Japan Journal of Industrial and Applied Mathematics
https://doi.org/10.1007/s13160-023-00604-8

ORIGINAL PAPER

A model-free shrinking-dimer saddle dynamics for finding saddle point and solution landscape

Lei Zhang[1] · **Pingwen Zhang**[2,3] · **Xiangcheng Zheng**[4]

Received: 22 December 2022 / Revised: 18 June 2023 / Accepted: 3 July 2023

Abstract
We propose a model-free shrinking-dimer saddle dynamics for finding any-index saddle points and constructing the solution landscapes, in which the force in the standard saddle dynamics is replaced by a surrogate model trained by the Gassian process learning. By this means, the exact form of the model is no longer necessary such that the saddle dynamics could be implemented based only on some observations of the force. This data-driven approach not only avoids the modeling procedure that could be difficult or inaccurate, but also significantly reduces the number of queries of the force that may be expensive or time-consuming. We accordingly develop a sequential learning saddle dynamics algorithm to perform a sequence of local saddle dynamics, in which the queries of the training samples and the update or retraining of the surrogate force are performed online and around the latent trajectory in order to improve the accuracy of the surrogate model and the value of each sampling. Numerical experiments are performed to demonstrate the effectiveness and efficiency of the proposed algorithm.

Keywords Model-free · Saddle point · Saddle dynamics · Solution landscape · Gaussian process · Surrogate model

✉ Lei Zhang
zhangl@math.pku.edu.cn

Pingwen Zhang
pzhang@pku.edu.cn

Xiangcheng Zheng
xzheng@sdu.edu.cn

[1] Beijing International Center for Mathematical Research, Center for Machine Learning Research, Center for Quantitative Biology, Peking University, Beijing 100871, China

[2] School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China

[3] School of Mathematical Sciences, Laboratory of Mathematics and Applied Mathematics, Peking University, Beijing 100871, China

[4] School of Mathematics, Shandong University, Jinan 250100, China

Published online: 18 July 2023

Mathematics Subject Classification 37M05 · 60G15 · 65D15 · 62L05

1 Introduction

Finding saddle points on complex energy landscapes or dynamical systems provides substantial physical and chemical properties and is thus an important problem in various fields such as nucleation and phase transformations in solid and soft matter [49, 50] and transition rates in chemical reactions and computational biology [31, 40, 42]. The saddle points could be classified by the Morse index characterized by the maximal dimension of a subspace on which the Hessian $H(x)$ is negative definite [29]. In practice, the index-1 saddle point represents the transition state connecting two local minima according to the transition state theory [28, 52]. The index-2 saddle points are particularly useful for providing trajectory information of chemical reactions in chemical systems [19]. The excited states in quantum systems can also be characterized as saddle point configurations [1].

There exist several algorithms of finding saddle points with low indexes [12, 14, 15, 25, 41]. In general, the computation of high-index (index>1) saddle point is more difficult as it has multiple unstable eigen-directions and receives less attention, though the number of high-index saddle points are much larger than the number of local minima and index-1 saddle points on the energy landscape [3, 22]. High-index shrinking-dimer saddle dynamics proposed in [44], which involves the force calculation (i.e. the negative gradient of the energy function), provides an effective means for finding any-index saddle points. This method is then combined with the downward or upward search algorithms [43, 45] to construct the solution landscapes of both energy systems and dynamical (non-gradient) systems [17, 18, 27, 36], which also provides a systemic approach to find the largest possible number of saddle points.

However, it is possible that the exact form of the force in high-index saddle dynamics is not given a priori in some cases such that we need to either investigate the modeling of the underlying processes or perform experiments in order to obtain the inquired values of the force in saddle dynamics. In practice, modeling complex problems may be difficult or inaccurate, which often restricts the computation of the saddle points and its applications.

To resolve this issue, we employ a data-driven approach to propose a model-free saddle dynamics, in which the force in the original saddle dynamics is replaced by a surrogate model trained by the Gassian process learning. In the past few decades, the Gaussian process has been widely employed in extensive applications for constructing the surrogate models from the training data [33, 34]. In particular, there exist some recent works on combining the Gaussian process with searching algorithms of index-1 saddle points [5, 16, 21]. Here we adopt this idea in the computation of high-index saddle points. By this means, the exact form of the model is no longer necessary such that the saddle dynamics could be implemented based only on some observations of the force. Furthermore, the number of the force calculations during training is generally smaller than that in performing the saddle dynamics.

Based on the proposed model-free saddle dynamics and its local nature, we adopt the sequential learning framework [32] to develop a sequential learning saddle dynamics algorithm in which the queries of the training samples and the update or retraining of the surrogate force are performed online and around the latent trajectory. In this way, the accuracy of the surrogate model and the value of each sampling could be improved such that the queries of the force could be further reduced. The proposed method could be further combined with the downward search algorithm for construction of the model-free solution landscape [43, 45], which provides a pathway map consisting of both saddle points and minima and avoids the sampling on energy landscape of the model system. Numerical experiments are performed to demonstrate the effectiveness and efficiency of the proposed algorithm in comparison with the standard saddle dynamics.

The rest of the paper is organized as follows: In Sect. 2 we propose a model-free shrinking-dimer saddle dynamics by incorporating the Gaussian process learning with the original saddle dynamics. In Sect. 3 we present a sequential learning saddle dynamics algorithm to compute high-index saddle points and construct the solution landscapes. Numerical experiments are performed in Sect. 4 and we draw concluding remarks in the last section.

2 Model-free saddle dynamics

We present a model-free shrinking-dimer saddle dynamics for finding high-index saddle points of complex energy functions or dynamical systems, in which the exact form of the force may not be given a priori and is recovered from its observations at discrete locations. By this means, the saddle dynamics could be implemented based on these observations from, e.g., experiments or simulations, even without knowing the formulations of the energy functions or dynamics. This data-driven approach not only avoids the modeling procedure that could be difficult or inaccurate, but could significantly reduce the number of queries of the true force that may be expensive or time-consuming in practical problems.

