at time t, $\langle\Psi\rangle$ the same expectation values in the unperturbed system and $\delta^{(n)}\langle\Psi\rangle$ is the perturbation term of order n in the perturbation. Until now we have focused on the $n = 1$ term. It can be shown that the general n^{th} term can be expressed by the following $n-$uple convolution integral [13, 49]:

$$\delta^{(n)}\langle\Psi\rangle(t) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} d\tau_1 ... d\tau_n \Gamma^{(n)}_{\Psi,G}(\tau_1, ..., \tau_n) e(t-\tau_1) ... e(t-\tau_n), \tag{2.103}$$

where we have introduced the n^{th} order response function. This response function is causal, i.e. it is vanishing if any of its arguments is negative, and it can be expressed as follows:

$$\begin{aligned}\Gamma^{(n)}_{\Psi,G}(\tau_1, ..., \tau_n) = \int \rho(dx)\Theta(\tau_1)...\Theta(\tau_n - \tau_{n-1})\times \\ \times G(x)\nabla(e^{(\tau_n-\tau_{n-1})L}...G(x)\nabla(e^{\tau_1 L}\Psi(x))),\end{aligned} \tag{2.104}$$

where e^{tL} is the operator of time evolution for the observables.

Let's see what happens in the frequency domain [13, 26]. First, we Fourier transform (2.103), obtaining:

$$\delta^{(n)}\langle\Psi\rangle(\omega) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} d\omega_1 ... d\omega_n \Gamma^{(n)}_{\Psi,G}(\omega_1, , , \omega_n) e(\omega_1) ... e(\omega_n)\, \delta\left(\omega - \sum_{k=1}^{n} \omega_k\right), \tag{2.105}$$

where the Dirac delta ensures that the sum of all the n frequencies ω_k is equal to the frequency ω of the perturbation. The Fourier transform of the n^{th} order response function is defined as follows:

$$\Gamma^{(n)}_{\Psi,G}(\omega_1, , , \omega_n) = \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} dt_1 ... dt_n \Gamma^{(n)}_{\Psi,G}(t_1, , , t_n) \exp\left(i \sum_{k=1}^{n} \omega_k t_k\right). \tag{2.106}$$

We remark that the fact that the Fourier transform can be performed boils down to the integrability of the response function. Ruelle [49] proved that in the $n = 1$ case showing that the response function is not exponentially increasing, since both the terms coming from the stable and unstable manifolds converge. In the non-linear case, i.e. for $n > 1$, we can argue that this line of thinking remains valid. since higher-order correlations (in the equilibrium case) are usually weaker and they decrease faster in time.

We can derive the generalized KK dispersion relations for the n^{th} order response function following and extending the same line of thinking used in the $n = 1$ case (2.85). First, we define the symmetric and asymmetric part of the response function:

$$S^{(n)}_{\Psi,G}(t_1, , , t_n) \equiv \Gamma^{(n)}_{\Psi,G}(t_1, , , t_n) + \Gamma^{(n)}_{\Psi,G}(-t_1, , , -t_n), \tag{2.107}$$

$$A^{(n)}_{\Psi,G}(t_1,,,t_n) \equiv \Gamma^{(n)}_{\Psi,G}(t_1,,,t_n) - \Gamma^{(n)}_{\Psi,G}(-t_1,,,-t_n), \tag{2.108}$$

which are different from zero for $t_n > ... > t_1 > 0$ or $t_n < .... < t_1 < 0$. Since $\Gamma^{(n)}_{\Psi,G}(\omega) = \left(\Gamma^{(n)}_{\Psi,G}\right)^* (-\omega)$, we have that $\mathrm{Im}\left[\Gamma^{(n)}_{\Psi,G}\right](\omega)$ is odd under the transformation $\omega \to -\omega$ (where ω is the vector with all the ω_k), while $\mathrm{Re}\left[\Gamma^{(n)}_{\Psi,G}\right](\omega)$ is even under that transformation, Performing the Fourier transform of $S^{(n)}_{\Psi,G}$ and $A^{(n)}_{\Psi,G}$ we have that:

$$2\,\mathrm{Re}\left[\Gamma^{(n)}_{\Psi,G}\right](\omega_1,...,\omega_n) = \int_{-\infty}^{\infty} dt_1...dt_n \exp\left(-i\sum_{k=1}^{n}\omega_k t_k\right) S^{(n)}_{\Psi,G}(t_1,,,t_n), \tag{2.109}$$

and:

$$2i\,\mathrm{Im}\left[\Gamma^{(n)}_{\Psi,G}\right](\omega_1,...,\omega_n) = \int_{-\infty}^{\infty} dt_1...dt_n \exp\left(-i\sum_{k=1}^{n}\omega_k t_k\right) A^{(n)}_{\Psi,G}(t_1,,,t_n). \tag{2.110}$$

We consider now the following identities, given by the causality property of the response function:

$$\begin{aligned}\Gamma^{(n)}_{\Psi,G}(t_1,,,t_n) &= \prod_{i=1}^{q}\Theta(t_{k_i})\Gamma^{(n)}_{\Psi,G}(t_1,,,t_n)\\ &= \prod_{i=1}^{q}\Theta(t_k)S^{(n)}_{\Psi,G}(t_1,,,t_n)\\ &= \prod_{i=1}^{q}\Theta(t_{k_i})A^{(n)}_{\Psi,G}(t_1,,,t_n),\end{aligned} \tag{2.111}$$

where $q \leq n$ and the indices $k_1...k_q$ run over a subset of the indices $1...n$. The relations (2.111) are valid for any of these subsets. Performing the Fourier transform of (2.111) we obtain:

$$\begin{aligned}\Gamma^{(n)}_{\Psi,G}(\omega_1,...,\omega_n) &= \prod_{i=1}^{q}\left(\frac{i}{\pi}P\left[\frac{1}{\omega_{k_i}}\right] + \delta(\omega_{k_i})\right) * \left(\frac{1}{2^q}\Gamma^{(n)}_{\Psi,G}(\omega_1,,,\omega_n)\right)\\ &= \prod_{i=1}^{q}\left(\frac{i}{\pi}P\left[\frac{1}{\omega_{k_i}}\right] + \delta(\omega_{k_i})\right) * \left(\frac{1}{2^{q-1}}\mathrm{Re}\left[\Gamma^{(n)}_{\Psi,G}\right](\omega_1,...,\omega_n)\right)\\ &= \prod_{i=1}^{q}\left(\frac{i}{\pi}P\left[\frac{1}{\omega_{k_i}}\right] + \delta(\omega_{k_i})\right) * \left(\frac{i}{2^{q-1}}\mathrm{Im}\left[\Gamma^{(n)}_{\Psi,G}\right](\omega_1,...,\omega_n)\right),\end{aligned} \tag{2.112}$$

where $*$ is a convolution product over the frequencies $\omega_{k_1}...\omega_{k_n}$. The relations (2.112) are the generalized KK dispersion relations. Even in the non-linear regime it is possible to reconstruct the full response of the system starting just from the real or the imaginary part of the response function, with the only assumption of causality.

2.5.1 Higher-order FDT

We now want to extend the FDT found in the linear case (2.30) to all the orders [26]. In order to do that we must assume that the invariant measure of the system is smooth $\rho(dx) = \rho(x)dx$, otherwise the contribution along the stable directions would spoil the

FDT. We start from the n^{th} order response function (2.104), rewriting it by means of the perturbation operator $L_1(.) = G \cdot \nabla(.)$ defined in (2.7):

$$\begin{aligned}\Gamma^{(n)}_{\Psi,G}(\tau_1,...,\tau_n) = \int \rho(dx)\Theta(\tau_1)...\Theta(\tau_n - \tau_{n-1})\times \\ \times L_1(e^{(\tau_n-\tau_{n-1})L}...L_1(e^{\tau_1 L}\Psi(x)))\end{aligned} \tag{2.113}$$

We simplify the notation introducing the evolution operator $\Lambda(t) = \exp(tL)$:

$$\begin{aligned}\Gamma^{(n)}_{\Psi,G}(\tau_1,...,\tau_n) = \int \rho(dx)\Theta(\tau_1)...\Theta(\tau_n - \tau_{n-1})\times \\ \times L_1\Lambda(\tau_n - \tau_{n-1})...L_1\Lambda(\tau_1)\Psi(x) \\ =\Theta(\tau_1)...\Theta(\tau_n - \tau_{n-1})\times \\ \times \langle \rho, L_1\Lambda(\tau_n - \tau_{n-1})...L_1\Lambda(\tau_1)\Psi(x)\rangle,\end{aligned} \tag{2.114}$$

where $\langle \psi_1, \psi_2 \rangle$ defines an integration operator in space for the product ψ_1 and ψ_2, for any couple of integrable functions. Given that, we recall the definition of the adjoint operator(2.28) of L_1: $\mathcal{L}_1(.) = -\nabla(G.)$ and of the operator L^* which defines the time-reversed motion (2.45). Moreover, we can define the adjoint operator of the evolution operator $\Lambda^\dagger(t) = \exp(tL^*)$, which evolves the operators backwards in time:

$$\Lambda^\dagger(t)[\psi(x)] = \psi(x(-t)), \tag{2.115}$$

for any function ψ. At this point, we derive from (2.114) the following relation, taking the adjoint of all the operators:

$$\begin{aligned}\Gamma^{(n)}_{\Psi,G}(\tau_1,...,\tau_n) =&\Theta(\tau_1)...\Theta(\tau_n - \tau_{n-1})\times \\ &\times \langle \Lambda^\dagger(\tau_1)\mathcal{L}_1...\Lambda^\dagger(\tau_n - \tau_{n-1})\mathcal{L}_1\rho, \Psi(x)\rangle,\end{aligned} \tag{2.116}$$

which is the generalized FDT extending the Kubo-Agarwal formula (2.30) for non-linear response functions.

2.6 Surrogate Response Theory

A new meaningful angle of the problem in RT has been introduced in [28] where response relations *between perturbed observables* are built. These relations can be useful in a large variety of contexts where the knowledge of the forcing is just partial, and we want to use perturbed observables to predict the state of other perturbed observables. Let's consider the perturbed system (2.5), where the perturbation has time pattern $e(t)$ and space pattern $G(x)$. We consider now the Fourier Transform of two linear response relation of the kind (2.17), for two generic observables Ψ_1 and Ψ_2:

$$\begin{cases}\delta\langle\Psi_1\rangle(\omega) =\Gamma_{\Psi_1,G}(\omega)e(\omega) \\ \delta\langle\Psi_2\rangle(\omega) =\Gamma_{\Psi_2,G}(\omega)e(\omega)\end{cases} \tag{2.117}$$

Now we take the ratio between the two equations in (2.117), getting:

$$\delta\langle\Psi_1\rangle(\omega) = \frac{\Gamma_{\Psi_1,G}(\omega)}{\Gamma_{\Psi_2,G}(\omega)}\delta\langle\Psi_2\rangle(\omega). \tag{2.118}$$

It is remarkable that this can be done for any time pattern of the perturbation $e(t)$. This is because we are focusing on the linear order terms, while for higher-order terms the time pattern enters in the Fourier transform of the change of the observables through a more complicated integral formula (2.106) (and hence it cannot be deleted). We can define the Fourier transform of the surrogate response function H_{12} between Ψ_1 and Ψ_2 as follows:

$$H_{12}(\omega) \equiv \frac{\Gamma_{\Psi_1,G}(\omega)}{\Gamma_{\Psi_2,G}(\omega)}. \tag{2.119}$$

If we go back in the time domain, performing the inverse Fourier transform of the relation (2.118), we have the following response relation:

$$\begin{aligned} \delta\langle\Psi_1\rangle(t) =& \int_{-\infty}^{\infty} H_{12}(t-\tau)\delta\langle\Psi_1\rangle(\tau) \\ =& H_{12}(t) * \delta\langle\Psi_2\rangle(t) \end{aligned} \tag{2.120}$$

H_{12} is called a surrogate response function because the perturbed observable Ψ_2 in (2.120) acts as a surrogate for the forcing to predict the state of the other perturbed observable Ψ_1. The surrogate response function depends on the spatial pattern of the forcing and on the given couple of observables which act as predictor and predictand. We remark that the observables can even be the same physical quantity but with different spatial supports (if the system is spatially extended): the knowledge of the state of the quantity in one location can effectively predict the state of the same quantity in another location, as we will see in chapter 4 in the climatic model Lorenz 96.

As discussed for the response functions Γ in (2.16), we want to deal with response functions which have predictive power, i.e. they are zero for negative times. Only if the surrogate response functions have this feature, they can be used for predictions purposes since the state of the perturbed observable Ψ_1 can be predicted with just the information about the evolution of Ψ_2 up to time t. Moreover, we observe that if the surrogate response function is predictive, it is possible to derive a corresponding set of KK dispersion relations. To have the surrogate response function H predictive, it is necessary that its Fourier transforms has no poles in the upper complex $\omega-$plane. In order to check that, we have to consider (2.119), looking for the zeros of the global response function $\Gamma_{\Psi_2,G}$ at the denominator (the poles of the response function at the numerator are not in the upper complex $\omega-$plane since Γ is predictive). The presence of zeroes in the Fourier transform $\Gamma_{\Psi,G}$ is not trivial to detect [73]. Physically, it is equivalent to stating that exists a frequency ω_1 such that $\delta\langle\Psi\rangle(\omega_1) = \Gamma_{\Psi,G}(\omega_1)e(\omega_1) = 0$. As a consequence, if we consider a mono-chromatic perturbation with frequency ω_1, there is no response of the system looking through the lens of $\delta\langle\Psi\rangle$: some information is lost and this can signal the possibility of the absence of predictive power of the observable Ψ. The dependence of the predicative power of the surrogate response function on the particular choice of the observables is physically reasonable: we expect that not all choices of predictors and predictands are equally fortunate. For instance, if there is a causal link in a feedback or a flow of information from one variable to another, there is an asymmetry in the predictive power, which can be observed in the magnitude of the non-causal part of the surrogate response functions H (i.e. for negative times). In the case the surrogate response function H_{12} has a non-causal component, predictions

must be performed using the following modified surrogate response function:

$$H'_{12}(t) \equiv \Theta(t) H_{12}(t). \tag{2.121}$$

On the other hand, if the problem at hand focuses on the reconstruction of a past signal from the data, also the non-causal part can be considered. In the context of control theory (which considers general linearized systems), it is possible to derive an effective criterion which can identify non-predictive dynamical variables, as we will see in chapter 3.

In order to get a physical insight of the structure of the surrogate response functions H_{12}, we will now make some assumption on its functional form. The following results are not general, but they are extremely useful in order to better understand the surrogate response theory. Let's approximate H_{12} with the following rational function:

$$H_{12}(\omega) = \frac{P_{12}(\omega)}{Q_{12}(\omega)} = \frac{a \prod_{j=1}^{N} (\omega - \omega_j)}{b \prod_{k=1}^{M} (\omega - v_k)}. \tag{2.122}$$

where N and M are integers which can also be very large, and the ω_j and v_k are in general distinct. Moreover, we assume that each of the the ω_j and v_k are of first order. For very large ω, we have that $H_{12}(\omega) \sim \frac{a}{b}\omega^{N-M}$. We can now rewrite (2.122), performing a division of the polynomials:

$$\begin{aligned} H_{12}(\omega) =& S_{12}(\omega) + \frac{R_{12}(\omega)}{Q_{12}(\omega)} \\ =& S_{12}(\omega) + K_{12}(\omega), \end{aligned} \tag{2.123}$$

where $S_{12}(\omega)$ is a polynomial of order $N - M$ (if $N \geq M$, otherwise that term is vanishing), while the order of R_{12} is strictly smaller than the order of Q_{12}. We can express S_{12} as follows:

$$S_{12}(\omega) = \sum_{j=0}^{N-M} c_j (-i\omega)^j. \tag{2.124}$$

The second term K_{12} can be written in the following way, performing a partial fraction expansion:

$$K_{12}(\omega) = \frac{a}{b} \sum_{k=1}^{M} \frac{\alpha_k}{\omega - v_k}, \tag{2.125}$$

where $\alpha_k = R_{12}(v_k) \left(\frac{dQ_{12}(x)}{dx}_{|x=v_k} \right)^{-1}$. We now want to go back in the time domain, performing the inverse Fourier transform of (2.123). We thus recall that the inverse Fourier transform of $(-i\omega)^j$ is $\delta^j(t)$, i.e. the j^{th} derivative of the Dirac delta. Moreover, we divide now the M poles v_k of K_{12} into M_p predictive poles p_k and M_{np} not predictive poles n_p, which have respectively negative and positive imaginary part. Performing the inverse Fourier transform, we have:

$$\begin{aligned} H_{12}(t) =& \sum_{j=1}^{N-M} c_j \delta^j(t) + \frac{a}{b}\Theta(t) \sum_{k=1}^{M_p} \alpha_k e^{-ip_k t} + \frac{a}{b}\Theta(-t) \sum_{p=1}^{M_{np}} \alpha_p e^{-in_p t} \\ =& S_{12}(t) + K_{12}(t), \end{aligned} \tag{2.126}$$

where the first term S_{12} gives a singular contribution, while the second is not-singular and it is further divided in a predictive term and a not-predictive term. Since the singular term has to be interpreted in a distributional way, we consider its action within the expression (2.120):

$$\delta\langle\Psi_1\rangle(t) = \sum_{j=0}^{N-M} (-1)^j c_j \frac{d^j \delta\langle\Psi_2\rangle(t)}{dt^j} + \\ + \int_{-\infty}^{\infty} d\tau \left[\Theta(\tau) \sum_{k=1}^{M_p} \alpha_k e^{-ip_k\tau} \delta\langle\Psi_2\rangle(t-\tau) + \Theta(-\tau) \sum_{p=1}^{M_{np}} \alpha_p e^{-in_p\tau} \delta\langle\Psi_2\rangle(t-\tau) \right]. \tag{2.127}$$

The first term in (2.127) links the values of the two observables $\langle\Psi_1\rangle$ and $\langle\Psi_1\rangle$ at the same time t, hence it is local in time. On the other hand, the non-singular terms give a non-local (in time) contribution, also called memory terms. The smaller is the imaginary part of the pole, the bigger will be its contribution in the asymptotic limit (for $t \to \pm\infty$, depending on the causal or not causal nature of the pole). The smallest predictive pole p and the smallest not-predictive pole n control the asymptotic behaviour of the surrogate response function. Moreover, note that if $M < N$, the surrogate response function $H_{12}(\omega)$ presents a singular term, while $H_{12}(\omega) = 1/H_{12}(\omega)$ does not, hence the local (in time) flow of information has a definite direction in this setting where we consider observables as surrogates of the forcing. The fact that the poles are different between H_{12} and H_{12} is relevant, because it shows that different observables keep memory of the system in different ways, with different time scales given by the imaginary terms of the poles. Note that (2.127) slightly generalizes the relation found in [28], since it considers also non-causal terms.

We can further the inspection of H_{12} as a rational function allowing poles or zeros of higher order. For example we can assume that one predictive pole v_1 (and just one) has order $s \geq 1$:

$$H_{12}(\omega) = \frac{P_{12}(\omega)}{Q_{12}(\omega)} = \frac{a \prod_{j=1}^{N} (\omega - \omega_j)}{b(v - v_1)^s \prod_{k=1}^{M-s} (\omega - v_k)}. \tag{2.128}$$

As a consequence, for some coefficients α_k and γ_k, we have that the non-singular component K assumes the following form:

$$K_{12}(\omega) = \sum_{k=1}^{M-s} \frac{\alpha_k}{\omega - v_k} + \sum_{k=1}^{s} \frac{\gamma_k}{(\omega - v_1)^k}. \tag{2.129}$$

The terms of the second sum in (2.129) will give a contribute in the time domain in $H_{12}(t)$ of the following form:

$$\Theta(t) \sum_{k=1}^{s} \gamma_k t^{k-1} e^{-iv_1 t}. \tag{2.130}$$

These terms, thanks to the polynomial factor t^{k-1}, are the cause of the emergence of peaks of the surrogate response function $H_{12}(t)$ away from $t = 0$.

2.6.1 More complex forcings: more surrogates

Let's considered the perturbed system (2.5) but with a more complex pattern of forcing, consisting in two independent perturbations:

$$F(x) \Rightarrow F(x) + e_1(t)G_1(x) + e_2(t)G_2(x), \tag{2.131}$$

where $G_{1,2}(x)$ are space patterns while $e_{1,2}(x)$ time patterns. We consider now three independent observables Ψ_1, Ψ_2 and Ψ_3. The linear change in their expectation value after the perturbation is the following (in the frequency domain):

$$\begin{cases} \delta\langle\Psi_1\rangle(\omega) = \Gamma_{\Psi_1,G_1}(\omega)e_1(\omega) + \Gamma_{\Psi_1,G_2}(\omega)e_2(\omega) \\ \delta\langle\Psi_2\rangle(\omega) = \Gamma_{\Psi_2,G_1}(\omega)e_1(\omega) + \Gamma_{\Psi_2,G_2}(\omega)e_2(\omega) \\ \delta\langle\Psi_3\rangle(\omega) = \Gamma_{\Psi_3,G_1}(\omega)e_1(\omega) + \Gamma_{\Psi_3,G_2}(\omega)e_2(\omega). \end{cases} \tag{2.132}$$

We want to surrogate the two forcings using two observables, as follows:

$$\delta\langle\Psi_3\rangle(\omega) = H_{31}(\omega)\delta\langle\Psi_1\rangle(\omega) + H_{32}(\omega)\delta\langle\Psi_2\rangle(\omega). \tag{2.133}$$

We plug the RHS in (2.132) into (2.133), obtaining the following (we drop the dependence on ω for clarity):

$$\begin{aligned} \Gamma_{\Psi_3,G_1}e_1 + \Gamma_{\Psi_3,G_2}e_2 = & H_{31}\Gamma_{\Psi_1,G_1}e_1 + H_{31}\Gamma_{\Psi_1,G_2}e_2 + \\ & + H_{32}\Gamma_{\Psi_2,G_1}e_1 + H_{32}\Gamma_{\Psi_2,G_2}e_2. \end{aligned} \tag{2.134}$$

We use now the fact that the time patters are arbitrary, since they do not appear in the definition of the linear response functions. We can thus express (2.134) in a matricial form:

$$\begin{pmatrix} \Gamma_{\Psi_1,G_1} & \Gamma_{\Psi_2,G_1} \\ \Gamma_{\Psi_1,G_2} & \Gamma_{\Psi_2,G_2} \end{pmatrix} \begin{pmatrix} H_{31} \\ H_{32} \end{pmatrix} = \begin{pmatrix} \Gamma_{\Psi_3,G_1} \\ \Gamma_{\Psi_3,G_2} \end{pmatrix}. \tag{2.135}$$

which can be recasted in the following form, in order to explicit the surrogate response functions:

$$\begin{pmatrix} H_{31} \\ H_{32} \end{pmatrix} = \begin{pmatrix} \Gamma_{\Psi_1,G_1} & \Gamma_{\Psi_2,G_1} \\ \Gamma_{\Psi_1,G_2} & \Gamma_{\Psi_2,G_2} \end{pmatrix}^{-1} \begin{pmatrix} \Gamma_{\Psi_3,G_1} \\ \Gamma_{\Psi_3,G_2}. \end{pmatrix}, \tag{2.136}$$

where we have assumed the invertibility of the considered matrix. Once we obtain the surrogate response H from the relation (2.136), we can go back in the time domain in order to compute $\delta\langle\Psi_3\rangle(t)$:

$$\delta\langle\Psi_3\rangle(t) = H_{31}(t) * \delta\langle\Psi_1\rangle(t) + H_{32}(t) * \delta\langle\Psi_2\rangle(t), \tag{2.137}$$

which do not depend on the time patterns $e_{1,2}(t)$ at all.

We can generalize the derivation of the surrogate response functions considering N independent forcings:

$$F(x) \Rightarrow F(x) + \sum_{l=1}^{N} e_l(t)G_l(x) \tag{2.138}$$

and $N+1$ independent observables $\Psi_{1,\dots,N+1}$. We want to express the linear change of the expectation value of an observable as a function of the other N observables;

$$\delta\langle\Psi_{N+1}\rangle(\omega) = \sum_{l=1}^{N} H_{N+1,l}(\omega)\delta\langle\Psi_l\rangle(\omega). \tag{2.139}$$

As before, we have that $\delta\langle\Psi_k\rangle(\omega) = \sum_{j=1}^{N} \Gamma_{\Psi_k,G_j}(\omega)e_j(\omega)$, where Γ_{Ψ_k,G_j} is the response function related to the observable Ψ_k and the space pattern G_j. We plug these relations into (2.139), obtaining (dropping the dependence on ω):

$$\begin{pmatrix} \Gamma_{\Psi_1,G_1} & \dots & \Gamma_{\Psi_N,G_1} \\ \vdots & \ddots & \vdots \\ \Gamma_{\Psi_1,G_N} & \dots & \Gamma_{\Psi_N,G_N} \end{pmatrix} \begin{pmatrix} H_{N+1,1} \\ \vdots \\ H_{N+1,N} \end{pmatrix} = \begin{pmatrix} \Gamma_{\Psi_{N+1},G_1} \\ \vdots \\ \Gamma_{\Psi_{N+1},G_N} \end{pmatrix} \tag{2.140}$$

$$\begin{pmatrix} H_{N+1,1} \\ \vdots \\ H_{N+1,N} \end{pmatrix} = \begin{pmatrix} \Gamma_{\Psi_1,G_1} & \dots & \Gamma_{\Psi_N,G_1} \\ \vdots & \ddots & \vdots \\ \Gamma_{\Psi_1,G_N} & \dots & \Gamma_{\Psi_N,G_N} \end{pmatrix}^{-1} \begin{pmatrix} \Gamma_{\Psi_{N+1},G_1} \\ \vdots \\ \Gamma_{\Psi_{N+1},G_N} \end{pmatrix}, \tag{2.141}$$

where the N Fourier transform of the surrogate response functions are explicited by means of an expression containing a priori all the possible response functions Γ_{Ψ_k,G_j}. As a consequence, the poles of the surrogate response functions are no more the poles of the response functions, but a function of them. Then, if few observables are not able to predict the state of another observable Φ, it can happen (considering more forcings) that a combination of these non predictive observables can succeed in predicting Φ, as observed in the climatic model Lorenz 96 in [28]. This is reasonable: forcing the system in more ways and considering more observables as surrogates of the perturbations we gain more information and thus more predictive power of the system. Once we have obtained the surrogate response functions, we plug them into (2.139), obtaining the following expression after an inverse Fourier transform:

$$\delta\langle\Psi_{N+1}\rangle(t) = \sum_{l=1}^{N} H_{N+1,l}(t) * \delta\langle\Psi_l\rangle(t). \tag{2.142}$$

The generality of the derivation of response theory and surrogate response theory enables them to be applied in several scientific disciplines, providing new insights and points of view. In the following two chapters we apply the surrogate response theory in two diverse contexts: control theory and in the climatic model Lorenz 96.

