

Stability of variable-step BDF2 and BDF3 methods

Zhaoyi Li* Hong-lin Liao†

January 4, 2022

Abstract

We prove that the two-step backward differentiation formula (BDF2) method is stable on arbitrary time grids; while the variable-step BDF3 scheme is stable if almost all adjacent step ratios are less than 2.553. These results relax the severe mesh restrictions in the literature and provide a new understanding of variable-step BDF methods. Our main tools include the discrete orthogonal convolution kernels and an elliptic-type matrix norm.

Keywords: variable-step BDF formula, stability, step-ratio condition, discrete orthogonal convolution kernels

AMS subject classifications: 65M06, 65M12

1 Introduction

In this paper, we revisit the stability of variable-step backward differentiation formulas (BDF) for the following initial value problem

$$\frac{dv}{dt} = f(t, v) \quad \text{for } 0 < t \leq T.$$

Choose time levels $0 = t_0 < t_1 < t_2 < \dots < t_N = T$ with the variable time-step $\tau_k := t_k - t_{k-1}$ for $1 \leq k \leq N$, and the maximum step size $\tau := \max_{1 \leq k \leq N} \tau_k$. Put the adjacent time-step ratios

$$r_k := \frac{\tau_k}{\tau_{k-1}} \quad \text{for } 2 \leq k \leq N \quad \text{with } r_1 := 0.$$

For any sequences $\{v^n\}_{n=0}^N$, denote $\partial_\tau v^n := (v^n - v^{n-1})/\tau_n$. Let $\Pi_{n,k}v$ ($k = 1, 2, 3$) be the interpolating polynomial of a function v over the nodes t_{n-k}, \dots, t_{n-1} and t_n . To describe the variable coefficients, define the following three functions

$$d_0(x, y) := \frac{1 + 2x}{1 + x} + \frac{xy}{1 + y + xy}, \quad (1.1)$$

$$d_1(x, y) := -\frac{x}{1 + x} - \frac{xy}{1 + y + xy} - \frac{xy^2}{1 + y + xy} \frac{1 + x}{1 + y}, \quad (1.2)$$

$$d_2(x, y) := \frac{xy^2}{1 + y + xy} \frac{1 + x}{1 + y}, \quad \text{for } x, y \geq 0. \quad (1.3)$$

*Department of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, P. R. China. Email: 1525170777@qq.com

†Corresponding author. ORCID 0000-0003-0777-6832. Department of Mathematics, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; Key Laboratory of Mathematical Modelling and High Performance Computing of Air Vehicles (NUAA), MIIT, Nanjing 211106, China. Hong-lin Liao (liaohl@nuaa.edu.cn and liaohl@csrc.ac.cn) is supported by a grant 12071216 from National Natural Science Foundation of China.

Taking $v^n = v(t_n)$, one has the BDF1 formula $D_1 v^n := (\Pi_{n,1} v)'(t_n) = d_0(0,0) \partial_\tau v^n$ for $n \geq 1$, and the variable-step BDF2 formula

$$D_2 v^n := (\Pi_{n,2} v)'(t_n) = d_0(r_n, 0) \partial_\tau v^n + d_1(r_n, 0) \partial_\tau v^{n-1} \quad \text{for } n \geq 2. \quad (1.4)$$

The variable-step BDF3 formula $D_3 v^n := (\Pi_{n,3} v)'(t_n)$ for $n \geq 3$,

$$D_3 v^n = d_0(r_n, r_{n-1}) \partial_\tau v^n + d_1(r_n, r_{n-1}) \partial_\tau v^{n-1} + d_2(r_n, r_{n-1}) \partial_\tau v^{n-2}. \quad (1.5)$$

As seen, the formula (1.5) also represents a variable-order variable-step BDF formula from the first-order to third-order. The theoretical results on (1.5) are naturally valid for any combinations of the BDF-k formulas if the time-step ratios are allowed to be zero in the associated coefficients.

We consider the variable-step BDF-k time-stepping scheme for the ODE problem,

$$D_k v^n = f(t_n, v^n) \quad \text{for } k \leq n \leq N. \quad (1.6)$$

Without losing the generality, assume that the starting values u^1, \dots, u^{k-1} are available by some other high-order approaches, such as Runge-Kutta methods. For the stability analysis, assume further that the perturbed solution \bar{v}^n solves the equation $D_k \bar{v}^n = f(t_n, \bar{v}^n) + \varepsilon^n$ with a bounded sequence $\{\varepsilon^n\}$. Then the difference $\tilde{v}^i = \bar{v}^i - v^i$ solves

$$D_k \tilde{v}^i = f(t_i, \bar{v}^i) - f(t_i, v^i) + \varepsilon^i \quad \text{for } k \leq i \leq N. \quad (1.7)$$

These well-known stiff solvers have been tested to be practically valuable for differential-algebraic problems [5, 9, 10] and for the hyperbolic systems with multiscale relaxation [1]. Also, the variable-step versions are computationally efficient in capturing the multi-scale time behaviors [2, 4, 10–13, 15] via adaptive time-stepping strategies. On the other hand, the stability and convergence analysis of variable-step BDF methods are difficult and remain incomplete, cf. [2, 4, 6–9, 15], because they involve multiple degrees of freedom (independent time-step sizes).

One of the main defects in the existing theory is the severe mesh condition required for stability. For examples, the zero-stability of variable-step BDF2 method requires a step-ratio condition $0 < r_k \leq 1 + \sqrt{2} \approx 2.414$, see [6, 7]. By considering a specific time grid with constant step-ratio $r_k = r$, it was also shown in [6] that the zero-stability of variable-step BDF3 method requires the maximum step-ratio limit $R_3 < (1 + \sqrt{5})/2 \approx 1.618$. By using an elliptic type norm, Calvo, Grande and Grigorieff [2] gave the zero-stability constraint $r_k \leq 1.476$ in 1990. Twenty years ago, Guglielmi and Zennaro [8] applied a spectral radius approach to find the step-ratio condition $r_k \leq 1.501$, which seems to be the best result for the variable-step BDF3 method.

We aim to relax the existing mesh conditions for the perturbed stability of variable-step BDF2 and BDF3 methods according to the following definition.

Definition 1 (Stability). *Assume that the nonlinear function $f(t, v)$ is a Lipschitz-continuous function with the Lipschitz constant $L_f > 0$. The k-step time-stepping (1.6) with unequal time-steps is stable if there exists a real number τ_0 and a fixed constant C (independent of the step sizes τ_n) such that the numerical solution of (1.7) fulfills*

$$\max_{k \leq n \leq N} |\tilde{v}^i| \leq C \left(\sum_{i=0}^{k-1} |\tilde{v}^i| + \tau \sum_{i=1}^{k-1} |\partial_\tau \tilde{v}^i| + \max_{k \leq i \leq N} |\varepsilon^i| \right) \quad \text{for all } \tau_n \leq \tau_0.$$

This definition is practically reasonable in the sense that one always uses some high-order scheme to start the k -step time-stepping (1.6). In the uniform time-step case, it reduces into the classical zero-stability, cf. [3, Definition 2.2],

$$\max_{k \leq n \leq N} |\tilde{v}^i| \leq C \left(\sum_{i=0}^{k-1} |\tilde{v}^i| + \max_{k \leq i \leq N} |\varepsilon^i| \right) \quad \text{for all } \tau_n \leq \tau_0.$$

Definition 1 would be as useful as the original zero-stability for the numerical analysis of variable-step, multi-step methods. Actually, it was proven in [14] that the step-ratio condition $r_k < 3.561$ is sufficient to the stability of variable-step BDF2 method for linear parabolic problem. In a recent work [12], this step-ratio condition was slightly improved to a new one, $r_k < 4.864$. These step-ratio conditions are weaker than the classical zero-stability restriction $r_k < 2.414$, but both of them are sufficient for the stability of BDF2 method in the sense of Definition 1.

