



HAL
open science

Ruelle-Pollicott Resonances of Stochastic Systems in Reduced State Space. Part I: Theory

Mickael D. D Chekroun, Alexis Tantet, Henk Anton Dijkstra, J. David D
Neelin

► **To cite this version:**

Mickael D. D Chekroun, Alexis Tantet, Henk Anton Dijkstra, J. David D Neelin. Ruelle-Pollicott Resonances of Stochastic Systems in Reduced State Space. Part I: Theory. 2019. hal-02396076v1

HAL Id: hal-02396076

<https://hal.science/hal-02396076v1>

Preprint submitted on 6 Dec 2019 (v1), last revised 20 Mar 2020 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Ruelle-Pollicott Resonances of Stochastic Systems in Reduced State Space. Part I: Theory

Mickaël D. Chekroun · Alexis Tantet · Henk A. Dijkstra · J. David Neelin

December 6, 2019

Abstract A theory of Ruelle-Pollicott (RP) resonances for stochastic systems is introduced. These resonances are defined as the eigenvalues of the generator (Kolmogorov operator) of a given stochastic system. By relying on the theory of Markov semigroups, decomposition formulas of correlation functions and power spectral densities (PSDs) in terms of RP resonances are then derived. These formulas describe, for a broad class of stochastic differential equations (SDEs), how the RP resonances characterize the decay of correlations as well as the signal's oscillatory components manifested by peaks in the PSD.

It is then shown that a notion reduced RP resonances can be rigorously defined, as soon as the dynamics is partially observed within a reduced state space V . These reduced resonances are obtained from the spectral elements of reduced Markov operators acting on functions of the state space V , and can be estimated from series. When the sampling rate (in time) at which the observations are collected is either sufficiently small or large, it is shown that the reduced RP resonances approximate the RP resonances of the generator of the conditional expectation in V , i.e. the optimal reduced system in V obtained by averaging out the contribution of the unobserved variables. The approach is illustrated on a stochastic slow-fast system for which it is shown that the reduced RP resonances allow for a good reconstruction of the correlation functions and PSDs, even when the time-scale separation is weak.

The companions articles, Part II [[TCND19a](#)] and Part III [[TCND19b](#)], deal with further practical aspects of the theory presented in this contribution. One important byproduct consists of the diagnosis usefulness of stochastic dynamics that RP resonances offer. This is illustrated in the case of a stochastic Hopf bifurcation in Part II. There, it is shown that such a bifurcation has a clear signature in terms of the geometric organization of the RP resonances in the left half plane. This analysis provides thus an unambiguous signature of nonlinear oscillations contained in a noisy signal and that can be extracted from time series. By relying then on the theory of reduced RP resonances presented in this contribution, Part III addresses then the question of detection and characterization of such oscillations in a high-dimensional stochastic system, namely the Cane-Zebiak model of El Niño-Southern Oscillation subject to noise modeling fast atmospheric fluctuations.

M.D. Chekroun

Department of Earth and Planetary Sciences, Weizmann Institute, Rehovot 76100, Israel; Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA 90095-1565, USA
E-mail: mchekroun@atmos.ucla.edu

A. Tantet

École Polytechnique, Laboratoire de Météorologie Dynamique, Palaiseau, France

H.A. Dijkstra

Institute of Marine and Atmospheric Research, Department of Physics and Astronomy, University of Utrecht, Utrecht, The Netherlands

J.D. Neelin

Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, CA 90095-1565, USA

Keywords Ruelle-Pollicott resonances; Conditional expectation; Correlation functions; Kolmogorov operator; Markov semigroups.

Contents

1	Introduction	2
2	Ruelle-Pollicott resonances and decay of correlations from stochastic differential equations	3
3	Reduced Ruelle-Pollicott resonances	9
4	Applications to a stochastic slow-fast system	16
A	Elements of stochastic analysis	19

1 Introduction

The determination of reduced equations that aims at mimicking in a reduced state space the dynamics issued from a system of *stochastic differential equations (SDEs)* has become a central question in several disciplines ranging from molecular dynamics [SFHD99, SH99, SHD01] to epidemic models [FS09, FBS09] and climate dynamics over the past decades; see e.g. [MTVE01, CLW15b, GCF17] and references therein. Various approaches have been proposed, and in most of the cases, different metrics associated with moment estimates, probability density functions or time marginals are used to compare the reduced dynamics obtained from the surrogate system to that of the original system of stochastic differential equations (SDEs).

Here, we adopt a completely different approach which consists of using correlation functions and power spectral densities as “metrics” of comparison. In that respect, we frame hereafter a theory of Ruelle-Pollicott (RP) resonances for stochastic systems. Initially introduced for discrete and continuous chaotic deterministic systems [Rue86, Pol86], these resonances are known to provide rigorous decomposition formulas of correlation functions; see e.g. [Bal00, BL07]. However, the derivation of such formulas in the deterministic context is made difficult by the singular character of the underlying invariant measures. In presence of noise, smoothing effects are known to occur for a broad class of SDEs (see Appendix A.2), and the derivation of such formulas is thus facilitated. Section 2 presents an extension of such decomposition formulas to the stochastic context.

The RP resonances are then defined naturally as the eigenvalues of the generator (Kolmogorov operator) of a given stochastic system. As shown in Sec. 2, these resonances allow for decomposition formulas of correlation functions and PSDs in terms of these resonances; see Corollary 1 and formula (2.23) below. The resulting formulas describe how the RP resonances characterize the decay of correlations as well as the signal’s oscillatory components manifested by peaks in the PSD. The obtention of these formulas relies on tools and concepts from stochastic analysis, based on the theory of Markov semigroups on one hand, and the spectral theory of semigroups, on the other. These tools are briefly surveyed in Appendix A, having in mind a wider audience in the geosciences and macroscopic physics.

From a practical viewpoint, our decomposition formulas benefit furthermore from a natural dynamical interpretation as they relate to the spectral elements of the Kolmogorov operator and thus to the SDE’s coefficients themselves. Such dynamical interpretations (and thus decompositions) are potentially useful for identifying physical processes responsible for power excess bumps or other broad band peaks in the PSD of noisy observations, a topic of active research in various areas of physics such as asteroseismology [KMH⁺10, GKWG09], supersonic flows [Tam95], or climate dynamics [GAD⁺02].

In Section 3, we present the main contribution of this article, by inquiring whether one can extract useful resonances from partial observations in a reduced state space, that still relate to the correlation functions and PSDs. First, we show that a notion reduced RP resonances can be indeed rigorously framed, as soon as the (stochastic) dynamics is partially observed within a reduced state space V . These reduced resonances are obtained from the spectral elements of reduced Markov operators acting on functions of the state space V , and can be estimated from series; see Sec. 3.3. The reduced Markov operators are ensured by Theorem 2 that extends [CNK⁺14, Theorem A] within the stochastic context of this article. *Mutatis mutandis*, Theorem 2 as [CNK⁺14, Theorem A], shows that — given a reduced state space V , a mapping h from the full state \mathcal{X} space onto V , and a Markov semigroup P_t possessing an invariant measure on \mathcal{X} — a family of reduced Markov operators (acting on functions of V) can be naturally associated with P_t , V and h . This family characterizes a coarse-graining in the reduced state space V — such as induced by

the map h — of the actual transition probabilities associated with P_t and taking place within the full state space \mathcal{X} . As a result, Theorem 2 provides, for instance, an extension to the stochastic context of Markov representations used in [SFHD99] for the description of coarse-grained dynamical features of Hamiltonian systems arising in molecular dynamics.

In a second step, we show, under ergodicity (or mixing) assumptions satisfied for a broad class of SDEs, that the reduced RP resonances approximate the RP resonances of the generator of the conditional expectation in V , i.e. the optimal reduced system in V obtained by averaging out the contribution of the unobserved variables; see Theorem 3. These resonances can be estimated from time series; see Sec. 3.3. Theorem 3 shows then that the relationship between the reduced RP resonances and the conditional expectation is ensured when the sampling rate (in time) at which the observations are collected, is either sufficiently small or large.¹ The fact that the reduced RP resonances relate to the conditional expectation offers useful insights for the investigation of whether non-Markovian effects *à la* Mori-Zwanzig should be included or not for the derivation of efficient reduced systems [WL12, WL13].

In that perspective, we analyze in Sec. 4 correlation functions and PSDs as reconstructed from reduced RP resonances of a stochastic slow-fast system, by using the formulas of Sec. 2 in which the reduced RP resonances replace the genuine eponymous resonances. In the case of a strong time-scale separation, without any surprise these reduced RP resonances, as associated with the conditional expectation — well-approximated by a slow manifold reduction for the example of Sec. 4 —, provide excellent reproduction of correlation functions and PSDs obtained by standard sample estimates. The surprise arises when the system is placed in a regime without time-scale separation. In this case, the reduced RP resonances allow still to reconstruct to a very good accuracy level the correlation functions and PSDs providing thus an unambiguous diagnosis that the conditional expectation is sufficient to close the system and that the inclusion of non-Markovian effects is here useless. Such a diagnosis regarding the conditional expectation drawn from RP resonance analysis offers thus promising perspectives in terms of reduced order modeling, and may add insights to reduction approaches exploiting conditional expectations such as in e.g. [WVE04, LL10, LVE14].

The usefulness of reduced RP resonance analysis is not limited to providing useful diagnosis for model reduction. The companions articles, Part II [TCND19a] and Part III [TCND19b], deal with further practical aspects of the theory presented in this article. One important byproduct consists of the diagnosis usefulness of stochastic dynamics that RP resonances themselves, offer. This is illustrated in the case of a stochastic Hopf bifurcation in Part II. There, it is shown that such a bifurcation has a clear signature in terms of the geometric organization of the RP resonances in the left half plane. This analysis provides thus an unambiguous signature of nonlinear oscillations contained in a noisy signal and that can be extracted from time series. By relying then on the theory of reduced RP resonances presented in this contribution, Part III addresses then the question of detection and characterization of such oscillations in a high-dimensional stochastic system, namely a system obtained from a semispectral approximation of the Cane-Zebiak model of El Niño-Southern Oscillation, with the addition of noise modeling fast atmospheric fluctuations.

2 Ruelle-Pollicott resonances and decay of correlations from stochastic differential equations

The authors in [CNK⁺14] have introduced a new mathematical framework to (i) understand and diagnose — through partial observations — the variability of chaotic flows, and (ii) to analyze parameter sensitivity that may occur in the modeling of such observations. The framework relied on the theory of Ruelle-Pollicott (RP) resonances introduced in the mid-80's [Rue86, Pol86] and that, at the time of the publication of [CNK⁺14], was known only by a little group of experts working in the field dynamical system theory and the mathematical study of scattering resonances [Zwo17]. Initially introduced for discrete and continuous chaotic deterministic systems (see also [BL07] for the case of Anosov flows), the RP resonances extend to stochastic systems. This section presents such an extension by relying on tools from stochastic analysis, based on the theory of Markov semigroups on one hand, and the spectral theory of semigroups, on the other; see also [Gas02, DZ15] for complementary approaches.

As we will see, the RP resonances characterize important features of the solution's variability (e.g. its oscillatory behavior), such as typically reflected in power spectra or correlation functions, in terms of the

¹ In practice however it is often observed that the reduced RP resonances still provide useful information for “intermediate” sampling rates; see Part III [TCND19b].

spectrum of the underlying Liouville operator for deterministic systems or the Fokker-Planck operator for stochastic systems, but are in general difficult to estimate especially if the dimension of the state space is large. Section 3 below addresses the implication of estimating resonances from partial observations in a reduced state space, and how these reduced resonances relate to the full RP resonances.

2.1 Generalities

We consider Stochastic Differential Equations (SDEs) of the form:

$$dX = F(X) dt + D(X) dW_t, \quad X \in \mathbb{R}^p. \quad (2.1)$$

Here $W_t = (W_t^1, \dots, W_t^q)$ denotes an \mathbb{R}^q -valued Wiener process (q not necessarily equal to p) whose components are mutually independent Brownian motions.

In Eq. (2.1), the drift part is provided by a (possibly nonlinear) vector field F of \mathbb{R}^p , and the (also possibly nonlinear) stochastic diffusion in its Itô version, given by $D(X) dW_t$, has its i^{th} -component ($1 \leq i \leq p$) given by

$$\left[D(X) dW_t \right]_i = \sum_{j=1}^q D_{ij}(X) dW_t^j, \quad q \geq 1. \quad (2.2)$$

Throughout this article, we assume that the vector field F and the matrix-valued function

$$D : \mathbb{R}^p \rightarrow \text{Mat}_{\mathbb{R}}(p \times q),$$

satisfy regularity conditions that guarantee the existence and the uniqueness of mild solutions, as well as the continuity of the trajectories; see, e.g. [Cer01,FGP10] for such conditions in the case of locally Lipschitz coefficients. Sometimes the resulting stochastic process solving Eq. (2.1) emanating from x at $t = 0$, will be denoted by X_t^x , when dependence on the initial datum would be stressed, otherwise simply X_t .

It is well-known that the evolution of the probability density of the stochastic process, X_t , solving Eq. (2.1), is governed by the *Fokker-Planck equation*

$$\partial_t \rho(X, t) = \mathcal{A} \rho(X, t) = -\text{div}(\rho(X, t) F(X)) + \frac{1}{2} \text{div} \nabla (\Sigma(X) \rho(X, t)), \quad X \in \mathbb{R}^p, \quad (2.3)$$

with $\Sigma(X) = D(X) D(X)^T$ denoting the diffusion tensor. In practice, one is interested in stationary solutions μ to Eq. (2.3) that provide a statistical description of the asymptotic behavior of the solutions to Eq. (2.1), and thus of probability density functions (PDFs), typically obtained as marginal distributions of μ .

What is less-known however is that the spectral properties of the 2nd-order differential operator, \mathcal{A} , informs about fundamental objects such as the power spectra or correlation functions computed typically along a stochastic path of Eq. (2.1). To understand these relationships, we derive in the next subsection, decomposition formulas of correlations and power spectra in terms of the spectral elements of an operator closely related to \mathcal{A} , namely the generator K of the associated Markov semigroup.

2.2 Ruelle-Pollicott (RP) resonances and the spectral decomposition of correlation functions

As recalled in Appendix A.3 (Theorem 4), the existence of an invariant measure μ ensures that the Markov semigroup P_t (see Appendix A.1) associated with Eq. (2.1), is a strongly continuous semigroup in $L_{\mu}^p(\mathcal{X})$, for any $p \geq 1$. This property alone allows us to get useful decomposition formulas of correlation functions $C_{f,g}(t)$ given by

$$C_{f,g}(t) = \int f \cdot P_t g d\mu - \int f d\mu \int g d\mu. \quad (2.4)$$

This is the content of Corollary 1 below. It makes use of elements of the spectral theory of strongly continuous semigroups that we recall below and apply to Markov semigroups.

In that respect, recall that the *essential growth bound* of a semigroup $\mathcal{T} = (T(t))_{t \geq 0}$ is defined as

$$\omega_{ess}(\mathcal{T}) = \inf_{t > 0} \frac{1}{t} \log \|T(t)\|_{ess}, \quad (2.5)$$

where

$$\|T(t)\|_{ess} = \inf \left\{ \|T(t) - \mathcal{C}\|_{\mathcal{L}(\mathcal{E})} : \mathcal{C} \text{ is a linear and compact operator of } \mathcal{E} = L_{\mu}^p(\mathcal{H}) \right\}. \quad (2.6)$$

In other words, $\|T(t)\|_{ess}$ measures the distance of $T(t)$ to the set of linear and compact operators of $L_{\mu}^p(\mathcal{X})$. In the case this distance approaches zero as $t \rightarrow \infty$, the strongly continuous semigroup \mathcal{T} is called *quasi-compact*. The theory of quasi-compact strongly continuous semigroups shows that the latter property is equivalent to $-\infty < \omega_{ess}(\mathcal{T}) < 0$; see [EN00, Prop. V.3.5].

