Commun. Comput. Phys.
doi: 10.4208/cicp.030614.180315a

Vol. **18**, No. 5, pp. 1352-1379
November 2015

A Minimum Action Method with Optimal Linear Time Scaling

Xiaoliang Wan*

Department of Mathematics, Center for Computation and Technology, Louisiana State University, Baton Rouge 70803, USA.

Received 3 June 2014; Accepted (in revised version) 18 March 2015

Abstract. In this work, we develop a minimum action method (MAM) with optimal linear time scaling, called tMAM for short. The main idea is to relax the integration time as a functional of the transition path through optimal linear time scaling such that a direct optimization of the integration time is not required. The Feidlin-Wentzell action functional is discretized by finite elements, based on which h-type adaptivity is introduced to tMAM. The adaptive tMAM does not require reparametrization for the transition path. It can be applied to deal with quasi-potential: 1) When the minimal action path is subject to an infinite integration time due to critical points, tMAM with a uniform mesh converges algebraically at a lower rate than the optimal one. However, the adaptive tMAM can recover the optimal convergence rate. 2) When the minimal action path is subject to a finite integration time, tMAM with a uniform mesh converges at the optimal rate since the problem is not singular, and the optimal integration time can be obtained directly from the minimal action path. Numerical experiments have been implemented for both SODE and SPDE examples.

AMS subject classifications: 49M25, 60H30, 65K10

Key words: Large deviation principle, small random perturbations, minimum action method, rare events, uncertainty quantification.

1 Introduction

Small random perturbations of dynamical systems can introduce rare but important events. For instance, the transitions between different stable equilibrium states of a deterministic dynamical system would be impossible if noise does not exist. Such noise-induced transitions are actually observed and critical in many physical, biological and chemical systems [16]. Examples include nucleation events of phase transitions, chemical reactions, regime change in climate, conformation changes of biomolecules, hydrodynamic instability, etc.

*Corresponding author. *Email address:* xlwan@math.lsu.edu (X. Wan)

When the amplitude of random perturbations is small, the Freidlin-Wentzell (F-W) theory of large deviations provides a rigorous mathematical framework to understand the effects of noise on dynamics [11]. The theory shows that the occurrence of rare events actually has a rather deterministic nature in the sense that a sharply peaked probability will be observed around the pathway that is least unlikely. The central object of F-W theory of large deviations is the F-W action functional. The minimizer of the F-W action functional provides the pathway of maximum likelihood and the corresponding minimum gives an estimate of the probability and the rate of occurrence of the rare events. The most important practical issue becomes how to seek the minimum and minimizer of the F-W action functional, which has to be addressed numerically for most cases.

In gradient systems, the most probable transition path is consistent with the minimum energy path (MEP), which is always parallel to the drift of the stochastic differential equation, and passes the separatrix through some saddle points with one-dimensional unstable manifold [19], i.e., transition states. Many algorithms have been developed to search the MEPs and transition states, including the string method [4, 6], the dimer method and its variants [14, 27], the nudged elastic band method [15], eigenvector-following-type methods [1], the gentlest ascent dynamics [7], etc., where the transition mechanism of gradient systems is usually employed to improve the numerical efficiency.

Unfortunately, there does not exists a definite transition mechanism in non-gradient systems [21, 22, 24, 26], where a direct minimization of the F-W action functional has to be considered. The F-W action functional is defined on a finite time interval, which can capture rare events defined on the specified time scale. The original minimum action method (MAM) [5] was constructed to deal with this case. However, the time scale of some random events increases exponentially as the amplitude of noise decreases, such as transitions between two critical points. Then we need to consider the quasi-potential, where the integration time in the F-W action functional becomes an optimization parameter and the optimal integration time can be infinite. There are several algorithms to deal with quasi-potential. The most general algorithm is the geometric MAM (gMAM) [13], which demonstrates that under proper constraints the F-W action functional can be reformulated with respect to a scaled arc length. The integration time then disappears in the optimization problem and will be recovered by the mapping between time and arc length after the most probable path is obtained. Another minimum action method, called adaptive MAM (aMAM) [20], was constructed for the F-W action functional formulated in time. The key observation of aMAM is that if the optimal integration time is infinite, the most probable transition path can be well resolved through reparametrization or remeshing using a finite but large integration time. Since the integration time must be prescribed, aMAM is not able to deal with the quasi-potential if the optimal integration time is finite. The aMAM was reformulated in [23, 25] within the framework of finite element method to enhance the numerical efficiency and robustness.

In this work, we develop a minimum action method to deal with quasi-potential formulated with respect to time instead of arc length. We choose to work with time instead of arc length because it can be more flexible, from the algorithm point of view, to deal

with the objective functional formulated with respect to time, especially for spatially extended systems. Due to the observation of aMAM [20], we will use a finite integration time for numerical approximation. We then consider a linear time scaling to map the time interval $[0,T]$ to $[0,1]$. The optimality condition shows that the optimal integration time is uniquely determined by the transition path defined on $[0,1]$. We then substitute such an optimality condition into the action functional to remove T from the optimization problem. The transition path is approximated by hp finite elements, where h indicates the element size and p the polynomial order. Such a minimum action method is called tMAM for short. Numerical experiments show that tMAM with equidistant linear elements converges algebraically at the optimal rate for the case that the optimal integration time is finite, and at a rate slightly lower than the optimal one for the case that the optimal integration time is infinite. To deal with the singularity induced by the infinite optimal integration time, we refine the finite element space using h-adaptivity, which recovers the optimal algebraic convergence rate. This way, the global reparametrization used in all available MAMs is avoided. The tMAM is mainly discussed with respect to a stochastic differential equation and can be generalized straightforwardly to spatially extended systems.

The paper is organized as follows. In Section 2, we give a brief description of the theoretical background. Our methodology is presented in Section 3 and the adaptive tMAM is developed in Section 4. We generalize tMAM to spatially extended systems in Section 5 without addressing the spatial discretization. Numerical experiments are given in Section 6 followed by a discussion section.

2 Problem

We consider small random perturbations of a dynamical system. The random process $\boldsymbol{X}_t=\boldsymbol{X}(t):\mathbb{R}_+\to\mathbb{R}^n$ is defined by the following stochastic ordinary differential equation (SODE):

$$d\boldsymbol{X}_t=\boldsymbol{b}(\boldsymbol{X}_t)dt+\sqrt{\varepsilon}d\boldsymbol{W}_t, \tag{2.1}$$

where $\boldsymbol{W}_t$ is a standard Wiener process in $\mathbb{R}^n$ and ε is a small positive parameter.

Let $\boldsymbol{\phi}(t)\in\mathbb{R}^n$ be an absolutely continuous function defined on $t\in[0,T]$. The Freidlin-Wentzell theory of large deviations [11] asserts that when ε is small enough the probability of $\boldsymbol{X}_t$ passing the δ-tube about $\boldsymbol{\phi}(t)$ on $[0,T]$ is

$$\Pr(\sup_{0\le t\le T}|\boldsymbol{X}_t-\boldsymbol{\phi}|<\delta)\approx\exp(-\varepsilon^{-1}S_T(\boldsymbol{\phi})), \tag{2.2}$$

where $|\cdot|$ indicates the ℓ_2 norm in $\mathbb{R}^n$, and $S_T(\boldsymbol{\phi})$ is called the action functional defined as

$$S_T(\boldsymbol{\phi})=\frac{1}{2}\int_0^T L(\boldsymbol{\phi},\dot{\boldsymbol{\phi}})dt=\frac{1}{2}\int_0^T|\dot{\boldsymbol{\phi}}-\boldsymbol{b}(\boldsymbol{\phi})|^2dt. \tag{2.3}$$

Eq. (2.2) implies the large deviation principle (LDP), which basically says that the probability of some random events can be estimated through a minimization if the random

perturbation is small enough. For example, if A is a Borel subset in $\mathbb{R}^n$, we have from LDP that

$$\lim_{\varepsilon\downarrow 0}\varepsilon\log\Pr(X_T\in A)=-\inf_{\substack{\boldsymbol{\phi}(0)=\boldsymbol{x},\\ \boldsymbol{\phi}(T)\in A}} S_T(\boldsymbol{\phi}). \tag{2.4}$$

This result can be explained as the probability of paths that start from $\boldsymbol{x}$ and end in A at time T is determined by the minimizer of the action functional.

When $\varepsilon\downarrow 0$, the time scale of some events will increase exponentially, e.g., exit of the domain of attraction of a stable equilibrium. We then need to generalize the fact that T is finite in Eq. (2.4) and define the quasi-potential from point $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ as

$$V(\boldsymbol{x}_1,\boldsymbol{x}_2)=\inf_{T>0}\inf_{\substack{\boldsymbol{\phi}(0)=\boldsymbol{x}_1,\\ \boldsymbol{\phi}(T)=\boldsymbol{x}_2}} S_T(\boldsymbol{\phi}). \tag{2.5}$$

The probability meaning of the quasi-potential is as follows

$$V(\boldsymbol{x}_1,\boldsymbol{x}_2)=\lim_{T\to\infty}\lim_{\delta\downarrow 0}\lim_{\varepsilon\downarrow 0}-\varepsilon\log\Pr(\tau_\delta\le T), \tag{2.6}$$

where τ_δ is the first entrance time of the δ-neighborhood of $\boldsymbol{x}_2$ for the process $\boldsymbol{X}_t$ starting from $\boldsymbol{x}_1$. In other words, the quasi-potential $V(\boldsymbol{x}_1,\boldsymbol{x}_2)$ characterizes the difficulty of passage from $\boldsymbol{x}_1$ to a small neighborhood of $\boldsymbol{x}_2$. For example, if there exist only two stable equilibrium states $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, the long-term stochastic dynamics can be reduced to a Markov chain on the state space $\{\boldsymbol{x}_1,\boldsymbol{x}_2\}$. According to the LDP [10], the transition rates are

$$k_{1\to 2}\asymp\exp\left(-\varepsilon^{-1}V(\boldsymbol{x}_1,\boldsymbol{x}_2)\right),\quad k_{2\to 1}\asymp\exp\left(-\varepsilon^{-1}V(\boldsymbol{x}_2,\boldsymbol{x}_1)\right), \tag{2.7}$$

where the arrow indicates the transition direction, and $f(\varepsilon)\asymp g(\varepsilon)$ if and only if $\frac{\log f(\varepsilon)}{\log g(\varepsilon)}\to 1$ as $\varepsilon\downarrow 0$.

3 Methodology

To address our methodology, we focus on the following optimization problem given by the LDP to characterize the transition from one state $\boldsymbol{a}_1$ to another state $\boldsymbol{a}_2$:

$$V(\boldsymbol{a}_1,\boldsymbol{a}_2)=S_{T^*}(\boldsymbol{\phi}^*)=\min_{T\in\mathbb{R}^+,\boldsymbol{\phi}\in A} S_T(\boldsymbol{\phi})=\min_{T\in\mathbb{R}^+,\boldsymbol{\phi}\in A}\int_0^T L(\dot{\boldsymbol{\phi}},\boldsymbol{\phi})dt, \tag{3.1}$$

where the constraints for the set A are

$$\boldsymbol{\phi}(0)=\boldsymbol{a}_1,\quad \boldsymbol{\phi}(T)=\boldsymbol{a}_2. \tag{3.2}$$

Here $\boldsymbol{\phi}^*$ indicates the minimizer, also called the minimal action path (MAP), subject to the optimal transition time $T^*\in(0,\infty]$.

Assumption 3.1. We assume that the transition path $\boldsymbol{\phi}(t)$ can be approximated by a finite dimensional approximation space $V=\{v: v\in\mathbb{R}^n, v_i\in \text{span}\{\psi_i(t)\}_{i=1}^M\}$ such that

$$\boldsymbol{\phi}(t)=\sum_{i=1}^{M}\boldsymbol{\phi}_i\psi_i(t),\quad \delta\boldsymbol{\phi}(t)=\sum_{i=1}^{M}\delta\boldsymbol{\phi}_i\psi_i(t), \tag{3.3}$$

where $\boldsymbol{\phi}_i,\delta\boldsymbol{\phi}_i\in\mathbb{R}^n$, and $\delta\boldsymbol{\phi}(t)$ indicates a perturbation function in V.

Assumption 3.2. If $T^*=\infty$, we assume that

$$S_T(\hat{\boldsymbol{\phi}}^*)\to S_{T^*=\infty}(\boldsymbol{\phi}^*)\quad \text{as } T\to\infty, \tag{3.4}$$

where $\hat{\boldsymbol{\phi}}^*$ is the minimizer of $S_T(\boldsymbol{\phi})$ for a fixed T.

Remark 3.1. The validity of Assumption 3.2 is based on the observations given by the adaptive minimum action method (aMAM) in [20], where it shows that the correct MAP and minimum action can be well approximated using a finite but large integration time T when $T^*=\infty$.

