

Escape over a saddle by colored noise: Theory and numerics

Jiayao Shao (邵家瑶)^{*}, Tobias Grafke[†], and Robert S. MacKay[‡]
Mathematics Institute, University of Warwick, Coventry CV4 7AL, United Kingdom



(Received 28 February 2026; accepted 9 July 2026; published 12 August 2026)

Stochastic dynamical systems allow modeling of transitions induced by random disturbances, in particular from an attracting equilibrium and crossing the stable manifold of a saddle. While the small-noise limit is well-described by the large deviation principle, existing computational methods often struggle with stochastic forcings any more complicated than nondegenerate Gaussian white noise, and with unbounded time intervals. The primary innovations of this work are extending the framework to cater for colored and degenerate forcing and unbounded time horizons—scenarios that are physically realistic but numerically challenging. We cater for degenerate noise by using the Hamiltonian optimal control method. We cater for a class of colored noises by using linear filters on white noise. We cater for infinite time horizon by introducing the method of division (MOD), a computational approach for approximating rare transition events, including their most likely paths and the exponential scaling laws of their transition rates. The effectiveness of MOD and the above approaches to colored noise, is demonstrated by illustration on two examples: an inverted double-well potential and a simplified roll-heave model for ship capsize.

DOI: [10.1103/w8dd-9cqq](https://doi.org/10.1103/w8dd-9cqq)

I. INTRODUCTION

Dynamical systems theory is widely used in fields ranging from physics and chemistry to biology, where in particular it provides a framework for modeling state transitions by representation of motion in phase spaces. The point in phase space completely characterizes the system's current state, so that the physical mechanism of a transition can be understood from its phase space trajectory.

A key topic within this field is the departure from a stable equilibrium under deterministic external disturbances, a phenomenon observed across a wide range of systems. Examples include the isomerization of molecular clusters [1] and the dynamic behavior of microbeams under electrostatic actuation [2]. Classical nonlinear oscillators under deterministic forcing have long served as canonical models for studying such escape phenomena [3,4]. A particularly important application arises in marine engineering: ship capsize problems mentioned in Naik and Ross [5], Thompson and Souza [6], are interpreted as escapes from basins of attraction, under periodic forcing. However, real forcing for a ship is not periodic. One can treat the effects of general bounded aperiodic forcing by extension of the theory of nonautonomous hyperbolic dynamics [7]. Within this deterministic context, the transition is often conceptualized via transition state theory (TST) [8],

which identifies a critical dividing surface—typically associated with a saddle point—that trajectories must cross to escape a potential well.

For many purposes, however, it is better to view the forcing as coming from a stochastic process, in particular, if one wishes to determine the scaling of the transition rates with the strength of the noise.

From the probabilistic viewpoint, disturbances are treated as random noise and methods are employed for stochastic differential equations. The study of noise-induced transitions traces back to Kramers [9], who provided a stochastic generalization of TST to account for the influence of thermal fluctuations and friction on escape rates. This refinement extends the purely geometric view of the transition to a dynamic one, where the "bottleneck" is crossed via noise-induced fluctuations. Subsequently, a more general and rigorous foundation was established through the large-deviation framework from Freidlin and Wentzell [10]. This framework provides a systematic approach for computing the exponential scaling laws of transition rates and has found wide applicability across various disciplines, ranging from physics (e.g., Kautz [11]) to the study of metastable systems (e.g., in computational neuroscience [12], atmospheric sciences [13], nonequilibrium phase transitions and active matter [14], and climate tipping points [15]). The theory was later extended to higher-dimensional systems and non-Markovian perturbations such as colored noise [16]. Beyond transition rates, more recent work has focused on computing the most probable transition paths. Numerical schemes such as the minimum action method and its variants [17,18] enable efficient computation of these optimal paths in many contexts, with applications in fields like hydrodynamics [19].

Based on deterministic and probabilistic approaches, this paper introduces a novel method for analyzing transitions from a stable equilibrium (for example, a well in a potential

^{*}Contact author: jiayao.shao@warwick.ac.uk

[†]Contact author: T.Grafke@warwick.ac.uk

[‡]Contact author: R.S.MacKay@warwick.ac.uk

landscape) to a saddle point. Our approach offers several key advantages: (i) it accommodates infinite transition time both analytically and computationally; (ii) it allows for degenerate noise (i.e., that does not act in all directions in the phase space); and (iii) it allows the external perturbation to be colored noise rather than white noise. Taken together, these three points result in an approach that is capable of handling physically relevant situations that are out of reach for previous methods. In safety and reliability engineering, one is generally interested in the rate of occurrence of very unlikely transitions, which can be computed by estimating the barrier height to cross in the infinite-time limit. Crucially, these infinite-time trajectories are not merely theoretical constructs. They are essential for practical estimation of mean first passage times and transition rates per unit time. For instance in the ship model, these metrics capture the dominant exponential contribution to the capsizing rate, providing a rigorous basis for long-term safety assessment. Existing methods to compute barrier heights on an infinite time horizon, e.g., Grafke *et al.* [20], Heymann and Vanden-Eijnden [21], rely on the invertibility of the noise covariance matrix, and thus do not apply to the degenerate noise case. These issues and the solutions provided by the Hamiltonian optimal control (HOC) method are discussed in detail in Sec. II B.

The paper contributes to the goal of Bujorianu *et al.* [22], namely to integrate approaches from stochastic analysis and from nonautonomous dynamical systems theory to the question of ship capsizing. It can be viewed as the stochastic companion to McSweeney-Davis *et al.* [7], which presents a development of the nonautonomous dynamical systems approach.

The outline of this paper is as follows. Section II explains the benefits and challenges of incorporating degenerate and filtered noise, particularly in the computation of the Freidlin-Wentzell action. Here, we apply the HOC method based on Grafke and Vanden-Eijnden [18] to address singularities arising from noise degeneracy. A new method called the method of division is presented in Sec. III, to allow numerical computation of infinite-time trajectories. On top of HOC, this approach isolates the two ends of the transition trajectory and linearizes the dynamics around each. In particular, it separately analyses motion in the stable and unstable directions at the saddle point. This greatly enhances computational performance and convergence efficiency by concentrating computational resources on the transition region. The linearized dynamics is extended over infinite time thus the combined trajectory is traversed in infinite time. Section IV applies the method of division to two case studies: an inverted double-well model and a ship capsizing model, illustrating its practical effectiveness and applicability.

II. STOCHASTIC DYNAMICAL SYSTEMS

A stochastic dynamical system introduces noise (usually Gaussian white) into a deterministic dynamical system, such that the system's behavior becomes random. This randomness can lead to phenomena like transitions from one potential well to another one, representing escapes from stable states due to stochastic effects. Many physical dynamical systems follow Newton's second law, leading to a phase-space for-

mulation where variables are partitioned into configuration and momentum coordinates. When stochasticity is introduced, external disturbances typically represent physical forces. Consequently, the noise should enter the system only through the equations for the momentum variables, leaving the configuration variables purely deterministic. This structural constraint is more realistic but leads to problems. Consider a generic stochastic differential equation (SDE),

$$dX = b(X)dt + \sigma dW,$$

where $X \in \mathbb{R}^n$ and $dW \in \mathbb{R}^m$. If noise is not applied across all dimensions, then the covariance matrix $C = \sigma\sigma^T$ becomes singular (noninvertible), a condition known as degenerate noise.

Degeneracy of the noise can arise in a second way. To enhance model realism, colored noise can be applied to the momentum dimensions, instead of white noise. A mathematically simple form of colored noise is given by filtering white noise, i.e., using the output ξ of a stochastic differential equation of the form

$$d\xi = A\xi dt + \sigma dW,$$

a multidimensional form of an equation originally introduced by Uhlenbeck and Ornstein [23]. In this formulation, A is a general stable matrix, meaning all its eigenvalues have negative real parts. The autocorrelation properties of higher-dimensional Ornstein-Uhlenbeck (OU) processes were analyzed in Van Kampen [24], which is essential for modeling mechanical systems with memory. Gardiner [25] further discussed the significance of filtered noise versus white noise in physical interpretations of SDEs. Adjusting the matrix A allows to explicitly choose correlation times and oscillation behavior of the random noise components ξ . While this attaches physical scales to the forcing, it is important to note that the filtered noise paths are continuous but nowhere differentiable with probability one [26]. Figure 1 provides the simplest example of passing white noise through a first-order low-pass filter. Note that it is impossible to plot a typical sample from white noise, because its value is nowhere defined, but what is plotted here is a sample of a discrete-time approximation.

Thus, a simple way to model the effect of colored noise is to augment the state space by a set of filter variables ξ with OU dynamics. It has the advantage of keeping the model in the class of standard SDEs. But a clear consequence of such modeling is that the white noise acts only on the filter variables, not the original system, so the noise is automatically degenerate. Note, however, that because the filter dynamics are linear, the resulting filtered noise is still Gaussian.

As discussed in the previous section, a common phenomenon in various domains is the occurrence of escape from a stable equilibrium, due to forcing. In this article, we will investigate a specific circumstance: transition from a stable equilibrium to a nearby saddle, under filtered noise. The motivation is transition to the region on the other side of the saddle's stable manifold, which is assumed to be of codimension one, but for small noise, the most likely route to the other side is via a set of paths that concentrate on paths that go to the saddle. In the limit of small noise, such transitions become rare and are difficult to observe via direct sampling, but their probability can be expressed by Large Deviation Theory.

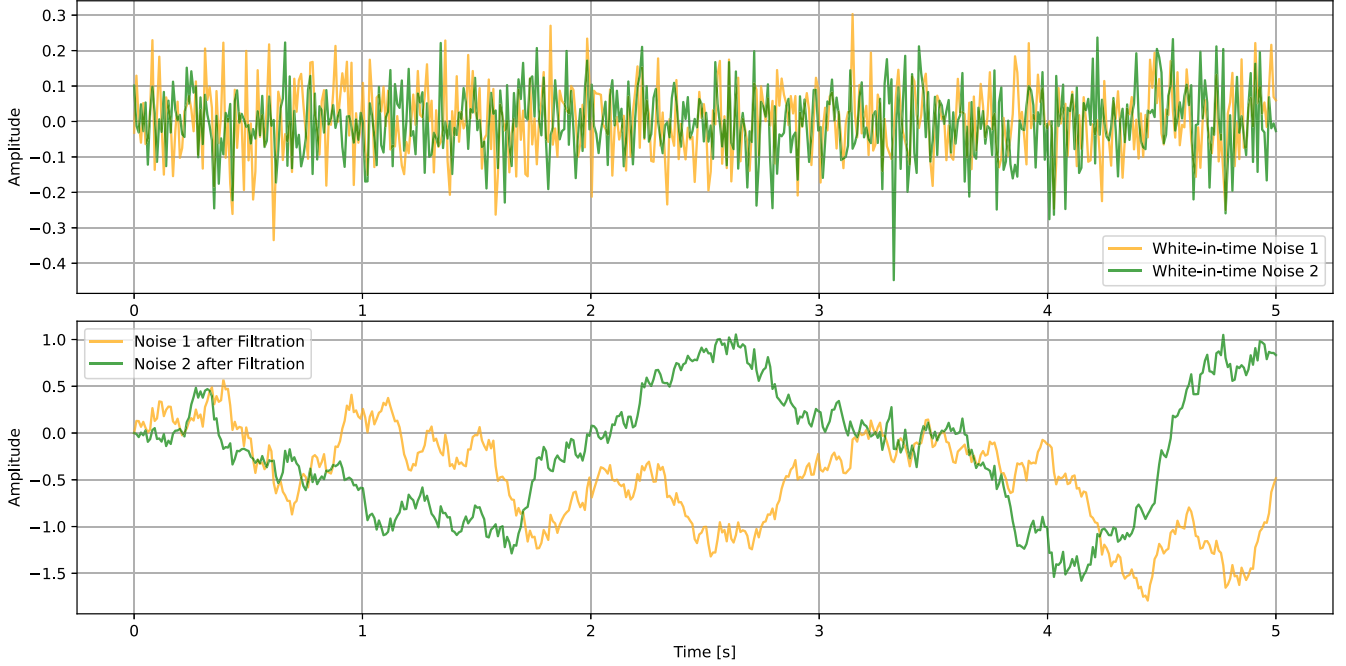


FIG. 1. Comparison between white noise and filtered noise in two-dimensional case. The top row shows increments dW_1 and dW_2 of independent white Gaussian noises. The lower row shows filtered noises ξ_1 and ξ_2 with $d\xi_1 = (-0.1\xi_1 + \xi_2)dt + dW_1$ and $d\xi_2 = (-\xi_1 - 0.1\xi_2)dt + dW_2$.

A. Large deviation theory

The theory of large deviations in probability theory deals with the behavior of extreme tails of sequences of probability distributions, going beyond the domain of applicability of central limit theory. It formalizes the concepts of concentration of measures and extends the notion of convergence of probability measures [27]. Here we follow the scheme from Ref. [18].

Given a generic SDE taking the form

$$dX_t = b(X_t)dt + \sqrt{\epsilon}\sigma(X_t)dW_t, \quad X_t \in \mathbb{R}^n,$$

with a noise strength ϵ , initial condition X_0 , total time T and deterministic drift $b(X_t)$, the probability that a sample path $\{X_t\}_{t=0}^T$ is located inside a cylinder of small radius δ around a chosen path $\phi(t)$ is said to satisfy a large deviation principle (LDP) if

$$\mathbb{P}\left(\sup_{t \in [0, T]} |X_t - \phi(t)| < \delta\right) \asymp e^{-\frac{S[\phi]}{\epsilon}} \quad (1)$$

when $\epsilon \rightarrow 0$, where $S[\phi]$ is a functional called a "rate function." The $\asymp$ denotes that the asymptotic behavior of both sides is equivalent when taking the logarithm:

$$\lim_{\epsilon \rightarrow 0} \epsilon \log \mathbb{P}\left(\sup_{t \in [0, T]} |X_t - \phi(t)| < \delta\right) = -S[\phi].$$

We can write this as $\mathbb{P}(\sup_{t \in [0, T]} |X_t - \phi(t)| < \delta) = K(\epsilon) \exp(-S[\phi]/\epsilon)$ with K being some prefactor that is subexponential. The rate function was derived by Freidlin and Wentzell [10] for the case of nondegenerate noise, i.e., $C = \sigma\sigma^T$ positive-definite.

Theorem 1 (Freidlin-Wentzell theorem). The sample paths of the stochastic differential equation

$$dX_t = b(X_t)dt + \sqrt{\epsilon}\sigma(X_t)dW_t,$$

with $X_t \in \mathbb{R}^n$, $t \in [0, T]$ satisfy a large deviation principle with rate functional

$$S[\phi] = \frac{1}{2} \int_0^T \langle \dot{\phi} - b(\phi), C^{-1}(\dot{\phi} - b(\phi)) \rangle dt, \quad (2)$$

where $C(X_t) = \sigma(X_t)\sigma^T(X_t)$ denotes the positive-definite covariance matrix. $S[\phi]$ can be interpreted as the "action" for the Euler-Lagrange dynamics of (2), and $\phi \in \mathbb{R}^n$ is any path to which an action value can be assigned.

