Unified analysis for variational time discretizations of higher order and higher regularity applied to non-stiff ODEs

Simon Becher¹ · Gunar Matthies¹

Received: 19 May 2021 / Accepted: 28 June 2021 / Published online: 12 August 2021
© The Author(s) 2021

Abstract

We present a unified analysis for a family of variational time discretization methods, including discontinuous Galerkin methods and continuous Galerkin–Petrov methods, applied to non-stiff initial value problems. Besides the well-definedness of the methods, the global error and superconvergence properties are analyzed under rather weak abstract assumptions which also allow considerations of a wide variety of quadrature formulas. Numerical experiments illustrate and support the theoretical results.

Keywords Discontinuous Galerkin · Continuous Galerkin–Petrov · Variational methods · Quadrature · Higher order · Superconvergence

Mathematics Subject Classification (2010) 65L05 · 65L20 · 65L60

1 Introduction

We study variational time discretization methods applied to non-stiff initial value problems of the form

$$u'(t) = f(t, u(t)), \quad u(t_0) = u_0 \in \mathbb{R}^d, \quad (1.1)$$

where f is supposed to be sufficiently smooth and to satisfy a Lipschitz condition on the second variable. Here, an initial value u_0 is given at $t = t_0$. The system of ordinary differential equations (ODEs) is considered on a time interval $I = (t_0, t_0 + T]$ with arbitrary but fixed length $T > 0$.

✉ Gunar Matthies
Gunar.Matthies@tu-dresden.de

Simon Becher
Simon.Becher@tu-dresden.de

¹ TU Dresden, Institute of Numerical Mathematics, 01062 Dresden, Germany

Probably the most popular variational discretization schemes for solving (1.1) numerically are discontinuous Galerkin (dG) methods, cf. [15], and continuous Galerkin–Petrov (cGP) methods, cf. [13]. Both types of methods have been studied in the literature for many decades. However, an important part for the design of a then fully computable discretization, namely the use of quadrature methods for approximate integration, is often only marginally considered or the assumptions on the quadrature formulas are quite restrictive and, thus, allow a small variability of quadrature formulas only. So, for example, the number of quadrature points is supposed to equal the dimension of the local ansatz space of the discrete solution, cf. [6, 7], or the number of independent variational conditions, cf. [13, 14]. Often also the investigations are restricted to special Gauss, Gauss–Radau, or Gauss–Lobatto formulas. A quite general setting is considered in [9] where different ways to approximate the integrals over products of f with test functions are studied. Since in [9] the dynamical behavior of the schemes is analyzed, it is assumed that at least for Dahlquist’s test problem $u' = \lambda u$ all integrals are integrated exactly which again is quite restrictive.

In this paper, we try to keep the requirements on the approximation operators involved in the definition of the discrete method as low as possible. Our scope is to figure out how the approximation properties of these operators affect the error behavior of the numerical solution. Thereby, our studies are not restricted to the dG or cGP method but cover the whole family of variational discretization methods recently introduced in [3, 4]. The used assumptions are quite general and abstract in order to enable the study of a wide variety of methods. Especially, we allow that the right-hand side is approximated by a composition of various interpolation operators (interpolation cascade) or that the quadrature formulas also use derivative values. Therefore, all variational methods considered in [3] and their convergence results can be verified by a rigorous and unified error analysis in the non-stiff case now.

The paper is structured as follows. In Section 2, the variational time discretization methods are formulated. The well-definedness of the method formulation is studied in Section 3 where the existence of unique discrete solutions is proven under appropriate conditions. Section 4 is devoted to the error analysis. Here, at first, the global error is bounded by a sum of rather abstract approximation error terms. Then, in order to give a better insight into the error behavior, the convergence orders of these approximation errors are estimated in terms of basic quantities describing the approximation properties of the involved operators. Thereafter, superconvergence results in the time mesh nodes are derived in Section 5. Finally, in Section 6, we use numerical experiments to illustrate the convergence behavior of the variational time discretization methods and show that the proven estimates are sharp.

In order not to obscure the view on the main arguments, we have reduced the proofs to the essential ideas. For the technical details and a more detailed tracing of the constants, we refer the reader to the preprint version [2]. Specific references are given at the beginning of each proof.

Notation: In this paper, C denotes a generic constant independent of the time mesh interval length. To describe the vector-valued case ($d > 1$) in an easy way, let $(\cdot, \cdot)$ be the standard inner product and $\|\cdot\|$ the Euclidean norm on $\mathbb{R}^d$, $d \in \mathbb{N}$.

For an arbitrary interval J and positive integers q , the spaces of continuous and k times continuously differentiable $\mathbb{R}^q$ -valued functions on J are written as $C(J, \mathbb{R}^q)$

and $C^k(J, \mathbb{R}^q)$, respectively. For non-negative integers s , we write $P_s(J, \mathbb{R}^q)$ for the space of $\mathbb{R}^q$ -valued polynomials of degree less than or equal to s . Moreover, $P_{-1}(J, \mathbb{R}^q) := \{0\}$. For $q = 1$, we sometimes omit $\mathbb{R}^q$. Further notation will be introduced later at the beginning of the sections where it is needed.

2 Formulation of the methods

The interval I is decomposed by

$$t_0 < t_1 < \cdots < t_{N-1} < t_N = t_0 + T$$

into N disjoint subintervals $I_n := (t_{n-1}, t_n]$, $n = 1, \dots, N$. Moreover, we set

$$\tau_n := t_n - t_{n-1}, \quad \tau := \max_{1 \leq n \leq N} \tau_n.$$

For convenience and to simplify the notation in some proofs, we assume $\tau_n \leq 1$ which is not really a restriction since we are interested in the asymptotic error behavior for $\tau \rightarrow 0$. For any piecewise continuous function v , we define by

$$v(t_n^+) := \lim_{t \rightarrow t_n+0} v(t), \quad v(t_n^-) := \lim_{t \rightarrow t_n-0} v(t), \quad [v]_n := v(t_n^+) - v(t_n^-)$$

the one-sided limits and the jump of v at t_n . Furthermore, standard notation for the floor function is used.

We now present a slight generalization of the variational time discretization methods $\mathbf{VTD}_k^r$ investigated in [3]. Let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$. Then, the local version (I_n -problem) of the numerical method reads as follows:

Given $U(t_{n-1}^-) \in \mathbb{R}^d$, find $U \in P_r(I_n, \mathbb{R}^d)$ such that

$$U(t_{n-1}^+) = U(t_{n-1}^-), \quad \text{if } k \geq 1, \quad (2.1a)$$

$$U^{(i+1)}(t_n^-) = \left. \frac{d^i}{dt^i} (f(t, U(t))) \right|_{t=t_n^-}, \quad \text{if } k \geq 2, i = 0, \dots, \left\lfloor \frac{k}{2} \right\rfloor - 1, \quad (2.1b)$$

$$U^{(i+1)}(t_{n-1}^+) = \left. \frac{d^i}{dt^i} (f(t, U(t))) \right|_{t=t_{n-1}^+}, \quad \text{if } k \geq 3, i = 0, \dots, \left\lfloor \frac{k-1}{2} \right\rfloor - 1, \quad (2.1c)$$

and

$$\mathcal{J}_n \left[(U', \varphi) \right] + \delta_{0,k} ([U]_{n-1}, \varphi(t_{n-1}^+)) = \mathcal{J}_n \left[(\mathcal{I}_n f(\cdot, U(\cdot)), \varphi) \right] \quad \forall \varphi \in P_{r-k}(I_n, \mathbb{R}^d) \quad (2.1d)$$

where $U(t_0^-) = u_0$ and $\delta_{i,j}$ is the Kronecker symbol.

Here, $\mathcal{J}_n$ denotes an integrator that typically represents either the integral over I_n or the application of a quadrature formula for approximate integration. Details will be described later on. Moreover, $\mathcal{I}_n$ could be used to model some projection/interpolation of f or the usage of some special quadrature rules even if $\mathcal{J}_n$ is just the integral.

The problem class (2.1) generalizes discontinuous Galerkin methods (corresponding to $k = 0$) and continuous Galerkin methods (corresponding to $k = 1$). The conditions (2.1b) and (2.1c) are higher order collocation conditions at the end

points of the subintervals involving derivatives of the considered system of ordinary differential equations. For more details, we refer to [3, Remark 2.1].

3 Existence and uniqueness

First of all, we agree that both $\mathcal{J}_n$ and $\mathcal{I}_n$ are local versions (obtained by transformation) of appropriate linear operators $\widehat{\mathcal{J}}$ and $\widehat{\mathcal{I}}$ given on the closure of the reference interval $(-1, 1]$. However, $\mathcal{J}_n$ is an approximation of the integral operator while $\mathcal{I}_n$ approximates the identity operator. Thus, the transformations scale quite differently. More precisely, let

$$T_n : (-1, 1] \rightarrow I_n, \quad \hat{t} \mapsto \frac{t_n + t_{n-1}}{2} + \frac{\tau_n}{2} \hat{t}, \quad (3.1)$$

denote the affine transformation that maps the reference interval $(-1, 1]$ on an arbitrary mesh interval $I_n = (t_{n-1}, t_n]$. Furthermore, let $k_{\mathcal{J}}$ and $k_{\mathcal{I}}$ be the smallest non-negative integers such that $\widehat{\mathcal{J}}$ and $\widehat{\mathcal{I}}$ are well-defined for functions in $C^{k_{\mathcal{J}}}([-1, 1])$ and $C^{k_{\mathcal{I}}}([-1, 1])$, respectively. Then, we have for all $\varphi \in C^{k_{\mathcal{J}}}(\bar{I}_n, \mathbb{R}^d)$ and for all $\psi \in C^{k_{\mathcal{I}}}(\bar{I}_n, \mathbb{R}^d)$ that

$$\mathcal{J}_n[\varphi] = \widehat{\mathcal{J}}[\varphi \circ T_n](T_n)' = \frac{\tau_n}{2} \widehat{\mathcal{J}}[\varphi \circ T_n] \quad \text{and} \quad \mathcal{I}_n\psi = (\widehat{\mathcal{I}}(\psi \circ T_n)) \circ T_n^{-1}.$$

Of course, these operators act component-wise when applied to vector-valued functions. Moreover, we suppose for all non-negative integers l that $\widehat{\mathcal{I}}\hat{v} \in C^l([-1, 1])$ holds for all $\hat{v} \in C^{\max\{k_{\mathcal{I}}, l\}}([-1, 1])$, i.e., $\widehat{\mathcal{I}}\hat{v}$ is at least as smooth as $\hat{v}$.

The study of existence and uniqueness of solutions to (2.1) as well as the later error analysis is strongly connected with the following operator. Let $\widehat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}} : C^{k_{\mathcal{J}}+1}([-1, 1]) \rightarrow P_r([-1, 1])$ where $k_{\mathcal{J}} := \max\{\lfloor \frac{k}{2} \rfloor - 1, k_{\mathcal{J}}, k_{\mathcal{I}}\}$ be defined by

$$(\widehat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}}\hat{v})^{(i)}(-1^+) = \hat{v}^{(i)}(-1^+), \quad \text{if } k \geq 1, i = 0, \dots, \lfloor \frac{k-1}{2} \rfloor, \quad (3.2a)$$

$$(\widehat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}}\hat{v})^{(i)}(+1^-) = \hat{v}^{(i)}(+1^-), \quad \text{if } k \geq 2, i = 1, \dots, \lfloor \frac{k}{2} \rfloor, \quad (3.2b)$$

$$\begin{aligned} \widehat{\mathcal{J}}[(\widehat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}}\hat{v})'\hat{\varphi}] + \delta_{0,k}\widehat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}}\hat{v}(-1^+)\hat{\varphi}(-1^+) \\ = \widehat{\mathcal{J}}[(\widehat{\mathcal{I}}(\hat{v}'))\hat{\varphi}] + \delta_{0,k}\hat{v}(-1^+)\hat{\varphi}(-1^+) \quad \forall \hat{\varphi} \in P_{r-k}([-1, 1]). \end{aligned} \quad (3.2c)$$

Assumption 1 As before let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$, be the parameters of the method. The integrator $\mathcal{J}$ is such that $\hat{\psi} \in P_{r-\max\{1,k\}}([-1, 1])$ and

$$\mathcal{J}\left[(1-\hat{t})^{\lfloor \frac{k}{2} \rfloor} (1+\hat{t})^{\lfloor \frac{k-1}{2} \rfloor} \hat{\psi}\hat{\varphi}\right] = 0 \quad \forall \hat{\varphi} \in P_{r-\max\{1,k\}}([-1, 1])$$

imply $\hat{\psi} \equiv 0$. Here, the absolute value in the exponent is needed only for $k = 0$.

Lemma 1 Let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$. Suppose that Assumption 1 holds. Then, $\widehat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}}$ given by (3.2) is well-defined. If $\widehat{\mathcal{I}}$ preserves polynomials up to degree $\tilde{r}$, then $\widehat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}}$ preserves polynomials up to degree $\min\{\tilde{r} + 1, r\}$.

Proof (for more details see [2, Lemma 3.1]) In order to show that the operator is well-defined, we need to verify that the

$$\left(\left\lfloor \frac{k-1}{2} \right\rfloor + 1\right) + \left\lfloor \frac{k}{2} \right\rfloor + (r-k+1) = \frac{k}{2} + \frac{k-1}{2} - \frac{1}{2} + 2 + r - k = r + 1$$

linear conditions of (3.2) uniquely define a polynomial in $P_r([-1, 1])$. In short, it remains to prove that $\hat{v} \equiv 0$ implies $\hat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}} \hat{v} \equiv 0$.

There are two cases. For $k = 0$ choosing in (3.2c) with $\hat{v} \equiv 0$ test functions of the form $\hat{\varphi} = (1 + \hat{t})\tilde{\varphi}$, $\tilde{\varphi} \in P_{r-1}([-1, 1])$, we get by Assumption 1 $(\hat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}} \hat{v})' \equiv 0$. From this and using (3.2c) for $\hat{v} \equiv 0$ with $\hat{\varphi} \equiv 1$, we then conclude $\hat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}} \hat{v}(-1^+) = 0$ and, thus, $\hat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}} \hat{v} \equiv 0$.

For $k \geq 1$ we see from (3.2a), (3.2b), both with $\hat{v} \equiv 0$, that $(\hat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}} \hat{v})'(\hat{t}) = (1 - \hat{t})^{\lfloor \frac{k}{2} \rfloor} (1 + \hat{t})^{\lfloor \frac{k-1}{2} \rfloor} \hat{\psi}(\hat{t})$ with $\hat{\psi} \in P_{r-k}([-1, 1])$. So, combining Assumption 1 and (3.2c) for $\hat{v} \equiv 0$ we gain $\hat{\psi} \equiv 0$. Thus, $(\hat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}} \hat{v})' \equiv 0$. So, using (3.2a) with $\hat{v} \equiv 0$ for $i = 0$, it again follows that $\hat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}} \hat{v} \equiv 0$.

Using the uniqueness of the approximation, the second statement can be easily verified. $\square$

Lemma 2 Let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$. Suppose that Assumption 1 holds. Then, there is a positive constant $\kappa_{r,k}$, independent of τ_n , such that

$$\begin{aligned} & \int_{I_n} \|\varphi'(t)\| \, dt \leq \tau_n \sup_{t \in I_n} \|\varphi'(t)\| \\ & \leq \kappa_{r,k} \left[\sum_{i=1}^{\lfloor \frac{k-1}{2} \rfloor} \left(\frac{\tau_n}{2}\right)^i \|\varphi^{(i)}(t_{n-1}^+)\| + \sum_{i=1}^{\lfloor \frac{k}{2} \rfloor} \left(\frac{\tau_n}{2}\right)^i \|\varphi^{(i)}(t_n^-)\| + \sum_{i=\delta_{0,k}}^{r-k} \|\mathcal{J}_n[\varphi'(t)(1 + T_n^{-1}(t))^i]\| \right] \end{aligned}$$

and

$$\sup_{t \in I_n} \|\varphi(t)\| \leq \min \left\{ \|\varphi(t_{n-1}^+)\|, \|\varphi(t_n^-)\| \right\} + \int_{I_n} \|\varphi'(t)\| \, dt$$

for all $\varphi \in P_r(I_n, \mathbb{R}^d)$.

