University of Reading

Edgeworth expansions for slow-fast systems with finite time scale separation

Article

Accepted Version

Wouters, J. ORCID: https://orcid.org/0000-0001-5418-7657 and Gottwald, G. A. (2019) Edgeworth expansions for slow-fast systems with finite time scale separation. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 475 (2223). 20180358. ISSN 1364-5021 doi: 10.1098/rspa.2018.0358 Available at https://centaur.reading.ac.uk/81424/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1098/rspa.2018.0358

Publisher: Royal Society Publishing

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

PROCEEDINGS A

rspa.royalsocietypublishing.org

Research

Article submitted to journal

Subject Areas:
70K70, 65C20, 37A50, 60F05

Keywords:
multi-scale systems, homogenization, Edgeworth expansion, stochastic limit systems

Author for correspondence:
Jeroen Wouters
e-mail: j.wouters@reading.ac.uk

Edgeworth expansions for slow-fast systems with finite time scale separation

Jeroen Wouters[1,2] Georg A. Gottwald[3]

[1]Department of Mathematics and Statistics, University of Reading, Reading, United Kingdom
[2]Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark
[3]School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia

We derive Edgeworth expansions that describe corrections to the Gaussian limiting behaviour of slow-fast systems. The Edgeworth expansion is achieved using a semi-group formalism for the transfer operator, where a Duhamel-Dyson series is used to asymptotically determine the corrections at any desired order of the time scale parameter ε. The corrections involve integrals over higher-order auto-correlation functions. We develop a diagrammatic representation of the series to control the combinatorial wealth of the asymptotic expansion in ε and provide explicit expressions for the first two orders. At a formal level, the expressions derived are valid in the case when the fast dynamics is stochastic as well as when the fast dynamics is entirely deterministic. We corroborate our analytical results with numerical simulations and show that our method provides an improvement on the classical homogenization limit which is restricted to the limit of infinite time scale separation.

1. Introduction

Many systems in the natural sciences feature a time scale separation between slowly and rapidly evolving variables. Examples range from molecular drug design where a protein interacts with a target molecule in the presence of a rapidly fluctuating environment [1], to climate dynamics where the slowly evolving ocean dynamics is driven by rapidly evolving atmospheric weather systems [2].

Such systems can often be modeled by multi-scale systems of the form

THE ROYAL SOCIETY PUBLISHING

© The Author(s) Published by the Royal Society. All rights reserved.

$$\mathrm{d}x = \frac{1}{\varepsilon} f_0(x,y)\mathrm{d}t + f_1(x,y)\mathrm{d}t \tag{1.1}$$

$$\mathrm{d}y = \frac{1}{\varepsilon^2} g_0(y)\mathrm{d}t + \frac{1}{\varepsilon}\beta(y)\mathrm{d}W_t + \frac{1}{\varepsilon} g_1(x,y)\mathrm{d}t. \tag{1.2}$$

with $x \in \mathbb{R}^d$, $y \in \mathbb{R}^m$ and $\beta \in \mathbb{R}^{m\times l}$ and l-dimensional Wiener process W_t. The fast dynamics can be stochastic with $\beta \neq 0$ or deterministic with $\beta \equiv 0$. In the stochastic case we assume that the fast dynamics $\mathrm{d}y = g_0\mathrm{d}t + \beta\mathrm{d}W_t$ is an ergodic process with an absolutely continuous measure μ and the full system (1.1)-(1.2) admits an absolutely continuous measure $\mu^{(\varepsilon)}$. In the purely deterministic case $\beta \equiv 0$ we assume that the fast dynamics $\mathrm{d}y = g_0\mathrm{d}t$ admits a unique invariant physical measure μ on $\mathbb{R}^m$ and the full system (1.1)-(1.2) admits a unique invariant physical measure $\mu^{(\varepsilon)}$ on $\mathbb{R}^{d+m}$ [1]. Here $\varepsilon \ll 1$ denotes the degree of time scale separation between the slow and fast variables, x and y, respectively. Often only the slow variables are of interest in such systems, and one seeks reduced equations for the slow dynamics only.

In the limit of infinite time scale separation $\varepsilon \to 0$, closed diffusion equations for the slow variables can be obtained by the method of homogenization [3–9]. The diffusive behaviour emerges as the integrated effect of the fast dynamics, reminiscent of the summation of random variables in the central limit theorem (CLT). This method applies to slow-fast systems, where the fast dynamics can be either stochastic or deterministic with $\beta \equiv 0$. In the deterministic case the fast dynamics needs to be sufficiently chaotic. If the leading order slow vector field averages to zero, i.e. if $\int f_0(x,y)\mu(\mathrm{d}y) = 0$, the slow dynamics is approximated on time scales of $\mathcal{O}(1)$ by a stochastic differential equation; see [5–7] for the stochastic case and [10–12] and [13–15] for the deterministic case. Homogenization has been used to design efficient numerical multi-scale integrators such as equation-free projection [16,17] and the heterogeneous multi-scale method [18,19], and has been used for stochastic parameterization in the climate sciences [20–26].

In realistic physical systems, however, the time scale separation is always finite. In systems without a clear time scale separation classical homogenization theory may fail and may not be able to reliably approximate the stochastic long-time behaviour of the slow dynamics. For finite values of ε there is an intricate feedback between the evolution of the slow x variables and the fast y variables which prevents the approximation of the integrated fast dynamics by Brownian motion. Similar issues arise in the CLT, where the Gaussian distribution is only an accurate approximation for sums with a sufficiently large number of summands. For the CLT techniques exist to obtain a more accurate description of the distribution of finite sums than provided by the limiting Gaussian. A classical technique is the Edgeworth expansion, which provides an expansion of the distributions of sums, asymptotic in $1/\sqrt{n}$, where n is the length of the sum [27,28].

In the case of a multi-scale system such as (1.1)-(1.2) the small parameter controlling the limit is now ε, instead of $1/\sqrt{n}$, and the random variables converging to a Gaussian are the increments over time of the slow variable x. The multi-scale system (1.1)-(1.2) features three distinct time scales [29]: the fast time scale of $\mathcal{O}(\varepsilon^2)$, an intermediate time scale of $\mathcal{O}(\varepsilon)$ on which the slow dynamics is trivial but the fast dynamics has equilibrated, and the long diffusive time scale of $\mathcal{O}(1)$ on which the slow dynamics exhibits nontrivial diffusive behaviour. The corrections to Gaussianity occur on the intermediate time scale; it is sufficiently long for the integrated noise on x to become nearly Gaussian, but not long enough for the slow dynamics to dominate. To focus on the statistical behaviour on the intermediate time scale we consider the transition probabilities

[1] An ergodic measure is called physical if for a set of initial conditions of nonzero Lebesgue measure the temporal average of a continuous observable converges to the spatial average over this measure.

of the slow variable x

$$\pi_\varepsilon(\mathrm{x},t,x_0)=\mathbb{P}\left(\frac{x(t)-x(0)}{\sqrt{t}}\in(\mathrm{x},\mathrm{x}+\mathrm{dx})\middle|\, x(0)=x_0,\, y(0)\sim\mu^{(\varepsilon)}_{x_0}\right),$$

where $\mu^{(\varepsilon)}_{x_0}$ is the invariant measure of (1.1)-(1.2) conditioned on $x=x_0$ and $x(t)$ the slow variable of a solution of the multi-scale system (1.1)-(1.2).

Homogenization dictates that on the intermediate time scale and in the limit $\varepsilon\to 0$ of infinite time scale separation, $\pi_\varepsilon(t=\varepsilon)$ becomes Gaussian. We will refer to this convergence as the CLT in the context of slow-fast systems. For small but finite ε, the deviations from Gaussianity of π_ε will be small. We can therefore expand $\pi_\varepsilon(\mathrm{x},\varepsilon,x_0)$ in $\sqrt{\varepsilon}$. This expansion is the equivalent of the classical Edgeworth expansion for slow-fast systems.

Whereas the limiting Gaussian probability, implied by homogenization theory, only involves the two-time statistics, higher-order Edgeworth expansions involve higher-order time correlations, containing more information about the dynamics. This constitutes our main result:

(a) Main result

We assume that the fast dynamics of the multi-scale system (1.1)-(1.2) decorrelates sufficiently rapidly, such that higher-order correlation functions are integrable and all expectation values appearing in the formulae below exist. Furthermore we assume that the conditional invariant measure $\mu^{(\varepsilon)}_{x_0}$ obeys linear response w.r.t ε, and $\int f_0\,\mu(dy)=0$. Then the expansion is given up to $\mathcal{O}(\varepsilon^{\frac{3}{2}})$ in the limit $\varepsilon\to 0$ with $t=\varepsilon\ll 1$ by

$$\begin{aligned}\pi_\varepsilon(\mathrm{x},t=\varepsilon,x_0)=\mathbf{n}_{0,\sigma^2}(\mathrm{x})\Bigg(1+\sqrt{\varepsilon}\left(\frac{c^{(1)}_{\frac12}}{\sigma}H_1\left(\frac{\mathrm{x}}{\sigma}\right)+\frac{c^{(3)}_{\frac12}}{3!\sigma^3}H_3\left(\frac{\mathrm{x}}{\sigma}\right)\right)\\ +\varepsilon\left(\frac{c^{(2)}_1+{c^{(1)}_{\frac12}}^2}{2\sigma^2}H_2\left(\frac{\mathrm{x}}{\sigma}\right)+\frac{c^{(4)}_1+4c^{(1)}_{\frac12}c^{(3)}_{\frac12}}{4!\sigma^4}H_4\left(\frac{\mathrm{x}}{\sigma}\right)+\frac{{c^{(3)}_{\frac12}}^2}{2(3!\sigma^3)^2}H_6\left(\frac{\mathrm{x}}{\sigma}\right)\right)\Bigg)\\ +\mathcal{O}(\varepsilon^{\frac{3}{2}}),\end{aligned}\tag{1.3}$$

with $c^{(p)}_{\frac12}=c^{(p)}_{0,\frac12}+c^{(p)}_{1,-\frac12}$ and $c^{(p)}_1=c^{(p)}_{0,1}+c^{(p)}_{1,0}+c^{(p)}_{2,-1}$, with expressions given in equations (4.9), (4.11), (4.12), (4.13), (4.15), (4.16), (4.18), (4.19) and (4.20), and where $H_n(\mathrm{x})=(\mathrm{x}-\frac{\mathrm{d}}{\mathrm{dx}})^n 1$ are Hermite polynomials of degree n.

Establishing this expansion involves expanding the first four cumulants $c^{(p)}$ of $(x(t)-x(0))/\sqrt{t}$ with $p\le 4$, resulting in their expansion coefficients $c^{(p)}_k$ with $k\in\{\frac12,1\}$ (see Section 4). These coefficients only involve the leading order measure $\mu^{(0)}_{x_0}=\mu$ and, in particular, do not involve the linear response correction to $\mu^{(\varepsilon)}_{x_0}$. It should be noted that the expressions for the cumulant expansion coefficients $c^{(p)}_{(i,j)}$ as derived below determine the functional form of the expansion, but are not sufficient to show that an Edgeworth expansion holds for a given class of dynamical systems.

(b) Plan of the paper

The paper is organized as follows. In Section 2 we briefly review homogenization theory highlighting the rôle of infinite time scale separation in convergence of transition probabilities to Gaussian distributions. Section 3 reviews Edgeworth expansions for discrete stochastic systems. Section 4 introduces Edgeworth expansions and finite time scale separation corrections to the CLT for dynamical multi-scale systems and presents the explicit expression of the Edgeworth

expansion; the lengthy and involved derivation is provided in the Appendix A. In Section 5 we present a numerical example corroborating our analytical main result for the transition probability and show that the Edgeworth expansion improves on the CLT. We conclude with a summary and an outlook in Section 6.

2. Homogenization

Homogenization describes the integrated effect of fast (either stochastic or deterministic chaotic) dynamics on slow variables as noise. Initially developed for stochastic multi-scale systems [5–7], homogenization has been extended recently to deterministic multi-scale systems when the fast dynamics evolves on a compact attractor $\Lambda \subset \mathbb{R}^m$ with an ergodic invariant measure μ. It was shown rigorously that for sufficiently chaotic fast dynamics the emergent stochastic long time behaviour of the slow dynamics is given by stochastic differential equations driven by Brownian motion [10–12,30]. While homogenization for stochastic systems has been proven in a very general setting, the results in the deterministic case are so far limited to the skew product case with $g_1 = 0$. The assumptions on the chaoticity of the fast subsystem are mild, including Axiom A diffeomorphisms and flows, Hénon-like attractors and Lorenz attractors.