2.1 Saddle dynamics and its implementation

We begin by introducing the high-index saddle dynamics proposed in [44] to find an index-k ($1 \le k \in \mathbb{N}$) saddle point of an energy function $E(x)$

$$\begin{cases} \dfrac{dx}{dt} = \beta\left(I - 2\sum\limits_{j=1}^{k} v_j v_j^\top\right)F(x), \\ \dfrac{dv_i}{dt} = \gamma\left(I - v_i v_i^\top - 2\sum\limits_{j=1}^{i-1} v_j v_j^\top\right)H(x)v_i, \quad 1 \le i \le k. \end{cases} \tag{1}$$

Here the force $F : \mathbb{R}^N \to \mathbb{R}^N$ is generated from the energy $E(x)$ by $F(x) = -\nabla E(x)$, $H(x) := -\nabla^2 E(x)$ corresponds to the Hessian of $E(x)$, β, $\gamma > 0$ are relaxation

parameters, x represents the state variable and direction variables $\{v_i\}_{i=1}^k$ form a basis for the unstable subspace of the Hessian at x.

More generally, (1) could be applied to find the saddle points of the dynamical systems $\dot{x} = F(x)$ which could be non-gradient, i.e., $F(x)$ is not a negative gradient of some energy function $E(x)$. In this case, the Hessian in (1) should be replaced by the Jacobian of $F(x)$ or its symmetrization [45]. Without loss of generality, we focus on the saddle dynamics of gradient systems in this paper.

As it is often expensive to calculate and store the Hessian, the dimer method [20, 47, 51] is applied to approximate the multiplication of the Hessian and the vector v as follows

$$H(x)v \approx \hat{H}(x, v, l) := \frac{F(x + lv) - F(x - lv)}{2l} \tag{2}$$

where $2l$ refers to the dimer length for some $l > 0$. Invoking this dimer approximation in (1) leads to the shrinking-dimer saddle dynamics [44]

$$\begin{cases} \dfrac{dx}{dt} = \beta\left(I - 2\displaystyle\sum_{j=1}^{k} v_j v_j^\top\right)F(x), \\ \dfrac{dv_i}{dt} = \gamma\left(I - v_i v_i^\top - 2\displaystyle\sum_{j=1}^{i-1} v_j v_j^\top\right)\hat{H}(x, v_i, l), \quad 1 \le i \le k, \\ \dfrac{dl}{dt} = -\dfrac{dG}{dl}. \end{cases} \tag{3}$$

Here an auxiliary function $G(l)$ defined on $[0, \infty)$ with $l = 0$ being its global minimum is introduced to control the dimer length. Following [47], we take $G(l) = l^2/2$ to get an exponential decay of the dimer length, while other choices such as $G(l) = l^3$ lead to a more gradual polynomial decay [48].

It is proved in [45] that a linearly stable stationary point x^* of the saddle dynamics (3) is an index-k saddle point satisfying $F(x^*) = 0$. For practical implementation, we follow [44, 45] to get the scheme of the numerical solutions $\{x_n\}$, $\{v_{i,n}\}_{i=1}^k$ and $\{l_n\}$ for $n \ge 1$

$$\begin{cases} x_n = x_{n-1} + \tau\beta\left(I - 2\displaystyle\sum_{j=1}^{k} v_{j,n-1} v_{j,n-1}^\top\right)F(x_{n-1}), \\ \tilde{v}_{i,n} = v_{i,n-1} + \tau\gamma\hat{H}(x_{n-1}, v_{i,n-1}, l_{n-1}), \quad 1 \le i \le k, \\ \{v_{i,n}\}_{i=1}^k = \text{GramSchmidt}(\{\tilde{v}_{i,n}\}_{i=1}^k), \end{cases} \tag{4}$$

equipped with the prescribed initial values x_0, l_0 and (orthonormal) $\{v_{i,0}\}_{i=1}^k$. Here $l_{n-1} = l(\tau(n-1))$ for $n \ge 1$ are determined by solving the equation of l in (3) analytically under suitable choice of G, and the Gram-Schmidt orthonormalization procedure is applied in the third equation of (4) to ensure the computational accuracy [44]. The scheme (4) has been extensively applied to compute the saddle points and to construct the solution landscapes [17, 18, 43]. However, implementing this scheme

requires the values of F at x_{n-1} and $x_{n-1} \pm lv_{i,n-1}$ for $1 \le i \le k$ at each time step t_n, which may be difficult to evaluate if the model (e.g. the form of F) is unknown or the query of F is expensive or time-consuming.

2.2 Surrogate force via Gaussian process

To accommodate the concerns mentioned above, we intend to replace the true force $F(x) = [F_1(x), \cdots, F_N(x)]^\top$ in (4) by a surrogate force $\mathcal{F}(x) = [\mathcal{F}_1(x), \cdots, \mathcal{F}_N(x)]^\top$, which is learned from the observations of F at training locations via the multi-output Gaussian process, see e.g. [24]. Suppose $\mathcal{F}(x)$ satisfies the following Gaussian process

$$\mathcal{F}(x) \sim \mathcal{GP}(0, \mathcal{K}(x, x')) \tag{5}$$

where the multi-output covariance $\mathcal{K}(x, x') \in \mathbb{R}^{N\times N}$ is defined as

$$\mathcal{K}(x, x') = \begin{bmatrix} k_{1,1}(x, x') & \cdots & k_{1,N}(x, x') \\ \vdots & \ddots & \vdots \\ k_{N,1}(x, x') & \cdots & k_{N,N}(x, x') \end{bmatrix}$$

with the element $k_{p,q}$ $(1 \le p, q \le N)$ corresponding to the covariance between $\mathcal{F}_p(x)$ and $\mathcal{F}_q(x')$. Denote the hyper-parameters in $k_{p,q}$ $(1 \le p, q \le N)$ as θ.

Given the m observations $Y := \{y^1, \cdots, y^m\}$ of F at the training data set $X := \{x_1^t, \cdots, x_m^t\}$ where $y^j = [y_1^j, \cdots, y_N^j]^\top$ for $1 \le j \le m$ and the observation of the ith output y_i $(1 \le i \le N)$ is assumed to satisfy $y_i(x) = \mathcal{F}_i(x) + \varepsilon_i$ with the independent and identically distributed Gaussian noise $\varepsilon_i \sim \mathcal{N}(0, \sigma_{s,i}^2)$. Then the hyper-parameters θ and the variances $\{\sigma_{s,i}\}_{i=1}^N$ could be inferred by the standard maximum likelihood estimation of maximizing the marginal likelihood $p(Y|X;\theta, \{\sigma_{s,i}\}_{i=1}^N)$ [24, 26, 34].