2.6.2 A new ratio method to quantify unpredictability

As we have seen before, the surrogate response functions H_{12} (between two generic observables Ψ_1 and Ψ_2) could be non-predictive. In other words, they have a non-vanishing non-causal component with support on the negative times. The presence of the non-causal component hinders the prediction of Ψ_1 at time t using just the time behaviour of Ψ_2 up to time t. An interesting problem is to actually quantify the non-causal component of the surrogate response function. This problem can emerge in a variety of situations where we have more non-predictive surrogate response functions and we want to choose between them the one which provides the best prediction. For example, we could have a set of observables $\{\Psi_1, ..., \Psi_m\}$ and we want to find the one that better predicts another observable Ψ_a. If one observable Ψ_k is such that the surrogate response function $H_{a,k}$ is predictive, we choose Ψ_k without hesitations. The choice becomes hard when all the picked observables $\{\Psi_1, ..., \Psi_m\}$ are not predictive. We want to have a method which allows us to choose the best predictive one between them. Another setting where such a method could be helpful is that of two observables

Ψ_a and Ψ_b which try to predict each other. We would like to understand whether it is Ψ_a which predicts better Ψ_b or the contrary since a better prediction power could be linked to a causal link or a flow of information with a definite verse from an observable to the other. It could happen that both the two surrogate response functions $H_{a,b}$ and $H_{b,a}$ are non-predictive. In that case, the existence of the before-mentioned method would be crucial to identify the observable with more prediction power.

We now present such a method, developing the idea behind it. If the surrogate response function H_{12} is predictive, a set of KK relations could be derived from it. In particular, following the derivation presented in Section 2.4, for H_{12} we would have:

$$H_{12}(t) = \Theta(t) H_{12}^c(t), \tag{2.143}$$

where $H_c(t)$ is the causal part of H_{12}. Performing the Fourier transform of (2.143) we get:

$$\begin{aligned} H_{12}(\omega) =& \Theta(\omega) * H_{12}^c(\omega) \\ =& \left(\frac{1}{2}\delta(\omega) + \frac{i}{2\pi} P\left[\frac{1}{\omega}\right]\right) * H_{12}^c(\omega) \\ =& \frac{1}{2} H_{12}^c(\omega) + \frac{i}{2\pi} P\left[\frac{1}{\omega}\right] * H_{12}^c(\omega). \end{aligned} \tag{2.144}$$

Let's see how the KK relation (2.144) changes with the presence of a non-predictive component H_{nc} in the surrogate response function:

$$H_{12}(t) = \Theta(t) H_{12}^c(t) + \Theta(-t) H_{12}^{nc}(t). \tag{2.145}$$

We perform the Fourier transform of (2.145) obtaining the following modified KK relation:

$$\begin{aligned} H_{12}(\omega) =& \Theta(\omega) * H_{12}^c(\omega) + \Theta(-\omega) * H_{12}^{nc}(\omega) \\ =& \left(\frac{1}{2}\delta(\omega) + \frac{i}{2\pi} P\left[\frac{1}{\omega}\right]\right) * H_{12}^c(\omega) + \left(\frac{1}{2}\delta(\omega) - iP\left[\frac{1}{\omega}\right]\right) * H_{12}^{nc}(\omega). \end{aligned} \tag{2.146}$$

We can see in (2.146) that for each ω there are two contributes for the Fourier transform of the response function: a causal one and a not-causal one. We notice that in the limit of vanishing not-causal component, (2.146) reduces to the usual KK relation (2.144).

Looking at (2.146) we can deduce that the non-causal part is more important as the second term grows. As a consequence, we could assess the importance of the non-causal component with respect to the causal-component quantifying the norm of those two terms. In particular, we choose the L_2 norm to quantify the norm of a function, which is defined as follows for a complex function $g(x)$:

$$\|g\|_2 \equiv \int dx\, g(x) g^*(x) = \int dx\, |g(x)|^2. \tag{2.147}$$

Given that, we say that to assess the importance of the non-causal component of the response function H_{12}, we have to look at the following ratio:

$$R(H_{12}) \equiv \frac{\|\Theta(-\omega) * H_{12}^{nc}(\omega)\|_2}{\|\Theta(\omega) * H_{12}^c(\omega)\|_2}. \tag{2.148}$$

We can translate the ratio (2.148) in the time domain using the Plancherel Theorem, which states that the L_2 norm of a function $g(t)$ is equal to the L_2 norm of its Fourier transform:

$$\|g(t)\|_2 = \|g(\omega)\|_2. \tag{2.149}$$

As a consequence, we can say that the following ratio quantifies the unpredictable component of a surrogate response function:

$$R(H_{12}) \equiv \frac{\|\Theta(-t)H_{12}^{nc}(t)\|_2}{\|\Theta(t)H_{12}^{c}(t)\|_2}. \tag{2.150}$$

The practical use of the method (2.150) is the following: we take the non-causal component $\Theta(-t)H_{12}^{nc}(t)$ of the surrogate response function H_{12} and we compute its L_2 norm:

$$\|\Theta(-t)H_{12}^{nc}(t)\|_2 = \int_{-\infty}^{0} dt\, |H_{12}^{nc}|^2(t). \tag{2.151}$$

We do the same for the causal component:

$$\|\Theta(t)H_{12}^{c}(t)\|_2 = \int_{0}^{\infty} dt\, |H_{12}^{c}|^2(t), \tag{2.152}$$

then we take the ratio between these two quantities, obtaining (2.150). We can see that the method (2.150) is very straight-forward and easy to implement. Note that if there is no non-causal component, the ratio is vanishing: the surrogate response function is predictive. Moreover, since this method revolves around the response function, it does not depend on the chosen time pattern. The relation (2.150) provides a method a priori which permits to quantify the unpredictive power of the surrogate response function. In Section 4.2.2 we will actually see the effectiveness of this method.

Note that in the causal component $H^c(t)$ there could be an eventual singular component $S(t)$, as in (2.127). In that case the integral over time of the singular part $\int_0^\infty S(t)dt$ could be divergent since generally, it is a sum of derivatives of delta functions. One could think to use a method like (2.150) but just for the non-singular component of the surrogate response function.

A pedagogical example

We now consider a very simple example, to show the idea behind the effectiveness of the ratio method (2.150). We consider the following response function:

$$H(t) = \Theta(t)A\, e^{-at} + \Theta(-t)B\, e^{bt}, \tag{2.153}$$

where the coefficients a and b are positive while the constants A and B are real numbers. The non-causal component is $H^{nc} = B\, e^{bt}$ and the causal component is $H^c(t) = A\, e^{-at}$. Note that the time scale related to the causal component is $1/a$, while the time scale related to the non-causal component is $1/b$. The smaller is the coefficient, the bigger is the time scale and the importance of the related component in the response function. Note that the Fourier transform of this function is the following:

$$H(\omega) = \frac{iA}{\omega + ia} + \frac{(-i)B}{\omega - ib}, \tag{2.154}$$

hence it falls in the general category of rational functions (2.122). The ratio method (2.150) takes the following form for the surrogate response function (2.153):

$$\begin{aligned} R(H) =& \left(\frac{B^2}{2b}\right) / \left(\frac{A^2}{2a}\right) \\ =& \frac{B^2}{A^2} \cdot \frac{a}{b}, \end{aligned} \tag{2.155}$$

which depends on the interplay between the time scales of the causal and not causal components and on the interplay between the magnitude of these components. We can see from the ratio method (2.155) that if the time scale $1/b$ of the non-causal component H^{nc} is much smaller than the time scale $1/a$ of the causal component, the ratio goes to zero. This is reasonable: in that case the surrogate response function has a very important non-causal component. A similar conclusion can be drawn considering a non causal component whose magnitude B is much smaller than the magnitude A of the causal term. Also in that case the ratio method goes to zero, as expected by a predictive surrogate response function. In this very simple example we can see that the ratio method is very useful to quantify the relation between the causal and not-causal components of the response function. We can extend this line of thinking to the following more general surrogate response function:

$$H(t) = \Theta(t) \sum_{k=1}^{\infty} A_k \, e^{-a_k t} + \Theta(-t) \sum_{k=1}^{\infty} B_k \, e^{b_k t}, \tag{2.156}$$

where both components have infinite terms, each one with its time scale and its magnitude. The ratio method (2.150) takes the following form, which generalizes (2.155):

$$R(H) = \left(\sum_{k=1}^{\infty} \frac{B_k^2}{b_k}\right) / \left(\sum_{k=1}^{\infty} \frac{A_k^2}{a_k}\right), \tag{2.157}$$

where it is present a complicated relation between the various time scales and magnitudes of infinite causal and not causal terms. Each one of these terms is weighted by its magnitude and its time scale. The basic limits considered in the context of (2.155) can be taken also for the relation (2.157): if the time scale or the magnitude of one term vanishes, the importance of the relative component (be it causal or not causal) is drastically reduced with respect to the other.

Chapter 3

Surrogate RT and Control in Linear Systems

In Section 2.6 we have introduced the surrogate RT: a fairly general theory which focuses on understanding when perturbed observables can be used as surrogates of the actual forcing to predict the state of another perturbed observable. There are cases where the prediction is possible and other cases where it is not, due to the non-causal nature of the surrogate response function. In this chapter we apply such a theory in the engineering context of control theory. This chapter is organized as follows. In **Section 3.1** we present our setting, which consists in linearized models around asymptotically stable equilibrium points [29, 30, 64]. Then, we derive response functions (or transfer functions) and surrogate response functions in this linearized context. In **Section 3.2** we present the main result of the chapter: the Unpredictability Criterion (UC). This Criterion allows us to rule out a priori the dynamical variables of our system which cannot be used for prediction purposes. The UC takes a very explicit and simple form, which makes it amenable to use in many applications, as listed above in the contexts of control theory and rational approximation theory. Next, we provide the proof of this theorem and a simple yet hopefully illustrative numerical example of an application of the UC. In **Section 3.3** we add memory effects to the linear systems considered before and we see how the UC changes its form. We observe that because the variables know more of each other it is less likely that they are unpredictive. In Section 3.4 we take a different point of view, considering linear systems defined by random matrices. After discussing the stability conditions for these systems, we derive an average UC, which tells us on average whether a dynamical variable is unpredictive. In **Appendix B** we show a simple pedagogical example to show the idea beneath the UC.

3.1 Our setting

3.1.1 Linearized dynamical systems

Let's consider a generic dynamical system (sometimes referred to as process in the control theory literature [29, 30]):

$$\dot{\vec{y}} = \vec{F}(y), \tag{3.1}$$

where $\vec{y}$ is a vector of N dynamical variables which belongs to some phase space $\Omega \subseteq \mathbb{R}^N$. We now suppose that this dynamical system has an *equilibrium point* $\tilde{y}$, i.e. a point in the phase space such that $F(\tilde{y}) = 0$. We can now expand the function $\vec{F}$ around the steady-state point $\tilde{y}$, obtaining:

$$\dot{y}_i(t) = -\sum_j A_{ij}(y_j(t) - \tilde{y}_j) + ..., \qquad A_{ij} = -\frac{\partial F_i}{\partial y_j}_{|_{y=\tilde{y}}} \tag{3.2}$$

We introduce the variables $\vec{x} = \vec{y} - \tilde{y}$ in such a way to have the fixed point in 0:

$$\dot{x}_i(t) = -\sum_j A_{ij}x_j(t) + ..., \tag{3.3}$$

In our following analysis we will focus just on equilibrium points which are asymptotically stable, i.e. stable and attractive.[1] Doing so, the matrix A will have just eigenvalues with positive real part [74] .

3.1.2 Adding the input: perturbing the linearized dynamical system

Response functions

Let's consider the dynamical system (3.3), linearized around an asymptotically stable equilibrium $\tilde{x} = 0$. Starting from the system at the steady-state, we perturb the system with a small perturbation, with spatial pattern $\{G_i\}_i$ (with at least one $G_i \neq 0$) and time pattern $e(t)$:

$$\dot{x}_i(t) = -\sum_j A_{ij}x_j(t) + G_i e(t). \tag{3.4}$$

What we want to do now is to build the response functions $\Gamma_{i,\mathrm{gl}}$ which connect the perturbation $G_i e(t)$ (the input) with the response of the system seen through the dynamical variable x_i (the output). Those response functions are called transfer functions in the control theory literature [29]. In our setting, we can find those response functions with just a few steps of computations.

Let's take the Fourier transform of the equation (3.4), looking at it component by component:

$$-i\omega x_i(\omega) = -\sum_j A_{ij}x_j(\omega) + G_i e(\omega) \tag{3.5}$$

We can now solve this equation with respect to $x(\omega)$:

$$x_i(\omega) = \sum_j (A - i\omega\mathbb{1})^{-1}_{ij} G_j e(\omega), \tag{3.6}$$

where $(A - i\omega\mathbb{1})^{-1}_{ij}$ is the $(i,j)-$entry of the matrix $(A - i\omega\mathbb{1})^{-1}$. From equation (3.6) we can extract the Fourier transform of the response function which we refer to as

[1] We recall that an equilibrium point $\tilde{x}$ is *stable* if for each neighbourhood U of $\tilde{x}$ there is another neighbourhood U_0 of $\tilde{x}$ such that $\Phi^F_t(U_0) \subseteq U$ for all t, where Φ^F_t is the flow associated to the map F. Moreover, we recall that an equilibrium point is *attractive* if it exists a neighbourhood V of $\tilde{x}$ such that for any $\vec{x_0} \in V$, we have that $\lim_{t\longrightarrow\infty} \vec{x}(t) = \tilde{x}$.

the global response function because it takes into account all the contributes of the dynamical variables:

$$\Gamma_{i,\mathrm{gl}}(\omega) = \sum_{j=1}^{N} \Gamma_{ij}(\omega), \tag{3.7}$$

where:

$$\Gamma_{ij}(\omega) = G_j(A - i\omega\mathbb{1})^{-1}_{ij}. \tag{3.8}$$

It is important to remark that in (3.8) there is no summation over the index j, the constant Gj is just multiplied. We can rewrite (3.8) using the Cramer's rule:

$$\Gamma_{ij}(\omega) = \frac{G_j}{\det[A - i\omega\mathbb{1}]} C_{ji}(\omega), \tag{3.9}$$

where C_{ji} is the determinant of the minor M_{ji}, i.e. the matrix obtained removing the row j and the column i from the full matrix $A - i\omega\mathbb{1}$. If we perform the inverse Fourier transform of (3.6), we obtain the response relation [4]:

$$x_i(t) = \int_{-\infty}^{t} \Gamma_{i,\mathrm{gl}}(t - \tau) e(\tau) d\tau, \tag{3.10}$$

where the effect of the perturbation at time $\tau < t$ on the dynamical variable x_i at time t is mediated by the response functions Γ_{ij}, which depend just on the perturbed observable and the space pattern of the forcing (not on the time pattern of the forcing). The response function defined in (3.10) has support just for positive times. This is due to the fact that the poles of its Fourier transform (3.8) are all located in the lower complex $\omega-$plane, i.e. they all have negative imaginary part (this can be noticed very quickly observing that at the denominator there is the secular equation for the matrix A, which eigenvalues all have positive real part). As a consequence, the response function is predictive: we can predict the state of the perturbed dynamical variable $x_i(t)$ using just the perturbation $e(\tau)$ up to time t, as it is shown in (3.10).

Surrogate response functions

A new angle of the problem in RT has been introduced in [28] where, given a perturbation of a rather general statistical mechanical system,response relations *between perturbed observables* are built. Let's build them in our setting. We consider the Fourier Transform of two response relation of the kind (3.10), for the dynamical variables x_i and x_j:

$$\begin{cases} x_i(\omega) = \Gamma_{i,\mathrm{gl}}(\omega) e(\omega) \\ \\ x_j(\omega) = \Gamma_{j,\mathrm{gl}}(\omega) e(\omega) \end{cases} \tag{3.11}$$

Now we take the ratio between the two equations in (3.11), getting:

$$x_i(\omega) = \frac{\Gamma_{i,\mathrm{gl}}(\omega)}{\Gamma_{j,\mathrm{gl}}(\omega)} x_j(\omega). \tag{3.12}$$

We can define the Fourier transform of the surrogate response function H_{ij} between x_i and x_j as follows:

$$H_{ij}(\omega) \equiv \frac{\Gamma_{i,\mathrm{gl}}(\omega)}{\Gamma_{j,\mathrm{gl}}(\omega)}. \tag{3.13}$$

If we go back in the time domain, performing the inverse Fourier transform(3.12), we have the following response relation (we remark that in the following formula there is no implicit index summation):

$$x_i(t) = \int_{-\infty}^{\infty} H_{ij}(t-\tau)x_j(\tau). \tag{3.14}$$

H_{ij} is called a surrogate response function because the perturbed observable x_j in (3.14) acts as a surrogate for the forcing to predict the state of the other perturbed observable x_i. The variable x_j is reconstructing the state of the other dynamical variable x_i.

As discussed for the response functions Γ in section 3.1.2, we want to deal with response functions which have predictive power, i.e. they are zero for negative times. Only if the surrogate response functions have this feature, they can be used for predictions purposes. To have the surrogate response functions H predictive, it is necessary that their Fourier transforms have no poles in the upper complex $\omega-$plane. In order to check that, we have to consider (3.13), looking for the zeros of the global response function at the denominator (the poles of the global response function at the numerator are not in the upper complex $\omega-$plane, as discussed in 3.1.2).

3.2 Unpredictability Criterion

We may wonder whether we can find a criterion that allows us to discover *a priori* whether an observable j is actually able to make predictions by looking at the properties of the matrix A and of the vector G. We claim to have found in our setting such a criterion, which can be used to study the properties of all variables x_j such that the corresponding component G_j of the forcing is not vanishing. Note that, unless symmetries are present in the system, this limitation is only minor. We first define the following quantity:

$$SR_j[A] = \sum_{k=1}^{N} A_{jk}G_k, \tag{3.15}$$

i.e. the sum of all the elements in the row j of the matrix A of the linearized system (3.3), weighted by the spatial pattern of the forcing G_j. Given that definition, we present the claimed criterion:

Unpredictability Criterion (UC): We consider the perturbed problem (3.4) with a spatial pattern of the forcing $G_j \neq 0$ for some j. If the following inequality is valid for the perturbed dynamic variable x_j (with $G_j \neq 0$), then that variable is generally *not* able to predict the future state of other variables[2]:

$$SR_j(A) > G_j Tr[A]. \tag{3.17}$$

[2]A particular case where two not-predictive variables x_i and x_j can predict each other can happen when their global response function Γ present the same zeros in the complex upper $\omega-$plane. In this case the surrogate response function H_{ij} is *predictive* since the singularity is cancelled:

$$H_{ij} = \frac{\Gamma_{i,\mathrm{gl}}}{\Gamma_{j,\mathrm{gl}}} \sim \frac{\Pi_i \cancel{(\omega - \omega_i^{NP})}}{\Pi_i \cancel{(\omega - \omega_i^{NP})}} \tag{3.16}$$

This predictive power emerges from the particular combination of the observables i and j and the spatial pattern of the forcing G. As a consequence, the occurrence of such a zero/pole cancellation is

We can see that it is a very simple and efficient criterion. Using the UC we can rule out a priori some of the variables which cannot be used for prediction purposes.
Its physical meaning can be explained as follows. Let's rewrite the criterion (3.17), removing from both sides the single term $G_j A_{jj}$:

$$\sum_{k\neq j} A_{jk} G_k > G_j \sum_{h\neq j} A_{hh}. \tag{3.18}$$

The quantity at the LHS could be seen as the *linear feedback* from the system to the dynamical variable j, while the RHS as the total dissipation encountered in the system (as it can be seen in the context of (2.19)).

The LHS can be seen as measuring the total information received by x_j from the whole system, while the quantity at the RHS as the total information retained by the other variables x_h. We can say that the sense of the criterion is that the observable j is not able to predict the state of the other variables if too much information arrives at j from them. In Appendix B we provide an explicit derivation of the UC for the case of systems with 3 degrees of freedom. In Section 3.3 we show how the UC changes introducing memory effects in the dynamical system.

In control theory one of the detrimental problems which can emerge in designing controllers is the presence of zeros with positive imaginary part in the response functions. The dynamical systems which shows this feature are called non-minimum phase systems [29]. These zeros can be linked with delays of the response of the system to a perturbation [29]. We can see this very quickly considering a generic response function Γ whose Fourier transform has a zero $i\omega_{NP}$ with $\omega_{NP} > 0$ and performing a Padé approximation [75]. The Padé approximant of a function $f(x)$ is a ratio of polynomials of order p and q such that it better approximates $f(x)$ for increasing p and q:

$$f(x) \approx \frac{\sum_{i=1}^{p} a_i x^i}{\sum_{j=1}^{q} b_j x^j}, \tag{3.19}$$

where the coefficients a_i and b_j generally depend on the orders p and q. For the exponential function $f(x) = e^{-x}$ the Padé approximant is:

$$f(x) = e^{-x} \approx \frac{\sum_{i=1}^{p} \frac{(p+q-j)!p!}{(p+q)!j!(p-j)!}(-x)^i}{\sum_{j=1}^{q} \frac{(p+q-j)!q!}{(p+q)!j!(q-j)!}(x)^j}. \tag{3.20}$$

For small x, we can truncate at the first order considering $p = q = 1$:

$$f(x) = e^{-x} \approx \frac{1 - x/2}{1 + x/2}. \tag{3.21}$$

At this point we make the assumption that the non-predictive pole ω_{NP} has a large magnitude. This is equivalent to assume that the surrogate response functions whose

not generic and not robust. In fact, it is sufficient to slightly modify the system so that the two poles do not match anymore.

use the observable related to Γ as predictor have a small non-causal component, as explained in Section 2.6.2. We can thus use the Padé approximant (3.21) for the exponential function in the Fourier transform of the response function Γ:

$$\begin{aligned}\Gamma(\omega) =&(\omega - i\omega_{NP})R(\omega)\\ =&\frac{(\omega - i\omega_{NP})}{(\omega + i\omega_{NP})}(\omega + i\omega_{NP})R(\omega)\\ \approx& - e^{-2\left(\frac{\omega}{i\omega_P}\right)}(\omega + \omega_{NP})R(\omega)\end{aligned} \tag{3.22}$$

where $R(\omega)$ is the remaining part of the response function. The exponential factor in the frequency space is translated in a delay in the time domain. In fact, defining $\tilde{\Gamma}(\omega) = -R(\omega)(\omega + \omega_{NP})$, we have:

$$\begin{aligned}\Gamma(t) =& \int \frac{d\omega}{2\pi}\tilde{\Gamma}(\omega)e^{-i\omega t}\\ \approx& \int \frac{d\omega}{2\pi}e^{-2\left(\frac{\omega}{i\omega_{NP}}\right)}\tilde{\Gamma}(\omega)e^{-i\omega t}\\ \approx& \tilde{\Gamma}(t - \frac{2}{\omega_{NP}}).\end{aligned} \tag{3.23}$$

We can see that the smaller is the imaginary part ω_{NP} of a non-predictive zeros, the longer is the delay of the response and then the more difficult is the system to control (since it is slower to respond to the given input). For smaller ω_{NP} we assume that a delay is still present in the response function, generally different from $1/\omega_{NP}$. It is noticeable that the presence of a zero with positive imaginary part in a response function of an observable is the same condition which hampers the predictive power of that observables in the surrogate RT. As a consequence, if we find using the UC that some observable of a dynamical system cannot make predictions of other observables, we deduce that the considered system is necessarily a non-minimum one. The presence of a Criterion which can immediately establish whether a system is non-minimum or not can have relevant industrial applications in designing controllers, since those systems have controllability issues [29, 33].

3.2.1 Proof of the Unpredictability Criterion

Let's prove the UC for general matrices NxN, with the spatial pattern of the perturbation $G_i \neq 0$ for some i. We consider the global response function for the dynamical variable x_j, with $G_j \neq 0$:

$$\Gamma_{j,\text{gl}} = \sum_{k=1}^{N}\Gamma_{jk},$$

and we look for its zeros. We plug the relation (3.9) into each of the terms Γ_{jk}. Doing that, we have that looking for the zeros of $\Gamma_{j,\text{gl}}$ is equivalent to finding the $N-1$ roots of a polynomial in ω of order $N-1$, i.e. the following equation (the determinant at the denominator in (3.9) has been simplified):

$$(-i\omega)^{N-1} + \alpha_{N-2}(-i\omega)^{N-2} + ... + \alpha_0 = 0. \tag{3.24}$$

We make the rewriting $(-i\omega) = x$, obtaining:

$$x^{N-1} + \alpha_{N-2}x^{N-2} + ... + \alpha_0 = 0, \tag{3.25}$$

where (recalling that $G_j \neq 0$):

$$\alpha_{N-2} = \left(-\sum_{k\neq j} A_{jk}\frac{G_k}{G_j} + \sum_{h\neq j} A_{hh} \right). \tag{3.26}$$

Now we prove that if the coefficient α_{N-2} is negative, the equation (3.25) admits at least one solution with positive real part. There are two ways to prove it:

- Using the Vieta's formula [76]:

$$-\frac{\alpha_{N-2}}{\alpha_{N-1}} = x_1 + ... + x_{N-1}, \tag{3.27}$$

  where x_i is the $i-$th solution of (3.25) and α_{N-1} is the coefficient of the $(N-1)$-power term. At this point, since α_{N-2} is negative and $\alpha_{N-1} = 1$, we can observe that if all the solutions $\{x_i\}$ have a negative real part, the Vieta's equation (3.27) is violated. Hence there should be at least one solution x_k with positive real part.