Proof (for more details see [2, Lemma 3.4]) The second statement follows easily using the fundamental theorem of calculus and the triangle inequality.

In order to show the first statement, we pull back to the reference interval $(-1, 1]$. To this end, we use that via the affine transformation T_n of (3.1) any $\varphi \in P_{r-1}(I_n, \mathbb{R}^d)$ is associated with a $\hat{\varphi} \in P_{r-1}((-1, 1], \mathbb{R}^d)$ by

$$\hat{\varphi}(\hat{t}) := (\varphi \circ T_n)(\hat{t}) = \varphi(T_n(\hat{t})) \quad \forall \hat{t} \in (-1, 1].$$

On the finite dimensional function space $P_{r-1}((-1, 1], \mathbb{R}^d)$, however, we have by a norm equivalence that

$$\sup_{\hat{t} \in (-1, 1]} \|\hat{\psi}(\hat{t})\| \leq \frac{\kappa_{r,k}}{2} \left[\sum_{i=0}^{\lfloor \frac{k-1}{2} \rfloor - 1} \|\hat{\psi}^{(i)}(-1^+)\| + \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor - 1} \|\hat{\psi}^{(i)}(+1^-)\| + \sum_{i=\delta_{0,k}}^{r-k} \|\hat{\mathcal{J}}[\hat{\psi}(\hat{t})(1 + \hat{t})^i]\| \right]$$

for all $\widehat{\psi} \in P_{r-1}((-1, 1], \mathbb{R}^d)$ where $\kappa_{r,k} > 0$ is independent of τ_n . To see that the right-hand side terms really define a norm, note that there appear precisely those r conditions of (3.2) that determine $(\widehat{\mathcal{J}}^{\mathcal{J}, \text{Id}} \widehat{v})'$ uniquely which shows the positive definiteness on $P_{r-1}((-1, 1], \mathbb{R}^d)$. The wanted statement then follows from scaling and transformation arguments. $\square$

Lemma 3 *Let $r \in \mathbb{N}_0$. Suppose that Assumption 1 holds. Then, there is a positive constant κ_r , independent of τ_n , such that*

$$\sup_{t \in I_n} \|\varphi(t)\| \leq \kappa_r \sum_{i=0}^r \left\| \mathcal{J}_n \left[\varphi'(t) (1 + T_n^{-1}(t))^i \right] + \delta_{0,i} \varphi(t_{n-1}^+) \right\|$$

for all $\varphi \in P_r(I_n, \mathbb{R}^d)$.

Proof (for more details see [2, Lemma 3.5]) Similar to the proof of Lemma 2, the statement follows using a suitable norm equivalence on the finite dimensional space $P_r((-1, 1], \mathbb{R}^d)$ along with transformations via T_n given in (3.1) between the reference interval $(-1, 1]$ and the actual interval I_n . $\square$

In the following, we assume that the inverse inequalities

$$\sup_{t \in I_n} \|V^{(l)}(t)\| \leq C_{\text{inv}} \left(\frac{\tau_n}{2} \right)^{-l} \sup_{t \in I_n} \|V(t)\| \quad \forall V \in P_r(I_n, \mathbb{R}^d), \quad l \in \mathbb{N}, \quad (3.3)$$

hold true where $C_{\text{inv}} > 0$ is independent of τ_n but may depend on l and r . The proof is standard and uses transformations to together with norm equivalences on finite dimensional spaces on the reference interval.

For the proof of the existence and uniqueness of solutions of (2.1), we need some more assumptions.

Assumption 2 The reference integrator $\widehat{\mathcal{J}}$ satisfies

$$\|\widehat{\mathcal{J}}[\widehat{\varphi}]\| \leq \mathfrak{C}_0 \sum_{j=0}^{k_{\mathcal{J}}} \sup_{\hat{t} \in [-1, 1]} \|\widehat{\varphi}^{(j)}(\hat{t})\| \quad \forall \widehat{\varphi} \in C^{k_{\mathcal{J}}}([-1, 1], \mathbb{R}^d)$$

where, as before, $k_{\mathcal{J}} \geq 0$ is the smallest integer such that $\widehat{\mathcal{J}}$ is well-defined on $C^{k_{\mathcal{J}}}([-1, 1])$. This means by transformation that

$$\|\mathcal{J}_n[\varphi]\| \leq \mathfrak{C}_0 \frac{\tau_n}{2} \sum_{j=0}^{k_{\mathcal{J}}} \left(\frac{\tau_n}{2} \right)^j \sup_{t \in I_n} \|\varphi^{(j)}(t)\| \quad \forall \varphi \in C^{k_{\mathcal{J}}}(\bar{I}_n, \mathbb{R}^d)$$

holds for the local integrators $\mathcal{J}_n$, $1 \leq n \leq N$.

Assumption 3 Let $0 \leq l \leq k_{\mathcal{J}}$. The reference approximation operator $\widehat{\mathcal{L}}$ satisfies

$$\sup_{\hat{t} \in [-1, 1]} \left\| (\widehat{\mathcal{L}}\widehat{\varphi})^{(l)}(\hat{t}) \right\| \leq \mathfrak{C}_1 \sum_{j=0}^{\max\{k_{\mathcal{L}}, l\}} \sup_{\hat{t} \in [-1, 1]} \left\| \widehat{\varphi}^{(j)}(\hat{t}) \right\|$$

for $\widehat{\varphi} \in C^{\max\{k_{\mathcal{L}}, l\}}([-1, 1], \mathbb{R}^d)$ where, as before, $k_{\mathcal{L}} \geq 0$ is the smallest integer such that $\widehat{\mathcal{L}}$ is well-defined on $C^{k_{\mathcal{L}}}([-1, 1])$. This means by transformation that

$$\sup_{t \in I_n} \left\| (\mathcal{I}_n \varphi)^{(l)}(t) \right\| \leq \mathfrak{C}_1 \sum_{j=0}^{\max\{k_{\mathcal{L}}, l\}} \left(\frac{\tau_n}{2} \right)^{j-l} \sup_{t \in I_n} \left\| \varphi^{(j)}(t) \right\|$$

for $\varphi \in C^{\max\{k_{\mathcal{L}}, l\}}(\overline{I}_n, \mathbb{R}^d)$ holds for the local operators $\mathcal{I}_n$, $1 \leq n \leq N$.

Assumption 4 For $0 \leq i \leq k_{\mathcal{J}} = \max \left\{ \left\lfloor \frac{k}{2} \right\rfloor - 1, k_{\mathcal{J}}, k_{\mathcal{L}} \right\}$, it holds for sufficiently smooth functions v, w that

$$\left\| \frac{d^i}{dt^i} \left(f(t, v(t)) - f(t, w(t)) \right) \Big|_{t=s} \right\| \leq \mathfrak{C}_2 \sum_{l=0}^i \left\| (v - w)^{(l)}(s) \right\|$$

for almost every $s \in \overline{I} = [t_0, t_0 + T]$. Here $\mathfrak{C}_2$ depends on $k_{\mathcal{J}}$ and f .

Remark 4 Sufficient conditions for Assumption 4 would be

- (i) for $k_{\mathcal{J}} = 0$: f satisfies a Lipschitz condition on the second variable with constant $L > 0$,
- (ii) for $k_{\mathcal{J}} \geq 1$: f is affine linear in u , i.e., $f(t, u(t)) = A(t)u(t) + b(t)$, and $\|A(\cdot)\|_{C^{k_{\mathcal{J}}}} < \infty$. Then the inequality follows from Leibniz' rule for the i th derivative.
- (iii) In the literature, see [12, p. 74], there also appear conditions of the form

$$\sup_{t \in I, y \in \mathbb{R}^d} \left\| \frac{\partial}{\partial y} f^{(i)}(t, y) \right\| < \infty, \quad 0 \leq i \leq k_{\mathcal{J}},$$

where $f^{(i)}$ denotes the i th total derivative of f with respect to t in the sense of [12, p. 65]. These conditions may be weaker in some cases.

Since in general the constant $\mathfrak{C}_2$ is somewhat connected to the Lipschitz constant and, thus, to the stiffness of the ode system, the dependence on this constant is studied very thoroughly in the analysis.

Now, we investigate the solvability of the local problem (2.1).

Theorem 5 (Existence and uniqueness) *Let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$. We suppose that Assumptions 1, 2, 3, and 4 hold. Then, there is a constant $\gamma_{r,k} > 0$ multiplicatively depending on $\mathfrak{C}_2^{-1}$ but independent of n such that problem (2.1) has a unique solution for all $1 \leq n \leq N$ when $\tau_n \leq \gamma_{r,k}$.*

Proof (for more details see [2, Theorem 3.7]) Since the single local problems can be solved one after another, it suffices to prove that (2.1) determines a unique solution for a given initial value U_{n-1}^- .

In order to show this, we use the auxiliary mapping $g : P_r(I_n, \mathbb{R}^d) \rightarrow P_r(I_n, \mathbb{R}^d)$ given by

$$g(U)(t_{n-1}^+) = U_{n-1}^-, \quad \text{if } k \geq 1, \quad (3.4a)$$

$$g(U)^{(i+1)}(t_n^-) = \frac{d^i}{dt^i}(f(t, U(t)))|_{t=t_n^-}, \quad \text{if } k \geq 2, i = 0, \dots, \lfloor \frac{k}{2} \rfloor - 1, \quad (3.4b)$$

$$g(U)^{(i+1)}(t_{n-1}^+) = \frac{d^i}{dt^i}(f(t, U(t)))|_{t=t_{n-1}^+}, \quad \text{if } k \geq 3, i = 0, \dots, \lfloor \frac{k-1}{2} \rfloor - 1, \quad (3.4c)$$

$$\begin{aligned} \mathcal{J}_n[(g(U)'(t), \varphi(t))] + \delta_{0,k}(g(U)(t_{n-1}^+), \varphi(t_{n-1}^+)) \\ = \mathcal{J}_n[(\mathcal{I}_n f(t, U(t)), \varphi(t))] + \delta_{0,k}(U_{n-1}^-, \varphi(t_{n-1}^+)) \quad \forall \varphi \in P_{r-k}(I_n, \mathbb{R}^d). \end{aligned} \quad (3.4d)$$

To verify that g is well-defined, just follow the lines of the proof of Lemma 1.

It can be easily seen that for given U_{n-1}^- every fixed point of g is a solution of the local problem (2.1) and vice versa. To get the existence of a unique fixed point, Banach's fixed point theorem shall be applied. Therefore, we prove for $\tau_n \leq \gamma_{r,k}$ with $\gamma_{r,k} > 0$ to be defined that g is on $G_r(U_{n-1}^-) := \{U \in P_r(I_n, \mathbb{R}^d) : U(t_{n-1}^+) = U_{n-1}^- \text{ if } k \geq 1\}$ a contraction with respect to the supremum norm.

So, let $V, W \in G_r(U_{n-1}^-)$. Then, due to (3.4a) and Lemma 2 (if $k \geq 1$) or Lemma 3 (if $k = 0$), we have $(g(V) - g(W))(t_{n-1}^+) = 0$ (for $k \geq 1$) and

$$\begin{aligned} & \sup_{t \in I_n} \|g(V) - g(W)\| \\ & \leq C \left[\underbrace{\sum_{i=1}^{\lfloor \frac{k-1}{2} \rfloor} \left(\frac{\tau_n}{2}\right)^i \|(g(V) - g(W))^{(i)}(t_{n-1}^+)\|}_{\text{(I)}} + \underbrace{\sum_{i=1}^{\lfloor \frac{k}{2} \rfloor} \left(\frac{\tau_n}{2}\right)^i \|(g(V) - g(W))^{(i)}(t_n^-)\|}_{\text{(II)}} \right. \\ & \quad \left. + \underbrace{\sum_{i=0}^{r-k} \left\| \mathcal{J}_n \left[(g(V) - g(W))'(t) (1 + T_n^{-1}(t))^i \right] + \delta_{0,k} \delta_{0,i}(g(V) - g(W))(t_{n-1}^+) \right\|}_{\text{(III)}} \right]. \end{aligned} \quad (3.5)$$

In order to bound the summands of (I), we combine (3.4c), Assumption 4, and the inverse inequalities (3.3) to get for $1 \leq i \leq \lfloor \frac{k-1}{2} \rfloor$

$$\begin{aligned} & \|(g(V) - g(W))^{(i)}(t_{n-1}^+)\| = \left\| \frac{d^{i-1}}{dt^{i-1}}(f(t, V(t)) - f(t, W(t))) \right\|_{t=t_{n-1}^+} \\ & \leq \mathfrak{C}_2 \sum_{l=0}^{i-1} \|(V - W)^{(l)}(t_{n-1}^+)\| \leq \mathfrak{C}_2 C_{\text{inv}} i \left(\frac{\tau_n}{2}\right)^{-i+1} \sup_{t \in I_n} \|(V - W)(t)\|. \end{aligned}$$

The arguments for the summands of (II) are analogous. Altogether, we obtain

$$(I) + (II) \leq C\mathfrak{C}_2 \frac{\tau_n}{2} \sup_{t \in I_n} \|(V - W)(t)\|. \quad (3.6)$$

We now consider the third sum of (3.5). Exploiting (3.4d) we gain that

$$(III) = \sum_{i=0}^{r-k} \left\| \mathcal{I}_n \left[\left(\mathcal{I}_n f(t, V(t)) - \mathcal{I}_n f(t, W(t)) \right) (1 + T_n^{-1}(t))^i \right] \right\|.$$

Because of Assumption 2 and Leibniz' rule for the j th derivative, it follows

$$\begin{aligned} & \left\| \mathcal{I}_n \left[\left(\mathcal{I}_n f(t, V(t)) - \mathcal{I}_n f(t, W(t)) \right) (1 + T_n^{-1}(t))^i \right] \right\| \\ & \leq \mathfrak{C}_0 \frac{\tau_n}{2} \sum_{j=0}^{k_{\mathcal{I}}} \left(\frac{\tau_n}{2} \right)^j \sum_{l=0}^j \binom{j}{l} \sup_{t \in I_n} \left\| \frac{d^l}{dt^l} \left(\mathcal{I}_n f(t, V(t)) - \mathcal{I}_n f(t, W(t)) \right) \right\| \\ & \quad \sup_{t \in I_n} \left\| \left((1 + T_n^{-1}(t))^i \right)^{(j-l)} \right\| \\ & \leq C \frac{\tau_n}{2} \sum_{l=0}^{k_{\mathcal{I}}} \left(\frac{\tau_n}{2} \right)^l \sup_{t \in I_n} \left\| \frac{d^l}{dt^l} \left(\mathcal{I}_n f(t, V(t)) - \mathcal{I}_n f(t, W(t)) \right) \right\| \end{aligned} \quad (3.7)$$

where we used for the last inequality that

$$\sup_{t \in I_n} \left\| \left((1 + T_n^{-1}(t))^i \right)^{(j-l)} \right\| \leq \begin{cases} \left(\frac{2}{\tau_n} \right)^{j-l} \frac{i!}{(i-(j-l))!} 2^{i-(j-l)}, & 0 \leq j-l \leq i, \\ 0, & i < j-l. \end{cases}$$

