J Sci Comput manuscript No.
(will be inserted by the editor)

Error Analysis of Explicit Partitioned Runge-Kutta Schemes for Conservation Laws

Willem Hundsdorfer · David I. Ketcheson · Igor Savostianov

Received: date / Accepted: date

Abstract An error analysis is presented for explicit partitioned Runge-Kutta methods and multirate methods applied to conservation laws. The interfaces, across which different methods or time steps are used, lead to order reduction of the schemes. Along with cell-based decompositions, also flux-based decompositions are studied. In the latter case mass conservation is guaranteed, but it will be seen that the accuracy may deteriorate.

Keywords multirate methods · partitioned Runge-Kutta methods · conservation · stability · convergence

Mathematics Subject Classification (2000) 65L06 · 65M06 · 65M20

1 Introduction

Spatial discretization of partial differential equations (PDEs) lead to systems of ordinary differential equations (ODEs), the so-called semi-discrete systems. In this paper we will consider explicit time stepping schemes applied to conservation laws $u_t + \nabla \cdot f(u) = 0$ with a given spatial discretization. The CFL stability condition bounds the time step in terms of the ratio of local (spatial) mesh width and characteristic speeds. If either of these factors varies substantially, it is natural to use local

This work has been supported by Award No. FIC/2010/05 from King Abdullah University of Science and Technology (KAUST).

W. Hundsdorfer
CWI, PO Box 94079, Amsterdam, The Netherlands. E-mail: willem.hundsdorfer@cwi.nl

D.I. Ketcheson
Division Mathematical and Computer Sciences & Engineering, King Abdullah University of Science & Technolog,y (KAUST), P.O. Box 4700, Thuwal 23955, Saudi Arabia. E-mail: david.ketcheson@kaust.edu.sa

I. Savostianov
CWI, PO Box 94079, Amsterdam, The Netherlands. E-mail: igor.savostianov@gmail.com

time steps that match the local convective velocity or spatial mesh width. Schemes in which different time steps are used over different parts of the spatial grid are referred to as multirate schemes. Such schemes can be studied in the more general setting of partitioned or additive Runge-Kutta methods. There exist two natural spatial partitionings for conservation laws: one based on partitioning cells with solution values, and one based on partitioning fluxes. The latter approach ensures mass conservation.

The classical order of a numerical ODE solver is often larger when applied to non-stiff ODEs than when applied to PDEs, where one considers time step Δt and spatial mesh width Δx tending to zero simultaneously. This phenomenon, known as *order reduction*, often arises at the boundaries. In multirate methods, interfaces are created within the problem domain, at the points where the step size changes, and these interfaces act like artificial spatial boundaries. It is natural to ask whether, and to what extent, order reduction may arise at these interfaces.

In this paper we conduct a careful analysis of the local and global errors of partitioned multirate schemes based on partitioning of the cells or the fluxes, including discussion of the relation between local consistency and conservation. Throughout, we assume the solution to be sufficiently smooth. For conservation laws this means that accuracy is studied away from shocks. (For results of multirate schemes near shocks, where monotonicity properties and maximum principles are important, we refer to [6].) The convergence results are illustrated using some existing first- and second-order multirate schemes, applied to semi-discretizations of hyperbolic conservation laws.

The outline of this paper is as follows. Multirate methods are conveniently analyzed in the broader framework of partitioned and additive Runge-Kutta methods, which we review in Section 2. In Section 3, we present some multirate methods of order one and two, along with simple numerical tests showing some of their deficiencies. General expressions for the local errors, that can be used to derive error bounds when both Δt and Δx tend to zero, are given in Section 4. Detailed error bounds are found in Section 5 for cell-based decomposition and in Section 6 for flux-based decomposition. It will be seen that flux-based decompositions often lead to a lower order of convergence. Some conclusions and final remarks are given in Section 7.

2 Partitioned Runge-Kutta methods

Setting and notation

The system of ODEs in $\mathbb{R}^m$, with given initial value, will be written as

$$u'(t) = F(t, u(t)) , \qquad u(0) = u_0 . \tag{2.1}$$

In our applications, this ODE system will be a semi-discrete system obtained from a conservation law by a finite difference or finite volume discretization in space. Each component $u_j(t)$ of the vector $u(t) = [u_j(t)] \in \mathbb{R}^m$ then stands for an approximation at time t to the pointwise or average value of the PDE solution at x_j, $j = 1, 2, \ldots, m$, and F is the spatial discretization operator.

For the time integration of the semi-discrete system we will consider partitioned methods based on a decomposition of F,

$$F(t,v) = F_1(t,v) + F_2(t,v) + \cdots + F_r(t,v)\,, \tag{2.2}$$

where each $F_k : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ corresponds to the spatial discretization operator in a certain region Ω_k of the spatial PDE domain $\Omega = \Omega_1 \cup \cdots \cup \Omega_r$.

Let $\mathcal{I} = \mathcal{I}_1 \cup \cdots \cup \mathcal{I}_r$ be a partitioning of the index set $\mathcal{I} = \{1, 2, \ldots, m\}$, with $j \in \mathcal{I}_k$ if $x_j \in \Omega_k$. To define the schemes we consider corresponding diagonal matrices $I = I_1 + \cdots + I_r$, where I is the identity matrix and the I_k are diagonal with entries zero or one: the j-th diagonal entry of I_k equal to one iff $j \in \mathcal{I}_k$. Then $F_k = I_k F$ defines a *cell-based* decomposition; the function F_k contains those components of F that correspond to the spatial region Ω_k. Another possibility is to base the decomposition on fluxes, to ensure mass conservation; such *flux-based* decompositions will be discussed later in some detail. Our main interest is in methods that use different step sizes in each spatial domain Ω_k.

For a given decomposition (2.2), we consider partitioned Runge-Kutta methods, giving approximations $u_n \approx u(t_n)$ at the time levels $t_n = n\Delta t$, $n \geq 0$. A step from t_n to t_{n+1} with an s-stage method reads

$$v_{n,i} \;=\; u_n + \Delta t \sum_{k=1}^{r} \sum_{j=1}^{s} a_{ij}^{(k)} F_k(t_n + c_j \Delta t, v_{n,j})\,, \qquad i = 1, \ldots, s\,, \tag{2.3a}$$

$$u_{n+1} \;=\; u_n + \Delta t \sum_{k=1}^{r} \sum_{j=1}^{s} b_{j}^{(k)} F_k(t_n + c_j \Delta t, v_{n,j})\,. \tag{2.3b}$$

The internal stage vectors $v_{n,i}$, $i = 1, \ldots, s$, give approximations to $u(t_n + c_i \Delta t)$ at the intermediate time levels. For applications to conservation laws we will restrict ourselves to explicit methods, where $a_{ij}^{(k)} = 0$ if $j \geq i$.

For general decompositions $F = F_1 + \cdots + F_r$, method (2.3) is usually called an additive Runge-Kutta method, and the name partitioned Runge-Kutta method is often reserved for the case where the decomposition has a partitioned structure ($F_k = I_k F$). However, as noted in [1, p. 153], any partitioned method can be written as an additive one (and vice versa) by modifying the right hand side, so we do not distinguish these two classes of methods.

In this section we will briefly discuss some basic properties of the partitioned and additive methods. A more extensive discussion is found in [6].

Internal consistency and conservation

Let $c_i^{(k)} = \sum_{j=1}^{s} a_{ij}^{(k)}$, $i = 1, \ldots, s$. If we have

$$c_i^{(k)} = c_i^{(l)} \quad \text{for all } 1 \leq k, l \leq r \text{ and } 1 \leq i \leq s\,, \tag{2.4}$$

then the internal vectors $v_{n,i}$ are consistent approximations to $u(t_n + c_i \Delta t)$, and the method is *internally consistent*. As will be seen, this is an important property for the accuracy of the method when applied to ODEs obtained by semi-discretization.

If (2.4) holds, this gives an obvious choice for the abscissae c_i in (2.3). If (2.4) is not satisfied, then we take $c_i = c_i^{(r)}$, $1 \le i \le s$, where it is assumed that the r-th Runge-Kutta method used in (2.3) is the most 'refined' one. This is a natural choice for the multirate methods that will be considered; it may be somewhat arbitrary for general partitioned methods.

Apart from consistency, we will also study *conservation* of linear invariants; for example, mass conservation. Suppose that $h^T = [h_1, \ldots, h_m]$ is such that $h^T u(t) = \sum_j h_j u_j(t)$ is a conserved quantity for the ODE system (2.1). This will hold for an arbitrary initial value u_0 provided that

$$h^T F(t, v) = 0 \qquad \text{for all } t \ge 0,\ v \in \mathbb{R}^m\,. \tag{2.5}$$

For the partitioned Runge-Kutta scheme we then have

$$h^T u_{n+1} = h^T u_n + \Delta t \sum_{k \ne l} \sum_{j=1}^{s} \big(b_j^{(k)} - b_j^{(l)}\big) h^T F_k(t_n + c_j \Delta t, v_{n,j})\,,$$

for any $1 \le l \le r$. Therefore, as noted in [3], the discrete conservation property $h^T u_{n+1} = h^T u_n$ will be satisfied provided that

$$b_j^{(k)} = b_j^{(l)} \qquad \text{for all } 1 \le k, l \le r \text{ and } 1 \le j \le s\,. \tag{2.6}$$

If the h_j represent lengths, areas or volumes of cells, this is often called *mass conservation*. Of course, if $h^T F_k(t, v) \equiv 0$ for all $1 \le k \le r$, then the conservation property will always hold, even if (2.6) is not satisfied. This will be valid for decompositions of F that are based on fluxes.

