Analysis of explicit multirate and partitioned Runge-Kutta schemes for conservation laws

Centrum voor Wiskunde en Informatica

REPORT*RAPPORT*

MAS

Modelling, Analysis and Simulation

Modelling, Analysis and Simulation

Analysis of explicit multirate and partitioned Runge-Kutta schemes for conservation laws

W. Hundsdorfer, A. Mozartova, V. Savcenco

REPORT MAS-E0715 AUGUST 2007

Centrum voor Wiskunde en Informatica (CWI) is the national research institute for Mathematics and Computer Science. It is sponsored by the Netherlands Organisation for Scientific Research (NWO).
CWI is a founding member of ERCIM, the European Research Consortium for Informatics and Mathematics.

CWI's research has a theme-oriented structure and is grouped into four clusters. Listed below are the names of the clusters and in parentheses their acronyms.

Probability, Networks and Algorithms (PNA)

Software Engineering (SEN)

Modelling, Analysis and Simulation (MAS)

Information Systems (INS)

P.O. Box 94079, 1090 GB Amsterdam (NL)
Kruislaan 413, 1098 SJ Amsterdam (NL)
Telephone +31 20 592 9333
Telefax +31 20 592 4199

ISSN 1386-3703

Analysis of explicit multirate and partitioned Runge-Kutta schemes for conservation laws

ABSTRACT
Multirate schemes for conservation laws or convection-dominated problems seem to come in two flavors: schemes that are locally inconsistent, and schemes that lack mass-conservation. In this paper these two defects are discussed for one-dimensional conservation laws. Particular attention will be given to monotonicity properties of the multirate schemes, such as maximum principles and the total variation diminishing (TVD) property. The study of these properties will be done within the framework of partitioned Runge-Kutta methods.

2000 Mathematics Subject Classification: 65L06, 65M06, 65M20
Keywords and Phrases: multirate methods, partitioned Runge-Kutta methods, monotonicity, TVD, stability, convergence

Analysis of Explicit Multirate and Partitioned Runge-Kutta Schemes for Conservation Laws

W. Hundsdorfer, A. Mozartova*, V. Savcenco†

CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

Abstract

Multirate schemes for conservation laws or convection-dominated problems seem to come in two flavors: schemes that are locally inconsistent, and schemes that lack mass-conservation. In this paper these two defects are discussed for one-dimensional conservation laws.

Particular attention will be given to monotonicity properties of the multirate schemes, such as maximum principles and the total variation diminishing (TVD) property. The study of these properties will be done within the framework of partitioned Runge-Kutta methods.

2000 Mathematics Subject Classification: 65L06, 65M06, 65M20.
Keywords and Phrases: multirate methods, partitioned Runge-Kutta methods, monotonicity, TVD, stability, convergence.

1 Introduction

Multirate schemes for conservation laws that have appeared in the literature all seem to have one of the following defects: there are schemes that are *locally inconsistent*, e.g. [3, 4, 17, 18], and schemes that are *not mass-conservative*, e.g. [26]. In this paper these two defects are discussed for one-dimensional conservation laws $u_t + f(u)_x = 0$. We will mainly concentrate on time stepping aspects for simple schemes with one level of temporal refinement. The spatial grids are assumed to be given and fixed in time. Spatial discretization of a PDE (partial differential equation) then leads to a system of ODEs (ordinary differential equations), the so-called semi-discrete system. Particular attention will be given to monotonicity properties of the multirate time stepping schemes, such as maximum principles and the total variation diminishing (TVD) property.

After some preliminaries, we will present in Section 3 a detailed analysis of two multirate forward Euler schemes, due to Osher & Sanders [18] and Tang & Warnecke [26]. The first of these schemes is inconsistent at interface points, but it will be shown that convergence of order one can be still obtained in the maximum-norm. Furthermore, we will see that step size restrictions for monotonicity will depend on the type of monotonicity: in general the restrictions for maximum principles can be more relaxed than for the TVD property.

*The work of A. M. is supported by the Netherlands Organisation for Scientific Research NWO.

†The work of V. S. is supported by a Peterich Scholarship through the Netherlands Organisation for Scientific Research NWO.

In Section 4 we will present some multirate schemes that are based on a standard two-stage Runge-Kutta method. These multirate schemes were recently introduced by Tang & Warnecke [26], Constantinescu & Sandu [3], and Savcenco et al. [22]. For these schemes some results of numerical experiments for linear advection and Burgers' equation are discussed.

For the analysis of general multirate schemes it is convenient to write them in the form of partitioned Runge-Kutta methods. In Section 5 it will be seen that recent results for (standard and additive) Runge-Kutta methods of Higueras, Ferracina and Spijker [7, 10, 11, 25] can then be employed to obtain monotonicity results for the multirate schemes through the partitioned Runge-Kutta methods. As for the forward Euler multirate schemes, the step size restrictions for maximum-norm monotonicity and maximum principles are in general more relaxed than for the TVD property. Comparison of the theoretical results with the numerical tests indicates that the restrictions for maximum-norm monotonicity are more relevant in practice. This section also contains a discussion on local and global temporal errors for problems with smooth solutions. To understand the convergence behaviour of the schemes, the propagation of the local errors, with associated damping and cancellation effects, are to be taken into account.

2 Preliminaries

2.1 Forward Euler multirate schemes for the advection equation

2.1.1 Examples of simple schemes

Consider as a simple example the advection equation

$$u_t + u_x = 0 \tag{2.1}$$

on a one-dimensional spatial region $0 < x < 1$ with given initial value $u(x,0)$, and inflow boundary condition $u(0,t)$ or spatial periodicity. Spatial discretization is performed with the first-order upwind scheme on cells $\mathcal{C}_j = (x_j - \frac{1}{2}\Delta x_j, x_j + \frac{1}{2}\Delta x_j)$. This gives a semi-discrete system

$$u_j'(t) = \frac{1}{\Delta x_j}\big(u_{j-1}(t) - u_j(t)\big) \qquad \text{for } j \in \mathcal{I} = \{1, 2, \ldots, m\}\,, \tag{2.2}$$

where $u_j'(t) = \frac{d}{dt}u_j(t)$, and $u_j(t)$ approximates $u(x_j,t)$ or the average value over the surrounding cell $\mathcal{C}_j$.

Application of the forward Euler method with time step Δt gives the CFL stability condition $\nu_j \le 1$ for all j, where $\nu_j = \Delta t/\Delta x_j$ is the local Courant number. Suppose this stability condition is satisfied for $j \in \mathcal{I}_1$ but on $\mathcal{I}_2 = \mathcal{I} - \mathcal{I}_1$ we need to take two smaller steps with step size $\frac{1}{2}\Delta t$ to reach $t_{n+1} = t_n + \Delta t$.

