

## Adaptive minimum action method for the study of rare events

Xiang Zhou, Weiqing Ren, and Weinan E

Citation: *The Journal of Chemical Physics* **128**, 104111 (2008); doi: 10.1063/1.2830717

View online: <http://dx.doi.org/10.1063/1.2830717>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/jcp/128/10?ver=pdfcov>

Published by the [AIP Publishing](#)

---

### Articles you may be interested in

[Bézier curve string method for the study of rare events in complex chemical systems](#)

*J. Chem. Phys.* **141**, 074110 (2014); 10.1063/1.4893216

[Adaptive single replica multiple state transition interface sampling](#)

*J. Chem. Phys.* **139**, 044105 (2013); 10.1063/1.4813777

[Adaptively biased sequential importance sampling for rare events in reaction networks with comparison to exact solutions from finite buffer dCME method](#)

*J. Chem. Phys.* **139**, 025101 (2013); 10.1063/1.4811286

[Theoretical prediction of rare gas inserted hydronium ions:  \$\text{HRgOH}\_2 +\$](#)

*J. Chem. Phys.* **138**, 194308 (2013); 10.1063/1.4804623

[A direct method for locating minimum-energy crossing points \(MECPs\) in spin-forbidden transitions and nonadiabatic reactions](#)

*J. Chem. Phys.* **123**, 094711 (2005); 10.1063/1.2007708

---



**NEW Special Topic Sections**

**NOW ONLINE**  
Lithium Niobate Properties and Applications:  
Reviews of Emerging Trends

**AIP** | Applied Physics  
Reviews

## Adaptive minimum action method for the study of rare events

Xiang Zhou,<sup>1,a)</sup> Weiqing Ren,<sup>2,b)</sup> and Weinan E<sup>3,c)</sup>

<sup>1</sup>Program for Applied and Computational Mathematics, Princeton University, Princeton, New Jersey 08540, USA

<sup>2</sup>Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, USA

<sup>3</sup>Department of Mathematics and Program for Applied and Computational Mathematics, Princeton University, Princeton, New Jersey 08540, USA

(Received 21 September 2007; accepted 10 December 2007; published online 14 March 2008)

An adaptive minimum action method is proposed for computing the most probable transition paths between stable equilibria in metastable systems that do not necessarily have an underlying energy function, by minimizing the action functional associated with such transition paths. This new algorithm uses the moving mesh strategy to adaptively adjust the grid points over the time interval of transition. Numerical examples are presented to demonstrate the efficiency of the adaptive minimum action method. © 2008 American Institute of Physics. [DOI: 10.1063/1.2830717]

### I. INTRODUCTION

The behavior of stochastically perturbed dynamical systems is dominated by rare but important transition events between stable states. Such rare events play a major role in chemical reactions, conformational change of biomolecules, nucleation events, and the like.<sup>1–4</sup> Recently their relevance has also been realized in many new applications, such as species competition models in ecology and evolution,<sup>5</sup> genetic toggle switch in molecular biology,<sup>6</sup> stochastic catastrophe in electric, and computer networks.<sup>7,8</sup>

Theoretical understanding of such transition events and transition paths has attracted a lot of attention for many years. The theory of large deviations was developed for this purpose.<sup>9</sup> Most relevant to the discussion of this paper is the Freidlin–Wentzell theory,<sup>10</sup> which gives an estimate of the probability of the paths in terms of an action functional over the paths. One of the key conclusions of this theory is that the most probable path is the minimizer of the action functional associated with the random dynamical system, i.e., the most probable path is the *minimum action path* (MAP). For gradient systems for which the system is characterized by an underlying energy landscape, in the high friction limit, the minimum action path is simply the *minimum energy path* (MEP) which minimizes the energy barrier along the path. The minimum energy path can be characterized as the heteroclinic orbit connecting the two given equilibria. It is easy to show that such MEPs have to go through saddle points, and these saddle points define the barrier of the transition.<sup>11</sup> For nongradient systems, there is no such a clean characterization of the MAPs, and they may behave in a much more complex way.<sup>10,12</sup>

From a numerical viewpoint, one difficulty in finding such transition path is the disparity of the time scales associated with the frequency of the transition and the time that the transition event actually takes. This is partly reflected in

the fact that the transition path spends most of its time at the equilibria with fast transition between the equilibria. Therefore a key issue for computing such transition path is the numerical parametrization. For gradient systems, the string method<sup>13</sup> uses the arc length parametrization to compute the MEPs. This intrinsic parametrization is very effective, the problem associated with the time scale separation is eliminated.

For nongradient systems, the algorithm<sup>14</sup> proposed by E *et al.* is to minimize the action functional with respect to all possible paths connecting initial and final states of the system over a specified time interval of transition, using the physical time as parametrization of the paths. This type of algorithm is called the *minimum action method* (MAM). However, this original minimum action method uses a rather inefficient numerical parametrization of the paths, the physical time. Olender and Elber<sup>15</sup> were the first to realize that it is better to use the arc length parametrization. For gradient systems, they modified the energy functional by adding a penalty term that favors equal arc length parametrization. This strategy was adopted by the well-known nudged elastic band method.<sup>16</sup> These penalty-based approaches do not enforce strictly the equal arc length parametrization. The string method, on the other hand, is based on the philosophy that the minimum action path is a geometric curve and therefore one should view it as such and parametrize it using any intrinsic parametrization such as the arc length.<sup>13</sup> The string method is quite successful in that sense. However, it is limited to gradient systems with an underlying energy function. For general nongradient systems, Heymann and Vanden–Eijnden<sup>17</sup> proposed to reformulate the action functional in the space of geometric curves parametrized by arc length. One can then find the minimum action path using standard optimization techniques. Following them, we will call their modified new action *the geometric action* and their new method as the *geometric minimum action method*, or “gMAM.” The efficiency of the minimum action method is improved significantly as a consequence of this intrinsic parametrization.

<sup>a)</sup>Electronic mail: xiangz@math.princeton.edu.

<sup>b)</sup>Electronic mail: weiqing@cims.nyu.edu.

<sup>c)</sup>Electronic mail: weinan@math.princeton.edu.

In this paper, we take a different viewpoint. We will devise a very simple numerical strategy in which one still uses the Wentzell–Freidlin action and hence the physical time parameter as the variable but adaptively chooses the mesh using a moving mesh strategy. We call this method the adaptive minimum action method, or “aMAM.” The advantage of this approach is that one does not need to reformulate the original action functional and one can adjust the efficiency using the monitor function. In particular, in the limit as the parameter  $C \rightarrow +\infty$  in the monitor function [Eq. (11)] (see Sec. III), one recovers the arc length parametrization.

This paper is organized as follows. In the next section, we give a brief review of the Freidlin–Wentzell theory and the original minimum action method.<sup>14</sup> Then we present the application of the moving mesh method to the minimum action method and introduce the adaptive minimum action method in Sec. III. Numerical results from three examples are presented for discussions in Sec. IV. Some conclusions are drawn in the last section. We focus on the case when the transition time interval is finite, leaving the case of infinite transition time interval to the appendix.

## II. THE LEAST ACTION PRINCIPLE AND THE MINIMUM ACTION METHOD

Consider the following stochastic differential equation in  $\mathbb{R}^d$  with a small positive parameter  $\varepsilon$ :

$$dX_t = b(X_t)dt + \sqrt{\varepsilon}dW_t, \quad (1)$$

where  $W_t$  is a Wiener process on  $\mathbb{R}^d$ . The Wentzell–Freidlin theory<sup>10</sup> gives an estimate of the probability distribution of the solutions  $X$  of Eq. (1) over any fixed time interval  $[T_1, T_2]$ : For any small parameter  $\delta$  and a smooth path  $\varphi$  on  $[T_1, T_2]$ ,

$$\mathbf{P}\{\rho(X, \varphi) < \delta\} \approx \exp\left(-\frac{1}{\varepsilon}S_{T_1, T_2}[\varphi]\right), \quad (2)$$

where  $\rho$  is the distance in the space of continuous functions. The action functional  $S$  at a given path  $\varphi(t)$  ( $t \in [T_1, T_2]$ ) is given by

$$S_{T_1, T_2}[\varphi] = \frac{1}{2} \int_{T_1}^{T_2} |\dot{\varphi}(t) - b(\varphi(t))|^2 dt. \quad (3)$$

Given two states  $a_1$  and  $a_2$ , the most probable path that connects  $a_1$  and  $a_2$  over the time interval  $[T_1, T_2]$  is the minimizer of  $S_{T_1, T_2}[\varphi]$  with the constraint that  $\varphi(T_1) = a_1$  and  $\varphi(T_2) = a_2$ . Such a path is called a minimum action path or MAP. In practice, it is of particular interest to study such MAPs for the situation when  $a_1$  and  $a_2$  are stable stationary points of the dynamical system without noise. To find the average transition time from  $a_1$  to  $a_2$ , we should in addition minimize over  $T_1$  and  $T_2$ . This leads to the study of MAP over the infinite time interval,

$$S[\varphi] = \frac{1}{2} \int_{-\infty}^{+\infty} |\dot{\varphi}(t) - b(\varphi(t))|^2 dt, \quad (4)$$

subject to  $\varphi(-\infty) = a_1$  and  $\varphi(+\infty) = a_2$ .

