Commun Nonlinear Sci Numer Simulat 145 (2025) 108731

Contents lists available at ScienceDirect

Communications in Nonlinear Science and Numerical Simulation

journal homepage: www.elsevier.com/locate/cnsns

Research paper

Construction and analysis for orthonormalized Runge–Kutta schemes of high-index saddle dynamics

Shuai Miao [a], Lei Zhang [b], Pingwen Zhang [c], Xiangcheng Zheng [d,*]

[a] School of Mathematical Sciences, Peking University, Beijing, 100871, China
[b] Beijing International Center for Mathematical Research, Center for Quantitative Biology, Center for Machine Learning Research, Peking University, Beijing 100871, China
[c] School of Mathematics and Statistics, Wuhan University, Wuhan, 430072, China
[d] School of Mathematics, State Key Laboratory of Cryptography and Digital Economy Security, Shandong University, Jinan 250100, China

ARTICLE INFO

Keywords:
Saddle point
Saddle dynamics
Orthonormalized Runge–Kutta scheme
Error estimate

ABSTRACT

Saddle points are prevalent in complex systems and contain important information. The high-index saddle dynamics (HiSD) and the generalized HiSD (GHiSD) are two efficient approaches for determining saddle points of any index and for constructing the solution landscape. In this work, we first present an example to show that the orthonormality of directional vectors in saddle dynamics is critical in locating the saddle point. Then we construct two orthonormalized Runge–Kutta schemes tailored for the HiSD and GHiSD. We find that if a set of vectors are almost orthonormal with the error $O(\tau^\alpha)$ for some $\alpha > 0$, then the Gram–Schmidt process also applies an $O(\tau^\alpha)$ perturbation to orthonormalize them. We apply this and employ the structures of Runge–Kutta schemes to prove the almost orthonormality in numerical schemes and then prove their second-order accuracy with respect to the time step size. We substantiate the theoretical findings by several numerical experiments.

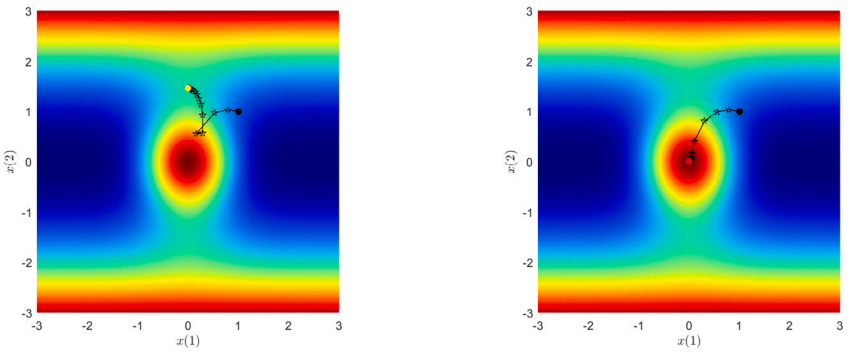
1. Introduction

Searching stationary points of complex systems is an important but challenging topic with wide applications across many scientific fields [1–16]. Over the past few decades, scholars have proposed a series of efficient algorithms for searching saddle points [5,17–30] and have conducted various numerical analyses [28,31–35]. In 2019, the high-index saddle dynamics (HiSD) method was introduced to calculate any index saddle points [36]. This method serves as an efficient tool for constructing the solution landscape, offering a pathway map that begins with a high-index saddle point and connects to lower-index saddle points and minimizers [37]. The solution landscape approach has been utilized to explore various physical systems [38–43].

Consider a twice Fréchet differentiable energy functional $E(\boldsymbol{x})$ with $\boldsymbol{x} \in \mathbb{R}^{\mathbb{N}}$ defined on a real Hilbert space $\mathcal{H}$, with the associated natural force $\boldsymbol{F}(\boldsymbol{x}) = -\nabla E(\boldsymbol{x})$ and the negative Hessian $\mathbb{G}(\boldsymbol{x}) = -\nabla^2 E(\boldsymbol{x})$. Clearly, $\mathbb{G}(\boldsymbol{x})$ is symmetric, i.e., $\mathbb{G}(\boldsymbol{x}) = \mathbb{G}(\boldsymbol{x})^\top$. An equilibrium (or critical) point $\bar{\boldsymbol{x}} \in \mathcal{H}$, where $\boldsymbol{F}(\bar{\boldsymbol{x}}) = \boldsymbol{0}$, could be categorized as either a local extremum or a saddle point based on whether $E(\bar{\boldsymbol{x}})$ represents a local extremum in the vicinity of $\bar{\boldsymbol{x}}$. Assuming $\bar{\boldsymbol{x}}$ is non-degenerate, meaning the Hessian $\nabla^2 E(\bar{\boldsymbol{x}})$ has a bounded inverse at $\bar{\boldsymbol{x}}$, we can define the index (Morse index) at $\bar{\boldsymbol{x}}$ via Morse theory [44]. This index is the highest dimension of the subspace on which $\nabla^2 E(\bar{\boldsymbol{x}})$ is negative definite. The HiSD for an index-k saddle point of the energy functional $E(\boldsymbol{x})$, where

* Corresponding author.
E-mail addresses: miaoshuai@math.pku.edu.cn (S. Miao), zhangl@math.pku.edu.cn (L. Zhang), pzhang@pku.edu.cn (P. Zhang), xzheng@sdu.edu.cn (X. Zheng).

https://doi.org/10.1016/j.cnsns.2025.108731
Received 28 December 2024; Received in revised form 19 February 2025; Accepted 21 February 2025
Available online 1 March 2025



(a) RK2-I for locating an index-2 saddle point (b) oRK2-I for locating an index-2 saddle point

Fig. 1. Comparison of the numerical results of RK2 and oRK2-I (Black dot: starting point; Yellow dot and red dot: end points.).

$1 \le k \in \mathbb{N}$, is expressed as follows [36]:

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = \beta\left(\mathbb{I} - 2\sum\limits_{j=1}^{k} \boldsymbol{v}_j\boldsymbol{v}_j^{\top}\right)\boldsymbol{F}(\boldsymbol{x}), \\ \dfrac{d\boldsymbol{v}_i}{dt} = \gamma\left(\mathbb{I} - \boldsymbol{v}_i\boldsymbol{v}_i^{\top} - 2\sum\limits_{j=1}^{i-1} \boldsymbol{v}_j\boldsymbol{v}_j^{\top}\right)\mathbb{G}(\boldsymbol{x})\boldsymbol{v}_i, \quad 1 \le i \le k, \end{cases} \tag{1}$$

where $\boldsymbol{x}$ represents a position variable, $\{\boldsymbol{v}_i\}_{i=1}^{k}$ are k directional variables, and β, $\gamma > 0$ are relaxation parameters.

The HiSD method extends beyond gradient systems and has been adapted for non-gradient systems through a generalized HiSD (GHiSD) approach. This extension enables the computation of any index saddle points and facilitates the exploration of solution landscapes in non-gradient systems [45]. Consider the autonomous dynamical system in $\mathbb{R}^n$:

$$\dot{\boldsymbol{x}} = \boldsymbol{F}(\boldsymbol{x}), \tag{2}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{F} : \mathbb{R}^n \to \mathbb{R}^n$ is a second-order continuously differentiable function. If a point x^* satisfies $\boldsymbol{F}(\boldsymbol{x}^*) = 0$, it is called an equilibrium point of system (2). Assuming $\boldsymbol{x}^*$ is a hyperbolic equilibrium point, the Jacobian $\mathbb{J}(\boldsymbol{x}^*) = \nabla\boldsymbol{F}(\boldsymbol{x}^*)$ has no eigenvalues with zero real part. If all eigenvalues of $\mathbb{J}(\boldsymbol{x}^*)$ have positive (or negative) real parts, $\boldsymbol{x}^*$ is a source (or sink); otherwise, it is a saddle point. The index of a saddle point is defined by the number of eigenvalues of $\mathbb{J}(\boldsymbol{x}^*)$ with positive real parts [46]. The GHiSD for an index-k saddle point is formulated as [45]:

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = \beta\left(\mathbb{I} - 2\sum\limits_{j=1}^{k} \boldsymbol{v}_j\boldsymbol{v}_j^{\top}\right)\boldsymbol{F}(\boldsymbol{x}), \\ \dfrac{d\boldsymbol{v}_i}{dt} = \gamma\left(\mathbb{I} - \boldsymbol{v}_i\boldsymbol{v}_i^{\top} - 2\sum\limits_{j=1}^{i-1} \boldsymbol{v}_j\boldsymbol{v}_j^{\top}\right)\left(\mathbb{J}(\boldsymbol{x}) + \mathbb{J}^{\top}(\boldsymbol{x})\right)\boldsymbol{v}_i, \quad 1 \le i \le k, \end{cases} \tag{3}$$

where $\boldsymbol{x}$ represents a position variable and $\{\boldsymbol{v}_i\}_{i=1}^{k}$ are k directional variables.

It is shown in [36,45] that if the initial values of $\{\boldsymbol{v}_i(t)\}_{i=1}^{k}$ satisfy the orthogonality condition, then $\{\boldsymbol{v}_i(t)\}_{i=1}^{k}$ keep orthonormal for any $t \ge 0$. Note that such orthogonality has been used to give the reflection operator in the dynamics of position variable, cf. the derivations of HiSd and GHiSD in [36,45]. Thus, if such orthogonality is not preserved in numerical methods, the searching direction of the position variable may gradually diverge, leading to the failure of the convergence to the saddle point. For instance, we calculate the index-2 saddle point of the Eckhardt surface [47] using both the midpoint formula (RK2-I), a second-order Runge–Kutta scheme, and the orthonormalized second-order Runge–Kutta scheme (oRK2-I) (31) of the index-2 saddle dynamics with the terminal time $T = 14$, the time step size $\tau = 0.14$ and initial values $\boldsymbol{x}(0) = (1, 1)$, $\boldsymbol{v}_1(0) = (1, 0)$ and $\boldsymbol{v}_2(0) = (0, 1)$. The results are shown in Fig. 1, which indicate that the RK2-I method fails to locate the saddle point, while the oRK2-I method converges to an index-2 saddle point. The better performance of the orthonormalized Runge–Kutta methods, along with the desire to further enhance its confidence, has prompted us to conduct their error estimates.

In recent years, some results have been achieved in the numerical analysis of HiSD. The rate of convergence of the numerical scheme for HiSD was analyzed in [48]. In [49], the error estimate for the explicit Euler discrete scheme of HiSD and GHiSD was conducted, demonstrating a first-order rate of convergence with respect to time for both pathway and eigenvectors. In [50], the error estimate for numerical discretization to the shrinking-dimer saddle dynamics by matching the dimer length and the time step size was analyzed. In [51], a semi-implicit discrete scheme for HiSD was constructed and proved to be first-order accurate. In [52] the mechanism of orthonormal preservation of directional variables in HiSD is revealed by numerical analysis. In this paper, we develop orthonormalized explicit Runge–Kutta schemes for HiSD and GHiSD in order to achieve high-order accuracy. The main difficulties we overcome lie in employing the structures of Runge–Kutta schemes to prove the almost orthonormality in numerical schemes.

The rest of this paper is organized as follows: In Section 2, we introduce a unified abstract formulation of the dynamics of directional variables for both HiSD and GHiSD and then provide an analysis of the errors arising from the orthogonalization process. In Section 3, we develop two orthonormalized second-order Runge–Kutta schemes and carry out detailed error estimations for each. Section 4 presents several numerical experiments to illustrate the performance of these schemes. Finally, Section 5 concludes with a comprehensive summary of our findings.