Then for this simple situation, the scheme of Osher and Sanders [18] can be written as

$$u_j^{n+\frac{1}{2}} = \begin{cases} u_j^n & \text{for } j \in \mathcal{I}_1\,, \\ u_j^n + \frac{1}{2}\nu_j(u_{j-1}^n - u_j^n) & \text{for } j \in \mathcal{I}_2\,, \end{cases} \tag{2.3a}$$

$$u_j^{n+1} = u_j^n + \tfrac{1}{2}\nu_j(u_{j-1}^n - u_j^n) + \tfrac{1}{2}\nu_j(u_{j-1}^{n+\frac{1}{2}} - u_j^{n+\frac{1}{2}}) \quad \text{for } j \in \mathcal{I}\,. \tag{2.3b}$$

As observed in [26], the scheme (2.3) is not consistent at the interface: if $i-1 \in \mathcal{I}_1$ and $i \in \mathcal{I}_2$ then

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

which is consistent for fixed Courant number ν_i with the equation

$$u_t + (1 - \tfrac{1}{4}\nu_i)u_x = \mathcal{O}(\Delta t) + \mathcal{O}(\Delta x_i)\,,$$

rather than the original advection equation (2.1).

To overcome this inconsistency, Tang and Warnecke [26] therefore proposed the modified scheme

$$u_j^{n+\frac{1}{2}} = u_j^n + \tfrac{1}{2}\nu_j(u_{j-1}^n - u_j^n) \quad \text{for } j \in \mathcal{I}\,, \tag{2.4a}$$

$$u_j^{n+1} = u_j^{n+\frac{1}{2}} + \begin{cases} \tfrac{1}{2}\nu_j(u_{j-1}^n - u_j^n) & \text{for } j \in \mathcal{I}_1\,, \\ \tfrac{1}{2}\nu_j(u_{j-1}^{n+\frac{1}{2}} - u_j^{n+\frac{1}{2}}) & \text{for } j \in \mathcal{I}_2\,. \end{cases} \tag{2.4b}$$

This scheme, however, is not mass conserving at the interface. If $i-1 \in \mathcal{I}_1$ and $i \in \mathcal{I}_2$ then the flux at $x_{i-1/2}$ that leaves cell $\mathcal{C}_{i-1}$ over the time interval $[t_n, t_{n+1}]$ equals u_{i-1}^n, whereas the flux that enters $\mathcal{C}_i$ is given by $\frac{1}{2}(u_{i-1}^n + u_{i-1}^{n+1/2})$.

It should be noted that except for interface points the schemes (2.3) and (2.4) are identical. For example, if $\mathcal{I}_1 = \{j : j < i\}$ and $\mathcal{I}_2 = \{j : j \geq i\}$, then (2.3) and (2.4) give in one step the same result for $j \neq i$. It will be shown next that, also with larger interface regions, the properties of internal consistency and mass conservation cannot be combined.

2.1.2 Incompatibility of consistency and mass conservation

Consider the first-order upwind discretization (2.2) for the advection equation with spatial periodicity. Then

$$M = \sum_{j \in \mathcal{I}} \Delta x_j u_j(t)\,.$$

is a conserved quantity. If the u_j are densities, this is global mass conservation.

Now suppose that for $j \leq k_1$ we use forward Euler with step size Δt, for $j > k_2$ we apply forward Euler with step size $\frac{1}{2}\Delta t$, and on the interface region $k_1 < j \leq k_2$ we take any combination of a number of forward Euler steps with Δt and $\frac{1}{2}\Delta t$ together with interpolation or extrapolation. The result can be written as

$$u_j^{n+1} = \begin{cases} u_j^n + \nu_j(u_{j-1}^n - u_j^n)\,, & 1 \leq j \leq k_1\,, \\ u_j^n + \nu_j(u_{j-1}^n - u_j^n) + \nu_j^2 \displaystyle\sum_{k=1}^{m} \alpha_{jk}\, u_k^n\,, & k_1 < j \leq k_2\,, \\ u_j^n + \nu_j(u_{j-1}^n - u_j^n) + \tfrac{1}{4}\nu_j^2(u_{j-2}^n - 2u_{j-1}^n + u_j^n)\,, & k_2 < j \leq m\,, \end{cases} \tag{2.5}$$

with unspecified coefficients α_{jk}, and with $u_0 = u_m$ due to spatial periodicity. The interface at $x = 0, 1$ poses no problem here. We will show that this scheme cannot be both mass conservative and consistent, no matter how the scheme is defined on the interface region $k_1 < j \leq k_2$. For convenience it can be assumed that the spatial grid is uniform, $\nu_j = \nu = \Delta t/\Delta x$, and we set $\alpha_{jk} = 0$ for $j \leq k_1$ and $j > k_2$.

Insertion of exact solution values in the scheme gives for $k_1 < j \leq k_2$ the truncation error

$$\frac{1}{\Delta t}\big(u(x_j, t_{n+1}) - u(x_j, t_n)\big) - \frac{1}{\Delta x}\big(u(x_{j-1}, t_n) - u(x_j, t_n)\big) - \frac{\Delta t}{\Delta x^2}\sum_{k=1}^{m} \alpha_{jk} u(x_k, t_n)\,.$$

For consistency, that is, truncation error $\mathcal{O}(\Delta t)+\mathcal{O}(\Delta x)$, we obtain by Taylor expansion the conditions

$$\sum_k \alpha_{jk} = 0\,, \quad \sum_k (k-j)\alpha_{jk} = 0 \qquad \text{for } k_1 < j \le k_2\,. \tag{2.6}$$

On the other hand, we have

$$\begin{aligned}\Delta x \sum_j u_j^{n+1} - \Delta x \sum_j u_j^n \;&= \frac{\Delta t^2}{\Delta x} \sum_j \sum_k \alpha_{jk}\, u_k^n + \frac{\Delta t^2}{4\Delta x} \sum_{j>k_2} \big(u_{j-2}^n - 2u_{j-1}^n + u_j^n\big) \\ &= \frac{\Delta t^2}{\Delta x} \sum_k \Big(\sum_j \alpha_{jk}\Big) u_k^n + \frac{\Delta t^2}{4\Delta x} u_{k_2-1}^n - \frac{\Delta t^2}{4\Delta x} u_{k_2}^n\,,\end{aligned}$$

from which it seen that the requirement of mass conservation leads to

$$\sum_j \alpha_{jk} = \begin{cases} 0 & \text{if } k \ne k_2-1, k_2\,, \\ -\frac{1}{4} & \text{if } k = k_2-1\,, \\ \frac{1}{4} & \text{if } k = k_2\,. \end{cases} \tag{2.7}$$

However, the conditions (2.6) and (2.7) together lead to a contradiction:

$$\begin{aligned} 0 \;&=\; \sum_j \sum_k (k-j)\alpha_{jk} \;=\; \sum_j \sum_k \big((k-k_2+1)-(j-k_2+1)\big)\alpha_{jk} \\ &= \sum_j (j-k_2+1) \sum_k \alpha_{jk} - \sum_k (k-k_2+1) \sum_j \alpha_{jk} \;=\; \sum_j \alpha_{jk_2} \;=\; \tfrac{1}{4}\,. \end{aligned}$$

This shows that consistency and mass conservation cannot be valid at the same time.