Order conditions

The order conditions for partitioned Runge-Kutta methods applied to non-stiff problems are found e.g. in [5, Thm. I.15.9] for $r = 2$. This order will be denoted by p. As we will see, it often does not correspond to the order of convergence for semi-discrete ODE systems, and therefore p is usually referred to as the *classical order*.

To write the order conditions in a compact way, let the coefficients of the method be contained in $A_k = [a_{ij}^{(k)}] \in \mathbb{R}^{s \times s}$ and $b_k = [b_i^{(k)}] \in \mathbb{R}^s$, and set $e = [1, \ldots, 1]^T \in \mathbb{R}^s$. The conditions for order p up to 3 are

$$p = 1: \quad b_k^T e = 1 \qquad \text{for } \ k = 1, \ldots, r\,, \tag{2.7a}$$

$$p = 2: \quad b_k^T A_l\, e = \tfrac{1}{2} \qquad \text{for } \ k, l = 1, \ldots, r\,, \tag{2.7b}$$

$$p = 3: \quad b_k^T C_{l_1} A_{l_2} e = \tfrac{1}{3}\,, \quad b_k^T A_{l_1} A_{l_2} e = \tfrac{1}{6} \quad \text{for } \ k, l_1, l_2 = 1, \ldots, r\,, \tag{2.7c}$$

where $C_l = \mathrm{diag}(A_l e)$.

For semi-discrete ODE systems obtained from a PDE, the accuracy of the internal stage vectors $v_{n,i} \approx u(t_n + c_i \Delta t)$ is also of importance. The component-wise powers of $c = [c_i] = A_r e$ are denoted by $c^j = [c_i^j]$, and $c^0 = e$. The method is said to have *stage order* q if

$$A_k\, c^j \;=\; \tfrac{1}{j+1}\, c^{j+1} \qquad \text{for } \ j = 0, \ldots, q-1 \ \text{ and } \ k = 1, \ldots, r\,. \tag{2.8}$$

A method is internally consistent if it has stage order $q \geq 1$. Furthermore, it is easy to see that an explicit method cannot have $q > 1$.

Finally, we mention that a necessary condition for having order p is

$$b_k^T c^j = \tfrac{1}{j+1} \qquad \text{for } j = 0, \ldots, p \text{ and } k = 1, \ldots, r\,. \tag{2.9}$$

3 Multirate methods

An important class of methods contained in (2.3) are the multirate methods. We will consider multirate methods that are based on a single Runge-Kutta method, such that if $\mathcal{I}_k = \mathcal{I}$ and the other $\mathcal{I}_l$ are empty, then (2.3) reduces to m_k applications of this base method (with step-size $\Delta t/m_k$, $m_1 = 1 < m_2 < \cdots < m_r$). It was shown in [6] that the conditions for internal consistency (2.4) and conservation of linear invariants (2.6) are incompatible for such multirate schemes.

3.1 Examples

We consider some explicit multirate schemes that were discussed in [6]; additional examples can be found e.g. in [3,15,16,19]. The schemes in this paper are either based on the forward Euler method

$$u_{n+1} = u_n + \Delta t F(t_n, u_n)\,,$$

or the explicit trapezoidal rule (modified Euler method)

$$u^*_{n+1} = u_n + \Delta t F(t_n, u_n)\,, \qquad u_{n+1} = u_n + \tfrac{1}{2}\Delta t F(t_n, u_n) + \tfrac{1}{2}\Delta t F(t_{n+1}, u^*_{n+1})\,.$$

Furthermore, we take $r = 2$, $m_1 = 1$, $m_2 = 2$, that is, the local time step is Δt on $\mathcal{I}_1$ and $\frac{1}{2}\Delta t$ on $\mathcal{I}_2$. The coefficients of the schemes are represented by a tableau

$$\begin{array}{c|c|c} c & A_1 & A_2 \\ \hline & b_1^T & b_2^T \end{array}$$

with $A_k = [a_{ij}^{(k)}] \in \mathbb{R}^{s\times s}$, $b_k = [b_i^{(k)}] \in \mathbb{R}^s$ and $c = [c_i] = A_2 e \in \mathbb{R}^s$.

The scheme with $s = 2$, $p = 1$, $q = 0$, given by the tableau

$$\begin{array}{c|cc|cc} 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 1/2 & 0 \\ \hline & 1/2 & 1/2 & 1/2 & 1/2 \end{array} \tag{3.1}$$

is a simple example from Osher & Sanders [11], applied here with only one level of temporal refinement. We refer to this as the OS1 scheme.

Another scheme based on forward Euler was given by Tang & Warnecke [18]. It has $s = 2$, $p = q = 1$,

$$\begin{array}{c|cc|cc} 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 1/2 & 0 \\ \hline & 1 & 0 & 1/2 & 1/2 \end{array} \tag{3.2}$$

This scheme is internally consistent but does not conserve linear invariants because $b_1 \neq b_2$. We will refer to (3.2) as the TW1 scheme.

A second-order scheme of Tang & Warnecke [18], referred to as the TW2 scheme, is based on the explicit trapezoidal rule. It has $s = 4$, $p = 2$, $q = 1$,

$$\begin{array}{c|cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 1/2 & 1/4 & 1/4 & 0 & 0 & 1/4 & 1/4 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1/4 & 1/4 & 1/2 & 0 \\ \hline & 1/2 & 0 & 0 & 1/2 & 1/4 & 1/4 & 1/4 & 1/4 \end{array} \tag{3.3}$$

A related scheme, due to Constantinescu & Sandu [3], with $s = 4$, $p = 2$, $q = 0$, is given by

$$\begin{array}{c|cccc|cccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1/4 & 1/4 & 1/2 & 0 \\ \hline & 1/4 & 1/4 & 1/4 & 1/4 & 1/4 & 1/4 & 1/4 & 1/4 \end{array} \tag{3.4}$$

This scheme is conservative, but not internally consistent. We will refer to (3.4) as the CS2 scheme.

As a final example we consider the following scheme with $s = 5$, $p = 2$, $q = 1$,

$$\begin{array}{c|ccccc|ccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1/2 & 3/8 & 1/8 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 \\ 1/2 & 3/8 & 1/8 & 0 & 0 & 0 & 1/4 & 0 & 1/4 & 0 & 0 \\ 1 & 1/2 & 1/2 & 0 & 0 & 0 & 1/4 & 0 & 1/4 & 1/2 & 0 \\ \hline & 1/2 & 1/2 & 0 & 0 & 0 & 1/4 & 0 & 1/4 & 1/4 & 1/4 \end{array} \tag{3.5}$$

We will refer to this as the SH2 scheme. This scheme has been described in [6]; it was obtained by adaptation of an implicit (Rosenbrock) scheme from [14]. Although it looks already a bit complicated, the idea is simple: first a coarse Δt step is taken with the explicit trapezoidal rule on the whole index set $\mathcal{I}$, and then two refined $\frac{1}{2}\Delta t$ steps are taken on $\mathcal{I}_2$, using information from the coarse step by quadratic (Hermite) interpolation at the time level $t_n + \frac{1}{2}\Delta t$.

It is important to note that the number of stages s is not a good measure for the work-load per step. For example, with the SH2 scheme we have $s = 5$, but neglecting the (small) interface region only two F_1 evaluations and four F_2 evaluations are needed per step.

3.2 Numerical tests in 1D

3.2.1 Advection with smooth solution

A convergence analysis of the above multirate schemes, in the framework of partitioned Runge-Kutta methods, will be given in the next two sections. Here we present some simple numerical results for the second-order schemes that will motivate the analysis.

To test the accuracy of the schemes we consider the linear advection equation $u_t + u_x = 0$ on the spatial interval $\Omega = [0, 1]$ with periodic boundary conditions,

and time interval $0 < t \leq T = 1$. For test purposes a uniform spatial grid is taken, to ensure that interface effects are not related to the spatial discretization. The WENO5 finite difference scheme is used; see e.g. [17]. Further we employ a fixed Courant number $\nu = \Delta t/\Delta x = 0.5$, $\Delta x = 1/m$, and cell-based splitting $F = I_1F + I_2F$, with $\mathcal{I}_2 = \{i : x_i \in [\frac{1}{8}, \frac{3}{8}] \cup [\frac{5}{8}, \frac{7}{8}]\}$.

For this accuracy test a smooth solution $u(x,t) = \sin^2(\pi(x-t))$ is considered. The errors in the maximum norm ($\|v\|_\infty = \max_j |v_j|$) and discrete L_1-norm ($\|v\|_1 = \sum_j \Delta x_j |v_j|$) are presented in Table 3.1. The entries in the table are the total (absolute) errors with respect to the exact PDE solution. In this test the spatial errors are much smaller than the errors due to time integration with the multirate methods.

Table 3.1 Results for the smooth advection problem with the CS2, TW2 and SH2 schemes. Maximum errors and L_1-errors at final time $T = 1$ for various m with fixed Courant number $\Delta t/\Delta x = 0.5$, $\Delta x = 1/m$. The approximate order of convergence is also given.

m	100	200	400	800	Order
CS2, $\|\varepsilon_N\|_\infty$	$8.22 \cdot 10^{-4}$	$2.75 \cdot 10^{-4}$	$1.46 \cdot 10^{-4}$	$8.37 \cdot 10^{-5}$	1
CS2, $\|\varepsilon_N\|_1$	$2.85 \cdot 10^{-4}$	$7.81 \cdot 10^{-5}$	$2.09 \cdot 10^{-5}$	$5.73 \cdot 10^{-6}$	2
TW2, $\|\varepsilon_N\|_\infty$	$3.12 \cdot 10^{-4}$	$8.04 \cdot 10^{-5}$	$2.02 \cdot 10^{-5}$	$5.05 \cdot 10^{-6}$	2
TW2, $\|\varepsilon_N\|_1$	$1.98 \cdot 10^{-4}$	$5.12 \cdot 10^{-5}$	$1.28 \cdot 10^{-5}$	$3.21 \cdot 10^{-6}$	2
SH2, $\|\varepsilon_N\|_\infty$	$3.13 \cdot 10^{-4}$	$8.06 \cdot 10^{-5}$	$2.02 \cdot 10^{-5}$	$5.05 \cdot 10^{-6}$	2
SH2, $\|\varepsilon_N\|_1$	$1.99 \cdot 10^{-4}$	$5.13 \cdot 10^{-5}$	$1.28 \cdot 10^{-5}$	$3.21 \cdot 10^{-6}$	2

It is seen that with the CS2 scheme we have only first-order convergence in the maximum norm. The largest errors are localized near the interface points; the L_1-errors are still second-order. For the schemes TW2 and SH2 we have order two convergence also in the maximum norm.

