# Sample-path large deviations for stochastic evolutions driven by the square of a Gaussian process

Freddy Bouchet*
*Laboratoire de Physique ENS de Lyon and CNRS, 46 Alley d'Italie, F-69364 Lyon Cedex 07, France*

Roger Tribe† and Oleg Zaboronski‡
*Department of Mathematics, University of Warwick, Coventry CV4 7AL, United Kingdom*

(Received 5 November 2021; revised 8 June 2022; accepted 30 January 2023; published 7 March 2023)

Recently, a number of physical models have emerged described by a random process with increments given by a quadratic form of a fast Gaussian process. We find that the rate function which describes sample-path large deviations for such a process can be computed from the large domain size asymptotic of a certain Fredholm determinant. The latter can be evaluated analytically using a theorem of Widom which generalizes the celebrated Szegő-Kac formula to the multidimensional case. This provides a large class of random dynamical systems with timescale separation for which an explicit sample-path large-deviation functional can be found. Inspired by problems in hydrodynamics and atmosphere dynamics, we construct a simple example with a single slow degree of freedom driven by the square of a fast multivariate Gaussian process and analyze its large-deviation functional using our general results. Even though the noiseless limit of this example has a single fixed point, the corresponding large-deviation effective potential has multiple fixed points. In other words, it is the addition of noise that leads to metastability. We use the explicit answers for the rate function to construct instanton trajectories connecting the metastable states.

DOI: 10.1103/PhysRevE.107.034111

## I. INTRODUCTION

Large-deviation theory recently became a key theoretical tool for the statistical mechanics of nonequilibrium systems. Describing sample-path large deviations for the dynamics of effective degrees of freedom leads to a precise understanding of typical and rare trajectories of physical, biological, or economic processes. A paradigm example for the effective descriptions of complex systems using large-deviation theory is the macroscopic fluctuation theory of systems of interacting particles [1]. However, for genuine nonequilibrium processes, without local detailed balance, the class of systems for which the rate function can be found explicitly is extremely limited.

In this paper, we consider a class of systems for which the effective dynamics has increments which are given by a quadratic form of a fast Gaussian process. This type of stochastic driving is relevant for many applications. Quadratic interactions are common in many physical examples, such as hydrodynamics, plasmas described by the Vlasov equation, magnetohydrodynamics, self-gravitating systems, the Kardar-Parisi-Zhang equation, and quadratic networks (for instance, heat transfer across quadratic networks [2]), to cite just a few. For all these systems with quadratic nonlinearities, in some regime a separation of timescales exists and the effective degrees of freedom are coupled to fast-evolving Gaussian processes. This is the case, for example, for the kinetic theories of plasma [3,4], self-gravitating systems [5], geostrophic turbulence [6], and wave turbulence [7] for some specific dispersion relations, among many other examples. From a theoretical and mathematical perspective, modeling the driver of the effective degrees of freedom by a quadratic form of a fast Gaussian process proves to be a decisive simplification. With this assumption, we will be able to write explicit formulas for the sample-path large-deviation rate function and proceed to its analysis in many interesting examples.

The study of a slow process coupled to a fast one is a classical paradigm of physics and mathematics, the celebrated Kapitza pendulum [8] being a canonical example. For such fast and slow dynamics, one can study the averaging of the effect of the fast variable on the slow one (law of large numbers), or the typical fluctuations (stochastic averaging [9]), or the rare fluctuations described by the large-deviation theory [10]. The latter is a natural tool for describing the evolution of metastable systems consisting of long periods spent near an equilibrium point interspersed by rare transitions to a distinct equilibrium along an almost deterministic "instanton" trajectory. A number of systems with timescale separation and drift quadratic in fast variables exhibit metastability (see, e.g., Refs. [11,12] and references therein).

The large-deviation theory has been developed for slow and fast Markov processes [13,14] or deterministic systems [15,16]. Unfortunately, there are not many examples of fast and slow systems for which the large-deviation rate function is known explicitly, which would enable the study of detailed properties of the systems such as the equilibrium points and the transition trajectories connecting them.

*freddy.bouchet @ ens-lyon.fr
†r.p.tribe@warwick.ac.uk
‡o.v.zaboronski@warwick.ac.uk

 

A pedagogical review of the large-deviation theory for systems of stochastic differential equations (SDEs) with two well-separated timescales can be found in Ref. [17]. The theory is illustrated by a class of examples such that the drift for the slow process is given by a second degree polynomial of the fast process. The corresponding large-deviation principle is expressed in terms of the solution of a matrix Ricatti equation. Unfortunately, the resulting expression is not explicit enough to study the practically important phenomenon of *stochastically* generated metastability: all metastable models considered in Ref. [17] already possess multiple fixed points in the noiseless limit. The turbulent models discussed in Ref. [18] suffer from the same flaw: metastability appears to be due to a careful choice of the potential rather than being generated dynamically.

The main contribution of the current work is twofold. First, we apply the asymptotic theory of Fredholm determinants to the calculation of the large deviation rate function; this results is an explicit formula for the rate function which characterizes sample-path large deviations for the slow process in terms of a finite-dimensional determinant of a matrix of the size equal to the number of fast degrees of freedom. Essentially, Szegő's theory of Fredholm determinants is used to build an asymptotic solution to the matrix Ricatti equation of Ref. [17].

Second, we introduce a concrete illustrative example with stochastically generated metastability. This is a system of stochastic differential equations with a single slow variable and a multidimensional fast variable for which all the drifts are quadratic such that in the noiseless limit there is a unique fixed point. However, the addition of noise leads to the appearance of multiple fixed points for the effective Hamiltonian dynamics describing the sample-path deviations. In other words the noisy system exhibits metastability. We use our explicit knowledge of the rate function to construct transition paths (instanton trajectories) between the fixed points.

A third result of our paper is of a more technical nature: as it turns out, it is enough to characterize the fast variables as a Gaussian process with the autocorrelation function which decays sufficiently fast, for example, exponentially. In particular, it is not necessary to require that the fast process be Markovian. Relaxing this assumption opens up a possibility of using our results in turbulence modeling in the following way: the autocorrelation functions of the small-scale turbulence are measured experimentally and used to model the small-scale fluctuations as a fast Gaussian process. Then the large-deviation properties of the large-scale turbulence can be studied theoretically using the theory described below. A rigorous validation of the large-deviation principle without assuming Markovianity is also a very natural question for the probability theory.

It is worth stressing that the current paper does not deal with applications of the developed theory to specific physical systems. However, it has already been proven useful for the study of large deviations in a mean field model of plasma kinetics (see the Ref. [19] for details). We also hope that we can apply the explicit formulas found here to understand the hydrodynamic bistability discussed in Refs. [11,12].

The rest of the paper is organized as follows. We start with the definition of the model in Sec. II and give a heuristic derivation of the corresponding large-deviation principle in Sec. III. The highlight of this section is the application of Widom's theorem for the asymptotics of Fredholm determinants to the calculation of the rate function. In Sec. IV we show the emergence of metastability for a particular representative of our class of models and study the corresponding "instanton" trajectories. Brief conclusions are presented in Sec. V. Appendices A, B, and C contain some technical derivations for Sec. III. Appendix D contains a review of Widom's theorem.

## II. SLOW DYNAMICS QUADRATICALLY DRIVEN BY A FAST GAUSSIAN PROCESS

Consider the following stochastic model:

$$\dot{X}(t) = Y^T\left[\frac{t}{\epsilon}, X(t)\right] M Y\left[\frac{t}{\epsilon}, X(t)\right] - \nu X(t),$$
$$X(0) = x_0, \tag{1}$$

where $\{X(t)\}_{t\geqslant 0}$ is an $\mathbb{R}^n$-valued random process; $\epsilon$ is a parameter which determines the timescale separation between the processes $X$ and $Y$, $0 < \epsilon << 1$; and for a fixed $x \in \mathbb{R}^n$, $[Y(t, x), t, x \in \mathbb{R}]$ is an $N$-dimensional time-stationary centered Gaussian process with the autocorrelation function (covariance matrix)

$$C_{ij}(\tau, x, y) = \mathbb{E}[Y_i(t, x)Y_j(t+\tau, y)],$$
$$\text{where } \tau \geqslant 0,\ 1 \leqslant i, j \leqslant N, \tag{2}$$

which is assumed to be continuous in all the arguments $\tau$, $x$, and $y$. As we will see, only $C(\tau, x, x)$ enters the final expression for the large-deviation rate function, which justifies our shorthand notation $C(\tau, x) := C(\tau, x, x)$. Finally, $M$ is an $n \times N \times N$ matrix, symmetric with respect to the permutation of the last two indices, and $\nu > 0$ is a parameter. Notice that the $(X, Y)$ process need not be Markovian.

We assume that $C(\tau, x)$ decays at least exponentially with $\tau$, perhaps uniformly with respect to $x$. Then, in the limit of $\epsilon \to 0$, the slow random process $X$ stays near the solution to the deterministic equation:

$$\dot{x}(t) = \mathrm{tr}\{MC[0, x(t)]\} - \nu x(t),$$
$$x(0) = x_0, \tag{3}$$

where tr is the trace over $N$ "fast" indices. Equation (3) is a consequence of the ergodic average applied to the integral form of Eq. (1). The *typical* fluctuations of $X(t)$ around $x(t)$ are Gaussian, with covariance of order $\epsilon$ (more precisely, the distribution of $\lim_{\epsilon\to 0}\frac{X(t)-x(t)}{\sqrt{\epsilon}}$ is centered Gaussian). Here we are interested in the statistics of *large deviations* of $X(t)$ when $X(t) - x(t) = O(1)$, which are no longer Gaussian in general.

## III. LARGE-DEVIATION PRINCIPLE FOR PATHS OF THE SLOW PROCESS

If the fast process were Markov, the starting point for our analysis would be the known large-deviation principle for fast and slow Markov systems expressed in terms of the Legendre transform of the cumulant generating functional:

$$Z_T[x, \lambda] = \log \mathbb{E}_Y \exp\left\{\int_0^T dt \lambda(t) f\{x(t), Y[t, x(t)]\}\right\}, \tag{4}$$

where $f$ is the right-hand side of Eq. (1) for the slow degrees of freedom (see Ref. [17] for a review).