2.2 General formulations

In this paper we will discuss monotonicity properties and temporal convergence of multirate schemes for general semi-discrete problems in $\mathbb{R}^m$,

$$u'(t) = F(u(t))\,, \qquad u(0) = u_0\,. \tag{2.8}$$

The approximations to $u(t_n) = [u_j(t_n)] \in \mathbb{R}^m$ will be denoted by $u_n = [u_j^n] \in \mathbb{R}^m$. As above, we consider partitioning $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2$. Corresponding to these sets $\mathcal{I}_k$, let I_1, I_2 be $m \times m$ diagonal matrices with diagonal entries 0 or 1, such that $(I_k)_{jj} = 1$ for $j \in \mathcal{I}_k$, $k = 1, 2$. We have $I_1 + I_2 = I$, the identity matrix.

The semi-discrete system (2.2) obviously fits in this form with linear F. The general system (2.8) allows nonlinear problems and nonlinear discretizations. For such systems the Osher-Sanders scheme (2.3) becomes

$$\begin{cases} u_{n+\frac{1}{2}} = u_n + \frac{1}{2}\Delta t I_2 F(u_n)\,, \\ u_{n+1} = u_n + \frac{1}{2}\Delta t F(u_n) + \frac{1}{2}\Delta t F(u_{n+\frac{1}{2}})\,, \end{cases} \tag{2.9}$$

and the Tang-Warnecke scheme (2.4) reads

$$\begin{cases} u_{n+\frac{1}{2}} = u_n + \frac{1}{2}\Delta t F(u_n)\,, \\ u_{n+1} = u_n + \Delta t I_1 F(u_n) + \frac{1}{2}\Delta t I_2 \big(F(u_n) + F(u_{n+\frac{1}{2}})\big)\,. \end{cases} \tag{2.10}$$

In the following we will refer to (2.9) as the OS1 scheme, and to (2.10) as the TW1 scheme. We note that in [18] and [26] the number of sub-steps on the index set $\mathcal{I}_2$ was allowed to be larger than two for these schemes. More general formulations will be considered in Section 5.

2.3 Monotonicity assumptions

Consider a suitable convex function,[1] semi-norm or norm $\|v\|$ for $v = [v_j] \in \mathbb{R}^m$. Interesting examples are the maximum-norm

$$\|v\|_\infty = \max_{1\le j\le m} |v_j|\,, \tag{2.11}$$

or the total variation semi-norm

$$\|v\|_{\rm TV} = \sum_{j=1}^{m} |v_{j-1} - v_j| \qquad \text{with } v_0 = v_m\,, \tag{2.12}$$

arising from one-dimensional scalar PDEs with spatial periodicity.

The basic monotonicity assumption on the semi-discrete system that will be used in this section is

$$\|v + \tau_1 I_1 F(v) + \tfrac{1}{2}\tau_2 I_2 F(v)\| \le \|v\| \qquad \text{for all } v \in \mathbb{R}^m \text{ and } 0 \le \tau_1, \tau_2 \le \tau_0\,, \tag{2.13}$$

where $\tau_0 > 0$ is a problem dependent parameter. For the multirate schemes we shall determine factors C such that we have the monotonicity property

$$\|u_{n+1}\| \le \|u_n\| \quad \text{whenever } \Delta t \le C\tau_0\,. \tag{2.14}$$

For a given scheme, the optimal C will be called the threshold factor for monotonicity. In general, such monotonicity properties are intended to ensure that unwanted overshoots or numerical oscillations will not arise. Following [23, 24] we will call a scheme total variation diminishing (TVD) if (2.14) holds with the semi-norm (2.12). If the (semi-)norm is not specified, methods that have a positive threshold C can be called strong stability preserving (SSP), as in [5] for standard, single-rate methods.

Example 2.1 Apart from (semi-)norms, such as $\|v\|_{\rm TV}$ and $\|v\|_\infty$, we can also consider convex functions. For example, following [25], consider

$$\|v\|_+ = \max_{1\le j\le m} v_j\,, \qquad \|v\|_- = -\min_{1\le j\le m} v_j\,.$$

Then, having (2.14) for both these convex functions amounts to the maximum principle

$$\min_{1\le i\le m} u_i^0 \;\le\; u_j^n \;\le\; \max_{1\le i\le m} u_i^0 \qquad \text{for all } n \ge 1 \text{ and } 1 \le j \le m\,.$$

In general, this is of course somewhat stronger than having monotonicity in the maximum-norm, $\|u_{n+1}\|_\infty \le \|u_n\|_\infty$, but for the schemes considered in this paper the associated threshold values C will be the same. ◇

Example 2.2 Consider a scalar conservation law $u_t + f(u)_x = 0$ with a periodic boundary condition, and with $0 \le f'(u) \le \alpha$. Spatial discretization in conservation form gives semi-discrete systems (2.8) with

$$F_j(v) = \frac{1}{\Delta x_j}\big(f(v_{j-\frac12}) - f(v_{j+\frac12})\big)$$

[1] Recall that $\phi : \mathbb{R}^m \to \mathbb{R}$ is a convex function if $\phi((1-\theta)v + \theta w) \le (1-\theta)\phi(v) + \theta\phi(w)$ for all $\theta \in [0,1]$ and $v, w \in \mathbb{R}^m$. If we have $\phi(v) \ge 0$, $\phi(v+w) \le \phi(v) + \phi(w)$ and $\phi(\lambda v) = |\lambda|\phi(v)$ for all $\lambda \in \mathbb{R}$, $v, w \in \mathbb{R}^m$, then ϕ is a semi-norm. If it holds in addition that $\phi(v) = 0$ only if $v = 0$, then ϕ is a norm.

where $v_{j\pm1/2}$ are the values at the cell boundaries, determined from the components of $v = [v_i] \in \mathbb{R}^m$. Using limiters in the discretization it can be guaranteed that

$$0 \;\le\; \frac{v_{j-\frac12} - v_{j+\frac12}}{v_{j-1} - v_j} \;\le\; 1+\mu$$

with a constant $\mu \ge 0$ determined by the limiter; see also formula (8) in [4]. This holds trivially for the first-order upwind discretization with $\mu = 0$; a detailed higher-order example will be given in Appendix A. It now follows that $F_j(v)$ can be written as

$$F_j(v) = \frac{a_j(v)}{\Delta x_j}\big(v_{j-1} - v_j\big)\,, \qquad j = 1,\dots,m\,,\;\; v_0 = v_m\,,$$

where

$$0 \le a_j(v) \le \alpha(1+\mu) \qquad \text{for all } j \text{ and } v \in \mathbb{R}^m\,.$$

Suppose that $\Delta x_j = h$ for $j \in \mathcal{I}_1$ and $\Delta x_j = \frac12 h$ for $j \in \mathcal{I}_2$. Then a well-known lemma of Harten [8, Lemma 2.2] shows that (2.13) will be valid for the total variation semi-norm (2.12) provided that

$$\frac{\alpha\tau_0}{h} \;\le\; \frac{1}{1+\mu}\,.$$

Moreover, it is easy to see that (2.13) will also hold in the maximum-norm under the same CFL restriction. ◇

3 Analysis of the forward Euler multirate schemes

3.1 Monotonicity results

3.1.1 Monotonicity results for scheme TW1

Standard (single-rate) schemes give the same step size restriction for various monotonicity properties. As we shall see, with the multirate schemes different step size restrictions are obtained for the maximum-norm or the total variation semi-norm.

