


Ritz method for transition paths and quasipotentials of rare diffusive events

Lukas Kikuchi ^{1,*}, Rajesh Singh,¹ M. E. Cates,¹ and R. Adhikari^{1,2}

¹*DAMTP, Centre for Mathematical Sciences, University of Cambridge, Cambridge CB3 0WA, United Kingdom*

²*The Institute of Mathematical Sciences-HBNI, CIT Campus, Chennai 600113, India*



(Received 29 January 2020; accepted 9 July 2020; published 6 August 2020)

The probability of trajectories of weakly diffusive processes to remain in the tubular neighborhood of a smooth path is given by the Freidlin-Wentzell-Graham theory of large deviations. The most probable path between two states (the instanton) and the leading term in the logarithm of the process transition density (the quasipotential) are obtained from the minimum of the Freidlin-Wentzell action functional. Here we present a Ritz method that searches for the minimum in a space of paths constructed from a global basis of Chebyshev polynomials. The action is reduced, thereby, to a multivariate function of the basis coefficients, whose minimum can be found by nonlinear optimization. For minimization regardless of path duration, this procedure is most effective when applied to a reparametrization-invariant “on-shell” action, which is obtained by exploiting a Noether symmetry and is a generalization of the scalar work [Olender and Elber, *J. Mol. Struct. THEOCHEM* **398**, 63 (1997)] for gradient dynamics and the geometric action [Heyman and Vanden-Eijnden (2008)] for nongradient dynamics. Our approach provides an alternative to chain-of-states methods for minimum energy paths and saddle points of complex energy landscapes and to Hamilton-Jacobi methods for the stationary quasipotential of circulatory fields. We demonstrate spectral convergence for three benchmark problems involving the Müller-Brown potential, the Maier-Stein force field, and the Egger weather model.

DOI: [10.1103/PhysRevResearch.2.033208](https://doi.org/10.1103/PhysRevResearch.2.033208)

I. INTRODUCTION

The theory of Freidlin and Wentzell [1] gives asymptotic probability estimates of rare events in dynamical systems perturbed by small noise [2–5]. Specifically, Freidlin-Wentzell theory yields estimates of the stationary distributions and mean first-passage times. Both these quantities are determined, in turn, by the asymptotic estimate of the probability of a stochastic trajectory to not deviate from a smooth path by more than a given amount in a given interval of time. The key result of Freidlin and Wentzell is that the limiting form of this probability, for small noise and small deviations, is given by a nonnegative functional of the smooth path. This functional is the Freidlin-Wentzell action and its minimum, for fixed initial and terminal states, determines both the stationary distributions and first-passage times. The smooth path minimizing the action is often called the Freidlin-Wentzell instanton. The theory is applicable to dynamical systems of both gradient and nongradient character and can so be used to study a wide variety of equilibrium and nonequilibrium systems modelled by Itô diffusions [6–15].

Determining the minimum of the Freidlin-Wentzell action is a problem in the calculus of variations. The Euler-Lagrange equations provide the necessary conditions for extrema of

variational problems and, unsurprisingly, have been the basis of the large literature devoted to the numerical computation of Freidlin-Wentzell instantons [6,16–18]. There exists, however, an alternative “direct” route for the solution of variational problems in which the functional is reduced, through finite-dimensional parametrizations of paths, to a multivariate function and then extremised by appropriate multivariate optimization methods [19,20]. To the best of our knowledge, the first use of the direct method for the Freidlin-Wentzell action, discretized by finite differences, appears in the work of Weinan, Ren, and Vanden-Eijnden [21].

Here we combine the direct method with a Ritz discretization [19,20,22] to minimize the Freidlin-Wentzell action. We analyze paths in a spectral basis of Chebyshev polynomials and use spectral quadrature to express the action as a multivariate function of the basis coefficients. Nonlinear optimization is used to obtain coefficients that give the least action from which the instanton is synthesized in the spectral basis. For minimization over paths regardless of their duration, this procedure is especially effective when applied to a reparametrization-invariant on-shell form of the action that follows from the time-translational invariance of the Lagrangian. This generalizes the scalar work functional of Olender and Elber (for gradient dynamics) and the geometric action of Heyman and Vanden-Eijnden [23] (for nongradient dynamics). Our method is efficient enough to robustly sample the logarithm of the asymptotic estimate of the stationary distribution, i.e., the quasipotential, avoiding the alternative, but numerically delicate, route of solving the Hamilton-Jacobi equation [24–26]. Our method is simple to use, converges rapidly, and is applicable to both equilibrium and nonequilibrium problems. Its implementation is

*ltk26@cam.ac.uk

freely available on GitHub as the open-source Python library PyRitz.¹

The remainder of this paper is organized as follows. In the next section, we recall key results of Freidlin-Wentzell theory from dual perspectives of Itô stochastic differential equations and the corresponding Fokker-Planck equations. In Sec. III we present the derivation of the on-shell form of the Freidlin-Wentzell action in a manner reminiscent of the Routh reduction procedure in classical mechanics and explain its relation to the scalar work and the geometric action. In Sec. IV we describe the direct method for the minimization of functionals, the Chebyshev spectral basis in which we construct smooth paths, the spectral quadrature rule we use to evaluate the action, and the multivariate nonlinear optimization methods we employ to find the minimum. In Sec. V we apply the direct method to three well-known diffusion processes and demonstrate convergence in each case.

A particular achievement of our approach is its relatively facile ability to calculate quasipotentials. This can be done with a sufficiently high density of sample points to construct effectively continuous maps of the quasipotential, which we do here for the same set of benchmark problems. We conclude with a discussion on extending the method to degenerate diffusion processes, systems with inertia and to the stochastic dynamics of fields.

II. LARGE-DEVIATION THEORY

We consider the autonomous dynamics of a d -dimensional coordinate $X = (X^1, \dots, X^d)$ in \mathbb{R}^d perturbed by configuration-dependent noise of intensity $\sqrt{\varepsilon}$ described by the Itô diffusion equation

$$dX^\mu = a^\mu(X)dt + \sqrt{\varepsilon}\sigma_\nu^\mu(X)dW^\nu, \quad (1)$$

governing the stochastic trajectory $X(t)$, where $a^\mu(X)$ is the drift vector, $\sqrt{\varepsilon}\sigma_\nu^\mu(X)$ is the volatility, $W^\nu(t)$ is a d -dimensional Wiener process, and repeated indices are summed over. The transition probability density of the process, $P_{1|1}(x, t|x_0) = P(X(t) = x|X(0) = x_0)$, obeys the Fokker-Planck equation $\partial_t P(x|x_0) = \mathcal{L}P(x|x_0)$, where the Fokker-Planck operator is

$$\mathcal{L}(x) = -\frac{\partial}{\partial x^\mu} a^\mu(x) + \frac{\varepsilon}{2} \frac{\partial^2}{\partial x^\mu \partial x^\nu} b^{\mu\nu}(x) \quad (2)$$

and $b^{\mu\nu}(x) = \sigma_\lambda^\mu(x)\sigma_\sigma^\nu(x)\delta^{\lambda\sigma}$ is the diffusion tensor. We assume it to be nondegenerate, positive definite, and invertible. The inverse, $b_{\mu\nu}(x)$, induces a Riemannian structure in \mathbb{R}^d with a norm $|x|_b = \sqrt{b_{\mu\nu}x^\mu x^\nu}$ that is distinct from the Euclidean norm $|x| = \sqrt{(x^1)^2 + \dots + (x^d)^2}$. We use the subscript b to indicate this second “diffusion” norm. The stationary density satisfies the time-independent Fokker-Planck equation $\mathcal{L}P_1(x) = 0$ and, when it exists, is reached asymptotically in time for arbitrary initial distributions, $\lim_{t \rightarrow \infty} P_{1|1}(x, t|x_0) = P_1(x)$. The stationary distribution is unique for ergodic systems Pavliotis [27].