After inferring these parameters, the posterior distribution of $\mathcal{F}$ at any test point x_* could be analytically derived as [24, 34]

$$\mathcal{F}(x_*)|X, Y, x_* \sim \mathcal{N}(\mu_*, \Sigma_*)$$

where the predicted mean and variance are presented as

$$\mu_* = K_*^\top (K(X, X) + \Sigma)^{-1} \bar{y}$$

and

$$\Sigma(x_*) = \mathcal{K}(x_*, x_*) - K_*^\top (K(X, X) + \Sigma)^{-1} K_*.$$

Here $\bar{y} \in \mathbb{R}^{mN\times 1}$ is a rearrangement of the observations in Y defined by

$$\bar{y} = [y_1^1, \cdots, y_1^m, y_2^1, \cdots, y_2^m, \cdots, y_N^1, \cdots, y_N^m]^\top,$$

$K_* \in \mathbb{R}^{mN\times N}$ is defined by

$$K_* = \begin{bmatrix} K_{1,1}(X, x_*) & \cdots & K_{1,N}(X, x_*) \\ \vdots & \ddots & \vdots \\ K_{N,1}(X, x_*) & \cdots & K_{N,N}(X, x_*) \end{bmatrix}$$

where

$$K_{p,q}(X, x_*) = [k_{p,q}(x_1^t, x_*), \cdots, k_{p,q}(x_m^t, x_*)]^\top$$

for $1 \le p, q \le N$, $\Sigma = \Sigma_s \bigotimes I_m \in \mathbb{R}^{mN \times mN}$ is the noise matrix with

$$\Sigma_s = \begin{bmatrix} \sigma_{s,1}^2 & & \\ & \ddots & \\ & & \sigma_{s,N}^2 \end{bmatrix},$$

and the symmetric block matrix $K(X, X) \in \mathbb{R}^{mN \times mN}$ is given as

$$K(X, X) = \begin{bmatrix} K_{1,1}(X, X) & \cdots & K_{1,N}(X, X) \\ \vdots & \ddots & \vdots \\ K_{N,1}(X, X) & \cdots & K_{N,N}(X, X) \end{bmatrix}$$

with

$$K_{p,q}(X, X) = \begin{bmatrix} k_{p,q}(x_1^t, x_1^t) & \cdots & k_{p,q}(x_1^t, x_m^t) \\ \vdots & \ddots & \vdots \\ k_{p,q}(x_m^t, x_1^t) & \cdots & k_{p,q}(x_m^t, x_m^t) \end{bmatrix}$$

for $1 \le p, q \le N$.

In practice, we choose μ_* as the value of $\mathcal{F}(x_*)$, i.e. we denote $\mathcal{F}(x_*) = \mu_*$, and thus the predicted value of $F(x_*)$, and the diagonal entries of $\Sigma(x_*)$ represent its variances. Then we replace the true force $F(x)$ by its surrogate model $\mathcal{F}(x)$ in the saddle dynamics (3) to get the following model-free saddle dynamics

$$\begin{cases} \dfrac{dx}{dt} = \beta \left(I - 2 \displaystyle\sum_{j=1}^{k} v_j v_j^\top \right) \mathcal{F}(x), \\ \dfrac{dv_i}{dt} = \gamma \left(I - v_i v_i^\top - 2 \displaystyle\sum_{j=1}^{i-1} v_j v_j^\top \right) \mathcal{H}(x, v_i, l), \quad 1 \le i \le k, \\ \dfrac{dl}{dt} = -\dfrac{dG}{dl} \end{cases} \tag{6}$$

with

$$\mathcal{H}(x, v_i, l) := \frac{\mathcal{F}(x + l v_i) - \mathcal{F}(x - l v_i)}{2l}.$$

The corresponding numerical scheme takes an analogous form as (4)

$$\begin{cases} x_n = x_{n-1} + \tau\beta\left(I - 2\sum_{j=1}^{k} v_{j,n-1}v_{j,n-1}^{\top}\right)\mathcal{F}(x_{n-1}), \\ \tilde{v}_{i,n} = v_{i,n-1} + \tau\gamma\mathcal{H}(x_{n-1}, v_{i,n-1}, l_{n-1}), \quad 1 \le i \le k, \\ \{v_{i,n}\}_{i=1}^{k} = \text{GramSchmidt}(\{\tilde{v}_{i,n}\}_{i=1}^{k}). \end{cases} \tag{7}$$

3 A sequential learning algorithm

We propose a sequential leaning algorithm involving the model-free saddle dynamics scheme (7) in practical computations of saddle points and solution landscapes. The motivation is that due to the local nature of the saddle dynamics, the training data located far from the dynamical path of saddle dynamics may have less contributions or even introduce errors in predicting the force $F(x)$. Ideally, the training data should be sampled near the dynamical path, which, however, is not known a priori that makes the task of generating the training data for training the surrogate force nontrivial. To accommodate this issue, we adopt the sequential learning technique to perform the model-free saddle dynamics through an active learning framework, where the acquisition of training samples and the validation and update of the surrogate force are performed "online" (during optimization). By this means, we expect to reduce the number of training points while preserving the accuracy of the surrogate model and thus the computed saddle points.

The basic idea of the sequential learning is to divide the learning-based optimization into several trust region optimization subproblems. For the jth subproblem, we perform the Gaussian process learning to train the force $F(x)$ in a hypercube trust region

$$Q(x_c^j, \Delta^j) := \{x | \|x - x_c^j\|_\infty \le \Delta^j\}$$

for the center x_c^j and trust region length Δ^j based on a training data set with N_{sam} samples $D := \{(x_1^t, y_1), \cdots, (x_{N_{sam}}^t, y_{N_{sam}})\}$ and then perform the surrogate-force-based saddle dynamics (7) inside this region for $n = 1, 2\cdots$ until the x_n gets out of this region for some $n = n_*$.

Then we check the reliability of $\mathcal{F}$ at x_{n_*} in order to determine the center x_c^{j+1} and the trust region length Δ^{j+1} of the $(j+1)$th subproblem. We compute the maximum norm of the diagonal of the covariance matrix $\Sigma(x_{n_*})$ of $\mathcal{F}(x_{n_*})$, that is,

$$r := \|\text{diag}\,(\Sigma(x_{n_*}))\|_\infty. \tag{8}$$

Given the lower and upper tolerances $0 < \text{tol}_l < \text{tol}_u$, we encounter three cases:

- If $r < \text{tol}_l$, which means that the surrogate model is fairly reliable, we could set x_c^{j+1} as x_{n_*} and enlarge the trust region length in the $(j+1)$th subproblem.