The following heuristic argument serves to show how the diffusive behaviour of the slow dynamics is linked to the CLT for time-integrated stationary processes in the case when the dynamical system is entirely deterministic. Consider the simplified version of the multi-scale system (1.1)-(1.2) with $x \in \mathbb{R}$ and $y \in \mathbb{R}^m$

$$\dot{x} = \frac{1}{\varepsilon} f_0(y)$$
$$\dot{y} = \frac{1}{\varepsilon^2} g_0(y).$$

Integrating the slow dynamics leads to

$$x(t) = x(0) + \frac{1}{\varepsilon} \int_0^t f_0(y(s))\, \mathrm{d}s.$$

Transforming the integrand to the fast time scale $\tau = s/\varepsilon^2$ we obtain

$$x(t) = x(0) + W_\varepsilon(t),$$

where we introduced

$$W_\varepsilon(t) = \varepsilon \int_0^{\frac{t}{\varepsilon^2}} f_0(y_0(\tau))\, \mathrm{d}\tau, \tag{2.1}$$

with rescaled fast dynamics $\dot{y}_0 = g_0(y_0)$. For $\int f_0(y)\mu(\mathrm{d}y) = 0$ the integral is collecting weakly dependent variables with mean zero, provided the fast dynamics is sufficiently chaotic. The integral term in (2.1) is formally of the form of the CLT where $n = 1/\varepsilon^2$ terms are integrated and then scaled by $1/\sqrt{n} = \varepsilon$, and, assuming the CLT holds for y, converges weakly to Brownian motion W_t on the long diffusive time scale $t = \mathcal{O}(1)$. The slow dynamics is approximated by

$$\mathrm{d}X = \sigma \mathrm{d}W_t.$$

By explicitly calculating $\lim_{\varepsilon \to 0} \int_\Lambda \mu(\mathrm{d}y) W_\varepsilon(t)^2$, we obtain a Green-Kubo formula for the variance with

$$\frac{1}{2}\sigma^2 = \int_0^\infty \mathrm{d}s \int_\Lambda \mu(\mathrm{d}y)\, f_0(y) f_0(\varphi^s y),$$

where φ^s denotes the flow map of the fast dynamics $\dot{y}_0 = g_0(y_0)$. Note that in the deterministic case, randomness is only introduced through the choice of the initial condition $y(0)$.

This heuristic argument can be made rigorous. It can be shown that a functional central limit theorem exists and solutions of the multi-scale system (1.1)-(1.2) converge weakly to solutions the

homogenized Itô stochastic differential equation

$$dX = F(X)\mathrm{d}t + \sigma(X)\,\mathrm{d}W_t\,, \tag{2.2}$$

where W_t denotes l-dimensional Brownian motion. The drift coefficient $F:\mathbb{R}^d\to\mathbb{R}^d$ is given by

$$F(x) = \int f_1(x,y)\,\mu(\mathrm{d}y) + \int_0^\infty \mathrm{d}s \int \mathbb{E}\left[f_0(x,y)\cdot\partial_x f_0(x,\varphi^t y)\right]\,\mu(\mathrm{d}y)$$
$$+ \int_0^\infty \mathrm{d}s \int \mathbb{E}\left[g_1(x,y)\cdot\partial_y\left(f_0(x,\varphi^t y)\right)\right]\,\mu(\mathrm{d}y), \tag{2.3}$$

with μ the ergodic invariant measure corresponding to $\mathrm{d}y = g_0(y)\mathrm{d}t + \beta(y)\mathrm{d}W_t$ and $\mathbb{E}$ the expectation value w.r.t. the Wiener measure on W_t in the stochastically driven case ($\beta\neq 0$). The diffusion coefficient $\sigma:\mathbb{R}^d\to\mathbb{R}^{d\times l}$ is given by the Green-Kubo formula

$$\sigma(x)\sigma^T(x) = \int_0^\infty \mathrm{d}s \int \mathbb{E}\left(f_0(x,y)\otimes f_0(x,\varphi^t y) + f_0(x,\varphi^t y)\otimes f_0(x,y)\right)\,\mu(\mathrm{d}y), \tag{2.4}$$

where the outer product between two vectors is defined as $(a\otimes b)_{ij} = a_i b_j$. These expressions for the drift and diffusion can be derived formally by an asymptotic expansion of the backward Kolmogorov equation [2]. We remark that one can add a stochastic driver to the slow dynamics in (1.1) which would lead to an additively increased diffusion (2.4) (cf. [52]). For simplicity of exposition we do not consider this case here.

In the deterministic case $\beta = 0$, homogenization results [10–12] assure that the family of solutions of the original multi-scale dynamical system converge weakly in the sup-norm topology to the unique solution X of the reduced stochastic differential equation (2.2) as $\varepsilon\to 0$. At the heart of diffusive limits of deterministic dynamical multi-scale systems lies a functional CLT which assures that $W_\varepsilon(t)\to_w W(t)$ in $C([0,\infty),\mathbb{R}^d)$ in the limit $\varepsilon\to 0$. The functional CLT implies the CLT but not vice versa. We are, however, only aware of a single example where a judiciously chosen unbounded observable of a deterministic dynamical system satisfies a CLT but not the associated functional CLT [31].

In the following we evaluate corrections to the CLT for increments of x by probing the finite ε corrections of the transition probability in an Edgeworth expansion.

3. The Edgeworth expansion and corrections to the central limit theorem

Before introducing Edgeworth expansions for continuous time multi-scale systems, it is instructive to briefly review the case of Edgeworth expansions in the discrete time stochastic case. The central limit theorem describes when appropriately scaled sums

$$S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n (y_i - m)$$

of variables y_i with mean m and variance σ^2 converge in distribution to a normal distribution in the sense that

$$\mathbb{P}(a\le S_n\le b)\to \frac{1}{\sqrt{2\pi\sigma^2}}\int_a^b e^{-\frac{s^2}{2\sigma^2}}\,\mathrm{d}s$$

as $n\to\infty$ [28,32]. It is valid for *i.i.d.* random variables, weakly dependent random variables [28, 32], as well as for a large class of dynamical systems [13,33–36]. Edgeworth expansions describe deviations from the CLT for finite n. We briefly review in the next subsection the well studied case of Edgeworth expansions for stochastic random variables.

[2]Strictly speaking, these formulae are only valid for correlation functions which are slightly more than integrable. For fast systems with decaying autocorrelation functions which are only integrable, one can find expressions for the drift and diffusion coefficients which are, however, more complicated; see [12] for details.

(a) Edgeworth expansions for stochastic random variables

In the stochastic context Edgeworth expansions are usually derived in the spectral framework by studying the characteristic function $\chi_n(\xi) = \mathbb{E}[\exp(i\xi S_n)]$ of the random variable S_n ($\mathbb{E}$ denotes in this subsection the expectation with respect to μ, the distribution of y). The characteristic function χ_n is related to the cumulants $c_n^{(j)}$ of S_n by $\chi_n(\xi) = \exp(\sum_{j=1}^{\infty} \frac{(i\xi)^j}{j!} c_n^{(j)})$. Assuming that $c_n^{(1)} = \mathbb{E}[S_n] = 0$, we have for weakly dependent random variables with uniform or strong mixing properties that $c_n^{(2)} = c_\infty^{(2)} + \frac{1}{n}\delta c^{(2)} + o(n^{-1})$, $c_n^{(3)} = \frac{1}{\sqrt{n}} c_\infty^{(3)} + o(n^{-1})$ and $c_n^{(4)} = \frac{1}{n} c_\infty^{(4)} + o(n^{-1})$, while higher cumulants are of higher order in $1/\sqrt{n}$. Specifically, up to order $o(n^{-1/2})$ the expansions of the cumulants are related to the correlation functions of the parent random variables y_i by

$$c_n^{(2)} = \sigma^2 - 2\frac{1}{n}\sum_{j=1}^{\infty} j\mathbb{E}[y_1 y_{j+1}]$$

$$\sqrt{n} c_n^{(3)} = \mathbb{E}[y_1^3] + 3\sum_{j=1}^{\infty}\left(\mathbb{E}[y_1^2 y_{j+1}] + \mathbb{E}[y_1 y_{j+1}^2]\right) + 6\sum_{i,j=1}^{\infty}\mathbb{E}[y_1 y_{1+i} y_{1+i+j}]$$

with $\sigma^2 = \mathbb{E}[y_1^2] + 2\sum_{j=1}^{\infty}\mathbb{E}[y_1 y_j]$ given by the Green-Kubo formula [37]. The above equations implicitly define the terms $c_\infty^{(2)}$, $\delta c^{(2)}$ and $c_\infty^{(3)}$. Inserting the expansions of the cumulants into the characteristic function, we obtain

$$\begin{aligned}
\chi_n(\xi) &= \exp\left(c_n^{(2)}\frac{(i\xi)^2}{2!} + c_n^{(3)}\frac{(i\xi)^3}{3!} + c_n^{(4)}\frac{(i\xi)^4}{4!} + \dots\right) \\
&= \exp\left(\frac{1}{\sqrt{n}} c_\infty^{(3)}\frac{(i\xi)^3}{3!} + \frac{1}{n}\left(\delta c^{(2)}\frac{(i\xi)^2}{2!} + c_\infty^{(4)}\frac{(i\xi)^4}{4!}\right) + \mathcal{O}\left(\frac{1}{n^{3/2}}\right)\right) \times \exp\left(-c_\infty^{(2)}\frac{\xi^2}{2}\right) \\
&= \left(1 + \frac{1}{\sqrt{n}} c_\infty^{(3)}\frac{(i\xi)^3}{3!} + \frac{1}{n}\left(\delta c^{(2)}\frac{(i\xi)^2}{2!} + c_\infty^{(4)}\frac{(i\xi)^4}{4!}\right)\right. \\
&\qquad \left. + \frac{1}{2}\left(\frac{1}{\sqrt{n}} c_\infty^{(3)}\frac{(i\xi)^3}{3!}\right)^2 + \mathcal{O}\left(\frac{1}{n^{3/2}}\right)\right) \times \exp\left(-c_\infty^{(2)}\frac{\xi^2}{2}\right).
\end{aligned}$$

The characteristic function therefore converges pointwise to $\exp(-c_\infty^{(2)}\xi^2/2)$, which is the characteristic function of a Gaussian with variance $c_\infty^{(2)} = \sigma^2$. Since, by Lévy's continuity theorem, pointwise convergence of the characteristic functions is equivalent to convergence in distribution, the CLT follows. An expansion in $\frac{1}{\sqrt{n}}$ of the probability density function (pdf) of S_n is obtained by the inverse Fourier transform of the characteristic function $\chi_n(\xi)$. Under the inverse Fourier transform the terms $(i\xi)^k$ in the expansion become k-th derivatives of the normal distribution, i.e. Hermite polynomials. Let ρ_n be the pdf of the normalized Birkhoff sum S_n. Then, the first two Edgeworth approximations $\rho_n^{(1)}$ and $\rho_n^{(2)}$ of the probability density ρ_n of S_n are

$$\rho_n^{(1)}(\mathrm{x}) = \mathbf{n}_{0,\sigma^2}(\mathrm{x})\left(1 + \frac{c_\infty^{(3)}}{6\sigma^3\sqrt{n}} H_3\left(\frac{\mathrm{x}}{\sigma}\right)\right), \tag{3.1}$$

and

$$\rho_n^{(2)}(\mathrm{x}) = \mathbf{n}_{0,\sigma^2}(\mathrm{x})\left(1 + \frac{c_\infty^{(3)}}{6\sigma^3\sqrt{n}} H_3\left(\frac{\mathrm{x}}{\sigma}\right) + \frac{\delta c^{(2)}}{2\sigma^2 n} H_2\left(\frac{\mathrm{x}}{\sigma}\right) + \frac{c_\infty^{(4)}}{24\sigma^4 n} H_4\left(\frac{\mathrm{x}}{\sigma}\right) + \frac{{c_\infty^{(3)}}^2}{72\sigma^6 n} H_6\left(\frac{\mathrm{x}}{\sigma}\right)\right),$$

where H_k is the k-th Hermite polynomial. We have $\rho_n(\mathrm{x}) = \rho_n^{(1)}(\mathrm{x}) + o\left(\frac{1}{\sqrt{n}}\right)$ and $\rho_n(\mathrm{x}) = \rho_n^{(2)}(\mathrm{x}) + o\left(\frac{1}{n}\right)$ uniformly in x [28]. The Edgeworth approximation generally yields an improved approximation of the pdf around the mean of the distribution [38]. Note that $\rho_n^{(p)}$ is no longer nonnegative or normalized and therefore the approximation $\rho_n^{(p)}$ is no longer a

probability density function. In contrast to the Gaussian distribution $\mathbf{n}_{0,\sigma^2}(\mathrm{x})$, the first Edgeworth approximation may have a non-zero third moment $c_\infty^{(3)}/\sqrt{n}$, which vanishes as $n \to \infty$. Higher order approximations $\rho_n^{(p)}$ can be derived by increasing the order of the Taylor series.