Furthermore, involving Assumptions 3 and 4, and the inverse inequalities (3.3) we conclude

$$\begin{aligned} (III) & \leq C \frac{\tau_n}{2} \sum_{j=0}^{\max\{k_{\mathcal{I}}, k_T\}} \left(\frac{\tau_n}{2} \right)^j \sup_{t \in I_n} \left\| \frac{d^j}{dt^j} \left(f(t, V(t)) - f(t, W(t)) \right) \right\| \\ & \leq C\mathfrak{C}_2 \frac{\tau_n}{2} \sum_{j=0}^{\max\{k_{\mathcal{I}}, k_T\}} \left(\frac{\tau_n}{2} \right)^j \sum_{l=0}^j \sup_{t \in I_n} \left\| (V - W)^{(l)}(t) \right\| \\ & \leq C\mathfrak{C}_2 \frac{\tau_n}{2} \sup_{t \in I_n} \|(V - W)(t)\|. \end{aligned} \quad (3.8)$$

Now, combining (3.5), (3.6), and (3.8), we get

$$\sup_{t \in I_n} \|g(V) - g(W)\| \leq C\mathfrak{C}_2 \frac{\tau_n}{2} \sup_{t \in I_n} \|(V - W)(t)\|.$$

Hence, for sufficiently small $\tau_n \leq \gamma_{r,k}$, where $\gamma_{r,k}$ multiplicatively depends on $\mathfrak{C}_2^{-1}$ but is independent of n , the mapping g is on $G_r(U_{n-1}^-)$ a contraction with respect to the supremum norm. By Banach's fixed-point theorem, we have the existence of a unique fixed point which is just the unique solution of the local problem (2.1). $\square$

4 Error analysis

In order to prove an error estimate, we reuse the operator $\widehat{\mathcal{J}}^{\mathcal{J}, \mathcal{I}}$ introduced in the previous section. By transformation, we get for $n = 1, \dots, N$ local approximation operators $\mathcal{J}_n^{\mathcal{J}, \mathcal{I}} : C^{k_{\mathcal{J}}+1}(\overline{I}_n, \mathbb{R}^d) \rightarrow P_r(I_n, \mathbb{R}^d)$ satisfying

$$(\mathcal{J}_n^{\mathcal{J}, \mathcal{I}} v)^{(i)}(t_{n-1}^+) = v^{(i)}(t_{n-1}^+), \quad \text{if } k \geq 1, i = 0, \dots, \lfloor \frac{k-1}{2} \rfloor, \quad (4.1a)$$

$$(\mathcal{J}_n^{\mathcal{J}, \mathcal{I}} v)^{(i)}(t_n^-) = v^{(i)}(t_n^-), \quad \text{if } k \geq 2, i = 1, \dots, \lfloor \frac{k}{2} \rfloor, \quad (4.1b)$$

$$\begin{aligned} \mathcal{J}_n \left[\left((\mathcal{J}_n^{\mathcal{J}, \mathcal{I}} v)'(t), \varphi(t) \right) \right] + \delta_{0,k} \mathcal{J}_n^{\mathcal{J}, \mathcal{I}} v(t_{n-1}^+) \varphi(t_{n-1}^+) \\ = \mathcal{J}_n \left[(\mathcal{I}_n(v')(t), \varphi(t)) \right] + \delta_{0,k} v(t_{n-1}^+) \varphi(t_{n-1}^+) \quad \forall \varphi \in P_{r-k}(I_n, \mathbb{R}^d). \end{aligned} \quad (4.1c)$$

Of course, these operators are well-defined under Assumption 1.

For convenience, we define for $v \in C^{k_{\mathcal{J}}+1}(\overline{I}, \mathbb{R}^d)$ a global approximation operator $\mathcal{J}^{\mathcal{J}, \mathcal{I}}$ by combining the local approximations on the mesh intervals, i.e., $(\mathcal{J}^{\mathcal{J}, \mathcal{I}} v)|_{I_n} = \mathcal{J}_n^{\mathcal{J}, \mathcal{I}}(v|_{I_n}) \in P_r(I_n, \mathbb{R}^d)$ for $n = 1, \dots, N$ and setting $\mathcal{J}^{\mathcal{J}, \mathcal{I}} v(t_0^-) = v(t_0^-) = v(t_0)$. Note that even for $k \geq 1$, in general $\mathcal{J}^{\mathcal{J}, \mathcal{I}} v$ is not globally continuous.

In order to prove the error estimate we need to strengthen the assumptions since derivatives can be handled given in certain points, but not their supremum. Therefore, compared to Theorem 5, we replace Assumption 3 by Assumptions 5a or 5b. Furthermore, we need an auxiliary interpolation operator $\mathcal{I}^{\text{app}}$, see Definition 7, which amongst others is based on these assumptions.

For brevity, the Assumptions 5a and 5b below are stated directly for the local operators $\mathcal{J}_n$ and $\mathcal{I}_n$. However, appropriate properties of $\widehat{\mathcal{J}}$ and $\widehat{\mathcal{I}}$ guarantee these assumptions by transformation, cf. Assumptions 2 and 3.

Assumption 5a For $1 \leq n \leq N$ and $0 \leq l \leq k_{\mathcal{J}}$ it holds $\mathcal{I}_n \varphi \in C^l(\overline{I}_n, \mathbb{R}^d)$ and there are disjoint points $\hat{t}_m^{\mathcal{I}}, m = 0, \dots, K^{\mathcal{I}}$, in the reference interval $[-1, 1]$ such that

$$\left(\frac{\tau_n}{2}\right)^l \sup_{t \in I_n} \left\| (\mathcal{I}_n \varphi)^{(l)}(t) \right\| \leq \mathfrak{C}_{1,1} \sum_{m=0}^{K^{\mathcal{I}}} \sum_{j=0}^{\tilde{K}_m^{\mathcal{I}}} \left(\frac{\tau_n}{2}\right)^j \left\| \varphi^{(j)}(t_{n,m}^{\mathcal{I}}) \right\| + \mathfrak{C}_{1,2} \sup_{t \in I_n} \|\varphi(t)\|$$

for $\varphi \in C^{k_{\mathcal{I}}}(\overline{I}_n, \mathbb{R}^d)$ where $t_{n,m}^{\mathcal{I}} := \frac{t_n + t_{n-1}}{2} + \frac{\tau_n}{2} \hat{t}_m^{\mathcal{I}}$. Note that then typically $k_{\mathcal{I}} = \max \{ \tilde{K}_m^{\mathcal{I}} : m = 0, \dots, K^{\mathcal{I}} \}$.

Assumption 5b For $1 \leq n \leq N$, there are disjoint points $\hat{t}_m^{\mathcal{J}}, m = 0, \dots, K^{\mathcal{J}}$, in the reference interval $[-1, 1]$ such that

$$\|\mathcal{J}_n[\varphi]\| \leq \tilde{\mathfrak{C}}_{0,1} \frac{\tau_n}{2} \sum_{m=0}^{K^{\mathcal{J}}} \sum_{j=0}^{\tilde{K}_m^{\mathcal{J}}} \left(\frac{\tau_n}{2}\right)^j \left\| \varphi^{(j)}(t_{n,m}^{\mathcal{J}}) \right\| + \tilde{\mathfrak{C}}_{0,2} \frac{\tau_n}{2} \sup_{t \in I_n} \|\varphi(t)\|$$

for $\varphi \in C^{k_{\mathcal{J}}}(\bar{I}_n, \mathbb{R}^d)$ where $t_{n,m}^{\mathcal{J}} := \frac{t_n + t_{n-1}}{2} + \frac{\tau_n}{2} \hat{t}_m^{\mathcal{J}}$. Note that then typically $k_{\mathcal{J}} = \max \{\tilde{K}_m^{\mathcal{J}} : m = 0, \dots, K^{\mathcal{J}}\}$.

Moreover, for $1 \leq n \leq N$ assume that there are disjoint points $\hat{t}_m^{\mathcal{I}}$, $m = 0, \dots, K^{\mathcal{I}}$, in the reference interval $[-1, 1]$ such that

$$\begin{aligned} & \sum_{m=0}^{K^{\mathcal{J}}} \sum_{l=0}^{\tilde{K}_m^{\mathcal{J}}} \left(\frac{\tau_n}{2} \right)^l \left\| (\mathcal{I}_n \varphi)^{(l)}(t_{n,m}^{\mathcal{J}}) \right\| + \sup_{t \in I_n} \|\mathcal{I}_n \varphi(t)\| \\ & \leq \tilde{\mathfrak{C}}_{1,1} \sum_{m=0}^{K^{\mathcal{I}}} \sum_{j=0}^{\tilde{K}_m^{\mathcal{I}}} \left(\frac{\tau_n}{2} \right)^j \left\| \varphi^{(j)}(\hat{t}_{n,m}^{\mathcal{I}}) \right\| + \tilde{\mathfrak{C}}_{1,2} \sup_{t \in I_n} \|\varphi(t)\| \end{aligned}$$

for $\varphi \in C^{\max\{k_{\mathcal{J}}, k_{\mathcal{I}}\}}(\bar{I}_n, \mathbb{R}^d)$ where $t_{n,m}^{\mathcal{I}} := \frac{t_n + t_{n-1}}{2} + \frac{\tau_n}{2} \hat{t}_m^{\mathcal{I}}$.

Remark 6 Assumption 5a is satisfied if $\mathcal{I}_n$ is a polynomial approximation operator whose defining degrees of freedom only use derivatives in certain points, as, for example, Hermite interpolation operators. Together with Assumption 2, the term $\|\mathcal{J}_n[\mathcal{I}_n \varphi]\|$ can be estimated by the supremum of $\|\varphi\|$ and certain point values of derivatives of φ .

However, Assumption 5a is not satisfied if $\mathcal{I}_n = \text{Id}$ and $k_{\mathcal{J}} > 0$. In order to enable a similar estimate for $\|\mathcal{J}_n[\mathcal{I}_n \varphi]\|$ also in this case, Assumption 5b is formulated. Here, the requirements on the integrator $\mathcal{J}_n$ are increased. Of course, the defining degrees of freedom for the integrator now should use derivatives in certain points only. In return, the requirements for $\mathcal{I}_n$ can be weakened such that they are met for example also by $\mathcal{I}_n = \text{Id}$.

Definition 7 (Auxiliary interpolation operator) For the error estimation, we introduce a special Hermite interpolation operator $\mathcal{I}_n^{\text{app}}$. Concretely, the operator should satisfy the following conditions: $\mathcal{I}_n^{\text{app}}$ preserves derivatives up to order $\lfloor \frac{k}{2} \rfloor - 1$ in t_n^- and up to order $\lfloor \frac{k-1}{2} \rfloor - 1$ in t_{n-1}^+ , i.e.,

$$(\mathcal{I}_n^{\text{app}} \varphi)^{(l)}(t_n^-) = \varphi^{(l)}(t_n^-) \quad \text{for } 0 \leq l \leq \lfloor \frac{k}{2} \rfloor - 1, \quad (4.2a)$$

$$(\mathcal{I}_n^{\text{app}} \varphi)^{(l)}(t_{n-1}^+) = \varphi^{(l)}(t_{n-1}^+) \quad \text{for } 0 \leq l \leq \lfloor \frac{k-1}{2} \rfloor - 1. \quad (4.2b)$$

Moreover, we demand that

$$(\mathcal{I}_n^{\text{app}} \varphi)^{(l)}(t_{n,m}^{\mathcal{I}}) = \varphi^{(l)}(t_{n,m}^{\mathcal{I}}) \quad \text{for } 0 \leq m \leq K^{\mathcal{I}}, 0 \leq l \leq \tilde{K}_m^{\mathcal{I}}, \quad (4.3a)$$

with $t_{n,m}^{\mathcal{I}} := \frac{t_n + t_{n-1}}{2} + \frac{\tau_n}{2} \hat{t}_m^{\mathcal{I}}$ where the points $\hat{t}_m^{\mathcal{I}}$ are those of Assumptions 5a or 5b, respectively. If (4.2) and (4.3a) provide r^{app} independent interpolation conditions and $r^{\text{app}} < r + 1$, then we choose $r + 1 - r^{\text{app}}$ further points $\hat{t}_m^{\mathcal{I}} \in (-1, 1) \setminus \{\hat{t}_j^{\mathcal{I}} : j = 0, \dots, K^{\mathcal{I}}\}$, $m = K^{\mathcal{I}} + 1, \dots, K^{\mathcal{I}} + r + 1 - r^{\text{app}}$, and demand

$$(\mathcal{I}_n^{\text{app}} \varphi)(t_{n,m}^{\mathcal{I}}) = \varphi(t_{n,m}^{\mathcal{I}}) \quad \text{for } K^{\mathcal{I}} + 1 \leq m \leq K^{\mathcal{I}} + r + 1 - r^{\text{app}} \quad (4.3b)$$

where again $t_{n,m}^{\mathcal{I}} := \frac{t_n + t_{n-1}}{2} + \frac{\tau_n}{2} \hat{t}_m^{\mathcal{I}}$. We agree that $\mathcal{I}_n^{\text{app}}$ is applied component-wise to vector-valued functions. Overall, conditions (4.2) and (4.3) uniquely define an Hermite-type interpolation operator of ansatz order $\max\{r^{\text{app}} - 1, r\}$.

Now, we are able to prove an abstract error estimate.

Theorem 8 *Let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$. We suppose that Assumptions 1, 2, and 4 hold. Moreover, let Assumptions 5a or 5b be satisfied. Denote by u and U the solutions of (1.1) and (2.1), respectively. Then, we have for $1 \leq n \leq N$, sufficiently small τ , and $l = 0, 1$ that*

$$\sup_{t \in I_n} \|(u - U)^{(l)}(t)\| \leq C \max_{1 \leq v \leq n-1} \tau_v^{-1} \|(u - \mathcal{J}_v^{\mathcal{J}, \mathcal{I}} u)(t_v^-)\| \\ + C \max_{1 \leq v \leq n} \left(\sup_{t \in I_v} \|(\text{Id} - \mathcal{I}_v^{\text{app}})u(t)\| + \sum_{j=0}^l \sup_{t \in I_v} \|(u - \mathcal{J}_v^{\mathcal{J}, \mathcal{I}} u)^{(j)}(t)\| \right)$$

where the C s in general exponentially depend on the product of T and $\mathfrak{C}_2$.

Proof (for more details see [2, Theorem 4.5]) Using the approximation $\mathcal{J}^{\mathcal{J}, \mathcal{I}} u$, for a definition see directly below (4.1), we split the error in two parts

$$\eta(t) := u(t) - \mathcal{J}^{\mathcal{J}, \mathcal{I}} u(t), \quad \zeta(t) := \mathcal{J}^{\mathcal{J}, \mathcal{I}} u(t) - U(t).$$

Then, the triangle inequality yields

$$E_n := \sup_{t \in I_n} \|u(t) - U(t)\| \leq \sup_{t \in I_n} \|\eta(t)\| + \sup_{t \in I_n} \|\zeta(t)\|. \quad (4.4)$$

The first term on the right-hand side already is the approximation error of the operator $\mathcal{J}^{\mathcal{J}, \mathcal{I}}$. So, it only remains to study the second term.

Using, among others, the fundamental theorem of calculus we find

$$\sup_{t \in I_n} \|\zeta(t)\| \leq \underbrace{\|\zeta(t_0^-)\|}_{=0} + \sum_{v=1}^n \left(\int_{I_v} \|\zeta'(s)\| \, ds + \|\zeta\|_{v-1} \right) \quad (4.5)$$

where $\zeta(t_0^-) = 0$ follows due to $U(t_0^-) = u_0 = u(t_0^-) = \mathcal{J}^{\mathcal{J}, \mathcal{I}} u(t_0^-)$, see (2.1) and directly below (4.1).

