High order schemes based on operator splitting and deferred corrections for stiff time dependent PDEs

Max Duarte, Matthew Emmett

High order schemes based on operator splitting and deferred corrections for stiff time dependent PDE's

Max Duarte*† Matthew Emmett*

April 1, 2016

Abstract

We consider quadrature formulas of high order in time based on Radau–type, L–stable implicit Runge–Kutta schemes to solve time dependent stiff PDEs. Instead of solving a large nonlinear system of equations, we develop a method that performs iterative deferred corrections to compute the solution at the collocation nodes of the quadrature formulas. The numerical stability is guaranteed by a dedicated operator splitting technique that efficiently handles the stiffness of the PDEs and provides initial and intermediate solutions to the iterative scheme. In this way the low order approximations computed by a tailored splitting solver of low algorithmic complexity are iteratively corrected to obtain a high order solution based on a quadrature formula. The mathematical analysis of the numerical errors and local order of the method is carried out in a finite dimensional framework for a general semi–discrete problem, and a time–stepping strategy is conceived to control numerical errors related to the time integration. Numerical evidence confirms the theoretical findings and assesses the performance of the method in the case of a stiff reaction–diffusion equation.

Keywords: High order time discretization, operator splitting, deferred corrections, error control.
AMS subject classifications: 65M12, 65M15, 65L04, 65M20, 65M70, 65G20.

1 Introduction

Operator splitting techniques were originally introduced with the main objective of saving computational costs compared to fully coupled techniques. A complex and potentially large problem could be then split into smaller parts or subproblems of different nature with a significant reduction of the algorithmic complexity and computational requirements. Operator splitting techniques have been used over the past years to carry out numerical simulations in several domains, from biomedical models to combustion or air pollution modeling applications. Moreover, these methods continue to be widely used due to their simplicity of implementation and their high degree of liberty in terms of the choice of dedicated numerical solvers for the split subproblems. In particular, they are suitable for stiff problems for which robust and stable methods that properly handle and damp out fast transients inherent to the different processes must be used. In most applications, first and second order splitting schemes are implemented, for which a general mathematical background is available (see, e.g., [31] for ODEs and [34] for PDEs) that characterizes splitting errors originating from the separate evolution of the split subproblems. Even though higher order splitting schemes have been also developed, more sophisticated numerical implementations are required and their applicability is currently limited to specific linear or nonstiff problems (see, e.g., [12, 10, 33, 6] and discussions therein).

*Center for Computational Sciences and Engineering, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd. MS 50A-1148, Berkeley, CA 94720, USA (MWEmmett@lbl.gov).

†*Present address:* CD-adapco, 200 Shepherds Bush Road, London W6 7NL, UK (max.duarte@cd-adapco.com).

In the past decades high order and stable implicit Runge–Kutta schemes have been developed and widely investigated to solve stiff problems modeled by ODEs (see [32] Sect. IV and references therein). These advances can be naturally exploited to solve stiff semi–discrete problems originating from PDEs discretized in space, as considered, for instance, in [5, 9, 4] to simulate compressible flows. Here we are, in particular, interested in a class of fully implicit Runge–Kutta schemes, built upon collocation methods that use quadrature formulas to numerically approximate integrals [29, 44]. However, the high performance of implicit Runge–Kutta methods for stiff ODEs is adversely affected when applied to large systems of nonlinear equations arising in the numerical integration of semi–discrete PDEs. Significant effort is thus required to achieve numerical implementations that solve the corresponding algebraic problems at reasonable computational expenses. As an alternative to building such a high order implicit solver we consider low order operator splitting schemes, specifically conceived to solve stiff PDEs, embedded in a classical iterative deferred correction scheme (see, e.g., [41]) to approximate the quadrature formulas of an s–stage implicit Runge–Kutta scheme.

In this work, the high order quadrature formulas over a time step Δt, corresponding to an s–stage implicit Runge–Kutta scheme, are evaluated using the numerical approximations computed by a splitting solver at the s intermediate collocation nodes. Such a dedicated splitting solver for stiff PDEs can be built based on the approach introduced in [18]. This approach relies on the use of one–step and high order methods with appropriate stability properties and time–stepping features for the numerical integration of the split subproblems. The splitting time step can be therefore defined independently of standard stability constraints associated with mesh size or stiff source time scales. All the numerical integration within a given Δt is thus performed by the splitting solver with no stability restrictions, while the fully coupled system without splitting is evaluated for the quadrature formulas. Starting from a low order splitting approximation, the stage solutions are then iteratively corrected to obtain a high order quadrature approximation. This is done in the same spirit of other iterative schemes that correct low order approximations to obtain higher order numerical solutions of ODEs or PDEs, like, for instance, the parareal algorithm [35] or SDC (Spectral Deferred Correction) schemes [20].

This paper is organized as follows. Time operator splitting as well as implicit Runge–Kutta schemes are briefly introduced in Section 2. The new deferred correction splitting algorithm is introduced in Section 3. A mathematical analysis of the numerical errors is conducted in Section 4, and a time step selection strategy with error control is subsequently derived in Section 5. Spatial discretization errors are briefly discussed in Section 6. Finally, theoretical findings are numerically investigated in Section 7 for a stiff reaction–diffusion problem.

2 Mathematical background

We consider a parabolic, time dependent PDE given by

$$\left.\begin{array}{ll} \partial_t u = F(\partial^2_{\boldsymbol{x}} u, \partial_{\boldsymbol{x}} u, u), & t > t_0,\, \boldsymbol{x} \in \mathbb{R}^d, \\ u(t_0, \boldsymbol{x}) = u_0(\boldsymbol{x}), & t = t_0,\, \boldsymbol{x} \in \mathbb{R}^d, \end{array}\right\} \tag{1}$$

where $u : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^m$ and $F : \mathbb{R}^m \to \mathbb{R}^m$. For many physically inspired systems, the right hand side F can be split into components according to

$$F(\partial^2_{\boldsymbol{x}} u, \partial_{\boldsymbol{x}} u, u) = F(u) = F_1(u) + F_2(u) + \ldots, \tag{2}$$

where the $F_i(u)$, $i = 1, \ldots$, represent different physical processes. For instance, a scalar nonlinear reaction–diffusion equation with $u : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}$ would be split into $F_1(u) = -\partial_{\boldsymbol{x}} \cdot (D(u)\partial_{\boldsymbol{x}} u)$ and $F_2(u) = f(u)$ for some diffusion coefficient $D : \mathbb{R} \to \mathbb{R}$, and nonlinear function $f : \mathbb{R} \to \mathbb{R}$. Within

the general framework of (1), we consider the semi–discrete problem

$$\left.\begin{array}{ll} \mathrm{d}_t \boldsymbol{u} = \boldsymbol{F}(\boldsymbol{u}), & t > t_0, \\ \boldsymbol{u}(0) = \boldsymbol{u}_0, & t = t_0, \end{array}\right\} \tag{3}$$

corresponding to problem (1) discretized on a grid $\mathbf{X}$ of size N; $\boldsymbol{u}$, $\boldsymbol{u}_0 \in \mathbb{R}^{m\cdot N}$ and $\boldsymbol{F} : \mathbb{R}^{m\cdot N} \to \mathbb{R}^{m\cdot N}$. Analogous to (2), we split $\boldsymbol{F}(\boldsymbol{u})$ into components $\boldsymbol{F}_i(\boldsymbol{u})$.

2.1 Time operator splitting techniques

Assuming for simplicity that $\boldsymbol{F}(\boldsymbol{u}) = \boldsymbol{F}_1(\boldsymbol{u}) + \boldsymbol{F}_2(\boldsymbol{u})$, we denote the solution of the subproblem

$$\left.\begin{array}{ll} \mathrm{d}_t \boldsymbol{v} = \boldsymbol{F}_1(\boldsymbol{v}), & t > t_0, \\ \boldsymbol{v}(0) = \boldsymbol{v}_0, & t = t_0, \end{array}\right\}$$

by $X^t\boldsymbol{v}_0$, and the solution of the subproblem

$$\left.\begin{array}{ll} \mathrm{d}_t \boldsymbol{w} = \boldsymbol{F}_2(\boldsymbol{w}), & t > t_0, \\ \boldsymbol{w}(0) = \boldsymbol{w}_0, & t = t_0. \end{array}\right\}$$

by $Y^t\boldsymbol{w}_0$. The Lie (or Lie–Trotter [43]) splitting approximations to the solution of problem (3) are then given by

$$\mathcal{L}_1^t\boldsymbol{u}_0 = X^tY^t\boldsymbol{u}_0, \qquad \mathcal{L}_2^t\boldsymbol{u}_0 = Y^tX^t\boldsymbol{u}_0. \tag{4}$$

Lie approximations are of first order in time; second order can be achieved by using symmetric Strang (or Marchuk [36]) formulas [42] to obtain

$$\mathcal{S}_1^t\boldsymbol{u}_0 = X^{t/2}Y^tX^{t/2}\boldsymbol{u}_0, \qquad \mathcal{S}_2^t\boldsymbol{u}_0 = Y^{t/2}X^tY^{t/2}\boldsymbol{u}_0. \tag{5}$$

Similar constructions follow for more than two subproblems in (2). Let us denote by $\mathcal{S}^{\Delta t}\boldsymbol{u}_0$ any of these four splitting approximations, where Δt represents the *splitting time step* (i.e., the overall time–marching algorithm would compute $\boldsymbol{u}_{n+1} = \mathcal{S}^{\Delta t}\boldsymbol{u}_n$). In this work we consider splitting approximations built in practice under the precepts established in [18]. In particular, dedicated solvers with time–stepping features are implemented to separately integrate each split subproblem during successive splitting time steps such that the numerical stability of the method is always guaranteed.

2.2 Time implicit Runge–Kutta schemes

Let us now consider an implicit s–stage Runge–Kutta scheme to discretize (3) in time. Given a time step Δt, the solution $\boldsymbol{u}(t_0 + \Delta t)$ is approximated by $\boldsymbol{u}_1^{\mathrm{RK}}$, computed as

$$\boldsymbol{u}_i = \boldsymbol{u}_0 + \Delta t \sum_{j=1}^{s} a_{ij}\boldsymbol{F}\left(t_0 + c_j\Delta t, \boldsymbol{u}_j\right), \qquad i = 1, \ldots, s; \tag{6}$$

$$\boldsymbol{u}_1^{\mathrm{RK}} = \boldsymbol{u}_0 + \Delta t \sum_{j=1}^{s} b_j\boldsymbol{F}\left(t_0 + c_j\Delta t, \boldsymbol{u}_j\right). \tag{7}$$

The arrays $\boldsymbol{b}$, $\boldsymbol{c} \in \mathbb{R}^s$ gather the various coefficients $\boldsymbol{b} = (b_1, \ldots, b_s)^T$ and $\boldsymbol{c} = (c_1, \ldots, c_s)^T$, and $\boldsymbol{A} \in \mathcal{M}_s(\mathbb{R})$ such that $\boldsymbol{A} = (a_{ij})_{1\le i,j\le s}$. These coefficients define the stability properties and the order conditions of the method, and are usually arranged in a Butcher tableau according to

$$\begin{array}{c|c} \boldsymbol{c} & \boldsymbol{A} \\ \hline & \boldsymbol{b}^T \end{array}.$$

Recall that, if all the elements of the matrix of coefficients $\boldsymbol{A}$ are nonzero, then the scheme is said to be a *fully IRK scheme* [32]. Moreover, if

$$a_{sj} = b_j, \quad j = 1, \ldots, s, \tag{8}$$

then the last stage $\boldsymbol{u}_s$ corresponds to the solution $\boldsymbol{u}_1^{\mathrm{RK}}$ according to (6)–(7).

Methods satisfying (8) are called *stiffly accurate* [39] and are particularly appropriate for the solution of (stiff) singular perturbation problems and for differential–algebraic equations [32]. Fully IRK schemes with a number of stages inferior to its approximation order can be built based on collocation methods [29, 44], together with the simplified order conditions introduced by Butcher [7]. In this case, the coefficients $\boldsymbol{b}$ correspond to the quadrature formula of order p such that $\int_0^1 \pi(\tau)\,\mathrm{d}\tau = \sum_{j=1}^s b_j \pi(c_j)$ for polynomials $\pi(\tau)$ of degree $\leq p-1$. Moreover, the coefficients in $\boldsymbol{c}$ and $\boldsymbol{A}$, together with conditions for the stage order q, imply that at every stage $i = 1, \ldots, s$ the quadrature formula $\int_0^{c_i} \pi(\tau)\,\mathrm{d}\tau = \sum_{j=1}^s a_{ij}\pi(c_j)$ holds for polynomials $\pi(\tau)$ of degree $\leq q-1$. Depending on the quadrature formula considered, for instance, Gauss, Radau or Lobatto, different families of implicit Runge–Kutta methods can be constructed (for more details, see [32] Sect. IV.5).