However, it turns out that assuming the Gaussianity of $Y$ and the exponential decay of the corresponding autocorrelation function it is possible to arrive at a counterpart of Eq. (4) without assuming Markovianity [see Eq. (5) below]. As we already explained in the Introduction, by extending the range of possible drivers we open up a possibility of applying our results to turbulent modeling.

The following is essentially a computation of the functional integral measure for the slow variable $X$, which we have to use instead of the Martin-Siggia-Rose method [20], which is only applicable to the Markov case. It is not a proof, but rather a heuristic argument devised to give an intuitive feel for the conjectured form of the large-deviation principle. Let us fix the final time $t > 0$, choose a large integer $P \in \mathbb{N}$ and a positive number $\eta$, and define

$$\Delta t = \frac{t}{P},\ b_\eta(x) = \prod_{\alpha=1}^{n}[x_\alpha - \eta, x_\alpha + \eta],$$

where $x$ is a point in $\mathbb{R}^n$ and $\prod$ stands for the direct product of intervals. Geometrically, $b_\eta(x)$ is a hypercube in $\mathbb{R}^n$ centered on $x$ with side $2\eta$. Let $(\lambda_1, \lambda_2, \dots, \lambda_P)$ and $(x_1, x_2, \dots, x_P)$ be two sequences of $n$-dimensional vectors. Let $\mathbb{P}$ be the probability distribution for the process $(X, Y)$. Let $\mathbb{E}$ be the corresponding expectation. We are interested in the probability that at the times $k\Delta t$ the corresponding values of the slow process $X(k\Delta t)$ are near the points $x_k$, $1 \leqslant k \leqslant P$. A computation exploiting Chebyshev's inequality shows that for any sequence of $\lambda$'s

$$\begin{aligned}\epsilon \log \mathbb{P}[X(k\Delta t) &\in b_\eta(x_k),\\ k = 1, \dots, P] &\leqslant \sum_{k=1}^{P}\big(\lambda_k^T(x_{k-1} - x_k)\\ &+ \epsilon \log \mathbb{E}[e^{\lambda_k^T F(Y,x_{k-1})}]\big) + R(\epsilon, \Delta t, \eta),\end{aligned} \tag{5}$$

where

$$F(Y, x) = \int_0^{\Delta t/\epsilon} d\tau Y^T(\tau, x)MY(\tau, x) - \nu x\Delta t/\epsilon, \tag{6}$$

and $R$ is an error term depending on $\epsilon$, $\eta$, and $\Delta t$ such that for $\eta = \mu\Delta t$

$$\lim_{\mu\to 0}\lim_{\Delta t\to 0}\lim_{\epsilon\to 0} R(\epsilon, \Delta t, \mu\Delta t) = 0. \tag{7}$$

The derivation of Eq. (5) is based on the approximation of $Y$ by a bounded process with a finite dependency range. It is carried out in Appendix A. Here we would only like to point out that the dependence on $\lambda$ in the right-hand side of Eq. (5) appears to be due to the repeated use of Chebyshev's inequality. Intuitively, the sequence $(\lambda_k)$ is the discretized counterpart of the response field appearing in the Martin-Siggia-Rose computation.

The next aim is to compute the expectation

$$\mathbb{E}[e^{\lambda^T F(Y,x)}] = e^{-\frac{\Delta t}{\epsilon}\nu\lambda^T x}\mathbb{E}[e^{\lambda^T[\int_0^{\Delta t/\epsilon} d\tau Y^T(\tau,x)MY(\tau,x)]}],$$

which can be done using the fact that for a fixed $x \in \mathbb{R}^n$ the process $Y(\cdot, x)$ is stationary and Gaussian. What follows is the key computation of the paper, linking averaging over fast Gaussian fields with the asymptotic of certain Fredholm determinants. Let us define $m := \lambda^T M$, an $N \times N$ symmetric matrix. It can be decomposed as $m = S^T S$, where $S$ is a possibly complex Cholesky factor of $m$. Rewrite

$$\begin{aligned}&\exp\left[\lambda^T\left(\int_0^{\Delta t/\epsilon} d\tau Y^T(\tau, x)MY(\tau, x)\right)\right]\\ &\quad = \int\prod_\tau \mathfrak{D}q(\tau)e^{-\frac{1}{4}\int_0^T d\tau q^T(\tau)q(\tau)+\int_0^{\Delta t/\epsilon} d\tau q^T(\tau)SY(\tau,x)}\end{aligned}$$

(the Hubbard-Stratonovich transformation). Then, for sufficiently small components of $\lambda$,

$$\begin{aligned}&\mathbb{E}\exp\left[\lambda^T\left(\int_0^{\Delta t/\epsilon} d\tau Y^T(\tau, x)MY(\tau, x)\right)\right]\\ &= \int\prod_\tau \mathfrak{D}q(\tau)e^{-\frac{1}{4}\int_0^{\frac{\Delta t}{\epsilon}} d\tau q^T(\tau)q(\tau)}\mathbb{E}(e^{\int_0^{\frac{\Delta t}{\epsilon}} d\tau q^T(\tau)SY(\tau,x)})\\ &= \int\prod_\tau \mathfrak{D}q(\tau)\exp\left[-\frac{1}{4}\int_0^{\Delta t/\epsilon} d\tau q^T(\tau)q(\tau)\right.\\ &\quad \left.+\frac{1}{2}\int_0^{\Delta t/\epsilon} d\tau_1\int_0^{\Delta t/\epsilon} d\tau_2 q^T(\tau_1)SC(\tau_1 - \tau_2, x)S^T q(\tau_2)\right]\\ &= \mathrm{Det}^{-\frac{1}{2}}(I - 2S\hat{C}_{\Delta t/\epsilon}(x)S^T) = \mathrm{Det}^{-\frac{1}{2}}(I - 2m\hat{C}_{\Delta t/\epsilon}(x)).\end{aligned} \tag{8}$$

Here $m\hat{C}_{\Delta t/\epsilon}(x)$ is an integral operator acting on (square integrable) $\mathbb{R}^N$-valued functions on $[0, \Delta/\epsilon]$ as follows:

$$\begin{aligned}f_\alpha(t) \mapsto m\hat{C}_{\Delta t/\epsilon}(x)(f)_\alpha(t) = \sum_{\beta,\delta=1}^{N}\int_0^{\Delta t/\epsilon} d\tau m_{\alpha\beta}\\ \times C_{\beta,\delta}(t - \tau, x)f_\delta(\tau),\end{aligned} \tag{9}$$

for all $\alpha = 1, \dots, N$; $t \in [0, \Delta t/\epsilon]$.

In what follows we will use capital Det and Tr to denote operator determinant and trace, and reserve the lowercase det and tr for the determinant and the trace of finite-dimensional matrices.

The calculation of Eq. (8) in the limit $\epsilon \to 0$ requires the asymptotic analysis of the Fredholm determinant of an integral operator acting on functions defined on a large interval. Fortunately, such an asymptotic can be computed using Widom's theorem, which generalizes the celebrated Szegő-Kac formula for Fredholm determinants (see Ref. [21]): for a sufficiently small $m$ (e.g., with respect to a matrix norm),

$$\begin{aligned}\log\mathrm{Det}(I - 2m\hat{C}_{\Delta t/\epsilon}(x)) = \frac{\Delta t}{\epsilon}\int_{\mathbb{R}}\frac{dk}{2\pi}\log\det\\ \times (I - 2m\tilde{C}(k, x)) + O(1),\end{aligned} \tag{10}$$

where $\tilde{C}(k, x) = \int_{\mathbb{R}} d\tau e^{ik\tau}C(\tau, x)$ is the Fourier transform of the autocorrelation function $C(\tau, x)$. This remarkable statement is reviewed in Appendix D. Substituting Eqs. (8) and (10) into Eq. (5), we find

$$\epsilon \log \mathbb{P}[X(p\Delta t) \in b_\eta(x_p),\ p=1,\dots,P] \overset{(5)}{\leqslant} \sum_{p=1}^{P} \Delta t \lambda_p^T \left(\frac{x_{p-1}-x_p}{\Delta t} - \nu x_{p-1}\right)$$
$$+ \sum_{p=1}^{P} \epsilon \log \mathbb{E} \exp\left[\int_0^{\Delta t/\epsilon} d\tau Y^T(\tau, x_{p-1}) m Y(\tau, x_{p-1})\right] + R$$
$$\overset{(8)}{=} \sum_{p=1}^{P} \Delta t \lambda_p^T \left(\frac{x_{p-1}-x_p}{\Delta t} - \nu x_{p-1}\right) - \frac{1}{2}\sum_{p=1}^{P} \epsilon \log \mathrm{Det}(I - 2m\hat{C}_{\frac{\Delta t}{\epsilon}}(x_{p-1})) + R$$
$$\overset{(10)}{=} \sum_{p=1}^{P} \Delta t \lambda_p^T \left(\frac{x_{p-1}-x_p}{\Delta t} - \nu x_{p-1}\right) - \frac{1}{2}\sum_{p=1}^{P} \epsilon \left[\frac{\Delta t}{\epsilon}\int_{\mathbb{R}} \frac{dk}{2\pi} \log\det(I - 2m\tilde{C}(k, x_{p-1}))\right] + R + O(\epsilon P)$$
$$= \sum_{p=1}^{P} \Delta t \lambda_p^T \left(\frac{x_{p-1}-x_p}{\Delta t} - \nu x_{p-1}\right) - \frac{1}{2}\sum_{p=1}^{P} \Delta t \int_{\mathbb{R}} \frac{dk}{2\pi} \log\det\left(I - 2\lambda_p^T M\tilde{C}(k, x_{p-1})\right) + R + O(\epsilon P), \tag{11}$$

where the $O(\epsilon P)$ addition to the error term comes from the $O(1)$ term in Eq. (10). The expression (11) is an upper bound on the (discretization of) the functional integral measure for the process $X$.

The next step is akin to the calculation of a path integral for $\epsilon \to 0$ using the Laplace method. The question we ask is: what is the probability $\mathbb{P}[X \in D]$, where $D$ is a "nice" subset of the space $C([0,t], \mathbb{R}^n)$ of $\mathbb{R}^n$-valued functions on $[0,t]$?