Interestingly, the functional form of the expansion is universal, in the sense that the parent process y_i only enters through the asymptotic cumulants $c_\infty^{(2)} = \sigma^2$, $\delta c^{(2)}$, $c_\infty^{(3)}$ and $c_\infty^{(4)}$.

In the stochastic context, Edgeworth expansions have been obtained for time series such as ARMA processes [39], continuous-time diffusions [40] and, employing the Nagaev-Guivarc'h method for the characteristic function, for ergodic Markov chains [41]. Edgeworth expansions have, to the best of our knowledge, not been explored in the multi-scale context. The next sections aim at filling this gap.

4. Edgeworth expansions for continuous-time multi-scale systems

We now determine Edgeworth corrections for multi-scale systems. The formulae will be given in terms of the generator of the fast process

$$\mathcal{L}_0 = g_0(y) \cdot \partial_y + \frac{1}{2}\beta(y)\beta^T(y) : \partial_y \cdot \partial_y \tag{4.1}$$

with L^2-adjoint $\mathcal{L}_0^*$ which acts as

$$\mathcal{L}_0^*\nu = -\partial_y \cdot (g_0(y)\nu) + \frac{1}{2}\partial_y \cdot \partial_y \cdot \left(\beta(y)\beta^T(y)\,\nu\right). \tag{4.2}$$

As described in the introduction, multi-scale systems (1.1)-(1.2) are characterized by three distinct time scales: the fast time scale of $\mathcal{O}(\varepsilon^2)$, an intermediate time scale of $\mathcal{O}(\varepsilon)$ on which the slow dynamics is trivial but the fast dynamics has equilibrated, and the long diffusive time scale of $\mathcal{O}(1)$ on which the slow dynamics exhibits nontrivial diffusive behaviour. The particular rôle of the intermediate time scale to control the normality of the noise is formally reflected in the homogenized stochastic limit system (2.2) which evolves on the diffusive time scale of $\mathcal{O}(1)$; since $\mathrm{d}W_t$ scales like $\sqrt{t}$ the Brownian motion is dominant on time scales of $\mathcal{O}(\varepsilon)$. On this time scale, the transition probability π_h of the diffusive homogenized limit system (2.2) satisfies the CLT and converges according to

$$\pi_h(\mathrm{x}, t, x_0) = \mathbb{P}\left(\frac{X(t) - x_0}{\sqrt{t}} \in (\mathrm{x}, \mathrm{x} + \mathrm{dx}) \,\middle|\, X(0) = x_0\right) \to \mathbf{n}_{0,\sigma^2}(\mathrm{x}),$$

where we let $t \to 0$ since t is of $\mathcal{O}(\varepsilon)$ and where X solves the homogenized stochastic differential equation (2.2). To study deviations from the Gaussian behaviour of the limit $\varepsilon \to 0$, we develop in Appendix A a semi-group formalism to calculate the Edgeworth expansion of the intermediate-time transition probabilities of the slow variable x of the multi-scale system (1.1)-(1.2)

$$\pi_\varepsilon(\mathrm{x}, t, x_0) = \mathbb{P}\left(\frac{x(t) - x_0}{\sqrt{t}} \in (\mathrm{x}, \mathrm{x} + \mathrm{dx}) \,\middle|\, x(0) = x_0,\ y(0) \sim \mu_{x_0}^{(\varepsilon)}\right),$$

for $t = \varepsilon$, and expand $\pi_\varepsilon(\mathrm{x}, \varepsilon, x_0)$ in $\sqrt{\varepsilon}$. Here $\mu_{x_0}^{(\varepsilon)}(y)$ is the invariant measure $\mu^{(\varepsilon)}(x, y)$ of (1.1)-(1.2) conditioned on $x = x_0$. By the Rokhlin disintegration theorem [42] the conditional measure is essentially unique and we furthermore assume that $\mu_{x_0}^{(\varepsilon)}$ obeys linear response w.r.t. ε, such that the conditional measure can be expanded in ε around μ. Whereas the coefficients in the homogenized equation only involve the two-time statistics, higher-order Edgeworth expansions involve higher-order time correlations, containing more information about the dynamics.

As in the case of random variables described in the previous section, the Edgeworth expansion of the intermediate-time transition probability π_ε can be calculated by the asymptotic expansion of its associated characteristic function which is entirely determined by the cumulants. We therefore set out to asymptotically calculate the p^{th}-moments of the slow variables $(x(t) - x_0)/\sqrt{t}$

in orders of $\sqrt{\varepsilon}$

$$m^{(p)} = \mathbb{E}^{x_0,\mu}\left[\left(\frac{\hat{x}(t)}{\sqrt{t}}\right)^p\right], \tag{4.3}$$

with $\hat{x}(t) = x(t) - x_0$ on the intermediate time scale $t = \varepsilon$ and the corresponding cumulants $c^{(p)}$. The conditional average $\mathbb{E}^{x_0,\mu}$ is with respect to the product measure

$$\mu^{(\varepsilon)}(dx, dy) = \left(\delta_{x_0} \times \mu_{x_0}^{(\varepsilon)}\right)(dx, dy). \tag{4.4}$$

The measure $\mu_{x_0}^{(\varepsilon)}$ depends on ε for non-skew product systems where the fast dynamics depends on the slow dynamics. Assuming that $\mu_{x_0}^{(\varepsilon)}$ obeys linear response w.r.t. ε we expand the measure $\mu_{x_0}^{(\varepsilon)}$ as

$$\mu_{x_0}^{(\varepsilon)} = \mu_{x_0}^{(0)} + \varepsilon\mu_{x_0}^{(1)} + \mathcal{O}(\varepsilon^2), \tag{4.5}$$

with $\mu_{x_0}^{(0)} = \mu$ the invariant measure of the fast process $\mathrm{d}y = g_0(y)\mathrm{d}t + \beta(y)\mathrm{d}W_t$. We note that, while linear response has been proven for a wide class of deterministic and stochastic systems and has been used for model reduction [43–45], counter-examples do exist [46–51].

We seek expansions of the scaled moments $m^{(p)}(t)$ in ε and in t. We take the limit $t = \varepsilon \ll 1$ and $t/\varepsilon^2 \to \infty$ as $\varepsilon \to 0$. In this limit the random variable $\hat{x}(t)/\sqrt{t}$ converges to a normal distribution and the coefficients of the different powers of $\sqrt{\varepsilon}$ appearing in the expansions of its cumulants will provide the desired Edgeworth corrections coefficients. We expand the p^{th} moments as

$$m^{(p)} = m_0^{(p)} + \sum_{|\boldsymbol{\alpha}|>0} \varepsilon^{\alpha_\varepsilon} t^{\alpha_t} m_{\boldsymbol{\alpha}}^{(p)}, \tag{4.6}$$

where we use multi-index notation to denote the expansions in ε and t with $\boldsymbol{\alpha} = (\alpha_\varepsilon, \alpha_t)$ and with $|\boldsymbol{\alpha}| = \alpha_\varepsilon + \alpha_t$ being the combined order of the contribution. We similarly expand the cumulant as

$$c^{(p)} = c_0^{(p)} + \sum_{|\boldsymbol{\alpha}|>0} \varepsilon^{\alpha_\varepsilon} t^{\alpha_t} c_{\boldsymbol{\alpha}}^{(p)}. \tag{4.7}$$

Note that $\alpha_{\varepsilon,t} < 0$ and $\alpha_{\varepsilon,t}$ half-integer is allowed.

Our expansion of the transition probability $\pi_\varepsilon(\mathrm{x}, t = \varepsilon, x_0)$ (1.3) involves the first four cumulants. The derivation of those cumulant expansions can be found in Appendix A. Here we only state the resulting formulae. The first cumulant is given up to order $\mathcal{O}(\varepsilon^{\frac{3}{2}})$ by

$$c^{(1)} = \sqrt{t}\, c_{0,\frac{1}{2}}^{(1)}(x_0) + R_\varepsilon^{(1)}, \tag{4.8}$$

where the remainders $R_\varepsilon^{(j)} = \sum_{|\boldsymbol{\alpha}|>1} \varepsilon^{\alpha_\varepsilon} t^{\alpha_t} c_{\boldsymbol{\alpha}}^{(j)}$ for $j = 1, \cdots, p$ consist of higher order terms and

$$c_{0,\frac{1}{2}}^{(1)} = F(x) = \langle f_1 \rangle - \langle f_0 \mathcal{L}_{0\perp}^{-1} \partial_x f_0 \rangle - \langle (g_1 \partial_y) \mathcal{L}_{0\perp}^{-1} f_0 \rangle, \tag{4.9}$$

recovering the drift coefficient $F(x)$ of the homogenized equation (2.2) (cf. [3,52]). The angular brackets denote the conditional average with respect to $\mu_{x_0}^{(0)} = \mu$, i.e. $\langle A(x,y) \rangle = \int A(x,y)\mu(\mathrm{d}y)$. The second cumulant and moment is given up to order $\mathcal{O}(\varepsilon^{\frac{3}{2}})$ by

$$c^{(2)} = m^{(2)} = c_0^{(2)} + t\, c_{0,1}^{(2)} + \frac{\varepsilon^2}{t} c_{2,-1}^{(2)} + \varepsilon\, c_{1,0}^{(2)} + R_\varepsilon^{(2)}. \tag{4.10}$$

The $\mathcal{O}(1)$ contribution is given by the homogenized Green-Kubo formula (2.4)

$$c_0^{(2)} = \sigma^2 = -2\langle f_0 \mathcal{L}_{0\perp}^{-1} f_0 \rangle$$

and higher-order contributions are given by

$$c^{(2)}_{0,1} = \frac{1}{2}\,\sigma^2\left(\frac{\partial\sigma}{\partial x}\right)^2 + \frac{1}{2}\,\sigma^3\frac{\partial^2\sigma}{\partial x^2} + \sigma^2\frac{\partial F}{\partial x} + F\sigma\frac{\partial\sigma}{\partial x} + F^2 \tag{4.11}$$

$$c^{(2)}_{2,-1} = -2\langle f_0\mathcal{L}_{0\perp}^{-2}f_0\rangle \tag{4.12}$$

$$\begin{aligned} c^{(2)}_{1,0} = &-2\langle f_0\mathcal{L}_{0\perp}^{-1}f_1\rangle - 2\langle f_1\mathcal{L}_{0\perp}^{-1}f_0\rangle + 2\langle f_0\mathcal{L}_{0\perp}^{-1}\partial_x f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle \\ &+ 4\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}\partial_x f_0\rangle + 2\langle f_0\mathcal{L}_{0\perp}^{-1}(g_1\partial_y)\mathcal{L}_{0\perp}^{-1}f_0\rangle \\ &+ 2\langle (g_1\partial_y)\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle. \end{aligned} \tag{4.13}$$

Here $\mathcal{L}_{0\perp}^{-1}$ denotes the invertible operator whose inverse is the restriction of $\mathcal{L}_0$ to the space orthogonal to the projection onto the invariant measure $\mu^{(0)}_{x_0}$ (see Section (i) in the Appendix for more details; cf. (A 9)). Recall that $\mathcal{L}_0$, associated with the ergodic fast dynamics, has a nontrivial kernel, namely functions which are constant in y, and as such is noninvertible.