In this work we consider the family of RadauIIA quadrature formulas introduced by Ehle [22], based on [8], that consider Radau quadrature formulas [40] such that $p = 2s-1$ and $q = s$. These are A– and L–stable schemes that are stiffly accurate methods according to (8). Note that, even though Gauss methods attain a maximum order of $p = 2s$ [7, 21], they are neither stiffly accurate nor L–stable schemes, which are both important properties for stiff problems. Approximations of less order are obtained with Lobatto methods satisfying $p = 2s-2$ [7, 21, 11, 2]. In particular the collocation methods with $p = 2s-2$ and $q = s$, known as the LobattoIIIA methods, yield stiffly accurate schemes, but these are only A–stable. As such, we hereafter focus on the RadauIIA quadrature formula of order 5 given by ([32] Table IV.5.6)

$$\begin{array}{c|ccc} \frac{4-\sqrt{6}}{10} & \frac{88-7\sqrt{6}}{360} & \frac{296-169\sqrt{6}}{1800} & \frac{-2+3\sqrt{6}}{225} \\ \frac{4+\sqrt{6}}{10} & \frac{296+169\sqrt{6}}{1800} & \frac{88+7\sqrt{6}}{360} & \frac{-2-3\sqrt{6}}{225} \\ 1 & \frac{16-\sqrt{6}}{36} & \frac{16+\sqrt{6}}{36} & \frac{1}{9} \\ \hline & \frac{16-\sqrt{6}}{36} & \frac{16+\sqrt{6}}{36} & \frac{1}{9} \end{array} . \tag{9}$$

3 Iterative construction of high order schemes

With this background, we describe in what follows how a high order quadrature approximation can be computed through an iterative scheme. Following the IRK scheme (6), we define $t_i = t_0 + c_i \Delta t$, $i = 1, \ldots, s$ and introduce the set $\boldsymbol{U} = (\boldsymbol{u}_i)_{i=1,\ldots,s}$, which is comprised of the approximations to the solution $\boldsymbol{u}(t_i)$ of (3) at the intermediate times. Since $c_s = 1$ for stiffly accurate schemes, and in particular for RadauIIA quadrature formulas, $\boldsymbol{u}_s$ stands for the approximation to $\boldsymbol{u}(t_0 + \Delta t)$, which was denoted as $\boldsymbol{u}_1^{\mathrm{RK}}$ in (7). To simplify the discussion that follows, we define

$$I_{t_0}^{t_i}(\boldsymbol{U}) := \Delta t \sum_{j=1}^{s} a_{ij} \boldsymbol{F}\left(t_0 + c_j \Delta t, \boldsymbol{u}_j\right), \tag{10}$$

so that the IRK scheme (6) can be recast as

$$\boldsymbol{u}_i = \boldsymbol{u}_0 + I_{t_0}^{t_i}(\boldsymbol{U}), \qquad i = 1, \ldots, s, \tag{11}$$

or equivalently as

$$\boldsymbol{u}_i = \boldsymbol{u}_{i-1} + I_{t_{i-1}}^{t_i}(\boldsymbol{U}), \qquad i = 1, \ldots, s, \tag{12}$$

where $I_{t_{i-1}}^{t_i}(\boldsymbol{U}) := I_{t_0}^{t_i}(\boldsymbol{U}) - I_{t_0}^{t_{i-1}}(\boldsymbol{U})$. Notice that $\int_{c_i}^{c_l} \pi(\tau)\,\mathrm{d}\tau = \sum_{j=1}^{s}(a_{lj} - a_{ij})\pi(c_j)$ still holds for $l > i$ and polynomials $\pi(\tau)$ of degree $\leq q-1$; and therefore $I_{t_{i-1}}^{t_i}(\boldsymbol{U})$ retains the stage order q.

To compute $\boldsymbol{U}$, that is, the stage values of the IRK scheme (6), and hence the Runge–Kutta approximation (7), we need to solve a nonlinear system of equations of size $m \times N \times s$ given by (6) or (11). One of the most common ways to do this consists of implementing Newton's method (see, e.g., [32] Sect. IV.8). In this work, however, we approximate $\boldsymbol{U}$ by an iterative deferred correction technique.

3.1 Deferred correction splitting scheme

Given a provisional set of solutions denoted by $\widetilde{\boldsymbol{U}} = (\widetilde{\boldsymbol{u}}_i)_{i=1,\dots,s}$, we can compute an approximation $\widehat{\boldsymbol{U}} = (\widehat{\boldsymbol{u}}_i)_{i=1,\dots,s}$ to $\boldsymbol{U}$, following (12), according to

$$\widehat{\boldsymbol{u}}_i = \widetilde{\boldsymbol{u}}_{i-1} + I_{t_{i-1}}^{t_i}(\widetilde{\boldsymbol{U}}), \qquad i = 1, \dots, s, \tag{13}$$

with $\widetilde{\boldsymbol{u}}_0 = \boldsymbol{u}_0$. This approximation is then iteratively corrected by computing

$$\widetilde{\boldsymbol{u}}_i^{k+1} = \widehat{\boldsymbol{u}}_i^{k} + \boldsymbol{\delta}_i^{k}, \qquad i = 1, \dots, s, \tag{14}$$

which defines a new set of provisional solutions $\widetilde{\boldsymbol{U}}^{k+1}$ corresponding to iteration $k+1$, and a new approximate solution $\widehat{\boldsymbol{U}}^{k+1}$ according to (13). When the corrections $(\boldsymbol{\delta}_i)_{i=1,\dots,s}$ become negligible, we expect that $\widetilde{\boldsymbol{U}}$ approaches the quadrature formulas used in the IRK scheme (6) (or (11)), with $\widetilde{\boldsymbol{u}}_s$ approximating $\boldsymbol{u}_1^{\mathrm{RK}}$ in (7).

Taking into account that system (1) (or (3)) is stiff, we follow the approach of [18] and consider an operator splitting technique with dedicated time integration schemes for each subproblem to handle stiffness and guarantee numerical stability. We thus embed such an operator splitting technique ((4) or (5)) into the deferred correction scheme (13)–(14), yielding the deferred correction splitting (DC–S) algorithm. The first approximation $\widetilde{\boldsymbol{U}}^0$ is obtained directly by recursively applying an operator splitting scheme as follows:

$$\widetilde{\boldsymbol{u}}_1^0 = \mathcal{S}^{c_1\Delta t}\boldsymbol{u}_0, \qquad \widetilde{\boldsymbol{u}}_i^0 = \mathcal{S}^{(c_i - c_{i-1})\Delta t}\widetilde{\boldsymbol{u}}_{i-1}^0, \quad i = 2, \dots, s. \tag{15}$$

We then define the corrections in (14) as

$$\boldsymbol{\delta}_i^k = \mathcal{S}^{(c_i - c_{i-1})\Delta t}\widetilde{\boldsymbol{u}}_{i-1}^{k+1} - \mathcal{S}^{(c_i - c_{i-1})\Delta t}\widetilde{\boldsymbol{u}}_{i-1}^{k}, \qquad i = 2, \dots, s.$$

Noticing that $\boldsymbol{\delta}_1^k = \boldsymbol{0}$, (14) becomes

$$\widetilde{\boldsymbol{u}}_1^{k+1} = \widehat{\boldsymbol{u}}_1^{k}, \qquad \widetilde{\boldsymbol{u}}_i^{k+1} = \widehat{\boldsymbol{u}}_i^{k} + \mathcal{S}^{(c_i - c_{i-1})\Delta t}\widetilde{\boldsymbol{u}}_{i-1}^{k+1} - \mathcal{S}^{(c_i - c_{i-1})\Delta t}\widetilde{\boldsymbol{u}}_{i-1}^{k}, \quad i = 2, \dots, s. \tag{16}$$

The numerical time integration of problem (3) is thus performed using an operator splitting technique throughout the time step Δt. These results are subsequently used to approximate the solutions $(\widetilde{\boldsymbol{u}}_j^k)_{j=1,\dots,s}$ at the collocation nodes in the quadrature formula (13), corresponding in our case to a RadauIIA quadrature formula. The coefficients of matrix $\boldsymbol{A}$ are embedded in the definition of $I_{t_{i-1}}^{t_i}(\widetilde{\boldsymbol{U}})$ in (13) following (10), while coefficients $\boldsymbol{c}$ define the length of the time substeps within a given Δt. Note that the fully coupled $\boldsymbol{F}(\widetilde{\boldsymbol{u}}_j^k)$, $j = 1, \dots, s$, is evaluated in $I_{t_{i-1}}^{t_i}(\widetilde{\boldsymbol{U}}^k)$ (eq. (13)) to compute $\widehat{\boldsymbol{u}}_i^k$, $i = 1, \dots, s$. Denoting as $\widehat{p}$ the order of the operator splitting scheme, we will demonstrate in the following section that each iteration increases the order of the initial numerical approximation (15) by one, that is, the local error of the method (16) behaves like $\mathcal{O}(\Delta t^{\widehat{p}+k+1})$, potentially up to the order p of the quadrature formula and at least up to its stage order plus one: $q+1$. As previously said, the numerical stability of the time integration process is guaranteed by

the use of a dedicated operator splitting solver [18], whereas the quadrature formulas correspond to the L–stable RadauIIA–IRK scheme. Consequently, Δt is not subject to any stability constraint and a time–stepping criterion to select the time step size based on some error estimate should be introduced. In particular if the splitting solver considers only time–implicit L–stable methods, the overall DC–S scheme will also be L–stable. However, one might be interested in using less computationally expensive explicit methods within the splitting solver.

The relation of the DC–S scheme (15)–(16) with other iterative methods in the literature, namely, the parareal algorithm [35] and SDC schemes [20], is briefly discussed in Appendices A and B.

4 Analysis of numerical errors

We investigate the behavior of the numerical error associated with the DC–S scheme (15)–(16), considering problem (3) in a finite dimensional setting. Let X be a Banach space with norm $\|\cdot\|$ and $\boldsymbol{F}$ an unbounded nonlinear operator form $D(\boldsymbol{F}) \subset X$ to X. We assume that $\boldsymbol{u}(t)$, which is the solution of (3), also belongs to X and that the same follows for the solutions of the split subproblems. Assuming a Lipschitz condition for $\boldsymbol{F}(\boldsymbol{u})$ given by

$$\|\boldsymbol{F}(\boldsymbol{u}) - \boldsymbol{F}(\boldsymbol{v})\| \leq \kappa \|\boldsymbol{u} - \boldsymbol{v}\|, \tag{17}$$

we have, given (10), that

$$\left\| I_{t_{i-1}}^{t_i}(\boldsymbol{U}) - I_{t_{i-1}}^{t_i}(\boldsymbol{V}) \right\| \leq C_1 \Delta t \sum_{j=1}^{s} \|\boldsymbol{u}_j - \boldsymbol{v}_j\|, \qquad i = 1, \ldots, s. \tag{18}$$

Furthermore, considering (10) with the exact solution $\boldsymbol{u}(t)$ gives

$$I_{t_0}^{t_i}\left[(\boldsymbol{u}(t_j))_{j=1,\ldots,s}\right] = \Delta t \sum_{j=1}^{s} a_{ij} \boldsymbol{F}\left(t_0 + c_j \Delta t, \boldsymbol{u}(t_j)\right); \tag{19}$$

and hence, considering the quadrature formula for the IRK scheme, we obtain

$$\left\| \int_{t_{i-1}}^{t_i} \boldsymbol{F}(\boldsymbol{u}(\tau))\, \mathrm{d}\tau - I_{t_{i-1}}^{t_i}\left[(\boldsymbol{u}(t_j))_{j=1,\ldots,s}\right] \right\| \leq \eta_{i-1}^{i} := C_2 \Delta t^{q+1}, \qquad i = 1, \ldots, s, \tag{20}$$

and for a stiffly accurate method,

$$\left\| \int_{t_0}^{t_s} \boldsymbol{F}(\boldsymbol{u}(\tau))\, \mathrm{d}\tau - I_{t_0}^{t_s}\left[(\boldsymbol{u}(t_j))_{j=1,\ldots,s}\right] \right\| \leq \eta_0^s := C_3 \Delta t^{p+1},$$

recalling that q and p stand, respectively, for the stage order and global order of the collocation method. For the RadauIIA quadrature formulas, recall that $p = 2s - 1$ and $q = s$.