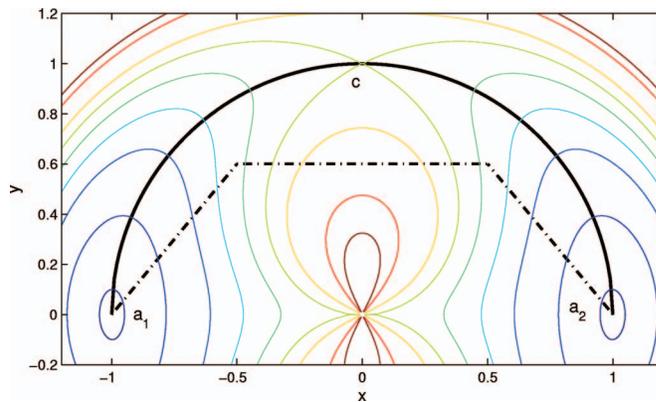


FIG. 1. (Color) The exact MAP in the upper half-plane (thicker solid curve) and initial path (dash-dotted curve) used in the minimum action method.

It is well known<sup>11</sup> that if the system [Eq. (1)] happens to be a gradient system and if  $a_1$  and  $a_2$  are two local minima of the underlying potential function  $V(x)$ , then the MAP  $\varphi$  is also the MEP.

Efficient numerical methods such as the nudged elastic band method<sup>16</sup> and the zero-temperature string method<sup>11,13,18</sup> have been developed for finding the MEPS. The most recent development of the string method<sup>18</sup> presents a new version of the string method that is not only simpler and easier to implement but also circumvents the need of the projection of the potential force.

For general nongradient systems, E *et al.* proposed the minimum action method.<sup>14</sup> To illustrate the MAM, we consider the following example for which the MAP can be found explicitly:<sup>18</sup>

$$\begin{cases} dx = -\partial_x V(x, y)dt + \sqrt{\varepsilon}dW_t^x \\ dy = -\partial_y V(x, y)dt + \sqrt{\varepsilon}dW_t^y, \end{cases} \quad (5)$$

with the potential

$$V(x, y) = (1 - x^2 - y^2)^2 + y^2/(x^2 + y^2). \quad (6)$$

The contour lines of  $V$  are shown in the Fig. 1. The potential  $V$  has two minima at  $a_1 = (-1, 0)$  and  $a_2 = (1, 0)$ , respectively. The saddle point in the upper half-plane is  $c = (0, 1)$ . The exact value of the corresponding minimum action functional is  $2 \times (V(c) - V(a_1)) = 2$ . The MAPs connecting the two minima are the upper and lower branches of the unit circle:  $x^2 + y^2 = 1$ . For convenience we will call Eq. (5) the “semi-circle” example.

It is useful to recall the original MAM.<sup>14</sup> It starts from the action functional [Eq. (3)]. Given a finite interval  $[T_1, T_2]$ , we divide it into  $m$  subintervals to form a mesh,

$$T_1 = t_0 < t_1 < \dots < t_m = T_2.$$

A path  $\varphi(t)$  is approximated by its values,  $\Phi_n$ , at  $t = t_n$  for  $n = 0, \dots, m$ . The action  $S$  of this path is approximated according to the midpoint rule,

$$S_{t_0, \dots, t_m}[\Phi_1, \dots, \Phi_{m-1}] = \frac{1}{2} \sum_{n=1}^m \left| \frac{\Phi_n - \Phi_{n-1}}{\Delta t_n} - b(\Phi_{n-1/2}) \right|^2 \Delta t_n, \quad (7)$$

where  $\Delta t_n = t_n - t_{n-1}$  and  $\Phi_{n-1/2} = (\Phi_n + \Phi_{n-1})/2$ . The two endpoints  $\Phi_0$  and  $\Phi_m$  are known to be  $a_1$  and  $a_2$ , respectively. In the original MAM, the mesh  $\{t_n\}$  does not change and thus the expression of the objective function [Eq. (7)] does not change either.

Steepest descent or other optimization methods can be applied to minimize the objective function [Eq. (7)]. The optimization algorithm used in Ref. 14 is the limited memory BFGS method (L-BFGS),<sup>19</sup> one of the most successful quasi-Newton methods. It requires the evaluation of the gradient of the objective function and achieves superlinear convergence. L-BFGS stores the  $M$  most recent updates ( $M$  levels of memory) of displacements and values of the objective function to construct an approximation of the Hessian. L-BFGS is used in this paper with  $M=5$ .

To get a feeling about the performance of MAM, we calculated the MAPs of Eq. (5) using a uniform mesh with 100 points over two time intervals  $[-15, 15]$  and  $[-0.5, 0.5]$ . The results are presented in Fig. 2. The first observation is that the MAP calculated over the time interval  $[-0.5, 0.5]$  is far away from the exact MAP. The MAP calculated over the time interval  $[-15, 15]$  is much closer to the exact MAP. This shows the importance of choosing a large enough time interval in order to capture the correct optimal transition path. Another notable observation in Fig. 2 is that the path spends most of its time waiting in the vicinity of the three fixed points; while during two small time subintervals around  $t = \pm 8$ , it hops rapidly from  $(-1, 0)$  to  $(0, 1)$  and then from  $(0, 1)$  to  $(1, 0)$ . The geometric shape of the MAP is determined mainly by the behavior over these two small time intervals, where, unfortunately, very few mesh points are allocated. The situation is even worse when the time interval  $[T_1, T_2]$  becomes larger. To see this, we take  $[T_1, T_2] = [-T/2, T/2]$  for different values of  $T$  and plot the corresponding MAPs in terms of the scaled time  $t/T$  in Fig. 3. We find that the larger  $T$ , the less portion of the time is spent for the transition.

### III. THE ADAPTIVE MINIMUM ACTION METHOD

To resolve the fast yet very important time scale in the transition, we use the moving mesh technique<sup>20</sup> to adaptively allocate the grid points. By now, the moving mesh technique is fairly well known and quite widely used.<sup>21–23</sup> Compared with adaptive mesh refinement,<sup>24–26</sup> the moving mesh technique has the advantage that the mesh is regular in the computational domain and smooth in the physical domain. For this reason, it should have better accuracy. The adaptive minimum action method is based on this idea. The goal is still to minimize the action functional [Eq. (3)],

$$S_{T_1, T_2}[\varphi] = \frac{1}{2} \int_{T_1}^{T_2} |\dot{\varphi}(t) - b(\varphi(t))|^2 dt,$$

but with a mesh that is changing. We seek a mapping from the physical domain  $t \in [T_1, T_2]$  to the computational domain