In the first stage of the TW1 scheme (2.10) we have of course

$$\|u_{n+\frac12}\| \le \|u_n\| \qquad \text{whenever } \Delta t \le \tau_0\,.$$

The second stage can be written in the form

$$u_{n+1} = (1-\theta)u_n + \theta\big(u_{n+\frac12} - \tfrac12\Delta t F(u_n)\big) + \Delta t I_1 F(u_n) + \tfrac12\Delta t I_2\big(F(u_n) + F(u_{n+\frac12})\big)\,,$$

with arbitrary $\theta \in [0,1]$. This leads to

$$\begin{aligned} u_{n+1} \;&=\; (1-\theta)\Big(u_n + \frac{2-\theta}{2(1-\theta)}\Delta t I_1 F(u_n) + \tfrac12\Delta t I_2 F(u_n)\Big) \\ &\quad + \theta\Big(u_{n+\frac12} + \frac{1}{2\theta}\Delta t I_2 F(u_{n+\frac12})\Big)\,. \end{aligned} \tag{3.1}$$

Under assumption (2.13) this gives the monotonicity property (2.14) with

$$C \;=\; \max_{0\le\theta\le1} \min\Big(1\,,\; \frac{2(1-\theta)}{2-\theta}\,,\; \theta\Big) \;=\; 2-\sqrt2\,. \tag{3.2}$$

This value $C \approx 0.58$ is valid for general semi-norms. So, in particular, it provides a TVD result for schemes with limiters.

Next, consider the maximum-norm. Then, by noting that the second stage can also be written as

$$u_{n+1} = I_1\big(u_n + \Delta t I_1 F(u_n)\big) + I_2\big(u_{n+\frac{1}{2}} + \tfrac{1}{2}\Delta t I_2 F(u_{n+\frac{1}{2}})\big)\,,$$

it directly follows (see also [26, Lemma 2.1]) that the threshold factor for max-norm monotonicity is

$$C = 1\,. \tag{3.3}$$

Note that this result has been obtained by using the inequality

$$\|I_1 v + I_2 w\| \le \max(\|v\|, \|w\|)\,, \tag{3.4}$$

which holds for the maximum-norm and for the convex functions $\|\cdot\|_{\pm}$ from Example 2.1, but not for general norms or semi-norms; in particular, it will not hold for the total variation semi-norm.

3.1.2 Monotonicity results for scheme OS1

In the first stage of the OS1 scheme (2.9) we directly obtain

$$\|u_{n+\frac{1}{2}}\| \le \|u_n\| \qquad \text{whenever } \Delta t \le \tau_0\,.$$

The second stage can be written as

$$u_{n+1} = (1-\theta)u_n + \theta\big(u_{n+\frac{1}{2}} - \tfrac{1}{2}\Delta t I_2 F(u_n)\big) + \tfrac{1}{2}\Delta t F(u_n) + \tfrac{1}{2}\Delta t F(u_{n+\frac{1}{2}})$$

with parameter $\theta \in [0,1]$. Hence

$$\begin{aligned} u_{n+1} \;=\; & (1-\theta)\Big(u_n + \frac{1}{2(1-\theta)}\Delta t I_1 F(u_n) + \tfrac{1}{2}\Delta t I_2 F(u_n)\Big) \\ & + \theta\Big(u_{n+\frac{1}{2}} + \frac{1}{2\theta}\Delta t F(u_{n+\frac{1}{2}})\Big)\,. \end{aligned} \tag{3.5}$$

It follows that under assumption (2.13) the monotonicity property (2.14) holds with

$$C \;=\; \max_{0\le\theta\le 1} \min\big(1\,,\,2(1-\theta)\,,\,\theta\big) \;=\; \tfrac{2}{3}\,. \tag{3.6}$$

Again, for the maximum-norm a better result can be obtained by considering $I_1 u_{n+1}$ and $I_2 u_{n+1}$ separately. Multiplication of (3.5) with I_1 and taking $\theta = \theta_1 = \frac{1}{2}$ gives

$$I_1 u_{n+1} = \tfrac{1}{2} I_1\big(u_n + \Delta t I_1 F(u_n)\big) + \tfrac{1}{2} I_1\big(u_{n+\frac{1}{2}} + \Delta t I_1 F(u_{n+\frac{1}{2}})\big)\,.$$

Likewise, with $\theta = \theta_2 = 1$, it follows that

$$I_2 u_{n+1} = I_2\big(u_{n+\frac{1}{2}} + \tfrac{1}{2}\Delta t I_2 F(u_{n+\frac{1}{2}})\big)\,.$$

Hence the threshold factor for max-norm monotonicity is

$$C = 1\,. \tag{3.7}$$

This result, formulated in terms of a maximum principle, was already obtained in [18] for first-order upwind spatial discretization and in [15] for a class of high-resolution discretizations. In these papers also TVD results were presented; this will be discussed below.

3.1.3 The TVD property for linear first-order upwind advection

For the linear advection equation $u_t + u_x = 0$ with spatial periodicity, the first-order upwind discretization (2.2) can be written as

$$u'(t) = Au(t)\,, \qquad A = H^{-1}(E - I)\,, \tag{3.8}$$

with $H = \mathrm{diag}(\Delta x_1, \ldots, \Delta x_m)$ and E the backward shift operator, $(Ev)_i = v_{i-1}$ for $i = 1, \ldots, m$ with $v_0 = v_m$. Consider also

$$\tilde{A} = H^{-1}(-I + E^T)\,.$$

This corresponds to first-order upwind discretization for $u_t - u_x = 0$. We denote $Z = \Delta t A$, $\tilde{Z} = \Delta t \tilde{A}$. Then

$$\tilde{Z} = H^{-1} Z^T H\,.$$

For the OS1 and TW1 schemes applied to (3.8) we have $u_{n+1} = Su_n$, where the amplification matrix S can be written as $S = R(Z)$ with

$$R(Z) = \begin{cases} R_{\text{OS1}}(Z) = I + Z + \frac{1}{4} Z I_2 Z\,, \\ R_{\text{TW1}}(Z) = I + Z + \frac{1}{4} I_2 Z^2\,. \end{cases}$$

Let $\tilde{R}$ be such that

$$\tilde{R}(Z)\, Z = Z\, R(Z)\,. \tag{3.9}$$

It is easily seen that $\tilde{R}_{\text{OS1}}(Z) = I + Z + \frac{1}{4} Z^2 I_2$ and $\tilde{R}_{\text{TW1}}(Z) = I + Z + \frac{1}{4} Z I_2 Z$. For both schemes it follows by some simple calculations that

$$R(\tilde{Z}) = H^{-1} \tilde{R}(Z)^T H\,. \tag{3.10}$$

As we saw above, both schemes OS1 and TW1 are such that

$$\|R(\tilde{Z})\|_\infty \le 1 \tag{3.11}$$

whenever $\nu_j = \Delta t / \Delta x_j \le k$ for $j = \mathcal{I}_k$, $k = 1, 2$. It will now be demonstrated that under the same CFL restriction we have

$$\|R(Z)v\|_{\text{TV}} \le \|v\|_{\text{TV}} \qquad \text{for all } v \in \mathbb{R}^m, \tag{3.12}$$

that is, the TVD property is valid with threshold $C = 1$ for the special case of first-order upwind advection discretization.