Denoting the exact solution to problem (3) as $\boldsymbol{u}(t) = \mathcal{T}^t \boldsymbol{u}_0$, we can in general write for the splitting approximation that

$$\mathcal{T}^{\Delta t}\boldsymbol{u}_0 - \mathcal{S}^{\Delta t}\boldsymbol{u}_0 = c_{\widehat{p}+1}(\boldsymbol{u}_0)\Delta t^{\widehat{p}+1} + c_{\widehat{p}+2}(\boldsymbol{u}_0)\Delta t^{\widehat{p}+2} + \ldots,$$

recalling that $\widehat{p}$ stands for the order of approximation of the splitting scheme (see, e.g., [34] Sect. IV.1 or [31] Sect. III). We then assume that the local truncation error of the splitting approximation is bounded according to

$$\left\| \mathcal{T}^{\Delta t}\boldsymbol{u}_0 - \mathcal{S}^{\Delta t}\boldsymbol{u}_0 \right\| \leq C_4 \Delta t^{\widehat{p}+1}, \tag{21}$$

and that the following bound

$$\left\| \mathcal{T}^{\Delta t}\boldsymbol{u}_0 - \mathcal{S}^{\Delta t}\boldsymbol{u}_0 - \left[\mathcal{T}^{\Delta t}\boldsymbol{v}_0 - \mathcal{S}^{\Delta t}\boldsymbol{v}_0\right] \right\| \leq C_5 \Delta t^{\widehat{p}+1} \|\boldsymbol{u}_0 - \boldsymbol{v}_0\| \tag{22}$$

is also valid. Moreover, we assume that $\mathcal{S}^t$ satisfies the Lipschitz condition

$$\left\|\mathcal{S}^{\Delta t}\boldsymbol{u}_0 - \mathcal{S}^{\Delta t}\boldsymbol{v}_0\right\| \leq (1 + C_6\Delta t)\|\boldsymbol{u}_0 - \boldsymbol{v}_0\|. \tag{23}$$

To simplify the discussion that follows, we define the error e_i^k at each collocation node i and iteration k according to

$$\left\|\boldsymbol{u}(t_i) - \widetilde{\boldsymbol{u}}_i^k\right\| \leq e_i^k, \quad i = 1, \ldots, s, \tag{24}$$

and the sum of these errors as

$$\Xi^{k,s} := \sum_{j=1}^{s} e_j^k. \tag{25}$$

With these definitions, we prove the following theorem.

Theorem 1. *Considering problem (3) with the Lipschitz condition for $\boldsymbol{F}(\boldsymbol{u})$ (17), the DC–S iteration (16) with a given time step Δt, and assumptions (22) and (23) for the splitting approximation, there are positive constants A, B, such that for $k = 1, 2, \ldots$,*

$$\Xi^{k,s} \leq A + B\Xi^{k-1,s}. \tag{26}$$

Proof. Defining $\Delta t_i := (c_i - c_{i-1})\Delta t$, $i = 2, \ldots, s$, we have from (16),

$$\begin{aligned}
\boldsymbol{u}(t_i) - \widetilde{\boldsymbol{u}}_i^{k+1} =& \boldsymbol{u}(t_i) - \widetilde{\boldsymbol{u}}_{i-1}^k - I_{t_{i-1}}^{t_i}(\widetilde{\boldsymbol{U}}^k) - \mathcal{S}^{\Delta t_i}\widetilde{\boldsymbol{u}}_{i-1}^{k+1} + \mathcal{S}^{\Delta t_i}\widetilde{\boldsymbol{u}}_{i-1}^k \\
=& \int_{t_{i-1}}^{t_i} \boldsymbol{F}(\widetilde{\boldsymbol{u}}^k(\tau))\,\mathrm{d}\tau - I_{t_{i-1}}^{t_i}(\widetilde{\boldsymbol{U}}^k) \\
&+ \mathcal{T}^{\Delta t_i}\boldsymbol{u}(t_{i-1}) - \mathcal{S}^{\Delta t_i}\boldsymbol{u}(t_{i-1}) - \left[\mathcal{T}^{\Delta t_i}\widetilde{\boldsymbol{u}}_{i-1}^k - \mathcal{S}^{\Delta t_i}\widetilde{\boldsymbol{u}}_{i-1}^k\right] \\
&+ \mathcal{S}^{\Delta t_i}\boldsymbol{u}(t_{i-1}) - \mathcal{S}^{\Delta t_i}\widetilde{\boldsymbol{u}}_{i-1}^{k+1},
\end{aligned}$$

with $k = 1, 2, \ldots$, after adding and subtracting $\mathcal{T}^{\Delta t_i}\widetilde{\boldsymbol{u}}_{i-1}^k = \widetilde{\boldsymbol{u}}_{i-1}^k + \int_{t_{i-1}}^{t_i} \boldsymbol{F}(\widetilde{\boldsymbol{u}}^k(\tau))\,\mathrm{d}\tau$, and $\mathcal{S}^{\Delta t_i}\boldsymbol{u}(t_{i-1})$; and similarly,

$$\begin{aligned}
\boldsymbol{u}(t_1) - \widetilde{\boldsymbol{u}}_1^{k+1} =& \boldsymbol{u}(t_1) - \boldsymbol{u}_0 - I_{t_0}^{t_1}(\widetilde{\boldsymbol{U}}^k) \\
=& \int_{t_0}^{t_1} \boldsymbol{F}(\boldsymbol{u}(\tau))\,\mathrm{d}\tau - I_{t_0}^{t_1}\left[(\boldsymbol{u}(t_j))_{j=1,\ldots,s}\right] \\
&+ I_{t_0}^{t_1}\left[(\boldsymbol{u}(t_j))_{j=1,\ldots,s}\right] - I_{t_0}^{t_1}(\widetilde{\boldsymbol{U}}^k),
\end{aligned}$$

after adding and subtracting $I_{t_0}^{t_1}\left[(\boldsymbol{u}(t_j))_{j=1,\ldots,s}\right]$.

Taking norms and considering (20), (22), and (23), there exist some α and β such that

$$\left\|\boldsymbol{u}(t_i) - \widetilde{\boldsymbol{u}}_i^{k+1}\right\| \leq \eta_{i-1}^i + \alpha\left\|\boldsymbol{u}(t_{i-1}) - \widetilde{\boldsymbol{u}}_{i-1}^k\right\| + \beta\left\|\boldsymbol{u}(t_{i-1}) - \widetilde{\boldsymbol{u}}_{i-1}^{k+1}\right\|, \tag{27}$$

with $\alpha := C_5\Delta t^{\widehat{p}+1}$ and $\beta := 1 + C_6\Delta t$. Similarly, from (18) and (19) there is a $\lambda := C_1\Delta t$ such that

$$\left\|\boldsymbol{u}(t_1) - \widetilde{\boldsymbol{u}}_1^{k+1}\right\| \leq \eta_0^1 + \lambda\sum_{j=1}^{s}\left\|\boldsymbol{u}(t_j) - \widetilde{\boldsymbol{u}}_j^k\right\|. \tag{28}$$

Summing (27)–(28) over i and considering the notation (24), we have for $i = 2, \ldots, s$:

$$\begin{aligned}
\Xi^{k+1,i} &= \eta_0^i + \lambda\Xi^{k,s} + \alpha\Xi^{k,i-1} + \beta\Xi^{k+1,i-1} \\
&\leq \eta_0^i + (\lambda + \alpha)\Xi^{k,s} + \beta\Xi^{k+1,i-1}.
\end{aligned} \tag{29}$$

In particular, from (28) we have

$$\Xi^{k,1} = e_1^k = \eta_0^1 + \lambda \Xi^{k-1,s} \leq \eta_0^1 + (\lambda + \alpha)\Xi^{k-1,s}, \tag{30}$$

and using the inequalities considered in (29) and (30) with $\mathcal{B}^i := \sum_{j=0}^{i} \beta^j$, we obtain by mathematical induction over i,

$$\Xi^{k,i} \leq \sum_{j=1}^{i} \beta^{i-j} \eta_0^j + (\lambda + \alpha)\mathcal{B}^{i-1}\Xi^{k-1,s}, \tag{31}$$

which proves (26). □

Bound (26) in Theorem 1 accounts for the approximation errors accumulated over the time subintervals at a given iteration and for the way the sum of these errors behaves from one iteration to the next one. The next Corollary investigates the accumulation of errors after a given number of iterations.

Corollary 1. *Considering Theorem 1, assumption (21) for the splitting approximation with* $\gamma := C_4 \Delta t^{\widehat{p}+1}$*, and* $\mathcal{B}^i := \sum_{j=0}^{i} \beta^j$ *with* $\beta := 1 + C_6 \Delta t$*, we have for* $k = 1, 2, \ldots,$

$$\Xi^{k,s} \leq A\left[1 + B + B^2 + \cdots + B^{k-1}\right] + s\gamma \mathcal{B}^{s-1} B^k, \tag{32}$$

with $A = \sum_{j=1}^{s} \beta^{s-j} \eta_0^j$ *and* $B = (\lambda + \alpha)\mathcal{B}^{s-1}$.

Proof. Considering (15) we have

$$\begin{aligned} \boldsymbol{u}(t_s) - \widetilde{\boldsymbol{u}}_s^0 &= \boldsymbol{u}(t_s) - \mathcal{S}^{\Delta t_s} \widetilde{\boldsymbol{u}}_{s-1}^0 \\ &= \mathcal{T}^{\Delta t_s} \boldsymbol{u}(t_{s-1}) - \mathcal{S}^{\Delta t_s} \boldsymbol{u}(t_{s-1}) + \mathcal{S}^{\Delta t_s} \boldsymbol{u}(t_{s-1}) - \mathcal{S}^{\Delta t_s} \widetilde{\boldsymbol{u}}_{s-1}^0, \end{aligned}$$

and thus after taking norms,

$$e_s^0 = \gamma + \beta e_{s-1}^0 = \gamma \mathcal{B}^{s-1}.$$

Noticing that

$$\Xi^{0,s} = \gamma\left[1 + \mathcal{B} + \mathcal{B}^2 + \cdots + \mathcal{B}^{s-1}\right] \leq s\gamma \mathcal{B}^{s-1},$$

bound (32) follows by mathematical induction over k using (26). □

With Theorem 1 and Corollary 1, the following can be proved.

Theorem 2. *For the DC–S scheme (15)–(16) with* $k = 1, 2, \ldots,$ *if* $C_1 < 1$ *in (18) and* $C_6 \Delta t < 1$ *in (23), the following bound holds:*

$$\left\| \boldsymbol{u}(t_0 + \Delta t) - \widetilde{\boldsymbol{u}}_s^k \right\| \leq c\, \Delta t^{\min[p+1, q+2, \widehat{p}+k+1]}, \tag{33}$$

with $c \geq \max\{C_3, C_2 C_6, s C_4 C_1\}$.

Proof. From (31) in Theorem 1 and considering that $\left\| \boldsymbol{u}(t_0 + \Delta t) - \widetilde{\boldsymbol{u}}_s^k \right\| \leq e_s^k = \Xi^{k,s} - \Xi^{k,s-1}$, we obtain

$$\begin{aligned} e_s^k &\leq \sum_{j=1}^{s} \beta^{s-j} \eta_{j-1}^j + (\lambda + \alpha)\beta^{s-1}\Xi^{k-1,s} \\ &\leq \sum_{j=1}^{s} \beta^{s-j} \eta_{j-1}^j + (\lambda + \alpha)\beta^{s-1}\left(A\left[1 + B + B^2 + \cdots + B^{k-2}\right]\right) \\ &\quad + (\lambda + \alpha)\beta^{s-1} s\gamma \mathcal{B}^{s-1} B^{k-1}, \end{aligned} \tag{34}$$

according to Corollary 1 with $A = \sum_{j=1}^{s} \beta^{s-j}\eta_0^j$ and $B = (\lambda+\alpha)\mathcal{B}^{s-1}$. Recalling that $\beta = 1+C_6\Delta t$, $\eta_{j-1}^j = C_2\Delta t^{q+1}$, $j = 1,\ldots,s$, and $\eta_0^s = C_3\Delta t^{p+1}$, and considering the binomial series, we have that

$$\begin{aligned}\sum_{j=1}^{s}\beta^{s-j}\eta_{j-1}^j &= \sum_{j=1}^{s}\left[\sum_{l=0}^{\infty}\left(\begin{array}{c} s-j \\ l \end{array}\right)(C_6\Delta t)^l\right]\eta_{j-1}^j \\ &= \eta_0^s + \sum_{j=1}^{s-1}\left[\sum_{l=1}^{\infty}\left(\begin{array}{c} s-j \\ l \end{array}\right)(C_6\Delta t)^l\right]\eta_{j-1}^j \\ &= C_3\Delta t^{p+1} + C_2\Delta t^{q+1}\sum_{j=1}^{s-1}\sum_{l=1}^{\infty}\left(\begin{array}{c} s-j \\ l \end{array}\right)(C_6\Delta t)^l; \end{aligned} \tag{35}$$

and similarly,

$$A = \sum_{j=1}^{s}\beta^{s-j}\eta_0^j = C_3\Delta t^{p+1} + C_2\Delta t^{q+1}(1+C_6\Delta t)\sum_{l=1}^{\infty}\left(\begin{array}{c} s-1 \\ l \end{array}\right)(C_6\Delta t)^{l-1},$$

where we have considered that

$$\mathcal{B}^{s-1} = \sum_{j=0}^{s-1}\beta^j = \frac{1-\beta^s}{1-\beta} = \sum_{l=1}^{\infty}\left(\begin{array}{c} s \\ l \end{array}\right)(C_6\Delta t)^{l-1}.$$