Associated with the Itô process is the Freidlin-Wentzell “action” functional [1,28,29],

$$S[x(t)] = \frac{1}{2} \int_0^T |\dot{x} - a(x)|_b^2 dt, \quad (3)$$

which gives an asymptotic estimate for the logarithm of the probability of trajectories $X(t)$ to remain in the tubular neighborhood of a smooth path $x(t)$ over the duration $0 \leq t \leq T$. We write this as

$$P_{\text{tube}}[x(t)] \asymp \exp \left\{ -\frac{1}{\varepsilon} S[x(t)] \right\}, \quad (4)$$

which, in terms of limits, means

$$S[x(t)] = \lim_{\delta \rightarrow 0} \lim_{\varepsilon \rightarrow 0} -\varepsilon \ln P \left[\sup_{0 \leq t \leq T} |X(t) - x(t)|_b < \delta \right].$$

The limits must be taken in the order above as they do not commute. Equation (4) is a large-deviation principle for trajectories of Itô processes, due to Wentzell and Freidlin and Graham [30].

For reasons described below, it is of interest to obtain the mode of the tube probability over the set of continuous paths

$$\gamma_T = \{x(t) | x(0) = x_1, x(T) = x_2, 0 \leq t \leq T\},$$

which have fixed termini x_1 and x_2 and are of duration T . This is equivalent to the variational problem of minimizing the Freidlin-Wentzell action. The minimum value of the action,

$$V_T(x_2|x_1) = \min_{\gamma_T} S[x(t)], \quad (5)$$

is called the quasipotential. The path attaining the minimum,

$$x_T^*(t) = \arg \min_{\gamma_T} S[x(t)], \quad (6)$$

is called the instanton. We emphasize that this path describes the smooth centerline of the tube of maximum probability and not a nondifferentiable trajectory of the diffusion process. It is the most probable dynamical path connecting two points in configuration space.

The instanton and the quasipotential are central objects in Freidlin-Wentzell-Graham theory and relate to the eikonal approximation of the Fokker-Planck equation [31]. Assuming the following form, known as the JWKB approximation [9], of the transition density,

$$P_{1|1}(x, t|x_0) \sim \exp \left[\frac{1}{\varepsilon} \sum_{n=0}^{\infty} \varepsilon^n \phi_n(x, x_0; t) \right],$$

with prefactors suppressed, substituting in the Fokker-Planck equation and matching terms gives a Hamilton-Jacobi equation for the lowest-order contribution,

$$\partial_t \phi_0 + \frac{1}{2} b^{\mu\nu} \partial_\mu \partial_\nu \phi_0 + a^\mu \partial_\mu \phi_0 = 0. \quad (7)$$

This corresponds to the Hamiltonian system

$$\begin{aligned} H(x, p) &= \frac{1}{2} b^{\mu\nu} p_\mu p_\nu + a^\mu p_\mu \\ \dot{x}^\mu &= + \frac{\partial H}{\partial p_\mu} = b^{\mu\nu} p_\nu + a^\mu \\ \dot{p}_\mu &= - \frac{\partial H}{\partial x^\mu} = - \frac{\partial b^{\nu\lambda}}{\partial x^\mu} p_\nu p_\lambda - \frac{\partial a^\nu}{\partial x^\mu} p_\nu, \end{aligned} \quad (8)$$

¹<https://github.com/lukastk/PyRitz>.

whose solutions define an equivalent variational problem of extremizing an action with the Lagrangian

$$\begin{aligned} L(x, \dot{x}) &= p_\mu \dot{x}^\mu - H(x, p) \\ &= \frac{1}{2}(\dot{x}^\mu - a^\mu) b_{\mu\nu} (\dot{x}^\nu - a^\nu). \\ &= \frac{1}{2} |\dot{x} - a(x)|_b^2. \end{aligned} \tag{9}$$

Thus, the rays of the Hamilton-Jacobi equation that determine the lowest-order contribution to the eikonal are local maxima of the tube probability, or in other words, $\phi_0(x, x_0; T) = V_T(x|x_0)$. The large-deviation principle of Freidlin and Wentzell and the theory of the nonequilibrium potential of Graham [28,29] thus appear as elegant reformulations of the JWKB approximation [31].

The correspondence with the JWKB approximation yields the asymptotic form of the transition density,

$$P_{1|1}(x, T|x_0) \asymp \exp\left[-\frac{1}{\varepsilon} V_T(x|x_0)\right], \tag{10}$$

and, in the $T \rightarrow \infty$ limit of the above, the asymptotic form of the stationary distribution,

$$P_1(x) \asymp \lim_{T \rightarrow \infty} \exp\left[-\frac{1}{\varepsilon} V_T(x|x_0)\right], \tag{11}$$

where x and x_0 belong to the same basin of attraction of an attractor \mathcal{A} . As is indicated by Eq. (11), it can be shown that this limit is independent of the initial coordinate,

$$\lim_{T \rightarrow \infty} V_T(x|x_0) = V_\infty^{\mathcal{A}}(x), \tag{12}$$

where $V_\infty^{\mathcal{A}}$ is equal, to within a constant, to the stationary quasipotential $V_\infty(x)$ in the basin of attraction of \mathcal{A} . For a system with multiple attractors \mathcal{A}_i , the global quasipotential is

$$V_\infty(x) = \min_i [V_\infty^{\mathcal{A}_i}(x) + C^{\mathcal{A}_i}], \tag{13}$$

where $C^{\mathcal{A}_i}$ is an additive constant. The constants are fixed by requiring

$$V_\infty^{\mathcal{A}_i}[x_s^{(i,j)}] + C^{\mathcal{A}_i} = V_\infty^{\mathcal{A}_j}[x_s^{(i,j)}] + C^{\mathcal{A}_j} \tag{14}$$

for attractors \mathcal{A}_i and \mathcal{A}_j with adjacent basins of attraction, where $x_s^{(i,j)}$ is the saddle with the lowest value on the separatrix between the basins [29]. The stationary quasipotential determines the mean persistence time of a trajectory in a basin of attraction which generalizes the Arrhenius law to systems out of equilibrium.

The $T \rightarrow \infty$ limit involved in the definition of the stationary quasipotential presents considerable numerical difficulties in the minimization of the Freidlin-Wentzell action. A more numerically amenable route to determining the stationary quasipotential is by the minimization of the action over paths that start at an attractor and end at a point in its basin, regardless of the duration. We show in the next section that the solution of this second variational problem does, indeed, yield the stationary quasipotential and derive an alternative form of the Freidlin-Wentzell action that is adapted to computing instantons regardless of their duration.

