Commun. Comput. Phys.
doi: 10.4208/cicp.OA-2017-0025

Vol. **23**, No. 2, pp. 408-439
February 2018

An hp-Adaptive Minimum Action Method Based on a Posteriori Error Estimate

Xiaoliang Wan[1,*], Bin Zheng[2] and Guang Lin[3]

[1] *Department of Mathematics, Center for Computation and Technology, Louisiana State University, Baton Rouge 70803, USA.*
[2] *Pacific Northwest National Laboratory, Richland, WA 99352, USA.*
[3] *Department of Mathematics & School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA.*

Received 27 January 2017; Accepted (in revised version) 9 May 2017

Abstract. In this work, we develop an hp-adaptivity strategy for the minimum action method (MAM) using a posteriori error estimate. MAM plays an important role in minimizing the Freidlin-Wentzell action functional, which is the central object of the Freidlin-Wentzell theory of large deviations for noise-induced transitions in stochastic dynamical systems. Because of the demanding computation cost, especially in spatially extended systems, numerical efficiency is a critical issue for MAM. Difficulties come from both temporal and spatial discretizations. One severe hurdle for the application of MAM to large scale systems is the global reparametrization in time direction, which is needed in most versions of MAM to achieve accuracy. We recently introduced a new version of MAM in [22], called tMAM, where we used some simple heuristic criteria to demonstrate that tMAM can be effectively coupled with h-adaptivity, i.e., the global reparametrization can be removed. The target of this paper is to integrate hp-adaptivity into tMAM using a posteriori error estimation techniques, which provides a general adaptive MAM more suitable for parallel computing. More specifically, we use the zero-Hamiltonian constraint to define an indicator to measure the error induced by linear time scaling, and the derivative recovery technique to construct an error indicator and a regularity indicator for the transition paths approximated by finite elements. Strategies for hp-adaptivity have been developed. Numerical results are presented.

AMS subject classifications: 60H35, 65C20, 65N20, 65N30

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

*Corresponding author. *Email addresses:* `xlwan@math.lsu.edu` (X. Wan), `Bin.Zheng@pnnl.gov` (B. Zheng), `guanglin@purdue.edu` (G. Lin)

1 Introduction

Small random perturbations of dynamical systems can introduce rare but important events, e.g., the transitions between different stable equilibrium states of a deterministic dynamical system. Such noise-induced transitions have been observed on both small and large scales, and are critical in many physical, biological and chemical systems. Examples include nucleation events of phase transitions, chemical reactions, regime change in climate, conformation changes of biomolecules, hydrodynamic instability, etc.

The Freidlin-Wentzell (F-W) theory of large deviations provides a rigorous mathematical framework to understand the transitions induced by small noise in general dynamical systems. The key object of the F-W theory of large deviations is the F-W action functional, and the critical quantities are the minimizer of the F-W action functional and the associated minimum value. Starting from [8], the large deviation principle given by the F-W theory has been approximated numerically, and the numerical methods are, in general, called minimum action method (MAM).

Consider an ordinary differential equations perturbed by small white noise

$$d\boldsymbol{X}_t = \boldsymbol{b}(\boldsymbol{X}_t)dt + \sqrt{\varepsilon}d\boldsymbol{W}_t, \tag{1.1}$$

where ε is a small positive parameter. We are interested in two types of problems:

$$\text{Problem I:} \quad \inf_{\substack{\boldsymbol{\phi}(0)=\boldsymbol{x}_1,\\ \boldsymbol{\phi}(T)=\boldsymbol{x}_2}} \left[S_T(\boldsymbol{\phi}) = \frac{1}{2}\int_0^T |\dot{\boldsymbol{\phi}} - \boldsymbol{b}(\boldsymbol{\phi})|^2 dt \right] \tag{1.2}$$

and

$$\text{Problem II:} \quad V(\boldsymbol{x}_1,\boldsymbol{x}_2) = \inf_{T>0} \inf_{\substack{\boldsymbol{\phi}(0)=\boldsymbol{x}_1,\\ \boldsymbol{\phi}(T)=\boldsymbol{x}_2}} S_T(\boldsymbol{\phi}), \tag{1.3}$$

where $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are two points in the phase space, $S_T(\boldsymbol{\phi})$ is called the action functional, and $V(\boldsymbol{x}_1,\boldsymbol{x}_2)$ the quasi-potential from point $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$. Here $\boldsymbol{\phi}(t)$ is a transition path connecting $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ on the time interval $[0,T]$. The minimizers of Problem I and II characterize the difficulty of the noise-induced transition from $\boldsymbol{x}_1$ to the vicinity of $\boldsymbol{x}_2$, see Eqs. (2.4) and (2.5). In Problem I, the transition is restricted to a certain time scale T, which is relaxed in Problem II. Let $\boldsymbol{\phi}^*(t)$ be the minimizer of either Problem I or Problem II, which is also called the minimal action path (MAP), or the instanton in physical literature related to path integral. From the application point of view, solving Problem I and II is important. For example, the minimizer of F-W action functional can be used to construct an asymptotically efficient estimator in important sampling, where optimization problems like Problem I and II need to be solved effectively [5, 17]. The MAM can help to explore a high-dimensional phase space [18, 27]. Another example is the nonlinear instability of wall-bounded shear flows, which can be modelled as a rare event of Navier-Stokes equations perturbed by small noise [20,23]. The most probable transition path provides useful information that is difficult or impossible to obtain in a deterministic way.

The original MAM given in [8] was for Problem I, where the action functional is discretized by the finite difference method and the optimization problem is solved by the L-BFGS method. From then on, many efforts have been made to obtain an effective MAM for Problem II. One main difficulty is that the quasi-potential will be achieved at an optimal integration time $T^*=\infty$ if there exists at least one critical point on the MAP from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$. Assume that Problem II can be well approximated by assigning a fixed but sufficiently large T when $T^*=\infty$. If we discretize the time interval $[0,T]$ uniformly, it is observed that along the path most grid points are clustered around the critical points due to the slow dynamics there. Only a few grid points can contribute to capture the MAP, which is mainly determined by fast dynamics. We call this a problem of clustering.

The problem of clustering was alleviated by geometric MAM (gMAM) [13], and adaptive MAM (aMAM) [26], which was improved recently by coupling WENO interpolation in [28]. The formulations of aMAM and gMAM are quite different. First of all, aMAM is formulated with respect to time while gMAM is formulated with respect to normalized arc length through a change of variable. GMAM assumes that the total arc length of the MAP is finite although $T^*=\infty$. Second, aMAM redistributes the grid points every a number of iteration steps by moving some grid points from the region of slow dynamics to the region of fast dynamics such that the non-equidistant time mesh is nearly uniform with respect to arc length while gMAM uses a uniform mesh with respect to arc length. Both aMAM and gMAM have some limitations. First, aMAM works for Problem I and partially for Problem II while gMAM works for Problem II only. AMAM is not able to deal with Problem II with a finite T^* since T must be prescribed. In gMAM, the integration time is determined by a mapping between time and arc length, implying that gMAM does not work for Problem I where T can be arbitrarily chosen. Second, both aMAM and gMAM need global reparametrization, which can become a bottleneck of parallel computing [21] and limit the application of MAM especially in spatially extended systems. A more flexible MAM is then expected, where both Problem I and II can be dealt with and the global reparametrization is not needed. The tMAM developed in [22] shows that such an MAM can be achieved. In tMAM, the action functional is formulated with respect to time and the discretization is based on finite element method [19]. The key idea of tMAM is to generate a sequence of T's to approximate T^* though optimal linear time scaling and h-type adaptivity, i.e., element refinement.

One common component in all MAMs formulated with respect to time is adaptivity [19, 22, 26]. Furthermore, all the adaptivity criteria are based on the simple physical intuition that the region of fast dynamics should be well resolved. However, the effectiveness of physically based adaptivity criteria is often limited from the algorithm point of view because they are not able to provide an accurate error or regularity estimate. In this paper, we intend to construct rigorous adaptivity criteria for MAM that are consistent with the approximation theory. More specifically, we use a posteriori error estimation techniques to measure the quality of the mesh given by finite element discretization, which results in an hp-adaptive tMAM. The adaptivity is used to control two types of approximation errors: model approximation and path approximation. The model approxi-

mation indicates the assumption that Problem II can be well approximated by a sequence of finite integration time T's when $T^*=\infty$. The path approximation indicates the finite element discretization of the MAP [19]. For the error of model approximation, we use the deviation from the surface $H(\boldsymbol{\phi},\boldsymbol{p})=0$ (see Eq. (2.24)) to define an error indicator [13,22]. For the error of path approximation, we use the derivative recovery technique of adaptive finite element method to define an elementwise error indicator and an elementwise regularity indicator. The error indicator helps to choose the elements for refinement, and the regularity indicator decides that h-refinement or p-refinement should be employed. Here h-refinement means to reduce the element size while p-refinement means to increase the polynomial order. We choose the derivative recovery technique mainly based on the following two observations: 1) The Euler-Lagrange equation associated with the action functional is a nonlinear elliptic equation, and 2) The derivative recovery technique does not depend on the nonlinearity of the system, which makes it more flexible and general.

For both Problem I and II, we rescale the time interval from $\Gamma_T=[0,T]$ to $\Gamma_1=[0,1]$ by a simple linear scaling $\tau=t/T$. Then the solutions of Problem I and Problem II with a finite T^* will be defined on the space $H^1(\Gamma_1;\mathbb{R}^n)$. For Problem II with an infinite T^*, a commonly used strategy is to use a finite but large T for approximation. We will show that in our formulation T does not need to be explicitly given. Instead, the discrete problem can find an optimal T, which is always finite and depends on the finite element mesh. This can be regarded as a regularization provided by discretization. This way, we only need to focus on how to choose a sequence of finite element meshes on Γ_1. The error indicators will be defined on $H^1(\Gamma_1;\mathbb{R}^n)$. For Problem I and Problem II with a finite T^*, the adaptive MAM is effective because the solution, which may have interior layers if T or T^* is large, is located in $H_1(\Gamma_1;\mathbb{R}^n)$. For problem II with an infinite T^*, the adaptive MAM also works very well although the solution is not located in $H^1(\Gamma_1;\mathbb{R}^n)$. This is because the adaptive finite element meshes can generate a minimizing sequence and the main approximation errors are from the region of fast dynamics subject to a finite transition time. The convergence of the minimizing sequence is addressed in [25]. We will demonstrate numerically that the h-adaptive MAM can recover the optimal convergence rate $\mathcal{O}(N^{-2p})$ with respect to the error of action functional and the hp-adaptive MAM can recover the exponential convergence.

This paper is organized as follows. We briefly review the minimum action method with optimal linear time scaling in Section 2. We then develop our adaptivity strategy in Section 3. Numerical results are given in Section 4 followed by a discussion section.

2 Minimum action method (MAM)

2.1 The large deviation principle (LDP)

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):

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

where $\boldsymbol{W}$ is a standard Wiener process in $\mathbb{R}^n$ and ε is a small positive parameter. We assume that the drift vector $\boldsymbol{b}$ and the diffusion tensor $\boldsymbol{a}=\sigma\sigma^{\mathsf{T}}$ are bounded and uniformly continuous, and $\boldsymbol{a}$ is uniformly elliptic, i.e., there exist $C>0$ such that $\langle v,\boldsymbol{a}(x)v\rangle\geq C|v|^2$, for any $v\in\mathbb{R}^n$. In other words, for any $x\in\mathbb{R}^n$, $\boldsymbol{a}(x)$ is symmetric and positive definite (SPD). Here $\langle\cdot,\cdot\rangle$ indicates the inner product of two vectors and $|\cdot|$ the ℓ_2 norm. For any $u,v,x\in\mathbb{R}^n$, we define a new inner product $\langle u,v\rangle_{\boldsymbol{a}}=\langle u,\boldsymbol{a}^{-1}(x)v\rangle$, and the associated norm $|u|_{\boldsymbol{a}}=\langle u,u\rangle_{\boldsymbol{a}}^{1/2}$.

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 [12] asserts that the probability of $\boldsymbol{X}$ passing the δ-tube about $\boldsymbol{\phi}(t)$ on $[0,T]$ is

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

when ε is small enough, and $S_T(\boldsymbol{\phi})$ is called the action functional defined as

$$S_T(\boldsymbol{\phi})=\frac{1}{2}\int_0^T|\dot{\boldsymbol{\phi}}-\boldsymbol{b}(\boldsymbol{\phi})|_{\boldsymbol{a}}^2dt=\frac{1}{2}\langle\dot{\boldsymbol{\phi}}-\boldsymbol{b}(\boldsymbol{\phi}),\dot{\boldsymbol{\phi}}-\boldsymbol{b}(\boldsymbol{\phi})\rangle_{\boldsymbol{a},t}, \tag{2.3}$$

where $\dot{\boldsymbol{\phi}}$ indicates the derivative with respect to t and the subscript t indicates the integration with respect to $t\in[0,T]$. We also define a weighted L^2 norm $\|\boldsymbol{u}\|_{\boldsymbol{a},t}=\langle\boldsymbol{u},\boldsymbol{u}\rangle_{\boldsymbol{a},t}^{1/2}$ for $\boldsymbol{u}(t)\in\mathbb{R}^n$, and $t\in[0,T]$. The fact given in Eq. (2.2) implies the large deviation principle (LDP), which says that the probability of some random events can be estimated asymptotically if the noise amplitude is small enough. For example, if A is a Borel subset in $\mathbb{R}^n$, we have the LDP that

$$\lim_{\varepsilon\downarrow0}\varepsilon\log\Pr(\boldsymbol{X}(0)=x,\boldsymbol{X}(T)\in A)=-\inf_{\substack{\boldsymbol{\phi}(0)=x,\\ \boldsymbol{\phi}(T)\in A}}S_T(\boldsymbol{\phi}), \tag{2.4}$$

which means that the transition probability from x to A at time T is determined asymptotically by the minimizer of the action functional.