The third moment and its cumulant are given up to order $\mathcal{O}(\varepsilon^{\frac{3}{2}})$ by

$$c^{(3)} = m^{(3)} = \sqrt{t}\,c^{(3)}_{0,\frac{1}{2}} + \frac{\varepsilon}{\sqrt{t}}\,c^{(3)}_{1,-\frac{1}{2}} + R^{(3)}_\varepsilon \tag{4.14}$$

with

$$c^{(3)}_{0,\frac{1}{2}} = 6\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle\frac{\partial}{\partial x}\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle \tag{4.15}$$

$$c^{(3)}_{1,-\frac{1}{2}} = 6\left\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\right\rangle \tag{4.16}$$

and the fourth cumulant is given up to order $\mathcal{O}(\varepsilon^{\frac{3}{2}})$ by

$$c^{(4)} = t\,c^{(4)}_{0,1} + \varepsilon\,c^{(4)}_{1,0} + \frac{\varepsilon^2}{t}\,c^{(4)}_{2,-1} + R^{(4)}_\varepsilon, \tag{4.17}$$

with

$$c^{(4)}_{0,1} = -24\,\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle\left(\frac{\partial}{\partial x}\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle\right)^2 - 16\,\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle^2\frac{\partial^2}{\partial x^2}\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle \tag{4.18}$$

$$c^{(4)}_{1,0} = -24\,\langle\frac{\partial}{\partial x}f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle - 36\,\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle\frac{\partial}{\partial x}\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle \tag{4.19}$$

$$c^{(4)}_{2,-1} = 24\left(\langle f_0\mathcal{L}_{0\perp}^{-2}f_0\rangle\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle - \langle f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle\right). \tag{4.20}$$

Higher cumulants give rise to terms of at least $\mathcal{O}(\varepsilon^{\frac{3}{2}})$. As expected from the CLT, the only $\mathcal{O}(1)$ contribution to the cumulants (4.8)-(4.17) appears in the second cumulant (4.10). The higher-order terms determine the Edgeworth corrections. At $\mathcal{O}(\sqrt{\varepsilon})$ – describing the lowest order correction to the CLT – the first and third cumulant (4.14) feature, and, if non-zero, give rise to skewness. Corrections to the variance and the fourth-order cumulant $c^{(4)}$ start to contribute at order $\mathcal{O}(\varepsilon)$. Note that the feedback of the slow dynamics on the approach to Gaussianity via vector field f_1 only appears in the corrections to the first moment.

In this formulation of the cumulant expansion we have not yet substituted $t = \varepsilon$. This allows us to separate the contributions which pertain in the limit $\varepsilon \to 0$ and encode the Edgeworth corrections to the transition probabilities of the homogenized slow diffusion equation (2.2), namely those with $\alpha_\varepsilon = 0$ (see also [40]).

Corollary 1. *For the transition probabilities $\pi_h(\mathrm{x},t,x_0)$ of the homogenized system, we have the following expansion for small t*

$$\pi_h(\mathrm{x},t,x_0)=\mathbf{n}_{0,\sigma^2}(\mathrm{x})\left(1+\sqrt{t}\left(\frac{c^{(1)}_{0,\frac{1}{2}}}{\sigma}H_1\left(\frac{\mathrm{x}}{\sigma}\right)+\frac{c^{(3)}_{0,\frac{1}{2}}}{3!\sigma^3}H_3\left(\frac{\mathrm{x}}{\sigma}\right)\right)\right. \tag{4.21}$$
$$\left.+t\left(\frac{c^{(2)}_{0,1}+{c^{(1)}_{0,\frac{1}{2}}}^2}{2\sigma^2}H_2\left(\frac{\mathrm{x}}{\sigma}\right)+\frac{c^{(4)}_{0,1}+4c^{(1)}_{0,\frac{1}{2}}c^{(3)}_{0,\frac{1}{2}}}{4!\sigma^4}H_4\left(\frac{\mathrm{x}}{\sigma}\right)+\frac{{c^{(3)}_{0,\frac{1}{2}}}^2}{2(3!\sigma^3)^2}H_6\left(\frac{\mathrm{x}}{\sigma}\right)\right)\right)$$
$$+\mathcal{O}(t^{\frac{3}{2}}),$$

where $H_n(\mathrm{x})=(\mathrm{x}-\frac{\mathrm{d}}{\mathrm{dx}})^n 1$ are Hermite polynomials of degree n.

The remaining contributions, (4.12),(4.13),(4.16),(4.19) and (4.20) involve intricate correlations between the slow and the fast dynamics which do not vanish on the intermediate time scale for $\varepsilon\to 0$ and $t\neq 0$. In particular the homogenized limit $\varepsilon\to 0$ does not involve the slow vector field f_1.

In homogenization the knowledge of the drift F and the Green-Kubo diffusion σ^2 is sufficient to determine the effective reduced stochastic slow dynamics (cf. (2.2)). Going beyond the CLT requires higher-order Edgeworth corrections involving indefinite integrals over multi-point correlation functions. It is instructive to note that the vector field $f_1(x,y)$ nontrivially enters the corrections to the variance (4.10). The detailed proof of our main result (1.3) can be found in Appendix A.

5. Numerical results

We now corroborate our main result on the explicit formula (1.3) for the transition probability $\pi_\varepsilon(\mathrm{x},t=\varepsilon,x_0)$ with numerical simulations. We show that the Edgeworth approximation of the intermediate-time transition probability provides a much better approximation to the true transition probability of the multi-scale system with finite time scale separation than the Gaussian limiting distribution implied by the assumption of infinite-time-scale separation and homogenization theory.

We consider here as an example of the general multi-scale system (1.1)-(1.2), the following skew-product slow-fast system for a one-dimensional slow variable

$$\dot{x}=\frac{1}{\varepsilon}f_0(y)-V'(x) \tag{5.1}$$

driven by a fast Lorenz system

$$\dot{y_1}=\frac{s_l}{\varepsilon^2}(y_2-y_1) \tag{5.2}$$
$$\dot{y_2}=\frac{1}{\varepsilon^2}(y_1(r_l-y_3)-y_2) \tag{5.3}$$
$$\dot{y_3}=\frac{1}{\varepsilon^2}(y_1y_2-b_ly_3), \tag{5.4}$$

where we choose the classical parameters $s_l=10$, $r_l=28$ and $b_l=8/3$ [53]. The Lorenz system is rapidly mixing with exponentially decaying correlation [54]. The fast variables drive the slow variable x via a function $f_0(y)=\cos(y_2/2)\exp(y_1/40)-\bar{f}_0$, generating skewed deterministic noise. Here $\bar{f}_0$ is a constant chosen such that the expectation value of f_0 under the fast dynamics is zero and the system (5.1)–(5.4) satisfies the centering condition. We consider here a harmonic potential $V(x)=\alpha x^2$ (i.e. $f_1=V'(x)$). In the absence of coupling to the fast Lorenz system the slow dynamics settles to the stable fixed point $x=0$. With nontrivial coupling f_0, the fast dynamics induces fluctuations around the stable fixed point. An example of a slow trajectory

is given in Figure 1. On large enough time scales we observe seemingly stochastic behaviour (left-hand figure), similar to that of the limiting homogenized system. On shorter time scales, the smoothness of the noise is however still discernible (right-hand figure).

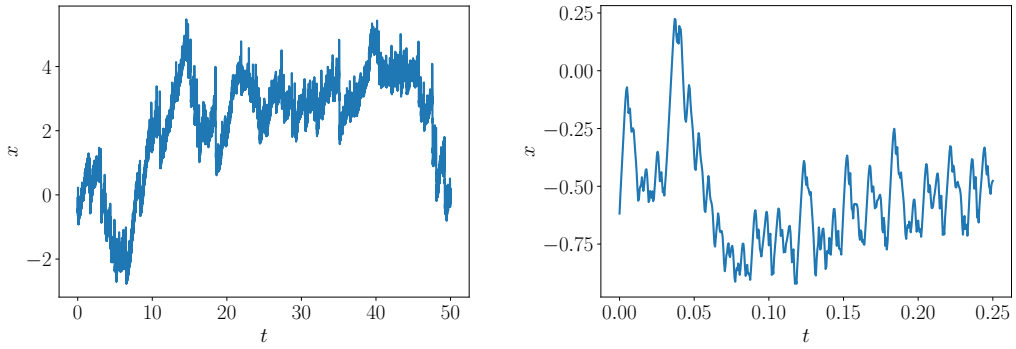

Figure 1. Trajectory of the slow variable x of the slow-fast system (5.1)-(5.4) with $\alpha = 0.1$ and $\varepsilon = 0.1$.

To numerically estimate the transition probability $\pi_\varepsilon(\mathrm{x}, t = \varepsilon, x_0)$ we perform ensemble averages of the slow variable $x(t = \varepsilon)$ using $M = 10^5$ long time simulations of 6×10^7 time units using a Dormand-Prince order $4/5$ Runge-Kutta method [55,56]. The simulations have the same initial condition $x(0) = 1$ of the slow variable but differ in the initial conditions for the fast Lorenz system which are chosen randomly from the Lorenz attractor (this is assured by letting random initial conditions settle on the attractor after a transient period of 10 time units).

The limiting Gaussian transition probability π_h can be analytically determined from the homogenized limiting equation $\mathrm{d}X = -\alpha X \mathrm{d}t + \sigma \mathrm{d}W$ which is an Ornstein-Uhlenbeck process where the diffusion coefficient σ^2 is given by the Green-Kubo formula. The limiting Gaussian transition probability has mean $x_0 \exp(-\alpha t)$ and variance by $(1 - \exp(-2\alpha t))\sigma^2/(2\alpha)$. The Edgeworth corrections to this limiting distribution are given by (1.3) and the first four cumulants (4.8), (4.10), (4.14) and (4.17). The terms $c_{2,1}^{(2)}$ (4.12), $c_{1,\frac{1}{2}}^{(3)}$ (4.16) and $c_{2,-1}^{(4)}$ (4.20) are estimated numerically by setting the gradient force $V'(x) \equiv 0$ in the long time integrations. Setting $f_1 = V'(x) \equiv 0$ implies that the Edgeworth expansion only involves (4.12), (4.16) and (4.20) which in turn can be accurately estimated without being numerically dominated by the other terms involving f_1. For details on the calculations of these coefficients see [57].

Figure 2 shows a comparison of the transition probabilities of the full deterministic multi-scale system with $\varepsilon = 0.1$, the limiting Gaussian implied by the homogenized limit and the transition probability given by the Edgeworth expansion. It is clearly seen that the the homogenized limit system is not able to capture the inherent skewness of the slow dynamics, whereas the Edgeworth approximation capture the transition probability remarkably well. We also see that the Edgeworth approximation is only a good approximation of the transition probability for the slow variable near the mode and exhibits deviations for x-values far from the mean with unphysical negative values.

6. Summary and outlook

In this article we derived Edgeworth expansions that describe corrections to the Gaussian limiting behaviour of slow-fast systems for finite time scale separation. The Edgeworth expansion is achieved using a semi-group formalism for the transfer operator, where a Duhamel-Dyson series is used to asymptotically determine the corrections at any desired order of the time scale parameter ε. The corrections appear on the intermediate time scale $\mathcal{O}(\varepsilon)$ and the asymptotics requires the limit $t = \varepsilon \ll 1$ and $t/\varepsilon^2 \to \infty$. We developed a diagrammatic representation of

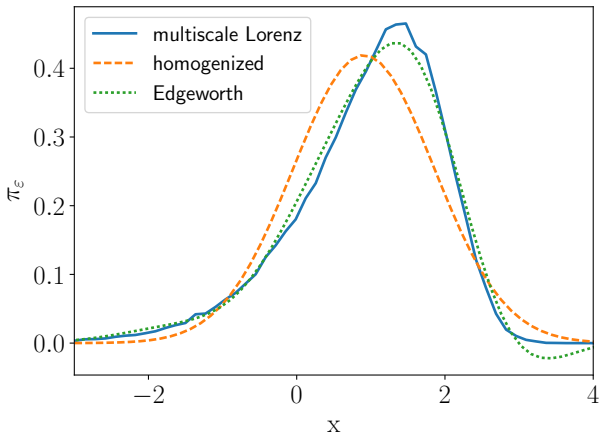

Figure 2. Transition probability $\pi_\varepsilon(\mathrm{x}, t, x_0)$ of the slow variable for the full multi-scale Lorenz system (5.1)-(5.4) with $\alpha = 0.1$ and $\varepsilon = 0.1$ (blue solid), transition probability $\pi_h(\mathrm{x}, t, x_0)$ of the corresponding homogenized system (orange dashed) and $\pi_\varepsilon(\mathrm{x}, t, x_0)$ of the Edgeworth expansion (1.3) (green dotted). The transition probabilities are evaluated at $t = 1$ and were initiated at $x_0 = 1$.

higher-order correlation integrals to control the combinatorial wealth of their asymptotic expansion in ε. To obtain our explicit formula for the transition probability (1.3) we required mixing assumptions on the integrability of higher-order autocorrelation functions. It is pertinent to mention that homogenization theory does not rely on mixing and the diffusive limit equations exist for non-mixing systems (albeit not with a Green-Kubo formula (2.4) for the diffusion), and similarly there are mixing systems which do not allow for a diffusive limit. We expect that similarly one can derive an Edgeworth expansion without these strong mixing assumptions, at the cost of not having compact explicit expressions for the drift and diffusion.