One strategy to deal with problem (3.1) is the geometric minimum action method (gMAM), where the action functional is reformulated with respect to a new parameter α, i.e., the scaled arc length, such that the integration time T is removed by a nonlinear mapping $\alpha(t)$ between $t\in[0,T]$ and $\alpha\in[0,1]$. The optimization in gMAM is with respect to curves that connect $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ parametrized by α, and the integration time T is recovered by the inverse mapping $t(\alpha)$.

In this work, we will address the optimization problem (3.1) with respect to time instead of arc length. Due to Assumption 3.2, we separate T from the integration by a simple linear scaling in contrast to the nonlinear mapping $\alpha(t)$ used in gMAM. More specifically, we consider a scaled time variable $s=\frac{t}{T}\in[0,1]$. Let $\frac{\delta S_T}{\delta\boldsymbol{\phi}}$ be the functional derivative of S_T with respect to $\boldsymbol{\phi}$, which satisfies

$$\delta S_T(\boldsymbol{\phi})=\langle\frac{\delta S_T(\boldsymbol{\phi})}{\delta\boldsymbol{\phi}},\delta\boldsymbol{\phi}\rangle_t=\lim_{\epsilon\to0}\frac{S_T(\boldsymbol{\phi}+\epsilon\delta\boldsymbol{\phi})-S_T(\boldsymbol{\phi})}{\epsilon},\quad T \text{ is fixed}, \tag{3.5}$$

with $\delta\boldsymbol{\phi}$ being an arbitrary perturbation testing function. Here $\langle\boldsymbol{f}_1(t),\boldsymbol{f}_2(t)\rangle_t$ indicates the inner product of vector functions $\boldsymbol{f}_1(t),\boldsymbol{f}_2(t)\in\mathbb{R}^n$ on the interval $t\in[0,T]$ and $\langle\boldsymbol{g}_1(s),\boldsymbol{g}_2(s)\rangle_s$ the inner product of vector functions $\boldsymbol{g}_1(s),\boldsymbol{g}_2(s)\in\mathbb{R}^n$ on the interval $s\in[0,1]$. We will use $\langle\boldsymbol{v},\boldsymbol{w}\rangle$ to indicate the inner product of vectors $\boldsymbol{v},\boldsymbol{w}\in\mathbb{R}^n$. From now on, we use $\dot{\boldsymbol{\phi}}$ to indicate the derivative with respect to t and $\boldsymbol{\phi}'$ the derivative with respect to s. We then have

$$\langle\boldsymbol{f}(t),\boldsymbol{g}(t)\rangle_t=\int_0^T\langle\boldsymbol{f}(t),\boldsymbol{g}(t)\rangle dt=T\int_0^1\langle\boldsymbol{f}(s),\boldsymbol{g}(s)\rangle ds=T\langle\boldsymbol{f}(s),\boldsymbol{g}(s)\rangle_s.$$

Lemma 3.1. *For any given $\boldsymbol{\phi}(s)\in V$, we have*

$$\frac{\partial S_T}{\partial T}=\langle \boldsymbol{b},\boldsymbol{b}\rangle_s - T^{-2}\langle \boldsymbol{\phi}',\boldsymbol{\phi}'\rangle_s = T^{-1}(\langle \boldsymbol{b},\boldsymbol{b}\rangle_t - \langle \dot{\boldsymbol{\phi}},\dot{\boldsymbol{\phi}}\rangle_t). \tag{3.6}$$

Proof. We consider the action functional in an abstract form

$$S_T(\boldsymbol{\phi})=\frac{T}{2}\int_0^1 L(T^{-1}\boldsymbol{\phi}',\boldsymbol{\phi})ds. \tag{3.7}$$

For a perturbation δT of T and a fixed path $\boldsymbol{\phi}$, we have

$$\begin{aligned} S_{T+\delta T}(\boldsymbol{\phi}) &= \frac{T+\delta T}{2}\int_0^1 L((T+\delta T)^{-1}\boldsymbol{\phi}',\boldsymbol{\phi})ds \\ &\approx \frac{T+\delta T}{2}\int_0^1 \left(L(T^{-1}\boldsymbol{\phi}',\boldsymbol{\phi}) - \frac{\partial L}{\partial(T^{-1}\boldsymbol{\phi}')}\cdot(T^{-2}\boldsymbol{\phi}')\delta T\right) ds, \end{aligned}$$

which results in the partial derivative

$$\frac{\partial S_T}{\partial T}=\frac{1}{2}\int_0^1 L(T^{-1}\boldsymbol{\phi}',\boldsymbol{\phi})ds - \frac{1}{2}\int_0^1 \frac{\partial L}{\partial(T^{-1}\boldsymbol{\phi}')}\cdot(T^{-1}\boldsymbol{\phi}')ds. \tag{3.8}$$

Substituting $L(T^{-1}\boldsymbol{\phi}',\boldsymbol{\phi})=\langle T^{-1}\boldsymbol{\phi}'-b,T^{-1}\boldsymbol{\phi}'-b\rangle$ into the above equation, we obtain the conclusion. □

For any given $\phi(s)$, Eq. (3.6) shows that S_T has a unique minimum at

$$T=\left(\frac{\langle \boldsymbol{\phi}',\boldsymbol{\phi}'\rangle_s}{\langle \boldsymbol{b},\boldsymbol{b}\rangle_s}\right)^{1/2}, \tag{3.9}$$

which is equivalent to the constraint

$$\langle \boldsymbol{b}(\boldsymbol{\phi}),\boldsymbol{b}(\boldsymbol{\phi})\rangle_t = \langle \dot{\boldsymbol{\phi}},\dot{\boldsymbol{\phi}}\rangle_t. \tag{3.10}$$

Note that if the following stronger constraint

$$\langle \boldsymbol{b}(\boldsymbol{\phi}),\boldsymbol{b}(\boldsymbol{\phi})\rangle = \langle \dot{\boldsymbol{\phi}},\dot{\boldsymbol{\phi}}\rangle, \quad \forall t \tag{3.11}$$

holds, $\partial S_T/\partial T=0$ for sure. The geometric minimum action method (gMAM) [13] shows that the constraint (3.11) is actually satisfied by the MAP under the assumption that the total arc length of the transition path is finite. In this paper, we will refer to Eq. (3.11) as the arc length constraint of the MAP. The constraint (3.10) is actually a weak version of the arc length constraint reflecting the effect of the linear time scaling.

Remark 3.2. The arc length constraint can be illustrated as follows. We have

$$\begin{aligned}\frac{1}{2}\int_0^T \langle \dot{\boldsymbol{\phi}}-b,\dot{\boldsymbol{\phi}}-b\rangle dt &= \frac{1}{2}\int_0^T (\langle \dot{\boldsymbol{\phi}},\dot{\boldsymbol{\phi}}\rangle + \langle \boldsymbol{b},\boldsymbol{b}\rangle - 2\langle \dot{\boldsymbol{\phi}},\boldsymbol{b}\rangle) dt \\ &\geq \int_0^T (|\dot{\boldsymbol{\phi}}||\boldsymbol{b}| - \langle \dot{\boldsymbol{\phi}},\boldsymbol{b}\rangle) dt = \int_0^T (|\dot{\boldsymbol{\phi}}||\boldsymbol{b}| - |\dot{\boldsymbol{\phi}}||\boldsymbol{b}|\cos\eta) dt \\ &= 2\int_0^T |\dot{\boldsymbol{\phi}}||\boldsymbol{b}|\sin^2\frac{\eta(t)}{2} dt = 2\int_0^C |\boldsymbol{b}|\sin^2\frac{\eta}{2} d\alpha,\end{aligned}$$

where α is the arc length, C the total arc length of the curve and η the angle between $\dot{\boldsymbol{\phi}}$ and $\boldsymbol{b}$. It is seen that if the arc length constraint (3.11) holds, the inequality in the above equation becomes equality and the action functional will depend on arc length instead of time. We refer to [13] for more details of gMAM.

To this end, we have some general comments about the approximation of problem (3.1). First, this problem has only been addressed by gMAM. The elegance of gMAM is that it proves that the pointwise arc length constraint (3.11) can be imposed such that the action functional only depends on a curve. This is an important observation because for any path connecting $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ the optimal way to associate an integration time T is given by the arc length constraint. Then the action functional can be minimized with respect to curves and the optimal integration time T^* is recovered by the nonlinear mapping between time and arc length. Second, to address problem (3.1) with respect to time instead of arc length, we must have Assumption 3.2 since T^* can be infinite. As long as Assumption 3.2 holds, we can consider the linear time scaling, which yields a weak version of the arc length constraint, i.e, Eq. (3.10). However, we expect that the arc length constraint (3.11) will be approached as the approximation space V is being refined.

Overall, due to Assumption 3.2, it is not necessary to reformulate the action functional with respect to arc length, however, the arc length constraint (3.11) given by gMAM will be used implicitly to compensate the time truncation in Assumption 3.2 through adaptive refinement of V (see Section 4 for algorithm development). This way, the procedure is simplified, which makes it more flexible to develop efficient solvers to minimize the action functional.

4 Minimum action method with optimal linear time scaling

In this section we will develop our minimum action method for problem (3.1) with respect to time.

4.1 Define a finite element approximation space V_h

We define the partition $\mathcal{T}_h$ for $s\in[0,1]$

$$\mathcal{T}_h: \quad 0=s_0<s_1<\cdots<s_N=1$$

and the associated finite element approximation space V_h

$$V_h^K=\{v: v\circ F_k^{-1}\in\mathscr{P}_p(R)\},$$
$$V_h=\{\boldsymbol{v}:\boldsymbol{v}\in\mathbb{R}^n, v_i\in H_0^1([0,1]), v_i|_K\in V_h^K, K\in\mathcal{T}_h\},$$

where F_K is the mapping function for the element $K=[s_i,s_{i+1}]$, $i=0,1,\cdots,N$, mapping the reference element $R=[-1,1]$ to element K, and $\mathscr{P}_p(R)$ denotes the set of polynomials of degree up to p over R. On the reference element R, we assume that $\mathscr{P}_p(R)$ consists of linear combinations of the following basis functions [17]:

$$\hat{\psi}_i(\tau)=\begin{cases}\frac{1-\tau}{2}, & i=0,\\ \frac{1-\tau}{2}\frac{1+\tau}{2}P_{i-1}^{1,1}(\tau), & 0<i<p,\\ \frac{1+\tau}{2}, & i=p,\end{cases} \tag{4.1}$$

where $P_i^{1,1}$ denote orthogonal Jacobi polynomials of degree i with respect to the weight function $(1-\tau)(1+\tau)$. $\hat{\psi}_0(\tau)$, and $\hat{\psi}_p(\tau)$ are consistent with linear finite element basis, and $\hat{\psi}_i(\tau)$, $0<i<p$, are introduced for high-order approximation. Note that $\hat{\psi}_i(\pm1)=0$ for $0<i<p$. We call $\hat{\psi}_0(\tau)$ and $\hat{\psi}_p(\tau)$ boundary modes, and $\hat{\psi}_i(\tau)$, $0<i<p$, interior modes.

The approximation space for $\boldsymbol{\phi}(s)$ is chosen as $V=V_h$. The discrete optimization problem is then defined as

$$S_{T_h^*}(\boldsymbol{\phi}_h^*)=\min_{T\in\mathbb{R}^+,\boldsymbol{\phi}_h\in A}\left[S_T(\boldsymbol{\phi}_h)=\frac{1}{2}T\langle T^{-1}\boldsymbol{\phi}_h'-\boldsymbol{b},T^{-1}\boldsymbol{\phi}_h'-\boldsymbol{b}\rangle_s\right], \tag{4.2}$$

where $\boldsymbol{\phi}_h^*$ is the approximated MAP and T_h^* the approximation of T^*.

4.2 Remove T from the objective functional

Noting that Eq. (3.9) provides an optimal integral time T for any given path $\boldsymbol{\phi}(s)$, we can treat T in the optimization problem (4.2) as a functional of $\boldsymbol{\phi}(s)$ such that we can remove T from the objective functional. In other words, we consider the following equivalent problem

$$\min_{\boldsymbol{\phi}_h\in A}\left[S_{T(\boldsymbol{\phi}_h)}(\boldsymbol{\phi}_h)=\frac{1}{2}T(\boldsymbol{\phi}_h)\langle T^{-1}(\boldsymbol{\phi}_h)\boldsymbol{\phi}_h'-\boldsymbol{b}(\boldsymbol{\phi}_h),T^{-1}(\boldsymbol{\phi}_h)\boldsymbol{\phi}_h'-\boldsymbol{b}(\boldsymbol{\phi}_h)\rangle_s\right], \tag{4.3}$$

where T is given by Eq. (3.9) and

$$\boldsymbol{\phi}_h(0)=\boldsymbol{a}_1,\quad \boldsymbol{\phi}_h(1)=\boldsymbol{a}_2.$$

This way, the optimization will be only with respect to the path $\boldsymbol{\phi}(s)$.