- Using the Routh–Hurwitz stability criterion [77, 78]. This method counts the number of roots with negative real part or positive looking at the coefficients of the polynomial. This method could be useful to count how many zeros actually are in the upper complex ω plane if we compute the whole series of coefficients in (3.25). In particular, it predicts that if the first two coefficients α_{N-1} and α_{N-2} of the polynomial have opposite sign, then there is at least one root x_k with a positive real part.

We have found that if the coefficient of the $(N-2)$-power term α_{N-2} is negative, there is at least one root x_k with a positive real part. This implies that the corresponding zero $\omega_k = i\,x_k$ of the global response function has an imaginary part positive. This gives the UC (3.17).

3.2.2 Numerical tests

In this section, we will test the effectiveness of the UC on a system of the type (3.4). We will see that these numerical test will support the predictions made by the UC regarding the unpredictive dynamical variables.

A remark: how to plot the surrogate response functions

We can compute the Fourier transform of the response functions $\Gamma_{i,\mathrm{gl}}(t)$ of systems of the type (3.4) using the relations (3.7) and (3.9). Taking the ratio of two response functions related to two dynamical variables x_i and x_j, we get the Fourier transform of the surrogate response function $H_{ij}(\omega)$ (3.13).
At this point, in order to clearly plot the surrogate response functions, we make the following consideration. Since the ratio between global response functions $\Gamma_{i,\mathrm{gl}}(t)$ is just a ratio between polynomials (3.9), we can decompose the Fourier transform of

the surrogate response function as the sum of two terms after performing the division between polynomials (2.123):

$$H_{ij}(\omega) = S_{ij}(\omega) + K_{ij}(\omega), \tag{3.28}$$

where S is just a polynomial whilst K is a fraction where the polynomial at the denominator has a smaller order with respect to the polynomial at the numerator. S is called the *singular* part, because if we perform the inverse Fourier transform on it we obtain a sum of Dirac delta and its derivatives. This singular part would make the plot of $H_{ij}(t)$ not easy to analyze. In order to remove it we study the behaviour of $H_{ij}(\omega)$, and we notice that it converges to a constant (we assume that the space pattern of the forcing is different from zero for both dynamical variable x_i and x_j, which will be the case in our example):

$$\lim_{\omega\to\infty} H_{ij}(\omega) = \lim_{\omega\to\infty} \frac{\sum_k G_k C_{ki}(\omega)}{\sum_k G_k C_{kj}(\omega)} = \frac{G_i}{G_j}. \tag{3.29}$$

In order to see that, we can notice that the dominant term in $\sum_k G_k C_{ki}(\omega)$ for large ω is $(-i\omega)^{N-1} G_i$, as we can see in (3.24) and, more explicitly, in the 3×3 case (B.2).

As a consequence, removing the constant $\frac{G_i}{G_j}$ from $H_{ij}(\omega)$, we remain with the non-singular part of the surrogate response function (3.28), whose plot does not have any delta function, making it easier to visualize. In all the plots of the following sections, just the non-singular part of the surrogate response functions is shown.

Examples of applicability of the Unpredictability Criterion

We consider the linear dynamical system (3.3) defined by the following 5×5 matrix:

$$A = \begin{pmatrix} 80 & 90 & 90 & 90 & 90 \\ 60 & 80 & 20 & 30 & 0 \\ 0 & 0 & 80 & 20 & 30 \\ 0 & 0 & 40 & 80 & 40 \\ 0 & 0 & 50 & 0 & 80 \end{pmatrix}, \tag{3.30}$$

This matrix has only eigenvalues with positive real part, hence the fixed point around which we are linearizing the system is asymptotically stable. We now consider two possible spatial pattern for the perturbation:

- $G_1^{(1)} = 3$ and $G_j^{(1)} = 1$ for $j \neq 1$. In this case the UC cannot be applied any dynamical variable x_1.
- $G_1^{(2)} = 0.5$ and $G_j^{(2)} = 1$ for $j \neq 1$. In this case the UC can be applied to the dynamical variable x_1. We can infer that this variable is not usable for prediction purposes, as we will show it explicitly in a numerical example below.

We want to see in both cases whether the dynamical variable x_1 can or cannot be used for prediction purposes, in order to test the UC. We first compute the global response functions of the dynamic variables $\Gamma_{1,3,5\text{gl}}$ using the relations (3.9) and (3.7). The results for both $G^{(1)}$ and $G^{(2)}$ are shown in Figure 3.1. Using the definition (3.13) we can compute the surrogate response functions H_{13}, H_{31}, H_{15} and H_{51}, for both spatial

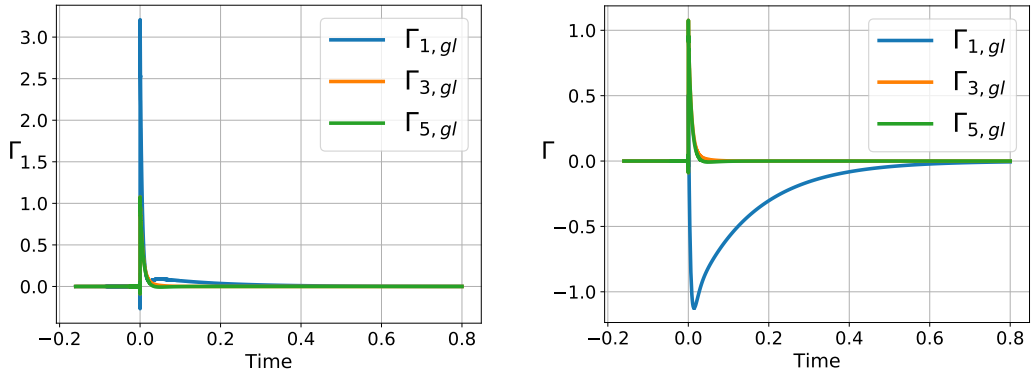


Figure 3.1: Global response functions $\Gamma_{1,3,5,\mathrm{gl}}$ in the time domain for x_1, x_3 and x_5 in the linear system defined by the 5×5 matrix (3.30) and space pattern of the perturbation $G^{(1)}$ (a) and $G^{(2)}$ (b).

pattern in Figure 3.2. We can notice that for $G^{(2)}$ the surrogate response functions H_{21} and H_{51} are not zero for negative times, hence they cannot be used for prediction purposes, as predicted by the UC.

Next, we test the effectiveness of the predictions made by our response functions and surrogate response functions for both $G^{(1)}$ and $G^{(2)}$. We perturb the system (3.30) with a perturbation with the following time pattern:

$$e(t) = \Theta(t)\left[0.05\sin(\pi\, t) - 0.1\sin\left(\frac{6}{7}\pi\, t\right) + 0.04\sin(8\pi\, t)\right]$$

We focus on the response of the dynamical variable x_3. We predict its time behaviour using $\Gamma_{3,\mathrm{gl}}$, making the convolution of it with the new time pattern as in (3.10). We can see in Figure 3.3 that the comparison between this prediction and the time behaviour of x_3 is excellent for both the spatial patterns of the forcing. We now predict the time behaviour of x_3 using the surrogate response functions H_{35} and H_{31}, making the convolution of them with with x_5 and x_1 respectively. We want to use the formula (3.14), but it does not provide causality, since it considers all the trajectory in time of x_5 and x_1. Since it is a prediction, we have to enforce causality in the formula (3.14), considering the trajectory of x_1 and x_5 up to time t to predict the state of x_3 at time t, using the modified surrogate response function (2.121):

$$x_3(t) = \int_{-\infty}^{t} H_{31}(t-\tau)x_1(\tau). \tag{3.31}$$

We can see in Figure 3.3(a) and (c) that the comparison between the time behaviour of x_3 and the prediction with H_{35} is excellent for both $G^{(1)}$ and $G^{(2)}$, whilst the comparison between x_3 and the prediction with H_{31} is bad for $G^{(2)}$ and excellent for $G^{(1)}$, as expected. This was expected since we found before, using the UC, that x_1 cannot be used for prediction purposes for $G^{(2)}$. In the panels (b) and (d) the predictions are made without enforcing causality, hence they work well.

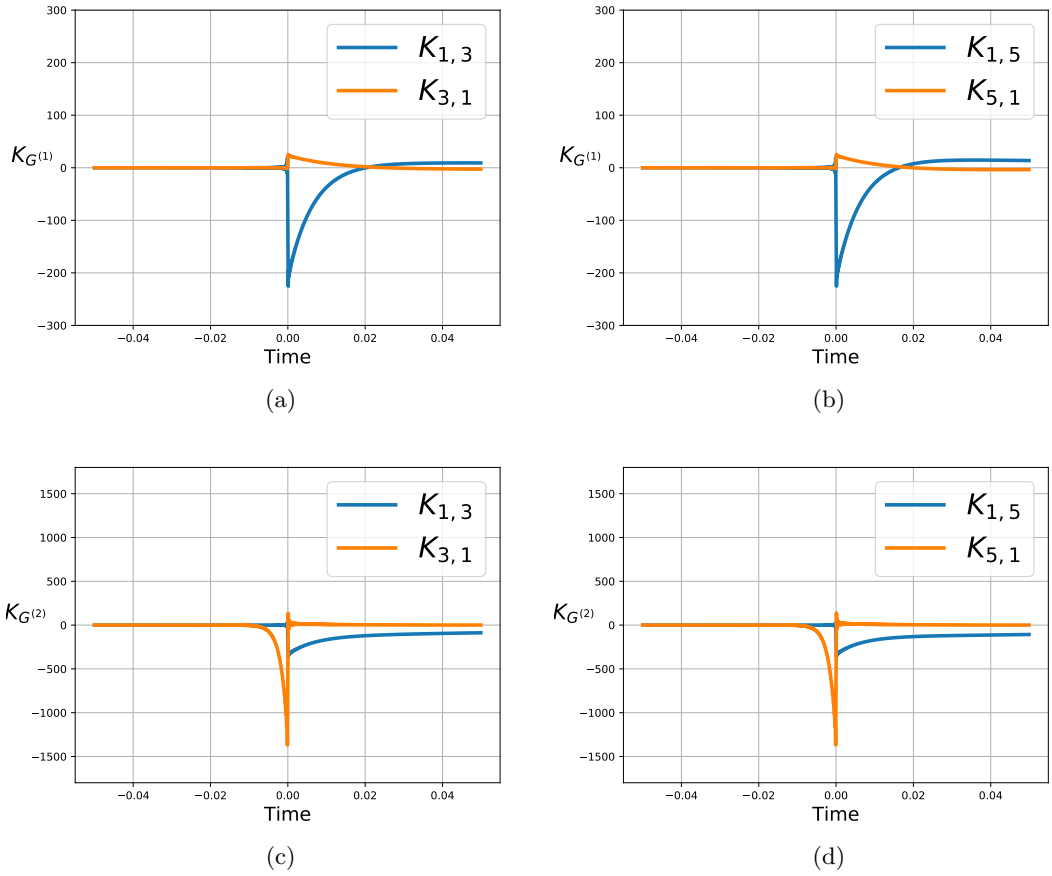


Figure 3.2: We consider the linear system defined by the 5×5 matrix (3.30). (a) Non singular components K_{13} and K_{31} using $G^{(1)}$. (b) Non singular components K_{15} and K_{51} using $G^{(1)}$. (c) shows the non singular components K_{13} and K_{31} using $G^{(2)}$. (d) shows the non singular components K_{15} and K_{51} using $G^{(2)}$. We can notice that in the case of $G^{(2)}$ we have that x_1 is not predictive since the support of K_{31} and K_{51} is not restricted to the positive semi-domain.

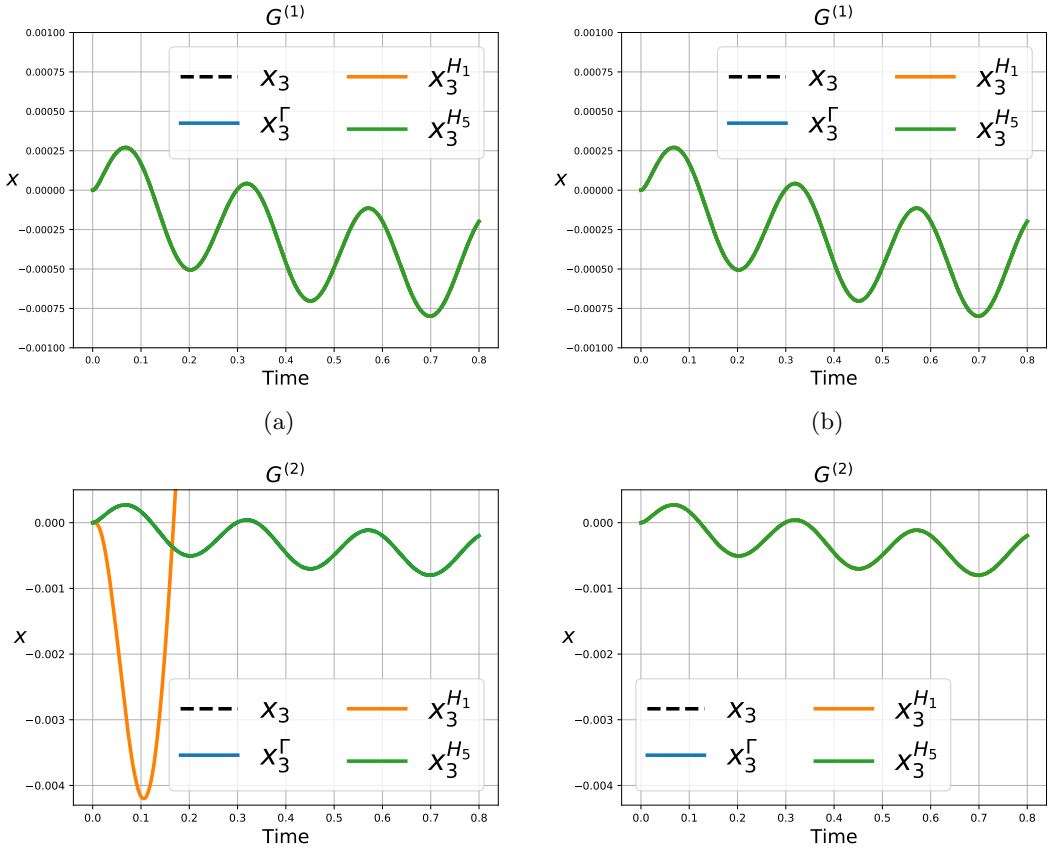


Figure 3.3: (a) and (b): Time behaviour of the response of x_3 (dashed black line) to the perturbation with time pattern $e(t) = 0.05\sin(\pi\, t) - 0.1\sin(\frac{6}{7}\pi\, t) + 0.04\sin(8\pi\, t)$ and spatial pattern $G^{(1)}$. in the system defined by the 5×5 matrix (3.30). The coloured lines predict the response of x_3 using the response function $\Gamma_{3,\mathrm{gl}}$ (blue line), the surrogate response functions H_{31} (orange line) and H_{35} (green line). The prediction with H_{31} in (a) is made enforcing causality, as in (3.31), whilst the one in (b) is made with the full (3.14). (c) and (d): same graphs but with spatial pattern of the perturbation $G^{(2)}$. We can see in (c) that in this case x_1 cannot be used to predict x_3.

3.3 Memory effects and UC

The goal of this section is to examine how the UC (3.17) changes considering the perturbed linear dynamical system (3.4) with additional Non-Markovian memory terms:

$$\dot{x}_i = -A_{ij}x_j + G_i e(t) + \int_{-\infty}^{t} M_{ij}(t-s)x_j(s), \tag{3.32}$$

where the M_{ij} are the memory kernels. We choose the following form for the memory kernel (like in the Mori-Zwanzig formalism [21–23, 34]):

$$M_{ij}(t) = \Theta(t)\alpha_{ij}e^{-\beta_{ij}t}, \qquad \alpha_{ij} \in \mathbb{R}, \quad \beta_{ij} > 0, \tag{3.33}$$

with Fourier transform:

$$M_{ij}(\omega) = \frac{i\alpha_{ij}}{\omega + i\beta_{ij}}. \tag{3.34}$$

The condition $\beta_{ij} > 0$ is necessary in order to have the memory kernel causal, i.e. to have the Heaviside theta in (3.33). At this point we Fourier transform the perturbed linear system looking for the Fourier transform of the response function, as we have done in Sec 3.1.2:

$$-i\omega x_i = -\sum_j A_{ij}x_j + G_i e(\omega) + \sum_j M_{ij}(\omega)x_j$$

$$\Rightarrow x_i(\omega) = \sum_j (A - i\omega\mathbb{1} - M(\omega))^{-1}_{ij} G_j e(\omega).$$

Hence we obtain the Fourier transform of the global response function for the dynamical variable x_i (we recall that the factor G_j is just multiplied to the element of the inverse matrix):

$$\Gamma_{i,\mathrm{gl}} = \sum_{j=1}^{N} G_j (A - i\omega\mathbb{1} - M(\omega))^{-1}_{ij}. \tag{3.35}$$

We want now to see how the UC (3.17) changes with the introduction of the memory effects. To see that, we follow the proof of the UC in Sec 3.2.1, looking for the zeroes of the global response function of a dynamical variable x_j such that $G_j \neq 0$. Finding the zeroes of $\Gamma_{i,\mathrm{gl}}$ means to look for the roots of the following equation, where $-i\omega = x$ and L is an integer which depends on the number N of dynamical variables and on how many memory kernels are present (i.e. how many kernels are such that $\alpha_{jk} \neq 0$):

$$x^L + c_{L-1}x^{L-1} + ... + c_0 = 0, \tag{3.36}$$

where (recalling $G_j \neq 0$):

$$c_{L-1} = \left(-\sum_{k\neq j} A_{jk}\frac{G_k}{G_j} + \sum_{h\neq j} A_{hh} + \sum_{\substack{r=1,...N \\ v\neq j}} \beta_{rv}(1 - \delta_{\alpha_{rv}}) \right), \tag{3.37}$$

where the factor $(1 - \delta_{\alpha_{rv}})$ accounts for the fact that the terms which contribute to the last summation of (3.37) are related to memory kernels different from zero (hence

such that $\alpha_{rv} \neq 0$). Using the Vieta's formula or the Routh-Hurwitz method, as in Sec 3.2.1, we are able to say that if the coefficient c_{M-1} is negative, then there is at least one root of equation (3.36) with positive real part. This gives the following form of the UC.

UC and Memory effects: We consider the perturbed problem (3.32). We consider one dynamic variable x_j such that $G_j \neq 0$. If the following inequality is satisfied for such x_j, then that variable is not predictive of the other dynamical variables:

$$SR_j(A) > G_j \left(Tr[A] + \sum_{\substack{r=1,\dots N \\ v \neq j}} \beta_{rv}(1 - \delta_{\alpha_{rv}}) \right). \tag{3.38}$$

As in (3.17), the LHS can be seen as the linear feedback seen by x_j whilst the RHS can be considered as the total dissipation occurring within the system. We observe that considering memory effects the UC is more difficult to satisfy with respect to the case without memory effects (3.17) since the RHS is greater. This is reasonable: due to the addition of memory effects, each variable knows more of the other variables. As a consequence, it is more difficult that the perturbed variables are not able to predict the other ones.
Following the same steps, it is possible to derive a different form of the UC for more complicated memory kernels with Fourier transform like the following (where the roots have a negative imaginary part, in order to have a causal kernel):

$$M_{ij}(\omega) = \frac{\mathrm{Pol}_n(\omega)}{\mathrm{Pol}_d(\omega)}, \qquad n < d, \tag{3.39}$$

where Pol_m stands for a polynomial of order m.

3.4 UC and random matrices

3.4.1 Stability

We want now to apply the UC to random matrices. The introduction of stochasticity in the entries of the matrix A can model a general context where the knowledge of the linearized system is only partial or subjugated to noise. If we want to construct a meaningful RT in this context, the first thing we have to check is that of the asymptotic stability of the system. If the system is asymptotically stable (i.e. if the eigenvalues of the random matrix all have a negative real part), the response functions are predictive, as noted in Sec. 3.1. In order to understand when that happens, let's start from the following fundamental result for the stability of random matrices, called Circular law [79].

Circular law: If A is a $N \times N$ matrix with entries independent and identically distributed (iid), which follow a distribution p with mean 0 and standard deviation (std)

1, then the eigenvalues of $A/\sqrt{N}$ are uniformly distributed over the unit disk in the complex plane in the limit of $N \to \infty$.

An immediate consequence of the Circular law is about the minumum value of the real part of the eigenvalues. If we call $\lambda_{\min}$ the eigenvalue with the smallest real part, then we have that:

$$\mathrm{Re}[\lambda_{\min}] \sim -\sqrt{N} \tag{3.40}$$

We can generalize that result with a generic value for the std σ, obtaining [80]:

$$\mathrm{Re}[\lambda_{\min}] \sim -\sigma\sqrt{N} \tag{3.41}$$

Moreover, we can introduce the connectance $C \in [0, 1]$ of the matrix A: for each entry A_{ij} we have that with probability C the entry A_{ij} is distributed as p, while with probability $(1 - C)$ it is vanishing [80]. As a consequence, we can make the following substitution in (3.41):

$$N \to CN, \tag{3.42}$$

obtaining:

$$\mathrm{Re}[\lambda_{\min}] \sim -\sigma\sqrt{CN} < 0. \tag{3.43}$$

The quantity CN can also be called complexity of the system. As it can be noticed in (3.43), we have that the linearized system (3.3) with random matrix A is not stable on average, since it is probable that some eigenvalues have real part negative. In order to stabilize the system, we can introduce a parameter $d > 0$, also called self-competition or selg-interaction parameter, such that the entries A_{ii} are iid (and independent also from the off-diagonal entries) with a translated distribution such that [80, 81]:

$$\mathbb{E}A_{ii} = d. \tag{3.44}$$

If we do that, the average value $\langle\lambda\rangle$ of the eigenvalues (i.e. the center of the disk) is translated by d towards positive values:

$$\begin{aligned} \langle\lambda\rangle =& \frac{\sum_{i=1}^{N} \lambda_i}{N} \\ =& \frac{\mathrm{Tr}[M]}{N} \\ =& \frac{\sum_{i=1}^{N} A_{ii}}{N} \\ =& d. \end{aligned} \tag{3.45}$$

Not only the center of the disk is translated, but also all the disk is rigidly translated of d towards positive values [82]. We can notice that looking at the secular equation of the matrix A with diagonal (3.44):

$$\begin{aligned} \det(\lambda\mathbb{1} - (A + d\mathbb{1})) =& \det((\lambda - d)\mathbb{1} - A) \\ =& \det(\lambda_A\mathbb{1} - A), \end{aligned} \tag{3.46}$$

hence:

$$\lambda = \lambda_A + d, \tag{3.47}$$

where λ_A are the eigenvalues of the matrix A without the translated diagonal (3.44) and λ with that diagonal. At this point the condition for the stability $\mathrm{Re}[\lambda_{\min}] > 0$ of the translated matrix A is the following [81]:

$$\sigma\sqrt{CN} < d, \tag{3.48}$$

hence the higher is the self-competition d parameter and the more stable is the system.

We can further generalize these stability results considering a distribution for the off-diagonal entries of M with non-vanishing mean μ, as done in [81]. As a consequence, the expectation value of these entries is the following:

$$\mathbb{E}[A_{ij}] = C\mu, \tag{3.49}$$

where C is the connectance. The introduction of a non-vanishing mean μ has consequences also on the stability criterion (3.48). To see this, let's consider the vector $v = (1, ..., 1)$ and then we look at its eigenproblem with the matrix A with eigenvalue λ:

$$Av = \lambda_1 v. \tag{3.50}$$

The equation (3.50) can have no solution. But if we consider that equation on average, we have that the eigenvalue λ_1 is:

$$\begin{aligned} \lambda_1 =& \mathbb{E}\left[\sum_{j=1}^{N} A_{ij}\right] \\ =& d + (N-1)C\mu, \end{aligned} \tag{3.51}$$

where we have used that the off-diagonal elements A_{ij} (along the same row) are iid and $\mathbb{E}A_{ii} = d$. Notice that $\lambda_1 \sim N$, while the radius of the eigenvalues disk goes like $\mu\sqrt{N}$. Moreover, we observe that the disk has translated, and its center is [82]:

$$\begin{aligned} \langle\lambda\rangle =& \frac{\sum_{i=1}^{N}\lambda_i}{N} \\ =& \frac{\sum_{i=1}^{N} A_{ii} - \lambda_1}{N} \\ =& \frac{dN - (d + (N-1)C\mu)}{N} \\ \approx& d - C\mu, \end{aligned} \tag{3.52}$$

where in the last step we have considered that we are in the limit of big N. The last element we need before we can write the new stability criterion for non-vanishing μ is the exact form of the radius of the eigenvalues disk (apart for λ_1). Thanks to (3.41), we have that the radius goes like $\sqrt{(N-1)v}$, where v is the variance of the entries A_{ij}:

$$\begin{aligned} v =& \mathbb{E}[A_{ij}^2] - (\mathbb{E}[A_{ij}])^2 \\ =& C(\sigma^2 + \mu^2) - C^2\mu^2 \\ =& C(\sigma^2 + (1-C)\mu^2). \end{aligned} \tag{3.53}$$

Then we have that the radius R of the disk goes like:

$$R = \sqrt{CN(\sigma^2 + (1-C)\mu^2)}. \tag{3.54}$$

As a consequence, depending on the sign of μ and on how big is N, the stability criterion of the random matrix A for the system (3.3) becomes [81]:

$$\max\{\sqrt{CN(\sigma^2+(1-C)\mu^2)}+C\mu, -(N-1)C\mu\} < d, \tag{3.55}$$

where the first term is related to the minimum eigenvalue within the disk, whilst the second term is related to λ_1. We can consider (3.55) as the stability criterion for the RT.