Since $\lambda = C_1\Delta t$ and $\alpha = C_5\Delta t^{\widehat{p}+1}$, we know that

$$(\lambda+\alpha)\beta^{s-1} = (C_1\Delta t + C_5\Delta t^{\widehat{p}+1})\sum_{l=0}^{\infty}\left(\begin{array}{c} s-1 \\ l \end{array}\right)(C_6\Delta t)^l$$

and

$$B = (\lambda+\alpha)\mathcal{B}^{s-1} = (C_1\Delta t + C_5\Delta t^{\widehat{p}+1})\sum_{l=1}^{\infty}\left(\begin{array}{c} s \\ l \end{array}\right)(C_6\Delta t)^{l-1};$$

we thus have with $\gamma = C_4\Delta t^{\widehat{p}+1}$ that

$$(\lambda+\alpha)\beta^{s-1}s\gamma\mathcal{B}^{s-1}B^{k-1} = sC_4C_1^k\Delta t^{\widehat{p}+k+1} + \mathcal{O}(C_1^k\Delta t^{\widehat{p}+k+2}), \tag{36}$$

which together with (35) and (34) prove (33) with $C_1 < 1$ into (36), taking into account that the expression $(\lambda+\alpha)\beta^{s-1}A\left[1+B+B^2+\cdots+B^{k-2}\right]$ yields $\mathcal{O}\left(\Delta t^{\min[p+2,q+2]}\right)$ plus higher order terms. □

For RadauIIA quadrature formulas, bound (33) reads $c\,\Delta t^{\min[2s,s+2,\widehat{p}+k+1]}$. The impact of approximating the solution at the collocation nodes of the quadrature formula can be seen, for instance, in (35), where a lower order might be attained as a consequence of the intermediate stage approximations. In particular the DC–S scheme (16) does not necessarily converge to the IRK scheme. On the other hand, (36) accounts for the iterative corrections performed on the initial low order approximation. Notice that in practice a limited number of iterations k will be performed, certainly less than p, and therefore the scheme would behave like $\mathcal{O}\left(\Delta t^{\min[2s,s+2,\widehat{p}+k+1]}\right)$ even for a finite $C_1 \geq 1$ in Theorem 2. The following Corollary gives us some further insight into the behavior of the DC–S scheme.

Corollary 2. *There is a maximum time step $\Delta t_{\max}$ such that for a given $\Delta t < \Delta t_{\max}$, there is a positive and bounded $C(k)$ such that for $k = 1,2,\ldots,\min[p-\widehat{p}, q-\widehat{p}+1]$, the following holds:*

$$\left\|\boldsymbol{u}(t_0+\Delta t) - \widetilde{\boldsymbol{u}}_s^k\right\| \leq C(k)\left\|\widetilde{\boldsymbol{u}}_s^k - \widetilde{\boldsymbol{u}}_s^0\right\|. \tag{37}$$

Proof. For successive iterations k and $k-1$ such that $k \leq \min[p-\widehat{p}, q-\widehat{p}+1]$, we note that (34) reduces to (36) and hence there is a positive constant ζ_k such that

$$\left\|\boldsymbol{u}(t_0+\Delta t)-\widetilde{\boldsymbol{u}}_s^k\right\| = \zeta_k \Delta t \left\|\boldsymbol{u}(t_0+\Delta t)-\widetilde{\boldsymbol{u}}_s^{k-1}\right\|, \tag{38}$$

and hence

$$\left\|\boldsymbol{u}(t_0+\Delta t)-\widetilde{\boldsymbol{u}}_s^k\right\| = \sigma_k (\Delta t)^k \left\|\boldsymbol{u}(t_0+\Delta t)-\widetilde{\boldsymbol{u}}_s^0\right\| \tag{39}$$

where $\sigma_k = \Pi_{j=1}^k \zeta_j$. Combining (39) with

$$\left\|\boldsymbol{u}(t_0+\Delta t)-\widetilde{\boldsymbol{u}}_s^0\right\| \leq \left\|\boldsymbol{u}(t_0+\Delta t)-\widetilde{\boldsymbol{u}}_s^k\right\| + \left\|\widetilde{\boldsymbol{u}}_s^k-\widetilde{\boldsymbol{u}}_s^0\right\|,$$

yields (37) with

$$C(k) = \frac{\sigma_k (\Delta t)^k}{1-\sigma_k (\Delta t)^k} \tag{40}$$

and $\Delta t_{\max} = (\max_k \zeta_k)^{-1}$ such that $C(k) > 0$ for any given $\Delta t < \Delta t_{\max}$. □

Notice that (38) explicitly reflects the increase of the approximation order with every correction iteration established in Theorem 2. A maximum time step per iteration $\Delta t_{\max,k}$ can be also defined as $\Delta t_{\max,k} = (\zeta_k)^{-1}$, which in particular implies that no correction is expected for $\Delta t \geq \Delta t_{\max,k}$ according to (38). The maximum time step is thus given by $\Delta t_{\max} = \min_k \Delta t_{\max,k}$ in Corollary 2. In practice we note that $\Delta t_{\max,k}$ can be larger than $\Delta t_{\max}$ during a given iteration.

5 Time stepping and error control

Since the numerical integration within the DC–S method is in practice performed by a splitting solver with no stability constraints, we introduce a time step selection strategy based on a user–defined accuracy tolerance. Denoting $err_k := \|\boldsymbol{u}(t_0+\Delta t)-\widetilde{\boldsymbol{u}}_s^k\|$, as the approximation error of the DC–S scheme (16) after a time step Δt, Corollary 2 gives us an estimate of err_k, where in particular err_0 stands for the initial splitting error. In practice we can numerically estimate the error by computing (37) with (40), for which we first need to estimate the σ_k coefficients.

Let us introduce the following approximation to $\boldsymbol{u}(t_0+\Delta t)$ according to (11), based on the original s–stage Runge–Kutta scheme:

$$\bar{\boldsymbol{u}}_s^k = \boldsymbol{u}_0 + I_{t_0}^{t_s}(\widetilde{\boldsymbol{U}}^k). \tag{41}$$

Notice that $\bar{\boldsymbol{u}}_s^k$ is in general different from $\widehat{\boldsymbol{u}}_s^k$ computed according to (13). In particular, for sufficiently small Δt, $\bar{\boldsymbol{u}}_s^0$ should be a better approximation to $\boldsymbol{u}(t_0+\Delta t)$ than the splitting solution $\widetilde{\boldsymbol{u}}_s^0$ given by (15). Therefore, we assume that

$$\left\|\boldsymbol{u}(t_0+\Delta t)-\widetilde{\boldsymbol{u}}_s^0\right\| \leq \left\|\boldsymbol{u}(t_0+\Delta t)-\bar{\boldsymbol{u}}_s^0\right\| + \left\|\bar{\boldsymbol{u}}_s^0-\widetilde{\boldsymbol{u}}_s^0\right\| \approx \left\|\bar{\boldsymbol{u}}_s^0-\widetilde{\boldsymbol{u}}_s^0\right\|, \tag{42}$$

and suppose that the same property holds for the corrective iterations, taking also into account that the fully coupled system $\boldsymbol{F}(\widetilde{\boldsymbol{u}}^k(t))$ is evaluated in (41) for $\bar{\boldsymbol{u}}_s^k$. Introducing $\overline{err}_k = \|\bar{\boldsymbol{u}}_s^k-\widetilde{\boldsymbol{u}}_s^k\|$, we approximate ζ_k in (38) by

$$\widetilde{\zeta}_k = \frac{1}{\Delta t}\frac{\overline{err}_k}{\overline{err}_{k-1}}, \qquad k=1,2,\ldots,$$

and σ_k by $\widetilde{\sigma}_k = \Pi_{j=1}^k \widetilde{\zeta}_j$. We then estimate the error of the DC–S method, $\widetilde{err}_k$, according to (37):

$$\widetilde{err}_k = \left[\frac{\widetilde{\sigma}_k (\Delta t)^k}{1-\widetilde{\sigma}_k (\Delta t)^k}\right] \left\|\widetilde{\boldsymbol{u}}_s^k-\widetilde{\boldsymbol{u}}_s^0\right\|, \qquad k=1,2,\ldots. \tag{43}$$

The initial error, that is, the splitting error err_0 can be directly approximated by $\overline{err}_0$, following (42), and thus $\widetilde{err}_0 = \overline{err}_0$.

Having an estimate of the approximation error (43), we can define an accuracy tolerance η such that for a user–defined η, no further correction iterations are performed if $\widetilde{err}_k \leq \eta$ is satisfied. Therefore, by supposing that

$$\eta = \left[\frac{\widetilde{\sigma}_k(\Delta t_{\text{new},k})^k}{1-\widetilde{\sigma}_k(\Delta t_{\text{new},k})^k}\right]\left\|\widetilde{\boldsymbol{u}}_s^k - \widetilde{\boldsymbol{u}}_s^0\right\|$$

and comparing it to (43), we have that

$$\Delta t_{\text{new},k} = \left[\frac{\eta}{[1-\widetilde{\sigma}_k(\Delta t)^k]\,\widetilde{err}_k + \widetilde{\sigma}_k(\Delta t)^k\eta}\right]^{1/k}\Delta t. \tag{44}$$

We can thus estimate the new time step as $\Delta t_{\text{new}} = \nu \times \min(\Delta t_{\text{new},k}, \Delta t_{\text{max},k})$, after k iterations, where $\Delta t_{\text{max},k} = (\widetilde{\zeta}_k)^{-1}$ and ν $(0 < \nu \leq 1)$ is a security factor. Note that this new time step supposes that k correction iterations will also be used for the next time step. By defining a maximum number of iterations $k_{\max}$, another way of estimating the new time step supposes that

$$\eta = \left[\frac{\widetilde{\sigma}_{k_{\max}}(\Delta t_{\text{new},k_{\max}})^{k_{\max}}}{1-\widetilde{\sigma}_{k_{\max}}(\Delta t_{\text{new},k_{\max}})^{k_{\max}}}\right]\left\|\widetilde{\boldsymbol{u}}_s^k - \widetilde{\boldsymbol{u}}_s^0\right\|, \tag{45}$$

and hence,

$$\Delta t_{\text{new},k_{\max}} = \left[\frac{\eta}{[1-\widetilde{\sigma}_k(\Delta t)^k]\,\widetilde{err}_k + \widetilde{\sigma}_k(\Delta t)^k\eta}\right]^{1/k_{\max}}\left[\frac{\widetilde{\sigma}_k}{\widetilde{\sigma}_{k_{\max}}}\right]^{1/k_{\max}}\Delta t^{k/k_{\max}}, \tag{46}$$

where $\widetilde{\sigma}_{k_{\max}}$ is approximated by $(\widetilde{\zeta}_k)^{k_{\max}-k}\widetilde{\sigma}_k$. In this case the new time step can be computed as $\Delta t_{\text{new}} = \nu \times \min(\Delta t_{\text{new},k_{\max}}, \Delta t_{\text{max},k})$, with the maximum number of iterations given, for instance, by $k_{\max} = \min[p-\widehat{p}, q-\widehat{p}+1]$. Notice that the time step $\Delta t_{\text{new},k_{\max}}$ is potentially larger than $\Delta t_{\text{new},k}$ for $k < k_{\max}$, since the former needs more iterations to attain an accuracy of η according to (45). The best choice between $\Delta t_{\text{new},k}$ and $\Delta t_{\text{new},k_{\max}}$ depends largely on the problem and, in particular, on the corresponding computational costs associated, for instance, with the function evaluations or with the splitting procedure. In general both approaches look for approximations of accuracy η, where smaller time steps involve fewer correction iterations and *vice–versa*. We could also consider, for instance, $\Delta t_{\text{new}} = \nu \times \min(\Delta t_{\text{new},k}, \Delta t_{\text{new},k_{\max}}, \Delta t_{\text{max},k})$.