III. ON-SHELL ACTION

We consider the variational problem of minimizing the Freidlin-Wentzell action over paths with fixed termini but of arbitrary duration,

$$\min_T \min_{\gamma_T} S[x(t)] = \min_T \min_{\gamma_T} \int_0^T L(x, \dot{x}) dt, \tag{15}$$

where both the initial and final points are in the basin of the attraction \mathcal{A} and the Freidlin-Wentzell Lagrangian following from Eq. (9) is

$$L(x, \dot{x}) = \frac{1}{2} b_{\mu\nu} \dot{x}^\mu \dot{x}^\nu - b_{\mu\nu} a^\mu \dot{x}^\nu + \frac{1}{2} b_{\mu\nu} a^\mu a^\nu. \tag{16}$$

This variational problem can be solved by introducing a parametrization u for both the coordinate and time,

$$x = x(u), \quad x' = dx/du; \quad t = t(u), \quad t' = dt/du,$$

that allows the shape of the path,

$$\sigma = \{x(u) \mid x(u_1) = x_1, x(u_2) = x_2, u_1 \leq u \leq u_2\},$$

to be varied independently of its duration,

$$T = \int_{u_1}^{u_2} t' du.$$

Coordinates x and time t are dependent variables in the reparametrized action,

$$S[x(u)] = \int_{u_1}^{u_2} L\left(x, \frac{x'}{t'}\right) t' du, \tag{17}$$

in which the time dependence appears only through the derivative t' . Therefore, t is a cyclic (or ignorable) coordinate and Noether's theorem implies that the corresponding conjugate momentum is conserved [32]:

$$-\frac{\partial(Lt')}{\partial t'} = \frac{1}{2} b_{\mu\nu} \frac{x'^\mu x'^\nu}{(t')^2} - \frac{1}{2} b_{\mu\nu} a^\mu a^\nu = E. \tag{18}$$

This defines submanifolds of the dynamics labeled by the "energy" E which we shall call shells. The bound $2E + |a|_b^2 \geq 0$ for the energy follows immediately from the positive-definiteness of the diffusion tensor.

Solving the first integral for t' gives

$$t' = \frac{dt}{du} = \frac{|x'|_b}{\sqrt{2E + |a|_b^2}}, \tag{19}$$

from which the duration of the path is obtained to be

$$T_E = \int_{u_1}^{u_2} \frac{|x'|_b}{\sqrt{2E + |a|_b^2}} du. \tag{20}$$

This shows that paths γ_{T_E} (of duration T_E) are equivalent to shapes σ_E (of energy E), where the latter is the restriction of shapes in σ to the shell of constant energy. Then, minimization over paths γ_T regardless of their duration is equivalent to minimization over shapes σ_E regardless of their energy, or

$$\min_T \min_{\gamma_T} S[x(t)] = \min_E \min_{\sigma_E} S[x(u)].$$

The action for shapes restricted to σ_E is obtained by eliminating t' between Eqs. (17) and (19). This gives the "on-shell"

form of the Freidlin-Wentzell action,

$$S_E[x(u)] = \int_{u_1}^{u_2} \left[\frac{E + |a|_b^2}{\sqrt{2E + |a|_b^2}} |x'|_b - b^{\mu\nu} a_\mu x'_\nu \right] du,$$

which is a functional of the shape $x(u)$, a function of the energy E , and allows for independent variations of both. It is straightforward to see that the integrand and, therefore, the action is minimized when $E = 0$. Therefore, most probable paths, regardless of their duration, are obtained by minimizing

$$S_0[x(u)] = \int_{u_0}^{u_1} (|a|_b |x'|_b - b^{\mu\nu} a_\mu x'_\nu) du \quad (21)$$

over shapes restricted to the zero-energy shell. The duration on the zero-energy shell,

$$T_0 = \int_{u_1}^{u_2} \frac{|x'|_b}{|a|_b} du, \quad (22)$$

shows that paths that leave, cross, or terminate at points of vanishing drift, $a^\mu(x) = 0$, are necessarily of infinite duration. The corresponding shapes σ_0^A can then be taken to start at a fixed point and end at another point x in the basin of attraction. The quasipotential is determined by a minimization over such shapes σ_0^A ,

$$V_\infty^A(x) = \min_{\sigma_0^A} S_0[x(u)], \quad (23)$$

and the shape attaining the minimum,

$$x_\infty^*(u) = \arg \min_{\sigma_0^A} S_0[x(u)], \quad (24)$$

is the stationary instanton. The time on the instanton path can be obtained by integrating $t' = |x'|_b/|a|_b$. The utility of the on-shell form of the action is that it provides the shape of the path independently of its duration. The latitude of obtaining the shape from a parametrization over a finite interval, even for paths of infinite duration, is extremely useful in numerical work.

The on-shell action is related to, but distinct from, the Jacobi action in mechanics [33,34], which, following a Routh reduction [32], would in this case be

$$\hat{S}[x(u)] = \int_{u_1}^{u_2} \left[\sqrt{2E + |a|_b^2} |x'|_b - a^\mu x'_\mu \right] du.$$

Though both the on-shell and Jacobi action agree on the zero-energy shell, only the former supports the interpretation as least action for nonzero energies. Furthermore, variations of the on-shell action have to respect the on-shell condition Eq. (19) (in other words, the solutions of its Euler-Lagrange equations does not coincide with its extrema). On the other hand, the Jacobi action can be varied using the standard Euler-Lagrange approach.

For gradient dynamics, that is $a^\mu = b^{\mu\nu} \partial U / \partial x^\nu$, the on-shell action generalizes the “scalar work” functional of Olander and Elber [35] to nonzero energies and configuration-dependent diffusion tensors. For nongradient dynamics, where the drift cannot be so expressed, the on-shell action generalizes the geometric action of Heyman and Vanden-Eijnden [17] to nonzero energies. The nonzero energy shell $|\dot{x}|_b^2 = |a|_b^2 + E$ admits the most general path consistent with time-translation

invariance, in contrast to the zero-energy shell where the magnitude of the velocity is always equal to that of the drift, $|\dot{x}|_b^2 = |a|_b^2$. Such general paths determine the quasipotential and the asymptotic form of the transition density for finite times and will be examined in detail in future work. Accordingly, we set $E = 0$ below. The derivation of the on-shell action only requires time-translation invariance of the Lagrangian and not, as in Refs. [17,35], their positive definiteness. Thus, it can be applied to stochastic actions whose Lagrangians are not necessarily positive definite, as for example the Onsager-Machlup action [36,37]. We now describe the Ritz method by which we minimize actions.

IV. RITZ METHOD

The direct method in the calculus of variations consists of constructing a sequence of extremization problems for a function of a finite number of variables that, in the passage to the limit of an infinite number of variables, yields the solution to the variational problem. The two main families of direct methods are finite differences and Ritz methods [19,20,22,38]. In the latter, the solution of the variational problem is sought in a sequence of functions

$$\varphi_1(t), \varphi_2(t), \dots, \varphi_n(t), \dots,$$

each of which satisfies end point conditions. The path is expressed as a linear combination of these functions

$$x_n(t) = \alpha_1 \varphi_1(t) + \dots + \alpha_n \varphi_n(t), \quad (25)$$

which transforms the action from a functional of the path into a function of the expansion coefficients,

$$\begin{aligned} S(\alpha_1, \dots, \alpha_n) &= \int_0^T L(x_n, \dot{x}_n) dt \\ &= \int_0^T L\left(\sum_{i=1}^n \alpha_i \varphi_i, \sum_{i=1}^n \alpha_i \dot{\varphi}_i\right) dt. \end{aligned} \quad (26)$$

Action minimization now becomes a search for a set of coefficients, α_i^* such that $S(\alpha_1^*, \dots, \alpha_n^*) < S(\alpha_1, \dots, \alpha_n)$. The necessary condition for this is the vanishing of the gradient,

$$\frac{\partial S}{\partial \alpha_i} = 0 \quad (i = 1, 2, \dots, n), \quad (27)$$

which is the Ritz system of nonlinear equations. Coefficients satisfying these conditions can be obtained by nonlinear optimization. The n th approximation to the minimum action path, $x_T^*(t)$, and the minimum value of the action, $S[x_T^*(t)]$, are obtained from these values of the coefficients. It is generally the case that this sequence of approximations converges to the minimum of the variational problem as $n \rightarrow \infty$ [19,20].