By analogy with the finite-dimensional Laplace method, one needs to minimize the functional integral measure (11) over $D$. The details of this computation can be found in Appendix B. Here we just state the answer after taking the continuous limit $\Delta t \to 0$:

$$\limsup_{\epsilon\to 0} \epsilon \mathbb{P}[X \in D] \leqslant -\inf_{x\in D}\left[\int_0^t d\tau \lambda_p^T(\tau)[\dot{x}(\tau) + \nu x(\tau)] + \frac{1}{2}\int_{\mathbb{R}} \frac{dk}{2\pi} \log\det(I - 2\lambda^T(\tau) M\tilde{C}[k, x(\tau)])\right]. \tag{12}$$

The derived bound is valid for an arbitrary function $\lambda$. Taking the infimum of the right-hand side of Eq. (12) over this function, one gets the optimal upper bound on $\mathbb{P}[X \in D]$:

$$\limsup_{\epsilon\to 0} \epsilon \mathbb{P}[X \in D] \leqslant -\sup_{\lambda}\inf_{x\in D}\left[\int_0^t d\tau \lambda_p^T(\tau)[\dot{x}(\tau) + \nu x(\tau)] + \frac{1}{2}\int_{\mathbb{R}} \frac{dk}{2\pi} \log\det(I - 2\lambda^T(\tau) M\tilde{C}[k, x(\tau)])\right]. \tag{13}$$

Staying at the similar level of rigor and using the same set of assumptions about the fast process as above, one can show that the right-hand side of Eq. (13) is also a lower bound on $\liminf_{\epsilon\to 0} \epsilon \mathbb{P}[X \in D]$. The corresponding calculation is based on a standard trick of deforming the probability distribution in such a way that the low-probability event at hand becomes almost inevitable (see, e.g., Ref. [22] for a short introduction). The details are given in Appendix C.

Therefore, it is natural to conjecture that the slow process $X$ satisfies the large-deviation principle with the rate $\epsilon$ and the explicit rate function given by

$$S_{\mathrm{eff}}[\lambda, x] = \int_0^t d\tau \lambda^T(\tau)[\dot{x}(\tau) + \nu x(\tau)] + \frac{1}{2}\int_0^t d\tau \times \int_{\mathbb{R}} \frac{dk}{2\pi} \log\det(I - 2\lambda^T(\tau) M\tilde{C}[k, x(\tau)]), \tag{14}$$

provided that $t$ is not too large. Less formally one can write

$$\mathbb{P}[X \in D] \sim e^{-\frac{1}{\epsilon}\sup_\lambda \inf_{x\in D} S_{\mathrm{eff}}[\lambda, x]}. \tag{15}$$

A typical application of the rate functional guessed above is the estimation of the probability of transitioning between fixed points of the typical evolution (3). If $x_0$ and $x_1$ are two such points, then

$$\mathbb{P}[X(t) \in dx_1 \mid X(0) = x_0] \sim e^{-\frac{1}{\epsilon}\sup_\lambda \inf_x S_{\mathrm{eff}}[\lambda, x]}, \tag{16}$$

where the "inf" and the "sup" are taken over the functions $x$ and $\lambda$ on $[0,t]$ such that $x(0) = x_0$ and $x(t) = x_1$.

When analyzing specific examples, it is often convenient to think of Eq. (14) as the action functional for a mechanical system with generalized coordinates $x$ and generalized momenta $\lambda = \frac{\delta S_{\mathrm{eff}}}{\delta \dot{x}}$. This system is Hamilton's with the Hamiltonian

$$H_{\mathrm{eff}}(\lambda, x) = -\nu\lambda^T x - \int_{\mathbb{R}} \frac{dk}{4\pi} \log\det(I - 2\lambda^T M\tilde{C}(k, x)) \tag{17}$$

(see Ref. [8] for details of the map between the Lagrangian and the Hamiltonian formalisms). As a self-consistency check, let us verify that the average evolution equation (3) appears as an equation for a typical trajectory for the large-deviation principles (16) and (14). A typical trajectory $(\lambda_c, x_c)_{0\leqslant\tau\leqslant t}$ is a solution to Euler-Lagrange equations associated with $S_{\mathrm{eff}}$ such that

$$S_{\mathrm{eff}}[\lambda_c, x_c] = 0.$$

Examining the derivation of the large-deviation principle, it is reasonable to expect that $\lambda_c = 0$. Expanding Eq. (14) around

$\lambda = 0$, we find

$$S_{\rm eff} = \int_0^t d\tau \lambda^T(\tau)(\dot{x}(\tau) + \nu x(\tau) - \mathrm{Tr}\{MC[0, x(t)]\}) + O(\lambda^2),$$

where we used that $\int_{\mathbb{R}} \frac{dk}{2\pi} \tilde{C}(k, x) = C(0, x)$. Therefore, $\lambda = 0$ solves the Euler-Lagrange equations if

$$\dot{x}(\tau) + \nu x(\tau) - \mathrm{Tr}\{MC[0, x(t)]\} = 0,\ x(0) = x_0,$$

which coincides with Eq. (3). In particular, the fixed points of the slow dynamics are solutions to

$$\nu x = \mathrm{Tr}[MC(0, x)]. \tag{18}$$

**Remarks**

(i) If $N = 1$, and $Y$ solves an Ornstein-Uhlenbeck SDE with an $X$-dependent drift, the corresponding large-deviation principle is as derived in Ref. [11] and is consistent with conjecture (13) for all values of $\lambda$. However, in general, one has to check that the optimal $\lambda$ belongs to the domain of applicability of Widom's theorem, which is one of the challenges for the rigorous justification of the conjecture. A natural guess is that the minimizer must be small enough to ensure positive definiteness of the quadratic form in the functional integral (8).

(ii) If $Y$ appears as a solution to an Ornstein-Uhlenbeck system of stochastic differential equations, then Eq. (13) can be viewed as a solution to the matrix Riccati problem for the rate function derived in Ref. [17].

(iii) In the context of modeling of two-dimensional turbulent flows, Eq. (1) can be interpreted as follows: $Y$ is a Gaussian model of fast small-scale velocity field whose evolution depends on the static background created by $X$; $X$ is a large-scale velocity field slowly evolving under the influence of $Y$. Thus, the model can be thought of as a nonlinear generalization of the passive vector advection model. The shape of $C$ reflects the nature of the small-scale turbulent flow (compressibility, isotropy, etc.).

## IV. AN EXAMPLE INSPIRED BY MULTISTABILITY IN HYDRODYNAMIC AND GEOSTROPHIC TURBULENCE

The aim of this section is to present an example of the use of the large-deviation principle (14). We are specifically interested in metastability phenomena observed in two-dimensional [11] and geostrophic [12] turbulent flows. In previous works, we have studied metastability for geostrophic dynamics [18], in cases when the turbulent flows are forced by white noises and the stochastic process is an equilibrium one with detailed balance or generalized detailed balance. The large-deviation principle (14) opens the possibility for studying metastability for turbulent flows modeled as a nonequilibrium process. As a first step, we now demonstrate the stochastic generation of metastability for systems with timescale separation using the simplest example of the system of stochastic differential equations with the quadratic drift for the slow variable.

To formulate the example, it is easier to use complex notations. The fast variable $Y(\cdot, x) \in \mathbb{C}^N$ is an analog of the set of Fourier components that describe the turbulent fluctuations. $Y$ is the stationary solution of the complex Ornstein-Uhlenbeck process. The SDEs for the full fast and slow system are as follows:

$$dY(t, x) = -\Gamma(x)Y(t, x)dt + \sigma dW(t),$$
$$dX(t) = Y[t/\epsilon, X(t)] * MY[t/\epsilon, X(t)]dt - \nu X(t)dt, \tag{19}$$

where $M$ is an $n \times N \times N$ matrix self-adjoint with respect to the last two indices; $dW$ is the $\mathbb{C}^N$-valued Brownian motion, with the nontrivial covariance

$$d\overline{W}_i dW_j = \delta_{ij}dt; \tag{20}$$

and $\Gamma(x)$ is a complex matrix, whose eigenvalues have positive real parts,

$$\Gamma(x) = \Gamma^{(0)} + ix^T\Gamma^{(1)}, \tag{21}$$

where $\Gamma^{(0)}$ is a real positive definite $N \times N$ matrix and $\Gamma^{(1)}$ is a real $n \times N \times N$ matrix. The former describes dissipation, whereas the latter corresponds to the "rotational" advection of $Y$ by the slow field $X$. All the coefficients are polynomials of degree at most one in $x$.

The model (19) is a representative of the class of models (1) and (2) treated in this paper: the slow field $X$ is driven by a quadratic form of $Y(\cdot, x)$ which, as follows from the first of the SDEs (19), is Gaussian with the exponentially decaying autocorrelation function. In particular, the large-deviation rate function can be derived from Eq. (14) of the previous section. Finally, notice that the process $(X, Y)$ defined by Eq. (19) is Markovian. As pointed out at the end of Sec. III, the task of justifying the large-deviation principle (14) reduces in this case just to the check of the applicability of Widom's theorem, whereas the important intermediate result (5) can be established rigorously (see Ref. [14]).

The structure of the system of SDEs (19) resembles that of the quasilinear approximation to the Navier-Stokes equation or quasigeostrophic equations (see Ref. [6] for details): the nonlinearity in the right-hand side is quadratic, the evolution of the slow variable is driven by the term quadratic in the fast variable, and the drift of the fast variable resembles advection by the slow field $X$. Let us stress that the model does not have any artificial "built-in" nonlinearity: the noiseless limit of Eq. (19) has a unique critical point $X = Y = 0$. The metastability described below is a purely stochastic effect.