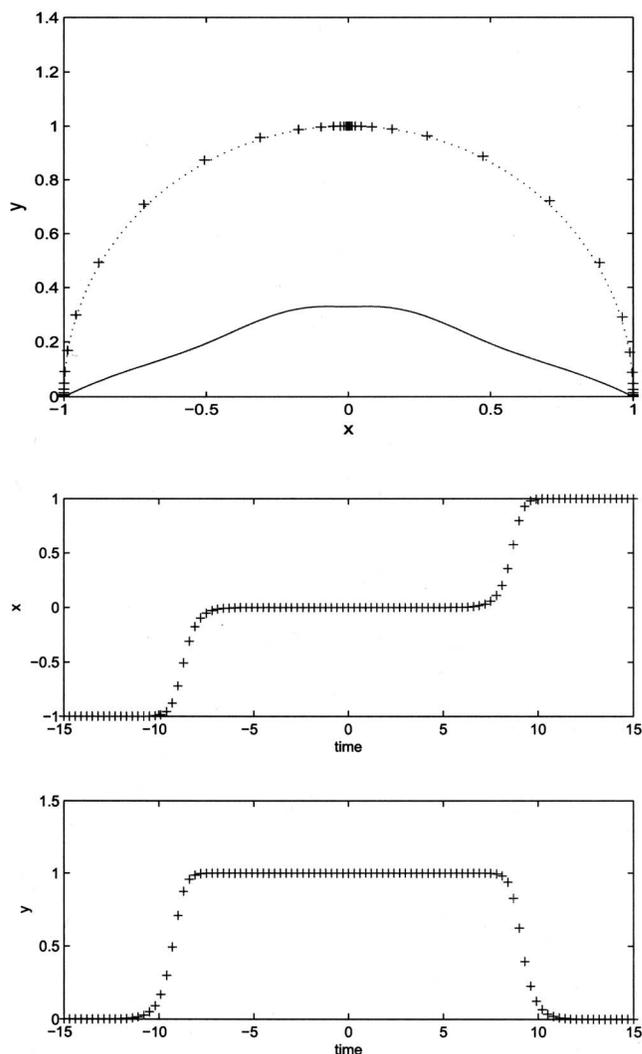


FIG. 2. The calculated MAPs with the uniform mesh ( $m=100$ ) over the time interval  $[T_1, T_2] = [-15, 15]$  (marked with “+”) and  $[T_1, T_2] = [-0.5, 0.5]$  (solid line). The top panel is the geometric curve of the calculated MAP (the exact MAP is also shown as the dotted curve). The bottom two panels show  $x$  and  $y$  components of  $\varphi(t)$  as a function of  $t$ . Almost 85% of points are wasted near three fixed points:  $(\pm 1, 0)$  and  $(0, 1)$ . Only about 15% of them contribute to the geometric shape of MAP.

$\alpha \in [0, 1]$  such that in the new variable  $\alpha$ , the large variation of the path  $\varphi$  at fast transitions between long-lived states is reduced and the uniform mesh can be used in the variable  $\alpha$ . The commonly used mesh generation techniques are based on a variational approach. Winslow<sup>23</sup> suggested a functional of the form

$$E(\alpha) = \int_{T_1}^{T_2} \frac{1}{w} \left( \frac{d\alpha}{dt} \right)^2 dt, \quad (8)$$

whose Euler–Lagrange equation is

$$\begin{cases} \frac{d}{dt} \left( \frac{1}{w(t)} \frac{d\alpha}{dt} \right) = 0, & t \in (T_1, T_2) \\ \alpha(T_1) = 0, & \alpha(T_2) = 1. \end{cases} \quad (9)$$

The coefficient  $w(t)$  is the monitor function which controls the distribution of image points along the path. To see how Winslow’s equation works, we integrate Eq. (9) and get

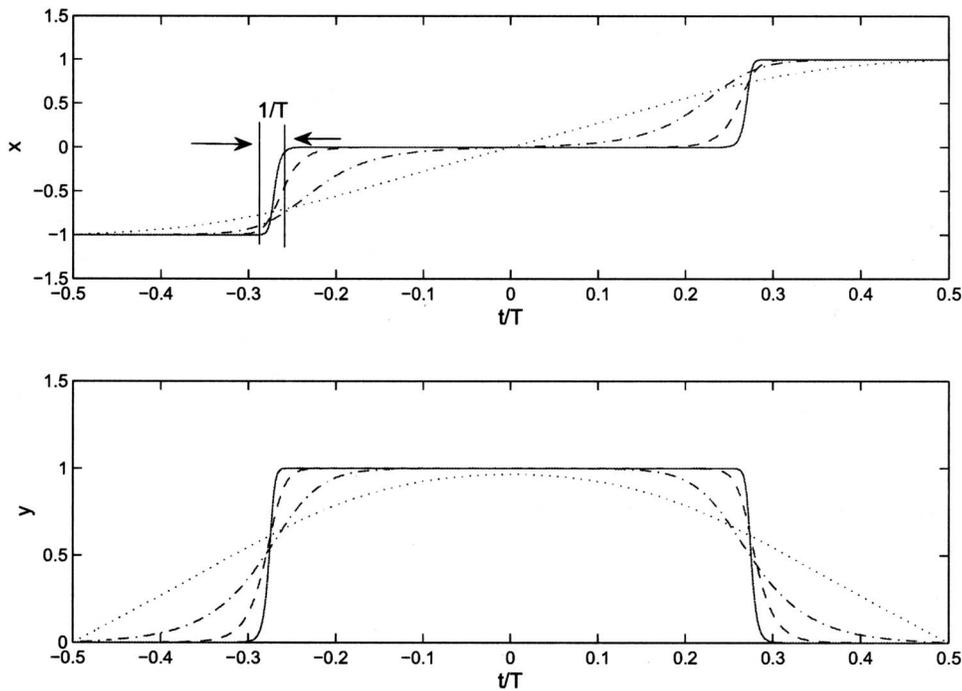


FIG. 3. The MAPs in terms of the scaled time  $t/T$  for different values of the truncated transition time  $T$ : 2 (dotted), 10 (dash-dotted), 30 (dashed), and 100 (solid).

$$\frac{1}{w(t)} \frac{d\alpha}{dt} = C_0$$

or

$$w(t) \frac{dt}{d\alpha} = \frac{1}{C_0},$$

where the constant  $C_0 = (\int_{T_1}^{T_2} w(t) dt)^{-1}$ . We integrate the above again from  $\alpha_{n-1}$  to  $\alpha_n$  and get

$$\Delta\alpha_n \int_{T_1}^{T_2} w(t) dt = \int_{t_{n-1}}^{t_n} w(t) dt \approx w(t_{n-1/2}) \Delta t_n. \quad (10)$$

In computations, a uniform mesh in  $\alpha$  is used, thus  $\Delta\alpha_n \equiv 1/m$ . Equation (10) indicates that the node distance  $\Delta t_n$  is smaller in the region where  $w(t)$  is larger. As a result,  $w$  can

control the allocation of the mesh points for the time variable  $t$ .

The key ingredient in the moving mesh method is a good monitor function  $w(t)$  in Eq. (9). Since the disparity of time scales leads to large variation of  $\varphi(t)$  in small effective transition time regions (see Fig. 2), the time derivative of the path  $|\dot{\varphi}_t|$  is a good candidate of the monitor function. However, this function is a bit singular. Therefore, we use the following monitor function:

$$w(t) = \sqrt{1 + C|\dot{\varphi}_t|^2}, \quad (11)$$

where  $C$  is a positive constant which controls the maximum time steps [Eq. (10) implies that the maximum time step is  $\Delta\alpha \int_{T_1}^{T_2} w(t) dt$ ]. When  $C$  goes to infinity,  $w(t) \sim |\dot{\varphi}_t|$ , thus the new mesh essentially generates the equal arc length param-

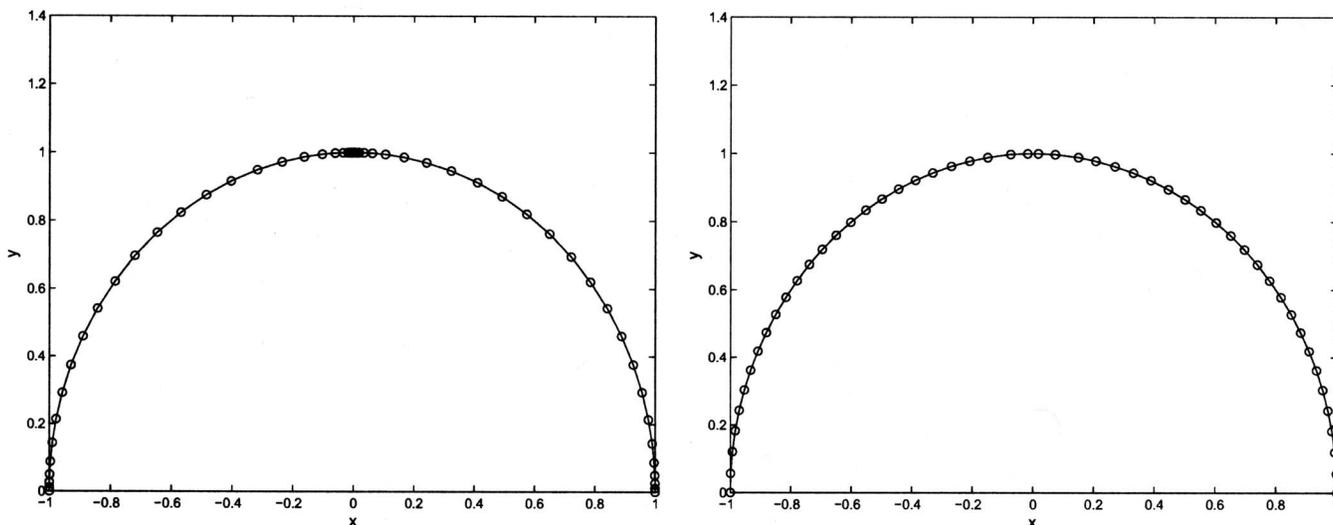


FIG. 4. The MAPs computed with different constants  $C$  in the monitor function [Eq. (11)].  $C=8$  in the left panel, and  $C=8000$  in the right panel. For larger  $C$ , the points along the path tend to be evenly distributed.

etrization of the path. Figure 4 presents numerical results for  $C=8$  and  $C=8000$  when the adaptive minimum action method is applied to the semicircle example [Eq. (5)].