3.4.2 The average UC

In the section before we have found a well-defined random matrices context where there is an effective criterion (3.55) to see if a matrix is obtained linearizing around an asymptotically stable fixed point (in the limit of big N). Now we want to see which form the UC assumes in this context. We add another layer of stochasticity assuming that the spatial pattern of the perturbation G is distributed with a PDF $g(G)$ with mean μ_G and std σ_G. As a consequence, the vector with entries G_i is composed of N realizations of the random variable G. We consider the UC in the following form for the dynamical variable x_j:

$$\sum_{k=1}^{N} A_{jk}G_k > G_j \sum_{h=1}^{N} A_{hh}. \tag{3.56}$$

Now we take the average of (3.56), recalling that along the diagonal the entries A_{ii} have average value d (and that N is very large):

$$\mathbb{E}[A_{kj}G_k] > d\, G_j. \tag{3.57}$$

We assume that there is no correlation between the values of G and the values of A_{ij}, i.e. we assign these values without correlating them. As a consequence, we have that:

$$\mathbb{E}[A_{kj}G_k] = \mathbb{E}[A_{kj}]\mathbb{E}[G_k], \tag{3.58}$$

and the UC becomes:

$$C\mu\, \mu_G > d\, G_j. \tag{3.59}$$

Note that (3.59) is valid for any variable x_j. If we average over them, we obtain:

$$C\mu\, \mu_G > d\, \mu_G. \tag{3.60}$$

If we assume $\mu_G \neq 0$ we obtain the following relation, which we call **average UC**:

$$C\mu > d, \tag{3.61}$$

which tells us that on average if the UC is satisfied. In particular, if the average value μ of the off-diagonal entries A_{ij} is too much positive and it overcomes the self-interaction term d, the relation (3.61) is satisfied. As a consequence, on average the perturbed variables cannot predict other perturbed variables in that case. The meaning of that fact is the following. If in the system there much more negative feedbacks (links between i and j such that $(-A_{ij}) < 0$ [3], also called mutualistic interactions) than

[3] Recall that in the equation of the dynamics (3.3) there is a minus in front of the matrix A.

positive feedbacks (such that $(-A_{ij}) > 0$, also called competitive interactions), the perturbed dynamical variables cannot predict each other on average. In other words, the full state $\{x_j\}_{j=1,\dots,N}$ cannot be fully reconstructed from knowing just a few of the dynamical variables. We can observe again what we had observed for the UC, but on average: if the average linear feedback μ from the system to the dynamical variable j is greater than the average self-interaction d within the system, the perturbed dynamical variables cannot predict other perturbed dynamical variables on average. If we want to have the possibility to have the prediction possible on average, it is necessary to change the distribution of the entries in such a way to increase the positive feedbacks, to obtain $C\mu_G < d$.

If we have $\mu_G = 0$ the line of thinking drastically change. The relation (3.59) becomes:

$$0 > d\, G_j, \tag{3.62}$$

which does not depend on the average value of the entries μ, but just on the distribution of G. We have that with probability $p_- = \int_{-\infty}^{0} g(G_j) dG_j$ a generic dynamical variable x_j cannot make predictions of the other variables. As a consequence we observe that if we perturb the system with a stochastic perturbation which contains both positive and negative contributes (in such a way that on average they cancel each other), the UC does not depend on how much positive or negative feedbacks are already present in the system.

3.4.3 Adding correlations

We may wonder what happens to the results above if we introduce correlations within the system defined by the matrix A. In particular, given two dynamical variables x_i and x_j we say that there is a correlation between A_{ij} and A_{ji}. For example, we can impose that A_{ij} and A_{ji} have the same sign, or that they have opposite sign [81]. In that case, the circular law changes as follows [83].

Elliptic Law: If A is a $N \times N$ real matrix such that:

- The pairs of entries $(A_{ij}, A_{ji})_{i \neq j}$ are iid such that $\mathbb{E}A_{ij} = \mathbb{E}A_{ji} = 0$ and $\mathbb{E}A_{ij}^2 = \mathbb{E}A_{ji}^2 = 1$,
- $\max(\mathbb{E}|A_{ij}|^4, \mathbb{E}|A_{ji}|^4) \leq M_4$ for a $M_4 \geq 0$,
- $\mathbb{E}(A_{ij}, A_{ji}) = \tilde{\rho}$, with $\tilde{\rho} \leq 1$,
- The diagonal entries A_{ii} are iid, and independent from the off-diagonal entries, such that $\mathbb{E}A_{ii} = 0$ and $\mathbb{E}A_{ii}^2 < \infty$,

then for $N \to \infty$ the eigenvalues of $A/\sqrt{N}$ converge to an uniform PDF over an ellipse in the complex plane with center $(0, 0)$, horizontal axis of length $(1 + \tilde{\rho})$ and vertical axis of length $(1 - \tilde{\rho})$.

Note that the elliptic law reduces to the circular law for $\tilde{\rho} = 0$. As we have done for the circular law, we can generalize this result [81], considering a non-vanishing mean μ and a non-unitary std σ for the off-diagonal entries A_{ij}, the connectance C of the

matrix and shifting the average value of the diagonal entries A_{ii} of d. Given that, the correlation parameter $\tilde{\rho}$ has to be multiplied for the square std:

$$\tilde{\rho} \to \sigma^2 \tilde{\rho}. \tag{3.63}$$

Moreover, the parameter which enters in the axes of the ellipse is the following ρ:

$$\begin{aligned} \rho =& \frac{\mathbb{E}(A_{ij}A_{ji}) - \mathbb{E}^2 A_{ij}}{v} \\ =& \frac{(\tilde{\rho}\sigma^2 + \mu^2) - C\mu^2}{\sigma^2 + (1-C)\mu^2}, \end{aligned} \tag{3.64}$$

where we have used the variance v of the off-diagonal terms A_{ij} defined in (3.53). Due to these generalizations, we have that the ellipse is shifted rigidly of $(d - C\mu)$ towards positive values and there is a eigenvalue λ_1 whose value goes like N, as in the circular case. The stability criterion (3.55) for the system (3.3) becomes [81]:

$$\max\{\sqrt{CN(\sigma^2 + (1-C)\mu^2)}(1+\rho) + C\mu, -(N-1)C\mu\} < d. \tag{3.65}$$

Remarkably, since the correlations introduced are between the entries A_{ij} and A_{ji}, the average UC (3.60) does not change. This is because its formulation is focused on the entries of the matrix A along the same row, and there are no couples of entries $(A_{ij}, A_{ji})_{i \neq j}$ along the same row. We can deduce that the average UC applies with the same form for both systems without and with correlations between $(A_{ij}, A_{ji})_{i \neq j}$.

Chapter 4

Surrogate RT and Lorenz 96

In this chapter we will apply the surrogate RT 2.6 on a climate model: the Lorenz 96 (L96) model [41–43]. In **Section 4.1** we will present this model and we will analyze its instabilities and the presence of travelling waves within it. The L96 has already been subjected to RT investigations in [6], showing successfully the applicability of the Ruelle's RT. The first application of the surrogate RT on the L96 model has been proposed in the seminal paper [28], where the surrogate RT has been introduced for the first time. In that paper, both the perturbation and the observables of interest were global quantities. In this chapter instead, we want to perturb locally the system and we want to predict the behaviour of local quantities using the surrogate RT. In **Section 4.2** we will show how to numerically retrieve the response functions and the surrogate response functions. We will discuss how their asymptotic behaviour for small times after the application of the local perturbation permits to classify the dynamical variables in a hierarchy in terms of their predictive ability. Moreover, we will introduce the concepts of anticipated and retarded surrogate response functions, discarding the anticipated ones as non-physical. Then, we will quantify the predictive power of the surrogate response functions using the ratio method derived in Section 2.6, which shows the presence of a definite verse of the flow of information. To show the effectiveness of the RT and the surrogate RT in this context, we will perform predictions of the actual response, pointing out the successful and unsuccessful cases. Lastly, we will improve the unsuccessful predictions provided by local observables, using the information provided by an additional forcing.

4.1 The L96 model

The L96 model [41–43] is a paradigmatic climate model which can be defined as follows. Let's consider a circle at a constant latitude on the globe and let's divide it into N sectors $\{i = 1, ..., N\}$. Then we consider a generic atmospheric variable x (e.g. the pressure or the vorticity) and we assign a value of x to each of the N sectors: $\{x_i\}_{i=1,...,N}$. Since we are on a latitudinal circle, the index i is cyclic: $x_{i-N} = x_i = x_{i+N}$. The L96 model is the following dynamical system model, which describes the evolution in time of these N coupled dynamical variables:

$$\dot{x}_i = x_{i-1}(x_{i+1} - x_{i-2}) - \gamma x_i + F, \tag{4.1}$$

where $\{i = 1, ..., N\}$, F is a generic forcing term and γ is the viscosity coefficient which we will take unitary $\gamma = 1$. Note that the statistical properties of each of the variables are the same since the system is invariant under translations: $i \to i + 1$. Even though it is a very simple model, all the basic physical processes are represented: the first quadratic term in (4.1) stands for a non-trivial advection process of the quantity x, the second linear term simulates the mechanical damping while the third constant term is an external forcing. The quadratic term which stands for the advection process is such that, in the unviscid and unforced regime ($F = \gamma = 0$), it allows for the conservation of the energy of the system, expressed as the sum of the squares of the variables:

$$E = \sum_{i=1}^{N} x_i^2. \tag{4.2}$$

To see this it is sufficient to derive E with respect time t and then plug (3.14) inside the derivative (in the unviscid and unforced regime)

$$\begin{aligned} \frac{dE}{dt} =& \sum_{i=1}^{N} x_i \dot{x}_i \\ =& \sum_{i=1}^{N} x_i x_{i-1} x_{i+1} - \sum_{i=1}^{N} x_i x_{i-1} x_{i+2} \\ =& 0, \end{aligned} \tag{4.3}$$

in the last step we have used the fact that periodic boundary conditions are imposed in the system (3.14). In the viscid and forced regime, on the other hand, we have:

$$\frac{dE}{dt} = -\sum_{i=1}^{N} x_i x_i + F \sum_{i=1}^{N} x_i, \tag{4.4}$$

hence the energy varies in time. The L96 dynamical system shows different behaviours accordingly to the values of F and N, which modify the geometry of the attractor. It can be shown that, when the dissipation parameter is unitary as in (4.1), the dynamics of the model is chaotic for $F \geq 6$ and $N \geq 20$, since some Lyapunov exponents turn positive [84]. This is intuitive: F forces energy in into the system so we expect that with its increase turbulence should appear in the system. We can see the emergence of the chaotic behaviour in the contour plots (a), (c) and (e) in Fig. 4.1, realized for $N = 10$, where the top and middle panels show the L96 system in the non-chaotic regime for respectively $F = 0.1$ and $F = 1$, while the bottom panels are in the chaotic regime since they are realized for $F = 8$. It is noticeable that in the top and middle left panel we can observe some waves travelling from right to left. These waves are roughly still present in the chaotic regime in the bottom left panel.

4.1.1 L96 instability analysis

As observed before, in Fig. 4.1 we can observe the presence of travelling waves from right to left in the L96 systema for a certain range of parameters. We now study their emergence due the modulation of the forcing parameter F, fixing the number of variables to $N > 4$ [41]. We notice that a fixed point for the L96 system (4.1) is given

by $x_j = F$ for each $j \in \{1, ..., N\}$. We want to understand when this steady state is stable or unstable. To do that we consider the following solution type, formed by a mono-chromatic wave perturbing the steady state $x_i = F$:

$$x_j = F + y_j, \qquad y_j = a\, e^{\sigma t} e^{i(kj - \omega t)}, \tag{4.5}$$

with momentum $k = 2\pi/N$, frequency ω, constant $a \ll F$ and $\sigma \in \mathbb{R}$. If $\sigma > 0$ the fixed point $x_j = F$ is unstable, while it is stable for $\sigma < 0$. If we substitute (4.5) into (4.1) we obtain the following relation:

$$(\sigma - i\omega) = F\left(e^{ik} - e^{-2ik}\right) - 1. \tag{4.6}$$

Considering separately the real and the imaginary part we obtain:

$$\begin{aligned} \sigma &= F(\cos(k) - \cos(2k)) - 1 \\ \omega &= -F(\sin(k) + \sin(2k)). \end{aligned} \tag{4.7}$$

We want to derive the minimum value of the forcing term F_c such that $\sigma > 0$ for $F \geq F_c$ (and hence the fixed point $x_j = F$ is unstable). We observe that the maximum value of the difference $(\cos(k) - \cos(2k))$ in (4.7) is $9/8$ and it is realized for $\overline{k} = \arccos(1/4) = \arg_k \max(\cos(k) - \cos(2k))$ [1]. As a consequence, we have that σ becomes positive for values of F bigger or equal than $F = F_c = 8/9$, making unstable the fixed point $x_j = F$. We have thus that in the stable steady state a travelling wave appears, with frequency given by ω in (4.7). In particular, for $k = \overline{k}$ and $F = F_c$ we have that $\omega_c = 1.291$ and the corresponding period is $T_c = |2\pi/\omega_c| \approx 4.8668$. Note that here we have disregarded the discrete nature of k. It is remarkable that the crest of these waves goes from right to left, since the phase velocity is:

$$v_p = \frac{\omega_c}{\overline{k}} \approx -0.98 < 0. \tag{4.8}$$

This fact confirms the observations made above regarding Fig. 4.1. Interestingly, it has to be noted that the group velocity is given by:

$$\begin{aligned} v_g =& \frac{\partial \omega_c}{\partial k}_{|k=k_c} \\ =& -F_c(\cos(\overline{k}) + 2\cos(2\overline{k})) \\ \approx& 1.333 > 0, \end{aligned} \tag{4.9}$$

hence the wave packets propagate from left to right. This is remarkable, since the group velocity marks the direction of the propagation of information within the system. There is a distinct verse of the flow of information. We show the emergence of the travelling waves with the increase of F in Fig. 4.1. In the top panels there are a contour plot for a realization of the L96 system (4.1) for a random initial condition and $N = 10$ and $F = 0.1$, and the time behaviour of two dynamic variables. We can see that there are no travelling waves and the dynamical variables converge to the fixed point $x_j = F = 0.1$. In the middle panels there are a contour plot $N = 10$ and $F = 1$ and

[1] Since the cosine is even in k, we could have chosen $-\overline{k}$. The main results afterwards would not have changed with that choice.

the time behaviour of the same two dynamical variables considered in the top panels. In this case we observe the presence of the travelling waves in the contour plot and in the fact that the dynamical variables evolve periodically in time. Lastly, in the bottom panels, realized for $N = 10$ and $F = 8$, we are in the chaotic regime, as it can be observed looking at the time behaviour of the dynamical variables in the right panel. In the left panel we can see that the travelling waves are roughly preserved.

4.2 Numerical tests

4.2.1 Linear response functions

We will now apply the formalism of the surrogate RT, presented in Section 2.6, to the L96 dynamical system (4.1). We will pose ourselves in the chaotic regime, choosing $N = 36$ for the number of dynamical variables x_j and $F = 8$ for the forcing term. In Fig. 4.2 we can explicitly see that we are in the chaotic regime. In the paper [28], where the surrogate RT has been introduced for the first time, the author focused on global perturbations, affecting all the x_i. We want now to take a different route, focusing on local perturbations. In particular, we will choose as a spatial pattern of the forcing the following:

$$G_i(x) = \delta_{ik} G, \tag{4.10}$$

where δ_{ik} is the Kronecker delta which is different from zero just for $i = k$, and G is a real number which measures the magnitude of the perturbation. In other words, the perturbation (4.10) perturbs the forcing term in the equations of L96. We will consider as observables of interest local quantities too: the dynamical variables x_j. We remark that this approach is feasible since the L96 system is an extended system in space. In other words, the problem we are addressing amounts to asking to what extent a perturbed variable i can predict the future state of another perturbed variable j after the system has been perturbed locally in a location k.

Numerical estimation

We call $\Gamma_{i,k}$ the response function of the perturbed variable x_i to the perturbation with spatial pattern (4.10), located in x_k. In order to compute these response function we proceed as follows [28, 85]. First, we run a long simulation with a random initial condition. We discard an initial transient and then we create an ensemble of M states, taking a state each T_{gap} time units. We use this ensemble of L96 states on the attractor in order to compute the response functions. We consider now a perturbation with spatial pattern (4.10) and instantaneous time pattern $e(t) = \delta(t)$. Looking at the response relation (2.14) we can see that doing so the response function $\Gamma_{i,\text{gl}}$ is just proportional to the perturbation of the dynamical variable $x_i(t)$:

$$\delta\langle x_i\rangle(t) = \Gamma_{i,k}(t). \tag{4.11}$$

As a consequence, we can compute the response function $\Gamma_{i,k}$ as an average over the ensemble of M initial conditions within the steady state regime built before:

$$\Gamma_{i,k}(t) = \frac{1}{M} \sum_{k=1}^{M} \delta x_i^{(k)}(t), \tag{4.12}$$

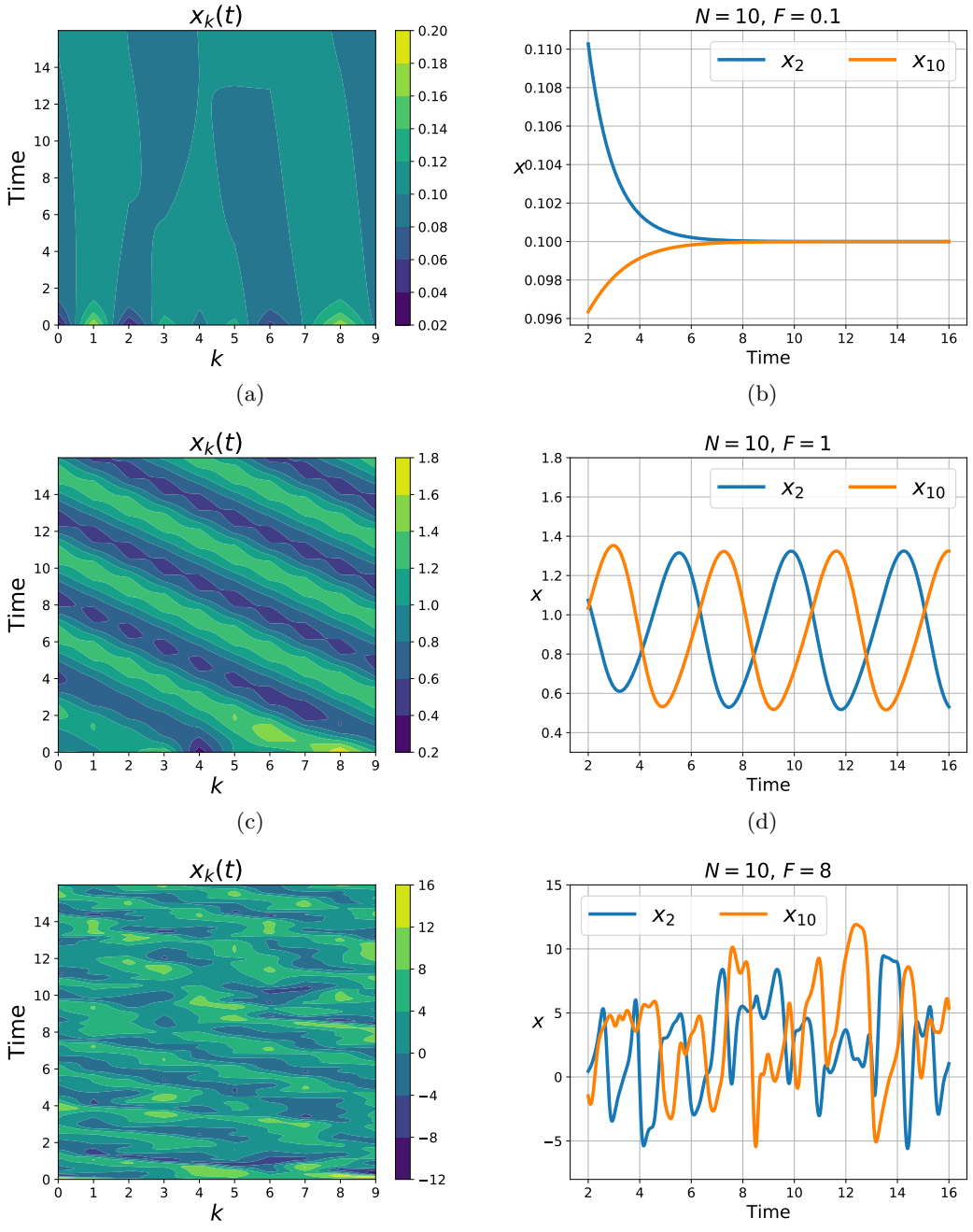
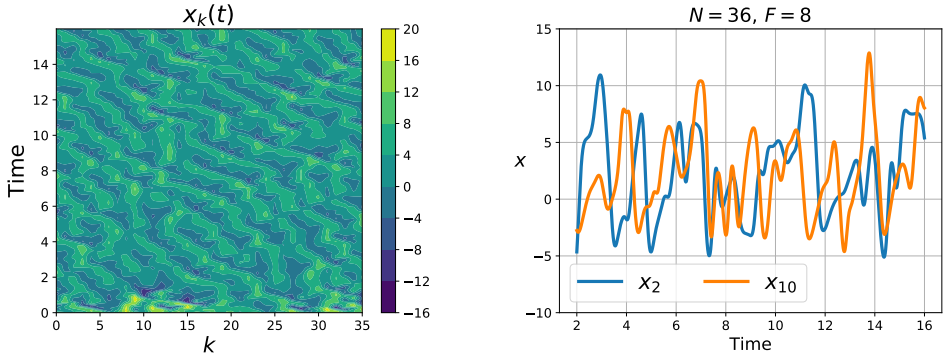


Figure 4.1: (a) and (b): the left panel shows a contour plot for a realization of the L96 system (4.1) for a random initial condition, for $N = 10$ and $F = 0.1$. On the horizontal axis there is the spatial index $\{i = 1, ..., N\}$, while on the vertical axis there is the time t. The colour of a point (i, t) shows the magnitude of the value of x_i at time t. The right panel shows the behaviour in time of two variables x_2 and x_{10}. They both converge to the fixed point $x_j = F = 0.1$. (c) and (d): the same figures, for $N = 10$ and $F = 1$. The periodic travelling wave can be recognized visually in the left panel and from the periodic behaviour of the dynamical variables showed in the right panel.(e) and (f): the same figures, for $N = 10$ and $F = 8$. Even though the travelling waves are roughly preserved in the left panel, we are in the chaotic regime as it can be seen in the time behaviour of the dynamical variables in the right panel.



(a) (b)

Figure 4.2: (a): Contour plots of a realization of the L96 system (4.1) for a random initial condition for $N = 36$ and $F = 8$. On the horizontal axis there is the spatial index $\{i = 1, ..., N\}$, while on the vertical axis there is the time t. The colour of a point (i, t) shows the magnitude of the value of x_i at time t. (b): Behaviour in time of two variables x_2 and x_{10}. We can see that the system is in the chaotic regime, but still the travelling waves from right to left can be seen in the right panel.

where $\delta x_i^k(t)$ is computed as follows. We pick an initial condition, i.e. an ensemble member k. We integrate the perturbed system and the unperturbed system for a certain number T of time steps, starting from the picked initial condition. We remark that, due to the chosen instantaneous time pattern, the perturbation enters in the evolution of the perturbed system with just a modification of the initial condition x_j^k picked in the ensemble:

$$x_j^k(0) \to x_j^k(0) + G. \tag{4.13}$$

Then, we compute $\delta x_i^k(t)$ taking the difference between the value of x_i at time t for the perturbed trajectory and the unperturbed trajectory. We do that for any t between 0 and T. Moreover, in order to compute the global response function in a more precise way, removing second-order corrections, we adopt the following procedure [6]. Given the initial condition picked in the ensemble, we consider the trajectory integrated using the dynamical system perturbed with a time pattern $e(t)$ and the trajectory integrated perturbing the system with a time pattern with amplitude $-e(t)$. We compute the ensemble average (4.12) taking the semi-difference between the value of x_i at time t between these two trajectories.

The first thing we have to confirm in our numerical study is that of we are in the linear regime. To check that, we have used the procedure described above to numerically compute the response functions $\Gamma_{i,k}$ for $i = \{k-2, k-1, k, k+1, k+2\}$, for different values of the forcing. The smaller is the value of the forcing G and the better is the description of the system by the linear RT. On the other hand, a smaller G implies that the perturbation is less felt by the system and the signal-to-noise ratio decreases. As a consequence, the smaller is G and the more ensemble members M we need to have clear response functions. Given that, we choose G as the biggest value of the forcing such that we are in the linear regime. Note that we have to check to be in the linear regime not only for the response function of the directly perturbed site $\Gamma_{k,k}$, but also for the response functions of the other sites. This is due to the fact that the perturbation is

less felt in those sites: for the same value of the forcing G it could happen a priori that the linear regime succeeds in describing the system in x_k but fail in these sites. In that case, it is necessary to resort to a smaller value of G. We varied the forcing parameter G in the range $G = \{0.5, 0.75, 1, 2, 2.5, 3, 6, 12\}$, obtaining the results showed in Fig. 4.3, where we plotted $\Gamma_{i,k}/G$. We can observe in that figure that in all the locations $i = \{k-2, k-1, k, k+1, k+2\}$ the curves overlap for G smaller or equal than 1. For higher values of the forcing G, the other non-linear orders of the RT start to matter more. As a consequence, we continuously deviate from the linear regime increasing G. Then, we choose $G = 1$ as the value of the forcing that we will use in our investigations thereafter. The response functions $\Gamma_{i,k}$ for $i = \{k-2, k-1, k, k+1, k+2\}$ and $G = 1$ are showed in Fig. 4.4, which are obtained for $M = 2 \cdot 10^6$ ensemble members. We have used such a large ensemble because, as we will see later, in order to retrieve the surrogate response functions H it is necessary to obtain the Fourier transform of the response functions. As a consequence, we need a large signal-to-noise ratio at each individual frequency ω, thus we need a large ensemble. If we stay in the time domain, this problem does not come up, since we integrate the response function in a convolution and hence the specific value of the response function at a time t is not relevant per se [40]. We remark that we have chosen to focus on just a few dynamical variables around the perturbation site. This is due to the fact that the perturbation is felt way less as more we move from the directly perturbed variable, and as a consequence the quality of the response functions decreases.