We start analyzing the integral term of (4.5). For $1 \leq n \leq N$ and because $\zeta|_{I_n} \in P_r(I_n, \mathbb{R}^d)$ we can apply Lemma 2 to get

$$\int_{I_n} \|\zeta'(t)\| \, dt \leq \tau_n \sup_{t \in I_n} \|\zeta'(t)\| \quad (4.6) \\ \leq C \left[\underbrace{\sum_{i=1}^{\lfloor \frac{k-1}{2} \rfloor} \left(\frac{\tau_n}{2}\right)^i \|\zeta^{(i)}(t_{n-1}^+)\|}_{(Ia)} + \underbrace{\sum_{i=1}^{\lfloor \frac{k}{2} \rfloor} \left(\frac{\tau_n}{2}\right)^i \|\zeta^{(i)}(t_n^-)\|}_{(Ib)} + \underbrace{\sum_{i=\delta_{0,k}}^{r-k} \left\| \mathcal{J}_n \left[\zeta'(t) (1 + T_n^{-1}(t))^i \right] \right\|}_{(II)} \right].$$

The right-hand side terms are now studied separately.

The arguments for the sums (Ia) and (Ib) are quite analogous. We therefore only consider the latter in detail. From (4.1b), (2.1b), the $(i - 1)$ th derivative of the ode system (1.1), and Assumption 4 we obtain for $1 \leq i \leq \lfloor \frac{k}{2} \rfloor$

$$\begin{aligned} \|\zeta^{(i)}(t_n^-)\| &= \|(\mathcal{J}_n^{\mathcal{I}} \mathcal{I}u - U)^{(i)}(t_n^-)\| = \|(u - U)^{(i)}(t_n^-)\| \\ &= \left\| \frac{d^{i-1}}{dt^{i-1}} \left(f(t, u(t)) - f(t, U(t)) \right) \right\|_{t=t_n^-} \leq \mathfrak{C}_2 \sum_{l=0}^{i-1} \|(u - U)^{(l)}(t_n^-)\|. \end{aligned}$$

Exploiting the definition of $\mathcal{I}_n^{\text{app}}$, especially (4.2a), as well as invoking the inverse inequality (3.3), we find for $0 \leq l \leq i - 1 \leq \lfloor \frac{k}{2} \rfloor - 1$

$$\begin{aligned} \|(u - U)^{(l)}(t_n^-)\| &= \|(\mathcal{I}_n^{\text{app}} u - U)^{(l)}(t_n^-)\| \leq C_{\text{inv}} \left(\frac{\tau_n}{2} \right)^{-l} \sup_{t \in I_n} \|(\mathcal{I}_n^{\text{app}} u - U)(t)\| \\ &\leq C_{\text{inv}} \left(\frac{\tau_n}{2} \right)^{-l} \left(\sup_{t \in I_n} \|(\mathcal{I}_n^{\text{app}} u - u)(t)\| + \sup_{t \in I_n} \|(u - U)(t)\| \right). \end{aligned} \quad (4.7)$$

So, overall we conclude

$$(Ia) + (Ib) \leq C \mathfrak{C}_2 \frac{\tau_n}{2} \left(\sup_{t \in I_n} \|(u - \mathcal{I}_n^{\text{app}} u)(t)\| + E_n \right). \quad (4.8)$$

We now consider (II) in (4.6). First of all, from (4.1c), (2.1d), (1.1), and the continuity of u it follows for all $\varphi \in P_{r-k}(I_n, \mathbb{R}^d)$

$$\begin{aligned} &\mathcal{J}_n [\zeta'(t), \varphi(t)] + \delta_{0,k} (\zeta(t_{n-1}^+), \varphi(t_{n-1}^+)) \\ &= \mathcal{J}_n [(\mathcal{I}_n(u')(t) - U'(t), \varphi(t))] + \delta_{0,k} ((u - U)(t_{n-1}^+), \varphi(t_{n-1}^+)) \\ &= \mathcal{J}_n [(\mathcal{I}_n f(t, u(t)) - \mathcal{I}_n f(t, U(t)), \varphi(t))] + \delta_{0,k} ((u - U)(t_{n-1}^-), \varphi(t_{n-1}^+)). \end{aligned} \quad (4.9)$$

Therefore, by a component-wise derivation we get

$$(II) = \sum_{i=\delta_{0,k}}^{r-k} \left\| \mathcal{J}_n \left[(\mathcal{I}_n f(t, u(t)) - \mathcal{I}_n f(t, U(t))) (1 + T_n^{-1}(t))^i \right] \right\|. \quad (4.10)$$

For the summands on the right-hand side, we consider two different cases. If Assumption 5a holds, we conclude that

$$\begin{aligned} & \left\| \mathcal{J}_n \left[\left(\mathcal{I}_n f(t, u(t)) - \mathcal{I}_n f(t, U(t)) \right) (1 + T_n^{-1}(t))^i \right] \right\| \\ & \leq C \frac{\tau_n}{2} \sum_{l=0}^{k_{\mathcal{J}}} \left(\frac{\tau_n}{2} \right)^l \sup_{t \in I_n} \left\| \frac{d^l}{dt^l} \left(\mathcal{I}_n f(t, u(t)) - \mathcal{I}_n f(t, U(t)) \right) \right\| \\ & \leq C \frac{\tau_n}{2} \left(\mathfrak{C}_{1,1} \sum_{m=0}^{K^{\mathcal{I}}} \sum_{j=0}^{\tilde{K}_m^{\mathcal{I}}} \left(\frac{\tau_n}{2} \right)^j \left\| \frac{d^j}{dt^j} \left(f(t, u(t)) - f(t, U(t)) \right) \right\|_{t=t_{n,m}^{\mathcal{I}}} \right. \\ & \quad \left. + \mathfrak{C}_{1,2} \sup_{t \in I_n} \|f(t, u(t)) - f(t, U(t))\| \right) \end{aligned}$$

where the first inequality follows, as in (3.7), by Assumption 2 and Leibniz' rule for the j th derivative. Using quite similar arguments, under Assumption 5b we gain that

$$\begin{aligned} & \left\| \mathcal{J}_n \left[\left(\mathcal{I}_n f(t, u(t)) - \mathcal{I}_n f(t, U(t)) \right) (1 + T_n^{-1}(t))^i \right] \right\| \\ & \leq \tilde{\mathfrak{C}}_{0,1} C \frac{\tau_n}{2} \sum_{m=0}^{K^{\mathcal{J}}} \sum_{l=0}^{\tilde{K}_m^{\mathcal{J}}} \left(\frac{\tau_n}{2} \right)^l \left\| \frac{d^l}{dt^l} \left(\mathcal{I}_n f(t, u(t)) - \mathcal{I}_n f(t, U(t)) \right) \right\|_{t=t_{n,m}^{\mathcal{J}}} \\ & \quad + \tilde{\mathfrak{C}}_{0,2} C \frac{\tau_n}{2} \sup_{t \in I_n} \left\| \left(\mathcal{I}_n f(t, u(t)) - \mathcal{I}_n f(t, U(t)) \right) \right\| \\ & \leq C \frac{\tau_n}{2} \left(\tilde{\mathfrak{C}}_{1,1} \sum_{m=0}^{K^{\mathcal{I}}} \sum_{j=0}^{\tilde{K}_m^{\mathcal{I}}} \left(\frac{\tau_n}{2} \right)^j \left\| \frac{d^j}{dt^j} \left(f(t, u(t)) - f(t, U(t)) \right) \right\|_{t=t_{n,m}^{\mathcal{I}}} \right. \\ & \quad \left. + \tilde{\mathfrak{C}}_{1,2} \sup_{t \in I_n} \left\| \left(f(t, u(t)) - f(t, U(t)) \right) \right\| \right). \end{aligned}$$

So, either way applying Assumption 4 gives

$$\begin{aligned} & \left\| \mathcal{J}_n \left[\left(\mathcal{I}_n f(t, u(t)) - \mathcal{I}_n f(t, U(t)) \right) (1 + T_n^{-1}(t))^i \right] \right\| \\ & \leq C \mathfrak{C}_2 \frac{\tau_n}{2} \left(\sum_{m=0}^{K^{\mathcal{I}}} \sum_{l=0}^{\tilde{K}_m^{\mathcal{I}}} \left(\frac{\tau_n}{2} \right)^l \left\| (u - U)^{(l)}(t_{n,m}^{\mathcal{I}}) \right\| + \sup_{t \in I_n} \|(u - U)(t)\| \right). \end{aligned}$$

Recalling the definition of $\mathcal{I}_n^{\text{app}}$, especially (4.3a), the terms that include derivatives can be estimated similar to (4.7) which implies that

$$\begin{aligned} & \left\| \mathcal{J}_n \left[\left(\mathcal{I}_n f(t, u(t)) - \mathcal{I}_n f(t, U(t)) \right) (1 + T_n^{-1}(t))^i \right] \right\| \\ & \leq C \mathfrak{C}_2 \frac{\tau_n}{2} \left(\sup_{t \in I_n} \|(\text{Id} - \mathcal{I}_n^{\text{app}}) u(t)\| + E_n \right). \end{aligned} \quad (4.11)$$

Hence, together with (4.10) we obtain

$$(II) \leq C\mathfrak{C}_2 \frac{\tau_n}{2} \left(\sup_{t \in I_n} \|(\text{Id} - \mathcal{I}_n^{\text{app}}) u(t)\| + E_n \right). \quad (4.12)$$

So, combining (4.6) with (4.8) and (4.12), we have already shown that

$$\int_{I_n} \|\zeta'(t)\| dt \leq \tau_n \sup_{t \in I_n} \|\zeta'(t)\| \leq C\mathfrak{C}_2 \frac{\tau_n}{2} \left(\sup_{t \in I_n} \|(\text{Id} - \mathcal{I}_n^{\text{app}}) u(t)\| + E_n \right) \quad (4.13)$$

for all $1 \leq n \leq N$.

Next, we analyze the jump term of (4.5). First, we have a closer look at $[\zeta]_{n-1}$ for $1 \leq n \leq N$. There are two cases. If $k \geq 1$, the discrete solution U is globally continuous due to (2.1a). So, by (4.1a) and the continuity of u

$$[\zeta]_{n-1} = \left[\mathcal{J}^{\mathcal{J}, \mathcal{I}} u \right]_{n-1} = u(t_{n-1}^+) - \mathcal{J}^{\mathcal{J}, \mathcal{I}} u(t_{n-1}^-) = (u - \mathcal{J}^{\mathcal{J}, \mathcal{I}} u)(t_{n-1}^-).$$

Otherwise, if $k = 0$ we exploit (4.9) to rewrite the jump term as follows

$$[\zeta]_{n-1} = (u - \mathcal{J}^{\mathcal{J}, \mathcal{I}} u)(t_{n-1}^-) + \mathcal{J}_n [\mathcal{I}_n f(t, u(t)) - \mathcal{I}_n f(t, U(t))] - \mathcal{J}_n [\zeta'(t)].$$

Hence, using (4.11) with $i = 0$ to bound $\|\mathcal{J}_n [\mathcal{I}_n f(t, u(t)) - \mathcal{I}_n f(t, U(t))]\|$ and combining Assumption 2, the inverse inequality (3.3), and (4.13) to estimate $\|\mathcal{J}_n [\zeta'(t)]\|$, we have in both cases

$$\|[\zeta]_{n-1}\| \leq \left\| (u - \mathcal{J}^{\mathcal{J}, \mathcal{I}} u)(t_{n-1}^-) \right\| + \delta_{0,k} C\mathfrak{C}_2 \frac{\tau_n}{2} \left(\sup_{t \in I_n} \|(\text{Id} - \mathcal{I}_n^{\text{app}}) u(t)\| + E_n \right). \quad (4.14)$$

Summarizing, we get from (4.4), (4.5), (4.13), and (4.14) for $1 \leq n \leq N$

$$E_n \leq \sup_{t \in I_n} \|\eta(t)\| + \underbrace{C\mathfrak{C}_2}_{=\tilde{C}} \sum_{v=1}^n \frac{\tau_v}{2} \left(\sup_{t \in I_v} \|(\text{Id} - \mathcal{I}_v^{\text{app}}) u(t)\| + E_v \right) + \sum_{v=1}^n \left\| (u - \mathcal{J}^{\mathcal{J}, \mathcal{I}} u)(t_{v-1}^-) \right\|.$$

Note that $\tilde{C}$ is independent of T but especially depends multiplicatively on the Lipschitz constant of f (hidden in $\mathfrak{C}_2$). For τ_n sufficiently small ($\tilde{C}\tau_n/2 < 1$), the E_n term of the right-hand side can be absorbed on the left.

Applying a variant of the discrete Gronwall lemma, see [17, Lemma 1.4.2, p. 14], and using that $u(t_0^-) = u_0 = \mathcal{J}^{\mathcal{J}, \mathcal{I}} u(t_0^-)$, we easily obtain the wanted statement for $l = 0$ where the arising error constant exponentially depends on T and the Lipschitz constant of f (hidden in $\tilde{C}, \mathfrak{C}_2$).

Finally, we derive a bound for the first derivative of the error. Recalling the estimate for ζ' in (4.13) it follows

$$\begin{aligned} \sup_{t \in I_n} \|(u - U)'(t)\| &\leq \sup_{t \in I_n} \left\| (u - \mathcal{J}_n^{\mathcal{J}, \mathcal{I}} u)'(t) \right\| + \sup_{t \in I_n} \left\| (\mathcal{J}_n^{\mathcal{J}, \mathcal{I}} u - U)'(t) \right\| \\ &\leq \sup_{t \in I_n} \left\| (u - \mathcal{J}_n^{\mathcal{J}, \mathcal{I}} u)'(t) \right\| + C\mathfrak{C}_2 \left(\sup_{t \in I_n} \|(\text{Id} - \mathcal{I}_n^{\text{app}}) u(t)\| + E_n \right). \end{aligned}$$

Using the already known estimate for E_n , the statement for $l = 1$ follows. $\square$

Remark 9 Based on Theorem 8 and using an inverse inequality we can also prove abstract estimates for higher order derivatives of the error. We find

$$\begin{aligned} \sup_{t \in I_n} \|(u - U)^{(l)}(t)\| &\leq \sup_{t \in I_n} \left\| \left(u - \mathcal{J}_n^{\mathcal{J}, \mathcal{I}} u \right)^{(l)}(t) \right\| \\ &\quad + C_{\text{inv}} \left(\frac{\tau_n}{2} \right)^{-l} \left(\sup_{t \in I_n} \left\| \left(\mathcal{J}_n^{\mathcal{J}, \mathcal{I}} u - u \right)(t) \right\| + \sup_{t \in I_n} \|(u - U)(t)\| \right). \end{aligned}$$

However, since we only have a non-local error estimate for $\sup_{t \in I_n} \|(u - U)(t)\|$, we cannot expect that $\left(\frac{\tau_n}{2}\right)^{-l}$ can be compensated in general. So, usually we additionally need to assume that $\tau_\nu \leq \tau_{\nu+1}$ for all ν or alternatively that the mesh is quasi uniform ($\tau/\tau_\nu \leq C$ for all ν) to obtain a proper estimate.

Remark 10 In the proof of Theorem 8 stiffness of the problem would be critical at several points.

Indeed, for large Lipschitz constants (hidden in $\mathfrak{C}_2$ and so in $\tilde{C}$) the needed inequality $\tilde{C}\tau_n/2 < 1$ would force very small time step lengths. For semidiscretizations in space of time-space problems, where the Lipschitz constant is typically proportional to h^{-2} with h denoting the spatial mesh parameter, this would cause upper bounds on the time step length with respect to h similar to CFL conditions.

Moreover, since the error constant C exponentially depends on $\mathfrak{C}_2$, it would be excessively large for stiff problems. So, the estimate would be useless then.