To see that the largest errors for the CS2 scheme occur indeed at the interfaces, the errors as function of x at the final time $t_n = T = 1$ are displayed in Figure 3.1 for $m = 400$. The (relatively) large errors for the CS2 scheme at the interface points are clearly visible. In contrast, the errors for the TW2 scheme show no visible interface effects; the errors for SH2 were almost the same as those for TW2 in this test.

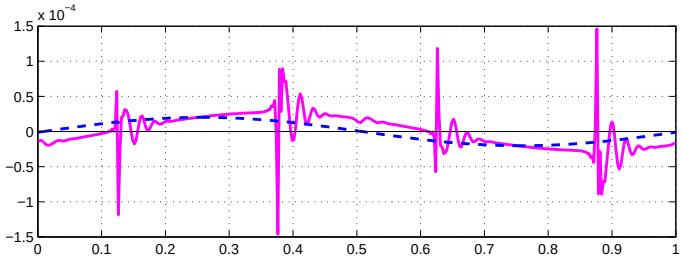

Fig. 3.1 Linear advection test. Errors versus $x_j \in [0,1]$ at final time $T = 1$ for the schemes CS2 (solid line) and TW2 (dashed line), $m = 400$.

3.2.2 Shock speeds with Burgers' equation

The main topic studied in this paper is convergence for smooth solutions. Mass conservation will play only a minor role. This conservation property, or the lack of it, is of course important for problems with discontinuous solutions.

To illustrate this, we apply the CS2, TW2 and SH2 schemes with cell-based decomposition to Burgers' equation

$$u_t + f(u)_x = 0\,, \qquad f(u) = \tfrac{1}{2}u^2\,, \tag{3.6}$$

with periodic boundary conditions on the spatial region $\Omega = [0,1]$, and the initial block profile $u(x,0) = 1$ if $x \in [0, \frac{1}{2}]$, $u(x,0) = 0$ if $x \in [\frac{1}{2}, 1]$. The shock that starts at $x = \frac{1}{2}$ should be located at $x = \frac{3}{4}$ at the output time $T = \frac{1}{2}$.

A conservative spatial discretization $u_i' = \frac{1}{\Delta x}(f_{i-1/2}(u) - f_{i+1/2}(u))$ is used with local Lax-Friedrichs fluxes

$$f_{j+\frac{1}{2}}(u) \;=\; \tfrac{1}{2}\Big(f(u^-_{j+\frac{1}{2}}) + f(u^+_{j+\frac{1}{2}}) + \alpha_{j+\frac{1}{2}}(u^-_{j+\frac{1}{2}} - u^+_{j+\frac{1}{2}})\Big)\,, \tag{3.7}$$

where $\alpha_{j+\frac{1}{2}} = \max|f'(v)|$ with v ranging between the states $u^-_{j+1/2}$, $u^+_{j+1/2}$ to the left and the right of the cell boundaries, computed from $u = [u_i]$ with the WENO5 scheme, as in [17].

The index sets $\mathcal{I}_k$ are changing in time, moving along with the shock. Let $u_n = [u_i^n]$ denote the numerical solution at time t_n; we take $\mathcal{I}_1 = \mathcal{I}_1^n = \{i : u_i^n < \frac{1}{8}\}$. So in the regions with small values of u_i^n, corresponding to small local Courant numbers $\Delta t|f'(u_i^n)|/\Delta x$, we use step-size Δt; elsewhere the step-size is $\frac{1}{2}\Delta t$.

The results with the non-conservative schemes TW2 and SH2 are shown in Figure 3.2. It is obvious that the lack of conservation leads to a shock that moves with a wrong speed; furthermore, the shock does not converge to the correct location upon refinement of the grid. For the conservative CS2 scheme the shock location is correct, of course, in accordance with the Lax-Wendroff theorem.

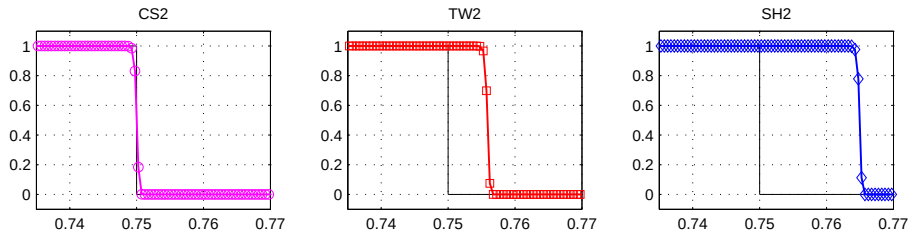

Fig. 3.2 Shock location for Burgers' equation at time $T = \frac{1}{2}$ for the schemes CS2 (left), TW2 (middle) and SH2 (right), with $\mathcal{I}_1 = \mathcal{I}_1^n = \{i : u_i^n < \frac{1}{8}\}$, $\Delta t = \Delta x = 1/m$, $m = 2000$.

4 Local and global discretization errors

The local discretization errors of the partitioned methods (2.3) will be expressed in terms of derivatives of the functions

$$\varphi_k(t) \;=\; F_k(t, u(t))\,. \tag{4.1}$$

The discretization errors can be studied for nonlinear problems; see Remark 4.2. However, to avoid unnecessary technical complications we consider only linear problems with constant coefficients,

$$u'(t) \;=\; Lu(t) + g(t)\,. \tag{4.2}$$

For the error analysis in this section we will follow an approach which is common for splitting methods; see [8, Chap. IV], for example. Let $Z_k = \Delta t L_k$, corresponding to the splitting $Z = \Delta t L = Z_1 + \cdots + Z_r$. Below some rational or polynomial expressions in the matrices Z_j will arise. For this we will use the notation

$$\underline{Z} \;=\; (Z_1, Z_2, \ldots, Z_r)\,. \tag{4.3}$$

4.1 Perturbed schemes

To derive recursions for the global errors, it is convenient to first study the effect of perturbations on the stages. Along with (2.3) we consider a perturbed scheme

$$\tilde v_{n,i} \;=\; \tilde u_n + \Delta t \sum_{k=1}^{r} \sum_{j=1}^{s} a_{ij}^{(k)} F_k(t_n + c_j \Delta t, \tilde v_{n,j}) \;+\; \rho_{n,j}\,, \qquad i = 1, \ldots, s\,, \tag{4.4a}$$

$$\tilde u_{n+1} \;=\; \tilde u_n + \Delta t \sum_{k=1}^{r} \sum_{j=1}^{s} b_{j}^{(k)} F_k(t_n + c_j \Delta t, \tilde v_{n,j}) \;+\; \sigma_n\,. \tag{4.4b}$$

The perturbations can be used to define residual, local errors per stage.

For the vector $c = [c_i] \in \mathbb{R}^s$ we denote its j-th power per component as $c^j = [c_i^j]$ for $j \geq 1$, with $c^0 = e = [1, \ldots, 1]^T \in \mathbb{R}^s$. To make the dimensions fitting we will use the Kronecker products $\boldsymbol{A}_k = A_k \otimes I$, $\boldsymbol{b}_k^T = b_k^T \otimes I$, $\boldsymbol{c^j} = c^j \otimes I$ and $\boldsymbol{e} = e \otimes I$ with $m \times m$ identity matrix $I = I_{m\times m}$. To make the notation consistent, the $ms \times ms$ identity matrix is denoted by $\boldsymbol{I}$. Furthermore, we let $\boldsymbol{Z}_k = I \otimes Z_k$, $Z_k = \Delta t L_k$, with $I = I_{s\times s}$.

To write the difference of (4.4) and (2.3) in a compact form, let also $\boldsymbol{\rho}_n = [\rho_{n,i}]$ and $\boldsymbol{v}_n = [v_{n,i}]$, $\tilde{\boldsymbol{v}}_n = [\tilde v_{n,i}] \in \mathbb{R}^{sm}$. Then

$$\tilde{\boldsymbol{v}}_n - \boldsymbol{v}_n \;=\; \boldsymbol{e}(\tilde u_n - u_n) + \sum_{k=1}^{r} \boldsymbol{A}_k \boldsymbol{Z}_k (\tilde{\boldsymbol{v}}_n - \boldsymbol{v}_n) + \boldsymbol{\rho}_n\,, \tag{4.5a}$$

$$\tilde u_{n+1} - u_{n+1} \;=\; \tilde u_n - u_n + \sum_{k=1}^{r} \boldsymbol{b}_k^T \boldsymbol{Z}_k (\tilde{\boldsymbol{v}}_n - \boldsymbol{v}_n) + \sigma_n\,. \tag{4.5b}$$