2. Preliminary results

2.1. A unified formulation

Let

$$\boldsymbol{g}_i = \left(\mathbb{I} - \boldsymbol{v}_i\boldsymbol{v}_i^\top - 2\sum_{j=1}^{i-1}\boldsymbol{v}_j\boldsymbol{v}_j^\top\right)\mathbb{G}(\boldsymbol{x})\boldsymbol{v}_i$$

or

$$\boldsymbol{g}_i = \left(\mathbb{I} - \boldsymbol{v}_i\boldsymbol{v}_i^\top - 2\sum_{j=1}^{i-1}\boldsymbol{v}_j\boldsymbol{v}_j^\top\right)\left(\mathbb{J}(\boldsymbol{x}) + \mathbb{J}^\top(\boldsymbol{x})\right)\boldsymbol{v}_i,$$

then we unify the dynamics of directional variables in HiSD and GHiSD into the following abstract form:

$$\begin{cases}\dfrac{d\boldsymbol{v}_i}{dt} = \boldsymbol{g}_i, \\ \boldsymbol{v}_i|_{t=0} = \boldsymbol{v}_{i,0},\end{cases} \qquad i = 1, 2, \dots, k. \tag{4}$$

Here $\{\boldsymbol{v}_{i,0}\}_{i=1}^k \in \mathbb{R}^d$ represent the initial values of $\{\boldsymbol{v}_i\}_{i=1}^k$ satisfying the orthonormalization condition $\boldsymbol{v}_{i,0}^\top\boldsymbol{v}_{j,0} = \delta_{i,j}$ where $\delta_{i,j}$ is the Dirac notation defined as

$$\delta_{i,j} = \begin{cases}1, & i = j, \\ 0, & i \neq j.\end{cases}$$

Then $\boldsymbol{v}_i^\top\boldsymbol{v}_j = \delta_{i,j}$ is valid for any $t \geq 0$ [36,45].

By carefully analyzing the error estimate process for HiSD in [49], we notice that the main difficulty lies in the numerical analysis of directional variables due to the additional orthonormalization, while the error estimate for the position variable follows the standard procedure and could be carried out similarly as, e.g. [49, Theorem 4.3] where the errors of $\boldsymbol{x}$ is controlled by the errors of $\{\boldsymbol{v}_i\}_{i=1}^k$ and higher-order terms of τ. Furthermore, under the Lipschitz condition of $\boldsymbol{F}(\boldsymbol{x})$, the position variable $\boldsymbol{x}$ is bounded for $0 \leq t \leq T$ for some terminal time T [49]. Thus, we mainly focus on the dynamics (4) of $\{\boldsymbol{v}_i\}_{i=1}^k$ and omit its dependence on $\boldsymbol{x}$ (detailed numerical schemes of HiSD and GHiSD can be found in Appendix). Finally, if $\mathbb{G}$ (or $\mathbb{J}$) is bounded for bounded $\boldsymbol{x}$, there exist positive constants L_g and C_g such that for $0 \leq t \leq T$ and $1 \leq i \leq k$

$$\|\boldsymbol{g}_i(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k) - \boldsymbol{g}_i(\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_k)\| \leq L_g \max_{1\leq j\leq k}\|\boldsymbol{v}_j - \boldsymbol{u}_j\|, \tag{5}$$

$$\|\boldsymbol{g}_i(\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_k)\| \leq C_g, \tag{6}$$

$$\forall\,\{\boldsymbol{v}_l\}_{l=1}^k,\ \{\boldsymbol{u}_l\}_{l=1}^k \text{ such that } \boldsymbol{v}_l^\top\boldsymbol{v}_j = \boldsymbol{u}_l^\top\boldsymbol{u}_j = \delta_{l,j} \text{ for } 1 \leq l, j \leq k. \tag{7}$$

Here $\|\cdot\|$ represents the standard l^2 norm of a vector or a matrix.

2.2. Analysis of orthogonalization perturbation

To numerically solve the equations, we generate a uniform grid over the interval $[0, T]$ with a step size τ. The time nodes are denoted as $t_n = n\tau$ for $n = 0, 1, 2, \dots, N$ with $N = T/\tau$. To ensure the orthonormality of numerical solutions of directional vectors, a common approach is to perform Gram–Schmidt orthogonalization at each time step. Specifically, for a set of k vectors $\{\tilde{\boldsymbol{v}}_i\}_{i=1}^k$, the Gram–Schmidt orthogonalization reads:

$$\boldsymbol{v}_i = \frac{1}{Y_i}\left(\tilde{\boldsymbol{v}}_i - \sum_{j=1}^{i-1}\left(\tilde{\boldsymbol{v}}_i^\top\boldsymbol{v}_j\right)\boldsymbol{v}_j\right), \quad 1 \leq i \leq k,$$

where the normalization factor is calculated as

$$Y_i = \left\|\tilde{\boldsymbol{v}}_i - \sum_{j=1}^{i-1}\left(\tilde{\boldsymbol{v}}_i^\top\boldsymbol{v}_j\right)\boldsymbol{v}_j\right\| = \left(\|\tilde{\boldsymbol{v}}_i\|^2 - \sum_{j=1}^{i-1}\left(\tilde{\boldsymbol{v}}_i^\top\boldsymbol{v}_j\right)^2\right)^{1/2}.$$

Them the resulting vectors $\{\boldsymbol{v}_i\}_{i=1}^k$ are orthonormal. In the subsequent numerical schemes, we denote the orthonormalization as

$$\{\boldsymbol{v}_i\}_{i=1}^k = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_i\}_{i=1}^k).$$

Incorporating orthonormalization into numerical schemes introduces new challenges. Nevertheless, based on the specific structure of the Gram–Schmidt process, the difference of vectors before and after orthonormalization has the following tight relation.

Lemma 1. *Given a set of vectors $\{\tilde{\boldsymbol{v}}_i\}_{i=1}^k$, let $\{\boldsymbol{v}_i\}_{i=1}^k$ be orthonormal vectors generated from $\{\tilde{\boldsymbol{v}}_i\}_{i=1}^k$ via the Gram–Schmidt process. If*

$$|\tilde{\boldsymbol{v}}_i^\top \tilde{\boldsymbol{v}}_j - \delta_{i,j}| \le M\tau^\alpha, \quad 1 \le i, j \le k,$$

for some $\alpha > 0$ where M is a positive constant, then for a sufficiently small time step τ, there exists a positive constant Q such that

$$\|\tilde{\boldsymbol{v}}_i - \boldsymbol{v}_i\| \le Q\tau^\alpha.$$

Proof. The proof of this lemma is a generalization of the proof of Lemma 4.2 in [49], and is thus omitted. □

This lemma indicates that if a set of vectors before orthonormalization is almost orthonormal with the error of $O(\tau^\alpha)$, the orthonormalization will only apply an $O(\tau^\alpha)$ perturbation to orthonormalize them. In the rest of the work, we will present two second-order Runge–Kutta schemes, namely the midpoint method and the improved Euler method, incorporating the orthonormalization steps and then rigorously analyze the impact of orthonormalization on the convergence order of these schemes.

3. Two orthonormalized Runge–Kutta schemes

We propose and analyze two orthonormalized second-order Runge–Kutta methods for (4).

3.1. Orthonormalized midpoint formula

Given the numerical solution $\boldsymbol{v}_{i,n}$ at the nth time step, the numerical solution at the $(n+1)$th time step of (4) can be obtained using the following midpoint formula:

$$\begin{cases} \boldsymbol{v}_{i,n+1/2} = \boldsymbol{v}_{i,n} + \dfrac{1}{2}\tau \boldsymbol{g}_{i,n}, \\ \boldsymbol{v}_{i,n+1} = \boldsymbol{v}_{i,n} + \tau \boldsymbol{g}_{i,n+1/2}, \end{cases}$$

where $\boldsymbol{v}_{i,n+1/2}$ is an intermediate variable approximating the value of the solution at the half time step $\boldsymbol{v}_i(t_{n+1/2})$, $t_{n+1/2} = (t_n + t_{n+1})/2$, and

$$\boldsymbol{g}_{i,n} = \left(\mathbb{I} - \boldsymbol{v}_{i,n}\boldsymbol{v}_{i,n}^\top - 2\sum_{j=1}^{i-1} \boldsymbol{v}_{j,n}\boldsymbol{v}_{j,n}^\top\right)\mathbb{M}\boldsymbol{v}_{i,n}$$

where $\mathbb{M} = \mathbb{G}$ or $\mathbb{J}^\top + \mathbb{J}$.

Then the orthonormalized midpoint formula (oRK2-I) for (4) is as follows:

$$\begin{cases} \tilde{\boldsymbol{v}}_{i,n+1/2} = \boldsymbol{v}_{i,n} + \dfrac{1}{2}\tau \boldsymbol{g}_{i,n}, \\ \{\boldsymbol{v}_{i,n+1/2}\}_{i=1}^k = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+1/2}\}_{i=1}^k), \\ \tilde{\boldsymbol{v}}_{i,n+1} = \boldsymbol{v}_{i,n} + \tau \boldsymbol{g}_{i,n+1/2}, \\ \{\boldsymbol{v}_{i,n+1}\}_{i=1}^k = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+1}\}_{i=1}^k). \end{cases} \tag{8}$$

In this case, the numerical solution at each step satisfies the orthogonality condition $\boldsymbol{v}_{i,n}^\top \boldsymbol{v}_{j,n} = \delta_{i,j}$.

We analyze the orthonormalization error of the intermediate variables.

Lemma 2. *In the oRK2-I, the numerical solution for (4) before orthonormalization satisfies the following conditions for $1 \le n \le N-1$:*

$$\left|\tilde{\boldsymbol{v}}_{i,n+1/2}^\top \tilde{\boldsymbol{v}}_{j,n+1/2}^\top - \delta_{i,j}\right| \le C\tau^2,$$
$$\left|\tilde{\boldsymbol{v}}_{i,n+1}^\top \tilde{\boldsymbol{v}}_{j,n+1} - \delta_{i,j}\right| \le C\tau^3.$$

Here and hereafter, C denotes a positive constant that is independent of τ, n, and N.

Proof. By the orthonormality of $\boldsymbol{v}_{i,n}$, which implies

$$\begin{aligned} \boldsymbol{v}_{i,n}^\top \boldsymbol{g}_{j,n} + \boldsymbol{v}_{j,n}^\top \boldsymbol{g}_{i,n} &= \boldsymbol{v}_{i,n}^\top \left(\mathbb{I} - \boldsymbol{v}_j\boldsymbol{v}_j^\top - 2\sum_{k=1}^{j-1} \boldsymbol{v}_k\boldsymbol{v}_k^\top\right)\mathbb{M}(\boldsymbol{x})\boldsymbol{v}_j + \boldsymbol{v}_{j,n}^\top \left(\mathbb{I} - \boldsymbol{v}_i\boldsymbol{v}_i^\top - 2\sum_{j=1}^{i-1} \boldsymbol{v}_j\boldsymbol{v}_j^\top\right)\mathbb{M}(\boldsymbol{x})\boldsymbol{v}_i \\ &= 0. \end{aligned}$$

Then we get

$$\begin{aligned} \tilde{\boldsymbol{v}}_{i,n+1/2}^\top \tilde{\boldsymbol{v}}_{j,n+1/2} &= (\boldsymbol{v}_{i,n} + \frac{1}{2}\tau \boldsymbol{g}_{i,n})^\top (\boldsymbol{v}_{j,n} + \frac{1}{2}\tau \boldsymbol{g}_{j,n}) \\ &= \boldsymbol{v}_{i,n}^\top \boldsymbol{v}_{j,n} + \frac{1}{2}\tau \boldsymbol{v}_{i,n}^\top \boldsymbol{g}_{j,n} + \frac{1}{2}\tau \boldsymbol{v}_{j,n}^\top \boldsymbol{g}_{i,n} + \frac{1}{4}\tau^2 \boldsymbol{g}_{i,n}^\top \boldsymbol{g}_{j,n} \\ &= \delta_{i,j} + \frac{1}{4}\tau^2 \boldsymbol{g}_{i,n}^\top \boldsymbol{g}_{j,n}. \end{aligned}$$

