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

Mickaël D. Chekroun, Alexis Tantet, Henk Dijkstra, J. David D Neelin

Noname manuscript No.
(will be inserted by the editor)

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

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

March 16, 2020

Abstract A theory of Ruelle-Pollicott (RP) resonances for stochastic differential systems is presented. 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. They inform us about the spectral elements of some coarse-grained version of the SDE generator. When the time-lag at which the transitions are collected from partial observations in V, is either sufficiently small or large, it is shown that the reduced RP resonances approximate the (weak) 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 [TCND20] and Part III [TCND19], 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 provide. 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 manifestation in terms of a geometric organization of the RP resonances along discrete parabolas in the left half plane. Such geometric features formed by (reduced) RP resonances are extractable from time series and allow thus for providing an unambiguous "signature" of nonlinear oscillations embedded within a stochastic background. By relying then on the theory of reduced RP resonances presented in this contribution, Part III addresses 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

The determination of reduced equations that aims at mimicking in a reduced state space the dynamics (reduced systems for short) issued from a system of *stochastic differential equations (SDEs)* posed in a higher-dimensional state space has become a central question in several disciplines ranging from molecular dynamics [SFHD99, SH99, SHD01, ZHS17, DLP+18] to epidemic models [FS09, FBS09] and climate dynamics over the past decades; see e.g. [PS95, MTVE01, GKS04, CLW15b, GCF17, BAB+17] 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); see e.g. [LL10, LLO17, ZHS17, DLP+18, CLM19].

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 differential systems. Initially introduced for discrete and continuous chaotic deterministic systems [Rue86, Pol86], these resonances are known to provide rigorous decomposition formulas of correlation functions and power spectra; see e.g. [BER89, Fro97, Bal00, MG07, GLP13, NZ15] and references therein. However, the rigorous 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 power spectral densities (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 that take place within the full state space $\mathcal{X}$. Theorem 2 can be viewed as providing the analogue, in the SDE context dealt with in this article, of coarse-grained Markov representations used for describing coarse-grained dynamics of Hamiltonian systems [SFHD99] or more general Markov state models [SS13]. We emphasize that, in the context of homogenization of SDEs [PS08], a notion of reduced RP resonances was also considered in [CVE11] as eigenvalues of a coarse-grained diffusion processes obtained from multiscale data.

In a second step, under ergodicity (or mixing) assumptions satisfied for a broad class of SDEs, we show that the reduced RP resonances relate to 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 and Corollary 2. Thus, reduced RP resonances inform us about the spectral elements of some coarse-grained version of the SDE generator. By relying on Theorems 2 and 3, these resonances can be estimated from time series by estimation of transition probabilities at a time-lag τ; see Sec. 3.3. Corollary 2 shows that an approximation relationship between the reduced RP resonances and those of the conditional expectation generator is ensured when the time-lag τ at which the transitions are collected from partial observations in V, is either sufficiently small or large.[1] In the context of homogenization of SDEs, the reduced RP resonances approximate even the RP resonances of the full SDE system; see [CVE11] and Remark 4-(i). The facts that the reduced RP resonances can be estimated from time series and relate to the conditional expectation, provide 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; see e.g. [GKS04, CH06, WL12, WL13, KCG15, CLM17, CLM19].

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 for reconstructing to a very good accuracy level the correlation functions and PSDs. Thus here the conditional expectation is sufficient to close the system (at least for reconstructing correlations) and the inclusion of non-Markovian effects for model reduction can be neglected. Such a diagnosis regarding the conditional expectation drawn from RP resonance analysis offers thus promising perspectives in terms of reduced-order modeling, and provide useful insights to reduction approaches of SDEs exploiting conditional expectations such as in e.g. [WVE04, LL10, LVE14, ZHS17, DLP+18].

The usefulness of reduced RP resonance analysis is not limited to providing useful diagnosis for model reduction. The companions articles, Part II [TCND20] and Part III [TCND19], 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, these resonances forming in particular discrete parabolas whose characteristics describe the frequency of the underlying oscillations and whose associated modes inform about the interactions between the noise and nonlinear effects; see [TCND20].

As pointed out in Sec. 3.3 and substantiated in Part III, such features are furthermore useful to identify and qualify from time series, the presence of nonlinear oscillations embedded within a stochastic background. Indeed by relying on the theory of reduced RP resonances presented in this contribution, Part III addresses 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[2], with the addition of noise modeling fast atmospheric fluctuations. Finally, we mention that the characterization of nonlinear oscillations in terms of (reduced) RP resonances has already demonstrated its usefulness for data-driven modeling

[1] In practice however it is often observed that the reduced RP resonances still provide useful information for "intermediate" time-lags; see Part III [TCND19].

[2] We refer to [CCHT19] for a mathematical analysis of the related Jin-Neelin model.

purposes of multiscale datasets arising in climate dynamics. Indeed by leveraging on the understanding gained here and in Part II [TCND20] as well as by exploiting the framework of [CK17] to decompose the multiscale variability, successful applications to the emulation of wind-driven ocean gyres dynamics have been reported in [KCB18] by means of frequency-ranked networks of stochastic Stuart-Landau oscillators. The modeling of Arctic sea ice concentration and extent has been also addressed within this framework; see [KCYG18, KCG18]. We envision a further program of using reduced RP resonances for the stochastic inverse modeling of climate datasets along with the analysis of the nearing to tipping points [TLLD18] from such datasets, in the context of climate change.

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 group of experts working in the field of dynamical system theory [Bal00] and the mathematical study of scattering resonances [Zwo17]. Initially introduced for discrete and continuous chaotic deterministic systems (see also [GLP13] for the case of Anosov flows), the RP resonances extend to stochastic differential 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 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:

$$\mathrm{d}X = F(X)\,\mathrm{d}t + D(X)\,\mathrm{d}W_t, \qquad X \in \mathbb{R}^d. \tag{2.1}$$

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

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

$$\Big[D(X)\,\mathrm{d}W_t\Big]_i = \sum_{j=1}^{q} D_{ij}(X)\,\mathrm{d}W_t^j, \quad q \geq 1. \tag{2.2}$$

Throughout this article, we assume that the vector field F and the (non-zero rank) matrix-valued function

$$D : \mathbb{R}^d \to \mathrm{Mat}_{\mathbb{R}}(d \times q),$$

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*

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

with $\mathbf{\Sigma}(X) = D(X)D(X)^T$ 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, $\mathcal{A}$, inform 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 $L^p_\mu(\mathbb{R}^d)$, for any $p \geq 1$. This property alone allows us to get useful decomposition formulas of correlation functions $C_{f,g}(t)$ given by

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

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

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

where

$$\|T(t)\|_{ess} = \inf \Big\{ \|T(t) - \mathcal{C}\|_{\mathcal{L}(\mathcal{E})} \, : \, \mathcal{C} \text{ is a linear and compact operator of } \mathcal{E} = L^p_\mu(\mathbb{R}^d) \Big\}. \tag{2.6}$$

In other words, $\|T(t)\|_{ess}$ measures the distance of $T(t)$ to the set of linear and compact operators of $L^p_\mu(\mathbb{R}^d)$. In the case this distance approaches zero as $t \to \infty$, 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 infinity[3].

In all cases, we have the following decomposition theorem of strongly continuous semigroups [EN00, Theorem 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 $L^2_\mu(\mathbb{R}^d)$. Let $\lambda_1, \cdots, \lambda_N$ be in $\sigma(K)$ satisfying $\mathrm{Re}\,\lambda_1, \cdots, \mathrm{Re}\,\lambda_N > \omega_{ess}(\mathcal{P})$. Then the following properties hold:*

(i) The λ_j are isolated spectral values of K with finite algebraic multiplicity.

(ii) If $\Pi_1, \cdots, \Pi_N$ denote the corresponding spectral projections and $m_1, \cdots, m_N$ the corresponding orders of poles of the resolvent of K, then

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

where

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

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

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

$$\|R_N(t)\|_{\mathcal{L}(L^2_\mu(\mathbb{R}^d))} \leq M e^{(\omega^*_N+\epsilon)t},\ \forall\, t \geq 0, \tag{2.9}$$

with

$$\omega^*_N = \sup\{\omega_{ess}(\mathcal{P})\} \cup \{\mathrm{Re}\,\lambda\ :\ \lambda \in \sigma(K)\backslash\{\lambda_1, \cdots, \lambda_N\}\}. \tag{2.10}$$

In what follows we denote by $\langle\cdot,\cdot\rangle_\mu$ the inner product in L^2_μ. 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^2_μ. Let f and g in L^2_μ such that $\langle f\rangle_\mu = \langle g\rangle_\mu = 0$, then given the λ_js and the Π_js as defined in Theorem 1, the correlation function $C_{f,g}(t)$ associated with the observables f and g has the following expansion*

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

with

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

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

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

where ψ_j denotes the L^2_μ-eigenfunction associated with λ_j and ψ^*_j, the associated L^2_μ-eigenfunction of the adjoint operator K^* of K.