By developing a new framework with discrete orthogonal convolution (DOC) kernels, we apply Definition 1 to examine the ratio-step conditions for the stability of BDF2 and BDF3 methods in solving the ODE problems. The main contribution is two-fold:

- Theorem 2.1 shows that the BDF2 method is stable on arbitrary time grids.
- Theorem 3.1 shows that the variable-step BDF3 method is stable if the adjacent step ratios $r_k < R_3$, where $R_3 \approx 2.553$ is the unique positive root of

$$R_3^6 + R_3^5 - 4R_3^4 - 8R_3^3 - 10R_3^2 - 6R_3 - 2 = 0.$$

We start from a novel viewpoint by writing the BDF- k ($k = 2, 3$) formula as a convolution summation of local difference quotients (this viewpoint is quite different from that in previous works [12, 14])

$$D_k v^n := \sum_{j=1}^n d_{n-j}^{(k,n)} \partial_\tau v^j \quad \text{for } n \geq k, \quad (1.8)$$

where the first superscript index k in the discrete kernels $d_{n-j}^{(k,n)}$ denotes the order of method. From the BDF2 formula (1.4) with $n \geq 2$, we have

$$d_j^{(2,n)} := d_j(r_n, 0) \quad \text{for } j = 0, 1, \quad \text{and} \quad d_j^{(2,n)} := 0 \quad \text{for } n \geq j + 1 \geq 3. \quad (1.9)$$

Similarly, from the BDF3 formula (1.5) with $n \geq 3$, we have

$$d_j^{(3,n)} := d_j(r_n, r_{n-1}) \quad \text{for } j = 0, 1, 2, \quad \text{and} \quad d_j^{(3,n)} := 0 \quad \text{for } n \geq j + 1 \geq 4. \quad (1.10)$$

Here and hereafter, assume that the summation $\sum_{k=i}^j \cdot$ to be zero and the product $\prod_{k=i}^j \cdot$ to be one if the index $i > j$. As for the BDF- k kernels $d_{n-j}^{(k,n)}$ with any fixed indexes n , we recall a class of discrete orthogonal convolution (DOC) kernels $\{\vartheta_{n-j}^{(k,n)}\}_{j=k}^n$ by a recursive procedure, also see [12, 14],

$$\vartheta_0^{(k,n)} := \frac{1}{d_0^{(k,n)}} \quad \text{and} \quad \vartheta_{n-j}^{(k,n)} := -\frac{1}{d_0^{(k,j)}} \sum_{i=j+1}^n \vartheta_{n-i}^{(k,n)} d_{i-j}^{(k,i)} \quad \text{for } k \leq j \leq n-1. \quad (1.11)$$

2 Unconditional stability of BDF2 method

For the BDF2 method, the associated DOC kernels are positive and decay exponentially.

Lemma 2.1. *The DOC kernels $\vartheta_{n-j}^{(2,n)}$ in (1.11) have an explicit formula*

$$\vartheta_{n-j}^{(2,n)} = \frac{1}{d_0^{(2,j)}} \prod_{i=j+1}^n \frac{r_i}{1+2r_i} > 0 \quad \text{for } 2 \leq j \leq n,$$

and satisfy

$$\sum_{j=2}^n \vartheta_{n-j}^{(2,n)} = 1 - \prod_{i=2}^n \frac{r_i}{1+2r_i} < 1 \quad \text{for } n \geq 2.$$

Proof. For any fixed j , by taking $n = j$ in the identity (1.13), one has $\vartheta_0^{(2,j)} = 1/d_0^{(2,j)}$ for $j \geq 2$. According to the definition (1.9), one has $d_{n-i}^{(2,n)} = 0$ for $n \geq i + 2$. The identity (1.13) gives

$$\vartheta_{m-j}^{(2,m)} = -\frac{d_1^{(2,m)}}{d_0^{(2,m)}} \vartheta_{m-1-j}^{(2,m-1)} = -\frac{d_1(r_m, 0)}{d_0(r_m, 0)} \vartheta_{m-1-j}^{(2,m-1)} = \frac{r_m}{1+2r_m} \vartheta_{m-1-j}^{(2,m-1)}$$

for $m \geq j + 1 \geq 2$. It yields immediately

$$\vartheta_{n-j}^{(2,n)} = \vartheta_{n-2-j}^{(2,n-2)} \prod_{i=n-1}^n \frac{r_i}{1+2r_i} = \dots = \vartheta_0^{(2,j)} \prod_{i=j+1}^n \frac{r_i}{1+2r_i} \quad \text{for } 2 \leq j \leq n.$$

It leads to the first result. By taking $v^n = t_n$ in (1.8), one has $\partial_\tau t_n = 1$ and $D_2 t_n = 1$ for $n \geq 2$. Thus applying the equality (1.14) with $k = 2$ yields

$$\begin{aligned} \sum_{j=2}^n \vartheta_{n-j}^{(2,n)} &= \sum_{j=2}^n \vartheta_{n-j}^{(2,n)} D_2 t_j = (\partial_\tau t_1) \sum_{i=2}^n \vartheta_{n-i}^{(2,n)} d_{i-1}^{(2,i)} + \partial_\tau t_n \\ &= 1 + \vartheta_{n-2}^{(2,n)} d_1^{(2,2)} = 1 - \prod_{i=2}^n \frac{r_i}{1+2r_i} < 1 \quad \text{for } n \geq 1, \end{aligned} \quad (2.1)$$

where the explicit formula of $\vartheta_{n-2}^{(2,n)}$ and the definition (1.9) were used. It completes the proof. \square

Theorem 2.1. *If $\tau \leq 1/(4L_f)$, the BDF2 solution of (1.7) satisfies*

$$|\tilde{v}^n| \leq 2 \exp(4L_f t_{n-1}) \left(|\tilde{v}^1| + 2\tau |\partial_\tau \tilde{v}^1| + 2t_n \max_{2 \leq i \leq n} |\varepsilon^i| \right) \quad \text{for } 2 \leq n \leq N.$$

Thus the variable-step BDF2 scheme is stable on arbitrary time meshes.

Proof. Take $k = 2$. Multiplying both sides of (1.16) with $2\tau_n \tilde{v}^n$, one applies Lemma 2.1 (the positivity and boundedness of DOC kernels) to obtain that

$$\begin{aligned} |\tilde{v}^n|^2 - |\tilde{v}^{n-1}|^2 + |\tau_n \partial_\tau \tilde{v}^n|^2 &= -2\tau_n \tilde{v}^n \mathcal{I}_2^n[\tilde{v}] + 2\tau_n \tilde{v}^n \sum_{i=2}^n \vartheta_{n-i}^{(2,n)} [f(t_i, \bar{v}^i) - f(t_i, v^i) + \varepsilon^i] \\ &\leq 2\tau_n |\tilde{v}^n| |\mathcal{I}_2^n[\tilde{v}]| + 2\tau_n |\tilde{v}^n| \sum_{i=2}^n \vartheta_{n-i}^{(2,n)} |f(t_i, \bar{v}^i) - f(t_i, v^i)| + 2\tau_n |\tilde{v}^n| \max_{1 \leq i \leq n} |\varepsilon^i| \sum_{i=2}^n \vartheta_{n-i}^{(2,n)} \\ &\leq 2\tau_n |\tilde{v}^n| |\mathcal{I}_2^n[\tilde{v}]| + 2L_f \tau_n |\tilde{v}^n| \sum_{i=2}^n \vartheta_{n-i}^{(2,n)} |\tilde{v}^i| + 2|\tilde{v}^n| \tau_n \max_{1 \leq i \leq n} |\varepsilon^i|, \end{aligned}$$

where the Lipschitz continuous property of the nonlinear function f has been used. By dropping the nonnegative term at the left side and summing n from $n = 2$ to m , we have

$$|\tilde{v}^m|^2 \leq |\tilde{v}^1|^2 + 2 \sum_{j=2}^m \tau_j |\tilde{v}^j| |\mathcal{I}_2^j[\tilde{v}]| + 2L_f \sum_{j=2}^m \tau_j |\tilde{v}^j| \sum_{i=2}^j \vartheta_{j-i}^{(2,j)} |\tilde{v}^i| + 2 \sum_{j=2}^m \tau_j |\tilde{v}^j| \max_{1 \leq i \leq j} |\varepsilon^i|.$$

Let m_0 ($1 \leq m_0 \leq m$) be an integer such that $|\tilde{v}^{m_0}| = \max_{1 \leq k \leq m} |\tilde{v}^k|$. Now we take $m := m_0$ in the above inequality and get

$$|\tilde{v}^{m_0}|^2 \leq |\tilde{v}^1| |\tilde{v}^{m_0}| + 2 |\tilde{v}^{m_0}| \sum_{j=2}^{m_0} \tau_j |\mathcal{I}_2^j[\tilde{v}]| + 2L_f |\tilde{v}^{m_0}| \sum_{j=2}^{m_0} \tau_j |\tilde{v}^j| + 2 |\tilde{v}^{m_0}| \max_{2 \leq i \leq m_0} |\varepsilon^i| \sum_{j=2}^{m_0} \tau_j.$$

It leads to

$$|\tilde{v}^m| \leq |\tilde{v}^{m_0}| \leq |\tilde{v}^1| + 2 \sum_{j=2}^m \tau_j |\mathcal{I}_2^j[\tilde{v}]| + 2L_f \sum_{j=2}^m \tau_j |\tilde{v}^j| + 2t_m \max_{2 \leq i \leq m} |\varepsilon^i| \quad \text{for } 2 \leq m \leq N. \quad (2.2)$$