Recalling that the splitting error can be estimated as $\widetilde{err}_0 = \overline{err}_0 = \|\bar{\boldsymbol{u}}_s^0 - \widetilde{\boldsymbol{u}}_s^0\|$, we can also define a new way to adapt the splitting time step for stiff problems, as previously considered in [14]. Given the splitting approximation (15), one needs only to compute $\bar{\boldsymbol{u}}_s^0$, according to (41). Defining η_{split} as the accuracy tolerance for the splitting approximation, we can compute the splitting time step $\Delta t_{\text{new,split}}$ as

$$\Delta t_{\text{new,split}} = \nu\left[\frac{\eta_{\text{split}}}{\widetilde{err}_0}\right]^{1/(\widehat{p}+1)}\Delta t, \tag{47}$$

assuming from (21) that $\widetilde{err}_0 = C_4\Delta t^{\widehat{p}+1}$ and $\eta = C_4\Delta t_{\text{new,split}}^{\widehat{p}+1}$, where $\widehat{p}$ stands for the order of the splitting scheme. The same technique may be applied to any other low order time integration method. In particular by setting $\eta_{\text{split}} = \eta$, we can estimate the splitting time step $\Delta t_{\text{new,split}}$ that would be required to attain accuracy η through (47), in contrast to $\Delta t_{\text{new},k}$ or $\Delta t_{\text{new},k_{\max}}$.

5.1 Predicting approximation errors

If $k_{\max}$ correction iterations have been performed and the error estimate is still too large ($\widetilde{err}_{k_{\max}} > \eta$), then the DC–S scheme defined in (16), together with the initial splitting approximation (15),

must be restarted with a new time step Δt_{new}, computed as previously defined. In this case, these $k_{\max}$ iterations are wasteful, and hence we derive a method for predicting $\widetilde{err}_{k_{\max}}$ based on the current error estimate $\widetilde{err}_k$. Denoting the predicted final error estimate as $\widetilde{err}^*_{k_{\max}}$, we approximate it according to

$$\widetilde{err}^*_{k_{\max}} = (\widetilde{\zeta}_k \Delta t)^{k_{\max}-k} \widetilde{err}_k, \qquad k = 1, \ldots, k_{\max} - 1,$$

following (38). If for some $k < k_{\max} - 1$, we have that $\widetilde{err}^*_{k_{\max}} > \eta$, we restart the scheme (15)–(16) with a new time step given by

$$\Delta t^*_{\text{new},k_{\max}} = \nu \left[\frac{\eta}{(\widetilde{\zeta}_k)^{k_{\max}-k} \widetilde{err}_k} \right]^{1/k_{\max}} \Delta t^{k/k_{\max}},$$

supposing that $\widetilde{err}_{0,\text{new}} = \widetilde{err}_0$ in

$$\eta = \widetilde{\sigma}_{k_{\max}} (\Delta t^*_{\text{new},k_{\max}})^{k_{\max}} \left[\frac{\widetilde{err}_k}{\widetilde{\sigma}_k (\Delta t)^k} \right],$$

where $\widetilde{\sigma}_{k_{\max}} = (\widetilde{\zeta}_k)^{k_{\max}-k} \widetilde{\sigma}_k$. For the initial approximation $k = 0$, we predict $\widetilde{\sigma}_{k_{\max}}$ based on the previous $\widetilde{\sigma}_{k_{\max},\text{old}}$ and time step Δt_{old} according to

$$\widetilde{err}^*_{k_{\max}} = \widetilde{\sigma}^*_{k_{\max}} (\Delta t)^{k_{\max}} \widetilde{err}_0, \qquad \widetilde{\sigma}^*_{k_{\max}} = \widetilde{\sigma}_{k_{\max},\text{old}} \left[\frac{\Delta t}{\Delta t_{\text{old}}} \right]^{k_{\max}};$$

with the corresponding time step

$$\Delta t^*_{\text{new},k_{\max}} = \nu \left[\frac{\eta}{\widetilde{\sigma}^*_{k_{\max}} \widetilde{err}_0} \right]^{1/k_{\max}},$$

supposing again that $\widetilde{err}_{0,\text{new}} = \widetilde{err}_0$.

6 Space discretization errors

So far, only temporal discretization errors were investigated. For the sake of completeness we briefly study in the following the impact of spatial discretization errors in the approximations computed with the DC–S method. Contrary to Section 4, only an empirical approach is considered. Denoting by $u(t_0 + \Delta t, \boldsymbol{x})|_{\mathbf{X}}$, the analytic solution of problem (1) projected on the grid $\mathbf{X}$ at time $t_0 + \Delta t$, and by $\boldsymbol{u}(t_s)$, the exact solution of the semi–discrete problem (3) at time $t_s = t_0 + \Delta t$; we can assume that for the approximation $\widetilde{\boldsymbol{u}}^k_s$ coming from (16), the following bound is satisfied

$$\left\| u(t_0 + \Delta t, \boldsymbol{x})|_{\mathbf{X}} - \widetilde{\boldsymbol{u}}^k_s \right\| \leq \| u(t_0 + \Delta t, \boldsymbol{x})|_{\mathbf{X}} - \boldsymbol{u}(t_s) \| + \left\| \boldsymbol{u}(t_s) - \widetilde{\boldsymbol{u}}^k_s \right\|. \tag{48}$$

That is, the approximation error of the DC–S method is bounded by the sum of space discretization errors, $\| u(t_0 + \Delta t, \boldsymbol{x})|_{\mathbf{X}} - \boldsymbol{u}(t_s) \|$, and time discretization errors, $\| \boldsymbol{u}(t_s) - \widetilde{\boldsymbol{u}}^k_s \|$. As previously established in Theorem 2, the local order of approximation of the latter is given by $\min[p+1, q+2, \widehat{p}+k+1]$. Therefore, given a semi–discrete problem (3) with solution $\boldsymbol{u}(t)$, high order approximations can be computed by means of (16) within an accuracy tolerance of η. In particular, for sufficiently fine grids and/or if high order space discretization schemes are used, space discretization errors may become small enough and the approximation error of the DC–S method would be estimated as η, also with respect to the analytic solution of problem (1) according to (48).

6.1 Introducing high order space discretization

Spatial resolution is a critical aspect for many problems, and high order space discretization schemes are often required. We consider high order schemes to discretize $F(u)$ in space for problem (1), and denote such an approximation as $\boldsymbol{F}^{\mathrm{HO}}(\boldsymbol{u})$, so that $\boldsymbol{u}^{\mathrm{HO}}(t)$ stands for the solution of the corresponding semi–discrete problem. We can thus use scheme (16) to approximate $\boldsymbol{u}^{\mathrm{HO}}(t)$ by considering $\boldsymbol{F}^{\mathrm{HO}}(\boldsymbol{u})$ in the computation of the $\widehat{\boldsymbol{u}}_i^k$'s in (13), and $\boldsymbol{F}_i^{\mathrm{HO}}(\boldsymbol{u})$, $i = 1, 2, \ldots$, for the splitting approximations in (16) and (15). Let us call this approximation $(\widetilde{\boldsymbol{u}}_s^k)^{\mathrm{HO}}$. Recalling that the numerical error $\|\boldsymbol{u}^{\mathrm{HO}}(t_s) - (\widetilde{\boldsymbol{u}}_s^k)^{\mathrm{HO}}\|$ can be controlled by an accuracy tolerance η, a high order space discretization also results in better error control with respect to the analytic solution $u(t, \boldsymbol{x})$, following the discussion in the previous section.

A more efficient alternative, however, uses $\boldsymbol{F}^{\mathrm{HO}}(\boldsymbol{u})$ only in the quadrature formulas for the $\widehat{\boldsymbol{u}}_i^k$'s in (13), and low order discrete $\boldsymbol{F}_i(\boldsymbol{u})$'s, $i = 1, 2, \ldots$ for the splitting approximations in (16) and (15). The latter procedure results in an approximation $(\widetilde{\boldsymbol{u}}_s^k)^{\widetilde{\mathrm{HO}}}$ to $(\widetilde{\boldsymbol{u}}_s^k)^{\mathrm{HO}}$. Since low order discretization in space is used to evaluate the corrections, that is, to predict the solutions at the collocation nodes of the quadrature formulas, we expect that $(\widetilde{\boldsymbol{u}}_s^k)^{\widetilde{\mathrm{HO}}}$ becomes equivalent to $(\widetilde{\boldsymbol{u}}_s^k)^{\mathrm{HO}}$ after some iterations; if this does happen, then $(\widetilde{\boldsymbol{u}}_s^k)^{\widetilde{\mathrm{HO}}}$ becomes an approximation of maximum order $p = 2s - 1$ in time, according to Theorem 2, and of high order in space to the analytic solution $u(t, \boldsymbol{x})$ of problem (1). Nevertheless, taking into account the hybrid structure of $(\widetilde{\boldsymbol{u}}_s^k)^{\widetilde{\mathrm{HO}}}$ in terms of space discretization errors, a simple decomposition of errors like (48) is no longer possible. Moreover, we cannot expect that each correction iteration increases the order of approximation by one and consequently, the error estimates previously established in §5 are no longer valid for this configuration.

7 Numerical illustrations: The Belousov–Zhabotinski reaction

Let us consider the numerical approximation of a model for the Belousov–Zhabotinski (BZ) reaction, a catalyzed oxidation of an organic species by acid bromated ion (see [24] for more details and illustrations). The present mathematical formulation [25, 28] takes into account three species: hypobromous acid $HBrO_2$, bromide ions Br^-, and cerium (IV). Denoting by $a = [\mathrm{Ce(IV)}]$, $b = [HBrO_2]$, and $c = [Br^-]$, we obtain a very stiff system of three PDEs given by

$$\left.\begin{aligned} \partial_t a - D_a\, \partial^2_{\boldsymbol{x}} a &= \frac{1}{\mu}(-qa - ab + fc), \\ \partial_t b - D_b\, \partial^2_{\boldsymbol{x}} b &= \frac{1}{\varepsilon}\left(qa - ab + b(1-b)\right), \\ \partial_t c - D_c\, \partial^2_{\boldsymbol{x}} c &= b - c, \end{aligned}\right\} \tag{49}$$

where $\boldsymbol{x} \in \mathbb{R}^d$, with real, positive parameters: f, small q, and small ε and μ, such that $\mu \ll \varepsilon \ll 1$. In this study: $\varepsilon = 10^{-2}$, $\mu = 10^{-5}$, $f = 1.6$, $q = 2 \times 10^{-3}$; with diffusion coefficients: $D_a = 2.5 \times 10^{-3}$, $D_b = 2.5 \times 10^{-3}$, and $D_c = 1.5 \times 10^{-3}$. The dynamical system associated with this problem models reactive, excitable media with a large time scale spectrum (see [28] for more details). The spatial configuration with the addition of diffusion develops propagating wavefronts with steep spatial gradients.

7.1 Numerical analysis of errors

We consider problem (49) in a 1D configuration, discretized on a uniform grid of 1001 points over a space region of $[0, 1]$. A standard, second order, centered finite differences scheme is employed for the

diffusion term. To obtain an initial condition, we initialize the problem with a discontinuous profile close to the left boundary; we then integrate until the BZ wavefronts are fully developed. Figure 1 shows the time evolution of the propagating waves for a time window of $[0, 1]$. For comparison, we can solve the semi–discrete problem associated with (49) using a dedicated solver for stiff ODEs. We consider the Radau5 solver [32], based on a fifth order, implicit Runge–Kutta scheme built upon the RadauIIA coefficients in (9). The reference solution for the semi–discrete problem is thus given by the solution obtained with Radau5 over the time interval $[0, 1]$, computed with a fine tolerance: $\eta_{\text{Radau5}} = 10^{-14}$.