Lemma 3.1 *If (3.10) and (3.11) are valid, then (3.12) is also satisfied.*

Proof. Along with the discrete L_1-norm on $\mathbb{R}^m$, $\|v\|_1 = \sum_{j=1}^m \Delta x_j |v_j|$, we also consider the ℓ_1-norm $\|v\|_{\ell_1} = \sum_{j=1}^m |v_j|$, together with the induced matrix norms. Then we have $\|W\|_\infty = \|W^T\|_{\ell_1}$ for any $W \in \mathbb{R}^{m \times m}$; see for example [12]. Moreover, it is easily seen that $\|W^T\|_{\ell_1} = \|H^{-1} W^T H\|_1$, and therefore

$$\|W\|_\infty = \|H^{-1} W^T H\|_1\,.$$

Hence (3.10) and (3.11) imply

$$\|\tilde{R}(Z)\|_1 \le 1\,. \tag{3.13}$$

Further we have

$$\|v\|_{\mathrm{TV}} = \sum_{j=1}^{m} |v_{j-1} - v_j| = \|Av\|_1 = \frac{1}{\Delta t}\|Zv\|_1 .$$

Consequently, for a scheme $u_{n+1} = R(Z)u_n$ the TVD property (3.12) is equivalent to

$$\|ZR(Z)v\|_1 = \|\tilde{R}(Z)Zv\|_1 \le \|Zv\|_1 .$$

This is satisfied because $\|\tilde{R}(Z)w\|_1 \le \|w\|_1$ for any $w \in \mathbb{R}^m$, in view of (3.13). □

The above result is not new for the OS1 scheme. In fact, already in [18] the result was given for the case of first-order upwind discretization for non-linear problems. In [15] this was extended to a class of high-resolution spatial discretizations. The proofs of these more general results for the OS1 scheme are more technical than the above.

3.2 Convergence for smooth problems

In this section bounds for the global errors $e_n = u(t_n) - u_n$ will be derived. It will be assumed that the problem (2.8) is sufficiently smooth. Both the schemes OS1 and TW1 are covered by the formula

$$\begin{aligned} u_{n+\frac{1}{2}} &= u_n + \kappa\Delta t I_1 F(u_n) + \tfrac{1}{2}\Delta t I_2 F(u_n) , \\ u_{n+1} &= u_n + \tfrac{1}{2}\Delta t\big(F(u_n) + F(u_{n+\frac{1}{2}})\big) + \kappa\Delta t I_1\big(F(u_n) - F(u_{n+\frac{1}{2}})\big) , \end{aligned} \tag{3.14}$$

with parameter value $\kappa = 0$ for OS1 and $\kappa = \frac{1}{2}$ for TW1.

If we insert exact ODE values $u(t_n)$, $u(t_{n+1/2})$, $u(t_{n+1})$ into the stages of (3.14) we obtain residuals $\rho_{n+1/2}$ and ρ_{n+1}, respectively. By Taylor expansions it is easily found that

$$\begin{aligned} \rho_{n+\frac{1}{2}} &= u(t_{n+\frac{1}{2}}) - u(t_n) - \kappa\Delta t I_1 u'(t_n) - \tfrac{1}{2}\Delta t I_2 u'(t_n) \\ &= \big(\tfrac{1}{2} - \kappa\big)\Delta t I_1 u'(t_n) + \tfrac{1}{8}\Delta t^2 u''(t_n) + \mathcal{O}(\Delta t^3) , \end{aligned}$$

$$\begin{aligned} \rho_{n+1} &= u(t_{n+1}) - u(t_n) - \big(\tfrac{1}{2}I + \kappa I_1\big)\Delta t u'(t_n) - \big(\tfrac{1}{2}I - \kappa I_1\big)\Delta t u'(t_{n+\frac{1}{2}}) \\ &= \Delta t^2\big(\tfrac{1}{4}I + \tfrac{1}{2}\kappa I_1\big)u''(t_n) + \mathcal{O}(\Delta t^3) . \end{aligned}$$

Let $Z_\ell \in \mathbb{R}^{m\times m}$ be such that

$$Z_\ell\big(u(t_\ell) - u_\ell\big) = \Delta t\big(F(u(t_\ell)) - F(u_\ell)\big) \tag{3.15}$$

for all $\ell = n, n+\frac{1}{2}$, $n \ge 0$. If F is differentiable we can take Z_ℓ as the integrated Jacobian matrix

$$Z_\ell = \int_0^1 \Delta t F'(\theta u(t_\ell) + (1-\theta)u_\ell)\, d\theta .$$

For the errors in the two stages of (3.14) it follows that

$$\begin{aligned} e_{n+\frac{1}{2}} &= e_n + \kappa I_1 Z_n e_n + \tfrac{1}{2} I_2 Z_n e_n + \rho_{n+\frac{1}{2}} , \\ e_{n+1} &= e_n + \tfrac{1}{2} Z_n e_n + \tfrac{1}{2} Z_{n+\frac{1}{2}} e_{n+\frac{1}{2}} + \kappa I_1\big(Z_n e_n - Z_{n+\frac{1}{2}} e_{n+\frac{1}{2}}\big) + \rho_{n+1} . \end{aligned}$$

Eliminating $e_{n+1/2}$ we thus obtain a recursion for the global errors of the form

$$e_{n+1} = S_n e_n + d_n , \qquad n = 0, 1, \ldots , \tag{3.16}$$

with amplification matrix S_n and local discretization error d_n. The resulting expressions are given below for $\kappa = 0, \frac{1}{2}$. The recursion (3.16) will be the basis for the subsequent analysis. The method is called *consistent* of order p if $\|d_n\| = \mathcal{O}(\Delta t^{p+1})$, and *convergent* of order p if $\|e_n\| = \mathcal{O}(\Delta t^p)$ for all n.

Since we want to study convergence at all grid points, including the interface points, the natural norm is the maximum-norm. For stability it will be assumed that

$$\|I + I_1 Z_\ell + \tfrac{1}{2} I_2 Z_\ell\|_\infty \le 1\,, \tag{3.17}$$

for all $\ell = n, n+\frac{1}{2}$. It is easily seen that we then have $\|I + \theta_1 I_1 Z_\ell + \frac{1}{2}\theta_2 I_2 Z_\ell\|_\infty \le 1$ whenever $0 \le \theta_j \le 1$. This is of the same form as (2.13), with $F(v)$ replaced by $Z_\ell v$.