A semigroup is *eventually compact* if there exists $t_0 > 0$ such that $T(t_0)$ is compact. A semigroup that is eventually compact satisfies thus $\omega_{ess}(\mathcal{T}) = -\infty$. At the same time, the definition of (2.5) allows for semigroups that are not eventually compact while their essential growth bound is negative infinity².

In all cases, we have the following decomposition theorem of strongly continuous semigroups [EN00, Thm. V.3.1] that we apply to the case of Markov semigroups possessing an invariant measure.

Theorem 1 *Let μ be an invariant measure of a Markov semigroup $\mathcal{P} = (P_t)_{t \geq 0}$ and let K be the corresponding generator in L_{μ}^2 . Let $\lambda_1, \dots, \lambda_N$ be in $\sigma(K)$ satisfying $\operatorname{Re} \lambda_1, \dots, \operatorname{Re} \lambda_N > \omega_{ess}(\mathcal{P})$. Then the following properties hold:*

- (i) *The λ_j 's are isolated spectral values of K with finite algebraic multiplicity.*
- (ii) *If Π_1, \dots, Π_N denote the corresponding spectral projections and m_1, \dots, m_N the corresponding orders of poles of the resolvent of K , then*

$$P_t = \sum_{j=1}^N T_j(t) + R_N(t) \quad (2.7)$$

where

$$T_j(t) = \left[\sum_{k=0}^{m_j-1} \frac{t^k}{k!} (K - \lambda_j Id)^k \right] e^{\lambda_j t} \Pi_j, \quad (2.8)$$

and where for every $\epsilon > 0$, there exists $M > 0$ such that

$$\|R_N(t)\|_{\mathcal{L}(L_{\mu}^2(\mathcal{X}))} \leq M e^{(\omega_N^* + \epsilon)t}, \quad \forall t \geq 0, \quad (2.9)$$

with

$$\omega_N^* = \sup\{\omega_{ess}(\mathcal{P})\} \cup \{\operatorname{Re} \lambda : \lambda \in \sigma(K) \setminus \{\lambda_1, \dots, \lambda_N\}\}. \quad (2.10)$$

In what follows we denote by $\langle \cdot, \cdot \rangle_{\mu}$ the inner product in L_{μ}^2 . We have then the following corollary regarding the decomposition of correlation functions.

Corollary 1 *Let μ be an invariant measure of a Markov semigroup $\mathcal{P} = (P_t)_{t \geq 0}$ and let K be the corresponding generator in L_{μ}^2 . Let f and g in L_{μ}^2 such that $\langle f \rangle_{\mu} = \langle g \rangle_{\mu} = 0$, then given the λ_j s and the Π_j s as defined in Theorem 1, the correlation function $C_{f,g}(t)$ associated with the observables f and g has the following expansion*

$$C_{f,g}(t) = \sum_{j=1}^N \left[\sum_{k=0}^{m_j-1} \frac{t^k}{k!} (K - \lambda_j Id)^k \right] e^{\lambda_j t} \int f(x) (\Pi_j g)(x) d\mu(x) + \mathcal{Q}_N(t), \quad (2.11)$$

with

$$|\mathcal{Q}_N(t)| \leq M \|f\|_{L_{\mu}^2} \|g\|_{L_{\mu}^2} e^{(\omega_N^* + \epsilon)t}, \quad \forall t \geq 0. \quad (2.12)$$

² For instance any semigroup \mathcal{T} such that $\|T(t)\|_{ess} \leq M \exp(-ct^{\alpha})$, with $\epsilon > 0$ and $0 < \alpha, M < 1$.

In the particular case of a discrete spectrum, $\sigma(K) = \{\lambda_1, \lambda_2, \dots\}$, with $\operatorname{Re} \lambda_{j+1} \leq \operatorname{Re} \lambda_j$ for all $j \in \mathbb{N}^*$, and $\operatorname{Re} \lambda_j \rightarrow -\infty$, if $m_j = 1$ for all $j \geq 1$, then the expansion (2.11) takes the form:

$$C_{f,g}(t) = \sum_{j=1}^{\infty} e^{\lambda_j t} \langle f, \psi_j \rangle_{\mu} \langle \psi_j^*, g \rangle_{\mu}, \quad (2.13)$$

where ψ_j denotes the L_{μ}^2 -eigenfunction associated with λ_j and ψ_j^* , the associated L_{μ}^2 -eigenfunction of the adjoint operator K^* of K .

Proof The decomposition (2.11) is a direct consequence of (2.7)-(2.9). The proof of (2.13) consists of noting for all $j \geq 1$

$$\begin{aligned} \int f(x)(\Pi_j g)(x) d\mu(x) &= \int f(x) \langle \psi_j^*, g \rangle_{\mu} \psi_j(x) d\mu(x), \\ &= \langle f, \psi_j \rangle_{\mu} \langle \psi_j^*, g \rangle_{\mu}, \end{aligned} \quad (2.14)$$

and that $\mathcal{Q}_N(t) \xrightarrow[N \rightarrow \infty]{} 0$ in virtue of our assumptions about $\sigma(K)$ and of (2.9) and (2.10).

The λ_j 's such that $\operatorname{Re} \lambda_j > \omega_{\text{ess}}(\mathcal{P})$, are called the **RP resonances**. In other words they correspond to the point spectrum of K . Note that there exist stochastic processes for which $R_N = 0$ (i.e. no essential spectrum), for instance Ornstein-Uhlenbeck processes [MPP02]. Remark 4-(ii) points out other (nonlinear) stochastic processes that do not have an essential spectrum. Panel (a) of Fig. 1 shows a schematic of a general case, where the RP resonances are contained within a strip of the complex plane, away from the imaginary axis and the essential spectrum of K . The rate of decay of correlations is controlled by the spectral gap, $\tau = \operatorname{gap}(K)$ defined in (A.27). Appendix A.5 reports on a broad class of Markov semigroups (and thus SDEs) whose generator possesses a spectral gap; see also [Fro97, Bal00, BL07, CNK⁺14] for results in the deterministic context.

Remark 1

- (i) **Decay of correlations.** Under the conditions of Corollary 1, one obtains that $C_{f,g}(t) \xrightarrow[t \rightarrow \infty]{} 0$, for any f, g in L_{μ}^2 such that $\int f d\mu = \int g d\mu = 0$, and without assuming P_t to be strong Feller and irreducible.
- (ii) **Transfer operator.** By working in the Hilbert space L_{μ}^2 allows us to define the transfer operator \mathfrak{L}_t acting on functions in L_{μ}^2 , by the duality formula:

$$\int f \cdot P_t g d\mu = \int \mathfrak{L}_t f \cdot g d\mu, \quad f, g \in L_{\mu}^2, \quad t \geq 0. \quad (2.15)$$

In a similar fashion that P_t is related to the Kolmogorov operator \mathcal{K} via its generator (see (iii) below), the transfer operator \mathfrak{L}_t is related to the Fokker-Planck operator \mathcal{A} defined in (2.3).

- (iii) **RP resonances and Kolmogorov operator.** There are relationships between the abstract operator K and the concrete Kolmogorov operator \mathcal{K} defined in (A.4). In that respect, a theorem of Dynkin ensures (e.g. [RW00, p. 258]) that if $(P_t)_{t \geq 0}$ is a Feller semigroup of a Markov process with continuous paths such that $C_0^{\infty}(\mathbb{R}^N)$ is contained in the domain $D(K)$ of the generator, then there exist functions a_{ij} , b_i , and c in $C(\mathbb{R}^N)$ ($i, j = 1, \dots, N$) such that for any x , the matrix $(a_{ij}(x))_{1 \leq i, j \leq N}$ is non-negative definite, $c(x) \leq 0$,³ and

$$Kf(x) = \sum_{i,j=1}^N a_{ij}(x) \partial_{ij}^2 f + \sum_{i=1}^N b_i(x) \partial_i f + c(x)f, \quad f \in C_0^{\infty}(\mathbb{R}^N). \quad (2.16)$$

Furthermore,

$$b_i(x) = F_i(x), \quad a_{ij}(x) = \frac{1}{2} (D(x)D(x)^T)_{ij}. \quad (2.17)$$

³ Furthermore if the process is non-explosive then $c \equiv 0$. This excludes the cases for which the underlying Markov process leaving at time 0 from x in \mathbb{R}^N escapes to infinity at some finite time $t > 0$. This article is not concerned with explosive stochastic processes.

For such reasons we will sometimes refer to the RP resonances as the RP resonances of the Kolmogorov operator K , or the Fokker-Planck operator \mathcal{A} (by duality). This language is often adopted in Part III, more oriented towards applications than the present Part I.

However, the spectral analysis of general 2nd-order operators, typically non-self adjoint such as given in (2.16), is not an easy task in practice, especially when N is large; see [Dav96, Dav07]. We refer nevertheless to [EH03, HN04] for a detailed (and enlightening) study of the “shape” of the spectrum (in L^2_μ for [HN04] and Sobolev spaces for [EH03]) of a broad class of hypoelliptic operators such as arising in the theory of Langevin dynamics; see also [OPPS12].

Remark 2

- (i) Note that the sum in (2.13) starts actually at $j = 2$ since $\psi_1 = \mathbb{1}_X$ (since $1 \in \sigma(P_t)$) and $\langle f \rangle_\mu = 0$, by assumption.
- (ii) When $\sigma(K) = \{\lambda_1, \lambda_2, \dots\}$ with $\operatorname{Re} \lambda_{j+1} \leq \operatorname{Re} \lambda_j$ for all $j \in \mathbb{N}^*$, and $\operatorname{Re} \lambda_j \rightarrow -\infty$, is only assumed (without requiring $m_j = 1$) then (2.11) holds with $N = \infty$ and $\mathcal{Q}_N = 0$. Such an hypothesis on the spectrum of K is systematically satisfied for the Markov semigroups that are eventually compact and possess an invariant measure μ ; see [EN00, Cor. V.3.2]. A large subclass of such Markov semigroups are the compact ones. It includes a broad class of SDEs such as the following gradient dynamics on \mathbb{R}^d subject to noise:

$$dx = -\nabla V(x) dt + \sqrt{2\beta^{-1}} dW_t, \quad \beta > 0, \quad (2.18)$$

where W_t is a d -dimensional Wiener process. The diffusion is in that case elliptic, so it is straightforward that the assumptions of Hörmander’s theorem are satisfied, and the corresponding transition probabilities are smooth, provided that V is smooth. In fact for a broad class of smooth potentials V that satisfy appropriate growth conditions, the compactness of P_t is ensured; e.g. [LB06, Thm. 8.5.3] and [GRS99, Lem. 1.2]. See also [MPW02] for conditions ensuring the compactness of Markov semigroups that are not necessarily associated with gradient dynamics.

- (iii) Note that if P_t is eventually compact and possesses a unique ergodic invariant measure μ , then there exists $\alpha > 0$ and $M \geq 1$ and

$$\|P_t \varphi - \langle \varphi \rangle_\mu\|_{L^2_\mu} \leq M e^{-\alpha t} \|\varphi - \langle \varphi \rangle_\mu\|_{L^2_\mu}. \quad (2.19)$$

Note that if P_t has a regularizing effect sending L^2_μ into the Sobolev space $W_\mu^{1,2}$ in finite time and $W_\mu^{1,2}$ is compactly embedded into L^2_μ then the Poincaré inequality holds [LB06, Prop. 8.6.1] and P_t is eventually compact. The proof of (2.19) is then a consequence of the theory of eventually compact semigroups [EN00, Cor. V.3.3] and the existence of a unique ergodic measure which ensures that $\mathbb{1}_X$ is the unique L^2_μ -eigenfunction of P_t (up to a multiplication factor) associated with the eigenvalue 1.

2.3 Decomposition of the power spectrum

Given an observable $h : \mathbb{R}^p \rightarrow \mathbb{R}$ for the system (2.1), we recall that the correlation spectrum $S_h(f)$ is obtained by taking the Fourier transform of the correlation function $C_h(t)$, namely

$$S_h(f) = \widehat{C_h}(f), \quad (2.20)$$

where C_h given by (2.4) (for a given invariant measure μ) with $f = g = h$, therein.

For a broad class of SDEs that possess an ergodic probability distribution μ , the spectrum in L^2_μ , $\sigma(K)$, of the Markov semigroup generator, is typically contained in the left-half complex plane, $\{z \in \mathbb{C} : \operatorname{Re}(z) \leq 0\}$ and its resolvent $R(z) = (z\operatorname{Id} - K)^{-1}$, is a well-defined linear operator that satisfies

$$S_h(f) = \int_{\mathbb{R}^p} h(X) [R(if)h](X) d\mu. \quad (2.21)$$

In (2.21), the frequency f lies in the complex plane \mathbb{C} , and the poles of the resolvent $R(if)$ —which correspond to the RP resonances —introduce singularities into $S_h(f)$. Once the power spectral density (PSD) is calculated, i.e. once $|S_h(f)|$ is computed with f taken to be real, these poles manifest themselves

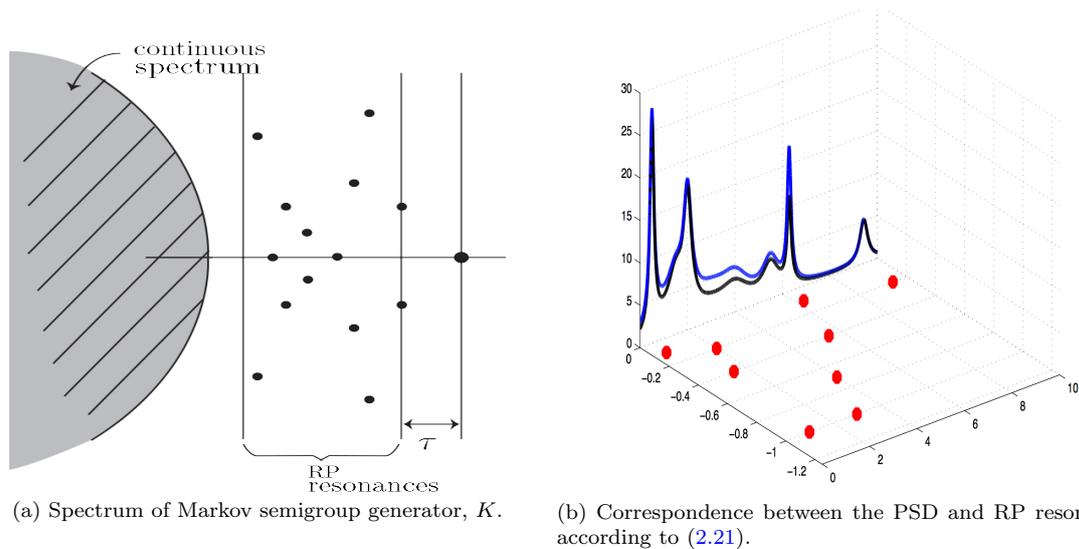


Fig. 1: **Panel (a):** The Ruelle-Pollicott (RP) resonances are the isolated eigenvalues of the Markov semigroup generator, K ; they are represented by red dots in Panel (b) and by black dots here. The rightmost vertical line represents the imaginary axis above which the power spectrum lies; see Panel (a) for another perspective. The rate of decay of correlations is controlled by the spectral gap τ ; see Appendix A.5. **Panel (b):** [Courtesy of Maciej Zworski] The imaginary part of the RP resonances corresponds to the location of a peak in the PSD (black curve lying above the imaginary axis) and the real part to its width. In blue is represented a reconstruction of the PSD based on RPs; a discrepancy is shown here to emphasize that in practice the RPs are very often only estimated/approximated; see [CNK⁺14].

as peaks that stand out over a continuous background at the frequency f if the corresponding RP resonances with imaginary part f (or nearby) are close enough to the imaginary axis. The continuous background may have different origins. In the case of a pure point spectrum, it is due to RP resonances located far from the imaginary axis. In the presence of a continuous spectrum and RP resonances (such as shown in Panel (a) of Fig. 1), if the latter are close to the imaginary axis, then the continuous background of the PSD is mainly due to the contribution of the continuous part of $\sigma(K)$ lying typically in a sector $\{z \in \mathbb{C} : \text{Re}(z) \leq -\gamma\}$, for some $\gamma > 0$.⁴

Formula (2.21) establishes furthermore a useful correspondence between PSD and RP resonances. Indeed, from (2.21) we see that the imaginary part of the RP resonances corresponds to the location of a peak in the PSD and the absolute value of the real part to its width; see Panel (b) of Fig. 1.