Lemma 4.1. *Let*

$$\boldsymbol{\phi}_h(s)=\sum_{i=1}^{M}\boldsymbol{\phi}_i\psi_i(s)\in V_h. \tag{4.4}$$

We have

$$\frac{\partial S_T}{\partial \phi_{i,j}} = T\langle T^{-1}\boldsymbol{\phi}_h' - \boldsymbol{b}, T^{-1}\psi_i' \boldsymbol{e}_j - \nabla_{\boldsymbol{\phi}_h}\boldsymbol{b}\psi_i \boldsymbol{e}_j\rangle_s, \tag{4.5}$$

where $T(\boldsymbol{\phi}_h)$ is given by Eq. (3.9).

Proof. For a perturbation $\delta\boldsymbol{\phi}_h$, the linear perturbation of $\boldsymbol{b}(\boldsymbol{\phi}_h+\delta\boldsymbol{\phi}_h)$ is

$$\boldsymbol{b}(\boldsymbol{\phi}_h+\delta\boldsymbol{\phi}_h) \approx \boldsymbol{b}(\boldsymbol{\phi}_h) + \nabla_{\boldsymbol{\phi}_h}\boldsymbol{b}\delta\boldsymbol{\phi}_h, \quad \forall s.$$

Let

$$\delta T = T(\boldsymbol{\phi}_h+\delta\boldsymbol{\phi}_h) - T(\boldsymbol{\phi}_h),$$

which yields that

$$\delta(T^{-1}) = -T^{-2}\delta T.$$

Then the variation of $S_T(\boldsymbol{\phi}_h)$ is

$$\begin{aligned}
&S_{T+\delta T}(\boldsymbol{\phi}_h+\delta\boldsymbol{\phi}_h) - S_T(\boldsymbol{\phi}_h)\\
=&2^{-1}(T+\delta T)\Big\langle (T^{-1}-T^{-2}\delta T)(\boldsymbol{\phi}_h'+\delta\boldsymbol{\phi}_h') - \boldsymbol{b} - \nabla_{\boldsymbol{\phi}_h}\boldsymbol{b}\delta\boldsymbol{\phi}_h,\\
&(T^{-1}-T^{-2}\delta T)(\boldsymbol{\phi}_h'+\delta\boldsymbol{\phi}_h') - \boldsymbol{b} - \nabla_{\boldsymbol{\phi}_h}\boldsymbol{b}\delta\boldsymbol{\phi}_h\Big\rangle_s - 2^{-1}T\langle T^{-1}\boldsymbol{\phi}_h' - \boldsymbol{b}, T^{-1}\boldsymbol{\phi}_h' - \boldsymbol{b}\rangle_s\\
&+ \text{high-order terms}\\
=&T\langle T^{-1}\boldsymbol{\phi}_h' - \boldsymbol{b}, T^{-1}\delta\boldsymbol{\phi}_h' - T^{-2}\boldsymbol{\phi}_h'\delta T - \nabla_{\boldsymbol{\phi}_h}\boldsymbol{b}\delta\boldsymbol{\phi}_h\rangle_s + 2^{-1}\delta T\langle T^{-1}\boldsymbol{\phi}_h' - \boldsymbol{b}, T^{-1}\boldsymbol{\phi}_h' - \boldsymbol{b}\rangle_s\\
&+ \text{high-order terms}.
\end{aligned}$$

We look at the terms related to δT and have

$$\begin{aligned}
&-T\langle T^{-1}\boldsymbol{\phi}_h' - b, T^{-2}\boldsymbol{\phi}_h'\delta T\rangle_s + 2^{-1}\delta T\langle T^{-1}\boldsymbol{\phi}_h' - b, T^{-1}\boldsymbol{\phi}_h' - \boldsymbol{b}\rangle_s\\
=&-2^{-1}\delta T\langle T^{-1}\boldsymbol{\phi}_h' - \boldsymbol{b}, T^{-1}\boldsymbol{\phi}_h' + \boldsymbol{b}\rangle_s\\
=&-2^{-1}\delta T(\langle T^{-1}\boldsymbol{\phi}_h', T^{-1}\boldsymbol{\phi}_h'\rangle_s - \langle \boldsymbol{b}, \boldsymbol{b}\rangle_s) = 0,
\end{aligned}$$

where in the last step we used Eq. (3.9). Thus, we have

$$\delta S_T = T\langle T^{-1}\boldsymbol{\phi}_h' - \boldsymbol{b}, T^{-1}\delta\boldsymbol{\phi}_h' - \nabla_{\boldsymbol{\phi}_h}\boldsymbol{b}\delta\boldsymbol{\phi}_h\rangle_s.$$

We then obtain the desired gradient by choosing $\delta\boldsymbol{\phi}_h = \delta\phi_{i,j}\psi_i\boldsymbol{e}_j$. □

To this end, a gradient-type optimization algorithm, such as nonlinear conjugate gradient and BFGS, etc., can be employed to solve the discretized problem (4.3) [18]. Once the MAP $\boldsymbol{\phi}_h^*$ is obtained, the optimal integration time is approximated as $T_h^* = T(\boldsymbol{\phi}_h^*)$ by Eq. (3.9). From now on, we will use tMAM to indicate the proposed minimum action method with optimal linear time scaling.

Remark 4.1. The formula given in Eq. (4.5) has the same form as the gradient of action functional for a fixed integration time T [23, 25], except that $T(\boldsymbol{\phi}_h)$ is now a functional of $\boldsymbol{\phi}_h$. This implies that the optimization problems given by Eq. (2.4) with a fixed T and Eq. (2.5) with a free T can both be dealt with by tMAM using the same optimization algorithm subject to a slight modification that T is updated with respect to the transition path or not. We note that gMAM is not able to minimize the action functional for a given T and aMAM is not able to deal with quasi-potential subject to a finite optimal integration time. In this sense, tMAM is more flexible than both aMAM and gMAM.

Lemma 4.2. *Let $\boldsymbol{\phi}^*$ be a local minimizer of problem* (3.1). *Assume that $\boldsymbol{\phi}^* \in H^1(\mathbb{R}^n;[0,1]) \cap C^1(\mathbb{R}^n;[0,1])$ and $\boldsymbol{b}(\boldsymbol{\phi}) \in C^2(\mathbb{R}^n;[0,1])$. The optimal convergence rate given by V_h with a fixed polynomial order p is*

$$|S_{T^*}(\boldsymbol{\phi}^*) - S_{T(\boldsymbol{\phi}_h^*)}(\boldsymbol{\phi}_h^*)| \sim \mathcal{O}(N^{-2p}), \tag{4.6}$$

where N is the number of elements.

Proof. We only need to focus on the case that T^* is finite, where $S_{T^*}(\boldsymbol{\phi}^*) = S_{T(\boldsymbol{\phi}^*)}(\boldsymbol{\phi}^*)$. Let $\boldsymbol{\phi}_h^* = \boldsymbol{\phi}^* + (\boldsymbol{\phi}_h^* - \boldsymbol{\phi}^*) = \boldsymbol{\phi}^* + \delta\boldsymbol{\phi}^*$.

$$S_{T(\boldsymbol{\phi}_h^*)}(\boldsymbol{\phi}_h^*) - S_{T^*}(\boldsymbol{\phi}^*) = S_{T(\boldsymbol{\phi}^* + \delta\boldsymbol{\phi}^*)}(\boldsymbol{\phi}^* + \delta\boldsymbol{\phi}^*) - S_{T(\boldsymbol{\phi}^*)}(\boldsymbol{\phi}^*).$$

Since $\boldsymbol{\phi}^*$ is a local minimizer, $\delta S_T = T\langle T^{-1}(\boldsymbol{\phi}^*)' - \boldsymbol{b}, T^{-1}(\delta\boldsymbol{\phi}^*)' - (\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})(\delta\boldsymbol{\phi}^*)\rangle_s = 0$. We need to consider the second-order variation. We first perturb δS_T using a perturbation function $\delta\boldsymbol{\psi}^*$. It can be shown that

$$\begin{aligned}
&\delta S_T(\boldsymbol{\phi}^* + \delta\boldsymbol{\psi}^*, \delta\boldsymbol{\phi}^*) - \delta S_T(\boldsymbol{\phi}^*, \delta\boldsymbol{\phi}^*) \\
=&T\langle T^{-1}(\delta\boldsymbol{\psi}^*)' - (\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})\delta\boldsymbol{\psi}^*, T^{-1}(\delta\boldsymbol{\phi}^*)' - (\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})\delta\boldsymbol{\phi}^*\rangle_s \\
&- T\langle T^{-1}(\boldsymbol{\phi}^*)' - \boldsymbol{b}, B(\boldsymbol{b}, \delta\boldsymbol{\phi}^*)\delta\boldsymbol{\psi}^*\rangle_s - T^{-1}\langle \boldsymbol{b}, \boldsymbol{b}\rangle_s \delta_T(\delta\boldsymbol{\psi}^*)\delta_T(\delta\boldsymbol{\phi}^*),
\end{aligned}$$

where $B(\boldsymbol{b}, \delta\boldsymbol{\phi}^*)\delta\boldsymbol{\psi}^*$ indicates the first-order perturbation of the term $\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b}\delta\boldsymbol{\phi}^*$ induced by $\delta\boldsymbol{\psi}^*$, and $B(\boldsymbol{b}, \delta\boldsymbol{\phi}^*) \in \mathbb{R}^{n\times n}$, and

$$B_{i,j} = \sum_{k=1}^{n} (\nabla_{\boldsymbol{\phi}^*}\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})_{ikj}\delta\phi_k^* = \sum_{k=1}^{n} \frac{\partial^2 b_i}{\partial\phi_k^*\partial\phi_j^*}\delta\phi_k^*, \quad i,j = 1,\cdots,n,$$

and

$$\delta_T(\delta\boldsymbol{\phi}^*) = T(\boldsymbol{\phi}^* + \delta\boldsymbol{\phi}^*) - T(\boldsymbol{\phi}^*) = \frac{\langle(\boldsymbol{\phi}^*)', (\delta\boldsymbol{\phi}^*)'\rangle_s - T^2\langle \boldsymbol{b}, (\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})(\delta\boldsymbol{\phi}^*)\rangle_s}{T\langle \boldsymbol{b}, \boldsymbol{b}\rangle_s}.$$

Then the second-order variation is

$$\begin{aligned}
2\delta^2 S_T =& T\langle T^{-1}(\delta\boldsymbol{\phi}^*)' - (\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})\delta\boldsymbol{\phi}^*, T^{-1}(\delta\boldsymbol{\phi}^*)' - (\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})\delta\boldsymbol{\phi}^*\rangle_s \\
&- T\langle T^{-1}(\boldsymbol{\phi}^*)' - \boldsymbol{b}, B(\boldsymbol{b}, \delta\boldsymbol{\phi}^*)\delta\boldsymbol{\phi}^*\rangle_s - T^{-1}\langle \boldsymbol{b}, \boldsymbol{b}\rangle_s (\delta_T(\delta\boldsymbol{\phi}^*))^2.
\end{aligned}$$

From the approximation theory, we know that

$$\langle \delta\boldsymbol{\phi}^*, \delta\boldsymbol{\phi}^* \rangle_s = \langle \boldsymbol{\phi}_h^* - \boldsymbol{\phi}^*, \boldsymbol{\phi}_h^* - \boldsymbol{\phi}^* \rangle_s \sim \mathcal{O}(N^{-2(p+1)}), \quad \langle (\delta\boldsymbol{\phi}^*)', (\delta\boldsymbol{\phi}^*)' \rangle_s \sim \mathcal{O}(N^{-2p}).$$

Using Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&|\langle T^{-1}(\delta\boldsymbol{\phi}^*)', T^{-1}(\delta\boldsymbol{\phi}^*)' \rangle_s| \sim \mathcal{O}(N^{-2p}),\\
&|\langle T^{-1}(\delta\boldsymbol{\phi}^*)', (\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})\delta\boldsymbol{\phi}^* \rangle_s| \sim \mathcal{O}(N^{-2p-1}),\\
&|\langle (\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})\delta\boldsymbol{\phi}^*, (\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})\delta\boldsymbol{\phi}^* \rangle_s| = |\langle \delta\boldsymbol{\phi}^*, (\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})^{\mathsf{T}}(\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})\delta\boldsymbol{\phi}^* \rangle_s| \sim \mathcal{O}(N^{-2p-2}),\\
&|\langle T^{-1}(\boldsymbol{\phi}^*)' - \boldsymbol{b}, B(\boldsymbol{b}, \delta\boldsymbol{\phi}^*)\delta\boldsymbol{\phi}^* \rangle_s| = |\langle \delta\boldsymbol{\phi}^*, A(\boldsymbol{\phi}^*)\delta\boldsymbol{\phi}^* \rangle_s| \sim \mathcal{O}(N^{-2p-2})\\
&|\delta_T| \sim |\langle (\boldsymbol{\phi}^*)', (\delta\boldsymbol{\phi}^*)' \rangle_s| \sim \mathcal{O}(N^{-p}),
\end{aligned}$$

where we assume that the matrices $\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b}$, $(\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})^{\mathsf{T}}(\nabla_{\boldsymbol{\phi}^*}\boldsymbol{b})$ and $A(\boldsymbol{\phi}^*)$ are uniformly bounded. Overall, we then have

$$|\delta^2 S_T(\boldsymbol{\phi}^*)| \sim \mathcal{O}(N^{-2p}).$$

The proof is completed. □

4.3 Why does tMAM with equidistant elements work

We know that for the case $T^* = \infty$, the original MAM [5] with equidistant elements does not work due to the clustering problem although a large T is used, which is a fundamental numerical difficulty induced by the separation of slow and fast dynamics.