By using Lemma 2.1, the starting error $\mathcal{I}_2^n[\tilde{v}]$ in (1.15) reads

$$\mathcal{I}_2^n[\tilde{v}] = \partial_\tau \tilde{v}^1 \sum_{i=2}^n \vartheta_{n-i}^{(2,n)} d_{i-1}^{(2,i)} = \vartheta_{n-2}^{(2,n)} d_1^{(2,2)} (\partial_\tau \tilde{v}^1) = -(\partial_\tau \tilde{v}^1) \prod_{i=2}^n \frac{r_i}{1+2r_i} \quad \text{for } n \geq 2$$

such that

$$\sum_{j=2}^m \tau_j |\mathcal{I}_2^j[\tilde{v}]| \leq \tau |\partial_\tau \tilde{v}^1| \sum_{j=2}^m \frac{1}{2^{j-1}} \leq \tau |\partial_\tau \tilde{v}^1| \quad \text{for } m \geq 2.$$

It follows from (2.2) that

$$|\tilde{v}^n| \leq |\tilde{v}^1| + 2\tau |\partial_\tau \tilde{v}^1| + 2L_f \sum_{j=2}^n \tau_j |\tilde{v}^j| + 2t_n \max_{2 \leq i \leq n} |\varepsilon^i| \quad \text{for } 2 \leq n \leq N.$$

Assuming that $\tau \leq 1/(4L_f)$, one has

$$|\tilde{v}^n| \leq 2|\tilde{v}^1| + 4\tau |\partial_\tau \tilde{v}^1| + 4L_f \sum_{j=2}^{n-1} \tau_j |\tilde{v}^j| + 4t_n \max_{2 \leq i \leq n} |\varepsilon^i| \quad \text{for } 2 \leq n \leq N.$$

The standard discrete Grönwall inequality, e.g. [14, Lemma 3.1], completes the proof. \square

3 Stability analysis of BDF3 method

3.1 Decaying of DOC kernels

Note that, the variable-step BDF3 method (1.8) of $k = 3$ is exact for the linear polynomial $v = t$. Taking $v^n = t_n$ in (1.8), one can find that $D_3 t_n = 1$ for $n \geq 3$. As done in (2.1), we can apply the discrete equality (1.14) with (1.15) to get

$$\sum_{j=3}^n \vartheta_{n-j}^{(3,n)} = \sum_{j=1}^2 \sum_{i=3}^n \vartheta_{n-i}^{(3,n)} d_{i-j}^{(3,i)} + 1 \quad \text{for } n \geq 3.$$

However, it does not provide enough information for the subsequent stability analysis because no explicit formulas of DOC kernels $\vartheta_{n-j}^{(3,n)}$ are available. Furthermore, the DOC kernels are not always positive, see Remark 4 below, we turn to bound $\sum_{j=3}^n |\vartheta_{n-j}^{(3,n)}|$ for any time-levels t_n by imposing certain step-ratio constraint.

To do so, introduce the modified DOC kernels

$$\widehat{\vartheta}_{i-j}^{(3,i)} := \vartheta_{i-j}^{(3,i)} d_0^{(3,j)} \quad \text{for } i \geq j \geq 3. \quad (3.1)$$

The discrete identity (1.13) gives

$$\sum_{i=j}^n d_{n-i}^{(3,n)} \widehat{\vartheta}_{i-j}^{(3,i)} \equiv \delta_{nj} d_0^{(3,j)} \quad \text{for } n \geq j \geq 3. \quad (3.2)$$

For any fixed index $j \geq 3$, by taking $n = j$ and $n = j + 1$ in the identity (3.2), respectively, one can derive that

$$\widehat{\vartheta}_0^{(3,j)} = 1 \quad \text{and} \quad \widehat{\vartheta}_1^{(3,j+1)} = -\frac{d_1^{(3,j+1)}}{d_0^{(3,j+1)}} \quad \text{for } j \geq 3. \quad (3.3)$$

According to the definition (1.10), one has $d_{n-i}^{(3,n)} = 0$ for $n \geq i + 3$. The identity (1.13) gives

$$\widehat{\vartheta}_{m-j}^{(3,m)} + \frac{d_1^{(3,m)}}{d_0^{(3,m)}} \widehat{\vartheta}_{m-1-j}^{(3,m-1)} + \frac{d_2^{(3,m)}}{d_0^{(3,m)}} \widehat{\vartheta}_{m-2-j}^{(3,m-2)} = 0 \quad \text{for } m \geq j + 2 \geq 5,$$

or the difference equation

$$\widehat{\vartheta}_{m-j}^{(3,m)} - \alpha_m \widehat{\vartheta}_{m-1-j}^{(3,m-1)} + \beta_m \widehat{\vartheta}_{m-2-j}^{(3,m-2)} = 0 \quad \text{for } m \geq j + 2 \geq 5, \quad (3.4)$$

where, by using (1.1)-(1.3) and (A.2)-(A.3), the coefficients α_m and β_m are defined by

$$\alpha_m := -\frac{d_1^{(3,m)}}{d_0^{(3,m)}} = \alpha(r_m, r_{m-1}) \quad \text{and} \quad \beta_m := \frac{d_2^{(3,m)}}{d_0^{(3,m)}} = \beta(r_m, r_{m-1}) \quad \text{for } m \geq 3. \quad (3.5)$$

To evaluate the asymptotic behaviors of the DOC kernels $\widehat{\vartheta}_{m-j}^{(3,m)}$, introduce the following vector and companion matrix

$$\vec{u}_m = \begin{pmatrix} \widehat{\vartheta}_{m-j}^{(3,m)} \\ \widehat{\vartheta}_{m-1-j}^{(3,m-1)} \end{pmatrix} \quad \text{and} \quad A_m := \begin{pmatrix} \alpha_m & -\beta_m \\ 1 & 0 \end{pmatrix}. \quad (3.6)$$

For any fixed index $j \geq 1$, the difference equation (3.4) reads

$$\vec{u}_m = A_m \vec{u}_{m-1} \quad \text{for } m \geq j + 2 \geq 5,$$

or

$$\vec{u}_n = A_n \vec{u}_{n-1} = A_n A_{n-1} \vec{u}_{n-2} = \cdots = \prod_{i=j+2}^n A_i \vec{u}_{j+1} \quad \text{for } n \geq j + 2 \geq 5. \quad (3.7)$$

The associated initial vector $\vec{u}_{j+1} = (\alpha_{j+1}, 1)^T$ according to (3.3).

If all of step-ratios $0 < r_k < \hat{R}_3 \approx 3.4405$ ($k \geq 2$), Lemmas A.1-A.2 show that

$$\beta_i < \alpha_i < 1 + \beta_i < 2 \quad \text{for } i \geq 3. \quad (3.8)$$

They imply that the roots of $\lambda^2 - \alpha_i \lambda + \beta_i = 0$ satisfy $|\lambda| < 1$. Thus the eigenvalues of A_i have the module less than 1 and the spectral radius $\rho(A_i) < 1$ for any $i \geq 3$. In general, it is not sufficient to ensure the global decaying of DOC kernels $\hat{v}_{n-j}^{(3,n)}$.

It needs to examine the joint companion matrix $\prod_{i=j+2}^n A_i$ for $n \geq j+2 \geq 5$. Let $\|\cdot\|_\infty$ be the maximum norm on the space \mathbb{C}^2 and let the same symbol $\|\cdot\|_\infty$ denote also the corresponding induced matrix norm. We adopt the technique of Calvo *et al* [2] by considering a complex constant μ with $\text{Im}(\mu) \neq 0$ and the following elliptic type norm

$$\|A_i\|_{H,\infty} := \|H^{-1}A_iH\|_\infty \quad \text{with} \quad H := \begin{pmatrix} \mu & \bar{\mu} \\ 1 & 1 \end{pmatrix}. \quad (3.9)$$

Then one can bound the DOC kernels via

$$\|\vec{u}_n\|_\infty \leq \|H\|_\infty \prod_{i=j+2}^n \|A_i\|_{H,\infty} \|H^{-1}\|_\infty \|\vec{u}_{j+1}\|_\infty \quad \text{for } n \geq j+2 \geq 5. \quad (3.10)$$

To process the analysis, it is to determine a fixed parameter μ . We will choose a proper constant $\tilde{R}_3 < \hat{R}_3 \approx 3.4405$ to ensure that

$$\|A_i\|_{H,\infty} < 1 \quad \text{for } i \geq 3, \quad (3.11)$$

if all of step-ratios $r_k < \tilde{R}_3$ ($k \geq 2$). It is easy to derive that

$$\|A_i\|_{H,\infty} = \frac{|\mu^2 - \alpha_i \mu + \beta_i| + |\mu|^2 - \alpha_i \mu + \beta_i}{|\mu - \bar{\mu}|} \quad \text{for } i \geq 3. \quad (3.12)$$