Proof The decomposition (2.11) is a direct consequence of the definition of $C_{f,g}$ in (2.4) and of (2.7)-(2.9). The proof of (2.13) consists of noting that when $m_j = 1$ in (2.11), the integrals therein reduce to

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

and that $\mathcal{Q}_N(t) \underset{N\to\infty}{\longrightarrow} 0$ in virtue of our assumptions about $\sigma(K)$ and the reminder estimate (2.9). □

The λ_j such that $\mathrm{Re}\,\lambda_j > \omega_{ess}(\mathcal{P})$, 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]. Remarks 2-(ii)/(iii) point 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, $\tau = \mathrm{gap}(K)$ 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 e.g. [Fro97, Bal00, GLP13, CNK+14] and references therein for results in the deterministic context.

Remark 1

(i) **Decay of correlations**. *Under the conditions of Corollary 1, one obtains that $C_{f,g}(t) \underset{t\to\infty}{\longrightarrow} 0$, for any f, g in L^2_μ such that $\int f\,\mathrm{d}\mu = \int g\,\mathrm{d}\mu = 0$, and without assuming P_t to be strong Feller and irreducible.*

(ii) **Transfer operator**. *By working in the Hilbert space L^2_μ allows us to define the transfer operator $\mathfrak{L}_t$ acting on functions in L^2_μ, by the duality formula:*

$$\int f \cdot P_t g \, \mathrm{d}\mu = \int \mathfrak{L}_t f \cdot g \, \mathrm{d}\mu, \ f, g \in L^2_\mu, \ t \geq 0. \tag{2.15}$$

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 $C_0^\infty(\mathbb{R}^d)$ is contained in the domain $D(K)$ of the generator K, then there exist functions a_{ij}, b_i, and c in $C(\mathbb{R}^d)$ $(1 \leq i, j \leq d)$ such that for any x, the matrix $\big(a_{ij}(x)\big)_{1\leq i,j\leq d}$ is non-negative definite, $c(x) \leq 0$,[4] and*

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

Furthermore,

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

For such reasons we sometimes refer to the RP resonances as the RP resonances of the Kolmogorov operator $\mathcal{K}$, 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.
We mention that 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 p is large; see [Dav96, Dav07]. Instead, the notion of reduced RP resonances based on Corollary 2 below (see also Sec. 3) informs us about the spectral elements of some coarse-grained version of $\mathcal{K}$; see Theorem 3 and Remark 4-(iii). 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

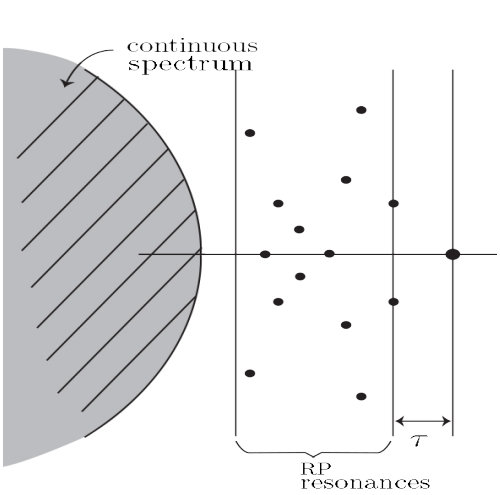
(i) Note that the sum in (2.13) starts actually at $j = 2$ since $\psi_1 = \mathbb{1}_{\mathbb{R}^d}$ (since $1 \in \sigma(P_t)$) and $\langle f \rangle_\mu = 0$, by assumption.

(ii) When $\sigma(K) = \{\lambda_1, \lambda_2, \cdots\}$ with $\mathrm{Re}\,\lambda_{j+1} \leq \mathrm{Re}\,\lambda_j$ for all $j \in \mathbb{N}^*$, and $\mathrm{Re}\,\lambda_j \to -\infty$, 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:

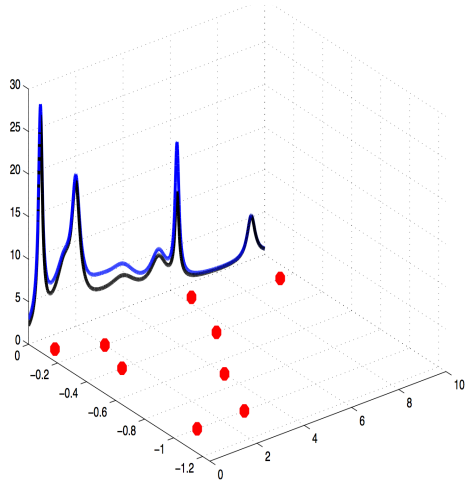
$$\mathrm{d}x = -\nabla V(x)\,\mathrm{d}t + \sqrt{2\beta^{-1}}\,\mathrm{d}W_t, \ \beta > 0, \tag{2.18}$$

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, Theorem 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.

[4] 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 $\mathbb{R}^d$ escapes to infinity at some finite time $t > 0$. This article is not concerned with explosive stochastic processes.



(a) Spectrum of Markov semigroup generator, K.



(b) Correspondence between the PSD and RP resonances according to (2.21).

Fig. 1: **Panel (a):** The Ruelle-Pollicott (RP) resonances are shown here in the left half plane. 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].

(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

$$\|P_t\varphi - \langle\varphi\rangle_\mu\|_{L^2_\mu} \leq Me^{-\alpha t}\|\varphi - \langle\varphi\rangle_\mu\|_{L^2_\mu}. \tag{2.19}$$

Note that if P_t has a regularizing effect sending L^2_μ into the Sobolev space $W^{1,2}_\mu$ in finite time and $W^{1,2}_\mu$ 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 $\mathbb{1}_{\mathbb{R}^d}$ 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 $h : \mathbb{R}^d \to \mathbb{R}$ 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

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

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, $\{z \in \mathbb{C} \ : \ \mathrm{Re}\,(z) \leq 0\}$ and its resolvent $R(z) = (z\mathrm{Id} - K)^{-1}$, is a well-defined linear operator that satisfies

$$S_h(f) = \int_{\mathbb{R}^d} h(X)\big[R(if)h\big](X)\,\mathrm{d}\mu. \tag{2.21}$$

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 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 $\{z \in \mathbb{C} \,:\, \mathrm{Re}\,(z) \leq -\gamma\}$, for some $\gamma > 0$.[5]

Formula (2.21) establishes furthermore a useful correspondence between PSD and RP resonances. Indeed, from (2.21) we infer 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 $\alpha_j^k(h)$ the coefficients resulting as integration with respect to μ in (2.11), the latter decomposition formula writes

$$C_h(t) = \sum_{j=1}^{N} \left(\sum_{k=0}^{m_j-1} \frac{t^k}{k!} \alpha_j^k(h) \right) e^{\lambda_j t} + \mathcal{Q}_N(t), \tag{2.22}$$

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 do not depend on the observable h, but that the $\alpha_j^k(h)$ do.

If we assume that $\mathrm{Re}\,(\lambda_j) < 0$ for $j > 0$, that each λ_j is simple ($m_j = 1$) 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)), where $\alpha_j^k = \alpha_j$ as given by (2.14), and the corresponding correlation spectrum $S_h(f)$ possesses itself a similar decomposition in terms of Lorentzian functions, namely:

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

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, NZ15]. Note that RP resonances appear also in the decomposition of the susceptibility function arising in linear response theory [Luc18, Eq. (5)]; see [Rue05, Ces07, Rue09].

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 [TCND19]. 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.