For example, we consider the ODE example in Section 6.1. If we use the two critical points $\boldsymbol{a}_1 = (-1,0)$ and $\boldsymbol{a}_3 = (0,1)$, where $\boldsymbol{a}_1$ is a stable fixed point and $\boldsymbol{a}_3$ is a saddle point located on the basin of attraction of $\boldsymbol{a}_1$, as the two ends of transition paths, T^* should be ∞. However, if we move $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$ slightly to $\hat{\boldsymbol{a}}_1 = (\cos(\pi - 0.1), \sin(\pi - 0.1))$ and $\hat{\boldsymbol{a}}_3 = (\cos(\pi/2 + 0.1), \sin(\pi/2 + 0.1))$, respectively, along the MAP connecting $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$, T^* for the MAP connecting $\hat{\boldsymbol{a}}_1$ and $\hat{\boldsymbol{a}}_3$ becomes from ∞ to about 2.3 (see Fig. 3) while the corresponding action only drops from 2 to 1.96. Thus resolving the region of fast dynamics between $\hat{\boldsymbol{a}}_1$ and $\hat{\boldsymbol{a}}_3$ is critical to capture the MAP connecting $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$. If we choose T=100 and 100 equidistant linear finite elements to discretize the MAP connecting $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$, each element has a length 1 and we only have two elements to cover the region of fast dynamics between $\hat{\boldsymbol{a}}_1$ and $\hat{\boldsymbol{a}}_3$ and all other points will be clustered around $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$.

To resolve the clustering problem for a fixed T, the adaptive MAM (aMAM) was developed in [20]. The key idea of aMAM is to redistribute the grid points with respect to arc length such that more points will be moved into the region of fast dynamics. Numerical experiments show that such a strategy is crucial for accuracy. However, the reparametrization is a global operation involving re-meshing and projection of the path from the old mesh to the new one, which becomes a bottleneck for scalability if parallel computing is considered [25].

For tMAM, we consider the optimization problem (4.3). The most important difference of tMAM from the original MAM and aMAM is that the integration time T is decided optimally by the path itself instead of being fixed. There are several benefits about relaxing the integration time T. First, for any given ϕ_h, we have

$$S_{T(\boldsymbol{\phi}_h)}(\boldsymbol{\phi}_h) \leq S_{T=C}(\boldsymbol{\phi}_h),$$

where the equality only holds if $T(\boldsymbol{\phi}_h)=C$. If $\boldsymbol{\phi}_h^*$ is the MAP returned by aMAM, we can immediately obtain a smaller action as

$$S_{T(\boldsymbol{\phi}_h^*)}(\boldsymbol{\phi}_h^*) \leq S_{T=C}(\boldsymbol{\phi}_h^*).$$

In other words, although the assumption that T should be large enough for numerical approximation is consistent with the real dynamics, it does not provide the most effective numerical way to reduce the value of the action functional. Second, The clustering problem can be alleviated by the optimal linear time scaling. We still consider the aforementioned SODE example. If we use 100 equidistant linear finite elements to minimize $S_{T(\boldsymbol{\phi}_h)}(\boldsymbol{\phi}_h)$, we obtain $T_h^*=5.67$ for the MAP from $\boldsymbol{a}_1$ to $\boldsymbol{a}_3$. In other words, we have about $2.3/(5.67/100)\approx 40$ elements to cover the region of fast dynamics between $\hat{\boldsymbol{a}}_1$ and $\hat{\boldsymbol{a}}_3$. Thus, tMAM is able to put quite a number of elements into the region of fast dynamics by the optimal linear time scaling.

Numerical experiments show that tMAM with equidistant elements converges algebraically even for the case $T^*=\infty$. However, due to the singularity induced by $T^*=\infty$, tMAM with equidistant elements is not able to reach the optimal convergence rate given by V_h (see Lemma 4.2). For example, if we use linear finite elements, the optimal convergence rate should be $\mathcal{O}(N^{-2})$ for the error of $S_{T(\boldsymbol{\phi}_h^*)}(\boldsymbol{\phi}_h^*)$, which can be achieved by both aMAM and gMAM with global reparametrization. The tMAM with equidistant elements is converging at a rate $\mathcal{O}(N^{-\alpha})$ with $\alpha<2$ (see Section 6.1 for detailed numerical experiments). This is the price we need to pay to just consider a linear time scaling. Then the main task of this paper is to verify if an adaptive refinement of V_h is able to recover the optimal convergence rate.

We now give an intuitive argument about the comparison between gMAM and tMAM. We still use the aforementioned MAP connecting $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$ as an example, which does not pass any other critical points. The reason that T^* is infinite is that the Jacobin between time and arc length is singular at critical points. According to the arc length constraint, we have $d\alpha=|\dot{\boldsymbol{\phi}}^*|dt=|\boldsymbol{b}(\boldsymbol{\phi}^*)|dt$ and

$$T^*=\int_{\boldsymbol{\phi}^*}\frac{d\alpha}{|\boldsymbol{b}(\boldsymbol{\phi})|}, \tag{4.7}$$

where α is the arc length. If $\boldsymbol{b}(\boldsymbol{\phi})\in C^1(\mathbb{R}^n)$, we then have $|\boldsymbol{b}(\boldsymbol{\phi})|\sim C\alpha$ around $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$, which implies that $T^*=\infty$. We now split the MAP into three pieces $\boldsymbol{a}_1\to\hat{\boldsymbol{a}}_1\to\hat{\boldsymbol{a}}_3\to\boldsymbol{a}_3$. We connect $\boldsymbol{a}_1$ and $\hat{\boldsymbol{a}}_1$ using a linear path and a finite time, say 1, and do the same thing for $\hat{\boldsymbol{a}}_3$ and $\boldsymbol{a}_3$. The MAP from $\hat{\boldsymbol{a}}_1$ to $\hat{\boldsymbol{a}}_3$ can be captured by tMAM since the optimal integration

time, say $\hat{T}^*$, is finite. We now have an artificial path connecting $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$ subject to a finite integration time. First, if we let $\hat{\boldsymbol{a}}_1$ go to $\boldsymbol{a}_1$ and $\hat{\boldsymbol{a}}_3$ to $\boldsymbol{a}_3$, this path converges to the MAP with a finite optimal integration time $\hat{T}^*+2$; Second, the action given by this path should be larger than that given by tMAM since it might not be the optimal linear time scaling. Considering that the linear time scaling in tMAM is not able to handle the details around $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$, the action given by tMAM should be larger than that given by gMAM. However, this error can be limited to the vicinity of $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$, where only a small action is needed since $\dot{\boldsymbol{\phi}} \approx 0$ and $\boldsymbol{b}(\boldsymbol{\phi}) \approx 0$ around critical points. We also note that although gMAM preserves exactly the nonlinear mapping between time and arc length, the fact that the Jacobin between time and arc length is singular at critical points brings troubles to numerical computations especially when unknown critical points are located on the MAP [13]. For tMAM, this is not an issue because it uses a linear mapping.

4.4 Refine the approximation space adaptively

Before we address the adaptivity of tMAM, we clarify the difference between the adaptivity for tMAM and that for aMAM. In aMAM, the adaptivity means re-meshing, where a new global partition $\mathcal{T}_h$ will be generated for a fixed element number N. In the language of adaptive finite element method, it corresponds to the *r*-adaptivity. For tMAM, we will not consider re-meshing. Instead, we will consider *h*-adaptivity. In other words, we refine the current partition $\mathcal{T}_h$ by decomposing a certain number of elements. The reasoning of *h*-adaptivity is that tMAM can balance the element distribution in regions of fast and slow dynamics by optimal linear time scaling, and we do not expect that such a balance will change significantly. In other words, most of the elements in the region of fast dynamics will remain in the region of fast dynamics after $\mathcal{T}_h$ is refined. Similar arguments are valid for the region of slow dynamics.

4.4.1 An error indicator for the effect of linear time scaling

In this work, we will focus on *h*-refinement. It is proved in the gMAM [13] that the arc length constraint (3.11) should hold pointwisely for the quasi-potential. In contrast, the optimal integration time T of tMAM satisfies Eq. (3.9), which is weaker than the arc length constraint (3.11). We then propose to use the following quantity θ as an error indicator for $\boldsymbol{\phi}_h$

$$\theta^2 = T(\boldsymbol{\phi}_h)\int_0^1 (|\boldsymbol{b}(\boldsymbol{\phi}_h)| - T^{-1}(\boldsymbol{\phi}_h)|\boldsymbol{\phi}_h'|)^2 ds = \int_0^T (|\boldsymbol{b}(\boldsymbol{\phi}_h)| - |\dot{\boldsymbol{\phi}}_h|)^2 dt, \tag{4.8}$$

which measures the deviation of Assumption 3.2 to the arc length constraint.

Lemma 4.3. *Let* $V_h^0 = \{\boldsymbol{\phi}_h | \boldsymbol{\phi}_h \in V_h, \boldsymbol{\phi}_h(0) = \boldsymbol{a}_1, \boldsymbol{\phi}_h(1) = \boldsymbol{a}_2\}$. *If* $T^{-1}(\boldsymbol{\phi}_h)|\boldsymbol{\phi}_h'| = |\boldsymbol{b}(\boldsymbol{\phi}_h)|$ *does not hold pointwisely for any* $\boldsymbol{\phi}_h \in V_h^0$, θ *defined in Eq.* (4.8) *can serve as an error indicator in the sense that if* $S_{T(\boldsymbol{\phi}_h^*)}(\boldsymbol{\phi}_h^*) \to S_{T^*}(\boldsymbol{\phi}^*)$, *then* $\theta \to 0$.

Proof. Following is a general argument, where we will not focus on the mathematical details. In the Lemma, we assume that $T^{-1}(\boldsymbol{\phi}_h)|\boldsymbol{\phi}_h'| = |\boldsymbol{b}(\boldsymbol{\phi}_h)|$ does not hold pointwisely

for any $\boldsymbol{\phi}_h \in V_h^0$. Since $\boldsymbol{\phi}_h \in V_h^0$ is a piecewise polynomial, $\boldsymbol{\phi}_h'$ can be discontinuous at points $s = s_i$, $i = 1, \cdots, N-1$. If it is so, the continuity of $|\boldsymbol{b}(\boldsymbol{\phi}_h)|$ implies the arc length constraint will break in the vicinity of s_i. Within each element, $\boldsymbol{\phi}_h$ is a polynomial of finite order. If $|\boldsymbol{b}(\boldsymbol{\phi}_h)|^2$ is not a polynomial, the arc length constraint cannot hold since $T^{-2}|\boldsymbol{\phi}_h'|^2$ is a polynomial. Let $|\boldsymbol{b}(\boldsymbol{\phi}_h)|^2$ be a polynomial. If the arc length constraint holds, then they should belong to the same space $\{s^n\}_{n=0}^{2p-2}$, which is, in general, not true if $\boldsymbol{b}(\boldsymbol{\phi}_h)$ is linear or has a polynomial nonlinearity. In other words, the validity of the assumption is related to the definition of $\boldsymbol{b}(\boldsymbol{\phi}_h)$ and can be checked once it is specified. We will focus on the cases that our assumption holds.

We now consider the regular case where $T^* < \infty$. According to the definition of θ, it reaches to its minimum 0 when $|\boldsymbol{b}(\boldsymbol{\phi})| = T^{-1}(\boldsymbol{\phi})|\boldsymbol{\phi}'|$ pointwisely. At $\boldsymbol{\phi}^*$, $T(\boldsymbol{\phi}^*) = T^*$ and $|\boldsymbol{b}(\boldsymbol{\phi})| = (T^*)^{-1}|\boldsymbol{\phi}'|$ holds pointwisely due to the arc length constraint (3.11), which yields that $\theta(\boldsymbol{\phi}^*) = 0$. If $|\boldsymbol{b}(\boldsymbol{\phi})| \neq T^{-1}(\boldsymbol{\phi})|\boldsymbol{\phi}'|$ at a certain point, due to the continuity $|\boldsymbol{b}(\boldsymbol{\phi})| \neq T^{-1}(\boldsymbol{\phi})|\boldsymbol{\phi}'|$ holds in a small open interval containing this point, which results in $\theta(\boldsymbol{\phi}) \neq 0$. In other words, $\theta(\boldsymbol{\phi}_h) \neq 0$ for any $\boldsymbol{\phi}_h \in V_h^0$ due to the assumption.