Let us take $f = g = h$ in (2.11). By denoting by $a_j(h)$'s the coefficients (2.14) arising in (2.11), the latter decomposition formula writes

$$C_h(t) = \sum_{j=1}^{N_p} \left[\sum_{k=0}^{m_j-1} \frac{t^k}{k!} (K - \lambda_j \text{Id})^k \right] a_j(h) e^{\lambda_j t} + \mathcal{Q}_N(t), \quad (2.22)$$

where $\mathcal{Q}_N(t)$ exhibits typically a decay property associated with properties of the essential spectrum of \mathcal{A} . Note that the λ_j 's do not depend on the observable h , but that the $a_j(h)$'s do.

If we assume that $\text{Re}(\lambda_j) < 0$ for $j > 0$, each λ_j is simple and the absence of an essential spectrum for K , then the correlation $C_h(t)$ in (2.11) takes the simpler form of a weighted sum of complex exponentials (i.e. (2.13)), and the corresponding correlation spectrum $S_h(f)$ possesses itself a similar decomposition in terms of Lorentzian functions, namely:

$$S_h(f) = -\frac{1}{\pi} \sum_{j=1}^{\infty} a_j(h) \frac{\text{Re}(\lambda_j)}{(f - \text{Im}(\lambda_j))^2 + \text{Re}(\lambda_j)^2}, \quad f \in \mathbb{C}. \quad (2.23)$$

⁴ While we recall that in such a case, the RP resonances are the isolated eigenvalues of finite multiplicity, lying within a strip $-\gamma < \text{Re}(z) \leq 0$; see Panel (a) of Fig. 1.

It is noteworthy that such Lorentzian decompositions of the PSD are frequently encountered in applications such as for instance in spectroscopy due to the presence of resonances; e.g. [GKWG09, KMH⁺10].

In summary, the decompositions (2.11) and (2.23) inform us about the following features:

- (i) Each RP resonance is associated with an exponential contribution to the decay of correlation.
- (ii) The closer an eigenvalue to imaginary axis, the slower the decay.
- (iii) In the limit of purely imaginary eigenvalues, the associated contributions to the correlation functions are purely oscillatory and prevent the decay of correlations.
- (iv) The angular frequency at which each contribution oscillates is given by the imaginary part of the associated eigenvalue.
- (v) Eigenvalues close to the imaginary axis are associated with resonances (i.e. peaks) in the power spectrum. The spectral peak is located at the frequency given by the imaginary part of the eigenvalue and its width is proportional to the absolute value of the real part.
- (vi) The contribution of each eigenvalue to a correlation function or a power spectrum is weighted as in (2.14), corresponding to the projection of the observables h onto the eigenfunctions of K and its adjoint.

3 Reduced Ruelle-Pollicott resonances

We present in this section the main results that serve us as a foundational basis for the applications discussed hereafter and in Part III [TCND19b]. The goal is to characterize the resonances that can be extracted from low-dimensional observations in a reduced state space and how these resonances relate to the RP resonances introduced in Sec. 2.2 above.

The results of this section rely substantially on the general disintegration theorem of probability measures; see e.g. [DM78, p. 78]. The disintegration theorem states that given a probability measure μ on \mathbb{R}^N , a finite-dimensional Euclidean space V for which $1 \leq \dim(V) < N$, and a map $h : \mathbb{R}^N \rightarrow V$ (Borel-measurable), then there exists a uniquely determined family of probability measures $\{\mu_v\}_{v \in V}$ such that, for \mathfrak{m} -almost all ⁵ v in V , μ_v is concentrated on the pre-image $h^{-1}(\{v\})$ of v , i.e. $\mu_v(\mathbb{R}^N \setminus h^{-1}(\{v\})) = 0$, and such that for every Borel-measurable function $\phi : \mathbb{R}^N \rightarrow \mathbb{R}$,

$$\int_{\mathbb{R}^N} \phi(x) d\mu(x) = \int_V \left(\int_{x \in h^{-1}(\{v\})} \phi(x) d\mu_v(x) \right) d\mathfrak{m}(v). \quad (3.1)$$

Here \mathfrak{m} denotes the push-forward in V of the measure μ in \mathbb{R}^N , by the map h .

Hereafter, we apply this result for the proof of Theorem 3 when the reduced state space, V , is a subspace of \mathbb{R}^N , and the mapping h is the projector Π_V onto V . In this case, a decomposition analogous to (3.1) holds for the measure μ itself, namely

$$\mu(B \times F) = \int_F \mu_v(F) d\mathfrak{m}(v), \quad B \times F \in \mathcal{B}(V) \otimes \mathcal{B}(W), \quad (3.2)$$

with W being the subspace such that $V \oplus W = \mathbb{R}^N$. For later notation convenience, we denote by \mathcal{X}_v the pre-image $h^{-1}(\{v\})$.

3.1 Reduced Markov operators from partial observations of stochastic systems

We start with Theorem 2 below that extends [CNK⁺14, Theorem A] within the stochastic context of this article. *Mutatis mutandis*, Theorem 2 as [CNK⁺14, Theorem A], shows that — given a reduced state space V , a continuous map h from \mathbb{R}^N to V , and a Markov semigroup $(P_t)_{t \geq 0}$ possessing an invariant measure on \mathbb{R}^N — a family of reduced Markov operators (acting on functions of V) can be naturally associated with $(P_t)_{t \geq 0}$, V and h . This family characterizes a coarse-graining in the reduced state space V — such as induced by the map h — of the actual transition probabilities associated with $(P_t)_{t \geq 0}$ and taking place within the full state space \mathbb{R}^N . As a result, Theorem 2 provides, for instance, an extension to the stochastic

⁵ i.e. up to an exceptional set of null measure with respect to \mathfrak{m} .

context of Markov representations used in [SFHD99] for the description of coarse-grained dynamical features of Hamiltonian systems arising in molecular dynamics. Furthermore, empirical estimates of the transition probabilities appearing in (3.3) of Theorem 2 are important for applications, as briefly discussed in Section 3.3 below and in more details in Part III [TCND19b].

Theorem 2 *Let $(P_t)_{t \geq 0}$ be a Markov semigroup that possesses an invariant measure μ , and let V be a reduced state space for which $1 \leq \dim(V) < \dim(\mathbb{R}^N)$. Let $h : \mathbb{R}^N \rightarrow V$ be a continuous surjective function and let us denote by $\mathfrak{m} = h_*\mu$, the push-forward of the measure μ by h .*

Then there exists a time-dependent family of Markov operators T_t acting on $L_{\mathfrak{m}}^1(V)$ such that, for any Borel sets B and C of V , and any $t > 0$,

$$\frac{\langle T_t \mathbb{1}_B, \mathbb{1}_C \rangle_{1, \infty}}{\mathfrak{m}(B)} = \text{Prob}(S(t, \cdot)x \in h^{-1}(C) | x \in h^{-1}(B)), \quad (3.3)$$

where $S(t, \omega)$ denotes the stochastic flow associated with P_t .

Furthermore, if μ is ergodic then for any f in $L_{\mathfrak{m}}^2(V)$, it holds

$$\frac{1}{t} \int_0^t \int_V T_s f(v) \, \text{d}\mathfrak{m}(v) \, \text{d}s \xrightarrow[t \rightarrow \infty]{} \int_V f(v) \, \text{d}\mathfrak{m}(v). \quad (3.4)$$

Proof Let us introduce for any $t > 0$ the operator T_t which maps f in $L_{\mathfrak{m}}^1(V)$ to $T_t f$ given by

$$T_t f(v) = \int_{x \in \mathcal{X}_v} P_t[f \circ h](x) \, \text{d}\mu_v(x), \quad (3.5)$$

where \mathcal{X}_v denotes the pre-image (in \mathbb{R}^N) of v by h , and μ_v denotes the disintegrated probability measure associated with h , such as given by the disintegration theorem recalled above. The bracket-notation, $[\cdot]$, is used here in (3.5) to emphasize that the formula should be read as follows: P_t is first applied on $f \circ h$, then the integration with respect to μ_v , is undertaken. Note that T_t acts on observables of V , i.e. on functions $\phi(v)$, with v lying within the reduced state space.

By integrating (3.5) with respect to the measure \mathfrak{m} , and applying (3.1) with $\phi = P_t[f \circ h]$, one notes that

$$\int_V T_t f(v) \, \text{d}\mathfrak{m}(v) = \int_{\mathbb{R}^N} P_t[f \circ h](x) \, \text{d}\mu(x). \quad (3.6)$$

This last identity allows us to get the following estimates

$$\begin{aligned} \left| \int_V T_t f(v) \, \text{d}\mathfrak{m}(v) \right| &\leq \int_{\mathbb{R}^N} |P_t[f \circ h](x)| \, \text{d}\mu(x) \\ &\leq \int_{\mathbb{R}^N} |f \circ h(x)| \, \text{d}\mu(x), \end{aligned} \quad (3.7)$$

the last inequality resulting from (A.10) applied with $p = 1$ and the invariance of μ . We have thus the estimate

$$\|T_t f\|_{L_{\mathfrak{m}}^1(V)} \leq \int_V |f(v)| \int_{\mathcal{X}_v} \mathbb{1}_V \, \text{d}\mu_v \, \text{d}\mathfrak{m}(v) = \|f\|_{L_{\mathfrak{m}}^1(V)}, \quad (3.8)$$

which shows that T_t maps $L_{\mathfrak{m}}^1(V)$ into itself, i.e. T_t is a Markov operator.

The rest of the proof is also straightforward, and consists of noting that for any Borel sets B and C in V , one has

$$\begin{aligned} \langle T_t \mathbb{1}_B, \mathbb{1}_C \rangle_{L_{\mathfrak{m}}^1, L_{\mathfrak{m}}^\infty} &= \int_V \mathbb{1}_C(v) \left[\int_{\mathcal{X}_v} P_t[\mathbb{1}_B \circ h](y) \, \text{d}\mu_v(y) \right] \, \text{d}\mathfrak{m}(v) \\ &= \int_{\mathbb{R}^N} \mathbb{1}_{h^{-1}(C)}(x) P_t[\mathbb{1}_B \circ h](x) \, \text{d}\mu(x) \\ &= \int_{\mathbb{R}^N} \mathfrak{L}_t \mathbb{1}_{h^{-1}(C)} \cdot \mathbb{1}_B \circ h \, \text{d}\mu \\ &= \int_{\mathbb{R}^N} \mathfrak{L}_t \mathbb{1}_{h^{-1}(C)} \cdot \mathbb{1}_{h^{-1}(B)} \, \text{d}\mu \end{aligned} \quad (3.9)$$

Since $\mathfrak{m} = h_*\mu$, we deduce that $\mathfrak{m}(B) = \mu(h^{-1}(B))$, and thus by dividing by $\mathfrak{m}(B)$, one obtains that

$$\frac{\langle T_t \mathbb{1}_B, \mathbb{1}_C \rangle_{1,\infty}}{\mathfrak{m}(B)} = \text{Prob}(S(t, \cdot)x \in h^{-1}(C) | x \in h^{-1}(B)). \quad B, C \in \mathcal{B}(V). \quad (3.10)$$

Finally the proof of (3.4) consists of applying again (3.1) the characterization (iii) of the ergodicity recalled in Definition 1.

Remark 3 For any f and g in $L^2_{\mathfrak{m}}(V)$, one can define correlation functions (in the reduced state space V) associated with the family of Markov operators T_t by

$$\tilde{C}_{f,g}(t) = \int_V f \cdot T_t g \, d\mathfrak{m} - \int_V f \, d\mathfrak{m} \int_V g \, d\mathfrak{m}. \quad (3.11)$$

These correlations inherit a nice property resulting from the disintegration formula (3.1), namely

$$\tilde{C}_{f,g}(t) = C_{f \circ h, g \circ h}(t), \quad (3.12)$$

where $C_{f \circ h, g \circ h}(t)$ is given by (2.4) in which f (resp. g) is replaced by $f \circ h$ (resp. $g \circ h$). In particular the decomposition formulas of Corollary 1 apply to the observables $f \circ h$ and $g \circ h$, and provide in turn a spectral decomposition of correlation functions in the reduced state space V , in terms of the spectral elements of the generator K of the Markov semigroup $(P_t)_{t \geq 0}$.

The issue, however, is that such a decomposition involves a spectral analysis of K that is typically out of reach numerically when the dimension of state space is large, and is difficult analytically in the general case. One would like instead to have at our disposal a decomposition in terms of the spectral elements associated with the reduced Markov operators $(T_t)_{t \geq 0}$ rather than with full Markov semigroup $(P_t)_{t \geq 0}$. The stumbling block is that $(T_t)_{t \geq 0}$ does not form a semigroup in general i.e.

$$T_{t+s} \neq T_t T_s, \quad (3.13)$$

(see [TvdBD15]) and one cannot rely on the spectral theory of semigroups such as used in Section 2.2, to reach this goal. The next sections analyzes what type of useful information that may still be extracted from T_t .

3.2 Pseudo generators of the family of reduced Markov operators and asymptotic behavior

We begin with the following Lemma that characterizes the pseudo-generators G_t that can be associated with the family of reduced Markov operators $(T_t)_{t \geq 0}$, although the latter family does not form a semigroup in general. For related material, we refer to the recent works [BKJ15, BHJK15] concerned with the metastability analysis of Langevin dynamics. In a certain sense, Theorem 3 below along with the notion of reduced RP resonances proposed in Sec. 2.2, provide a complementary approach for the analysis of reduced Markov operators for situations not necessarily limited to Langevin dynamics. Theorem 3 shows in particular that the asymptotic behavior of G_t as $t \rightarrow \infty$ (in a weak sense) is governed by the generator of a reduced SDE in which the dependence on the unobserved variables (lying outside of the reduced state space V) has been averaged out; see (3.22) below. In other words, Theorem 3 provides a useful relationship between the reduced non-Markovian process associated with $(T_t)_{t \geq 0}$ and the Markov process associated with the conditional expectation induced by the observable h .

Lemma 1 Let K denote the generator of P_t in $L^2_{\mu}(\mathbb{R}^N)$. Then, the Markov operator T_t of Theorem 2 possesses for each time t , a pseudo-generator, i.e. for any f in $L^2_{\mathfrak{m}}(V)$ for which $f \circ h$ lies in $D(K)$, $\lim_{s \rightarrow 0} (T_{t+s}f - T_t f)/s$ exists and is given by

$$G_t f(v) = \int_{\mathcal{X}_v} P_t K[f \circ h](x) \, d\mu_v(x), \quad (3.14)$$

whose domain is thus

$$D(G_t) = \{f \in L^2_{\mathfrak{m}}(V) : f \circ h \in D(K)\}. \quad (3.15)$$

Proof The proof is elementary and consists of noting that for any f in $L_m^2(V)$ for which $f \circ h$ lies in $D(K)$, we have

$$\frac{T_{t+s}f - T_t f}{s} = \int_{\mathcal{X}_v} \frac{P_{t+s} - P_t}{s} [f \circ h](x) d\mu_v(x) \xrightarrow{s \rightarrow 0} \int_{\mathcal{X}_v} P_t K[f \circ h](x) d\mu_v(x), \quad (3.16)$$

from the basic properties of a strongly continuous semigroup such as applied to P_t (e.g. [EN00, Lem. II.1.3]) and the dominated convergence theorem.