When $\varepsilon\downarrow0$, 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 between two points $x_1,x_2\in\mathbb{R}^n$

$$V(x_1,x_2)=\inf_{T>0}\inf_{\substack{\boldsymbol{\phi}(0)=x_1,\\ \boldsymbol{\phi}(T)=x_2}}S_T(\boldsymbol{\phi}). \tag{2.5}$$

The probabilistic meaning of the quasi-potential (2.5) is as follows

$$V(x_1,x_2)=\lim_{T\to\infty}\lim_{\delta\downarrow0}\lim_{\varepsilon\downarrow0}-\varepsilon\log\Pr(\tau_\delta\leq T), \tag{2.6}$$

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, let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ be two stable equilibrium states, then we have the LDP [12] that

$$\lim_{\delta\downarrow 0}\lim_{\varepsilon\downarrow 0}\varepsilon\log\Pr(\text{transition from } \boldsymbol{x}_1 \text{ to } B_\delta(\boldsymbol{x}_2)) = -V(\boldsymbol{x}_1,\boldsymbol{x}_2), \tag{2.7}$$

where $B_\delta(\boldsymbol{x}_2)$ is a ball with radius δ and center $\boldsymbol{x}_2$. It is seen that the LDPs given in Eqs. (2.4) and (2.6) define Problem I and II, respectively.

2.2 A MAM based on optimal linear time scaling (tMAM)

In this section, we define our numerical strategy for approximating Problem I and II. The approximation space will be given by finite elements. Since we focus on algorithm in this work, the related convergence issues will be summarized into two assumptions.

2.2.1 Approximation space for transition paths

Consider the time interval $\Gamma_T = [0,T]$, where T is finite. The approximation space will be defined on $\Gamma_1 = [0,1]$ with respect to a linearly scaled variable $\tau = t/T$. Consider a partition of Γ_1

$$\mathcal{T}_h: \quad 0 = \tau_0 < \tau_1 < \cdots < \tau_N = 1.$$

Let $R = [-1,1]$ be a reference element and F_{e_i} an affine mapping from the element $e_i = [\tau_i,\tau_{i+1}]$, $i = 0,1,\cdots,N-1$, to the reference element R. Then in each element e_i, we can define a linear space spanned by polynomials

$$W_{e_i}^{(p)} = \{v : v\circ F_{e_i}^{-1} \in \mathscr{P}_p(R)\}, \tag{2.8}$$

where $\mathscr{P}_p(R)$ denotes the set of polynomials of degree up to p over R. In particular, we choose $\mathscr{P}_p(R) = \text{span}\{\hat{\psi}_i(\hat{\tau})\}_{i=0}^m$, where

$$\hat{\psi}_i(\hat{\tau}) = \begin{cases} \dfrac{1-\hat{\tau}}{2}, & i=0, \\ \dfrac{1+\hat{\tau}}{2}, & i=1, \\ \dfrac{1-\hat{\tau}}{2}\dfrac{1+\hat{\tau}}{2}P_{i-2}^{1,1}(\hat{\tau}), & 2\le i\le m, \end{cases} \tag{2.9}$$

where $P_i^{1,1}(\hat{\tau})$ denotes orthogonal Jacobi polynomials of degree i with respect to the weight function $(1-\hat{\tau})(1+\hat{\tau})$ [14]. Note the polynomial order of $\hat{\psi}_i$ is equal to i for $i \ge 2$. For the partition $\mathcal{T}_h$, we define the following finite element approximation space

$$W_h^{(p)} = \left\{\boldsymbol{v} : \boldsymbol{v}\in\mathbb{R}^n, v_i \in H^1(\Gamma_1), v_i|_{e_j} \in W_{e_j}^{(p)}, \boldsymbol{v}(0) = \boldsymbol{x}_1, \boldsymbol{v}(1) = \boldsymbol{x}_2\right\} \subset H^1(\Gamma_1;\mathbb{R}^n), \tag{2.10}$$

where $i=1,\cdots,n$, and $j=0,\cdots,N-1$, and the regularity requirement is due to the definition of action functional.

It is seen that $W_h^{(p)}$ consists of piecewise polynomials up to degree p and associated with the partition $\mathcal{T}_h$. $\hat{\psi}_0(\hat{\tau})$ and $\hat{\psi}_1(\hat{\tau})$ are consistent with linear finite element basis, and $\hat{\psi}_i(\hat{\tau})$, $2\leq i\leq m$, are introduced for high-order approximation. Note that $\hat{\psi}_i(\pm 1)=0$ for $2\leq i\leq m$. We call $\hat{\psi}_0(\hat{\tau})$ and $\hat{\psi}_1(\hat{\tau})$ boundary modes, and $\hat{\psi}_i(\hat{\tau})$, $2\leq i\leq m$, interior modes. According to the definition of the finite element basis (2.9), we have dual convergence paths: h- and p-convergence.

2.2.2 Two assumptions

It is seen in Section 2.1 that the critical issue for the application of the quasi-potential is to address the following optimization problem, i.e., Problem II

$$V(\boldsymbol{x}_1,\boldsymbol{x}_2)=S_{T^*}(\boldsymbol{\phi}^*)=\min_{\substack{T\in\mathbb{R}^+,\\ \boldsymbol{\phi}(0)=\boldsymbol{x}_1,\\ \boldsymbol{\phi}(T)=\boldsymbol{x}_2}}\left[S_T(\boldsymbol{\phi})=\frac{1}{2}\|\dot{\boldsymbol{\phi}}-\boldsymbol{b}(\boldsymbol{\phi})\|^2_{\boldsymbol{\alpha},t}\right], \tag{2.11}$$

where the minimizer $\boldsymbol{\phi}^*$ subject to the optimal transition time T^* is the minimal action path (MAP). T^* can be finite or infinite, depending on $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and the structure of the phase space. For instance, if either $\boldsymbol{x}_1$ or $\boldsymbol{x}_2$ is a critical point, then $T^*=\infty$. To approximate the quasi-potential, all transition paths will be approximated by the finite element space $W_h^{(p)}$, which will always correspond to a finite optimal integration time (see Lemma 2.1). We have two related convergence issues, both of which are from commonly used techniques in numerical approximation. We summarize them into two assumptions.

The first assumption is from truncation, where a large but finite integration time is used to deal with the case that $T^*=\infty$. This assumption has been used in [15, 26] for algorithm construction.

Assumption 2.1. Consider the transition from $\boldsymbol{x}_1$ to $\boldsymbol{x}_2$ in the phase space. If $T^*=\infty$ for the quasi-potential $V(\boldsymbol{x}_1,\boldsymbol{x}_2)$, we assume that $\{(T_k,\boldsymbol{\phi}_k^*)\}_{k=1}^{\infty}$ provides a minimizing sequence and $\{\boldsymbol{\phi}_k^*\}$ contains a subsequence that converges to $\boldsymbol{\phi}^*$ with respect to an appropriate norm, where $\{T_k\}$ is an increasing sequence going to ∞ and $\boldsymbol{\phi}_k^*$ is the minimizer of $S_{T_k}(\boldsymbol{\phi})$.

Let

$$V(t,\boldsymbol{x}_1,\boldsymbol{x}_2)=\min_{\substack{\boldsymbol{\phi}(0)=\boldsymbol{x}_1,\\ \boldsymbol{\phi}(t)=\boldsymbol{x}_2}} S_t(\boldsymbol{\phi}).$$

The Hamilton-Jacobi equation for $V(t,\boldsymbol{x}_1,\boldsymbol{x}_2)$ takes the form [1]

$$\frac{\partial V}{\partial t}+H(\boldsymbol{\phi},\nabla_{\boldsymbol{\phi}}V)=0,$$

where $H(\boldsymbol{\phi},\boldsymbol{p})$ is the Hamiltonian, see Eq. (2.22). If $T^*=\infty$, the Hamilton-Jacobi equation has a steady solution. Assumption 2.1 can be regarded as that the Hamilton-Jacobi solution evolves up to a large but finite time.

From the definition of action functional, it is seen that the rescaled MAP $\boldsymbol{\phi}^*$ should be searched in $H^1(\Gamma_1;\mathbb{R}^n)$ if T is prescribed. For numerical approximation, we need a subspace of $H^1(\Gamma_1;\mathbb{R}^n)$, which introduces our second assumption.

Assumption 2.2. Let $\hat{W}_\beta \subset H^1(\Gamma_1;\mathbb{R}^n)$ and $\hat{W}_\beta \to H^1(\Gamma_1;\mathbb{R}^n)$ as $\beta \to 0$, where β should be regarded as a parameter indicating the convergence of $\hat{W}_\beta$ to $H^1(\Gamma_1;\mathbb{R}^n)$ (For a fixed p, if we choose $W_h^{(p)}$ defined in Section 2.2.1, the parameter β can be regarded as h). Let $\boldsymbol{\phi}_\beta^*$ and T_β^* be the approximations of $\boldsymbol{\phi}^*$ and T^*, respectively, where $\boldsymbol{\phi}^* \in H^1(\Gamma_1;\mathbb{R}^n)$, and $\boldsymbol{\phi}_\beta^* \in \hat{W}_\beta$. We assume that $\boldsymbol{\phi}_\beta^* \to \boldsymbol{\phi}^*$ as $\beta \to 0$.

Assumption 2.2 is a very general assumption, which basically says that the numerical solution $\boldsymbol{\phi}_\beta^*$ is consistent with the direct approximation of $\boldsymbol{\phi}^*$ given by $\hat{W}_\beta$. The main reason that we make such an assumption is that the convergence analysis of MAM has not been fully implemented although many versions of discretization have been developed, including finite difference method [9,26] and finite element method [19]. So far the development of MAM has been mainly focused on the problem of clustering induced by the case that $T^* = \infty$, which corresponds to the fact that the Euler-Lagrange equation is a singularly perturbed problem for a large T (see Eq. (3.2)). Hence, a uniform discretization does not have the optimal convergence rate. The main goal of this paper is to recover the optimal convergence rate more effectively through adaptivity.

2.2.3 Generate a minimizing sequence by optimal linear time scaling

We rewrite the action functional with respect to the rescaled time $\tau = t/T$ [22]

$$S_T(\boldsymbol{\phi}) = \frac{T}{2}\left\| T^{-1}\boldsymbol{\phi}'(\tau) - \boldsymbol{b}(\boldsymbol{\phi})\right\|_{\boldsymbol{\alpha},\tau}^2, \tag{2.12}$$

where $\boldsymbol{\phi}'(\tau)$ is the derivative with respect to τ. According to Assumption 2.1, we can consider the following constrained optimization problem

$$\min_{\substack{T\in[0,M],\\ \boldsymbol{\phi}(0)=\boldsymbol{x}_1,\\ \boldsymbol{\phi}(1)=\boldsymbol{x}_2}} S_T(\boldsymbol{\phi}), \tag{2.13}$$

where M is a large but finite number. If the minimizer is not reached at the boundary $T = M$, the following optimality conditions should be satisfied:

$$\frac{\partial S_T}{\partial T} = 0, \quad \frac{\delta S_T}{\delta \boldsymbol{\phi}} = 0, \tag{2.14}$$

where we used functional derivative in the second equation. It is easy to see that for any given $\boldsymbol{\phi}$, $\partial S_T/\partial T = 0$ has a unique solution [22]

$$T = \hat{T}(\boldsymbol{\phi}) = \frac{\|\boldsymbol{\phi}'\|_{\boldsymbol{\alpha},\tau}}{\|\boldsymbol{b}(\boldsymbol{\phi})\|_{\boldsymbol{\alpha},\tau}}. \tag{2.15}$$

Then problem (2.13) can be simplified as

$$\min_{\substack{\boldsymbol{\phi}(0)=\boldsymbol{x}_1,\\ \boldsymbol{\phi}(1)=\boldsymbol{x}_2}} S_{\hat{T}(\boldsymbol{\phi})}(\boldsymbol{\phi}), \tag{2.16}$$

where the optimality condition with respect to T is enforced. This simplification is true if T^* is finite since there always exists an $M > T^*$. However, when $T^* = \infty$, the minimizer of problem (2.13) will be reached at $T = M$, which makes the linear scaling not valid.

To verify the effectiveness of linear scaling for the case that $T^* = \infty$, we need to look into the discrete version of problem (2.16). Let $\boldsymbol{\phi}_h \in W_h^{(p)}$ be an approximation of $\boldsymbol{\phi}$, where we fixed p (this does not affect our following argument). We write the discrete version of problem (2.16) as

$$\min_{\substack{\boldsymbol{\phi}_h(0)=\boldsymbol{x}_1,\\ \boldsymbol{\phi}_h(1)=\boldsymbol{x}_2}} S_{\hat{T}(\boldsymbol{\phi}_h)}(\boldsymbol{\phi}_h). \tag{2.17}$$

The key observation is the following lemma [25]:

Lemma 2.1. *If $\boldsymbol{\phi}_h^* \in W_h^{(p)}$ is a (local) minimizer of problem* (2.17)*, we have*

$$\hat{T}(\boldsymbol{\phi}_h^*) < \infty. \tag{2.18}$$

Proof. We argue by contradiction. Note that

$$\begin{aligned}\hat{T}^2(\boldsymbol{\phi}_h^*) &= \frac{\|(\boldsymbol{\phi}_h^*)'\|_{\boldsymbol{\alpha},\tau}^2}{\|\boldsymbol{b}(\boldsymbol{\phi}_h^*)\|_{\boldsymbol{\alpha},\tau}^2} \\ &\le \frac{\sup_{\tau\in[0,1]} \|\boldsymbol{\alpha}^{-1}(\boldsymbol{\phi}_h^*)\| \|(\boldsymbol{\phi}_h^*)'\|_\tau^2}{\|\boldsymbol{b}(\boldsymbol{\phi}_h^*)\|_{\boldsymbol{\alpha},\tau}^2} \le C \frac{\sup_{\tau\in[0,1]} \|\boldsymbol{\alpha}^{-1}(\boldsymbol{\phi}_h^*)\| \|\boldsymbol{\phi}_h^*\|_\tau^2}{\|\boldsymbol{b}(\boldsymbol{\phi}_h^*)\|_{\boldsymbol{\alpha},\tau}^2},\end{aligned}$$

where the last inequality is from the inverse inequality of finite element discretization and the constant C only depends on mesh [6]. Due to the properties of $\boldsymbol{\alpha}$ and the continuity of $\boldsymbol{\phi}_h$, $\sup_{\tau\in[0,1]} \|\boldsymbol{\alpha}^{-1}\| < \infty$. If $\hat{T}(\boldsymbol{\phi}_h^*) = \infty$, we have two possible cases: $\|\boldsymbol{b}(\boldsymbol{\phi}_h^*)\|_{\boldsymbol{\alpha},\tau} = 0$ or $\|\boldsymbol{\phi}_h^*\|_\tau = \infty$. The first case implies that $\boldsymbol{\phi}_h^*$ is a fixed point almost everywhere for $\tau \in [0,1]$, which is not true. The second case implies that $\boldsymbol{\phi}_h^*$ must go to infinity somewhere due to the continuity. It is reasonable to assume that infinity is not a saddle point, which means that this case is also impossible if $\boldsymbol{\phi}_h^*$ is a (local) minimizer. □

Since $\hat{T}(\boldsymbol{\phi}_h^*) < \infty$, problem (2.17) is well defined. In other words, although problems (2.13) and (2.16) are, in general, not equivalent, their discrete versions are equivalent in the sense that there always exists an M such that $M > \hat{T}(\boldsymbol{\phi}_h^*)$. Thus, the discrete version of the original problem (2.11), i.e.,

$$\min_{\substack{T\in\mathbb{R}^+,\\ \boldsymbol{\phi}_h(0)=\boldsymbol{x}_1,\\ \boldsymbol{\phi}_h(T)=\boldsymbol{x}_2}} S_T(\boldsymbol{\phi}_h) \tag{2.19}$$

can only reach its minimum at a finite time although $T^*=\infty$, i.e., Lemma 2.1 establishes the equivalence between problems (2.19) and (2.17). It can be shown that $\hat{T}(\boldsymbol{\phi}_h^*)$ goes to infinity as the mesh is refined, see Remark 2.1 for an example. This way, we do not need to choose M explicitly as in problem (2.13). Instead we consider a sequence of refined meshes for problem (2.17), and obtain a minimizing sequence $\{(\hat{T}(\boldsymbol{\phi}_h^*),\boldsymbol{\phi}_h^*)\}$ as h goes to zero.