Elimination of $\tilde{\boldsymbol{v}}_n - \boldsymbol{v}_n$ leads to

$$\tilde{u}_{n+1} - u_{n+1} \;=\; R(\underline{Z})(\tilde{u}_n - u_n) + \boldsymbol{r}(\underline{Z})^T \boldsymbol{\rho}_n + \sigma_n \,, \tag{4.6}$$

where the amplification matrix $R(\underline{Z}) \in \mathbb{R}^{m\times m}$ and $\boldsymbol{r}(\underline{Z})^T \in \mathbb{R}^{m\times ms}$ are defined by

$$R(\underline{Z}) \;=\; I + \boldsymbol{r}(\underline{Z})^T \boldsymbol{e} \,, \tag{4.7a}$$

$$\boldsymbol{r}(\underline{Z})^T \;=\; \big[r_1(\underline{Z}), \ldots, r_s(\underline{Z})\big] = \Big(\sum_{k=1}^{r} \boldsymbol{b}_k^T \boldsymbol{Z}_k \Big) \Big(\boldsymbol{I} - \sum_{k=1}^{r} \boldsymbol{A}_k \boldsymbol{Z}_k \Big)^{-1} . \tag{4.7b}$$

The $r_j(\underline{Z})$ are polynomial expressions (for explicit methods) or rational expressions (for implicit methods) in $Z_1, Z_2, \ldots, Z_r$. It will be assumed that these expressions are bounded,

$$\|r_j(\underline{Z})\| \;\le\; M \qquad \text{for } j = 1, \ldots, r \,, \tag{4.8}$$

with some fixed $M > 0$. For explicit methods, this will be ensured by requiring that the matrices Z_j are bounded. Moreover, it may be assumed that the functions $r_j(\underline{Z})$ are not linearly dependent:

$$\sum_{j=1}^{s} \gamma_j \, r_j(\underline{Z}) \;=\; 0 \quad (\text{for all } \underline{Z} = (Z_1, \ldots, Z_r)) \quad \Longrightarrow \quad \gamma_1 = \gamma_2 = \cdots = \gamma_s = 0 \,. \tag{4.9}$$

Violation of this last condition would mean there are perturbations in (4.4) of the form $\rho_{n,j} = \gamma_j \rho_0$, with $\rho_0 \in \mathbb{R}^m$, which do not influence the outcome, no matter how large $\|\rho_0\|$ is. This indicates a redundancy (reducibility) in the scheme.

4.2 Error recursions

Let $\varepsilon_n = u(t_n) - u_n$ be the global discretization error at time level t_n, $n \ge 0$. As we will see, these global errors satisfy a recursion

$$\varepsilon_{n+1} \;=\; R(\underline{Z})\, \varepsilon_n \;+\; \delta_n \,, \qquad n \ge 0 \,, \tag{4.10}$$

where δ_n is a local discretization error, introduced in the step from t_n to t_{n+1}.

Lemma 4.1 *Suppose the functions $\varphi_k(t) = F_k(t, u(t))$ $(1 \le k \le r)$ are l times continuously differentiable, and (4.8) is valid. Then the local error δ_n in (4.10) is given by*

$$\delta_n \;=\; \sum_{j=1}^{l} \frac{\Delta t^j}{j!} \sum_{k=1}^{r} d_{j,k}(\underline{Z}) \varphi_k^{(j-1)}(t_n) \;+\; \mathcal{O}(\Delta t^{l+1}) \max_{k,t} \|\varphi_k^{(l)}(t)\| \,, \tag{4.11}$$

where

$$d_{j,k}(\underline{Z}) = \big(I - j \boldsymbol{b}_k^T \boldsymbol{c}^{j-1} \big) + \boldsymbol{r}(\underline{Z})^T \big(\boldsymbol{c}^j - j \boldsymbol{A}_k \boldsymbol{c}^{j-1} \big) \,. \tag{4.12}$$

Proof. Consider the perturbed scheme (4.4) with $\tilde{u}_n = u(t_n)$ and $\tilde{v}_{n,i} = u(t_n + c_i \Delta t)$, $i = 1, \ldots, s$. This choice for the $\tilde{v}_{n,i}$ defines the perturbations $\rho_{n,i}$ and σ_n, and we obtain by Taylor expansion

$$\boldsymbol{\rho}_n = \sum_{k=1}^{r} \sum_{j \geq 1} \frac{\Delta t^j}{j!} \big(\boldsymbol{c}^j - j \boldsymbol{A}_k \boldsymbol{c}^{j-1} \big) \varphi_k^{(j-1)}(t_n) \,,$$
$$\sigma_n = \sum_{k=1}^{r} \sum_{j \geq 1} \frac{\Delta t^j}{j!} \big(I - j \boldsymbol{b}_k^T \boldsymbol{c}^{j-1} \big) \varphi_k^{(j-1)}(t_n) \,.$$

If the φ_k are l times continuously differentiable, the sum over j can be truncated, with j ranging from 1 to l and with a remainder term at the Δt^{l+1} level, involving the $\|\varphi_k^{(l)}(t)\|$ with t between t_n and $t_n + c_i \Delta t$, $i = 1, \ldots, s$.

Subtraction of (2.3) from (4.4), leads to the error recursion (4.10) with

$$\delta_n = \boldsymbol{r}(\underline{Z})^T \boldsymbol{\rho}_n + \sigma_n \,.$$

Insertion of the Taylor expansions for $\boldsymbol{\rho}_n$ and σ_n, with assumption (4.8), thus lead to the expressions (4.11), (4.12) for the local errors. □

For a method with classical order p and stage order $q \leq p$, we have

$$d_{j,k}(\underline{Z}) \;=\; 0 \qquad \text{if } \; j \leq q \,, \tag{4.13a}$$

$$d_{j,k}(\underline{Z}) \;=\; \boldsymbol{r}(\underline{Z})^T \big(\boldsymbol{c}^j - j \boldsymbol{A}_k \boldsymbol{c}^{j-1} \big) \qquad \text{if } \; q < j \leq p \,. \tag{4.13b}$$

Note that $\boldsymbol{r}(\underline{Z}) = 0$ if all $Z_l = 0$, and so the same property holds for the functions $d_{j,k}(\underline{Z})$, $q < j \leq p$. In fact, since we know that $\delta_n = \mathcal{O}(\Delta t^{p+1})$ for non-stiff problems [5], it follows that $d_{j,k}(\underline{Z}) = \mathcal{O}(\Delta t^{p+1-j})$ if all $Z_l = \mathcal{O}(\Delta t)$, $1 \leq l \leq r$. The above properties will be used in the analysis of the local discretization errors.

Remark 4.2 The above derivations can also be performed for nonlinear problems (2.1), essentially by replacing occurring Kronecker products such as $\boldsymbol{Z} = I \otimes Z$, $Z = \Delta t L$, with the varying block-diagonal matrix $\boldsymbol{Z} = \mathrm{Diag}(Z_{n,i}) \in \mathbb{R}^{ms \times ms}$ where $Z_{n,i}(\tilde{v}_{n,i} - v_{n,i}) = \Delta t \big(F(\tilde{v}_{n,i}) - F(v_{n,i}) \big)$, with changes over the steps and the stages. However, this leads to more complicated notation, and it does not give additional insight. ◇

5 Error analysis for cell-based splittings

From now on, we restrict our attention to explicit methods. In this section it will be assumed that the splitting (2.2) is cell-based, $F_j = I_j F$ for $j = 1, \ldots, r$. Then we have $\varphi_k(t) = I_k u'(t)$, which is bounded in the maximum norm uniformly in the spatial mesh width. For flux-based splittings, considered in Section 6, this last property will not be valid.

Throughout the remaining sections we will denote by $\mathcal{O}(\Delta t^q)$ a scalar or vector for which all components can be bounded $K \Delta t^q$, for $\Delta t > 0$ small enough, with K not depending on the mesh widths Δx_j in the spatial discretization.

5.1 Order reduction

In this section we derive bounds for the discretization errors that are valid for semi-discrete systems with smooth solutions. The classical, non-stiff order conditions are then no longer sufficient to obtain convergence of order p. This so-called order reduction is due to the fact that F contains negative powers of the mesh widths Δx_j in space. We will accept a restriction on $\Delta t/\Delta x_j$ for stability, but the resulting error bounds should not contain negative powers of Δx_j.

For the partitioned methods we want to see the effects of the partitioning on the errors. We will therefore study the errors in the maximum norm, assuming stability of the scheme:

$$\sup_{n\geq 0} \|R(\underline{Z})^n\|_\infty \leq K\,. \tag{5.1}$$

Sufficient conditions for having (5.1) with $K = 1$ have been derived in [6] for nonlinear problems. For explicit methods, a necessary stability condition is boundedness of the Z_j, and therefore (4.8) will be satisfied.

Let $Z_k = \Delta t L_k$, corresponding to the splitting $L = L_1 + \cdots + L_r$. If L is a discretized convection operator, and a CFL restriction $\Delta t/\Delta x_j \leq \nu$ is satisfied with some fixed ν, then $\|Z_k\|_\infty = \mathcal{O}(1)$. It will be tacitly assumed that the exact solution is smooth, so that derivatives of $u(t)$ are $\mathcal{O}(1)$. If the splitting is cell-based, then $\varphi_k(t) = I_k u'(t)$, so any term $\varphi_k(t)$ and its time derivatives will then be $\mathcal{O}(1)$. Note, however, that $\varphi_k(t)$ is not a smooth grid function: there will be jumps over the interfaces of the spatial components, and therefore we will in general only have $\|Z\varphi_k(t)\|_\infty = \mathcal{O}(1)$ instead of $\|Z\varphi_k(t)\|_\infty = \mathcal{O}(\Delta t)$.

If the stability assumption (5.1) holds, it follows directly that consistency of order q (i.e., $\|\delta_n\|_\infty = \mathcal{O}(\Delta t^{q+1})$) implies convergence of order q (i.e., $\|\varepsilon_n\|_\infty = \mathcal{O}(\Delta t^q)$), but we will see that the order of convergence can also be one larger than the order of consistency.