Some standard computations lead to formulas for the correlation and autocorrelation functions, $C(0, x) := \mathbb{E}[Y(0, x) \otimes Y^*(0, x)]$ and $C(\tau, x) := \mathbb{E}[Y(\tau, x) \otimes Y^*(0, x)]$. Here $\otimes$ denotes the tensor product: for vectors $a, b \in \mathbb{C}^n$, $a \otimes b$ is an $n \times n$ matrix such that $(a \otimes b)_{ij} = a_i b_j$. $C(0, x)$ solves the Lyapunov equation

$$\Gamma(x)C(0, x) + C(0, x)\Gamma^*(x) = \sigma\sigma^*, \tag{22}$$

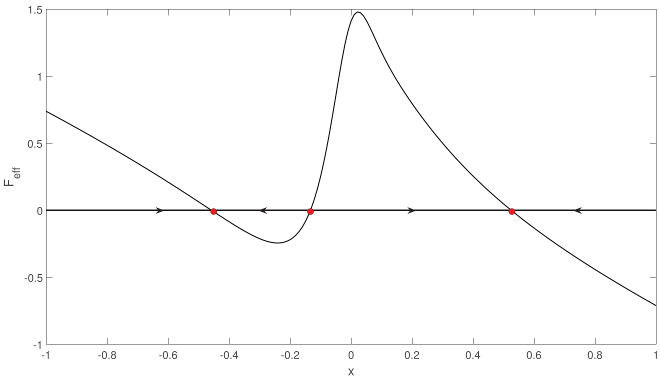


FIG. 1. The effective force $F_{\rm eff}(x) := \frac{\partial H_{\rm eff}}{\partial\lambda}(0,x)$ for the model (27). Notice a pair of stable fixed points of the averaged dynamics separated by an unstable fixed point.

whereas

$$C(\tau, x) = e^{-\Gamma(x)\tau}C(0, x),\ \ \tau \geqslant 0. \tag{23}$$

If $\tau < 0$, then $C(\tau, x) = C(0, x)e^{\Gamma^*(x)\tau}$. The effective Hamiltonian (17) rewritten in complex terms is

$$H_{\rm eff}(\lambda, x) = -\lambda^T \nu x - \int_{\mathbb{R}} \frac{dk}{4\pi} \log\det(I - 2\lambda^T M\tilde{C}(k, x)), \tag{24}$$

where

$$\begin{aligned}\tilde{C}(k, x) &:= \int_{\mathbb{R}} d\tau e^{ik\tau} C(\tau)\\ &= [\Gamma(x) - ik]^{-1}C(0, x) + C(0, x)[\Gamma^*(x) + ik]^{-1},\end{aligned} \tag{25}$$

is the Fourier transform of the autocorrelation function.

Keeping matters as simple as possible, let us choose $\Gamma^{(0)}$ and $\Gamma^{(1)}$ to be the diagonal matrices with real entries $\{\gamma_p^{(0)}, \gamma_p^{(1)}\}_{1\leqslant p\leqslant N}$, where $\gamma^{(0)}$'s are all positive. The fixed-point equation (18) takes the form

$$\begin{aligned}&\sum_{j,k=1}^{N} \frac{(\sigma\sigma^*)_{jk}(M_\alpha)_{kj}}{\{\gamma_j^{(0)} + \gamma_k^{(0)} + i\sum_{\beta=1}^{n}\big[(\gamma_\beta^{(1)})_j - (\gamma_\beta^{(1)})_k\big]x_\beta\}}\\ &\quad = \nu x_\alpha,\ \ 1 \leqslant \alpha \leqslant n.\end{aligned} \tag{26}$$

Notice that if either the noise covariance matrix $\sigma\sigma^*$ or the interaction matrix $M_\alpha$ is diagonal, there is a unique solution for the $\alpha$th component of the fixed point. Indeed, if $(M_\alpha)_{kj} = 0$ for all $k \neq j$, then the left-hand side of Eq. (26) becomes $x$ independent and the equation becomes linear with respect to $x_\alpha$. The same remark applies if $\sigma\sigma^*$ is diagonal. Similarly, the fixed point is unique if $(\gamma_\beta^{(1)})_j - (\gamma_\beta^{(1)})_k = 0$ for all $j$, $k$, and $\beta$. However, for general correlated noise, interaction, and an inhomogeneous rotation matrix $\gamma^{(1)}$, there are typically multiple solutions to Eq. (26).

We therefore conclude with the simplest nontrivial example such that Eq. (26) has multiple real solutions (see Fig. 1). For this example, $n = 1$, $N = 3$, $\nu = I_3$, and it has two stable and one unstable fixed points. The appearance of powers of 2 and $\pi$ in the following parametrization has no special

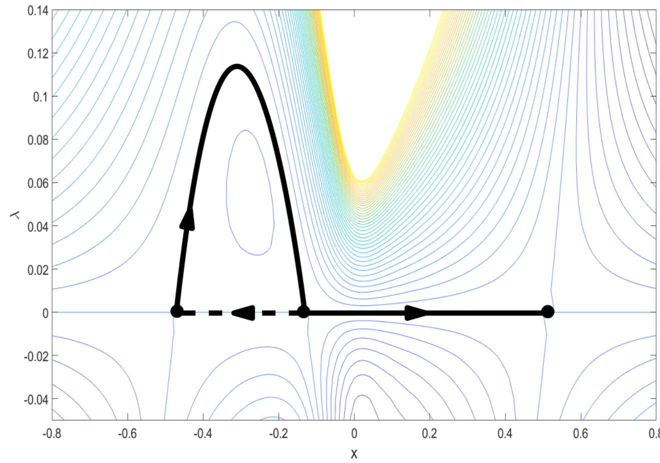


FIG. 2. Contour lines of $H_{\rm eff}$ for the model (27). The contour lines in the upper half plane serve as optimal trajectories for transitions between the stable fixed points in a finite time. The wide curve is the infinite-time optimal transition curve. The dashed segment marks the typical trajectory connecting the unstable and stable fixed points.

meaning:

$$\sigma = \frac{1}{2^{1/4}}\begin{pmatrix}-1-i & 0 & 0\\ 1-i & -1-i & 0\\ -1-i & 1-i & -1-i\end{pmatrix},$$

$$M = \begin{pmatrix}0 & 1 & 1\\ 1 & 0 & 1\\ 1 & 1 & 0\end{pmatrix},\ \gamma^{(0)} = \begin{pmatrix}1\\1\\1\end{pmatrix},\ \gamma^{(1)} = \pi^2\begin{pmatrix}1\\2\\3\end{pmatrix}. \tag{27}$$

The choice $M_{ii} = 0$ and $M_{ij} = \text{const}$ for $i \neq j$ reflects some properties of the interaction matrix for the two-dimensional Navier-Stokes equation, but it is also not essential for the appearance of multiple equilibria.

The fact that multiple equilibria appear naturally in the model (19) together with its link to quasilinear hydrodynamics explained above makes us hope that the large-deviation principle (14) might prove useful in studying realistic hydrodynamic phenomena of metastability, such as the zonal-dipole transition discovered in Ref. [11]. Euler-Lagrange equations associated with the effective action functional (14) are Hamilton's with the Hamiltonian (24). Therefore, each solution lies on a constant energy surface $H_{\rm eff}(\lambda, x) = E$. If there is a single slow variable, the trajectories coincide with constant energy surfaces. This allows one to determine a family of the most likely transition paths between the fixed points (the instanton trajectories) by building the contour plot of $H_{\rm eff}$ numerically (see Fig. 2).

## V. CONCLUSIONS AND OUTLOOK

Motivated by hydrodynamic applications, we have considered a model with two timescales, where the slow variable is driven by a quadratic function of a fast Gaussian process with rapidly decaying autocorrelations. A natural question of computing the probabilities of rare events in this model reduces to the computation of large-interval asymptotics for a certain Fredholm determinant. To the leading order, such a computation can be easily carried out using Widom's theorem. To apply the resulting large-deviation principle, we considered a special case of the fast field being a complex Ornstein-Ohlenbeck process with the rotational component of the drift given by a linear function of the slow process. As it turns out, the average slow dynamics for such a model exhibits multiple equilibria, the transitions between which can be studied using large-deviation theory.

There are many natural further questions to ask. First, it should be a straightforward task to furnish a rigorous proof or provide a counterexample to the statement of the conjecture (14). Second, for the cases, when the fast process conditional on the value of the slow process is an Ornstein-Uhlenbeck process, it might be interesting to consider finite-$\epsilon$ corrections to the leading-order answer. Albeit known, the subleading terms in the Widom asymptotic are only characterized as solutions to a certain matrix Wiener-Hopf integral equation. There is, however, a chance of finding these corrections rather more explicitly as solutions to the time-dependent Riccati equations derived in Ref. [17].

Finally, the model considered has the general structure of many equations of hydrodynamics, plasma dynamics, self-gravitating systems, wave turbulence, or other physical systems with quadratic couplings or interactions. It would, therefore, be extremely interesting to analyze metastability for such physical systems, in the presence of timescale separation, using the findings of the present paper.

## APPENDIX A: THE DERIVATION OF EQ. (5)

In the calculation below we rely on the observation that the Gaussian process $Y(\cdot, x)$ with exponentially decaying autocorrelation function should be well approximated by a bounded process with a finite dependency range. In other words, we assume that there are constants $C$ and $\delta$: $Y^T(t,x)Y(t,x) \leqslant C$ for all $(t,x)$, and the processes $[Y(t,x)]_{t<T}$ and $[Y(t,x)]_{t>T+\delta}$ are independent for any $T \in \mathbb{R}$ [23].