Remark 2.1. Consider a special case that $\boldsymbol{\phi}_h$ is defined on a quasi-uniform mesh, and $\boldsymbol{\alpha}$ is an identity matrix, and $\boldsymbol{b}(\boldsymbol{\phi})$ is linear with respect to $\boldsymbol{\phi}$. We have

$$\hat{T}(\boldsymbol{\phi}_h^*)=\frac{\|\left(\boldsymbol{\phi}_h^*\right)'\|_\tau}{\|\boldsymbol{b}(\boldsymbol{\phi}_h^*)\|_\tau}\sim\frac{\|\left(\boldsymbol{\phi}_h^*\right)'\|_\tau}{\|\boldsymbol{\phi}_h^*\|_\tau}\sim\frac{1}{h}<\infty,$$

where h^{-1} is from the inverse inequality of finite element discretization. For this case, $\hat{T}(\boldsymbol{\phi}_h^*)$ is of $\mathcal{O}(h^{-1})$, and goes to infinity as $h\to 0$.

Remark 2.2. According to Assumption 2.1, we assume that $\boldsymbol{\phi}_h^*\to\boldsymbol{\phi}^*$ with respect to a proper norm. From the proof of Lemma 2.1, it is seen that $\hat{T}(\boldsymbol{\phi}_h^*)\to\infty$ implies $\|\left(\boldsymbol{\phi}_h^*\right)'\|_\tau\to\infty$ since $\|\boldsymbol{b}(\boldsymbol{\phi}_h^*)\|_{\boldsymbol{\alpha},\tau}>0$. This means that $\boldsymbol{\phi}^*\notin H^1(\Gamma_1;\mathbb{R}^n)$ although $\boldsymbol{\phi}_h^*\in H^1(\Gamma_1;\mathbb{R}^n)$ for any h, which is not surprising since we are not able to implement linear scaling on an infinite time interval. So the convergence of $\boldsymbol{\phi}_h^*$ to $\boldsymbol{\phi}^*$ needs to be addressed in a larger space than $H^1(\Gamma_1;\mathbb{R}^n)$, e.g., the space of absolutely continuous functions [25].

2.2.4 The zero-Hamiltonian constraint

Although we can use problem (2.17) to generate a minimizing sequence, the fact that problem (2.16) is not well defined for $T^*=\infty$ implies that the minimizer $\boldsymbol{\phi}^*$ should satisfy a stronger condition than the optimal linear time scaling. Consider a change of variable in general, say $s=s(t)$, we have (see Lemma 3.1, Ch. 4 in [12])

$$S_T(\boldsymbol{\phi})\geq S(\tilde{\boldsymbol{\phi}})=\int_{s(0)}^{s(T)}(|\tilde{\boldsymbol{\phi}}'|_{\boldsymbol{\alpha}}|\boldsymbol{b}|_{\boldsymbol{\alpha}}-\langle\tilde{\boldsymbol{\phi}}',\boldsymbol{b}\rangle_{\boldsymbol{\alpha}})ds, \tag{2.20}$$

where $\tilde{\boldsymbol{\phi}}(s)=\boldsymbol{\phi}(t(s))$, $\tilde{\boldsymbol{\phi}}'$ is the derivative with respect to s, and the equality holds if the mapping from t to s is given by the condition

$$|\dot{\boldsymbol{\phi}}|_{\boldsymbol{\alpha}}=|\boldsymbol{b}(\boldsymbol{\phi})|_{\boldsymbol{\alpha}},\quad\forall t. \tag{2.21}$$

Another way to obtain this constraint is to consider the Hamiltonian

$$H(\boldsymbol{\phi},\boldsymbol{p})=\langle\boldsymbol{b}(\phi),\boldsymbol{p}\rangle+\frac{1}{2}\langle\boldsymbol{p},\boldsymbol{\alpha}\boldsymbol{p}\rangle, \tag{2.22}$$

which is obtained from the Legendre transform of the Lagrangian

$$L(\boldsymbol{\phi},\dot{\boldsymbol{\phi}})=\frac{1}{2}\langle\dot{\boldsymbol{\phi}}-\boldsymbol{b}(\boldsymbol{\phi}),\dot{\boldsymbol{\phi}}-\boldsymbol{b}(\boldsymbol{\phi})\rangle_{\boldsymbol{\alpha}} \tag{2.23}$$

with respect to $\dot{\boldsymbol{\phi}}$. Since the Hamiltonian does not depend on time, we know it is a first integral, i.e., $H \equiv cnst$. If $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are located on a trajectory of the unperturbed system, the minimizer $\boldsymbol{\phi}^*$ is nothing but the trajectory implying that $\boldsymbol{p} = \frac{\partial L}{\partial \dot{\boldsymbol{\phi}}} = \boldsymbol{\alpha}^{-1}(\dot{\boldsymbol{\phi}} - \boldsymbol{b}(\boldsymbol{\phi})) = 0$ and $H(\boldsymbol{\phi}, \boldsymbol{p}) = 0$. We then have

$$H\left(\boldsymbol{\phi}, \frac{\partial L}{\partial \dot{\boldsymbol{\phi}}}\right) = \frac{1}{2}\langle \dot{\boldsymbol{\phi}}, \dot{\boldsymbol{\phi}} \rangle_{\boldsymbol{\alpha}} - \frac{1}{2}\langle \boldsymbol{b}(\boldsymbol{\phi}), \boldsymbol{b}(\boldsymbol{\phi}) \rangle_{\boldsymbol{\alpha}} = 0, \tag{2.24}$$

which yields the constraint (2.21). We refer to it as zero-Hamiltonian constraint in this paper. According to Maupertuis' principle of least action, the minimizer $\boldsymbol{\phi}^*$ can be found in the geometrically fixed curves lying on the surface $H(\boldsymbol{\phi}, \boldsymbol{p}) = 0$ and connecting $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ [1], where the geodesic lines are defined by the metric $ds = |\dot{\boldsymbol{\phi}}|_{\boldsymbol{\alpha}} dt = |\boldsymbol{b}(\boldsymbol{\phi})|_{\boldsymbol{\alpha}} dt$. Instead of using transition paths parametrized by time, the gMAM uses the geodesic lines on the surface $H(\boldsymbol{\phi}, \boldsymbol{p}) = 0$ [13].

It is easy to see that the functional $\hat{T}(\boldsymbol{\phi})$ is consistent with the constraint $H = 0$ for any fixed T. Note that if

$$\left|\frac{d\boldsymbol{\phi}}{dt}\right|_{\boldsymbol{\alpha}} = \left|\frac{d\boldsymbol{\phi}}{d\tau}\right|_{\boldsymbol{\alpha}} T^{-1} = |\boldsymbol{b}(\boldsymbol{\phi})|_{\boldsymbol{\alpha}}, \quad \forall \tau \in [0,1],$$

we can obtain $T = \hat{T}(\boldsymbol{\phi})$ by integrating both sides on Γ_1. This means that if T^* is finite, it can be recovered by $\boldsymbol{\phi}^*$ as $T^* = \hat{T}(\boldsymbol{\phi}^*)$.

The deviation of $\boldsymbol{\phi}_h^*$ (the solution of problem (2.17)) from the surface $H(\boldsymbol{\phi}, \boldsymbol{p}) = 0$ is twofold: First, $\boldsymbol{\phi}_h$ cannot satisfy Eq. (2.21) pointwisely due to the non-differentiation of boundary modes (see Eq. (2.9)). Second, when $T^* = \infty$ the integration time is truncated by the optimal linear time scaling enforced in problem (2.17). However, as the mesh is refined, the deviation from the surface $H(\boldsymbol{\phi}, \boldsymbol{p}) = 0$ should decrease. Since $H = 0$ is a necessary condition satisfied by the MAP $\boldsymbol{\phi}^*$, we also want to minimize the deviation of $\boldsymbol{\phi}_h^*$ from the surface $H(\boldsymbol{\phi}, \boldsymbol{p}) = 0$ through mesh refinement.

2.2.5 An outline of our algorithm

To this end, we are able to provide an outline of our algorithm:

(1) Choose an approximation space $W_h^{(p)}$ for $\boldsymbol{\phi}_h$.

(2) Solve the discrete problem (2.17) to obtain $\boldsymbol{\phi}_h^*$.

(3) Examine the quality of the mesh. Refine the mesh and go to step (2).

The loop generated by step (2) and (3) will be stopped when a certain criterion is reached. The rest of this paper will focus on step (3). In particular, we will consider *hp*-refinement of the mesh.

3 An hp-adaptive tMAM

3.1 Properties of dynamics

We first derive the Euler-Lagrange equation of problem (2.17), and then illustrate the necessity of adaptivity using a typical scenario for noise-induced transitions.

Lemma 3.1. *The Euler-Lagrange equation of problem* (2.17) *is*

$$\hat{T}^{-1}\left(\boldsymbol{\alpha}^{-1}(\hat{T}^{-1}\boldsymbol{\phi}_h'-\boldsymbol{b})\right)'+\left((\nabla_{\boldsymbol{\phi}_h}\boldsymbol{b})^{\mathsf{T}}\boldsymbol{\alpha}^{-1}-\frac{1}{2}\boldsymbol{B}^{\mathsf{T}}\right)(\hat{T}^{-1}\boldsymbol{\phi}_h'-\boldsymbol{b})=0. \tag{3.1}$$

Here $\boldsymbol{B}\in\mathbb{R}^{n\times n}$ *satisfies*

$$\delta\left(\boldsymbol{\alpha}^{-1}\right)(\hat{T}^{-1}\boldsymbol{\phi}_h'-\boldsymbol{b})=\boldsymbol{B}\delta\boldsymbol{\phi}_h,$$

where $\delta\left(\boldsymbol{\alpha}^{-1}\right)\in\mathbb{R}^{n\times n}$ *is the linear perturbation of* $\boldsymbol{\alpha}^{-1}$ *subject to a perturbation function* $\delta\boldsymbol{\phi}_h\in\mathbb{R}^n$. *For a fixed integration time, we can replace* $\hat{T}$ *with* T.

Proof. Let $\delta\boldsymbol{\phi}_h$ be a arbitrary perturbation function with $\delta\boldsymbol{\phi}_h(0)=\delta\boldsymbol{\phi}_h(1)=0$. Let $\delta\hat{T}$ be the first-order variation of $\hat{T}(\boldsymbol{\phi}_h)$ and $\delta(\boldsymbol{\alpha}^{-1})$ be the linear perturbation of the matrix $\boldsymbol{\alpha}^{-1}$. Let $\boldsymbol{f}=\hat{T}^{-1}\boldsymbol{\phi}_h'-\boldsymbol{b}$. Then the linear part of $\hat{S}_{\hat{T}(\boldsymbol{\phi}_h+\delta\boldsymbol{\phi}_h)}(\boldsymbol{\phi}_h+\delta\boldsymbol{\phi}_h)-\hat{S}_{\hat{T}(\boldsymbol{\phi}_h)}(\boldsymbol{\phi}_h)$ is

$$\delta\hat{S}_{\hat{T}}=\hat{T}\langle\boldsymbol{f},\delta\boldsymbol{f}\rangle_{\boldsymbol{\alpha},\tau}+\frac{1}{2}\hat{T}\langle\boldsymbol{f},\delta(\boldsymbol{\alpha}^{-1})\boldsymbol{f}\rangle_{\tau}+\frac{1}{2}\delta\hat{T}\langle\boldsymbol{f},\boldsymbol{f}\rangle_{\boldsymbol{\alpha},\tau}.$$