The method, then, has three parts: first, the choice of basis functions $\varphi_i(t)$; second, the quadrature rule that integrates the Lagrangian to obtain the action as a function of the expansion coefficients; and, third, the optimization that yields the coefficients at the minimum. Since each part is only loosely dependent on the others, Ritz methods come in many varieties [39]. Our choices are centered around Chebyshev polynomials as described below. Approximation by Chebyshev polynomials and their optimality for the purpose are described in Refs. [40–42].

A. Basis functions

We consider a path $x(u)$ that is a Lipschitz continuous function of the parameter u in the interval $[-1, 1]$. Then, it has an absolutely and uniformly convergent Chebyshev expansion,

$$x(u) = \sum_{k=0}^{\infty} a_k T_k(u), \quad a_k = \frac{2}{\pi} \int_{-1}^1 \frac{x(u) T_k(u)}{\sqrt{1-u^2}} du,$$

where $T_k(u)$ are Chebyshev polynomials of the first kind and the integral must be halved for $k = 0$. A suitable sequence of paths can be constructed from the first n terms of this infinite series. However, it is computationally more convenient, for reasons that will be clear below, to construct the sequence from n th degree polynomials that interpolate the path at the $n + 1$ Chebyshev points,

$$u_j = -\cos(j\pi/n), \quad (j = 0, 1, \dots, n). \quad (28)$$

The n th degree interpolant can be expressed in standard form as a sum of Lagrange cardinal polynomials $\ell_j(u)$ or as a linear combination of Chebyshev polynomials,

$$x_n(u) = \sum_{j=0}^n \alpha_j \ell_j(u) = \sum_{k=0}^n c_k T_k(u). \quad (29)$$

The coefficients c_k are aliased versions of the coefficients a_k . Since the cardinal polynomials have the property

$$\ell_j(u_k) = \begin{cases} 1, & j = k, \\ 0, & \text{otherwise,} \end{cases} \quad (j, k = 0, \dots, n),$$

$x_n(u_k) = \alpha_k$, that is, the expansion coefficients α_k are path coordinates at the Chebyshev points. Expressing the entire path in terms its discrete coordinates has the advantage that end point conditions can be imposed by setting

$$\alpha_0 = x(u_0) = x_0, \quad \alpha_n = x(u_n) = x_1. \quad (30)$$

Admissible paths of degree n are, then, parametrized by the $n - 1$ independent coefficients $\alpha_1, \dots, \alpha_{n-1}$. In contrast, imposing end point conditions in series form leads to a numerically inconvenient linear dependence between the coefficients c_k . The derivative of the path is a polynomial of degree $n - 1$ that can be expressed in terms of the interpolant as

$$x'_n(u) = \sum_{j=0}^n \alpha_j \ell'_j(u) = \sum_{j=0}^n \beta_j \ell_j(u) \quad (31)$$

with the two sets of expansion coefficients related by the Chebyshev spectral differentiation matrix

$$\beta_j = D_{jk} \alpha_k, \quad D_{jk} = \ell'_k(u_j). \quad (32)$$

We use the barycentric form of the Lagrange polynomials [43]

$$\ell_j(u) = \frac{w_j}{u - u_j} \bigg/ \sum_{k=0}^n \frac{w_k}{u - u_k}. \quad (33)$$

with weights [44]

$$w_j = \begin{cases} \frac{1}{2}, & j = 0 \\ (-1)^j, & j = 1, \dots, n - 1 \\ \frac{1}{2} \cdot (-1)^n, & j = n. \end{cases}$$

This form is both numerically stable and, costing no more than $O(n)$ operations, efficient to evaluate [45].

Chebyshev interpolants converge exponentially for analytic functions and algebraically for functions with a finite number of derivatives. More precisely, for an analytic path, $\|x - x_n\| = O(\rho^{-n})$ for some $\rho > 1$ as $n \rightarrow \infty$. For a path with ν derivatives and ν th derivative of bounded variation K , $\|x - x_n\| = O(Kn^{-\nu})$ as $n \rightarrow \infty$. These estimates are in the supremum norm $\|a\|$, that is, the maximum of the absolute value of a in the interval $[-1, 1]$. In contrast, finite-difference methods can only achieve polynomial, but never exponential, rates of convergence, even for analytic paths [40,41].

B. Quadrature

To reduce the action to a multivariate function of the coefficients it is necessary to evaluate the integral

$$S(\alpha_1, \dots, \alpha_n) = \int_{-1}^1 L(x_n(u), x'_n(u)) du \quad (34)$$

using a quadrature rule. For instance, quadrature at the Chebyshev points u_j gives

$$\begin{aligned} S(\alpha_1, \dots, \alpha_n) &= \sum_{j=0}^n \omega_j L(x_n(u_j), x'_n(u_j)) \\ &= \sum_{j=0}^n \omega_j L(\alpha_j, \beta_j) \\ &= \sum_{j=0}^n \omega_j L(\alpha_j, D_{jk} \alpha_k), \end{aligned}$$

where ω_j are the quadrature weights. However, standard quadrature rules at this set of n Chebyshev points, which integrate a polynomial of degree less than or equal to n exactly, will generally be inaccurate. The reason is that the Lagrangian has polynomial degree different from, and usually greater than, the polynomial degree of the path. For instance, when b_{ij} is a constant, the term quadratic in the velocities has twice the polynomial degree of the path. Therefore, if the Lagrangian is to be integrated accurately, then the order of the quadrature must be different from, and in general greater than, the polynomial degree of the path.

Therefore, we define a second set of $n_q > n$ Chebyshev points

$$v_j = -\cos(j\pi/n_q), \quad (j = 0, 1, \dots, n_q), \quad (35)$$

and interpolate the path at these points. This is done efficiently by matrix multiplication with a $(n_q + 1) \times (n + 1)$ matrix

$$x_n(v_j) = \sum_{k=0}^n B_{jk} \alpha_k, \quad (36)$$

$$x'_n(v_j) = \sum_{k=0}^n B_{jk} \beta_k, \quad (37)$$

whose elements are derived from the barycentric interpolant

$$B_{jk} = \frac{w_k}{v_j - u_k} \bigg/ \sum_{l=0}^{n_q} \frac{w_l}{v_l - u_l}. \quad (38)$$

The Lagrangian is evaluated at these second set of points after which Clenshaw-Curtis quadrature [40,41] is used to evaluate

the action,

$$\begin{aligned}
 S(\alpha_1, \dots, \alpha_n) &= \sum_{j=0}^{n_q} \omega_j L(x_n(v_j), x'_n(v_j)) \\
 &= \sum_{j=0}^{n_q} \omega_j L(B_{jk}\alpha_k, B_{jk}\beta_k) \\
 &= \sum_{j=0}^{n_q} \omega_j L(B_{jk}\alpha_k, B_{jk}D_{kl}\alpha_l) \\
 &\equiv \sum_{j=0}^{n_q} \omega_j L(B_{jk}\alpha_k, C_{jk}\alpha_k). \quad (39)
 \end{aligned}$$

As with interpolation, Clenshaw-Curtis quadrature converges exponentially for Lagrangians that are analytic in u and algebraically for Lagrangians with a finite number u derivatives. Precise estimates are given in Ref. [42]. For fixed values of n and n_q , the matrices B_{ij} and C_{ij} in the above expression are constant and can be precomputed and stored. The multiplications require $O(nn_q)$ operations, and so there is a linear cost, for fixed n , to increase the order of the quadrature. For Lagrangians of polynomial order n_L , the number of quadrature points must be $n_q > (n+1)n_L$. For nonpolynomial Lagrangians, n_q has to be chosen to ensure that the n_q th Chebyshev coefficient is suitably small. Well-defined procedures exist for the adaptive truncation of Chebyshev series [46] but here we use a simple rule of thumb and set $n_q = 10n$ leaving the implementation of more efficient truncations to future work. We note that in the direct finite-difference method, introduced in Ref. [21], the path is interpolated at uniformly spaced points by a quadratic polynomial and the Lagrangian is integrated using the trapezoidal rule. This combination can exactly evaluate the action for Lagrangians that are at most quadratic polynomials.