To estimate a probability in terms of exponential moments, we follow the logic of the Chernoff bound (the exponential version of Chebyshev's inequality) and notice the following elementary inequality: for any $\lambda \in \mathbb{R}$,

$$\mathbb{1}[x \in (y-\eta, y+\eta)] \leqslant e^{\lambda(x-y)+\eta|\lambda|}\mathbb{1}[x \in (y-\eta, y+\eta)].$$

The corresponding $n$-dimensional generalization is

$$\mathbb{1}[x \in b_\eta(y)] \leqslant e^{\lambda^T(x-y)+\eta n||\lambda||_\infty}\mathbb{1}[x \in b_\eta(y)], \tag{A1}$$

where $\lambda, x, y \in \mathbb{R}^n$ and $||\lambda||_\infty = \max_{1\leqslant\alpha\leqslant n}|\lambda_\alpha|$. Using Eq. (A1),

$$\begin{aligned}\mathbb{P}[X(k\Delta t) \in b_\eta(x_k), 1 \leqslant k \leqslant P] &= \mathbb{E}\left[\prod_{k=1}^P \mathbb{1}[X(k\Delta t) \in b_\eta(x_k)]\right]\\ &\leqslant \mathbb{E}\left[e^{\sum_{l=1}^P[\frac{\lambda_l^T}{\epsilon}[X(l\Delta t)-x_l]+\eta n||\lambda_l||_\infty]}\right.\\ &\quad\left.\times \prod_{k=1}^P \mathbb{1}[X(k\Delta t) \in b_\eta(x_k)]\right].\end{aligned} \tag{A2}$$

Next we need to calculate $X(k\Delta t) - x_k$ for each $k$ by solving Eq. (1) over the time interval $[(k-1)\Delta t, k\Delta t]$: Denote the right-hand side of the equation for $\dot{X}(\tau)$ by $f[\tau/\epsilon, X(\tau)]$. Our assumptions imply that both $||f||_\infty$ and $||\nabla_2 f||_\infty$ are bounded by some constants, which will be denoted by $M_0$ and $M_1$, correspondingly. Expanding $f$ in the Taylor series in the second argument, one finds

$$\lambda^T\dot{X}(\tau) = \lambda^T f(\tau/\epsilon, x_{k-1}) + \rho_k,\ \ \tau \in [(k-1)\Delta t, k\Delta t], \tag{A3}$$

where $\rho_k = \lambda^T\nabla_2 f[\tau/\epsilon, (1-c)x_{k-1} + cX(\tau)][X(\tau) - x_{k-1}]$, for some $c \in (0,1)$. Here we used the mean value form of the remainder for the Taylor series. Using the bound $||\nabla_2 f||_\infty \leqslant M_1$ and noticing that $\nabla_2 f$ is an $n \times n$ matrix,

$$|\rho_k| \leqslant n^2||\lambda||_\infty M_1||X(\tau) - x_{k-1}||_\infty. \tag{A4}$$

The estimate of the size of $X(\tau) - x_{k-1}$ uses the equation for $X$ once more:

$$\begin{aligned}||X(\tau) - x_{k-1}||_\infty &\leqslant ||X(\tau) - X[\Delta t(k-1)]||_\infty\\ &\quad + ||X[\Delta t(k-1)] - x_{k-1}||_\infty \leqslant \eta + ||\\ &\quad\times \int_{(k-1)\Delta t}^{\tau} d\tau' f[\tau'/\epsilon, X(\tau')]||_\infty \leqslant \eta + \Delta t M_0.\end{aligned}$$

The penultimate step uses the fact that the indicators under the sign of the expectation in Eq. (A2) enforce the constraint $||X[(k-1)\Delta t] - x_{k-1}||_\infty < \eta$; the last step uses the bound $||f||_\infty < M_0$. Putting it all together we find that

$$|\rho_k| \leqslant n^2||\lambda||_\infty(\eta + \Delta t M_0)M_1. \tag{A5}$$

Integrating Eq. (A3) over the interval $[(k-1)\Delta t, k\Delta t]$, we conclude that

$$\begin{aligned}\lambda_k^T[X(k\Delta t) - x_k] &= \lambda_k^T(x_{k-1} - x_k)\\ &\quad + \lambda_k^T\int_{(k-1)\Delta t}^{k\Delta t} d\tau f(\tau/\epsilon, x_{k-1}) + \tilde{\rho}_k,\end{aligned} \tag{A6}$$

where $||\tilde{\rho}_k||_\infty \leqslant ||\lambda_k||_\infty[n\eta + n^2(\eta + \Delta t M_0)M_1\Delta t]$. Notice the extra contribution to the error term coming from one more application of the bound $||X[(k-1)\Delta t] - x_{k-1}||_\infty \leqslant \eta$. Substituting Eq. (A6) into Eq. (A2) and upper-bounding the product of the indicators by 1, one arrives at the following intermediate result:

$$\begin{aligned}&\mathbb{P}[X(k\Delta t) \in b_\eta(x_k), k = 1, \ldots, P]\\ &\leqslant e^{\sum_{k=1}^P \frac{\lambda_k^T}{\epsilon}(x_{k-1}-x_k)}\mathbb{E}\left[\prod_{k=1}^P e^{\frac{\lambda_k^T}{\epsilon}\left[\int_{(k-1)\Delta t}^{k\Delta t} d\tau f(\tau/\epsilon, x_{k-1}) + R_k\right]}\right],\end{aligned} \tag{A7}$$

where $R_k = ||\lambda_k||_\infty[2n\eta + n^2(\eta + \Delta t M_0)M_1\Delta t]$. It remains to approximate the expectation in Eq. (A7) by the product of expectations. To this end we write

$$\lambda_k^T \int_{(k-1)\Delta t}^{k\Delta t} d\tau f(\tau/\epsilon, x_{k-1})$$
$$= \epsilon\lambda_k^T \int_{(k-1)\Delta t/\epsilon}^{k\Delta t/\epsilon} d\tau f(\tau, x_{k-1})$$
$$= \epsilon\lambda_k^T \int_{(k-1)\Delta t/\epsilon+\delta}^{k\Delta t/\epsilon-\delta} d\tau f(\tau, x_{k-1}) + E_k, \quad \text{(A8)}$$

where the bound $||f||_\infty \leqslant M_0$ implies that $|E_k| \leqslant 2\epsilon n||\lambda_k||_\infty M_0\delta$. Crucially, notice that $f(\tau/\epsilon, x)$ depends on $Y(\tau/\epsilon, x)$ only. Therefore, the random variables $\epsilon\lambda_k^T \int_{(k-1)\Delta t/\epsilon-\delta}^{k\Delta t/\epsilon+\delta} d\tau f(\tau, x_{k-1})$ and $1 \leqslant k \leqslant P$ are mutually independent due to the finite dependency range $\delta$ of the process $Y$. Substituting Eq. (A8) into Eq. (A7) and exploiting the independence, one finds

$$\mathbb{P}[X(k\Delta t) \in b_\eta(x_k),\ k = 1, \dots, P] \leqslant e^{\sum_{k=1}^P \frac{\lambda_k^T}{\epsilon}(x_{k-1}-x_k+\tilde{R}_k)}$$
$$\times \prod_{k=1}^P \mathbb{E}[e^{\lambda_k^T \int_{(k-1)\Delta t/\epsilon}^{k\Delta t/\epsilon} d\tau f(\tau, x_{k-1})}],$$

where $\tilde{R}_k = ||\lambda_k||_\infty[4\epsilon n M_0\delta + 2n\eta + n^2(\eta + \Delta t M_0)M_1\Delta t]$. In the last expression we extended the integration interval back to $[(k-1)\Delta t, k\Delta t]/\epsilon$, which explains the doubling of the $\delta$-dependent contribution to the error term. Finally, let us notice that the total error term $R(\epsilon, \Delta t, \eta) := \sum_{k=1}^P \tilde{R}_k$ has the following property:

$$\lim_{\mu\to 0}\lim_{\Delta t\to 0}\lim_{\epsilon\to 0} R(\epsilon, \Delta t, \mu\Delta t) = 0. \quad \text{(A9)}$$

The derivation of Eq. (5) is complete.

## APPENDIX B: THE DERIVATION OF EQ. (12)

First of all, let us explain what we mean by a "nice" set of functions $D$. To this end, we need to introduce one more notation. For $f \in C([0, T], \mathbb{R}^n)$, let

$$B_\eta(f) = \{g \in C([0, t], \mathbb{R}^n) :$$
$$\times \sup_{\tau\in[0,t], 1\leqslant\alpha\leqslant n} |f_\alpha(\tau) - g_\alpha(\tau)| < \eta\}. \quad \text{(B1)}$$

This is an infinite-dimensional generalization of the hypercube $b_\eta$ introduced above. We say that set $D$ is nice if for any $\eta > 0$ we can find finitely many smooth functions $x^{(1)}, x^{(2)}, \dots, x^{(M)} \in C([0, T], \mathbb{R}^n)$:

$$D \subset \bigcup_{j=1}^M B_\eta(x^{(j)}). \quad \text{(B2)}$$

In other words $D$ can be covered by finitely many hypercubes of any positive "linear size" [24]. Let $x_k^{(j)} = x^{(j)}(k\Delta t)$ and $1 \leqslant k \leqslant P$. Then

$$\mathbb{P}[X \in D] \overset{(B2)}{\leqslant} \mathbb{P}\left[X \in \bigcup_{j=1}^M B_\eta(x^{(j)})\right] = \mathbb{P}[\exists j \leqslant M : X \in B_\eta(x^{(j)})] = \mathbb{P}[\exists j \leqslant M : X(\tau) \in b_\eta(x^{(j)}(\tau)),\ \tau \in [0, t]]$$
$$\leqslant \mathbb{P}\big[\exists j \leqslant M : X(k\Delta t) \in b_\eta\big(x_k^{(j)}\big),\ 1 \leqslant k \leqslant P\big]$$
$$\overset{(*)}{\leqslant} \sum_{k=1}^M \mathbb{P}\big[X(k\Delta t) \in b_\eta\big(x_k^{(j)}\big),\ 1 \leqslant k \leqslant P\big]$$
$$\leqslant M \max_{1\leqslant j\leqslant M} \mathbb{P}\left[\bigcap_{k=1}^P \big\{X(k\Delta t) \in b_\eta\big(x_k^{(j)}\big)\big\}\right].$$

All of the above steps should be self-explanatory, let us just notice that the inequality $(*)$ is the union bound. Taking the logarithm of both sides of the derived inequality and using the bound (11), one finds

$$\epsilon\mathbb{P}[X \in D] \leqslant \epsilon \log M + \epsilon \max_{1\leqslant j\leqslant M} \log \mathbb{P}\left[\bigcap_{k=1}^P \big\{X(k\Delta t) \in b_\eta\big(x_k^{(j)}\big)\big\}\right] \leqslant \epsilon \log M + \epsilon \sup_{x\in D} \log \mathbb{P}\left[\bigcap_{k=1}^P \{X(k\Delta t) \in b_\eta(x_k)\}\right]$$
$$\overset{(11)}{\leqslant} \epsilon \log M + R + O(\epsilon P) + \sup_{x\in D}\left[\sum_{p=1}^P \Delta t \lambda_p^T \left(\frac{x_{p-1} - x_p}{\Delta t} - \nu x_{p-1}\right) - \frac{1}{2}\sum_{p=1}^P \Delta t \int_{\mathbb{R}} \frac{dk}{2\pi} \log\det\big(I - 2\lambda_p^T M\tilde{C}(k, x_{p-1})\big)\right].$$