Noticing that the following fact

$$|\mu|^2 - \alpha_i \mu + \beta_i - |\mu^2 - \alpha_i \mu + \beta_i|^2 = \beta_i |\mu - \bar{\mu}|^2,$$

one has

$$\|A_i\|_{H,\infty} = \frac{|\mu^2 - \alpha_i \mu + \beta_i|}{|\mu - \bar{\mu}|} + \sqrt{\beta_i + \frac{|\mu^2 - \alpha_i \mu + \beta_i|^2}{|\mu - \bar{\mu}|^2}} \quad \text{for } i \geq 3. \quad (3.13)$$

Then the requirement (3.11) is equivalent to

$$\frac{|\mu^2 - \alpha_i \mu + \beta_i|}{|\mu - \bar{\mu}|} < \frac{1}{2}(1 - \beta_i) \quad \text{for } i \geq 3. \quad (3.14)$$

Then, taking into account the coefficient relationship (3.8), we can find that the condition (3.14) can be written equivalently in the following form, also see [2, Eq. (14)],

$$\left| \mu - \frac{1}{2}\alpha_i - \frac{i}{2}\sqrt{(1 + \beta_i)^2 - \alpha_i^2} \right| < \frac{1}{2}(1 - \beta_i) \quad \text{for } i \geq 3. \quad (3.15)$$

Note that, the other branch requiring $|\mu - \frac{1}{2}\alpha_i + \frac{i}{2}\sqrt{(1 + \beta_i)^2 - \alpha_i^2}| < \frac{1}{2}(1 - \beta_i)$ is omitted here since it is equivalent to (3.15) by replacing the undetermined parameter μ with $\bar{\mu}$. Since the coefficients α_i, β_i are dependent on the ratios r_i and r_{i-1} , The inequalities in (3.15) define a family of complex disks $\mathfrak{D}(r_i, r_{i-1})$ centered at $(\frac{1}{2}\alpha_i, \frac{i}{2}\sqrt{(1 + \beta_i)^2 - \alpha_i^2})$ with the radius $\frac{1}{2}(1 - \beta_i)$.

A heuristic analysis is considered firstly to obtain a rough estimate for the complex parameter μ , while the mathematical proof is left to next subsection. To ensure (3.15), one should choose a fixed μ_0 such that the intersection set

$$\bigcap_{r_i, r_{i-1} \in [0, \tilde{R}_3]} \mathfrak{D}(r_i, r_{i-1}) \quad \text{is not empty.}$$

Reminding the increasing functions (A.2)-(A.3), we see that the disk $\mathfrak{D}(0, 0)$ centered at $(0, \frac{1}{2})$ has the maximum radius $\frac{1}{2}$, while the disk $\mathfrak{D}(\tilde{R}_3, \tilde{R}_3)$ centered at

$$\left(\frac{1}{2}\alpha(\tilde{R}_3, \tilde{R}_3), \frac{1}{2}\sqrt{[1 + \beta(\tilde{R}_3, \tilde{R}_3)]^2 - \alpha^2(\tilde{R}_3, \tilde{R}_3)} \right)$$

has the minimum radius $\frac{1}{2}(1 - \beta(\tilde{R}_3, \tilde{R}_3))$. Obviously, the largest value of \tilde{R}_3 may be determined from the fact that the smallest disk $\mathfrak{D}(\tilde{R}_3, \tilde{R}_3)$ is tangential to the largest one $\mathfrak{D}(0, 0)$ at the tangential point μ_0 . This necessary condition holds if and only if

$$\frac{1}{4}\alpha^2(\tilde{R}_3, \tilde{R}_3) + \left(\frac{1}{2}\sqrt{[1 + \beta(\tilde{R}_3, \tilde{R}_3)]^2 - \alpha^2(\tilde{R}_3, \tilde{R}_3)} - \frac{1}{2} \right)^2 = \frac{1}{4}[2 - \beta(\tilde{R}_3, \tilde{R}_3)]^2.$$

which leads to

$$\alpha^2(\tilde{R}_3, \tilde{R}_3) + 8\beta^2(\tilde{R}_3, \tilde{R}_3) - 8\beta(\tilde{R}_3, \tilde{R}_3) = 0, \quad (3.16)$$

or equivalently,

$$9\tilde{R}_3^6 - 2\tilde{R}_3^5 - 35\tilde{R}_3^4 - 42\tilde{R}_3^3 - 22\tilde{R}_3^2 - 4\tilde{R}_3 + 1 = 0.$$

This polynomial equation has two positive roots $\tilde{R}_3 \approx 2.5808$ and $\tilde{R}_3 \approx 0.1304$ (it is checked that this situation also occurs in [2]). We throw away the small root $\tilde{R}_3 \approx 0.1304$ and choose the large one $\tilde{R}_3 \approx 2.5808$ with the corresponding tangential point $\tilde{\mu}_* \approx 0.4979 + 0.5454i$.

Remark 1. *It is to mention that, by following the joint spectral radius approach in [8], one can obtain a slightly improved step-ratio constraint $0 < r_k < 2.705$ to ensure the boundedness of DOC kernels; however, it seems difficult to obtain the desired decaying behavior of DOC kernels theoretically from the extremal norm estimate of the family of companion matrices $\{A_i | i \geq 3\}$.*

Remark 2. *The condition (3.15) is sufficient for (3.11), or $\|A_i\|_{H, \infty} < 1$; but different choices of the parameter μ leads to different values of $\|A_i\|_{H, \infty}$. Consider the simple case with $r_{i-1} \equiv 0$ for any $i \geq 3$, corresponding to the variable-step BDF2 scheme. By using the definition (3.5) together with (1.1)-(1.3), one has $\alpha_i = \frac{r_i}{1+2r_i}$, $\beta_i = 0$ and the reduced class of complex discs*

$$\mathfrak{D}(r_i, 0) = \left\{ \mu \mid \left| \mu - \frac{1}{2}\alpha(r_i, 0) - \frac{i}{2}\sqrt{1 - \alpha^2(r_i, 0)} \right| < \frac{1}{2} \right\} \quad \text{for } i \geq 3.$$

Recalling the fact $(\alpha_i + \frac{1}{2})(\alpha_i - 1) < 0$ for any $r_i \geq 0$, one can check that

$$\left| \frac{1}{4} - \frac{1}{2}\alpha(r_i, 0) + \frac{\iota}{2} \left(\frac{\sqrt{3}}{2} - \sqrt{1 - \alpha^2(r_i, 0)} \right) \right| < \frac{1}{2}.$$

That is, the distance between the center $(\frac{1}{4}, \frac{\sqrt{3}}{4})$ of $\mathfrak{D}(\infty, 0)$ and the center $(\frac{1}{2}\alpha_i, \frac{1}{2}\sqrt{1 - \alpha_i^2})$ of $\mathfrak{D}(r_i, 0)$ is always less than $\frac{1}{2}$. Thus these complex discs are always have a common region for any $r_i > 0$, that is, $\cap_{r_i > 0} \mathfrak{D}(r_i, 0)$ is not empty. There exists a fixed parameter μ such that $\|A_i\|_{H, \infty} < 1$ for any $r_i > 0$. By taking $\mu_1 = \iota/2$ in (3.13), one has

$$\|A_i\|_{H, \infty} = \frac{2|\mu_1| |\mu_1 - \alpha_i|}{|\mu_1 - \bar{\mu}_1|} = \sqrt{\alpha_i^2 + \frac{1}{4}} \quad \text{for } r_i > 0.$$

Since $\alpha_i \in (0, \frac{1}{2})$ for any $r_i > 0$, one gets $\|A_i\|_{H, \infty} \in (\frac{1}{2}, \frac{1}{\sqrt{2}})$. On the other hand, taking $\mu_2 = 1/4 + \iota/4$ in (3.13) arrives at

$$\|A_i\|_{H, \infty} = \frac{2|\mu_2| |\mu_2 - \alpha_i|}{|\mu_2 - \bar{\mu}_2|} = \sqrt{2\left(\frac{1}{4} - \alpha_i\right)^2 + \frac{1}{8}} \quad \text{for } r_i > 0.$$

For any $r_i > 0$, we have $\|A_i\|_{H, \infty} \in (\frac{\sqrt{2}}{4}, \frac{1}{2})$, which are less than $1/2$. Always, the estimate (3.10) implies the globally asymptotic decaying of DOC kernels, as shown by Lemma 2.1.