Substituting $\delta\boldsymbol{f}=\hat{T}^{-1}\delta\boldsymbol{\phi}_h'-\hat{T}^{-2}\boldsymbol{\phi}_h'\delta\hat{T}-\nabla_{\boldsymbol{\phi}_h}\boldsymbol{b}\delta\boldsymbol{\phi}_h$ and $\delta(\boldsymbol{\alpha}^{-1})\boldsymbol{f}=\boldsymbol{B}\delta\boldsymbol{\phi}_h$, we have $\delta\hat{S}_{\hat{T}}=I_1+I_2$, where I_2 includes all terms related to $\delta\hat{T}$. We have

$$\begin{aligned}I_2&=\frac{1}{2}\delta\hat{T}\langle\boldsymbol{f},\boldsymbol{f}\rangle_{\boldsymbol{\alpha},\tau}-\delta\hat{T}\langle\boldsymbol{f},\hat{T}^{-1}\boldsymbol{\phi}_h'\rangle_{\boldsymbol{\alpha},\tau}\\&=-\frac{1}{2}\delta\hat{T}(\hat{T}^{-2}\|\boldsymbol{\phi}_h'\|_{\boldsymbol{\alpha},\tau}^2-\|\boldsymbol{b}\|_{\boldsymbol{\alpha},\tau}^2)=0,\end{aligned}$$

where we used Eq. (2.15). I_1 can be expressed as

$$\begin{aligned}I_1&=\langle\boldsymbol{f},\delta\boldsymbol{\phi}_h'\rangle_{\boldsymbol{\alpha},\tau}-\hat{T}\langle\boldsymbol{f},\nabla_{\boldsymbol{\phi}_h}\boldsymbol{b}\delta\boldsymbol{\phi}_h\rangle_{\boldsymbol{\alpha},\tau}+\frac{1}{2}\hat{T}\langle\boldsymbol{f},\boldsymbol{B}\delta\boldsymbol{\phi}_h\rangle_{\tau}\\&=-\langle(\boldsymbol{\alpha}^{-1}\boldsymbol{f})',\delta\boldsymbol{\phi}_h\rangle_{\tau}-\hat{T}\langle(\nabla_{\boldsymbol{\phi}_h}\boldsymbol{b})^{\mathsf{T}}(\boldsymbol{\alpha}^{-1})\boldsymbol{f},\delta\boldsymbol{\phi}_h\rangle_{\tau}+\frac{1}{2}\hat{T}\langle\boldsymbol{B}^{\mathsf{T}}\boldsymbol{f},\delta\boldsymbol{\phi}_h\rangle_{\tau},\end{aligned}$$

where integration by parts is applied. Due to the arbitrariness of $\delta\boldsymbol{\phi}_h$, $\delta\hat{S}_{\hat{T}}=0$ yields that

$$(\boldsymbol{\alpha}^{-1}\boldsymbol{f})'+\hat{T}\left((\nabla_{\boldsymbol{\phi}_h}\boldsymbol{b})^{\mathsf{T}}(\boldsymbol{\alpha}^{-1})-\frac{1}{2}\boldsymbol{B}^{\mathsf{T}}\right)\boldsymbol{f}=0.$$

The proof is completed. □

Let $\boldsymbol{\alpha}=\boldsymbol{I}$ with $\boldsymbol{I}$ being an identity matrix. The E-L equation (3.1) becomes

$$\hat{T}^{-2}\boldsymbol{\phi}_h''+\hat{T}^{-1}\left((\nabla_{\boldsymbol{\phi}_h}\boldsymbol{b})^{\mathsf{T}}-\nabla_{\boldsymbol{\phi}_h}\boldsymbol{b}\right)\boldsymbol{\phi}_h'-(\nabla_{\boldsymbol{\phi}_h}\boldsymbol{b})^{\mathsf{T}}\boldsymbol{b}=0, \tag{3.2}$$

which corresponds to a nonlocal nonlinear elliptic boundary value problem. If $T^*=\infty$, we know that $\hat{T}(\phi_h^*)\to\infty$ as $h\to 0$. Then Eq. (3.2) should have similar properties with the following singularly perturbed equation

$$\epsilon^2\boldsymbol{\phi}''+\epsilon\left((\nabla_{\boldsymbol{\phi}}\boldsymbol{b})^{\mathsf{T}}-\nabla_{\boldsymbol{\phi}}\boldsymbol{b}\right)\boldsymbol{\phi}'-(\nabla_{\boldsymbol{\phi}}\boldsymbol{b})^{\mathsf{T}}\boldsymbol{b}=0,$$

where $0<\epsilon\ll 1$. It is well known that a singularly perturbed problem has internal/boundary layers, which need to be resolved by adaptivity, i.e., fine mesh in the internal/boundary layers and coarse mesh elsewhere. For our problem, the internal/boundary layers are related to critical points.

Consider a MAP $\boldsymbol{\phi}_{1\to s}^*$ from a fixed point $\boldsymbol{x}_1$ to a saddle point $\boldsymbol{x}_s$ without any other critical points along the MAP. Apparently, the dynamics is slow around $\boldsymbol{x}_1$ and $\boldsymbol{x}_s$, and fast elsewhere. Let $\boldsymbol{\phi}_{T,1\to s}^*$ be an approximation of $\boldsymbol{\phi}_{1\to s}^*$ given by a finite T. With respect to the scaled parameter $\tau\in[0,1]$, we have the following picture of $\boldsymbol{\phi}_{T,1\to s}^*$. When T is large enough, the zero-Hamiltonian constraint should hold approximately, i.e.,

$$T^{-1}|(\boldsymbol{\phi}_{T,1\to s}^*)'|_{\boldsymbol{\alpha}}\approx|\boldsymbol{b}(\boldsymbol{\phi}_{T,1\to s}^*)|_{\boldsymbol{\alpha}},\quad\forall\tau. \tag{3.3}$$

In the region of fast dynamics, $|\boldsymbol{b}(\boldsymbol{\phi}_{T,1\to s}^*)|_{\boldsymbol{\alpha}}=\mathcal{O}(1)$, which implies that $|(\boldsymbol{\phi}_{T,1\to s}^*)'|_{\boldsymbol{\alpha}}=\mathcal{O}(T)$. In other words, the time taken in the region of fast dynamics is $\mathcal{O}(T^{-1})$. Then, on the time interval $[0,1]$, we can have an internal boundary layer of width $\mathcal{O}(T^{-1})$, which contributes most to the transition and must be well resolved.

3.2 Indicators of hp-adaptivity

When we move from the original problem (2.11) to the discrete problem (2.17), we include two types of approximation:

- Model approximation: When T^* is finite, the optimal linear time scaling is a global effect of the zero-Hamiltonian constraint defined pointwisely. However, this is not true for the case that $T^*=\infty$. We regard the deviation from the surface $H(\boldsymbol{\phi},\boldsymbol{p})=0$ as the error of model approximation.

- Path approximation: We use problem (2.17) to generate a minimizing sequence. When $T^*=\infty$, there always exists a larger integration time such that the corresponding MAP yields a smaller action. We regard the deviation of current MAP from the MAP given by a larger integration time as error of path approximation (see Remark 3.1).

To measure the quality of the numerical MAP, we will define an error indicator for each type of approximation. We will also define a regularity indicator to help us make a choice between h- and p-refinement.

Remark 3.1. For the path approximation, if T^* is finite, $\boldsymbol{\phi}^* \in H^1(\Gamma_1;\mathbb{R}^n)$ and we define the difference between $\boldsymbol{\phi}_h^*$ and $\boldsymbol{\phi}^*$ as the error of path approximation. When $T^*=\infty$, we cannot do this because $\boldsymbol{\phi}_h^* \in H^1(\Gamma_1;\mathbb{R}^n)$ while $\boldsymbol{\phi}^* \notin H^1(\Gamma_1;\mathbb{R}^n)$. The convergence of $\boldsymbol{\phi}_h^*$ to $\boldsymbol{\phi}^*$ cannot be addressed in $H^1(\Gamma_1;\mathbb{R}^n)$ but in the space of absolutely continuous functions. This issue is not practically important. We can always consider an approximation $\boldsymbol{\phi}^{M,*}$ of $\boldsymbol{\phi}^*$ subject to an integration time M that is larger than $\hat{T}(\boldsymbol{\phi}_h^*)$. As long as M is finite, we know that $\boldsymbol{\phi}^{M,*} \in H^1(\Gamma_1;\mathbb{R}^n)$ subject to a linear time scaling. We can use the difference between $\boldsymbol{\phi}^{M,*}$ and $\boldsymbol{\phi}_h^*$ to measure the error of path approximation.

3.2.1 Adaptivity in aMAM and tMAM

Adaptivity is a crucial component for both the aMAM [26] and tMAM [22]. In aMAM, a moving mesh technique is employed to redistribute all grid points every a certain number of iteration steps such that the grid points are nearly uniform with respect to arc length. In tMAM, h-adaptivity is employed, where the elements of the largest arc length are refined. In other words, both aMAM and tMAM use the arc length as an indicator for the region that needs a better resolution. In gMAM [13], no adaptivity is used because it is formulated directly with respect to rescaled arc length $\alpha \in [0,1]$. The internal/boundary layers are thus removed and a uniform mesh with respect to α can be used. However, it is not surprising that the nonlinear mapping between α and t is not well behaved around critical points, e.g., the Jacobian is singular at critical points. If there exist unknown critical points along the MAP, they must be identified to avoid deterioration of accuracy [13]. This implies that h-refinement might be necessary for gMAM to capture the unknown saddle points while maintaining the accuracy. In [28], it is confirmed that a non-uniform mesh with respect to arc length yields a better accuracy assuming that the saddle point on the MAP is not explicitly specified.

Simply speaking, in current MAMs the most effective way to improve accuracy is to use non-uniform meshes given by physically based adaptivity criteria. The drawback of current adaptivity criteria is that they do not take into account the regularity of MAP with respect to either time or arc length, which is particularly important for high-order approximation. Thus, new strategies are needed when considering hp-adaptivity.

3.2.2 Indicator for the zero-Hamiltonian constraint

We first look at the model approximation. The zero-Hamiltonian constraint is a necessary condition satisfied by the minimizer. Since the consistency between the linear time scaling and the zero-Hamiltonian constraint can only be reached when T^* is finite, it is important to measure the deviation of $\boldsymbol{\phi}_h$ from the zero-Hamiltonian constraint when

$T^*=\infty$. We define the following indicator on element $e_i:=[\tau_i,\tau_{i+1}]$ [22]:

$$\begin{aligned}\theta_{e_i}^2&=\int_{\hat{T}\tau_i}^{\hat{T}\tau_{i+1}}(|\dot{\boldsymbol{\phi}}_h|_{\boldsymbol{\alpha}}-|\boldsymbol{b}|_{\boldsymbol{\alpha}})^2dt\\&=\hat{T}\int_{\tau_i}^{\tau_{i+1}}(\hat{T}^{-1}|\boldsymbol{\phi}_h'|_{\boldsymbol{\alpha}}-|\boldsymbol{b}|_{\boldsymbol{\alpha}})^2d\tau,\quad i=0,\cdots,N-1.\end{aligned}\tag{3.4}$$

We use θ_{e_i} as an indicator of the error of model approximation, which should decrease to zero as the mesh is refined. Let $\theta_{\max}$ and $\theta_{\min}$ be the maximum and minimum values of θ_{e_i} respectively. If $\theta_{\max}/\theta_{\min}>\beta_\theta$, we will apply h-refinement to elements with the largest θ_{e_i}, where β_θ is a prescribed threshold. In other words, we expect that the element-wise distribution of the deviation from the zero-Hamiltonian constraint will be uniform. We do not consider p-refinement for model approximation.

3.2.3 Error indicator based on derivative recovery

We now look at the path approximation, i.e., solving problem (2.17). We will construct error indicators using the derivative recovery technique [3, 4] in this work. Other techniques of a posteriori error estimate, such as residual-based estimate, can also be considered. The motivation we use derivative recovery is its simplicity and flexibility, where we only work with the numerical path and do not need the information of $\boldsymbol{b}(\boldsymbol{\phi})$.

For any vector function $\boldsymbol{\phi}\in L^2(\Gamma_1;\mathbb{R}^n)$, we define the projection operator $\mathcal{Q}_h$ satisfying

$$\langle\mathcal{Q}_h\boldsymbol{\phi},\boldsymbol{\psi}_h\rangle_\tau=\langle\boldsymbol{\phi},\boldsymbol{\psi}_h\rangle_\tau,\quad\forall\boldsymbol{\psi}_h\in W_h^{(1)},\tag{3.5}$$

where we simply do an L^2 projection onto linear finite element space for each component of $\boldsymbol{\phi}$. We also define a smoothing operator $\mathcal{S}_h=\mathcal{I}-\lambda^{-1}\mathcal{A}_h$, where $\mathcal{I}$ is an identity operator and $\mathcal{A}_h:W_h^{(1)}\to W_h^{(1)}$ is uniquely determined by

$$\langle\mathcal{A}_h\boldsymbol{\phi}_h,\boldsymbol{\psi}_h\rangle_\tau=\langle\boldsymbol{\phi}_h',\boldsymbol{\psi}_h'\rangle_\tau+\langle\boldsymbol{\phi}_h,\boldsymbol{\psi}_h\rangle_\tau,\quad\forall\boldsymbol{\phi}_h,\boldsymbol{\psi}_h\in W_h^{(1)}.\tag{3.6}$$

Here $\lambda=\rho(\mathcal{A}_h)\simeq h^{-2}$ with h being the element size. We employ the smoothing operator $\mathcal{S}_h$ mainly because that the E-L equation (3.1) is a (nonlinear) elliptic equation.

Let $\boldsymbol{\phi}_{hp}\in W_h^{(p)}$ be a finite element approximation of a vector function $\boldsymbol{\phi}\in\mathbb{R}^n$. The p-th derivative $\boldsymbol{\phi}_{hp}^{(p)}$ is a piecewise constant vector. Then the recovered p-th derivative is defined as $\mathcal{R}\boldsymbol{\phi}_{hp}^{(p)}=\mathcal{S}_h^m\mathcal{Q}_h\boldsymbol{\phi}_{hp}^{(p)}$. Our error estimator will be based on the following two theorems.

Theorem 3.1. *Let ϕ_i be the i-th component of $\boldsymbol{\phi}_{hp}$, $i=1,\cdots,n$. Assume that $\phi_i\in H^{p+2}(\Gamma_1)\cap W^{p+1,\infty}(\Gamma_1)$, and $\boldsymbol{\phi}_{hp}\in W_h^{(p)}$ be an approximation of $\boldsymbol{\phi}$ satisfying*

$$\|\boldsymbol{\phi}-\boldsymbol{\phi}_{hp}\|_{h,p-1,\Gamma_1}\lesssim h^2|\boldsymbol{\phi}|_{p+1,\Gamma_1},\tag{3.7}$$

$$\|\boldsymbol{\phi}-\boldsymbol{\phi}_{hp}\|_{h,p-1,\infty,\Gamma_1}\lesssim h^2|\boldsymbol{\phi}|_{p+1,\infty,\Gamma_1},\tag{3.8}$$

where $\|*\|_{h,\cdot,\Gamma_1}$ *indicates the discrete norm defined as* $\|*\|_{h,\cdot,\Gamma_1}=\sum_{i=0}^{N-1}\|*\|_{\cdot,e_i}$. *Then*

$$\|\boldsymbol{\phi}^{(p)}-\mathcal{R}\boldsymbol{\phi}_{hp}^{(p)}\|_{0,\Gamma_1}\lesssim h(mh^{1/2}+\epsilon_m)|||\boldsymbol{\phi}|||, \tag{3.9}$$

$$\|(\boldsymbol{\phi}^{(p)}-\mathcal{R}\boldsymbol{\phi}_{hp}^{(p)})'\|_{0,\Gamma_1}\lesssim (mh^{1/2}+\epsilon_m)|||\boldsymbol{\phi}|||, \tag{3.10}$$

where $|||\boldsymbol{\phi}|||=\|\boldsymbol{\phi}\|_{p+2,\Gamma_1}+|\boldsymbol{\phi}|_{p+1,\infty,\Gamma_1}$, $\epsilon_m=(1-\kappa^{-1})^m$ *with* $\kappa=(Ch^2)\lambda$ *for some constant* C *and small* $m\in\mathbb{N}$.