The stated theorem provides a LDP for paths. We are interested in probability of transition, regardless of the path. The probability to hit (close to) the saddle also fulfills an LDP via a "contraction principle" [28], where the rate function is given by the minimum over all paths that reach the saddle. The only required ingredient is that the mapping from path to endpoint is continuous (which it is), the rest follows directly via contraction.

As a consequence, LDT replaces the ineffective rare event sampling problem with a deterministic optimization problem to compute the minimum action value S^* . Under the assumption of a good rate function [10], this minimum is attained by at least one optimal transition path ϕ^* (though not necessarily unique). This reframing has multiple advantages: (i) the computational cost is roughly independent of the rareness of the event, whereas direct sampling of rare events becomes typically exponentially harder in the tails, (ii) the action at the minimum allows to quantify the tail scaling of the expected waiting time to observe the event, and its explicit dependence on the noise strength is known, and (iii) the minimizing trajectory ϕ^* (possibly nonunique) represents the most likely way the rare transition is realized, allowing to identify physical mechanisms responsible for, or involved in, the transition.

When the noise is degenerate, however, the classical expression for the action functional (2) becomes formally ill-posed due to the singularity of the noise covariance matrix. To extend the LDP to such degenerate settings, Azencott [29] introduced the use of Hörmander's condition, ensuring that the system is hypoelliptic so that the noise propagates through the deterministic coupling. This theoretical framework was later formalized and refined by Deuschel and Stroock [30].

In this work, rather than relying on the probabilistic derivation of LDP, we adopt the Hamiltonian optimal control approach. This method provides an equivalent characterization of the rate functional by identifying the minimum energy required to drive the system along a specific trajectory. By reformulating the problem as a deterministic variational task in more dimensions, we avoid inversion of the noise covariance matrix, thereby naturally accommodating the degeneracy of the noise. This approach does not require explicit preconditioning of the system and has broader applicability, extending beyond strictly hypoelliptic systems.

B. Hamiltonian optimal control method

Minimizing $S(\phi)$ is equivalent to an optimal control problem. We will first introduce the general optimal control result for a system of ODEs with nondegenerate forcing. This reproduces the same expression as the rate functional (2). Then we discuss the Hamiltonian optimal control method, which introduces a conjugate variable for each component of the state of the system, and has no requirement for inverting the covariance matrix. This section builds upon the MSRJD formalism [31–33], while adopting a classical mechanics perspective as developed by Graham and Tél [34].

For any trajectory ϕ of the dynamics $\dot{\phi} = b(\phi) + \sigma\eta$ defined on $t \in [0, T]$, where $b(\phi)$ is a deterministic vector field in n dimensions, σ is invertible and η is a square-integrable forcing function, we wish to minimize its action $\frac{1}{2} \int_0^T |\eta|^2 dt$ subject to the boundary conditions $\phi(0) = \phi_0$ and $\phi(T) = \phi_T$. This can be written as

$$\begin{aligned} \frac{1}{2} \int_0^T |\eta|^2 dt &= \frac{1}{2} \int_0^T \langle \sigma^{-1}(\dot{\phi} - b(\phi)), \sigma^{-1}(\dot{\phi} - b(\phi)) \rangle dt \\ &= \frac{1}{2} \int_0^T (\dot{\phi} - b(\phi))^T C^{-1} (\dot{\phi} - b(\phi)) dt, \end{aligned}$$

where $C = \sigma\sigma^T$. This expression yields the Freidlin-Wentzell theorem for its rate functional and corresponds to the standard optimal control approach for Lagrangian

$$\mathcal{L}(\phi, \dot{\phi}) = \frac{1}{2} \langle \dot{\phi} - b(\phi), C^{-1}(\dot{\phi} - b(\phi)) \rangle.$$

A necessary condition for an optimal solution is that it satisfy the Euler-Lagrange equations for this Lagrangian.

The above derivation holds only under the condition that the covariance matrix C is invertible. However, in many contexts, this does not hold, as in the example we mentioned above. In such cases, the Hamiltonian optimal control method (which goes back to Pontryagin and is, for example, outlined in Ref. [18]) provides a solution to address the issue of non-invertibility. We would like to minimize $\frac{1}{2} \int_0^T |\eta|^2 dt$ subject to the constraint $\dot{\phi} = b(\phi) + \sigma\eta$ and the boundary conditions $\phi(0) = \phi_0$ and $\phi(T) = \phi_T$. By introducing a Lagrange

multiplier function μ and vector β (both of dimension n), followed by integration by parts, the objective function can be chosen to be

$$\begin{aligned} J &= \frac{1}{2} \int_0^T \langle \eta, \eta \rangle dt + \int_0^T \langle \mu, \dot{\phi} - b(\phi) - \sigma\eta \rangle dt \\ &\quad + \langle \beta, (\phi(T) - \phi_T) \rangle \\ &= \frac{1}{2} \int_0^T (\eta^T \eta - \dot{\mu}^T \phi - \mu^T b(\phi) - \mu^T \sigma \eta) dt \\ &\quad + \mu(T)^T \phi(T) - \mu(0)^T \phi_0 + \beta^T (\phi(T) - \phi_T). \end{aligned}$$

Varying with respect to η and ϕ gives the following necessary conditions for finding stationary points of the above optimization problem:

$$\begin{aligned} \frac{\delta J}{\delta \eta} &= \eta^T - \mu^T \sigma = 0, \\ \frac{\delta J}{\delta \phi} &= \dot{\mu}^T + \mu^T \nabla b(\phi) = 0, \\ \frac{\delta J}{\delta \phi(T)} &= \mu(T) + \beta^T = 0. \end{aligned}$$

Assume the objective function J has a unique minimizer η^* with the corresponding Lagrange multiplier μ^* . Then, at the optimum the first equation becomes $\eta^* = \sigma^T \mu^*$. With $C = \sigma\sigma^T$ we obtain the equations for the optimized trajectory ϕ^* and its adjoint μ^* :

$$\dot{\phi}^* = b(\phi^*) + C\mu^* = \nabla_{\mu} \mathcal{H}(\phi^*, \mu^*), \quad (3a)$$

$$\dot{\mu}^* = -(\nabla b(\phi^*))^T \mu^* = -\nabla_{\phi} \mathcal{H}(\phi^*, \mu^*). \quad (3b)$$

This pair of equations can be viewed as the Hamilton's equations of motion for

$$\mathcal{H}(\phi, \mu) = \langle b(\phi), \mu \rangle + \frac{1}{2} \langle \mu, C\mu \rangle,$$

which is the Legendre-Fenchel transform of the Lagrangian $\mathcal{L}(\phi, \dot{\phi})$ with respect to $\dot{\phi}$. One can solve Eq. (3a) forward in time from $\phi(0) = \phi_0$ and solve Eq. (3b) backward in time from $\mu(T) = -\beta$, and hope to vary β until the desired final condition $\phi(T) = \phi_T$ is obtained. In our case the endpoint conditions could be the stable equilibrium (ϕ_0) and the saddle (ϕ_T) correspondingly, or points near these, as we will be interested in the limiting case of infinite time from sink to saddle. Moreover, by the forward equation (3a), the action takes a simpler form:

$$S^* = \frac{1}{2} \int_0^T \langle C\mu^*, \mu^* \rangle dt. \quad (4)$$

With this reformulation, the inverse of C is no longer required and the problem is well-defined even if σ is degenerate, namely, to minimize Eq. (4) subject to Eq. (3a) and end conditions.

The above approach can be further improved by applying the augmented Lagrangian method [35]. The point is that finding the right value of β is not always easy and it is better to convexify the objective function. This method is based on the Lagrange Multiplier concept but replaces the above constrained optimization problem with a series of unconstrained problems, by adding a penalty term to the objective function.

In our notation, this gives

$$\frac{1}{2} \int_0^T \langle \eta, \eta \rangle dt + \int_0^T \langle \mu, \dot{\phi} - b(\phi) - \sigma \eta \rangle dt + \langle \beta, (\phi(T) - \phi_T) \rangle + \lambda \|\phi(T) - \phi_T\|^2.$$

Here β is to mimic a Lagrange multiplier and λ is the penalty parameter. The limit of λ to infinity would guarantee the convergence to the ‘true’ solution but an algorithm for updating β

$$\beta \leftarrow \beta + \lambda \cdot c(\phi)$$

makes it not necessary to take λ to infinity, just large enough, thereby avoiding ill-conditioning. Here $c(\phi)$ is the associated constraint on ϕ which equals $c(\phi) = (\phi(T) - \phi_T)$ in our example. The improved version avoids numerical instabilities and leads to strong theoretical convergence. Detailed derivation is given in Chapter 17 of Ref. [36]. This method only affects the pair of Hamilton equations through the boundary condition for μ in the backward equation (3b) via $\phi(T)$ and λ .

III. METHOD OF DIVISION

While the previous section addresses the question of how to overcome complications of degenerate and colored noise, we now turn our attention to the problem of an infinite time horizon. To this effect, we introduce a new method in a general context, with one illustrative example and a more realistic but still simplified real-world application discussed in the subsequent section. This approach is specifically designed for handling transitions from a stable equilibrium to a saddle point, defined to be a hyperbolic equilibrium point with one-dimensional unstable manifold.

We would like to determine the probability rate for transition by computing the minimum of the rate functional for $T \rightarrow \infty$. However, numerical simulation over an infinite time interval is infeasible. Approximating it with a finite time interval leads to very different behavior (if the time interval is chosen too short) or is computationally prohibitive, especially for high-dimensional systems. Also, any transition trajectory on an infinite time interval spends almost all its time near the fixed points, and only briefly passes through the transitional region. Any equidistant discretization of the computational time interval thus constitutes an extremely inefficient use of computational resources, where most effort is spent on regions where the dynamics are almost linear.

Furthermore, considering an infinite time horizon introduces a numerical issue from time-translation degeneracy. Since the action functional is invariant under arbitrary shifts along an infinite time-axis, any time-shift of an optimal trajectory remains optimal. Numerically, this degeneracy causes the discretized solution to slide freely along the time domain, thereby severely undermining the robustness and reproducibility of the results and leading to significant instability.

The proposed method, called the method of division, tackles the above challenges by focusing the numerics exclusively on the nonlinear transitional regime, using analytical approximations for the semi-infinite intervals near the two equilibria via their linearizations, and requiring the transitional phase

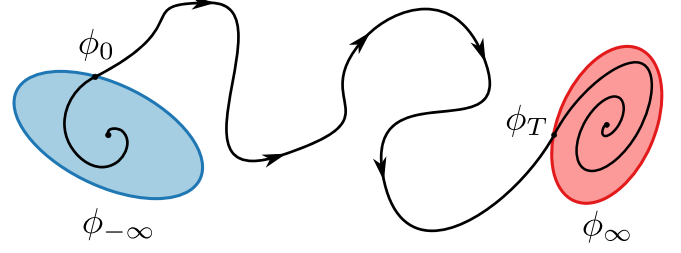


FIG. 2. Sketch of the method of division, where the transition process is divided into three parts: linearized dynamics around the sink $\phi_{-\infty}$ (in blue) and the saddle ϕ_{∞} (in red), and the transition from ϕ_0 to ϕ_T .

to end on a specified surface. This approach eliminates unnecessary computation near the equilibria, thus enhancing performance without sacrificing accuracy. Furthermore, the approximate dynamics around the endpoints are analyzed under an infinite timescale, which allows us to study the asymptotic behavior as $\sqrt{\epsilon} \rightarrow 0$ and $T \rightarrow \infty$ in equation (1). This makes it possible to investigate the limiting characteristics of the transition, yielding more meaningful insights into rare-event dynamics. Imposing a terminal constraint removes the time-translation degeneracy. Specifically, we constrain the trajectory to cross a predefined ellipsoid around the saddle at the end of the transitional phase. Instead of an ellipsoid, any separating surface between the two equilibria could work, provided that the optimal transition trajectory intersects it transversally at a single point. Our choice had the advantage of helping the trajectory to find the saddle.

The method of division divides the whole transition process into three parts:

(1) Initial escape (ϕ_0): The process involves the system escaping from the neighborhood of the stable equilibrium and moving to a point, ϕ_0 , near the equilibrium. This is allowed to take an infinite amount of time for interval $(-\infty, 0]$.

(2) Transitional phase: The core transitional phase from ϕ_0 to somewhere on the ellipsoid around the saddle, denoted as ϕ_T . This part occurs over the finite, effective transition time $[0, T]$.

(3) Final approach: The final part in which the system moves from ϕ_T to the saddle. Like the initial escape, this part can also take an infinite amount of time for interval $[T, \infty)$.

Rather than finding the optimal trajectory connecting the two equilibrium points, the method begins and terminates the optimal trajectory close enough to these points such that a linear approximation of the dynamics is accurate enough to extend it to the two equilibria. See Fig. 2 as a model sketch. $\phi_{-\infty}$ and ϕ_{∞} represent the sink and the saddle correspondingly. The part in blue is the initial escape and the part in red is the final approach. The trajectory in between spends a finite time T for effective transition. Consequently, the total action now consists of three components: the action associated with escaping the neighborhood of attraction, the action for the core transition as defined by the Hamiltonian optimal control outlined in Sec. II B, and lastly the action of overcoming the unstable direction of the saddle. Importantly, the optimization problem for the two linear parts can be solved analytically. The method of division handles a more difficult question: we

find not only the optimal trajectory for transition but also the optimal starting point ϕ_0 and ending point ϕ_T . We will explain in detail the analysis for these three processes.

Note that by a trajectory ϕ we mean a solution of the dynamics

$$\dot{\phi} = b(\phi) + \sigma\eta,$$

where $b(\phi)$ is the deterministic vector field, and $\sigma\eta$ represents the "forcing" that is to be controlled and optimized to give the smallest overall action.

A. Initial escape: Sink

Assume without loss of generality that the sink is at the origin: $\phi_{-\infty} = \mathbf{0}$. From the sink $\phi_{-\infty}$ to a nearby ϕ_0 , the dynamics can be approximated by a linearized dynamics with derivative $A_{-\infty} \in \mathbb{R}^{n \times n}$, obtained by setting $A_{-\infty} = \nabla b(\mathbf{0})$. The equation of motion is then $\dot{\phi} = A_{-\infty}\phi + \sigma\eta$, with boundary conditions $\phi(-\infty) = 0$ and $\phi(0) = \phi_0$. We can interpret the action as the energy needed to overcome friction, with external force influencing the system. Thus, the action for this part is: $\frac{1}{2} \int_{-\infty}^0 |\eta(s)|^2 ds$, subject to the constraint that this trajectory reached the point ϕ_0 .

The matrix $A_{-\infty}$ is assumed to be stable, i.e., all eigenvalues have negative real parts. By Duhamel's principle (e.g., in Zabczyk [37]),

$$\phi(t) = \int_{-\infty}^t e^{A_{-\infty}(t-s)} \sigma \eta(s) ds.$$

In particular, we have

$$\phi(0) = \int_{-\infty}^0 e^{-A_{-\infty}s} \sigma \eta(s) ds.$$

To clarify the notations used: $\phi(-\infty)$ and $\phi(0)$ refer to the starting and ending points of process 1, where $\phi(-\infty)$ is assumed to be zero and $\phi(0)$ is computed using Duhamel's principle. However, ϕ_0 represents a chosen endpoint of process 1 and is also the starting point for transitional process 2.