As a result,

$$\limsup_{\epsilon\to 0} \epsilon\mathbb{P}[X \in D] \leqslant \lim_{\epsilon\to 0} R + \sup_{x\in D} \sum_{p=1}^P \Delta t\left[\lambda_p^T\left(\frac{x_{p-1} - x_p}{\Delta t} - \nu x_{p-1}\right) - \frac{1}{2}\int_{\mathbb{R}} \frac{dk}{2\pi} \log\det\big(I - 2\lambda_p^T M\tilde{C}(k, x_{p-1})\big)\right]. \quad \text{(B3)}$$

Finally, notice that the left-hand-side of Eq. (B3) does not depend on $\Delta t$ and $\eta$. Let $\lambda$ and $x$ be a pair of $\mathbb{R}^n$-valued functions on $[0, t]$ such that

$$\lambda(k\Delta t) = \lambda_k,\ x(k\Delta t) = x_k,\ 1 \leqslant k \leqslant P.$$

Setting $\eta = \mu\Delta t$, applying $\lim_{\mu\to 0}\lim_{\Delta t\to 0}$ to both sides of Eq. (B3), and using the property (7) of the error term, one arrives at Eq. (12).

## APPENDIX C: LOWER BOUND ON $\log \mathbb{P}[X \in D]$

For the lower bound, let us take the pair $x, \lambda \in C([0,t],\mathbb{R}^n)$ to be the solution to the Euler-Lagrange equations describing the critical points of Eq. (14) and assume that the solution is unique and smooth. The boundedness of the right-hand side of the equation for $\dot{x}$ means that there is $M_2 > 0$ such that $||\dot{x}(t)||_\infty < M_2$ for all $t \in [0,t]$.

Let us fix $\Delta t > 0$. Using the above bound on $\dot{x}$ and the bound $||f||_\infty < M_0$ discussed in the text above Eq. (A3), it is easy to establish the following: if $X(t) \in b_\eta(x(t))$ at some time $t$, then for all $\tau \in [t, t+\Delta t]$,

$$||X(\tau) - x(\tau)||_\infty \leqslant \eta + (M_0 + M_2)\Delta t. \tag{C1}$$

Let $x_k = x(k\Delta t)$, $\lambda_k = \lambda(k\Delta t)$, and $1 \leqslant k \leqslant P$, where $P = \lfloor \frac{t}{\Delta t} \rfloor$. Choose $\rho > 0$: $B_\rho(x) \subset D$. Then

$$\mathbb{P}[X \in D] \geqslant \mathbb{P}[X \in B_\rho(x)] \geqslant \mathbb{P}[X(k\Delta t) \in b_\eta(x_k), \; 1 \leqslant k \leqslant P], \tag{C2}$$

provided $\eta > 0$ and $\Delta t > 0$ are such that $\eta + (M_0 + M_2)\Delta t < \rho$: given such a choice, the estimate (C1) implies the inclusion of events $\cap_{k=1}^P \{X(k\Delta t) \in b_\eta(x_k)\} \subset \{X \in B_\rho(x)\}$, which leads to the claimed inequality in Eq. (C2).

Following the steps which led to Eq. (A6), one finds

$$X_k - X_{k-1} = \int_{(k-1)\Delta t}^{k\Delta t} f(\tau/\epsilon, x_{k-1}) d\tau + V_k, \tag{C3}$$

where $X_k := X(k\Delta t)$ and $||V_k||_\infty \leqslant nM_1(\eta + \Delta t M_0)\Delta t$. Let

$$F_k^{(\epsilon)} = \int_{(k-1)\Delta t+\epsilon\delta}^{k\Delta t} f(\tau/\epsilon, x_{k-1}) d\tau.$$

By the finite dependency assumption the random variables $(F_k^{(\epsilon)})_{k\geqslant 1}$ are independent. Define $F_{k-1} := F_{k-1}^{(0)}$. Then the right-hand side of Eq. (C3) is equal to $F_{k-1} + V_k$. Notice the following elementary inequality:

$$\mathbb{1}[X + v \in b_\eta(x)] \geqslant \mathbb{1}[X \in b_{\eta - ||v||_\infty}(x)]. \tag{C4}$$

The right-hand side is nonzero provided $||v||_\infty < \eta$. The following estimate is based on Eq. (C2), the independence of $(F_k^{(\epsilon)})_{k\geqslant 1}$, the inequality (C4), and the tower property of conditional probabilities:

$$\begin{aligned}
\mathbb{P}[X \in D] &\geqslant \mathbb{P}[X \in B_\rho(x)] \geqslant \mathbb{E}\left[\prod_{k=1}^P \mathbb{1}[X_k \in b_\eta(x_k)]\right] \\
&= \mathbb{E}\left\{\mathbb{E}\left[\prod_{k=1}^P \mathbb{1}[X_k \in b_\eta(x_k)] \Big| (Y_\tau)_{0\leqslant\tau\leqslant(P-1)\Delta t/\epsilon}\right]\right\} \\
&= \mathbb{E}\left\{\prod_{k=1}^{P-1} \mathbb{1}[X_k \in b_\eta(x_k)] \mathbb{E}\big[\mathbb{1}[X_P \in b_\eta(x_P)] \big| (Y_\tau)_{0\leqslant\tau\leqslant(P-1)\Delta t/\epsilon}\big]\right\} \\
&= \mathbb{E}\left(\prod_{k=1}^{P-1} \mathbb{1}[X_k \in b_\eta(x_k)] \mathbb{E}\big\{\mathbb{1}\big[X_{P-1} + F_P^{(\epsilon)} + v \in b_\eta(x_P)\big] \big| (Y_\tau)_{0\leqslant\tau\leqslant(P-1)\Delta t/\epsilon}\big\}\right) \\
&\geqslant \min_{y_{P-1}\in b_\eta(x_{P-1})} \mathbb{E}\left(\prod_{k=1}^{P-1} \mathbb{1}[X_k \in b_\eta(x_k)] \mathbb{E}\big\{\mathbb{1}\big[y_{P-1} + F_P^{(\epsilon)} + v \in b_\eta(x_P)\big] \big| (Y_\tau)_{0\leqslant\tau\leqslant(P-1)\Delta t/\epsilon}\big\}\right) \\
&= \min_{y_{P-1}\in b_\eta(x_{P-1})} \mathbb{E}\left[\prod_{k=1}^{P-1} \mathbb{1}[X_k \in b_\eta(x_k)]\right] \mathbb{E}\big\{\mathbb{1}\big[y_{P-1} + F_P^{(\epsilon)} + v \in b_\eta(x_P)\big]\big\} \geqslant \prod_{k=1}^P \min_{y_{k-1}\in b_\eta(x_{k-1})} \mathbb{E}\big\{\mathbb{1}\big[y_{k-1} + F_k^{(\epsilon)} + v \in b_\eta(x_k)\big]\big\} \\
&= \prod_{k=1}^P \min_{y_{k-1}\in b_\eta(x_{k-1})} \mathbb{E}\{\mathbb{1}[y_{k-1} + F_k + w \in b_\eta(x_k)]\} \geqslant \prod_{k=1}^P \min_{y_{k-1}\in b_\eta(x_{k-1})} \mathbb{E}\{\mathbb{1}[y_{k-1} + F_k \in b_{\eta-r}(x_k)]\}.
\end{aligned} \tag{C5}$$

Here $v$ and $w$ are a shorthand notation for random errors satisfying deterministic bounds on their norms, $||v||_\infty \leqslant nM_1(\eta + \Delta t M_0)\Delta t + \epsilon M_0\delta$, $||w||_\infty \leqslant nM_1(\eta + \Delta t M_0)\Delta t + 2\epsilon M_0\delta$, and $r = nM_1(\eta + \Delta t M_0)\Delta t + 2\epsilon M_0\delta$.

The following steps are standard in the context of the theory of large deviations: let $\eta' = \eta - r$. Then notice that

$$\begin{aligned}
\mathbb{E}\{\mathbb{1}[F_k \in b_\eta(x_k - y_{k-1})]\} &= \mathbb{E}\big\{e^{-\frac{1}{\epsilon}\lambda_k^T F_k} e^{\frac{1}{\epsilon}\lambda_k^T F_k} \mathbb{1}[F_k \in b_{\eta'}(x_k - y_{k-1})]\big\} \geqslant e^{-\frac{\lambda_k(x_k - y_{k-1}) + n||\lambda_k||_\infty \eta'}{\epsilon}}, \\
\mathbb{E} e^{\frac{1}{\epsilon}\lambda_k^T F_k} \mathbb{1}[F_k \in b_{\eta'}(x_k - y_{k-1})]\} &= e^{-\frac{\lambda_k(x_k - y_{k-1}) + n||\lambda_k||_\infty}{\epsilon}} \mathbb{E}\big[e^{\frac{1}{\epsilon}\lambda_k^T F_k}\big] \mathbb{E}^{(\lambda_k)}\{\mathbb{1}[F_k \in b_{\eta'}(x_k - y_{k-1})]\},
\end{aligned} \tag{C6}$$

where

$$\mathbb{E}^{(\lambda_k)}[\bullet] := \frac{\mathbb{E}[e^{\frac{1}{\epsilon}\lambda_k^T F_k}\bullet]}{\mathbb{E}[e^{\frac{1}{\epsilon}\lambda_k^T F_k}]}$$

is the expectation with respect to the probability measure tilted by the exponential factor $e^{\frac{1}{\epsilon}\lambda_k^T F_k}$. The derivation of Eqs. (C5) and (C6) did not use any assumptions about the sequence $(\lambda_1, \dots, \lambda_P)$. Now let us choose the sequence in such a way that

$$\lim_{\epsilon\to 0} \mathbb{E}^{(\lambda_k)}[F_k] = x_k - y_{k-1}, \ 1 \leqslant k \leqslant P, \tag{C7}$$

which coincides with the discretized version of the Euler-Lagrange equations $\delta S_{\text{eff}}/\delta\lambda(\tau) = 0,\ 0 < \tau < t$, if $y_k = x_k$ for all $k$'s. Equivalently,

$$\frac{\partial}{\partial\lambda_k} \lim_{\epsilon\to 0} \epsilon \log \mathbb{E}[e^{\frac{1}{\epsilon}\lambda_k^T F_k}] = x_k - y_{k-1}, \ 1 \leqslant k \leqslant P. \tag{C8}$$