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

The results of this section rely substantially on the general disintegration theorem of probability measures[6]; see e.g. [DM78, p. 78]. See also [AGS08, Theorem 5.3.1] and [CP97]. The disintegration theorem states that given a probability measure μ on $\mathbb{R}^d$, a finite-dimensional Euclidean space V for which $1 \leq \dim(V) < d$, and a map $h : \mathbb{R}^d \to V$ (Borel-measurable), then there exists a uniquely determined family of probability measures $\{\mu_v\}_{v\in V}$ such that, for $\mathfrak{m}$-almost all [7] v in V, μ_v is concentrated on the pre-image $h^{-1}(\{v\})$ of v, i.e. $\mu_v\left(\mathbb{R}^d \setminus h^{-1}(\{v\})\right) = 0$, and such that for every Borel-measurable function $\phi : \mathbb{R}^d \to \mathbb{R}$,

$$\int_{\mathbb{R}^d} \phi(x)\,\mathrm{d}\mu(x) = \int_V \Big(\int_{x\in h^{-1}(\{v\})} \phi(x)\,\mathrm{d}\mu_v(x) \Big)\,\mathrm{d}\mathfrak{m}(v). \tag{3.1}$$

Here $\mathfrak{m}$ denotes the push-forward in V of the measure μ in $\mathbb{R}^d$, 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 $\mathbb{R}^d$, and the mapping h is the projector $\mathit{\Pi}_V$ onto V. In this case, a decomposition analogous to (3.1) holds for the measure μ itself, namely

$$\mu(B \times F) = \int_F \mu_v(F)\,\mathrm{d}\mathfrak{m}(\xi), \qquad B \times F \in \mathcal{B}(V) \otimes \mathcal{B}(W), \tag{3.2}$$

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

$$\mathcal{X}_v = h^{-1}(\{v\}), \quad \forall\, v \in V. \tag{3.3}$$

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 $\mathbb{R}^d$ to V, and a Markov semigroup $(P_t)_{t\geq 0}$ possessing an invariant measure on $\mathbb{R}^d$ — 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 $(P_t)_{t\geq 0}$ that take place within the full state space $\mathbb{R}^d$. Theorem 2 can be viewed as providing the analogue, in the SDE context dealt with in this article, of coarse-grained Markov representations used for describing coarse-grained dynamics of Hamiltonian systems [SFHD99] or more general Markov state models [SS13]. We emphasize that, as for these contexts, empirical estimates of the transition probabilities appearing in (3.4) of Theorem 2 are important for applications, as briefly discussed in Section 3.3 below and in more details in Part III [TCND19].

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 $1 \leq dim(V) < p$. Let $h : \mathbb{R}^d \to V$ 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 $L^1_{\mathfrak{m}}(V)$ such that, for any Borel sets B and C of V, and any $t > 0$,

$$\frac{\langle T_t \mathbb{1}_B, \mathbb{1}_C\rangle_{1,\infty}}{\mathfrak{m}(B)} = \mathit{Prob}(S(t,\cdot)x \in h^{-1}(C) | x \in h^{-1}(B)), \tag{3.4}$$

where $S(t,\omega)$ denotes the stochastic flow associated with P_t, and $\mathbb{1}_E$ denotes the characteristic function of a Borel set E.

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

$$\frac{1}{t}\int_0^t \int_V T_s f(v)\,\mathrm{d}\mathfrak{m}(v)\,\mathrm{d}s \underset{t\to\infty}{\longrightarrow} \int_V f(v)\,\mathrm{d}\mathfrak{m}(v). \tag{3.5}$$

[6] Variation of this theorem is used in the study of spectral gaps for deterministic maps and is known as Rokhlin's disintegration theorem; see [GL20].

[7] i.e. up to an exceptional set of null measure with respect to $\mathfrak{m}$.

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

$$T_t f(v) = \int_{x \in \mathcal{X}_v} P_t[f \circ h](x)\, \mathrm{d}\mu_v(x), \tag{3.6}$$

where $\mathcal{X}_v$ denotes the pre-image (in $\mathbb{R}^d$) 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 (3.6) 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 (3.6) with respect to the measure $\mathfrak{m}$, and applying (3.1) with $\phi = P_t[f \circ h]$, one notes that

$$\int_V T_t f(v)\, \mathrm{d}\mathfrak{m}(v) = \int_{\mathbb{R}^d} P_t[f \circ h](x)\, \mathrm{d}\mu(x). \tag{3.7}$$

This last identity allows us to get the following estimates

$$\begin{aligned} \Big| \int_V T_t f(v)\, \mathrm{d}\mathfrak{m}(v) \Big| &\leq \int_{\mathbb{R}^d} |P_t[f \circ h](x)|\, \mathrm{d}\mu(x) \\ &\leq \int_{\mathbb{R}^d} |f \circ h(x)|\, \mathrm{d}\mu(x), \end{aligned} \tag{3.8}$$

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

$$\|T_t f\|_{L^1_{\mathfrak{m}}(V)} \leq \int_V |f(v)| \int_{\mathcal{X}_v} \mathbb{1}_V\, \mathrm{d}\mu_v\, \mathrm{d}\mathfrak{m}(v) = \|f\|_{L^1_{\mathfrak{m}}(V)}, \tag{3.9}$$

which shows that T_t maps $L^1_{\mathfrak{m}}(V)$ 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

$$\begin{aligned} \langle T_t \mathbb{1}_B, \mathbb{1}_C \rangle_{L^1_{\mathfrak{m}}, L^\infty_{\mathfrak{m}}} &= \int_V \mathbb{1}_C(v) \Big[\int_{\mathcal{X}_v} P_t[\mathbb{1}_B \circ h](y)\, \mathrm{d}\mu_v(y) \Big]\, \mathrm{d}\mathfrak{m}(v) \\ &= \int_{\mathbb{R}^d} \mathbb{1}_{h^{-1}(C)}(x) P_t[\mathbb{1}_B \circ h](x)\, \mathrm{d}\mu(x) \\ &= \int_{\mathbb{R}^d} \mathfrak{L}_t \mathbb{1}_{h^{-1}(C)} \cdot \mathbb{1}_B \circ h\ \mathrm{d}\mu \\ &= \int_{\mathbb{R}^d} \mathfrak{L}_t \mathbb{1}_{h^{-1}(C)} \cdot \mathbb{1}_{h^{-1}(B)}\ \mathrm{d}\mu \end{aligned} \tag{3.10}$$

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

$$\frac{\langle T_t \mathbb{1}_B, \mathbb{1}_C \rangle_{1,\infty}}{\mathfrak{m}(B)} = \mathrm{Prob}(S(t,\cdot)x \in h^{-1}(C) | x \in h^{-1}(B)). \quad B, C \in \mathcal{B}(V). \tag{3.11}$$

Finally the proof of (3.5) 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*

$$\widetilde{C}_{f,g}(t) = \int_V f \cdot T_t g\, \mathrm{d}\mathfrak{m} - \int_V f\, \mathrm{d}\mathfrak{m} \int_V g\, \mathrm{d}\mathfrak{m}. \tag{3.12}$$

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

$$\widetilde{C}_{f,g}(t) = C_{f \circ h, g \circ h}(t), \tag{3.13}$$

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 the 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 the 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.

$$T_{t+s}\neq T_tT_s, \tag{3.14}$$

(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 spectral information 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 along with the notion of reduced RP resonances based on Corollary 2 below (see also Sec. 3.3), 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 $t\to\infty$ (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 (3.23) below. In other words, Theorem 3 provides a useful relationship between the reduced non-Markovian process associated with the family $(T_t)_{t\geq 0}$ and the Markov process associated with the conditional expectation induced by the observable h.

Lemma 3.1 *Let K denote the generator of P_t in $L^2_\mu(\mathbb{R}^d)$ with domain $D(K)$ as defined in (A.13). Let $h:\mathbb{R}^d\to V$ be a continuous surjective observable. 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 $D(K)$, $\lim_{s\to 0}(T_{t+s}f-T_tf)/s$ exists and is given by*

$$G_tf(v)=\int_{\mathcal{X}_v}P_tK[f\circ h](x)\,\mathrm{d}\mu_v(x), \tag{3.15}$$

whose domain is thus

$$D(G_t)=\{f\in L^2_{\mathfrak{m}}(V)\,:\,f\circ h\in D(K)\}. \tag{3.16}$$

Proof. The proof is elementary and consists of noting that for any f in $L^2_{\mathfrak{m}}(V)$ for which $f\circ h$ lies in $D(K)$, we have

$$\frac{T_{t+s}f-T_tf}{s}=\int_{\mathcal{X}_v}\frac{P_{t+s}-P_t}{s}[f\circ h](x)\,\mathrm{d}\mu_v(x)\underset{s\to 0}{\longrightarrow}\int_{\mathcal{X}_v}P_tK[f\circ h](x)\,\mathrm{d}\mu_v(x), \tag{3.17}$$

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 $L^2_{\mathfrak{m}}(V)$ such that for any f in $L^2_{\mathfrak{m}}(V)$*

$$\frac{1}{t}\int_0^t\int_V G_tf(v)\,\mathrm{d}\mathfrak{m}(v)\underset{t\to\infty}{\longrightarrow}\int_V\mathcal{G}f(v)\,\mathrm{d}\mathfrak{m}(v), \tag{3.18}$$

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

$$\mathrm{Re}\,\langle\mathcal{G}f,f\rangle_{L^2_{\mathfrak{m}}(V)}\leq 0,\ \forall\, f\in D(\mathcal{G}). \tag{3.19}$$