Our work points to several applications and directions, planned for further research. We derive here the Edgeworth expansion for continuous-time multi-scale systems in the case where at leading order the slow dynamics does not couple back into the fast dynamics, i.e. $g_0 = g_0(y)$. We also expect a similar expansion to hold when $g_0 = g_0(x, y)$ and the slow dynamics couples back into the fast dynamics at leading order. A complicating issue here is the potential breakdown of linear response when the fast invariant measure does not depend smoothly on the slow variables.

Edgeworth corrections to the CLT are not restricted to systems where the noise originates as the accumulative effect of rapidly decorrelating fast variables, as we have described here for slow-fast systems. Sums of uncorrelated deterministic variables also appear in weak coupling limits where a distinguished degree of freedom is weakly coupled to a bath of N degrees of freedom. Here stochastic limit systems arise in the limit of an infinitely large bath with $N \to \infty$ [3]. Again, corrections for finite values of N can be studied using an Edgeworth expansion.

The universal form of the deviations from Gaussianity given by the Edgeworth expansion suggests that one can devise stochastic parametrizations and effective diffusive dynamics for the slow variables for finite time scale separation by substituting the fast dynamics with a surrogate system with the same Edgeworth coefficients, improving on the classical homogenization limit SDEs. In particular, the universal character allows for a data-driven approach where the cumulants are numerically estimated to build a stochastic model for the observed variables.

Data Accessibility. This article has no additional data.

Authors' Contributions. JW designed the research and performed the numerical experiments. All authors contributed to the research and the writing of the paper.

Competing Interests. There are no competing interests.

Funding. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° PIOF-GA-2013-626210. GAG is partially supported by the ARC grant DP180101385.

Acknowledgements. We thank Ben Goldys and Françoise Pène for enlightening discussions and comments.

A. Appendix: Derivation of the cumulant expansion

Outline of the derivation: The Edgeworth expansion involves an expansion of the cumulants in orders of ε. The derivation of the Edgeworth expansion proceeds in a number of steps. We first consider the Laplace transform of the moments, when expressed in terms of the Koopman operator, and perform a subsequent expansion using eigenfunctions of the fast generator $\mathcal{L}_0$ (Section i). Introducing a two-dimensional diagrammatic representation we perform a combinatorial accounting of the sequences appearing in the expansion of the Laplace transform (Section ii). In particular, this allows us to identify those sequences which vanish and those which will give non-trivial contributions. This will allow us to provide the explicit expressions of the cumulants $c^{(p)}$ with $p \leq 4$ stated in (4.8), (4.10), (4.14) and (4.17) to capture the Edgeworth corrections up to $\mathcal{O}(\varepsilon^{\frac{3}{2}})$ (Section iii). We then show that it is sufficient for the estimation of the cumulants up to $\mathcal{O}(\varepsilon^{\frac{3}{2}})$ to perform averages with respect to the fast measure $\mu_{x_0}^{(0)}$ only, discarding contributions from the linear response term $\mu_{x_0}^{(1)}$ (Lemma 1). We conclude with Lemma 2 showing that higher cumulants $c^{(p)}$ with $p \geq 5$ do not contribute at order $\mathcal{O}(\varepsilon^{\frac{3}{2}})$ to the Edgeworth expansion and the transition probability $\pi_\varepsilon(\mathrm{x}, t = \varepsilon, x_0)$.

Derivation of the cumulant expansion: We express the conditional average in (4.3) in terms of the Koopman operator $e^{\mathcal{L}t}$ associated with the multi-scale system (1.1)-(1.2) with the generator

$$\mathcal{L} = \frac{1}{\varepsilon^2}\mathcal{L}_0 + \frac{1}{\varepsilon}\mathcal{L}_1 + \mathcal{L}_2, \tag{A 1}$$

where

$$\mathcal{L}_0 = g_0(y)\cdot\partial_y + \frac{1}{2}\beta(y)\beta^T(y) : \partial_y\cdot\partial_y, \quad \mathcal{L}_1 = f_0(x,y)\cdot\partial_x + g_1(x,y)\cdot\partial_y, \quad \mathcal{L}_2 = f_1(x,y)\cdot\partial_x. \tag{A 2}$$

We seek expansions in t and ε of the moments

$$m^{(p)} = \frac{1}{t^{p/2}}\mathbb{E}^{x_0,\mu}\left[e^{t\left(\frac{\mathcal{L}_0}{\varepsilon^2}+\frac{\mathcal{L}_{12}}{\varepsilon}\right)}z^{(p)}(x)\right] = \frac{1}{t^{p/2}}\int\int e^{t\left(\frac{\mathcal{L}_0}{\varepsilon^2}+\frac{\mathcal{L}_{12}}{\varepsilon}\right)}z^{(p)}(x)\left(\delta_{x_0}\times\mu_{x_0}^{(\varepsilon)}\right)(\mathrm{d}x,\mathrm{d}y), \tag{A 3}$$

where $z^{(p)}(x) = (x - x_0)^p$ and where we introduce for convenience $\mathcal{L}_{12} = \mathcal{L}_1 + \varepsilon\mathcal{L}_2$. Using (4.5), we split the conditional average over the invariant measure $\mu^{(\varepsilon)}$ (see (4.4)) according to

$$m^{(p)} = \frac{1}{t^{p/2}}\delta_{x_0}\left((\mu_{x_0}^{(0)} + \varepsilon\mu_{x_0}^{(1)} + \mathcal{O}(\varepsilon^2))e^{t\mathcal{L}}z^{(p)}(x)\right) = \frac{1}{t^{p/2}}\delta_{x_0}\left(\mathcal{A}^{(p)} + \varepsilon\,\mathcal{B}^{(p)} + \mathcal{O}(\varepsilon^2)\right), \tag{A 4}$$

where we define

$$\mathcal{A}^{(p)}(t) = \left\langle e^{t\left(\frac{\mathcal{L}_0}{\varepsilon^2}+\frac{\mathcal{L}_{12}}{\varepsilon}\right)}z^{(p)}(x)\right\rangle \quad \text{and} \quad \mathcal{B}^{(p)}(t) = \mu_{x_0}^{(1)}\left(e^{t\left(\frac{\mathcal{L}_0}{\varepsilon^2}+\frac{\mathcal{L}_{12}}{\varepsilon}\right)}z^{(p)}(x)\right).$$

As before, angular brackets denote the average with respect to $\mu_{x_0}^{(0)}$.

We first calculate the first term in (A 4) associated with the average over $\mu_{x_0}^{(0)}$ before proving that averages with respect to $\mu_{x_0}^{(1)}$ do not contribute in Lemma 1.

(i) Laplace transform of the moments and expansion using eigenfunctions of $\mathcal{L}_0$

We assume that the spectrum of the generator $\mathcal{L}$ is contained in the left-half complex plane, $\{z \in \mathbb{C} : \mathrm{Re}(z) \leq 0\}$ (see [58] for Anosov flows). We can then take the Laplace transform of $\mathcal{A}^{(p)}(t)$

$$\mathfrak{L}\{\mathcal{A}^{(p)}\}(s) = \left\langle \left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2} - \frac{\mathcal{L}_{12}}{\varepsilon} \right)^{-1} z^{(p)}(x) \right\rangle \tag{A 5}$$

for $\mathrm{Re}(s) > 0$.

Using the operator identity

$$(C-D)^{-1} = C^{-1} + C^{-1}D(C-D)^{-1} = C^{-1} + C^{-1}DC^{-1} + C^{-1}DC^{-1}DC^{-1} + \dots \tag{A 6}$$

with $C = s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2}$ and $D = \frac{\mathcal{L}_{12}}{\varepsilon}$ we have

$$\begin{aligned}
\mathfrak{L}\{\mathcal{A}^{(p)}\}(s) = \Bigg\langle \Bigg(& \left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2} \right)^{-1} + \left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2} \right)^{-1} \frac{\mathcal{L}_{12}}{\varepsilon} \left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2} \right)^{-1} \\
& + \left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2} \right)^{-1} \frac{\mathcal{L}_{12}}{\varepsilon} \left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2} \right)^{-1} \frac{\mathcal{L}_{12}}{\varepsilon} \left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2} \right)^{-1} + \dots \Bigg) z^{(p)} \Bigg\rangle \\
= & \frac{z^{(p)}}{s} + \frac{1}{s^2} \Bigg\langle \Bigg(\frac{\mathcal{L}_{12}}{\varepsilon} + \frac{\mathcal{L}_{12}}{\varepsilon} \left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2} \right)^{-1} \frac{\mathcal{L}_{12}}{\varepsilon} \\
& + \frac{\mathcal{L}_{12}}{\varepsilon} \left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2} \right)^{-1} \frac{\mathcal{L}_{12}}{\varepsilon} \left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2} \right)^{-1} \frac{\mathcal{L}_{12}}{\varepsilon} + \dots \Bigg) z^{(p)} \Bigg\rangle ,
\end{aligned} \tag{A 7}$$

where we used in the last equality that $\mathcal{L}_0 A(x) = \mathcal{L}_0^* \mu_{x_0}^{(0)} = 0$.

We assume here that the resolvent is compact. This is the case on L_p-spaces if the generator includes diffusion. For the purely advective case of general deterministic multi-scale systems under consideration here, there is currently no theory available on what function spaces the resolvent is compact[3]. Under this assumption let us decompose $s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2}$ into projectors p_i on the eigenfunctions of $\mathcal{L}_0$, where $p_i \bullet = (r_i, \bullet) l_i$ and $p_0 \bullet = (\mu^{(0)}, \bullet) 1$, such that $\mathcal{L}_0 = \sum_{i=0}^{\infty} \lambda_i p_i$. The pairing (a, b) is defined as $(a, b) = \int a(dy) b$. The eigenfunctions l_i and r_i satisfy $(r_i, l_j) = \delta_{i,j}$ and correspond to the eigenvalue λ_i with $\lambda_0 = 0$ [60, Chapter 7]. Then

$$s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2} = \sum_{i=0}^{\infty} \left(s - \frac{\lambda_i}{\varepsilon^2} \right) p_i = s p_0 + \left(s - \frac{\lambda_1}{\varepsilon^2} \right) p_1 + \left(s - \frac{\lambda_2}{\varepsilon^2} \right) p_2 + \dots$$

and therefore

$$\begin{aligned}
\left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2} \right)^{-1} &= \frac{1}{s} p_0 + \left(s - \frac{\lambda_1}{\varepsilon^2} \right)^{-1} p_1 + \left(s - \frac{\lambda_2}{\varepsilon^2} \right)^{-1} p_2 + \dots = \frac{1}{s} p_0 + \left(s\mathbb{1}_\perp - \frac{\mathcal{L}_{0\perp}}{\varepsilon^2} \right)^{-1} \\
&= \frac{1}{s} p_0 - \varepsilon^2 (\mathcal{L}_{0\perp})^{-1} - s\varepsilon^4 (\mathcal{L}_{0\perp})^{-2} - s^2 \varepsilon^6 (\mathcal{L}_{0\perp})^{-3} - \dots .
\end{aligned} \tag{A 8}$$

where we define the operators restricted to the space orthogonal to the projection onto the invariant measure $\mu_{x_0}^{(0)}$ as $\mathcal{L}_{0\perp} = \lambda_1 p_1 + \lambda_2 p_2 + \dots$ and $\mathbb{1}_\perp = p_1 + p_2 + \dots$. We note that formally

$$\mathcal{L}_{0,\perp}^{-1} = -\int_0^\infty \mathrm{d}\tau \left(e^{\tau \mathcal{L}_0} - p_0 \right) \tag{A 9}$$

and therefore formally $\langle A \mathcal{L}_{0\perp}^{-1} f_0 \rangle = \langle A \mathcal{L}_0^{-1} f_0 \rangle$ and $\langle f_0 \mathcal{L}_{0\perp}^{-1} A \rangle = \langle f_0 \mathcal{L}_0^{-1} A \rangle$ for any A since $p_0 f_0 = \langle f_0 \rangle = 0$.