Recall that an explicit formula derived with the help of Widom's theorem shows that $\lim_{\epsilon\to 0} \epsilon \log \mathbb{E}[e^{\frac{1}{\epsilon}\lambda_k^T F_k}]$ is finite [see Eqs. (8) and (10)]. Calculating the second $\lambda$-derivative of $\log \mathbb{E}[e^{\frac{1}{\epsilon}\lambda^T F_k}]$, one finds that

$$\lim_{\epsilon\to 0} \mathrm{Cov}^{(\lambda_k)}[F_k] = \lim_{\epsilon\to 0} \epsilon^2 \partial_\lambda \otimes \partial_\lambda \log \mathbb{E}\big[e^{\frac{1}{\epsilon}\lambda^T F_k}\big] = 0. \tag{C9}$$

Expressions (C7) and (C9) and Chebyshev's inequality imply that

$$\lim_{\epsilon\to 0} \mathbb{E}^{(\lambda_k)}\{\mathbb{1}[F_k \in b_{\eta'}(x_k - y_{k-1})]\} = 1.$$

Using this observation in Eqs. (C5) and (C6), one finds that

$$\begin{aligned}
&\liminf_{\epsilon\to 0} \epsilon \log \mathbb{P}[X \in D] \\
&\geqslant \sum_{k=1}^{P} \min_{y_{k-1}\in b_\eta(x_{k-1})} \big\{\big[\lambda_k^T(y_{k-1} - x_k) - n||\lambda_k||_\infty \eta''\big] \\
&+ \lim_{\epsilon\to 0} \epsilon \log \mathbb{E}[e^{\frac{1}{\epsilon}\lambda_k^T F_k}]\big\},
\end{aligned} \tag{C10}$$

provided the sequence $(\lambda_k)$ solves Eq. (C7). Here

$$\eta'' = \eta' \mid_{\epsilon=0} = \eta - r \mid_{\epsilon=0} = \eta - nM_1(\eta + \Delta t M_0)\Delta t.$$

As the right-hand side of Eq. (C10) does not depend on $\eta$ and $\Delta t$, one can set $\eta = a(\Delta t)^\mu$, where $a > 0$ and $\mu > 1$, and take the limit $\Delta t \to 0$. In the limit, using that $y_k \in b_\eta(x_k)$, one finds that the system of equations (C8) becomes

$$\frac{\partial}{\partial\lambda(\tau)} \lim_{\Delta t, \epsilon\to 0} \frac{\epsilon}{\Delta t} \log \mathbb{E}\big[e^{\frac{1}{\epsilon}\lambda_k^T F_k}\big] \,|_{k=\frac{\tau}{\Delta t}} = \dot{x}(\tau), \ \tau \in [0, t].$$

Due to Eqs. (8) and (10), the above equation coincides with the Euler-Lagrange equation $\frac{\delta S_{\text{eff}}}{\delta\lambda(\tau)} = 0$, where $S_{\text{eff}}$ is given by Eq. (14). The bound (C10) becomes

$$\begin{aligned}
&\liminf_{\epsilon\to 0} \epsilon \log \mathbb{P}[X \in D] \\
&\geqslant \int_0^t d\tau \bigg( - \lambda^T(\tau)x(\tau) + \lim_{\Delta t\to 0} \lim_{\epsilon\to 0} \frac{\epsilon}{\Delta t} \\
&\times \log \mathbb{E}[e^{\frac{1}{\epsilon}\lambda_k^T F_k}]\big|_{k\Delta t = \tau}\bigg),
\end{aligned} \tag{C11}$$

where the functions $\lambda$ and $x$ solve the Euler-Lagrange equations associated with the effective action functional $S_{\text{eff}}$. Furthermore, the right-hand side of Eq. (C11) coincides with the effective action functional (14). Therefore, by the assumed uniqueness of the solution to the Euler Lagrange equations, the right-hand side of Eq. (C11) must coincide with Eq. (13). The lower bound is derived.

## APPENDIX D: WIDOM'S THEOREM

In Ref. [21], Widom simply formulates the theorem, states that it can be easily verified by taking the continuous limit of the corresponding statement for large Toeplitz matrices, and then moves on to the main topic of the paper: the asymptotic of Fredholm determinants for operators acting on spaces of functions of several variables. Thus, there is a gap in the story, which we partially fill in the present Appendix by deriving Eq. (10). In our proof we use the probabilistic method developed in the original paper by Kac *et al.* [25].

*Theorem.* Let $K : \mathbb{R} \to \mathbb{R}^{N\times N}$ be an $N \times N$-matrix-valued function of one variable. Assume that $K$ is even $[K(t) = K(-t)$, for any $t \in \mathbb{R}]$ and non-negative $[K_{ij}(t) \geqslant 0$ for any $t \in \mathbb{R}$ and $1 \leqslant i, j \leqslant N]$. Assume in addition that

$$\int_{\mathbb{R}} |t| K(t) dt < \infty, \tag{D1}$$

$$\int_{\mathbb{R}} \sum_{k=1}^{N} K_{ki} \leqslant 1, \ \ 1 \leqslant i \leqslant N. \tag{D2}$$

The function $K$ can be regarded as a kernel of an integral operator $\hat{K}$ acting on square-integrable functions from $\mathbb{R}$ to $\mathbb{R}^N$,

$$f \mapsto \hat{K}f(t) = \int_{\mathbb{R}} d\tau K(t-\tau) f(\tau), \ t \in \mathbb{R}. \tag{D3}$$

Then there is $\lambda_{\max} > 0$ such that for any $\lambda : |\lambda| < \lambda_{\max}$ the Fredholm determinant $\mathrm{Det}(I - \lambda\hat{K}_T)$ exists and

$$\log \mathrm{Det}(I - \lambda\hat{K}_T) = T \int_{\mathbb{R}} \frac{dk}{2\pi} \log \det[1 - \lambda\tilde{K}(k)] + O(T^0), \tag{D4}$$

where $\hat{K}_T$ is the restriction of $\hat{K}$ to functions on $[0, T]$ and

$$\tilde{K}(k) = \int_{\mathbb{R}} dx e^{-ikx} K(x), \ k \in \mathbb{R}. \tag{D5}$$

Let us sketch the proof of the theorem using, as we already mentioned, the probabilistic method used in Ref. [25] to prove a continuous version of Szegö's formula for the asymptotics of Toeplitz determinants. For a sufficiently small $|\lambda|$, we can calculate the Fredholm determinant using the trace-log formula,

$$\log \mathrm{Det}(I - \lambda\hat{K}_T) = -\sum_{n=1}^{\infty} \frac{1}{n} \lambda^n \mathrm{Tr}\hat{K}_T^n, \tag{D6}$$

where

$$\begin{aligned}
\mathrm{Tr}\hat{K}_T^n &= \int_{[0,T]^n} dx_1 dx_2 \dots dx_n, \\
&\mathrm{tr}K(x_1 - x_2)K(x_2 - x_3)\dots K(x_n - x_1).
\end{aligned}$$

Using the cyclic property of trace and the fact that the function $K$ is even, we find

$$\frac{d}{dT}\mathrm{Tr}\hat{K}_T^n = n\int_{[0,T]^n} dx_2dx_3\dots dx_n,$$
$$\mathrm{tr}K(-x_2)K(x_2-x_3)\dots K(x_{n-1}-x_n)K(x_n). \quad \text{(D7)}$$

Consider the following discrete-time Markov chain $\{X_n, S_n\}_{n\geqslant 0}$ on the state space $\mathbb{R}\times\{1,2,\dots,N\}$:

(i) $(X_0, S_0)\sim(\delta_0, U_N)$, where $U_N$ is the uniform distribution on $\{1,2,\dots,N\}$;

(ii) at each time step, the transition $(x,i)\to(y,k)$ happens with probability $K_{ki}(y-x)dy$.

Notice that this is a Markov chain with killing, the survival probability when transitioning from state $(x,i)$ is $g_i(x) := \sum_{k=1}^N \int_{\mathbb{R}} K_{ki}(y-x)dy \leqslant 1$. Examining the expression (D7) for the derivative of the trace of the $n$th power of $\hat{K}$, we see that it can be interpreted as the following expectation with respect to the law of the chain $\{X_n, S_n\}_{n\geqslant 0}$:

$$\frac{d}{dT}\mathrm{Tr}\hat{K}_T^n = Nn\mathbb{E}[\mathbb{1}(X_n\in d0)\mathbb{1}(S_n=S_0)\mathbb{1}(\tau=n)], \quad \text{(D8)}$$

where $\tau$ is the first exit time of the chain from the interval $(0,T)\times\{1,2,\dots,N\}$. To derive the above expression we exploited the identity $\mathbb{1}(X_n\in d0)\mathbb{1}(\tau\geqslant n)=\mathbb{1}(X_n\in d0)\mathbb{1}(\tau=n)$. Substituting Eq. (D8) into Eq. (D7) and then into Eq. (D6), we find that

$$\frac{d}{dT}\log\mathrm{Det}(I-\lambda\hat{K}_T) = -N\mathbb{E}[\lambda^\tau\mathbb{1}(X_\tau\in d0)\mathbb{1}(S_\tau=S_0)]$$
$$= -N\mathbb{E}[\lambda^{\tau_0}\mathbb{1}(X_{\tau_0}\in d0)$$
$$\times\ \mathbb{1}(M_{\tau_0}<T)\mathbb{1}(S_{\tau_0}=S_0)],$$

where $\tau_0$ is the first exit time from $(0,\infty)\times\{1,2,\dots,N\}$, $M_{\tau_0}=\max_{1\leqslant n<\tau_0}(X_n)$. As $\log\det(I-\lambda\hat{K}_0)=0$, we can integrate the last expression to find

$$\log\mathrm{Det}(I-\lambda\hat{K}_T) = -N\mathbb{E}[\lambda^{\tau_0}\mathbb{1}(X_{\tau_0}\in d0)(T-M_{\tau_0})_+$$
$$\times\ \mathbb{1}(S_{\tau_0}=S_0)],$$

where $(x)_+ := \max(x,0)$. Noticing that $T-(T-M)_+ = \min(T,M)$, we can rearrange the above expression as follows:

$$\log\mathrm{Det}(I-\hat{K}_T) = -NT\mathbb{E}[\lambda^{\tau_0}\mathbb{1}(X_{\tau_0}\in d0)\mathbb{1}(S_{\tau_0}=S_0)]$$
$$+N\mathbb{E}[\lambda^{\tau_0}\mathbb{1}(X_{\tau_0}\in d0)\min(T,M_{\tau_0})$$
$$\times\ \mathbb{1}(S_{\tau_0}=S_0)].$$