The mesh equation [Eq. (9)] is solved together with an optimization algorithm. We summarize the complete procedure of the aMAM over the finite time interval  $[T_1, T_2]$  in Algorithm 1. In this algorithm, the path is characterized by the mesh  $\{t_n\}$  and the corresponding image points  $\{\Phi_n\}$  in the configuration space,  $n=0, \dots, m$  where  $m$  is the total number of grid points.

Algorithm 1 (“aMAM”)

step 0. Set  $k=0$ . Set the uniform mesh  $\{\alpha_n\}$  over the interval  $[0, 1]$ .

Choose an initial (uniform) mesh  $\{t_n\}^{(k)}$  over  $[T_1, T_2]$  and the initial path  $\{\Phi_n\}^{(k,0)}$ .

Step 1. do  $i=1, \dots, R$

$$\{\Phi_n\}^{(k,i)} = \{\Phi_n\}^{(k,i-1)} + \beta_{i-1}^{(k)} p_{i-1}^{(k)}$$

enddo

where  $\beta$  and  $p$  are, respectively, the search steps and directions in the optimization algorithm for minimizing Eq. (7).

Step 2. Calculate the discrete monitor function

$\{w_n\}^{(k)}$  by Eq. (11) using

$\{\Phi_n\}^{(k,R)}$ . Compute the

quantity

$$Q = \frac{\max_n w_n^{(k)} \Delta t_n}{\min_n w_n^{(k)} \Delta t_n}$$

if  $Q < Q'$

$\{\Phi_n\}^{(k,0)} = \{\Phi_n\}^{(k,R)}$ ; go to step 1.

else

go to step 3.

endif

Step 3. Solve the mesh equation [Eq. (9)] using  $\{w_n\}^{(k)}$  on the mesh  $\{t_n\}^{(k)}$  to obtain  $\{\alpha_n\}^{(k)}$ . Interpolate the mapping between  $\{\alpha_n\}^{(k)}$  and  $\{t_n\}^{(k)}$  to get the new mesh  $\{t_n\}^{(k+1)}$  corresponding to  $\{\alpha_n\}$ . Interpolate the path ( $\{\Phi_n\}^{(k,R)}, \{t_n\}^{(k)}$ ) on the new mesh  $\{t_n\}^{(k+1)}$  to get  $\{\Phi_n\}^{(k+1,0)}$ .  $k=k+1$ ; go to step 1.

In this algorithm, Step 1 is the traditional optimization program. The parameter  $R$  is the iteration number.  $R$  should be larger than  $M$  for  $M$ -level L-BFGS solver since all levels of memory have to be discarded after each mesh adjustment. Step 2 is to check the mesh quality by looking at  $Q$ . Since  $\{\alpha_n\}$  is uniform,  $Q$  should be close to 1 by Eq. (10) if the grid points in the time variable are allocated in the desired way. The threshold  $Q'$  is typically chosen to be 2–5 in practice. Step 3 is the mesh moving procedure. After the mesh equation is solved, two interpolations are needed. The first one must conserve the monotonicity of the mapping between  $\alpha$  and  $t$ . We use the piecewise linear interpolation for this purpose. The second interpolation of the path on the new mesh requires sufficient high order of accuracy to match the accuracy of the optimization algorithm. We use the cubic spline interpolation in this work.

The moving mesh strategy can be extended to deal with the infinite time interval where  $T_1 \rightarrow -\infty$  and  $T_2 \rightarrow +\infty$ . One

strategy is to introduce a mapping between the infinite time interval and the region  $[0, 1]$  of a new variable  $s$  such that the action functional is well defined in terms of new variable  $s$ . See Appendix for details.

## IV. NUMERICAL EXAMPLES

We demonstrate the adaptive minimum action method using three examples. The first one is the semicircle example [Eq. (5)], and the second is a gradient system with Mueller potential<sup>27</sup> whose MAP has more than two stable equilibria. The last is a simple nongradient system. We use five level limited memory BFGS as the optimization algorithm with the preconditioner  $(-\partial_{tt})^{-1}$ .

### A. Semicircle example

The initial path is located in the upper half-plane as shown in Fig. 1. The termination criteria in L-BFGS is that the relative error between two successively calculated actions is less than  $10^{-4} \times m^{-2}$ . The error of calculated MAP is defined to be

$$e(m) = \max_n |(x_n^2 + y_n^2)^{1/2} - 1|.$$

The monitor function is  $w(t) = (1 + 8000 \times |\varphi_t|^2)^{1/2}$ . Three finite time intervals  $[-T/2, T/2]$  with  $T=1, 2, 10$  are chosen and the number of points  $m=50$ . The numerical results of the MAPs are shown in Fig. 5. The comparison between the original minimum action method and the adaptive minimum action is illustrated in Fig. 6 for different time intervals.

Figure 5 confirms that the adaptive minimum action method can distribute grid points much more evenly along the MAP than the original minimum action method. The effect of the monitor function has been discussed in Fig. 4, Sec. III. As far as the accuracy is concerned, Fig. 6 shows that the adaptive minimum method lowers the errors both in the action and in the MAP, especially for large time intervals. When the time interval  $T$  is reasonably small (e.g.,  $T \lesssim 10$  in this case), the uniform time step  $T/m$  is small and the moving mesh strategy does not take much effect; the main part of the error is due to the finite time interval truncation. For example, in Fig. 5, the numerical MAP with  $T=1$  deviates a lot from the correct MAP, and the calculated minimal action with  $T=1$  is 3.92, while the exact value is 2.00. When  $T$  increase to 10, the error in the action and the path in the adaptive minimum action method decrease to the orders of  $10^{-3}$  and  $10^{-4}$ , respectively, even though the number of grid points  $m=50$  does not change. To illustrate how the finite time interval truncation affects the numerical results, we use different  $T$  from 2 to 120 for a very large  $m=5000$  in the adaptive minimum action method. The error is shown in Fig. 7. In the fast decay part of the plotted figure ( $T < 20$  roughly saying), the error is dominated by the finite size effect of the time interval. When  $T$  is larger than 10, the error due to finite time interval truncation is smaller compared with the errors in Fig. 6 and the error due to time step discretization begins to be the main source of the total error. Indeed, the adaptive minimum action method decreases this type of error due to discrete time step by properly adjusting time mesh. Thus,

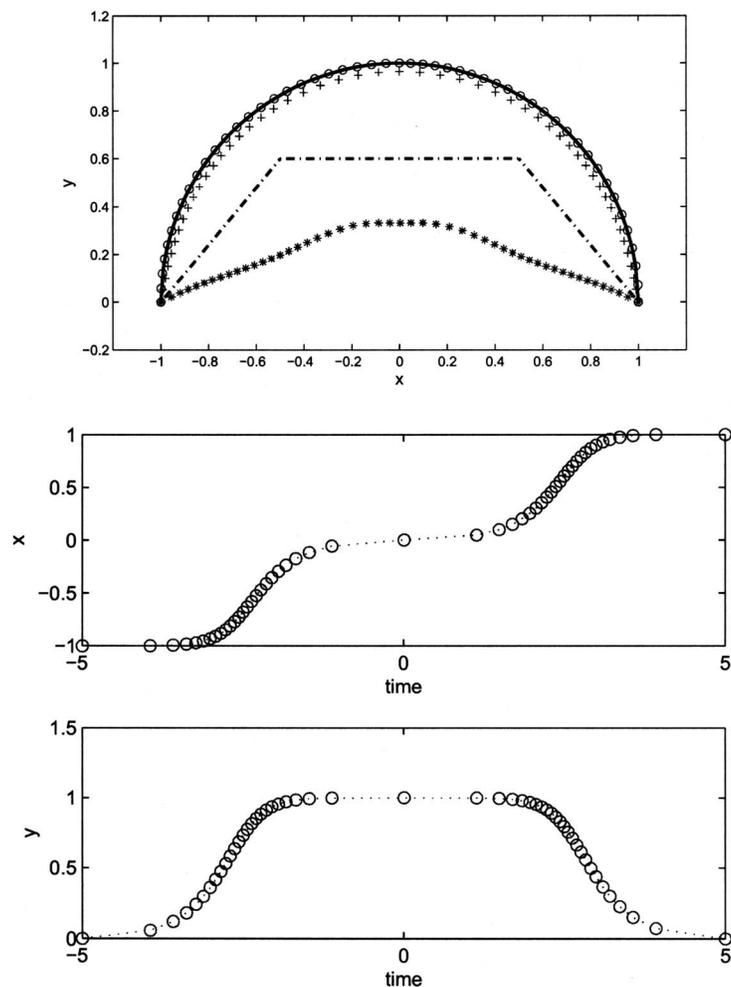


FIG. 5. Numerical results of Eq. (5) calculated by the adaptive minimum action method in three finite time intervals with different values of  $T$  (the number of discrete points is  $m=50$ ). Top panel: the initial path (dash-dotted line), the exact path (thick solid line) and the calculated MAPs with  $T=10$  ( $\circ$ ),  $2$  ( $+$ ), and  $1$  ( $*$ ). Bottom two panels: the  $x$  and  $y$  components of the calculated MAP with  $T=10$  in terms of the time  $t$ .