5.2 Local error analysis

To analyze the order of the local errors we will distinguish various cases , depending whether the method is internally consistent or not (stage order $q \geq 1$ or $q = 0$).

Stage order zero: Let us first consider a method with classical order $p \geq 1$ and stage order $q = 0$, that is, the method is not internally consistent: $A_k e \neq A_l e$ for some k, l. Then the leading term in the local error is

$$\delta_n = \Delta t \sum_{k=1}^{r} d_{1,k}(\underline{Z})\varphi_k(t_n) + \mathcal{O}(\Delta t^2)\,, \qquad d_{1,k}(\underline{Z}) = \boldsymbol{r}(\underline{Z})^T(\boldsymbol{c} - \boldsymbol{A}_k\boldsymbol{e})\,. \tag{5.2}$$

Since $\varphi_k(t_n) = \mathcal{O}(1)$, this gives an $\mathcal{O}(\Delta t)$ local error bound in the maximum norm, which is of course quite poor. After all, δ_n is the error that results after one step if $\varepsilon_n = 0$. However, it will be seen that this still can lead to convergence of order one.

Stage order one: Next assume $q \geq 1$, that is, the internal consistency condition (2.4) is satisfied: $A_k e = A_l e$ for $1 \leq k, l \leq r$. If $p = 1$ it follows directly that $\|\delta_n\|_\infty = \mathcal{O}(\Delta t^2)$.

If $p \geq 2$ the leading term in the local discretization errors is given by

$$\delta_n = \tfrac{1}{2}\Delta t^2 \sum_{k=1}^{r} d_{2,k}(\underline{Z})\varphi_k'(t_n) + \mathcal{O}(\Delta t^3)\,, \qquad d_{2,k}(\underline{Z}) = \boldsymbol{r}(\underline{Z})^T\big(\boldsymbol{c}^2 - 2\boldsymbol{A}_k\boldsymbol{c}\big)\,. \tag{5.3}$$

This still gives only consistency of order one, that is, an error $\mathcal{O}(\Delta t^2)$ after one step, but we will discuss below damping and cancellation effects that can lead to convergence with order two in this case.

Higher orders: For explicit methods it is not possible to have $c^2 = 2A_k c$. With (4.9) this implies that the functions $d_{2,k}(\underline{Z})$ cannot be identically equal to zero. Yet there are exceptional cases where (5.3) can give consistency of order larger than one, under the assumption that $Lu''(t) = \mathcal{O}(1)$. If all $d_{2,k}(\underline{Z})$ are equal, say

$$d_{2,k}(\underline{Z}) \;=\; Q(\underline{Z}) \qquad \text{for } \; k = 1, \ldots, r\,, \tag{5.4}$$

with $Q(\underline{Z})$ a polynomial expression in the Z_k, then we have $\delta_n = \frac{1}{2}\Delta t^2 Q(\underline{Z})u''(t_n) + \mathcal{O}(\Delta t^3)$. Since the constant term in $Q(\underline{Z})$ is zero if $p \geq 2$, and $Z_k u''(t) = \Delta t I_k L u''(t) = \mathcal{O}(\Delta t)$, this gives indeed $\|\delta_n\|_\infty = \mathcal{O}(\Delta t^3)$. It should be noted, however, that (5.4) will only occur if

$$A_k c = A_l c \qquad \text{for all } \; k, l = 1, \ldots, r\,. \tag{5.5}$$

This will hold, of course, for the case that all coefficient matrices A_k are equal, but it is not valid for the multirate methods from Section 3.1.

Methods with equal coefficient matrices A_k have been studied in [9]. For such methods the above arguments can be simplified, see Theorem 2.1 in [9] and Remark 6.2 in the present paper, since the internal stages then only use the complete function F rather than the F_k from the decomposition (2.2).

The above expressions for the local errors show that *order reduction* is to be expected: the accuracy will primarily depend on the stage order q, rather than on the classical order p. This order reduction will appear primarily at interface points on the spatial grid, where the grid-functions $\varphi_k(t)$ have jumps.

Further we note that these expressions for the local errors are similar to those given, for example, in [7] for implicit-explicit Runge-Kutta methods, and in [12] for a class of implicit additive Runge-Kutta methods for parabolic problems with domain decomposition.

5.3 Global error analysis

Based on the local error behaviour, one would expect convergence with order one for the TW2 and SH2 schemes, and lack of convergence for the scheme CS2. This is not what was seen in the numerical test for advection with a smooth solution. To obtain the correct (observed) order of convergence, we need to study the propagation of the leading term in the local error.

In the following result we consider a partitioned method (2.3) with classical order p and stage order q. For the leading local error terms, it will be assumed that there is a matrix $W \in \mathbb{R}^{m\times m}$ such that

$$\big(\boldsymbol{r}(\underline{Z})^T\boldsymbol{e}\big)W \;=\; \sum_{k=1}^{r} d_{q+1,k}(\underline{Z})I_k\,. \tag{5.6}$$

Theorem 5.1 *Assume that $p \geq q+1$ and the stability condition (5.1) holds. Assume furthermore that (5.6) holds with $\|W\|_\infty = \mathcal{O}(1)$. Then the method is convergent with order $q+1$ in the maximum norm.*

Proof. Let $\xi_n = \frac{1}{(q+1)!}\Delta t^{q+1} W u^{(q+1)}(t_n)$. Then $\|\xi_n\|_\infty = \mathcal{O}(\Delta t^{q+1})$, $\|\xi_{n+1} - \xi_n\|_\infty = \mathcal{O}(\Delta t^{q+2})$, and the local error δ_n from (4.11) can be decomposed as

$$\delta_n \;=\; (R(\underline{Z}) - I)\xi_n + \eta_n \,,$$

where $\eta_n = \mathcal{O}(\Delta t^{q+2})$ contains the higher-order terms. Introducing $\hat{\varepsilon}_n = \varepsilon_n - \xi_n$, we get the recursion

$$\hat{\varepsilon}_{n+1} \;=\; R(\underline{Z})\hat{\varepsilon}_n + \hat{\delta}_n \,, \qquad \hat{\delta}_n \;=\; \xi_{n+1} - \xi_n + \eta_n \,.$$

In the standard way, it is seen from (5.1) that $\|\hat{\varepsilon}_n\|_\infty \leq K(\|\hat{\varepsilon}_0\|_\infty + \sum_{j=0}^n \|\hat{\delta}_j\|_\infty)$. Since $\varepsilon_0 = 0$ we obtain

$$\|\varepsilon_n\|_\infty \leq \|\xi_n\|_\infty + K\|\xi_0\|_\infty + \sum_{k=0}^{n} K\big(\|\xi_{k+1} - \xi_k\|_\infty + \|\eta_k\|_\infty\big) \,,$$

from which the convergence result now follows. □

This result and its proof is similar as for standard Runge-Kutta methods where order reduction may arise due to boundary conditions; see e.g. [2] or the review in [8, Sect. II.2]. With the partitioned Runge-Kutta methods and multirate schemes, we are creating interfaces that act like (internal) boundaries with time-dependent boundary conditions.

5.4 Examples: multirate methods with cell-based splittings

For the simple multirate examples from Section 3 we now study the order of convergence in the maximum norm. It will be assumed that

$$\|I + Z_1\|_\infty \;\leq 1 \,, \qquad \|I + \tfrac{1}{2}Z_2\|_\infty \;\leq\; 1 \,. \tag{5.7}$$

These conditions (or, rather, the nonlinear counterparts) were used in [6] to prove the stability condition (5.1) with $K = 1$ for the multirate schemes.

5.4.1 First-order multirate schemes OS1, TW1

For the TW1 scheme, we have $p = q = 1$, giving local errors $\|\delta_n\|_\infty = \mathcal{O}(\Delta t^2)$ from which we obtain in the standard way convergence with order 1 in the maximum norm.

Consider the OS1 scheme, with $p = 1$ but $q = 0$. Here $\|\delta_n\|_\infty = \mathcal{O}(\Delta t)$ only. Still, first-order convergence can be shown. For this, it is assumed, in addition to (5.7), that

$$\|Z_2\|_\infty \;\leq\; 4\theta \;<\; 4 \,. \tag{5.8}$$

From (5.7) it follows already that $\|Z_2\|_\infty \leq 4$, and consequently (5.8) is only a minor strengthening of the assumptions (5.7).

For this OS1 scheme we have, with $Z = Z_1 + Z_2$,

$$\boldsymbol{r}(\underline{Z})^T = \left[\tfrac{1}{2}Z(I + \tfrac{1}{2}Z_2) \quad \tfrac{1}{2}Z\right], \qquad c - A_1 e = \begin{bmatrix} 0 \\ \frac{1}{2} \end{bmatrix}, \qquad c - A_2 e = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

Hence

$$\boldsymbol{r}(\underline{Z})^T \boldsymbol{e} = Z(I + \tfrac{1}{4}Z_2), \qquad d_{1,1}(\underline{Z}) = \tfrac{1}{4}Z, \qquad d_{1,2}(\underline{Z}) = 0,$$

and (5.6) reads

$$(I + \tfrac{1}{4}Z_2)W = I_1 .$$

In view of (5.8) we have $\|(I + \frac{1}{4}Z_2)^{-1}\|_\infty \leq (1-\theta)^{-1}$, which ensures the bound $\|W\|_\infty \leq (1-\theta)^{-1}$. Application of Theorem 5.1 shows that the OS1 scheme will indeed converge with order 1 under the assumptions (5.7), (5.8).