The expansion (A 7) contains terms with both positive and negative powers of s. According to the residue theorem, only the residues contribute to the inverse Laplace transform. By the way it is defined in Eq. (A 7), the function $\mathfrak{L}\{\mathcal{A}^{(p)}\}(s)$ can only have poles at zero or at $\frac{\lambda_i}{\varepsilon^2}$. The poles at

[3] For hyperbolic maps one has good statistical properties when considering anisotropic Banach spaces (see [59] and references therein).

$\frac{\lambda_i}{\varepsilon^2}$ decay exponentially upon inverse Laplace transform (recall that $t/\varepsilon^2 \to \infty$), and we therefore only need to consider the poles at zero.

(ii) Combinatorial analysis of the expansion at each order in ε and t

The expansion of the Laplace transform (A 5) contains powers of ε and of s. To facilitate the combinatorial problem of accounting for the non-zero contributions at a specified order of ε and s (or t, respectively) we introduce the following diagrammatic representation. We classify all terms in the expansion (A 7) as sequences $S(a,b)$ with $a_i \in \mathbb{N}_0$, $b_j \in \{1,2\}$ of the form

$$S(a,b) := \frac{1}{s^2}\langle B_{b_1} A_{a_1} B_{b_2} A_{a_2} \dots B_{b_l} A_{a_l} B_{b_{l+1}} z^{(p)} \rangle, \tag{A 10}$$

where $A_0 = \frac{1}{s} p_0$, $A_1 = -\varepsilon^2 (\mathcal{L}_{0\perp})^{-1}$, $A_2 = -s\varepsilon^4 (\mathcal{L}_{0\perp})^{-2}$, etc. represent the terms in the series (A 8), and $B_1 = \mathcal{L}_1/\varepsilon$ and $B_2 = \mathcal{L}_2$. We introduce the shorthand notation for a sequence of length l, $\begin{pmatrix} & a_1 & \dots & a_l & \\ b_1 & & \dots & & b_{l+1} \end{pmatrix}$ for $S(a,b)$. We also introduce a product $\otimes$ of two subsequences, with

$$\begin{pmatrix} & a_1 & \dots & a_l & \\ b_1 & & \dots & & b_{l+1} \end{pmatrix} \otimes \begin{pmatrix} & c_1 & \dots & c_l & \\ d_1 & & \dots & & d_{l+1} \end{pmatrix}$$
$$= \begin{pmatrix} & a_1 & \dots & a_l & & 0 & & c_1 & \dots & c_l & \\ b_1 & & \dots & & b_{l+1} & & d_1 & & \dots & & d_{l+1} \end{pmatrix}.$$

For example

$$\frac{1}{s^2}\left\langle \mathcal{L}_2 \frac{p_0}{s} \mathcal{L}_1 (-\varepsilon^2 (\mathcal{L}_{0\perp})^{-1}) \mathcal{L}_1 z^{(p)} \right\rangle = S((0,1),(2,1,1))$$
$$= \begin{pmatrix} & 0 & & 1 & \\ 2 & & 1 & & 1 \end{pmatrix} = \begin{pmatrix} & \\ 2 & \end{pmatrix} \otimes \begin{pmatrix} & 1 & \\ 1 & & 1 \end{pmatrix}.$$

The order of ε of a given sequence is given by

$$O_\varepsilon = 2\sum_{i=1}^{l} a_i + \sum_{j=1}^{l+1} (b_j - 2) = \left(2\sum_{i=1}^{l} a_i + \sum_{j=1}^{l+1} b_j \right) - 2(l+1),$$

and the order of s is given by

$$O_s = -2 + \sum_{i=1}^{l} (a_i - 1),$$

or, equivalently, the order of t in the inverse Laplace transform is given by

$$O_t = 1 - \sum_{i=1}^{l} (a_i - 1) = 1 + l - \sum_{i=1}^{l} a_i.$$

The combined order in t and ε of the sequence is

$$O_{\varepsilon,t} = \sum_{i=1}^{l} a_i + \sum_{j=1}^{l+1} b_j - l - 1. \tag{A 11}$$

Note that the combined order of a sequence does not change upon commutation of the subsequences in a product. Finally, the sign of the sequence is given by $(-1)^\theta$ where $\theta = \sum_{i=1}^{l} \delta_{0,a_i}$ is the number of non-zero elements among the a_i where we used the Kronecker δ.

We can readily identify certain sequences which do not contribute. The centering condition $\langle \mathcal{L}_1 \rangle = 0$ implies that sequences containing a subsequence $\begin{pmatrix} \dots & 0 & & 0 & \dots \\ & & 1 & & \end{pmatrix}$ vanish. For

the same reason, sequences starting with $\begin{pmatrix} & 0 & \cdots \\ 1 & & \end{pmatrix}$ or ending in $\begin{pmatrix} \cdots & 0 & \\ & & 1 \end{pmatrix}$ vanish. Furthermore, sequences containing an insufficient number of derivatives with respect to x do not contribute; if the number of derivatives is less than the order p, the sequence $S(a,b)$ averages to zero since $\delta_{x_0}\left(\partial_x^n (x-x_0)^p\right) = 0$ for $n < p$. Hence only those sequences with $l+1 \geqslant p$ contribute since the operators $\mathcal{L}_1$ and $\mathcal{L}_2$ contain at most one x-derivative. This implies that for fixed moment order p we have

$$l \geq l_{\min} = p - 1 \tag{A 12}$$

Recall that moments are rescaled by $t^{-\frac{p}{2}}$ (cf. (4.3)). We therefore require

$$O_{\varepsilon,t} - \lfloor \frac{p}{2} \rfloor \leq 1 \tag{A 13}$$

to retain only terms that contribute to first order after normalization. This implies

$$\sum_{i=1}^{l} a_i + \sum_{j=1}^{l+1} b_j - l - 1 \leqslant q + 1. \tag{A 14}$$

We now identify all those sequences which are compatible with the constraints (A 14) and (A 12). For sake of exposition, we treat even and odd moments separately. We first determine the possible sequences of different lengths l for even moments $m^{(2q)}$ where $l \geqslant 2q-1$.

Case $l = 2q-1$ Equation (A 14) implies $\sum_{i=1}^{l} a_i + \sum_{j=1}^{l+1} b_j \leqslant 3q+1$. The sequences of length $2q-1$ have in total $2l+1 = 4q-1$ elements. Each sequence must therefore contain at least $q-2$ zero elements. Since sequences with subsequence $\begin{pmatrix} \cdots & 0 & & 0 & \cdots \\ & & 1 & & \end{pmatrix}$ average to zero, there are 6 possible types of sequence that fulfill the inequality, namely

$$\begin{pmatrix} & 1 & \\ 1 & & 1 \end{pmatrix}^{\otimes q} \qquad \begin{array}{l}\text{permutations: } 1 \\ O_{\varepsilon,t} = q\end{array} \tag{A 15}$$

$$\begin{pmatrix} \\ 2 \end{pmatrix}^{\otimes 2} \otimes \begin{pmatrix} & 1 & \\ 1 & & 1 \end{pmatrix}^{\otimes (q-1)} \qquad \begin{array}{l}\text{permutations: } \binom{q+1}{2} \\ O_{\varepsilon,t} = q+1\end{array} \tag{A 16}$$

$$\begin{pmatrix} 1 & & 1 & & 1 & \\ & 1 & & 1 & & 1 \end{pmatrix} \otimes \begin{pmatrix} & 1 & \\ 1 & & 1 \end{pmatrix}^{\otimes (q-2)} \qquad \begin{array}{l}\text{permutations: } \binom{q-1}{1} \\ O_{\varepsilon,t} = q+1\end{array} \tag{A 17}$$

$$\begin{pmatrix} & 1 & \\ 2 & & 1 \end{pmatrix} \otimes \begin{pmatrix} & 1 & \\ 1 & & 1 \end{pmatrix}^{\otimes (q-1)} \qquad \begin{array}{l}\text{permutations: } \binom{q}{1} \\ O_{\varepsilon,t} = q+1\end{array} \tag{A 18}$$

$$\begin{pmatrix} & 2 & \\ 1 & & 1 \end{pmatrix} \otimes \begin{pmatrix} & 1 & \\ 1 & & 1 \end{pmatrix}^{\otimes (q-1)} \qquad \begin{array}{l}\text{permutations: } \binom{q}{1} \\ O_{\varepsilon,t} = q+1\end{array} \tag{A 19}$$

$$\begin{pmatrix} & 1 & \\ 1 & & 2 \end{pmatrix} \otimes \begin{pmatrix} & 1 & \\ 1 & & 1 \end{pmatrix}^{\otimes (q-1)} \qquad \begin{array}{l}\text{permutations: } \binom{q}{1} \\ O_{\varepsilon,t} = q+1\end{array} \tag{A 20}$$

with all possible permutations of subsequences. For $q = 2$, we can verify by simple enumeration that these are all the sequences possible if we restrict to $b_j < 3$. All other possible nonzero terms with $b_j > 3$ can be obtained by substituting a b_j in one of these sequences, but this operation increases the order. For $q > 2$ the only way to extend a series without increasing the order is to append an element $\begin{pmatrix} & 1 & \\ 1 & & 1 \end{pmatrix}$.

Case $l = 2q$ Equation (A 14) implies $\sum_{i=1}^{l} a_i + \sum_{j=1}^{l+1} b_j \leqslant 3q + 2$. The sequences of length $2q$ have in total $2l + 1 = 4q + 1$ elements. Each sequence must therefore contain at least $q - 1$ zero elements. There is only 1 possible type of sequence that fulfills the inequality, namely

$$\begin{pmatrix} & 1 & & 1 & \\ 1 & & 1 & & 1 \end{pmatrix} \otimes \begin{pmatrix} & 1 & \\ 1 & & 1 \end{pmatrix}^{\otimes(q-1)} \qquad \begin{array}{l} \text{permutations: } \binom{q}{1} \\ O_{\varepsilon,t} = q + 1 \end{array} \tag{A 21}$$

with all possible permutations of subsequences.

Case $l = 2q + 1$ Equation (A 14) implies $\sum_{i=1}^{l} a_i + \sum_{j=1}^{l+1} b_j \leqslant 3q + 3$. The sequences of length $2q + 1$ have in total $2l + 1 = 4q + 3$ elements. Each sequence must therefore contain at least q zero elements. There is only 1 possible type of sequence that fulfills the inequality, namely

$$\begin{pmatrix} & 1 & \\ 1 & & 1 \end{pmatrix}^{\otimes(q+1)} \qquad \begin{array}{l} \text{permutations: } 1 \\ O_{\varepsilon,t} = q + 1 \end{array} \tag{A 22}$$

with all possible permutations of subsequences.

Case $l \geqslant 2q + 2$ Nonzero sequences of length $2q + 2$ are of at least order $q + 2$, so they do not contribute.

For odd moments $m^{(2q+1)}$, the inequality (A 14) together with the condition on the number of derivatives $l \geqslant 2q$ implies that only sequences (A 21)-(A 22) contribute.