C. Optimization

To minimize the action over the expansion coefficients $\alpha_1, \dots, \alpha_{n-1}$ we use both gradient-free and gradient-based algorithms. For gradient-free algorithms we provide Eq. (39) directly. For algorithms that require the gradient, the chain rule gives

$$\begin{aligned}
 \nabla_{\alpha_i} S &= \frac{\partial}{\partial \alpha_i} \left[\sum_{j=1}^{n_q} \omega_j L(B_{jk}\alpha_k, C_{jk}\alpha_k) \right] \\
 &= \sum_{j=1}^{n_q} \left[\frac{\partial L}{\partial x_n(v_j)} B_{ji}^* + \frac{\partial L}{\partial x'_n(v_j)} C_{ji}^* \right], \quad (40)
 \end{aligned}$$

where $B_{ij}^* = \omega_i B_{ij}$ and $C_{ij}^* = \omega_i C_{ij}$. These matrices, too, can be precomputed and stored and only the partial derivatives of the Lagrangian need to be computed for given values of the coefficients. For the examples presented below, we use *NEWUOA* [47] for gradient-free optimization and the *SLSQP* algorithm [48] for gradient-based optimization, both of which are implemented in the *NLOPT* numerical optimization package [49]. For nonequilibrium systems, instantons lose smoothness when passing through fixed points. For such paths, convergence is still achieved but at less than spectral

rates. Spectral convergence can be recovered if paths are evaluated piecewise, taking care to isolate the points of derivative discontinuities. This is feasible because fixed points are the only locations where Freidlin-Wentzell instantons can lose smoothness [29].

V. NUMERICAL RESULTS

In this section, we apply the Ritz method to three diffusion processes that are widely used to benchmark rare event algorithms. The first is overdamped Brownian motion in a complex energy landscape, the second is overdamped Brownian motion under the influence of a circulatory force, and the third is a model of the weather. All three models have configuration-independent diffusion tensors for which it is not necessary to distinguish between covariant and contravariant indices. Python codes for each of these examples are freely available on GitHub.

A. Brownian dynamics in a complex potential

Our first example considers the overdamped Brownian motion in a two-dimensional potential with a constant friction. The usual equations of Brownian dynamics can be recast into Itô form,

$$\begin{aligned}
 dX_1 &= -\mu \partial_1 U dt + \sqrt{2\mu\varepsilon} dW_1 \\
 dX_2 &= -\mu \partial_2 U dt + \sqrt{2\mu\varepsilon} dW_2,
 \end{aligned}$$

where μ is the mobility and $\varepsilon = k_B T$ is the temperature. The Freidlin-Wentzell action for a smooth path with two-dimensional coordinate $x = (x_1, x_2)$ is

$$S[x] = \frac{1}{2} \int_0^T \frac{1}{2\mu} |\dot{x} + \mu \nabla U|^2 dt,$$

where $\nabla U = (\partial_1 U, \partial_2 U)$. The minimum of the zero-energy action,

$$S_0[x] = \int_{-1}^1 |\nabla U(x)| |x'| du + [U(x)]_{-1}^1,$$

provides the most probable shape and the stationary quasipotential. The second term does not affect the minimization and can be discarded. The resulting reduced action

$$\tilde{S}[x] = \int_{-1}^1 |\nabla U| |x'| du \quad (41)$$

is of the same form as Fermat's principle for optical rays, where $|\nabla U(x)|$ plays the role of the refractive index and $|x'| du = ds$ is the arc length of the ray. In geometric optics, Fermat's principle is equivalent to Huygen's principle and its "wavelet equation,"

$$\partial_i U = |\nabla U| \frac{dx_i}{ds}. \quad (42)$$

This can be easily verified by differentiating it with respect to arc length, to obtain the eikonal equation,

$$\partial_i |\nabla U| = \frac{d}{ds} \left[|\nabla U| \frac{dx_i}{ds} \right],$$

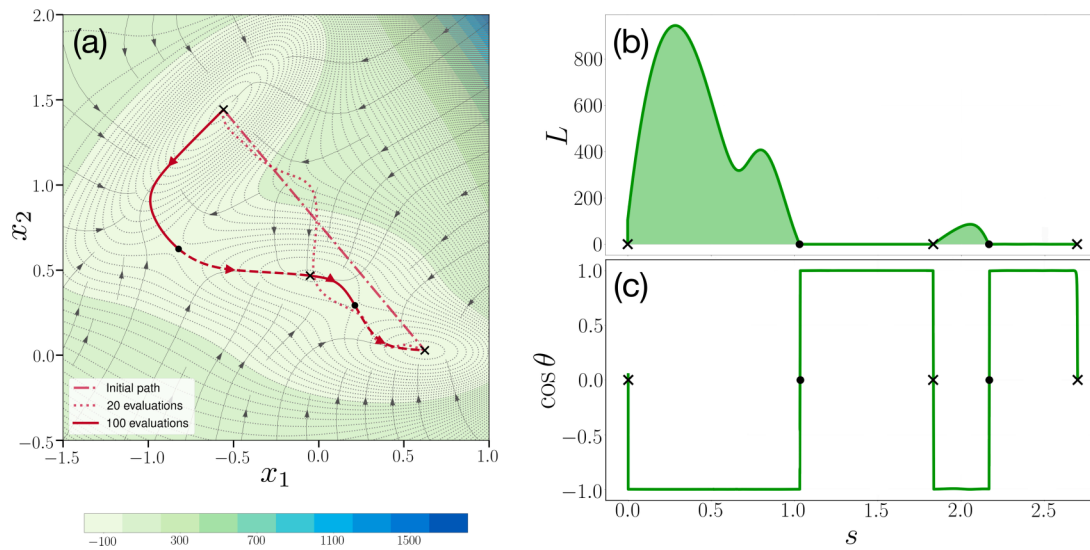


FIG. 1. Ritz method for overdamped motion in the Müller-Brown potential, which has three minima (crosses) and two saddle points (dots). The initial path is the straight line connecting two minima and the instanton is the solid line, with broken segments showing motion along the force. The instanton automatically locates and passes through both saddles. A typical path before convergence to the minimum is shown as a dotted line. (b) The value of the Lagrangian as a function of Euclidean arc length of the instanton. The action vanishes to machine precision on segments of the path where motion is along the force. (c) The cosine of the angle θ between the tangent and the force is always ± 1 , i.e., the instanton is a minimum energy path. The instanton is represented by a polynomial of degree $n = 10$.

which is identical to the Euler-Lagrange equation of the zero-energy action. The wavelet equation implies that the tangent $t = dx/ds$ to the path is parallel to the gradient of the potential, or equivalently, that rays are normal to contours of the potential. This is the well-known condition for a minimum energy path and was first derived variationally from the scalar work functional by Olender and Elber [35]. It provides a stringent test of the fidelity of the paths obtained by minimization.