In combination with the smoothness assumptions on the problem this stability result will easily lead to convergence for the TW1 scheme. Due to the inconsistency at interface points, the error build-up is more complicated for scheme OS1. It will still be possible to show convergence with order one under the following additional assumptions:

$$\|I_2 Z_\ell\|_\infty \le 4K < 4\,, \tag{3.18}$$

$$\|Z_{\ell+\frac{1}{2}} - Z_\ell\|_\infty \le L\Delta t\,, \tag{3.19}$$

for $\ell = n, n+\frac{1}{2}$, $n \ge 0$, with constants $K \in (0,1)$ and $L \ge 0$. Note that (3.18) may be slightly stronger than the local CFL condition implied by (3.17) on the index set $\mathcal{I}_2$.

3.2.1 Convergence of scheme TW1

For the TW1 scheme (2.10) we obtain from the above derivation, with $\kappa = \frac{1}{2}$, the expressions

$$S_n = I_1\big(I + Z_n\big) + I_2\big(I + \tfrac{1}{2} Z_{n+\frac{1}{2}}\big)\big(I + \tfrac{1}{2} Z_n\big)\,, \tag{3.20}$$

$$d_n = \tfrac{1}{2}\Delta t^2\big(I_1 + \tfrac{1}{2} I_2 + \tfrac{1}{8} I_2 Z_{n+\frac{1}{2}}\big) u''(t_n) + \mathcal{O}(\Delta t^3)\,. \tag{3.21}$$

As already noted above, (3.17) has the same form as (2.13). Therefore we can copy the derivation leading to (3.3) which now gives the bound

$$\|S_n\|_\infty \le 1 \tag{3.22}$$

for the amplification matrix.

Furthermore, (3.17) implies $\|I_2(I + \frac{1}{4} Z_\ell)\|_\infty \le 1$, which provides the local error bound

$$\|d_n\|_\infty \le \tfrac{1}{2}\Delta t^2 \|u''(t_n)\|_\infty + \mathcal{O}(\Delta t^3)\,.$$

Convergence now follows in a standard fashion. Summarizing, we have the following result:

Theorem 3.2 *Consider the TW1 scheme (2.10) with the time step restriction (3.17). Then $\|S\|_\infty \le 1$, and we have the error bound*

$$\|e_n\|_\infty \;\le\; \tfrac{1}{2} T \Delta t \max_{t\in[0,T]} \|u''(t)\|_\infty + \mathcal{O}(\Delta t^2)\,, \qquad 0 \le t_n \le T\,.$$

3.2.2 Convergence of scheme OS1

Also for the OS1 scheme (2.9) we can prove convergence with order one in the maximum-norm, in spite of the local inconsistencies. For this result, damping and cancellation effects are to be taken into account.

For the OS1 scheme we obtain from the above derivation, with $\kappa = 0$, the expressions

$$S_n = I + \tfrac{1}{2}Z_n + \tfrac{1}{2}Z_{n+\frac{1}{2}}\big(I + \tfrac{1}{2}I_2 Z_n\big)\,, \tag{3.23}$$

$$d_n = \tfrac{1}{4}\Delta t Z_{n+\frac{1}{2}} I_1 u'(t_n) + \tfrac{1}{4}\Delta t^2\big(I + \tfrac{1}{4}Z_{n+\frac{1}{2}}\big)u''(t_n) + \mathcal{O}(\Delta t^3)\,. \tag{3.24}$$

In the same way as above it follows that (3.22) is valid, showing stability of the error recursion. However, here we get only an $\mathcal{O}(\Delta t)$ bound for the local errors because $Z_\ell I_1 u'(t_n)$ will not be an $\mathcal{O}(\Delta t)$ term in general; this is due to the fact that $I_1 u'(t)$ is not a smooth grid function (jumps at the interfaces). To prove convergence we need to establish a relation between local errors and amplification factors.

We have

$$S_n - I = Z_{n+\frac{1}{2}}\big(I + \tfrac{1}{4}I_2 Z_n\big) - \tfrac{1}{2}\big(Z_{n+\frac{1}{2}} - Z_n\big)\,.$$

Hence

$$Z_{n+\frac{1}{2}} = (S_n - I)Q_n + \tfrac{1}{2}\big(Z_{n+\frac{1}{2}} - Z_n\big)Q_n\,, \qquad Q_n = \big(I + \tfrac{1}{4}I_2 Z_n\big)^{-1}\,.$$

It follows that we can decompose the local error as

$$d_n = (S_n - I)\xi_n + \eta_n\,, \tag{3.25}$$

with

$$\begin{aligned} \xi_n &= \tfrac{1}{4}\Delta t Q_n I_1 u'(t_n)\,, \\ \eta_n &= \tfrac{1}{8}\Delta t\big(Z_{n+\frac{1}{2}} - Z_n\big)Q_n I_1 u'(t_n) + \tfrac{1}{4}\Delta t^2\big(I + \tfrac{1}{4}Z_{n+\frac{1}{2}}\big)u''(t_n) + \mathcal{O}(\Delta t^3)\,. \end{aligned} \tag{3.26}$$

Such a decomposition can be used to show convergence for scheme OS1; the arguments are the same as in [14, p. 216] for constant $S_n = S$. Let us define $\hat{e}_n = e_n + \xi_n$ for $n \geq 0$. Then

$$\hat{e}_{n+1} = S_n \hat{e}_n + \hat{d}_n\,, \qquad \hat{d}_n = \xi_{n+1} - \xi_n + \eta_n\,,$$

for $n \geq 0$. Hence

$$\|\hat{e}_n\|_\infty \leq \|\hat{e}_0\|_\infty + \sum_{k=0}^{n} \|\hat{d}_k\|_\infty\,.$$

Since $e_0 = 0$ we obtain

$$\|e_n\|_\infty \leq \|\xi_0\|_\infty + \|\xi_n\|_\infty + \sum_{k=0}^{n}\big(\|\xi_{k+1} - \xi_k\|_\infty + \|\eta_k\|_\infty\big)\,. \tag{3.27}$$

It remains to bound the terms on the right-hand side. Under assumption (3.18) it is easily seen that

$$\|Q_k\|_\infty \leq (1 - K)^{-1}\,.$$

Moreover, we have

$$Q_{k+1} - Q_k = -\tfrac{1}{4}Q_k(I_2 Z_{k+1} - I_2 Z_k)Q_{k+1}\,,$$

$$\|Q_{k+1} - Q_k\|_\infty \le \tfrac{1}{2}\Delta t L (1-K)^{-2}\,.$$

It follows that

$$\|\xi_k\|_\infty \le \tfrac{1}{4}(1-K)^{-1}\Delta t \|u'(t_k)\|_\infty\,,$$
$$\|\xi_{k+1} - \xi_k\|_\infty \le \tfrac{1}{8}(1-K)^{-2} L \Delta t^2 \|u'(t_k)\|_\infty + \tfrac{1}{4}(1-K)^{-1}\Delta t^2 \|u''(t_k)\|_\infty + \mathcal{O}(\Delta t^3)\,,$$
$$\|\eta_k\|_\infty \le \tfrac{1}{8}(1-K)^{-1} L \Delta t^2 \|u'(t_k)\|_\infty + \tfrac{1}{4}\Delta t^2 \|u''(t_k)\|_\infty\,.$$

Insertion of these three estimates into (3.27) gives the following convergence result.