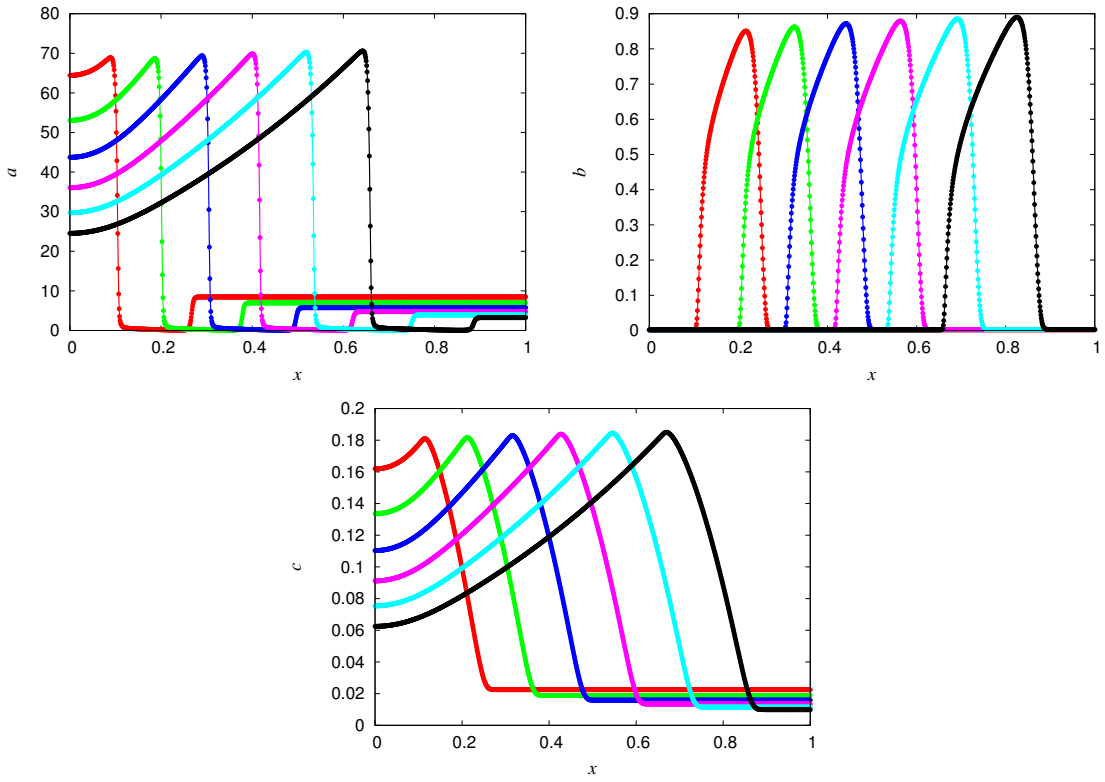

Figure 1: BZ propagating waves for variables a (top left), b (top right), and c (bottom), at time intervals of 0.2 within $[0, 1]$ from left to right.

We consider the splitting solver introduced in [18] for the initialization of the iterative algorithm (15) and the corrective stages in (16) for the DC–S scheme. Radau5 is used to integrate point by point the chemical terms in (49), while the diffusion problem is integrated with a fourth order, stabilized explicit Runge–Kutta method: ROCK4 [1]. (See [17, 19] and discussions therein for a more complex application considering the same splitting solver.) Tolerances of these solvers are set to a reasonable value of $\eta_{\text{Radau5}} = \eta_{\text{ROCK4}} = 10^{-5}$, taking into account that the splitting solver only provides approximations to the solutions at the collocation nodes used in the high order quadrature formulas. Again, the RadauIIA coefficients of order 5 given in (9), that is, with $s = 3$ nodes, are considered for the quadrature formulas in (13). In what follows we consider both Lie and Strang splitting schemes, (4) and (5), ending with the time integration of the reaction terms. The latter is particularly relevant for relatively large splitting time steps [15, 13].

Figure 2 illustrates local and global L^2–errors for various time steps Δt with respect to the reference solution, using the Lie scheme as the splitting solver into the DC–S scheme. The Lie

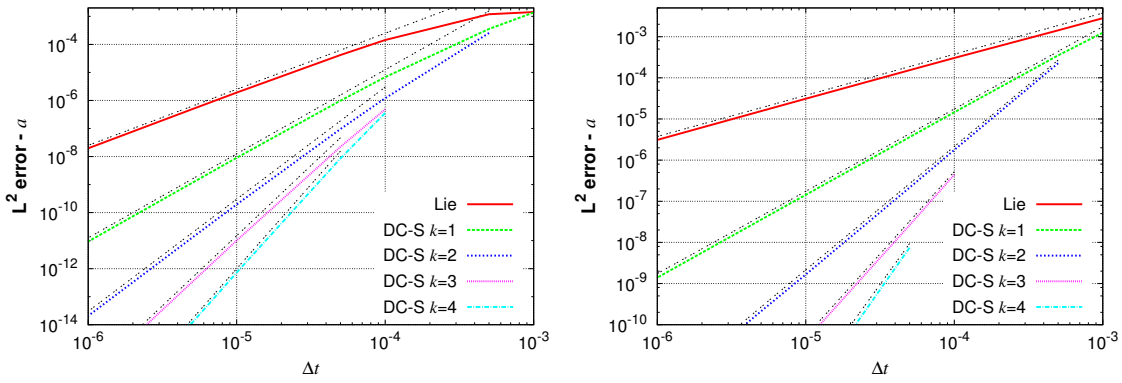

Figure 2: Local (left) and global (right) L^2–errors for the DC–S scheme with Lie splitting. Dashed lines of slopes 2 to 6 (left), and 1 to 5 (right) are also depicted.

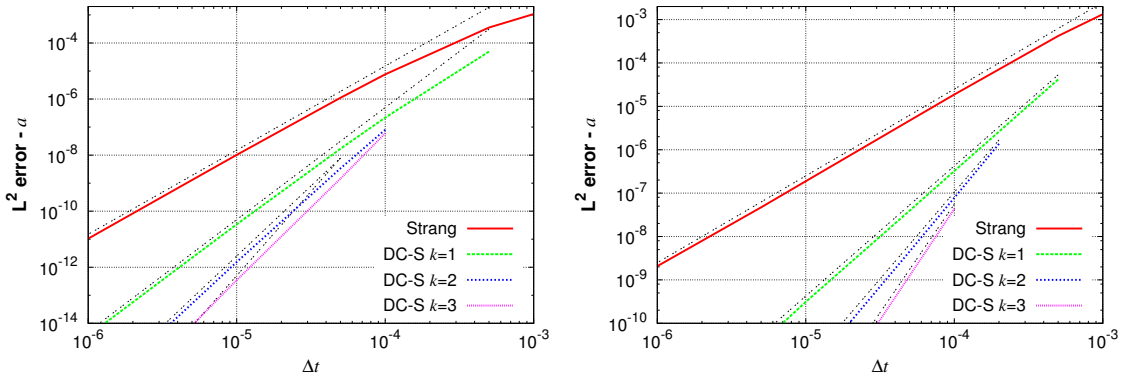

Figure 3: Local (left) and global (right) L^2–errors for the DC–S scheme with Strang splitting. Dashed lines of slopes 3 to 6 (left), and 2 to 5 (right) are also depicted.

solution in Figure 2 corresponds to the splitting approximation with splitting time step $\Delta t_{\text{split}} = \Delta t$. Local errors are evaluated at $t_0 + \Delta t$, starting in all cases from the reference solution at $t_0 = 0.5$. Global errors are evaluated at final time, $t = 1$, after integrating in time with a constant time step Δt. All L^2–errors are scaled by the maximum norm of variable $a(t, x)$ at the corresponding time. For this particular problem the highest order $p = 2s - 1 = 5$ is attained and local errors behave like $\mathcal{O}(\Delta t^{\min[6,2+k]})$ according to Theorem 2, with $\widehat{p} = 1$ for the Lie solver. Notice that very fine tolerances do not need to be considered for the solvers used within the splitting method. For relatively large time steps we observe the well–known loss of order related to splitting schemes on stiff PDEs, as investigated in [13], which propagates throughout the DC–S iterations. Nevertheless, for this particular problem, this loss of order is substantially compensated during the time integration, as observed in the global errors at the final integration time. Same remarks can be made for the DC–S scheme with Strang splitting and $\widehat{p} = 2$, as seen in Figure 3. However, the iterative scheme attains a local order between $q + 2 = 5$ and $p + 1 = 6$, even though the global errors behave like $\mathcal{O}(\Delta t^{\min[5,2+k]})$.

Using the time–stepping procedure established in §5, Figure 4 shows the estimated values of the local error $\widetilde{err}_k$, according to (43), compared to the actual local errors illustrated in Figures

2 and 3. Recall that $\widetilde{err}_0$ stands for the splitting error. The estimated maximum time steps at each corrective iteration $\Delta t_{\max,k} = (\widetilde{\zeta}_k)^{-1}$, are also indicated. They represent the maximum time step at which corrections can be iteratively introduced by the DC–S scheme, which corresponds to the intersection of local errors at different iterations in Figures 2 and 3. Notice that the loss of order was not considered in §5 to derive these estimates; however, the computations can be safely performed because $\widetilde{err}_k$ overestimates the real local error for relatively large time steps, while the $\Delta t_{\max,k}$'s are also smaller than the real ones. The bound (37) in Corollary 2, on which $\widetilde{err}_k$ is based, is nevertheless formally guaranteed up to $k = \min[p - \widehat{p}, q - \widehat{p} + 1]$, that is, $k = 3$ and $k = 2$ for the Lie and Strang splitting, respectively.

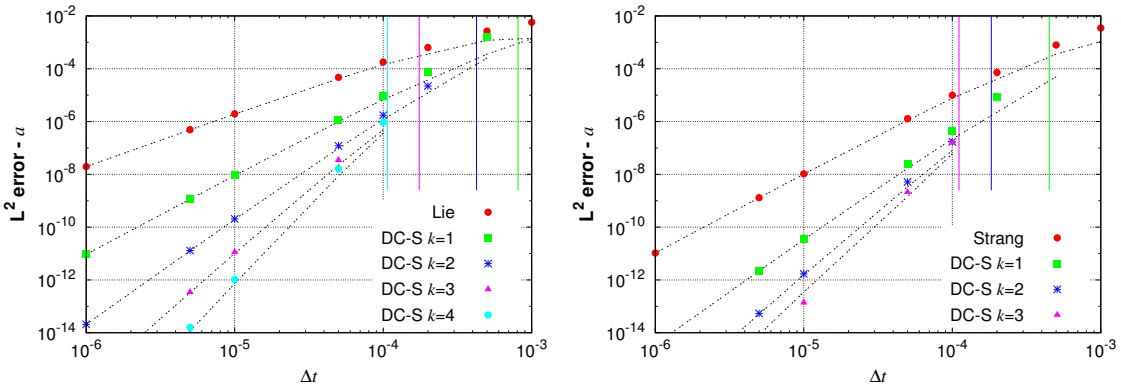

Figure 4: Local error estimates $\widetilde{err}_k$ for different time steps Δt for the DC–S scheme with Lie (left) and Strang (right) splitting. Solid vertical lines stand for the maximum time steps $\Delta t_{\max,k}$. Dashed lines correspond to local L^2–errors from Figures 2 and 3.

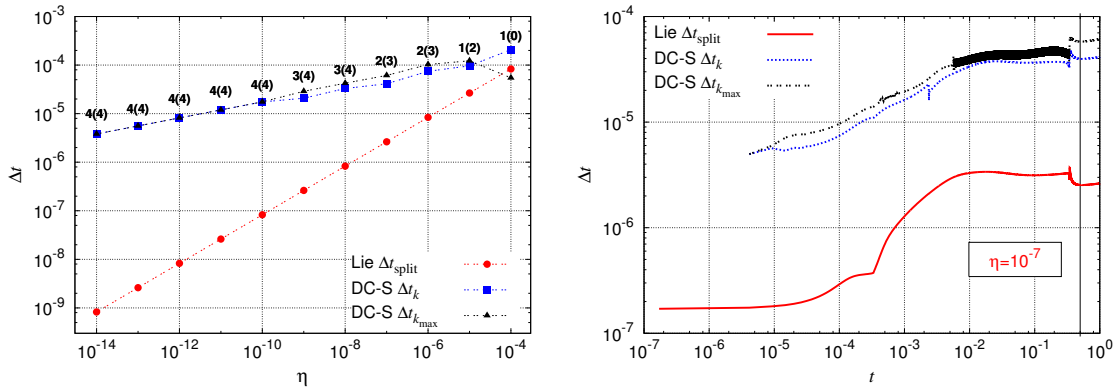

Figure 5: Left: time steps Δt at $t = 0.5$ for different accuracy tolerances η for the Lie and DC–S (Lie) scheme using $\Delta t_{\mathrm{new},k}$ (44) and $\Delta t_{\mathrm{new},k_{\max}}$ (46) for time–stepping. Number of iterations performed are indicated, the case $\Delta t_{\mathrm{new},k_{\max}}$ in parenthesis. Right: time–stepping for $\eta = 10^{-7}$; the solid vertical line corresponds to time $t = 0.5$.