Of course, Theorem 8 provides an abstract bound for the error of the variational time discretization method. However, the order of convergence still is not clear. Since $\mathcal{I}^{\text{app}}$ is an Hermite-type interpolator of ansatz order larger than or equal to r its approximation order is at least $r + 1$. It remains to prove suitable bounds on the error of the approximation operator $\mathcal{J}^{\mathcal{J}, \mathcal{I}}$.

Definition 11 (Approximation orders of $\mathcal{J}_n$ and $\mathcal{I}_n$) Let $r_{\text{ex}}^{\mathcal{J}}$, $r_{\text{ex}}^{\mathcal{I}}$, $r_{\mathcal{I}}$, and $r_{\mathcal{I},i}^{\mathcal{J}} \in \mathbb{N}_0 \cup \{-1, \infty\}$ denote the largest numbers such that

$$\begin{aligned} \int_{I_n} \varphi(t) \, dt &= \mathcal{J}_n[\varphi] & \forall \varphi \in P_{r_{\text{ex}}^{\mathcal{J}}}(I_n), & \quad \varphi = \mathcal{I}_n \varphi \quad \forall \varphi \in P_{r_{\mathcal{I}}}(I_n), \\ \int_{I_n} \varphi(t) \, dt &= \int_{I_n} \mathcal{I}_n \varphi(t) \, dt & \forall \varphi \in P_{r_{\text{ex}}^{\mathcal{I}}}(I_n), \\ \mathcal{J}_n[\varphi \psi_i] &= \mathcal{J}_n[(\mathcal{I}_n \varphi) \psi_i] & \forall \varphi \in P_{r_{\mathcal{I},i}^{\mathcal{J}}}(I_n), \psi_i \in P_i(I_n). \end{aligned}$$

Here, $P_{-1}(I_n)$ is interpreted as $\{0\}$, in which case the respective operator does not provide the corresponding approximation property. For convenience, set $r_{\mathcal{I}}^{\mathcal{J}} := r_{\mathcal{I},r-k}^{\mathcal{J}}$. Note that $r_{\text{ex}}^{\mathcal{I}} \geq r_{\mathcal{I},i}^{\mathcal{I}} \geq r_{\mathcal{I}}$ and $r_{\mathcal{I},i}^{\mathcal{J}} \geq r_{\mathcal{I}}$ hold by definition.

Using standard techniques the above quantities can be connected with certain approximation estimates. For example, let $\check{r} \in \mathbb{N}_0$ then together with Assumption 2 we find that

$$\left\| \int_{I_n} \varphi(t) dt - \mathcal{J}_n[\varphi] \right\| \leq C \frac{\tau_n}{2} \sum_{j=\min\{\check{r}, r_{\text{ex}}^{\mathcal{J}}+1\}}^{\max\{k_{\mathcal{J}}, \min\{\check{r}, r_{\text{ex}}^{\mathcal{J}}+1\}\}} \left(\frac{\tau_n}{2}\right)^j \sup_{t \in I_n} \|\varphi^{(j)}(t)\| \quad (4.15)$$

for arbitrary $\varphi \in C^{\max\{k_{\mathcal{J}}, \min\{\check{r}, r_{\text{ex}}^{\mathcal{J}}+1\}\}}(\bar{I}_n, \mathbb{R}^d)$.

Lemma 12 *Let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$, and suppose that Assumptions 1, 2, and 3 hold. Furthermore, let $l, \check{r} \in \mathbb{N}_0$ and define*

$$j_{\min, \check{r}} := \min \left\{ \check{r}, r + 1, r_{\mathcal{I}}^{\mathcal{J}} + 2 \right\}, \quad j_{\max, \check{r}} := \max \left\{ k_{\mathcal{J}} + 1, l, j_{\min, \check{r}} \right\}.$$

If $v \in C^{j_{\max, \check{r}}}(\bar{I}_n, \mathbb{R}^d)$ then the error estimate

$$\sup_{t \in I_n} \left\| \left(v - \mathcal{J}_n^{\mathcal{J}, \mathcal{I}} v \right)^{(l)}(t) \right\| \leq C \sum_{j=j_{\min, \check{r}}}^{j_{\max, \check{r}}} \left(\frac{\tau_n}{2}\right)^{j-l} \sup_{t \in I_n} \|v^{(j)}(t)\| =: \mathcal{T}_{(4.16)}^{\check{r}, l}[v] \quad (4.16)$$

holds with a constant C independent of τ_n .

Proof The estimate follows from standard approximation theory since under the given assumptions $\mathcal{J}_n^{\mathcal{J}, \mathcal{I}}$ preserves polynomials up to degree $\min\{r, r_{\mathcal{I}}^{\mathcal{J}} + 1\}$. Also note that a stability estimate for $\mathcal{J}_n^{\mathcal{J}, \mathcal{I}}$, cf. [2, Lemma 4.2], which then motivates the upper summation bound $j_{\max, \check{r}}$, can be easily proven by inverse inequalities, estimation on the reference interval, and transformation. $\square$

In many cases, the approximation error of $\mathcal{J}^{\mathcal{J}, \mathcal{I}}$ in the mesh points t_n^- behaves much better than the pointwise estimate of Lemma 12 suggests. However, to see this, we need some further knowledge on the approximation property connected with the quantity $r_{\mathcal{I}, i}^{\mathcal{J}}$. Besides, the respective result presented in the following lemma will be also used later in the superconvergence analysis.

Lemma 13 *Let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$. Suppose that the Assumptions 2 and 3 hold. Moreover, assume that $\psi_i \in P_i(I_n, \mathbb{R})$, $i \in \mathbb{N}_0$, satisfies $\sup_{t \in I_n} |\psi_i^{(j)}(t)| \leq C \tau_n^{-j}$ for all $j \in \mathbb{N}_0$. Let $\check{r} \in \mathbb{N}_0$ and define*

$$j_{\min, i, \check{r}}^* := \min \left\{ \check{r}, r_{\mathcal{I}, i}^{\mathcal{J}} + 1 \right\}, \quad j_{\max, i, \check{r}}^* := \max \left\{ k_{\mathcal{J}}, k_{\mathcal{I}}, j_{\min, i, \check{r}}^* \right\}.$$

Then, we have for $\varphi \in C^{j_{\max, i, \check{r}}^}(\bar{I}_n, \mathbb{R}^d)$ the bound*

$$\|\mathcal{J}_n[(\varphi - \mathcal{I}_n \varphi)(t) \psi_i(t)]\| \leq C \frac{\tau_n}{2} \sum_{j=j_{\min, i, \check{r}}^*}^{j_{\max, i, \check{r}}^*} \left(\frac{\tau_n}{2}\right)^j \sup_{t \in I_n} \|\varphi^{(j)}(t)\|$$

with a constant C independent of τ_n .

Proof (for more details see [2, Lemma 4.10]) Let $\tilde{\varphi} \in P_{\min\{\check{r}-1, r_{\mathcal{I},i}^{\mathcal{J}}\}}(I_n, \mathbb{R}^d)$. We start rewriting the left-hand side of the wanted inequality as follows

$$\begin{aligned} & \mathcal{J}_n [(\varphi - \mathcal{I}_n \varphi)(t) \psi_i(t)] \\ &= \mathcal{J}_n [(\varphi - \tilde{\varphi})(t) \psi_i(t)] + \underbrace{\mathcal{J}_n [(\tilde{\varphi} - \mathcal{I}_n \tilde{\varphi})(t) \psi_i(t)]}_{=0} + \mathcal{J}_n [\mathcal{I}_n(\tilde{\varphi} - \varphi)(t) \psi_i(t)], \end{aligned}$$

where the term in the middle vanishes by definition of $r_{\mathcal{I},i}^{\mathcal{J}}$. Exploiting Assumption 2, the Leibniz rule for the j th derivative, and the given bound for $\psi_i^{(j)}$ to estimate the two remaining terms, we gain

$$\begin{aligned} & \|\mathcal{J}_n [(\varphi - \mathcal{I}_n \varphi)(t) \psi_i(t)]\| \\ & \leq C \frac{\tau_n}{2} \sum_{l=0}^{k_{\mathcal{J}}} \left(\frac{\tau_n}{2}\right)^l \left(\sup_{t \in I_n} \|(\varphi - \tilde{\varphi})^{(l)}(t)\| + \sup_{t \in I_n} \|(\mathcal{I}_n(\varphi - \tilde{\varphi}))^{(l)}(t)\| \right). \end{aligned}$$

Furthermore, using Assumption 3 to bound the latter summands, we conclude that

$$\|\mathcal{J}_n [(\varphi - \mathcal{I}_n \varphi)(t) \psi_i(t)]\| \leq C \frac{\tau_n}{2} \sum_{j=0}^{\max\{k_{\mathcal{J}}, k_{\mathcal{I}}\}} \left(\frac{\tau_n}{2}\right)^j \sup_{t \in I_n} \|(\varphi - \tilde{\varphi})^{(j)}(t)\|$$

for all $\tilde{\varphi} \in P_{\min\{\check{r}-1, r_{\mathcal{I},i}^{\mathcal{J}}\}}(I_n, \mathbb{R}^d)$. So, choosing $\tilde{\varphi}$ as the Taylor polynomial of φ at $(t_{n-1} + t_n)/2$ of degree $\min\{\check{r} - 1, r_{\mathcal{I},i}^{\mathcal{J}}\}$, the wanted statement follows by standard error estimates for the Taylor polynomial. $\square$

Now, we can state and prove the improved estimate for the approximation error of $\mathcal{J}^{\mathcal{J}, \mathcal{I}}$ in the mesh points t_n^- .

Lemma 14 *Let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$. Suppose that the Assumptions 1, 2, and 3 hold. Moreover, assume that $\max\{r_{\text{ex}}^{\mathcal{J}}, r_{\mathcal{I}}^{\mathcal{J}} + 1\} \geq r - 1$. Let $\check{r} \in \mathbb{N}_0$ and define*

$$\begin{aligned} j_{\min, \check{r}}^{\diamond} &:= \min\left\{\check{r}, \max\left\{r_{\text{ex}}^{\mathcal{J}} + 1, \min\left\{r, r_{\mathcal{I}}^{\mathcal{J}} + 1\right\}\right\} + 1, r_{\mathcal{I},0}^{\mathcal{J}} + 2\right\}, \\ j_{\max, \check{r}}^{\diamond} &:= \max\left\{k_{\mathcal{J}} + 1, j_{\min, \check{r}}^{\diamond}\right\}. \end{aligned}$$

Then, provided $v \in C^{j_{\max, \check{r}}^{\diamond}}(\bar{I}_n, \mathbb{R}^d)$, the estimate

$$\left\| (v - \mathcal{J}_n^{\mathcal{J}, \mathcal{I}} v)(t_n^-) \right\| \leq C \sum_{j=j_{\min, \check{r}}^{\diamond}}^{j_{\max, \check{r}}^{\diamond}} \left(\frac{\tau_n}{2}\right)^j \sup_{t \in I_n} \|v^{(j)}(t)\| =: \mathcal{T}_{(4.17)}[v] \quad (4.17)$$

holds for $1 \leq n \leq N$ where the constant C is independent of τ_n .

Proof (for more details see [2, Lemma 4.11]) From the fundamental theorem of calculus and exploiting the definition (4.1) of $\mathcal{J}_n^{\mathcal{I}}$, we gain

$$\begin{aligned} (v - \mathcal{J}_n^{\mathcal{I}} v)(t_n^-) &= (v - \mathcal{J}_n^{\mathcal{I}} v)(t_{n-1}^+) + \int_{I_n} (v - \mathcal{J}_n^{\mathcal{I}} v)'(t) dt \\ &= -\mathcal{J}_n \left[\mathcal{I}_n v' - (\mathcal{J}_n^{\mathcal{I}} v)' \right] + \int_{I_n} (v - \mathcal{J}_n^{\mathcal{I}} v)'(t) dt \\ &= \mathcal{J}_n [v' - \mathcal{I}_n v'] + \left(\int_{I_n} \cdot dt - \mathcal{J}_n \cdot \right) [v'(t)] - \left(\int_{I_n} \cdot dt - \mathcal{J}_n \cdot \right) \left[(\mathcal{J}_n^{\mathcal{I}} v)'(t) \right]. \end{aligned}$$

The first and the second difference on the right-hand side then can be estimated by Lemma 13 with $i = 0$, $\psi_0 \equiv 1$, and (4.15), respectively. The third difference vanishes if $r_{\text{ex}}^{\mathcal{I}} \geq r - 1$ since then $(\mathcal{J}_n^{\mathcal{I}} v)' \in P_{r-1}(I_n, \mathbb{R}^d)$ is integrated exactly. Otherwise, if $r_{\mathcal{I}}^{\mathcal{I}} + 1 \geq r - 1$, i.e., $\mathcal{J}_n^{\mathcal{I}}$ preserves polynomials up to degree $r - 1$, it can be suitably bounded combining (4.15) and (4.16).

Altogether, recalling (4.16) which in some cases may provide a better estimate, we conclude the wanted statement. Here, also note that always $r_{\mathcal{I},0}^{\mathcal{I}} \geq r_{\mathcal{I},r-k}^{\mathcal{I}} = r_{\mathcal{I}}^{\mathcal{I}} \geq \min \{r - 1, r_{\mathcal{I}}^{\mathcal{I}}\}$. $\square$

Finally, summarizing the above results, we now want to list the proven convergence orders.

Corollary 15 *Let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$, and $l \in \{0, 1\}$. Suppose that Assumptions 1, 2, 3, and 4 hold. Moreover, let Assumption 5a or 5b be satisfied. Denoting by u and U the solutions of (1.1) and (2.1), respectively, we have for $1 \leq n \leq N$*

$$\sup_{t \in I_n} \|(u - U)^{(l)}(t)\| \leq C \tau^{\min\{r, r_{\mathcal{I}}^{\mathcal{I}} + 1\}}, \quad (4.18)$$

with $r_{\mathcal{I}}^{\mathcal{I}}$ as defined in Definition 11. If in addition $\max \{r_{\text{ex}}^{\mathcal{I}}, r_{\mathcal{I}}^{\mathcal{I}} + 1\} \geq r - 1$, then we even have

$$\sup_{t \in I_n} \|(u - U)(t)\| \leq C \tau^{\min\{r+1, r_{\mathcal{I}}^{\mathcal{I}} + 2, r_{\mathcal{I},0}^{\mathcal{I}} + 1, \max\{r_{\text{ex}}^{\mathcal{I}} + 1, \min\{r, r_{\mathcal{I}}^{\mathcal{I}} + 1\}\}}\} \quad (4.19)$$

as an improved L^∞ estimate.

If $\max \{r_{\text{ex}}^{\mathcal{I}}, r_{\mathcal{I}}^{\mathcal{I}} + 1\} \geq r - 1$ is satisfied, we obtain formally

$$\sup_{t \in I_n} \|(u - U)'(t)\| \leq C \tau^{\min\{r, r_{\mathcal{I}}^{\mathcal{I}} + 1, r_{\mathcal{I},0}^{\mathcal{I}} + 1, \max\{r_{\text{ex}}^{\mathcal{I}} + 1, \min\{r, r_{\mathcal{I}}^{\mathcal{I}} + 1\}\}}\}$$

for the $W^{1,\infty}$ seminorm. However, this also gives (4.18) only.

Remark 16 Since the quantity $r_{\mathcal{I}}^{\mathcal{J}} = r_{\mathcal{I},r-k}^{\mathcal{J}}$ used in the lemmas and the corollary above is quite abstract, we want to provide lower bounds for $r_{\mathcal{I},i}^{\mathcal{J}}$ based on the more familiar quantities $r_{\mathcal{I}}$, $r_{\text{ex}}^{\mathcal{I}}$, and $r_{\text{ex}}^{\mathcal{J}}$. For a proof of these bounds, we refer to [2, Lemma 4.13].