5.4.2 Second-order multirate schemes CS2, TW2, SH2

For the second-order methods, the expressions for $R(\underline{Z})$ and the $d_{q+1,k}(\underline{Z})$ are already rather complicated. Therefore, instead of a detailed analysis of (5.6), we will only present here some experimental results for the semi-discrete system

$$u_j'(t) = \frac{1}{\Delta x_j}\big(u_{j-1}(t) - u_j(t)\big) \qquad \text{for } j \in \mathcal{I} = \{1, 2, \ldots, m\}, \tag{5.9}$$

with $u_0(t) = 0$, corresponding to first-order upwind discretization of the advection equation $u_t + u_x = 0$ with homogeneous inflow condition $u(0,t) = 0$.

We take a partitioning $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2 = \{1, 2, \ldots, m\}$ with $\mathcal{I}_2 = \{j : \frac{1}{4}m < j \leq \frac{3}{4}m\}$, and mesh widths $\Delta x_j = h$ if $j \in \mathcal{I}_1$, $\Delta x_j = \frac{1}{2}h$ if $j \in \mathcal{I}_2$, with $h = 4/(3m)$. In Figure 5.1 the norm $\|W\|_\infty$ is plotted as function of $m = 20, 40, \ldots, 640$ for various values of $\nu = \Delta t/h$ for the schemes TW2 and CS2; the results for SH2 were similar to those of TW2. In this example, the matrix $r(\underline{Z})^T \boldsymbol{e}$ is nonsingular, and it is well-conditioned for $\nu \leq 1$. We see that $\|W\|_\infty = \mathcal{O}(1)$ provided that $\nu < 1$, whereas $\|W\|_\infty \sim m$ if $\nu = 1$. Other partitionings $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2$ produced similar results.

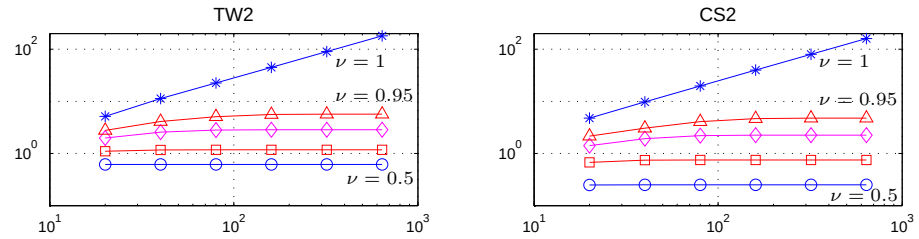

Fig. 5.1 Norm $\|W\|_\infty$ versus $m = 20, 40, \ldots, 640$ for various values of $\nu = \Delta t/h$ with the schemes TW2 (left) and CS2 (right). Markers: ○ for $\nu = 0.5$, □ for $\nu = 0.75$, ◇ for $\nu = 0.9$, △ for $\nu = 0.95$ and ∗ for $\nu = 1$.

The combination of Theorem 5.1 and these experimental bounds for first-order advection discretization does provide a heuristic explanation for the advection test results in Section 3.2, where we observed convergence of the schemes TW2 and SH2 with order two in the maximum norm, and with order one for the CS2 scheme.

5.5 Numerical test: 2D advection

The numerical test in Section 3.2 for 1D advection was highly artificial, because there was no practical need to refine on subintervals. Below we will present a more relevant test for advection in 2D.

As before, we will use the WENO5 scheme for the spatial discretization. This spatial scheme combines high accuracy with a good behaviour near discontinuities. Since the focus in this paper is temporal accuracy, we will use linear advection problems with smooth initial profiles in the tests. Due to the WENO5 spatial discretization, the semi-discrete ODE system is still nonlinear.

As a test example we consider here the two-dimensional advection equation

$$u_t + (a_1 u)_x + (a_2 u)_y \;=\; 0 \tag{5.10a}$$

for $0 < x, y < 1$, $0 < t \leq 1$, with divergence-free velocity field given by

$$a_1(x,y) = 2\pi(y - \tfrac{1}{2})\,, \qquad a_2(x,y) = -2\pi(x - \tfrac{1}{2})\,, \tag{5.10b}$$

and initial profile

$$u(x,y,0) \;=\; e^{-10\left((x-\frac{1}{2})^2+(y-\frac{1}{4})^2\right)}\,. \tag{5.10c}$$

The wind field gives a uniform clock-wise rotation around the center of the domain. We take end time $T = \frac{1}{3}$, giving a rotation of the initial profile over an angle $\frac{2}{3}\pi$. At the inflow boundaries Dirichlet conditions are described, corresponding to the exact solution.

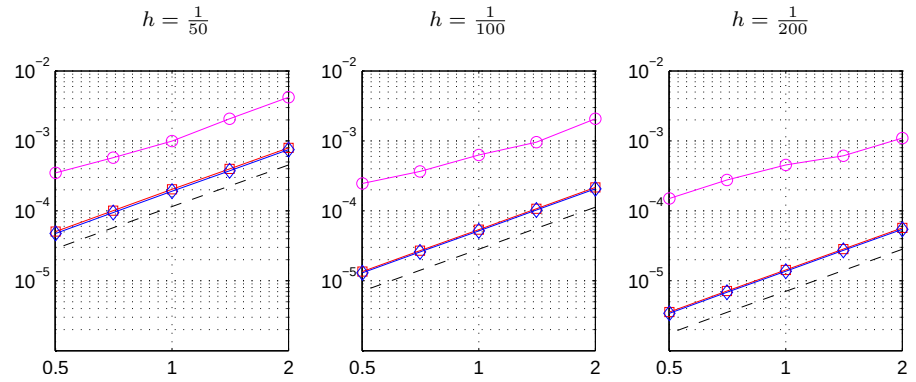

Fig. 5.2 Test (5.10) with cell-based decomposition. Maximum errors for the schemes TW2 (□ marks), CS2 (○ marks) and SH2 (◇ marks) as function of Courant numbers $\nu = 2\pi\Delta t/h$, between $\frac{1}{2}$ and 2, for the grids with $h = \frac{1}{50}$ (left), $h = \frac{1}{100}$ (middle) and $h = \frac{1}{200}$ (right).

We consider a partitioning where $\mathcal{I}_1$ corresponds to the grid points in the region where $|x - \frac{1}{2}| + |y - \frac{1}{2}| \leq \frac{1}{3}$. This is a natural partitioning since the velocity field increases towards the corners of the domain. In the test we compare the solutions obtained by the multirate schemes with an accurate semi-discrete solution, obtained with a Runge-Kutta method with small step-size.

The results on three uniform grids, with $\Delta x = \Delta y = h$, $h = \frac{1}{50}, \frac{1}{100}, \frac{1}{200}$, are presented in Figure 5.2. There, for each separate grid, the maximum errors are plotted for various Courant numbers $\nu = \Delta t \max_{x,y}(|a_1| + |a_2|)/h = 2\pi\Delta t/h$. The dashed line in the figures gives the result for the scheme where in each time step the explicit trapezoidal rule is applied twice, with step-size $\frac{1}{2}\Delta t$, over the whole region.

On any fixed grid the three schemes are second-order convergent (classical order two), but it is clear that the CS2 scheme has a large error constant, affected by h. Comparing the errors on the three grids for the same Courant number shows indeed a very slow convergence for the CS2 scheme.

6 Decomposition based on fluxes

For conservation laws, the semi-discrete system (2.1) will in general be in conservative form. In 1D, for example, we will have

$$u_j'(t) = \frac{1}{\Delta x_j}\big(f_{j-\frac{1}{2}}(u(t)) - f_{j+\frac{1}{2}}(u(t))\big)\,, \qquad j \in \mathcal{I} = \{1, 2, \ldots, m\}\,. \tag{6.1}$$

Multirate methods can be based on these numerical fluxes $f_{j\pm 1/2}(u)$ rather than in terms of the components of $F(u)$.

A decomposition $F = F_1 + F_2 + \cdots + F_r$ can be based on fluxes in the following way. The conservative semi-discrete ODE system (6.1) has right-hand side function

$$F(v) = H^{-1} D\, \Phi(v) \tag{6.2}$$

with $H = \mathrm{diag}(\Delta x_j)$, bi-diagonal difference matrix D, and flux vector $\Phi(v) = [\Phi_j(v)]$, $\Phi_j(v) = f_{j+1/2}(v)$. If J_k corresponds to a discrete indicator function for a region Ω_k of the PDE domain $\Omega = \Omega_1 \cup \ldots \cup \Omega_r$, then

$$F_k(v) = H^{-1} D\, J_k\, \Phi(v)\,, \qquad k = 1, \ldots, r\,, \tag{6.3}$$

gives a flux-based decomposition of F.

As an example, suppose that $r = 2$, $\Omega_1 = \{x : x \le x_i\}$ and $\Omega_2 = \{x : x \ge x_i\}$. Then the jth component of the vector functions F_1 and F_2 is given by

$$\left.\begin{aligned}
F_{1,j}(v) &= \frac{1}{\Delta x_j}\big(f_{j-\frac{1}{2}}(v) - f_{j+\frac{1}{2}}(v)\big)\,, & F_{2,j}(v) &= 0 \quad \text{for } j < i\,,\\
F_{1,j}(v) &= \frac{1}{\Delta x_i} f_{i-\frac{1}{2}}(v)\,, & F_{2,j}(v) &= \frac{-1}{\Delta x_i} f_{i+\frac{1}{2}}(v) \quad \text{for } j = i\,,\\
F_{1,j}(v) &= 0\,, & F_{2,j}(v) &= \frac{1}{\Delta x_j}\big(f_{j-\frac{1}{2}}(v) - f_{j+\frac{1}{2}}(v)\big) \quad \text{for } j > i\,.
\end{aligned}\right\} \tag{6.4}$$

Since we are here dealing with fluxes, mass-conservation is guaranteed for any stage. However, there some serious drawbacks as well.