Following Ref. [35], we choose the Müller-Brown potential in Ref. [50] as an example of a complex energy landscape. The potential and its stationary points are shown in Fig. 1. The three minima are marked by crosses and two saddle points by dots. The instanton is computed by requiring the path to start at the minimum on the top left and terminate at the minimum on the bottom right. The initial straight-line shape, an intermediate shape and the converged instanton are shown in Fig. 1(a). The minimization automatically locates the two saddle points and makes the the instanton pass through them. The action cost along the path is shown in Fig. 1(b), where the vanishing of the action on segments of the path along the force is clearly seen. The cosine of the angle between the tangent and force is shown in Fig. 1(c) and the condition for a minimum energy path is clearly fulfilled. We emphasize that the condition is not imposed separately but is satisfied automatically at the minimum. The Ritz method provides an alternative to chain-of-states methods for finding minimum energy paths. It does not need the Hessian of the potential, which makes it suitable for problems where such evaluations are expensive. Unlike Ref. [17], our parametrization has no unit-speed constraint and the minimization, accordingly, is unconstrained. The method applies without change to dynamics with configuration-dependent friction.

B. Brownian dynamics in a circulatory field

Our second example consider, in contrast to the first, Brownian motion in a force field that cannot be derived from a potential and, as such, necessarily has a nonvanishing curl. Choosing the force field of Maier and Stein [9] gives

$$dX_1 = (X_1 - X_1^3 - \beta X_1 X_2^2)dt + \sqrt{\epsilon}dW_1$$

$$dX_2 = -(1 + X_1^2)X_2dt + \sqrt{\epsilon}dW_2$$

for the overdamped motion of the two-dimensional coordinate $X = (X_1, X_2)$, where β is a parameter. The force field $f(x_1, x_2) = [x_1 - x_1^3 - \beta x_1 x_2^2, -(1 + x_1^2)x_2]$ is smooth, and f_1 is odd in x_1 and even in x_2 , while for f_2 the converse holds. There are two stable fixed points at $x_a = (-1, 0)$ and $x_b = (1, 0)$, and a saddle point at $x_s = (0, 0)$. The force field admits a potential only for $\beta = 1$, when it can be written as $f = -\nabla U$, with $U(x_1, x_2) = -\frac{1}{2}x_1^2 + \frac{1}{4}x_1^4 + \frac{1}{2}(1 + x_1^2)x_2^2$. The force field is shown in the first panel of Fig. 2 for $\beta = 10$ together with the instanton moving from x_a to x_b . As before, solid (dashed) segments represent motion against (along) the vector field. The instanton moving from x_b to x_a is obtained by reflection about the x_1 axis showing that that fluctuational and relaxational paths are not identical in a nongradient field.

The middle panels shows the stationary quasipotential $V_\infty^{A_i}(x)$ with respect to the attractors at $(-1, 0)$ and $(1, 0)$, respectively. The quasipotential is sampled on a 128×128 grid by computing instantons between a point on the grid and the relevant attractor. The contours of the quasipotential and its heatmap are obtained from these discrete samples. To the best of our knowledge, all prior estimations of the quasipotential for this problem (and more generally, for circulatory forces) have required numerical solutions of the Hamilton-Jacobi equation. Our method of direct sampling provides

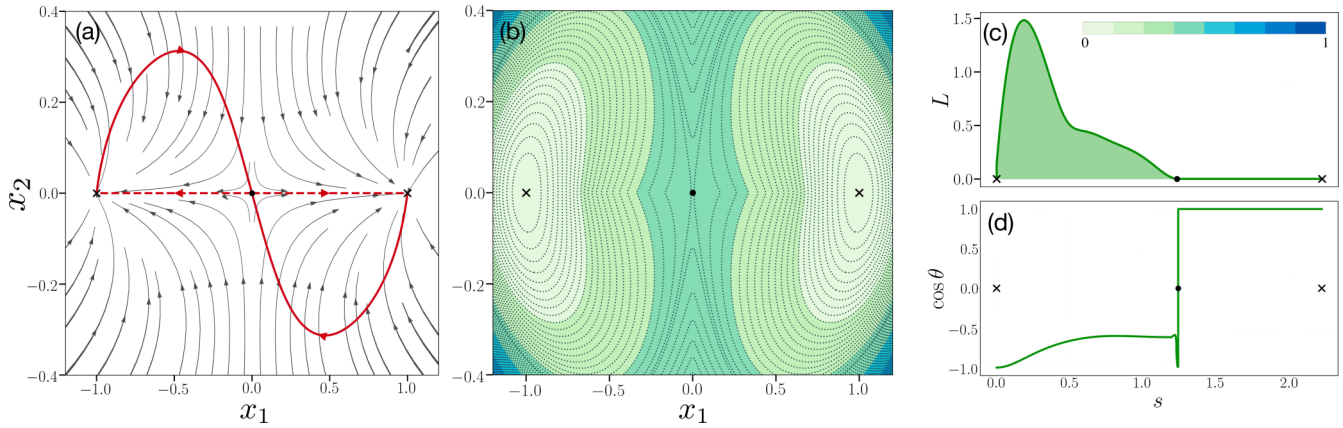


FIG. 2. Ritz method for overdamped motion in a circulatory (i.e., nongradient) force field. The instanton is in red with solid (dashed) segments showing motion against (along) the force field. The instanton is reflected about the horizontal axis for motion starting on the right, showing the inequivalence of fluctuational and relaxational paths for nongradient dynamics. (b) The quasipotential, computed using Eq. (13), with a caustic at the unstable fixed point. (c) The value of the Lagrangian as a function of the Euclidean arc length of the instanton. As in the potential case, the action vanishes to machine precision on segments where motion is along the force. (d) The cosine of the angle θ between the tangent and the force is, unlike in the potential case, not always ± 1 . The instanton is represented by a polynomial of degree $n = 8$.

an alternative to this route of computing the quasipotential. The right panel shows the Lagrangian as a function of arc length along the instanton. As in the previous example, the Lagrangian vanishes along segments of the path where motion is along the force. For motion against the force, the tangent to the path is no longer parallel to the force, as shown by the variation of the cosine of the angle θ between the tangent and the force. We note that our method is agnostic to the existence, or not, of a potential for the drift and treats both these cases on equal footing.

C. Egger model of weather

Our final example is a reduced model of the weather for a three-dimensional coordinate $X = (X_1, X_2, X_3)$ that has a

circulatory drift,

$$\begin{aligned} dX_1 &= \left[kX_2 \left(X_3 - \frac{\beta}{k^2} \right) - \gamma X_1 \right] dt + \sqrt{\epsilon} dW_1 \\ dX_2 &= \left[kX_1 \left(\frac{\beta}{k^2} - X_3 \right) - \gamma X_2 + \frac{HX_3}{k} \right] dt + \sqrt{\epsilon} dW_2 \\ dX_3 &= \left[-\frac{1}{2} HkX_2 - \gamma (X_3 - U_0) \right] dt + \sqrt{\epsilon} dW_3, \end{aligned} \quad (43)$$

where $k, \beta, \gamma, U_0,$ and H are constants. This model is due to Egger [51]. It is not particularly illuminating to visualize the three-dimensional vector field describing this dynamics but we note that it has two stable fixed points, marked by crosses in Fig. 3, and a saddle fixed point marked by a dot.