Thus, we obtain the first result in the lemma

$$\left|\tilde{\boldsymbol{v}}_{i,n+1/2}^{\top}\tilde{\boldsymbol{v}}_{j,n+1/2}^{\top}-\delta_{i,j}\right|\leq C\tau^2.$$

According to Lemma 1, we have

$$\|\tilde{\boldsymbol{v}}_{i,n+1/2}-\boldsymbol{v}_{i,n+1/2}\|\leq C\tau^2. \tag{9}$$

On the other hand, in the numerical scheme (8), we have

$$\begin{aligned}\tilde{\boldsymbol{v}}_{i,n+1}^{\top}\tilde{\boldsymbol{v}}_{j,n+1}&=(\boldsymbol{v}_{i,n}+\tau\boldsymbol{g}_{i,n+1/2})^{\top}(\boldsymbol{v}_{j,n}+\tau\boldsymbol{g}_{j,n+1/2})\\&=\boldsymbol{v}_{i,n}^{\top}\boldsymbol{v}_{j,n}+\tau\boldsymbol{v}_{i,n}^{\top}\boldsymbol{g}_{j,n+1/2}+\tau\boldsymbol{v}_{j,n}^{\top}\boldsymbol{g}_{i,n+1/2}+\tau^2\boldsymbol{g}_{i,n+1/2}^{\top}\boldsymbol{g}_{j,n+1/2}.\end{aligned} \tag{10}$$

According to the orthonormality of $\boldsymbol{v}_{i,n+1/2}$, the numerical solution at the half step satisfies

$$\boldsymbol{v}_{i,n+1/2}^{\top}\boldsymbol{g}_{j,n+1/2}+\boldsymbol{v}_{j,n+1/2}^{\top}\boldsymbol{g}_{i,n+1/2}=0.$$

Thus, using (9), we can estimate the cross terms in (10) as

$$\begin{aligned}\boldsymbol{v}_{i,n}^{\top}\boldsymbol{g}_{j,n+1/2}+\boldsymbol{v}_{j,n}^{\top}\boldsymbol{g}_{i,n+1/2}=&(\boldsymbol{v}_{i,n}-\boldsymbol{v}_{i,n+1/2})^{\top}\boldsymbol{g}_{j,n+1/2}+(\boldsymbol{v}_{j,n}-\boldsymbol{v}_{j,n+1/2})^{\top}\boldsymbol{g}_{i,n+1/2}\\=&(\boldsymbol{v}_{i,n}-\tilde{\boldsymbol{v}}_{i,n+1/2})^{\top}\boldsymbol{g}_{j,n+1/2}+(\boldsymbol{v}_{j,n}-\tilde{\boldsymbol{v}}_{j,n+1/2})^{\top}\boldsymbol{g}_{i,n+1/2}+O(\tau^2).\end{aligned} \tag{11}$$

In the numerical scheme (8), we have

$$\boldsymbol{v}_{i,n}-\tilde{\boldsymbol{v}}_{i,n+1/2}=-\frac{1}{2}\tau\boldsymbol{g}_{i,n}. \tag{12}$$

Then we get

$$\begin{aligned}&(\boldsymbol{v}_{i,n}-\tilde{\boldsymbol{v}}_{i,n+1/2})^{\top}\boldsymbol{g}_{j,n+1/2}+(\boldsymbol{v}_{j,n}-\tilde{\boldsymbol{v}}_{j,n+1/2})^{\top}\boldsymbol{g}_{i,n+1/2}\\=&-\frac{1}{2}\tau(\boldsymbol{g}_{i,n}^{\top}\boldsymbol{g}_{j,n+1/2}+\boldsymbol{g}_{j,n}^{\top}\boldsymbol{g}_{i,n+1/2}).\end{aligned} \tag{13}$$

Using (10), (11) and (13), we have

$$\begin{aligned}&\tilde{\boldsymbol{v}}_{i,n+1}^{\top}\tilde{\boldsymbol{v}}_{j,n+1}\\=&\boldsymbol{v}_{i,n}^{\top}\boldsymbol{v}_{j,n}-\frac{1}{2}\tau^2(\boldsymbol{g}_{i,n}^{\top}\boldsymbol{g}_{j,n+1/2}+\boldsymbol{g}_{j,n}^{\top}\boldsymbol{g}_{i,n+1/2})+\tau^2\boldsymbol{g}_{i,n+1/2}^{\top}\boldsymbol{g}_{j,n+1/2}+O(\tau^3)\\=&\delta_{i,j}+\frac{1}{2}\tau^2(\boldsymbol{g}_{i,n+1/2}-\boldsymbol{g}_{i,n})^{\top}\boldsymbol{g}_{j,n+1/2}+\frac{1}{2}\tau^2(\boldsymbol{g}_{j,n+1/2}-\boldsymbol{g}_{j,n})^{\top}\boldsymbol{g}_{i,n+1/2}+O(\tau^3).\end{aligned} \tag{14}$$

Using the Lipschitz condition (5), (6), (9) and (12), we have

$$\begin{aligned}\|\boldsymbol{g}_{i,n+1/2}-\boldsymbol{g}_{i,n}\|&\leq C\max_{1\leq i\leq k}\|\boldsymbol{v}_{i,n+1/2}-\boldsymbol{v}_{i,n}\|\\&\leq C\max_{1\leq i\leq k}(\|\boldsymbol{v}_{i,n+1/2}-\tilde{\boldsymbol{v}}_{i,n+1/2}\|+\|\tilde{\boldsymbol{v}}_{i,n+1/2}-\boldsymbol{v}_{i,n}\|)\\&=C\max_{1\leq i\leq k}\|\boldsymbol{v}_{i,n+1/2}-\tilde{\boldsymbol{v}}_{i,n+1/2}\|+\frac{1}{2}C\tau\max_{1\leq i\leq k}\|\boldsymbol{g}_{i,n}\|\\&\leq C\tau^2+C\tau\leq C\tau.\end{aligned} \tag{15}$$

Using (15), (14) and (6), we obtain the estimate for the orthogonality error

$$\begin{aligned}&\left|\tilde{\boldsymbol{v}}_{i,n+1}^{\top}\tilde{\boldsymbol{v}}_{j,n+1}-\delta_{i,j}\right|\\\leq&\frac{1}{2}\tau^2\|\boldsymbol{g}_{i,n+1/2}-\boldsymbol{g}_{i,n}\|\|\boldsymbol{g}_{j,n+1/2}\|+\frac{1}{2}\tau^2\|\boldsymbol{g}_{j,n+1/2}-\boldsymbol{g}_{j,n}\|\|\boldsymbol{g}_{i,n+1/2}\|+C\tau^3\\\leq&C\tau^3\|\boldsymbol{g}_{j,n+1/2}\|+C\tau^3\|\boldsymbol{g}_{i,n+1/2}\|+C\tau^3\leq C\tau^3,\end{aligned}$$

which proves the second result in the lemma. □

For the numerical solution $\boldsymbol{v}_{i,n}$ obtained through the oRK2-I (8), define the error of the numerical solution as

$$\begin{aligned}\boldsymbol{e}_{i,n+1/2}&=\boldsymbol{v}_{i,n+1/2}-\boldsymbol{v}_i(t_{n+1/2}),\\\boldsymbol{e}_{i,n}&=\boldsymbol{v}_{i,n}-\boldsymbol{v}_i(t_n).\end{aligned}$$

Using the perturbation introduced by orthonormalization, we can bound the error of the scheme as follows.

Lemma 3. *For sufficiently smooth $\{\boldsymbol{v}_i\}_{i=1}^k$ in (4), the error $\boldsymbol{e}_{i,n+1}$ in the oRK2-I (8) satisfies the following inequality:*

$$\max_{1\leq i\leq k}\|\boldsymbol{e}_{i,n+1}\|\leq\left(1+C\tau+C\tau^2\right)\max_{1\leq i\leq k}\|\boldsymbol{e}_{i,n}\|+C\tau^3,\quad i=1,2,\ldots,k,\quad 1\leq n\leq N-1. \tag{16}$$

Proof. For the continuous Eqs. (4), given sufficient smoothness, we can use Taylor expansion to obtain

$$\begin{aligned}\boldsymbol{v}_i(t_{n+1/2})&=\boldsymbol{v}_i(t_n)+\frac{1}{2}\tau\boldsymbol{g}_i(\{\boldsymbol{v}_i(t_n)\}_{i=1}^k)+O(\tau^2),\\\boldsymbol{v}_i(t_{n+1})&=\boldsymbol{v}_i(t_n)+\tau\boldsymbol{g}_i(\{\boldsymbol{v}_i(t_{n+1/2})\}_{i=1}^k)+O(\tau^3).\end{aligned} \tag{17}$$

Subtracting the expanded results (17) from the numerical scheme (8), we obtain

$$\begin{aligned}\tilde{\boldsymbol{v}}_{i,n+1/2}-\boldsymbol{v}_i(t_{n+1/2})&=\boldsymbol{e}_{i,n}+\frac{1}{2}\tau(\boldsymbol{g}_{i,n}-\boldsymbol{g}_i(\{\boldsymbol{v}_i(t_n)\}_{i=1}^k))+O(\tau^2),\\ \tilde{\boldsymbol{v}}_{i,n+1}-\boldsymbol{v}_i(t_{n+1})&=\boldsymbol{e}_{i,n}+\tau(\boldsymbol{g}_{i,n+1/2}-\boldsymbol{g}_i(\{\boldsymbol{v}_i(t_{n+1/2})\}_{i=1}^k))+O(\tau^3).\end{aligned}$$

Taking the norm on both sides, we get

$$\begin{aligned}\|\tilde{\boldsymbol{v}}_{i,n+1/2}-\boldsymbol{v}_i(t_{n+1/2})\|&\le\|\boldsymbol{e}_{i,n}\|+\frac{1}{2}\tau\|\boldsymbol{g}_{i,n}-\boldsymbol{g}_i(\{\boldsymbol{v}_i(t_n)\}_{i=1}^k)\|+O(\tau^2),\\ \|\tilde{\boldsymbol{v}}_{i,n+1}-\boldsymbol{v}_i(t_{n+1})\|&\le\|\boldsymbol{e}_{i,n}\|+\tau\|\boldsymbol{g}_{i,n+1/2}-\boldsymbol{g}_i(\{\boldsymbol{v}_i(t_{n+1/2})\}_{i=1}^k)\|+O(\tau^3).\end{aligned}\tag{18}$$

Using the Lipschitz condition of $\boldsymbol{g}_i$ in (5), we have

$$\begin{aligned}\|\boldsymbol{g}_{i,n}-\boldsymbol{g}(\{\boldsymbol{v}_i(t_n)\}_{i=1}^k)\|&\le L_g\max_{1\le i\le k}\|\boldsymbol{v}_{i,n}-\boldsymbol{v}_i(t_n)\|\\ &=L_g\max_{1\le i\le k}\|\boldsymbol{e}_{i,n}\|,\\ \|\boldsymbol{g}_{i,n+1/2}-\boldsymbol{g}(\{\boldsymbol{v}_i(t_{n+1/2})\}_{i=1}^k)\|&\le L_g\max_{1\le i\le k}\|\boldsymbol{v}_{i,n+1/2}-\boldsymbol{v}_i(t_{n+1/2})\|\\ &=L_g\max_{1\le i\le k}\|\boldsymbol{e}_{i,n+1/2}\|,\end{aligned}\tag{19}$$

Substituting (19) into (18), we obtain

$$\begin{aligned}\|\tilde{\boldsymbol{v}}_{i,n+1/2}-\boldsymbol{v}_i(t_{n+1/2})\|&\le\|\boldsymbol{e}_{i,n}\|+\frac{1}{2}L_g\tau\max_{1\le i\le k}\|\boldsymbol{e}_{i,n}\|+O(\tau^2),\\ \|\tilde{\boldsymbol{v}}_{i,n+1}-\boldsymbol{v}_i(t_{n+1})\|&\le\|\boldsymbol{e}_{i,n}\|+L_g\tau\max_{1\le i\le k}\|\boldsymbol{e}_{i,n+1/2}\|+O(\tau^3).\end{aligned}\tag{20}$$