Theorem 3 *Assume that the assumptions of Theorem 2 hold. If μ is ergodic, then there exists a closed linear operator \mathcal{G} acting on $L_m^2(V)$ such that for any f in $L_m^2(V)$*

$$\frac{1}{t} \int_0^t \int_V G_t f(v) dm(v) \xrightarrow{t \rightarrow \infty} \int_V \mathcal{G}f(v) dm(v), \quad (3.17)$$

and for which \mathcal{G} is dissipative in the sense that

$$\operatorname{Re} \langle \mathcal{G}f, f \rangle_{L_m^2(V)} \leq 0, \quad \forall f \in D(\mathcal{G}). \quad (3.18)$$

Furthermore, if μ is strongly mixing then

$$\lim_{t \rightarrow \infty} \int_V G_t f(v) dm(v) = \int_V \mathcal{G}f(v) dm(v). \quad (3.19)$$

Recall that the Markov semigroup P_t from which G_t is defined, is associated to Eq. (2.1). If h is a projector, the operator \mathcal{G} is densely defined, and if P_t is a Feller semigroup for which $C_0^\infty(\mathbb{R}^N) \subset D(K)$, then \mathcal{G} possesses the following differential expression

$$\mathcal{G}f(v) = \frac{1}{2} \sum_{i,j=1}^{\dim(V)} \bar{\Sigma}_{ij}(v) \partial_{ij}^2 f + \sum_{i=1}^{\dim(V)} \bar{F}_i(v) \partial_i f, \quad v \in V, \quad f \in C_0^2(V). \quad (3.20)$$

where

$$\bar{\Sigma}_{ij}(v) = \int_{\mathcal{X}_v} \Sigma_{ij}(x) d\mu_v(x), \quad \bar{F}_i(v) = \int_{\mathcal{X}_v} F_i(x) d\mu_v(x), \quad x \in \mathbb{R}^N, \quad v \in V, \quad (3.21)$$

where F and Σ denote respectively the drift part and the diffusion tensor associated with Eq. (2.1).

In other words, \mathcal{G} is the generator of the Markov process associated with the **conditional expectation** induced by h , namely with the following reduced SDE in V :

$$\boxed{dv = \bar{F}(v) dt + \sigma(v) dW_t^V, \quad v \in V,} \quad (3.22)$$

with W_t^V denoting a Brownian motion in V and for $1 \leq i, j \leq \dim(V)$,

$$\bar{\Sigma}_{ij}(v) = (\sigma(v)\sigma(v)^T)_{ij} = \sum_{k=1}^M \overline{D_{ik}(x)D_{jk}(x)}, \quad \text{for some } M \geq 1, \quad (3.23)$$

where the $D_{\ell k}$'s are the diffusion coefficients of the SDE (2.1).

Proof Step 1: Proof of (3.17) and (3.18). Let f be in \mathcal{X} given in (3.15). Then

$$\begin{aligned} \int_V G_t f(v) dm(v) &= \int_V \int_{\mathcal{X}_v} P_t K[f \circ h](x) d\mu_v(x) dm(v), \\ &= \int_{\mathbb{R}^N} P_t K[f \circ h](x) d\mu(x), \end{aligned} \quad (3.24)$$

by application of the disintegration formula (3.1).

Assuming the measure μ to be ergodic, the characterization (iii) of Definition 1 allows us to infer (3.17) with \mathcal{G} given by

$$\mathcal{G}f(v) = \int_{\mathcal{X}_v} K[f \circ h](x) d\mu_v(x), \quad (3.25)$$

by application once more of the disintegration formula.

The same formula ensures furthermore that

$$\int_V \mathcal{G}f \cdot f \, d\mathbf{m} = \int_{\mathbb{R}^N} K[f \circ h] \cdot [f \circ h] \, d\mu \quad (3.26)$$

and by taking the real part, (3.18) follows from (A.14).

Step 2: \mathcal{G} is closed. Let $\{f_n\}$ be a sequence in $D(\mathcal{G})$ converging to f in $L_m^2(V)$ such that $\mathcal{G}f_n \rightarrow g$ in $L_m^2(V)$ as $n \rightarrow \infty$.

Then from (3.25), one has

$$\begin{aligned} \int_V |\mathcal{G}f_n - g|^2 \, d\mathbf{m} &= -2 \int_{\mathbb{R}^N} K[f_n \circ h](x) \cdot (g \circ h(x)) \, d\mu + \\ &\quad \int_{\mathbb{R}^N} |K[f_n \circ h](x)|^2 \, d\mu + \int_{\mathbb{R}^N} |g \circ h(x)|^2 \, d\mu(x). \end{aligned} \quad (3.27)$$

Now since K is closed in $L_\mu^2(\mathbb{R}^N)$, $K[f_n \circ h] \xrightarrow{n \rightarrow \infty} K[f \circ h]$ and $f \circ h \in D(K)$, namely $f \in D(\mathcal{G})$. This shows furthermore that one can pass to the limit in (3.27), which gives

$$2 \int_{\mathbb{R}^N} K[f \circ h](x) \cdot (g \circ h(x)) \, d\mu = \int_{\mathbb{R}^N} |K[f \circ h](x)|^2 \, d\mu + \int_{\mathbb{R}^N} |g \circ h(x)|^2 \, d\mu(x), \quad (3.28)$$

The identity (3.28) says nothing else than

$$\|K[f \circ h] - g \circ h\|_{L_\mu^2}^2 = 0, \quad (3.29)$$

i.e. $K[f \circ h] = g \circ h$, μ -almost everywhere, which gives

$$\int_{\mathcal{X}_v} K[f \circ h](x) \, d\mu_v(x) = \int_{\mathcal{X}_v} g \circ h(x) \, d\mu_v(x) = g(v), \quad \text{for } \mathbf{m}\text{-almost every } v \text{ in } V. \quad (3.30)$$

We have thus proved $\mathcal{G}f = g$, with f in $D(K)$, i.e. that \mathcal{G} is closed.

Step 3: \mathcal{G} is densely defined if h is a projector. The purpose is to prove that $D(\mathcal{G})$ is dense in $L_m^2(V)$. Assume by contradiction that there exist g in $L_m^2(V)$ for which any sequence in $D(\mathcal{G})$ would stay at a certain distance (in $L_m^2(V)$) from g or would have g as a point of accumulation, but not unique.

On the other hand, since $D(K)$ is dense in $L_\mu^2(\mathbb{R}^N)$ here exists a sequence $\{\varphi_n\}$ in $D(K)$ that converges to $g \circ h$ in $L_\mu^2(\mathbb{R}^N)$. Let us define a sequence $\{f_n\}$ (in $D(\mathcal{G})$) from the semiconjugacy relation

$$f_n \circ h = h \circ \varphi_n, \quad (3.31)$$

i.e. such that the following diagram is commutative

$$\begin{array}{ccc} \mathbb{R}^N & \xrightarrow{\varphi_n \in D(K)} & \mathbb{R}^N \\ \downarrow h & & \downarrow h \\ V & \xrightarrow{f_n \in D(\mathcal{G})} & V \end{array}$$

Now let us note that $h \circ \varphi_n \xrightarrow{n \rightarrow \infty} h \circ g \circ h = g \circ h$ if h is a projector, and therefore

$$\int_{\mathbb{R}^N} |(f_n \circ h)(x) - (g \circ h)(x)|^2 \, d\mu(x) = \int_V |f_n(v) - g(v)|^2 \, d\mathbf{m}(v) \xrightarrow{n \rightarrow \infty} 0, \quad (3.32)$$

by application of the disintegration formula, leading thus to a contradiction.

Step 4: \mathcal{G} provides the conditional expectation. This step is a straightforward consequence of the Dynkin theorem (see Remark 1-(iii)) and the representation formula (3.25) which leads to the expression (3.20) of \mathcal{G} , by integration with respect to μ_v .

Remark 4

(i) \mathcal{G} describes long- and short-time behaviors in V . Due to (3.13), the stochastic process associated with T_t in the reduced state space V is in general non-Markovian. Theorem 3 shows however that \mathcal{G} plays an important role in the description of the asymptotic behavior of this stochastic process. Noteworthy is that \mathcal{G} as defined in (3.25) is exactly the pseudo-generator G_t of T_t , as $t \rightarrow 0$.

The operator \mathcal{G} relates thus the long time asymptotics ($t \rightarrow \infty$) of the reduced stochastic process to its infinitesimal characteristics which describe the short-time behavior ($t \rightarrow 0$), as for time homogeneous Markov processes.

(ii) **Reduced Kolmogorov operator.** Given a projector $h = \Pi_V$ onto V , another useful object related to the conditional expectation (3.22) is the following reduced Kolmogorov operator:

$$\bar{\mathcal{K}}f(v) = \int_{\mathcal{X}_v} \mathcal{K}[f \circ h](x) d\mu_v(x), \quad v \in V, \quad f \in C_0^2(V), \quad (3.33)$$

where \mathcal{K} is the Kolmogorov operator defined in Appendix A.3 for the full SDE.

This operator has an interesting interpretation: although it is not the “generator” of the non-Markovian stochastic process v_t associated with T_t , it provides the generator of the Markov stochastic process describing the average behavior of v_t , when the averaging is taken over the unobserved variables (i.e. those lying outside of V).

(iii) For the sake of clarity, Theorem 3 has been articulated (in part) for the case where h is a projection. Of course Theorem 3 can be proved for more general mappings h , and for V that is not necessarily a subspace of \mathbb{R}^N . In such cases, the corresponding reduced Kolmogorov operator, $\bar{\mathcal{K}}$, becomes typically a non-local operator. Such considerations about the choice of reduced state space has its importance in practice. Indeed non-local features of $\bar{\mathcal{K}}$ can intuitively help keep a “trace” in the reduced state space of certain interactions between the observed and unobserved variables that would be otherwise averaged out by the conditional expectation by using standard projections.

3.3 Reduced RP resonances and conditional expectation

As pointed out earlier, RP resonances although useful to describe power spectra and correlation functions, are difficult to estimate in practice when the dimension of the state space gets large. In practice, only partial observations of the solutions to Eq. (2.1) are available, e.g. few solution’s components. Theorem 2 shows that from partial observations of a complex system that lie within a reduced state space V and are collected at a sampling rate τ , a (reduced) Markov operator T_τ with state space V can be inferred from these observations. Theorem 2 shows then that this operator T_τ characterizes the coarse-graining in V of the transition probabilities in the full state space. At an intuitive level if a dominant recurrent behavior occurs within an irregular background, then T_τ must still “feel” this recurrent behavior within V , in case this dominant behavior is reflected in V . As pointed out already in [CNK⁺14] such a recurrent behavior is manifested by eigenvalues of T_τ distributed evenly along an inner circle typically close to the unit circle, or by forming a parabola in the complex plane depending on the representation adopted; see Part II [TCND19a] and Part III [TCND19b].

An issue though, is that in general T_τ does not come from a Markov semigroup (acting on functions of the reduced state space V), and thus one cannot *per se* rely on the theory of RP resonances presented in Sec. 2.2 to decompose e.g. correlation functions of V ; see Remark 3. Theorem 3 on the other hand, shows that T_τ has a pseudo-generator G_τ that is close to the generator of the conditional expectation (3.22), when either τ is sufficiently small or large; see Remark 4-(ii). Thus, in such cases, the resonances that can be estimated from data collected at the sampling rate τ in the reduced state space have an interesting interpretation. They approximate the RP resonances associated with the reduced system (3.22), and are called the **reduced RP resonances**. As such, these resonances describe the solution’s variability captured by the conditional expectation (3.22).

In practice, the dimension of V is kept low so that T_τ can be efficiently estimated via a maximum likelihood estimator (MLE). Note that the reduced state space V should be also chosen such that the observed dynamics in V carry relevant information on e.g. the variability of interest contained in the original system’s solutions.

We detail below our estimation procedure of reduced RP resonances. First a bounded domain \mathcal{D} of V should be chosen large enough so that “most realizations” of the stochastic process X_t solving Eq. (2.1) fall inside \mathcal{D} after application of the observable $h : \mathbb{R}^p \rightarrow V$, i.e. \mathcal{D} must be chosen so that $h(X_t)$ belongs to \mathcal{D} for many realizations of the noise in Eq. (2.1). This domain is then discretized as the union of M disjoint boxes B_j , forming thus a partition.

We assume that our observations are made out of the stochastic process X_t (solving Eq. (2.1)) at discrete time instants $t = t_n$, given as multiple of a sampling time τ , i.e. $t_n = n\tau$ with $1 \leq n \leq N$, with N assumed to be large. These observations made in the observation space V are denoted by $Y_n = h(X_{t_n})$. By adapting the material contained in the Supporting Information of [CNK⁺14] (see also [TvdBD15]), the Markov operator T_τ is approximated by the $M \times M$ transition matrix Γ_τ whose entries are given

$$(\Gamma_\tau)_{ij} = \frac{\#\left\{\left(Y_n \in B_j\right) \wedge \left(Y_{n+1} \in B_i\right)\right\}}{\#\left\{Y_n \in B_j\right\}}, \quad (3.34)$$

where the B_j 's form a partition (composed of M disjoint boxes) of the aforementioned domain \mathcal{D} in V ; see e.g. [SFHD99, CVE06, CNK⁺14, TvdBD15]. In (3.34), the notation $\#\{(Y_n \in B_k)\}$ gives the number of observations Y_n visiting the box B_k , and the logical symbol “ \wedge ” means “and.” The leading eigenvalues of the transition matrix Γ_τ can then be computed with an iterative algorithm such as ARPACK [LSY97].

The reduced RP resonances are then obtained as the eigenvalues $\lambda_k(\tau)$ obtained from the eigenvalues $\zeta_k(\tau)$ of the Markov matrix T_τ , according to

$$\lambda_k(\tau) = \frac{\log(|\zeta_k(\tau)|)}{\tau} + i \frac{\arg(\zeta_k(\tau))}{\tau}, \quad 1 \leq k \leq M, \quad (3.35)$$

where $\arg(z)$ (resp. $\log(z)$) denotes the principal value of the argument (that we adopt to lie in $[-\pi, \pi]$ in this article) (resp. logarithm) of the complex number z . At a basic level, the motivation behind (3.35) is that the eigenvalues of Γ_τ as the eigenvalues of a Markov matrix, lie within the unit circle (representation that was adopted in [CNK⁺14]) whereas we want here to relate these eigenvalues with the RP resonances associated with the generator K of the original Eq. (2.1). This way, the $\lambda_k(\tau)$'s given by (3.35) lie naturally within the left-half complex plane.

For τ sufficiently small or large, i.e. when the generator of the conditional expectation is well approximated by pseudo-generator of T_τ , one can thus proceed as follows to measure the amount of solution's variability captured by the conditional expectation:

- (C₁) Given an observable and a reduced state space V , one forms the right-hand side (RHS) of (2.13) in which the λ_j 's are replaced by the $\lambda_j(\tau)$'s given by (3.35), and the eigenfunctions by those of the Markov operator T_τ .
- (C₂) One compares the resulting “reduced” correlation function obtained in step 1, to the correlation function as estimated from standard techniques. If the correlations are close, one can conclude that the conditional expectation provides actually a good reduced system and that the non-Markovian effects are negligible to obtain an efficient closure in V , for this observable.