We then consider θ in the space $\hat{V} = \boldsymbol{\phi}^* \oplus V_h^0$, where $\boldsymbol{\phi}^*$ is the unique minimizer of θ in $\hat{V}$ since no element in V_h^0 satisfies the arc length constraint (3.11). Let $\boldsymbol{\phi}_h^* \in V_h^0$ be the approximation of $\boldsymbol{\phi}^*$ given by the optimization algorithm. We now look at

$$\theta^2(\boldsymbol{\phi}_h^*) = \int_0^{T(\boldsymbol{\phi}_h^*)} (|\boldsymbol{b}(\boldsymbol{\phi}_h^*)| - |\dot{\boldsymbol{\phi}}_h^*|)^2 dt.$$

We regard $\boldsymbol{\phi}_h^*$ as $\boldsymbol{\phi}_h^* = \boldsymbol{\phi}^* - (\boldsymbol{\phi}^* - \boldsymbol{\phi}_h^*) = \boldsymbol{\phi}^* + \delta\boldsymbol{\phi}_h^*$, where $\delta\boldsymbol{\phi}_h^* = \boldsymbol{\phi}_h^* - \boldsymbol{\phi}^* \in \hat{V}$ is treated as a small perturbation determined by the accuracy of the discrete optimization problem (4.3). Then we have

$$\theta^2(\boldsymbol{\phi}_h^*) = \theta^2(\boldsymbol{\phi}^* + \delta\boldsymbol{\phi}_h^*) - 0 = \theta^2(\boldsymbol{\phi}^* + \delta\boldsymbol{\phi}_h^*) - \theta^2(\boldsymbol{\phi}^*).$$

Since $\boldsymbol{\phi}^* \in \hat{V}$ is the minimizer of θ^2, it is not difficult to verify that $\delta\theta^2(\boldsymbol{\phi}^*) = 0$. Then $\theta^2(\boldsymbol{\phi}_h^*)$ is determined by the second-order variation $\delta^2\theta^2(\boldsymbol{\phi}^*)$, which should be always positive if $\delta\boldsymbol{\phi}_h^*$ is not trivial. Let $P_{V_h^0}\boldsymbol{\phi}^*$ indicates the projection of $\boldsymbol{\phi}^*$ onto V_h^0, then

$$\delta\boldsymbol{\phi}_h^* = P_{V_h^0}\boldsymbol{\phi}^* - \boldsymbol{\phi}_h^* + (I - P_{V_h^0})\boldsymbol{\phi}^*,$$

where I is an identity operator. Let $\delta\Phi$ indicates all the coefficients of $(P_{V_h^0}\phi^* - \phi_h^*) \in V_h^0$. In the matrix form, we have

$$\theta^2(\boldsymbol{\phi}_h^*) = \delta\Phi^{\mathsf{T}} A \delta\Phi + \mathcal{O}(|\delta\Phi|^3),$$

where A is a positive definite matrix. In other words, as V_h^0 is refined and $\boldsymbol{\phi}_h^*$ goes to $P_{V_h^0}\boldsymbol{\phi}^*$, θ goes to zero correspondingly because A is positive definite. In this sense, θ can serve as an error indicator.

We now consider $T^* = \infty$. We first assume that $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are critical points, and the MAP does not pass any other critical points. We then break the MAP into three parts:

$\boldsymbol{a}_1 \to \hat{\boldsymbol{a}}_1 \to \hat{\boldsymbol{a}}_2 \to \boldsymbol{a}_2$, where $\hat{\boldsymbol{a}}_1$ and $\hat{\boldsymbol{a}}_2$ are two arbitrary points satisfying $|\boldsymbol{a}_1 - \hat{\boldsymbol{a}}_1| = \rho$ and $|\boldsymbol{a}_2 - \hat{\boldsymbol{a}}_2| = \rho$. If we let $\rho \to 0$, the MAP from $\boldsymbol{a}_1$ to $\boldsymbol{a}_2$ will be recovered. Note that T^* for the MAP from $\hat{\boldsymbol{a}}_1$ to $\hat{\boldsymbol{a}}_2$ is finite, the aforementioned argument is valid for this part. For the MAPs $\boldsymbol{a}_1 \to \hat{\boldsymbol{a}}_1$ and $\hat{\boldsymbol{a}}_2 \to \boldsymbol{a}_2$, we can just connect them using straight lines with a fixed time $T = 1$. As $\rho \to 0$, θ goes to zero, because $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ are critical points. If the MAP passes a critical point, such a strategy is also valid and we only need to break the MAP to more parts to separate the critical points. □

4.4.2 An adaptivity criterion

Although θ can be used to measure the deviation from the arc length constraint, it is, in general, not an effective measure of the distance between $\boldsymbol{\phi}_h^*$ and $\boldsymbol{\phi}^*$. Thus we still need an indicator of the error $\boldsymbol{\phi}^* - \boldsymbol{\phi}_h^*$. In this paper, we consider a simple indicator based on the following physical intuition [20]. When $T^* = \infty$, there exist two main difficulties to achieve a good accuracy in action functional. First, fast dynamics contributes most to the action functional, which needs to be well resolved. Second, slow dynamics introduces singularity in T, but fortunately its contribution to the action functional is small.

When the action functional is formulated with respect to time, elements in the region of fast dynamics are more likely to have a large arc length. To well capture the fast dynamics, we need to refine these elements. The arc length of element $[s_i, s_{i+1}]$ is

$$\hat{s}_i = \int_{Ts_i}^{Ts_{i+1}} |\dot{\boldsymbol{\phi}}_h| dt = \int_{s_i}^{s_{i+1}} |\boldsymbol{\phi}_h'| ds. \tag{4.9}$$

We define

$$\hat{s}_{\max} = \max\{\hat{s}_i\}, \quad i = 0, 1, \cdots, N-1, \tag{4.10}$$

where $\hat{s}_{\max}$ is the largest elementwise arc length. We then define a threshold $0 < r_{\hat{s}} < 1$. For element i, if $\hat{s}_i \geq r_{\hat{s}} \hat{s}_{\max}$, we will refine the element to capture fast dynamics better.

The arc length does not make much sense for slow dynamics. However, due to the singularity in T, the slow dynamics can be important for accuracy although its contribution is relatively small. To deal with the slow dynamics, we use the parameter θ. We define for each element

$$\theta_i^2 = T(\phi_h) \int_{s_i}^{s_{i+1}} (|\boldsymbol{b}(\boldsymbol{\phi}_h)| - T^{-1}(\boldsymbol{\phi}_h) |\boldsymbol{\phi}_h'|)^2 ds, \quad i = 0, \cdots, N-1, \tag{4.11}$$

and

$$\theta_{\max} = \max\{\theta_i\}, \quad i = 0, 1, \cdots, N-1. \tag{4.12}$$

An element will be refined if $\theta_i \geq r_\theta \theta_{\max}$, where the threshold $0 < r_\theta < 1$.

Elements associated with fast dynamics may have large θ_i and $\hat{s}_i$ at the same time while elements associated with slow dynamics are more likely to have a large θ_i but a small $\hat{s}_i$. To deal with both fast and slow dynamics, we simply decompose element i to two equidistant elements if the following criterion

$$\hat{s}_i \geq r_s \hat{s}_{\max} \quad \text{or} \quad \theta_i \geq r_\theta \theta_{\max} \tag{4.13}$$

is satisfied. Our main concern is that whether the adaptive tMAM can recover the optimal h-convergence rate given by the approximation space V_h.

Algorithm 1 Adaptive tMAM

Initialize $\mathcal{T}_h^{(0)}$ using a relatively coarse partition of the interval $[0,1]$.
Start a loop for adaptivity in terms of the index k.

- Solve optimization problem (4.3) for $\boldsymbol{\phi}_h^{*,(k)}$ using current $\mathcal{T}_h^{(k)}$.
- Compute elementwisely $\{\hat{s}_i\}_{i=0}^{N-1}$ and $\{\theta_i\}_{i=0}^{N-1}$.
- Check the adaptivity criterion (4.13) for mesh refinement to obtain $\mathcal{T}_h^{(k+1)}$.
- Project the current path $\boldsymbol{\phi}_h^{*,(k)}$ onto $\mathcal{T}_h^{(k+1)}$ as the initial path.

Stop the loop if one of the following conditions is satisfied:

- The index k reaches its prescribed maximum.
- The number of degrees of freedom becomes too large.
- The difference between $S_{T(\boldsymbol{\phi}_h^{*,(k)})}(\boldsymbol{\phi}_h^{*,(k)})$ and $S_{T(\boldsymbol{\phi}_h^{*,(k+1)})}(\boldsymbol{\phi}_h^{*,(k+1)})$ is small enough.

Remark 4.2. A more general definition of $\hat{s}_i$ is

$$\hat{s}_i = \int_{Ts_i}^{Ts_{i+1}} w(t)dt = \int_{Ts_i}^{Ts_{i+1}} \sqrt{1+C|\dot{\boldsymbol{\phi}}_h|^2}dt, \tag{4.14}$$

where C is a positive number [20]. When C goes to infinity, $w(t) \sim |\dot{\boldsymbol{\phi}}_h|$.

5 Spatially extended systems

We consider

$$\partial_t u(t,\boldsymbol{x}) + \mathcal{G}u(t,\boldsymbol{x}) = \sqrt{\varepsilon}\dot{W}(t,\boldsymbol{x}), \tag{5.1}$$

where $\boldsymbol{x} \in D \subset \mathbb{R}^d$ is the space variable, $\mathcal{G}$ is a spatial differentiation operator, and $W(\boldsymbol{x},t)$ is space-time white noise. The F-W action functional is defined as [3,9]

$$S_T(u) = \frac{1}{2}\int_0^T \|\partial_t u - \mathcal{G}u\|^2 dt = \frac{1}{2}\langle \partial_t u - \mathcal{G}u, \partial_t u - \mathcal{G}u\rangle_{\boldsymbol{x},t}, \tag{5.2}$$

where $\|\cdot\|$ indicates the L_2 norm for $\boldsymbol{x} \in D$ and $\langle\cdot,\cdot\rangle_{\boldsymbol{x},t}$ the inner product with respect to both $\boldsymbol{x}$ and t. We consider the optimization problem [8]

$$\min_{T\in\mathbb{R}^+, u\in A} S_T(u), \tag{5.3}$$

where the set A contains all transition paths such that

$$u(0,\boldsymbol{x})=u_0(\boldsymbol{x}),\quad u(T,\boldsymbol{x})=u_1(\boldsymbol{x}).$$

We will generalize the strategy for ODE systems to deal with problem (5.3). Considering the linear scaling $s=t/T$, we have the optimization problem

$$\min_{T\in\mathbb{R}^+,u\in A}\left[S_T(u)=\frac{1}{2}T\langle T^{-1}\partial_s u-\mathcal{G}u,T^{-1}\partial_s u-\mathcal{G}u\rangle_{\boldsymbol{x},t}\right], \tag{5.4}$$

which yields that

$$\frac{\partial S_T}{\partial T}=\langle\mathcal{G}u,\mathcal{G}u\rangle_{\boldsymbol{x},s}-T^{-2}\langle\partial_s u,\partial_s u\rangle_{\boldsymbol{x},s}=0, \tag{5.5}$$

i.e.,

$$T(u)=\left(\frac{\langle\partial_s u,\partial_s u\rangle_{\boldsymbol{x},s}}{\langle\mathcal{G}u,\mathcal{G}u\rangle_{\boldsymbol{x},s}}\right)^{1/2}. \tag{5.6}$$

We then regard T as a functional of u and remove it from the optimization problem to obtain

$$\min_{u\in A}\left[S_T(u)=\frac{1}{2}T(u)\langle T^{-1}(u)\partial_s u-\mathcal{G}u,T^{-1}(u)\partial_s u-\mathcal{G}u\rangle_{\boldsymbol{x},s}\right], \tag{5.7}$$

where

$$u(0,\boldsymbol{x})=u_0(\boldsymbol{x}),\quad u(1,\boldsymbol{x})=u_1(\boldsymbol{x}).$$

We define the perturbation operator $\hat{\mathcal{G}}$ satisfying

$$\mathcal{G}(u+\delta u)=\mathcal{G}u+\hat{\mathcal{G}}\delta u+\mathcal{O}(\delta^2 u).$$

It can be shown that

$$\begin{aligned}\delta S_T&=T\left\langle(T^{-1}\partial_s-\mathcal{G})u,(T^{-1}\partial_s-\hat{\mathcal{G}})\delta u\right\rangle_{\boldsymbol{x},s}+\frac{\delta T}{2}(\langle\mathcal{G}u,\mathcal{G}u\rangle_{\boldsymbol{x},s}-T^{-2}\langle\partial_s u,\partial_s u\rangle_{\boldsymbol{x},s})\\&=T\left\langle(T^{-1}\partial_s-\mathcal{G})u,(T^{-1}\partial_s-\hat{\mathcal{G}})\delta u\right\rangle_{\boldsymbol{x},s},\end{aligned} \tag{5.8}$$

where the condition (5.5) is applied. To this end, we obtain the following lemma:

Lemma 5.1. *Let $\{h_i(\boldsymbol{x})\}_{i=1}^{N_x}$ and $\{\psi_i(t)\}_{i=1}^{N_t}$ the approximation bases for space and time, respectively. Then $u(t,\boldsymbol{x})$ has the following approximation*

$$u(t,\boldsymbol{x})\approx u_h(t,\boldsymbol{x})=\sum_{i=1}^{N_x}\sum_{j=1}^{N_t}u_{i,j}h_i(\boldsymbol{x})\psi_j(t),$$

and the gradient of S_T is given as

$$\frac{\partial S_T}{\partial u_{i,j}}=\delta S_T|_{\delta u=h_i(\boldsymbol{x})\psi_j(t)}, \tag{5.9}$$

where δS_T is given by Eq. (5.8).