Furthermore, if μ is strongly mixing then

$$\lim_{t\to\infty}\int_V G_t f(v)\,\mathrm{d}\mathfrak{m}(v) = \int_V \mathcal{G}f(v)\,\mathrm{d}\mathfrak{m}(v). \tag{3.20}$$

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 $C_0^\infty(\mathbb{R}^d) \subset D(K)$, then $\mathcal{G}$ possesses the following differential expression

$$\mathcal{G}f(v) = \frac{1}{2}\sum_{i,j=1}^{\dim(V)} \overline{\mathbf{\Sigma}}_{ij}(v)\partial_{ij}^2 f + \sum_{i=1}^{\dim(V)} \overline{F}_i(v)\partial_i f, \quad v\in V, \quad f\in C_0^2(V). \tag{3.21}$$

where

$$\overline{\mathbf{\Sigma}}_{ij}(v) = \int_{\mathcal{X}_v} \mathbf{\Sigma}_{ij}(x)\,\mathrm{d}\mu_v(x), \quad \overline{F}_i(v) = \int_{\mathcal{X}_v} F_i(x)\,\mathrm{d}\mu_v(x), \quad x\in\mathbb{R}^d,\ v\in V, \tag{3.22}$$

where F and $\mathbf{\Sigma}$ 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 a reduced ("effective") SDE in V of the form:*

$$\boxed{\mathrm{d}v = \overline{F}(v)\,\mathrm{d}t + \sigma(v)\,\mathrm{d}W_t^V, \quad v\in V,} \tag{3.23}$$

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

$$\overline{\mathbf{\Sigma}}_{ij}(v) = (\sigma(v)\sigma(v)^T)_{ij} = \sum_{k=1}^{q} \overline{D_{ik}(x)D_{jk}(x)}, \quad x\in\mathbb{R}^d,\ v\in V, \tag{3.24}$$

where the $D_{\ell k}$ are the diffusion coefficients of the original SDE (2.1), and $\overline{(\cdot)}$ denotes the averaging over the disintegrated measure μ_v.

Proof. **Step 1: Proof of (3.18) and (3.19).** Let f be in $D(G_t)$ given in (3.16). Then

$$\begin{aligned}\int_V G_t f(v)\,\mathrm{d}\mathfrak{m}(v) &= \int_V\int_{\mathcal{X}_v} P_t K[f\circ h](x)\,\mathrm{d}\mu_v(x)\,\mathrm{d}\mathfrak{m}(v),\\ &= \int_{\mathbb{R}^d} P_t K[f\circ h](x)\,\mathrm{d}\mu(x),\end{aligned} \tag{3.25}$$

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 (3.18) with $\mathcal{G}$ given by

$$\mathcal{G}f(v) = \int_{\mathcal{X}_v} K[f\circ h](x)\,\mathrm{d}\mu_v(x), \tag{3.26}$$

by application once more of the disintegration formula.

The same formula ensures furthermore that

$$\int_V \mathcal{G}f\cdot f\,\mathrm{d}\mathfrak{m} = \int_{\mathbb{R}^d} K[f\circ h]\cdot[f\circ h]\,\mathrm{d}\mu \tag{3.27}$$

and by taking the real part, (3.19) follows from (A.14).

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

Then from (3.26), one has

$$\begin{aligned}\int_V |\mathcal{G}f_n - g|^2\,\mathrm{d}\mathfrak{m} = -2\int_{\mathbb{R}^d} K[f_n\circ h](x)\cdot(g\circ h(x))\,\mathrm{d}\mu +\\ \int_{\mathbb{R}^d} |K[f_n\circ h](x)|^2\,\mathrm{d}\mu + \int_{\mathbb{R}^d} |g\circ h(x)|^2\,\mathrm{d}\mu(x).\end{aligned} \tag{3.28}$$

Now since K is closed in $L^2_\mu(\mathbb{R}^d)$, $K[f_n \circ h] \underset{n\to\infty}{\longrightarrow} K[f \circ h]$ and $f \circ h \in D(K)$, namely $f \in D(\mathcal{G})$. This shows furthermore that one can pass to the limit in (3.28), which gives

$$2\int_{\mathbb{R}^d} K[f\circ h](x)\cdot(g\circ h(x))\,\mathrm{d}\mu = \int_{\mathbb{R}^d} |K[f\circ h](x)|^2\,\mathrm{d}\mu + \int_{\mathbb{R}^d} |g\circ h(x)|^2\,\mathrm{d}\mu(x), \tag{3.29}$$

The identity (3.29) says nothing else than

$$\|K[f\circ h] - g\circ h\|^2_{L^2_\mu} = 0, \tag{3.30}$$

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

$$\int_{\mathcal{X}_v} K[f\circ h](x)\,\mathrm{d}\mu_v(x) = \int_{\mathcal{X}_v} g\circ h(x)\,\mathrm{d}\mu_v(x) = g(v), \quad \text{for } \mathfrak{m}\text{-almost every } v \text{ in } V. \tag{3.31}$$

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 $L^2_\mathfrak{m}(V)$. Assume by contradiction that there exist g in $L^2_\mathfrak{m}(V)$ for which any sequence in $D(\mathcal{G})$ would stay at a certain distance (in $L^2_\mathfrak{m}(V)$) from g or would have g as a point of accumulation, but not unique.

Let us take $V = \mathbb{R}^q$ with $1 \le q < d$ (since h is a projector) and define G the mapping from V to V such that $\Pi_j G = g$, for all $1 \le j \le q$. Since $D(K)$ is dense in $L^2_\mu(\mathbb{R}^d)$ ([Paz83, Cor. 2.5 p. 5]) there exists a sequence $\{\varphi_n\}$ in $(D(K))^d$ that converges to $G\circ h$ in $(L^2_\mu(\mathbb{R}^d))^d$, such that in particular $\|\varphi_n - G\circ h\|$ (norm in $\mathbb{R}^d$) converges to zero in L^2_μ. Let us define, almost everywhere (a.e.), a sequence $\{f_n\}$ in $(D(\mathcal{G}))^q$ from the semiconjugacy relation[8]

$$f_n \circ h = h \circ \varphi_n, \tag{3.32}$$

i.e. such that the following diagram is commutative a.e.

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

Furthermore, since h is a projector

$$\int \|h(\varphi_n(x)) - h(G\circ h(x))\|^2\,\mathrm{d}\mu(x) \le \int \|\varphi_n(x) - G\circ h(x)\|^2\,\mathrm{d}\mu(x) \tag{3.33}$$

which, thanks to (3.32), is equivalent to

$$\int \|f_n\circ h(x) - G\circ h(x)\|^2\,\mathrm{d}\mu(x) \le \int \|\varphi_n(x) - G\circ h(x)\|^2\,\mathrm{d}\mu(x). \tag{3.34}$$

On the other hand, by application of the disintegration formula

$$\int \left|\Pi_j\Big(f_n\circ h(x) - G\circ h(x)\Big)\right|^2 \mathrm{d}\mu(x) = \int_V |\Pi_j f_n(v) - g(v)|^2\,\mathrm{d}\mathfrak{m}(v), \quad 1 \le j \le q. \tag{3.35}$$

We have thus proved, thanks to (3.34) that for any $1 \le j \le q$,

$$\int_V |\Pi_j f_n(v) - g(v)|^2\,\mathrm{d}\mathfrak{m}(v) \le \int \|\varphi_n(x) - G\circ h(x)\|^2\,\mathrm{d}\mu(x). \tag{3.36}$$

Now since $\|\varphi_n - G\circ h\|$ converges to zero in L^2_μ by construction, we conclude that $\Pi_j f_n$, which belongs to $D(\mathcal{G})$ (also by construction), converges to g in $L^2_\mathfrak{m}(V)$, 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 (3.26) which leads to the expression (3.21) of $\mathcal{G}$, by integration with respect to μ_v. □

[8] Note that f_n defined by (3.32) implies that $\Pi_j f_n \circ h = \Pi_j \varphi_n$ belongs to $D(K)$ by construction, and thus $\Pi_j f_n$ belongs to $D(\mathcal{G})$ for every $1 \le j \le q$, since the RHS of (3.16) is also the domain of $\mathcal{G}$ as (3.16) is independent on t.

We have then the immediate corollary.

Corollary 2 *Under the assumptions of Theorem 3, if μ is strongly mixing, the eigenvalues of G_t provide approximations of the weak eigenvalues of $\mathcal{G}$ in the sense that if (λ_k^t, ψ_k^t) denotes an eigenpair of G_t, then*

$$\int_V \mathcal{G}\psi_k^t \, \mathrm{d}\mathfrak{m} = \lambda_k^t \int_V \psi_k^t \, \mathrm{d}\mathfrak{m} + \epsilon(t), \tag{3.37}$$

with $\epsilon(t) \to 0$, as $t \to \infty$.