3.2 Stability analysis

To avoid the undesired equation (3.16) which has two different positive roots, and to simplify the subsequent mathematical derivations, we fix the complex parameter

$$\mu = \mu_* := 1/2 + \iota/2,$$

which is very close to the optimal tangential point $\tilde{\mu}_* \approx 0.4979 + 0.5454\iota$ in the heuristic analysis. Reminding the condition (3.15), we determine the maximum step-ratio R_3 by

$$\left| 1 - \alpha(R_3, R_3) + \iota \left(1 - \sqrt{(1 + \beta(R_3, R_3))^2 - \alpha(R_3, R_3)^2} \right) \right| = 1 - \beta(R_3, R_3),$$

which leads to the polynomial equation

$$R_3^6 + R_3^5 - 4R_3^4 - 8R_3^3 - 10R_3^2 - 6R_3 - 2 = 0.$$

It has a unique positive root $R_3 \approx 2.553$. Now we prove the following result.

Lemma 3.1. *If the step ratios $r_k < R_3$ ($k \geq 2$), there exists a constant $C_3 > 0$, independent of the time t_n , such that the DOC kernels $\vartheta_{n-j}^{(3,n)}$ in (1.11) satisfy*

$$\sum_{j=3}^n |\vartheta_{n-j}^{(3,n)}| \leq C_3 \quad \text{and} \quad \sum_{j=i}^n |\vartheta_{j-i}^{(3,j)}| \leq C_3 \quad \text{for } n \geq 3 \ (i \geq 3).$$

Proof. The definitions (1.1) and (1.10) give $d_0^{(3,j)} = d_0(r_j, r_{j-1}) \geq 1$ for $j \geq 3$. By using (3.1), (3.3) and the coefficient bound (3.8), one gets

$$\vartheta_0^{(3,j)} = \frac{1}{d_0^{(3,j)}} \leq 1 \quad \text{and} \quad \vartheta_1^{(3,j+1)} = \frac{\alpha_{j+1}}{d_0^{(3,j)}} \leq \frac{2}{d_0^{(3,j)}} \leq 2, \quad \text{for } j \geq 3.$$

That is, the first two kernels are always positive and bounded so that

$$\sum_{j=3}^n |\vartheta_{n-j}^{(3,n)}| \leq 3 \quad \text{and} \quad \sum_{j=i}^n |\vartheta_{j-i}^{(3,j)}| \leq 3 \quad \text{for } n = 3, 4 \ (i \geq 3).$$

Now consider the general case $n \geq 5$. Taking $\mu = \mu^*$ in the condition (3.14) yields

$$|2\beta_i - \alpha_i + \iota(1 - \alpha_i)| < 1 - \beta_i \quad \text{for } i \geq 3,$$

or equivalently,

$$2\alpha_i^2 + 3\beta_i^2 - 4\alpha_i\beta_i - 2\alpha_i + 2\beta_i < 0 \quad \text{for } i \geq 3.$$

Thanks to Lemma A.3, they are valid if all of step-ratios $0 < r_k < R_3$ ($k \geq 2$). According to (3.6) and (3.13), the elliptic norm $\|A_i\|_{H,\infty}$ is a continuous function with respect to $r_i, r_{i-1} \in [\epsilon, R_3 - \epsilon]$ for arbitrary small $\epsilon > 0$. Thus there exists a constant $\delta \in (0, 1)$ such that

$$\|A_i\|_{H,\infty} \leq \delta < 1 \quad \text{for } i \geq 3.$$

By taking the parameter $\mu = \mu^*$ in (3.9) and (3.13), one has $\|H\|_\infty = 2$ and $\|H^{-1}\|_\infty = 1 + \frac{\sqrt{2}}{2}$. Moreover, by (A.4)-(A.5), one has $\|\vec{u}_{j+1}\|_\infty \leq \max\{\alpha(R_3, R_3), 1\} = \alpha(R_3, R_3)$. Thus, the estimate (3.10) gives

$$|\vartheta_{n-j}^{(3,n)}| d_0^{(3,j)} \leq \|\vec{u}_n\|_\infty \leq c_R \delta^{n-j-1} \quad \text{for } n \geq j+2 \geq 5,$$

where $c_R = (2 + \sqrt{2})\alpha(R_3, R_3)$, and then

$$|\vartheta_{n-j}^{(3,n)}| \leq c_R \delta^{n-j-1} \quad \text{for } n \geq j+2 \geq 5. \quad (3.17)$$

Therefore it follows that

$$\begin{aligned} \sum_{j=3}^n |\vartheta_{n-j}^{(3,n)}| &\leq c_R \sum_{j=3}^n \delta^{n-j-1} < \frac{c_R}{\delta(1-\delta)} \quad \text{for } n \geq 5, \\ \sum_{j=i}^n |\vartheta_{j-i}^{(3,j)}| &\leq c_R \sum_{j=i}^n \delta^{j-i-1} < \frac{c_R}{\delta(1-\delta)} \quad \text{for } n \geq i \geq 3. \end{aligned}$$

We obtain the desired estimates by taking a constant $C_3 := \frac{c_R}{\delta(1-\delta)}$, which is independent of the time-level index n . This proof is complete. \square

Remark 3. The elliptic type norm in (3.9) admits some other nonsingular matrices H . Here consider a simple case with

$$H_0 := \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \quad \text{such that} \quad H_0^{-1} A_i H_0 = \begin{pmatrix} \alpha_i - \beta_i & -\beta_i \\ 1 - \alpha_i + \beta_i & \beta_i \end{pmatrix}$$

Let $R_{3,0} := \frac{1}{3} + \frac{1}{3} \sqrt[3]{19 + 3\sqrt{33}} + \frac{1}{3} \sqrt[3]{19 - 3\sqrt{33}} \approx 1.839$ be the unique positive root of the cubic equation $R_{3,0}^3 - R_{3,0}^2 - R_{3,0} - 1 = 0$. If the step-ratios satisfy $0 < r_k < R_{3,0}$ (it is also superior to the existing mesh conditions in the literature [2, 6, 8]), one can follow the proofs of Lemmas A.1-A.2 to check that

$$2\beta_i < \alpha_i < 1 \quad \text{for } i \geq 3.$$

It follows that

$$\|A_i\|_{H_0, \infty} = \max\{\alpha_i, 1 - \alpha_i + 2\beta_i\} < 1 \quad \text{for } i \geq 3,$$

and then the DOC kernels $\vartheta_{n-j}^{(3,n)}$ are globally decaying. Although the proof of Lemma 3.1 is technically complex, the matrix H in (3.9) would be optimal in the sense that the companion matrix A_i always has a pair of conjugate complex eigenvalues. Actually, the inequality $\alpha_i^2 < 4\beta_i$ holds when the adjacent step ratios r_i, r_{i-1} are close to 1, cf. Remark 4.

Remark 4. Consider the uniform time-step $\tau_k = \tau$, the BDF3 kernels (1.10) give

$$d_0^{(3,n)} = \frac{11}{6}, \quad d_1^{(3,n)} = -\frac{7}{6}, \quad d_2^{(3,n)} = \frac{1}{3} \quad \text{and} \quad d_j^{(3,n)} = 0 \quad \text{for } j \geq 3.$$

The difference equation (3.4) becomes

$$\widehat{\vartheta}_{n-j}^{(3,n)} - \frac{7}{11} \widehat{\vartheta}_{n-1-j}^{(3,n-1)} + \frac{2}{11} \widehat{\vartheta}_{n-2-j}^{(3,n-2)} = 0 \quad \text{for } n \geq j + 2 \geq 5.$$

Since the associated characteristic equation $\lambda^2 - \frac{7}{11}\lambda + \frac{2}{11} = 0$ has two roots $\lambda_{1,2} = (7 \pm i\sqrt{39})/22$, we have the following explicit formula of DOC kernels

$$\vartheta_{n-j}^{(3,n)} = \frac{11i}{d_0^{(3,j)}\sqrt{39}} \left[\left(\frac{7 - i\sqrt{39}}{22} \right)^{n-j+1} - \left(\frac{7 + i\sqrt{39}}{22} \right)^{n-j+1} \right] \quad \text{for } n \geq j + 2 \geq 5.$$

In Figure 1, we fix $n = 30$ and list the values of DOC kernels $\vartheta_j^{(3,n)}$ for different choices (let $0 < \epsilon_k < 1$ be the uniformly distributed random number):

- (a) the uniform mesh with $r_k = 1$ for $2 \leq k \leq 30$;
- (b) the random mesh with $r_k := R_3 \epsilon_k$ for $2 \leq k \leq 30$;
- (c) the random mesh with $r_k := 3\epsilon_k$ for $2 \leq k \leq 30$.

As observed, the DOC kernels $\vartheta_{n-j}^{(3,n)}$ are not always positive, but they always decay fast, as predicted by (3.17). It is interesting to find that, from the case (c), the DOC kernels maintain the fast damping property as a whole although there are many of step ratios greater than R_3 (about 15% in the above test).