Given ϕ_0 , the endpoint for the process 1, as a constraint, there is the following objective function by applying the method of Lagrange multipliers:

$$\begin{aligned} F[\phi, \eta, v] &= \frac{1}{2} \int_{-\infty}^0 |\eta(s)|^2 ds + \langle v, (\phi_0 - \phi(0)) \rangle \\ &= \frac{1}{2} \int_{-\infty}^0 |\eta(s)|^2 ds + \langle v, (\phi_0 - \int_{-\infty}^0 e^{-A_{-\infty}s} \sigma \eta(s) ds) \rangle \\ &= \frac{1}{2} \int_{-\infty}^0 \eta^T \eta ds + v^T (\phi_0 - \int_{-\infty}^0 e^{-A_{-\infty}s} \sigma \eta(s) ds), \end{aligned}$$

which leads to

$$\begin{aligned} \frac{\delta F}{\delta \eta} &= \eta^T - v^T e^{-A_{-\infty}s} \sigma = 0, \\ \eta^T &= v^T e^{-A_{-\infty}s} \sigma. \end{aligned}$$

Therefore, under the optimal trajectory and forcing:

$$\begin{aligned} \phi(0) &= \int_{-\infty}^0 e^{-A_{-\infty}s} \sigma \sigma^T e^{-A_{-\infty}^T s} v ds = Qv, \\ \frac{1}{2} \int_{-\infty}^0 |\eta(s)|^2 ds &= \frac{1}{2} \int_{-\infty}^0 v^T e^{-A_{-\infty}s} \sigma \sigma^T e^{-A_{-\infty}^T s} v ds \\ &= \frac{1}{2} v^T Q v, \end{aligned}$$

where

$$Q = \int_{-\infty}^0 e^{-A_{-\infty}s} \sigma \sigma^T e^{-A_{-\infty}^T s} ds.$$

A convenient way to compute Q is to note that it satisfies the Lyapunov equation $A_{-\infty}Q + QA_{-\infty}^T = -\sigma\sigma^T$. As $A_{-\infty}$ is contracting, the solution to this Lyapunov equation exists and is unique [38] and can be computed by standard routines. Furthermore, assuming controllability of the linearized dynamics around the sink ensures that Q (called the controllability Gramian) is invertible [37]. Since at the constraint $\phi(0) = \phi_0$, we can express v as $v = Q^{-1}\phi_0$. This leads to a further simplification of the action expression:

$$\frac{1}{2} \int_{-\infty}^0 |\eta(s)|^2 ds = \frac{1}{2} \phi_0^T Q^{-1} \phi_0.$$

Given a chosen ϕ_0 , we have the optimal η for the trajectory from the stable equilibrium $\phi(-\infty)$ to ϕ_0 . By substituting $s = t + \tau$, we can explicitly express $\phi(t)$ as

$$\begin{aligned} \phi(t) &= \int_{-\infty}^t e^{A_{-\infty}(t-s)} \sigma \eta(s) ds \\ &= e^{A_{-\infty}t} \int_{-\infty}^t e^{-A_{-\infty}s} \sigma \sigma^T e^{-A_{-\infty}^T s} v ds \\ &= e^{A_{-\infty}t} \int_{-\infty}^0 e^{-A_{-\infty}(t+\tau)} \sigma \sigma^T e^{-A_{-\infty}^T (t+\tau)} v d\tau \\ &= \int_{-\infty}^0 e^{-A_{-\infty}\tau} \sigma \sigma^T e^{-A_{-\infty}^T \tau} e^{-A_{-\infty}^T t} v d\tau \\ &= Q e^{-A_{-\infty}^T t} v \\ &= Q e^{-A_{-\infty}^T t} Q^{-1} \phi_0. \end{aligned} \tag{5}$$

The final equation allows for an easy sketch of the linearized dynamics. Equivalently, the optimally controlled dynamics for the linearized system around the sink is given by

$$\dot{\phi} = -QA_{-\infty}^T Q^{-1} \phi.$$

B. Final approach: Saddle

Similar to the case of a sink, we can linearly approximate the dynamics around the saddle s using a matrix A_{∞} obtained by evaluating $\nabla b(\phi)$ at the saddle s . For simplicity, let us first shift the saddle to the origin $\mathbf{0}$. An equation that includes the effect of shifting the saddle will be given at the end of this subsection. The saddle has $n - 1$ eigenvalues with negative real parts and one positive eigenvalue $\lambda_+ > 0$.

It is convenient to analyze the linearized system in a coordinate system that separates off the unstable direction. Thus, we

define a projection P by $P = rl^T$ where r and l are the corresponding right and left eigenvectors of the positive (unstable) eigenvalue λ_+ (normalized to have $l^T r = 1$). This projection decomposes the space into $E_s \oplus E_u$, where $E_u = \text{Im}(P)$ and $E_s = \text{Ker}(P)$. At least one component of l is nonzero, without loss of generality the first. Define a transformation matrix U as

$$U = \begin{pmatrix} r_1 & -\frac{l_2}{l_1} & -\frac{l_3}{l_1} & \dots & -\frac{l_n}{l_1} \\ r_2 & 1 & 0 & \dots & 0 \\ r_3 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ r_n & 0 & 0 & \dots & 1 \end{pmatrix}. \quad (6)$$

The image by U of the first unit vector is r , and $l^T U$ is the row vector $[1, 0, \dots, 0]$. This transforms A_∞ into a matrix $\tilde{A} = U^{-1}A_\infty U$ taking the block form

$$\tilde{A} = \left(\begin{array}{c|c} \tilde{A}_u & 0 \\ \hline 0 & \tilde{A}_s \end{array} \right) = \left(\begin{array}{c|c} \lambda_+ & 0 \\ \hline 0 & \tilde{A}_s \end{array} \right),$$

where $\tilde{A}_s \in \mathbb{R}^{(n-1) \times (n-1)}$ is a stable matrix. Compared to diagonalization, this transformation defined by U is more efficient (it does not require computation of the other eigenvectors) and robust (diagonalization is unstable near cases of repeated eigenvalues, and in general impossible if there are any repeated eigenvalues). There is freedom to apply any invertible coordinate change to the $(n-1) \times (n-1)$ block, if desired, but we did not use that.

Thus, by changing the coordinates and writing $U^{-1}(\phi - s) = \tilde{\phi}$, the linearized dynamics is then equivalent to

$$\dot{\tilde{\phi}} = \tilde{A}\tilde{\phi} + \tilde{\sigma}\eta,$$

where $\tilde{\sigma} = U^{-1}\sigma$. Since $\tilde{A}$ is in block form, this can be decomposed as

$$\begin{aligned} \dot{\tilde{\phi}}^1 &= \lambda_+ \tilde{\phi}^1 + \tilde{\sigma}^1 \eta, \\ \dot{\tilde{\phi}}^i &= (\tilde{A}_s \tilde{\phi})^i + \tilde{\sigma}^i \eta, \quad \text{for } 2 \leq i \leq n, \end{aligned}$$

where $\tilde{\sigma}^i$ denotes the i th row of $\tilde{\sigma}$. Note that, to travel from a point ϕ_T to the saddle, external force is only required for the ϕ^1 component. For the components along the stable manifold, the dynamics will naturally flow toward the saddle, as long as the optimal forcing η associated with the unstable component goes to zero.

Hence, we aim to find an optimal η such that for $T \leq t \leq \infty$, the dynamics satisfies $\tilde{\phi}^1(T) = \tilde{\phi}_T^1$ and $\tilde{\phi}^1(\infty) = 0$. Since the time interval is infinite, let us shift it to $0 \leq t \leq \infty$ for simplicity in expression and calculation. The boundary condition becomes $\tilde{\phi}^1(0) = \tilde{\phi}_T^1$. By Duhamel's Principle, $\tilde{\phi}^1(0) = -\int_0^\infty e^{-\lambda_+ s} \tilde{\sigma}^1 \eta(s) ds$. Minimizing $\frac{1}{2} \int_0^\infty |\eta|^2 dt$ subject to the boundary conditions employing the Lagrangian multiplier method gives objective function:

$$G[\tilde{\phi}_T^1, \eta, \beta] = \frac{1}{2} \int_0^\infty |\eta(s)|^2 ds + \beta(\tilde{\phi}_T^1 - \tilde{\phi}^1(0)).$$

By similar computation as in the sink case,

$$\begin{aligned} G[\tilde{\phi}_T^1, \eta, \beta] &= \frac{1}{2} \int_0^\infty |\eta(s)|^2 ds \\ &\quad + \beta \left(\tilde{\phi}_T^1 + \int_0^\infty e^{-\lambda_+ s} \tilde{\sigma}^1 \eta(s) ds \right), \end{aligned}$$

$$\frac{\delta G}{\delta \eta} = \eta^T + \beta e^{-\lambda_+ s} \tilde{\sigma}^1 = 0,$$

$$\eta = -(\tilde{\sigma}^1)^T e^{-\lambda_+ s} \beta,$$

$$\beta = \frac{\tilde{\phi}_T^1}{q},$$

$$\frac{1}{2} \int_0^\infty |\eta|^2 dt = \frac{1}{2} q^{-1} (\tilde{\phi}_T^1)^2,$$

where $q \in \mathbb{R}$ is given by

$$q = \int_0^\infty e^{-\lambda_+ s} (\tilde{\sigma}^1)(\tilde{\sigma}^1)^T e^{-\lambda_+ s} ds = \frac{(\tilde{\sigma}^1)(\tilde{\sigma}^1)^T}{2\lambda_+}.$$

Assuming controllability of the linearized dynamics around the saddle, q is nonzero. Actually, we don't need complete controllability for this, because q is just the (1,1)-component of the controllability Gramian for infinite time. Hence, with a given $\tilde{\phi}_T$ and the optimal noise η , the controlled dynamics is

$$\hat{A} = \left(\begin{array}{c|c} -\lambda_+ & 0 \\ \hline -q^{-1}(\tilde{\sigma}^1)(\tilde{\sigma}^1)^T & \tilde{A}_s \end{array} \right), \quad (7)$$

where the i th entry of the first column is given by $-(\tilde{\sigma}^i)(\tilde{\sigma}^1)^T q^{-1}$ for $2 \leq i \leq n$. $\hat{A}$ is now stable. The trajectory for the unstable component takes the form

$$\tilde{\phi}^1(t) = \tilde{\phi}_T^1 e^{-\lambda_+ t}, \quad (8)$$

and the trajectories for the stable components are

$$\begin{aligned} \tilde{\phi}^i(t) &= e^{\tilde{A}_s t} \tilde{\phi}^i(0) \\ &\quad + (\tilde{A}_s + \lambda_+ I)^{-1} (e^{-\lambda_+ t} I - e^{\tilde{A}_s t}) \tilde{\sigma}^i (\tilde{\sigma}^1)^T q^{-1} \tilde{\phi}^1(0) \end{aligned} \quad (9)$$

for $2 \leq i \leq n$. These trajectories can be transformed back to the original coordinate by $\phi = U\tilde{\phi} + s$. Detailed computation for the analysis around the saddle is in the Appendix A.

C. Transitional phase

After knowing a point ϕ_0 near the sink and a point ϕ_T near the saddle, the Hamiltonian optimal control method allows us to find an optimal trajectory from ϕ_0 to ϕ_T in time T , as already described. However, as discussed before, the overall problem of traveling from the sink at time $-\infty$ to the saddle at time $+\infty$ has a time-translation degeneracy and is problematic for numerical algorithms. This time symmetry issue can be removed by requiring the trajectory at a specified time to lie on a predefined separating surface. We choose to require $\phi(T)$ to be on an ellipsoid around the saddle, transverse to the optimum flow for reaching the saddle, defined by $\hat{A}$. One could instead require $\phi(0)$ to be on a suitable ellipsoid around the sink, or $\phi(T/2)$ to be on some surface "half-way" in between the sink and saddle, but we made this choice to focus the computational method on paths that approach the saddle.

Consider an ellipsoid centered at the saddle, defined by $(\tilde{\phi}_T)^T M (\tilde{\phi}_T) - r^2 = 0$ for some positive-definite matrix M , such that the optimally controlled dynamics is transverse to the ellipsoid. For r small, this ellipsoid bounds a region around the saddle that is small enough for the accuracy of the linearization approximation. The overall optimization problem finds an optimal point on the ellipsoid to enter this domain and

the linearized motion within this domain is purely analytical. To achieve the transversality, we choose the matrix M to satisfy the Lyapunov equation $\hat{A}^T M + M \hat{A} + I_n = 0$, where I_n is the n -dimensional identity matrix, but could be any other positive-definite matrix, and $\hat{A}$ [see Eq. (7)] is the transformed linearized dynamics under control. Because $\hat{A}$ is strictly stable, the resulting M is guaranteed to be positive definite.

It is important to distinguish between the two versions of the Lyapunov equation used for the sink and saddle analyses. Careful attention is required during implementation to ensure the correct form is applied. We provide a rigorous derivation and contextual explanation for each case in Appendix B.

By restricting the endpoint $\phi(T)$ to this transverse ellipsoid, the continuous time-translation symmetry is successfully broken. Consequently, the nonlinear optimization phase over the finite interval $[0, T]$ becomes well-conditioned, removing the numerical ambiguity of path-sliding and allowing the optimal trajectory to be solved robustly. We have no guarantee of uniqueness, but we have made the determination of an optimal trajectory robust.

D. Overall problem

We now combine the three phases into one overall problem, which involves allowing ϕ_0 to vary near the sink, and ϕ_T to vary on the chosen ellipsoid around the saddle. The action is the sum of three parts for the three phases. Its optimization is carried out with respect to variations of the transition path and these endpoints. As mentioned previously, the reason for requiring ϕ_T to be on the ellipsoid is to remove a time-translation degeneracy for the optimum, which would otherwise occur.

Firstly, the optimization problem is to minimize the overall action functional:

$$S[\phi_0, \eta] = \frac{1}{2} \phi_0^T Q^{-1} \phi_0 + \frac{1}{2} \int_0^T \langle \eta, \eta \rangle dt + \frac{1}{2} q^{-1} [(U^{-1}(\phi(T) - \mathbf{s}))^1]^2,$$

where $(U^{-1}(\phi(T) - \mathbf{s}))^1$ represents the unstable component of the trajectory's endpoint relative to the saddle $\mathbf{s}$ in the transformed coordinate system. Within this formulation, the endpoint $\phi(T)$ is uniquely determined by the choice of the initial state ϕ_0 and the control path η . Specifically, by enforcing the initial condition constraint:

$$\phi_0 = \phi(0), \quad (10)$$

and integrating the forward equation of the Hamiltonian method:

$$\dot{\phi} = b(\phi) + \sigma \eta, \quad (11)$$

over the time horizon $t \in [0, T]$, the entire trajectory and thus its terminal state $\phi(T)$ is fully specified.