Theorem 3.2. *Assume that the hypotheses of Theorem* 3.1 *hold and there exists a positive constant* $C(\boldsymbol{\phi})$ *such that*

$$\|\boldsymbol{\phi}^{(p)}-\boldsymbol{\phi}_{hp}^{(p)}\|_{0,\Gamma_1}\geq C(\boldsymbol{\phi})h. \tag{3.11}$$

Then

$$\left|\frac{\|(\mathcal{I}-\mathcal{R})\boldsymbol{\phi}_{hp}^{(p)}\|_{0,\Gamma_1}}{\|\boldsymbol{\phi}^{(p)}-\boldsymbol{\phi}_{hp}^{(p)}\|_{0,\Gamma_1}}-1\right|\lesssim (mh^{1/2}+\epsilon_m)|||\boldsymbol{\phi}|||. \tag{3.12}$$

Remark 3.2. Theorem 3.1 mostly is a restatement of Theorems 2.5 and 2.6 in [3] with respect to appropriate norms for the vector function $\boldsymbol{\phi}$. Theorem 3.2 can be easily derived from Theorem 3.1, also see [4]. In [3], the operator $\mathcal{R}=\mathcal{S}_h^m\mathcal{Q}_h$ was studied for high-order derivatives of a finite element solution defined on unstructured but shape regular triangulations. The results remain valid for one-dimensional functions. Theorem 3.1 can be proved by following the proof of Theorem 2.5 in [3], where the key modification is to replace trace theorem with Sobolev inequality since we have point values on the boundary instead of a manifold. Due to this observation, we omit the tedious proof here.

Theorems 3.1 investigates the superconvergence of $\mathcal{R}\boldsymbol{\phi}_{hp}^{(p)}$ to the true derivative $\boldsymbol{\phi}^{(p)}$ in an asymptotically exact sense, i.e, the right-hand sides of inequality (3.9) and (3.10) goes to zero as $h\to 0$ and $m\to\infty$ in a appropriate fashion, also see [2]. Theorem 3.1 implies that $(\mathcal{I}-\mathcal{R})\boldsymbol{\phi}_{hp}^{(p)}$ can be used as an estimator of the error $\boldsymbol{\phi}^{(p)}-\boldsymbol{\phi}_{hp}^{(p)}$. The operator $\mathcal{R}=\mathcal{S}_h^m\mathcal{Q}_h$ consists of two consecutive steps: L^2 projection given by $\mathcal{Q}_h$ and multigrid-like smoothing given by $\mathcal{S}_h^m$. Note that $\mathcal{Q}_h$ always projects a piecewise constant vector onto $W_h^{(1)}$, which does not depend on the polynomial order p. In other words, we can apply $\mathcal{R}$ to $\boldsymbol{\phi}_{hp}^{(p)}$ repeatedly starting from $p=1$ such that the error of high-order derivatives can be estimated and p-adaptivity becomes feasible.

We now focus on the polynomial approximation $\boldsymbol{\phi}_p$ on element e_i. For simplicity, we here neglect the subscript h. Let ψ_i be the i-th local basis function, where the subscript i is equal to the polynomial order for $i\geq 2$, see Eq. (2.9). We consider the following polynomial extension $\tilde{\boldsymbol{\phi}}_{p+1}$ of degree $p+1$:

$$\tilde{\boldsymbol{\phi}}_{p+1}-\boldsymbol{\phi}_p=\mathrm{diag}(\boldsymbol{c})(\boldsymbol{\psi}_{p+1}-\mathcal{P}_p\boldsymbol{\psi}_{p+1}), \tag{3.13}$$

where $\boldsymbol{c}\in\mathbb{R}^n$, $\boldsymbol{\psi}_{p+1}=\psi_{p+1}[1,1,\cdots,1]^\mathsf{T}$, and $\mathcal{P}_p$ is a projection operator for each component of $\boldsymbol{\psi}_{p+1}$ onto the space $\text{span}\{\psi_i\}_{i=0}^p$ with $p\geq 1$. Since ψ_{p+1} is not orthogonal to $\text{span}\{\psi_i\}_{i=0}^p$, its projection needs to be removed from the extension. Taking the $(p+1)$-th derivative of both sides of Eq. (3.13), we have

$$c_i=\left(\tilde{\boldsymbol{\phi}}_{p+1}^{(p+1)}\right)_i\left(\boldsymbol{\psi}_{p+1}^{(p+1)}\right)_i^{-1},\quad i=1,\cdots,n. \tag{3.14}$$

We then use $(R\boldsymbol{\phi}_p^{(p)})'$ to approximate $\tilde{\boldsymbol{\phi}}_{p+1}^{(p+1)}$ to obtain $\boldsymbol{c}$. Noting the formula

$$\frac{d^j}{d\hat{\tau}^j}P_i^{\alpha,\beta}(\hat{\tau})=\frac{\Gamma(i+j+\alpha+\beta+1)}{2^j\Gamma(i+\alpha+\beta+1)}P_{i-j}^{\alpha+j,\beta+j}(\hat{\tau}), \tag{3.15}$$

where the function $\Gamma(\cdot)$ satisfies $\Gamma(m+1)=m!$ for $m\in\mathbb{N}_0$, we can obtain that

$$\begin{aligned}\frac{d^{p+1}}{d\hat{\tau}^{p+1}}\hat{\psi}_{p+1}(\hat{\tau})&=\frac{d^{p+1}}{d\hat{\tau}^{p+1}}\left(\frac{1-\hat{\tau}}{2}\frac{1+\hat{\tau}}{2}P_{p-1}^{1,1}(\hat{\tau})\right)\\&=\binom{p+1}{2}\frac{d^2}{d\hat{\tau}^2}\left(\frac{(1-\hat{\tau})(1+\hat{\tau})}{4}\right)\frac{d^{p-1}}{d\hat{\tau}^{p-1}}P_{p-1}^{1,1}(\hat{\tau})\\&=-\frac{\Gamma(2p+1)}{2^{p+1}\Gamma(p)}.\end{aligned}$$

Taking into account the scaling factor $(\tau_{i+1}-\tau_i)^{p+1}/2^{p+1}$, we have

$$\boldsymbol{c}=-(R\boldsymbol{\phi}_p^{(p)})'\frac{\Gamma(p)(\tau_{i+1}-\tau_i)^{p+1}}{\Gamma(2p+1)}. \tag{3.16}$$

Let $\boldsymbol{\epsilon}_{e_i}=\tilde{\boldsymbol{\phi}}_{p+1}-\boldsymbol{\phi}_p$. According to Theorem 3.2, we expect that

$$\|(\mathcal{I}-\mathcal{R})\boldsymbol{\phi}_{hp}^{(p)}\|_{0,e_i}\approx\|\boldsymbol{\epsilon}_{e_i}^{(p)}\|_{0,e_i},$$

if the $(p+1)$-th derivative can be well recovered. In the region near singularity or with rapid changes, we do not expect that the $(p+1)$-th derivative can be well approximated and need to consider the following scaling factor [4]

$$\alpha_{e_i}=\frac{\|(\mathcal{I}-\mathcal{R})\boldsymbol{\phi}_{hp}^{(p)}\|_{0,e_i}}{\|\boldsymbol{\epsilon}_{e_i}^{(p)}\|_{0,e_i}}.$$

We will use α_{e_i} as a regularity indicator in element e_i. Generally speaking, we consider p-refinement if α_{e_i} is close to 1 and h-refinement if α_{e_i} is far from 1.

We also use $\boldsymbol{\epsilon}_{e_i}$ to define a local error indicator as

$$\eta_{e_i}=|\alpha_{e_i}\boldsymbol{\epsilon}_{e_i}|_{1,e_i}\approx|\boldsymbol{\phi}^*-\boldsymbol{\phi}_{hp}|_{1,e_i},\quad i=0,\cdots,N-1. \tag{3.17}$$

Then the total error is

$$\eta = \left(\sum_{i=0}^{N-1} \eta_{e_i}^2 \right)^{1/2}. \tag{3.18}$$

Note that η_{e_i} is just for the approximation of problem (2.17), which is not aware of the error induced by Assumption 2.1.

The following lemma shows that η can also be used to control the error of action functional, where we remove the error of model approximation by considering a fixed T.

Lemma 3.2. *Assume that $\boldsymbol{b}(\boldsymbol{\phi}) \in C^2(\mathbb{R}^n)$, $\boldsymbol{\phi}^* \in C^1(\Gamma_1;\mathbb{R}^n) \cap H^1(\Gamma_1;\mathbb{R}^n)$ and $\boldsymbol{\alpha}^{-1} = \boldsymbol{I}$. For a fixed integration time T, the approximation $\boldsymbol{\phi}_{hp}$ of the (local) minimizer $\boldsymbol{\phi}^*$ satisfies*

$$|S_T(\boldsymbol{\phi}_{hp}) - S_T(\boldsymbol{\phi}^*)| \le C|\boldsymbol{\phi}^* - \boldsymbol{\phi}_{hp}|_1^2 \approx C\eta^2, \tag{3.19}$$

where C is a constant.

Proof. We have the second-order variation around the MAP $\boldsymbol{\phi}^*$:

$$S_T(\boldsymbol{\phi}^* + \delta\boldsymbol{\phi}) - S_T(\boldsymbol{\phi}^*) = \delta S_T(\boldsymbol{\phi}^*, \delta\boldsymbol{\phi}) + \delta^2 S_T(\boldsymbol{\phi}^*, \delta\boldsymbol{\phi}) + \mathcal{O}(|\delta\boldsymbol{\phi}|^3).$$

Due to the fact that $\boldsymbol{\phi}^*$ corresponds to a local minimum of S_T, we have

$$\delta S_T(\boldsymbol{\phi}^*, \delta\boldsymbol{\phi}) = 0 \quad \delta^2 S_T(\boldsymbol{\phi}^*, \delta\boldsymbol{\phi}) > 0, \quad \forall \delta\boldsymbol{\phi} \neq 0.$$

Let $F(\boldsymbol{\phi}, \boldsymbol{\phi}') = \frac{1}{2} T \langle T^{-1}\boldsymbol{\phi}' - \boldsymbol{b}(\boldsymbol{\phi}), T^{-1}\boldsymbol{\phi}' - \boldsymbol{b}(\boldsymbol{\phi}) \rangle$. The second-order variation can be written as

$$\delta^2 S_T = \frac{1}{2} \left[\langle F_{\boldsymbol{\phi}\boldsymbol{\phi}} \delta\boldsymbol{\phi}, \delta\boldsymbol{\phi} \rangle_\tau + 2\langle F_{\boldsymbol{\phi}'\boldsymbol{\phi}} \delta\boldsymbol{\phi}, \delta\boldsymbol{\phi}' \rangle_\tau + \langle F_{\boldsymbol{\phi}'\boldsymbol{\phi}'} \delta\boldsymbol{\phi}', \delta\boldsymbol{\phi}' \rangle_\tau \right],$$

where

$$F_{\phi_i'\phi_j'} = \frac{1}{2T}\delta_{ij}, F_{\phi_i'\phi_j} = -\partial_{\phi_j} b_i, F_{\phi_i\phi_j} = -T\langle \partial_{\phi_i}\boldsymbol{b}, \partial_{\phi_j}\boldsymbol{b} \rangle - T\langle T^{-1}(\boldsymbol{\phi}^*)' - \boldsymbol{b}, \partial_{\phi_i\phi_j}\boldsymbol{b} \rangle.$$

Due to the regularity assumption, $F_{\boldsymbol{\phi}\boldsymbol{\phi}}$, $F_{\boldsymbol{\phi}'\boldsymbol{\phi}}$ and $F_{\boldsymbol{\phi}'\boldsymbol{\phi}'}$ are continuous with respect to $\tau \in [0,1]$. Let $\delta\boldsymbol{\phi} = \boldsymbol{\phi}^* - \boldsymbol{\phi}_{hp}$. We then have

$$\begin{aligned} |S_T(\boldsymbol{\phi}_{hp}) - S_T(\boldsymbol{\phi}^*)| \le{}& \frac{1}{2}|\boldsymbol{\phi}^* - \boldsymbol{\phi}_{hp}|_1^2 \max_{\tau\in[0,1]} \|F_{\boldsymbol{\phi}'\boldsymbol{\phi}'}\| + \frac{1}{2}|\boldsymbol{\phi}^* - \boldsymbol{\phi}_{hp}|_0^2 \max_{\tau\in[0,1]} \|F_{\boldsymbol{\phi}\boldsymbol{\phi}}\| \\ &+ |\boldsymbol{\phi}^* - \boldsymbol{\phi}_{hp}|_1 |\boldsymbol{\phi}^* - \boldsymbol{\phi}_{hp}|_0 \max_{\tau\in[0,1]} \|F_{\boldsymbol{\phi}'\boldsymbol{\phi}}\|. \end{aligned}$$

Using the Poincaré inequality, we obtain Eq. (3.19). ∎

Remark 3.3. To compute $\epsilon_{e_i}^{(p)}$, we also need to compute $\psi_{p+1}^{(p)}$, where the following formula can be used: for $p \geq 2$,

$$\begin{aligned}
\frac{d^p}{d\hat{\tau}^p}\hat{\psi}_{p+1}(\hat{\tau}) &= \frac{d^p}{d\hat{\tau}^p}\left(\frac{1-\hat{\tau}}{2}\frac{1+\hat{\tau}}{2}P_{p-1}^{1,1}(\hat{\tau})\right) \\
&= \binom{p}{1}\left(\frac{1-\hat{\tau}^2}{4}\right)'\frac{d^{p-1}}{d\hat{\tau}^{p-1}}P_{p-1}^{1,1}(\hat{\tau}) \\
&\quad + \binom{p}{2}\left(\frac{1-\hat{\tau}^2}{4}\right)''\frac{d^{p-2}}{d\hat{\tau}^{p-2}}P_{p-1}^{1,1}(\hat{\tau}) \\
&= -\frac{p\Gamma(2p+1)}{2^p\Gamma(p+2)}\hat{\tau} - \binom{p}{2}\frac{\Gamma(2p)}{2^{p-1}\Gamma(p+2)}P_1^{p-1,p-1}(\hat{\tau}) \\
&= -\frac{\Gamma(2p+1)}{2^{p+1}\Gamma(p)}\hat{\tau},
\end{aligned}$$

where in the last step we use the fact that

$$P_1^{p-1,p-1}(\hat{\tau}) = p\hat{\tau}.$$

Remark 3.4. When dealing with elements of different polynomial orders, we first project the solution elementwisely to $\text{span}\{\psi_i\}_{i=0}^{p}$ and then apply the derivative recovery technique to obtain η_{e_i} for the elements of polynomial order p. The projection is necessary because ψ_i are not orthogonal to each other. After p is looped from the lowest order to the highest one, η_{e_i} will be obtained for all elements.