The next section illustrates such situations. Note that by replacing the RP resonances by the reduced RP resonances in (2.23), power spectral densities can also be approximated in the same fashion.

When τ is neither small or large, the precise relationships between the $\lambda_k(\tau)$'s and the actual RP resonances are non-trivial to characterize in general. Nevertheless, in certain cases, as shown in Part III, the reduced RP resonances are very useful to diagnose and characterize important dynamical features such as nonlinear oscillations embedded within a stochastic background.

4 Applications to a stochastic slow-fast system

4.1 The model

We consider the following stochastic system

$$\begin{cases} dx = (\lambda x - fy - \gamma xz) dt + \sigma dW_t^1 & (4.1a) \\ dy = (fx + \lambda y - \gamma yz) dt + \sigma dW_t^2 & (4.1b) \\ dz = -\frac{1}{\epsilon}(z - x^2 - y^2) dt + \frac{\sigma}{\sqrt{\epsilon}} dW_t^3. & (4.1c) \end{cases}$$

The stochastic processes W_t^1, W_t^2, W_t^3 are independent Brownian motions. The parameters λ, f, γ and ϵ are specified below. In the case $\sigma = 0$, system (4.1) arises in fluid dynamics and is investigated in [NAM⁺03] as low-dimensional reduced model for a flow past a circular cylinder. Hereafter, we deal exclusively with the stochastic case, i.e. when $\sigma > 0$. The reduction problem of this system is analyzed rigorously in [CLPR19] using a different approach inspired from [CLW15b]. In particular it is proved that this system generates a Markov semigroup which is strong Feller and irreducible that possesses a unique ergodic invariant measure μ and thus the theory of RP resonances presented in this article applies.

System (4.1) is a slow-fast system driven by additive noise. The theory of slow-fast systems in the deterministic case, i.e. when $\sigma = 0$ in system (4.1), is well established when the time-scale separation is strong; see for instance [Jon95, NS13] or the recent monograph [Kue15] and references therein. Indeed, a typical behavior of such systems is characterized by a separation of time scales between the so-called "slow" and "fast" variables, as controlled by the parameter ϵ . For our present system, as ϵ gets small, the z -variable exhibits fast fluctuations on timescales over which the x - and y -variables vary more and more slowly.

Central to the study of such systems in the deterministic context, is the existence of a *slow manifold* which expresses often a (possibly approximate) slaving relationships between the slow- and fast-variables; the latter lying typically in an " ϵ -neighborhood" (for ϵ sufficiently small) to the graph of a function h of the slow variables. In the case of system (4.1) the slow manifold is explicitly given, for $\sigma = 0$, by

$$\begin{aligned} \Phi : \mathbb{R} \times \mathbb{R} &\longrightarrow \mathbb{R}, \\ (x, y) &\mapsto x^2 + y^2. \end{aligned} \quad (4.2)$$

More precisely, for $\sigma = 0$, any solution $(x(t), y(t), z(t))$ to (4.1) is attracted exponentially fast to the manifold given by $\mathcal{M}_\epsilon = \text{graph}(\Phi) + \mathcal{O}(\epsilon)$, and in particular the near to slaving relationship $z(t) = h(x^2(t) + y^2(t)) + \mathcal{O}(\epsilon)$ holds for t sufficiently large. Foundations of such a dynamical behavior are found in [Fen79, FM71] for much more general systems than (4.1), and we refer to [NS13, Chap. 10] for an introduction to the theory of invariant manifolds (IMs) for singularly perturbed ordinary differential equations (ODEs); see also [Kue15].

In what follows we determine the reduced RP resonances as described in Sec. 3.3 not only for the system (4.1), but also for the reduced system based on the slow manifold Φ , namely by integrating

$$\begin{aligned} du &= (\lambda u - fv - \gamma u(u^2 + v^2)) dt + \sigma dW_t^1 \\ dv &= (fv + \lambda v - \gamma v(u^2 + v^2)) dt + \sigma dW_t^2. \end{aligned} \quad (4.3)$$

The RP resonances of this system, a Hopf normal form subject to an additive white noise, are analyzed in details in Part II [TCND19a]. This system is a *paragon* of nonlinear oscillations in presence of noise.

The reduced state space V is taken to be the (x, y) -plane, being thus only a reduced state space for the original system. As shown below, the estimation of (reduced) RP resonances allow also for comparing the original system with its slow manifold reduction (4.3).

4.2 Numerical results

Both systems are numerically integrated via an Euler-Maruyama scheme with a time step of size $\delta t = 10^{-5}$. In each experiment, the systems are simulated up to $T = 8 \times 10^4$ after removal of a transient dynamics of length $T_s = 10^3$. For a given sampling rate τ (specified below) and $t_n = n\tau$, the variables collected are

$Y_n = (x_n, y_n)$ where $x_n = x(t_n)$ and $y_n = y(t_n)$ for system (4.1), on one hand, and $Y_n = (u_n, v_n)$ where $u_n = u(t_n)$ and $v_n = v(t_n)$ for system (4.3), on the other. The domain \mathcal{D} used for estimating the Markov matrix I_τ in (3.34), is taken to be $\mathcal{D} = [-6, 6] \times [-6, 6]$, decomposed into a uniform grid constituted of 300×300 cells.

Three parameter regimes are considered hereafter: two regimes with a strong time-scale separation (Cases I and II), and one with no apparent time-scale separation (Case III); see Tables 1 and 2. The sampling rate τ is chosen depending on these regimes as follows: $\tau = 10^{-3}$ for Case I, and $\tau = 10^{-2}$ for Cases II and III.

Within this experimental protocol, we first estimate the reduced RP resonances. To do so, we first estimate the Markov matrix entries of I_τ according to (3.34), and then determine the reduced RP resonances according to (3.35). For the three parameter regimes considered here, these resonances are shown by blue '+' signs in each Panel (a) of Figures 2, 3, and 4. The resonances for the slow manifold system (4.3) are estimated according the same procedure. They are no longer reduced but genuine RP resonances since the system is two-dimensional. Thus, the RP resonances associated with system (4.3) are shown by orange 'x' signs in each Panel (a) of Figures 2, 3, and 4. In each of these Panels (a), the rightmost narrow panel shows the spectral reconstructions of the PSD of $u(t)$ (resp. $x(t)$) based on RP resonances (resp. reduced RP resonances) according to procedure (C₁) described in Sec. 3.3, by a dashed orange (resp. blue) curve, while the black curve shows its standard PSD sample estimate. In each Panel (b) of Figures 2, 3, and 4 are shown the autocorrelation function (ACF) of $u(t)$ (resp. $x(t)$) based on RP resonances (resp. reduced RP resonances) based on the procedure (C₁), and here again the black curve its standard ACF sample estimate.

The conclusions of these numerical experiments are without ambiguity. First, in the case of a strong time-scale separation between the observed variables (x, y) and the unobserved variable (z) , i.e. in Cases I and II, the reduced RP resonances allow for an almost exact reconstruction of the ACF and PSD of $x(t)$ (as for $y(t)$, not shown). This is explained from the theoretical understanding provided by Secs. 3.2 and 3.3. Indeed, as explained therein, since the sampling rate τ is small here, it is expected that the reduced RP resonances provide a good approximation of the resonances associated with the generator of the conditional expectation (3.22). On the other hand, the conditional expectation is known to provide a (very) good approximation of the dynamics of the slow variables when ϵ is small, and that it coincides with the slow manifold reduced system (4.3) as $\epsilon \rightarrow 0$; see [CLPR19]. Thus without any surprise the RP resonances of system (4.3) coincide with the reduced RP resonances of (4.1) for a strong time-scale separation. The fact that these resonance are organized along parabolas in the left half plane, is thoroughly analyzed in Part II [TCND19a].

The conclusion for the case of no time-scale separation (Case III) is more subtle, but reveals an interesting usefulness of the RP resonance analysis. For this case, the RP resonances of the slow manifold reduced system (4.3) differ from the reduced RP resonances (see Panel (a) in Fig. 4) which comes here with a noticeable mismatch at the level of ACFs although not revealed at the level of PSDs. On the contrary, the reduced RP resonances allow for an almost perfect reconstruction of the ACF. As a consequence, one can conclude that the slow manifold is no longer a valid parameterization of the z -variable in terms of the x - and y -variables (also observed in [CLPR19] for this regime) but that still, the conditional expectation (3.22) provides a valid approximation for the reduced dynamics.

Such a diagnosis regarding the conditional expectation drawn from RP resonance analysis offers promising perspectives in terms of reduced order modeling. As illustrated here, the latter resonances could indeed serve to determine whether efforts on non-Markovian effects must be invested or not. In that respect, we mention the recent variational approach relying on optimal manifolds minimizing a parameterization defect [CLM19] which offers new perspectives to approximate analytically the conditional expectation; see also [CLM17]. Applied to Case III, this approach provides an analytical substitute to the slow manifold which allows in turn for the derivation of an efficient 2D reduced system of SDEs (without non-Markovian terms), confirming the conclusions drawn from RP resonance analysis.

Acknowledgements This work has been partially supported by the Office of Naval Research (ONR) Multidisciplinary University Research Initiative (MURI) grant N00014-12-1-0911 and N00014-16-1-2073 (MDC), by the National Science Foundation grants DMS-1616981 (MDC) and AGS-1540518 (JDN), by the LINC project (No. 289447) funded by EC's Marie-Curie ITN (FP7-PEOPLE-2011-ITN) program (AT and HD) and by the Utrecht University Center for Water, Climate and Ecosystems (AT).

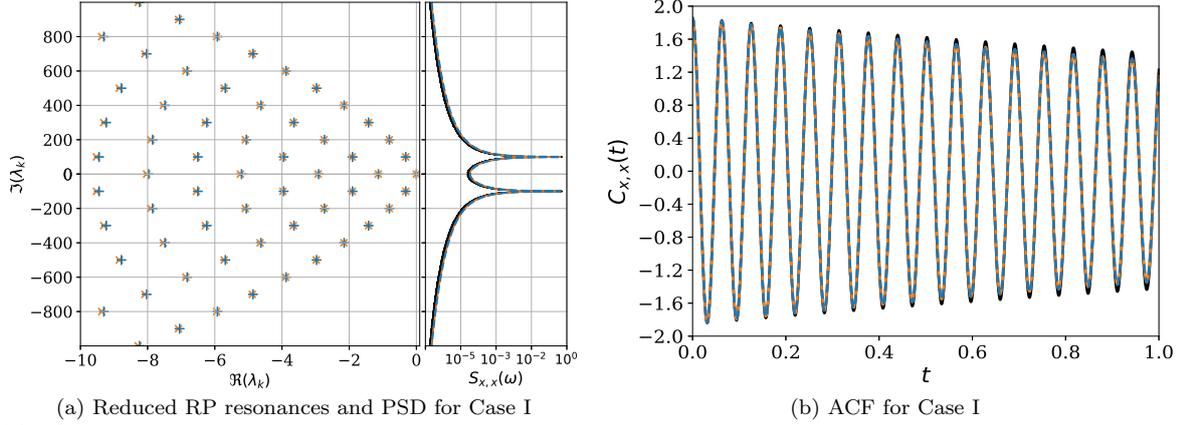


Fig. 2: **Leftmost frame in Panel (a):** RP resonances of (4.3) (orange 'x' sign) and reduced RP resonances of (4.1) (blue '+' sign) in the (x, y) -plane. **Rightmost frame in Panel (a):** Power spectral densities (PSDs). **Panel (b):** Autocorrelation function (ACF) of $x(t)$ (resp. $u(t)$). The spectral reconstructions of the PSD and ACF of $u(t)$ (resp. $x(t)$) based on RP resonances (resp. reduced RP resonances) according to procedure (C_1) described in Sec. 3.3, are shown by a dashed orange (resp. blue) curve, while the black curves show their standard sample estimates.

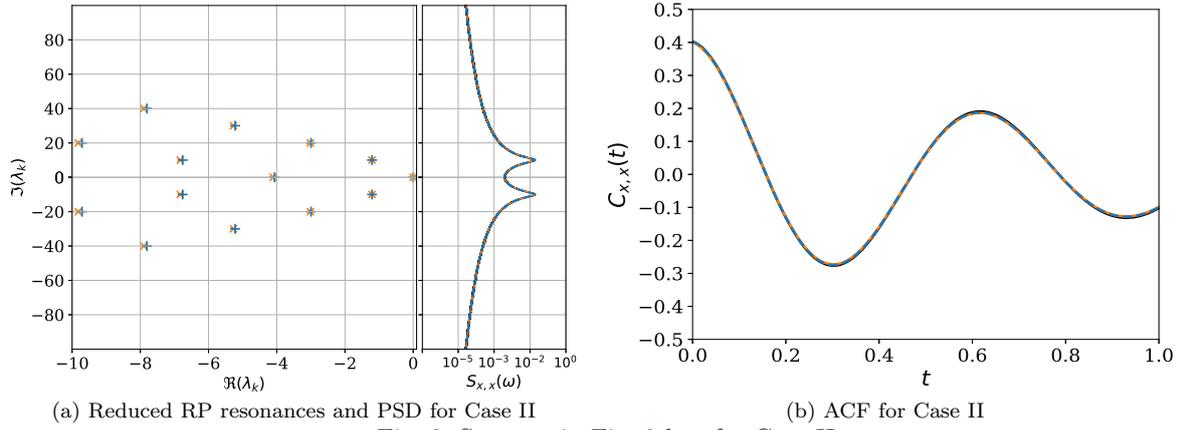


Fig. 3: Same as in Fig. 2 but for Case II.

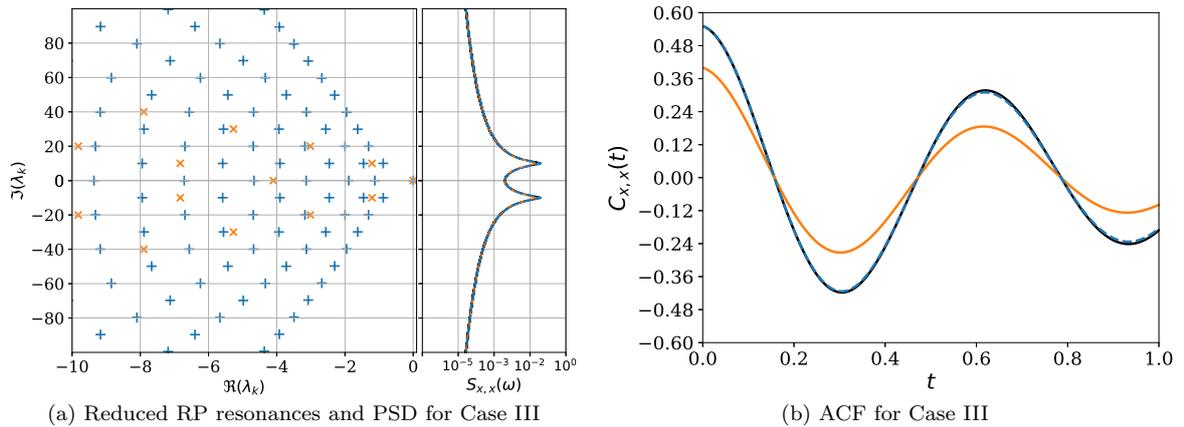


Fig. 4: Same as in Fig. 2 but for Case III.

Table 1: Parameter regimes: Case I and Case II

	λ	f	γ	ϵ	σ
Case I	10^{-3}	10^2	5.6×10^{-2}	10^{-2}	0.55
Case II	10^{-3}	10	1	10^{-2}	0.2

Table 2: Parameter regime: Case III

	λ	f	γ	ϵ	σ
Case IV	10^{-3}	10	1	10	0.3

A Elements of stochastic analysis

In this appendix we present a short survey of elements of stochastic analysis used in the Main Text. The main objective is to introduce the key concepts and tools of stochastic analysis for stochastic differential equations (SDEs), to a wider audience in the geosciences and macroscopic physics.