The elementwise error indicator θ_i is defined as

$$\theta_i^2 = T^{-1}(u)\int_{s_i}^{s_{i+1}} \left(T^{-1}(u)\|\partial_s u\| - \|\mathcal{G}u\|\right)^2 ds, \quad i=0,1,\cdots,N-1, \tag{5.10}$$

and the adaptivity criterion (4.13) can be generalized straightforwardly for spatially extended systems.

Remark 5.1. It is well known that Eq. (5.1) can be ill-posed if $W(\boldsymbol{x},t)$ is space-time white noise especially for $d \geq 2$. Then a finite correlation length l_c in space is usually introduced in space to obtain well-posedness. In other words, the spatial correlation should be included in the definition of action functional [11]. However, studies [2, 12] show that under certain conditions, such as $0<\varepsilon \ll l_c \ll 1$, the correlation length l_c can disappear at the level of large deviations, i.e., a regular L_2 norm in space can be used in the definition of action functional. Roughly speaking, under certain conditions the action functional defined in Eq. (5.2) can be used when the spatial correlation length is small enough. In fact, tMAM can be readily generalized to deal with noise that is correlated in space and white in time. We here only use the action functional (5.2) to demonstrate tMAM for spatially extended systems.

6 Numerical results

In this section, we implement some numerical experiments to examine the performance of (adaptive) tMAM. Two ODE examples and one PDE example are considered. The optimization problem (4.3) is solved by nonlinear conjugate gradient method [23].

6.1 An SODE example: a gradient system

We consider the following SODE example:

$$\begin{cases} dx = -\partial_x V(x,y)dt + \sqrt{\varepsilon}dW_t^x, \\ dy = -\partial_y V(x,y)dt + \sqrt{\varepsilon}dW_t^y, \end{cases} \tag{6.1}$$

where W_t^x and W_t^y are independent Wiener processes, and the potential $V(x,y)$ is

$$V(x,y) = (1-x^2-y^2)^2 + y^2/(x^2+y^2). \tag{6.2}$$

This problem was used to study the adaptive MAM (aMAM) in [20], which is a gradient system. The dynamical system has two stable fixed points $\boldsymbol{a}_1 = (-1,0)$ and $\boldsymbol{a}_2 = (1,0)$, which are local minima of the potential $V(x,y)$. We consider the MAP in the upper half-plane connecting $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ through the saddle point $\boldsymbol{a}_3 = (0,1)$. Then the explicit form of this MAP is the upper branch of the unit circle: $x^2+y^2=1$. The exact action functional is $2\times(V(\boldsymbol{a}_3)-V(\boldsymbol{a}_1))=2$.

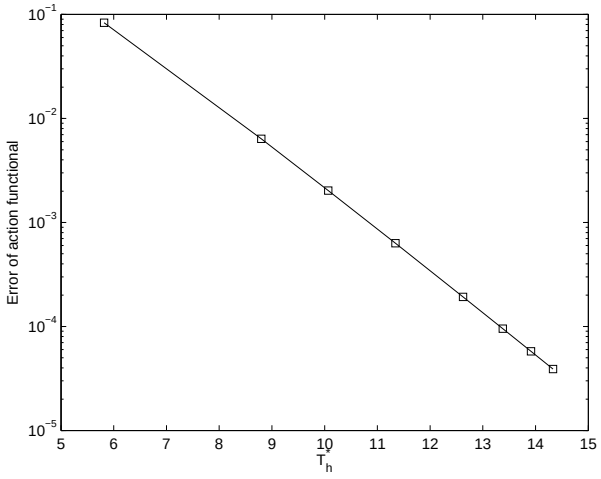

Figure 1: Error of action functional with respect to the optimal integration time $T_h^*(\boldsymbol{\phi}_h^*)$. All partitions $\mathcal{T}_h$ are given by uniform linear elements.

We first look at the relation between the error of action functional and $T_h^*(\phi_h^*)$. We use the two stable fixed points $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ as the ends of the transition paths and uniform linear elements for the partition $\mathcal{T}_h$. In Fig. 1, we plot the error of action functional with respect to T_h^*. First, although $T^*=\infty$, T_h^* can be relatively small to reach a good accuracy for the action functional. Second, T_h^* depends on the resolution. When the resolution is lower, T_h^* is smaller. As the space V_h is refined, T_h^* increases correspondingly. Third, as the integration time T_h^* increases, the error of action functional decreases exponentially. See Section 6.2 for more numerical results about T_h^*.

We now compare the performance of tMAM and aMAM for this problem. In Fig. 2, we the plot the approximated MAPs given MAM, aMAM and tMAM subject to 50 linear finite elements, where $T=50$ for MAM and aMAM. It is seen that for MAM, most of the grid points are clustered around simple fixed points $\boldsymbol{a}_1$, $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$. The re-meshing strategy of aMAM produces a more uniform distribution of the grids with respect to arc length, which also improves the accuracy significantly. The grid distribution of tMAM is uniform with respect to time. Although the grids of tMAM are not uniform with respect to arc length, the element size decreases gradually to the simple fixed points instead of dropping suddenly as in MAM. We now compare the accuracy of aMAM and tMAM for the same resolution. For the resolution of 50 linear finite elements, the aMAM with T=50 and tMAM yield errors 9.24e-3 and 6.37e-3, respectively, in action functional, which are comparable. By varying T, the best accuracy of aMAM we can achieve is 5.32e-3 at $T=18$. However, the best accuracy of aMAM is difficult to obtain in reality because of the following two conflicting issues: 1) $T^*=\infty$ implies that the integration time T should be large enough; 2) The approximation error will increase for a given time mesh as T increases. We usually do not know how to choose a T in advance to balance these two

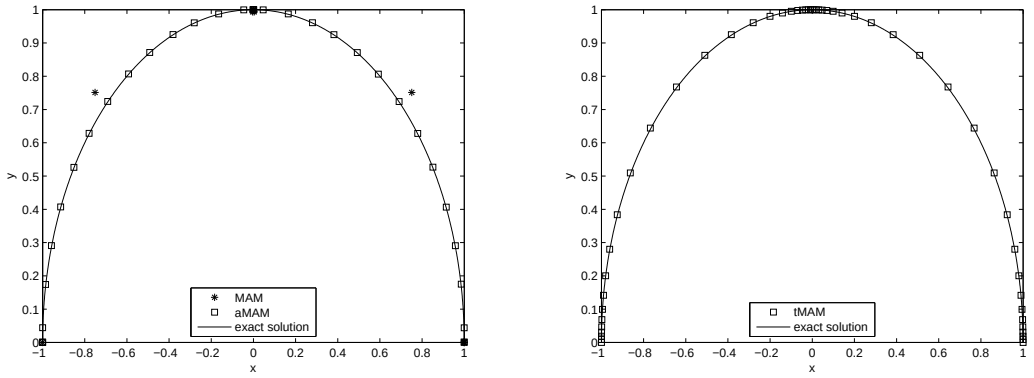

Figure 2: The MAP is approximated using 50 linear finite elements. Left: approximated MAPs given by MAM and aMAM with a fixed $T=50$; Right: approximated MAP given by tMAM.

issues for aMAM. In tMAM, the choice of T is determined by the algorithm, which yields the smallest action for a fixed mesh.

In Fig. 3, we plot the h-convergence of aMAM and tMAM with respect to the number of elements, where a singular case is considered in the left plot using $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ as the two ends of transition paths, and a non-singular case is considered in the right plot using $\hat{\boldsymbol{a}}_1=(\cos(\pi-0.1),\sin(\pi-0.1))$ and $\hat{\boldsymbol{a}}_3=(\cos(\pi/2+0.1),\sin(\pi/2+0.1))$ as the two ends of transition paths. Note that $\hat{\boldsymbol{a}}_1$ and $\hat{\boldsymbol{a}}_3$ are located on the MAP connecting $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$, and close to $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$, respectively. More specifically, the arc length of the MAP between $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$ is $\pi/2\approx 1.57$ and the arc length of the MAP between $\hat{\boldsymbol{a}}_1$ and $\hat{\boldsymbol{a}}_3$ is about 1.37. We first look at the singular case. First, it is seen that aMAM and tMAM converge algebraically at a rate of $\mathcal{O}(N^{-2})$ and $\mathcal{O}(N^{-1.7})$, respectively. In other words, tMAM with a uniform time mesh is not able to reach the optimal convergence rate, which is a common phenomena of finite element method for singular problems. Second, tMAM with a uniform time mesh, in general, provides a better accuracy than aMAM when the number of degrees of freedom is small, and will be outperformed by aMAM when the resolution is fine enough due to the smaller convergence rate. Third, the adaptive tMAM can recover the optimal convergence rate of $\mathcal{O}(N^{-2})$. Compared to aMAM, adaptive tMAM only refine the elements which have a relatively large arc length $\hat{s}_i$ or θ_i. In other words, reparametrization is not necessary. We also have a sequence of hierarchical meshes, which provides the information about the convergence. We now look at the non-singular case, where there do not exist critical points on the MAP. It is seen that tMAM converges at the optimal rate of $\mathcal{O}(N^{-2})$ while aMAM is not able to deal with this case since the integration time T must be given in aMAM. According to the adaptive tMAM with 1946 linear elements, the minimum action is 1.960133 reached at $T_h^*=2.299242$ for the MAP connecting $\hat{\boldsymbol{a}}_1$ and $\hat{\boldsymbol{a}}_3$. It is known that if we use $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$ as the two ends of transition paths, the problem will be singular corresponding to $T^*=\infty$ and a minimum action 2. It is seen that although the minimum action drops slightly, the integration time T^* drops substantially.

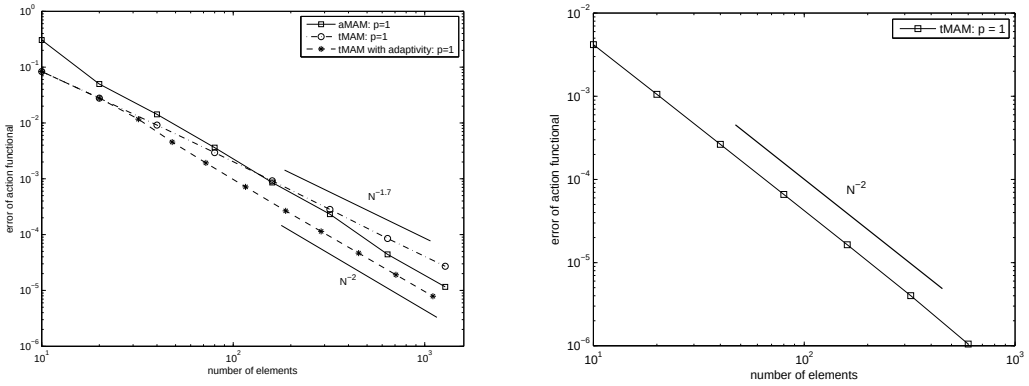

Figure 3: h-convergence of aMAM, tMAM and adaptive tMAM for linear elements. $r_\theta = 0.8$ and $r_\alpha = 0.5$. Left: A singular case is considered with $(x(0),y(0)) = \boldsymbol{a}_1$ and $(x(T),y(T)) = \boldsymbol{a}_2$. The exact MAP is the upper semicircle of radius 1 corresponding to an action 2; Right: A non-singular case is considered with $(x(0),y(0)) = (\cos(\pi-0.1),\sin(\pi-0.1))$ and $(x(T),y(T)) = (\cos(\pi/2+0.1),\sin(\pi/2+0.1))$. The exact MAP is part of the upper semicircle of radius 1 connecting the two ends. The reference action is given by the adaptive tMAM with 1946 linear elements, which is 1.960133 reached at $T_h^* = 2.299242$.