Due to the definition of G_t in (3.15) and of $\mathcal{G}$ in (3.26), we have also that (3.37) holds with $\epsilon(t) \to 0$, as $t \to 0$.

The eigenvalues of G_t in Corollary 2 are called that the **reduced RP resonances** for the time-lag t.

Remark 4

(i) **Approximation of RP resonances**. *In the context of homogenization of SDEs, explicit formulas for the generator $\mathcal{G}$ of the conditional expectation are available; see e.g. [PS08, Chaps. 11 & 18]. Within this context and adopting the language of this article, the estimates [CVE11, Eqns. (5.12)-(5.13)] show that the RP resonances (resp. the eigenfunctions in L^2_μ) of the full generator K are approximated by the eigenvalues (resp. the eigenfunctions in $L^2_{\mathfrak{m}}$ after multiplication by the μ_x-density) of $\mathcal{G}$, when the time-scale separation between the observed and unobserved variables is sufficiently large. Thus, for such situations, the reduced RP resonances provide (weak) approximations to the actual RP resonances, due to Corollary 2.*

In the general case, error estimates between the RP resonances and the reduced RP resonances are difficult to derive. We refer though to Sec. 4 for an example for which the reduced RP resonances allow for very good approximation of correlation functions, even when the time-scale separation is weak. There, it is the decomposition (2.13) that is approximated when the RP resonances (and corresponding modes) therein are replaced by the spectral elements of G_t.

(ii) **$\mathcal{G}$ describes long- and short-time behaviors in V.** *Due to (3.14), 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. As pointed out in Corollary 2, the operator $\mathcal{G}$ as defined in (3.26) is exactly the pseudo-generator G_t of T_t, as $t \to 0$.*

The operator $\mathcal{G}$ relates thus the long time assymptotics ($t \to \infty$) of the reduced stochastic process to its infinitesimal characteristics which describe the short-time behavior ($t \to 0$), as for time homogeneous Markov processes. We refer to [LL10, LLO17, ZHS17, DLP+18] for error estimates between the coarse-grained dynamics and the effective dynamics involving conditional expectation, in the context of Langevin or overdamped Langevin dynamics.

(iii) **Coarse-grained Kolmogorov operator.** *Given a projector $h = \Pi_V$ onto V, another useful object related to the conditional expectation (3.23) is the following coarse-grained Kolmogorov operator:*

$$\overline{\mathcal{K}} f(v) = \int_{\mathcal{X}_v} \mathcal{K}[f \circ h](x) \, \mathrm{d}\mu_v(x), \quad v \in V, \quad f \in C_0^2(V), \tag{3.38}$$

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 the family $(T_t)_{t\geq 0}$, it provides the generator of the Markov process describing the average behavior of v_t, when the averaging is taken over the unobserved variables, i.e. over the variables lying outside of V.

(iv) For the sake of clarity, the formulas (3.21)-(3.22) and (3.23)-(3.24) in Theorem 3 have been articulated for the case where h is a projection. Of course such formulas can be generalized to more general mappings h, and for V that is not necessarily a subspace of $\mathbb{R}^d$. In such cases, the corresponding coarse-grained Kolmogorov operator, $\overline{\mathcal{K}}$, 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 $\overline{\mathcal{K}}$ 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. At the same time, a "bad" choice of the observable h can lead to a very poor domain $D(G_t)$ in Lemma 3.1. We refer to [BKL17] for an illuminating discussion regarding the impact of the choice of observable (and their regularity) in the context of linear response theory for (deterministic) Axiom A diffeomorphisms.

3.3 Empirical reduced RP resonances and reconstruction of correlations

As pointed out above, 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 system that lie within a reduced state space V and whose transitions are collected at a time-lag τ ($t = \tau$ in (3.4)), 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 [TCND20] and Part III [TCND19] .

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, roughly speaking, that T_τ has a pseudo-generator G_τ that is close (in a weak sense) to the generator $\mathcal{G}$ of the conditional expectation (3.23), when either τ is sufficiently small or large; see Remarks 4-(ii). Thus, in such cases, the resonances estimated from data whose transitions are collected at a time-lag τ in the reduced state space have an interesting interpretation. They approximate the (weak) RP resonances associated with the generator $\mathcal{G}$ of the reduced system (3.23) (Corollary 2), and are called the **empirical reduced RP resonances**. Two factors play an important role in this approximation: (a) the partition of V used to approximate the Markov operator T_τ by a transition matrix Γ_τ, and (b) the amount of data used for the estimation of the entries of Γ_τ (see (3.39) below); see [CVE06, CVE09]. When the latter is sufficiently large we may infer that the dominant part of the spectrum of Γ_τ provides a good approximation of that of T_τ, for a sufficiently fine partition of V.[9] Another important factor in this approximation is, as mentioned above, related to the time-lag τ as it impacts whether the pseudo-generator G_τ is a good (weak) approximation of the generator $\mathcal{G}$ of the conditional expectation. If all these factors are favorable, the empirical reduced RP resonances, as good approximation of the generator $\mathcal{G}$'s RP resonances, describe the solution's variability captured by the conditional expectation (3.23); Corollary 1 applied to the Markov semigroup generated by Eq. (3.23). These estimated reduced RP resonances inform us thus about the spectral elements of some coarse-grained version of the original SDE generator, but does not provide in general a sharp approximation of the RP resonances of the full generator K. We refer though to [CVE11] for estimates in the context of homogenized diffusion processes. Translated in the language of this article, it is shown there that the reduced RP resonances (associated with the homogenized diffusion operator) are asymptotically close to the RP resonances of the underlying multiscale diffusion operator, when the time-scale separation is sufficiently large.

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 and h should be also chosen such that the observed dynamics in V via h 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 $h : \mathbb{R}^d \to V$, 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 in V are made out of the stochastic process X_t (solving Eq. (2.1) in $\mathbb{R}^d$) at discrete time instants $t = t_n$, given as multiple of a sampling time δt, i.e. $t_n = n\delta t$ with $1 \leq n \leq N$, with N assumed to be large. We also assume the time-lag τ to be equal to a multiple of δt, i.e. $\tau = \ell\delta t$. These observations made in the observation space V are denoted by $Y_n = h(X_{t_n})$. By adapting the material contained in e.g. the Supporting Information of [CNK+14] (see also [TvdBD15]), the Markov operator T_τ is approximated by the

[9] We refer to [SS13] for useful error bounds regarding the dominant eigenvalues for certain types of coarse-graining maps.

$M \times M$ transition matrix Γ_τ whose entries are given

$$(\Gamma_\tau)_{ij} = \frac{\#\Big\{\Big(Y_n \in B_j\Big) \wedge \Big(Y_{n+\ell} \in B_i\Big)\Big\}}{\#\Big\{Y_n \in B_j\Big\}}, \tag{3.39}$$

where the B_j form a partition (composed of M disjoint boxes) of the aforementioned domain $\mathcal{D}$ in V; see also [CVE06, SS13] and references therein. In (3.39), 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]. We mention that the Markov operator T_τ can be also approximated — following the route paved by Ulam [Ula64] — from many short simulations instead of using long time series like here; see [Fro01, DFJ01]. Similarly, the Ulam's method can be adapted for the generator itself to estimate, when the dimension of the problem permits, the generator's spectral elements without trajectory integration, see [FJK13].

The empirical 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

$$\lambda_k(\tau) = \frac{\log\big(|\zeta_k(\tau)|\big)}{\tau} + i\frac{\arg\big(\zeta_k(\tau)\big)}{\tau}, \qquad 1 \le k \le M, \tag{3.40}$$

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 (3.40) 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 $\lambda_k(\tau)$ given by (3.40) lie naturally within the left-half complex plane.

For τ sufficiently small or large, i.e. when the generator of the conditional expectation is reasonably well approximated by the pseudo-generator of T_τ (in a weak sense, see Theorem 3 and Remark 4-(ii)), one can thus proceed as follows to measure the amount of solution's variability captured by the conditional expectation:

- $(\mathbf{C_1})$ Given an observable and a reduced state space V, one forms the right-hand side (RHS) of (2.13) in which the λ_j are replaced by the empirical reduced RPs, i.e. $\lambda_j(\tau)$ given by(3.40), and the eigenfunctions by those of the Markov operator T_τ. This way, one forms what we call the "reduced" correlation function.
- $(\mathbf{C_2})$ One compares the resulting reduced correlation function obtained in step $(\mathbf{C_1})$, 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.