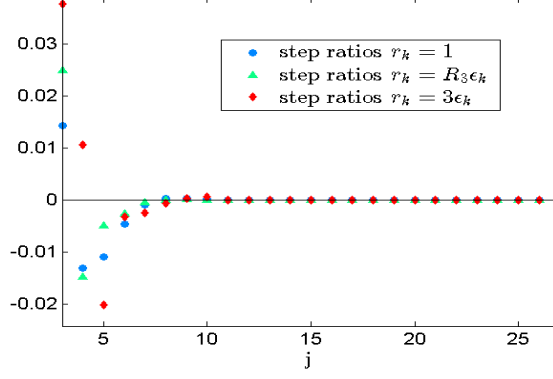


Figure 1: The DOC kernels $\vartheta_j^{(3,n)}$ ($j \geq 2$) for different step-ratio patterns.

Theorem 3.1. *If $\tau \leq 1/(4C_3L_f)$, the BDF3 solution of (1.7) satisfies*

$$|\tilde{v}^n| \leq 2 \exp(4C_3L_f t_{n-1}) \left(|\tilde{v}^2| + 5C_3\tau |\partial_\tau \tilde{v}^2| + 2C_3\tau |\partial_\tau \tilde{v}^1| + 2C_3t_n \max_{3 \leq i \leq n} |\varepsilon^i| \right) \quad \text{for } 3 \leq n \leq N.$$

Thus the variable-step BDF3 scheme is stable if $0 < r_k < R_3$ for $k \geq 2$.

Proof. Multiplying both sides of (1.16) with $2\tau_n \tilde{v}^n$, one applies Lemma 3.1 to obtain that

$$\begin{aligned} |\tilde{v}^n|^2 - |\tilde{v}^{n-1}|^2 &\leq -2\tau_n \tilde{v}^n \mathcal{I}_3^n[\tilde{v}] + 2\tau_n \tilde{v}^n \sum_{i=3}^n \vartheta_{n-i}^{(3,n)} [f(t_i, \bar{v}^i) - f(t_i, v^i) + \varepsilon^i] \\ &\leq 2\tau_n |\tilde{v}^n| |\mathcal{I}_3^n[\tilde{v}]| + 2L_f \tau_n |\tilde{v}^n| \sum_{i=3}^n |\vartheta_{n-i}^{(3,n)}| |\tilde{v}^i| + 2C_3 |\tilde{v}^n| \tau_n \max_{3 \leq i \leq n} |\varepsilon^i|. \end{aligned}$$

By summing the above inequality for n from $n = 3$ to m , we have

$$|\tilde{v}^m|^2 \leq |\tilde{v}^2|^2 + 2 \sum_{j=3}^m \tau_j |\tilde{v}^j| |\mathcal{I}_3^j[\tilde{v}]| + 2L_f \sum_{j=3}^m \tau_j |\tilde{v}^j| \sum_{i=3}^j |\vartheta_{j-i}^{(3,j)}| |\tilde{v}^i| + 2C_3 \sum_{j=3}^m |\tilde{v}^j| \tau_j \max_{3 \leq i \leq j} |\varepsilon^i|.$$

Taking some integer m_1 ($2 \leq m_1 \leq m$) such that $|\tilde{v}^{m_1}| = \max_{2 \leq k \leq m} |\tilde{v}^k|$. Now we take $m := m_1$ in the above inequality and get

$$|\tilde{v}^m| \leq |\tilde{v}^{m_1}| \leq |\tilde{v}^2| + 2\tau \sum_{j=3}^m |\mathcal{I}_3^j[\tilde{v}]| + 2C_3L_f \sum_{j=3}^m \tau_j |\tilde{v}^j| + 2C_3t_m \max_{3 \leq i \leq m} |\varepsilon^i| \quad (3.18)$$

for $3 \leq m \leq N$. It remains to evaluate the error $\sum_{j=3}^m |\mathcal{I}_3^j[\tilde{v}]|$ from the starting values. Taking the index $k = 3$ in (1.15) gives

$$\mathcal{I}_3^n[\tilde{v}] = \partial_\tau \tilde{v}^2 \sum_{i=3}^n \vartheta_{n-i}^{(3,n)} d_{i-2}^{(3,i)} + \partial_\tau \tilde{v}^1 \sum_{i=3}^n \vartheta_{n-i}^{(3,n)} d_{i-1}^{(3,i)} \quad \text{for } n \geq 3.$$

Recalling the definition (1.10) of BDF3 kernels with the increasing property (A.1), it is easy to check that $d_2^{(3,3)} \leq d_2(R_3, R_3) \leq 2$ and $|d_1^{(3,3)}| + d_2^{(3,4)} \leq -d_1(R_3, R_3) + d_2(R_3, R_3) \leq 5$. Thus we apply Lemma 3.1 to get

$$\begin{aligned}
\sum_{j=3}^m |\mathcal{I}_3^j[\tilde{v}]| &\leq |\partial_\tau \tilde{v}^2| \sum_{j=3}^m \sum_{i=3}^j |\vartheta_{j-i}^{(3,j)}| |d_{i-2}^{(3,i)}| + |\partial_\tau \tilde{v}^1| \sum_{j=3}^m \sum_{i=3}^j |\vartheta_{j-i}^{(3,j)}| |d_{i-1}^{(3,i)}| \\
&= |\partial_\tau \tilde{v}^2| \sum_{i=3}^m |d_{i-2}^{(3,i)}| \sum_{j=i}^m |\vartheta_{j-i}^{(3,j)}| + |\partial_\tau \tilde{v}^1| \sum_{i=3}^m |d_{i-1}^{(3,i)}| \sum_{j=i}^m |\vartheta_{j-i}^{(3,j)}| \\
&\leq C_3 |\partial_\tau \tilde{v}^2| (|d_1^{(3,3)}| + d_2^{(3,4)}) + C_3 |\partial_\tau \tilde{v}^1| d_2^{(3,3)} \\
&\leq 5C_3 |\partial_\tau \tilde{v}^2| + 2C_3 |\partial_\tau \tilde{v}^1| \quad \text{for } m \geq 3.
\end{aligned}$$

It follows from (3.18) that

$$|\tilde{v}^n| \leq |\tilde{v}^2| + 5C_3\tau |\partial_\tau \tilde{v}^2| + 2C_3\tau |\partial_\tau \tilde{v}^1| + 2C_3L_f \sum_{j=3}^n \tau_j |\tilde{v}^j| + 2C_3t_n \max_{3 \leq i \leq n} |\varepsilon^i|.$$

Assuming that $\tau \leq 1/(4C_3L_f)$, one has

$$|\tilde{v}^n| \leq 2|\tilde{v}^2| + 10C_3\tau |\partial_\tau \tilde{v}^2| + 4C_3\tau |\partial_\tau \tilde{v}^1| + 4C_3L_f \sum_{j=3}^{n-1} \tau_j |\tilde{v}^j| + 4C_3t_n \max_{3 \leq i \leq n} |\varepsilon^i|$$

for $3 \leq n \leq N$. The standard Grönwall inequality completes the proof. \square

From the proof of Lemma 3.1, the uniform boundedness of $\sum_{j=1}^n |\vartheta_{n-j}^{(3,n)}|$ and $\sum_{j=i}^n |\vartheta_{j-i}^{(3,j)}|$ does not require all DOC kernels to decay rapidly. It always allows a finite number of bounded DOC kernels. As seen from the numerical tests in the next section, the imposed step-ratio condition $0 < r_k < R_3$ in Theorem 3.1 is sufficient, but far from necessary. A practical step-ratio constraint for stability is that

$$\textit{almost all step ratios satisfy } 0 < r_k < R_3 \textit{ for } 2 \leq k \leq N.$$

4 Numerical tests

Consider an ODE model $v' = 2v - 3e^{-t}$ for $0 < t \leq 1$ with a smooth solution $v = \exp(-t)$ in our numerical tests. We run the variable-step BDF2 and BDF3 schemes in two scenarios:

- (a) The graded meshes $t_k = (k/N)^\gamma$ for $0 \leq k \leq N$ with grading parameters $\gamma > 1$. The maximum step-ratio is $r_{\max} = r_2 = 2^\gamma - 1$ and $\tau/\tau_1 = \tau_N/\tau_1 = N^\gamma - (N-1)^\gamma \approx \gamma N^{\gamma-1}$.
- (b) The random meshes with time-steps $\tau_k = \epsilon_k / \sum_{k=1}^N \epsilon_k$, where $\epsilon_k \in (0, 1)$ are uniformly distributed random numbers.