when  $T$  continues to increase, the adaptive minimum action can keep the error nearly unchanged, compared with the original minimum action method (see the very slight increasing part of the solid and broken lines marked with “\*” in Fig. 6).

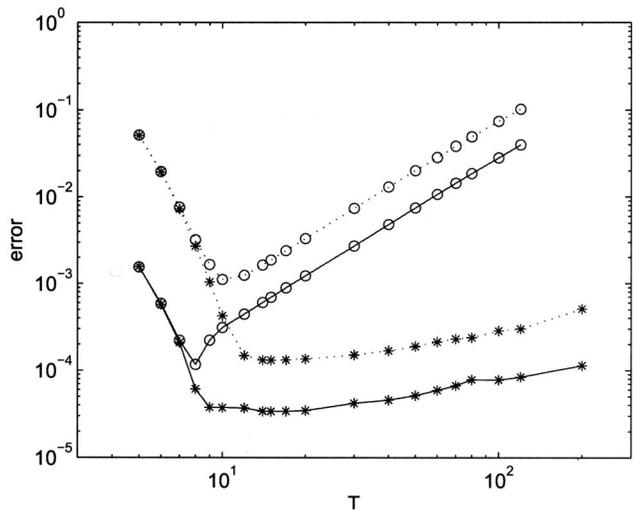


FIG. 6. The errors of the calculated actions and MAPs of Eq. (5) with different transition time intervals  $[-T/2, T/2]$  (dotted line: the error of action, solid line: the error of MAP) for fixed number of points  $m=200$ . The errors of the original minimum action method (uniform mesh) are marked with  $\circ$ , and the errors of moving mesh method are marked with  $*$ .

The numerical scheme [Eq. (7)] and the L-BFGS give a convergence rate  $O(\Delta t^2)$ , i.e.,  $O(m^{-2})$ , for a fixed time interval. It can be demonstrated numerically that the moving mesh strategy in the adaptive minimum action method does not lower the order of accuracy. We choose two time intervals  $[T_1, T_2]=[-5, 5]$  and  $[T_1, T_2]=[-50, 50]$  to study the rate

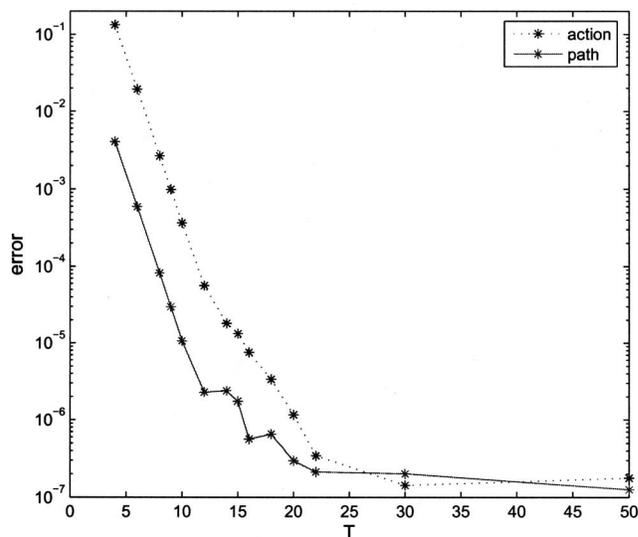


FIG. 7. The error for finite transition time  $[-T/2, T/2]$  for different values of  $T$  from 2 to 120.  $m=5000$ .

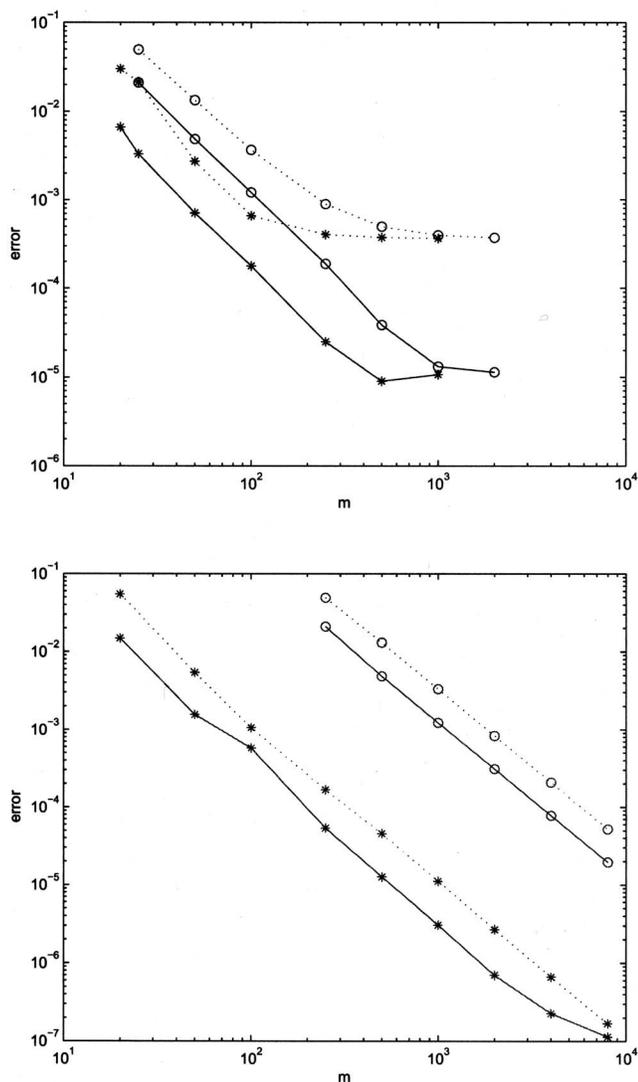


FIG. 8. The error of the action (dotted line) and MAP (solid line) for  $[T_1, T_2] = [-5, 5]$  (top) and  $[T_1, T_2] = [-50, 50]$  (bottom). The results marked with  $\circ$  are from the original minimum action method and the results marked with  $*$  are from the adaptive minimum action method.

of convergence as  $m$  increases. See Fig. 8. The orders of the original minimum action method are 1.92 and 2.05 for the action and the MAP, respectively; the orders of the moving mesh method are 2.08 and 2.03 for the action and the MAP, respectively (based on the data from  $[T_1, T_2] = [-50, 50]$ ).

If we allow the finite time interval  $[T_1, T_2] = [-\sqrt{m}/2, \sqrt{m}/2]$  to depend on the number of points,  $m$ , in order to diminish the error of finite time truncation, the convergence rate of the original minimum action method becomes  $O(\Delta t^2) = O(1/m)$  since the uniform time step size  $\Delta t = \sqrt{m}/m = 1/\sqrt{m}$ . While in the adaptive minimum action method, this convergence rate can be improved to nearly  $O(1/m^2)$  because in the dense region of the time mesh which mainly decides the accuracy, the time step size is approximately of the order of  $1/m$ , thanks to the moving mesh strategy. The numerical result in Fig. 9 confirms this.