- For the case $r > \text{tol}_u$, then the surrogate model may not be reliable such that the jth subproblem should be solved again using the retrained surrogate force under the updated training data set with N_{new} newly added samples. In this case, we set x_c^{j+1} back to x_c^j and shrink the trust region length to improve the accuracy.
- For $\text{tol}_l \leq r \leq \text{tol}_u$, we set x_c^{j+1} as x_{n_*} and keep the trust region length unchanged to solve the $(j+1)$th subproblem.

In all three cases, the training data obtained before will be inherited if they locate in the updated region, which helps to improve the accuracy of the surrogate model. The algorithm will be terminated if the error of two adjacent numerical solutions of x is smaller than the tolerance tol_x or the total number of steps N_t exceeds the prescribed value N_m. We summarize this algorithm in Algorithm 1, which computes an index-k saddle point and count the total number N_f of queries of F.

Algorithm 1 Sequential learning algorithm of finding an index-k saddle point

Step 1: Initialization
Initial data: x_0, $\{v_{i,0}\}_{i=1}^k$, l_0, β, γ, τ, tol_l, tol_u, N_m, $N_t = 0$, $N_f = 0$, tol_x, N_{sam}, N_{new}, x_c, Δ.
Initial sampling: Sample N_{sam} points $X = \{x_1^t, \cdots, x_{N_{sam}}^t\}$ in the trust region $Q(x_c, \Delta)$ and then evaluate the corresponding values of F as $Y = \{F(x_1^t), \cdots, F(x_{N_{sam}}^t))\}$ to get the training date set $D = \{X, Y\}$.
$N_f \Leftarrow N_f + N_{sam}$.
Step 2: Implement model-free saddle dynamics within $Q(x_c, \Delta)$
Infer the hyper-parameters θ and the variances $\{\sigma_{s,i}\}_{i=1}^N$ by maximum likelihood estimation with the training set D to obtain $\mathcal{F}$.
Implement saddle dynamics (7) using $\mathcal{F}$ until one of the following cases occurs:
(**a**) $x_n \notin Q(x_c, \Delta)$ for some $n = n_*$; (**b**) $N_t > N_m$; (**c**) $\|x_n - x_{n-1}\|_\infty \leq \text{tol}_x$ for some n.
For the cases (**b**) and (**c**), terminate the algorithm and output the latest numerical solution of x and N_f.
Step 3: Update of x_c, Δ **and** D
Compute r in (8).
if $r < \text{tol}_l$ **then** $x_c \Leftarrow x_{n_*}$ and $\Delta \Leftarrow 2\Delta$;
else
    if $r > \text{tol}_u$ **then** $\Delta \Leftarrow \Delta/2$;
    else $x_c \Leftarrow x_{n_*}$.
    end if
end if
Resample N_{new} training points in $Q(x_c, \Delta)$ as in **Step 1** to get a training data set D_{new}.
$N_f \Leftarrow N_f + N_{new}$, $D \Leftarrow D|_{x \in Q(x_c, \Delta)} \cup D_{new}$ and jump to **Step 2**.

Based on Algorithm 1, we could apply the downward or upward search algorithms proposed in [43, 45] for construction of solution landscapes. For instance, the downward search algorithm aims to search lower-index saddles and stable states from a high-index saddle point. Given an index-k saddle point and its k unstable directions denoted by $(x^*, v_1, \cdots, v_k)$ as a parent state, we could apply this algorithm

to search for an index-m saddle point with $0 \le m < k$. Typically, the initial value is chosen as a perturbation of the parent state x^* along the direction v_i for some $m < i \le k$ with the directions $\{v_j\}_{j=1}^m$ as the initial vectors to start the saddle dynamics. By repeating this procedure on the newly found saddle points, the landscape under the given index-k saddle point could be constructed completely.

4 Numerical experiments

We present numerical examples to substantiate the effectiveness and the computational cost of the sequential-learning Gaussian process saddle dynamics (GPSD) proposed by **Algorithm 1** in finding saddle points, in comparison with the standard saddle dynamics (SD). Here the computational cost is characterized by the number N_f of the force evaluations in order to demonstrate that the GPSD is more efficient than the SD in practical problems when the query of the force is expensive or time-consuming. In numerical experiments, we follow the conventional treatment to take $k_{p,q}$ for $1 \le p, q \le N$ as the squared exponential covariance function with the hyper-parameters $\theta := (\sigma_f, \sigma_l)$

$$k(x, x') := \sigma_f^2 \exp\Big(-\frac{\|x - x'\|_2^2}{2\sigma_l^2}\Big).$$

4.1 A Rosenbrock type function

We compute the saddle points of the following four-dimensional Rosenbrock type function

$$E(x_1, x_2, x_3, x_4) := a(x_4 - x_3^2)^2 + b(x_3 - x_2^2)^2 + c(x_2 - x_1^2)^2 + d(1 - x_1)^2.$$

We set $\tau = l_0 = 0.01$, tol_l=0.05, tol_u=0.15, $N_m = 20000$, tol_x=1×10^{-6}, $N_{sam} = N_{new} = 100$, $x_0 = (0.7, 0.8, 1.2, 0.7)$, $\Delta = 0.025$ and the dimer length with polynomial decay

$$l_n = \frac{l_0}{1 + (n\tau)^2}.$$

The initial values $\{v_{i,0}\}_{i=1}^k$ are chosen as the orthonormal eigenvectors of the Hessian of E at x_0. The Latin hypercube sampling technique [38] is applied to generate the training data locations. Under different coefficients a, b, c and d, the index of the saddle point (1, 1, 1, 1) of E is different, and we perform both methods for the four cases in Table 1.

Numerical results are presented in Tables 2, 3, which indicate that both GPSD and SD generate satisfactory approximations for the saddle point (1, 1, 1, 1) with different indexes, while the GPSD requires much less queries of the true force F, which demonstrate the potential of the GPSD in practical problems where the acquisition of the force is expensive or time-consuming.