To start the multi-step schemes, a two-stage third-order singly diagonally implicit Runge-Kutta method is employed. We record the maximum error $e(N) := \max_{1 \leq n \leq N} |v(t_n) - v^n|$ in each run and compute the convergence order by

$$\text{Order} \approx \frac{\log(e(N)/e(2N))}{\log(\tau(N)/\tau(2N))}$$

Table 1: Numerical results of BDF2 method on graded time meshes

N	$\frac{\tau}{\tau_1}$	$\gamma = 2, r_{\max} = 3$		$\frac{\tau}{\tau_1}$	$\gamma = 3, r_{\max} = 7$		$\frac{\tau}{\tau_1}$	$\gamma = 4, r_{\max} = 15$	
		$e(N)$	Order		$e(N)$	Order		$e(N)$	Order
40	7.9e+01	5.28e-04	–	4.7e+03	8.77e-04	–	2.5e+05	1.35e-03	–
80	1.6e+02	1.34e-04	1.97	1.9e+04	2.25e-04	1.96	2.0e+06	3.49e-04	1.95
160	3.2e+02	3.39e-05	1.99	7.6e+04	5.72e-05	1.98	1.6e+07	8.91e-05	1.97
320	6.4e+02	8.52e-06	1.99	3.1e+05	1.44e-05	1.99	1.3e+08	2.25e-05	1.98
640	1.3e+03	2.14e-06	2.00	1.2e+06	3.61e-06	1.99	1.1e+09	5.66e-06	1.99
1280	2.6e+03	5.34e-07	2.00	4.9e+06	9.06e-07	2.00	8.4e+09	1.42e-06	2.00

Table 2: Numerical results of BDF3 method on graded time meshes

N	$\frac{\tau}{\tau_1}$	$\gamma = 2, r_{\max} = 3$		$\frac{\tau}{\tau_1}$	$\gamma = 3, r_{\max} = 7$		$\frac{\tau}{\tau_1}$	$\gamma = 4, r_{\max} = 15$	
		$e(N)$	Order		$e(N)$	Order		$e(N)$	Order
40	7.9e+01	1.27e-05	–	4.7e+03	2.94e-05	–	2.5e+05	5.73e-05	–
80	1.6e+02	1.65e-06	2.95	1.9e+04	3.91e-06	2.91	2.0e+06	7.85e-06	2.87
160	3.2e+02	2.10e-07	2.97	7.6e+04	5.05e-07	2.96	1.6e+07	1.03e-06	2.94
320	6.4e+02	2.65e-08	2.99	3.1e+05	6.41e-08	2.98	1.3e+08	1.31e-07	2.97
640	1.3e+03	3.32e-09	2.99	1.2e+06	8.07e-09	2.99	1.1e+09	1.66e-08	2.98
1280	2.6e+03	4.16e-10	3.00	4.9e+06	1.01e-09	2.99	8.4e+09	2.08e-09	2.99

where $\tau(N)$ denotes the maximum time-step size for total N subintervals.

Tables 1-2 list the numerical results of the BDF2 and BDF3 methods on graded meshes with three grading parameters, respectively. We see that, although the ratio τ/τ_1 increases as fast as $O(N^{\gamma-1})$, the numerical solutions remains stable and convergent with full accuracy.

Table 3: Numerical results of BDF2 method on random time meshes

N	$\tau(N)$	$e(N)$	Order	r_{\max}	τ/τ_1
40	4.49e-02	5.86e-04	–	28.30	1.22
80	2.40e-02	1.43e-04	2.03	91.41	1.06
160	1.18e-02	4.09e-05	1.81	32.54	19.88
320	6.18e-03	1.03e-05	1.99	418.41	1.21
640	3.01e-03	2.51e-06	2.04	604.02	2.66
1280	1.55e-03	6.63e-07	1.92	1963.80	1.02

Tables 3-4 record the numerical results on random time meshes. Table 3 suggests that the variable-step BDF2 method is robust with respect to the step-ratios r_k , and supports the theoretical findings in Theorem 2.1. Reminding the step-ratio restriction $0 < r_k < 2.533$ in Theorem 3.1 for the BDF3 method, we also record the number (denoted by N_1 in Table 4)

Table 4: Numerical results of BDF3 method on random time meshes

N	$\tau(N)$	$e(N)$	Order	r_{\max}	N_1	τ/τ_1
40	4.60e-02	8.20e-06	–	7.58	9	1.57
80	2.46e-02	1.18e-06	2.79	361.49	18	1.93
160	1.22e-02	1.83e-07	2.69	1682.18	35	4.51
320	6.11e-03	2.40e-08	2.93	79.90	60	2.91
640	3.27e-03	3.40e-09	2.82	5765.00	146	1.20
1280	1.56e-03	4.16e-10	3.03	9677.92	250	6.64

of time levels with $r_k \geq 2.533$. It is seen that the variable-step BDF3 method is stable and third-order convergent on random meshes, even if there are many of (about 20% in our tests) step-ratios do not meet our theoretical condition. Nonetheless, it remains mysterious to us.

5 Concluding remarks

The stability of BDF2 and BDF3 methods with unequal time-steps is investigated by a new theoretical framework using the discrete orthogonal convolution kernels. Thanks to the global analysis, that is, the present technique formulates the current BDF solution as a convolution summation of all previous information with DOC kernels as the convolutional weights, see the form (1.16), we improved the setp-ratio constraints for the stability. It is to mention that, the global decaying estimates of DOC kernels are also critical in the numerical analysis of BDF methods for parabolic equations, cf. [11,12,14]. The stability and convergence theory of variable-step BDF3 scheme for these stiff problems will be addressed in forthcoming reports.

For the high-order BDF-k ($k = 4, 5, 6$) time-stepping methods, we can also define the associated DOC kernels $\vartheta_{n-j}^{(k,n)}$ via the recursive procedure (1.11). From the proof of Theorem 3.1, one needs a result similar to Lemma 3.1 under certain step-ratio condition. Actually, the uniform boundedness (there exists a constant C_k independent of the time-level indexes n) of the absolute summations of DOC kernels

$$\sum_{j=k}^n |\vartheta_{n-j}^{(k,n)}| \leq C_k \quad \text{and} \quad \sum_{j=i}^n |\vartheta_{j-i}^{(k,j)}| \leq C_k \quad \text{for any } n \geq i \geq k,$$

is sufficient to the stability of BDF-k schemes for initial value problems. However, this issue seems to be theoretically challenging for $k \geq 4$ and remains open to us.

Acknowledgements

The authors would like to thank Dr. Ji Bingquan for his valuable discussions.

A Two auxiliary functions for the BDF3 method

By using the definitions (1.1)-(1.3), it is easy to check that

$$\frac{\partial}{\partial x} |d_\nu(x, y)| > 0 \quad \text{and} \quad \frac{\partial}{\partial y} |d_\nu(x, y)| > 0 \quad \text{for } \nu = 0, 1, 2 \text{ and } x, y > 0. \quad (\text{A.1})$$

That is, the functions $d_0(x, y)$, $-d_1(x, y)$ and $d_2(x, y)$ are increasing with respect to $x, y > 0$. We consider two auxiliary functions for the analysis of variable-step BDF3 method,

$$\alpha(x, y) := -\frac{d_1(x, y)}{d_0(x, y)} = \frac{x(x^2y^2 + 4xy^2 + 3y^2 + 2xy + 3y + 1)}{(y+1)(3x^2y + 4xy + 2x + y + 1)}, \quad (\text{A.2})$$

$$\beta(x, y) := \frac{d_2(x, y)}{d_0(x, y)} = \frac{x(x+1)^2y^2}{(y+1)(3x^2y + 4xy + 2x + y + 1)}. \quad (\text{A.3})$$

Their properties will be examined with respect to two independent variables $x, y > 0$ due to the facts $\alpha_m = \alpha(r_m, r_{m-1})$ and $\beta_m = \beta(r_m, r_{m-1})$ for $m \geq 2$ according to (3.5).

The above functions α and β are strictly increasing with respect to $x, y > 0$. Actually, by simple but tedious calculations for (A.2) -(A.3), it is not difficult to check that

$$\frac{\partial \alpha}{\partial x} = \frac{(x+1)^2(3x^2 + 2x + 3)y^3 + 2(2x^3 + 5x^2 + 6x + 3)y^2 + (x+2)^2y + 1}{(y+1)(3x^2y + 4xy + 2x + y + 1)^2}, \quad (\text{A.4})$$

$$\frac{\partial \alpha}{\partial y} = \frac{x(x+1)^2(xy + y + 1)(3xy + 3y + 1)}{(y+1)^2(3x^2y + 4xy + 2x + y + 1)^2}, \quad (\text{A.5})$$

and

$$\frac{\partial \beta}{\partial x} = \frac{(x+1)y^2(3x^3y + 5x^2y + 4x^2 + 3xy + 3x + y + 1)}{(y+1)(3x^2y + 4xy + 2x + y + 1)^2}, \quad (\text{A.6})$$

$$\frac{\partial \beta}{\partial y} = \frac{x(x+1)^2y(3x^2y + 6xy + 4x + 2y + 2)}{(y+1)^2(3x^2y + 4xy + 2x + y + 1)^2} \quad \text{for } x, y > 0. \quad (\text{A.7})$$

We have the following results. Some of proofs are technically complex and the mathematical derivations have been checked carefully by a symbolic calculation software.