To quantitatively compare the computational efficiency of the uniform mesh method and the moving mesh method, we let  $[T_1, T_2] = [-15, 15]$  and compute the errors of the

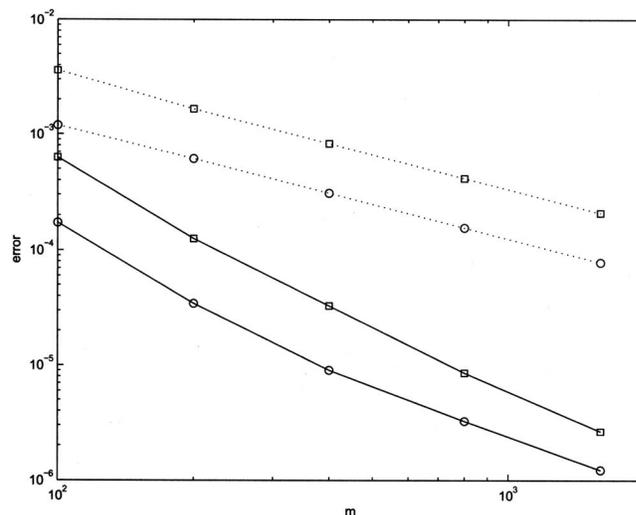


FIG. 9. The error of the action ( $\square$ ) and MAP ( $\circ$ ) for  $[T_1, T_2] = [-\sqrt{m}/2, \sqrt{m}/2]$ . The dotted lines (the slopes are 1.02 and 0.99, respectively) are results from the original minimum action method and the solid lines (the slopes are 1.97 and 1.77, respectively) are from the adaptive minimum action method.

MAP and the action for different  $m$  and the necessary number of evaluations of the objective function and its gradient,  $n_{\text{ev}}(m)$ , in the iterations of the optimization algorithm. Since each evaluation of the objective function [Eq. (7)] and its gradient requires summation of  $m$  terms, we use  $m \times n_{\text{ev}}(m)$  as the measurement of the computational cost for the original minimum action method, and  $m \times (n_{\text{ev}}(m) + K)$  for the adaptive minimum action method, where  $K$  is the number of mesh adjustments. Then we plot the costs versus the error in Fig. 10. It is clear that the adaptive minimum action method outperforms the original minimum action method.

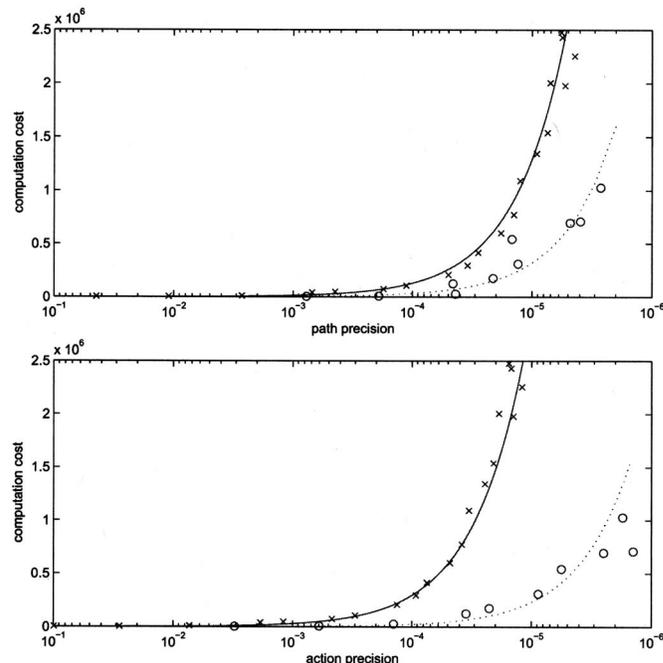


FIG. 10. The computational cost of the uniform mesh method ( $\times$ , fitted by the solid line) and the moving mesh method ( $\circ$ , fitted the dotted line) in order to reach the prescribed precisions (in the horizontal axis).

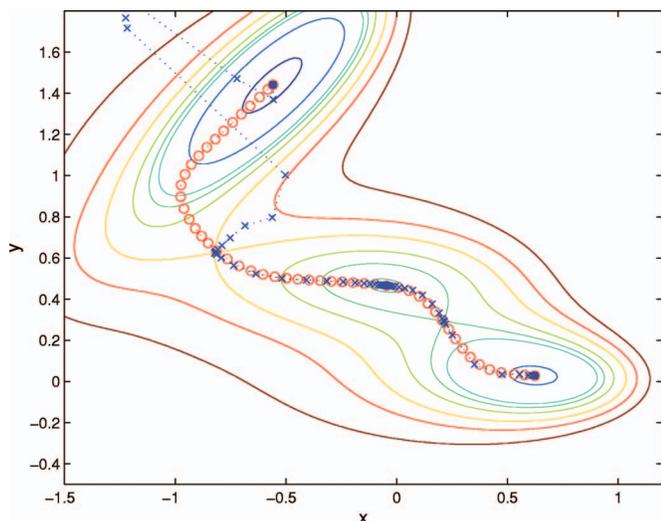


FIG. 11. (Color) Calculated MAPs ( $\circ$  from moving mesh,  $\times$  from uniform mesh) with  $m=50$  and  $T=12$ . The contour lines of the Mueller potential [Eq. (12)] are shown and the three solid dots are its stable equilibria.

### B. Case 2: Mueller potential system

The Mueller potential is of the form

$$V(x, y) = \sum_{i=1}^4 D_i \exp(A_i(x - x'_i)^2 + B_i(x - x'_i)(y - y'_i) + C_i(y - y'_i)^2), \quad (12)$$

where the parameters are ( $D=(D_1, D_2, D_3, D_4)$ , etc.):  $D=(-2, -1, -1.7, 0.15)$ ,  $A=(-1, -1, -6.5, 0.7)$ ,  $B=(0, 0, 11, 0.6)$ ,  $C=(-10, -10, -6.5, 0.7)$ ,  $x'=(1, 0, -0.5, -1)$ , and  $y'=(0, 0.5, 1.5, 1)$ .

Figure 11 shows the MAPs obtained using the adaptive minimum action method and the original minimum action method with  $m=50$ . The original minimum action method cannot resolve the exit path from the first equilibrium with 50 points. The moving mesh method not only produces the good approximation of the exact MAP with 50 points, but

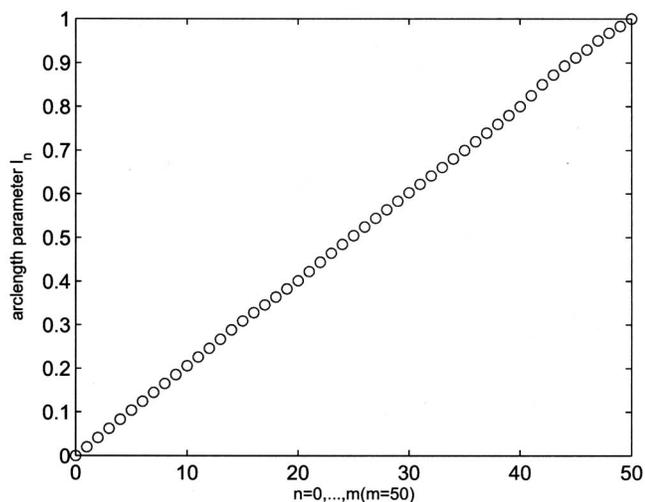


FIG. 12. The arc length (normalized)  $l_n$  as a function of the  $n$ th grid point for  $C=4000$  in the monitor function [Eq. (11)].  $l_n$  is defined to be  $\sum_{i=1}^n |\Phi_i - \Phi_{i-1}| / \sum_{i=1}^m |\Phi_i - \Phi_{i-1}|$  for  $n=1, \dots, m$ .

also parametrizes the path in nearly equal arc length distribution: see Fig. 12 for the arc length parametrization of the points along the numerical MAP.

### C. A simple nongradient system

We consider the following simple model:

$$\begin{cases} \dot{x} = x(3 - x - 2y) + \sqrt{\varepsilon} \dot{W}_1 \\ \dot{y} = y(3 - y - 2x) + \sqrt{\varepsilon} \dot{W}_2. \end{cases} \quad (13)$$

The stable steady states are  $(0,3)$  and  $(3,0)$  and the saddle point is  $(1,1)$ . The separatrix is the line  $y=x$  which is shown in Fig. 13 in solid lines with the heteroclinic orbits connecting the two equilibria. The calculated MAP from  $(0,3)$  to  $(3,0)$  is presented as circles in Fig. 13 using the adaptive minimum action method over the infinite time interval. When  $m=100$  is used, the actions from the original and adaptive minimum action methods for different  $T$  are shown in Fig. 14.

### V. CONCLUDING REMARKS

In this paper, we presented an adaptive minimum action method by adopting the moving mesh strategy to the original minimum action method. Compared with the original minimum action method, the adaptive minimum action method distributes the grid points more evenly along the geometric path. Thus it requires much fewer grid points to represent the minimum action path. Compared with the geometric minimum action method of Heymann and Vanden-Eijnden, the adaptive minimum action method has the feature that it works with the original action functional, but its numerical representation is very close to one that uses the arclength parametrization.