The advantages of considering a high order approximation in time can be inferred from Figure 5, especially when fine accuracies are required. For instance, to attain an accuracy of $\eta = 10^{-12}$ the DC–S scheme based on Lie splitting needs a time step of $\Delta t \approx 10^{-5}$, whereas considering the Lie approximation would require a splitting time step 1000 times smaller. In particular small variations

of Δt in the DC–S scheme can greatly improve the accuracy of the time integration at roughly the same computational cost. Moreover, in order to guarantee a splitting solution of accuracy η, the numerical solution of the split subproblems must also be performed with the same level of accuracy [18]. For the splitting solver considered here, the latter involves $\eta_{\text{Radau5}} \leq \eta$ and $\eta_{\text{ROCK4}} \leq \eta$, whereas this is not required for the DC–S scheme. As previously noted, using $\Delta t_{\text{new},k_{\max}}$ (46) for the time–stepping involves slightly larger time steps because more iterations are performed. The dynamical time–stepping is illustrated in Figure 5 (right), using $\Delta t_{\text{new},k}$ (44) and $\Delta t_{\text{new},k_{\max}}$ (46) for the DC–S scheme, and $\Delta t_{\text{new,split}}$ (47) to adapt the splitting time step for the Lie solver. For this particular problem, a roughly constant time step is attained, after some initial transients, consistent with the quasi self–similar propagation of the wavefronts, as depicted in Figure 1. In all cases the time–stepping procedure guarantees numerical approximations within a user–defined accuracy.

7.2 Space discretization errors

We now investigate the spatial discretization errors in the previous approximations. We consider both a second and a fourth order centered finite differences scheme for the Laplacian operator, as well as various resolutions given by 501, 1001, 2001, 4001, 8001, and 16001 grid points. The DC–S solution (with Lie splitting and $k = 4$) computed with the finest space resolution of 16001 points is taken as the reference solution. Figure 6 (left) shows the numerical errors for different discretizations with respect to the reference solution. All approximations are initialized from the same solution represented on the 16001–points reference grid; then, an integration time step is computed on all grids with the same DC–S integration scheme and a time step of 10^{-6}. The goal is to illustrate only the space discretization errors, as seen in Figure 6 (left).

Considering the same results obtained with the Lie and the corresponding DC–S scheme on a 1001–points grid, previously shown in Figure 2, we compute the numerical errors with respect to the reference solution on 16001 points. Figure 6 (right) illustrates the errors arising from spatial and temporal numerical errors after one time step. According to (48) we see that for sufficiently large time steps the temporal integration error is mainly responsible for the total approximation error, whereas the accuracy of the method is limited by the space discretization errors for small time steps. Considering a high order space discretization scheme increases the region for which time integration errors control the accuracy of the approximation. For a relatively coarse grid of 1001 grid points, it can be seen that an order of magnitude can be already gained in terms of accuracy by increasing the order of the space discretization scheme, as seen in Figure 6. Greater improvements are expected for finer spatial resolutions as inferred from Figure 6 (left).

Coming back to a spatial resolution of 1001 points, Figure 7 (left) displays numerical errors related to the time integration, as in Figure 2 (left), this time with the fourth order space discretization. As before the reference solution is obtained with Radau5 and $\eta_{\text{Radau5}} = 10^{-14}$, considering now the fourth order space discretization scheme. The same previous conclusions apply as for the results in Figure 2, now with high order approximations in both time and space. Similar behaviors are observed for the Strang case. Nevertheless, a high order scheme in space can be more time and memory consuming, since it usually requires larger stencils and thus more computational work. One alternative, previously discussed in §6.1, uses high order space discretizations for the quadrature formulas, which in practice involve matrix–vector multiplications; while the time integrators that generate the iterative corrections are performed with low order discretization schemes in both time and space. In Figure 7 (right) we evaluate the temporal errors for this configuration, using the same reference solution as in Figure 7 (left). It can be seen that the space discretization errors are progressively eliminated until we obtain the same results observed in Figure 7 (left). That is, the hybrid implementation with second/fourth order in space converges to the high order quadrature formula evaluated with a high order space discretization. However, the time stepping strategy with error control established in §5 is no longer valid since a decomposition of time and space discretization errors such as (48) is no longer possible.

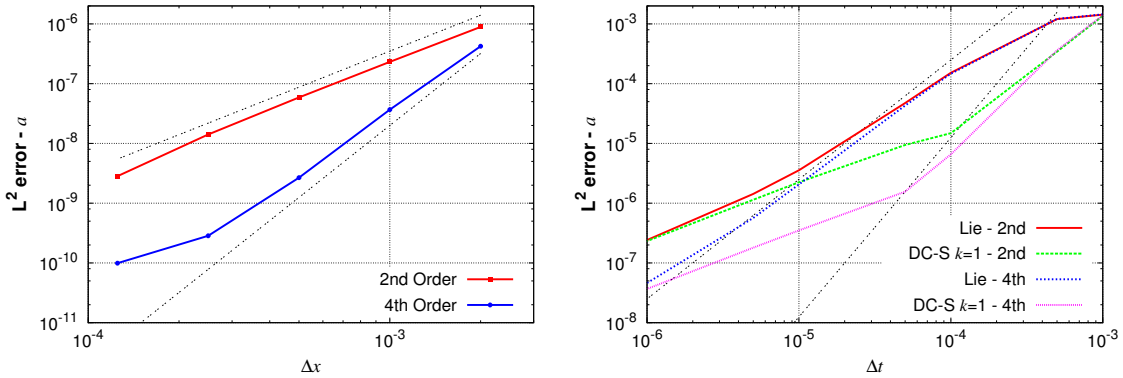

Figure 6: Left: space discretization errors after one time step with the same time integration scheme. Right: local L^2–errors for the DC–S scheme with Lie splitting on a 1001–points grid, accounting for both time and space numerical errors. Dashed lines of slopes 2 and 4 (left), and 2 and 3 (right) are also depicted. Space discretizations of second and fourth order are considered.

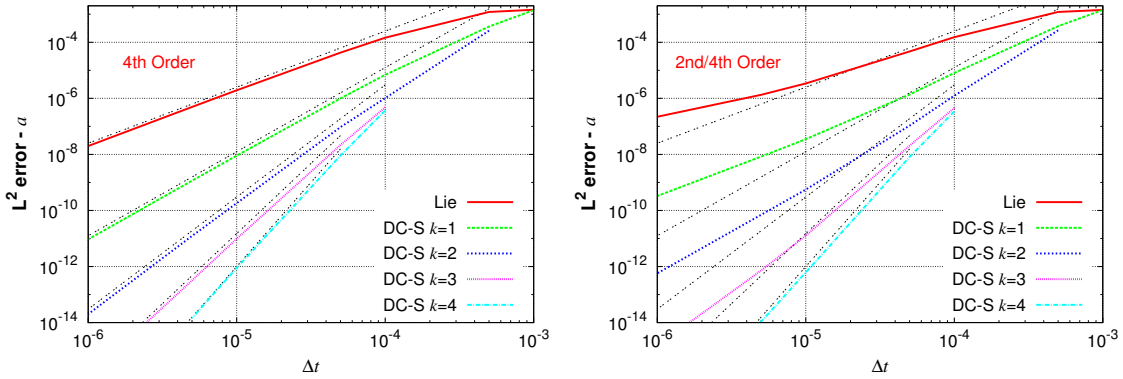

Figure 7: Local L^2–errors for the DC–S scheme with Lie splitting with fourth order (left) and second/fourth order (right) space discretizations. Dashed lines of slopes 3 to 6 are also depicted.

8 Concluding remarks

We have introduced a new numerical scheme that achieves high order approximations in time for the solution of time dependent stiff PDEs. The method exploits the advantages of operator splitting techniques to handle the stiffness associated with different processes modeled by the PDEs, while high order in time is attained through an iterative procedure based on a standard deferred correction technique. In this way the initially low order approximation computed by a tailored splitting solver is iteratively corrected to obtain a high order approximation based on a quadrature formula. The latter is based on Radau collocation nodes according to the RadauIIA, s–stage quadrature formulas. Moreover, the splitting solver uses dedicated methods in terms of numerical stability and independent time–stepping features within every splitting time step to separately advance each subproblem originating from the PDE. Consequently, no time step restriction needs to be observed in the deferred correction splitting technique to guarantee the numerical stability of the numerical integration. The traditionally low algorithmic complexity and efficiency of a splitting approach are therefore preserved, while high order approximations in time are achieved by including

extra function evaluations at the collocation nodes and by performing matrix–vector multiplications according to the quadrature formulas.

A mathematical analysis of the method was also conducted in a general finite dimensional space with standard assumptions on the PDEs and the splitting approximations. In this context it was proved that the local error of the iterative method behaves like $\Delta t^{\min[p+1,q+2,\widehat{p}+k+1]}$, $k = 1, 2, \ldots$, where p, q, and $\widehat{p}$ stand for the global and stage orders of the quadrature formulas and the global order of the splitting approximation, respectively. A maximum time step was also formally identified beyond which the iterative scheme yields no additional correction. Based on these theoretical results, a time–stepping strategy was derived to monitor the approximation errors related to the time integration. Both the definition of the maximum time step as well as the time–stepping procedure remain valid regardless of the integration scheme used to approximate the solution at the collocation nodes. Numerical results confirmed the theoretical findings in terms of time integration errors and the orders attained. In particular, given a stiff PDE discretized on a given grid, the time–stepping technique yields numerical time integrations within a user–defined accuracy tolerance. This error control remains effective for the overall accuracy of the numerical approximations for a sufficiently fine space resolution and/or when high order space discretization schemes are considered.

Finally, a hybrid approach wherein low order spatial discretizations are used to advance the correction equation while fully coupled high order spatial discretizations are used to evaluate the quadrature formula was introduced. The numerical tests performed herein demonstrate that this approach converged, for the particular problem considered, to the same solution obtained using the high order spatial discretization throughout. The advantage to the hybrid approach is that the correction equation is solved using less computationally expensive spatial operators, although the theoretical results concerning adaptive time–stepping no longer apply.

A Relation with the parareal algorithm

Considering problem (3) over the time domain $[t_0, t_n]$, it can be decomposed into n subdomains $[t_{i-1}, t_i[$, $i = 1, \ldots, n$, and $\Delta t_i := t_i - t_{i-1}$. The parareal algorithm [35] is based on two propagation operators: $\mathcal{G}^t \boldsymbol{u}_0$ and $\mathcal{F}^t \boldsymbol{u}_0$, that provide, respectively, a coarse and a more accurate (fine) approximation to the solution of problem (3). The algorithm starts with an initial approximation $(\widetilde{\boldsymbol{u}}_i^0)_{i=1,\ldots,n}$, given by the sequential computation:

$$\widetilde{\boldsymbol{u}}_i^0 = \mathcal{G}^{\Delta t_i} \widetilde{\boldsymbol{u}}_{i-1}^0, \quad i = 1, \ldots, n. \tag{50}$$

It then performs the correction iterations:

$$\widetilde{\boldsymbol{u}}_i^{k+1} = \mathcal{F}^{\Delta t_i} \widetilde{\boldsymbol{u}}_{i-1}^k + \mathcal{G}^{\Delta t_i} \widetilde{\boldsymbol{u}}_{i-1}^{k+1} - \mathcal{G}^{\Delta t_i} \widetilde{\boldsymbol{u}}_{i-1}^k, \quad i = 1, \ldots, n. \tag{51}$$

The coarse approximations, which are computed sequentially, are thus iteratively corrected in order to get a more accurate solution based on the fine solver, performed in parallel.

The coarse propagator can be thus defined as a low order operator splitting technique as considered in [16]. Moreover, if the n time intervals are defined according to the collocation nodes of an IRK quadrature formula, then the fine propagator could be defined following (13). The iterative DC–S scheme (15)–(16) can be thus formally seen as a particular implementation of the parareal algorithm (50)–(51). The parareal algorithm can be actually recast as a deferred correction scheme [27]. The combination of time parallelism and deferred correction techniques was recently investigated in [38, 23]. However, in the DC–S scheme the deferred corrections are performed over a single time step instead of the entire time domain of interest. The latter involves a departure from the time parallel capabilities of the parareal algorithm, as it was originally conceived. Furthermore, the quadrature formula in the DC–S scheme that considers the approximations $(\widetilde{\boldsymbol{u}}_i^k)_{i=1,\ldots,s}$ to compute $\widehat{\boldsymbol{u}}_i^k$ at each node cannot be viewed as the standard fine propagator in the parareal framework which would rather perform a numerical time integration from $\widetilde{\boldsymbol{u}}_{i-1}^k$. The latter is especially relevant when

one studies the method since a common approach in the analysis of the parareal algorithm considers the fine propagator as the semiflow corresponding to the exact solution (see, e.g., [35, 3, 27, 26]).