This is an exact expression for the Fredholm determinant as an expectation with respect to the law of the Markov chain we defined. In many cases it allows for an efficient computation of the large-$T$ expansion of the Fredholm determinant using purely probabilistic methods. For us it is sufficient to check that $\lim_{T\to\infty}\min(T,M_{\tau_0})=M_{\tau_0}$, which implies that

$$\log\mathrm{Det}(I-\lambda\hat{K}_T) = -NT\mathbb{E}[\lambda^{\tau_0}\mathbb{1}(X_{\tau_0}\in d0)\mathbb{1}(S_{\tau_0}=S_0)]$$
$$+O(T^0). \quad \text{(D9)}$$

To calculate the expectation entering the leading term we use the following combinatorial lemma (see, e.g., Ref. [26]): Let $(0, R_1, R_1+R_2, \dots, R_1+R_2+\dots+R_{n-1}, 0)$ be the first $n$ $\mathbb{R}$-projections of the states of the chain with $\tau_0=n$. Then

$$\sum_{p=0}^{n-1}\prod_{k=1}^{n-1}\mathbb{1}(R_{1+p}+R_{2+p}+\dots+R_{k+p}>0)$$
$$=1 \text{ a.s.},\ 0\leqslant p\leqslant n-1. \quad \text{(D10)}$$

The addition of subscripts in the above formula should be understood modulo $n$. The above statement is very general and relies only on the absence of atoms in the transition probabilities $K(y-x)dy$.

In this case, for any sequence $(0, R_1, R_1+R_2, \dots, R_1+R_2+\dots+R_{n-1}, 0)$, its graph will almost surely have a unique global minimum, so there will be a unique cyclic permutation $(0, R_{1+p}, R_{1+p}+R_{2+p}, \dots, R_{1+p}+R_{2+p}+\dots+R_{n-1+p}, 0)$, whose graph will stay positive between times 1 and $n-1$. Then

$$N\mathbb{E}[\lambda^{\tau_0}\mathbb{1}(X_{\tau_0}\in d0)\mathbb{1}(S_{\tau_0}=S_0)] = N\sum_{n=1}^\infty\lambda^n\mathbb{E}[\mathbb{1}(X_{\tau_0}\in d0)\mathbb{1}(S_{\tau_0}=S_0)\mathbb{1}(\tau_0=n)]$$
$$=\sum_{n=1}^\infty\lambda^n\int_{\mathbb{R}^n}dr_1\dots dr_n\mathrm{tr}[K(r_1)\dots K(r_n)]\delta(r_1+\dots+r_n)\prod_{k=1}^{n-1}\mathbb{1}(r_1+\dots+r_k>0)$$
$$=\sum_{n=1}^\infty\frac{\lambda^n}{n}\int_{\mathbb{R}^n}dr_1\dots dr_n\mathrm{tr}[K(r_1)\dots K(r_n)]\delta(r_1+\dots+r_n)\sum_{p=0}^{n-1}\prod_{k=1}^{n-1}\mathbb{1}(r_{1+p}+\dots+r_{k+p}>0)$$
$$=\sum_{n=1}^\infty\frac{\lambda^n}{n}\int_{\mathbb{R}^n}dr_1\dots dr_n\mathrm{tr}[K(r_1)\dots K(r_n)]\delta(r_1+\dots+r_n)$$
$$=\sum_{n=1}^\infty\frac{\lambda^n}{n}\int_{\mathbb{R}}\frac{dk}{2\pi}\int_{\mathbb{R}^n}dr_1\dots dr_n e^{-ik(r_1+\dots+r_n)}\mathrm{tr}[K(r_1)\dots K(r_n)]$$
$$=\sum_{n=1}^\infty\frac{\lambda^n}{n}\int_{\mathbb{R}}\frac{dk}{2\pi}\mathrm{tr}[\tilde{K}(k_1)\dots\tilde{K}(k_n)] = -\int_{\mathbb{R}}\frac{dk}{2\pi}\log\det[I-\lambda\tilde{K}(k)]. \quad \text{(D11)}$$

The third inequality is the symmetrization of the integrand with respect to all cycling permutations, and the fourth inequality is due to the combinatorial lemma (D10). Substituting Eq. (D11) into Eq. (D9), we arrive at the statement (D4) of Widom's theorem.

**Remarks**

(i) In Ref. [21], Widom presents a stronger version of the above statement which characterizes the $O(T^0)$ term fully. For the current paper we only need the leading term.

(ii) The actual statement of Widom's theorem does not require the positivity of the kernel. In fact, all steps of the proof presented below go through for signed kernels as well, but the probabilistic intuition guiding these steps is lost. See also Ref. [25] for similar remarks about the original proof of Szegő's theorem by M. Kac *et al.*

(iii) It is possible to give an alternative derivation of Eq. (10) based on the resummation of the cumulant expansion for the expectation of a quadratic function of a Gaussian process. The downside of such a derivation is difficulty in controlling the subleading terms.


[1] L. Bertini, A. De Sole, D. Gabrielli, G. Jona-Lasinio, and C. Landim, Macroscopic fluctuation theory, Rev. Mod. Phys. **87**, 593 (2015).

[2] K. Saito and A. Dhar, Generating function formula of heat transfer in harmonic networks, Phys. Rev. E **83**, 041121 (2011).

[3] E. M. Lifshitz and L. P. Pitaevskii, *Physical Kinetics*, Course of Theoretical Physics (Pergamon, Oxford, 1981).

[4] D. Nicholson, *Introduction to Plasma Theory* (Wiley, New York, 1983).

[5] J. Binney and S. Tremaine, *Galactic Dynamics* (Princeton University, Princeton, NJ, 1987), p. 747.

[6] F. Bouchet, C. Nardini, and T. Tangarife, Kinetic theory of jet dynamics in the stochastic barotropic and 2D Navier-Stokes equations, J. Stat. Phys. **153**, 572 (2013).

[7] S. Nazarenko, *Wave Turbulence*, Lecture Notes in Physics, Vol. 825 (Springer, Berlin, 2011).

[8] L. D. Landau and E. M. Lifshitz, *Mechanics*, Course of Theoretical Physics, Vol. 1 (Butterworth-Heinemann, Oxford, 1976).

[9] G. Pavliotis and A. Stuart, *Multiscale Methods: Averaging and Homogenization* (Springer Science & Business Media, New York, 2008).

[10] M. I. Freidlin, J. Szücs, and A. D. Wentzell, *Random Perturbations of Dynamical Systems* (Springer, Berlin, 2012).

[11] F. Bouchet and E. Simonnet, Random Changes of Flow Topology in Two-Dimensional and Geophysical Turbulence, Phys. Rev. Lett. **102**, 094504 (2009).

[12] F. Bouchet, J. Rolland, and E. Simonnet, Rare Event Algorithm Links Transitions in Turbulent Flows with Activated Nucleations, Phys. Rev. Lett. **122**, 074502 (2019).

[13] M. I. Freidlin, The averaging principle and theorems on large deviations, Russ. Math. Surv. **33**, 117 (1978).

[14] A. Y. Veretennikov, On large deviations for SDEs with small diffusion and averaging, Stochastic Processes Their Appl. **89**, 69 (2000).

[15] Y. Kifer, Averaging in dynamical systems and large deviations, Inventiones Math. **110**, 337 (1992).

[16] Y. Kifer, Averaging principle for fully coupled dynamical systems and large deviations, Ergodic Theory Dyn. Syst. **24**, 847 (2004).

[17] F. Bouchet, T. Grafke, T. Tangarife, and E. Vanden-Eijnden, Large deviations in fast–slow systems, J. Stat. Phys. **162**, 793 (2016).

[18] F. Bouchet, J. Laurie, and O. Zaboronski, Langevin dynamics, large deviations and instantons for the quasi-geostrophic model and two-dimensional Euler equations, J. Stat. Phys. **156**, 1066 (2014).

[19] O. Feliachi and F. Bouchet, Dynamical large deviations for homogeneous systems with long range interactions and the Balescu–Guernsey–Lenard equation, J. Stat. Phys. **186**, 22 (2022).

[20] P. C. Martin, E. Siggia, and H. Rose, Statistical dynamics of classical systems, Phys. Rev. A **8**, 423 (1973).

[21] H. Widom, Szegö's limit theorem: The higher-dimensional matrix case, J. Funct. Anal. **39**, 182 (1980).

[22] R. Durrett, *Probability: Theory and Examples*, Cambridge Series in Statistical and Probabilistic Mathematics, Vol. 49 (Cambridge University, Cambridge, England, 2019).

[23] It is the absence of error estimates associated with this approximation which makes our present discussion nonrigorous.

[24] A mathematician would say that $D$ is a totally bounded subset of the complete space $C([0, T], \mathbb{R}^n, ||\cdot||_\infty)$ and is therefore compact.

[25] M. Kac *et al.*, Toeplitz matrices, translation kernels and a related problem in probability theory, Duke Math. J. **21**, 501 (1954).

[26] W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2 (Wiley & Sons, New York, 2008).