Additionally, from Lemmas 1 and 2, we have

$$\begin{aligned}\|\tilde{\boldsymbol{v}}_{i,n+1/2}-\boldsymbol{v}_{i,n+1/2}\|&\le C\tau^2,\\ \|\tilde{\boldsymbol{v}}_{i,n+1}-\boldsymbol{v}_{i,n+1}\|&\le C\tau^3.\end{aligned}$$

That is,

$$\begin{aligned}\|\boldsymbol{e}_{i,n+1/2}\|&\le\|\tilde{\boldsymbol{v}}_{i,n+1/2}-\boldsymbol{v}(t_{i,n+1/2})\|+C\tau^2,\\ \|\boldsymbol{e}_{i,n+1}\|&\le\|\tilde{\boldsymbol{v}}_{i,n+1}-\boldsymbol{v}(t_{i,n+1})\|+C\tau^3.\end{aligned}\tag{21}$$

Substituting (20) into (21), we get

$$\begin{aligned}\|\boldsymbol{e}_{i,n+1/2}\|&\le\|\boldsymbol{e}_{i,n}\|+C\tau\max_{1\le i\le k}\|\boldsymbol{e}_{i,n}\|+C\tau^2,\\ \|\boldsymbol{e}_{i,n+1}\|&\le\|\boldsymbol{e}_{i,n}\|+C\tau\max_{1\le i\le k}\|\boldsymbol{e}_{i,n+1/2}\|+C\tau^3.\end{aligned}$$

Combining the above two inequalities and eliminating the intermediate variable, we obtain

$$\max_{1\le i\le k}\|\boldsymbol{e}_{i,n+1}\|\le\max_{1\le i\le k}\|\boldsymbol{e}_{i,n}\|+C\tau\max_{1\le i\le k}\|\boldsymbol{e}_{i,n}\|+C\tau^2\max_{1\le i\le k}\|\boldsymbol{e}_{i,n}\|+C\tau^3,$$

that is,

$$\max_{1\le i\le k}\|\boldsymbol{e}_{i,n+1}\|\le\left(1+C\tau+C\tau^2\right)\max_{1\le i\le k}\|\boldsymbol{e}_{i,n}\|+C\tau^3.$$

Thus, we complete the proof. □

After obtaining the above result, we can use the Gronwall inequality to obtain the error estimate.

Theorem 1. *For sufficiently smooth $\{\boldsymbol{v}_i\}_{i=1}^k$ in (4), the error in the oRK2-I (8) is second-order. That is,*

$$\max_{1\le i\le k}\|\boldsymbol{e}_{i,n}\|\le C\tau^2,\quad 1\le n\le N.$$

Proof. For any $1\le n_*\le N-1$, summing Eq. (16) in Lemma 3 from 1 to n_* gives

$$\sum_{n=1}^{n_*}\max_{1\le i\le k}\|\boldsymbol{e}_{i,n+1}\|\le\sum_{n=1}^{n_*}\left(1+C\tau+C\tau^2\right)\max_{1\le i\le k}\|\boldsymbol{e}_{i,n}\|+C\sum_{n=1}^{n_*}\tau^3,\quad i=1,2,\dots,k.$$

Further calculations yield

$$\max_{1\le i\le k}\|\boldsymbol{e}_{i,n_*+1}\|\le\max_{1\le i\le k}\|\boldsymbol{e}_{i,1}\|+\sum_{n=1}^{n_*}\left(C\tau+C\tau^2\right)\max_{1\le i\le k}\|\boldsymbol{e}_{i,n}\|+C\tau^2,\quad i=1,2,\dots,k.$$

Similar to the derivation in (9), $\max_{1\le i\le k}\|\boldsymbol{e}_{i,1}\|\le C\tau^2$ can be easily obtained. By combining $\max_{1\le i\le k}\|\boldsymbol{e}_{i,1}\|\le C\tau^2$ and using discrete Gronwall inequality, we can then derive

$$\max_{1\le i\le k}\|\boldsymbol{e}_{i,n_*+1}\|\le C\tau^2\exp\Big(\sum_{n=1}^{n_*}\left(C\tau+C\tau^2\right)\Big)\le C\tau^2,\quad i=1,2,\dots,k.$$

Due to the arbitrariness of n_*, the theorem is proved. □

For the usual midpoint method, the accuracy is second order. The above conclusion shows that for the oRK2-I, the order of error remains second order. This theoretically ensures that the orthonormalization operation does not affect the overall error order of the midpoint method.

3.2. Orthonormalized improved Euler formula

Given the numerical solution at the nth time step, the numerical solution at the $(n+1)$th time step for (4) can be computed using the following improved Euler formula:

$$\begin{cases} \boldsymbol{v}_{i,n+*} = \boldsymbol{v}_{i,n} + \tau \boldsymbol{g}_{i,n}, \\ \boldsymbol{v}_{i,n+1} = \boldsymbol{v}_{i,n} + \frac{1}{2}\tau(\boldsymbol{g}_{i,n} + \boldsymbol{g}_{i,n+*}), \end{cases}$$

The core idea of the improved Euler formula is the predictor–corrector method. The orthonormalized improved Euler numerical scheme (oRK2-II) of (4) is then given by:

$$\begin{cases} \tilde{\boldsymbol{v}}_{i,n+*} = \boldsymbol{v}_{i,n} + \tau \boldsymbol{g}_{i,n}, \\ \{\boldsymbol{v}_{i,n+*}\}_{i=1}^{k} = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+*}\}_{i=1}^{k}), \\ \tilde{\boldsymbol{v}}_{i,n+1} = \boldsymbol{v}_{i,n} + \frac{1}{2}\tau(\boldsymbol{g}_{i,n} + \boldsymbol{g}_{i,n+*}), \\ \{\boldsymbol{v}_{i,n+1}\}_{i=1}^{k} = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+1}\}_{i=1}^{k}). \end{cases} \tag{22}$$

We analyze the orthonormalization error for the intermediate variables in the following lemma.

Lemma 4. *In the oRK2-II (22), the numerical solutions of (4) before orthonormalization satisfy the following conditions for* $1 \le n \le N-1$*:*

$$\left|\tilde{\boldsymbol{v}}_{i,n+*}^{\top}\tilde{\boldsymbol{v}}_{j,n+*} - \delta_{i,j}\right| \le C\tau^2,$$
$$\left|\tilde{\boldsymbol{v}}_{i,n+1}^{\top}\tilde{\boldsymbol{v}}_{j,n+1} - \delta_{i,j}\right| \le C\tau^3.$$

Proof. For the intermediate variables, the first step of the oRK2-II is equivalent to using the forward Euler method. According to conclusions from the forward Euler method and the orthonormality of $\boldsymbol{v}_{i,n}$, we have

$$\begin{aligned} \tilde{\boldsymbol{v}}_{i,n+*}^{\top}\tilde{\boldsymbol{v}}_{j,n+*} &= (\boldsymbol{v}_{i,n} + \tau \boldsymbol{g}_{i,n})^{\top}(\boldsymbol{v}_{j,n} + \tau \boldsymbol{g}_{j,n}) \\ &= \boldsymbol{v}_{i,n}^{\top}\boldsymbol{v}_{j,n} + \tau \boldsymbol{v}_{i,n}^{\top}\boldsymbol{g}_{j,n} + \tau \boldsymbol{v}_{j,n}^{\top}\boldsymbol{g}_{i,n} + \tau^2 \boldsymbol{g}_{i,n}^{\top}\boldsymbol{g}_{j,n} \\ &= \delta_{i,j} + \tau^2 \boldsymbol{g}_{i,n}^{\top}\boldsymbol{g}_{j,n}. \end{aligned}$$

Thus, we obtain

$$\left|\tilde{\boldsymbol{v}}_{i,n+*}^{\top}\tilde{\boldsymbol{v}}_{j,n+*} - \delta_{i,j}\right| \le C\tau^2.$$

This proves the first part of the lemma.

According to Lemma 1, we have

$$\|\tilde{\boldsymbol{v}}_{i,n+*} - \boldsymbol{v}_{i,n+*}\| \le C\tau^2. \tag{23}$$

For the numerical scheme (22), we have

$$\begin{aligned} &\tilde{\boldsymbol{v}}_{i,n+1}^{\top}\tilde{\boldsymbol{v}}_{j,n+1} \\ =&(\boldsymbol{v}_{i,n} + \frac{1}{2}\tau(\boldsymbol{g}_{i,n} + \boldsymbol{g}_{i,n+*}))^{\top}(\boldsymbol{v}_{j,n} + \frac{1}{2}\tau(\boldsymbol{g}_{i,n} + \boldsymbol{g}_{i,n+*})) \\ =&\boldsymbol{v}_{i,n}^{\top}\boldsymbol{v}_{j,n} + \frac{1}{2}\tau \boldsymbol{v}_{i,n}^{\top}(\boldsymbol{g}_{j,n} + \boldsymbol{g}_{j,n+*}) + \frac{1}{2}\tau \boldsymbol{v}_{j,n}^{\top}(\boldsymbol{g}_{i,n} + \boldsymbol{g}_{i,n+*}) + \frac{1}{4}\tau^2(\boldsymbol{g}_{i,n} + \boldsymbol{g}_{i,n+*})^{\top}(\boldsymbol{g}_{j,n} + \boldsymbol{g}_{j,n+*}) \\ =&\boldsymbol{v}_{i,n}^{\top}\boldsymbol{v}_{j,n} + \frac{1}{2}\tau \boldsymbol{v}_{i,n}^{\top}\boldsymbol{g}_{j,n+*} + \frac{1}{2}\tau \boldsymbol{v}_{j,n}^{\top}\boldsymbol{g}_{i,n+*} + \frac{1}{4}\tau^2(\boldsymbol{g}_{i,n} + \boldsymbol{g}_{i,n+*})^{\top}(\boldsymbol{g}_{j,n} + \boldsymbol{g}_{j,n+*}). \end{aligned} \tag{24}$$

According to the orthonormality of $\boldsymbol{v}_{i,n+*}$, the numerical solution at the predicted value satisfies

$$\boldsymbol{v}_{i,n+*}^{\top}\boldsymbol{g}_{j,n+*} + \boldsymbol{v}_{j,n+*}^{\top}\boldsymbol{g}_{i,n+*} = 0. \tag{25}$$

Thus using (23), we obtain:

$$\boldsymbol{v}_{i,n}^{\top}\boldsymbol{g}_{j,n+*} + \boldsymbol{v}_{j,n}^{\top}\boldsymbol{g}_{i,n+*} = (\boldsymbol{v}_{i,n} - \tilde{\boldsymbol{v}}_{i,n+*})^{\top}\boldsymbol{g}_{j,n+*} + (\boldsymbol{v}_{j,n} - \tilde{\boldsymbol{v}}_{j,n+*})^{\top}\boldsymbol{g}_{i,n+*} + O(\tau^2). \tag{26}$$

In the numerical scheme (22), we have

$$\boldsymbol{v}_{i,n} - \tilde{\boldsymbol{v}}_{i,n+*} = -\tau \boldsymbol{g}_{i,n}, \tag{27}$$

Substituting (27) into (26), we get

$$\boldsymbol{v}_{i,n}^{\top}\boldsymbol{g}_{j,n+*} + \boldsymbol{v}_{j,n}^{\top}\boldsymbol{g}_{i,n+*} = -\tau \boldsymbol{g}_{i,n}^{\top}\boldsymbol{g}_{j,n+*} - \tau \boldsymbol{g}_{j,n}^{\top}\boldsymbol{g}_{i,n+*} + O(\tau^2). \tag{28}$$