Let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$, and $i \in \mathbb{N}_0$. Then, $r_{\mathcal{I},i}^{\mathcal{J}} \geq r_{\mathcal{I}}$. So, for $r_{\mathcal{I}} = \infty$ the bound cannot be improved further. Otherwise, supposing that $\mathcal{I}_n$ is a projection onto the space of polynomials of maximal degree $r_{\mathcal{I}} < \infty$, i.e., $\mathcal{I}_n : C^{k_{\mathcal{I}}}(\bar{I}_n) \rightarrow P_{r_{\mathcal{I}}}(\bar{I}_n)$ and $\mathcal{I}_n \varphi = \varphi$ for all $\varphi \in P_{r_{\mathcal{I}}}(\bar{I}_n)$, we even get

$$r_{\mathcal{I},i}^{\mathcal{J}} \geq \max \left\{ r_{\mathcal{I}}, \min \left\{ r_{\text{ex}}^{\mathcal{J}} - i, r_{\mathcal{I},i}^{\mathcal{J}} \right\} \right\}.$$

Of course, it holds $r_{\mathcal{I},0}^{\mathcal{J}} = r_{\text{ex}}^{\mathcal{I}}$. In order to simplify the term on the right-hand side for $i \geq 1$, we additionally assume that $\mathcal{I}_n$ satisfies

$$\int_{I_n} \mathcal{I}_n((\varphi - \mathcal{I}_n \varphi) \psi)(t) \, dt = 0 \quad \forall \psi \in P_i(I_n) \quad (4.20)$$

for all $\varphi \in C^{k_{\mathcal{I}}}(\bar{I}_n)$. Then, we simply have

$$r_{\mathcal{I},i}^{\mathcal{J}} \geq \max \left\{ r_{\mathcal{I}}, \min \left\{ r_{\text{ex}}^{\mathcal{J}}, r_{\text{ex}}^{\mathcal{I}} \right\} - i \right\}$$

since then $r_{\mathcal{I},i}^{\mathcal{J}} \geq \max \{ r_{\mathcal{I}}, r_{\text{ex}}^{\mathcal{I}} - i \}$.

Furthermore, under the weaker assumption that $\mathcal{I} = \mathcal{I}^1 \circ \dots \circ \mathcal{I}^l$ is a composition of several projection operators $\mathcal{I}^j$, $1 \leq j \leq l$, that all satisfy (4.20), we still find

$$r_{\mathcal{I},i}^{\mathcal{J}} \geq \min_{j \in \mathcal{M}_i \cup \{l\}} \left\{ \max \left\{ r_{\mathcal{I}^j}, r_{\text{ex}}^{\mathcal{I}^j} - i \right\} \right\}$$

where $\mathcal{M}_i := \left\{ j \in \mathbb{N} \mid 1 \leq j \leq l-1, \max \left\{ r_{\mathcal{I}^j}, r_{\text{ex}}^{\mathcal{I}^j} - i \right\} < \min_{j+1 \leq m \leq l} \{ r_{\mathcal{I}^m} \} \right\}$.

5 Superconvergence analysis

In order to prove superconvergence in time mesh points, we exploit a special representation of the discrete problem (2.1). To this end, we need one further approximation operator $\widehat{\mathcal{P}}^{\mathcal{J},\mathcal{I}}$, see [2, (5.1)], which is connected to $\widehat{\mathcal{J}}^{\mathcal{J},\mathcal{I}}$ in such a way that $\widehat{\mathcal{P}}^{\mathcal{J},\mathcal{I}}(\widehat{v}') = (\widehat{\mathcal{J}}^{\mathcal{J},\mathcal{I}}\widehat{v})'$ holds true. For brevity, we directly consider its local version $\mathcal{P}_n^{\mathcal{J},\mathcal{I}}$. Then, analogously to the respective statements for $\mathcal{J}_n^{\mathcal{J},\mathcal{I}}$, we get the following.

Lemma 17 (Approximation operator) *Let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$. Moreover, suppose that Assumption 1 holds. Then, the operator $\mathcal{P}_n^{\mathcal{J}, \mathcal{I}} : C^{k, \mathcal{J}}(\bar{I}_n, \mathbb{R}^d) \rightarrow P_{r-1}(I_n, \mathbb{R}^d)$, $1 \leq n \leq N$, given by*

$$\left(\mathcal{P}_n^{\mathcal{J}, \mathcal{I}} v\right)^{(i)}(t_{n-1}^+) = v^{(i)}(t_{n-1}^+), \quad \text{if } k \geq 3, i = 0, \dots, \lfloor \frac{k-1}{2} \rfloor - 1, \quad (5.1a)$$

$$\left(\mathcal{P}_n^{\mathcal{J}, \mathcal{I}} v\right)^{(i)}(t_n^-) = v^{(i)}(t_n^-), \quad \text{if } k \geq 2, i = 0, \dots, \lfloor \frac{k}{2} \rfloor - 1, \quad (5.1b)$$

$$\mathcal{J}_n \left[\left(\mathcal{P}_n^{\mathcal{J}, \mathcal{I}} v(t), \varphi(t)\right) \right] = \mathcal{J}_n \left[(\mathcal{I}_n v(t), \varphi(t)) \right] \quad \forall \varphi \in P_{r-k}(I_n, \mathbb{R}^d) \text{ with } \delta_{0,k} \varphi(t_{n-1}^+) = 0, \quad (5.1c)$$

is well-defined.

Furthermore, let Assumptions 2 and 3 hold. For $l, \check{r} \in \mathbb{N}_0$, we define

$$j_{\min, \check{r}}^\bullet := \min \left\{ \check{r}, r, r_{\mathcal{I}}^{\mathcal{J}} + 1 \right\}, \quad j_{\max, l, \check{r}}^\bullet := \max \left\{ k_{\mathcal{J}}, l, j_{\min, \check{r}}^\bullet \right\}.$$

Then, we have for $v \in C^{j_{\max, l, \check{r}}^\bullet}(\bar{I}_n, \mathbb{R}^d)$ the error estimates

$$\sup_{t \in I_n} \left\| \left(v - \mathcal{P}_n^{\mathcal{J}, \mathcal{I}} v \right)^{(l)}(t) \right\| \leq C \sum_{j=j_{\min, \check{r}}^\bullet}^{j_{\max, l, \check{r}}^\bullet} \left(\frac{\tau_n}{2} \right)^{j-l} \sup_{t \in I_n} \left\| v^{(j)}(t) \right\| =: \mathcal{T}_{(5.2)}^{\check{r}, l}[v] \quad (5.2)$$

with a constant C independent of τ_n .

Inspecting the definition of U in (2.1) and (5.1), we find that

$$U'|_{I_n} = \mathcal{P}_n^{\mathcal{J}, \mathcal{I}} f(\cdot, U(\cdot)). \quad (5.3)$$

Therefore, since for $k \geq 1$ furthermore $U(t_{n-1}^+) = U_{n-1}^-$ with U_{n-1}^- given, the discretization method fits into the unified framework of [1] for $k \geq 1$.

Theorem 18 (Superconvergence estimate) *Let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$. Suppose that the Assumptions 1, 2, and 3 hold. Moreover, denote by u and U the solutions of (1.1) and (2.1), respectively. Suppose that (for τ sufficiently small) the global error $\sup_{t \in I} \|(u - U)(t)\|$, as well as U and all of its derivatives, can be bounded independent of the mesh parameter. Then, we have*

$$\|(u - U)(t_N^-)\| \leq C(f, u) \left(\sup_{t \in I} \|(u - U)(t)\|^2 + \tau^{r_{\text{super}}} \right) \quad (5.4a)$$

with

$$r_{\text{super}} := \min \left\{ 2r - k + 1, r_{\text{var}}^{\mathcal{J}, \mathcal{I}} + 1, \max \left\{ r_{\text{ex}}^{\mathcal{J}} + 1, \min \left\{ r, r_{\mathcal{I}}^{\mathcal{J}} + 1 \right\} \right\} \right\} \quad (5.4b)$$

where $r_{\text{var}}^{\mathcal{J}, \mathcal{I}} := \min_{0 \leq i \leq r-k} \left\{ r_{\mathcal{I}, i}^{\mathcal{J}} + i \right\}$.

Proof (for more details see [2, Theorem 5.3]) In order to prove the wanted statement, we firstly derive an estimate for the error $e(t) = (u - U)(t)$ at t_n^- provided that we already have a suitable bound at t_{n-1}^- . To this end, we adapt some basic ideas known from the superconvergence proof for collocation methods, see [10, Theorem II.1.5,

p. 28]. So, we consider the local discrete solution $U|_{I_n}$ as solution $\tilde{U}$ of the perturbed initial value problem

$$\tilde{U}'(t) = f(t, \tilde{U}(t)) + \text{def}(t) \quad \text{in } I_n, \quad \tilde{U}(t_{n-1}) = U(t_{n-1}^+),$$

where $\text{def}(t) := U'(t) - f(t, U(t))$ denotes the defect. Since u solves (1.1), we find for all $t \in I_n$ after linearization that

$$U'(t) - u'(t) = \frac{\partial}{\partial v} f(t, u(t)) (U(t) - u(t)) + \text{def}(t) + \text{rem}(t)$$

where f is interpreted as function of (t, v) and the remainder term is given by

$$\text{rem}(t) := \int_0^1 \int_0^s (U(t) - u(t))^T \frac{\partial^2}{\partial v^2} f(t, u(t) + \tilde{s}(U(t) - u(t))) (U(t) - u(t)) \, d\tilde{s} \, ds.$$

The variation of constants formula, cf. [11, Theorem I.11.2, p. 66], yields

$$e(t) = R(t, t_{n-1}^+) e(t_{n-1}^+) - \int_{t_{n-1}}^t R(t, s) (\text{def}(s) + \text{rem}(s)) \, ds \quad \forall t \in I_n$$

where $R(t, s)$ is the resolvent of the homogeneous differential equation $y'(t) = \frac{\partial}{\partial v} f(t, u(t)) y(t)$ for initial values given at s .

Using that u is continuous as well as U for $k \geq 1$, we conclude

$$\begin{aligned} e(t_n^-) &= R(t_n^-, t_{n-1}^+) e(t_{n-1}^-) \\ &\quad - \delta_{0,k} R(t_n^-, t_{n-1}^+) [U]_{n-1} - \int_{I_n} R(t_n^-, s) (\text{def}(s) + \text{rem}(s)) \, ds. \end{aligned} \quad (5.5)$$

The right-hand side now is split and the single terms are studied separately.

First, for the term including $e(t_{n-1}^-)$ we gain because of $R(t_{n-1}^+, t_{n-1}^+) = \text{Id}$

$$\|R(t_n^-, t_{n-1}^+) e(t_{n-1}^-)\| \leq \left(1 + \tau_n \sup_{\tilde{s} \in I_n} \left\| \frac{\partial}{\partial t} R(\tilde{s}, t_{n-1}^+) \right\| \right) \|e(t_{n-1}^-)\|. \quad (5.6)$$

The term including the remainder term $\text{rem}(\cdot)$ can be bounded as follows

$$\left\| \int_{I_n} R(t_n^-, s) \text{rem}(s) \, ds \right\| \leq C \tau_n \sup_{s \in I_n} \|e(s)\|^2 \quad (5.7)$$

where we also exploited that $u(s) - \tilde{s}e(s)$ is in a bounded neighborhood of $u(s)$ for all $s \in I_n$, $\tilde{s} \in [0, 1]$, since by assumption $\|e(s)\| \leq C$.

Finally, the remaining terms are considered. A Taylor series expansion of $R(t_n^-, s)$ with respect to s in t_{n-1}^+ motivates the following decomposition

$$\begin{aligned} & \delta_{0,k} R(t_n^-, t_{n-1}^+) [U]_{n-1} + \int_{I_n} R(t_n^-, s) \text{def}(s) \, ds \\ &= \sum_{i=0}^{r-k} \frac{1}{i!} \frac{\partial^i}{\partial s^i} R(t_n^-, t_{n-1}^+) \underbrace{\left(\int_{I_n} (s - t_{n-1})^i (U'(s) - f(s, U(s))) \, ds + \delta_{0,k} \delta_{0,i} [U]_{n-1} \right)}_{=(I)} \\ &+ \underbrace{\int_{I_n} \int_{t_{n-1}}^s \frac{(\tilde{s} - t_{n-1})^{r-k}}{(r-k)!} \frac{\partial^{r-k+1}}{\partial s^{r-k+1}} R(t_n^-, \tilde{s}) \, d\tilde{s} (U'(s) - f(s, U(s))) \, ds}_{=(II)}. \end{aligned} \quad (5.8)$$

We start with (II). Using (5.3) and (5.2) with $v(\cdot) = f(\cdot, U(\cdot))$, we obtain

$$\begin{aligned} \|(II)\| &\leq \tau_n \frac{\tau_n^{r-k+1}}{(r-k+1)!} \sup_{\tilde{s} \in I_n} \left\| \frac{\partial^{r-k+1}}{\partial s^{r-k+1}} R(t_n^-, \tilde{s}) \right\| \mathcal{T}_{(5.2)}^{\min\{\bar{r}, r, r_{\mathcal{L}}^{\mathcal{J}} + 1\}, 0} [f(\cdot, U(\cdot))] \\ &\leq C \frac{\tau_n}{2} \tau_n^{\min\{\bar{r}, r, r_{\mathcal{L}}^{\mathcal{J}} + 1\} + r - k + 1} G_n^{(II)} \end{aligned}$$

where $G_n^{(II)}$ depends on derivatives of R and $f(\cdot, U(\cdot))$.

The term (I) with $0 \leq i \leq r - k$ can be rewritten using (2.1d)

$$\begin{aligned} (I) &= \left(\int_{I_n} \cdot \, ds - \mathcal{J}_n \cdot \right) \left[(s - t_{n-1})^i (U'(s) - f(s, U(s))) \right] \\ &+ \mathcal{J}_n \left[(s - t_{n-1})^i (\mathcal{I}_n - \text{Id}) f(s, U(s)) \right]. \end{aligned} \quad (5.9)$$

The first term on the right-hand side can be bounded applying (4.15), Leibniz' rule for the j th derivative, and if beneficial again (5.3) as well as (5.2) with $v(\cdot) = f(\cdot, U(\cdot))$. We then find

$$\left\| \left(\int_{I_n} \cdot \, ds - \mathcal{J}_n \cdot \right) \left[(s - t_{n-1})^i (U'(s) - f(s, U(s))) \right] \right\| \leq C \frac{\tau_n}{2} \tau_n^{r_i^{(Ia)}} G_n^{(Ia)}$$

with

$$r_i^{(Ia)} := \max \left\{ \min \left\{ \bar{r}, r_{\text{ex}}^{\mathcal{J}} + 1 \right\}, \min \left\{ \bar{r}, r, r_{\mathcal{L}}^{\mathcal{J}} + 1 \right\} + i \right\}$$

where $G_n^{(Ia)}$ depends on (partial) derivatives of U and $f(\cdot, U(\cdot))$. Factoring out τ_n^i , the second term on the right-hand side of (5.9) can be analyzed by Lemma 13 with $\varphi(t) = f(t, U(t))$ and $\psi_i(t) = \left(\frac{t - t_{n-1}}{\tau_n} \right)^i$. Then we obtain

$$\left\| \mathcal{J}_n \left[(s - t_{n-1})^i (\text{Id} - \mathcal{I}_n) f(s, U(s)) \right] \right\| \leq C \frac{\tau_n}{2} \tau_n^{\min\{\bar{r}, r_{\mathcal{L}, i}^{\mathcal{J}} + 1\} + i} G_n^{(Ib)}$$

where $G_n^{(Ib)}$ depends on derivatives of $f(\cdot, U(\cdot))$. So, combining (5.8) with the estimates for the single parts of (I) and (II), we gain for $\bar{r}$ sufficiently large

$$\left\| \delta_{0,k} R(t_n^-, t_{n-1}^+) [U]_{n-1} + \int_{I_n} R(t_n^-, s) \text{def}(s) \, ds \right\| \leq C \frac{\tau_n}{2} \tau_n^{r_{\text{super}}} G_n \quad (5.10)$$

with

$$r_{\text{super}} = \min \left\{ 2r - k + 1, r_{\text{var}}^{\mathcal{J}, \mathcal{I}} + 1, \max \left\{ r_{\text{ex}}^{\mathcal{J}} + 1, \min \left\{ r, r_{\mathcal{I}}^{\mathcal{J}} + 1 \right\} \right\} \right\}$$

where $r_{\text{var}}^{\mathcal{J}, \mathcal{I}} = \min_{0 \leq i \leq r-k} \left\{ r_{\mathcal{I}, i}^{\mathcal{J}} + i \right\}$ and G_n depends on (partial) derivatives of R , U , and $f(\cdot, U(\cdot))$. Here note that $r_{\text{var}}^{\mathcal{J}, \mathcal{I}} \leq r_{\mathcal{I}, r-k}^{\mathcal{J}} + r - k = r_{\mathcal{I}}^{\mathcal{J}} + r - k$ was used to simplify the exponent r_{super} . Therefore, incorporating (5.6), (5.7), and (5.10) in (5.5) gives

$$\|e(t_n^-)\| \leq (1 + \tau_n \lambda_n) \|e(t_{n-1}^-)\| + C \tau_n \sup_{s \in I_n} \|e(s)\|^2 + C \frac{\tau_n}{2} \tau_n^{r_{\text{super}}} G_n$$

with $\lambda_n = \sup_{\tilde{s} \in I_n} \left\| \frac{\partial}{\partial t} R(\tilde{s}, t_{n-1}^+) \right\|$. A variant of the discrete Gronwall lemma, see [8, Proposition 3.3], together with $(1 + x) \leq e^x$ then yields

$$\|e(t_n^-)\| \leq C(t_n - t_0) \exp \left((t_n - t_0) \max_{v=1, \dots, n} \lambda_v \right) \max_{v=1, \dots, n} \left(\sup_{s \in I_v} \|e(s)\|^2 + \tau_v^{r_{\text{super}}} G_v \right)$$

where we also used $e(t_0^-) = 0$.