First, monotonicity assumptions such as $\|v + \tau F_k(v)\| \le \|v\|$ will not be valid in the maximum norm with this decomposition. This can be seen already quite easily for the first-order upwind advection discretization (5.9) with $r = 2$. Writing this system as $u'(t) = Lu(t)$, the above decomposition would correspond to $L = LI_1 + LI_2$, that

is, $F_k = LI_k$, but it is easy to show that $\|I+\tau LI_k\|_\infty$ is larger than one for any $\tau > 0$. Consequently, stability assumptions like (5.7) are also no longer relevant.

Secondly, such a flux-based decomposition of F can easily lead to inconsistencies, since we do not have $F_k(u(t)) = \mathcal{O}(1)$, no matter how smooth the solution is. For example, for the first-order upwind system (5.9), using these F_1 and F_2 in the OS1 scheme gives a completely inconsistent scheme. This issue of accuracy will be discussed next.

6.1 Error analysis

We will discuss here the effect of flux-based decompositions on the local errors. The transition of local to global errors is similar to the cell-based decompositions. Note that the formulas (5.2) and (5.3) are still correct for the leading term, with $\varphi_k(t) = F_k(t, u(t))$. However, now $\|\varphi_k(t)\|_\infty$ will be proportional to $1/\Delta x$, see e.g. formula (6.4) with $j = i$, and therefore we only have $\|\Delta t\varphi_k(t)\|_\infty = \mathcal{O}(1)$ under a CFL restriction on $\Delta t/\Delta x$. This may lead to smaller orders of consistency/convergence than for the cell-based splittings. We will discuss various cases, leading to convergence with order zero, one, or more, separately.

Stage order zero: If the method is not internally consistent, the principal local error term is $\Delta t \sum_k d_{1,k}(\underline{Z})\varphi_k(t_n)$, see (5.2). Since we now only have $\Delta t\varphi_k(t) = \mathcal{O}(1)$, the error after one step may not tend to zero as $\Delta t \to 0$.

Example 6.1 For the advection equation $u_t + u_x = 0$, consider (6.4) with first-order upwind fluxes $f_{j+1/2}(v) = v_j$, and denote the components of the vector u_n as $u_j^n \approx u(x_j, t_n)$. A little calculation shows that at the interface point the scheme (3.1) gives

$$u_i^{n+1} = u_i^n + \frac{\Delta t}{\Delta x_i}(u_{i-1}^n - u_i^n) + \tfrac{1}{4}\Big(\frac{\Delta t}{\Delta x_i}\Big)^2 u_i^n \,.$$

Already after one step, starting with $u_j^0 = u(x_j, t_0)$, this gives an error $u_i^1 - u(x_i, t_1) = \frac{1}{4}\nu_i^2 u(x_i, t_0) + \mathcal{O}(\Delta t)$ if $\Delta t \to 0$ while $\nu_i = \Delta t/\Delta x_i$ is held fixed, leading to an $\mathcal{O}(1)$ error in the maximum norm if $u(x_i, t_0) \neq 0$. ◇

Stage order one: For methods that are internally consistent, with stage order $q = 1$, the the principal local error term is $\frac{1}{2}\Delta t^2 \sum_k d_{2,k}(\underline{Z})\varphi_k'(t_n)$, see (5.3). Since $\Delta t\varphi_k'(t) = \mathcal{O}(1)$, this gives an error proportional to Δt in the maximum norm after one step. Due to damping and cancellation effects, we can still have convergence with order 1. In some numerical tests this will be seen to hold for the multirate methods from Section 3.1.

Higher orders: If we have an internally consistent method, for which all $d_{2,k}(\underline{Z})$ are equal, say $d_{2,k}(\underline{Z}) = Q(\underline{Z})$ for $k = 1, \ldots, r$, as in (5.4), then $\|\delta_n\|_\infty = \mathcal{O}(\Delta t^2)$, because $\sum_k \varphi_k(t) = u'(t)$, which is a smooth, bounded grid function, unlike the individual $\varphi_k(t)$ terms. As noted before, this requires (5.5), which does not hold for the multirate methods from Section 3.1.

For general partitioned methods, if we have, instead of (5.4), the stronger assumption

$$d_{2,k}(\underline{Z}) \;=\; P(\underline{Z}) \cdot Z \qquad \text{for } k = 1, \ldots, r\,, \tag{6.5}$$

then the leading term in $\|\delta_n\|_\infty$ will even be $\mathcal{O}(\Delta t^3)$, because in this case

$$\tfrac{1}{2}\Delta t^2 \sum_{k=1}^{r} d_{2,k}(\underline{Z})\varphi_k'(t_n) = \tfrac{1}{2}\Delta t^2 P(\underline{Z})\, Z u''(t_n)\,,$$

and $Zu''(t) = \Delta t L u''(t) = \mathcal{O}(\Delta t)$ if $Lu''(t) = \mathcal{O}(1)$, which will be valid if the PDE solution is smooth with boundary conditions that are constant in time. The assumption (6.5) will hold if $p \geq 3$ and all coefficient matrices A_k are equal.

Remark 6.2 As noted above, having a partitioned method with equal coefficient matrices A_k will often be beneficial with respect to the accuracy. In fact, it was shown in [9] that for such methods the order of consistency will be p for cell-based splittings and $p-1$ for flux-based splittings. This can also be demonstrated from the local error expansions that are used in this paper.

If $A_k = A$ for all k, then leading term in the local error is given by

$$\delta_n = \tfrac{1}{2}\Delta t^2 Q(\underline{Z})u''(t_n) + \mathcal{O}(\Delta t^3)\,, \qquad Q(\underline{Z}) \;=\; \boldsymbol{r}(\underline{Z})^T(\boldsymbol{c}^2 - 2\boldsymbol{A}\boldsymbol{c})\,,$$

and we have

$$\boldsymbol{r}(\underline{Z})^T \;=\; (\textstyle\sum_k \boldsymbol{b}_k^T \boldsymbol{Z}_k)(\boldsymbol{I} - \boldsymbol{A}\boldsymbol{Z})^{-1} \;=\; (\textstyle\sum_k \boldsymbol{b}_k^T \boldsymbol{Z}_k)(\boldsymbol{I} + \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{A}^2\boldsymbol{Z}^2 + \cdots)\,.$$

Hence

$$Q(\underline{Z}) \;=\; \textstyle\sum_k Z_k(q_{1k} + q_{2k}Z + q_{3k}Z^2 + \cdots)\,, \qquad q_{jk} = b_k^T A^{j-1}(c^2 - 2Ac)\,.$$

If the method has order p we have $q_{jk} = 0$ for $0 \leq j \leq p-2$, $1 \leq k \leq r$. Therefore, neglecting the higher order terms,

$$\delta_n \;=\; \tfrac{1}{2}\Delta t^p \textstyle\sum_k Z_k(q_{p-1,k} + q_{p,k}Z + \cdots)L^{p-2}u''(t_n)\,. \tag{6.6}$$

Assuming $L^{p-2}u''(t) = \mathcal{O}(1)$, which is an assumption on the boundary conditions for the PDE solution, this gives $\delta_n = \mathcal{O}(\Delta t^p)$. For cell-based splitting, $Z_k = I_k L$, we get

$$\delta_n \;=\; \tfrac{1}{2}\Delta t^{p+1} \textstyle\sum_k I_k(q_{p-1,k} + q_{p,k}Z + \cdots)L^{p-1}u''(t_n)\,. \tag{6.7}$$

Under the assumption $L^{p-1}u''(t) = \mathcal{O}(1)$ it now follows that $\delta_n = \mathcal{O}(\Delta t^{p+1})$, which is the classical order of consistency. ◇

6.2 Numerical tests

To show the effect of flux-based decompositions, the previous tests are repeated with the CS2, TW2 and SH2 schemes. It should be noted that, since the CS2 scheme is always conservative, there is actually no need to apply this scheme with such flux-based decompositions. Instead, the more accurate cell-based decompositions can be used for this scheme.

6.2.1 1D advection

We consider once more the simple problem $u_t + u_x = 0$ with periodic boundary conditions and $u(x,0) = \sin^2(\pi x)$, that was already used in Section 3.2 with cell-based splittings. The set-up of the test is the same as before, with WENO5 spatial discretization, temporal refinement on the domain $\{x : |x-\frac{1}{4}| \le \frac{1}{8}\} \cup \{x : |x-\frac{3}{4}| \le \frac{1}{8}\}$, and a fixed Courant number $\nu = \Delta t/\Delta x = 0.5$, only now we consider a flux-based splitting of $F = F_1 + F_2$. The results are given in Table 6.1.

Table 6.1 Flux-based splitting. Results for the smooth advection problem with the CS2, TW2 and SH2 schemes. Maximum errors and L_1-errors at final time $t_N = 1$ for various m with fixed Courant number $\Delta t/\Delta x = 0.5$, $\Delta x = 1/m$. The approximate order of convergence is also given.

m	100	200	400	800	Order
CS2, $\|\varepsilon_N\|_\infty$	$3.98 \cdot 10^{-2}$	$3.65 \cdot 10^{-2}$	$3.54 \cdot 10^{-2}$	$3.52 \cdot 10^{-2}$	0
CS2, $\|\varepsilon_N\|_1$	$4.43 \cdot 10^{-3}$	$1.48 \cdot 10^{-3}$	$5.12 \cdot 10^{-4}$	$2.09 \cdot 10^{-4}$	1
TW2, $\|\varepsilon_N\|_\infty$	$8.20 \cdot 10^{-4}$	$4.20 \cdot 10^{-4}$	$2.45 \cdot 10^{-4}$	$1.31 \cdot 10^{-4}$	1
TW2, $\|\varepsilon_N\|_1$	$2.45 \cdot 10^{-4}$	$6.57 \cdot 10^{-5}$	$1.80 \cdot 10^{-5}$	$5.08 \cdot 10^{-6}$	2
SH2, $\|\varepsilon_N\|_\infty$	$3.73 \cdot 10^{-4}$	$1.30 \cdot 10^{-4}$	$6.69 \cdot 10^{-5}$	$3.77 \cdot 10^{-5}$	1
SH2, $\|\varepsilon_N\|_1$	$2.07 \cdot 10^{-4}$	$5.29 \cdot 10^{-5}$	$1.36 \cdot 10^{-5}$	$3.49 \cdot 10^{-6}$	2

From this table, we make the following observations. In the maximum norm there is no convergence for the CS2 scheme, and only first-order (approximately) convergence for the TW2 and SH2 schemes. In the L_1-norm these orders of convergence are one higher, due to the fact that the largest errors are confined to relatively small spatial regions, near the interface points.