Asymptotic behaviour

We remark that it is possible to obtain the behaviour of the response functions $\Gamma_{i,k}$ for $t \to 0^+$, hence just after the system is perturbed, as follows. Let's consider the response function (2.16) in our case:

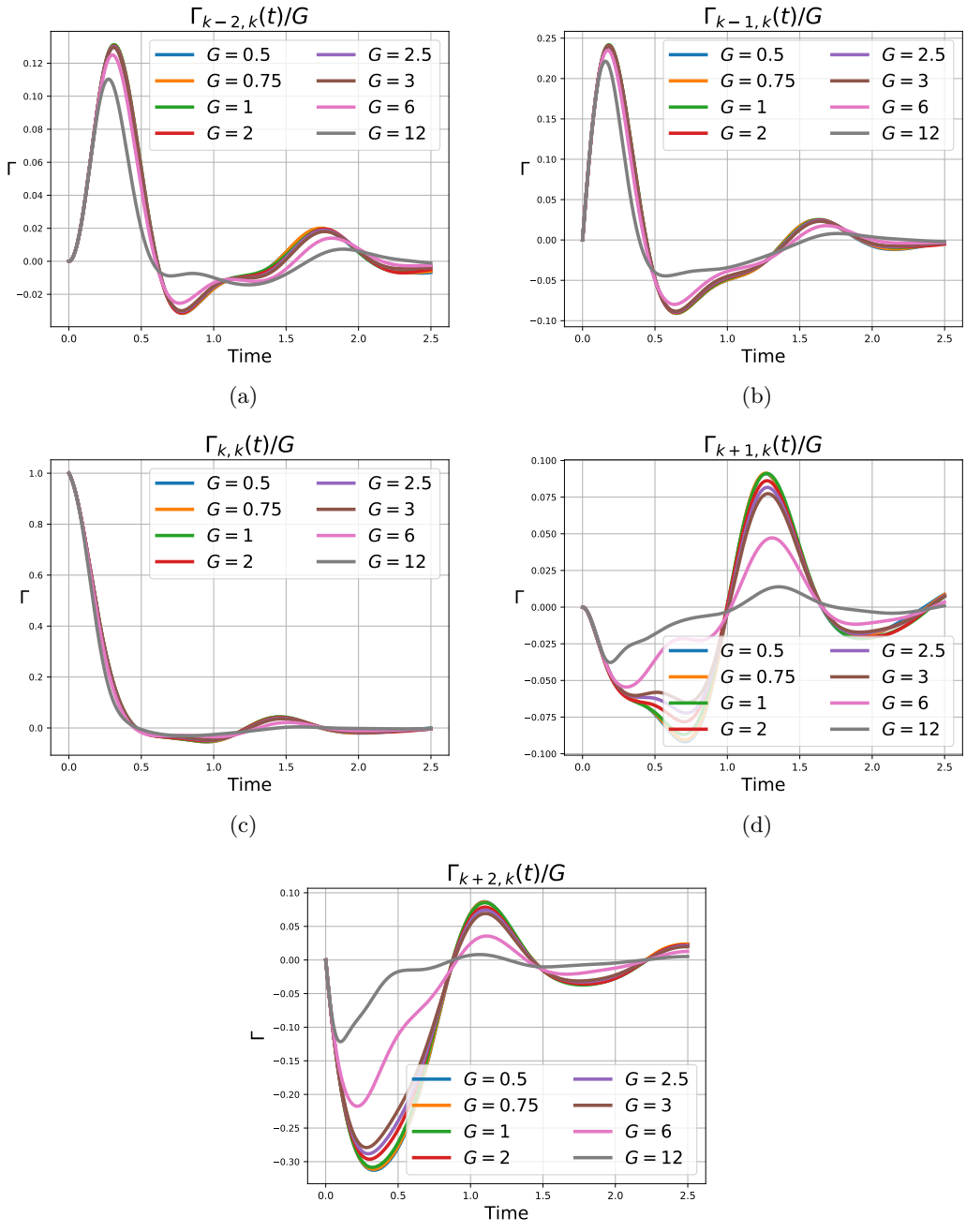
$$\begin{aligned} \Gamma_{i,k} =& \Theta(t) \int \rho(dx) \sum_l G\, \delta_{kl} \partial_l (x_i(t)) \\ =& \Theta(t) \int \rho(dx)\, G\, \partial_k (x_i(t)), \end{aligned} \tag{4.14}$$

where ∂_k stands for the derivation with respect to $x_{k(0)}$ and ρ is the steady-state distribution over the initial condition $x(0)$ from which the trajectory $x(t)$ starts. We now expand $x_i(t)$ around $t = 0^+$ and we exploit the equation of dynamics (4.1):

$$\partial_k (x_i(t)) \approx \delta_{i,k} + t(C^{(1)}_{k-1}\delta_{i,k-1} + C^{(1)}_{k+1}\delta_{i,k+1} + C^{(1)}_{k+2}\delta_{i,k+2} - \delta_{ik}) + ..., \tag{4.15}$$

where $C^{(a)}_p$ is the coefficient related to the dynamical variable p of the term of order a in the expansion above (we have that $C^{(0)k} = \delta_{ik}$). Note that the leading term in the expansion (4.15) depends on the distance between the considered dynamical variable x_i and the location of the perturbation (x_k): the leading term is t^q if x_i is distant q from x_k from the left or $2q-1$ or $2q$ from the right. If we compute the coefficients $C^{(a)}_p$, further details emerge. The coefficients for the linear terms ($k = 1$) give:

$$\begin{aligned} C^{(1)}_{k-1} =& x_{k-2}(0) \\ C^{(1)}_{k+1} =& (x_{k+2}(0) - x_{k-1}(0)) \\ C^{(1)}_{k+2} =& -x_{k+1}(0). \end{aligned} \tag{4.16}$$



(a) (b) (c) (d) (e)

Figure 4.3: Plots of the response functions $\Gamma_{i,k}/G$ for $i = \{k-2, k-1, k, k+1, k+2\}$, showed respectively in (a), (b), (c), (d) and (e). We used different values of the forcing, ranging in $G = \{0.5, 0.75, 1, 2, 2.5, 3, 6, 12\}$. Note that for G smaller or equal than 1 we are in the linear regime. For higher values of the forcing, the response functions start to change in a continuous way, deviating from the linear regime.

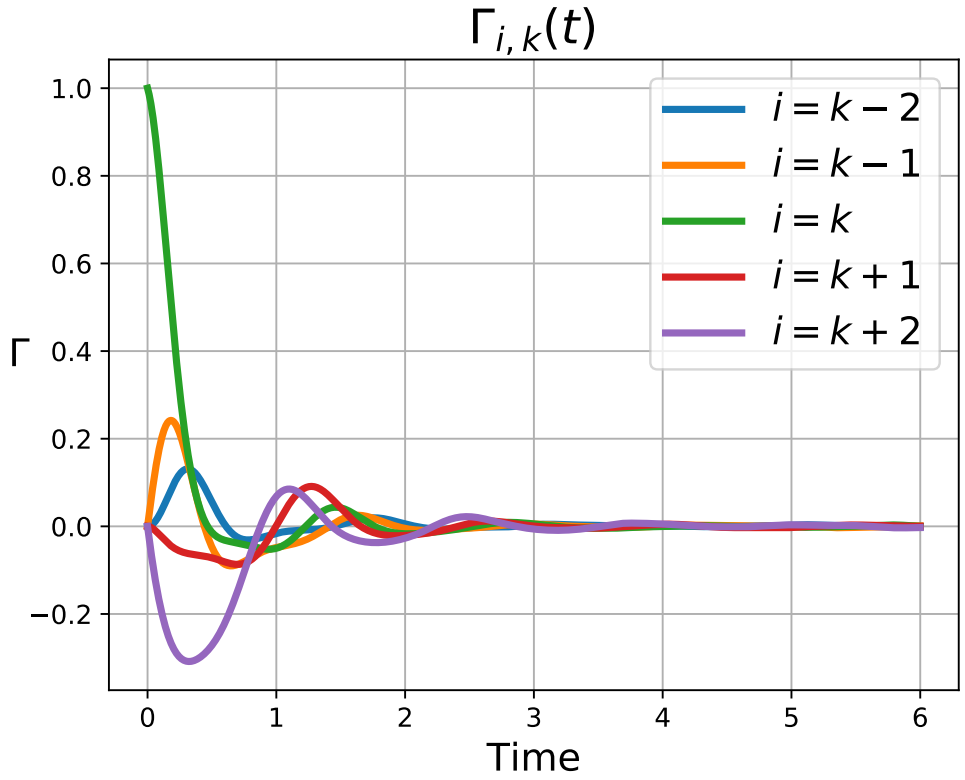


Figure 4.4: Plots of the response functions $\Gamma_{i,k}$ for $i = \{k-2, k-1, k, k+1, k+2\}$ and $G = 1$.

As a consequence, if we consider the expansions of the response function (4.14) which are linear in the limit $t \to 0^+$ we obtain the following relations:

$$\begin{aligned}
\Gamma_{k-1,k} =& \Theta(t) \left(G \int \rho(dx) x_{k-2}(0) \right) t + ... \\
\Gamma_{k+1,k} =& \Theta(t) \left(G \int \rho(dx) (x_{k+2}(0) - x_{k-1}(0)) \right) t + ... \\
\Gamma_{k+2,k} =& \Theta(t) \left(-G \int \rho(dx) x_{k+1}(0) \right) t + ...
\end{aligned} \tag{4.17}$$

We can observe in (4.17) that the coefficient $\langle C^{(1)}_{k+1} \rangle$ (the average is over the measure ρ) of the linear term in the expansion of $\Gamma_{k+1,k}$ is vanishing, since:

$$\int \rho(dx)\, x_{k+2}(0) = \int \rho(dx)\, x_{k-1}(0). \tag{4.18}$$

The relation (4.18) is due to the fact that the statistical properties of the dynamical variables are the same, as noted above. We can notice that $\langle C^{(1)}_{k-1} \rangle = -\langle C^{(1)}_{k+2} \rangle$ for the same reason. As a consequence, the leading term in the expansion for small t of $\Gamma_{k+1,k}$ is a quadratic term t^2. We can repeat the same derivation for all orders k, observing that the leading term of $\Gamma_{k+2q-1,k}$ is t^{q+1} instead of t^q. Given that, we can say that the leading term of a generic $\Gamma_{i,k}$ is:

$$\Gamma_{i,k}(t) \underset{t \to 0^+}{\approx} \begin{cases} \Theta(t) G, & i = k \\ \Theta(t) \left(G \langle C^{(1)}_i \rangle \right) t. & i \in \{k-1, k+2\} \\ \Theta(t) \left(G \langle C^{(q)}_i \rangle \right) t^q, & i \in \{k-q, k+2q-3, k+2q\},\ q \geq 2, \end{cases} \tag{4.19}$$

where the averages are over the stationary measure ρ. In particular, we notice that at $t = 0$ the response function $\Gamma_{i,k}$ is equal to G if $i = k$ and it is vanishing for $i \neq k$. This is intuitive: at $t = 0$ the perturbation is felt in all its intensity in the spot which has been directly perturbed, while it is not felt at all in the other locations. As t increases, the perturbation propagates also to the other spots, with time scale which depends on the leading order (4.19) of the response function $\Gamma_{i,k}$ or, in other words, on the distance with respect to the directly perturbed x_k. Note that the perturbation propagates more at right than left, since for each dynamical variable x_{k-q} at the left of x_k with leading term t^q in the response function (4.19) there are two dynamical variables x_{k+2q-3} and x_{k+2q} (for $q > 1$) at its right with the same leading term in the response function (hence they feel the perturbation for $t \to 0^+$ after the same time scale). We can have a figurative intuition of the actual propagation of the signal in space and in time in Fig. 4.5. It is remarkable that there are some sites like $k+3$ or $k+5$ that, even though they are very close to the perturbation site k, they are not affected by the perturbation at time $t+2$ or $t+3$. This is because the advection taking place in the L96 system is not a pure one, since it is mixed with other physical processes.

As noted in the relations (2.93) and (2.94), the asymptotic behaviour of the response function in the time domain for small t is related to the asymptotic behaviour for big ω of the Fourier transform of the response function:

$$\Gamma_{i,k}(t) \approx \alpha_{i,k} \Theta(t) t^\beta$$

	a	b
$\Gamma_{k-1,k}$	2.033 ± 0.006	0.9940 ± 0.0006
$\Gamma_{k+1,k}$	-2.241 ± 0.007	1.9542 ± 0.0006

Table 4.1: Results of the fit of $\Gamma_{k-1,k}(t)$ and $\Gamma_{k+1,k}(t)$ for small t with the function $f(t) = a\, t^b$, with error given by standard deviation.

$$\Updownarrow$$

$$\Gamma_{i,k}(\omega) \approx (\alpha_{i,k}\, \beta!\, i^{\beta+1}) \frac{1}{\omega^{\beta+1}}$$

As a consequence, we can derive from (4.19) the asymptotic behaviour of the Fourier response of the response functions:

$$\Gamma_{j,k}(\omega) \underset{\omega\to\infty}{\approx} \begin{cases} i\,\dfrac{G}{\omega}, & j = k \\ -\dfrac{G\langle C_j^{(1)}\rangle}{\omega^2} & j \in \{k-1, k+2\} \\ (i^{q+1}q!)\dfrac{G\langle C_j^{(q)}\rangle}{\omega^{q+1}}, & j \in \{k-q, k+2q-3, k+2q\},\ q \geq 2, \end{cases} \tag{4.20}$$

We have tested the relation (4.20) for $\Gamma_{k-1,k}$, $\Gamma_{k,k}$ and $\Gamma_{k+1,k}$, whose leading term in the expansion (4.19) is respectively of order t, t^0 and t^2. We have fitted the response functions $\Gamma_{k-1,k}(t)$ and $\Gamma_{k+1,k}(t)$ for small t with the function $f(t) = a\, t^b$, obtaining the result showed in Table 4.1. We can notice that the exponent b gives the behaviour predicted by (4.20) in both cases. For $\Gamma_{k,k}$ the situation is simpler: in Fig. 4.4 we can see that in $t = 0$ we have $\Gamma_{k,k}(0) = G = 1$, as predicted. We now plug the coefficients in Table 4.1 obtained by the fits in (4.20), and compare the results with the asymptotic behaviour of the response functions for large ω. As we can observe in Fig. 4.6 that the comparison is very good in all the three cases.

Making predictions

Lastly, we test the effectiveness of the response functions we derived, predicting the response of the L96 system (4.1) perturbed as follows:

$$\dot{x}_i = x_{i-1}(x_{i+1} - x_{i-2}) - x_i + F + G\delta_{i,k}e(t), \tag{4.21}$$

where the spatial pattern of the perturbation is given by (4.10), while the time pattern is the following:

$$e(t) = \Theta(t) - \Theta(t-\tau), \tag{4.22}$$

with $\tau = 0.1$. The time perturbation is equal to 1 for $t \in [0, \tau]$, then it goes to zero. We integrate the perturbed system (4.21), taking the initial condition from an ensemble of $M = 2 \cdot 10^6$ members and then averaging the obtained response, as follows:

$$\delta\langle x_i\rangle(t) = \frac{1}{M}\sum_{q=1}^{M} \delta x_i^{(k)}(t), \tag{4.23}$$

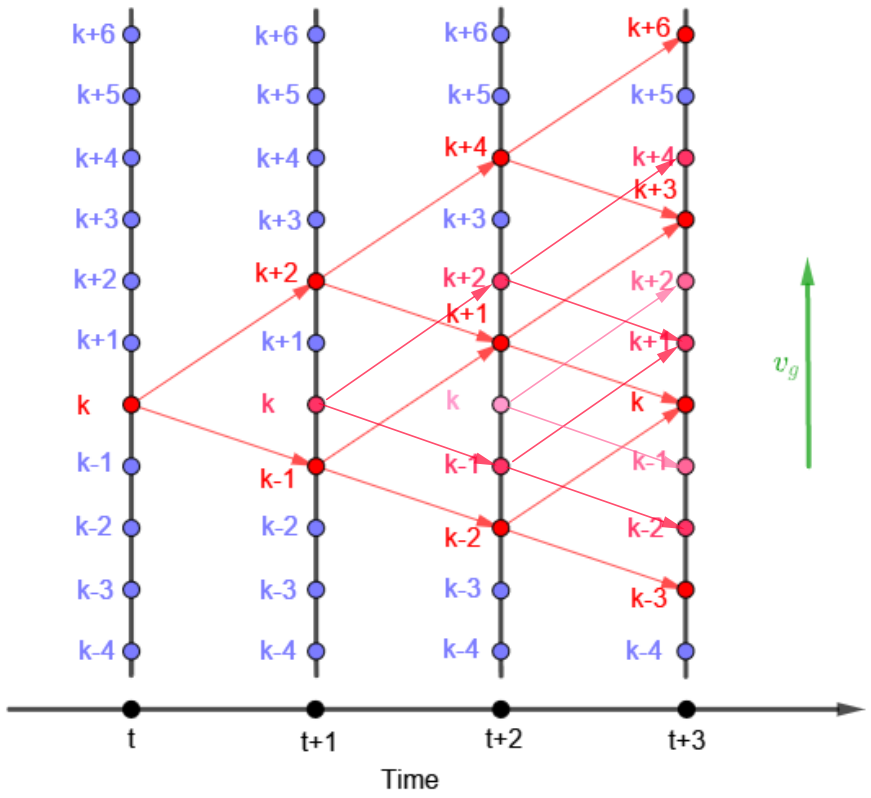


Figure 4.5: Propagation of the perturbation for small times in the L96 model perturbed locally in x_k starting from a time t. The vertical lines are the dynamical variable x_k taken at different time instants, while the horizontal line below is the time axis. We have discretized the time in unit time steps for clarity purposes. At a given time, we have coloured the dynamical variables which feel directly the perturbation from the perturbed variables of the instant before with red. The colour loses its intensity as time goes by. We can notice that there are more red-coloured variables above the site x_k than below. This is consistent with the fact that the information propagates from the dynamical variables x_{k+j} with $j < 0$ to the dynamical variables x_{k+j} with $j > 0$, with a velocity given by the group velocity v_g, shown with a green arrow in the figure.

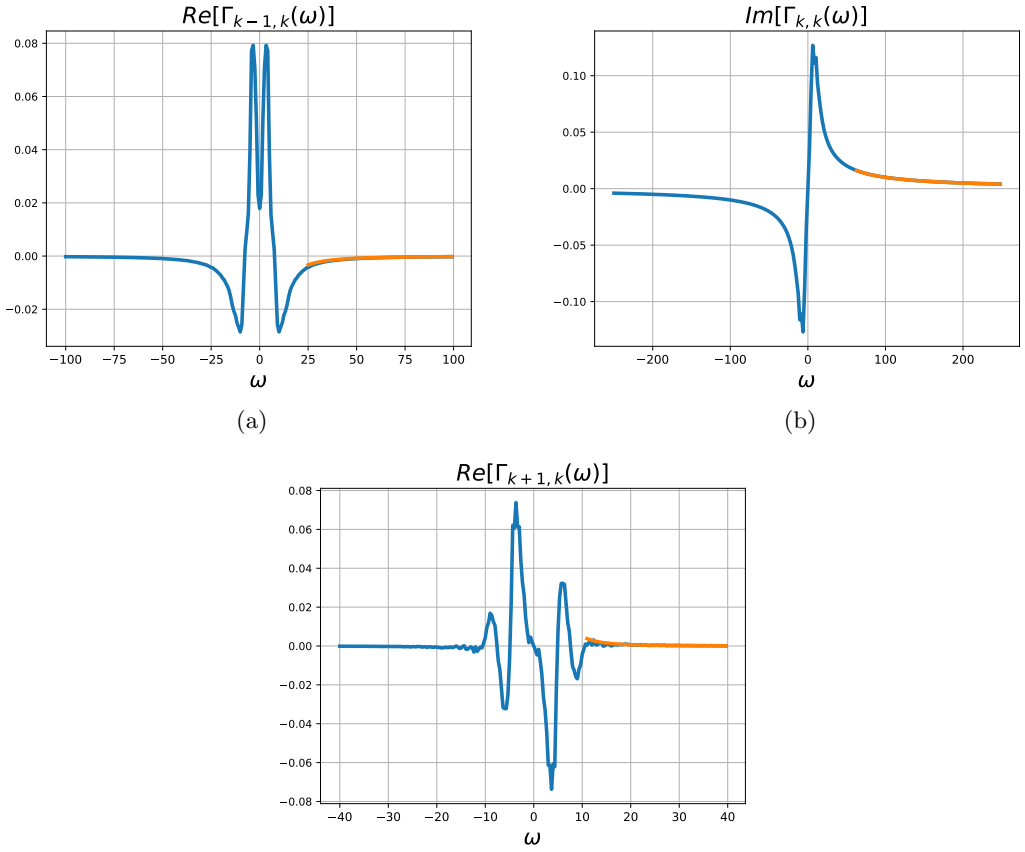


Figure 4.6: Comparison of the asymptotic behaviour of the Fourier transform of $\Gamma_{k-1,k}$, $\Gamma_{k,k}$ and $\Gamma_{k+1,k}$ with the predictions provided by (4.20), with the values provided by the fit results in Tab. 4.1. In (a) there is $\mathrm{Re}[\Gamma_{k-1,k}]$, in (b) $\mathrm{Im}[\Gamma_{k,k}]$ and in (c) $\mathrm{Im}[\Gamma_{k+1,k}]$.

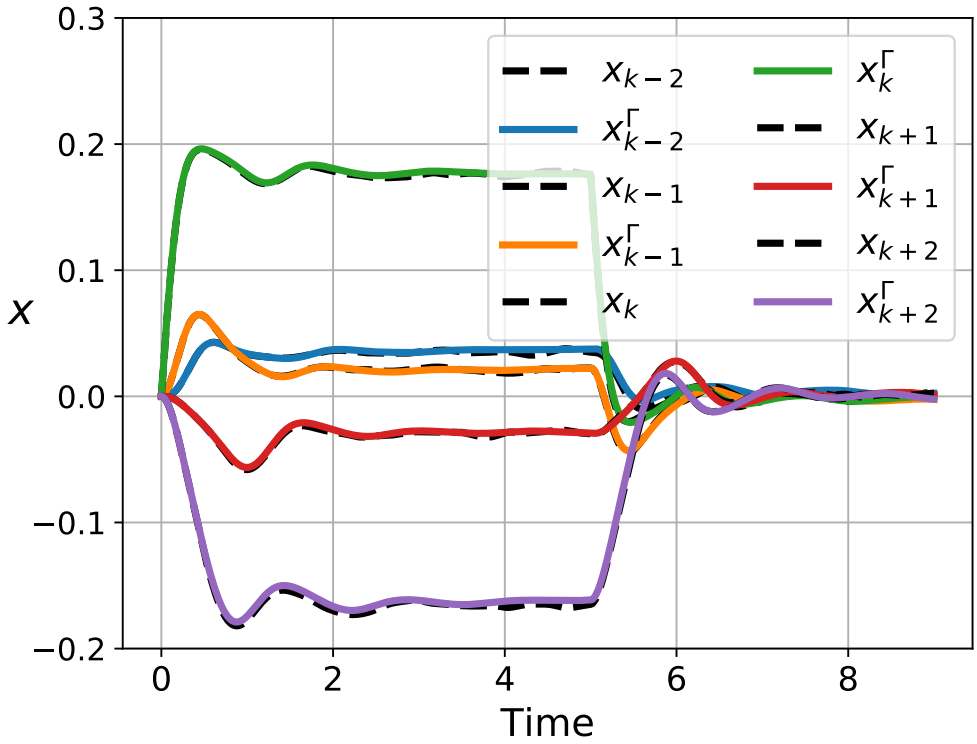


Figure 4.7: Integrated response of the perturbed L96 system (4.21) to the perturbation with time pattern $e(t) = (\Theta(t) - \Theta(t-\tau))$ with $\tau = 5$, compared to the prediction made by the response functions $\Gamma_{i,k}$ for $i \in \{k-2, k-1, k, k+1, k+2\}$. The integrated response is denoted by x_i, while the predicted one by x_i^{Γ}.

where $\delta x_i^{(k)}(t)$ is the difference between the trajectory integrated in the perturbed L96 system (4.21) and the trajectory integrated in the unperturbed L96 system (4.1). Both these trajectories start from the same initial condition, with index k, picked from the ensemble. To improve the results, as we have done for the response functions, we remove second-order contributions taking the semi-difference between the trajectory integrated using the L96 system perturbed with time pattern $e(t)$ and the trajectory integrated using the L96 system perturbed with time pattern $-e(t)$ [6].