The three terms in the objective function correspond to the action accumulated across the three distinct stages defined by the method of division, where only the middle term is explicitly expressed as a time integral. The first term accounts for the initial escape state near the sink, entirely characterized by the optimized starting point ϕ_0 . The third term represents the remaining cost during the final approach state, where only

the unstable component of the endpoint $\phi(T)$ is relevant. In the context of the large deviation principle, the minimized value of $S[\phi_0, \eta]$ dictates the exponential scaling law of the transition rate.

In addition to the initial condition and system dynamics, a terminal constraint is imposed on the endpoint $\phi(T)$, forcing it to lie on a predefined ellipsoid surrounding the saddle point:

$$r^2 = \tilde{\phi}(T)^T M \tilde{\phi}(T), \quad (12)$$

where $\tilde{\phi}(T) = U^{-1}(\phi(T) - \mathbf{s})$ is the endpoint expressed in the transformed coordinates. The optimization problem thus reduces to finding the minimum action value by varying ϕ_0 and η subject to the three constraints from Eq. (10) to Eq. (12).

Algorithm and numerical solutions

To solve this optimization problem, we apply the augmented Lagrangian method. The objective (cost) function is

$$\begin{aligned} J[\phi, \eta, \mu, \nu, \beta, \lambda] = & \frac{1}{2} \phi_0^T Q^{-1} \phi_0 + \frac{1}{2} \int_0^T \langle \eta, \eta \rangle dt \\ & + \frac{1}{2} q^{-1} (\tilde{\phi}(T)^1)^2 + \langle \nu, \phi_0 - \phi(0) \rangle \\ & + \int_0^T \langle \mu, \dot{\phi} - b(\phi) - \sigma \eta \rangle dt \\ & + \beta (\tilde{\phi}(T)^T M \tilde{\phi}(T) - r^2) \\ & + \lambda (\tilde{\phi}(T)^T M \tilde{\phi}(T) - r^2)^2. \end{aligned}$$

The parameters $\nu \in \mathbb{R}^n$, $\mu \in \mathbb{R}^n$, $\beta \in \mathbb{R}$ are the Lagrange multipliers and $\lambda \in \mathbb{R}$ is a penalty parameter. Note that $\phi(0)$ and $\phi(T)$ correspond to the initial point and endpoint of the trajectory in process 2. For convenience, we can substitute $\phi(0) = \phi_1$ and $\phi(T) = \phi_N$ when discretizing the trajectory ϕ into N points. A detailed discretization scheme is given in the Appendix D.

Varying the discretized cost function J with respect to the variables ϕ_0 , η , and ϕ gives

$$\begin{aligned} \frac{\partial J}{\partial \phi_0} &= Q^{-1} \phi_0 + \nu, \\ \frac{\partial J}{\partial \eta_i} &= \eta_i - (\mu^T \sigma)_i, \\ \frac{\partial J}{\partial \phi_1} &= -\nu - \mu_1 - \mu_1 \nabla b(\phi_1)^T \Delta t, \\ \frac{\partial J}{\partial \phi_i} &= \mu_{i-1} - \mu_i - \mu_i \nabla b(\phi_i)^T \Delta t, \\ \frac{\partial J}{\partial \phi_N} &= \mu_{N-1} + \frac{(l^T (\phi_N - \mathbf{s})) l}{q} + 2(U^{-T} M U^{-1})(\phi_N - \mathbf{s}) \beta \\ &\quad + 4\lambda(U^{-T} M U^{-1})(\phi_N - \mathbf{s})(\tilde{\phi}_N^T M \tilde{\phi}_N - r^2). \end{aligned}$$

This follows a very similar idea as discussed in the Hamiltonian optimal control method: the constraint and the penalty term on the ellipsoid together provide the boundary condition for a discretization of the backward Hamiltonian equation (3b) on μ . The optimized noise η can be found via gradient descent in the direction $\eta - \sigma^T \mu$ until $\eta = \sigma^T \mu$. The backward

equation on μ further gives a condition on ν so that ϕ_0 can be optimized. Hence, we have two variables to optimize: ϕ_0 and η where η is a function of time $t \in [0, T]$. The trajectory ϕ can be computed from ϕ_0 , knowing η , through the forward Hamilton's equation (3a), and the endpoint ϕ_N is forced on the ellipsoid that gives the smallest action. The algorithm is as follows:

(1) Initialize ϕ_0 , η , the Lagrange multiplier β and the penalty parameter λ . Choose convergence thresholds τ_{grad} and τ_β .

(2) For fixed β and λ , update ϕ_0 and η :

(a) Solve the forward Hamiltonian dynamics

$$\dot{\phi} = b(\phi) + \sigma\eta, \quad \phi(0) = \phi_0,$$

on $[0, T]$ to obtain $\phi(T)$.

(b) Solve the backward adjoint dynamics

$$\dot{\mu} = -\nabla b(\phi)^\top \mu,$$

with terminal condition

$$\mu_{N-1} = -\frac{(l^\top(\phi_N - s))l}{q} - 2(U^{-T}MU^{-1})(\phi_N - s)(\beta + 2\lambda(\tilde{\phi}_N^\top M \tilde{\phi}_N - r^2)),$$

to obtain μ and ν with

$$\nu = -\mu_1 - \mu_1 \nabla b(\phi_1)^\top \Delta t.$$

(c) Evaluate the gradients of the augmented cost functional with respect to the optimization arguments:

$$\nabla_\eta J = \eta - \sigma^\top \mu$$

$$\nabla_{\phi_0} J = Q\phi_0 + \nu,$$

and update η and ϕ_0 using a chosen optimization algorithm.

(d) Repeat steps (a)–(c) until the norm of the overall gradient is strictly smaller than τ_{grad} .

(3) Update the penalty parameter and the Lagrange multiplier with some chosen constant $1 < c \ll 2$:

$$\beta_{k+1} = \beta_k + \lambda_k(\tilde{\phi}^\top(T)M\tilde{\phi}(T) - r^2), \quad \lambda_{k+1} = c\lambda_k.$$

Repeat steps 2 and 3 until $|\beta_{k+1} - \beta_k| < \tau_\beta$ such that the constraint is satisfied, i.e., $\lambda_k(\tilde{\phi}^\top(T)M\tilde{\phi}(T) - r^2) \approx 0$. Output: Optimized η and ϕ_0 .

Note that for step 2(c) we utilized the nonlinear conjugate gradient method, as it exhibits significantly faster convergence for this path-optimization problem. A detailed discussion on the inner product definition and the resulting gradient norm is provided in Appendix E.

E. Linearization robustness

In this section, we analyze the linearization error from the initial escape phase and the final approach phase. It can be skipped on a first reading. We turn on the nonlinear part of the vector field:

$$b(\phi) = A\phi + \delta b(\phi),$$

where A is the linearized dynamics that could be either $A_{-\infty}$ or A_{∞} . We write the perturbed trajectory as $\phi = \bar{\phi} + \delta\phi$ and the perturbed forcing as $\eta = \bar{\eta} + \delta\eta$, where $\bar{\cdot}$ represents the

zeroth order term and $\delta\cdot$ is the first-order correction. The optimization problem becomes to minimize the action

$$S = \frac{1}{2} \int (\bar{\eta} + \delta\eta)^\top (\bar{\eta} + \delta\eta) dt,$$

subject to the constraints

$$\dot{\phi} = \dot{\bar{\phi}} + \delta\dot{\phi} = A(\bar{\phi} + \delta\phi) + \delta b(\bar{\phi} + \delta\phi) + \sigma(\bar{\eta} + \delta\eta)$$

and boundary conditions at the ends of the trajectory. Following a similar calculation via the Hamiltonian optimal control framework introduced earlier, where μ acts as the Lagrange multiplier, the first-order equations for the perturbations (taking the sink as an example; see Appendix C for details) are given by

$$\delta\dot{\phi} = A\delta\phi + \delta b(\bar{\phi}) + C\delta\mu,$$

$$\delta\dot{\mu} = -A^\top \delta\mu - \nabla \delta b^\top \bar{\mu},$$

with boundary condition $\delta\phi(-\infty) = \mathbf{0}$ and $\delta\phi(0) = 0$ such that $\bar{\phi}(0) + \delta\phi(0) = \phi_0$ is the boundary constraint. Similarly to the zeroth order case, we can forward integrate the $\delta\phi$ equation and use $\delta\phi(0) = 0$ as the boundary condition to backward integrate the $\delta\mu$ equation.

After solving the $\delta\mu$ equation, we can express the first-order change in the action as

$$\begin{aligned} \delta S &= \int_{-\infty}^0 \bar{\eta}^\top \delta\eta dt, \\ &= \int_{-\infty}^0 \bar{\mu}(s)^\top C\delta\mu(s) ds, \\ &= -\bar{\phi}_0^\top Q^{-1} \int_{-\infty}^0 e^{-As} \delta b(\bar{\phi}(s)) ds, \\ &= -\bar{\phi}_0^\top Q^{-1} \int_{-\infty}^0 e^{-As} \delta b(Qe^{-A^\top s} Q^{-1} \bar{\phi}_0) ds. \end{aligned}$$

This is of order $O(\bar{\phi}_0^3)$. By comparing it with the zeroth order action under linearization $\bar{S} = \frac{1}{2} \bar{\phi}_0^\top Q^{-1} \bar{\phi}_0$: the relative change is

$$\frac{\delta S}{\bar{S}} \approx \frac{2|\delta b(\bar{\phi}_0)|}{c|\bar{\phi}_0|},$$

where c is the slowest contraction rate for A . We have to check that $|\delta b(\bar{\phi}_0)| \ll c|\bar{\phi}_0|$ for a good linearization approximation.

In addition to determining an order of magnitude for the accuracy of the linear approximations, the first-order correction to the action derived here from linearization error can be added to the computed action to produce a more accurate value.

IV. EXAMPLES

A. Inverted double well

Consider an inverted double well model with potential $V(x) = \frac{1}{2}x^2 - \frac{1}{4}x^4$. The Newton's equation of motion for a particle with mass m in this potential is $m\ddot{x} = x^3 - x$. For simplicity, assume $m = 1$. To illustrate the method of division, together with our way of treating filtered noise, let us make

this a damped system and add the simplest filtered noise to the model:

$$\begin{aligned} dx &= y dt, \\ dy &= (x^3 - x - ky + z)dt, \\ dz &= (-z)dt + \sqrt{\epsilon} dW, \end{aligned}$$

where k is a positive constant denoting the damping coefficient, and z is the one-dimensional filtered noise added. White noise is only added to the last line, which gives degenerate noise for the system as a whole, i.e., the covariance matrix is singular.

The model is unrealistic in that $V(x) \rightarrow -\infty$ as $x \rightarrow \pm\infty$ and the dynamics suffers finite-time blowup, but we are interested here just in reaching one of the two saddles ($x = \pm 1$) from the sink at $x = 0$. It is easy to make more realistic models, e.g., $V(x) = a \cos x - b \cos 2x$ with $4b > a > 0$, but simpler to illustrate with a polynomial rather than a trigonometric function.

The above simultaneous equations could also be expressed in the generic SDE format as

$$d\phi = b(\phi)dt + \sigma dW,$$

with

$$\phi = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \quad b(\phi) = \begin{pmatrix} y \\ x^3 - x - ky + z \\ -z \end{pmatrix} \text{ and } \sigma = \begin{pmatrix} 0 \\ 0 \\ \sqrt{\epsilon} \end{pmatrix}.$$

The deterministic dynamics of this system have three equilibrium points: the stable equilibrium $(0,0,0)$ and the two saddles $\pm(1,0,0)$. Due to symmetry, consider only one saddle at $(1,0,0)$. The linearized dynamics ∇b around the two critical points are given by

$$A_{-\infty} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & -k & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad A_{\infty} = \begin{pmatrix} 0 & 1 & 0 \\ 2 & -k & 1 \\ 0 & 0 & 1 \end{pmatrix}. \quad (13)$$

Choose $k < 1$ such that the sink has one negative real eigenvalue and two complex eigenvalues with small negative real parts. The saddle has one positive real eigenvalue and two negative real eigenvalues.

The transition problem is to find the probability rate to escape over one of the saddles if the system starts at the origin, as well as the trajectory it would take. Figure 3 illustrates the result by the method of division. In Fig. 3(a), the optimal trajectory is shown on a background of level sets of the Hamiltonian $H(x, y) = V(x) + \frac{y^2}{2}$ for the undamped system. Note the Hamiltonian $H(x, y)$ is in the sense of mechanics and is different from the Hamiltonian of the optimal control method mentioned before. Figure 3(c) is a 3D plot of the trajectory, showing also the ellipsoid defined around the saddle. The dashed pink line is the linearized trajectory within the ellipsoid. Figures 3(b) and 3(d) show the filtered noise dimension and the corresponding Gaussian white noise. The linear part (light green) and nonlinear part (dark green) are seen to join smoothly, supporting the validity of use of the linearized approximations at the two ends. The computed action value is $S = 0.1330$.

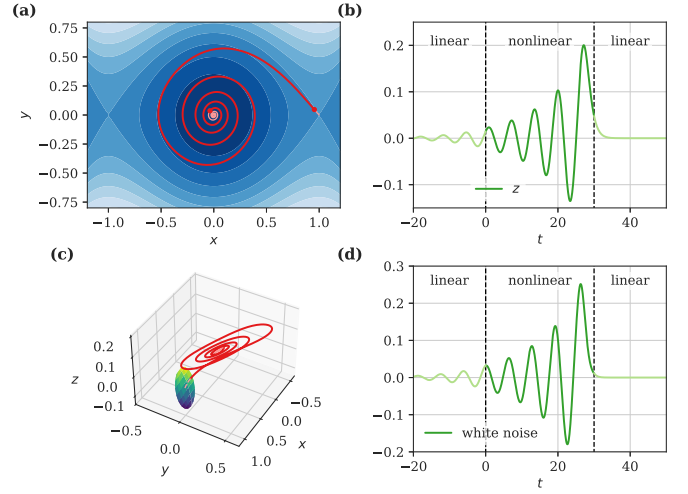


FIG. 3. Optimal trajectory from the sink to the saddle in three sections: linearized dynamics at both ends and nonlinear dynamics for the transition. The figures on the right show the optimal filtered noise and the corresponding realization of the white noise.

According to the large deviation principle (1), the probability rate scales asymptotically as $e^{-S/\epsilon}$. For instance, taking a weak disturbance with $\sqrt{\epsilon} = 0.1$, the probability rate is approximately $K \times 1.67 \times 10^{-6}$, where K denotes the prefactor in the LDP. We do not know the prefactor, but if we assume that it varies slowly with ϵ then we can deduce, for example, that if the noise amplitude $\sqrt{\epsilon}$ is increased from 0.1 to 0.2 then the transition rate increases by a factor of approximately $e^{0.133/75} \approx 21483$. In nongradient systems, the preexponential factor is not merely a constant but depends on the entire history of the dynamics along the most probable trajectory [39]. But we treat it as a constant in the example section due to the significant complexity involved in its analytical determination.