6.2.2 2D advection

Also the test for 2D advection (5.10) with rotational velocity field was performed again, but now with flux-based decomposition. Similar as before, for the cell-based splittings, we consider a partitioning of the region with step-size Δt in that part of the region where $|x-\frac{1}{2}| + |y-\frac{1}{2}| \le \frac{1}{3}$, and a step-size $\frac{1}{2}\Delta t$ is taken elsewhere. The errors in the solutions obtained by the multirate schemes are measured with respect to an accurate numerical solution of the semi-discrete system, so it is only the temporal error that is measured here.

The errors in the maximum norm are given in Figure 6.1 on three uniform grids, $\Delta x = \Delta y = h$ with $h = \frac{1}{50}, \frac{1}{100}, \frac{1}{200}$, and this is to be compared with the results in Figure 5.2. Again the dashed line in the figure gives the result for the scheme where in each time step the explicit trapezoidal rule is applied twice, with step-size $\frac{1}{2}\Delta t$, over the whole region.

Compared to Figure 5.2, the negative effect of the flux-based splitting on the accuracy of the CS2 scheme is obvious. Here it is to be noted that the vertical axis in Figure 6.1 has been shifted to include the error lines in the plots.

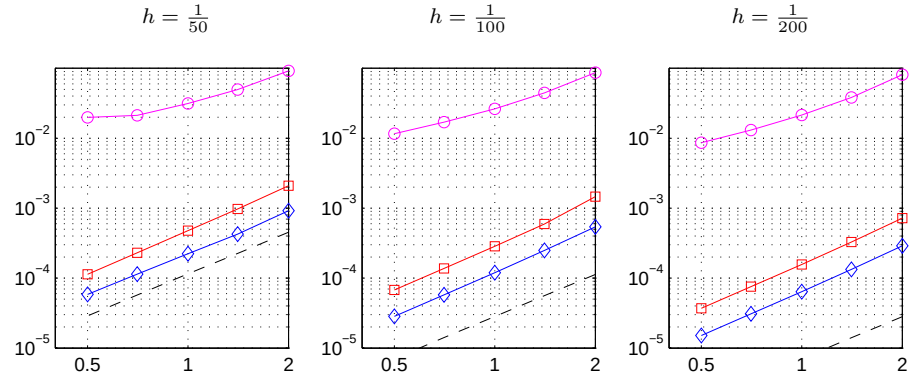

Fig. 6.1 Test (5.10) with flux-based decomposition. Maximum errors for the schemes TW2 (□ marks), CS2 (∘ marks) and SH2 (◇ marks) as function of Courant numbers $\nu = 2\pi\Delta t/h$, between $\frac{1}{2}$ and 2, for the grids with $h = \frac{1}{50}$ (left), $h = \frac{1}{100}$ (middle) and $h = \frac{1}{200}$ (right).

However, the accuracy of the TW2 and SH2 schemes has deteriorated as well, which is most clear by comparing these results with the ones for the explicit trapezoidal rule with small step-size $\frac{1}{2}\Delta t$ over the whole region, which we may consider here as a 'target' solution. In contrast to Figure 5.2, where the results of the TW2 and SH2 schemes were close to these reference solutions, now the errors with the TW2 and SH2 schemes are much larger, in particular on the finer grids.

Instead of errors $C\Delta t^2$ with a fixed constant C, the constants in front of the global errors are now proportional to h^{-1}, and comparing the results on the three grids for fixed ratios $\Delta t/h$, it can be observed that the order of convergence for TW2 and SH2 is now only one. So we still have convergence with these schemes in the maximum norm, but it is much slower than for the cell-based splittings.

The largest errors are found near the interfaces. Measuring the errors in the L_1-norm would yield convergence with one order higher, similar as for the 1D test in Table 6.1. Convergence with order two in the L_1-norm with the TW2 and SH2 schemes may be satisfactory for many applications.

7 Conclusions

In this paper the accuracy of partitioned Runge-Kutta methods has been studied, with applications to explicit multirate schemes. When such methods are applied to PDEs, it is not sufficient to look at the order for non-stiff problems. The interfaces between the regions where different methods – or different time steps – are applied act like time dependent boundary conditions, and order reduction is to be expected.

To see the effect of the partitioning at the interfaces, the accuracy of the schemes was mainly considered in the maximum norm. Convergence in the discrete L_1-norm is in general one order larger. This due to fact that the largest errors are confined to small spatial regions near the interfaces.

To guarantee mass conservation during all stages of the computation, a decomposition based on fluxes seems attractive. However, it was seen that the order of convergence for smooth problems will be smaller compared to cell-based splittings.

On the other hand, for partitioned Runge-Kutta methods with different weights, the cell-based splittings may lead to an incorrect propagation of discontinuities.

If a high accuracy is required, then one would like to use high-order methods, of course, and the decompositions considered in this paper do not seem to be very suited. Alternatives are the use of smooth partitions of unity, similar to the approach in [10, 13] for parabolic problems, or an approach with overlapping regions. The study of convergence and monotonicity/SSP properties of such methods is part of our current research.

Acknowledgements This paper originated from work of W. H. with Anna Mozartova and Valeriu Savcenco. The contributions of Mozartova and Savcenco, on the design of multirate methods and monotonicity properties of these methods, are contained in [6]. They are thanked for helpful comments on preliminary convergence results for first-order methods.

References

1. U.M. Ascher, S.J. Ruuth, R.J. Spiteri, *Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations*. Appl. Numer. Math. 25 (1997), 151–167.
2. P. Brenner, M. Crouzeix, V. Thomée, *Single step methods for inhomogeneous linear differential equations*. RAIRO Anal. Numer. 16 (1982), 5–26.
3. E.M. Constantinescu, A. Sandu, *Multirate timestepping methods for hyperbolic conservation laws*. J. Sci. Comput. 33 (2007), 239-278.
4. M. Günther, A. Kværnø, P. Rentrop, *Multirate partitioned Runge-Kutta methods*. BIT 41 (2001), 504–514.
5. E. Hairer, S.P. Nørsett, G. Wanner, *Solving Ordinary Differential Equations I – Nonstiff Problems*. Second edition, Springer Series Comput. Math. 8, Springer, 1993.
6. W. Hundsdorfer, A. Mozartova, V. Savcenco, *Monotonicity Conditions for Multirate and Partitioned Explicit Runge-Kutta Schemes*, In: Recent Developments in the Numerics of Nonlinear Hyperbolic Conservation Laws. Eds. R. Ansorge et al. NNFM 120, Springer, 2013, 177–195.
7. W. Hundsdorfer, S.J. Ruuth, *IMEX extensions of linear multistep methods with general monotonicity and boundedness properties*. J. Comput. Phys. 225 (2007), 2016–2042.
8. W. Hundsdorfer, J.G. Verwer, *Numerical Solution of Advection-Diffusion-Reaction Equations*. Springer Series Comput. Math. 33, Springer, 2003.
9. D.I. Ketcheson, C.B. Macdonald, S.J. Ruuth, *Spatially partitioned embedded Runge-Kutta methods*. SIAM J. Numer. Anal. (2013), to appear.
10. T.P. Mathew, P.L. Polyakov, G. Russo, J. Wang, *Domain decomposition operator splittings for the solution of parabolic equations*. SIAM. J. Sci. Comput. 19 (1998), 912–932.
11. S. Osher, R. Sanders, *Numerical approximations to nonlinear conservation laws with locally varying time and space grids*. Math. Comp. 41 (1983), 321–336.
12. L. Portero, J.C. Jorge, B. Bujanda, *Avoiding order reduction of fractional step Runge-Kutta discretizations for linear time dependent oefficient parabolic problems*. Appl. Numer. Math. 48 (2004), 409–424.
13. L. Portero, B. Bujanda, J.C. Jorge, *A combined fractional step domain decomposition method for the numerical integration of parabolic problems*, Proceedings PPAM2003, Eds. R. Wyrzykowski et al., LNCS 3019, Springer, 2004, 1034-1041.
14. V. Savcenco, W. Hundsdorfer, J.G. Verwer, *A multirate time stepping strategy for stiff ordinary differential equations*. BIT 47 (2007), 137–155.
15. M. Schlegel, O. Knoth, M. Arnold, R. Wolke, *Multirate Runge-Kutta schemes for advection equations*, J. Comp. Appl. Math. 226 (2009), 345–357.
16. M. Schlegel, O. Knoth, M. Arnold, R. Wolke, *Numerical solution of multiscale problems in atmospheric modeling*. Appl. Num. Math. 62 (2012), 1531–1543.
17. C.-W. Shu, *High Order Weighted Essentially Nonoscillatory Schemes for Convection Dominated Problems*. SIAM Review 51 (2009), 82–126.

18. H.-Z. Tang, G. Warnecke, *High resolution schemes for conservation laws and convection-diffusion equations with varying time and space grids*. J. Comput. Math. 24 (2006), 121–140.
19. J. Wensch, O. Knoth, A. Galant, *Multirate infinitesimal step methods for atmospheric flow simulation*, BIT Numer. Math. 49 (2009), 449–473.