It remains, a small technical detail, to verify that G_v can be uniformly bounded independent of τ_v . The term depends on partial derivatives of f , derivatives of R , and on the derivatives of the discrete solution U , thus, potentially also on the mesh parameter. However, U can be uniformly bounded by assumption. So, we are done. $\square$

Remark 19 Using an alternative argument (inspired by the proof of [11, Theorem II.7.9, pp. 212/213]) that is based on the application of the non-linear variation-of-constants formula [11, Corollary I.14.6, p. 97], it can be shown that for $1 \leq k \leq r$ the term $\sup_{t \in I} \|(u - U)(t)\|^2$ in (5.4) is not necessary and can be dropped.

However, for $k = 0$, the alternative proof is much more complicated and in general only guarantees a worse superconvergence estimate than Theorem 18. Moreover, for all k , the notation gets more involved.

Lemma 20 Suppose that Assumption 1 holds along with an estimate similar to (5.2) that at least guarantees approximation order $r - 1$ for $\mathcal{P}_n^{\mathcal{J}, \mathcal{I}}$ (e.g. if $r_{\mathcal{I}}^{\mathcal{J}} \geq r - 2$). In addition, let the solutions u of (1.1) and U of (2.1) satisfy $\sup_{t \in I_n} \|(u - U)(t)\| \leq C$ for some constant C independent of the mesh parameter. Then, we have

$$\sup_{t \in I_n} \|U^{(l)}(t)\| \leq C(f, u)$$

for all $0 \leq l \leq r$.

Proof (for more details see [2, Lemma 5.5]) Firstly, by assumption we have

$$\sup_{t \in I_n} \|U(t)\| \leq \sup_{t \in I_n} \|u(t)\| + \sup_{t \in I_n} \|u(t) - U(t)\| \leq C(f, u).$$

From this, the wanted estimates follow by induction. Indeed, by (5.3) it holds

$$\sup_{t \in I_n} \|U^{(l+1)}(t)\| \leq \sup_{t \in I_n} \left\| \frac{d^l}{dt^l} f(t, U(t)) \right\| + \sup_{t \in I_n} \left\| \frac{d^l}{dt^l} \left(\text{Id} - \mathcal{P}_n^{\mathcal{J}, \mathcal{I}} \right) f(t, U(t)) \right\|.$$

The first term on the right-hand side only contains derivatives of U up to maximal order l . For the second term, a similar upper bound (independent of τ_n) of terms that contain derivatives of U up to maximal order l can be derived by cleverly combining (5.2), a generalization of Faà di Bruno's formula, see [16, Theorem 2.1], and inverse inequalities. Hence, by the induction hypotheses we gain the upper bound $C(f, u)$. $\square$

Summarizing the above observations, our analysis guarantees the following estimates in the time mesh points.

Corollary 21 *Let $r, k \in \mathbb{Z}$, $0 \leq k \leq r$. Suppose that Assumptions 1, 2, 3, and 4 hold. Moreover, let Assumption 5a or 5b be satisfied. Let u and U denote the solutions of (1.1) and (2.1), respectively. Then, if $r_{\mathcal{I}}^{\mathcal{J}} \geq r - 2$, we have for $1 \leq n \leq N$*

$$\|(u - U)(t_n^-)\| \leq C \left(\tau^{r_{\text{super}}} + \delta_{0,k} \tau^{2r_{\mathcal{I}}^{\mathcal{J}} + 4} \right), \quad (5.11)$$

with r_{super} from (5.4b) and $r_{\mathcal{I}}^{\mathcal{J}} = r_{\mathcal{I}, r-k}^{\mathcal{J}}, r_{\mathcal{I}, i}^{\mathcal{J}}$ as defined in Definition 11.

If $r_{\mathcal{I}}^{\mathcal{J}} < r - 2$, we in general cannot ensure the uniform boundedness of U and its derivatives since Lemma 20 does not hold. Then we only have

$$\|(u - U)(t_n^-)\| \leq \sup_{t \in I_n} \|(u - U)(t)\|$$

where we refer to Corollary 15 for bounds on the right-hand side term.

6 Numerical experiments

We consider the initial value problem

$$\begin{pmatrix} u_1'(t) \\ u_2'(t) \end{pmatrix} = \begin{pmatrix} -u_1^2(t) - u_2(t) \\ u_1(t) - u_1(t)u_2(t) \end{pmatrix}, \quad t \in (0, 32), \quad u(0) = \begin{pmatrix} 1/2 \\ 0 \end{pmatrix},$$

of a system of non-linear ordinary differential equations which has

$$u_1(t) = \frac{\cos t}{2 + \sin t}, \quad u_2(t) = \frac{\sin t}{2 + \sin t}$$

as solution.

The non-linear systems within each time step were solved by Newton's method where we applied a Taylor expansion of the inherited data from the previous time interval to calculate an initial guess for all unknowns on the current interval. If higher order derivatives were needed at initial time $t_0 = 0$, we applied

$$\begin{aligned} u^{(0)}(t_0) &:= u_0, & u^{(2)}(t_0) &:= \partial_t f(t_0, u(t_0)) + \partial_u f(t_0, u(t_0))u^{(1)}(t_0), \\ u^{(1)}(t_0) &:= f(t_0, u(t_0)), & u^{(j)}(t_0) &:= \frac{d^{j-1}}{dt^{j-1}} f(t, u(t))|_{t=t_0}, \quad j \geq 3, \end{aligned}$$

based on the ode system (1.1) and its derivatives.

By considering different choices for $\mathcal{J}_n$ and $\mathcal{I}_n$, we will show that our theory provides sharp bounds on the convergence order. Since $\mathcal{J}_n$ and $\mathcal{I}_n$ are obtained from $\widehat{\mathcal{J}}$ and $\widehat{\mathcal{I}}$ via transformation, we only specify the reference operators.

Each integrator $\widehat{\mathcal{J}}$ that has been used in our calculations is based on Lagrangian interpolation with respect to a specific node set $P_{\widehat{\mathcal{J}}}$. Hence, we have $k_{\mathcal{J}} = 0$. The interpolation operator $\widehat{\mathcal{I}}$ is of Lagrange-type and uses the node set $P_{\widehat{\mathcal{I}}}$. This means that $k_{\mathcal{I}} = 0$. Both node sets are given for each of our test cases. Since often nodes of quadrature formulas are used, we also write for instance “left Gauss–Radau(k)” to indicate that the nodes of the left-sided Gauss–Radau formula with k points have been used. All upcoming settings fulfill Assumption 1.

For all test cases, the method **VTD**₃⁶, which is cGP- $C^1(6)$, was applied as discretization. All calculations were carried out with the software Julia [5] using the floating point data type `BigFloat` with 512 bits. Errors were measured in the norms

$$\|\varphi\|_{L^\infty} := \sup_{t \in I} \|\varphi(t)\|, \quad \|\varphi\|_{\ell^\infty} := \max_{1 \leq n \leq N} \|\varphi(t_n^-)\|$$

where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^d$.

6.1 Case group 1

Case 1 Choosing integration and interpolation according to

$$P_{\widehat{\mathcal{J}}} = \left\{ -\frac{3}{4}, -\frac{1}{4}, \frac{1}{4}, \frac{3}{4} \right\}, \quad P_{\widehat{\mathcal{I}}} = \text{left Gauss–Radau}(3)$$

leads to $r_{\text{ex}}^{\mathcal{J}} = 3$ and $r_{\mathcal{I}}^{\mathcal{J}} = 2$. Hence, the condition $\max \{r_{\text{ex}}^{\mathcal{J}}, r_{\mathcal{I}}^{\mathcal{J}} + 1\} \geq r - 1$ in Corollary 15 is violated and the expected convergence orders for both the L^∞ norm and the $W^{1,\infty}$ seminorm are given by $\min \{r, r_{\mathcal{I}}^{\mathcal{J}} + 1\} = 3$, see (4.18). It can be seen from Table 1 that the theoretical predictions are met by the numerical experiments. Moreover, in accordance with Corollary 21 the ℓ^∞ convergence order is also just 3. This means that the uniform boundedness of $\sup_{t \in I} \|U^{(l)}(t)\|$ required by Theorem 18, which cannot be guaranteed because of $r_{\mathcal{I}} + 1 < r - 1$, is violated since otherwise (5.4) would yield order 4.

Table 1 Errors and convergence orders in different norms for Case 1

N	$\ e\ _{L^\infty}$	ord	$\ e'\ _{L^\infty}$	ord	$\ e\ _{\ell^\infty}$	ord
32	1.648-03		1.126-02		9.439-04	
64	1.413-04	3.54	1.423-03	2.98	1.046-04	3.17
128	1.408-05	3.33	1.876-04	2.92	1.264-05	3.05
256	1.615-06	3.12	2.369-05	2.99	1.559-06	3.02
512	1.961-07	3.04	2.965-06	3.00	1.937-07	3.01
1024	2.426-08	3.01	3.711-07	3.00	2.415-08	3.00
theo		3		3		3

The condition $\max \left\{ r_{\text{ex}}^{\mathcal{J}}, r_{\mathcal{I}}^{\mathcal{J}} + 1 \right\} \geq r - 1$ of Corollary 15 will be fulfilled for all coming cases. Hence, the computations should show the convergence order given by (4.19) for the L^∞ norm.

6.2 Case group 2

This group of cases provides choices for $P_{\widehat{\mathcal{J}}}$ and $P_{\widehat{\mathcal{I}}}$ such that the L^∞ convergence order is limited by the maximum expression inside the outer minimum in (4.19). In addition, the presented cases will show that each of the three terms occurring in the maximum term can limit the convergence order. In the following, we indicate the limiting term in boldface.

Case 2a The choices

$$P_{\widehat{\mathcal{J}}} = \left\{ -\frac{3}{4}, -\frac{1}{4}, \frac{1}{4}, \frac{3}{4} \right\}, \quad P_{\widehat{\mathcal{I}}} = \text{Gauss}(5)$$

provide the L^∞ convergence order $\min\{7, 6, 6, \max\{4, \min\{6, \mathbf{5}\}\}\} = 5$ where the convergence order is limited by the second term inside the inner minimum. We see from Table 2 that the experimental order of convergence is 6, i.e., one order higher than expected. This behavior can be explained by a closer look to Lemma 14. The proof there guarantees in this case that $(v - \mathcal{J}_n^{\mathcal{J}, \mathcal{I}} v)(t_n^-) = 0$ for all $v \in P_5(I_n)$. However, due to symmetry reasons, it holds $\int_{I_n} (v - \mathcal{J}_n^{\mathcal{J}, \mathcal{I}} v)'(t) dt = 0$ for all $v \in P_6(I_n)$ which implies that $(v - \mathcal{J}_n^{\mathcal{J}, \mathcal{I}} v)(t_n^-) = 0$ even for $v \in P_6(I_n)$. Thus, the convergence order of the limiting term is actually better than predicted.

Taking the same setting for $P_{\widehat{\mathcal{J}}}$ but using

$$P_{\widehat{\mathcal{I}}} = \left\{ -\frac{5}{6}, -\frac{13}{23}, \frac{1}{10}, \frac{12}{17}, \frac{4}{5} \right\},$$

the convergence order predicted by (4.19) is $\min\{7, 6, 6, \max\{4, \min\{6, \mathbf{5}\}\}\} = 5$ again. The limitation is here also caused by the second argument of the inner minimum. Table 2 shows under Case 2a* that this convergence order is obtained in the numerical experiments. Note that here the interpolation points just were chosen such

Table 2 Errors and convergence orders in the L^∞ norm for the cases of group 2

N	Case 2a		Case 2a*		Case 2b		Case 2c	
	$\ e\ _{L^\infty}$	ord	$\ e\ _{L^\infty}$	ord	$\ e\ _{L^\infty}$	ord	$\ e\ _{L^\infty}$	ord
32	8.482-06		7.745e-06		8.910-07		1.996-06	
64	1.477-07	5.84	1.500e-07	5.69	9.394-09	6.57	2.792-08	6.16
128	2.398-09	5.95	3.498e-09	5.42	1.081-10	6.44	4.251-10	6.04
256	3.759-11	6.00	9.412e-11	5.22	1.465-12	6.21	6.604-12	6.01
512	5.862-13	6.00	2.775e-12	5.08	2.175-14	6.07	1.034-13	6.00
1024	9.160-15	6.00	8.508e-14	5.03	3.344-16	6.02	1.617-15	6.00
theo		5		5		6		6

that still $r_{\mathcal{I},0}^{\mathcal{J}} = 5 > 4 = r_{\mathcal{I}}^{\mathcal{J}}$, i.e., especially $\widehat{\mathcal{J}}[(\hat{v} - \widehat{\mathcal{J}}^{\mathcal{J},\mathcal{I}}\hat{v})'] = \widehat{\mathcal{J}}[\hat{v}' - \widehat{\mathcal{I}}(\hat{v}')] = 0$ for $\hat{v}(\hat{t}) = \hat{t}^6$, but $\int_{-1}^1 (\hat{v} - \widehat{\mathcal{J}}^{\mathcal{J},\mathcal{I}}\hat{v})'(\hat{t}) d\hat{t} \neq 0$ for $\hat{v}(\hat{t}) = \hat{t}^6$ which ensures that the estimate of Lemma 14 is sharp.

Case 2b If we set

$$P_{\widehat{\mathcal{J}}} = \left\{ -\frac{3}{4}, -\frac{1}{4}, \frac{1}{4}, \frac{3}{4} \right\}, \quad P_{\widehat{\mathcal{I}}} = \left\{ -\frac{3}{4}, -\frac{1}{4}, \frac{1}{4}, \frac{3}{4} \right\},$$

the expected convergence order is $\min\{7, \infty, \infty, \max\{4, \min\{6, \infty\}\}\} = 6$. Here, the limitation comes from the first argument of the inner minimum. The numerical results given in Table 2 clearly show this convergence order.