Substituting (28) into (24), we have

$$\begin{aligned}
&\tilde{\boldsymbol{v}}_{i,n+1}^{\top}\tilde{\boldsymbol{v}}_{j,n+1}\\
=&\boldsymbol{v}_{i,n}^{\top}\boldsymbol{v}_{j,n}-\frac{1}{2}\tau^2\boldsymbol{g}_{i,n}^{\top}\boldsymbol{g}_{j,n+*}-\frac{1}{2}\tau^2\boldsymbol{g}_{j,n}^{\top}\boldsymbol{g}_{i,n+*}\\
&+\frac{1}{4}\tau^2(\boldsymbol{g}_{i,n}+\boldsymbol{g}_{i,n+*})^{\top}(\boldsymbol{g}_{j,n}+\boldsymbol{g}_{j,n+*})+O(\tau^3)\\
=&\delta_{i,j}+\frac{1}{4}\tau^2(\boldsymbol{g}_{i,n+*}-\boldsymbol{g}_{i,n})^{\top}(\boldsymbol{g}_{j,n+*}-\boldsymbol{g}_{j,n})+O(\tau^3).
\end{aligned}\tag{29}$$

Using the Lipschitz condition (5) and boundedness of $\boldsymbol{g}_{i,n}$ (6), combined with (23) and (27), we have:

$$\begin{aligned}
\|\boldsymbol{g}_{i,n+*}-\boldsymbol{g}_{i,n}\|&\le C\max_{1\le i\le k}\|\boldsymbol{v}_{i,n+*}-\boldsymbol{v}_{i,n}\|\\
&\le C\max_{1\le i\le k}\|\boldsymbol{v}_{i,n+*}-\tilde{\boldsymbol{v}}_{i,n+*}\|+C\max_{1\le i\le k}\|\tilde{\boldsymbol{v}}_{i,n+*}-\boldsymbol{v}_{i,n}\|\\
&= C\max_{1\le i\le k}\|\boldsymbol{v}_{i,n+*}-\tilde{\boldsymbol{v}}_{i,n+*}\|+\tau C\max_{1\le i\le k}\|\boldsymbol{g}_{i,n}\|\\
&\le C\tau^2+C\tau.
\end{aligned}\tag{30}$$

Taking the norm on both sides of (29) and using (30), we obtain

$$\left|\tilde{\boldsymbol{v}}_{i,n+1}^{\top}\tilde{\boldsymbol{v}}_{j,n+1}-\delta_{i,j}\right|\le\frac{1}{4}\tau^2\|\boldsymbol{g}_{i,n+*}-\boldsymbol{g}_{i,n}\|\|\boldsymbol{g}_{j,n+*}-\boldsymbol{g}_{j,n}\|+C\tau^3\le C\tau^3.$$

This proves the second part of the lemma. □

Remark 1. It is worth mentioning that although the orthonormalization error $\left|\tilde{\boldsymbol{v}}_{i,n+1}^{\top}\tilde{\boldsymbol{v}}_{j,n+1}-\delta_{i,j}\right|$ of oRK2-II has only third-order accuracy according to the theoretical analysis, numerical experiments later on show that it achieves fourth-order accuracy. Such interesting superconvergence phenomenon deserves further investigation.

For the numerical solution $\boldsymbol{v}_{i,n}$ obtained through the numerical scheme, we define the error of the numerical solution as

$$\begin{aligned}
\boldsymbol{e}_{i,n+*}&=\boldsymbol{v}_{i,n+*}-\boldsymbol{v}_i(t_{n+1}),\\
\boldsymbol{e}_{i,n}&=\boldsymbol{v}_{i,n}-\boldsymbol{v}_i(t_n).
\end{aligned}$$

Similar to the derivations of Lemma 3 and Theorem 1, we could obtain analogous results for the oRK2-II (22) as follows (the proofs are thus omitted due to similarity).

Lemma 5. *For sufficiently smooth $\{\boldsymbol{v}_i\}_{i=1}^k$ in (4), the error $\boldsymbol{e}_{i,n+1}$ in the oRK2-II (22) satisfies the following inequality:*

$$\max_{1\le i\le k}\|\boldsymbol{e}_{i,n+1}\|\le(1+C\tau)\max_{1\le i\le k}\|\boldsymbol{e}_{i,n}\|+C\tau^3,\quad i=1,2,\dots,k,\quad 1\le n\le N-1.$$

Theorem 2. *For sufficiently smooth $\{\boldsymbol{v}_i\}_{i=1}^k$ in (4), the error in the oRK2-II (22) is second-order. That is,*

$$\max_{1\le i\le k}\|\boldsymbol{e}_{i,n}\|\le C\tau^2,\quad 1\le n\le N.$$

For the usual improved Euler formula, the scheme accuracy is second-order. The above conclusions indicate that for the oRK2-II, the error order of the scheme remains second-order. This theoretically ensures that the orthonormalization operation does not affect the overall error order of the improved Euler formula.

Remark 2. In the Appendix, we present the complete discrete schemes oRK2-I (31) and oRK2-II (32) for HiSD and GHiSD, as well as the third-order orthonormalized Runge–Kutta scheme (oRK3) (33) and the fourth-order orthonormalized Runge–Kutta scheme (oRK4) (34). To enhance clarity in our discussion, we focus solely on analyzing the errors of the discrete schemes oRK2-I and oRK2-II for the unified form of the second equation in HiSD and GHiSD. In fact, the error analysis for the variable $\boldsymbol{x}$ in the first equation is not fundamentally challenging. By referring to the analysis of $\boldsymbol{x}$ in [49], it can be shown that it has the same order of accuracy as $\boldsymbol{v}$. Under suitable assumptions on the data, we can follow the same procedure to analyze the errors of oRK3 and oRK4.

4. Numerical experiments

In this section, we test the accuracy of the discrete schemes oRK2-I, oRK2-II, oRK3, and oRK4 for HiSD and GHiSD, as well as the convergence paths of these discrete schemes. For comparison purposes, we also provided the numerical results in Euler scheme. To evaluate convergence, we define the following errors:

$$\begin{aligned}
\mathrm{Err}(x)&:=\max_{1\le n\le N_T}\|\boldsymbol{x}(t_n)-\boldsymbol{x}_n\|,\\
\mathrm{Err}(\boldsymbol{v}_i)&:=\max_{1\le n\le N_T}\|\boldsymbol{v}_i(t_n)-\boldsymbol{v}_{i,n}\|,\quad 1\le i\le k,\\
\mathrm{Err}(\tilde{\boldsymbol{v}}_i)&:=\max_{1\le n\le N_T}\left|\|\tilde{\boldsymbol{v}}_{i,n}\|-1\right|,\quad 1\le i\le k,
\end{aligned}$$

Table 1
Convergence rates of five different schemes in Example 1.

Scheme	τ	Err(x)	Conv. rate	Err(v_1)	Conv. rate	Err($\tilde{v}_1$)	Conv. rate
Euler	$1/2^3$	1.02E−00		3.70E−01		4.18E−02	
	$1/2^4$	6.25E−01	0.70	2.13E−01	0.80	1.37E−02	1.61
	$1/2^5$	3.43E−01	0.87	1.15E−01	0.89	4.73E−03	1.53
	$1/2^6$	1.78E−01	0.94	5.93E−02	0.96	1.31E−03	1.85
	$1/2^7$	9.03E−02	0.98	3.00E−02	0.98	3.42E−04	1.93
	$1/2^8$	4.53E−02	0.99	1.50E−02	1.00	8.75E−05	1.97
oRK2-I	$1/2^3$	3.41E−01		1.48E−01		2.95E−02	
	$1/2^4$	1.01E−01	1.75	3.92E−02	1.92	3.67E−03	3.01
	$1/2^5$	2.92E−02	1.80	1.09E−02	1.84	7.00E−04	2.39
	$1/2^6$	7.85E−03	1.90	2.90E−03	1.91	9.09E−05	2.95
	$1/2^7$	2.04E−03	1.95	7.46E−03	1.96	1.81E−05	2.95
	$1/2^8$	5.18E−04	1.97	1.90E−04	1.98	1.49E−06	2.99
oRK2-II	$1/2^3$	4.73E−01		1.76E−01		3.09E−02	
	$1/2^4$	1.69E−01	1.49	6.22E−02	1.50	2.65E−03	3.54
	$1/2^5$	5.06E−02	1.74	1.83E−02	1.76	1.46E−04	4.18
	$1/2^6$	1.40E−02	1.86	5.01E−03	1.87	8.16E−06	4.16
	$1/2^7$	3.67E−03	1.93	1.31E−03	1.93	5.59E−07	3.87
	$1/2^8$	9.44E−04	1.96	3.37E−04	1.96	3.72E−08	3.91
oRK3	$1/2^3$	2.64E−01		9.47E−02		1.36E−02	
	$1/2^4$	3.86E−02	2.77	1.31E−02	2.85	6.79E−04	4.33
	$1/2^5$	5.29E−03	2.87	1.79E−03	2.87	2.73E−05	4.64
	$1/2^6$	6.77E−04	2.96	2.29E−04	2.97	1.20E−06	4.51
	$1/2^7$	8.51E−05	2.99	2.88E−05	2.99	1.11E−07	3.43
	$1/2^8$	1.06E−05	3.00	3.60E−06	3.00	8.30E−09	3.74
oRK4	$1/2^3$	6.13E−02		2.64E−02		8.63E−03	
	$1/2^4$	5.58E−03	3.46	1.99E−03	3.73	2.71E−04	4.99
	$1/2^5$	4.47E−04	3.64	1.59E−04	3.65	8.54E−06	4.99
	$1/2^6$	3.15E−05	3.83	1.11E−05	3.83	3.35E−07	4.67
	$1/2^7$	2.09E−06	3.91	7.37E−07	3.92	1.13E−08	4.89
	$1/2^8$	1.34E−07	3.96	4.74E−08	3.96	3.62E−10	4.96

where $\boldsymbol{x}_n$, $\boldsymbol{v}_{i,n}$ and $\tilde{\boldsymbol{v}}_{i,n}$ are the numerical solutions at time t_n, while $\boldsymbol{x}(t_n)$ and $\boldsymbol{v}_i(t_n)$ are the exact solutions at time t_n. Since there are no exact solution for HiSD and GHiSD, the numerical solution calculated by $\tau = 2^{-20}$ is used as the reference solution, and for simplicity we set $\beta = \gamma = 1$. Unless otherwise specified, we always take $T = 1$.

Example 1 (*Accuracy Test under Eckhardt Surface*)**.** We consider the saddle dynamics for the Eckhardt surface [47]

$$E(x_1, x_2) = \exp(-x_1^2 - (x_2+1)^2) + \exp(-x_1^2 - (x_2-1)^2) + 4\exp\left(-3\frac{x_1^2+x_2^2}{2}\right) + \frac{x_2^2}{2}.$$

We set $\boldsymbol{x}_0 = (0.5, 0.7)^T$ with initial eigenvectors $\boldsymbol{v}_{1,0} = (1,2)^T/\sqrt{5}$, and select time steps $\tau = 1/2^3, 1/2^4, \ldots, 1/2^8$. Five discretization schemes are employed to discretize HiSD for calculating a index-1 saddle point. The numerical results are presented in

From the Table 1, it can be observed that, regarding Err($\boldsymbol{x}$) and Err($\boldsymbol{v}_1$), the Euler scheme achieves first-order accuracy, while oRK2-I and oRK2-II achieve second-order accuracy, oRK3 and oRK4 demonstrate third-order and fourth-order accuracy, respectively.

In terms of Err($\tilde{\boldsymbol{v}}_1$), the Euler scheme, oRK2-I, oRK3, and oRK4 achieve second-order, third-order, fourth-order, and fifth-order accuracy, respectively, with oRK2-II achieving fourth-order accuracy (i.e. the superconvergence phenomenon).

Next, we set $T = 8$ with a time step $\tau = 1/2^4$. Fig. 2 presents the convergence trajectory cloud plots for the five discretization schemes, with Fig. 2(b) showing the results for the Euler scheme at a time step $\tau = 1/2^7$. It can be observed that all schemes converge to the same index-1 saddle point.