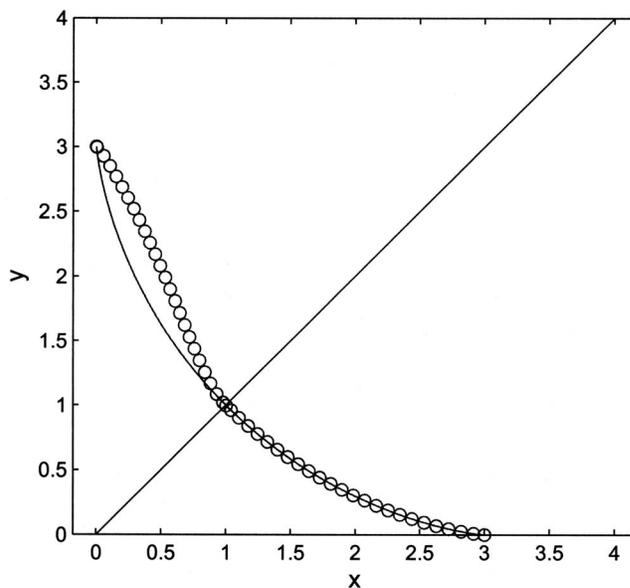


FIG. 13. Numerical MAP (circles) with  $m=50$ .

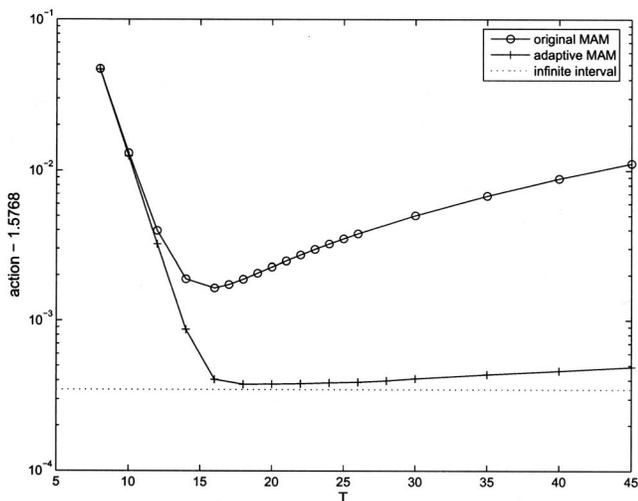


FIG. 14. Numerical results for the actions for  $m=100$ . The results marked with  $\circ$  are from original minimum action method and the results marked with  $+$  are from the adaptive minimum action method. Both methods are used over the finite time interval  $[-T/2, T/2]$ . The flat dotted line, from the adaptive minimum action method over the infinite interval (see Appendix,  $\lambda=0.5$ ), is plotted for comparison. (The value of 1.5768 is the numerical minimal action when  $m=3000$  is used; this value is used as a benchmark.)

## ACKNOWLEDGMENTS

We are grateful for valuable discussion with Tiejun Li. The work of W.E and X.Z. is supported in part by ONR Grant No. N00014-01-1-0674. The work of W.R. is partially supported by NSF Grant No. DMS-0604382.

## APPENDIX: THE MINIMUM ACTION METHOD OVER THE INFINITE TIME INTERVAL

We look at the action functional in time  $(-\infty, +\infty)$ ,

$$S[\varphi] = \frac{1}{2} \int_{-\infty}^{+\infty} |\dot{\varphi} - b(\varphi)|^2 dt, \quad (\text{A1})$$

subject to the constraints  $\varphi(-\infty)=a_1$  and  $\varphi(+\infty)=a_2$ . Our strategy is to change the physical time  $t$  into a new variable  $s$  by constructing a one-to-one mapping between  $[-\infty, +\infty]$  and  $[0,1]$ , and then expressing Eq. (A1) in terms of  $s$ . Such a mapping can be defined to be a solution of the following ODE:

$$\frac{ds}{dt} = g(s), \quad t \in (-\infty, +\infty) \quad (\text{A2})$$

where  $g$  is a non-negative function defined over  $[0,1]$  with two roots 0 and 1.

For simplicity, we take  $g(s)=\lambda s(1-s)$ , where  $\lambda$  is a positive constant. Then,

$$s(t) = \frac{1}{1 + e^{-\lambda t}}$$

or

$$t(s) = \frac{1}{\lambda} \ln \frac{s}{1-s}. \quad (\text{A3})$$

Thus, the action functional [Eq. (A1)] can be rewritten in  $s$ ,

$$S[\varphi] = \frac{1}{2} \int_0^1 |\varphi_s|^2 g(s) - 2\varphi_s \cdot b(\varphi) + \frac{|b(\varphi)|^2}{g(s)} ds, \quad (\text{A4})$$

where  $\varphi_s$  is the derivative of  $\varphi$  with respect to  $s$ . The constraints for the two endpoints become  $\varphi(s=0)=a_1$  and  $\varphi(s=1)=a_2$ .

The parameter  $\lambda$  should be carefully chosen to guarantee that each term in the integrand of Eq. (A4) goes to zero when  $s \rightarrow 0$  or  $s \rightarrow 1$ . The analysis of the asymptotic behavior of the MAP  $\varphi$  is necessary near fixed points. An ordinary differential equation can be derived to characterize the MAP. For Eq. (A1), the corresponding Euler–Lagrange equation is

$$\varphi_{tt} - J(\varphi)\varphi_t = -J^T(\varphi)(\varphi_t - b(\varphi)).$$

Define  $p(t) = \varphi_t(t) - b(\varphi(t))$ , then the above is transformed to a Hamiltonian system with Wentzell–Freidlin Hamiltonian  $H(\varphi, p) = \frac{1}{2}|p|^2 + b(\varphi) \cdot p$  (see Ref. 10),

$$\begin{cases} \dot{\varphi} = b(\varphi) + p \\ \dot{p} = -J(\varphi)^T p, \end{cases} \quad (\text{A5})$$

where the Jacobian  $J$  is  $J_{ij} = \partial b_i / \partial x_j$ . The asymptotic behavior of the MAP near the initial and final stable equilibria is governed by the linearization of Eq. (A5) at these two states. Let  $\lambda_i^{(1)}$  be the eigenvalues of the initial stable equilibrium  $a_1$  in the original deterministic system  $\dot{\varphi} = b(\varphi)$  and  $\lambda_i^{(2)}$  be the eigenvalues of the final stable equilibrium  $a_2$ , for  $i = 1, \dots, d$ . It is easy to check that the linearization of (A5) at fixed points  $(a_1, 0)$  and  $(a_2, 0)$  has the eigenvalues  $\pm \lambda_i^{(1)}$  and  $\pm \lambda_i^{(2)}$ , respectively. The path escaping from  $a_1$  lies on the unstable manifold of Eq. (A5) with nonzero  $p$ ; the path attracted to  $a_2$  lies in the stable manifold of  $(a_2, 0)$  of Eq. (A5) with zero  $p$ . Therefore, the linearized dynamics of the MAP  $\varphi$  near  $a_1$  and  $a_2$  is

$$\begin{cases} \varphi \sim e^{-\lambda_i^{(1)} t}, & t \rightarrow -\infty \\ \varphi \sim e^{\lambda_i^{(2)} t}, & t \rightarrow +\infty. \end{cases} \quad (\text{A6})$$

So, by the change of the variable [Eq. (A3)], we can have

$$\begin{aligned} \varphi_s &= \frac{\varphi_t}{g(s)} \sim -\frac{\lambda_i^{(1)} e^{-\lambda_i^{(1)} t}}{\lambda s(1-s)} \\ &= -\frac{\lambda_i^{(1)} (s/(1-s))^{\lambda_i^{(1)}/\lambda}}{\lambda s(1-s)}, \quad s \rightarrow 0, \end{aligned}$$

and

$$\begin{aligned}\varphi_s &= \frac{\varphi_t}{g(s)} \sim \frac{\lambda_i^{(2)} e^{\lambda_i^{(2)} t}}{\lambda s(1-s)} \\ &= \frac{\lambda_i^{(2)} ((1-s)/s)^{|\lambda_i^{(2)}|/\lambda}}{\lambda s(1-s)}, \quad s \rightarrow 1.\end{aligned}$$

These analyses show that  $\lambda$  should be less than  $\lambda_{\min} = \min\{|\lambda_i^{(1)}|, |\lambda_i^{(2)}|\}$  so that  $\varphi_s \rightarrow 0$  as  $s \rightarrow 0$  and 1 each term of  $i$  integrand of Eq. (A4) goes to zero at  $s=0, 1$  and thus Eq. (A4) is well defined in terms of new variable  $s \in [0, 1]$ .