B. Ship capsize model

A more complicated example is a simple two degree of freedom ship capsize model, proposed by Thompson and Souza [6]. This archetypal capsize equation of the ship is designed for the beam sea, where waves travel perpendicular to the body of the ship. It has two degrees of freedom: The heave z and the roll θ (see Fig. 4). The ship potential is given by

$$V(\theta, z) = \frac{1}{2} c \theta_c^2 \left[\left(\frac{\theta}{\theta_c} \right)^2 - \frac{1}{2} \left(\frac{\theta}{\theta_c} \right)^4 \right] + \frac{1}{2} h \left[z - \frac{1}{2} \gamma \theta^2 \right]^2.$$

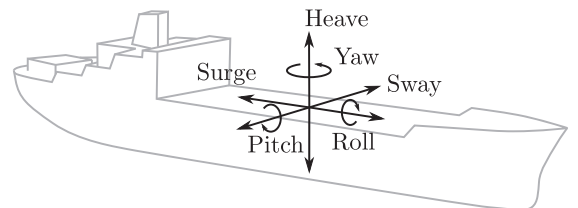


FIG. 4. Six degrees of freedom of a ship.

The roll and heave natural frequencies of the ship are proportional to $\sqrt{c}$ and $\sqrt{h}$, with coefficients $I^{-1/2}$, $m^{-1/2}$ respectively, where I is the moment of inertia for roll and m is the mass of the ship. Here, γ is a ship parameter that couples roll and heave, and θ_c denotes the static capsize angle. The restoring force in heave is $-\partial V/\partial z$ and the restoring moment in roll is $-\partial V/\partial \theta$. All variables are nondimensionalized. Then by distinguishing the momentum dimensions and the configuration dimensions and by adding damping to heave and roll, and adding filtered noise we obtain a six-dimensional stochastic dynamical system:

$$\begin{aligned}\dot{z} &= v_z, \\ \dot{v}_z &= \frac{h}{m} \left(\frac{1}{2} \gamma \theta^2 - z \right) - k_1 v_z + B_{11} \xi_1 + B_{12} \xi_2, \\ \dot{\theta} &= v_\theta, \\ \dot{v}_\theta &= \frac{c}{I} \theta \left(\frac{\theta^2}{\theta_c^2} - 1 \right) + \frac{h}{I} \gamma \theta \left(z - \frac{1}{2} \gamma \theta^2 \right) - k_2 v_\theta + B_{21} \xi_1 \\ &\quad + B_{22} \xi_2, \\ \dot{\xi}_1 &= A_{11} \xi_1 + A_{12} \xi_2 + \sqrt{\epsilon} \sigma_{11} \eta_1 + \sqrt{\epsilon} \sigma_{12} \eta_2, \\ \dot{\xi}_2 &= A_{21} \xi_1 + A_{22} \xi_2 + \sqrt{\epsilon} \sigma_{21} \eta_1 + \sqrt{\epsilon} \sigma_{22} \eta_2,\end{aligned}$$

where v_z and v_θ are the velocity coordinates. The variables ξ_i are the added filtered noise and η_i are the "derivative" of independent Wiener processes. This structure results in a system with filtered and degenerate noise, which provides a more realistic representation of ocean wave behavior. As demonstrated by Dimentberg [40], such a filtered noise approach captures parametric excitation far more accurately than a simple white noise model, as it accounts for the narrow-band spectral characteristics of the sea state.

Here we adopt a representative choice of the matrix A that determines the spectral properties of the filtered noise:

$$A = \begin{pmatrix} -\alpha & \omega \\ -\omega & -\alpha \end{pmatrix}.$$

This specific form yields complex eigenvalues $-\alpha \pm i\omega$, where ω is the characteristic oscillation frequency of the noise and α describes the decay of the noise's autocorrelation.

Similarly to the inverted double-well example, the unforced system has three equilibrium points, a stable equilibrium at the origin $\mathbf{0} \in \mathbb{R}^6$ and two saddle points at $(z, v_z, \theta, v_\theta, \xi_1, \xi_2) = (\frac{1}{2} \gamma \theta_c^2, 0, \pm \theta_c, 0, 0, 0)$ for some constant θ_c . The saddles have a one-dimensional unstable manifold and a five-dimensional stable manifold in the relevant parameter range. The structure and symmetry of the saddles fit with physical intuition: the ship loses balance and capsizes when it swings too severely to the left or the right. This instability comes only from the roll direction θ and $\pm \theta_c$ is the corresponding capsize angle.

Note that there exists a redundancy in the coefficient matrices B and σ , which can be removed by exploiting the rotation and scaling symmetries in the underlying noise spaces. Let $R \in SO(2)$ be a rotation matrix. The two-dimensional white noise η is invariant under this orthogonal rotation. Moreover, our choice of the drift matrix A commutes with such rotations (i.e., $R^{-1}AR = A$), allowing us to rotate the colored noise co-

TABLE I. Physical and stochastic parameters utilized in the ship capsize numerical simulation.

Parameter	Description	Value
m, I	Ship mass and moment of inertia	1.0, 1.0
k_1, k_2	Damping coefficients (heave and roll)	0.1, 0.1
h, c, γ, θ_c	Ship coefficients	$[0.4, 0.1, 1, \pi/4]$
B	Coupling matrix	$\begin{pmatrix} 1/\sqrt{5} & 0 \\ 0 & 2/\sqrt{5} \end{pmatrix}$
σ	Noise matrix	$\begin{pmatrix} 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} \end{pmatrix}$

ordinates ξ and the driving noise η independently. By utilizing these two rotational degrees of freedom, the coupling matrix σ can be reduced to a diagonal form with nonnegative entries. The scaling freedom of the smallness parameter ϵ further allows us to normalize these entries. Therefore, the matrix σ can be simplified to

$$\sigma = \begin{pmatrix} \cos \alpha & 0 \\ 0 & \sin \alpha \end{pmatrix},$$

with a single parameter $\alpha \in [0, \pi/2]$. Finally, the remaining scaling freedom in the transformation of ξ allows us to normalize the matrix B such that $\|B\| = 1$ under an appropriate matrix norm. These transformations effectively reduce the system to a general matrix B constrained by $\|B\| = 1$ and a single tuning angle α , thereby significantly decreasing the number of free parameters.

Figures 5 show representative results for the ship capsize problem. The specific physical parameters of the ship and noise parameters are systematically detailed in Table I, adopting the simplified choice after eliminating the redundancies. For the chosen parameter values, we have evaluated the controllability Gramian at the sink and its unstable-unstable component at the saddle as part of our numerical implementation, confirming their invertibility and the validity of the local controllability assumptions. The red curves indicate the optimal capsize trajectory—i.e., the most likely path to capsize—while the pink curves illustrate the linearized dynamics near the equilibrium points. In both cases, the linearized dynamics align well with the nonlinear dynamics at the boundaries. The blue and green curves represent noise: filtered noise applied to the heave and roll, and white noise applied to the filter coordinates.

In Fig. 5(a), the effective transition time is $T = 50$, whereas in Fig. 5(b) it is $T = 60$, with all other parameters held fixed. The action values are identical to three significant figures in both cases: $S = 0.0176$. The resulting trajectories are very similar as well, as indeed they should be; the main change as T increases is that the transition from the linearized to the nonlinear regime occurs closer to the sink. This highlights the numerical advantage of the method of division: one can identify a sufficient transition time T^* such that for any $T > T^*$, the trajectory $\phi[T]$ closely resembles $\phi[T^*]$, differing only by extending further within the linearized region near the sink. The action value also stabilizes for $T > T^*$, giving the "true" solution to the problem. This saves computation cost to a great extent when investigating the change to the trajectory for longer T . The leading corrections from the

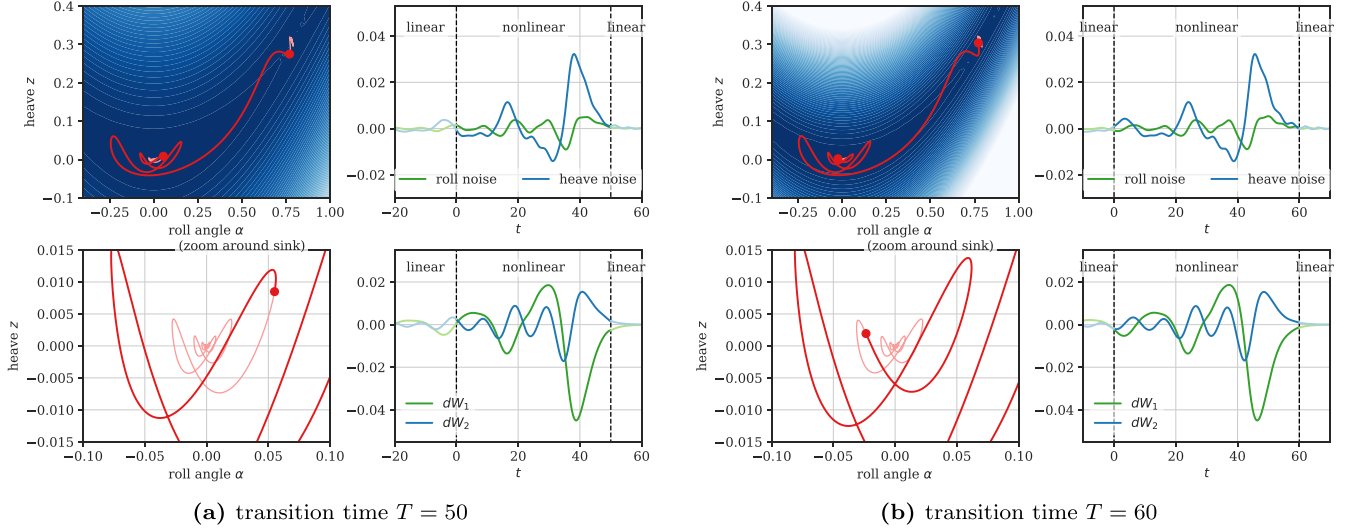


FIG. 5. Comparison between two ship capsizing trajectories and their corresponding optimal forces under different effective transition times (a) transition time $T = 50$ and (b) transition time $T = 60$. The forces on the filter variables are denoted as $\tilde{\eta}_i = \sigma_{i1}\eta_1 + \sigma_{i2}\eta_2$. Note that the two trajectories are virtually identical; changing T merely shifts the crossover point where the dynamics transitions from the linearized to the nonlinear regime.

linearization errors were found to be less than 0.1%, so it made little difference whether we added them in or not.

Parameter study

Different parameter choices lead to distinct capsizing behaviors, actions, and consequently, varying capsizing rates. In this section, we examine two primary factors influencing these outcomes: the spectral properties of the external disturbances and the characteristics of the ship itself. The external disturbances—mainly waves—are modeled using filtered noise with the characteristic oscillation frequency ω (see, for example, Fig. 1).

The transition rate is typically expressed via the LDP form $e^{-S/\epsilon}$ where the action S is evaluated relative to a constant white noise intensity ϵ . However, when the characteristic filter frequency varies, the effective intensity of the resulting colored noise also changes. Therefore, ϵ is no longer a suitable metric to characterize the size of the stochastic forcing. Instead of analyzing the frequency-dependence of capsizing rates at fixed ϵ , it is more meaningful to fix a quantification of the intensity of the forcing. For this purpose, we scale the action by a factor $\lambda_{\max}$ (see Appendix G for the definition and a detailed derivation). By comparing the normalized actions $S \cdot \lambda_{\max}$ across different wave frequencies ω , we effectively isolate the system's inherent dynamical sensitivity from frequency-dependent noise energy levels. This approach allows us to identify the resonance conditions under which the ship is most susceptible to capsize. Throughout this analysis, the heave and roll frequencies are kept fixed to isolate the spectral effects of the noise. However, to investigate the effect from different ship structures, we fix the filtered noise frequency and roll frequency while varying the ratio between heave and roll frequencies.

Figure 6(a) illustrates the variation of the reweighted action $S \cdot \lambda_{\max}$ over different wave frequencies, where the resonant heave and roll frequencies are marked by black dashed lines. Since a higher action value corresponds to a

lower capsizing probability, the local minimum observed near the heave frequency identifies the most critical frequency range for transition. In contrast, the plateau to the right of the roll frequency suggests a frequency-insensitive regime where nonlinear damping plays a dominant role. As the ship approaches roll resonance, the rapid increase in nonlinear dissipation can counter the additional energy input from waves, creating a temporary dynamic equilibrium that stabilizes the action.

The local minimum around the heave frequency indicates a specific resonance region where the ship's inherent dynamics synchronize with the incoming wave energy, leading to rapid energy accumulation and eventually capsizing. While capsizing is formally defined by the roll angle exceeding a critical threshold, our results demonstrate that resonance with heave motion is often more dangerous. This is consistent with the catastrophe model proposed by Zeeman [41], where vertical oscillations (heaving) redistribute buoyancy and induce parametric instability in the roll equilibrium.

In this analysis, the heave-roll frequency ratio is kept constant at 2, which is a typical ratio for vessels such as destroyers [41]. This configuration corresponds to a Fermi resonance regime. So it is interesting that the wave frequency for minimum weighted action coincides precisely with the heave frequency, confirming that heave oscillations can dominantly destabilize the ship and initiate severe rolling. Thus, the synchronization of external wave forcing with the heave mode represents the most susceptible state for the system.

Conversely, at higher wave frequencies (the right end of the plot), the ship's relatively longer natural period makes it less responsive to the rapid wave excitations. In this regime, the energy transfer from the waves to the roll motion becomes highly inefficient, resulting in high stability.

Figure 6(b) illustrates the action values under different ratios between heave and roll frequencies, defined as $r = \sqrt{h/c}$ (recall that the roll and heave frequencies are proportional to $\sqrt{c}$ and $\sqrt{h}$, respectively). In this experiment, the

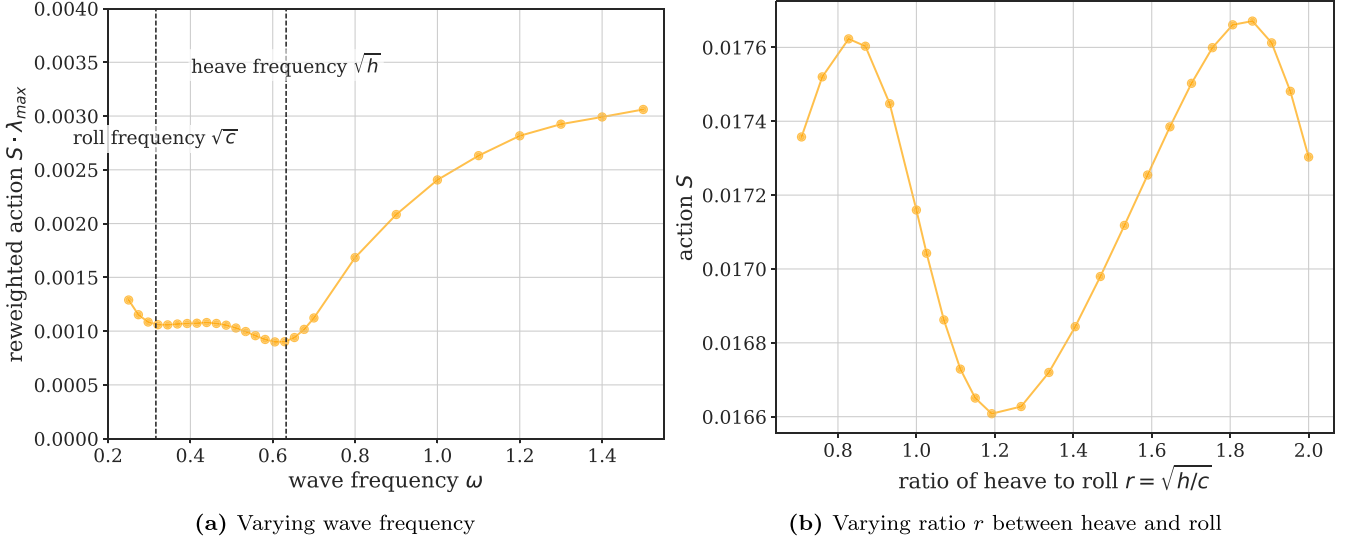


FIG. 6. Parameter studies on minimal action for escape (a) varying wave frequency ω , where the minimal action has been scaled to an appropriate quantification of the noise intensity; (b) varying ratio $r = \sqrt{h/c}$ between heave and roll natural frequencies.

characteristic frequency of filtered noise is held constant at 1.5 and the roll frequency is fixed at 1. By varying the ratio r from 0.7 to 2, we effectively span the most dynamically sensitive range, including both the simple resonance ($r = 1$) and the Fermi resonance ($r = 2$).