Theorem 3.3 *Consider the OS1 scheme (2.9) with the time step restriction (3.17). Then* $\|S\|_\infty \le 1$. *Under the additional assumption (3.18), (3.19) we have the error bound*

$$\|e_n\|_\infty \;\le\; (M_1 + M_2 T L)\Delta t \max_{t\in[0,T]} \|u'(t)\|_\infty + M_3 T \Delta t \max_{t\in[0,T]} \|u''(t)\|_\infty + \mathcal{O}(\Delta t^2)\,,$$

for $0 \le t_n \le T$, *with* M_1, M_2, M_3 *determined by* K.

3.2.3 Convergence of OS1 for linear first-order upwind advection

Consider the first-order upwind discretization (2.2) for linear advection. Then (3.17) will hold if

$$\frac{\Delta t}{\Delta x_j} \le 1 \quad \text{for } j \in \mathcal{I}_1\,, \qquad \frac{\Delta t}{2\Delta x_j} \le 1 \quad \text{for } j \in \mathcal{I}_2\,.$$

These are the usual restrictions on the local Courant numbers. To have (3.18) we get the restriction

$$\frac{\Delta t}{2\Delta x_j} \le K < 1 \quad \text{for } j \in \mathcal{I}_2\,.$$

However, for this first-order upwind advection case the condition (3.18) with $K < 1$ is not needed. Let $Z = \Delta t A$ with A as in (3.8). Suppose for simplicity that $\mathcal{I}_1 = \{j : j < i\}$, $\mathcal{I}_2 = \{j : j \ge i\}$ with given $i \in \mathcal{I}$. Consider

$$(S - I)\xi = Z I_1 v\,,$$

where $\xi = \xi_n$ and $v = v_n = \frac{1}{4}\Delta t u'(t_n)$ in the local error decomposition (3.25). The vector ξ will satisfy this relation if $(I + \frac{1}{4} I_2 Z)\xi = I_1 v$, that is

$$I_1 \xi = I_1 v\,, \qquad I_2\big(I + \tfrac{1}{4} Z\big)\xi = 0\,.$$

It is seen that $\xi = [\xi_j] \in \mathbb{R}^m$ is given by

$$\xi_j = v_j \quad (\text{for } j < i)\,, \qquad \xi_{i+k} = \Big(\frac{\nu_j}{\nu_j - 4}\Big)^{k+1} v_{i-1} \quad (\text{for } k \ge 0)\,,$$

where $\nu_j = \Delta t/\Delta x_j$. Therefore $\|\xi\|_\infty \le \|v\|_\infty$ if $\nu_j \le 2$ on $\mathcal{I}_2$.

It follows that for this linear advection case, the local error decomposition (3.25) will be valid under (3.17), with $\|\xi_n\|_\infty = \mathcal{O}(\Delta t)$, $\|\xi_{n+1} - \xi_n\|_\infty = \mathcal{O}(\Delta t^2)$, and with $\|\eta_n\|_\infty = \mathcal{O}(\Delta t^2)$ containing the higher-order terms in the local error, leading to convergence with order one.

4 Second-order schemes

In the literature, several second-order multirate schemes for conservation laws have been derived that are based on the standard two-stage Runge-Kutta method

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

The second stage can also be written as

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

Monotonicity properties are more clear with this form. The method is known as the explicit trapezoidal rule or the modified Euler method. In this section we consider some multirate schemes, based on this method, with one level of temporal refinement. Results on internal consistency and mass conservation are mentioned here, but a detailed discussion will only be given in Section 5.

The second-order scheme of Tang & Warnecke [26] reads

$$\begin{cases} u^*_{n+\frac{1}{2}} = u_n + \frac{1}{2}\Delta t F(u_n)\,, \\ u_{n+\frac{1}{2}} = \frac{1}{2}\big(u_n + u^*_{n+\frac{1}{2}} + \frac{1}{2}\Delta t F(u^*_{n+\frac{1}{2}})\big)\,, \\ u^*_{n+1} = I_1\big(u_n + \Delta t F(u_n)\big) + I_2\big(u_{n+\frac{1}{2}} + \frac{1}{2}\Delta t F(u_{n+\frac{1}{2}})\big)\,, \\ u_{n+1} = \frac{1}{2}I_1\big(u_n + u^*_{n+1} + \Delta t F(u^*_{n+1})\big) + \frac{1}{2}I_2\big(u_{n+\frac{1}{2}} + u^*_{n+1} + \frac{1}{2}\Delta t F(u^*_{n+1})\big)\,. \end{cases} \tag{4.1}$$

We will refer to this scheme as TW2. It will be shown below that this scheme is internally consistent but not mass-conserving.

Constantinescu & Sandu [3] introduced the following scheme, which will be referred to as CS2,

$$\begin{cases} u^*_{n+\frac{1}{2}} = u_n + \Delta t I_1 F(u_n) + \frac{1}{2}\Delta t I_2 F(u_n)\,, \\ u_{n+\frac{1}{2}} = u_n + \frac{1}{4}\Delta t I_2\big(F(u_n) + F(u^*_{n+\frac{1}{2}})\big)\,, \\ u^*_{n+1} = I_1\big(u_n + \Delta t I_1 F(u_{n+\frac{1}{2}})\big) + I_2\big(u_{n+\frac{1}{2}} + \frac{1}{2}F(u_{n+\frac{1}{2}})\big)\,, \\ u_{n+1} = u_n + \frac{1}{4}\Delta t\big(F(u_n) + F(u^*_{n+\frac{1}{2}}) + F(u_{n+\frac{1}{2}}) + F(u^*_{n+1})\big)\,. \end{cases} \tag{4.2}$$

This scheme is mass-conserving but not internally consistent. Nevertheless, we will see that it is still convergent (with order one) in the maximum-norm due to damping and cancellation effects. Note that for non-stiff ODE systems the scheme will be consistent and convergent with order two.

The related method of Dawson and Kirby [4] is also mass-conserving but not internally consistent. However in that scheme a limiter is applied which is adapted to the outcome of previous stages, so it does not fit in the framework of this paper where the semi-discrete system is supposed to be given a priori.