Next, we compare the result with the response predicted by the response functions $\Gamma(t)$ through the following convolution:

$$\delta\langle x_i(t)\rangle = \int_{-\infty}^{t} dt\, \Gamma_{i,k}(t-\tau)\, e(\tau), \qquad i \in \{k-2, k-1, k, k+1, k+2\} \tag{4.24}$$

for the temporal pattern $e(t)$ given by (4.22). The comparison between the actual response and the prediction for $\tau = 5$ is shown in Fig. 4.7. We can see that the prediction performed by the response function perfectly overlaps the integrated response in all its components: the initial kick, the following plateaux and the relaxation to the steady state after the time interval τ.

4.2.2 Surrogate linear response functions

As explained above, we have perturbed the L96 system locally in x_k and we have looked at the response in x_i, with $i \in \{k-2, k-1, k, k+1, k+2\}$ through the linear response functions $\Gamma_{i,k}$. We want now to numerically compute the surrogate response functions H_{ij}, with $i, j \in \{k-2, k-1, k, k+1, k+2\}$. In other words, we want to see if we are able to predict the response in time of one of these sites x_i using the response in another site x_j. To compute the surrogate response functions H_{ij}, as explained in Section 2.6, we have to perform the inverse Fourier transform of the following ratio between Fourier transforms of the response functions $\Gamma_{i,k}$ and $\Gamma_{j,k}$:

$$H_{ij}(\omega) = \frac{\Gamma_{i,k}(\omega)}{\Gamma_{j,k}(\omega)}. \tag{4.25}$$

Asymptotic behaviour

It is possible to obtain the asymptotic behaviour of the Fourier transform of the surrogate response function H_{ij} using the asymptotic behaviour of the response functions in (4.20). In particular, if $\Gamma_{i,k}$ goes like $1/\omega^{\alpha_i}$ and $\Gamma_{j,k}$ goes like $1/\omega^{\alpha_j}$, we have that:

$$\begin{aligned} H_{ij}(\omega) &\underset{\omega\to\infty}{\propto} \frac{1/\omega^{\alpha_i}}{1/\omega^{\alpha_j}} \\ &\underset{\omega\to\infty}{\propto} \omega^{\alpha_j - \alpha_i}. \end{aligned} \tag{4.26}$$

If $\alpha_i \leq \alpha_j$ we have that $H_{ij}(\omega)$ diverge for large ω. As a consequence, the surrogate response function $H_{ij}(t)$ in the time domain will have a singular component $S_{ij}(t)$ since, as shown in Section 2.6, we have that the inverse Fourier transform of $(-i\omega)^\alpha$ is $\delta^\alpha(t)$, i.e. the $j-$th derivative of the delta function $\delta(t)$. In that case, the surrogate response function can be written as:

$$H_{ij}(t) = S_{ij}(t) + K_{ij}(t), \tag{4.27}$$

where $K_{ij(t)}$ is the non-singular component and $S_{ij}(t)$ is given by huge peaks around $t = 0$. On the contrary, if $\alpha_i > \alpha_j$ the surrogate response function $H_{ij}(t)$ has no singular component S_{ij}.

A hierarchy between variables

As it is explained by the relations (3.22) and (3.23), a time delay in a response function $\Gamma_{i,k}(t)$ corresponds to a zero with positive imaginary part in the Fourier transform of the response function $\Gamma_{i,k}(\omega)$. The longer is the delay and then the smaller is the imaginary part of the pole. That fact has consequences on the predictive ability of the variable i: the smaller is the imaginary part of the pole and the bigger is the non-predictive component in the surrogate response functions $H^{nc}_{ji}(t)$ where i tries to predict other observables. This is reasonable: the more delayed is the response of a variable and the more information after a generic time t is required to perfectly reconstruct the state of another variable j at time t. In the specific case of a local perturbation in the L96 system, we have seen in (4.19) that different variables respond after different time scales after the perturbation in the site k: the bigger that time scale is and the less predictive that variable will be.

As a consequence, a local perturbation in the site k creates a hierarchy of dynamical variables based on their predictive power. The higher is the rank of a variable in that hierarchy and the more it is predictive. It is remarkable that a variable with a given rank cannot even try to predict the time behaviour of the variables of higher ranking. This is due to the fact that we would use a response with a given time delay to predict another response with a smaller delay. In other words, we would use an *anticipated* surrogate response function H_{ij}, since we would anticipate the variable with smaller delay. We can see that with the following computation. Let's consider the Fourier transform of a surrogate response function $H_{ij}(\omega)$ which is such that:

$$\begin{aligned} H_{ij}(\omega) =& \frac{\Gamma_{i,k}(\omega)}{\Gamma_{j,k}(\omega)} \\ =& \frac{(\omega - i\omega_i)}{(\omega - i\omega_j)} Z_{ij}(\omega), \end{aligned} \tag{4.28}$$

where both the Fourier transform of the response functions $\Gamma_{i,k}(\omega)$ and $\Gamma_{j,k}(\omega)$ have a non-predictive zero, i.e. $\omega_i > 0$ and $\omega_j > 0$. We have defined a suitable function $Z_{ij}(\omega)$ which contains all the rest of the surrogate response function. Similarly to what we have done in (4.19) [29], we perform a double Padé approximation using the Padè approximant [75] for the exponential function (3.21). To use this Padè approximant we assume that the both ω_i and ω_j are sufficiently big, hence that the non-causal parts of the surrogate response functions which use them as predictors are small (as explained in Section 2.6.2). As a consequence, we have that:

$$\begin{aligned} H_{ij}(\omega) =& \frac{(\omega - i\omega_i)}{(\omega + i\omega_i)} \frac{(\omega + i\omega_j)}{(\omega - i\omega_j)} \tilde{Z}_{ij}(\omega) \\ \approx& e^{-2\left(\frac{\omega}{i\omega_i} - \frac{\omega}{i\omega_j}\right)} \tilde{Z}_{ij}(\omega), \end{aligned} \tag{4.29}$$

where $\tilde{Z}$ is another function suitably defined. Now if we go in the time domain we see that the exponential factor in (4.29) gives a delay in the time domain:

$$\begin{aligned} H_{ij}(t) =& \int \frac{d\omega}{2\pi} H_{ij}(\omega) e^{-i\omega t} \\ \approx& \int \frac{d\omega}{2\pi} \tilde{Z}(\omega) e^{-i\omega\left(t - 2\left(\frac{1}{\omega_i} - \frac{1}{\omega_j}\right)\right)} \\ \approx& \tilde{Z}_{ij}\left(t - 2\left(\frac{1}{\omega_i} - \frac{1}{\omega_j}\right)\right). \end{aligned} \tag{4.30}$$

For smaller ω_i and ω_j we can go on with the Padé approximations, adding corrections to the delay $\left(\frac{1}{\omega_i} - \frac{1}{\omega_j}\right)$. As a consequence, the surrogate response function H_{ij} has a time delay which depends on the difference between the time delays of the response of the two variables used as predictor and predictand. We have two possibilities:

- The response of i has a longer delay than the one of j: $\left(\frac{1}{\omega_i} - \frac{1}{\omega_j}\right) > 0$. In this case we can refer to the surrogate response function as *retarded*.
- The response of i has a shorter delay than the one of j: $\left(\frac{1}{\omega_i} - \frac{1}{\omega_j}\right) < 0$. In this case the surrogate response function is called *anticipated*. We will not deal with this kind of surrogate response function since they violate the causality principle.

The choice to focus only on the retarded surrogate response functions implies that their asymptotic behavior of their Fourier transform (4.26) for large ω will be at most constant. In particular, the only case where a singular component $S_{ij}(t)$ can be present in the surrogate response function we consider is that of two dynamical variables which have the same asymptotic behavior for small t, hence they have the same rank. In that case the singular component is proportional to a delta function $\delta(t)$:

$$S_{ij}(t)) = s_{ij}\delta(t), \qquad s_{ij} \in \mathbb{R}. \tag{4.31}$$

Moreover, we observe in (4.30) that the larger is the difference between the delays of the response functions $\Gamma_{i,k}$ and $\Gamma_{j,k}$, i.e. the difference between their ranks, and the larger is the delay of the related surrogate response function H_{ij}. As a consequence, the non-causal component of the surrogate response function will grow larger. This can be seen more directly performing the inverse Fourier transform of (4.28) (with the simplifying assumption that $Z(\omega)$ has no poles):

$$H_{ij}(t) = (\omega_j - \omega_i)Z_{ij}(\omega_j)e^{t\omega_j}, \quad t < 0. \tag{4.32}$$

We can see from (4.32) that the non-causal component which emerges from the inverse Fourier transform is proportional to the difference between the non-predictive poles. We can thus deduce that the bigger is the difference between the time delays (proportional to the inverse of the pole), the larger in magnitude will be the non-causal component.

Now that we have introduced the hierarchy in the predictive power of the dynamical variables, let's see some examples of surrogate response functions, using the response functions we have numerically computed shown in Fig. 4.4. We can observe the actual propagation of the signal in Fig. 4.5, where different dynamical variables are perturbed after different times from the local perturbation in x_k, hence building the hierarchy discussed above. The most predictive variable (the one with the highest rank) is the one which is directly perturbed, hence x_k, since it responds without any delay to the perturbation, as it can be seen in (4.19). In fact, if we look at the surrogate response functions $H_{j,k}$ showed in Fig. 4.8, we notice that the non-causal component is practically vanishing with respect to the causal one. The second most predictive variables are x_{k-1} and x_{k+2}. As discussed above, we will discard the anticipated surrogate response function $H_{k,k-1}$ and $H_{k,k+2}$ since they are non-physical. The retarded surrogate response functions are shown in Fig. 4.9 and Fig. 4.10. We can observe that the surrogate response functions show a rather small non-causal component. Notice that just the non-singular component of the surrogate response functions $K_{k-1,k+2}$ and $K_{k+2,k-1}$ is shown in the figures, for clarity purposes. In order to remove the singular component, we have removed from the Fourier transform of the surrogate response function its asymptotic behavior. For $H_{k-1,k+2}$, the asymptotic behavior is:

$$\begin{aligned} H_{k-1,k}(\omega) &\underset{\omega\to\infty}{\to} \frac{-G\langle C^{(1)}_{k-1}\rangle/\omega^2}{-G\langle C^{(1)}_{k+2}\rangle/\omega^2} \\ H_{k-1,k+2}(\omega) &\underset{\omega\to\infty}{\to} -1, \end{aligned} \tag{4.33}$$

as we can deduce using (4.18) and (4.20). The asymptotic behavior of $H_{k+2,k-1}$ is just the inverse of (4.33), i.e -1. Hence we have $s_{k-1,k+2} = s_{k+2,k-1} - 1$, using the relation

$s_{k+2,k-1}$	$s_{k-1,k+2}$	$s_{k+1,k-2}$	$s_{k-2,k+1}$
-1	-1	-1/0.90	-0.90

Table 4.2: Constant which multiplies the delta function in the singular component of the surrogate response functions $H_{k+2,k-1}$, $H_{k-1,k+2}$, $H_{k+1,k-2}$ and $H_{k-2,k+1}$, as defined in (4.31).

(4.31). The third most predictive variables between the ones we have considered are x_{k-2} and x_{k+1}. In Fig. 4.11 we show the non-singular part of the surrogate response functions $K_{k-2,k+1}$ and $K_{k+1,k-2}$. To remove the constant limit behavior of their Fourier transform we proceeded as in (4.33):

$$\begin{aligned} H_{k-2,k+1}(\omega) &\underset{\omega\to\infty}{\to} \frac{-2iG\langle C^{(2)}_{k-2}\rangle/\omega^3}{-2iG\langle C^{(2)}_{k+1}\rangle/\omega^3} \\ H_{k-2,k+1}(\omega) &\underset{\omega\to\infty}{\to} \frac{\langle C^{(2)}_{k-2}\rangle}{\langle C^{(2)}_{k+1}\rangle}, \end{aligned} \tag{4.34}$$

where $\langle C^{(2)}_{k-2}\rangle$ and $\langle C^{(2)}_{k+1}\rangle$ can be computed directly from the expansion for small times (4.19) of the response functions $\Gamma_{k-2,k}$ and $\Gamma_{k+1,k}$:

$$\begin{aligned} \langle C^{(2)}_{k-2}\rangle &= \langle x_{k-3}x_{k-2}\rangle \\ \langle C^{(2)}_{k+1}\rangle &= -\langle x_{k-3}x_{k-2}\rangle - \langle x_{k-2}x_k\rangle. \end{aligned} \tag{4.35}$$

We can compute the expectation values in (4.35) directly from the unperturbed L96 system, obtaining the results shown in Tab. 4.2. The asymptotic behavior of $H_{k+1,k-2}(\omega)$ is given by $s_{k+1,k-2} = 1/s_{k-2,k+1}$. We can notice that the non-causal components of these non-singular parts of the surrogate response functions are more relevant than the ones considered before, as expected. It is noticeable the non-causal component of $K_{k-2,k+1}$ is significantly larger than the one of $K_{k+1,k-2}$. As discussed below, this is consistent with the fact that the group velocity v_g, which is the velocity of the flow of information, goes from left to right as showed in (4.9).

Quantifying the predictive power

We quantified the importance of the non-causal component of the surrogate response functions derived numerically by means of the ratio method (2.150) derived in Section 2.6. The results are presented in Table 4.3 for the retarded surrogate response functions between the dynamical variables x_{k-2}, x_{k-1}, x_{k+1} and x_{k+2}. For the surrogate response functions $H_{k-1,k+2}$, $H_{k-2,k+1}$ and their inverse, which display a singular component, we may wonder how much the numerical (and hence imperfect) representation of the delta functions impacts on the ratio method. In order to check that, we have also applied the ratio method to just the non-singular components of these surrogate response functions. The results are listed in Tab. 4.4. We can see that the ratio method is lowered considering just the non-singular component, in the large majority of cases.

Looking at the values provided by the ratio method, we observe that the weight of the non-causal part is bigger for the predictions made by $k-2$ and $k+1$, which are the less predictive variables between the ones we have considered. This observation is valid

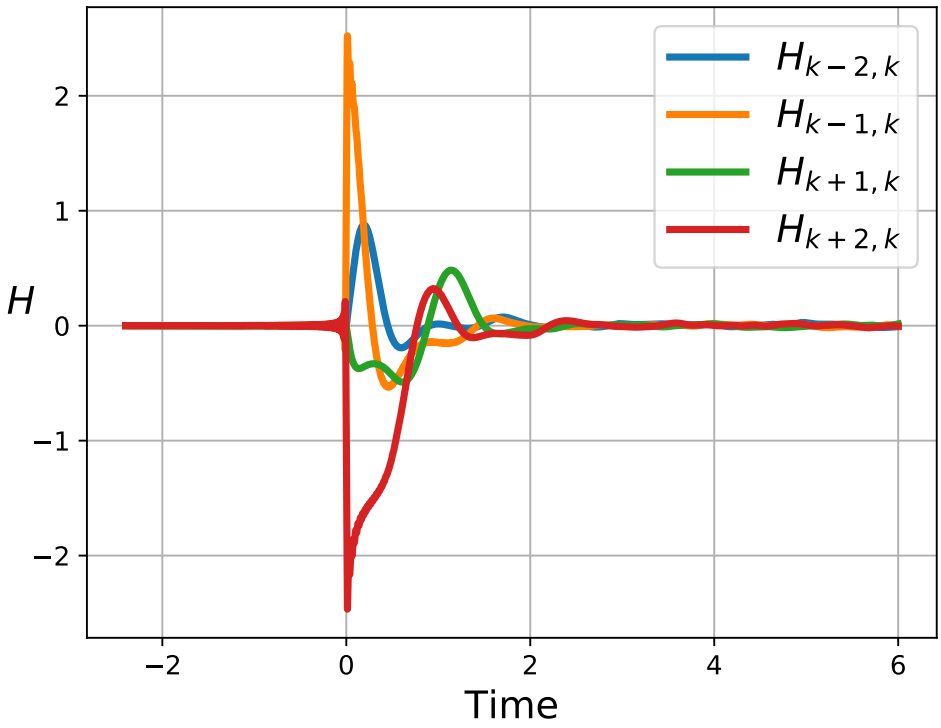
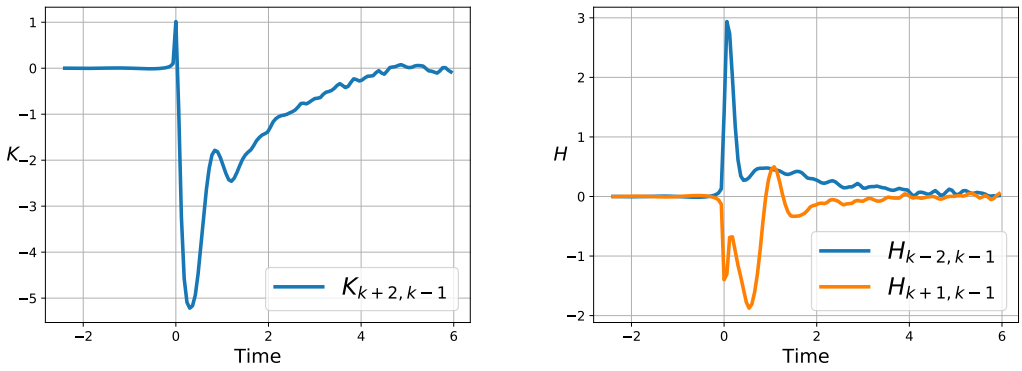
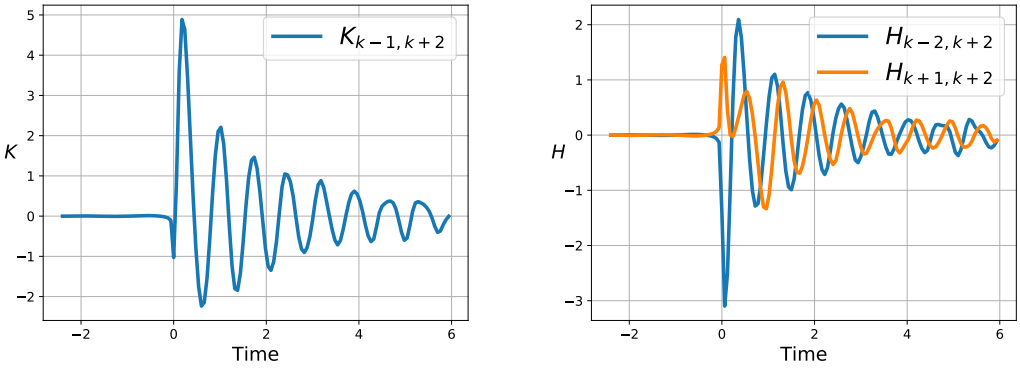


Figure 4.8: Surrogate response functions $H_{i,k}$ for $i \in \{k-2, k-1, k, k+1, k+2\}$.



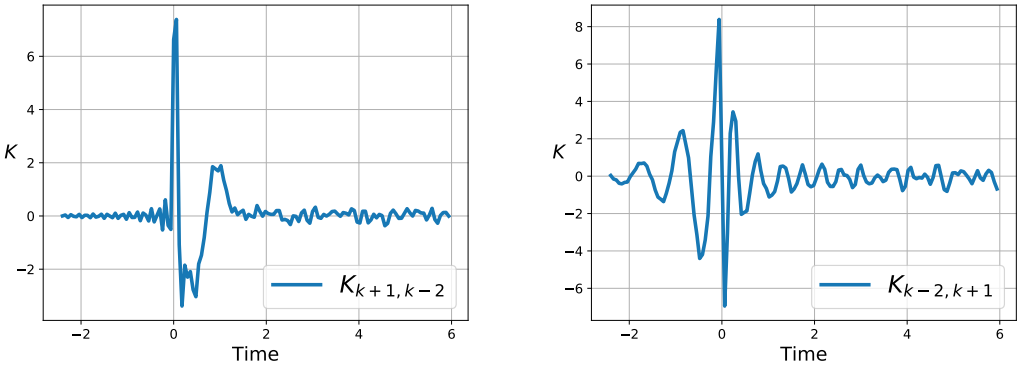
(a) (b)

Figure 4.9: (a) Non-singular part of the surrogate response function $K_{k-2,k+1}$. (b) Surrogate response functions $H_{k-2,k-1}$ and $H_{k+1,k-1}$.



(a) (b)

Figure 4.10: (a) Non-singular part of the surrogate response function $K_{k-1,k+2}$, for. (b) Surrogate response functions $H_{k-2,k+2}$ and $H_{k+1,k+2}$.



(a) (b)

Figure 4.11: (a) Non-singular part of the surrogate response function $K_{k+1,k-2}$. (b) Non-singular part of the surrogate response function $K_{k-2,k+1}$.

	x_{k-2}	x_{k-1}	x_{k+1}	x_{k+2}
x_{k-2}	·	·	0.30	·
x_{k-1}	0.013	·	0.019	0.12
x_{k+1}	0.68	·	·	·
x_{k+2}	0.0068	0.13	0.014	·

Table 4.3: Results of the ratio method (2.150) applied to the retarded surrogate response functions between the dynamical variables x_{k-2}, x_{k-1}, x_{k+1} and x_{k+2}. In the first column there are the predictors, while in the first row there are the predictands.

	H	K
$(k-1, k+2)$	0.12	0.0025
$(k+2, k-1)$	0.13	0.0037
$(k-2, k+1)$	0.68	0.66
$(k+1, k-2)$	0.30	0.11

Table 4.4: Result of the ratio method (2.150) applied to the retarded surrogate response functions which display a singular component. The method is applied to both the full response function (H) and just the non-singular component (K).

for both the full surrogate response function and just the non-singular component, so it is not due to the numerical representation of the delta function.

Another interesting observation we can make is the presence of a definite direction of the predictive power. If we take two variables x_i and x_j, we can ask ourselves whether it is better to use x_i to predict x_j or the other way around. Due to the fact that we have a hierarchy between variables in terms of their predictive power, we are able to say that it is the variable with higher rank the one which has to be used as predictor. Otherwise we would use an anticipated response function. The ambiguity lies in the case of two variables x_i and x_j which have the same rank. This is the case of x_{k-1} and x_{k+2} and of x_{k-2} and x_{k+1}. We can make use of the tables 4.3 and 4.4, in order to have a quantitative way to make our decision. We see that in both cases the variable at the left of perturbation site is the best predictor, since the ratio methods gives lower values for $H_{k-2,k+1}$ and $H_{k-1,k+2}$ with respect to the values for $H_{k+1,k-2}$ and $H_{k+2,k-1}$. There is a definite verse for the flow of information. This is consistent with the fact that the group velocity v_g of the travelling waves in the L96 system is positive (from left to tight) (4.9), since the group velocity is related to the information transport in the system.

Making predictions

Now we test the predictive ability of the surrogate response function computed above. We consider the L96 system (4.21) perturbed with a perturbation with spatial pattern $G(x) = G\delta_{i,k}$ given by (4.10) and time pattern (4.22):

$$e(t) = \Theta(t) - \Theta(t - \tau),$$

with $\tau = 5$. We start from the most predictive variable x_k, which tries to predict x_j, with $j \in \{k-2, k-1, k+1, k+2\}$ through the following convolution, where we forced the causality of the surrogate response function, as in (2.121):

$$\delta\langle x_i\rangle(t) = \int_{-\infty}^{t} d\tau\, H'_{ij}(t-\tau)\delta\langle x_j\rangle(\tau). \tag{4.36}$$

The responses of the dynamical variables $\delta\langle x_i\rangle$ and $\delta\langle x_j\rangle$ are computed using (4.23). The results are shown in Fig. 4.12. The comparison between the integrated response and the prediction made by the surrogate response functions is very good.

We now consider the second most predictive dynamical variables x_{k-1} and x_{k+2}. Since we are focusing on just the retarded surrogate response functions, x_{k-1} can predict only $x_{k-2,k+1,k+2}$ while x_{k+2} can predict only $x_{k-2,k-1,k+1}$. The predictions are showed in Fig. 4.13. We notice that the predictions work very well. We could have expected these good predictions by the fact that the non-singular component of the surrogate response functions are very small, as shown in Fig. 4.9 and 4.10. We remark that for the prediction made using $H_{k+2,k-1}$ and $H_{k-1,k+2}$, we have not directly used the surrogate response function obtained performing the inverse Fourier transform of the ratio of the response functions (4.25), in order to avoid errors due to the finite numerical representation of the delta function. We instead used the following trick, which is possible because we know that the singular component is proportional to just a delta function, as shown in (4.31). Knowing that, we can rewrite the convolution as follows (4.36):

$$\begin{aligned}\delta\langle x_i\rangle(t) &= \int_{-\infty}^{t} d\tau\, K'_{ij}(t-\tau)\delta\langle x_j\rangle(\tau) + \int_{-\infty}^{t} d\tau\, s_{ij}\delta(t-\tau)\delta\langle x_j\rangle(\tau) \\ &= \int_{-\infty}^{t} d\tau\, K'_{ij}(t-\tau)\delta\langle x_j\rangle(t) + s_{ij}\delta\langle x_j\rangle(t),\end{aligned} \tag{4.37}$$

where the convolution is made just using the non-singular component of the surrogate response function. The factor s_{ij} are listed in Tab. 4.2. It is remarkable that the worst prediction performed by x_{k-1} is the one about x_{k+2} and the the worst performed by x_{k+2} is the one about x_{k-1}. This is consistent with the quantitative results showed in Table 4.3.

Lastly, we take into account the predictions made the less predictive variables x_{k-2} and x_{k+1}, where we used the trick (4.37). The variables x_{k-2} and x_{k+1} can only predict each other since we are dealing with retarded response functions. As we can see in Fig. 4.14 the predictions do not work well, even though it has similar qualitative behaviour. We could have expected that from the fact that the non-causal component of the surrogate response functions $H_{k-2,k+1}$ and $H_{k+1,k-2}$ is relevant, as it can be seen in 4.11 and as it can be quantified in Tab. 4.4. Remarkably, we can observe in Fig. 4.14 the asymmetry between the predictive power of x_{k-2} and x_{k+1} we discussed before: x_{k+1} makes a prediction which differs more from the actual response than x_{k-2}, providing a hint for the presence of a definite verse of the flow of information within the L96 system.