A.1 Markov semigroups

Two approaches dominate the analysis of stochastic dynamics. We are here concerned with the approach rooted in Stochastic Analysis which, contrary to the random dynamical system (RDS) approach [Arn98, CSG11, CLW15a], does not substitute a deterministic (nonlinear) flow $S(t)$ by a stochastic flow $S(t, \omega)$ acting⁶ on the state space \mathcal{X} but rather by a family of linear operators P_t , acting on a space of observables of the state space, i.e. on functions of \mathcal{X} . A typical choice of observables is given by $\mathcal{C}_b(\mathcal{X})$, the space of bounded and continuous functions on \mathcal{X} .

More precisely, this family P_t reflects the (averaged) action of the stochastic flow at the level of functions and is given as the mapping which to each function ϕ in $\mathcal{C}_b(\mathcal{X})$ associates the function:

$$P_t \phi(x) = \mathbb{E}(\phi(S(t, \cdot)x)) = \int_{\Omega} \phi(S(t, \omega)x) d\mathbb{P}(\omega), \quad t \geq 0, x \in X. \quad (\text{A.1})$$

In (A.1), the function ϕ is the aforementioned *observable*. Its physical meaning could be, for instance, the potential vorticity or the temperature of a fluid at a given location or averaged over a volume. The RHS of (A.1) involves averaging over the realizations ω , i.e. expectation. For deterministic flow it reduces to $P_t \phi(x) = \phi(S(t)x)$ and is known as the Koopman operator. Note that P_t such as defined in (A.1) is not limited to stochastic flow, more generally $P_t \phi(x) = \mathbb{E}(\phi(X_t^x))$ where X_t^x denotes a stochastic process that solves Eq. (2.1) (as associated with P_t) and emanates from x in \mathcal{X} .

Under general assumptions on F and D , the stochastic process X_t solving Eq. (2.1) is Markovian (i.e. the future is determined only by the present value of the process) which translates at the level of P_t into the following semigroup property

$$P_0 = \text{Id}, \quad P_t P_s = P_{t+s}, \quad t, s \geq 0. \quad (\text{A.2})$$

A breakdown of (A.2) indicates thus that the underlying stochastic process is non-Markovian.

It is noteworthy to mention that even when P_t satisfies (A.2), it does not ensure that P_t is a *strongly continuous* semigroup [Paz83] on $\mathcal{C}_b(\mathcal{X})$. Nevertheless, $(P_t)_{t \geq 0}$ is extendable to a strongly continuous semigroup in L^2_μ as soon as μ is an invariant measure of the Markov semigroup; see Theorem 4 below. The spectral theory of such semigroups [EN00] is at the core of the description of mixing properties in L^2_μ , such as presented in Sec. 2.2 in the Main Text.

A.2 Ergodic invariant measures and the strong Feller-Irreducibility approach

The Fokker-Planck equation (2.3) may support several weak stationary solutions. An important question, is thus the identification of stationary measures that describe the asymptotic statistical behavior of the solutions of Eq. (2.1), in a typical fashion. The notion of ergodic invariant measures plays a central role in that respect, and relies on the following important characterization of ergodic measures for (stochastically continuous) Markov semigroups [DPZ96, Theorem 3.2.4].

Definition 1 *An invariant measure is ergodic if one of the following three equivalent statements holds:*

- i) *For any $f \in L^2_\mu(\mathcal{X})$, if $P_t f = f$, almost surely w.r.t μ (μ -a.s.) for all $t \geq 0$, then f is constant μ -a.s.*
- ii) *For any Borel set Γ of \mathcal{X} , if $P_t \mathbb{1}_\Gamma = \mathbb{1}_\Gamma$ μ -a.s. for all $t \geq 0$, then $\mu(\Gamma) = 0$ or 1.*
- iii) *For any $f \in L^2_\mu(\mathcal{X})$, $\frac{1}{T} \int_0^T P_s f ds \xrightarrow{T \rightarrow \infty} \int f d\mu$ in $L^2_\mu(\mathcal{X})$.*

⁶ ω labelling the noise realization.

In practice, an efficient approach to show the existence of an ergodic measure consists of showing the existence of a *unique* invariant measure, since in this case such an invariant measure is necessarily ergodic [DPZ96, Theorem 3.2.6]. Various powerful approaches exist to deal with the existence of a unique invariant measure. The next section discusses the classical approach based on the theory of *strong Feller Markov semigroups* and *irreducibility*.

The main interest of the strong Feller-Irreducibility approach lies in its usefulness for checking the conditions of the Doob-Khasminskii Theorem [Doo48, Kha60, DPZ96], the latter ensuring the existence of *at most* one ergodic invariant measure. This strategy requires the proof of certain smoothing properties of the associated Markov semigroup, and to show that any point can be (in probability) reached at any time instant by the process regardless of initial data. This property is known as *irreducibility*. It means that $P_t \mathbb{1}_U(x) > 0$ for all x in \mathcal{X} , every $t > 0$, and all non-empty open sets U of \mathcal{X} , which is equivalent to say that

$$\mathbb{P}(\|S(t, \cdot)x - z\| < \epsilon) > 0, \quad (\text{A.3})$$

for any z in \mathcal{X} , $\epsilon > 0$ and $t > 0$; see [Cer01, p. 67]. In other words the irreducibility condition expresses the idea that any neighborhood of any point z in \mathcal{X} , is reachable at each time, with a positive probability.

Remarkably, the irreducibility is usually inferred from the controllability of the associated control system $\dot{x} = F(x) + D(X)u(t)$; see [CLPR19] for a simple illustration. This approach is well-known and based on the support theorem of Stroock and Varadhan [SV71] (see also [IW14, Theorem 8.1]) that shows that several properties of the SDEs can be studied and expressed in terms of the control theory of ordinary differential equations (ODEs); see [DPZ96, Secns. 7.3 and 7.4] for the case of additive (non-degenerate) noise and [AK87, Kli87] for the more general case of nonlinear degenerate noise, i.e. in the case where the noise acts only on part of the system's equations, corresponding to $\ker(Q) \neq \{0\}$.

The strong Feller property means that the semigroup maps bounded measurable functions into bounded continuous functions. This property, related to a regularizing effect of the Markov semigroup $(P_t)_{t \geq 0}$, is a consequence of the *hypoellipticity* of the Kolmogorov operator \mathcal{K} defined on smooth functions ψ (of class C^2):

$$\mathcal{K}\psi(x) = \frac{1}{2} \text{Tr}(\Sigma \nabla^2 \psi(x)) + \langle F(x), \nabla \psi \rangle, \quad (\text{A.4})$$

where

$$\text{Tr}(\Sigma \nabla^2 \psi(\cdot)) = \sum_{i,j=1}^N [D(X)D(X)^T]_{ij} \partial_{ij}^2 \psi. \quad (\text{A.5})$$

Note that hypoelliptic operators include those that are uniformly elliptic for which the Weyl's smoothing lemma applies; e.g. [Dac04, Theorem 4.7]. Hypoellipticity allows nevertheless for dealing with the case of degenerate noise, which is important in applications.

A very efficient criteria for hypoellipticity is given by Hörmander's theorem [Hör67, Nor86]; see also [CSG11, Appendix C1] for a discussion on the related Hörmander's bracket condition and its implications to the existence of other types of meaningful measures for SDEs, namely the Sinai-Ruelle-Bowen (SRB) random measures. We refer also to Part II [TCND19a], for an instructive verification of the Hörmander's condition in the case of the Hopf normal form subject to additive noise.

From a geophysical perspective, it is noteworthy to mention that the strong Feller-Irreducibility approach allows for dealing with a broad class of truncations of fluid dynamics models that would be perturbed by noise, possibly degenerate. For instance, in the case of truncations of 2D or 3D Navier-Stokes equations, the strong Feller-Irreducibility approach has been shown to be applicable even for an additive noise that forces only very few modes [AS05, Rom04]. The delicate point of the analysis is the verification of the controllability (and thus irreducibility) of the associated control system, by techniques typically adapted from [JK85] or rooted in chronological calculus as in [AS05]. Whatever the approach, the analysis requires the appropriate translation into geometrical terms of the cascade of energy in which the nonlinear terms transmit the forcing from the few modes to all the others [Rom05].

A.3 Markov semigroups and mixing

We recall here standard results about Markov semigroups. It states that any Markov semigroup that is strong Feller and irreducible and for which an invariant measure exists (which is thus unique) is not only ergodic but also strongly mixing for the total variation norm of measures. Given two probability measures μ_1 and μ_2 on \mathcal{X} , we recall that the latter is defined as [Hai09, Eq. (3.1)]

$$\text{TV}(\mu_1, \mu_2) = \sup_{\substack{g \in \mathcal{B}_b(\mathcal{X}) \\ \|g\|_\infty \leq 1}} \left| \int g \, d\mu_1 - \int g \, d\mu_2 \right|, \quad (\text{A.6})$$

where $\mathcal{B}_b(\mathcal{X})$ denotes the set of Borel measurable and bounded functions on \mathcal{X} .

Theorem 4 *Let μ be an invariant measure of a Markov semigroup $(P_t)_{t \geq 0}$. For any $p \geq 1$ and $t \geq 0$, P_t is extendable to a linear bounded operator on $L_\mu^p(\mathcal{X})$ still denoted by P_t . Moreover*

- (i) $\|P_t\|_{\mathcal{L}(L_\mu^p(\mathcal{X}))} \leq 1$
- (ii) P_t is strongly continuous semigroup in $L_\mu^p(\mathcal{X})$.

If furthermore $(P_t)_{t \geq 0}$ is strong Feller and irreducible, then μ is ergodic and for any x in \mathcal{X} and g in L^1_μ

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T g(X_\tau^x) d\tau = \int_{\mathcal{X}} g(x) d\mu, \quad \mathbb{P}\text{-a.s.}, \quad (\text{A.7})$$

where X_t^x denotes the stochastic process solving the SDE associated with P_t .

In this case, the invariant measure μ is also **strongly mixing** in the sense that for any measure ν on \mathcal{X} , we have:

$$TV(\mathcal{L}_t \nu, \mu) \xrightarrow{t \rightarrow \infty} 0. \quad (\text{A.8})$$

For the definition of a strongly continuous semigroup also known as C_0 -semigroup we refer to [EN00, p. 36]; see also [CK17]. For an introduction to semigroup theory we refer to [vN12, EN06].

Proof We prove first (i). The proof is standard and can be found e.g. in [GZ03, Prop. 1.14] but is reproduced here for the reader's convenience. Let g be in $\mathcal{C}_b(\mathcal{X})$. By the Hölder inequality, we have

$$|P_t g(x)|^p \leq P_t(|g|^p)(x). \quad (\text{A.9})$$

If we now integrate both sides of this inequality with respect to μ , we obtain

$$\int_{\mathcal{X}} |P_t g(x)|^p \mu(dx) \leq \int_{\mathcal{X}} P_t(|g|^p)(x) \mu(dx) = \int_{\mathcal{X}} |g|^p(x) \mu(dx), \quad (\text{A.10})$$

the latter equality resulting from the invariance of μ . Since $\mathcal{C}_b(\mathcal{X})$ is dense in $L^p_\mu(\mathcal{X})$, the inequality (A.10) can be extended to any function in $L^p_\mu(\mathcal{X})$, and thus $(P_t)_{t \geq 0}$ can be uniquely extended to a contraction semigroup in $L^p_\mu(\mathcal{X})$, and property (i) is proved.

Let us show now that $(P_t)_{t \geq 0}$ is strongly continuous in $L^p_\mu(\mathcal{X})$. Since $(P_t)_{t \geq 0}$ is a Markov semigroup, for any g in $\mathcal{C}_b(\mathcal{X})$ and x in \mathcal{X} , we have that the mapping $t \mapsto P_t g(x)$ is continuous. Therefore by the dominated convergence theorem

$$\lim_{t \rightarrow 0} P_t g = g \text{ in } L^p_\mu(\mathcal{X}). \quad (\text{A.11})$$

The density of $\mathcal{C}_b(\mathcal{X})$ in $L^p_\mu(\mathcal{X})$ allows us to conclude that this convergence holds when g is in $L^p_\mu(\mathcal{X})$.

The ergodicity of μ results from the aforementioned Doob's theorem. The time-average property (A.7) and the mixing property (A.8) can be obtained as a consequence of e.g. [Sei97, Cor. 2.3]; see also [Ste94, Cor. 1].

A.4 Generator of a Markov semigroup

Recall that the generator A of any strongly continuous semigroup $(T(t))_{t \geq 0}$ on a Hilbert space \mathcal{H} is defined as the operator $A : D(A) \subset \mathcal{H} \rightarrow \mathcal{H}$, such that

$$A\varphi = \lim_{t \rightarrow 0^+} \frac{1}{t} (T(t)\varphi - \varphi), \quad (\text{A.12})$$

defined for every φ in the domain

$$D(A) = \{\varphi \in \mathcal{H} \mid \lim_{t \rightarrow 0^+} \frac{1}{t} (T(t)\varphi - \varphi) \text{ exists}\}. \quad (\text{A.13})$$

As any generator of a contraction semigroup, given an invariant measure μ , the generator K of the contraction semigroup $(P_t)_{t \geq 0}$ in L^2_μ (Theorem 4-(i)) is dissipative, which is equivalent to say, since L^2_μ is a Hilbert space, that

$$\operatorname{Re} \langle Kf, f \rangle_{L^2_\mu} \leq 0, \quad \forall f \in D(K), \quad (\text{A.14})$$

where $D(K)$ denotes the domain of K ; see e.g. [EN00, Prop. II.3.23]. The domain $D(K)$ is furthermore dense in L^2_μ and K is a closed operator; see [Paz83, Cor. 2.5 p. 5]. The isolated part of the spectrum of K provides the Ruelle-Pollicott resonances; see Sec. 2.2.

A.5 Return to equilibrium and spectral gap

We present here some useful results concerning (i) the exponential return to equilibrium for strong Feller and irreducible Markov semigroups, and (ii) spectral gap in the spectrum of the Markov semigroup generator K ; see Theorems 5 and 6 below. Theorem 5 deals with semigroups that become *quasi-compact* after a finite time, and Theorem 6 addresses the exponential L^2 -convergence and lower bound of the spectral gap. For Theorem 5, the approach is based on Lyapunov functions such as formulated in [RB06]. We propose a slightly different presentation for which we provide the main elements of the proof. We refer to [DFG09] for an efficient (and beautiful) generalization of such Lyapunov-type criteria allowing for *sub-exponential convergence* towards the equilibrium.

Recall that the *essential spectral radius* $\mathbf{r}_{ess}(T)$ of a linear bounded operator T on a Banach space \mathcal{X} satisfies [EN00, p. 249] the Hadamard formula

$$\mathbf{r}_{ess}(T) = \lim_{n \rightarrow \infty} \|T^n\|_{ess}^{1/n}, \quad (\text{A.15})$$

where

$$\|T\|_{ess} = \inf \left\{ \|T - \mathcal{C}\|_{\mathcal{L}(\mathcal{X})} : \mathcal{C} \text{ is a linear and compact operator of } \mathcal{X} \right\}. \quad (\text{A.16})$$

We have then the following convergence result.