The action attains a minimum of approximately 0.0166 at $r \approx 1.2$, whereas it reaches its maximum value of roughly 0.0177 at $r \approx 1.85$. Since the probability rate follows the asymptotic scaling $\mathbb{P} \sim K e^{-S/\epsilon}$ where the prefactor K depends subexponentially on the system parameters and noise parameter ϵ , we can estimate the relative likelihoods by assuming K remains constant between these two cases. The ratio of probability rates for capsizing is then expressed as

$$\frac{\mathbb{P}(\text{capsize rate when } r \approx 1.2)}{\mathbb{P}(\text{capsize rate when } r \approx 1.85)} = e^{-\Delta S/\epsilon},$$

where $\Delta S \approx 0.0011$ represents the action discrepancy. Due to the exponential dependence on ϵ , small changes in the noise intensity can lead to orders-of-magnitude changes in the relative likelihood of capsizing. For instance, at $\epsilon = 10^{-2}$, the ship is approximately 1.12 times more likely to capsize at $r \approx 1.2$ than at $r = 1.85$; at $\epsilon = 10^{-3}$, this factor increases to roughly 3; and at $\epsilon = 10^{-4}$, the probability increases drastically, reaching a likelihood approximately 6×10^4 times higher. These findings are consistent with results of Ref. [6], which suggested that heave-to-roll frequency ratio near 2 reduced susceptibility to capsize, in contrast to expectations based on parametric resonance.

V. CONCLUSION

In this paper, we introduced the method of division to address transition problems in stochastic dynamical systems over an unbounded time-scale and demonstrated its effectiveness through two examples: an inverted double well and a ship capsize problem. The proposed method accommodates both degenerate and filtered noise, making it more adaptable to real-world scenarios. It also enables an infinite timescale

by focusing computational effort on the nonlinear transitional phase. We illustrated the method with a transition from a stable sink to a saddle with a one-dimensional unstable manifold and discussed its broader applicability. This approach can be generalized to other disciplines with different stability structures, as transition problems arise across a wide range of fields.

Looking forward, several promising directions warrant further investigation. Firstly, the current geometric framework can be generalized to accommodate alternative and more complex stability structures, such as transitions between multiple saddles (saddle-to-saddle cases) or networks embedded with heteroclinic channels. Secondly, while the current Large Deviation framework successfully determines the leading-order exponential scaling behaviors, a highly desirable and important extension would be to resolve the subexponential prefactor. Quantifying this prefactor as developed in alternative contexts [42–45] represents a key avenue to transition from scaling laws to absolute transition probabilities. Finally, it would be highly interesting to extend the method of division to systems driven by non-Gaussian noises, such as t -processes or Lévy stable processes. Such extensions, which feature heavy tails and are characterized by more than two cumulants, would significantly broaden the applicability of our computational framework to a wider class of non-Gaussian physical and engineering systems.

ACKNOWLEDGMENTS

We are grateful to the reviewers and R. Deeley for useful comments. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission. J.S. is supported by the Warwick Mathematics Institute Centre for Doctoral Training, and gratefully acknowledges funding from the University of Warwick. T.G. acknowledges the support received from the EPSRC Projects No. EP/T011866/1 and No. EP/V013319/1.

DATA AVAILABILITY

The data and computational codes that support the findings of this study are available on GitHub at [46].

APPENDIX A: SADDLE ANALYSIS

The transformed linearized dynamics around the saddle are given by

$$\begin{aligned}\dot{\tilde{\phi}}^1 &= \lambda_+ \tilde{\phi}^1 + \tilde{\sigma}^1 \eta, \\ \dot{\tilde{\phi}}^i &= (\tilde{A}_s \tilde{\phi})^i + \tilde{\sigma}^i \eta \quad \text{for } 2 \leq i \leq n,\end{aligned}$$

where the first component is the unstable direction. We aim to find an optimal η such that for $0 \leq t \leq \infty$, the dynamics satisfies $\tilde{\phi}^1(0) = \tilde{\phi}_T^1$ and $\tilde{\phi}^1(\infty) = 0$. By Duhamel's principle

$$e^{-\lambda_+ t} \tilde{\phi}^1(t) = \tilde{\phi}^1(0) + \int_0^t e^{-\lambda_+ s} \tilde{\sigma}^1 \eta(s) ds.$$

Taking the limit $t \rightarrow \infty$ on both sides gives $\tilde{\phi}^1(0) = -\int_0^\infty e^{-\lambda_+ s} \tilde{\sigma}^1 \eta(s) ds$ since $\lim_{t \rightarrow \infty} e^{-\lambda_+ t} \tilde{\phi}^1(t) = 0$. Now, similar to the analysis for the sink, using the Lagrangian multiplier method:

$$\begin{aligned}G[\tilde{\phi}_T^1, \eta, \beta] &= \frac{1}{2} \int_0^\infty |\eta(s)|^2 ds + \beta (\tilde{\phi}_T^1 - \tilde{\phi}^1(0)) \\ &= \frac{1}{2} \int_0^\infty |\eta(s)|^2 ds \\ &\quad + \beta \left(\tilde{\phi}_T^1 + \int_0^\infty e^{-\lambda_+ s} \tilde{\sigma}^1 \eta(s) ds \right), \\ \frac{\delta G}{\delta \eta} &= \eta^T + \beta e^{-\lambda_+ s} \tilde{\sigma}^1 = 0, \\ \eta &= -\beta e^{-\lambda_+ s} (\tilde{\sigma}^1)^T.\end{aligned}$$

Then the action around the saddle is

$$\begin{aligned}\frac{1}{2} \int_0^\infty |\eta|^2 dt &= \frac{\beta^2}{2} \int_0^\infty e^{-\lambda_+ s} (\tilde{\sigma}^1)(\tilde{\sigma}^1)^T e^{-\lambda_+ s} ds \\ &= \frac{\beta^2}{2} q,\end{aligned}$$

where $q = \frac{(\tilde{\sigma}^1)(\tilde{\sigma}^1)^T}{2\lambda_+}$.

As under the optimal η , we have $\tilde{\phi}^1(0) = \tilde{\phi}_T^1$ with

$$\begin{aligned}\tilde{\phi}_T^1 &= \int_0^\infty e^{-\lambda_+ s} \tilde{\sigma}^1 (\tilde{\sigma}^1)^T e^{-\lambda_+ s} \beta ds \\ &= q\beta.\end{aligned}$$

Then $\beta = q^{-1} \tilde{\phi}_T^1$ and the action takes the form $\frac{1}{2} q^{-1} (\tilde{\phi}_T^1)^2$. Moreover, the dynamics of the unstable component at time t can be explicitly written as

$$\begin{aligned}\tilde{\phi}^1(t) &= - \int_t^\infty e^{\lambda_+(t-s)} \tilde{\sigma}^1 \eta ds \\ &= \int_t^\infty e^{\lambda_+(t-s)} \tilde{\sigma}^1 (\tilde{\sigma}^1)^T e^{-\lambda_+ s} \beta ds \\ &= e^{-\lambda_+ t} \int_0^\infty e^{-\lambda_+ \tau} \tilde{\sigma}^1 (\tilde{\sigma}^1)^T e^{-\lambda_+ \tau} \beta d\tau = \tilde{\phi}_T^1 e^{-\lambda_+ t},\end{aligned}$$

where we have substituted $s = t + \tau$ to shift the limit of the integral. This optimized η fully determines the trajectory for the stable components in the following manner, for $2 \leq i \leq n$:

$$\begin{aligned}\dot{\tilde{\phi}}^i &= (\tilde{A}_s \tilde{\phi})^i + \tilde{\sigma}^i \eta \\ &= (\tilde{A}_s \tilde{\phi})^i - \tilde{\sigma}^i (\tilde{\sigma}^1)^T e^{-\lambda_+ s} \tilde{\phi}_T^1 q^{-1} \\ &= (\tilde{A}_s \tilde{\phi})^i - \tilde{\sigma}^i (\tilde{\sigma}^1)^T q^{-1} \tilde{\phi}^1(t).\end{aligned}$$

Hence the linearized dynamics under control is given by

$$\hat{A} = \left(\begin{array}{c|c} -\lambda_+ & 0 \\ \hline -q^{-1} \tilde{\sigma}^i (\tilde{\sigma}^1)^T & \tilde{A}_s \end{array} \right),$$

where the i th entry of the first column is given by $-\tilde{\sigma}^i (\tilde{\sigma}^1)^T q^{-1}$ for $2 \leq i \leq n$. The trajectories of the stable components can be derived as follows:

$$\begin{aligned}\dot{\tilde{\phi}}^i - (\tilde{A}_s \tilde{\phi})^i &= -\tilde{\sigma}^i (\tilde{\sigma}^1)^T q^{-1} \tilde{\phi}^1(t), \\ \frac{d}{dt} (e^{-\tilde{A}_s t} \tilde{\phi}^i(t)) &= -e^{-\tilde{A}_s t} \tilde{\sigma}^i (\tilde{\sigma}^1)^T q^{-1} e^{-\lambda_+ t} \tilde{\phi}^1(0), \\ e^{-\tilde{A}_s t} \tilde{\phi}^i(t) - \tilde{\phi}^i(0) &= - \int_0^t e^{-(\tilde{A}_s + \lambda_+ I)s} ds \tilde{\sigma}^i (\tilde{\sigma}^1)^T q^{-1} \tilde{\phi}^1(0), \\ \tilde{\phi}^i(t) &= e^{\tilde{A}_s t} \tilde{\phi}^i(0) + (\tilde{A}_s + \lambda_+ I)^{-1} \\ &\quad \times (e^{-\lambda_+ t} I - e^{\tilde{A}_s t}) \tilde{\sigma}^i (\tilde{\sigma}^1)^T q^{-1} \tilde{\phi}^1(0),\end{aligned}$$

assuming that $-\lambda_+$ is not an eigenvalue of $\tilde{A}_s$. In the special case when $-\lambda_+$ is an eigenvalue of $\tilde{A}_s$, the integral involves powers of t as well as exponentials; the result depends on the multiplicity of $-\lambda_+$ so we do not write it out. The point $\tilde{\phi}_T$ lies on the ellipsoid and is to be optimized for the overall problem.

APPENDIX B: LYAPUNOV EQUATIONS

In the linearization analysis of the two endpoints, we employ two distinct forms of the Lyapunov equation. To clarify their physical significance, we derive both versions below and provide intuition regarding their respective roles in dynamical systems. The two versions are defined as follows:

$$A^T P + PA + Q = 0, \quad (\text{B1})$$

$$AP + PA^T + Q = 0. \quad (\text{B2})$$

Given a linear system $\dot{x} = Ax$ and a symmetric matrix Q , Eq. (B1) determines a symmetric matrix P such that the rate of change of $x^T P x$ is $-x^T Q x$. In particular if Q is chosen positive definite then $x^T P x$ decreases. If P turns out to be positive definite, then this provides quantitative bounds on the stability of the origin.

Consider a linear dynamical system $\dot{x} = Ax$ where A is stable. By defining a Lyapunov function $V(x) = x^T P x$ for some matrix P , its time derivative along the system trajectory is

$$\begin{aligned}\dot{V}(x) &= \dot{x}^T P x + x^T P \dot{x}, \\ &= x^T (A^T P + PA) x, \\ &= -x^T Q x.\end{aligned}$$

By choosing a positive definite matrix Q , $V(x)$ acts as an energylike function that strictly decreases along the trajectory over time. Geometrically, the solution P defines the level sets of the system energy. The ellipsoid $x^T P x \leq r^2$ characterizes a forward-invariant subset from which the trajectory cannot escape, with the flow always directed inward. In our saddle analysis, we apply this form to the optimally controlled dynamics $\dot{\phi} = A\phi$. By setting $Q = I$ (the identity matrix), we obtain a matrix $P = M$ that defines a forward-invariant ellipsoid $\phi^T M \phi \leq r^2$.

The version (B2) describes the steady-state covariance of a linear stable stochastic dynamical system. Consider a linear system driven by white noise: $\dot{x} = Ax + \sigma\eta$, where η is a zero-mean Gaussian white noise process and A is stable. The steady-state covariance matrix P can be expressed as the integral:

$$P = \int_{-\infty}^0 e^{-As} \sigma \sigma^T e^{-A^T s} ds, \\ = \int_{-\infty}^0 f(s) ds.$$

Let $f(s) = e^{-As} \sigma \sigma^T e^{-A^T s}$ be the integrand. Consider the time derivative of $f(s)$:

$$\frac{d}{ds} f(s) = -A e^{-As} \sigma \sigma^T e^{-A^T s} + e^{-As} \sigma \sigma^T e^{-A^T s} (-A^T), \\ = -A f(s) - f(s) A^T.$$

Then

$$\int_{-\infty}^0 \frac{d}{ds} f(s) ds = \int_{-\infty}^0 [-A f(s) - f(s) A^T] ds, \\ [e^{-As} \sigma \sigma^T e^{-A^T s}]_{-\infty}^0 = -A \int_{-\infty}^0 f(s) ds - \int_{-\infty}^0 f(s) ds A^T, \\ \sigma \sigma^T - 0 = -AP - PA^T,$$

by assuming A stable. Rearranging gives the Lyapunov equation $AP + PA^T + \sigma \sigma^T = 0$ with $Q = \sigma \sigma^T$. In our sink analysis, this form is used to characterize the stationary distribution of the stochastic process. Given the noise intensity (energy input) $Q = \sigma \sigma^T$, the solution P describes the invariant measure of the system state under persistent fluctuations.