We now look at the mechanism that aMAM and tMAM work for singular cases, i.e., $T^* = \infty$. For simplicity, we consider the MAP that connects $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$. Starting from 10 equidistant linear finite elements, we use adaptive tMAM to obtain a time mesh of 94 elements, where $r_\theta = 0.8$ and $r_\alpha = 0.5$. We then use aMAM to obtain a time mesh of 94 linear elements, where $T = 50$ is fixed. We will use the part of the MAP from $\hat{\boldsymbol{a}}_1$ to $\hat{\boldsymbol{a}}_3$ to characterize the fast dynamics. On the mesh given by aMAM, $\hat{\boldsymbol{a}}_1$ is passed on the element $t \in [23.74,23.95]$ and $\hat{\boldsymbol{a}}_3$ on the element $t \in [26.05,26.26]$, which implies that the time used from $\hat{\boldsymbol{a}}_1$ to $\hat{\boldsymbol{a}}_3$ is between 2.10 and 2.52. This is consistent with $T_h^* \approx 2.30$ predicted in the non-singular case. For the mesh given by adaptive tMAM, $\hat{\boldsymbol{a}}_1$ is passed on the time element $s \in [0.38125,0.3875]$ and $\hat{\boldsymbol{a}}_3$ on the element $s \in [0.6215,0.61875]$ subject to $T_h^* = 9.87$, which implies that the time used from $\hat{\boldsymbol{a}}_1$ to $\hat{\boldsymbol{a}}_3$ is between 2.22 and 2.34, which agrees with $T_h^* \approx 2.30$ predicted by the non-singular case. In other words, both aMAM and tMAM are able to adjust the time mesh to make sure that the part of the MAP given by the fast dynamics can be resolved well. For a reference, we also include the tMAM with 94 equidistant linear elements, where $\hat{\boldsymbol{a}}_1$ is passed on the element $s \in [0.287,0.298]$ and $\hat{\boldsymbol{a}}_3$ on the element $s \in [0.70,0.71]$ subject to a smaller $T_h^* = 5.61$, which implies that the time used from $\hat{\boldsymbol{a}}_1$ to $\hat{\boldsymbol{a}}_3$ is between 2.26 and 2.37. This also agrees with the prediction $T_h^* \approx 2.30$ given by the non-singular case. Another thing that is worthy of notice is that there are 44, 66 and 40 elements to cover the fast dynamics from $\hat{\boldsymbol{a}}_1$ to $\hat{\boldsymbol{a}}_3$ in aMAM, adaptive tMAM, and tMAM, respectively, which results in errors of action functional as 6.0e-4, 1.3-4, and 3.5e-4. It is seen that due to the variation of T in tMAM, the linear time scaling can put more elements to address the fast dynamics when no adaptivity is employed.

For adaptive tMAM and aMAM subject to 94 linear finite elements, we plot in Fig. 4 the distribution of $\hat{s}_i$ in the left plot and that of θ_i in the right plot, where the filled $\circ$ and $\diamond$ indicate the elements in which $\hat{\boldsymbol{a}}_1$ and $\hat{\boldsymbol{a}}_3$ are located, and serve as a delimiter

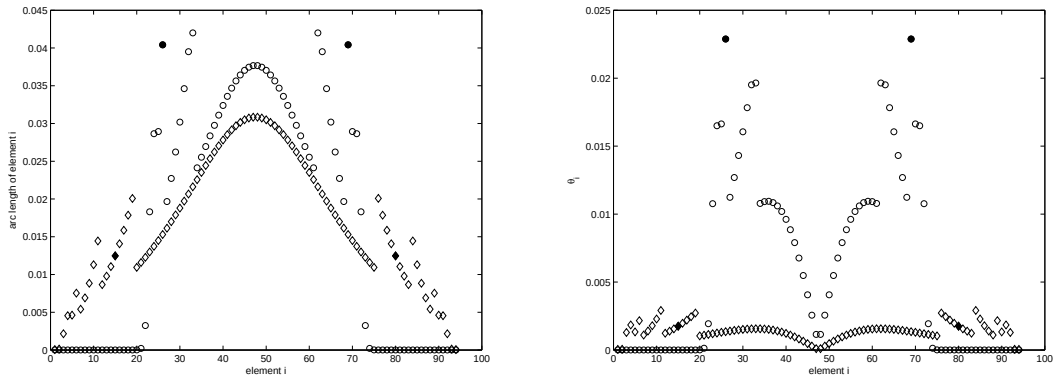

Figure 4: Elementwise distribution of $\hat{s}_i$ and θ_i. ○ is for adaptive tMAM and ◇ is for aMAM. The filled ○ and ◇ indicate the elements in which $\hat{a}_1$ and $\hat{a}_3$ are located. Left: $\hat{s}_i$; Right: θ_i.

between slow and fast dynamics. It is noticed that although aMAM tries to move the clustered points away from critical points $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$ to the region of fast dynamics, there are still quite a number of points clustered around $\boldsymbol{a}_1$ and $\boldsymbol{a}_3$, where these elements have a small arc length and are not effective for the approximation of the MAP. This is because that the chosen integration time $T=50$ is too large, and many redundant elements has to be pushed into the region of slow dynamics. The adaptive tMAM returns that $T_h^*=9.87$ in contrast to the fixed $T=50$ for aMAM. The adaptive tMAM has more effective elements in the region of slow dynamics. In the region of fast dynamics, the elements of adaptive tMAM have a smaller average arc length compared to those of aMAM. The adaptive tMAM has a much more uniform distribution of θ_i than aMAM in regions of both slow and fast dynamics due to the adaptivity criterion (4.13). It is noticed that θ_i can be relatively large in the region of slow dynamics, which implies that mesh refinement is also necessary here to improve the accuracy.

6.2 Another SODE example: a non-gradient system

We consider the following Maier-Stein model:

$$\begin{cases} dx=(x-x^3-\beta xy^2)dt+\sqrt{\varepsilon}dW_t^x, \\ dy=-(1+x^2)ydt+\sqrt{\varepsilon}dW_t^y, \end{cases} \tag{6.3}$$

where W_t^x and W_t^y are independent Wiener processes and $\beta>0$ is a parameter. This example was used to study gMAM in [13]. which is a non-gradient system except when $\beta=1$. This system has two stable fixed points: $\boldsymbol{b}_1=(-1,0)$ and $\boldsymbol{b}_2=(1,0)$, and a saddle point $\boldsymbol{b}_3=(0,0)$. For numerical experiments, we set $\beta=10$.

We consider the transition from $\boldsymbol{b}_1$ to $\boldsymbol{b}_2$. The MAP given by adaptive tMAM with 116 linear elements is plotted in the left plot of Fig. 5, where we also plot the element distribution with respect to arc length. The MAP passes the saddle point $\boldsymbol{b}_3$, but is not

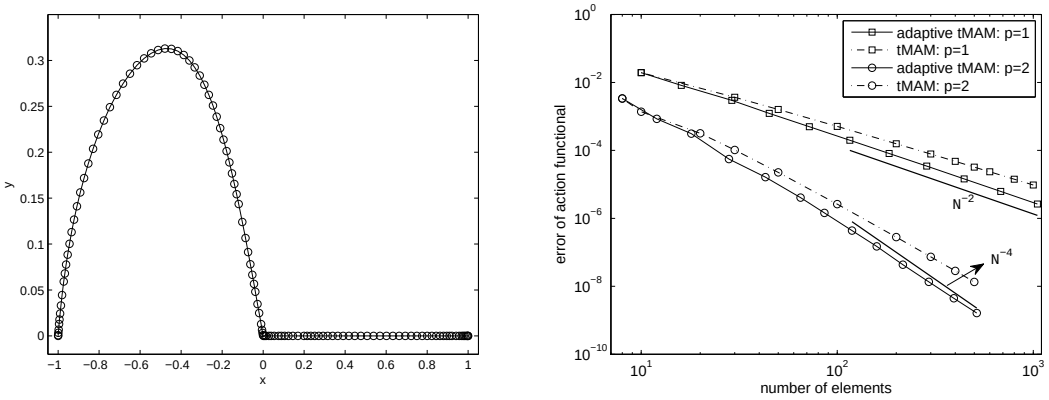

Figure 5: Left: The MAP from $\boldsymbol{b}_1$ to $\boldsymbol{b}_2$ with element distribution with respect to arc length for adaptive tMAM with 116 linear elements; Right: h-convergence of (adaptive) tMAM. The solid lines are for adaptive tMAMs and the dash dot lines are for tMAMs with a uniform partition. The reference solution is given by adaptive tMAM with 608 elements of polynomial order 4.

the heteroclinic orbit connecting $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ due to the fact that the system is non-gradient and β is large enough. In contrast to the MAP connecting $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ in the previous ODE example, which is smooth at the saddle point in terms of arc length, this MAP is continuous but not differentiable at the saddle point $\boldsymbol{b}_3$. In the right plot of Fig. 5, we plot the h-convergence of adaptive tMAM with linear and quadratic elements, where the reference solution is given by adaptive tMAM with 608 elements of polynomial order 4. It is seen that tMAM with a uniform partition converges algebraically at a rate slightly lower than the optimal one and the optimal convergence rate is eventually recovered by adaptivity regardless of the low regularity at the saddle point. In gMAM, special attention needs to be paid around the critical points, otherwise, we can only obtain a first-order convergence even a second-order scheme is used. One reason is that the arc length constraint holds pointwisely in gMAM, which makes the mapping between time and arc length sensitive to critical points; the other reason is that adaptive tMAM is defined in the framework of finite element method, which can deal with non-smooth functions more efficiently than the finite difference method used in gMAM.

In Fig. 6 we plot the x component with respect to s, where the left plot is for the adaptive tMAM with 116 linear elements and the right plot is for the adaptive tMAM with 109 elements of polynomial order 3. It is seen that the fast dynamics is resolved by a large number of small elements and the slow dynamics is captured by a small number of large elements and a long integration time. For a higher resolution, $T(\boldsymbol{\phi}_h^*)$ is larger and more time is assigned to slow dynamics, which makes it more consistent with real dynamics.

In Fig. 7, we plot the error of the action functional in terms of the integration time $T_h^*(\boldsymbol{\phi}_h^*)$, where the left plot is for tMAMs with a uniform partition and the right plot for adaptive tMAMs. Overall, the integration time T_h^* increases and the error decreases as the approximation space is refined. However, if a uniform refinement is considered, the error

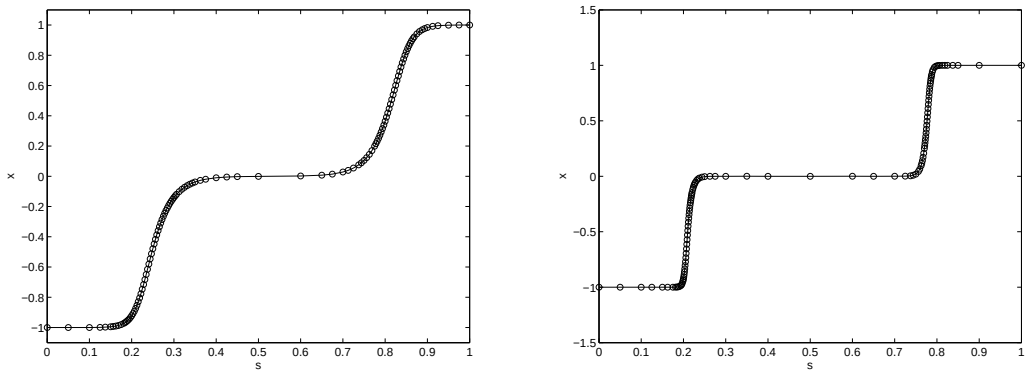

Figure 6: The x component with respect to time s. Left: the adaptive tMAM with 116 linear elements. $T(\boldsymbol{\phi}_h^*)=26.19$; Right: the adaptive tMAM with 109 elements of polynomial order 3. $T(\boldsymbol{\phi}_h^*)=127.01$.

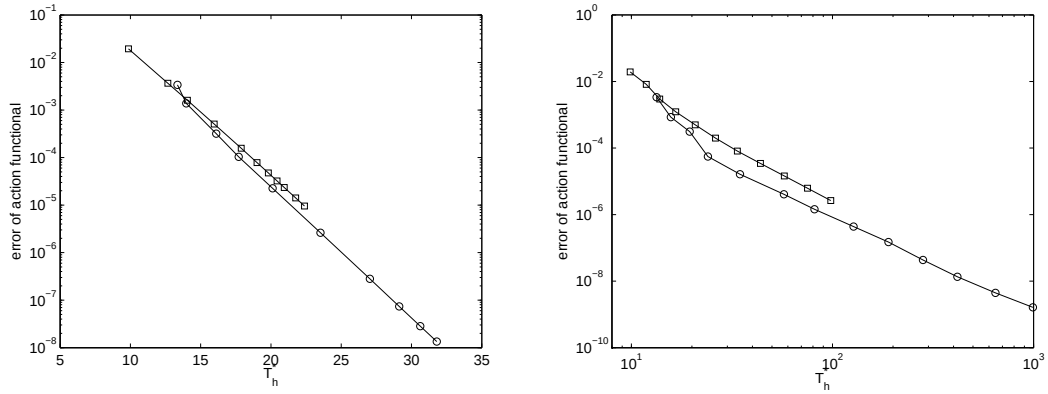

Figure 7: Error of action functional with respect to the optimal integration time $T_h^*(\phi_h^*)$. Left: Results given by tMAMs with a uniform partition. Right: Results given by adaptive tMAMs.

decays exponentially with respect to T_h^*; if the adaptive tMAM is considered, the error decays algebraically with respect to T_h^*. At this moment, it is still not clear why these two different convergence behaviors occur. It is seen that for a comparable accuracy, tMAM of a uniform partition needs a much smaller optimal integration time than the adaptive tMAM. Such an observation is consistent with the fact that the region of fast dynamics is critical to minimize the action functional while the region of slow dynamics is important to recover the optimal convergence rate.