Table 1 Index of the saddle point (1, 1, 1, 1) for E under different parameters

	a	b	c	d	Index
(i)	−0.5	0.5	0.5	2	1
(ii)	−0.5	0.5	−0.5	2	2
(iii)	−0.5	−0.5	−0.5	2	3
(iv)	−0.5	−0.5	−0.5	−2	4

Table 2 Comparison of SD and GPSD for approximating the saddle point (1, 1, 1, 1) of E

	SD	GPSD
(i)	(1.000, 0.999, 0.998, 0.995)	(1.022, 1.016, 1.018, 1.032)
(ii)	(1.000, 1.0000, 1.000, 1.000)	(1.011, 0.997, 1.004, 1.006)
(iii)	(1.000, 1.000, 1.000, 0.999)	(1.002, 0.991, 1.012, 0.999)
(iv)	(1.000, 0.999, 0.998, 0.995)	(0.990,1.002, 0.993, 0.991)

Table 3 Comparison of the number of force evaluation N_f in SD and GPSD for approximating the saddle point of E

	N_f in SD	N_f in GPSD
(i)	50673	5200
(ii)	13995	2700
(iii)	23548	3100
(iv)	155502	5200

4.2 A modified codesign problem

We compute the saddle point of the following modification of the codesign problem [2, 6] that simultaneously optimizes the plant design variable a and the control design variable $u(t)$ defined on [0, 0.1] in the plant design objective $J = J(a, u)$

$$J(a,u) := -\frac{a^2}{2} + \frac{1}{2}\int_0^{0.1} \xi^2(t)u^2(t)dt. \tag{9}$$

Different from the most conventional problems, in which the energy or cost functionals are exactly given in priori, there is an unknown state variable $\xi(t)$ in J that relates to both a and $u(t)$ via the underlying mechanism. In other words, for each query of J or the corresponding negative gradient $F = [a, -\xi(t)^2u(t)]^\top$ at some $(a, u(t))$ in each iteration of SD, we need to compute $\xi(t)$ by simulating or solving the governing process related to the current $(a, u(t))$, which is in general computationally expensive. Therefore, we expect to apply GPSD, which constructs a surrogate mapping from $(a, u(t))$ to F to partly circumvent queries of $\xi(t)$ in evaluating F, to reduce the number of simulating or computing the governing process during the SD iterations.

From $F = [a, -\xi(t)^2u(t)]^\top$ we find that $x^* := [0, 0]^\top$ is an index-1 saddle point of J and we will apply both SD and GPSD for its computation. For illustration, we follow

[6, Example 1] to select the governing process as a system of nonlinear differential equations on (0, 0.1]

$$\begin{cases} \dfrac{d\eta}{dt} = -a\eta + \xi^2, \\ \dfrac{d\xi}{dt} = \eta - 2a^2\xi - \eta^2 + u, \end{cases} \tag{10}$$

equipped with the initial conditions $\eta(0) = \xi(0) = 1$. In practice, much more complicated equations or processes than (10) could appear. We set $\tau = l_0 = 0.0025$, tol_l =0.05, tol_u=0.15, $N_m = 20000$, tol_x=1×10^{-6}, $N_{sam} = N_{new} = 300$, $\Delta = 0.1$ and the dimer length with polynomial decay as before. We discretize the domain [0, 0.1] of u, η and ξ by a uniform partition with mesh size 0.01, and approximate their values on this partition where the explicit difference method is used to solve (10). The initial value $v_{1,0}$ of saddle dynamics is chosen as the normalized vector of $[1, \cdots, 1]^\top$, and different initial values are applied as follows

$$(\text{i})\, x_0 = [0.2, \cdots, 0.2]^\top; (\text{ii})\, x_0 = [0.1, \cdots, 0.1]^\top; (\text{iii})\, x_0 = [0.05, \cdots, 0.05]^\top.$$

The Latin hypercube sampling technique [38] is applied to generate the training data locations. We measure the number N_s of simulating (10) and the errors between x^* and its output numerical approximation x_F in both methods in Table 4, which indicate that both methods generate satisfactory results while the GPSD performs much less simulations of the underlying process. Furthermore, as the initial value x_0 approaches x^* from (i) to (iii), the N_s of GPSD rapidly decreases while the N_s of SD does not decrease evidently. All these observations demonstrate the advantages of the proposed GPSD method.

4.3 A nonlocal phase-field model

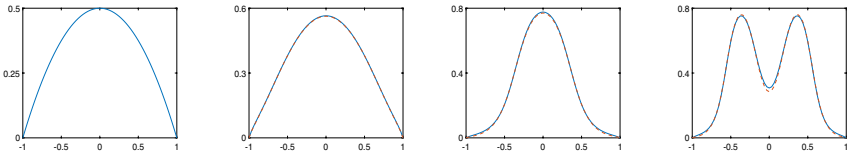
We employ SD and GPSD to compute saddle points of the following nonlocal phase-field equation [9, 37, 46]

$$\partial_t u = F(u) := -\kappa\left((-\Delta)^{\alpha/2}u + \frac{1}{\eta^2}(u-1)(u-1/2)u\right) \tag{11}$$

on $x \in (-1, 1)$, equipped with the initial and nonlocal boundary conditions

Table 4 Comparison of the number N_s of simulating the underlying process (10) and the errors $\|x_F - x^*\|_\infty$ in SD and GPSD for approximating the saddle point x^* of J

	N_s in SD	$\|x_F - x^*\|_\infty$ in SD	N_s in GPSD	$\|x_F - x^*\|_\infty$ in GPSD
(i)	10494	3.98×10^{-4}	4800	8.22×10^{-3}
(ii)	9618	3.98×10^{-4}	2700	9.47×10^{-4}
(iii)	8775	3.99×10^{-4}	1200	6.84×10^{-3}



(a) Initial value $u_0(x)$ (b) An index-1 saddle point (c) An index-2 saddle point (d) An index-3 saddle point

Fig. 1 Plots of $u_0(x)$ in (**a**) and the approximated saddle points by SD (blue solid) and GPSD (red dashed) for the cases (i)–(iii) in (**b**)–(**d**), respectively

Table 5 Comparison of the number of force evaluation N_f in SD and GPSD for computing the saddle points of the nonlocal phase field model (11)

	N_f in SD	N_f in GPSD
(i)	18225	960
(ii)	95125	3480
(iii)	117516	4200

$$u(x,0) = u_0(x), \quad x \in (-1,1); \quad u(x,t) = 0, \quad x \in (-\infty,-1] \cup [1,\infty), \quad t \geq 0.$$

Here κ represents the "elastic" relaxation time, η is the interface parameter and the fractional Laplacian operator is defined as [13, 23, 35]

$$(-\Delta)^{\alpha/2} u(x,t) := C_\alpha \ \text{P.V.} \int_{-\infty}^{\infty} \frac{u(x,t) - u(y,t)}{|x-y|^{1+\alpha}} dy$$

where

$$C_\alpha := \frac{2^{\alpha-1}\alpha\Gamma(\alpha/2+1/2)}{\pi^{1/2}\Gamma(1-\alpha/2)}.$$