APPENDIX C: LINEARIZATION ERROR ESTIMATION

The linearization error can be estimated by representing the nonlinear dynamics as a small correction to the linearized dynamics and treating it to first order:

$$b(\phi) = A\phi + \delta b(\phi).$$

Then,

$$\phi = \bar{\phi} + \delta\phi, \\ \eta = \bar{\eta} + \delta\eta.$$

We use $\bar{\cdot}$ on top of a variable to represent the zeroth order term and δ to the left of a variable to represent the first-order correction. The dynamics becomes

$$\dot{\phi} = \dot{\bar{\phi}} + \delta\dot{\phi} \\ = A(\bar{\phi} + \delta\phi) + \delta b(\bar{\phi} + \delta\phi) + \sigma(\bar{\eta} + \delta\eta).$$

Rearranging via different orders gives

$$\text{zero-th order : } \dot{\bar{\phi}} = A\bar{\phi} + \sigma\bar{\eta},$$

$$\text{first order : } \delta\dot{\phi} = A\delta\phi + \delta b(\bar{\phi}) + \sigma\delta\eta$$

and higher-order terms.

We treat the case of a sink, so we are interested in the time interval from $-\infty$ to 0. The perturbed objective function J is

$$\frac{1}{2} \int_{-\infty}^0 (\bar{\eta} + \delta\eta)^T (\bar{\eta} + \delta\eta) dt \\ + \int_{-\infty}^0 \langle (\bar{\mu} + \delta\mu), (\dot{\bar{\phi}} + \delta\dot{\phi}) - A(\bar{\phi} + \delta\phi) - \delta b(\bar{\phi} + \delta\phi) \\ - \sigma(\bar{\eta} + \delta\eta) \rangle dt + \langle v, \bar{\phi}(0) + \delta\phi(0) - \phi_0 \rangle.$$

Consider only the first order since the zeroth order is just the original problem:

$$\delta J = \int_{-\infty}^0 \bar{\eta}^T \delta\eta dt + \int_{-\infty}^0 \delta\mu^T \dot{\bar{\phi}} - \delta\mu^T A\bar{\phi} - \delta\mu^T \sigma\bar{\eta} dt \\ \times \int_{-\infty}^0 \bar{\mu}^T \delta\dot{\phi} - \bar{\mu}^T A\delta\phi - \bar{\mu}^T \delta b(\bar{\phi}) - \bar{\mu}^T \sigma\delta\eta dt \\ = \int_{-\infty}^0 \bar{\eta}^T \delta\eta - \int_{-\infty}^0 \dot{\bar{\mu}}^T \delta\phi dt \\ - \int_{-\infty}^0 (\delta\mu^T A\bar{\phi} + \delta\mu^T \sigma\bar{\eta}) dt + \int_{-\infty}^0 \delta\dot{\mu}^T \bar{\phi} dt \\ - \int_{-\infty}^0 (\bar{\mu}^T A\delta\phi + \bar{\mu}^T \delta b(\bar{\phi}) + \bar{\mu}^T \sigma\delta\eta) dt.$$

Taking variations,

$$\frac{\delta J}{\delta \bar{\phi}} = -\bar{\mu}^T \delta \nabla b - \delta\dot{\mu}^T - \delta\mu^T A, \\ \frac{\delta J}{\delta \bar{\eta}} = \delta\eta - \delta\mu^T \sigma.$$

That is, at the optimum,

$$\delta\dot{\mu} = -A^T \delta\mu - \nabla \delta b^T \bar{\mu}, \\ \delta\eta = \sigma^T \delta\mu.$$

The first-order action is then

$$\int_{-\infty}^0 \bar{\eta}^T \delta\eta dt = \int_{-\infty}^0 \bar{\mu}^T \sigma \sigma^T \delta\mu dt = \int_{-\infty}^0 \bar{\mu}^T C \delta\mu dt.$$

Also, the optimal first-order trajectory satisfies

$$\delta\dot{\phi} = A\delta\phi + \delta b(\bar{\phi}) + C\delta\mu,$$

where $C = \sigma \sigma^T$. For this forward equation, we know that $\delta\phi(-\infty) = 0$ and we want $\delta\phi(0) = 0$ so that the trajectory ends exactly at ϕ_0 with no correction term. Then by Duhamel's principle,

$$\delta\phi(0) = \int_{-\infty}^0 e^{-As} (\delta b(\bar{\phi}) + C\delta\mu) ds = 0, \quad (C1)$$

which gives

$$\int_{-\infty}^0 e^{-As} \delta b(\bar{\phi}) ds = - \int_{-\infty}^0 e^{-As} C \delta\mu ds.$$

Note that at zeroth order under optimized condition,

$$\begin{aligned}\dot{\bar{\phi}} &= A\bar{\phi} + C\bar{\mu}, \\ \dot{\bar{\mu}} &= -A^T \bar{\mu}.\end{aligned}$$

The forward ϕ -equation terminates at $\bar{\phi}(0)$ which provides the boundary condition for the backward μ -equation at $\bar{\mu}(0)$:

$$\begin{aligned}\bar{\phi}(0) &= \int_{-\infty}^0 e^{-As} C \bar{\mu}(s) ds \\ &= \int_{-\infty}^0 e^{-As} C e^{-A^T s} \bar{\mu}(0) ds = Q \bar{\mu}(0),\end{aligned}$$

where $Q = \int_{-\infty}^0 e^{-As} C e^{-A^T s} ds$ (this just reproduces the same Q as in the sink analysis section). Thus the boundary condition for $\bar{\mu}$ at $t = 0$ is

$$\bar{\mu}(0) = Q^{-1} \phi_0.$$

The first-order action can then be reformulated as

$$\begin{aligned}& \int_{-\infty}^0 \bar{\mu}(s)^T C \delta \mu(s) ds \\ &= \int_{-\infty}^0 \bar{\mu}(0)^T e^{-As} C \delta \mu(s) ds \\ &= \phi_0^T Q^{-1} \int_{-\infty}^0 e^{-As} C \delta \mu(s) ds \\ &= -\phi_0^T Q^{-1} \int_{-\infty}^0 e^{-As} \delta b(Q e^{-A^T s} Q^{-1} \phi_0) ds,\end{aligned}$$

using $(\star)$ and then $\bar{\phi}(s) = Q e^{-A^T s} Q^{-1} \phi_0$. This action correction is of $O(\phi_0^3)$. One can compare it with the zeroth order action under linearization $\frac{1}{2} \bar{\phi}_0^T Q^{-1} \phi_0$ and obtain an order of magnitude estimate for the integral:

$$\frac{\delta S}{S} \approx \frac{2|\delta b(\phi_0)|}{c|\phi_0|},$$

where c is the slowest contraction rate for A .

The above derivation is applicable to generic cases like the sink analysis in our method. It is similar for the saddle analysis but coordinate transformation should be taken into consideration (assume the saddle is at the origin):

$$\begin{aligned}U^{-1} \phi &= U^{-1} \bar{\phi} + U^{-1} \delta \phi, \\ \tilde{\phi} &= \bar{\phi} + \delta \phi,\end{aligned}$$

where U is the coordinate transformation matrix defined in Eq. (6) and the dynamics is

$$\begin{aligned}\tilde{b}(\tilde{\phi}) &= U^{-1} b(U \tilde{\phi}) = \tilde{A} \tilde{\phi} + \tilde{\delta} b(\tilde{\phi}), \\ \tilde{\delta} b &= U^{-1} \delta b(U \tilde{\phi}),\end{aligned}$$

where $\tilde{A}$ is in block form with the unstable component separated from the stable components. The first-order action in transformed coordinates is then

$$-\frac{\tilde{\phi}_T^1}{q} \int_0^\infty e^{-\lambda+s} \tilde{\delta} b^1(\tilde{\phi}(s)) ds,$$

where the dynamics of $\tilde{\phi}(s)$ is derived in Appendix A.

APPENDIX D: DISCRETIZED GRADIENT

The objective cost function J in its discrete version (using forward Euler method to discretize the integrals) takes the form

$$\begin{aligned}J[\phi, \eta, \mu, v, \beta, \lambda] &= \frac{1}{2} \sum_{i,j=1}^n \phi_0^i Q_{ij}^{-1} \phi_0^j + \frac{1}{2} \sum_{k=1}^{N-1} \sum_i (\eta_k^i)^2 \Delta t \\ &+ \frac{1}{2} q^{-1} (\tilde{\phi}(T)^1)^2 + \sum_{i=1}^n v_i (\phi_0^i - \phi(0)^i) \\ &+ \sum_{k=1}^{N-1} \sum_{i,j=1}^n \mu_k^i (\phi_k^i - b(\phi_k^i) - \sigma_{ij} \eta_k^j) \Delta t \\ &+ \beta \left(\sum_{i,j=1}^n \tilde{\phi}(T)^i M_{ij} \tilde{\phi}(T)^j - r^2 \right) \\ &+ \lambda \left(\sum_{i,j=1}^n \tilde{\phi}(T)^i M_{ij} \tilde{\phi}(T)^j - r^2 \right)^2,\end{aligned}$$

where n is the dimension of the SDE so $\phi \in \mathbb{R}^n$ and the total transition time T is discretized into N points such that $\phi_1 = \phi(t=0)$ and $\phi_N = \phi(t=T)$. The Lagrange multiplier μ is correspondingly discretized into $N-1$ points. The reason that μ has one point less than ϕ will be explained later. Also note that, under the transformation, $\tilde{\phi}_N = U^{-1}(\phi_N - \mathbf{s})$ where $\mathbf{s}$ is the saddle. Then the above equation is equivalent to

$$\begin{aligned}J[\phi] &= \frac{1}{2} \sum_{i,j=1}^n \phi_0^i Q_{ij}^{-1} \phi_0^j + \frac{1}{2} \sum_{k=1}^{N-1} \sum_i (\eta_k^i)^2 \Delta t \\ &+ \frac{1}{2} q^{-1} ((\tilde{\phi}_N)^1)^2 + \sum_{i=1}^n v_i (\phi_0^i - \phi_1^i) \\ &+ \sum_{k=1}^{N-1} \sum_{i,j=1}^n \mu_k^i \left(\frac{\phi_{k+1}^i - \phi_k^i}{\Delta t} - b(\phi_k^i) - \sigma_{ij} \eta_k^j \right) \Delta t \\ &+ \beta \left(\sum_{i,j=1}^n \tilde{\phi}_N^i M_{ij} \tilde{\phi}_N^j - r^2 \right) + \lambda \left(\sum_{i,j=1}^n \tilde{\phi}_N^i M_{ij} \tilde{\phi}_N^j - r^2 \right)^2.\end{aligned}$$

Varying J with respect to ϕ_0 and η gives

$$\frac{\partial J}{\partial \phi_0^i} = \sum_{j=1}^n Q_{ij} \phi_0^j + v_i,$$

$$\frac{\partial J}{\partial \eta_k^i} = \eta_k^i - \sum_{j=1}^n \sigma_{ij} \mu_k^j,$$

where the $\dot{\phi}$ is discretized via forward Euler method. Varying J with respect to ϕ gives

$$\frac{\partial J}{\partial \phi_1^i} = -v_i - \mu_1^i - \mu_1^i \nabla b(\phi_1^i)^T \Delta t,$$

$$\frac{\partial J}{\partial \phi_k^i} = \mu_{k-1}^i - \mu_k^i - \mu_k^i \nabla b(\phi_k^i)^T \Delta t.$$

For ϕ and μ , the superscript indicates the dimension from 1 to n and the subscript indicates the time index from 0 to N . Here note that

$$\tilde{\phi}_N^T M \tilde{\phi}_N = (\phi_N - \mathbf{s})^T U^{-T} M U^{-1} (\phi_N - \mathbf{s}),$$

where $\mathbf{s}$ is the saddle, so varying J with respect to ϕ_N gives

$$\begin{aligned} \frac{\partial J}{\partial \phi_N} &= \mu_{N-1} + \frac{(l^T (\phi_N - \mathbf{s}))l}{q} + 2(U^{-T} M U^{-1})(\phi_N - \mathbf{s})\beta \\ &\quad + 4\lambda(U^{-T} M U^{-1})(\phi_N - \mathbf{s})(\tilde{\phi}_N^T M \tilde{\phi}_N - r^2). \end{aligned}$$

Thus, in optimal conditions,

$$\begin{aligned} v &= -\mu_1 - \mu_1 \nabla b(\phi_1)^T \Delta t, \\ \mu_{i-1} &= \mu_i + \mu_i \nabla b(\phi_i)^T \Delta t, \\ \mu_{N-1} &= -\frac{(l^T (\phi_N - \mathbf{s}))l}{q} \\ &\quad - 2(U^{-T} M U^{-1})(\phi_N - \mathbf{s})(\beta + 2\lambda(\tilde{\phi}_N^T M \tilde{\phi}_N - r^2)). \end{aligned}$$

These μ equations describe the backward Hamilton's equation where ϕ_N provides the boundary information for μ_{N-1} and, iteratively, μ_k is determined by μ_{k+1} and ϕ_{k+1} . Therefore, we would only obtain $N - 1$ points for μ and that μ_1 and ϕ_1 together determines v so that the gradient $\frac{\partial J}{\partial \phi_0}$ is interpretable.

Varying J with respect to μ gives

$$\frac{\partial J}{\partial \mu_k^i} = \dot{\phi}_k^i - b(\phi_k^i) - \sum_{j=1}^n \sigma_{ij} \eta_k^j,$$

that is the forward Hamilton's equation when $\frac{\partial J}{\partial \mu_k^i} = 0$. We can interpret $\dot{\phi}$ by finite difference such that

$$\phi_{k+1}^i = \sum_{j=1}^n (b(\phi_k^i) + \sigma_{ij} \eta_k^j) \Delta t + \phi_k^i.$$

APPENDIX E: INNER PRODUCT FOR THE GRADIENT

In general, an inner product on a vector space V is defined via a positive-definite matrix g : $\langle u, v \rangle = u^i g_{ij} v^j$ for $u, v \in V$. Consider a smooth function $f : V \rightarrow \mathbb{R}$. The gradient of f at a general point x_0 is the vector ∇f defined by

$$\begin{aligned} \langle v, \nabla f \rangle &= df(v) \quad \forall v \in V, \\ v^i g_{ij} \nabla^j f &= v^i \frac{\partial f}{\partial x^i} \quad \forall v \in V \end{aligned}$$

as $df = \frac{\partial f}{\partial x^i} dx^i$. Thus,

$$\nabla^j f = g^{ji} \frac{\partial f}{\partial x^i},$$

where g^{ji} is the inverse matrix of g_{ij} .

With our model, define an inner product on the vector space $V = (\phi_0, \eta)$ by

$$\langle (\phi_0, \eta), (\tilde{\phi}_0, \tilde{\eta}) \rangle = \phi_0 \tilde{\phi}_0 + K \sum_{i=1}^{N-1} \eta_i \tilde{\eta}_i \Delta t$$

for some constant K (for simplicity let $K = 1$). Then the inner product is defined by a diagonal matrix G with $g_{11} = 1$ and

$g_{ii} = \Delta t$ for $2 \leq i$. We have the variation dJ :

$$dJ = (Q\phi_0 + v)d\phi_0 + \sum_{i=1}^{N-1} (\eta_i - \sigma^T \mu_i) \Delta t d\eta_i.$$

Then, the gradient of J is given by

$$\begin{aligned} \nabla J &= G^{-1} dJ \\ &= (Q\phi_0 + v, \eta_1 - \sigma^T \mu_1, \dots, \eta_{N-1} - \sigma^T \mu_{N-1}). \end{aligned}$$

Therefore, the gradient of J on ϕ_0 is

$$Q\phi_0 + v,$$

and the gradient of J on η is

$$\eta - \sigma^T \mu.$$

In practice, we update η and ϕ_0 via nonlinear conjugate gradient descent [36] with Fletcher-Reeves coefficients [47]:

$$\begin{aligned} \gamma &= \frac{(\nabla J)_k^T (\nabla J)_k}{(\nabla J)_{k-1}^T (\nabla J)_{k-1}}, \\ CG_{k+1} &= -(\nabla J)_k + \gamma \cdot CG_k, \\ (\nabla J)_{k+1} &= (\nabla J)_k + \alpha \cdot CG_{k+1}, \end{aligned}$$

where CG_i denote the conjugate gradient for the i th iteration and α is the step size determined by line search.