(iii) Calculation of the moments $m^{(p)}$

We have now derived all nontrivial sequences contributing to the moments $m^{(p)}$. We can obtain explicit expressions for the moments by calculating for each sequence the expectation value as prescribed in Eq. (A 10). In the process, subsequences become functions of x due to the projection p_0. Finally, we work out explicitly the x-derivatives in the sequence. For example, for the sequence $S((0,1),(2,1,1))$ with $p = 3$, we have that, as described above,

$$\begin{aligned} \mathfrak{L}^{-1} S((0,1),(2,1,1)) &= -\varepsilon^2 t^2 \left\langle \mathcal{L}_2 p_0 \mathcal{L}_1 \mathcal{L}_{0\perp}^{-1} \mathcal{L}_1 (x - x_0)^3 \right\rangle \\ &= -3\,\varepsilon^2 t^2 \left\langle f_1(x,y) \right\rangle \left\langle f_0(x,y) \mathcal{L}_{0\perp}^{-1} f_0(x,y) \right\rangle . \end{aligned}$$

Recalling the multi-index notation to denote the expansions in ε and t with $\boldsymbol{\alpha} = (\alpha_\varepsilon, \alpha_t)$ and with $|\boldsymbol{\alpha}| = \alpha_\varepsilon + \alpha_t$ being the combined order of the contribution, we obtain, in this way, from sequence (A 15) the leading homogenized term ($|\boldsymbol{\alpha}| = 0$) in the $2q$-th scaled moment $m^{(2q)}$

$$m_0^{(2q)} = \frac{1}{t^q} \frac{t^q}{q!} (2q)! \left(\frac{\sigma^2}{2} \right)^q , \tag{A 23}$$

where $\sigma^2 = -2\langle f_0 \mathcal{L}_{0\perp}^{-1} f_0 \rangle$ is the homogenized diffusion coefficient (2.4). The first factor $1/t^q$ is the normalization factor of the moment, the term $t^q/q!$ comes from the inverse Laplace transform of $1/s^{q+1}$ and $(2q)!$ comes from the $2q$-th derivative of $(x - x_0)^{2q}$. Finally, since only terms with $2q$ derivatives ∂_x yield nonzero contributions, only the term with f_0 in $\mathcal{L}_1$ contributes, resulting in $\left\langle f_0 e^{t\mathcal{L}_0} f_0 \right\rangle^q = \left(\sigma^2/2 \right)^q$.

Similar combinatorial accounting distills from the sequences (A 16) and (A 22) the homogenized first correction term ($\alpha_\varepsilon = 0$, $|\boldsymbol{\alpha}| = 1$) in the $2q$-th scaled moment $m^{(2q)}$

$$t\, m_{0,1}^{(2q)} = t\,\frac{(2q)!}{(p+1)!}\left(\binom{q+1}{2}F^2\left(\frac{\sigma^2}{2}\right)^{q-1} + F\left(\frac{\sigma^2}{2}\right)^{(q-1)}\sigma\partial_x\sigma\left(q^3+\frac{q^2}{2}-\frac{q}{2}\right)\right.$$
$$+ q(q+1)\left(\frac{\sigma^2}{2}\right)^{q}\partial_x F + \frac{\sigma^{(2q+1)}}{2^{(q+1)}}\partial_x^2\sigma\left(\frac{4}{3}q^3+q^2-\frac{q}{3}\right)$$
$$\left.+\frac{\sigma^{2q}}{2^{q+1}}(\partial_x\sigma)^2\frac{q}{3}(q+1)(6q^2-2q-1)\right),$$

where $F = \langle f_1\rangle - \langle (g_1\partial_y)\mathcal{L}_{0\perp}^{-1}f_0\rangle - \langle f_0\mathcal{L}_{0\perp}^{-1}\partial_x f_0\rangle$ is the homogenized drift coefficient (2.3).

From sequences (A 17)-(A 20) we obtain the non-homogenized corrections ($\alpha_\varepsilon > 0$, $|\boldsymbol{\alpha}| = 1$) to the $2q$-th scaled moment with $2q$ x-derivatives

$$\varepsilon\,\tilde{m}_{1,0}^{(2q)} + \frac{\varepsilon^2}{t}\,m_{2,-1}^{(2q)} = \frac{(2q)!}{t^q}\left(\frac{\varepsilon t^q}{q!}q(q-1)\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle\langle f_1\rangle\left(-\left\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\right\rangle\right)^{(q-2)}\right.$$
$$+\frac{\varepsilon t^q}{q!}q\left(-\langle f_1\mathcal{L}_{0\perp}^{-1}f_0\rangle\right)(-\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle)^{q-1}$$
$$+\frac{\varepsilon t^q}{q!}q\left(-\left\langle f_0\mathcal{L}_{0\perp}^{-1}f_1\right\rangle\right)(-\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle)^{(q-1)}$$
$$+\frac{\varepsilon^2 t^{q-1}}{(q-1)!}q\left(-\left\langle f_0\mathcal{L}_{0\perp}^{-2}f_0\right\rangle\right)(-\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle)^{q-1}$$
$$+\frac{\varepsilon^2 t^{q-1}}{(q-1)!}(q-1)\left(-\left\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\right\rangle\right)(-\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle)^{q-2}$$
$$\left.+\frac{\varepsilon^2 t^{q-1}}{(q-1)!}\binom{q-1}{2}\left\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\right\rangle^2(-\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle)^{q-3}\right).$$

From sequence (A 21) we obtain the non-homogenized corrections ($\alpha_\varepsilon > 0$, $|\boldsymbol{\alpha}| = 1$) to the $2q$-th moment with $2q+1$ x-derivatives

$$\varepsilon\,\tilde{\tilde{m}}_{1,0}^{(2q)} = (2q)!\frac{\varepsilon}{q!}\left(q\left(\langle(g_1\partial_y)\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle(-\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle)^{q-1}\right.\right.$$
$$+\langle f_0\mathcal{L}_{0\perp}^{-1}(g_1\partial_y)\mathcal{L}_{0\perp}^{-1}f_0\rangle(-\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle)^{q-1}$$
$$+(q-1)\left\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\right\rangle(-\langle(g_1\partial_y)\mathcal{L}_{0\perp}^{-1}f_0\rangle(-\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle)^{q-2})\Big)$$
$$+q(q-1)\left(-\left\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\right\rangle\right)^{q-1}\langle\partial_x f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle$$
$$+q^2\left(-\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle\right)^{q-1}\langle f_0\mathcal{L}_{0\perp}^{-1}\partial_x f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle$$
$$+q(q+1)(-\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle)^{q-1}\left\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}\partial_x f_0\right\rangle$$
$$+q\left(q^2-\frac{3}{2}q+\frac{1}{2}\right)(-\langle\partial_x f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle)\left(-\left\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\right\rangle\right)^{q-2}\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle$$
$$\left.+q\left(q^2-\frac{1}{2}q-\frac{1}{2}\right)\left(-\left\langle f_0\mathcal{L}_{0\perp}^{-1}\partial_x f_0\right\rangle\right)(-\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle)^{q-2}\langle f_0\mathcal{L}_{0\perp}^{-1}f_0\mathcal{L}_{0\perp}^{-1}f_0\rangle\right),$$

where $m_{1,0}^{(2q)} = \tilde{m}_{1,0}^{(2q)} + \tilde{\tilde{m}}_{1,0}^{(2q)}$. From sequence (A 22) we obtain the leading homogenized terms ($\alpha_\varepsilon = 0$, $|\boldsymbol{\alpha}| = \frac{1}{2}$) in the $(2q+1)$-th moment

$$\sqrt{t}\,m_{0,\frac{1}{2}}^{(2q+1)} = \sqrt{t}\frac{(2q+1)!}{q!}\frac{1}{2^q}(F\sigma^{2q} + q\sigma^{2q+1}\partial_x\sigma).$$

From sequence (A 21) we obtain the non-homogenized correction terms ($\alpha_\varepsilon > 0$, $|\boldsymbol{\alpha}| = \frac{1}{2}$) in the $(2q+1)$-th moment

$$\frac{\varepsilon}{\sqrt{t}}\, m^{(2q+1)}_{1,-\frac{1}{2}} = \frac{(2q+1)!}{t^{(2q+1)/2}}\varepsilon \frac{t^q}{q!} q \langle f_0 \mathcal{L}_{0\perp}^{-1} f_0 \mathcal{L}_{0\perp}^{-1} f_0 \rangle (-\langle f_0 \mathcal{L}_{0\perp}^{-1} f_0 \rangle)^{q-1}.$$

Summarizing for $q=0$ and $q=1$ we obtain the desired expressions for the first four cumulants (4.8), (4.10), (4.14) and (4.17). We have until now only performed the average with respect to the leading order contribution $\mu^{(0)}_{x_0}$. The following Lemma shows that this is sufficient

Lemma 1. *The asymptotic series of cumulants $c^{(p)}$ in ε and t up to the combined order $O_{\varepsilon,t} = \frac{3}{2}$ involves averages over $\mu^{(0)}_{x_0}$ only and the linear response term $\mu^{(1)}_{x_0}$ does not contribute.*

Proof. Analogously to (A 5), the Laplace transform of

$$\mathcal{B}^{(p)}(t) = \varepsilon \mu^{(1)}_{x_0}\left(e^{t\left(\frac{\mathcal{L}_0}{\varepsilon^2} + \frac{\mathcal{L}_{12}}{\varepsilon}\right)} z^{(p)}(x)\right)$$

can be expanded as

$$\begin{aligned}\mathfrak{L}\{\mathcal{B}^{(p)}\}(s) = \varepsilon \mu^{(1)}_{x_0}\Bigg(\Bigg(&\left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2}\right)^{-1} + \left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2}\right)^{-1}\frac{\mathcal{L}_{12}}{\varepsilon}\left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2}\right)^{-1} \\ &+ \left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2}\right)^{-1}\frac{\mathcal{L}_{12}}{\varepsilon}\left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2}\right)^{-1}\frac{\mathcal{L}_{12}}{\varepsilon}\left(s\mathbb{1} - \frac{\mathcal{L}_0}{\varepsilon^2}\right)^{-1} + \dots\Bigg) z^{(p)}\Bigg).\end{aligned}$$

Using our diagrammatic representation, the sequences are encoded as

$$S(a,b) = \begin{pmatrix} a_1 & & \dots & & a_l \\ & b_1 & \dots & b_{l-1} & \end{pmatrix}.$$

The order of ε of a given sequence is now $O_\varepsilon = 2\sum_{i=1}^{l} a_i + \sum_{j=1}^{l-1}(b_j - 2) + 1 = \left(2\sum_{i=1}^{l} a_i + \sum_{j=1}^{l-1} b_j\right) - 2(l-1) + 1$ and the order of t is $O_t = -1 - \sum_{i=1}^{l}(a_i - 1) = -1 + l - \sum_{i=1}^{l} a_i$. The combined order in t and ε of the sequence is then $O_{\varepsilon,t} = \sum_{i=1}^{l} a_i + \sum_{j=1}^{l+1} b_j - l + 2$. The only sequence that can contribute in this case is

$$\begin{pmatrix} 0 & & 1 & \\ & 1 & & 1 \end{pmatrix} \otimes \begin{pmatrix} & 1 & \\ 1 & & 1 \end{pmatrix}^{\otimes q-2} \otimes \begin{pmatrix} & 1 & & 0 \\ 1 & & 1 & \end{pmatrix}$$

at order $O_{\varepsilon,t} = q+1$. Because of the left-most projection p_0 in the sequence, such terms are functions of x only. Since $\mu^{(1)}_{x_0} = \lim_{\varepsilon\to 0}(\mu^{(\varepsilon)} - \mu^{(0)}_{x_0})/\varepsilon$ is a difference of two normalized measures, applying $\mu^{(1)}_{x_0}$ to a constant in y yields 0. Therefore the cumulants up to $\mathcal{O}(\varepsilon)$ only contain averages over the fast invariant measure $\mu^{(0)}_{x_0}$ and the average with respect to the linear response $\mu^{(1)}_{x_0}$ – the second term in (A 4) – vanishes. □

Remark 1. *The linear response term $\mu^{(1)}_{x_0}$ may become relevant for higher orders $O_{\varepsilon,t} \geqslant 2$.*

Finally, we prove that higher order cumulants $c^{(p)}$ with $p \geqslant 5$ do not contribute at $\mathcal{O}(\varepsilon)$, and we can be content finding expressions for the first 4 cumulants to determine the Edgeworth expression.

Lemma 2. $c^{(p)} = \mathcal{O}(\varepsilon^{\frac{3}{2}})$ *for* $p \geqslant 5$.