To discretize model (11), we adopt the numerical scheme of $(-\Delta)^{\alpha/2}$ in [10] with the mesh size $h = 2^{-5}$. Furthermore, we set $\alpha = 1.5$, tol_l=0.05, tol_u=0.15, $N_m = 20000$, tol_x=1×10^{-5}, $N_{sam} = N_{new} = 120$, $u_0(x) = 0.5(1-x^2)$ as shown in Fig. 1(a), $\kappa = 0.02$, $\Delta = 0.01$ and the shrinking dimer with exponential decay $l_n = e^{-n\tau} l_0$. The initial values $\{v_{i,0}\}_{i=1}^k$ are chosen as the orthonormal eigenvectors of the Jacobian of F at u_0. Appropriate smooth curves serve as the samples of the training set. We take $\tau = l_0 = 0.00025$ to compute the saddle points of model (11) under different values of $1/\eta^2$ and different indexes as follows

$$(\text{i})\ \frac{1}{\eta^2} = 30, \ \text{index} = 1; (\text{ii})\ \frac{1}{\eta^2} = 80, \ \text{index} = 2; (\text{iii})\ \frac{1}{\eta^2} = 120, \ \text{index} = 3.$$

Numerical results are presented in Fig. 1 and Table 5, which lead to the same conclusions as in the first example.

The we combine both SD and GPSD under $\tau = l_0 = 0.0005$ with the downward search algorithm proposed in [43, 45] to demonstrate the construction of the solution landscapes. We fix $1/\eta^2 = 120$ and $\kappa = 0.02$ and start from an index-3 saddle

point to construct the solution landscape of the nonlocal phase-field model (11). Numerical results are presented in Fig. 2, which indicate that both methods generate similar solution landscapes, which demonstrates the effectiveness of the GPSD in solution landscape constructions.

4.4 A particle cluster model

The Morse potential provides a model for analyzing the effect of strain and the range of the interparticle forces on structure, dynamics and thermodynamics [7, 8, 39]. In this example we employ GPSD to compute the solution landscape of the cluster of 5 particles in three space dimensions described by the energy

$$E = \sum_{1 \le i < j \le 5} V(r_{i,j})$$

where $r_{i,j}$ is the Euclidean distance between the ith and the jth particles and $V(r)$ is the Morse potential [30]

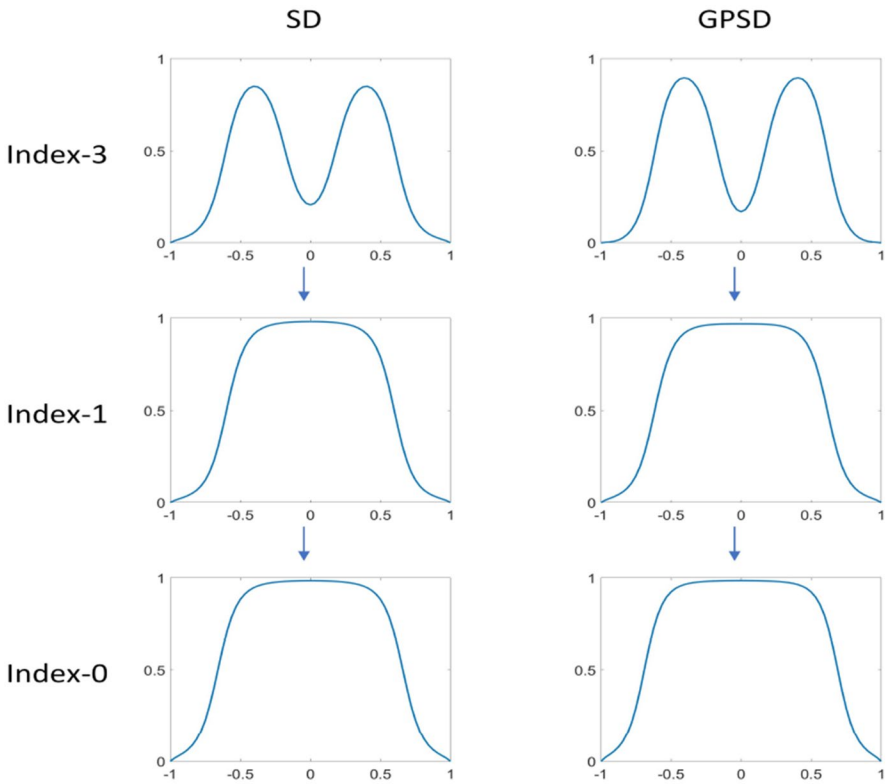


Fig. 2 Solution landscape of nonlocal phase-field model (11) constructed by (left) SD- and (right) GPSD-based downward search algorithm

$$V(r) = e^{-2\rho(r-1)} - 2e^{-\rho(r-1)}.$$

We select $\rho = 3$, tol_l=0.05, tol_u=0.15, $N_m = 20000$, tol_x=1×10^{-5}, $N_{sam} = 200$, $N_{new} = 200$, and $l_n = e^{-n\tau} l_0$. We combine the GPSD under $\tau = l_0 = 0.001$ with the downward search algorithm proposed in [43, 45] to construct the energy landscape from an index-3 saddle point in Fig. 3, which again indicates the effectiveness of the GPSD.

5 Concluding remarks

In this work we propose a model-free shrinking-dimer saddle dynamics and a corresponding sequential learning algorithm for finding any-index saddle points of complex models and constructing their solution landscapes. This data-driven approach avoids employing the exact form of the model and could significantly reduce the number of queries of the force that may be expensive or time-consuming, and the active learning mechanism along the latent trajectory increases the value of each sampling. More importantly, with the proposed algorithm, we can efficiently