APPENDIX F: LINEARIZED SHIP DYNAMICS

The linearized ship dynamics is analyzed in this section. Consider only the ship components with damping, but no forcing. The linearized dynamics is given by

$$\nabla b = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\frac{h}{m} & -k_1 & \frac{hy}{m}\theta & 0 \\ 0 & 0 & 0 & 1 \\ \frac{hy}{I}\theta & 0 & (\frac{c}{I\theta_c^2} - \frac{hy^2}{2I})3\theta^2 + (\frac{hy}{I}z - \frac{c}{I}) & -k_2 \end{pmatrix}.$$

Then the linearized dynamics around the sink $\mathbf{0} \in \mathbb{R}^4$ is

$$A_{-\infty} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\frac{h}{m} & -k_1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -\frac{c}{I} & -k_2 \end{pmatrix}.$$

The eigenvalues are

$$\begin{aligned} \lambda_1 &= \frac{-k_1 + \sqrt{k_1^2 - \frac{4h}{m}}}{2}, & \lambda_2 &= \frac{-k_1 - \sqrt{k_1^2 - \frac{4h}{m}}}{2}, \\ \lambda_3 &= \frac{-k_2 + \sqrt{k_2^2 - \frac{4c}{I}}}{2}, & \lambda_4 &= \frac{-k_2 - \sqrt{k_2^2 - \frac{4c}{I}}}{2}. \end{aligned}$$

If there is no damping, i.e., $k_1 = k_2 = 0$, then we would have completely imaginary eigenvalues $\pm \sqrt{\frac{h}{m}l}$ and $\pm \sqrt{\frac{c}{I}l}$. These correspond to the natural oscillation frequencies of the heave and the roll. We have damped oscillatory behavior around the sink if $0 < k_1 < \sqrt{\frac{4hl}{m}}$ and $0 < k_2 < \sqrt{\frac{4cl}{I}}$.

The linearized dynamics around the saddle $(\frac{1}{2}\gamma\theta_c^2, 0, \pm\theta_c, 0)$ are given by

$$A_\infty = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\frac{h}{m} & -k_1 & \frac{h\gamma\theta_c}{m} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{h\gamma\theta_c}{I} & 0 & (\frac{c}{I\theta_c^2} - \frac{h\gamma^2}{2I})3\theta_c^2 - \frac{c}{I} & -k_2 \end{pmatrix}.$$

This equilibrium point has a pair of complex eigenvalues with negative real part and a pair of real eigenvalues with one positive and one negative. The saddle has a one-dimensional unstable manifold and three-dimensional stable manifold.

APPENDIX G: NOISE INTENSITY

We wish to quantify the intensity of random forcing functions in a way that allows sensible comparison of the effects on a system as the probability distribution of the forcing is changed. For example, how does the capsizing rate vary as the characteristics of a filtered white noise are varied? The variance of the white noise is not necessarily the appropriate quantification of the filtered noise intensity.

First, choose an inner product on the state space so that the size of a force ζ at a given time is quantified by $\sqrt{\zeta^T K^T K \zeta}$ with K invertible (this allows, for example, to compare the effects of different components of the force, some of which might be linear forces, others torques). In our ship problem we have already nondimensionalized the ship variables so we just take K to be the identity matrix.

For a white noise forcing $\sqrt{\epsilon}\sigma\eta$, with η a standard multidimensional white noise, a suitable quantification of its intensity is then the ℓ_2 -norm of $\epsilon K\sigma\sigma^T K^T$. The matrix $\epsilon\sigma\sigma^T$ can be interpreted as the integrated autocorrelation function of the forcing, which is δ -correlated in this white case. This suggests the following quantification of the intensity of a general random forcing function.

Definition: The *intensity* $\mathcal{I}_\zeta$ of a random forcing function ζ on a system is the ℓ_2 -norm of KJK^T , where J is the integrated autocorrelation function of ζ .

Recalling that the ℓ_2 -norm of a symmetric matrix is its largest eigenvalue in absolute value, and that an integrated autocorrelation matrix J is always positive semi-definite, the intensity can alternatively be defined as the largest eigenvalue of KJK^T .

If $\zeta = B\xi$ with ξ a filtered white noise then the autocorrelation function of ξ is given in numerous texts (e.g., Caines [48], Gardiner [25]). In particular, if ξ is the solution $\xi(t) = \int_{-\infty}^t e^{A(t-s)}\sqrt{\epsilon}\sigma\eta(s)ds$ of

$$\dot{\xi} = A\xi + \sqrt{\epsilon}\sigma\eta,$$

for an asymptotically stable matrix A , then the autocorrelation function C_ξ of ξ is $C_\xi(\tau) = \Sigma e^{A^T\tau}$ for $\tau > 0$, $e^{-A\tau}\Sigma$ for $\tau < 0$, where Σ is the solution of the Lyapunov equation $A\Sigma + \Sigma A^T + \epsilon\sigma\sigma^T = 0$. It follows that its integral is

$$\int_{-\infty}^{\infty} C_\xi(\tau) d\tau = -\Sigma A^{-T} - A^{-1}\Sigma.$$

From the Lyapunov equation, this can be written as

$$\epsilon A^{-1}\sigma\sigma^T A^{-T}.$$

Then

$$J = \epsilon B A^{-1}\sigma\sigma^T A^{-T} B^T.$$

Define

$$\lambda_{\max} = \text{largest eigenvalue of } K B A^{-1}\sigma\sigma^T A^{-T} B^T K^T.$$

Then the intensity $\mathcal{I}_\zeta$ is $\epsilon\lambda_{\max}$. $\lambda_{\max}$ can equivalently be defined as the square of the largest singular value of the matrix $K B A^{-1}\sigma$.

Large deviation theory gives the capsizing probability rate as $r \asymp \exp(-S/\epsilon)$ for the minimum S of the action. We rewrite this as $\exp(-S\lambda_{\max}/\mathcal{I})$. So to make fair comparisons of the minimum action as filter parameters are varied, for equal forcing intensity, we plot the scaled action $S\lambda_{\max}$.

-
- [1] P. Collins, G. S. Ezra, and S. Wiggins, Isomerization dynamics of a buckled nanobeam, *Phys. Rev. E* **86**, 056218 (2012).
 - [2] S. Krylov and N. Dick, Dynamic stability of electrostatically actuated initially curved shallow micro beams, *Continuum Mech. Thermodyn.* **22**, 445 (2010).
 - [3] J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Applied Mathematical Sciences Vol. 42 (Springer, New York, NY, 1983).
 - [4] A. H. Nayfeh and D. T. Mook, *Nonlinear Oscillations* (Wiley, New York, NY, 1979).
 - [5] S. Naik and S. D. Ross, Geometry of escaping dynamics in nonlinear ship motion, *Commun. Nonlinear Sci. Numer. Simul.* **47**, 48 (2017).
 - [6] J. Thompson and J. Souza, Suppression of escape by resonant modal interactions: In shell vibration and heave-roll capsizing, *Proc. R. Soc. London A* **452**, 2527 (1996).
 - [7] A. McSweeney-Davis, R. S. MacKay, and S. Naik, Escape from a potential well under forcing and damping with application to ship capsizing, *Chaos Interdiscip. J. Nonlinear Sci.* **35**, 7 (2025).
 - [8] E. Wigner, The transition state method, *Trans. Faraday Soc.* **34**, 29 (1938).
 - [9] H. Kramers, Brownian motion in a field of force and the diffusion model of chemical reactions, *Physica* **7**, 284 (1940).
 - [10] M. I. Freidlin and A. D. Wentzell, *Random Perturbations of Dynamical Systems*, Grundlehren der mathematischen Wissenschaften (Springer, New York, NY, 2012).
 - [11] R. L. Kautz, Thermally induced escape: The principle of minimum available noise energy, *Phys. Rev. A* **38**, 2066 (1988).
 - [12] M. I. Rabinovich, R. Huerta, P. Varona, and V. S. Afraimovich, Transient cognitive dynamics, metastability, and decision making, *PLoS Comput. Biol.* **4**, e1000072 (2008).
 - [13] E. Simonnet, J. Rolland, and F. Bouchet, Multistability and rare spontaneous transitions in barotropic β -plane turbulence, *J. Atmos. Sci.* **78**, 1889 (2021).

- [14] T. Grafke, M. E. Cates, and E. Vanden-Eijnden, Spatiotemporal self-organization of fluctuating bacterial colonies, *Phys. Rev. Lett.* **119**, 188003 (2017).
- [15] J. Soons, T. Grafke, and H. A. Dijkstra, Most likely noise-induced tipping of the overturning circulation in a two-dimensional Boussinesq fluid model, *J. Fluid Mech.* **1009**, A53 (2025).
- [16] P. Hänggi, P. Talkner, and M. Borkovec, Reaction-rate theory: Fifty years after Kramers, *Rev. Mod. Phys.* **62**, 251 (1990).
- [17] W. E, W. Ren, and E. Vanden-Eijnden, Minimum action method for the study of rare events, *Commun. Pure Appl. Math.* **57**, 637 (2004).
- [18] T. Grafke and E. Vanden-Eijnden, Numerical computation of rare events via large deviation theory, *Chaos* **29**, 063118 (2019).
- [19] G. Dematteis, T. Grafke, M. Onorato, and E. Vanden-Eijnden, Experimental evidence of hydrodynamic instantons: The universal route to rogue waves, *Phys. Rev. X* **9**, 041057 (2019).
- [20] T. Grafke, T. Schäfer, and E. Vanden-Eijnden, Long term effects of small random perturbations on dynamical systems: Theoretical and computational tools, in *Recent Progress and Modern Challenges in Applied Mathematics, Modeling and Computational Science*, edited by R. Melnik, R. Makarov, and J. Belair (Springer, New York, NY, 2017), pp. 17–55.
- [21] M. Heymann and E. Vanden-Eijnden, The geometric minimum action method: A least action principle on the space of curves, *Commun. Pure Appl. Math.* **61**, 1052 (2008).
- [22] M. L. Bujorianu, R. S. MacKay, T. Grafke, S. Naik, and E. Boulougouris, A new stochastic framework for ship capsizing, [arXiv:2105.05965](https://arxiv.org/abs/2105.05965).
- [23] G. E. Uhlenbeck and L. S. Ornstein, On the theory of the Brownian motion, *Phys. Rev.* **36**, 823 (1930).
- [24] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (Elsevier, Amsterdam/New York, NY, 2008).
- [25] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences* (Springer, Berlin, Heidelberg, 2009).
- [26] R. J. Adler, *The Geometry of Random Fields*, Wiley Series in Probability and Mathematical Statistics (John Wiley & Sons, Chichester, UK, 1981).
- [27] S. Varadhan, Asymptotic probabilities and differential equations, *Commun. Pure Appl. Math.* **19**, 261 (1966).
- [28] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Stochastic Modelling and Applied Probability (Springer, Berlin, Heidelberg, 2009).
- [29] R. Azencott, Grandes déviations et applications, in *École d'Été de Probabilités de Saint-Flour VIII-1978*, Lecture Notes in Mathematics Vol. 774 (Springer, Berlin, Heidelberg, 1980), pp. 1–176.
- [30] J.-D. Deuschel and D. W. Stroock, *Large Deviations* (Academic Press, Boston, MA, 1989), Vol. 137.
- [31] P. C. Martin, E. D. Siggia, and H. A. Rose, Statistical dynamics of classical systems, *Phys. Rev. A* **8**, 423 (1973).
- [32] H. K. Janssen, On a Lagrangean for classical field dynamics and renormalization group calculations of dynamical critical properties, *Z. Phys. B* **23**, 377 (1976).
- [33] C. De Dominicis, Techniques de renormalization de la théorie des champs et dynamique des phénomènes critiques, *J. Phys. Colloques* **37**, C1-247 (1976).
- [34] R. Graham and T. Tél, Weak-noise limit of Fokker-Planck models and nonequilibrium potentials, *Phys. Rev. A* **31**, 1109 (1985).
- [35] M. R. Hestenes, Multiplier and gradient methods, *J. Optim. Theory Appl.* **4**, 303 (1969).
- [36] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed., Springer Series in Operations Research and Financial Engineering (Springer, New York, NY, 2006).
- [37] J. Zabczyk, *Mathematical Control Theory: An Introduction*, Systems & Control: Foundations & Applications (Springer International Publishing, Cham, 2020).
- [38] S. Boyd, Stanford University, lecture notes in EE363 (2008), <https://ee363.stanford.edu/archive/lectures/lyap.pdf>.
- [39] R. S. Maier and D. L. Stein, Escape problem for irreversible systems, *Phys. Rev. E* **48**, 931 (1993).
- [40] M. F. Dimentberg, *Statistical Dynamics of Nonlinear and Time-Varying Systems*, Mechanical Engineering Research Studies Vol. 1 (Research Studies Press, Taunton, Somerset, UK, 1988).
- [41] E. C. Zeeman, A catastrophe model for the stability of ships, in *Proceedings of the 10th Brazilian Mathematical Colloquium (Poços de Caldas, 1975)* (Inst. Mat. Pura Apl., Rio de Janeiro, 1977), Vol. 2, pp. 553–580.
- [42] F. Bouchet and J. Reygner, Generalization of the Eyring-Kramers transition rate formula to irreversible diffusion processes, *Ann. Henri Poincaré* **17**, 3499 (2016).
- [43] T. Grafke, T. Schaefer, and E. Vanden-Eijnden, Sharp asymptotic estimates for expectations, probabilities, and mean first passage times in stochastic systems with small noise, *Commun. Pure Appl. Math.* **77**, 2268 (2024).
- [44] T. Schorlepp, S. Tong, T. Grafke, and G. Stadler, Scalable methods for computing sharp extreme event probabilities in infinite-dimensional stochastic systems, *Stat. Comput.* **33**, 137 (2023).
- [45] E. R. Heller and D. T. Limmer, Evaluation of transition rates from nonequilibrium instantons, *Phys. Rev. Res.* **6**, 043110 (2024).
- [46] J. Shao, T. Grafke, and R. S. MacKay, Data and code for “Escape over a saddle by colored noise Theory and numerics”, GitHub repository, 2026, <https://github.com/JiayaoShao/Escape-over-a-saddle>.
- [47] R. Fletcher and C. M. Reeves, Function minimization by conjugate gradients, *Comput. J.* **7**, 149 (1964).
- [48] P. S. Caines, *Linear Stochastic Systems* (John Wiley & Sons, New York, NY, 1988).