Example 2 (*Three-dimensional Dynamical System*)**.** Consider a three-dimensional dynamical system [45]

$$\dot{\boldsymbol{x}} = -\begin{pmatrix} 0.6 & 0.1 & 0 \\ -0.1 & 0.6 & -0.05 \\ 0 & -0.1 & 0.6 \end{pmatrix}\boldsymbol{x} + 5\begin{pmatrix} (1+(x_1-5)^2)^{-1} \\ (1+(x_2-5)^2)^{-1} \\ (1+(x_3-5)^2)^{-1} \end{pmatrix}, \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3.$$

We employ five discretization schemes in GHiSD to locate a index-2 saddle point of this three-dimensional dynamical system. The initial point is set to $\boldsymbol{x}_0 = (3,4,5)^T$, with two orthogonal unit vectors $\boldsymbol{v}_{1,0} = (1,0,0)^T$ and $\boldsymbol{v}_{2,0} = (0,1,0)^T$ as the initial directions. We set $T = 8$ with a time step $\tau = 0.25$. Fig. 3 presents a 3D visualization of the computed trajectory, where all schemes eventually converge to an approximate index-2 saddle point at $\boldsymbol{x}^* = (4.1127, 3.2962, 5.7717)$. Using the solution from the Euler scheme with a time step of $\tau = 1/2^{20}$ as the reference solution, we observe that higher-order schemes exhibit smaller errors. Taking time steps $\tau = 1/2^3, 1/2^4, 1/2^5, \ldots, 1/2^7$, Figs. 4 and 5 illustrate the logarithmic relationship between Err($\boldsymbol{x}$), Err($\boldsymbol{v}_1$), and Err($\boldsymbol{v}_2$) and the

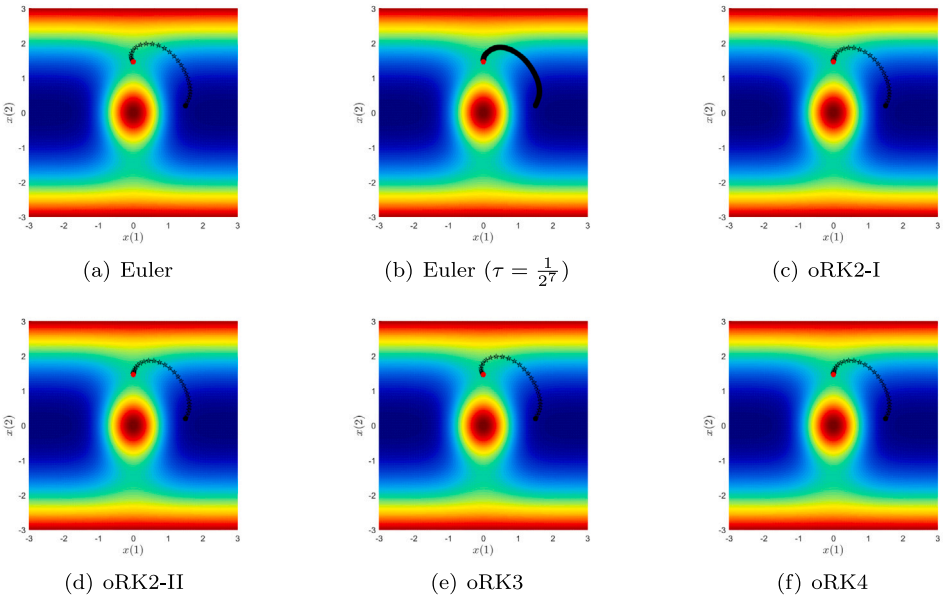


Fig. 2. The convergence trajectories under various schemes in Example 1.

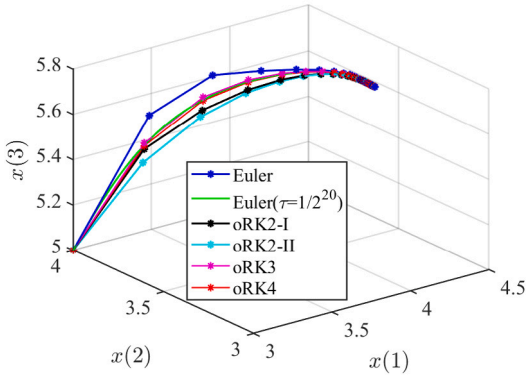


Fig. 3. The 3D convergence trajectories of Example 2 under different schemes.

reciprocal of the time step τ, indicating that the Euler scheme achieves first-order accuracy, oRK2-I and oRK2-II achieve second-order accuracy, while oRK3 and oRK4 reach third and fourth-order accuracy, respectively. Fig. 6 shows the logarithmic relationship of $\mathrm{Err}(\tilde{\boldsymbol{v}}_1)$ and $\mathrm{Err}(\tilde{\boldsymbol{v}}_2)$ with the reciprocal of the time step τ. The results indicate that the Euler scheme has second-order accuracy, oRK2-II achieves fourth-order accuracy, and oRK2-I, oRK3, and oRK4 achieve third, fourth, and fifth-order accuracy, respectively.

Example 3 (*Convergence of Dynamics under Modified Rosenbrock Function*)**.** In this study, we examine a high-dimensional modified Rosenbrock function defined as follows:

$$B_h(\boldsymbol{x}) = B(\boldsymbol{x}) - \sum_{i=1}^{h} s_i \arctan^2(x_i - x_i^*) + \sum_{j=h+1}^{d} s_j \arctan^2(x_j - x_j^*),$$

where $B(\boldsymbol{x})$ represents the d-dimensional Rosenbrock function given by

$$B(\boldsymbol{x}) = \sum_{i=1}^{d-1} [100(x_{i+1} - x_i^2)^2 + (1 - x_i)^2].$$

In our setup, we specify the dimension $d = 400$, set $h = 20$, with $s_i = 200$ for $i = 1, 2, \ldots, 20$, and $s_j = 1$ for $j = 21, 22, \ldots, d$. Here, $\boldsymbol{x}^* := [1, 1, \ldots, 1]^\top$ serves as an index-5 saddle point of $B_h(\boldsymbol{x})$.

We use five different discretization schemes of HiSD to compute the index-5 saddle point of $B_h(\boldsymbol{x})$. The initial point is set to $\boldsymbol{x}_0 = [1.05, 0.95, 1.05, 0.95, 1, 1, \ldots, 1]^\top$, and the initial set of eigenvectors is taken as the unit eigenvectors corresponding to the smallest

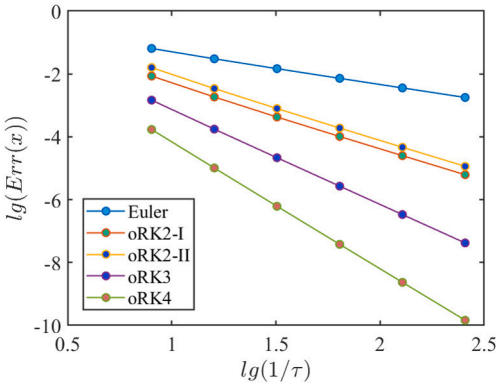


Fig. 4. The logarithmic plot of Err($\boldsymbol{x}$) with respect to the time step $1/\tau$ in Example 2.

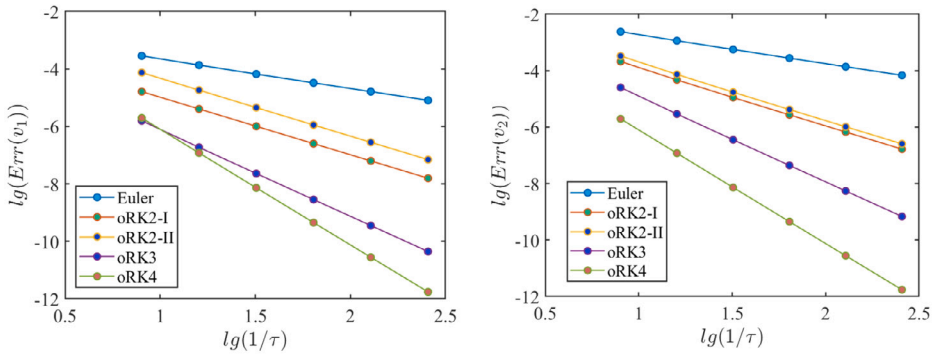


Fig. 5. The logarithmic plot of Err($\boldsymbol{v}_1$) and Err($\boldsymbol{v}_2$) against $1/\tau$ in Example 2.

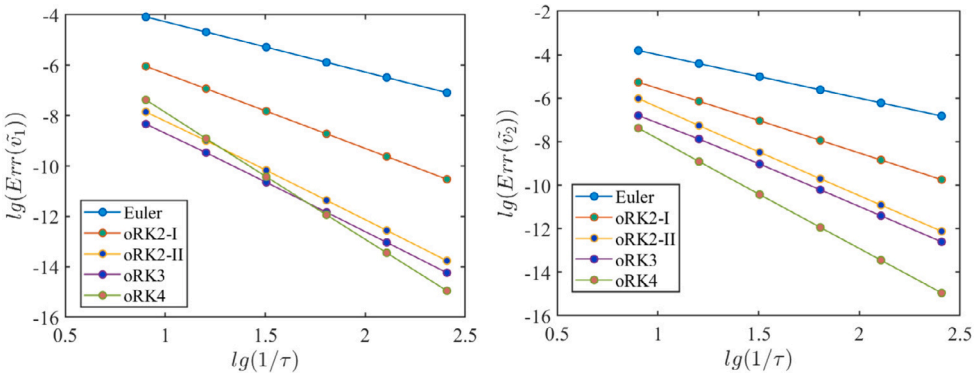


Fig. 6. The logarithmic plot of Err($\tilde{\boldsymbol{v}}_1$) and Err($\tilde{\boldsymbol{v}}_2$) against $1/\tau$ in Example 2.

five eigenvalues of the Hessian matrix of $B_h(\boldsymbol{x}_0)$. The time steps τ are set as $1/2^{12}$, $1/2^{13}$, …, $1/2^{17}$. From Table 2, we can see that regarding Err($\boldsymbol{x}$) and Err($\boldsymbol{v}_1$), the Euler scheme has only first-order accuracy, while oRK2-I and oRK2-II have second-order accuracy, oRK3 has third-order accuracy, and oRK4 has fourth-order accuracy. For the convergence of Err($\tilde{\boldsymbol{v}}_1$), the Euler scheme, oRK2-I, oRK3, and oRK4 have second-order, third-order, fourth-order, and fifth-order accuracy, respectively, while oRK2-II has fourth-order accuracy. This indicates that increasing the index of saddle points does not compromise the accuracy of oRK-type schemes. The table only lists the numerical results for Err($\boldsymbol{v}_1$) and Err($\tilde{\boldsymbol{v}}_1$); in fact, Err($\boldsymbol{v}_i$) and Err($\tilde{\boldsymbol{v}}_i$) ($i = 2, 3, 4, 5$) show similar results, which are omitted here.

Table 2
Convergence rates of five different schemes in Example 3.