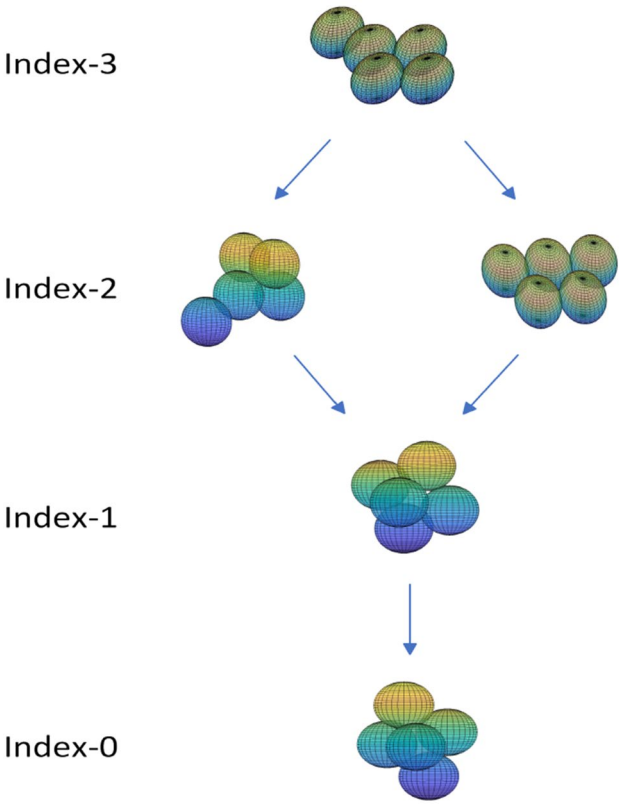


Fig. 3 Energy landscape of particle cluster constructed by the GPSD-based downward search algorithm

construct the model-free solution landscape, in which only the informative saddle points and minima are computed, so that one can avoid the sampling on the whole energy landscape of complex models.

As far as we know, high-index saddle dynamics method is the only method to compute high-index saddle points and to construct solution landscapes, though there exist several methods for finding saddle points with low indexes. For instance, the gentlest ascent method is proposed in [11] to compute the index-1 and index-2 saddle points. Some works also use the Newton's method to search saddle points, cf. [4], while the convergence behavior depends heavily on the selection of the initial value such that it is difficult to efficiently compute all saddle points. Thus, it is difficult to compare the proposed method with other ones in constructing solution landscapes.

There are several potential extensions of the current method. For instance, different covariance kernels in the Gaussian process or different learning methods such as the deep learning method could be applied to train the surrogate model. The proposed data-driven method could be further applied for constructing the solution landscapes of more complex problems, and how to utilize the structure of the solution landscape to further reduce the number of queries of the force is an interesting but challenging problem that will be considered later. Finally, there are some recent progresses on numerical analysis to high-index saddle dynamics [53–55], while it is not clear how to extend the developed methods to perform numerical analysis for the proposed model-free saddle dynamics due to the approximation of F via the surrogate model. We will investigate this interesting topic in the near future.

Acknowledgements We greatly appreciate Dr. Siwei Duo for providing the MATLAB code of computing the fractional Laplacian in [10].

Author Contributions LZ: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing. PZ: Funding acquisition, Supervision. XZ: Formal analysis, Funding acquisition, Investigation, Methodology, Visualization, Writing – original draft.

Funding This work was partially supported by the National Key R &D Program of China No. 2021YFF1200500, the Taishan Scholars Program of Shandong Province, and the National Natural Science Foundation of China No. 12225102, 12050002, 12288101.

Data availability No data was used for the research described in the article.

Declarations

Conflict of interest The authors have no conflict of interest.

References

1. Foresman, J., Head-Gordon, M., Pople, J., Frisch, M.: Toward a systematic molecular orbital theory for excited states. J. Phys. Chem. **96**, 135–149 (1992)
2. Allison, J., Guo, T., Han, Z.: Co-design of an active suspension using simultaneous dynamic optimization. ASME J. Mech. Des. **136**, 081003 (2014)
3. Chen, C., Xie, Z.: Search extension method for multiple solutions of a nonlinear problem. Comput. Math. Appl. **47**, 327–343 (2004)

4. Dauphin, Y., Pascanu, R., Gulcehre, C.: Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *2012 Proceedings of the 25th International Conference on Neural Information Processing Systems*, Cambridge, USA, 2933–2941 (2012)
5. Denzel, A., Kastner, J.: Gaussian process regression for transition state search. J. Chem. Theory Comput. **14**, 5777–5786 (2018)
6. Deshmukh, A., Allison, J.: Design of dynamic systems using surrogate models of derivative functions. ASME. J. Mech. Des. **139**, 101402 (2017)
7. Doye, J., Wales, D.: The Structure and stability of atomic liquids: From clusters to bulk. Science **271**, 484–487 (1996)
8. Doye, J., Wales, D., Berry, R.: The effect of the range of the potential on the structures of clusters. J. Chem. Phys. **103**, 4234–4249 (1995)
9. Duo, S., Wang, H.: A fractional phase-field model using an infinitesimal generator of α stable Lévy process. J. Comput. Phys. **384**, 253–269 (2019)
10. Duo, S., Zhang, Y.: Computing the ground and first excited states of the fractional schrödinger equation in an infinite potential well. Commun. Comput. Phys. **18**, 321–350 (2015)
11. E, W., Zhou, X.: The gentlest ascent dynamics. Nonlinearity **24**, 1831–1842 (2011)
12. E, W., Ren, W., Vanden-Eijnden, E.: String method for the study of rare events. Phys. Rev. B **66**, 052301 (2002)
13. D'Elia, M., Du, Q., Glusa, C., Gunzburger, M., Tian, X., Zhou, Z.: Numerical methods for nonlocal and fractional models. Acta Num. **29**, 1–124 (2020)
14. Gao, W., Leng, J., Zhou, X.: An iterative minimization formulation for saddle point search. SIAM J. Numer. Anal. **53**, 1786–1805 (2015)
15. Gould, N., Ortner, C., Packwood, D.: A dimer-type saddle search algorithm with preconditioning and linesearch. Math. Comp. **85**, 2939–2966 (2016)
16. Gu, S., Wang, H., Zhou, X.: Active learning for transition state calculation. J. Sci. Comput. **93**, 78 (2022)
17. Han, Y., Yin, J., Zhang, P., Majumdar, A., Zhang, L.: Solution landscape of a reduced Landau-de Gennes model on a hexagon. Nonlinearity **34**, 2048–2069 (2021)
18. Han, Y., Yin, J., Hu, Y., Majumdar, A., Zhang, L.: Solution landscapes of the simplified Ericksen-Leslie model and its comparison with the reduced Landau-de Gennes model. Proc. Roy. Soc. A-Math. Phy. **477**, 20210458 (2021)
19. Heidrich, D., Quapp, W.: Saddle points of index 2 on potential energy surfaces and their role in theoretical reactivity investigations. Theor. Chim. Acta **70**, 89–98 (1986)
20. Henkelman, G., Jónsson, H.: A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. J. Chem. Phys. **111**, 7010–7022 (1999)
21. Koistinen, O., Dagbjartsdóttir, F., Ásgeirsson, V., Vehtari, A., Jónsson, H.: Nudged elastic band calculations accelerated with Gaussian process regression. J. Chem. Phys. **147**, 152720 (2017)
22. Li, Y., Zhou, J.: A minimax method for finding multiple critical points and its applications to semilinear PDEs. SIAM J. Sci. Comput. **23**, 840–865 (2001)
23. Lischke, A., Pang, G., Gulian, M., Song, F., Glusa, C., Zheng, X., Mao, Z., Cai, W., Meerschaert, M., Ainsworth, M., Karniadakis, G.: What is the fractional Laplacian? A comparative review with new results. J. Comput. Phys. **404**, 109009 (2020)
24. Liu, H., Cai, J., Ong, Y.: Remarks on multi-output Gaussian process regression. Knowledge-Based Syst. **144**, 102–121 (2018)
25. Liu, W., Xie, Z., Yi, W.: Normalized Goldstein-type local minimax method for finding multiple unstable solutions of semilinear elliptic PDEs. Commun. Math. Sci. **19**, 147–174 (2021)
26. Liu, X., Zhu, Q., Lu, H.: Modeling multiresponse surfaces for airfoil design with multiple-output-Gaussian-process regression. J. Aircr. **51**, 740–747 (2014)
27. Luo, Y., Zheng, X., Cheng, X., Zhang, L.: Convergence analysis of discrete high-index saddle dynamics. SIAM J. Numer. Anal. **60**, 2731–2750 (2022)
28. Mehta, D.: Finding all the stationary points of a potential-energy landscape via numerical polynomial-homotopy-continuation method. Phys. Rev. E **84**, 025702 (2011)
29. Milnor, J. W.: *Morse Theory*, Princeton University Press, (1963)
30. Morse, P.: Diatomic molecules according to the wave mechanics II. Vib. Lev. Phys. Rev. **34**, 57–64 (1929)
31. Nie, Q., Qiao, L., Qiu, Y., Zhang, L., Zhao, W.: Noise control and utility: from regulatory network to spatial patterning. Sci. China Math. **63**, 425–440 (2020)