Note that by replacing the RP resonances by the (empirical) reduced RP resonances in (2.23), power spectral densities can also be approximated in the same fashion. One should however emphasize that such an analysis depends on the good choice of the time-lag τ at which the data are collected [TLLD18], and one should keep in mind this important aspect before drawing any conclusion.

When τ is neither small or large, the precise relationships between the $\lambda_k(\tau)$ and the actual RP resonances are non-trivial to characterize in general. Nevertheless, in certain cases, as shown in Part III [TCND19], the reduced RP resonances are very useful to diagnose and characterize important dynamical features such as nonlinear oscillations embedded within a stochastic background. The next section provides an example for which the non-Markovian effects are negligible to capture the slow variable's variability (in terms of correlation functions), whereas the time-scale separation between the "slow" and "fast" variables is not large.

4 Applications to a stochastic slow-fast system

4.1 The model

We consider the following stochastic system

$$\begin{cases} \mathrm{d}x = (\lambda x - fy - \gamma xz)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^1 & \text{(4.1a)} \\ \mathrm{d}y = (fx + \lambda y - \gamma yz)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^2 & \text{(4.1b)} \\ \mathrm{d}z = -\frac{1}{\epsilon}(z - x^2 - y^2)\,\mathrm{d}t + \frac{\sigma}{\sqrt{\epsilon}}\,\mathrm{d}W_t^3. & \text{(4.1c)} \end{cases}$$

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 in [CLPR19] 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

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

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 $\mathcal{M}_\epsilon = \mathrm{graph}(\Phi) + \mathcal{O}(\epsilon)$, 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 empirical 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

$$\begin{aligned} \mathrm{d}u &= \big(\lambda u - fv - \gamma u(u^2 + v^2)\big)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^1 \\ \mathrm{d}v &= \big(fv + \lambda v - \gamma v(u^2 + v^2)\big)\,\mathrm{d}t + \sigma\,\mathrm{d}W_t^2. \end{aligned} \tag{4.3}$$

The RP resonances of this system, a Hopf normal form subject to an additive white noise, are analyzed in details in Part II [TCND20]. 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 $T_s = 10^3$. With $t_n = n\delta t$, 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 Γ_τ in (3.39), 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 time-lag τ to estimate the transitions is chosen depending on these regimes as follows: $\tau = 10^{-3}$ for Case I, and $\tau = 10^{-2}$ for Cases II and 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

Within this experimental protocol, we first estimate the reduced RP resonances. To do so, we first estimate the Markov matrix entries of Γ_τ according to (3.39), and then determine the reduced RP resonances according to (3.40). 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 $(\mathbf{C_1})$ 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 $(\mathbf{C_1})$, and here again the black curve its standard ACF sample estimate.

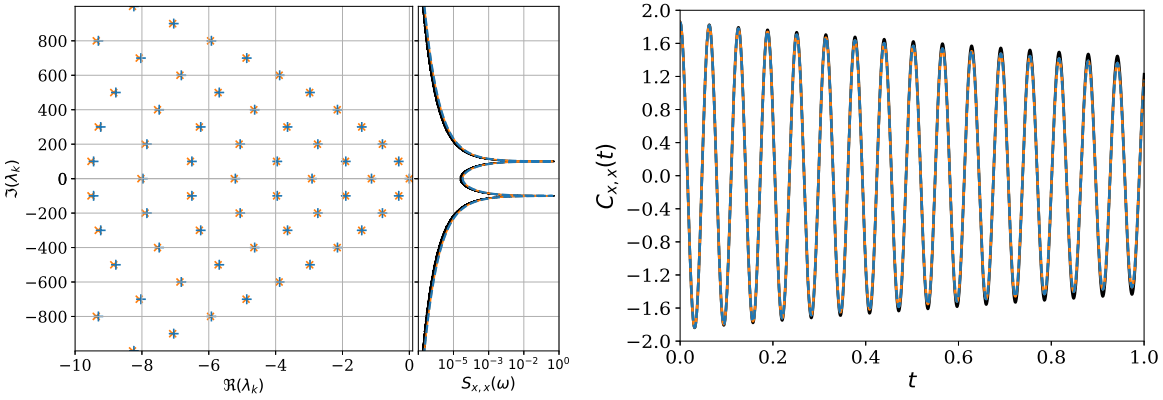
The conclusions of these numerical experiments are without ambiguity. First, in the case of a strong time-scale separation between the observed variables (x,y) and the unobserved variable (z), i.e. in Cases I and II, the reduced RP resonances allow for an almost exact reconstruction of the ACF and PSD of $x(t)$ (as for $y(t)$, not shown). This is explained from the theoretical understanding provided by Secns. 3.2 and 3.3. Indeed, as explained therein, since the time-lag 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 (3.23). 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 $\epsilon \to 0$; see [CLPR19]. Thus without any surprise the RP resonances of system (4.3) coincide with the reduced RP resonances of (4.1) (associated with the oberved variables (x, y)) 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 [TCND20].

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 come 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 (3.23) 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 from the original equations the conditional expectation; see also [CLM17]. Applied to Case III, this approach based on optimal parameterizing manifolds (OPMs) 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 the RP resonance analysis conducted here.

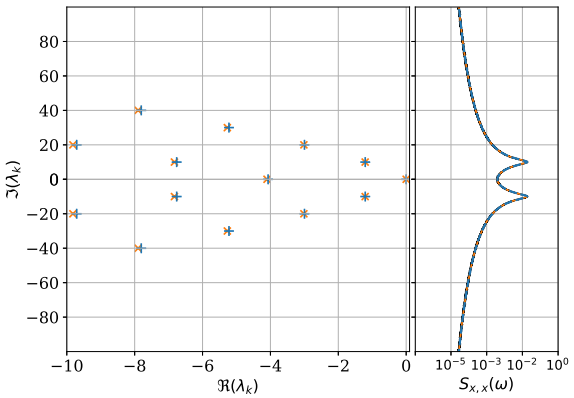
Acknowledgements The authors would like to thank the reviewers for their very useful and constructive comments. This work has been partially supported by the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement No. 810370 (MDC)), by the Office of Naval Research (ONR) Multidisciplinary University Research Initiative (MURI) grant N00014-16-1-2073 (MDC), by the National Science Foundation grants OCE-1658357, DMS-1616981(MDC), AGS-1540518 and AGS-1936810 (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).



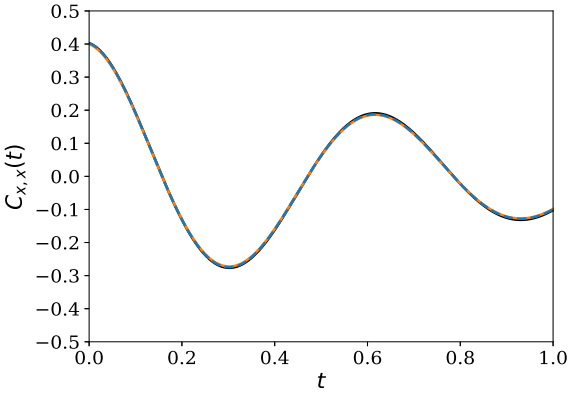
(a) Reduced RP resonances and PSD for Case I

(b) ACF for Case I

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 ($\mathbf{C_1}$) described in Sec. 3.3, are shown by a dashed orange (resp. blue) curve, while the black curves show their standard sample estimates.

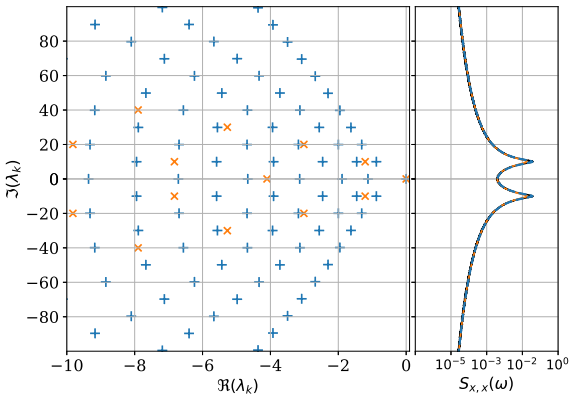


(a) Reduced RP resonances and PSD for Case II

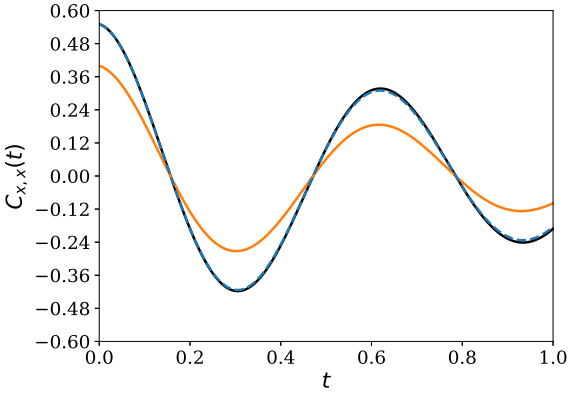


(b) ACF for Case II

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



(a) Reduced RP resonances and PSD for Case III



(b) ACF for Case III

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

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 $S(t,\omega)$ acting[10] 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 $\mathcal{X}$. In what follows $\mathcal{X}$ is a finite-dimensional Polish space.