Scheme	τ	Err($\boldsymbol{x}$)	Conv. rate	Err($\boldsymbol{v}_1$)	Conv. rate	Err($\tilde{\boldsymbol{v}}_1$)	Conv. rate
	$1/2^{12}$	6.24E−03		1.72E−03		9.09E−06	
	$1/2^{13}$	2.88E−03	1.12	7.97E−04	1.11	1.85E−06	2.29
Euler	$1/2^{14}$	1.38E−03	1.06	3.82E−04	1.06	4.23E−07	2.13
	$1/2^{15}$	6.69E−04	1.05	1.85E−04	1.05	1.01E−07	2.06
	$1/2^{16}$	3.21E−04	1.06	8.88E−05	1.06	2.48E−08	2.03
	$1/2^{17}$	1.49E−04	1.11	4.13E−05	1.11	6.14E−09	2.02
	$1/2^{12}$	7.08E−04		3.83E−04		1.31E−06	
	$1/2^{13}$	1.56E−04	2.18	7.64E−05	2.33	1.36E−07	3.27
oRK2-I	$1/2^{14}$	3.68E−05	2.09	1.72E−05	2.15	1.54E−08	3.14
	$1/2^{15}$	8.92E−06	2.04	4.08E−06	2.08	1.82E−09	3.08
	$1/2^{16}$	2.19E−06	2.03	9.88E−07	2.04	2.22E−10	3.03
	$1/2^{17}$	5.38E−07	2.03	2.41E−07	2.04	2.74E−11	3.02
	$1/2^{12}$	6.95E−04		3.82E−04		1.31E−06	
	$1/2^{13}$	1.54E−04	2.18	7.64E−05	2.32	8.20E−08	3.99
oRK2-II	$1/2^{14}$	3.63E−05	2.08	1.73E−05	2.15	5.14E−09	4.00
	$1/2^{15}$	8.81E−06	2.04	4.08E−06	2.08	3.21E−10	4.00
	$1/2^{16}$	2.17E−06	2.03	9.90E−07	2.04	2.01E−11	4.00
	$1/2^{17}$	5.31E−07	2.03	2.41E−07	2.04	1.26E−12	4.00
	$1/2^{12}$	5.49E−05		5.40E−05		2.29E−07	
	$1/2^{13}$	6.00E−06	3.19	5.56E−06	3.28	7.54E−09	4.92
oRK3	$1/2^{14}$	7.04E−07	3.09	6.30E−07	3.14	1.11E−09	2.77
	$1/2^{15}$	8.53E−08	3.05	7.49E−08	3.07	8.85E−11	3.65
	$1/2^{16}$	1.05E−08	3.02	9.14E−09	3.04	6.12E−12	3.85
	$1/2^{17}$	1.30E−09	3.01	1.13E−09	3.02	4.01E−13	3.93
	$1/2^{12}$	3.73E−06		5.26E−06		1.25E−07	
	$1/2^{13}$	2.04E−07	4.20	2.73E−07	4.27	4.22E−09	4.89
oRK4	$1/2^{14}$	1.19E−08	4.10	1.57E−08	4.13	1.39E−10	4.93
	$1/2^{15}$	7.20E−10	4.05	9.34E−10	4.07	4.47E−12	4.96
	$1/2^{16}$	4.42E−11	4.02	5.70E−11	4.03	1.41E−13	4.98
	$1/2^{17}$	2.74E−12	4.01	3.52E−12	4.02	4.22E−15	5.06

5. Conclusions

In this paper, we begin by identifying the common characteristics of the second equations in HiSD and GHiSD and abstract them into a unified form. Subsequently, we construct orthonormalized Runge–Kutta methods for HiSD and GHiSD and carefully analyze the error of the two second-order orthonormalized Runge–Kutta schemes. Finally, we verify the correctness of the theoretical results through several numerical experiments. The paper does not provide error estimates for the third-order and fourth-order orthonormalized Runge–Kutta schemes but presents numerical results. Under the assumption of a certain degree of smoothness in HiSD and GHiSD, the error estimates for the third-order and fourth-order orthonormalized Runge–Kutta schemes are expected to be similar.

CRediT authorship contribution statement

Shuai Miao: Writing – review & editing, Writing – original draft, Software, Conceptualization. **Lei Zhang:** Writing – review & editing, Writing – original draft, Conceptualization. **Pingwen Zhang:** Writing – review & editing, Writing – original draft, Conceptualization. **Xiangcheng Zheng:** Writing – review & editing, Writing – original draft, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (No. 12288101, 12225102, T2321001, 12301555, 12401506), the Taishan Scholars Program of Shandong Province (No. tsqn202306083), the National Key R&D Program of China (No. 2023YFA1008903), and the China Postdoctoral Science Foundation Grant (No. 2023M740103, GZB20230028).

Appendix. The orthonormalized Runge–Kutta schemes for HiSD and GHiSD

In this section, we present the second-order, third-order, fourth-order orthonormalized Runge–Kutta schemes for HiSD and GHiSD. To simplify the expressions, we denote $R(\{\boldsymbol{v}_j\}_{j=1}^{l}) = \mathbb{I} - 2\sum_{j=1}^{l}\boldsymbol{v}_j\boldsymbol{v}_j^{\top}$ for some $1 \le l \le k$. Given the initial condition $\boldsymbol{x}_0 = \boldsymbol{x}(0)$ and $\{\boldsymbol{v}_{i,0}\}_{i=1}^{k} = \{\boldsymbol{v}_i(0)\}_{i=1}^{k}$, the orthonormalized midpoint formula (oRK2-I) for HiSD is as follows:

$$\begin{cases} \boldsymbol{x}_{n+\frac{1}{2}} = \boldsymbol{x}_n + \frac{1}{2}\tau\beta R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_n), \\ \tilde{\boldsymbol{v}}_{i,n+\frac{1}{2}} = \boldsymbol{v}_{i,n} + \frac{1}{2}\tau\gamma(R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n}\boldsymbol{v}_{i,n}^{\top})\mathbb{G}(\boldsymbol{x}_n)\boldsymbol{v}_{i,n}, \quad 1 \le i \le k, \\ \{\boldsymbol{v}_{i,n+\frac{1}{2}}\}_{i=1}^{k} = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+\frac{1}{2}}\}_{i=1}^{k}), \\ \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \tau\beta R(\{\boldsymbol{v}_{j,n+\frac{1}{2}}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_{n+\frac{1}{2}}), \\ \tilde{\boldsymbol{v}}_{i,n+1} = \boldsymbol{v}_{i,n} + \tau\gamma(R(\{\boldsymbol{v}_{j,n+\frac{1}{2}}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n+\frac{1}{2}}\boldsymbol{v}_{i,n+\frac{1}{2}}^{\top})\mathbb{G}(\boldsymbol{x}_{n+\frac{1}{2}})\boldsymbol{v}_{i,n+\frac{1}{2}}, \quad 1 \le i \le k, \\ \{\boldsymbol{v}_{i,n+1}\}_{i=1}^{k} = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+1}\}_{i=1}^{k}). \end{cases} \tag{31}$$

With the same initial conditions as before, the orthonormalized improved Euler scheme (oRK2-II) for HiSD is as follows:

$$\begin{cases} \boldsymbol{x}_{n+*} = \boldsymbol{x}_n + \tau\beta R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_n), \\ \tilde{\boldsymbol{v}}_{i,n+*} = \boldsymbol{v}_{i,n} + \tau\gamma(R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n}\boldsymbol{v}_{i,n}^{\top})\mathbb{G}(\boldsymbol{x}_n)\boldsymbol{v}_{i,n}, \quad 1 \le i \le k, \\ \{\boldsymbol{v}_{i,n+*}\}_{i=1}^{k} = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+*}\}_{i=1}^{k}), \\ \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \frac{1}{2}\tau\beta[R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_n) + R(\{\boldsymbol{v}_{j,n+*}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_{n+*})], \\ \tilde{\boldsymbol{v}}_{i,n+1} = \boldsymbol{v}_{i,n} + \frac{1}{2}\tau\gamma[R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n}\boldsymbol{v}_{i,n}^{\top})\mathbb{G}(\boldsymbol{x}_n)\boldsymbol{v}_{i,n} + (R(\{\boldsymbol{v}_{j,n+*}\}_{j=1}^{i-1} - \boldsymbol{v}_{i,n+*}\boldsymbol{v}_{i,n+*}^{\top})\mathbb{G}(\boldsymbol{x}_{n+*})\boldsymbol{v}_{i,n+*}], \quad 1 \le i \le k, \\ \{\boldsymbol{v}_{i,n+1}\}_{i=1}^{k} = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+1}\}_{i=1}^{k}). \end{cases} \tag{32}$$

The orthonormalized third-order Runge–Kutta scheme (oRK3) for HiSD is as follows:

$$\begin{cases} \boldsymbol{x}_{n+\frac{1}{2}} = \boldsymbol{x}_n + \frac{1}{2}\tau\beta R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_n), \\ \tilde{\boldsymbol{v}}_{i,n+\frac{1}{2}} = \boldsymbol{v}_{i,n} + \frac{1}{2}\tau\gamma(R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{i-1})) - \boldsymbol{v}_{i,n}\boldsymbol{v}_{i,n}^{\top})\mathbb{G}(\boldsymbol{x}_n)\boldsymbol{v}_{i,n}, \quad 1 \le i \le k, \\ \{\boldsymbol{v}_{i,n+\frac{1}{2}}\}_{i=1}^{k} = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+\frac{1}{2}}\}_{i=1}^{k}), \\ \boldsymbol{x}_{n+*} = \boldsymbol{x}_n - \tau\beta R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_n) + 2\tau\beta R(\{\boldsymbol{v}_{j,n+\frac{1}{2}}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_{n+\frac{1}{2}}), \\ \tilde{\boldsymbol{v}}_{i,n+*} = \boldsymbol{v}_{i,n} - \tau\gamma(R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n}\boldsymbol{v}_{i,n}^{\top})\mathbb{G}(\boldsymbol{x}_n)\boldsymbol{v}_{i,n} + 2\tau\gamma(R(\{\boldsymbol{v}_{j,n+\frac{1}{2}}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n+\frac{1}{2}}\boldsymbol{v}_{i,n+\frac{1}{2}}^{\top})\mathbb{G}(\boldsymbol{x}_{n+\frac{1}{2}})\boldsymbol{v}_{i,n+\frac{1}{2}}, \quad 1 \le i \le k, \\ \{\boldsymbol{v}_{i,n+*}\}_{i=1}^{k} = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+*}\}_{i=1}^{k}), \\ \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \frac{1}{6}\tau\beta[R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_n) + 4R(\{\boldsymbol{v}_{j,n+\frac{1}{2}}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_{n+\frac{1}{2}}) + R(\{\boldsymbol{v}_{j,n+*}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_{n+*})], \\ \tilde{\boldsymbol{v}}_{i,n+1} = \boldsymbol{v}_{i,n} + \frac{1}{6}\tau\gamma[R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n}\boldsymbol{v}_{i,n}^{\top})\mathbb{G}(\boldsymbol{x}_n)\boldsymbol{v}_{i,n} + 4(R(\{\boldsymbol{v}_{j,n+\frac{1}{2}}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n+\frac{1}{2}}\boldsymbol{v}_{i,n+\frac{1}{2}}^{\top})\mathbb{G}(\boldsymbol{x}_{n+\frac{1}{2}})\boldsymbol{v}_{i,n+\frac{1}{2}} \\ +(R(\{\boldsymbol{v}_{j,n+*}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n+*}\boldsymbol{v}_{i,n+*}^{\top})\mathbb{G}(\boldsymbol{x}_{n+*})\boldsymbol{v}_{i,n+*}], \quad 1 \le i \le k, \\ \{\boldsymbol{v}_{i,n+1}\}_{i=1}^{k} = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+1}\}_{i=1}^{k}). \end{cases} \tag{33}$$