Theorem 5 *Let $\mathcal{P} = (P_t)_{t \geq 0}$ be a strong Feller and irreducible Markov semigroup generated by an SDE given by Eq. (2.1) for which F and G are locally Lipschitz. Assume that there exists a Lyapunov function⁷ U and a compact set \mathfrak{A} for which there exist $a > 0$, $0 < \kappa < 1$ and $b < \infty$, such that*

$$\mathcal{K}U \leq aU, \quad (\text{A.17a})$$

$$P_{t_0}U \leq \kappa U + b\mathbb{1}_{\mathfrak{A}}, \text{ for some } t_0 > 0, \quad (\text{A.17b})$$

where \mathcal{K} is the Kolmogorov differential operator generating the Markov process associated with \mathcal{P} . Then for all $t > t_0$, P_t becomes quasi-compact, i.e.

$$\mathbf{r}_{ess}(P_t) \leq \kappa, \quad (\text{A.18})$$

where the essential spectral radius is taken for P_t as acting on $\mathcal{X} = \mathcal{F}_U$ given by

$$\mathcal{F}_U = \{f : \mathbb{R}^N \rightarrow \mathbb{R} \mid f \text{ Borel measurable and } \|f\|_U < \infty\}, \quad (\text{A.19})$$

and endowed with the norm

$$\|f\|_U = \sup_{x \in \mathbb{R}^N} \frac{|f(x)|}{U(x)}. \quad (\text{A.20})$$

Furthermore $(P_t)_{t \geq 0}$ has a unique invariant measure μ , and the inequality (A.18) ensures that there exist $C > 0$ and $\lambda > 0$ such that for all f in \mathcal{F}_U ,

$$\left| P_t f(x) - \int f d\mu \right| \leq C e^{-\lambda t} U(x), \quad t > t_0, \quad \forall x \in \mathbb{R}^N. \quad (\text{A.21})$$

The proof of this result is found in Appendix A.6.

Remark 5 *The assumption (A.17b) is sometimes verified from moment estimates in practice. For instance if there exist $k_0 > 0$ and $k_1 > 0$ such that*

$$\mathbb{E}|X_t^x| \leq k_0 e^{-k_1 t} |x| + c, \quad t \geq 0, \quad (\text{A.22})$$

then for any $t \geq -\frac{1}{k_1} \log(\frac{1}{4k_0})$, we have $\mathbb{E}(|X_t^x| + 1) \leq \frac{1}{2}(|x| + 1) - \frac{1}{4}|x| + c + \frac{1}{2}$, which leads to

$$\mathbb{E}(|X_t^x| + 1) \leq \frac{1}{2}(|x| + 1) + (c + \frac{1}{2})\mathbb{1}_{B_r}, \quad (\text{A.23})$$

for all $r > 4(c + \frac{1}{2})$, and thus (A.17b) holds with $U(x) = |x| + 1$.

More generally, if

$$\mathcal{K}U \leq -\alpha U + \beta, \quad \text{with } \alpha > 0, \text{ and } 0 \leq \beta < \infty, \quad (\text{A.24})$$

then $\frac{d}{dt} P_t U(x) = P_t \mathcal{K}U(x) \leq -\alpha P_t U(x) + \beta$, leading to

$$\mathbb{E}[U(X_t^x)] \leq U(x) e^{-\beta t} + \frac{\beta}{\alpha} (1 - e^{-\alpha t}), \quad t > 0, \quad (\text{A.25})$$

and similarly (A.17b) holds. In addition, (A.24) implies (A.17a). Note that (A.24) and (A.25) are quite standard; see e.g. [DKZ12, Lemma 2.11].

Finally, note also that finding a Lyapunov function may be easier than proving inequalities of the form (A.22). For instance, if there is a Lyapunov function which grows polynomially like $\|p\|^q$, then one knows that the process has moments of order q ; see [MT93a, MT93b].

Finally, lower bounds of the spectral gap in L_μ^2 may be derived for a broad class of SDEs. Recall that the generator K has a spectral gap in L_μ^2 if there exists $\delta > 0$ such that

$$\sigma(K) \cap \{\lambda : \text{Re}(\lambda) > -\delta\} = \{0\}. \quad (\text{A.26})$$

The largest $\delta > 0$ with this property is denoted by $\text{gap}(K)$, namely

$$\text{gap}(K) = \sup\{\delta > 0 \text{ s.t. (A.26) holds}\}. \quad (\text{A.27})$$

The following result is a consequence in finite dimension of more general convergence results [GM05, Theorems 2.5 and 2.6]. Since $(P_t)_{t \geq 0}$ is a C_0 -semigroup in L_μ^2 , the theory of asymptotic behavior of a semigroup with a strictly dominant, algebraically simple eigenvalue (e.g. [vN12, Thm. 3.6.2]) implies the spectral gap property stated in the following.

⁷ Recall that a C^2 function U is called a Lyapunov function $U(x) \geq 1$ and $\lim_{|x| \rightarrow \infty} U(x) = \infty$, ensuring thus that the level sets $\{U \leq \alpha\}$ are compact.

Theorem 6 Assume that $(P_t)_{t \geq 0}$ is strong Feller and irreducible. Assume furthermore that the following ultimate bound holds for the associated stochastic process X_t^x , i.e. there exist $c, k, \alpha > 0$ such that

$$\mathbb{E}|X_t^x|^2 < k|x|^2 e^{-\alpha t} + c, \quad t \geq 0, \quad x \in \mathbb{R}^N. \quad (\text{A.28})$$

Then there exists a unique invariant measure μ for which the U -uniform ergodicity (A.21) holds with $U(x) = 1 + |x|^2$, as well as the following exponential L^2 -convergence

$$\|P_t \varphi - \int \varphi d\mu\|_{L_\mu^2} \leq C e^{-\lambda t} \|\varphi\|_{L_\mu^2}, \quad t \geq 0, \quad \varphi \in L_\mu^2, \quad (\text{A.29})$$

with C and λ positive constants independent of φ ; the latter rate of convergence being the same as that of (A.21). Furthermore, one has the following lower bound for the L_μ^2 -spectrum of the generator K :

$$0 < \lambda \leq \text{gap}(K). \quad (\text{A.30})$$

We will see in Part II [TCND19a] of this three-part article that Theorem 6 has important practical consequences. In particular it shows for a broad class of controllable ODEs, perturbed by a white noise process for which the Kolmogorov operator is hypoelliptic, that an L_μ^2 -spectral gap is naturally induced by the noise whereas in absence of the latter the gap may be zero, leading thus to a form of mixing enhancement by the noise. We finally mention [HSV14] for other conditions, ensuring an L_μ^2 -gap based on spectral gaps in Wasserstein distances, verifiable in practice by following the approach of [HMS11]. We show in the next section that Wasserstein distances are also suitable to measure the decay of correlations.

A.6 Proof of Theorem 5

Proof It is standard from the theory of Lyapunov functions that the existence of a unique invariant measure μ is ensured by the condition (A.17a) together with the irreducibility and strong Feller properties. The rest of the proof is thus concerned with (A.18) and the exponential convergence (A.21).

Step 1. First, note that the Itô formula gives

$$dU = \mathcal{K}U dt + \text{“Martingale”}, \quad (\text{A.31})$$

which leads (since $\mathcal{K}U \leq aU$) to

$$\mathbb{E}(U(x(t; x))) = P_t U(x) \leq e^{at} U(x), \quad (\text{A.32})$$

and therefore P_t is extendable to a linear operator on \mathcal{F}_U (defined in (A.19)) with norm $\|P_t\| \leq e^{at}$.

The second inequality in (A.17) ensures that for any $t > t_0$,

$$P_t U(x) \leq \kappa U(x) + b, \quad \forall x \in \mathbb{R}^N. \quad (\text{A.33})$$

By definition, a Markov semigroup is monotone, thus one may iterate (A.33) to obtain (by using $P_t \mathbb{1}_{\mathbb{R}^N} = \mathbb{1}_{\mathbb{R}^N}$),

$$P_{nt} U(x) \leq \kappa^n U(x) + \frac{b}{1 - \kappa}, \quad n \geq 1. \quad (\text{A.34})$$

Consider now an arbitrary compact set \mathfrak{B} in \mathbb{R}^N and f in \mathcal{F}_U , we have the bound

$$\begin{aligned} |P_{nt} f(x) - \mathbb{1}_{\mathfrak{B}}(x) P_{nt} \mathbb{1}_{\mathfrak{B}} f(x)| &\leq U(x) \sup_{y \in \mathcal{H} \setminus \mathfrak{B}} \frac{|P_{nt} f(y)|}{U(y)}, \\ &\leq U(x) \|f\|_U \sup_{y \in \mathcal{H} \setminus \mathfrak{B}} \frac{|P_{nt} U(y)|}{U(y)}, \end{aligned} \quad (\text{A.35})$$

where we have used the basic inequality (A.9) (with $p = 1$). This last inequality with (A.34) leads to

$$|P_{nt} f(x) - \mathbb{1}_{\mathfrak{B}}(x) P_{nt} \mathbb{1}_{\mathfrak{B}} f(x)| \leq U(x) \|f\|_U \left(\kappa^n + \frac{b}{1 - \kappa} \sup_{y \in \mathcal{H} \setminus \mathfrak{B}} \frac{1}{U(y)} \right). \quad (\text{A.36})$$

Since $\lim_{|x| \rightarrow \infty} U(x) = \infty$, given $\epsilon > 0$ and $n > 1$ one may thus choose a compact set \mathfrak{B}_n such that

$$\|P_{nt} f - \mathbb{1}_{\mathfrak{B}_n} P_{nt} \mathbb{1}_{\mathfrak{B}_n} f\|_U \leq \|f\|_U (\kappa + \epsilon)^n, \quad (\text{A.37})$$

which leads to

$$\|P_{nt} - \mathbb{1}_{\mathfrak{B}_n} P_{nt} \mathbb{1}_{\mathfrak{B}_n}\|_{\mathcal{L}(\mathcal{F}_U)} \leq (\kappa + \epsilon)^n. \quad (\text{A.38})$$

Step 2. Let us show that the linear operator

$$\Lambda = \mathbb{1}_{\mathfrak{B}} P_t \mathbb{1}_{\mathfrak{B}} : \mathcal{F}_U \longrightarrow \mathcal{F}_U, \quad (\text{A.39})$$

is compact for any compact set \mathfrak{B} of \mathbb{R}^N . This is equivalent to showing that for any sequence g_k in \mathcal{F}_U such that $\|g_k\|_U \leq 1$, one can extract a subsequence such that Ag_k is convergent in \mathcal{F}_U . Since P_t is strongly Feller and $\mathbb{1}_{\mathfrak{B}}g_k$ is bounded for each k , then $P_t\mathbb{1}_{\mathfrak{B}}g_k$ belongs to $\mathcal{C}_b(\mathfrak{B})$, by definition. Thus the sequence (Ag_k) lives in $\mathcal{C}(\mathfrak{B})$.

We have

$$|Ag_k(x)| \leq \|g_k\|_U P_t U(x) \leq \kappa U(x) + b \leq \kappa \sup_{y \in \mathfrak{B}} U(y) + b, \quad x \in \mathfrak{B}, \quad (\text{A.40})$$

which shows that $\{Ag_k\}$ is equibounded.

Furthermore, since P_t is strong Feller, it has a smooth kernel⁸ and we have for all $x, x' \in \mathfrak{B}$

$$\begin{aligned} |Ag_k(x) - Ag_k(x')| &\leq \int_{y \in \mathfrak{B}} |\mathfrak{p}_t(x, y) - \mathfrak{p}_t(x', y)| |f(y)| dy, \\ &\leq |x - x'| \sup_{u, v \in \mathfrak{B}} |\partial_u \mathfrak{p}_t(u, v)| \|g_k\|_U \int_{\mathfrak{B}} U(y) dy, \end{aligned} \quad (\text{A.41})$$

which shows that $\{Ag_k\}$ is equicontinuous.

Thus, the Ascoli-Arzelà theorem [Yos95, p. 85] applies and guarantees that a subsequence from Ag_k converges in $\mathcal{C}(\mathfrak{B})$ to g . Now since $U \geq 1$, the same extraction from Ag_k converges to $g\mathbb{1}_{\mathfrak{B}}$ in \mathcal{F}_U . We conclude that $\mathbb{1}_{\mathfrak{B}}P_t\mathbb{1}_{\mathfrak{B}}$ is a compact mapping for any compact set \mathfrak{B} of \mathbb{R}^N .

Step 3. Let \mathfrak{B}_n be a sequence of compact sets satisfying (A.37), and let us consider the compact operators (from Step 2) \mathfrak{C}_n defined by $\mathbb{1}_{\mathfrak{B}_n}P_{nt}\mathbb{1}_{\mathfrak{B}_n}$. We have then

$$\begin{aligned} \|P_{nt}\|_{ess} = \inf \left\{ \|P_{nt} - \mathfrak{C}\|_{\mathcal{L}(\mathcal{F}_U)} : \mathfrak{C} \text{ is a linear and compact operator of } \mathcal{F}_U \right\} &\leq \|P_{nt} - \mathfrak{C}_n\| \\ &\leq (\kappa + \epsilon)^n. \end{aligned} \quad (\text{A.42})$$

By applying to P_t the Hadamard formula recalled in (A.15), we have thus for $t > t_0$

$$\mathbf{r}_{ess}(P_t) = \lim_{n \rightarrow \infty} \|P_{nt}\|_{ess}^{1/n} \leq \kappa + \epsilon, \quad (\text{A.43})$$

for all $\epsilon > 0$, and we deduce (A.18).

The exponential convergence is then ensured by showing that there is no other eigenvalue than 1 on the unit disk (or outside the unit disk) and that 1 is a simple eigenvalue; see [RB06].

References

- AK87. L. Arnold and W. Kliemann, *On unique ergodicity for degenerate diffusions*, Stochastics: An international Journal of Probability and Stochastic Processes **21** (1987), no. 1, 41–61.
- Arn98. L. Arnold, *Random Dynamical Systems*, Springer Monographs in Mathematics, Springer, Berlin, 1998.
- AS05. A. A. Agrachev and A. V. Sarychev, *Navier–Stokes equations: controllability by means of low modes forcing*, Journal of Mathematical Fluid Mechanics **7** (2005), no. 1, 108–152.
- Bal00. V. Baladi, *Positive Transfer Operators and Decay of Correlations*, World Scientific, Singapore, 2000.
- BGL13. D. Bakry, I. Gentil, and M. Ledoux, *Analysis and Geometry of Markov Diffusion Operators*, vol. 348, Springer Science & Business Media, 2013.
- BHJK15. A. Bittracher, C. Hartmann, O. Junge, and P. Koltai, *Pseudo generators for under-resolved molecular dynamics*, The European Physical Journal Special Topics **224** (2015), no. 12, 2463–2490.
- BKJ15. A. Bittracher, P. Koltai, and O. Junge, *Pseudo generators of spatial transfer operators*, SIAM Journal on Applied Dynamical Systems **14** (2015), no. 3, 1478–1517.
- BL07. O. Butterley and C. Liverani, *Smooth Anosov flows: Correlation spectra and stability*, Journal of Modern Dynamics **1** (2007), no. 2, 301–322.
- Cer01. S. Cerrai, *Second-Order PDE's in Finite and Infinite Dimension: A Probabilistic Approach*, vol. 1762, Springer Science & Business Media, 2001.
- CK17. M. D. Chekroun and D. Kondrashov, *Data-adaptive harmonic spectra and multilayer Stuart-Landau models*, Chaos **27** (2017), no. 9, 093110.
- CLM17. M.D. Chekroun, H. Liu, and J.C. McWilliams, *The emergence of fast oscillations in a reduced primitive equation model and its implications for closure theories*, Computers & Fluids **151** (2017), 3–22.
- CLM19. M. D. Chekroun, H. Liu, and J. C. McWilliams, *Variational approach to closure of nonlinear dynamical systems: Autonomous case*, Journal of Stat. Physics. Accepted. (2019).
- CLPR19. M. D. Chekroun, J. S. W. Lamb, C. J. Pangerl, and M. Rasmussen, *A Girsanov approach to slow parameterizing manifolds in the presence of noise*, arXiv preprint, <https://arxiv.org/abs/1903.08598> (2019).
- CLW15a. M. D. Chekroun, H. Liu, and S. Wang, *Approximation of Stochastic Invariant Manifolds: Stochastic Manifolds for Nonlinear SPDEs I*, Springer Briefs in Mathematics, Springer, New York, 2015.