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:

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

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

[10] ω labelling the noise realization.

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

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

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 $f \in L^2_\mu(\mathcal{X})$, $\frac{1}{T}\int_0^T P_s f \,\mathrm{d}s \underset{T\to\infty}{\longrightarrow} \int f \,\mathrm{d}\mu$ in $L^2_\mu(\mathcal{X})$.

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

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

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 [SV72] (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 Markov 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 *Kolmogrorov operator* $\mathcal{K}$ defined on smooth functions ψ (of class C^2) as follows when $\mathcal{X} = \mathbb{R}^d$:

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

where

$$\mathrm{Tr}(\mathbf{\Sigma}\nabla^2\psi(\cdot)) = \sum_{i,j=1}^{d} \left[D(x)D(x)^T\right]_{ij}\partial^2_{ij}\psi. \tag{A.5}$$

Here Tr denotes the trace of a matrix. 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 Sinaï-Ruelle-Bowen (SRB) random measures. We refer also to Part II [TCND20], 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]. We mentioned however [MT16] for an example of a stochastic dynamical system which has the square of the Euclidean norm as the Lyapunov function, is hypoelliptic with nonzero noise forcing, and that yet fails to be reachable or ergodic.

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)]

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

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 $(P_t)_{t\ge 0}$. For any $p\ge 1$ and $t\ge 0$, P_t is extendable to a linear bounded operator on $L^p_\mu(\mathcal{X})$ still denoted by P_t. Moreover*

(i) $\|P_t\|_{\mathcal{L}(L^p_\mu(\mathcal{X}))} \le 1$

(ii) *P_t is strongly continuous semigroup in $L^p_\mu(\mathcal{X})$.*

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

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

where X^x_t 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:*

$$TV(\mathcal{L}_t\nu,\mu) \underset{t\to\infty}{\longrightarrow} 0. \tag{A.8}$$

For the definition of a strongly continuous semigroup also known as C_0-semigroup we refer to [EN00, p. 36]. 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

$$|P_t g(x)|^p \le P_t(|g|^p)(x). \tag{A.9}$$

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

$$\int_{\mathcal{X}} |P_t g(x)|^p \mu(\,\mathrm{d}x) \le \int_{\mathcal{X}} P_t(|g|^p)(x)\mu(\,\mathrm{d}x) = \int_{\mathcal{X}} |g|^p(x)\mu(\,\mathrm{d}x), \tag{A.10}$$

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 $(P_t)_{t\ge 0}$ can be uniquely extended to a contraction semigroup in $L^p_\mu(\mathcal{X})$, and property (i) is proved.

Let us show now that $(P_t)_{t\ge 0}$ is strongly continuous in $L^p_\mu(\mathcal{X})$. Since $(P_t)_{t\ge 0}$ 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

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

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 $A : D(A) \subset \mathcal{H} \to \mathcal{H}$, such that

$$A\varphi = \lim_{t\to 0^+} \frac{1}{t}\big(T(t)\varphi - \varphi\big), \tag{A.12}$$

defined for every φ in the domain

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

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

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

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 $\mathcal{E}$ satisfies [EN00, p. 249] the Hadamard formula

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

where

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

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 in $L^2_\mu(\mathbb{R}^d)$ ($\mathcal{X} = \mathbb{R}^d$) generated by an SDE given by Eq. (2.1) for which F and G are locally Lipschitz. Assume that there exists a Lyapunov function [11] U and a compact set $\mathfrak{A}$ for which there exist $a > 0$, $0 < \kappa < 1$ and $b < \infty$, such that*

$$\mathcal{K}U \leq aU, \tag{A.17a}$$

$$P_{t_0}U \leq \kappa U + b\mathbb{1}_{\mathfrak{A}},\ \ \textit{for some } t_0 > 0, \tag{A.17b}$$

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.

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

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

$$\mathcal{F}_U = \{f : \mathbb{R}^d \to \mathbb{R} \mid f \textit{ Borel measurable and } \|f\|_U < \infty\}, \tag{A.19}$$

and endowed with the norm

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

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 $\mathcal{F}_U$,

$$\left|P_t f(x) - \int f\,\mathrm{d}\mu\right| \leq Ce^{-\lambda t}U(x),\ \ t > t_0,\ \forall x \in \mathbb{R}^d. \tag{A.21}$$

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

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

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*

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

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*

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

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

More generally, if

$$\mathcal{K}U \leq -\alpha U + \beta,\ \ \textit{with}\ \alpha > 0,\ \textit{and}\ \ 0 \leq \beta < \infty, \tag{A.24}$$

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

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

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

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

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

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

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^2_μ, the theory of asymptotic behavior of a semigroup with a strictly dominant, algebraically simple eigenvalue (e.g. [vN12, Theorem. 3.6.2]) implies the spectral gap property stated in the following.

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*

$$\mathbb{E}\,|X_t^x|^2 < k|x|^2 e^{-\alpha t} + c,\ \ t \geq 0,\ \ x \in \mathbb{R}^d. \tag{A.28}$$

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*

$$\|P_t\varphi - \int \varphi\,\mathrm{d}\mu\|_{L^2_\mu} \leq Ce^{-\lambda t}\|\varphi\|_{L^2_\mu},\ t \geq 0,\ \varphi \in L^2_\mu, \tag{A.29}$$

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^2_μ*-spectrum of the generator* K*:*

$$0 < \lambda \leq \mathrm{gap}(K). \tag{A.30}$$

We will see in Part II [TCND20] 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^2_μ-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^2_μ-gap based on spectral gaps in Wasserstein distances, verifiable in practice by following the approach of [HMS11].

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

$$\mathrm{d}U = \mathcal{K}U\,\mathrm{d}t + \text{ “Martingale”}, \tag{A.31}$$

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

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

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 $t > t_0$,

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

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

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

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

$$\begin{aligned} |P_{nt} f(x) - \mathbb{1}_{\mathfrak{B}}(x) P_{nt} \mathbb{1}_{\mathfrak{B}} f(x)| &\leq U(x) \sup_{y \in \mathbb{R}^d \setminus \mathfrak{B}} \frac{|P_{nt} f(y)|}{U(y)}, \\ &\leq U(x) \|f\|_U \sup \frac{|P_{nt} U(y)|}{U(y)}, \end{aligned} \tag{A.35}$$

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

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

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

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

which leads to

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

Step 2. We show now that the linear operator

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

is compact for any compact set $\mathfrak{B}$ of $\mathbb{R}^d$. 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 Λg_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 (Λg_k) lies in $\mathcal{C}(\mathfrak{B})$.

We have

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

which shows that $\{\Lambda g_k\}$ is equibounded.

Furthermore, since P_t is strong Feller, it has a smooth kernel[12] and we have for all x and x' in $\mathfrak{B}$

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

which shows that $\{\Lambda g_k\}$ is equicontinuous.

Thus, the Ascoli-Arzelà theorem [Yos95, p. 85] applies and guarantees that a subsequence from Λg_k converges in $\mathcal{C}(\mathfrak{B})$ to g. Now since $U \geq 1$, the same extraction from Λg_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 $\mathbb{R}^d$.

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

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

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

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

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]. □

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

References

AGS08. L. Ambrosio, N. Gigli, and G. Savaré, *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Springer Science & Business Media, 2008.

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.

BAB+17. J. Berner, U. Achatz, L. BattÃľ, L. Bengtsson, A. CÃąmara, H.M. Christensen, M. Colangeli, D.R.B. Coleman, D. Crommelin, S.I. Dolaptchiev, C.L.E. Franzke, P. Friederichs, P. Imkeller, H. JÃďrvinen, S. Juricke, V. Kitsios, F. Lott, V. Lucarini, S. Mahajan, T.N. Palmer, C. Penland, M. Sakradzija, J.-S. von Storch, A. Weisheimer, M. Weniger, P.D. Williams, and J.-I. Yano, *Stochastic parameterization: Toward a new view of weather and climate models*, Bull. Amer. Met. Soc. **98** (2017), no. 3, 565–588.

Bal00. V. Baladi, *Positive Transfer Operators and Decay of Correlations*, World Scientific, Singapore, 2000.