3.3 Criteria for *hp*-adaptivity

After removing the global reparametrization, the adaptive tMAM shares the same scenario of adaptive finite element method:

$$\text{Solve} \to \text{Estimate} \to \text{Mark} \to \text{Refine}.$$

There exist many strategies to mark the elements that need refinement. We consider two strategies in this paper. The first strategy is based on [4], where a target number of degrees of freedom (DOFs) is specified for the new mesh; The second strategy is the Dölfler's strategy, where the elements that have the largest error will be refined.

3.3.1 Strategy one

Let p_{ave} be the average polynomial order. Before enriching the approximation space through *hp*-adaptivity, we can set up a target number of DOFs as

$$M_{hp} = M_{\text{old}}\left(\frac{1+p_{\text{ave}}}{p_{\text{ave}}}\right), \tag{3.20}$$

where M_{old} is the number of DOFs before adaptive refinement. It is seen that if $p_{\text{ave}}=1$, M_{new} matches the number of DOFs after a uniform h-refinement.

Let α_{ave} be the average of α_{e_i}. If a certain element needs to be refined, we choose either h- or p-refinement according to the following criterion:

$$Q_{hp}(e_i)=\begin{cases} 1, & \text{if } \alpha_{e_i}<2\min\{\alpha_{\text{ave}},\alpha_Q\} \quad (p\text{-refinement}),\\ 0, & \text{otherwise} \qquad\qquad\quad (h\text{-refinement}). \end{cases} \tag{3.21}$$

Note here that we define a threshold α_Q in addition to the average α_{ave}. The main motivation is that α can be large in most of the elements when the mesh is coarse. Then α_Q can help to trigger h-refinement instead of p-refinement. When the mesh is fine enough, α_{ave} will be close to one and α_Q will be not necessary any more.

Following is the first strategy of hp-adaptivity, where we first address the error of numerical approximation and then deal with the discrepancy between the optimal linear time scaling and the zero-Hamiltonian constraint:

(1) Refine e_i with the largest η_{e_i} according to $Q_{hp}(e_i)$ if $M<M_{hp}$, where M is the current number of DOFs. Here we may also consider multi-level h-refinement.

(2) If $\theta_{\max}/\theta_{\min}>\beta_\theta$ and $(M-M_{hp})<r_M(M_{hp}-M_{\text{old}})$, refine e_i with the largest θ_{e_i} by h-refinement, where $r_M>0$ is a prescribed constant. Here we only consider one-level h-refinement.

First of all, we use M_{hp} as a target number of DOFs for hp-adaptivity based on η_{e_i}. If θ_{e_i} varies significantly, we will trigger h-refinement to deal with the deviation from the zero-Hamiltonian constraint. Since the error of path approximation is usually dominant, we will only add a small number of additional DOFs for model approximation, in other words, we choose r_M to be biased to 0. Second, whenever h-refinement is implemented, the element is decomposed into two equidistant ones. Local indicators η_{e_i} and θ_{e_i} can be computed in child elements from the information that is inherited from the parent element. The parameter α_{e_i} will be passed directly from the parent element to child elements. The child elements can also be considered for h-refinement. This way, a multi-level h-refinement is allowed. Third, when p-refinement is implemented, the polynomial order will be increased by one from p to $p+1$ if a prescribed maximum polynomial order is not reached. Due to the lack of information of derivative of order $p+2$, we are not able to estimate the local error after p-refinement. Then elements after p-refinement will be not allowed to be refined any further.

Remark 3.5. Although the multi-level h-refinement works well if the starting mesh is relatively fine, we will focus on the case that the starting mesh has only a few time elements, e.g., two or three elements, and will not consider multi-level h-refinement in our numerical experiments.

Algorithm 1 hp-adaptive tMAM using strategy one

while $k \leq K_{\max}$ **or** $\epsilon > \epsilon_{\text{tol}}$ **do**
  Compute θ_{e_i}, η_{e_i}, α_{e_i} and M_{hp}.
  Build up a max-heap H_η of elements according to η_{e_i}.
  while $M < M_{\text{hp}}$ **do**
    Consider the element with largest η_{e_i}.
    if $Q_{hp}(e_i)$ **then**
      Do p-refinement.
      Set the local error $\eta_{e_i} = 0$.
    else
      Do h-refinement.
      if multi-level h-refinement is allowed **then**
        Update errors for new elements and add them to the heap H_η.
      else
        Set the local error $\eta_{e_i} = 0$ for child elements.
      end if
    end if
    Update M.
  end while
  if $\theta_{\max}/\theta_{\min} > \beta_\theta$ **then**
    while $M - M_{hp} < r_M(M_{hp} - M_{\text{old}})$ **do**
      Do h-refinement for the element with largest θ_{e_i}.
      Set the local indicator $\theta_{e_i} = 0$ for child elements.
    end while
  end if
  Solve problem (2.17) using new partition $\mathcal{T}_h^{k+1}$ to obtain MAP $\boldsymbol{\phi}_h^{*,k+1}$.
  $\epsilon \leftarrow \left(S_{T(\boldsymbol{\phi}_h^{*,k})}(\boldsymbol{\phi}_h^{*,k}) - S_{T(\boldsymbol{\phi}_h^{*,k+1})}(\boldsymbol{\phi}_h^{*,k+1}) \right) / S_{T(\boldsymbol{\phi}_h^{*,k+1})}(\boldsymbol{\phi}_h^{*,k+1})$.
  $k \leftarrow k+1$
end while

3.3.2 Strategy two

For the second strategy, we do not specify a target number of degrees of freedom for the hp-adaptivity based on η_{e_i}. Instead, we want to make sure that the error of path approximation will be reduced by a certain percentage. To do this, we consider Dölfler's strategy. Let $J = \{i | 0 \leq i \leq N-1\}$ be the index of all elements. We look for a subset $\hat{J} \subset J$ such that for $r_\eta \in (0,1]$,

$$r_\eta \sum_{i \in J} \eta_{e_i}^2 \leq \sum_{i \in \hat{J}} \eta_{e_i}^2. \tag{3.22}$$

To uniquely specify $\hat{J}$, we choose the elements that have the largest error, i.e.,

$$\min_{i\in\hat{J}}\eta_{e_i}\geq\max_{i\in J\setminus\hat{J}}\eta_{e_i}.$$

Once a certain element is marked, $Q_{hp}(e_i)$ is used as usual to make a decision for either h-refinement or p-refinement. We can also consider multi-level h-refinement in this strategy.

For the updated mesh, we will determine whether the refinement for model approximation is necessary by checking the ratio $\theta_{\max}/\theta_{\min}$. Whenever necessary, we employ the same criterion as in strategy one, i.e., we refine e_i with the largest θ_{e_i} if $M-M_{hp}<r_M(M_{hp}-M_{\text{old}})$, where M_{hp} is not a precomputed number any more but the number of DOFs of the updated mesh by hp-adaptivity. We will not consider multi-level h-refinement for θ_{e_i}.

3.3.3 Some implementation issues

First, we choose the integer $m=1$ or 2 for the smoothing operator $\mathcal{R}$. Theoretically, m should grow in a logarithmic-like fashion. The choice $m\leq 2$ is suggested in [2,4] based on empirical investigations. Numerical experiments show that this choice is also sufficient for our cases.

Second, the parameter λ in $\mathcal{S}_h$ will never be computed explicitly. Instead of we use m Jacobi-conjugate gradient iterations to solve the linear system $\boldsymbol{A}\boldsymbol{x}=0$ [2], where $\boldsymbol{A}$ is the stiffness matrix induced by $\mathcal{A}_h$ with an initial guess $\mathcal{Q}_h\boldsymbol{\phi}_{hp}^{(p)}$. Since m is small, the cost is $\mathcal{O}(pnN_e)$.

Third, two types of L^2 projection are involved for adaptivity: the projection induced by $\mathcal{Q}_h$ and the path projection from the old mesh to the new one. For the former case, we need to invert a N_e tridiagonal matrix pn times with a cost of $\mathcal{O}(pnN_e)$. For the latter case, we need to invert the local mass matrix for a certain number of elements with a cost of $\mathcal{O}(p^2N_e)$.

Overall, the cost for adaptivity is $\mathcal{O}(pnN_e)$ since the dimension n is usually comparable to or larger than the polynomial order p. Note that the total number of DOFs is $\mathcal{O}(pnN_e)$. So, the total cost for adaptivity is linear with respect to the total number of degrees of freedom, which is small compared to the cost for the optimization iterations.

4 Numerical experiments

In this section, we implement some numerical experiments to examine our hp adaptivity strategies. The optimization is based on the nonlinear conjugate gradient (CG) method [16,19].

Algorithm 2 hp-adaptive tMAM using strategy two

while $k \le K_{\max}$ **or** $\epsilon > \epsilon_{\text{tol}}$ **do**
  Compute θ_{e_i}, η_{e_i} and α_{e_i}.
  Build up a max-heap H_η of elements according to η_{e_i}.
  while $r_\eta \sum_{i\in J}\eta_{e_i}^2 > \sum_{i\in\hat{J}}\eta_{e_i}^2$ **do**
    Consider the element with largest η_{e_i}.
    if $Q_{hp}(e_i)$ **then**
      Do p-refinement.
      Set the local error $\eta_{e_i}=0$.
    else
      Do h-refinement.
      if multi-level h-refinement is allowed **then**
        Update errors for new elements and add them to the heap H_η.
      else
        Set the local error $\eta_{e_i}=0$ for child elements.
      end if
    end if
  end while
  if $\theta_{\max}/\theta_{\min} > \beta_\theta$ **then**
    Compute M_{hp} of the updated mesh.
    while $M - M_{hp} < r_M(M_{hp} - M_{\text{old}})$ **do**
      Do h-refinement for the element with largest θ_{e_i}.
      Set the local indicator $\theta_{e_i}=0$ for child elements.
    end while
  end if
  Solve problem (2.17) using new partition $\mathcal{T}_h^{k+1}$ to obtain MAP $\boldsymbol{\phi}_h^{*,k+1}$.
  $\epsilon \leftarrow \left(S_{T(\boldsymbol{\phi}_h^{*,k})}(\boldsymbol{\phi}_h^{*,k}) - S_{T(\boldsymbol{\phi}_h^{*,k+1})}(\boldsymbol{\phi}_h^{*,k+1})\right) / S_{T(\boldsymbol{\phi}_h^{*,k+1})}(\boldsymbol{\phi}_h^{*,k+1})$.
  $k \leftarrow k+1$
end while

4.1 A simple linear SODE system

We use the following linear SODE system to test some essential numerical issues:

$$d\boldsymbol{X}(t) = \boldsymbol{A}\boldsymbol{X}(t)dt + \sqrt{\varepsilon}d\boldsymbol{W}(t), \tag{4.1}$$

where $\boldsymbol{A}\in\mathbb{R}^{2\times 2}$ has real and negative eigenvalues such that $\boldsymbol{a}=(0,0)^{\mathsf{T}}$ is a stable fixed point. For simplicity, we choose $\boldsymbol{A}=\text{diag}\{\lambda_1,\lambda_2\}$ as a diagonal matrix, where $\lambda_i<0$, and $|\lambda_1|\ge|\lambda_2|$. For any give point $\boldsymbol{X}(0)=\boldsymbol{x}$ in phase space, there exist a trajectory $\boldsymbol{X}(t)=e^{tA}\boldsymbol{x}$ converging to $\boldsymbol{a}$ when $\varepsilon=0$. If we consider such a trajectory as a transition path from $\boldsymbol{x}$ to $\boldsymbol{a}$, we have $V(\boldsymbol{x},\boldsymbol{a})=0$ and the associated optimal integral time $T^*=\infty$. We then use the trajectory from $\boldsymbol{x}$ to $\boldsymbol{a}$ as a reference solution to test the effectiveness of our hp adaptivity

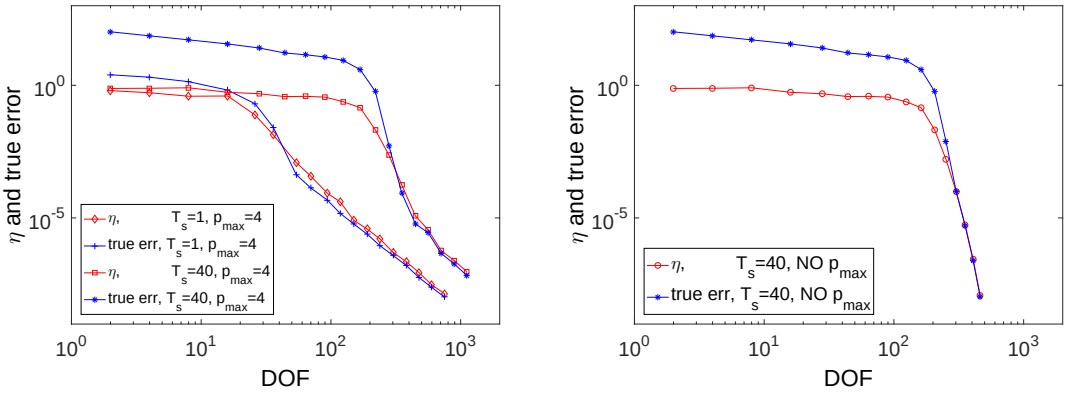

Figure 1: DOF versus η_{e_i} and true error for the linear SODE. All adaptive meshes start from two linear time elements. Left: hp- refinement with a constraint that $p\leq 4$; Right: hp-adaptivity with no constraint on p.

strategies.