32. Powel, M.: Algorithms for nonlinear constraints that use Lagrangian function. Math. Programm. **14**, 224–248 (1978)
33. Raissi, M., Perdikaris, P., Karniadakis, G.: Machine learning of linear differential equations using Gaussian processes. J. Comput. Phys. **348**, 683–693 (2017)
34. Rasmussen, C., Williams, C.: *Gaussian processes for machine learning*. MIT Press, (2006)
35. Sheng, C., Shen, J., Tang, T., Wang, L., Yuan, H.: Fast Fourier-like mapped Chebyshev spectral-Galerkin methods for PDEs with integral fractional Laplacian in unbounded domains. SIAM J. Numer. Anal. **58**, 2435–2464 (2020)
36. Shi, B., Han, Y., Zhang, L.: Nematic liquid crystals in a rectangular confinement: Solution landscape, and bifurcation. SIAM J. Appl. Math. **82**, 1808–1828 (2022)
37. Song, F., Xu, C., Karniadakis, G.: A fractional phase-field model for two-phase flows with tunable sharpness: algorithms and simulations. Comput. Methods Appl. Mech. Engrg. **305**, 376–404 (2016)
38. Tang, B.: Orthogonal array-based Latin hypercubes. J. Amer. Stat. Assoc. **88**, 1392–1397 (1993)
39. Wales, D.: Energy landscapes of clusters bound by short-ranged potentials. Chem. Phys. Chem. **11**, 2491–2494 (2010)
40. Wang, W., Zhang, L., Zhang, P.: Modelling and computation of liquid crystals. Acta Num. **30**, 765–851 (2021)
41. Xie, Z., Yuan, Y., Zhou, J.: On finding multiple solutions to a singularly perturbed Neumann problem. SIAM J. Sci. Comput. **34**, A395–A420 (2012)
42. Yin, J., Jiang, K., Shi, A.-C., Zhang, P., Zhang, L.: Transition pathways connecting crystals and quasicrystals. Proc. Natl. Acad. Sci. U.S.A. **118**, e2106230118 (2021)
43. Yin, J., Wang, Y., Chen, J., Zhang, P., Zhang, L.: Construction of a pathway map on a complicated energy landscape. Phys. Rev. Lett. **124**, 090601 (2020)
44. Yin, J., Zhang, L., Zhang, P.: High-index optimization-based shrinking dimer method for finding high-index saddle points. SIAM J. Sci. Comput. **41**, A3576–A3595 (2019)
45. Yin, J., Yu, B., Zhang, L.: Searching the solution landscape by generalized high-index saddle dynamics. Sci. China Math. **64**, 1801 (2021)
46. Yu, B., Zheng, X., Zhang, P., Zhang, L.: Computing solution landscape of nonlinear space-fractional problems via fast approximation algorithm. J. Comput. Phys. **468**, 111513 (2022)
47. Zhang, J., Du, Q.: Shrinking dimer dynamics and its applications to saddle point search. SIAM J. Numer. Anal. **50**, 1899–1921 (2012)
48. Zhang, J., Du, Q.: Constrained shrinking dimer dynamics for saddle point search with constraints. J. Comput. Phys. **231**, 4745–4758 (2012)
49. Zhang, L., Chen, L., Du, Q.: Morphology of critical nuclei in solid-state phase transformations. Phys. Rev. Lett. **98**, 265703 (2007)
50. Zhang, L., Chen, L., Du, Q.: Simultaneous prediction of morphologies of a critical nucleus and an equilibrium precipitate in solids. Commun. Comput. Phys. **7**, 674–682 (2010)
51. Zhang, L., Du, Q., Zheng, Z.: Optimization-based shrinking dimer method for finding transition states. SIAM J. Sci. Comput. **38**, A528–A544 (2016)
52. Zhang, L., Ren, W., Samanta, A., Du, Q.: Recent developments in computational modelling of nucleation in phase transformations. NPJ Comput. Mater. **2**, 16003 (2016)
53. Zhang, L., Zhang, P., Zheng, X.: Error estimates of Euler discretization to high-index saddle dynamics. SIAM J. Numer. Anal. **60**, 2925–2944 (2022)
54. Zhang, L., Zhang, P., Zheng, X.: Mathematical and numerical analysis to shrinking-dimer saddle dynamics with local Lipschitz conditions. CSIAM Trans. Appl. Math. **4**, 157–176 (2023)
55. Zhang, L., Zhang, P., Zheng, X.: Discretization and index-robust error analysis for constrained high-index saddle dynamics on high-dimensional sphere. Sci. China Math. (2023). https://doi.org/10.1007/s11425-022-2149-2