BER89. V. Baladi, J.-P. Eckmann, and D. Ruelle, *Resonances for intermittent systems*, Nonlinearity **2** (1989), no. 1, 119.

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.

BKL17. V. Baladi, T. Kuna, and V. Lucarini, *Linear and fractional response for the SRB measure of smooth hyperbolic attractors and discontinuous observables*, Nonlinearity **30** (2017), no. 3, 1204.

CCHT19. Y. Cao, M. D. Chekroun, A. Huang, and R. Temam, *Mathematical analysis of the Jin-Neelin model of El Niño-Southern Oscillation*, Chinese Annals of Mathematics, Series B **40** (2019), no. 1, 1–38.

Cer01. S. Cerrai, *Second-Order PDE's in Finite and Infinite Dimension: A Probabilistic Approach*, vol. 1762, Springer Science & Business Media, 2001.

Ces07. B. Cessac, *Does the complex susceptibility of the Hénon map have a pole in the upper-half plane? A numerical investigation*, Nonlinearity **20** (2007), no. 12, 2883.

CH06. A.J. Chorin and O.H. Hald, *Stochastic Tools in Mathematics and Science*, Surveys and Tutorials in the Applied Mathematical Sciences, no. 147, Springer New York, 2006.

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*, J. Stat. Phys. (2019), doi: 10.1007/s10955-019-02458-2.

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.

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.

CP97. J. T. Chang and D. Pollard, *Conditioning as disintegration*, Statistica Neerlandica **51** (1997), no. 3, 287–317.

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. D. Crommelin and E. Vanden-Eijnden, *Fitting time series by continuous-time Markov chains: A quadratic programming approach*, Journal of Computational Physics **217** (2006), no. 2, 782–805.

CVE09. ———, *Data-based inference of generators for Markov jump processes using convex optimization*, Multiscale Modeling & Simulation **7** (2009), no. 4, 1751–1778.

CVE11. ———, *Diffusion estimation from multiscale data by operator eigenpairs*, Multiscale Modeling & Simulation **9** (2011), no. 4, 1588–1623.

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.

DFJ01. M. Dellnitz, G. Froyland, and O. Junge, *The algorithms behind gaioâĂŤset oriented numerical methods for dynamical systems*, Ergodic theory, analysis, and efficient simulation of dynamical systems, Springer, 2001, pp. 145–174.

DKZ12. F. Dragoni, V. Kontis, and B. Zegarliński, *Ergodicity of Markov semigroups with Hörmander type generators in infinite dimensions*, Potential Analysis (2012), 1–29.

DLP+18. M.H. Duong, A. Lamacz, M.A. Peletier, A. Schlichting, and U. Sharma, *Quantification of coarse-graining error in Langevin and overdamped Langevin dynamics*, Nonlinearity **31** (2018), no. 10, 4517.

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 mathematiques **134** (2010), no. 4, 405–422.

FJK13. G. Froyland, O. Junge, and P. Koltai, *Estimating long-term behavior of flows without trajectory integration: the infinitesimal generator approach*, SIAM Journal on Numerical Analysis **51** (2013), no. 1, 223–247.

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.

Fro01. ———, *Extracting dynamical behavior via markov models*, Nonlinear dynamics and statistics, Springer, 2001, pp. 281–321.

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.

GKS04. D. Givon, R. Kupferman, and A. Stuart, *Extracting macroscopic dynamics: model problems and algorithms*, Nonlinearity **17** (2004), no. 6, R55.

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.

GL20. S. Galatolo and R. Lucena, *Spectral gap and quantitative statistical stability for systems with contracting fibers and Lorenz-like maps*, Disc. Cont. Dyn. Syst. A **40** (2020), no. 3, 1309.

GLP13. P. Giulietti, C. Liverani, and M. Pollicott, *Anosov flows and dynamical zeta functions*, Annals of Mathematics (2013), 687–773.

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. Zegarlinksi, *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.

KCB18. D. Kondrashov, M. D. Chekroun, and P. Berloff, *Multiscale stuart-landau emulators: Application to wind-driven ocean gyres*, Fluids **3** (2018), no. 21.

KCG15. D. Kondrashov, M. D. Chekroun, and M. Ghil, *Data-driven non-Markovian closure models*, Physica D: Nonlinear Phenomena **297** (2015), 33–55.

KCG18. D. Kondrashov, M. D. Chekroun, and M. Ghil, *Data-adaptive harmonic decomposition and prediction of Arctic sea ice extent*, Dynamics and Statistics of the Climate System **3** (2018), no. 1, 1–23.

KCYG18. D. Kondrashov, M. D. Chekroun, X. Yuan, and M. Ghil, *Data-adaptive harmonic decomposition and stochastic modeling of Arctic sea ice*, Advances in Nonlinear Geosciences (A. Tsonis, ed.), Springer, 2018, pp. 179–205.

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.

LLO17. F. Legoll, T. Lelièvre, and S. Olla, *Pathwise estimates for an effective dynamics*, Stochastic Processes and their Applications **127** (2017), no. 9, 2841–2863.

LSY97. R. 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.

Luc18. V. Lucarini, *Revising and extending the linear response theory for statistical mechanical systems: Evaluating observables as predictors and predictands*, J. Stat. Phys. **173** (2018), no. 6, 1698–1721.

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.

MG07. I. Melbourne and G.A. Gottwald, *Power spectra for deterministic chaotic dynamical systems*, Nonlinearity **21** (2007), no. 1, 179.

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.

MT16. A.J. Majda and X.T. Tong, *Ergodicity of truncated stochastic Navier Stokes with deterministic forcing and dispersion*, Journal of Nonlinear Science **26** (2016), no. 5, 1483–1506.

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.

NZ15. S. Nonnenmacher and M. Zworski, *Decay of correlations for normally hyperbolic trapping*, Inventiones mathematicae **200** (2015), no. 2, 345–438.

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.

PS95. C. Penland and P. D. Sardeshmukh, *The optimal growth of tropical sea surface temperature anomalies*, Journal of climate **8** (1995), no. 8, 1999–2024.

PS08. G. Pavliotis and A. Stuart, *Multiscale methods: Averaging and Homogenization*, Springer Science & Business Media, 2008.

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.

Rue05. ———, *Differentiating the absolutely continuous invariant measure of an interval map f with respect to f*, Communications in mathematical physics **258** (2005), no. 2, 445–453.

Rue09. ———, *A review of linear response theory for general differentiable dynamical systems*, Nonlinearity **22** (2009), no. 4, 855.

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. Deuflhard, *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. Deuflhard, *Transfer operator approach to conformational dynamics in biomolecular systems*, Ergodic theory, analysis, and efficient simulation of dynamical systems, Springer, 2001, pp. 191–223.

SS13. Ch. Schütte and M. Sarich, *Metastability and Markov State Models in Molecular Dynamics*, vol. 24, American Mathematical Soc., 2013.

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.

SV72. D. W. Stroock and S. R. S. 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, Volume 3: Probability Theory (Berkeley, Calif.), University of California Press, 1972, pp. 333–359.

Tam95. C.K.W. Tam, *Supersonic jet noise*, Annual Review of Fluid Mechanics **27** (1995), no. 1, 17–43.
TCND19. A. Tantet, M. D. Chekroun, J. D. Neelin, and H. A. Dijkstra, *Ruelle-Pollicott Resonances of Stochastic Systems in Reduced State Space. Part III: Application to El Niño-Southern Oscillation*, J. Stat. Phys. (2019), doi: 10.1007/s10955-019-02444-8.
TCND20. ———, *Ruelle-Pollicott Resonances of Stochastic Systems in Reduced State Space. Part II: Stochastic Hopf Bifurcation*, J. Stat. Phys. (Accepted) (2020).
TLLD18. A. Tantet, V. Lucarini, F. Lunkeit, and H.A. Dijkstra, *Crisis of the chaotic attractor of a climate model: a transfer operator approach*, Nonlinearity **31** (2018), no. 5, 2221.
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.
Ula64. S. M. Ulam, *Problems in Modern Mathematics*, science ed., Wiley, New York, 1964.
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. ———, *Multi-level dynamical systems: Connecting the Ruelle response theory and the Mori-Zwanzig approach*, J. Stat. Phys. **151** (2013), no. 5, 850–860.
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.
ZHS17. W. Zhang, C. Hartmann, and Ch. Schütte, *Effective dynamics along given reaction coordinates, and reaction rate theory*, Faraday discussions **195** (2017), 365–394.
Zwo17. M. Zworski, *Mathematical study of scattering resonances*, Bulletin of Mathematical Sciences **7** (2017), no. 1, 1–85.