B Relation with SDC methods

The iterative SDC method introduced in [20] to approximate the solution of problem (3) can be written in general as [37]

$$\widetilde{\boldsymbol{u}}_i^{k+1} = \widetilde{\boldsymbol{u}}_{i-1}^{k+1} + \int_{t_{i-1}}^{t_i} \left[\boldsymbol{F}(\widetilde{\boldsymbol{u}}^{k+1}(\tau)) - \boldsymbol{F}(\widetilde{\boldsymbol{u}}^{k}(\tau))\right] \mathrm{d}\tau + S_{t_{i-1}}^{t_i}(\widetilde{\boldsymbol{U}}^k), \quad i = 1, \ldots, n, \tag{52}$$

where $\Delta t = t_n - t_0$, and

$$S_{t_{i-1}}^{t_i}(\boldsymbol{U}) \approx \int_{t_{i-1}}^{t_i} \boldsymbol{F}(\boldsymbol{u}(\tau))\, \mathrm{d}\tau, \tag{53}$$

is a spectral integration operator defined by means of a quadrature formula that evaluates $\boldsymbol{F}(\boldsymbol{u}(t))$ at the n or $n+1$ (if time t_0 is included) collocation nodes. Depending on the stiffness of problem (1) (or (3)) either an explicit or implicit Euler approximation to the remaining integrals in (52) is considered, that is,

$$\widetilde{\boldsymbol{u}}_i^{k+1} = \widetilde{\boldsymbol{u}}_{i-1}^{k+1} + \Delta t_i \left[\boldsymbol{F}(\widetilde{\boldsymbol{u}}_{i-1}^{k+1}(\tau)) - \boldsymbol{F}(\widetilde{\boldsymbol{u}}_{i-1}^{k}(\tau))\right] + S_{t_{i-1}}^{t_i}(\widetilde{\boldsymbol{U}}^k),$$

or

$$\widetilde{\boldsymbol{u}}_i^{k+1} = \widetilde{\boldsymbol{u}}_{i-1}^{k+1} + \Delta t_i \left[\boldsymbol{F}(\widetilde{\boldsymbol{u}}_{i}^{k+1}(\tau)) - \boldsymbol{F}(\widetilde{\boldsymbol{u}}_{i}^{k}(\tau))\right] + S_{t_{i-1}}^{t_i}(\widetilde{\boldsymbol{U}}^k),$$

or a combination of both depending on the nature of each $\boldsymbol{F}_i(u)$ into $\boldsymbol{F}(u) = \boldsymbol{F}_1(u) + \boldsymbol{F}_2(u) + \ldots$ [37]. The 0–th iteration is computed by a low order time integration scheme. These approximations are thus iteratively corrected to obtain a high order quadrature rule solution: $\widetilde{\boldsymbol{u}}_n^{k+1} \approx \widetilde{\boldsymbol{u}}_{n-1}^{k+1} + S_{t_{n-1}}^{t_n}(\widetilde{\boldsymbol{U}}^k)$ following (52).

Introducing the following approximation into (52),

$$\widetilde{\boldsymbol{u}}_{i-1}^{k} + \int_{t_{i-1}}^{t_i} \boldsymbol{F}(\widetilde{\boldsymbol{u}}^{k}(\tau))\, \mathrm{d}\tau \approx \mathcal{S}^{(c_i - c_{i-1})\Delta t} \widetilde{\boldsymbol{u}}_{i-1}^{k},$$

with $\Delta t_1 = c_1 \Delta t$ and $\Delta t_i = (c_i - c_{i-1})\Delta t$, $i = 2, \ldots, n = s$, one can notice that (16) and (52) become equivalent, as long as the spectral operator $S_{t_{i-1}}^{t_i}(\boldsymbol{U})$ with its corresponding quadrature nodes are defined based on an IRK scheme relying on a collocation method, that is, $S_{t_{i-1}}^{t_i}(\boldsymbol{U}) \equiv I_{t_{i-1}}^{t_i}(\boldsymbol{U})$. The spectral operator (53) was built in [20] based on the Lagrange interpolant that evaluates $\boldsymbol{F}(\boldsymbol{u}(t))$ at the Gauss–Legendre nodes, which is equivalent to considering a Gauss quadrature formula for IRK schemes. The latter was later enhanced in [37] by employing Lobatto quadrature formulas. In this case the spectral operator is equivalent to that based on the LobattoIIIA–IRK scheme. These analogies between spectral operators in SDC schemes and quadrature formulas for IRK methods were investigated, for instance, in [30]. The DC–S scheme (16) can be thus recast as a variation of the SDC method in which numerical time integrations are performed within each time subinterval instead of approximating integrals into (52). The latter, however, involves important changes to the iterative procedure in (52) together with its practical implementation.

References

[1] A. Abdulle. Fourth order Chebyshev methods with recurrence relation. *SIAM J. Sci. Comput.*, 23(6):2041–2054 (electronic), 2002.

[2] O. Axelsson. A note on a class of strongly A–stable methods. *BIT Numer. Math.*, 12:1–4, 1972.

[3] G. Bal. On the convergence and the stability of the parareal algorithm to solve partial differential equations. In *Proceedings of the 15th International Domain Decomposition Conference, Lect. Notes Comput. Sci. Eng. 40*, pages 426–432. Springer, Berlin, 2003.

[4] H. Bijl and M.H. Carpenter. Iterative solution techniques for unsteady flow computations using higher order time integration schemes. *Int. J. Numer. Meth. Fluids*, 47:857–862, 2005.

[5] H. Bijl, M.H. Carpenter, V.N. Vatsa, and C.A. Kennedy. Implicit time integration schemes for the unsteady compressible Navier–Stokes equations: Laminar flow. *J. Comput. Phys.*, 179(1):313–329, 2002.

[6] S. Blanes, F. Casas, P. Chartier, and A. Murua. Optimized high–order splitting methods for some classes of parabolic equations. *Math. Comp.*, 82:1559–1576, 2013.

[7] J.C. Butcher. Implicit Runge–Kutta processes. *Math. Comp.*, 18:50–64, 1964.

[8] J.C. Butcher. Integration processes based on Radau quadrature formulas. *Math. Comp.*, 18:233–244, 1964.

[9] M.H. Carpenter, C.A. Kennedy, H. Bijl, S.A. Viken, and V.N. Vatsa. Fourth–order Runge–Kutta schemes for fluid mechanics applications. *J. Sci. Comput.*, 25:157–194, 2005.

[10] F. Castella, P. Chartier, S. Descombes, and G. Vilmart. Splitting methods with complex times for parabolic equations. *BIT Numer. Math.*, 49:487–508, 2009.

[11] F.H. Chipman. A–stable Runge–Kutta processes. *BIT Numer. Math.*, 11:384–388, 1971.

[12] S. Descombes. Convergence of a splitting method of high order for reaction–diffusion systems. *Math. Comp.*, 70(236):1481–1501, 2001.

[13] S. Descombes, M. Duarte, T. Dumont, F. Laurent, V. Louvet, and M. Massot. Analysis of operator splitting in the nonasymptotic regime for nonlinear reaction–diffusion equations. Application to the dynamics of premixed flames. *SIAM J. Numer. Anal.*, 52(3):1311–1334, 2014.

[14] S. Descombes, M. Duarte, T. Dumont, V. Louvet, and M. Massot. Adaptive time splitting method for multi-scale evolutionary partial differential equations. *Confluentes Math.*, 3(3):413–443, 2011.

[15] S. Descombes and M. Massot. Operator splitting for nonlinear reaction–diffusion systems with an entropic structure: Singular perturbation and order reduction. *Numer. Math.*, 97(4):667–698, 2004.

[16] M. Duarte, S. Descombes, and M. Massot. Parareal operator splitting techniques for multi–scale reaction waves: Numerical analysis and strategies. *ESAIM: Math. Model. Numer. Anal.*, 45:825–852, 2011.

[17] M. Duarte, M. Massot, S. Descombes, C. Tenaud, T. Dumont, V. Louvet, and F. Laurent. New resolution strategy for multi-scale reaction waves using time operator splitting and space adaptive multiresolution: Application to human ischemic stroke. *ESAIM: Proc.*, 34:277–290, 2011.

[18] M. Duarte, M. Massot, S. Descombes, C. Tenaud, T. Dumont, V. Louvet, and F. Laurent. New resolution strategy for multiscale reaction waves using time operator splitting, space adaptive multiresolution and dedicated high order implicit/explicit time integrators. *SIAM J. Sci. Comput.*, 34(1):A76–A104, 2012.

[19] T. Dumont, M. Duarte, S. Descombes, M.-A. Dronne, M. Massot, and V. Louvet. Simulation of human ischemic stroke in realistic 3D geometry. *Commun. Nonlinear Sci. Numer. Simul.*, 18(6):1539–1557, 2013.

[20] A. Dutt, L. Greengard, and V. Rokhlin. Spectral deferred correction methods for ordinary differential equations. *BIT Numer. Math.*, 40(2):241–266, 2000.

[21] B.L. Ehle. High order A–stable methods for the numerical solution of systems of DEs. *BIT Numer. Math.*, 8:276–278, 1968.

[22] B.L. Ehle. On Padé approximations to the exponential function and A–stable methods for the numerical solution of initial value problems. *Research Report CSRR 2010*, 1969.

[23] M. Emmett and M.L. Minion. Toward an efficient parallel in time method for partial differential equations. *Comm. App. Math. and Comp. Sci*, 7(1):105–132, 2012.

[24] I.R. Epstein and J.A. Pojman. *An Introduction to Nonlinear Chemical Dynamics*. Oxford University Press, 1998. Oscillations, Waves, Patterns and Chaos.

[25] R.J. Field, E. Koros, and R.M. Noyes. Oscillations in chemical systems. II. Thorough analysis of temporal oscillation in the bromate–cerium–malonic acid system. *J. Amer. Chem. Soc.*, 94(25):8649–8664, 1972.

[26] M. Gander and E. Hairer. Nonlinear convergence analysis for the parareal algorithm. In *Domain Decomposition Methods in Science and Engineering XVII*, pages 45–56. Springer, Berlin, 2008.

[27] M. Gander and S. Vandewalle. Analysis of the parareal time–parallel time–integration method. *SIAM J. Sci. Comput.*, 29(2):556–578, 2007.

[28] P. Gray and S.K. Scott. *Chemical Oscillations and Instabilites*. Oxford Univ. Press, 1994.

[29] A. Guillon and F.L. Soulé. La résolution numérique des problèmes différentiels aux conditions initiales par des méthodes de collocation. *RAIRO Anal. Numér. Ser. Rouge, v. R–3*, pages 17–44, 1969.

[30] T. Hagstrom and R. Zhou. On the spectral deferred correction of splitting methods for initial value problems. *Comm. App. Math. and Comp. Sci.*, 1(1):169–205, 2006.

[31] E. Hairer, C. Lubich, and G. Wanner. *Geometric Numerical Integration. Structure–Preserving Algorithms for Ordinary Differential Equations*. Springer–Verlag, Berlin, 2nd edition, 2006.

[32] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II. Stiff and Differential–Algebraic Problems*. Springer–Verlag, Berlin, 2nd edition, 1996.

[33] E. Hansen and A. Ostermann. High order splitting methods for analytic semigroups exist. *BIT Numer. Math.*, 49:527–542, 2009.

[34] W. Hundsdorfer and J.G. Verwer. *Numerical Solution of Time–Dependent Advection–Diffusion–Reaction Equations*. Springer–Verlag, Berlin, 2003.

[35] J.L. Lions, Y. Maday, and G. Turinici. Résolution d'EDP par un schéma en temps "pararéel". *C. R. Acad. Sci. Paris Sér. I Math.*, 332(7):661–668, 2001.

[36] G.I. Marchuk. Some application of splitting–up methods to the solution of mathematical physics problems. *Appl. Math.*, 13(2):103–132, 1968.

[37] M.L. Minion. Semi–implicit spectral deferred correction methods for ordinary differential equations. *Comm. Math. Sci.*, 1:471–500, 2003.

[38] M.L. Minion. A hybrid parareal spectral deferred corrections method. *Comm. App. Math. and Comp. Sci*, 5(2):265–301, 2010.

[39] A. Prothero and A. Robinson. On the stability and accuracy of one–step methods for solving stiff systems of ordinary differential equations. *Math. Comp.*, 28(125):145–162, 1974.

[40] R. Radau. Étude sur les formules d'approximation qui servent à calculer la valeur numérique d'une intégrale definie. *J. Math. Pures Appl.*, 6:283–336, 1880.

[41] R.D. Skeel. A theoretical framework for proving accuracy results for deferred corrections. *SIAM J. Numer. Anal.*, 19(1):171–196, 1982.

[42] G. Strang. On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.*, 5:506–517, 1968.

[43] H.F. Trotter. On the product of semi–groups of operators. *Proc. Am. Math. Soc.*, 10:545–551, 1959.

[44] K. Wright. Some relationships between implicit Runge–Kutta, collocation and Lanczos τ methods, and their stability properties. *BIT Numer. Math.*, 10:217–227, 1971.