6.3 An SPDE example

We consider the infinite dimensional analogue of the Maier-Stein model:

$$\begin{cases} \partial_t u=\kappa\partial_{xx}u+u-u^3-\beta uv^2+\sqrt{\varepsilon}\dot{W}_u(x,t), \\ \partial_t v=\kappa\partial_{xx}v-(1+u^2)v+\sqrt{\varepsilon}\dot{W}_v(x,t), \end{cases} \tag{6.4}$$

where $x \in [0,1]$, $\kappa > 0$ is a parameter, and $\dot{W}_u$ and $\dot{W}_v$ are space-time white noise. We assume that the above equations satisfy periodic boundary conditions in x direction. This problem was used to study geometric MAM (gMAM) in [13]. The system has two stable fixed points: $(u_-,v_-)=(-1,0)$ and $(u_+,v_+)=(+1,0)$. We consider the MAP from (u_-,v_-) to (u_+,v_+). When β is small enough, the system is close to a gradient system, and the MAP follows the graph of a heteroclinic orbit connecting (u_-,v_-) and (u_+,v_+); when β is large enough, the system is far from a gradient system and we then expect that the MAP connecting (u_-,v_-) to an unstable equilibrium point will be different from the heteroclinic orbit.

Let $\boldsymbol{\phi}(x,t)=(u,v)^{\mathsf{T}}$. Let $\mathcal{G}\boldsymbol{\phi}=\kappa\partial_{xx}\boldsymbol{\phi}+\boldsymbol{b}(\boldsymbol{\phi})$ with $\boldsymbol{b}(\boldsymbol{\phi})=(u-u^3-\beta uv^2,-(1+u^2)v)^{\mathsf{T}}$. The action functional can be written as

$$S_T(\boldsymbol{\phi})=\frac{1}{2}\langle\partial_t\boldsymbol{\phi}-\mathcal{G}\boldsymbol{\phi},\partial_t\boldsymbol{\phi}-\mathcal{G}\boldsymbol{\phi}\rangle_{x,t}. \tag{6.5}$$

The perturbation operator $\hat{\mathcal{G}}$ is obtained as

$$\hat{\mathcal{G}}\delta\boldsymbol{\phi}=(\kappa\partial_{xx}+\nabla_{\boldsymbol{\phi}}\boldsymbol{b}(\boldsymbol{\phi}))\delta\boldsymbol{\phi}. \tag{6.6}$$

We choose Fourier expansion in x direction due to the periodic boundary conditions, where $h_i(x)\in\text{span}\{e^{\mathbf{i}2\pi nx}\}_{|n|=0}^{16}$. Then Lemma 5.1 can be applied to compute the gradient of action functional.

We here only consider one case $\beta=1$ and $\kappa=0.024$ for the study of accuracy. More discussions of this SPDE generalization of the Maier-Stein model can be found in [13]. In Fig. 8, we plot the snapshots of u along the MAP from (u_-,v_-) to (u_+,v_+) on the left, and the partition $\mathcal{T}_h$ of the adaptive tMAM with 104 elements of polynomial order 3 on the right. It is seen that most of the elements go to the region of fast dynamics, where two

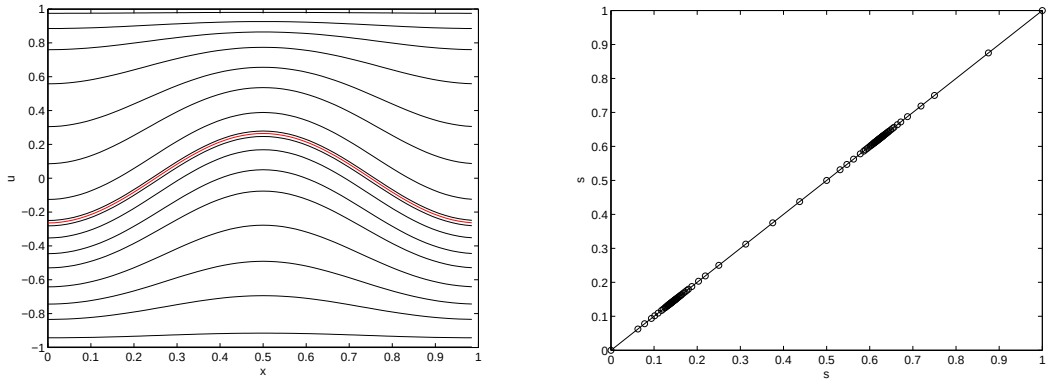

Figure 8: Left: snapshots of u along the MAP from (u_-,v_-) to (u_+,v_+). The red line indicates the unstable equilibrium points approximated by the MAP. $\beta=1$ and $\kappa=0.024$. Right: final partition $\mathcal{T}_h$ with 104 elements of polynomial order 3.

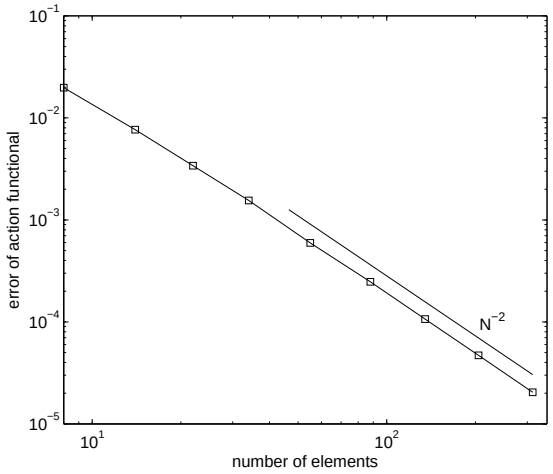

Figure 9: h-convergence of adaptive tMAM with linear elements. The reference solution is given by adaptive tMAM with 104 elements and polynomial order 3.

regions of fast dynamics are separated by the unstable equilibrium solution. In Fig. 9, we plot the convergence behavior of the adaptive tMAM with linear elements starting from 8 equidistant elements. It is seen that as the number of elements increases, the optimal convergence rate $\mathcal{O}(N^{-2})$ is eventually recovered by the h-adaptivity.

7 Conclusion

In this work we developed a minimum action method with optimal linear time scaling. The key idea is to treat the integration time as a functional through the weak arc length constraint (3.10) such that the integration time will be not optimized directly. A simple adaptivity criterion has also been constructed to enhance the convergence rate of tMAM. The adaptive tMAM has the following properties: (1) It is able to deal with quasi-potential with respect to time. If T^* is finite, its approximation can be obtained directly from the MAP. (2) The adaptive tMAM does not need re-meshing. Instead, we considered h-adaptivity, which produces a sequence of hierarchical partitions of the time interval. The hierarchical meshes provide the information of relative approximation error of the action functional. (3) When $T^*=\infty$, the adaptive tMAM can recover the optimal algebraic convergence rate for a fixed polynomial order p.

We have mainly focused on the convergence behavior of tMAM in this work, where the numerical efficiency depends on the underlying gradient-based optimization solver. There are a couple of ways to enhance the numerical efficiency, which is particularly important for spatially extended systems. First, the adaptive criterion can be refined. For instance, the threshold r_θ and $r_{\hat{s}}$ are not necessarily fixed and can be linked to the current MAP $\boldsymbol{\phi}_h^*$ and a targeted number of degrees of freedom for the new partition, e.g.,

we can consider multi-level h-refinement of a certain element. Furthermore, although we only considered h-adaptivity in this work, tMAM is actually flexible to couple with hp-adaptivity. Second, error indicators that are consistent with the approximation theory can be constructed. The indicator we used for the error $\boldsymbol{\phi}^* - \boldsymbol{\phi}_h^*$ is just based on physical intuition, which is not related to the best approximation of V_h for $\boldsymbol{\phi}^*$. Furthermore, if p-adaptivity is included, a regularity indicator is also needed. Third, preconditioning needs to be taken into account. Fourth, the Euler-Lagrange equation of the action functional can be solved by a Newton solver. For the third and fourth issues, we need to look into the structure of the spatial differentiation operator when spatially extended systems are considered. The study about these issues will be reported elsewhere.

Acknowledgments

This work was supported by NSF grant DMS-1115632 and AFOSR grant FA9550-15-1-0051.

References

[1] C. Cerjan and W. Miller, *On finding transition states*, J. Chem. Phys. **75(6)** (1981), 2800-2806.
[2] S. Cerrai and M. Freidlin, *Approximation of quasi-potential and exit problems for multidimensional RDE's with noise*, Trans. Amer. Math. Soc., 363(7) (2011), 3853–3892.
[3] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, Encyclopedia Math. Appl. 44, Cambridge University Press, Cambridge, UK, 1992.
[4] W. E, W. Ren and E. Vanden-Eijnden, *String method for the study of rare events*, Phys. Rev. B, **66** (2002), 052301.
[5] W. E, W. Ren and E. Vanden-Eijnden, *Minimum action method for the study of rare events*, Commun. Pure Appl. Math., **57** (2004), 637-565.
[6] W. E, W. Ren and E. Vanden-Eijnden, *Simplified and improved string method for computing the minimum energy paths in barrier-crossing events*, J. Chem. Phys., **126** (2007), 164103.
[7] W. E and X. Zhou, *The gentlest ascent dynamics*, Nonlinearity, **24(6)** (2011), 1831.
[8] W. E, X. Zhou and X. Cheng, *Subcritical bifurcation in spatially extended systems*, Nonlinearity, **25(3)** (2012), 761.
[9] W. Faris and G. Jona-Lasinio, *Large fluctuations for a nonlinear hear equation with noise*, J. Phys. A: Math. Gen. **15** (1982), 3025-3055.
[10] M. Freidlin, *Random perturbations of reaction-diffusion equations: the quasideterministic approximation*, Trans. Amer. Math. Soc., **305** (1988), 665–697.
[11] M. Freidlin and A. Wentzell, *Random Perturbations of Dynamical Systems*, second ed., Springer-Verlag, New York, 1998.
[12] M. Hairer and H. Weber, *Large deviations for white-noise driven, nonlinear stochastic PDEs in two and three dimensions*, preprint, 2014.
[13] 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** (2008), 1052-1117.
[14] G. Henkelman and H. Jónsson, *A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives*, J. Chem. Phys., **111(15)** (1999), 7010-7022.

[15] H. Jònsson, G. Mills and K. Jacobsen, *Nudged elastic band method for finding minimum energy paths of transitions*, Classical and Quantum Dynamics in Condensed Phase Simulations, Ed. B. Berne, G. Ciccotti and D. Coker, 1998.
[16] N. van Kampen, *Stochastic Processes in Physics and Chemistry*, North-Holland, 1981.
[17] G. Karniadakis and S. Sherwin, *Spectral/hp Element Methods for Computational Fluid Dynamics*, Oxford University Press, 2nd Edition, 2005.
[18] J. Nocedal and S. Wright, *Numerical Optimization*, Springer Series in Operations Research, Springer, New York, 1999.
[19] L. Onsager and S. Machlup, *Fluctuations and irreversible processes*, Phys. Rev., **91** (1953), 1505-1512.
[20] X. Zhou, W. Ren and W. E, *Adaptive minimum action method for the study of rare events*, J. Chem. Phys., **128** (2008), 104111.
[21] X. Zhou and W. E, *Study of noise-induced transitions in the Lorenz system using the minimum action method*, Comm. Math. Sci., **8(2)** (2010), 341-355.
[22] X. Wan, X. Zhou and W. E, *Study of the noise-induced transition and the exploration of the configuration space for the Kuramoto-Sivashinsky equation using the minimum action method*, Nonlinearity, **23** (2010), 475-493.
[23] X. Wan, *An adaptive high-order minimum action method*, J. Compt. Phys., **230** (2011), 8669-8682.
[24] X. Wan, *A minimum action method for small random perturbations of two-dimensional parallel shear flows*, J. Compt. Phys., **235** (2013), 497–514.
[25] X. Wan and G. Lin, *Hybrid parallel computing of minimum action method*, Parallel Computing, **39** (2013), 638–651.
[26] X. Wan, H. Yu and W. E, *Model the nonlinear instability of wall-bouned shear flows as a rare event: A study on two-dimensional Poiseuille flows*, Nonlinearity, in press.
[27] J. ZHANG AND Q. DU, *Shrinking dimer dynamics and its applications to saddle point search*, SIAM Journal on Numerical Analysis, 50(4) (2012), 1899–1921.