The orthonormalized fourth-order Runge–Kutta scheme (oRK4) for HiSD is as follows:

$$\begin{cases} \boldsymbol{x}_{n+\frac{1}{2}} = \boldsymbol{x}_n + \frac{1}{2}\tau\beta R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_n), \\ \tilde{\boldsymbol{v}}_{i,n+\frac{1}{2}} = \boldsymbol{v}_{i,n} + \frac{1}{2}\tau\gamma(R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n}\boldsymbol{v}_{i,n}^{\top})\mathbb{G}(\boldsymbol{x}_n)\boldsymbol{v}_{i,n}, \ 1 \le i \le k, \\ \{\boldsymbol{v}_{i,n+\frac{1}{2}}\}_{i=1}^{k} = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+\frac{1}{2}}\}_{i=1}^{k}), \\ \boldsymbol{x}_{n+*} = \boldsymbol{x}_n + \frac{1}{2}\tau\beta R(\{\boldsymbol{v}_{j,n+\frac{1}{2}}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_{n+\frac{1}{2}}), \\ \tilde{\boldsymbol{v}}_{i,n+*} = \boldsymbol{v}_{i,n} + \frac{1}{2}\tau\gamma(R(\{\boldsymbol{v}_{j,n+\frac{1}{2}}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n+\frac{1}{2}}\boldsymbol{v}_{i,n+\frac{1}{2}}^{\top})\mathbb{G}(\boldsymbol{x}_{n+\frac{1}{2}})\boldsymbol{v}_{i,n+\frac{1}{2}}, \ 1 \le i \le k, \\ \{\boldsymbol{v}_{i,n+*}\}_{i=1}^{k} = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+*}\}_{i=1}^{k}), \\ \boldsymbol{x}_{n+**} = \boldsymbol{x}_n + \frac{1}{2}\tau\beta R(\{\boldsymbol{v}_{j,n+*}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_{n+*}), \\ \tilde{\boldsymbol{v}}_{i,n+**} = \boldsymbol{v}_{i,n} + \frac{1}{2}\tau\gamma(R(\{\boldsymbol{v}_{j,n+*}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n+*}\boldsymbol{v}_{i,n+\frac{1}{2}}^{\top})\mathbb{G}(\boldsymbol{x}_{n+*})\boldsymbol{v}_{i,n+*}, \ 1 \le i \le k, \\ \{\boldsymbol{v}_{i,n+**}\}_{i=1}^{k} = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+**}\}_{i=1}^{k}), \\ \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \frac{1}{6}\tau\beta[R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_n) + 2R(\{\boldsymbol{v}_{j,n+\frac{1}{2}}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_{n+\frac{1}{2}}) + 2R(\{\boldsymbol{v}_{j,n+*}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_{n+*}) + R(\{\boldsymbol{v}_{j,n+**}\}_{j=1}^{k})\boldsymbol{F}(\boldsymbol{x}_{n+**})], \\ \tilde{\boldsymbol{v}}_{i,n+1} = \boldsymbol{v}_{i,n} + \frac{1}{6}\tau\gamma[(R(\{\boldsymbol{v}_{j,n}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n}\boldsymbol{v}_{i,n}^{\top})\mathbb{G}(\boldsymbol{x}_n)\boldsymbol{v}_{i,n} + 2(R(\{\boldsymbol{v}_{j,n+\frac{1}{2}}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n+\frac{1}{2}}\boldsymbol{v}_{i,n+\frac{1}{2}}^{\top})\mathbb{G}(\boldsymbol{x}_{n+\frac{1}{2}})\boldsymbol{v}_{i,n+\frac{1}{2}} \\ +2(R(\{\boldsymbol{v}_{j,n+*}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n+*}\boldsymbol{v}_{i,n+*}^{\top})\mathbb{G}(\boldsymbol{x}_{n+*})\boldsymbol{v}_{i,n+*} + (R(\{\boldsymbol{v}_{j,n+**}\}_{j=1}^{i-1}) - \boldsymbol{v}_{i,n+**}\boldsymbol{v}_{i,n+**}^{\top})\mathbb{G}(\boldsymbol{x}_{n+**})\boldsymbol{v}_{i,n+**}], \quad 1 \le i \le k, \\ \{\boldsymbol{v}_{i,n+1}\}_{i=1}^{k} = \mathrm{orth}(\{\tilde{\boldsymbol{v}}_{i,n+1}\}_{i=1}^{k}). \end{cases} \tag{34}$$

The oRK2-I, oRK2-II, oRK3 and oRK4 for GHiSD can be obtained by replacing $\mathbb{G}$ with $\mathbb{J}+\mathbb{J}^{\top}$ in above equations.

Data availability

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

References

[1] Baker J. An algorithm for the location of transition states. J Comput Chem 1986;7:385–95.
[2] Cheng X, Lin L, E W, Zhang P, Shi A. Nucleation of ordered phases in block copolymers. Phys Rev Lett 2010;104:148301.
[3] Doye J, Wales D. Saddle points and dynamics of lennard-jones clusters, solids, and supercooled liquids. J Chem Phys 2002;116:3777–88.
[4] E W, Ma C, Wu L. A comparative analysis of optimization and generalization properties of two-layer neural network and random feature models under gradient descent dynamics. Sci China Math 2020;63:1235–58.
[5] E W, Vanden-Eijnden E. Transition-path theory and path-finding algorithms for the study of rare events. Annu Rev Phys Chem 2010;61:391–420.
[6] Gu S, Lin L, Zhou X. Projection method for saddle points of energy functional in H^{-1} metric. J Sci Comput 2021;89:12.
[7] Gu S, Zhou X. Simplified gentlest ascent dynamics for saddle points in non-gradient systems. Chaos 2018;28:123106.
[8] Han Y, Hu Y, Zhang P, Majumdar A, Zhang L. Transition pathways between defect patterns in confined nematic liquid crystals. J Comput Phys 2019;396:1–11.
[9] Li Y, Zhou J. A minimax method for finding multiple critical points and its applications to semilinear PDEs. SIAM J Sci Comput 2001;23:840–65.
[10] Li Z, Zhang F, Zhou J. Partial Newton-correction method for multiple fixed points of semi-linear differential operators by Legendre-Gauss–Lobatto pseudospectral method. J Sci Comput 2023;97:32.
[11] Mallamace F, Corsaro C, Mallamace D, Vasi S, Vasi C, Baglioni P, Buldyrev S, Chen S, Stanley H. Energy landscape in protein folding and unfolding. Proc Natl Acad Sci USA 2016;113:3159–63.
[12] Onuchic J, Luthey-Schulten Z, Wolynes P. Theory of protein folding: The energy landscape perspective. Ann Rev Phys Chem 1997;48:545–600.
[13] Xie Z, Yuan Y, Zhou J. On finding multiple solutions to a singularly perturbed Neumann problem. SIAM J Sci Comput 2012;34:A395–420.
[14] Xie Z, Yuan Y, Zhou J. On solving semilinear singularly perturbed Neumann problems for multiple solutions. SIAM J Sci Comput 2022;44:A501–23.
[15] Zhang L, Chen L, Du Q. Morphology of critical nuclei in solid-state phase transformations. Phys Rev Lett 2007;98:265703.
[16] Zhang Y, Yang X, Zhang L, Li Y, Zhang T, Sun S. Energy landscape analysis for two-phase multi-component NVT flash systems by using ETD type high-index saddle dynamics. J Comput Phys 2023;477:111916.
[17] Bofill J, Quapp W, Caballero M. Locating transition states on potential energy surfaces by the gentlest ascent dynamics. Chem Phys Lett 2013;583:203–8.
[18] Cancés E, Legoll F, Marinica M-C, Minoukadeh K, Willaime F. Some improvements of the activation-relaxation technique method for finding transition pathways on potential energy surfaces. J Chem Phys 2009;130:114711.
[19] Cui G, Jiang K, Zhou T. An efficient saddle search method for ordered phase transitions involving translational invariance. Comput Phys Comm 2025;306:109381.
[20] Du Q, Zhang L. A constrained string method and its numerical analysis. Commun Math Sci 2009;7:1039–51.
[21] E W, Ren W, Vanden-Eijnden E. String method for the study of rare events. Phys Rev B 2002;66:052301.
[22] E W, Zhou X. The gentlest ascent dynamics. Nonlinearity 2011;24:1831–42.
[23] Henkelman G, Jónsson H. A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. J Chem Phys 1999;111:7010–22.
[24] Henkelman G, Jónsson H. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. J Chem Phys 2000;113:9978–85.
[25] Henkelman G, Jóhannesson G, Jónsson H. Methods for finding saddle points and minimum energy paths, theoretical methods in condensed phase chemistry. Dordrecht: Springer Netherlands; 2002, p. 269–302.
[26] Liu W, Xie Z, Yi W. Normalized goldstein-type local minimax method for finding multiple unstable solutions of semilinear elliptic PDEs. Commun Math Sci 2021;19:147–74.
[27] Sheppard D, Terrell R, Henkelman G. Optimization methods for finding minimum energy paths. J Chem Phys 2008;128:134106.
[28] Zhang J, Du Q. Shrinking dimer dynamics and its applications to saddle point search. SIAM J Numer Anal 2012;50:1899–921.
[29] Zhang L, Ren W, Samanta A, Du Q. Recent developments in computational modelling of nucleation in phase transformations. Npj Comput Mater 2016;2:16003.
[30] Zhang L, Du Q, Zheng Z. Optimization-based shrinking dimer method for finding transition states. SIAM J Sci Comput 2016;38:A528–44.
[31] Grantham W. Gradient transformation trajectory following algorithms for determining stationary min–max saddle points. In: Advances in dynamic game theory. Boston, MA: Birkhäuser Boston; 2007, p. 639–57.
[32] Gao W, Leng J, Zhou X. An iterative minimization formulation for saddle point search. SIAM J Numer Anal 2015;53:1786–805.
[33] Gould N, Ortner C, Packwood D. A dimer-type saddle search algorithm with preconditioning and linesearch. Math Comp 2016;85:2939–66.
[34] Levitt A, Ortner C. Convergence and cycling in walker-type saddle search algorithms. SIAM J Numer Anal 2017;55:2204–27.
[35] Miao S, Liu Z, Zhang L, Zhang P, Zheng X. Construction and analysis for Adams explicit discretization of high-index saddle dynamics. CSIAM Trans Appl Math 2025. to appear.
[36] Yin J, Zhang L, Zhang P. High-index optimization-based shrinking dimer method for finding high-index saddle points. SIAM J Sci Comput 2019;41:A3576–95.
[37] Yin J, Wang Y, Chen J, Zhang P, Zhang L. Construction of a pathway map on a complicated energy landscape. Phys Rev Lett 2020;124:090601.
[38] 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 A 2021;477:20210458.
[39] Shi B, Han Y, Zhang L. Nematic liquid crystals in a rectangular confinement: solution landscape, and bifurcation. SIAM J Appl Math 2022;82:1808–28.
[40] Xu Z, Han Y, Yin J, Yu B, Nishiura Y, Zhang L. Solution landscapes of the diblock copolymer-homopolymer model under two-dimensional confinement. Phys Rev E 2021;104:014505.
[41] Yin J, Huang Z, Cai Y, Du Q, Zhang L. Revealing excited states of rotational Bose–Einstein condensates. Innov 2024;5:100546.
[42] Yin J, Jiang K, Shi A-C, Zhang P, Zhang L. Transition pathways connecting crystals and quasicrystals. Proc Natl Acad Sci USA 2021;118:e2106230118.
[43] Yin J, Zhang L, Zhang P. Solution landscape of Onsager functional identifies non-axisymmetric critical points. Phys D: Nonlinear Phenom 2022;430:133081.
[44] Milnor J. Morse theory. Princeton University Press; 1963.
[45] Yin J, Yu B, Zhang L. Searching the solution landscape by generalized high-index saddle dynamics. Sci China Math 2021;64:1801.
[46] Wiggins S. Introduction to applied nonlinear dynamical systems and chaos. New York: Springer-Verlag; 2003.
[47] Eckhardt B. Irregular scattering. Phys D 1988;33:89–98.
[48] Luo Y, Zheng X, Cheng X, Zhang L. Convergence analysis of discrete high-index saddle dynamics. SIAM J Numer Anal 2022;60:2731–50.
[49] Zhang L, Zhang P, Zheng X. Error estimates for Euler discretization of high-index saddle dynamics. SIAM J Numer Anal 2022;60:2925–44.
[50] Zhang L, Zhang P, Zheng X. Mathematical and numerical analysis to shrinking-dimer saddle dynamics with local Lipschitz conditions. CSIAM Trans Appl Math 2023;4:157–76.
[51] Luo Y, Zhang L, Zhang P, Zhang Z, Zheng X. Semi-implicit method of high-index saddle dynamics and application to construct solution landscape. Numer Meth. PDEs 2024;40:e23123.
[52] Zhang L, Zhang P, Zheng X. Understanding high-index saddle dynamics via numerical analysis. Commun Math Sci 2025;23:541–60.