After choosing the parameter  $\lambda$ , we can extend the numerical scheme and moving mesh method in Sec. II B to Eq. (A4) by reformulating the action functional in  $s$ . Given a mesh of  $\{s_n\}$  in  $[0, 1]$ ,  $0=s_0 < s_1 < \dots < s_{m-1} < s_m=1$ , the discrete sum of Eq. (A4) is

$$\begin{aligned}S_{s_0, \dots, s_m}[\Phi_1, \dots, \Phi_{m-1}] &= \frac{1}{2} \sum_{n=1}^m \frac{|\Phi_n - \Phi_{n-1}|^2}{\Delta s_n} g(s_{n-1/2}) \\ &\quad + \frac{1}{2} \sum_{n=1}^m (\Phi_n - \Phi_{n-1}) \cdot b(\Phi_{n-1/2}) \\ &\quad + \frac{1}{2} \sum_{n=1}^m \frac{|b(\Phi_{n-1/2})|^2}{g(s_{n-1/2})} \Delta t_n,\end{aligned}\quad (\text{A7})$$

where  $\Delta s_n = s_n - s_{n-1}$  and  $\Phi_{n-1/2} = (\Phi_n + \Phi_{n-1})/2$ . The two endpoints  $\Phi_0$  and  $\Phi_m$  are known.

The moving mesh strategy is used for the variable  $s$  because the change of the variable [Eq. (A3)] does not help reducing the singularity in  $s$  due to the time scale separations. As before, the mesh equation for moving the mesh is

$$\frac{d}{ds} \left( \frac{1}{w(s)} \frac{d\alpha}{ds} \right) = 0, \quad s \in (0, 1), \quad (\text{A8})$$

where  $w(s) = \sqrt{1 + C|\varphi_s|^2}$ .

In Fig. 15, we present the numerical results for the “semi-circle” example for three choices of  $\lambda$  and give errors for different  $m$ , compared with the results for the adaptive minimum action method over the finite (but sufficient large) time interval  $[-15, 15]$ .

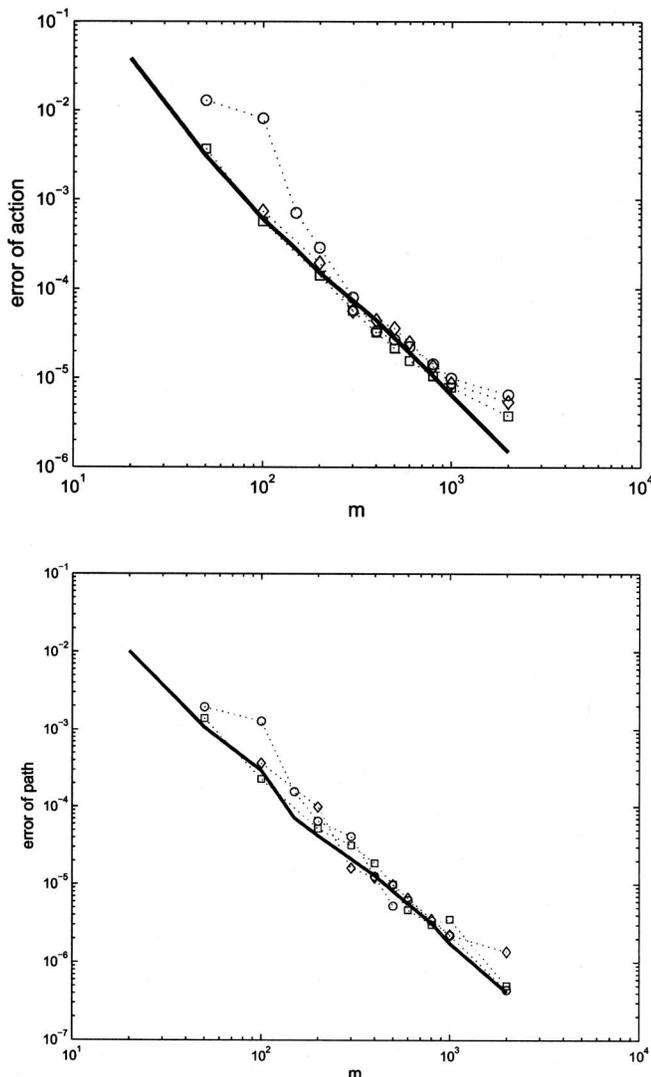


FIG. 15. The convergence rate for different values of  $\lambda$ , compared with that of the adaptive minimum action method over the interval  $[-15, 15]$  (thick solid line).  $\lambda=1$  ( $\circ$ ),  $\lambda=0.3$  ( $\square$ ),  $\lambda=0.05$  ( $\diamond$ ).  $\lambda_{\min}=2$ .

- <sup>1</sup>H. A. Karmers, *Physica (Amsterdam)* **7**, 284 (1940).
- <sup>2</sup>N. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
- <sup>3</sup>W. E, W. Ren, and E. Vanden-Eijnden, *Energy Landscapes and Rare Events*, Proc. Intl. Congress of Mathematicians (Higher Education Press, Beijing, 2002), Vol. 1, p. 621.
- <sup>4</sup>W. Ren, E. Vanden-Eijnden, P. Maragakis, and W. E, *J. Chem. Phys.* **123**, 134109 (2005).
- <sup>5</sup>M. Mangel, *Theor. Popul. Biol.* **45**, 16 (1994).
- <sup>6</sup>T. S. Gardner, C. R. Cantor, and J. J. Collins, *Nature (London)* **403**, 39 (2000).
- <sup>7</sup>C. L. DeMarco, *A Phase Transition Model for Cascading Network Failure*, *Icee Control Systems Magazine*, 21, 40 (2001).
- <sup>8</sup>R. Nelson, *J. Assoc. Comput. Mach.* **34**, 661 (1987).
- <sup>9</sup>S. Varadhan, *Large Deviations and Applications* (Society for Industrial Mathematics, Philadelphia, PA, 1984).
- <sup>10</sup>M. Freidlin and A. Wentzell, *Random Perturbations of Dynamical Systems*, 2nd ed. (Springer, New York, 1998).
- <sup>11</sup>W. Ren, Ph.D. thesis, New York University, 2002.
- <sup>12</sup>R. S. Maier and D. L. Stein, *Phys. Rev. E* **48**, 931 (1993).
- <sup>13</sup>W. E, W. Ren, and E. Vanden-Eijnden, *Phys. Rev. B* **66**, 052301 (2002).
- <sup>14</sup>W. E, W. Ren, and E. Vanden-Eijnden, *Commun. Pure Appl. Math.* **57**, 637 (2004).
- <sup>15</sup>R. Olender and R. Elber, *J. Mol. Struct.: THEOCHEM* **398–399**, 63 (1997).
- <sup>16</sup>H. Jónsson, G. Mills, and K. W. Jacobsen, *Nudged Elastic Band Method for Finding Minimum Energy Paths of Transitions*, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, Singapore, 1998), p. 385.
- <sup>17</sup>M. Heymann and E. Vanden-Eijnden, “The Geometric Minimum Action Method: A Least Action Principle on the Space of Curves,” *Commun. Pure Appl. Math.* (to be published).
- <sup>18</sup>W. E, W. Ren, and E. Vanden-Eijnden, *J. Chem. Phys.* **126**, 164103 (2007).
- <sup>19</sup>J. Nocedal, *Numerical Optimization*, Springer Series in Operations Research (Springer-Verlag, Berlin, 1999).
- <sup>20</sup>T. Tang, *Contemp. Math.* **383**, 141 (2005).
- <sup>21</sup>R. Li, T. Tang, and P. W. Zhang, *J. Comput. Phys.* **170**, 562 (2001).
- <sup>22</sup>W. Ren and X.-P. Wang, *J. Comput. Phys.* **159**, 246 (2000).

<sup>23</sup>A. M. Winslow, *J. Comput. Phys.* **1**, 149 (1967).

<sup>24</sup>I. Babuska and W. Rheinboldt, *Comput. Methods Appl. Mech. Eng.* **17**, 519 (1979).

<sup>25</sup>R. E. Bank, A. H. Sherman, and A. Weiser, in *Scientific Computing*

(*Applications of Mathematics and Computing to the Physical Sciences*), edited by R. S. Stepleman (North-Holland, Amsterdam, 1983), p. 3.

<sup>26</sup>M. J. Berger and J. Oliger, *J. Comput. Phys.* **53**, 484 (1984).

<sup>27</sup>R. Olender and R. Elber, *J. Chem. Phys.* **105**, 9299 (1996).