Case 2c Choosing

$$P_{\widehat{\mathcal{J}}} = \left\{ -1, -\frac{3}{5}, -\frac{1}{5}, \frac{1}{5}, \frac{3}{5}, 1 \right\}, \quad P_{\widehat{\mathcal{I}}} = \left\{ -1, -\frac{3}{5}, -\frac{1}{5}, \frac{1}{5}, \frac{3}{5}, 1 \right\},$$

results in the convergence order $\min\{7, \infty, \infty, \max\{6, \min\{6, \infty\}\}\} = 6$. The first argument of the maximum acts as limitation. The numerical results in Table 2 show that the expected convergence order is obtained. Note that it is not possible that $r_{\text{ex}}^{\mathcal{J}} + 1$ is the only limiting term since the structure of (4.19) implies that $\min\{r + 1, r_{\mathcal{I}}^{\mathcal{J}} + 2\} \geq r_{\text{ex}}^{\mathcal{J}} + 1 \geq \min\{r, r_{\mathcal{I}}^{\mathcal{J}} + 1\}$ if $r_{\text{ex}}^{\mathcal{J}} + 1$ is limiting. Hence, the integer $r_{\text{ex}}^{\mathcal{J}} + 1$ coincides either with $\min\{r + 1, r_{\mathcal{I}}^{\mathcal{J}} + 2\}$ or $\min\{r, r_{\mathcal{I}}^{\mathcal{J}} + 1\}$.

6.3 Case group 3

This group of cases studies the convergence orders in the L^∞ norm and the $W^{1,\infty}$ seminorm. The presented choices will show that each of the first three expressions in the outer minimum in (4.19) can bound the L^∞ convergence order. Moreover, the cases will demonstrate that the convergence order in the $W^{1,\infty}$ seminorm can be limited by both occurring terms in (4.18). Again, the limiting numbers are given in boldface.

Case 3a The choice

$$P_{\widehat{\mathcal{J}}} = \text{Gauss}(6), \quad P_{\widehat{\mathcal{I}}} = \left\{ -1, -\frac{1}{2}, \frac{1}{4}, \frac{3}{4}, 1 \right\}$$

results in

$$\begin{aligned} L^\infty \text{ order} &= \min\{7, 6, \mathbf{5}, \max\{12, \min\{6, 5\}\}\} = 5, \\ W^{1,\infty} \text{ order} &= \min\{6, \mathbf{5}\} = 5. \end{aligned}$$

Hence, the third argument in the outer minimum determines the convergence order for the L^∞ norm while the second argument of the minimum limits the convergence order of the $W^{1,\infty}$ seminorm. The numerical results in Table 3 provide the predicted convergence orders.

Table 3 Errors and convergence orders in different norms for the Cases 3a and 3b

N	Case 3a				Case 3b			
	$\ e\ _{L^\infty}$	ord	$\ e'\ _{L^\infty}$	ord	$\ e\ _{L^\infty}$	ord	$\ e'\ _{L^\infty}$	ord
32	2.306-05		3.193-04		1.242-03		1.717-02	
64	3.786-07	5.93	1.044-05	4.94	8.223-05	3.92	2.169-03	2.98
128	7.266-09	5.70	3.411-07	4.94	5.037-06	4.03	2.842-04	2.93
256	1.716-10	5.40	1.072-08	4.99	3.069-07	4.04	3.581-05	2.99
512	4.688-12	5.19	3.353-10	5.00	1.886-08	4.02	4.479-06	3.00
1024	1.400-13	5.07	1.048-11	5.00	1.168-09	4.01	5.604-07	3.00
theo		5		5		4		3

Case 3b Setting

$$P_{\widehat{\mathcal{J}}} = \text{Gauss}(6), \quad P_{\widehat{\mathcal{I}}} = \text{left Gauss-Radau}(3)$$

gives

$$\begin{aligned} L^\infty \text{ order} &= \min\{7, \mathbf{4}, 5, \max\{12, \min\{6, 3\}\}\} = 4, \\ W^{1,\infty} \text{ order} &= \min\{6, \mathbf{3}\} = 3. \end{aligned}$$

The convergence order in the $W^{1,\infty}$ seminorm is again determined by the second argument of the corresponding minimum. The limitation of the convergence order of the L^∞ norm is caused by the second term. We clearly see from Table 3 that the expected convergence orders are obtained by the numerical simulations.

Case 3c If we take

$$P_{\widehat{\mathcal{J}}} = \text{Gauss}(6), \quad P_{\widehat{\mathcal{I}}} = \text{Gauss}(5),$$

we get

$$\begin{aligned} L^\infty \text{ order} &= \min\{\mathbf{7}, 8, 10, \max\{12, \min\{6, 7\}\}\} = 7, \\ W^{1,\infty} \text{ order} &= \min\{\mathbf{6}, 7\} = 6, \\ \ell^\infty \text{ order} &= \min\{\mathbf{10}, \mathbf{10}, \max\{12, \min\{6, 7\}\}\} = 10. \end{aligned}$$

The convergence orders of the L^∞ norm and the $W^{1,\infty}$ seminorm are limited by the first argument in the corresponding minimum expressions. The numerical results in Table 4 indicate that the predicted orders are achieved. The additionally presented results in the ℓ^∞ show also the predicted behavior. Note that all three error expressions show the optimal convergence orders that are also obtained if exact integration is used and $\mathcal{I}$ is the identity operator.

6.4 Case group 4

This group of cases studies the superconvergence. Hence, we will restrict ourselves to cases where the convergence order in the ℓ^∞ norm suggested by (5.11) is strictly

Table 4 Errors and convergence orders in different norms for Case 3c

N	$\ e\ _{L^\infty}$	ord	$\ e'\ _{L^\infty}$	ord	$\ e\ _{\ell^\infty}$	ord
32	7.603-07		1.669-05		7.587-10	
64	6.524-09	6.86	2.870-07	5.86	7.087-13	10.06
128	5.181-11	6.98	4.556-09	5.98	6.862-16	10.01
256	4.068-13	6.99	7.155-11	5.99	6.689-19	10.00
512	3.181-15	7.00	1.119-12	6.00	6.529-22	10.00
1024	2.486-17	7.00	1.749-14	6.00	6.375-25	10.00
theo		7		6		10

greater than the convergence order in the L^∞ norm given by (4.19). We will show for this situation that the first two arguments in the minimum in (5.11) and the first argument inside the maximum there can limit the ℓ^∞ convergence order. We remind that the limiting term is written in boldface.

Case 4a The choice

$$P_{\widehat{\mathcal{J}}} = \text{Gauss}(6), \quad P_{\widehat{\mathcal{I}}} = \text{Gauss-Lobatto}(5)$$

leads to

$$\begin{aligned} L^\infty \text{ order} &= \min\{7, \mathbf{6}, 8, \max\{12, \min\{6, 5\}\}\} = 6, \\ \ell^\infty \text{ order} &= \min\{10, \mathbf{8}, \max\{12, \min\{6, 5\}\}\} = 8, \end{aligned}$$

see (4.19) and (5.11). The convergence order in ℓ^∞ is bounded by the second argument of the minimum expression. As viewable in Table 5, the expected convergence orders are obtained.

Case 4b Setting

$$P_{\widehat{\mathcal{J}}} = \text{Gauss}(6), \quad P_{\widehat{\mathcal{I}}} = \text{Gauss}(6),$$

the convergence orders

$$\begin{aligned} L^\infty \text{ order} &= \min\{\mathbf{7}, \infty, \infty, \max\{12, \min\{6, \infty\}\}\} = 7, \\ \ell^\infty \text{ order} &= \min\{\mathbf{10}, \infty, \max\{12, \min\{6, \infty\}\}\} = 10 \end{aligned}$$

are expected by our theory. Hence, the convergence order in the ℓ^∞ norm is limited by the first term inside the minimum in (5.11). The numerical results coincide with our predictions.

Case 4c Taking

$$P_{\widehat{\mathcal{J}}} = \text{Gauss}(4), \quad P_{\widehat{\mathcal{I}}} = \text{Gauss}(4)$$

provides

$$\begin{aligned} L^\infty \text{ order} &= \min\{\mathbf{7}, \infty, \infty, \max\{8, \min\{6, \infty\}\}\} = 7, \\ \ell^\infty \text{ order} &= \min\{10, \infty, \max\{\mathbf{8}, \min\{6, \infty\}\}\} = 8, \end{aligned}$$

Table 5 Errors and convergence orders in different norms for the cases of group 4

N	Case 4a				Case 4b			
	$\ e\ _{L^\infty}$	ord	$\ e\ _{\ell^\infty}$	ord	$\ e\ _{L^\infty}$	ord	$\ e\ _{\ell^\infty}$	ord
32	1.058-05		8.552-08		7.560-07		5.899-10	
64	1.717-07	5.94	3.348-10	8.00	6.529-09	6.86	5.610-13	10.04
128	2.835-09	5.92	1.310-12	8.00	5.183-11	6.98	5.452-16	10.01
256	4.464-11	5.99	5.120-15	8.00	4.068-13	6.99	5.318-19	10.00
512	6.981-13	6.00	2.005-17	8.00	3.181-15	7.00	5.192-22	10.00
1024	1.092-14	6.00	7.833-20	8.00	2.486-17	7.00	5.070-25	10.00
theo		6		8		7		10
N	Case 4c				Case 4d			
	$\ e\ _{L^\infty}$	ord	$\ e\ _{\ell^\infty}$	ord	$\ e\ _{L^\infty}$	ord	$\ e\ _{\ell^\infty}$	ord
32	1.617-06		4.037-08		7.603-04		1.867-04	
64	1.387-08	6.86	1.654-10	7.93	4.955-05	3.94	1.132-05	4.04
128	1.101-10	6.98	6.522-13	7.99	3.219-06	3.94	7.017-07	4.01
256	8.654-13	6.99	2.565-15	7.99	2.028-07	3.99	4.377-08	4.00
512	6.759-15	7.00	1.002-17	8.00	1.268-08	4.00	2.735-09	4.00
1024	5.283-17	7.00	3.916-20	8.00	7.929-10	4.00	1.710-10	4.00
theo		7		8		4		4

compare (4.19) and (5.11). The limitation of the ℓ^∞ convergence is caused by the first argument of the maximum in (5.11). The results in Table 5 show clearly the superconvergence since the convergence order in the ℓ^∞ norm is one higher than the convergence order in the L^∞ norm.

Case 4d Choosing

$$P_{\widehat{\mathcal{J}}} = \text{Gauss}(6), \quad P_{\widehat{\mathcal{I}}} = \text{Gauss}(3),$$

the estimates (4.19) and (5.11) suggest

$$\begin{aligned} L^\infty \text{ order} &= \min\{7, \mathbf{4}, 6, \max\{12, \min\{6, 3\}\}\} = 4, \\ \ell^\infty \text{ order} &= \min\{10, \mathbf{6}, \max\{12, \min\{6, 3\}\}\} = 6. \end{aligned}$$

However, since the uniform boundedness of $\sup_{t \in I} \|U^{(l)}(t)\|$ assumed in Theorem 18 cannot be ensured by Lemma 20 due to $r_{\widehat{\mathcal{I}}} + 1 = 3 < 5 = r - 1$, we actually do not expect any superconvergence. These expectations are confirmed by the numerical results given in Table 5. They show that for both the L^∞ and the ℓ^∞ norm convergence order 4 is obtained.

Table 6 Errors and convergence orders in different norms for all cases

Case	$\ e\ _{L^\infty}$		$\ e'\ _{L^\infty}$		$\ e\ _{\ell^\infty}$	
	eoc	theo	eoc	theo	eoc	theo
1	3.04	3	3.00	3	3.01	3
2a	6.00	5	5.00	5	6.00	5
2a*	5.08	5	5.00	5	5.00	5
2b	6.07	6	6.00	6	6.00	6
2c	6.00	6	5.99	6	6.00	6
3a	5.19	5	5.00	5	5.02	5
3b	4.02	4	3.00	3	4.12	4
3c	7.00	7	6.00	6	10.00	10
4a	6.00	6	5.00	5	8.00	8
4b	7.00	7	6.00	6	10.00	10
4c	7.00	7	6.00	6	8.00	8
4d	4.00	4	3.00	3	4.00	4

6.5 Summary

The experimentally obtained and theoretically predicted convergence orders for all cases and all considered norms are collected in Table 6. The experimental orders of convergence were calculated using the results obtained for 256 and 512 time steps.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Akrivis, G., Makridakis, C.H., Nochetto, R.H.: Galerkin and Runge-Kutta methods: unified formulation, a posteriori error estimates and nodal superconvergence. *Numer. Math.* **118**, 429–456 (2011)
2. Becher, S., Matthies, G.: Unified analysis for variational time discretizations of higher order and higher regularity applied to non-stiff odes. *arXiv:2105.06862v1* (2021)
3. Becher, S., Matthies, G.: Variational time discretizations of higher order and higher regularity. *BIT Numer. Math.* <https://doi.org/10.1007/s10543-021-00851-6> (2021)
4. Becher, S., Matthies, G., Wenzel, D.: Variational methods for stable time discretization of first-order differential equations. In: Georgiev, K., Todorov, M., Ivan, G. (eds.), *Advanced Computing in Industrial Mathematics: BGSIAM 2017*, pp. 63–75. Springer International Publishing, Cham (2019)

5. Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B.: Julia: a fresh approach to numerical computing. *SIAM Rev.* **59**(1), 65–98 (2017)
6. Delfour, M., Hager, W., Trochu, F.: Discontinuous Galerkin methods for ordinary differential equations. *Math. Comp.* **36**(154), 455–473 (1981)
7. Delfour, M.C., Dubeau, F.: Discontinuous polynomial approximations in the theory of one-step, hybrid and multistep methods for nonlinear ordinary differential equations. *Math. Comp.* **47**(175), 169–189 (1986)
8. Emmrich, E.: Discrete versions of Gronwall's lemma and their application to the numerical analysis of parabolic problems. Preprint 637-1999, Preprint series of the Institute of Mathematics, Technische Universität Berlin (1999)
9. Estep, D., Stuart, A.: The dynamical behavior of the discontinuous Galerkin method and related difference schemes. *Math. Comp.* **71**(239), 1075–1103 (2002)
10. Hairer, E., Lubich, C.h., Wanner, G.: Geometric numerical integration. Springer-Verlag, New York (2002). Corrected 2nd printing 2004
11. Hairer, E., Nørsett, S.P., Wanner, G. Solving ordinary differential equations I, 2nd edn. Springer-Verlag, New York (1993). Corrected 3rd printing 2008
12. Henrici, P.: Discrete variable methods in ordinary differential equations. Wiley, New York (1962)
13. Hulme, B.L.: Discrete Galerkin related one-step methods for ordinary differential equations. *Math. Comp.* **26**(120), 881–891 (1972)
14. Hulme, B.L.: One-step piecewise polynomial Galerkin methods for initial value problems. *Math. Comp.* **26**(118), 415–426 (1972)
15. Lasaint, P., Raviart, P.-A.: On a finite element method for solving the neutron transport equation. In: Mathematical aspects of finite elements in partial differential equations. (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1974), pp. 89–123 (1974)
16. Mishkov, R.L.: Generalization of the formula of Faà di Bruno for a composite function with a vector argument. *Int. J. Math. & Math. Sci.* **24**(7), 481–491 (2000)
17. Quarteroni, A., Valli, A.: Numerical approximation of partial differential equations. Springer Series in Computational Mathematics. Springer, Berlin (2008)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.