⁸ A probability kernel \mathfrak{T}_t allows for representing the Markov semigroup P_t as $P_t f(x) = \int \mathfrak{T}_t(x, dy) f(y)$; e.g. [BGL13, Prop. 1.2.3]. Having a smooth kernel means that $\mathfrak{T}_t(x, dy) = \mathfrak{p}_t(x, y) dy$ with \mathfrak{p}_t infinitely differentiable, i.e. smooth.

- CLW15b. ———, *Stochastic Parameterizing Manifolds and Non-Markovian Reduced Equations: Stochastic Manifolds for Nonlinear SPDEs II*, Springer Briefs in Mathematics, Springer, New York, 2015.
- CNK⁺14. M. D. Chekroun, J.D. Neelin, D. Kondrashov, J.C. McWilliams, and M. Ghil, *Rough parameter dependence in climate models: The role of Ruelle-Pollicott resonances*, Proc. Natl. Acad. Sci **111** (2014), no. 5, 1684–1690.
- CSG11. M. D. Chekroun, E. Simonnet, and M. Ghil, *Stochastic climate dynamics: Random attractors and time-dependent invariant measures*, Physica D **240** (2011), no. 21, 1685–1700.
- CVE06. DT Crommelin and Eric Vanden-Eijnden, *Fitting time series by continuous-time markov chains: A quadratic programming approach*, Journal of Computational Physics **217** (2006), no. 2, 782–805.
- Dac04. B. Dacorogna, *Introduction to the Calculus of Variations*, vol. 13, World Scientific, 2004.
- Dav96. E. B. Davies, *Spectral Theory and Differential Operators*, vol. 42, Cambridge University Press, 1996.
- Dav07. ———, *Linear Operators and their Spectra*, vol. 106, Cambridge University Press, 2007.
- DFG09. R. Douc, G. Fort, and A. Guillin, *Subgeometric rates of convergence of f-ergodic strong Markov processes*, Stochastic Processes and Their Applications **119** (2009), no. 3, 897–923.
- DKZ12. F. Dragoni, V. Kontis, and B. Zegarliński, *Ergodicity of Markov semigroups with Hörmander type generators in infinite dimensions*, Potential Analysis (2012), 1–29.
- DM78. C. Dellacherie and P.-A. Meyer, *Probabilities and Potential*, vol. 29, North-Holland Publishing Co., 1978.
- Doo48. J. L. Doob, *Asymptotic properties of Markoff transition probabilities*, Transactions of the American Mathematical Society **63** (1948), no. 3, 393–421.
- DPZ96. G. Da Prato and J Zabczyk, *Ergodicity for Infinite Dimensional Systems*, Cambridge University Press, Cambridge, 1996.
- DZ15. S. Dyatlov and M. Zworski, *Stochastic stability of Pollicott–Ruelle resonances*, Nonlinearity **28** (2015), no. 10, 3511.
- EH03. J.-P. Eckmann and M. Hairer, *Spectral properties of hypoelliptic operators*, Communications in Mathematical Physics **235** (2003), no. 2, 233–253.
- EN00. K.-J. Engel and R. Nagel, *One-parameter Semigroups for Linear Evolution Equations*, vol. 194, Springer Science & Business Media, 2000.
- EN06. ———, *A short course on operator semigroups*, Springer Science & Business Media, 2006.
- FBS09. E. Forgoston, L. Billings, and I. B. Schwartz, *Accurate noise projection for reduced stochastic epidemic models*, Chaos: An Interdisciplinary Journal of Nonlinear Science **19** (2009), no. 4, 043110.
- Fen79. N. Fenichel, *Geometric singular perturbation theory for ordinary differential equations*, Journal of Differential Equations **31** (1979), no. 1, 53–98.
- FGP10. F. Flandoli, M. Gubinelli, and E. Priola, *Flow of diffeomorphisms for SDEs with unbounded Holder continuous drift*, Bulletin des sciences mathématiques **134** (2010), no. 4, 405–422.
- FM71. N. Fenichel and J. K. Moser, *Persistence and smoothness of invariant manifolds for flows*, Indiana University Mathematics Journal **21** (1971), no. 3, 193–226.
- Fro97. G. Froyland, *Computer-assisted bounds for the rate of decay of correlations*, Communications in Mathematical Physics **189** (1997), no. 1, 237–257.
- FS09. E. Forgoston and I. B. Schwartz, *Escape rates in a stochastic environment with multiple scales*, SIAM J. Appl. Dyn. Syst. **8** (2009), no. 3, 1190–1217.
- GAD⁺02. M. Ghil, M. R. Allen, M.D. Dettinger, K. Ide, D. Kondrashov, M. E. Mann, A. W. Robertson, A. Saunders, Y. Tian, F. Varadi, et al., *Advanced spectral methods for climatic time series*, Rev. Geophys **40** (2002), no. 1, 1003.
- Gas02. P. Gaspard, *Trace formula for noisy flows*, Journal of Statistical Physics **106** (2002), no. 1-2, 57–96.
- GCF17. G. A. Gottwald, D. T. Crommelin, and C. L. E. Franzke, *Stochastic climate theory*, Nonlinear and Stochastic Climate Dynamics (C. L. E. Franzke and T. J. O’Kane, eds.), Cambridge University Press, 2017, pp. 209–240.
- GKWG09. M. Gruberbauer, T. Kallinger, W. W. Weiss, and D.B. Guenther, *On the detection of Lorentzian profiles in a power spectrum: a Bayesian approach using ignorance priors*, Astronomy & Astrophysics **506** (2009), no. 2, 1043–1053.
- GM05. B. Goldys and B. Maslowski, *Exponential ergodicity for stochastic reaction-diffusion equations*, Stochastic Partial Differential Equations and Applications-VII, Lect. Notes Pure Appl. Math., 245, Chapman Hall/CRC, Boca Raton, FL, 2006, 2005, pp. 115–131.
- GRS99. H. Ganidis, B. Roynette, and F. Simonot, *Convergence rate of some semi-groups to their invariant probability*, Stochastic processes and their applications **79** (1999), no. 2, 243–263.
- GZ03. A. Guionnet and B. Zegarliński, *Lectures on logarithmic Sobolev inequalities*, Séminaire de Probabilités XXXVI, Springer, 2003, pp. 1–134.
- Hai09. M. Hairer, *An introduction to stochastic PDEs*, arXiv preprint arXiv:0907.4178 (2009).
- HMS11. M. Hairer, J. C. Mattingly, and M. Scheutzow, *Asymptotic coupling and a general form of Harris theorem with applications to stochastic delay equations*, Probability Theory and Related Fields **149** (2011), no. 1-2, 223–259.
- HN04. F. Hérau and F. Nier, *Isotropic hypoellipticity and trend to equilibrium for the Fokker-Planck equation with a high-degree potential*, Archive for Rational Mechanics and Analysis **171** (2004), no. 2, 151–218.
- Hör67. L. Hörmander, *Hypoelliptic second order differential equations*, Acta Mathematica **119** (1967), no. 1, 147–171.
- HSV14. M. Hairer, A. M. Stuart, and S. J. Vollmer, *Spectral gaps for a Metropolis–Hastings algorithm in infinite dimensions*, The Annals of Applied Probability **24** (2014), no. 6, 2455–2490.
- IW14. N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, vol. 24, Elsevier, 2014.
- JK85. V. Jurdjevic and I. Kupka, *Polynomial control systems*, Mathematische Annalen **272** (1985), no. 3, 361–368.
- Jon95. C. K. R. T. Jones, *Geometric Singular Perturbation Theory*, Springer, Berlin Heidelberg, 1995.
- Kha60. R. Z. Khasminskii, *Ergodic properties of recurrent diffusion processes and stabilization of the solution to the Cauchy problem for parabolic equations*, Theory of Probability & Its Applications **5** (1960), no. 2, 179–196.

- Kli87. W. Kliemann, *Recurrence and invariant measures for degenerate diffusions*, The Annals of Probability (1987), 690–707.
- KMH⁺10. T. Kallinger, B. Mosser, S. Hekker, D. Huber, D. Stello, S. Mathur, S. Basu, T. R. Bedding, W. J. Chaplin, J. De Ridder, et al., *Asteroseismology of red giants from the first four months of Kepler data: Fundamental stellar parameters*, Astronomy & Astrophysics **522** (2010), A1.
- Kue15. C. Kuehn, *Multiple Time Scale Dynamics*, Applied Mathematical Sciences, vol. 191, Springer, 2015.
- LB06. L. Lorenzi and M. Bertoldi, *Analytical Methods for Markov Semigroups*, CRC Press, Taylor & Francis Group, 2006.
- LL10. F. Legoll and T. Lelièvre, *Effective dynamics using conditional expectations*, Nonlinearity **23** (2010), no. 9, 2131.
- LSY97. Richard B. Lehoucq, D. C. Sorensen, and C. Yang, *ARPACK Users' Guide: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, 1997, pp. xv + 137.
- LVE14. J. Lu and E. Vanden-Eijnden, *Exact dynamical coarse-graining without time-scale separation*, The Journal of Chemical Physics **141** (2014), no. 4, 044109.
- MPP02. G. Metafune, D. Pallara, and E. Priola, *Spectrum of Ornstein-Uhlenbeck operators in L^p spaces with respect to invariant measures*, Journal of Functional Analysis **196** (2002), no. 1, 40–60.
- MPW02. G. Metafune, D. Pallara, and M. Wacker, *Compactness properties of Feller semigroups*, Studia Mathematica **153** (2002), no. 2, 179–206.
- MT93a. S. P. Meyn and R. L. Tweedie, *Stability of Markovian processes II: Continuous-time processes and sampled chains*, Advances in Applied Probability (1993), 487–517.
- MT93b. ———, *Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes*, Advances in Applied Probability (1993), 518–548.
- MTVE01. A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden, *A mathematical framework for stochastic climate models*, Commun. Pure Appl. Math. **54** (2001), 891–974.
- NAM⁺03. B. R. Noack, K. Afanasiev, M. Morzyński, G. Tadmor, and F. Thiele, *A hierarchy of low-dimensional models for the transient and post-transient cylinder wake*, Journal of Fluid Mechanics **497** (2003), 335–363.
- Nor86. J. Norris, *Simplified Malliavin calculus*, Séminaire de Probabilités XX 1984/85, Springer, 1986, pp. 101–130.
- NS13. K. Nipp and D. Stoffer, *Invariant Manifolds in Discrete and Continuous Dynamical Systems*, vol. 21, European Mathematical Society, 2013.
- OPPS12. M. Ottobre, G.A. Pavliotis, and K. Pravda-Starov, *Exponential return to equilibrium for hypoelliptic quadratic systems*, Journal of Functional Analysis **262** (2012), no. 9, 4000–4039.
- Paz83. A Pazy, *Semigroups of Linear Operators and Application to Partial Differential Equations*, Springer, New York Berlin Heidelberg Tokyo, 1983.
- Pol86. M. Pollicott, *Meromorphic extensions of generalised zeta functions*, Inventiones Mathematicae **85** (1986), no. 1, 147–164.
- RB06. L. Rey-Bellet, *Ergodic properties of Markov processes*, Open Quantum Systems II, Springer, 2006, pp. 1–39.
- Rom04. M. Romito, *Ergodicity of the finite dimensional approximation of the 3D Navier–Stokes equations forced by a degenerate noise*, Journal of Statistical Physics **114** (2004), no. 1-2, 155–177.
- Rom05. ———, *A geometric cascade for the spectral approximation of the Navier-Stokes equations*, Probability and Partial Differential Equations in Modern Applied Mathematics, Springer, 2005, pp. 197–212.
- Rue86. D. Ruelle, *Locating resonances for axiom a dynamical systems*, Journal of Statistical Physics **44** (1986), no. 3-4, 281–292.
- RW00. L. G. Rogers and D. Williams, *Diffusions, Markov Processes and Martingales: Volume 1, Foundations*, vol. 2, Cambridge university press, 2000.
- Sei97. J. Seidler, *Ergodic behaviour of stochastic parabolic equations*, Czechoslovak Mathematical Journal **47** (1997), no. 2, 277–316.
- SFHD99. Ch. Schütte, A. Fischer, W. Huisinga, and P. Deuffhard, *A direct approach to conformational dynamics based on hybrid Monte Carlo*, Journal of Computational Physics **151** (1999), no. 1, 146–168.
- SH99. Ch. Schütte and W. Huisinga, *On conformational dynamics induced by Langevin processes*, Proceedings of the International Conference on Differential Equations, vol. 1, World Scientific, 1999, p. 7.
- SHD01. Ch. Schütte, W. Huisinga, and P. Deuffhard, *Transfer operator approach to conformational dynamics in biomolecular systems*, Ergodic theory, analysis, and efficient simulation of dynamical systems, Springer, 2001, pp. 191–223.
- Ste94. L. Stettner, *Remarks on Ergodic Conditions for Markov Processes on Polish Spaces*, Bulletin of the Polish Academy of Sciences-Mathematics **42** (1994), no. 2, 103–114.
- SV71. D.W. Stroock and S.R. Varadhan, *On the support of diffusion processes with applications to the strong maximum principle*, Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Univ. California, Berkeley, Calif., 1970/1971, vol. 3, 1971, pp. 333–359.
- Tam95. C.K.W. Tam, *Supersonic jet noise*, Annual Review of Fluid Mechanics **27** (1995), no. 1, 17–43.
- TCND19a. A. Tantet, M. D. Chekroun, J. D. Neelin, and H. A. Dijkstra, *Ruelle-Pollicott Resonances of Stochastic Systems in Reduced State Space. Part II: Stochastic Hopf Bifurcation*, J. Stat. Phys. (Submitted) (2019).
- TCND19b. ———, *Ruelle-Pollicott Resonances of Stochastic Systems in Reduced State Space. Part III: Application to El Niño-Southern Oscillation*, J. Stat. Phys. (Submitted) (2019).
- TvdBD15. A. Tantet, F. R. van der Burgt, and H. A. Dijkstra, *An early warning indicator for atmospheric blocking events using transfer operators*, Chaos: An Interdisciplinary Journal of Nonlinear Science **25** (2015), no. 3, 036406.
- vN12. J. van Neerven, *The Asymptotic Behaviour of Semigroups of Linear Operators*, vol. 88, Birkhäuser, 2012.
- WL12. J. Wouters and V. Lucarini, *Disentangling multi-level systems: averaging, correlations and memory*, J. Stat. Mech. (2012), P03003.

-
- WL13. Jeroen Wouters and Valerio Lucarini, *Multi-level dynamical systems: connecting the Ruelle response theory and the Mori-Zwanzig approach*, J. Stat. Phys. **151** (2013), no. 5, 1–14.
- WVE04. E Weinan and E. Vanden-Eijnden, *Metastability, conformation dynamics, and transition pathways in complex systems*, Multiscale modelling and simulation, Springer, 2004, pp. 35–68.
- Yos95. K. Yosida, *Functional analysis. Reprint of the sixth (1980) edition. Classics in Mathematics*, Springer-Verlag, Berlin **11** (1995), 501.
- Zwo17. Maciej Zworski, *Mathematical study of scattering resonances*, Bulletin of Mathematical Sciences **7** (2017), no. 1, 1–85.