We first verify the effectiveness of the error indicator η_{e_i}. To do it, we consider a finite integration time, i.e., we consider the transition from $\boldsymbol{x}$ to $e^{t\boldsymbol{A}}\boldsymbol{x}$ for a fixed t. To mimic the scale separation of fast and slow dynamics, we choose t large enough such that $|e^{t\boldsymbol{A}}\boldsymbol{x}|\ll 1$. Note that in the definition of the smoother operator S_h, each component of $\boldsymbol{\phi}_h$ is dealt with independently using the same formula. We then choose $|\lambda_1|\gg|\lambda_2|$ to test the effectiveness of η_{e_i}. We construct the matrix $\boldsymbol{A}$ as

$$\boldsymbol{A}=\boldsymbol{B}^{-1}\boldsymbol{J}\boldsymbol{B}=\begin{bmatrix} a & -b \\ b & a \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}\begin{bmatrix} a & b \\ -b & a \end{bmatrix},$$

where $a,b\in\mathbb{R}$ and $a^2+b^2=1$. We have

$$e^{t\boldsymbol{A}}=\boldsymbol{B}^{-1}\mathrm{diag}\{e^{\lambda_1 t},e^{\lambda_2 t}\}\boldsymbol{B}.$$

Let $\lambda_1=-10$, $\lambda_2=-0.01$, and $a=1/3$. We choose the starting point $\boldsymbol{x}=(1,1)$ and the ending point $e^{T_s\boldsymbol{A}}\boldsymbol{x}$. The larger T_s is, the closer $e^{T_s\boldsymbol{A}}\boldsymbol{x}$ is to the fixed point $\boldsymbol{a}$. We consider hp-adaptivity for numerical approximation starting from two linear time elements. Because we use a finite T_s, we switch off the h-adaptivity for model approximation.

In Fig. 1, we plot η and the true error of adaptive meshes with a constraint $p\leq 4$ on the left, and without a constraint on the polynomial order on the right. It is seen that there exists a large discrepancy between η and the true error when the mesh is coarse and η captures the true error well once the number of DOFs is large enough. If there is no constraint on the polynomial order, we observe p-convergence with respect to DOF; if the polynomial order is limited by $p_{\max}$, we observe a transition from p-convergence to h-convergence with an optimal rate $\mathcal{O}(N^{-p_{\max}})$ [22].

4.2 A non-gradient SODE system

We consider the following Maier-Stein model [13]:

$$\begin{cases} dX = (X - X^3 - \beta XY^2)dt + \sqrt{\varepsilon}dW^x, \\ dY = -(1+X^2)Ydt + \sqrt{\varepsilon}dW^y, \end{cases} \tag{4.2}$$

where W^x and W^y are independent Wiener processes and $\beta > 0$ is a parameter. This is a non-gradient system except when $\beta = 1$. This system has two stable fixed points: $\boldsymbol{a}_1 = (-1,0)^\mathsf{T}$ and $\boldsymbol{a}_2 = (1,0)^\mathsf{T}$, and a saddle point $\boldsymbol{a}_3 = (0,0)^\mathsf{T}$. For numerical experiments, we set $\beta = 10$ such that the system is not gradient.

We then consider the transition from $\boldsymbol{a}_1$ to $\boldsymbol{a}_2$. The MAP passes the saddle point $\boldsymbol{a}_3$, but is not the heteroclinic orbit connecting $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ due to the fact that the system is non-gradient. In Fig. 2, we plot the convergence behavior of h- and hp-refinement given by strategy one and two. For both strategies, we set $r_M = 0.1$. In other words, once the adaptivity is done for the numerical approximation, we seek 10% extra DOFs to deal with model approximation. We set $\alpha_Q = 4.0$, $\beta_\theta = 10.0$ and $r_\eta = 0.9$. It is seen that both strategy one and two show similar convergence behavior. The optimal convergence rate $\mathcal{O}(N^{-2p})$ is obtained for h-refinement in terms of the error of quasi-potential [22]. We also observe p-convergence for the hp-refinement. For strategy two, the ratio r_η controls the increment of DOFs. We have tested several $r_\eta \in [0.5, 0.9]$, which result in similar convergence behavior. The main difference is that more intermediate meshes will be generated if r_η is smaller. Fig. 2 shows that the singularity induced by $T^* = \infty$ can be effectively dealt with by hp-adaptivity.

We next look at the distribution of element size and polynomial order of adaptive meshes. In Fig. 3, we plot the distribution of element size of a final h-adaptive mesh with respect to time and arc length respectively. The elements are split into two groups: the first group indicated by blue color corresponds to the MAP from $\boldsymbol{a}_1$ to $\boldsymbol{a}_3$ and the

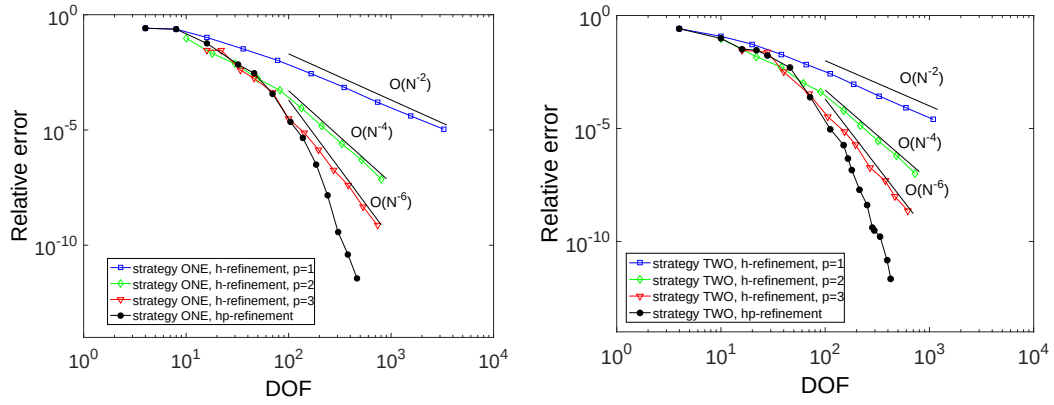

Figure 2: Convergence behavior of adaptive meshes in terms of the error of action functional. All adaptive meshes start from three equidistant linear elements. Left: strategy one; Right: strategy two.

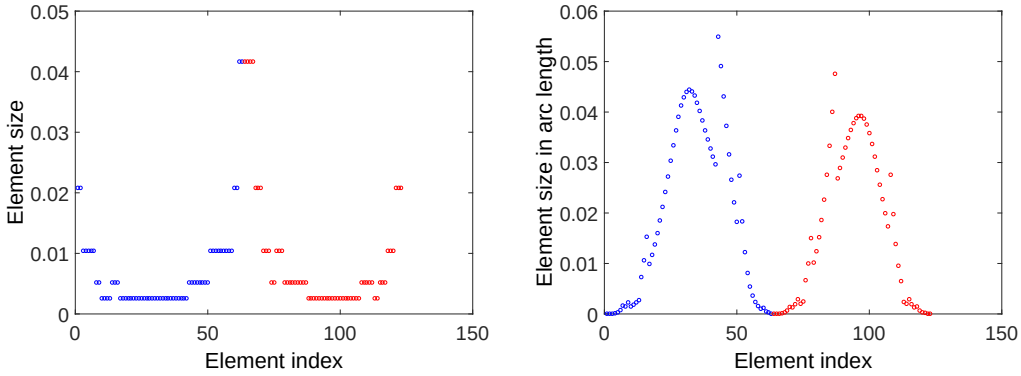

Figure 3: Distribution of element size with respect to time (left plot) or arc length (right plot). This is the final mesh in the left plot of Fig. 2 given by h-refinement with $p=3$. Blue color indicates the MAP from $\boldsymbol{a}_1$ to $\boldsymbol{a}_3$ and red color indicates the MAP from $\boldsymbol{a}_3$ to $\boldsymbol{a}_2$.

second group indicated by red color corresponds to the MAP from $\boldsymbol{a}_3$ to $\boldsymbol{a}_2$. In the region of slow dynamics, i.e., around the critical points, the element size is relatively larger with respect to time. However, with respect to arc length, the element size is relatively larger in the region of fast dynamics. This is consistent with the observation in [19] that a time mesh that is more uniform with respect to arc length may not result in a better accuracy in aMAM. We refer to [28] for more studies about this phenomena. The existence of outliers in the right plot is simply because only a portion of elements in the region of fast dynamics will trigger the adaptivity criterion. To illustrate this, we compare, in Fig. 4, the distribution of element size with respect to arc length from four consecutive h-adaptive meshes. In Fig. 5, we plot the distribution of element size and polynomial order for a final hp-adaptive mesh. It is seen that the distribution of element size has the same trend as that of h-adaptive meshes. The polynomial order is between 5 and 9, indicating that high-order approximation is efficient. It appears that in the region of slow dynamics a higher polynomial order is preferred. This is not surprising. We look at the first component $x(\tau)$ of the transition path. $x(\tau)$ looks like a step function jumping from -1 to 0, and then to 1. The flat part corresponds to the region of slow dynamics and the part of sharp transition corresponds to the region of fast dynamics. Thus it is more easier for the regularity indicator to trigger p-refinement in the region of slow dynamics and h-refinement in the region of fast dynamics.

We now switch off the adaptivity for the model approximation. The convergence behavior of adaptive meshes is shown in Fig. 6 for strategy one. It is seen that the refinement for model approximation does not affect h-adaptivity much, but it has a major impact on hp-adaptivity. Without the adaptivity for model approximation, the error of hp-adaptive meshes decays slowly in the early stage and plummets suddenly at a certain point. Starting from three elements, η_{e_i} mainly triggered h-refinement of the middle element and p-refinement for the first and last elements. Without the h-refinement from θ_{e_i}, the polynomial order in the first and last elements keeps increasing until the specified

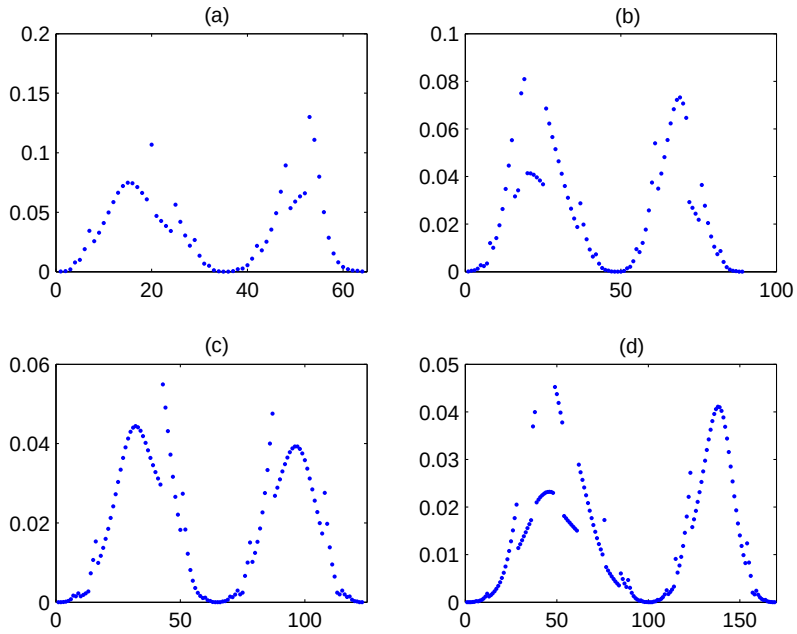

Figure 4: Evolution of element size with respect to arc length. The x-axis is element index and the y-axis is the element size in arc length. The meshes are based on h-adaptivity with a fixed polynomial order 3. The meshes are refined from (a) to (d). The plot (c) corresponds to the right plot of Fig. 3.

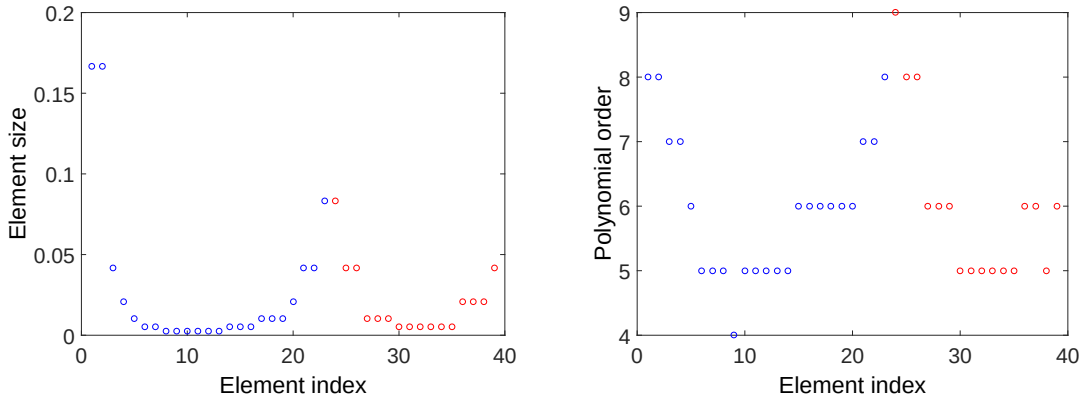

Figure 5: Distributions of element size with respect to time (left plot) and polynomial order (right plot). This is the final mesh in the left plot of Fig. 2 given by hp-refinement. Blue color indicates the MAP from $\boldsymbol{a}_1$ to $\boldsymbol{a}_3$ and red color indicates the MAP from $\boldsymbol{a}_3$ to $\boldsymbol{a}_2$.

maximum polynomial order 15 is reached. Once h-refinement is triggered in the first and last elements, the error drops quickly. This experiment demonstrates that h-refinement for the model approximation is necessary for robustness.