Proof. Using the recursion formula for cumulants

$$c^{(p)} = m^{(p)} - \sum_{m=1}^{p-1} \binom{p-1}{m-1} m^{(p-m)} c^{(m)},$$

we have by recursion at order $\mathcal{O}(1)$ for $q \geqslant 2$ that

$$c^{(2q)} = m_0^{(2q)} - \binom{2q-1}{1} m_0^{(2q-2)} c_0^{(2)} + \mathcal{O}(\varepsilon) = c_0^{(2q)} + \mathcal{O}(\varepsilon),$$

where, due to (A 23),

$$c_0^{(2q)} = \frac{(2q)!}{q!} \frac{\sigma^{2q}}{2^q} - \frac{(2q-1)!}{(2q-2)!} \frac{(2q-2)!}{(q-1)!} \frac{\sigma^{2q-2}}{2^{q-1}} \sigma^2 = 0.$$

Similarly, for the combined order 1 contribution to the even cumulant $\delta^{(1)} c^{(2q)} = c_{1,0}^{(2q)} + c_{0,1}^{(2q)} + c_{2,-1}^{(2q)}$, we find that for $q \geqslant 3$

$$\begin{aligned}
\delta^{(1)} c^{(2q)} = \delta^{(1)} m^{(2q)} &- \binom{2q-1}{0} \left(m_{0,\frac{1}{2}}^{(2q-1)} + m_{1,-\frac{1}{2}}^{(2q-1)} \right) c_0^{(1)} \\
&- \binom{2q-1}{1} (m_0^{(2q-2)} \delta^{(1)} c^{(2)} + \delta^{(1)} m^{(2q-2)} c_0^{(2)}) \\
&- \binom{2q-1}{2} \left(m_{0,\frac{1}{2}}^{(2q-3)} + m_{1,-\frac{1}{2}}^{(2q-3)} \right) c_0^{(3)} - \binom{2q-1}{3} m_0^{(2q-4)} \delta^{(1)} c^{(4)} = 0, \qquad \text{(A 24)}
\end{aligned}$$

where $\delta^{(1)} m^{(2q)} = m_{1,0}^{(2q)} + m_{0,1}^{(2q)} + m_{2,-1}^{(2q)}$. For the odd cumulants $c^{(2q+1)}$ we find recursively that the $\mathcal{O}(\sqrt{\varepsilon})$ contribution for $q \geqslant 2$ is

$$\begin{aligned}
c_{0,\frac{1}{2}}^{(2q+1)} + c_{1,-\frac{1}{2}}^{(2q+1)} = &\left(m_{0,\frac{1}{2}}^{2q+1} + m_{1,-\frac{1}{2}}^{2q+1} \right) - m_0^{2q} c_{0,\frac{1}{2}}^{(1)} - (2q) \left(m_{0,\frac{1}{2}}^{(2q+1)} + m_{1,-\frac{1}{2}}^{(2q+1)} \right) c_0^{(2)} \\
&- \binom{2q}{2} m_0^{(2q)} \left(c_{0,\frac{1}{2}}^{(3)} + c_{1,-\frac{1}{2}}^{(3)} \right) = 0.
\end{aligned}$$

□

References

1. Kamerlin SCL, Vicatos S, Dryga A, Warshel A. 2011 Coarse-Grained (Multiscale) Simulations in Studies of Biophysical and Chemical Systems. *Annual Review of Physical Chemistry* **62**, 41–64.
2. Imkeller P, von Storch JS. 2001 *Stochastic Climate Models*. Birkhäuser.
3. Givon D, Kupferman R, Stuart A. 2004 Extracting macroscopic dynamics: Model problems and algorithms. *Nonlinearity* **17**, R55–127.
4. Pavliotis G, Stuart A. 2008 *Multiscale Methods Averaging and Homogenization*. Texts in Applied Mathematics **53**, Springer.
5. Khasminsky RZ. 1966 On stochastic processes defined by differential equations with a small parameter. *Theory of Probability and its Applications* **11**, 211–228.
6. Kurtz TG. 1973 A limit theorem for perturbed operator semigroups with applications to random evolutions. *Journal of Functional Analysis* **12**, 55–67.
7. Papanicolaou GC. 1976 Some probabilistic problems and methods in singular perturbations. *Rocky Mountain Journal of Mathematics* **6**, 653–674.
8. Beck C. 1990 Brownian motion from deterministic dynamics. *Phys. A* **169**, 324–336.
9. Just W, Kantz H, Rödenbeck C, Helm M. 2001 Stochastic modelling: replacing fast degrees of freedom by noise. *J. Phys. A* **34**, 3199–3213.
10. Melbourne I, Stuart A. 2011 A note on diffusion limits of chaotic skew-product flows. *Nonlinearity* **24**, 1361–1367.
11. Gottwald GA, Melbourne I. 2013 Homogenization for deterministic maps and multiplicative noise. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science* **469**.

12. Kelly D, Melbourne I. 2017 Deterministic homogenization for fast–slow systems with chaotic noise. *Journal of Functional Analysis* **272**, 4063–4102.
13. Dolgopyat D. 2004 Limit theorems for partially hyperbolic systems. *Trans. Amer. Math. Soc.* **356**, 1637–1689).
14. De Simoi J, Liverani C. 2015 The Martingale Approach after Varadhan and Dolgopyat. In *Hyperbolic dynamics, fluctuations and large deviations* vol. 89*Proc. Sympos. Pure Math.* pp. 311–339. Amer. Math. Soc., Providence, RI.
15. De Simoi J, Liverani C. 2016 Statistical properties of mostly contracting fast-slow partially hyperbolic systems. *Invent. Math.* **206**, 147–227.
16. Gear C, Kevrekidis I. 2003 Projective methods for differential equations. *SIAM J. Sci. Comp.* **24**, 1091–1106.
17. Kevrekidis IG, Gear CW, Hyman JM, Panagiotis GK, Runborg O, Theodoropoulos C. 2003 Equation-free, coarse-grained multiscale computation: Enabling microscopic simulators to perform system-level analysis. *Comm. Math. Sci.* **1**, 715–762.
18. E W. 2003 Analysis of the heterogeneous multiscale method for ordinary differential equations. *Comm. Math. Sci.* **1**, 423–436.
19. E W, Engquist B, Li X, Ren W, Vanden-Eijnden E. 2007 Heterogeneous multiscale methods: A review. *Comm. Comp. Phys.* **2**, 367–450.
20. Majda AJ, Timofeyev I, Vanden-Eijnden E. 1999 Models for stochastic climate prediction. *Proceedings of the National Academy of Sciences* **96**, 14687–14691.
21. Majda AJ, Timofeyev I, Vanden-Eijnden E. 2001 A mathematical framework for stochastic climate models. *Communications on Pure and Applied Mathematics* **54**, 891–974.
22. Majda AJ, Timofeyev I, Vanden-Eijnden E. 2002 A priori tests of a stochastic mode reduction strategy. *Phys. D* **170**, 206–252.
23. Majda AJ, Timofeyev I, Vanden-Eijnden E. 2003 Systematic strategies for stochastic mode reduction in climate. *Journal of the Atmospheric Sciences* **60**, 1705–1722.
24. Monahan AH, Culina J. 2011 Stochastic Averaging of Idealized Climate Models. *Journal of Climate* **24**, 3068–3088.
25. Culina J, Kravtsov S, Monahan AH. 2011 Stochastic Parameterization Schemes for Use in Realistic Climate Models. *Journal of the Atmospheric Sciences* **68**, 284–299.
26. Gottwald G, Crommelin D, Franzke C. 2017 Ensemble-based Atmospheric Data Assimilation. In Franzke CLE, O'Kane TJ, editors, *Nonlinear and Stochastic Climate Dynamics* pp. 209–240. Cambridge: Cambridge University Press.
27. Bhattacharya RN, Rao RR. 2010 *Normal Approximation and Asymptotic Expansions* vol. 64*Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
28. Feller W. 1957 *An Introduction to Probability Theory and Its Applications*. New York: Wiley 2d ed edition.
29. de Simoi J, Liverani C, Poquet C, Volk D. 2017 Fast–Slow Partially Hyperbolic Systems Versus Freidlin–Wentzell Random Systems. *Journal of Statistical Physics* **166**, 650–679.
30. Dolgopyat D. 2005 Averaging and invariant measures. *Mosc. Math. J.* **5**, 537–576, 742.
31. Gouëzel S. 2004 Central limit theorem and stable laws for intermittent maps. *Probability Theory and Related Fields* **128**, 82–122.
32. Çınlar E. 2011 *Probability and Stochastics*. Number 261 in Graduate Texts in Mathematics. New York ; London: Springer.
33. Melbourne I, Nicol M. 2005 Almost sure invariance principle for nonuniformly hyperbolic systems. *Commun. Math. Phys.* **260**, 131–146.
34. Melbourne I, Nicol M. 2008 Large deviations for nonuniformly hyperbolic systems. *Trans. Amer. Math. Soc.* **360**, 6661–6676.
35. Melbourne I, Nicol M. 2009 A vector-valued almost sure invariance principle for hyperbolic dynamical systems. *Annals of Probability* **37**, 478–505.
36. Gouëzel S. 2010 Almost sure invariance principle for dynamical systems by spectral methods. *Ann. Probability* **38**, 1639–1671.
37. Götze F, Hipp C. 1983 Asymptotic Expansions for Sums of Weakly Dependent Random Vectors. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **64**, 211–239.
38. Field C, Ronchetti E. 1990 *Small Sample Asymptotics*. Institute of Mathematical Statistics.
39. Götze F, Hipp C. 1994 Asymptotic distribution of statistics in time series. *Ann. Statist.* **22**, 2062–2088.

40. Aït-Sahalia Y. 2002 Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach. *Econometrica* **70**, 223–262.
41. Hervé L, Pène F. 2010 The Nagaev-Guivarc'h method via the Keller-Liverani theorem. *Bull. Soc. Math. France* **138**, 415–489.
42. Rokhlin VA. 1949 On the fundamental ideas of measure theory. *Matematicheskii Sbornik* **67**, 107–150.
43. Wouters J, Lucarini V. 2013 Multi-level Dynamical Systems: Connecting the Ruelle Response Theory and the Mori-Zwanzig Approach. *Journal of Statistical Physics* **151**, 850–860.
44. Wouters J, Lucarini V. 2012 Disentangling multi-level systems: averaging, correlations and memory. *Journal of Statistical Mechanics: Theory and Experiment* **2012**, P03003.
45. Wouters J, Dolaptchiev SI, Lucarini V, Achatz U. 2016 Parameterization of stochastic multiscale triads. *Nonlin. Processes Geophys.* **23**, 435–445.
46. Baladi V, Smania D. 2008 Linear response formula for piecewise expanding unimodal maps. *Nonlinearity* **21**, 677.
47. Baladi V, Smania D. 2010 Alternative proofs of linear response for piecewise expanding unimodal maps. *Ergodic Theory and Dynamical Systems* **30**, 1–20.
48. Baladi V. 2014 Linear Response, or else. In *ICM Seoul 2014, Proceedings, Volume III* pp. 525–545.
49. Baladi V, Benedicks M, Schnellmann D. 2015 Whitney-Hölder continuity of the SRB measure for transversal families of smooth unimodal maps. *Invent. Math.* **201**, 773–844.
50. de Lima A, Smania D. 2016 Central limit theorem for the modulus of continuity of averages of observables on transversal families of piecewise expanding unimodal maps. *Journal of the Institute of Mathematics of Jussieu* pp. 1–61.
51. Gottwald GA, Wormell JP, Wouters J. 2016 On spurious detection of linear response and misuse of the fluctuation-dissipation theorem in finite time series. *Phys. D* **331**, 89–101.
52. Pavliotis GA, Stuart AM. 2008 *Multiscale Methods: Averaging and Homogenization*. New York: Springer.
53. Lorenz EN. 1963 Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* **20**, 130–141.
54. Araújo V, Melbourne I, Varandas P. 2015 Rapid Mixing for the Lorenz Attractor and Statistical Limit Laws for Their Time-1 Maps. *Communications in Mathematical Physics* **340**, 901–938.
55. ODE.jl v2.1 commit 8d4827b93609118633478acf09a74247f47cd97e. https://github.com/JuliaDiffEq/ODE.jl.
56. Dormand J, Prince P. 1980 A Family of Embedded Runge-Kutta Formulae. *Journal of Computational and Applied Mathematics* **6**, 19–26.
57. Wouters J, Gottwald GA. 2018 Stochastic model reduction for slow-fast systems with moderate time-scale separation. *arXiv:1804.09537 [cond-mat, physics:nlin]*. arXiv: 1804.09537.
58. Butterley O, Liverani C. 2007 Smooth Anosov flows: correlation spectra and stability. *Journal of Modern Dynamics* **1**, 301–322.
59. Baladi V. 2017 The quest for the ultimate anisotropic Banach space. *J. Stat. Phys.* **166**, 525–557.
60. Lasota A, Mackey MC. 1985 *Probabilistic Properties of Deterministic Systems*. Cambridge University Press, Cambridge.