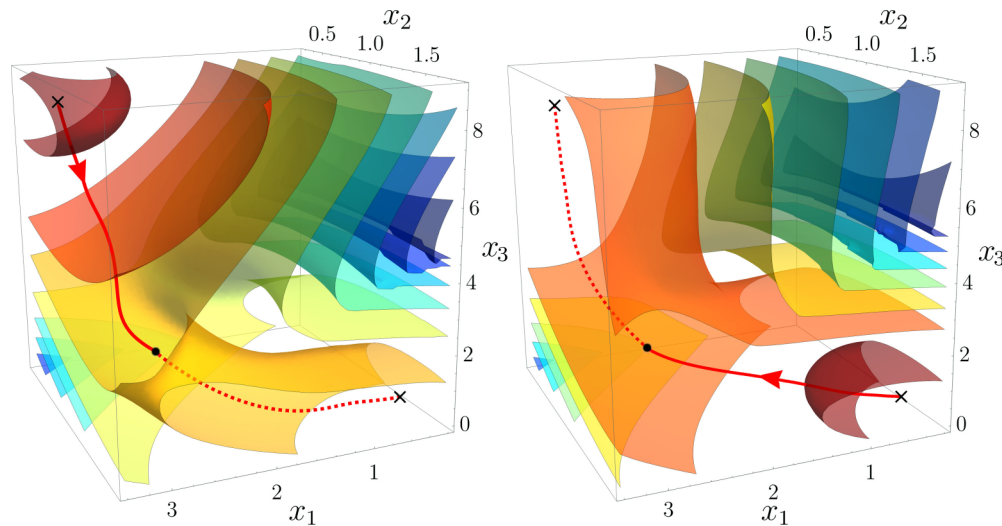


FIG. 3. Instantons and quasipotentials of the Egger model. The instantons are shown in red with solid (dashed) lines representing motion against (along) the vector field. The left and right panels are forward and reverse instantons. Isosurfaces of the quasipotential with respect to each attractor is shown in the respective panels. Isovalues increase from light red to blue in the range $\{1, 7, 11, 16, 21, 26, 31, 36\}$. Parameter values are $k = 2, \beta = 1.25, \gamma = 2, U_0 = 10.5,$ and $H = 12$. The instanton is represented by a polynomial of degree $n = 10$.

TABLE I. Convergence of the action S_n for a path of polynomial order n . The abbreviations M-B and M-S refer to Brownian dynamics in the Müller-Brown potential and the Maier-Stein force field respectively. The first five columns show the difference $S_n - S_{n+1}$ while the next five show the difference $S_n - S_{50}$. A 10th-order polynomial typically gives at least six digits of accuracy.

Model	$S_1 - S_2$	$S_3 - S_4$	$S_7 - S_8$	$S_{15} - S_{16}$	$S_{31} - S_{32}$	$S_2 - S_{50}$	$S_4 - S_{50}$	$S_8 - S_{50}$	$S_{16} - S_{50}$	$S_{32} - S_{50}$
M-B	2	3×10^{-4}	1×10^{-7}	5×10^{-14}	1×10^{-13}	4×10^{-2}	4×10^{-5}	8×10^{-8}	1×10^{-12}	3×10^{-13}
M-S	2×10^{-3}	3×10^{-7}	2×10^{-12}	1×10^{-16}	5×10^{-16}	2×10^{-4}	7×10^{-11}	2×10^{-13}	1×10^{-16}	2×10^{-16}
Egger	8×10^{-3}	1×10^{-3}	6×10^{-7}	4×10^{-9}	8×10^{-13}	5×10^{-3}	1×10^{-4}	1×10^{-6}	1×10^{-8}	2×10^{-10}

The instanton moving between these points is shown as before in the left and right panels of the figure. Also shown are isosurfaces of the quasipotential with respect to the stable fixed points, with isovalues increasing from red to blue. To the best of our knowledge, this is the first computation of the quasipotential for this model. We provide this example primarily to demonstrate the feasibility of sampling quasipotentials in dimensions greater than two with our method.

VI. NUMERICAL CONVERGENCE

We briefly recall the convergence properties of the Ritz method, comprising that of the basis functions, the quadrature, and the optimization. The Chebyshev interpolant converges to the most probable path, assuming that it is Lipschitz continuous, at a rate that increases with the number of derivatives the path admits and is exponential for a smooth path. Likewise, the Clenshaw-Curtis quadrature is guaranteed to converge to the minimum of the action, assuming that the Lagrangian is Lipschitz continuous. The optimal number of quadrature points for accuracy to machine precision can be obtained by following the decay of the Chebyshev coefficients of the Lagrangian and truncating at that value beyond which the coefficients vanish to machine precision. The optimization has lesser theoretical guarantees than the interpolation and quadrature, as is generally the case with search in high-dimensional spaces. However, the residual of the Ritz system provides an empirical measure for how closely the minimum has been located. In all three examples (and in others not presented here) we have found both gradient-free and gradient-based optimization to robustly locate the minima, and gradient-based methods to yield faster convergence. We note that for equilibrium problems, the gradient-free method does not require the Hessian of the energy function, which can be of significant computational advantage. In Table I we show the spectral convergence of the action with increasing polynomial order of the path for each of our examples.

VII. DISCUSSION

We have presented an efficient and accurate numerical method for computing most probable transitions paths and quasipotentials of rare diffusive events. The method directly minimizes the Freidlin-Wentzell action and thus provides a unified approach for transition paths in both equilibrium and nonequilibrium systems. Our reparametrization-invariant form of the action, derived using a Noether symmetry, is

well-suited for numerical work and is a generalization of the geometric action. This frees us from the constraints of the commonly used arc-length path parametrization and offers the maximum flexibility in choosing the space of polynomials in which action is minimized. Thus our method is not limited to the Chebyshev polynomials in $[-1, 1]$ used here but easily admits trigonometric polynomials and, more generally, any global basis. Numerical quadrature reduces the action to a multivariate function of coefficients of the path polynomial whose minimum is obtained by both gradient-free and gradient-based optimization. This gives, simultaneously, both the minimum value of the action and the most probable path. This efficiency of the method allows us to repeatedly compute minimum action paths between an attractor and a point in its basin of attraction and, thereby, map out the quasipotential. The quasipotential in a nonequilibrium steady state has the same significance as the Gibbs distribution in equilibrium and our method provides a robust way of obtaining it without the need to numerically solve the Hamilton-Jacobi partial differential equation.

The direct method used here consists of a discretization of the action followed by a search for the minimum in the resulting finite-dimensional space, expressed schematically in Fig. 4. In contrast, the majority of methods impose the vanishing variation of the action and then search for the solution of the Euler-Lagrange equation in a finite-dimensional space.

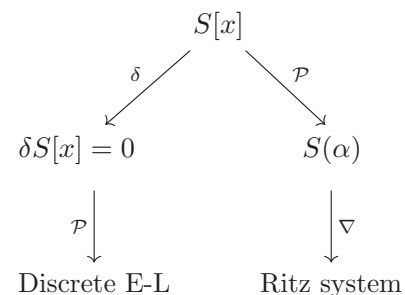


FIG. 4. Inequivalence of the direct and Euler-Lagrange routes to numerical action minimization. Here, $\delta \rightarrow$ functional variation, $\mathcal{P} \rightarrow$ finite-dimensional projection, and $\nabla \rightarrow$ function minimization. On the left branch, the action is first varied to obtain the Euler-Lagrange equation and then projected onto a finite-dimensional basis for numerical solution. On the right branch, the action is first projected onto a finite-dimensional basis and then minimized to obtain the Ritz system. The finite-dimensional projection of the Euler-Lagrange equations is, in general, not identical to the Ritz system.