4.2.3 Making predictions with more observables

We have seen that some local variables cannot predict other local variables, due to how the perturbation propagates. We want to show now that these local variables can

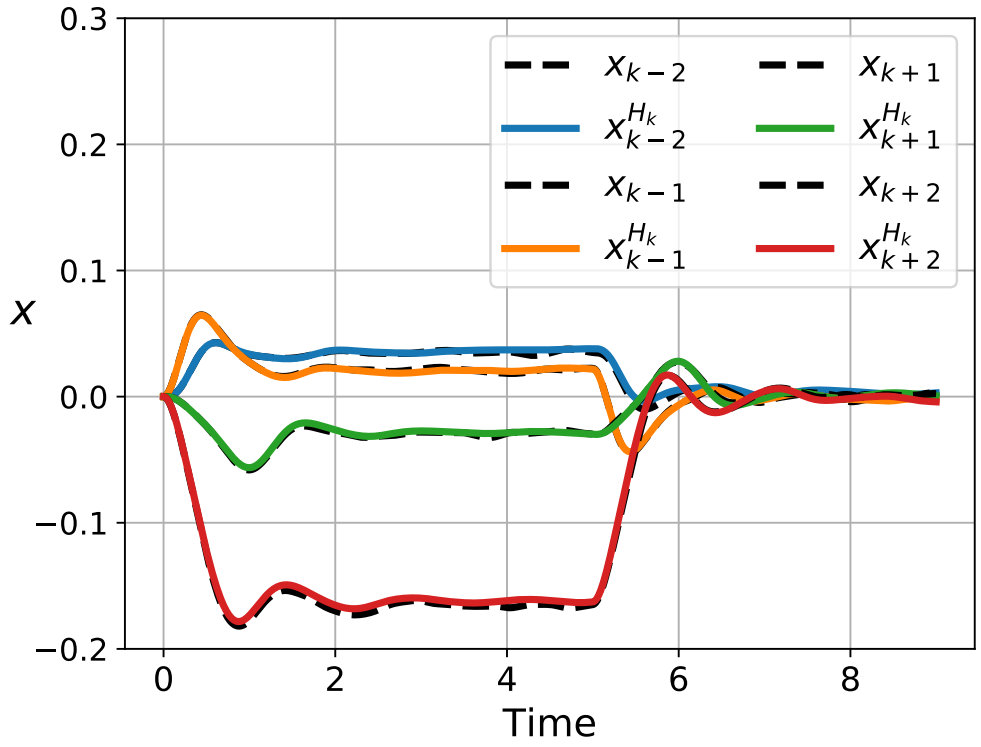
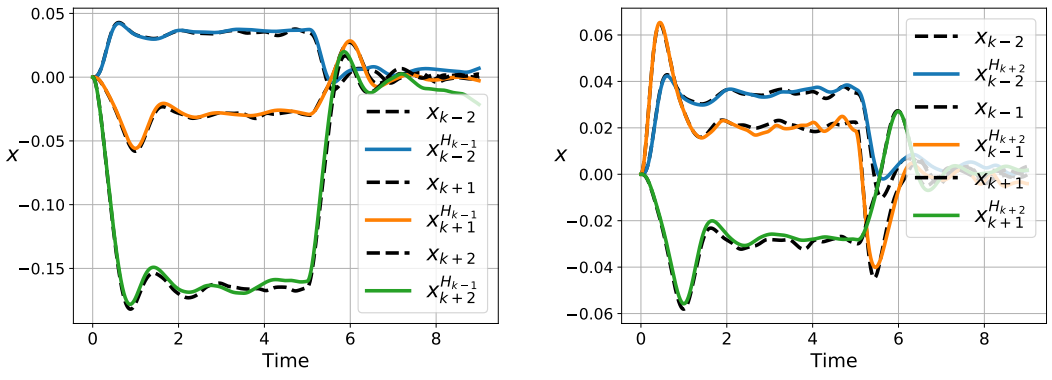
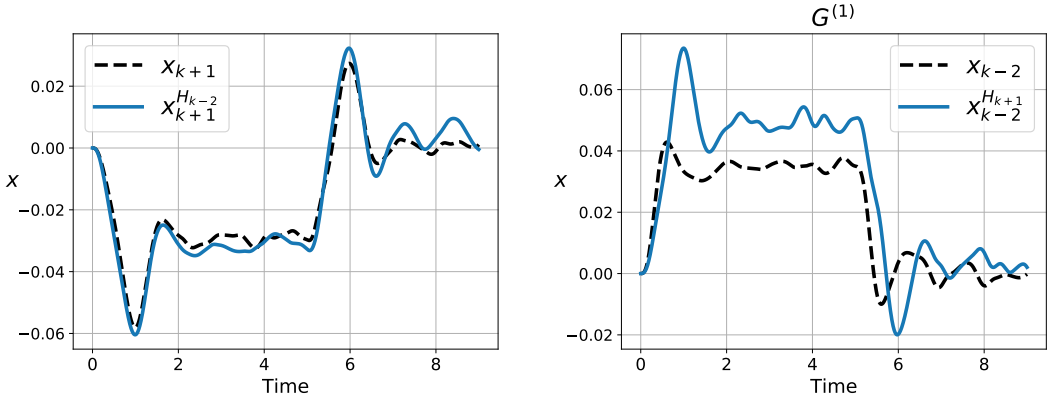


Figure 4.12: Integrated response of the perturbed L96 system (4.21) to the perturbation with time pattern $e(t) = (\Theta(t) - \Theta(t-\tau))$ with $\tau = 5$, compared with the predictions made by the surrogate response functions $H_{i,k}$ for $i \in \{k-2, k-1, k+1, k+2\}$. The integrated response is denoted by x_i, while the predicted one by $x_i^{H_k}$.



(a) (b)

Figure 4.13: Integrated response of the perturbed L96 system (4.21) to the perturbation with time pattern $e(t) = (\Theta(t) - \Theta(t-\tau))$ with $\tau = 5$, compared with the predictions made by the surrogate response functions $H_{i,k-1}$ for $i = \{k-2, k+1, k+2\}$ in (a) and with the predictions made by the surrogate response functions $H_{i,k+2}$ for $i = \{k-2, k-1, k+1\}$ in (b). The integrated response is denoted by x_i while the predicted one by $x_i^{H_{k-1}}$ and $x_i^{H_{k+2}}$.



(a) (b)

Figure 4.14: Integrated response of x_{k+1} in (a) and x_{k-2} in (b) of the perturbed L96 system (4.21) to the perturbation with time pattern $e(t) = (\Theta(t) - \Theta(t-\tau))$ with $\tau = 5$, compared with the predictions made by the surrogate response functions $H_{k+1,k-2}$ and $H_{k-2,k+1}$. The integrated response is denoted by x_i while the predicted one by $x_i^{H_{k-2}}$ and $x_i^{H_{k+1}}$.

still make predictions of other local variables, using also other observables to make predictions, as explained in Section 2.6. In particular, we want to improve the local variables' prediction using an additional local perturbation. The situation we have in mind is that of a local observer who wants to predict the state of an observable in other zones of the globe. First, he/she tries to use just the information provided by a local perturbation of the system, in a given zone of the globe. He/she can be unlucky and he/she can be in a location which in principle has no predictive power. In this case, he/she can improve his predictions considering the information given by another perturbation. We consider the particular case of x_{k+1} which tries to predict x_{k-2}. Due to the relevant non-causal part of the surrogate response function $H_{k-2,k+1}$, which can be seen in Fig. 4.11 (b), this prediction, shown in Fig. 4.14 (b) is unsuccessful. To improve the prediction, we will consider the following global observable, defined over the whole set of dynamical variables:

$$\Psi_1(t) \equiv \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{4.38}$$

which can be seen as the average energy of the L96 system at a time t. In order to use two observables as predictors, we perturb the system with two different forcings as in (2.131). The first one is the local perturbation of the forcing (4.10) used before, with time pattern given by (4.22):

$$\begin{cases} G_i^{(1)}(x) = \delta_{ik} G_1 \\ e^{(1)}(t) = \Theta(t) - \Theta(t-\tau_1), \end{cases} \tag{4.39}$$

with $G_1 = 1$ and $\tau_1 = 5$, as before. The second forcing we consider is a local perturbation in the viscosity of the system, which is the following:

$$\begin{cases} G_i^{(2)}(x) = -\, x_i\, \delta_{ik} G_2 \\ e^{(2)}(t) = \Theta(t) - \Theta(t-\tau_2), \end{cases} \tag{4.40}$$

with $G_2 = 0.1$ and $\tau_2 = 3$. At this point we can use the relation for the Fourier transform of the surrogate response function (2.141) for three observables:

$$\begin{aligned} H_{k-2,\Psi_1}(\omega) =& \frac{\left(\Gamma_{k+1,G^{(2)}}\Gamma_{k-2,G^{(1)}} - \Gamma_{k+1,G^{(1)}}\Gamma_{k-2,G^{(2)}}\right)}{\left(\Gamma_{k+1,G^{(2)}}\Gamma_{\Psi_1,G^{(1)}} - \Gamma_{\Psi_1,G^{(2)}}\Gamma_{k+1,G^{(1)}}\right)}(\omega) \\ H_{k-2,k+1}(\omega) =& \frac{\left(-\Gamma_{\Psi_1,G^{(2)}}\Gamma_{k-2,G^{(1)}} + \Gamma_{\Psi_1,G^{(1)}}\Gamma_{k-2,G^{(2)}}\right)}{\left(\Gamma_{k+1,G^{(2)}}\Gamma_{\Psi_1,G^{(1)}} - \Gamma_{\Psi_1,G^{(2)}}\Gamma_{k+1,G^{(1)}}\right)}(\omega) \end{aligned} \tag{4.41}$$

where H_{k-2,Ψ_1} links the global predictor Ψ_1 and the predictand x_{k-2}, while $H_{k-2,k+1}$ links the local predictor x_{k+1} and the predictand x_{k-2}. We can see that these surrogate response functions have in general different poles than the ones previously defined in the case of just one forcing. We now want to investigate the eventual presence of singular components in these surrogate response functions. To do that, we take the limit for $\omega \to \infty$ of the relations (4.41) and we use the limits obtained for the response functions which appear in these relations, listed in Tab. 4.5. These limits are obtained taking the limit for $t \to 0^+$ of the response functions in the time domain and using the relations (2.93) and (2.94). Given these results, the limit behaviour of (4.41) is the following:

$$\begin{aligned} \lim_{\omega\to\infty} H_{k-2,\Psi_1}(\omega) =& 0 \\ \lim_{\omega\to\infty} H_{k-2,k+1}(\omega) =& s_{k-2,k+1}, \end{aligned} \tag{4.42}$$

where $s_{k-2,k+1}$ is a constant which can be obtained as follows:

$$\lim_{\omega\to\infty} H_{k-2,k+1}(\omega) = \frac{-\langle x\rangle\langle C^{(2)}_{k-2}\rangle + \langle D^{(2)}_{k-2}\rangle}{\langle D^{(2)}_{k+1}\rangle - \langle x\rangle\langle C^{(2)}_{k+1}\rangle}, \tag{4.43}$$

where $\langle C^{(2)}_{k-2}\rangle$ and $\langle C^{(2)}_{k+1}\rangle$ are given by (4.35), while $\langle D^{(2)}_{k-2}\rangle$ and $\langle D^{(2)}_{k+1}\rangle$ can be computed from the expansion for small times of $\Gamma_{k+1,G^{(2)}}$ and $\Gamma_{k-2,G^{(2)}}$, obtaining:

$$\begin{aligned} \langle D^{(2)}_{k-2}\rangle =& \langle x_{k-3}x_{k-2}x_k\rangle \\ \langle D^{(2)}_{k+1}\rangle =& -3\langle x_{k-2}x_k\rangle + 3\langle x_{k-1}x_k\rangle + 2\langle x_k x_{k+1}x_{k+3}\rangle - 2\langle x_k^2 x_{k-1}\rangle - 2\langle x_k^2 x_{k-2}\rangle. \end{aligned} \tag{4.44}$$

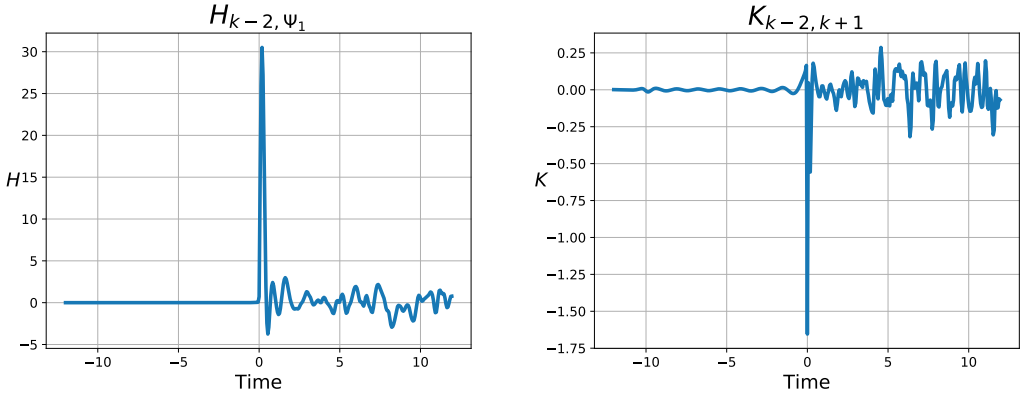
The expectation values on (4.44) can be found directly from the unperturbed L96 system, and they gives $s_{k-2,k+1} = 0.071$. Looking at (4.42) we can deduce that H_{k+1,Ψ_1} has no singular component since it vanishes for $\omega \to \infty$, while $H_{k+1,k-2}$ has a singular component proportional to a delta:

$$S_{k+1,k-2}(t) = s_{k-2,k+1}\delta(t). \tag{4.45}$$

In Fig. 4.15 the surrogate response function H_{k+1,Ψ_1} and the non singular component $K_{k-2,k+1}$ of the surrogate response function $H_{k+1,k-2}$ are shown. We can notice that the non-causal component is much smaller than the one displayed in Fig. 4.11 for both the surrogate response functions. We can quantify that observation using the ratio method (2.150), obtaining the results shown in Tab. 4.6, where we can see that the relative importance of the non-causal component is lowered a lot considering two

	$\omega \to \infty$
$\Gamma_{\Psi_1, G^{(1)}}$	$(iG_1/(N\omega))$
$\Gamma_{k+1, G^{(1)}}$	$-2iG_1\langle C^{(2)}_{k+1}\rangle/\omega^3$
$\Gamma_{k-2, G^{(1)}}$	$-2iG_1\langle C^{(2)}_{k-2}\rangle/\omega^3$
$\Gamma_{\Psi_1, G^{(2)}}$	$-(iG_2/(N\omega)) \cdot \langle x\rangle$
$\Gamma_{k+1, G^{(2)}}$	$2iG_2\langle D^{(2)}_{k+1}\rangle/\omega^3$
$\Gamma_{k-2, G^{(2)}}$	$2iG_2\langle D^{(2)}_{k-2}\rangle/\omega^3$

Table 4.5: Limit behaviour for $\omega \to \infty$ of the Fourier transform of the response functions for the observable Ψ_1, x_{k-2} and x_{k+1} and the perturbation with local spatial pattern $G^{(1)}$ (4.39) and global spatial pattern $G^{(2)}$. The constants $C^{(2)}_{k+1}$ and $C^{(2)}_{k-2}$ are defined as in (4.19), while the constants $D^{(2)}_{k+1}$ and $D^{(2)}_{k-2}$ are defined in a similar way for the perturbation $G^{(2)}$. The average $\langle x\rangle$ can be computed directly by an ensemble average.



(a) (b)

Figure 4.15: (a) Surrogate response function H_{k-2,Ψ_1}. (b) Non-singular part of the surrogate response function $K_{k-2,k+1}$.

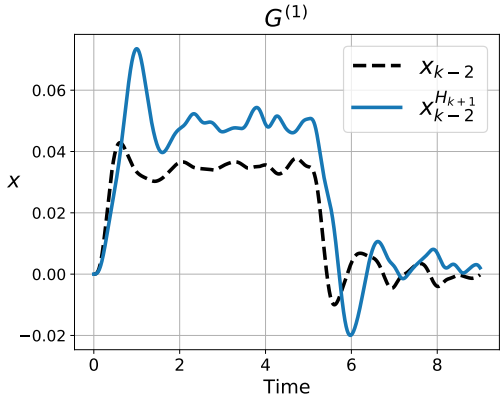
forcings. This is reasonable, as we extract more information from the system looking at its response to two different perturbations. Given the fact that $H_{k+1,k-2}$ has a singular component, we compute the response of x_{k-2} using the trick (4.37):

$$\delta\langle x_{k-2}\rangle(t) = H_{k-2,\Psi_2}(t) * \delta\langle\Psi_2\rangle(t) + K_{k-2,k+1}(t) * \delta\langle x_{k+1}\rangle(t) + s_{k-2,k+1}\delta\langle x_{k+1}\rangle(t). \tag{4.46}$$

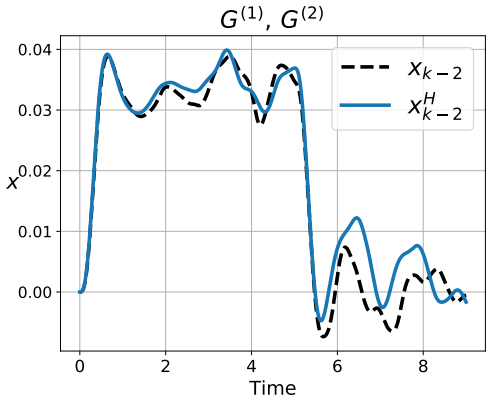
We can see in Fig. 4.16 that the prediction performed by the surrogate response functions in the case of two forcings (bottom panel) is much better than the one made using just one forcing (top panel). We can see that using more forcings and hence more observable to make predictions, we gain access to more information and we can make better predictions of what happens in the other locations of the system.

$K_{k-2,k+1}$	H_{k-2,Ψ_1}	$K_{k-2,k+1}$
0.66	0.0012	0.086

Table 4.6: In the first column there is the result of the ratio method applied to the surrogate response function $K_{k-2,k+1}$ when just one forcing $G^{(1)}$ is used. In the last two columns there are the results of the ratio method (2.150) applied to the surrogate response functions H_{k-2,Ψ_1} and the non-causal component K_{k-2,Ψ_1} when two forcings $G^{(1)}$ and $G^{(2)}$ are used.



(a)



(b)

Figure 4.16: Integrated response of x_{k-2} of the perturbed L96 system (4.21), compared (a) with the predictions made by the surrogate response functions $H_{k-2,k+1}$ with just one forcing (4.39) and (b) with the predictions made by the surrogate response functions H_{k-2,Ψ_1} and $H_{k-2,k+1}$ with the two forcings and (4.40) . The integrated response is denoted by x_i while the predicted one by x_i^H.

Conclusions

In our work we have first built a framework where all the possible approaches to RT have been included, reviewing the literature regarding RT. We started from the Kubo's Theory [1, 2], which studies how to predict the out of equilibrium response of a system which was at its thermodynamic equilibrium. In particular, the Kubo's theory is able to link through the FDT the spontaneous fluctuations occurring in the unperturbed steady state with the non-equilibrium response of the system. We have seen that the effectiveness of RT lies in the fact that it is a theory fundamentally built on ensembles of trajectories [4]. We have expanded the regime of applicability of RT to systems which display a NESS, showing that there are two main pathways to tackle the problem to predict the out-of-equilibrium response of the system with just the information provided by the steady state and the applied forcing [17].

The first approach assumes the smoothness of the invariant measure ρ of the unperturbed system and it derives the Kubo-Agarwal response formula [17], which reduces to the Kubo's response formula assuming the time-reversal property for the trajectory of the considered system. Remarkably, the Kubo's response formula structure is preserved in the Kubo-Agarwal formula, together with the FDT. This approach is particularly suited for systems which have a noise term in their equation of motion, which can emerge from a coarse-graining procedure or from the projection of the invariant microscopic measure along certain relevant directions [20].

There are cases where it is not possible to assume the smoothness of the invariant measure. This is the case of chaotic dissipative deterministic systems, which display a NESS since there is dissipation within the system. In such systems, the trajectories reach in the long time limit a set of points called strange attractor, which is usually fractal and with a dimension smaller than the one of the phase space [64]. As a consequence, the invariant measure is singular with respect to the Lebesgue measure, since it is supported on this attractor. Ruelle, in order to well define a RT for these systems, considered Axiom A systems. These systems are special because their invariant measure is an SRB measure, which is such that their tangent space can be split between an unstable manifold, a neutral manifold and a stable manifold. Ruelle showed in its seminal works [10, 67] that the Kubo formula is preserved along the unstable and central directions, where the measure is smooth, while the contribution along the stable manifold is deeply different and it is also harder to compute in practice than the unstable and central contributions [15, 17]. The main consequence of this fact is that the FDT holds just partially since the stable manifold contribution has no counterpart in the spontaneous fluctuations of the unperturbed steady state. Note that Axiom A

systems are a class of systems of a certain physical relevance, thanks to the chaotic hypothesis devised by Gallavotti and Cohen [11,12], which states that the reduced system composed by macroscopic observables in high-dimensional systems is practically an Axiom A system.

In all approaches to the RT presented above, the purpose is to predict the response of the perturbed system using the information given by the steady state of the unperturbed system and the forcing applied to the system, focusing on a few observable of interest. In Section 2.6, we have presented a rather new and insightful point of view in RT, introduced by Lucarini in his work [28]. Starting from the usual RT, Lucarini showed that it is possible to use the response of some perturbed observables as surrogates of the actual forcing to predict the response of other perturbed observables. This is achieved using some newly defined response functions, called surrogate response functions. We have seen that one of the main issues raised by the surrogate RT is about the presence of a non-causal component in the surrogate response functions, since some poles of their Fourier transform could be collocated in the upper complex $\omega-$ plane. This non-causal component could seriously hamper the predictive power of some perturbed observables, leading to the impossibility to predict the state of another perturbed observable at a time t using just the information provided up to time t. A meaningful question which can emerge from this fact is the following: can we actually quantify this loss of information? This question becomes important in cases where we have to choose the most predictive observable from a set of observables $\{\Psi_1, ..., \Psi_k\}$, to predict another observable Ψ. All the surrogate response functions H_{Ψ,Ψ_j} might display a non-causal component. In Section 2.6 we have derived for the first time a quantitative method which measures how much a predictor is unpredictive, starting from a generalized form of the KK relations for non-causal response functions. This new method computes the relevance of the non-causal component of the surrogate response function with respect to its causal component and it is very easy to numerically implement. Furthermore, this method can unveil the potential presence of information flows or causal links between couples of non-predictive observables. We have shown that in the L96 system in Chapter 4, where we have found results consistent with the group velocity present in the system, which is associated with the transport of information.

In this thesis we have shown the effectiveness of the surrogate RT, providing new results and insights in linear systems and a chaotic spatially extended non-linear dynamical system. In Chapter 3 we have implemented the surrogate RT formalism in the context of linearized systems, building response functions between different perturbed observables. We have devised and proved a criterion, called the Unpredictability Criterion (3.17), which can rule out a priori the perturbed observables which are generally not able to make predictions. It turns out, considering dynamical variables x_j such that the spatial pattern of the perturbation G_j is different from 0, that if the linear feedback seen by x_j is greater than the dissipation which occurs within the system, x_j cannot be generally used for prediction purposes. The derivation of the UC is a first step towards a general method which can allow us to discriminate a priori between predictive and not-predictive observables, just looking at the equation of the dynamics. We have provided numerical tests in a non-trivial case, which supported and confirmed the UC predictions. Next, we have plugged memory effects in the dynamical systems considered in the main body of the chapter, studying their impact on the UC. We have discovered that it is

more difficult that the variables are unpredictive in this case because the variables know more about each other. Lastly, we have extended the UC on dynamical systems defined through a random matrix, deriving in this context the average UC, which tells us if a dynamical variable is predictive on average. In particular, the average UC states, consistently with what observed before for the UC, that if there are too much negative feedbacks between the variables with respect to the self-interactions terms, on average they fail in their role of predictors. This is because the dynamical variables know way less about each other in this regime, and hence they cannot make faithful predictions.

A relevant first application of the UC can be found in the context of non-linear rational approximation [86]. In that context, the goal is to find the best rational function that fits the relation between two variables in the frequency domain [87]. The Criterion could provide some indications on the nature of the poles we expect to find after the fitting procedure and detect possible errors.

Moreover, we envision applications of our work in the context of control theory, where linearized systems are extensively studied and used. The UC can be useful in overcoming one of the main problems which emerges while designing a controller: the reconstruction of the state of the dynamical system starting from the sensed partial information. To do that, it is crucial to detect the most useful variables which can reconstruct the full state of the system. The surrogate RT could provide a set of tools to actually identify those variables, looking at their predictive ability. Our UC could be used as a method to immediately rule out the unpredictive variables looking directly at the equation of the dynamics.

Finally, our work can be used in control theory to understand whether a system is non-minimum [29]. These systems have controllability problems caused by the presence of zeros of the response functions with positive imaginary part [33] [88]. These zeros can be linked to delays of the response of the system to a perturbation [29]: the closer the imaginary part of a non-predictive zero is to zero, the longer is the delay of the response and then the more difficult is the system to control. This is extremely reasonable since a non-predictive zero in a response function Γ_i (of a dynamical variable x_i) with a very small imaginary part provides a substantial non-causal component of the surrogate response function H_{ji}, hence the variable x_i is a bad predictor of the other variables x_j. We can see that the presence of controllability problems (maybe due to a delay in the response) can be linked with the presence of unpredictability issues of the dynamical variables. As a consequence, if we can employ our UC to say whether some dynamical variable x_j is not predictive, we can infer that the analyzed system is necessarily a non-minimum one. This method could be of relevance for the design of controllers since those systems can have controllability issues derived from using non-minimum systems. Some industrial case studies can be found in [33] and [89]: a SISO industrial evaporator, a SISO isothermal CSTR, a MIMO non-isothermal CSTR and FCC reactors units. We envision applications of our work in more contexts where linearized systems are important. We have in mind species interaction networks, where it is difficult to know the precise state of all species each time. Lastly, a very significant direction would be to move out of linearized systems. Other paradigmatic non-linear models could be studied, for example models where a limit cycle is present in the steady state.