Lemma A.1. *It holds that*

$$0 < \beta(x, y) < \alpha(x, y) < 1 + \beta(x, y) \quad \text{for } x, y > 0.$$

Proof. Simple calculations lead to

$$\alpha(x, y) - \beta(x, y) = \frac{x(2xy + 2y + 1)}{3x^2y + 4xy + 2x + y + 1},$$

$$1 - \alpha(x, y) + \beta(x, y) = \frac{(x+1)(xy + y + 1)}{3x^2y + 4xy + 2x + y + 1}.$$

It completes the proof. □

Lemma A.2. *It holds that*

$$0 < \beta(x, y) < 1 \quad \text{for } 0 < x, y < \hat{R}_3 \approx 3.4405,$$

where \hat{R}_3 is the unique positive root of $\hat{R}_3^4 - 2\hat{R}_3^3 - 4\hat{R}_3^2 - 3\hat{R}_3 - 1 = 0$.

Proof. According to (A.6)-(A.7), it holds that

$$\beta(x, y) - 1 < \beta(\hat{R}_3, \hat{R}_3) - 1 = \frac{\hat{R}_3^4 - 2\hat{R}_3^3 - 4\hat{R}_3^2 - 3\hat{R}_3 - 1}{3\hat{R}_3^3 + 4\hat{R}_3^2 + 3\hat{R}_3 + 1} = 0 \quad \text{for } 0 < x, y < \hat{R}_3.$$

It completes the proof. □

Lemma A.3. *For $0 < x, y < R_3 \approx 2.553$, it holds that*

$$2\alpha^2(x, y) + 3\beta^2(x, y) - 4\alpha(x, y)\beta(x, y) - 2\alpha(x, y) + 2\beta(x, y) < 0.$$

Proof. Consider an auxiliary function, see Figure 2,

$$\begin{aligned} g(x, y) &:= \frac{1}{x(x+1)} \left[2\alpha(x, y)^2 + 3\beta(x, y)^2 - 4\alpha(x, y)\beta(x, y) - 2\alpha(x, y) + 2\beta(x, y) \right] \\ &= \frac{(3x^2y + 4xy + 2x + y + 1)}{\gamma(x, y)} \left[(x+1)^2 (x^2 + x - 4) y^4 - 2(4x^2 + 11x + 7) y^3 \right. \\ &\quad \left. - 2(2x^2 + 10x + 9) y^2 - 2(3x + 5)y - 2 \right] \quad \text{for } 0 \leq x, y < R_3, \end{aligned}$$

where

$$\gamma(x, y) := (y+1)^2 (3x^2y + 4xy + 2x + y + 1)^3.$$

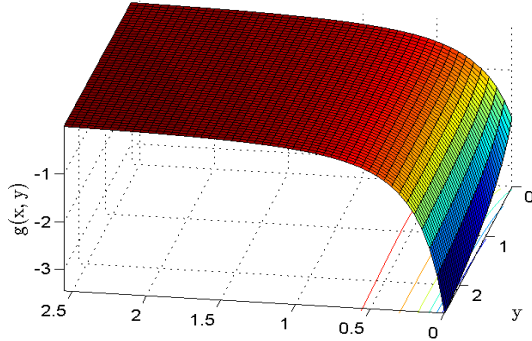


Figure 2: The surface of g on $[0, R_3]^2$.

Obviously, it is sufficient to show that $g(x, y) < 0$ for $0 < x, y < R_3$. Simple but tedious calculations yield

$$\begin{aligned} \frac{\partial g}{\partial x} &= \frac{1}{\gamma(x, y)} \left[2(27x^2 + 88x + 63) y^2 + 2(12x^3 + 90x^2 + 158x + 79) y^3 \right. \\ &\quad \left. + (4x^4 + 58x^3 + 207x^2 + 252x + 99) y^4 + (25 - x)(x+1)^3 y^5 + (36x + 50)y + 8 \right]. \end{aligned}$$

Obviously, we have $\frac{\partial g}{\partial x} > 0$ such that g is increasing with respect to x for $0 < x, y < R_3$. Thus it has no extreme points over the open square $(0, R_3)^2$. It remains to consider the maximum value along the four sides of $[0, R_3]^2$:

- (i) Along the side $y = 0$, we have $g(x, 0) = -\frac{2}{(2x+1)^2} < 0$ for $0 < x < R_3$.
- (ii) Along the side $y = R_3$, $g(x, R_3) < g(R_3, R_3) \approx -0.0000277 < 0$ for $0 < x < R_3$.
- (iii) Along the side $x = 0$, we have $g(0, y) = -\frac{4y+2}{y+1} < 0$ for $0 < y < R_3$.
- (iv) Along the side $x = R_3$, one has

$$g(R_3, y) \approx \frac{0.06763y^4 - 0.12922y^3 - 0.100507y^2 - 0.0267488y - 0.002113}{(y+1)^2(y+0.19847)^2} \quad \text{for } 0 < y < R_3.$$

It has a unique minimum point at $y \approx 0.103652$, while

$$g(R_3, y) < \max\{g(R_3, 0), g(R_3, R_3)\} = g(R_3, R_3) < 0 \quad \text{for } 0 < y < R_3.$$

In summary, we have $g(x, y) < 0$ for $0 < x, y < R_3$. It completes the proof. \square

References

- [1] G. ALBI, G. DIMARCO AND L. PARESCHI, *Implicit-explicit multistep methods for hyperbolic systems with multiscale relaxation*, SIAM J. Sci. Comput., 42(4) (2020) A2402–A2435 .
- [2] M. CALVO, T. GRANDE AND R.D. GRIGORIEFF, *On the zero stability of the variable order variable stepsize BDF-formulas*, Numer. Math., 57 (1990), 39–50.
- [3] M. CROUZEIX AND F.J. LISBONA, *The convergence of variable-stepsize, variable formula, multistep methods*, SIAM J. Numer. Anal., 21 (1984), 512–534.
- [4] V. DECARIA, A. GUZEL, W. LAYTON AND Y. LI, *A variable stepsize, variable order family of low complexity*, SIAM J. Sci. Comp., 43(3) (2021), A2130–A2160.
- [5] G. DIMARCO AND L. PARESCHI, *Implicit-explicit linear multistep methods for stiff kinetic equations*, SIAM J. Numer. Anal., 55 (2017), 664–690.
- [6] R.D. GRIGORIEFF, *Stability of multistep-methods on variable grids*, Numer. Math., 42 (1983), 359–377.
- [7] R.D. GRIGORIEFF AND P.J. PAES-LEME, *On the zero-stability of the 3-step BDF-formula on nonuniform grids*, BIT (1984), 24, 85–91.
- [8] N. GUGLIELMI AND M. ZENNARO, *On the zero-stability of variable stepsize multistep methods: the spectral radius approach*, Numer. Math., 88 (2001), 445–458.
- [9] E. HAIRER, S.P. NØRSETT AND G. WANNER, *Solving Ordinary Differential Equations I: Nonstiff Problems*, Volume 8 of Springer Series in Computational Mathematics, Second Edition, Springer-Verlag, 1992.

- [10] E. HAIRER AND G. WANNER, *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*, Springer Series in Computational Mathematics Volume 14, Second Edition, Springer-Verlag, 2002.
- [11] H.-L. LIAO, B. JI AND L. ZHANG, *An adaptive BDF2 implicit time-stepping method for the phase field crystal model*, IMA J. Numer. Anal., 2020, doi:10.1093/imanum/draa075.
- [12] H.-L. LIAO, B. JI, L. WANG AND Z. ZHANG, *Mesh-robustness of an energy stable BDF2 scheme with variable steps for the Cahn-Hilliard model*, arXiv:2102.03731v1, 2021.
- [13] H.-L. LIAO, T. TANG AND T. ZHOU, *On energy stable, maximum-bound preserving, second-order BDF scheme with variable steps for the Allen-Cahn equation*, SIAM J. Numer. Anal., 58(4) (2020), 2294-2314.
- [14] H.-L. LIAO AND Z. ZHANG, *Analysis of adaptive BDF2 scheme for diffusion equations*, Math. Comp., 90 (2021), 1207–1226.
- [15] D. WANG AND S.T. RUUTH, *Variable step-size implicit-explicit linear multistep methods for time-dependent partial differential equations*, J. Comput. Math., 26(6) (2008), 838–855.