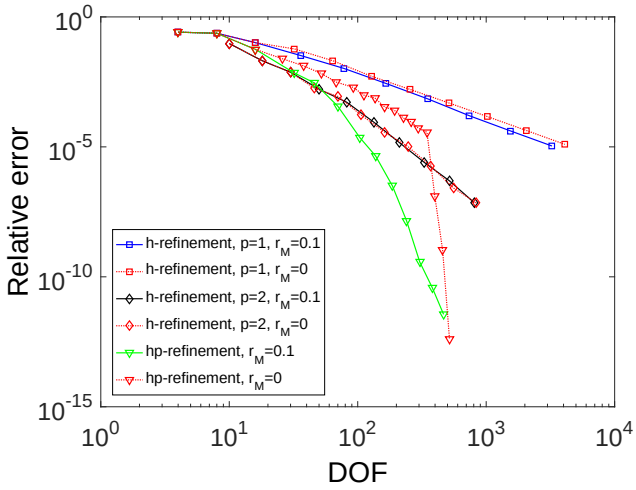

Figure 6: Convergence behavior of adaptive meshes with and without the h-adaptivity for model approximation in terms of the error of action functional, where all adaptive meshes are based on strategy one.

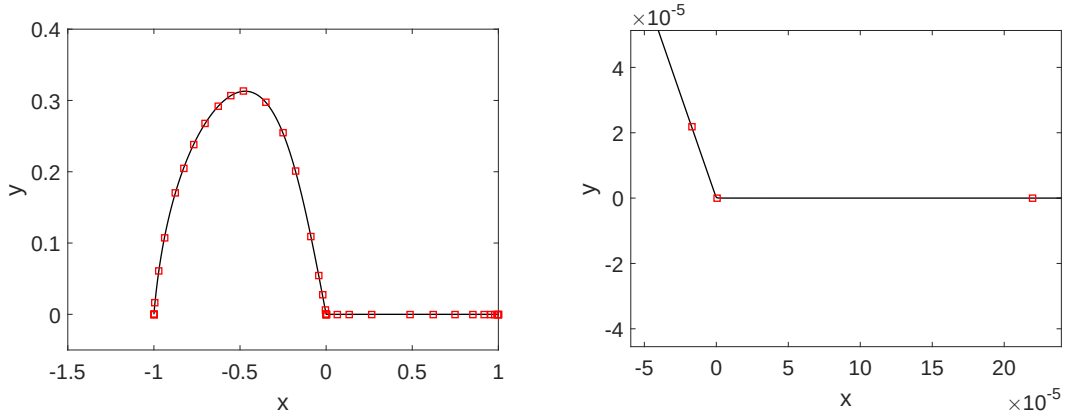

Figure 7: Behavior of the approximate MAP around the saddle point $\boldsymbol{a}_3$. This is the final mesh in the left plot of Fig. 2 given by hp-refinement. Left: The whole MAP with squares corresponding to the finite element discretization of $\Gamma_1=[0,1]$; Right: Close-up view around the saddle point $\boldsymbol{a}_3=(0,0)$.

Finally, we look at the behavior of the approximate MAP around the saddle point $\boldsymbol{a}_3$. It is known that in aMAM [26] there exists a tangling phenomena meaning that a lot of grid points are in a tangle around fixed points especially for a large integration time T. These tangled grid points do not contribute to the path approximation. Such a tangling phenomena is removed in [28] by an improved moving mesh technique coupling with WENO interpolation. We plot the approximate MAP given by the hp-adaptive mesh in Fig. 7 including a close-up view around the saddle point $\boldsymbol{a}_3$. It is seen that the abrupt turn at the saddle point $\boldsymbol{a}_3$ is well captured and no tangling phenomena is observed.

4.3 A non-gradient SPDE system

The algorithm developed can also be applied to study small random perturbations of spatially extended systems [7, 10, 11]. We subsequently consider the following infinite dimensional analogue of the Maier-Stein model [13]:

$$\begin{cases} \partial_t u = \kappa \partial_{xx} u + u - u^3 - \beta u v^2 + \sqrt{\varepsilon}\dot{W}_u(x,t), \\ \partial_t v = \kappa \partial_{xx} v - (1+u^2)v + \sqrt{\varepsilon}\dot{W}_v(x,t), \end{cases} \tag{4.3}$$

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. 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_+). Compared to the finite-dimensional Maier-Stein model, we have an additional parameter κ. When $\kappa < \kappa_{\text{crit}} = 1/(4\pi^2) \approx 0.0253$, there exist non-constant saddle points of the form $(u(x), 0)$. More details about the profile of $u(x)$ can be found in [13]. We here consider the case $\beta = 10$ and $\kappa = 0.01$ for numerical experiments.

Consider Eq. (4.3) is a spatially extended system, we consider a modified action functional. Let $\boldsymbol{\phi} = [u, v]^{\mathsf{T}} \in \mathbb{R}^2$. The deterministic dynamical system, i.e., $\varepsilon = 0$, is

$$\partial_t \boldsymbol{\phi} = \kappa \partial_{xx} \boldsymbol{\phi} + \boldsymbol{f}(\boldsymbol{\phi}), \tag{4.4}$$

where $\boldsymbol{f}(\boldsymbol{\phi})$ indicates the nonlinear terms in Eq. (4.3). To solve Eq. (4.4), we need spatial discretization. If we consider periodic boundary conditions, we often use the Fourier expansion of $\boldsymbol{\phi}$, say $\boldsymbol{\phi}_h$. For numerical approximation, we need to choose a procedure, e.g., Galerkin projection, to obtain a closed system of ordinary differential equations for the Fourier coefficients of $\boldsymbol{\phi}_h$. Let $\mathcal{P}_h$ indicates such a procedure and $\mathcal{P}_h \boldsymbol{f}$ the projection of $\boldsymbol{f}(\boldsymbol{\phi}_h)$ onto the truncated Fourier space. Since the dynamic solver solves an approximated system

$$\partial_t \boldsymbol{\phi}_h = \kappa \partial_{xx} \boldsymbol{\phi}_h + \mathcal{P}_h \boldsymbol{f}(\boldsymbol{\phi}_h), \tag{4.5}$$

it is more appropriate to consider the action functional induced by Eq. (4.5) instead of Eq. (4.4). In other words, we modify the action functional slightly as

$$S_T(\boldsymbol{\phi}_h) = \frac{1}{2}\int_0^T \|\partial_t \boldsymbol{\phi}_h - \partial_{xx}\boldsymbol{\phi}_h - \mathcal{P}_h \boldsymbol{f}(\boldsymbol{\phi}_h)\|_2^2 dt, \tag{4.6}$$

where $\|\cdot\|_2$ indicates L^2 norm in physical space. Note that $\boldsymbol{f}(\boldsymbol{\phi}_h)$ is replaced by $\mathcal{P}_h \boldsymbol{f}(\boldsymbol{\phi}_h)$, and such a replacement is often necessary from the numerical point of view [24]. In this work we use the discrete Fourier transform (DFT) to approximate $\mathcal{P}_h \boldsymbol{f}$. Depending on the nonlinearity of $\boldsymbol{f}$, the number of grid points for DFT can be chosen to reduce or remove the aliasing error.

In Fig. 8, we plot the convergence behavior of adaptive meshes given by strategies one and two. It is observed that the optimal convergence rate of h-convergence has been recovered and the hp-refinement provides fast (exponential) convergence.

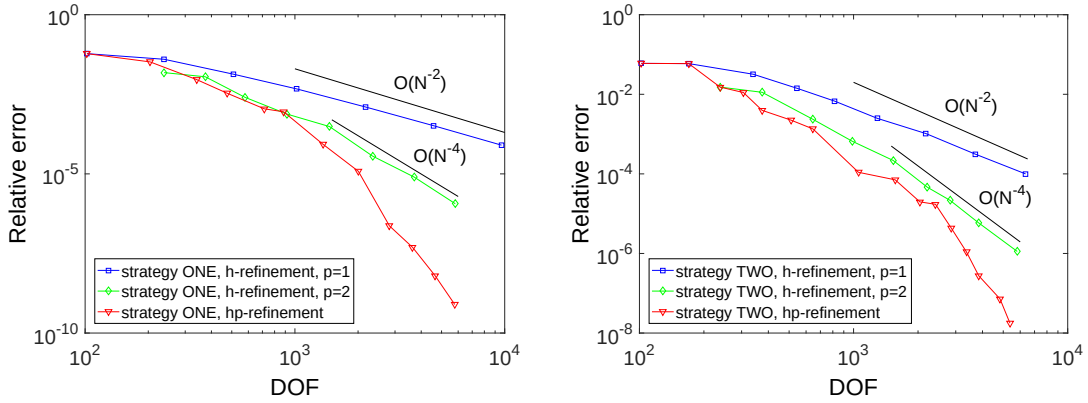

Figure 8: Convergence behavior of adaptive meshes in terms of the error of action functional. All adaptive meshes start from four equidistant time elements. Left: strategy one; Right: strategy two.

5 Conclusion and discussions

In this work, we have developed new criteria for the *hp*-adaptivity of the minimum action method using a posteriori error estimation techniques. The final adaptive MAM is a general algorithm for both SODEs and SPDEs. To alleviate the singularity $T^*=\infty$, we use the discrete problem (2.17) to generate a minimizing sequence, where the optimal linear time scaling is enforced. We need to address two types of numerical errors, which are from model approximation and path approximation, respectively. We used the Hamiltonian to define an error indicator for the model approximation. Using the derivative recovery technique, we constructed an error indicator and an regularity indicator for a finite element discretization of transition paths. Based on the proposed error and regularity indicators, strategies have been developed for *hp*-adaptivity. It has been demonstrated by numerical experiments that the approximation errors can be effectively dealt with by *h*- or *hp*-adaptivity, which means that global reparametrization is not necessary. For the first time we have addressed the adaptivity of minimum action method based on indicators that are consistent with the approximation theory. The numerical results are very promising. Considering that the minimum action method is computationally demanding for spatially extended systems and the theory of a posteriori error estimate of finite element methods has been well developed, it is a very interesting problem to seek more effective adaptivity strategies for the minimum action method when applied to large scale problems, such as Navier-Stokes equations [23].

Acknowledgments

X. Wan was partially supported by NSF Grant DMS-1620026 and AFOSR Grant FA9550-15-1-0051. B. Zheng and G. Lin would like to thank the support by the U.S. Department

of Energy, Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program as part of the Collaboratory on Mathematics for Mesoscopic Modeling of Materials. PNNL is operated by Battelle for the DOE under Contract DE-AC05-76RL01830. G. Lin would also like to thank the support of the DOE Multifaceted Mathematics for Complex Energy Systems (M^2ACS) project and NSF Grant DMS-1115887.

References

[1] V. I. Arnold, *Mathematical Methods of Classical Mechanics*, 2nd Edition, Springer, 1989.
[2] R. E. Bank and J. Xu, *Asymptotically exact a posteriori error estimators, Part II: general unstructured grids*, SIAM J. Numer. Anal., **41(6)** (2003), 2313–2332.
[3] R. Bank, J. Xu, and B. Zheng, *Superconvergent derivative recovery for Lagrange triangular elements of degree p on unstructured grids*, SIAM J. Numer. Anal. **45(5)** (2007), 2032–2046.
[4] R. Bank and H. Nguyen, *hp adaptive finite elements based on derivative recovery and superconvergence*, Comput. Visual Sci., **14(6)** (2012), 287–299.
[5] J. A. Bucklew, *Introduction to Rare Event Simulation*, Springer, 2014.
[6] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, SIAM, 2002.
[7] G. Da Prato and J. Zabczyk, *Stochastic Equations in Infinite Dimensions*, Encyclopedia Math. Appl. 44, Cambridge University Press, Cambridge, UK, 1992.
[8] W. E, W. Ren and E. Vanden-Eijnden, *String method for the study of rare events*, Phys. Rev. B, **66** (2002), 052301.
[9] 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.
[10] W. Faris and G. Jona-Lasinio, *Large fluctuations for a nonlinear hear equation with noise*, J. Phys. A: Math. Gen. **15** (1982), 3025-3055.
[11] M. Freidlin, *Random perturbations of reaction-diffusion equations: the quasideterministic approximation*, Trans. Amer. Math. Soc., **305** (1988), 665–697.
[12] M. Freidlin and A. Wentzell, *Random Perturbations of Dynamical Systems*, second ed., Springer-Verlag, New York, 1998.
[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. Karniadakis and S. Sherwin, *Spectral/hp Element Methods for Computational Fluid Dynamics*, Oxford University Press, 2nd Edition, 2005.
[15] B. S. Lindley and I. B. Schwartz, *An iterative action minimizing method for computing optimal paths in stochastic dynamical systems*, Physica D, **255** (2013), 22-30.
[16] J. Nocedal and S. Wright, *Numerical Optimization*, Springer Series in Operations Research, Springer, New York, 1999.
[17] E. Vanden-Eijnden and J. Weare, *Rare event simulation of small noise diffusions*, Commun. Pure Appl. Math., **65** (2012), 1770-1803.
[18] 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.
[19] X. Wan, *An adaptive high-order minimum action method*, J. Compt. Phys., **230** (2011), 8669-8682.
[20] X. Wan, *A minimum action method for small random perturbations of two-dimensional parallel shear flows*, J. Compt. Phys., **235** (2013), 497–514.

[21] X. Wan and G. Lin, *Hybrid parallel computing of minimum action method*, Parallel Computing, **39** (2013), 638–651.
[22] X. Wan, *A minimum action method with optimal linear time scaling*, Comm. Compt. Phys., **18(5)** (2015), 1352–1379.
[23] 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, **28** (2015), 1409–1440.
[24] X. Wan and H. Yu, *A dynamic-solver-consistent minimum action method: With an application to 2D Navier-Stokes equations*, J. Compt. Phys., **331(15)** (2017), 209–226.
[25] X. Wan, H. Yu and J. Zhai, *Convergence analysis of finite element approximation of large deviation principle*, preprint.
[26] X. Zhou, W. Ren and W. E, *Adaptive minimum action method for the study of rare events*, J. Chem. Phys., **128** (2008), 104111.
[27] 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.
[28] Y. Sun and X. Zhou, *An improved adaptive minimum action method for the calculation of transition path in non-gradient systems*, submitted.