The resulting discretized Euler-Lagrange equations is, in general, not identical to the Ritz system; in other words, these two methods of reducing an infinite-dimensional problem to a finite-dimensional one are not equivalent. In contrast to mechanics, where Newton's equations of motion are considered primary and the action derived, here it is the tube probability and hence the Freidlin-Wentzell action that is primary and the Euler-Lagrange equation for the most probable path that is derived. It appears more natural to us to discretize the primary, rather than the derived, object directly. Our approach is algorithmically simple and the only adjustable parameters are the polynomial order n and the quadrature order n_q . This simplicity does not compromise accuracy or efficiency, as confirmed by our examples.

The rapid convergence of the method holds promise for its application to problems involving the stochastic dynamics of fields, with both scalar and Lie group-valued order parameters. We also expect the method to apply to stochastic dynamics with degenerate diffusion tensors and to stochastic systems with inertia. These will be addressed in forthcoming work.

ACKNOWLEDGMENTS

We thank Julian Kappler for many helpful discussions. Work funded in part by the European Research Council under the Horizon 2020 Programme, ERC Grant No. 740269. R.S. is funded by a Royal Society-SERB Newton International Fellowship. M.E.C. is funded by the Royal Society.

-
- [1] A. D. Wentzell and M. I. Freidlin, *Russ. Math. Surv.* **25**, 1 (1970).
- [2] P. G. Bolhuis, D. Chandler, C. Dellago, and P. L. Geissler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).
- [3] R. J. Allen, P. B. Warren, and P. R. ten Wolde, *Phys. Rev. Lett.* **94**, 018104 (2005).
- [4] R. J. Allen, C. Valeriani, and P. R. ten Wolde, *J. Phys.: Condens. Matter* **21**, 463102 (2009).
- [5] L. Ebener, G. Margazoglou, J. Friedrich, L. Biferale, and R. Grauer, *Chaos* **29**, 063102 (2019).
- [6] L. Paninski, *J. Comput. Neurosci.* **21**, 71 (2006).
- [7] S. Huang, *Bioessays* **34**, 149 (2012).
- [8] F. Bouchet and J. Reygner, *Annales Henri Poincaré* **17**, 3499 (2016).
- [9] R. S. Maier and D. L. Stein, *J. Stat. Phys.* **83**, 291 (1996).
- [10] P. Wolynes, J. Onuchic, and D. Thirumalai, *Science* **267**, 1619 (1995).
- [11] B. C. Noltong and K. C. Abbott, *Ecology* **97**, 850 (2016).
- [12] M. Mangel, *Theor. Popul. Biol.* **45**, 16 (1994).
- [13] T. S. Gardner, C. R. Cantor, and J. J. Collins, *Nature* **403**, 339 (2000).
- [14] C. L. DeMarco, *IEEE Control Syst. Mag.* **21**, 40 (2001).
- [15] R. Nelson, *J. ACM* **34**, 661 (1987).
- [16] E. Weinan, W. Ren, and E. Vanden-Eijnden, *Phys. Rev. B* **66**, 052301 (2002).
- [17] M. Heymann and E. Vanden-Eijnden, *Commun. Pure Appl. Math.* **61**, 1052 (2008).
- [18] T. Grafke, T. Schäfer, and E. Vanden-Eijnden, in *Recent Progress and Modern Challenges in Applied Mathematics, Modeling and Computational Science* (Springer, New York, 2017), pp. 17–55.
- [19] I. Gelfand and S. Fomin, *Calculus of Variations*, Dover Books on Mathematics (Dover, London, 2012).
- [20] L. V. Kantorovich and V. I. Krylov, *Approximate Methods of Higher Analysis* (P. Noordhoff, Groningen, 1958).
- [21] E. Weinan, W. Ren, and E. Vanden-Eijnden, *Commun. Pure Appl. Math.* **57**, 637 (2004).
- [22] W. Ritz, *J. Math.* **135**, 1 (1909).
- [23] E. Vanden-Eijnden and M. Heymann, *J. Chem. Phys.* **128**, 061103 (2008).
- [24] M. Cameron, *Physica D* **241**, 1532 (2012).
- [25] S. Yang, S. F. Potter, and M. K. Cameron, *J. Comput. Phys.* **379**, 325 (2019).
- [26] D. Dahiya and M. Cameron, *J. Sci. Comput.* **75**, 1351 (2018).
- [27] G. A. Pavliotis, *Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations*, Vol. 60 (Springer, Berlin, 2014).
- [28] R. Graham, in *Springer Tracts in Modern Physics* Vol. 66 (Springer, Berlin, Heidelberg, 1973), pp. 1–97.
- [29] R. Graham, in *Fluctuations and Stochastic Phenomena in Condensed Matter*, edited by L. Garrido (Springer, Berlin, Heidelberg, 1987), pp. 1–34.
- [30] H. Touchette, *Phys. Rep.* **478**, 1 (2009).
- [31] D. Ludwig, *SIAM Rev.* **17**, 605 (1975).
- [32] E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge Mathematical Library (Cambridge University Press, Cambridge, 1988).
- [33] L. D. Landau and E. M. Lifshitz, *Mechanics*, 3rd ed. (Butterworth-Heinemann, Oxford, 1959).
- [34] F. Gantmacher, *Lectures in Analytical Mechanics*, translated by G. Yankovsky (Mir, Moscow, 1970).
- [35] R. Olender and R. Elber, *J. Mol. Struct: THEOCHEM* **398**, 63 (1997).
- [36] L. Onsager and S. Machlup, *Phys. Rev.* **91**, 1505 (1953).
- [37] R. L. Stratonovich, in *Noise in Nonlinear Dynamical Systems*, edited by F. Moss and P. V. E. McClintock, Vol. 1 (Cambridge University Press, Cambridge, 1989), pp. 16–71.
- [38] L. E. Elsgolts and G. Yankovsky, *Differential Equations and the Calculus of Variations* (Mir, Moscow, 1973).
- [39] M. J. Gander and G. Wanner, *SIAM Rev.* **54**, 627 (2012).
- [40] L. N. Trefethen, *Spectral Methods in MATLAB*, Vol. 10 (SIAM, Philadelphia, PA, 2000).
- [41] J. P. Boyd, *Chebyshev and Fourier Spectral Methods* (Courier Corporation, Mineola, NY, 2001).
- [42] L. N. Trefethen, *Approximation Theory and Approximation Practice*, Vol. 128 (SIAM, Philadelphia, PA, 2013).
- [43] R. Hamming, *Numerical Methods for Scientists and Engineers* (Courier Corporation, New York, 2012).
- [44] H. E. Salzer, *Comput. J.* **15**, 156 (1972).
- [45] J.-P. Berrut and L. N. Trefethen, *SIAM Rev.* **46**, 501 (2004).
- [46] J. L. Aurentz and L. N. Trefethen, *ACM Trans. Math. Softw.* **43**, 33 (2017).

- [47] M. J. D. Powell, in *Large-scale Nonlinear Optimization*, edited by G. Di Pillo and M. Roma (Springer US, Boston, MA, 2006), pp. 255–297.
- [48] D. Kraft, *ACM Transactions on Mathematical Software* **20**, 262 (1994).
- [49] S. G. Johnson, The nlopt nonlinear-optimization package (2014), Version 2.6.2, <http://github.com/stevengj/nlopt>.
- [50] K. Müller and L. D. Brown, *Theor. Chim. Acta* **53**, 75 (1979).
- [51] J. Egger, *J. Atmos. Sci.* **38**, 2606 (1981).