In Chapter 4 we have applied the Surrogate RT to a more complex model than the

ones considered in Chapter 3: the L96 model, which is a chaotic and spatially extended deterministic model [41–43], with travelling waves within it. It has been derived in climate studies, where it is used to describe the evolution in time of a meteorological quantity along a circle along the globe at constant latitude. The L96 system contains all the basic needed processes: advection of the meteorological quantity, dissipation and forcing. Since it is a spatially extended system, we have chosen to perturb it locally in a given location x_k and to observe the response of the system through the lens of the dynamical variables x_i of the system. The problem we studied amounts to understand when a local dynamical variable is able to predict another local dynamical variable. We have shown that the local perturbation in x_k introduces a hierarchy of the dynamical variable x_i in terms of their predictive power, which is strictly related to the delay in their response to the perturbation (the bigger the delay and the larger is the non-causal component of the surrogate response function). The higher is the rank of a variable in this hierarchy and the more predictive it is, starting from the directly perturbed variable x_k. Moreover, we have divided the surrogate response functions into anticipated surrogate response functions and retarded surrogate response functions. If the predictor feels the perturbation with a smaller or equal delay than the predictand, the related surrogated response function is retarded, while on the contrary it is anticipated. We have restricted ourselves just to the retarded surrogate response function, while we have discarded the anticipated surrogate response functions since they violate the causality principle and hence they are non-physical. As a consequence, variables of a given rank can predict the behaviour of other dynamical variables of the same rank or below. We have numerically derived some retarded surrogate response functions related to a few dynamical variables around the directly perturbed x_k, showing that the non-causal component increases as we move from x_k, together with the quality of the predictions, while the non-causal component is practically vanishing when x_k is the predictor. Furthermore, we have observed that for two variables which have the same rank in the hierarchy, the variable at the left of the perturbation can predict the behaviour of the variable at the right better than the reverse. This can be observed looking at the non-causal component of the related surrogate response functions, and it can be quantified using the ratio method devised in Section 2.6. This observation is consistent with the fact that in the system there is a group velocity v_g which goes from the left of the perturbation to its right. We can see that the surrogate RT and the ratio method in particular can unveil the presence of information flows within the system. Lastly, we have shown that it is possible to drastically improve the predictions of some non-predictive local dynamical variables using more than one forcing. This is reasonable: using more information from the system we can make better predictions about it. The situation we have in mind is that of a local observer which can have access to local and global information provided by a few forcings in order to predict the behaviour of a few observables of interest in other parts of the globe.

Our study of the L96 system employing the surrogate RT, with local perturbations and local observables, is meant to be a benchmark for a general methodology to study spatially extended systems:

- We locally perturb the system in a given location or a few selected close locations.
- We study how the perturbation propagates in the system for small times looking at the response functions of the local observables, showing that different local

observables react to the perturbations with different time scales.

- We build a hierarchy of these local variables in terms of their predictive power, i.e. the delay of their response to the perturbation. In that way, we can understand which are the most useful local observables to use for predictions of other local observables.
- We divide the surrogate response functions between these observables into anticipated and retarded, focusing just on the retarded ones. This means that a given local observable can make predictions just of other local observables with a lower or equal rank in the hierarchy.
- We numerically derive the surrogate response functions and we test their predictive ability perturbing the system with a perturbation with the same spatial pattern but different time pattern.
- We use the ratio method to quantify how much a surrogate response function is unpredictive, measuring the relative importance of its non-causal component. Using these numbers, we try to see if there is some systematic behaviour, with the purpose of unveiling potential causal links or flows of information between local variables.

We can see that the application of this rather general method can provide useful insights into the emergent properties of the system at hand. Furthermore, it provides a solid and flexible theory which points out which variables are better suited to be predictors and predictand, allowing predictions all over the system.

Appendix A

Kolgomorov's Criterion

In this appendix we show in the context of the Markov Chains a characterization of equilibrium steady state [44, 50, 90]: the Kolgomorov's Criterion. In particular, the criterion states that an equilibrium steady state, characterized by the DBE (1.30), is given if and only if the probability of a path in the phase space is equal to the reversed path. We first have to define some properties in order to introduce the theorem, considering a discrete Markov chain with transition matrix W, with state space $S = \{s_i\}_{i=1,\dots,N}$ and steady-state distribution p_{s_i} [44, 50, 90].

First, we say that a state s_i is accessible by another state s_j if there is a path of finite length which connects s_i and s_j, i.e. it exists a sequence of states $\{\alpha_k\}_{k=1,\dots,M}$, with M finite, such that its probability:

$$W_{s_i,\alpha_1}....W_{\alpha_M,s_j} > 0. \tag{A.1}$$

Then, we say that a Markov chain is irreducible when all the states are accessible by another state. Second, we define a state s_i as transient if, starting from s_i, there is a non-vanishing probability that the chain will never come back to s_i. If this probability is vanishing, the state s_i is said to be recurrent. Then, we further characterize the recurrent states s_i, by means of the mean hitting time M_i for s_i:

$$M_i = \mathbb{E}[T_i] = \sum_{n=1}^{\infty} n \cdot p^{(n)}_{s_i,s_i}, \tag{A.2}$$

where T_i is the time required from a generic path which starts from s_i to come back in s_i and $p^{(n)}_{s_i,s_i}$ is the probability to have such a path of length $T_i = n$. A recurrent state s_i is said to be positive if M_i is finite and null otherwise. A Markov chain is called positive (null) recurrent when all its states are positive (null) recurrent. Lastly, we define the period k of a state s_i as the greatest common divisor of the lengths of the paths that start from s_i and end in s_i:

$$k = \gcd\{n > 0 : p^{(n)}_{s_i,s_i} > 0\}. \tag{A.3}$$

If the period of a state s_i is one, the state is called aperiodic. In an irreducible Markov chain, if a state is aperiodic, all the states are aperiodic and the Markov chain is called aperiodic [50].

We now have all the tools to introduce the Kolgomorov's criterion:

Kolgomorov's criterion: An irreducible, positive recurrent, aperiodic Markov chain with transition matrix W has a steady state which satisfies the DBE (1.30) if and only if W satisfies:

$$W_{s_1,s_2}....,W_{s_n,s_1} = W_{s_1,s_n}....,W_{s_2,s_1}, \tag{A.4}$$

for all finite sequence of states $s_1,...,s_n \in S$.

In particular, notice that (A.4) means that a loop in the phase space can be travelled with equal probability in a verse or in the opposite. This implies the condition (1.32), hence that the probability to walk a path in the forward direction $\{s_i\}_{i=1,...,n}$ and in the reverse direction $\{s_i\}_{i=n,...,1}$ is equal. We can prove this last statement using the DBE (to see this it is sufficient to multiply (A.4) by p_{s_1} and then use the DBE). The Criterion can be extended in continuous-time Markov Chain, using the same ideas of the discrete case proof presented below.

The proof of the Criterion is the following. We start from proving the DBE (1.30) imply (A.4). We can rewrite the LHS of (A.4) as follows, using the DBE for the first factor W_{s_1,s_2}:

$$W_{s_1,s_2}....,W_{s_n,s_1} = (\frac{p_{s_2}}{p_{s_1}}W_{s_2,s_1})....,W_{s_n,s_1} \tag{A.5}$$

We can do the same for each of the factors, obtaining in the end the RHS of (A.4), hence proving this verse of the equivalence. which gives the time inversal condition (1.29).

The other verse of the proof, i.e. that the relation (A.4) implies the DBE (1.30), is the following. We fix two states s and t. Then, we have, thanks to (A.4):

$$W_{s,s_1}...W_{s_n,t} = \frac{W_{st}}{W_{ts}}W_{t,s_n}W_{i_1,s}, \tag{A.6}$$

for a choice of states $s_1,...,s_n$. We sum both sides of (A.6) for all possible choices of $s_1,...,s_n$, obtaining:

$$p^{(n)}_{s,t} = \frac{W_{st}}{W_{ts}}p^{(n)}_{t,s}. \tag{A.7}$$

Now we use the fact that, since the chain is positive recurrent, aperiodic and irreducible, we have that [50]:

$$\lim_{n\to\infty} p^{(n)}_{s_i,s_j} = p_{s_j}, \tag{A.8}$$

hence the steady-state probability that the chain is in s_j can be computed from the probability to start in a generic state s_i, ending in s_j after an infinite number of moves. Given that, we take $n \to \infty$ in (A.7), obtaining the DBE and proving the Criterion:

$$\frac{p_t}{p_s} = \frac{W_{st}}{W_{ts}}. \tag{A.9}$$

Appendix B

The pedagogical case of 3x3 matrices for the UC

We provide an explicit derivation of the UC for the instructive case of systems with 3 degrees of freedom. with a perturbation with spatial pattern $G_3 \neq 0$. Let's find the matrix with entries $\Gamma_{jk} = G_k(A - i\omega\mathbb{1})^{-1}_{jk}$ which describes the response functions:

$$\frac{1}{\det[A - i\omega\mathbb{1}]} \begin{pmatrix} (-A_{23}A_{32} + (A_{22} - i\omega)(A_{33} - i\omega))G_1 & (A_{13}A_{32} - A_{12}(A_{33} - i\omega))G_2 & (A_{12}A_{23} - A_{13}(A_{22} - i\omega))G_3 \\ (A_{23}A_{31} - A_{21}(A_{33} - i\omega))G_1 & (-A_{13}A_{31} + (A_{11} - i\omega)(A_{33} - i\omega))G_2 & (A_{13}A_{21} - (A_{11} - i\omega)A_{23})G_3 \\ (A_{21}A_{32} - A_{31}(A_{22} - i\omega))G_1 & (A_{12}A_{31} - A_{32}(A_{11} - i\omega))G_2 & (-A_{12}A_{21} + (A_{11} - i\omega)(A_{22} - i\omega))G_3 \end{pmatrix} \tag{B.1}$$

Let's look at the global response function of x_3: $\Gamma_{3,\mathrm{gl}} = \Gamma_{31} + \Gamma_{32} + \Gamma_{33}$. Recalling that $G_3 \neq 0$ and we look for the zeros of this global response function:

$$-G_3\omega^2 - i\omega(-A_{31}G_1 - A_{32}G_2 + G_3A_{11} + G_3A_{22}) - f[A] = 0 \tag{B.2}$$

where $f[A]$ is just a constant which is a function of the matrix A. Now we look for the roots of this equation, finding:

$$\omega_{1,2} = -\frac{i}{2}\left[(-A_{31}\frac{G_1}{G_3} - A_{32}\frac{G_2}{G_3} + A_{11} + A_{22}) \mp \sqrt{\left(-A_{31}\frac{G_1}{G_3} - A_{32}\frac{G_2}{G_3} + A_{11} + A_{22}\right)^2 + 4\,f[A]}\right]$$

We can see that if the linear coefficient of (B.2) is negative, the global response function $\Gamma_{3,\mathrm{gl}}$ has a zero in the upper complex $\omega-$plane, hence the variable x_j cannot be used for prediction purposes. The linear coefficient is negative if the following inequality holds:

$$A_{31}G_1 + A_{32}G_2 > (A_{11} + A_{22})G_3, \tag{B.3}$$

confirming in this case the UC (3.17).

Bibliography

[1] Kubo R.: *Statistical-mechanical theory of irreversible processes. i. General theory and simple applications to magnetic and conduction problems.* J. Phys. Soc. Japan 12 570–86 (1957).

[2] Kubo R.: *The fluctuation–dissipation theorem.* Rep. Prog. Phys. 29, 255–284 (1966).

[3] Van Kampen N.: *The case against linear response theory.* Phys. Nor. 5 279 (1971).

[4] Marini Bettolo Marconi U., Puglisi A., Rondoni L. and Vulpiani A.: *Fluctuation-Dissipation: Response Theory in Statistical Physics.* Phys. Rep. 461 111 (2008).

[5] Lucarini V., Saarinen J.J., Peiponen K.-E. and Vartiainen E.M.: *Kramers-Kronig Relations in Optical Materials Research.* Springer (2006).

[6] Lucarini V, Sarno S.: *A statistical mechanical approach for the computation of the climatic response to general forcings*, Nonlin. Processes Geophys., 18, 7-28, (2011).

[7] Ruelle D.: *Chaotic evolution and strange attractors: The statistical analysis of time series for deterministic nonlinear systems*, ambridge University Press, Cambridge, UK (1989).

[8] Ruelle D.: *Differentiation of SRB states.* Commun. Math. Phys. 187 227–41 (1997).

[9] Ruelle D.: *Nonequilibrium statistical mechanics near equilibrium: computing higher-order terms.* Nonlinearity 11 5–18 (1998).

[10] Ruelle D.: *A review of linear response theory for general differentiable dynamical systems.* Nonlinearity 22 855–70 (2009).

[11] Gallavotti G.: *Chaotic hypothesis: Onsager reciprocity and fluctuation-dissipation theorem.* Stat. Phys. 84 899–925 (1996).

[12] Gallavotti G. and Cohen E. G. D.: *Dynamical ensembles in stationary states.*J. Stat. Phys. 80 931–70 (1995).

[13] Lucarini V.: *Response theory for equilibrium and non-equilibrium statistical mechanics: causality and generalized Kramers-Kronig relations.* Stat. Phys. 131 543–58 (2008)

[14] Lucarini V.: *Evidence of dispersion relations for the nonlinear response of the Lorenz 63 system.* J. Stat. Phys. 134 381–400 (2009).

[15] Abramov R. V. and Majda A. J.: *Blended response algorithms for linear fluctuation-dissipation for complex nonlinear dynamical systems.* Nonlinearity 20 2793–821 (2007).

[16] Abramov R. V. and Majda A. J.: *New approximations and tests of linear fluctuation–response for chaotic nonlinear forced-dissipative dynamical systems*, J. Nonlinear Sci. 18 303–41 (2008).

[17] Baiesi M., Maes C.: *An update on the nonequilibrium linear response.* New J. Phys. 15, 013004 (2013).

[18] Seifert U. and Speck T.: *Fluctuation-dissipation theorem in nonequilibrium steady states.* EPL (Europhysics Letters), Volume 89, Number 1 (2010).

[19] Hänggi P. and Thomas H.: *Linear response and fluctuation theorems for nonstationary stochastic processes.* Z. Phys. B 22 295–300 (1975).

[20] Colangeli M,, Rondoni L. and Vulpiani F.: *luctuation–dissipation relation for chaotic non-Hamiltonian systems.* J. Stat. Mech. L04002 (2012).

[21] Zwanzig R., *Memory Effects in Irreversible Thermodynamics*, Phys. Rev. 124, 983-992 (1961)

[22] Mori H., *Transport, Collective Motion, and Brownian Motion*, Progress of Theoretical Physics 3, 423-455, 1965

[23] Wouters J., Lucarini, V.: *Multi-level dynamical systems: connecting the ruelle response theory and the Mori-Zwanzig Approach.* J. Stat. Phys. 151, 850–860 (2013).

[24] Chekroun M.D., Liu, H. and Wang S., *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs I*, Springer, 2015a.

[25] Chekroun M.D., Liu H. and Wang S., *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs II*, Springer, 2015b.

[26] Lucarini, V. and Colangeli M.: *Beyond the linear fluctuation-dissipation theorem: the role of causality.* J. Stat. Mech. P05013 (2012).

[27] Zubarev D., Gray P, and Shepherd P.J.: *Nonequilibrium Statistical Thermodynamics.* Springer US, Studies in Soviet Science: Physical Sciences (1974).

[28] Lucarini V.: *Revising and Extending the Linear Response Theory for Statistical Mechanical Systems: Evaluating Observables as Predictors and Predictants.* Journal of Statistical Physics (2018) 173: 1698-1721.

[29] Astrom K. J. and Murray R. M.: *Feedback systems: an introduction for scientists and engineers*, Princeton University Press, 2008.

[30] Terrell W. *Stability and Stabilization: An Introduction.* Princeton University Press., 2009.

[31] Franklin G. F., Powell J. D. and Emami-Naeini A..: *Feedback control of dynamic systems*, Addison-Wesley Publishing Company, 1993.

[32] Khalil H. K.: *Non-Linear systems*, Prentice Hall, 1995.

[33] Trickett , Bogle I.D.L., Lewin D.R.: *Methodology to Quantify Inverse Response of Nonlinear Systems*, IFAC Proceedings Volumes, Volume 27, Issue 7, Pages 141-146, (June 1994).

[34] Ghil M., Lucarini V.: *The Physics of Climate Variability and Climate Change*. Rev. Mod. Phys. 92, 35002 (2020).

[35] Peixoto, J. P. and Oort, A. H.: *Physics of Climate*, AIP Press, New York, (1992).

[36] Lucarini, V. et al.: *Mathematical and physical ideas for climate science*. Reviews of Geophysics 52(4), 809–859 (2014).

[37] Gritsun, A. and Lucarini, V.: *Fluctuations, response, and resonances in a simple atmospheric model*. Physica D: Nonlinear Phenomena 349, 62–76 (2017).

[38] Lucarini, V., Ragone, F. and Lunkeit, F.: *Predicting Climate Change Using Response Theory: Global Averages and Spatial Patterns*. Journal of Statistical Physics 166(3), 1036–1064 (2017).

[39] Ragone, F., Lucarini, V. and Lunkeit, F.: *A new framework for climate sensitivity and prediction: a modelling perspective*. Climate Dynamics 46(5–6), 1459–1471 (2016).

[40] Lembo, V., Lucarini, V. and Ragone, F. Beyond Forcing Scenarios: *Predicting Climate Change through Response Operators in a Coupled General Circulation Model*. Sci Rep 10, 8668 (2020).

[41] Lorenz, E.: *Predictability-A problem partly solved. Proc. Seminar on Predictability, Vol.1*, Reading, Berkshire, United Kingdom, ECMWF, 1-18, (1996).

[42] Lorenz, E., Emanuel, K.: *Optimal sites for supplementary weather observations*. J. Atmos. Sci. 55, 399-414, (1998).

[43] Lorenz, E.: *Designing Chaotic Models*. J. Atmos. Sci. 62, 1574-1587 (2004).

[44] Livi R. and Politi P.:*Nonequilibrium Statistical Physics: A Modern Perspective*. Cambridge University Press, 2017.

[45] Risken H-: *The Fokker–Planck Equation*. Berlin Springer 2nd edn (1989).

[46] Chaichian. M. and Demichev A.: *Path integrals in physics. Volume I: Stochastic processes and Quantum Mechanics*. CRC Press, 2001.

[47] De Cruz L., Schubert S., Demaeyer J., Lucarini V. and Vannitsem S.: *Exploring the Lyapunov instability properties of high-dimensional atmospheric and climate models*. Nonlinear Processes in Geophysics Discussions (2017).

[48] Oseledets V. I.: *A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems*, Transactions of the Moscow Mathematical Society, 19, 179–210 (1968).

[49] Ruelle D.: *Nonequilibrium statistical mechanics near equilibrium: computing higher order terms*. Nonlinearity 11, 5–18 (1998).

[50] Brémaud P.: *Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues*. Springer-Verlag New York Inc. (2001).

[51] Gardiner C. W.: *Handbook of Stochastic Methods.* Springer Series in Synergetics (2010).

[52] Avena L., Blondel O. and Faggionato A.: *Analysis of random walks in dynamic random environments via L^2-perturbations*, arXiv:1602.06322 (2016).

[53] Engel K.-J. and Nagel R.: *One-Parameter Semigroups for Linear Evolution Equations*, Springer-Verlag New York (2000).

[54] Lucarini V., Pavliotis G. A. and Zagli N.: *Response Theory and Phase Transitions for the Thermodynamic Limit of Interacting Identical Systems*, arXiv:2008.10912 (2020).

[55] Gutierrez M. S. and Lucarini V.: *Response and Sensitivity Using Markov Chains.* J. Stat. Phys. (2020).

[56] Evans D. and Rondoni L.: *Comments on the Entropy of Nonequilibrium Steady States*, J. Stat. Phys. 109, 3/4 (2002).

[57] Bonetto F., Kupiainen A. and Lebowitz J.L.: *Absolute continuity of projected SRB measures of coupled Arnold cat map lattices*, Ergod. Th. & Dynam. Sys. 25, 59 (2005).

[58] Boffetta G., Lacorata G., Musacchio S. and Vulpiani A.:*Relaxation of finite perturbations: beyond the fluctuation dissipation relation*, Chaos, 13, 3 (2003).

[59] E W, and Lu J.: *Multiscale modeling*, Scholarpedia 6 (10), 11527, revision #91540 (2011).

[60] Oono Y. and Paniconi M.: *Steady State Thermodynamics*, Progress of Theoretical Physics Supplement, Volume 130, (1998).

[61] Pavliotis G. A.: *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, Texts in applied mathematics 60 (2014).

[62] Prost J., Joanny J. F. and Parrondo J. M.: *Generalized fluctuation–dissipation theorem for steady-state systems,* Phys. Rev. Lett. 103 090601 (2009).

[63] Mézard M. and Montanari A.: *Information, physics and computation*, Oxford University press, (2009).

[64] Cencini M., Cecconi F., Vulpiani A.: *Chaos, from simple models to complex systems.* World Scientific, (Series on Advances in Statistical Mechanics), 2009.

[65] Choudhuri A. R.: *The Physics of Fluids and Plasmas: An Introduction for Astrophysicists*, Cambridge University Press (2008).

[66] Young L.-S.: *What are SRB measures and which dynamical systems have them?*, J. Stat. Phys. 108 733–54 (2002).

[67] Ruelle D.:*General linear response formula in statistical mechanics and the fluctuation–dissipation theorem far from equilibrium,* Phys. Lett. A 245 220–4 (1998).

[68] Reed M. and Simon B.: *Methods of Mathematical Physics II*, Academic Press, New York, Section X.8 (1975).

[69] Encyclopedic Dictionary of Mathematics (2nd English ed.) (MIT Press, Cambridge MA), Section 378B (1987).

[70] Reichl L. E.: *A Modern Course in Statistical Physics*, 2nd edn (New York: Wiley,1998).

[71] Lucarini V., Saarinen J.J., Peiponen K.-E. and Vartiaine E.M.: *Kramers-Kronig relations in Optical Materials Research*, Springer, Heidelberg (2005).

[72] Lucarini, V., Bassani, F., Saarinen, J.J., Peiponen, K.-E.: *Dispersion theory and sum rules in linear and nonlinear optics*. Rivista Nuovo Cimento 26, 1-120 (2003)

[73] Kravanja, P., Sakurai, T., Van Barel, M.: *On locating clusters of zeros in analytic functions*. BIT Num. Math. 39, 646–682 (1999).

[74] Arnold V. I.: *Ordinary Differential Equations*. The MIT Press, Cambridge, USA, 1978.

[75] Baker, G. A. Jr. and Graves-Morris P.: *Padé Approximants*. Cambridge U.P., 1996.

[76] Hazewinkel M.: *Viète theorem*, Encyclopedia of Mathematics, Springer Science+Business Media B.V. / Kluwer Academic Publishers, 1994.

[77] Routh, E. J.:*A Treatise on the Stability of a Given State of Motion: Particularly Steady Motion*. Macmillan, 1877.

[78] Hurwitz, A.: *Ueber die Bedingungen, unter welchen eine Gleichung nur Wurzeln mit negativen reellen Theilen besitzt*. Math. Ann. 46 (2): 273–284. doi:10.1007/BF01446812, 1895.

[79] Tao T. and Vu V.: *Random matrices: Universality of ESD and the Circular Law*, Annals of Probability. 38 (5): 2023–2065 (2010).

[80] May, R.: *Stability and complexity in model ecosystems*. Vol. 6. Princeton university press (2001).

[81] Allesina S. and Tang S.:*Stability criteria for complex ecosystems*. Nature 483, 205–208 (2012).

[82] Rothblum U.G. and Tan C.P.: *Upper bounds on the maximum modulus of subdominant eigen-values of non negative matrices*. Linear Algebra Appl. 66, 45–86 (1985).

[83] Naumov A.: *Elliptic law for real random matrices*, arXiv:1201.1639v5 [math.PR].

[84] Gallavotti, G., Lucarini, V.: *Equivalence of non-equilibrium ensembles and representation of friction in turbulent flows: the Lorenz 96 Model*. J. Stat. Phys. 156, 1027–1065 (2014).

[85] Lacorata G., Vulpiani A.: *Fluctuation-Response Relation and modeling in systems with fast and slow dynamics*, Nonlin. Processes Geophys., 14, 681–694, 2007.

[86] Gustavsen B.: *Rational approximation of frequency domain responses by vector fitting*. IEEE Transanctions on Power Delivery, Vol. 14, No. 3, July 1999.

[87] Berljafa M. and Güttel S.: *The RKFIT Algorithm for nonlinear rational approximation*. SIAM J. Sci. Comput., Vol. 39, No. 5, pp. A2049-A2071, 2017.

[88] Lewin, D. R., Faigon, M., Fuchs, A., and Semiat, R.: *Modelling and Control of Two-Phase Systems*, Proc. ESCAPE-1, Elsinore, Denmark, S149-S156, 1992.

[89] Yescas R. M., Bogle D. and Isunza F. L.: *Steady state and dynamic simulation of FCC reactor units*, Dynamics and Control of Chemical Reactors (DYCORD'95), Copenhagen, Denmark, (1995).

[90] Kelly, Frank P. *Reversibility and Stochastic Networks*. Wiley, Chichester. (1979).