In Savcenco [21] several other multirate schemes of order two can be found for stiff (parabolic) problems. These are Rosenbrock-type schemes that contain a parameter γ, and setting $\gamma = 0$ yields an explicit scheme. We consider here the scheme that was introduced in [22]; it will be referred to as SHV2. In this scheme, first a prediction $\bar{u}_{n+1}$ is computed, followed by refinement steps on $\mathcal{I}_2$ using interpolated values $\bar{u}_{n+1/2}$

on $\mathcal{I}_1$. The scheme reads

$$\begin{cases} \bar{u}^*_{n+1} = u_n + \Delta t F(u_n)\,, \\ \bar{u}_{n+1} = \frac{1}{2}u_n + \frac{1}{2}\bar{u}^*_{n+1} + \frac{1}{2}\Delta t F(\bar{u}^*_{n+1})\,, \\ \bar{u}_{n+\frac{1}{2}} = \frac{1}{2}u_n + \frac{1}{4}\bar{u}_{n+1} + \frac{1}{4}\bar{u}^*_{n+1}\,, \\ u^*_{n+\frac{1}{2}} = I_1\bar{u}_{n+\frac{1}{2}} + I_2\big(u_n + \frac{1}{2}\Delta t F(u_n)\big)\,, \\ u_{n+\frac{1}{2}} = I_1\bar{u}_{n+\frac{1}{2}} + I_2\big(\frac{1}{2}u_n + \frac{1}{2}u^*_{n+\frac{1}{2}} + \frac{1}{4}\Delta t F(u^*_{n+\frac{1}{2}})\big)\,, \\ u^*_{n+1} = I_1\bar{u}_{n+1} + I_2\big(u_{n+\frac{1}{2}} + \frac{1}{2}\Delta t F(u_{n+\frac{1}{2}})\big)\,, \\ u_{n+1} = I_1\bar{u}_{n+1} + I_2\big(\frac{1}{2}u_{n+\frac{1}{2}} + \frac{1}{2}u^*_{n+1} + \frac{1}{4}\Delta t F(u^*_{n+1})\big)\,. \end{cases} \tag{4.3}$$

This scheme will be seen to be internally consistent but not mass-conserving. We note that (4.3) could be written with fewer stages; there are no function evaluations of $\bar{u}_{n+1}$ and $\bar{u}_{n+\frac{1}{2}}$, so these vectors are just included for notational convenience. Further we note that this scheme was not intended originally as used here. Instead, the prediction values $\bar{u}^*_{n+1}$ and $\bar{u}_{n+1}$ were used in [22] to estimate local errors, and based on this estimate the partitioning $\mathcal{I} = \mathcal{I}_1 \cup \mathcal{I}_2$ was adjusted. For the schemes in the present paper the partitioning is supposed to be given, based on local Courant numbers.

The interpolation step in (4.3) can be written as

$$\bar{u}_{n+\frac{1}{2}} = \frac{3}{4}u_n + \frac{1}{4}\bar{u}_{n+1} + \frac{1}{4}\Delta t F(u_n)\,, \tag{4.4}$$

which corresponds to quadratic Hermite interpolation. As an alternative we can also consider linear interpolation

$$\bar{u}_{n+\frac{1}{2}} = \frac{1}{2}u_n + \frac{1}{2}\bar{u}_{n+1}\,, \tag{4.5}$$

but in the numerical tests (4.4) gave somewhat better results (errors approximately 5% smaller) in general.

In practical applications, for systems of conservation laws, evaluation of the function components $F_j(v)$ will be the main computational work. Note that if $I_k F(v)$ is needed then v should be known on $\mathcal{I}_k$ and on a few additional points near the interface (how many points depends on the stencil of the spatial discretization). If we ignore these interface points, and assume that $\mathcal{I}_k$ contains m_k points, $m_1 + m_2 = m$, then we can easily estimate the amount of work per step with the schemes. For the schemes TW2 and SHV2 this is $2(m + m_2)\mu_{\mathrm{w}}$, and for the CS2 scheme it is $4m\mu_{\mathrm{w}}$, where μ_{w} is the measure of work for a single component $F_j(v)$. Therefore, if $m_2 \ll m_1$, that is, temporal refinement is only needed at few points, then the CS2 scheme will be approximately twice as expensive as the other two schemes.

4.1 Numerical tests

An analysis of the above second-order schemes will be given in the next section in the framework of partitioned Runge-Kutta methods. Here we already present some numerical results that will serve as benchmarks for the analysis.

4.1.1 Linear advection with smooth solution

As a first test on the accuracy of the schemes we consider the linear advection equation (2.1) on the spatial interval $0 < x < 1$ with periodic boundary conditions, and time interval $0 < t \le T = 1$. For test purposes a uniform spatial grid is taken, so that interface

Table 1: Results for the smooth advection problem with the CS2, TW2 and SHV2 schemes. Maximum errors and L_1-errors at final time $t_N = T$ for various m with fixed Courant number $\nu = 0.4$.

m	100	200	400	800
CS2, $\|e_N\|_\infty$	$1.97 \cdot 10^{-3}$	$5.64 \cdot 10^{-4}$	$1.88 \cdot 10^{-4}$	$9.96 \cdot 10^{-5}$
CS2, $\|e_N\|_1$	$7.11 \cdot 10^{-4}$	$1.84 \cdot 10^{-4}$	$4.85 \cdot 10^{-5}$	$1.28 \cdot 10^{-5}$
TW2, $\|e_N\|_\infty$	$6.08 \cdot 10^{-4}$	$1.57 \cdot 10^{-4}$	$3.98 \cdot 10^{-5}$	$9.99 \cdot 10^{-6}$
TW2, $\|e_N\|_1$	$2.85 \cdot 10^{-4}$	$7.35 \cdot 10^{-5}$	$1.86 \cdot 10^{-5}$	$4.66 \cdot 10^{-6}$
SHV2, $\|e_N\|_\infty$	$6.10 \cdot 10^{-4}$	$1.57 \cdot 10^{-4}$	$3.95 \cdot 10^{-5}$	$9.90 \cdot 10^{-6}$
SHV2, $\|e_N\|_1$	$2.91 \cdot 10^{-4}$	$7.40 \cdot 10^{-5}$	$1.86 \cdot 10^{-5}$	$4.66 \cdot 10^{-6}$

effects are certainly not due to the spatial discretization, for which the WENO5 scheme is chosen; the formulas for this discretization can be found for example in [23]. Temporal refinement is used at the union of spatial intervals $\mathcal{D}_k = \{x : |x - k/10| \leq 1/40\}$, $k = 1, \ldots, 9$, and we consider a fixed Courant number $\nu = \Delta t/\Delta x = 0.4$.

For this accuracy test a smooth solution $u(x,t) = \sin^2(\pi(x-t))$ is considered. The errors in the maximum-norm and discrete L_1-norm ($\|v\|_1 = \sum_j \Delta x_j |v_j|$) are presented in Table 1. It is seen that with the CS2 scheme we have only first-order convergence in the maximum-norm, due to the interface points; the L_1-errors are still second-order. For the schemes TW2 and SHV2 we see an order two convergence also in the maximum-norm. The entries in Table 1 are the total (absolute) errors with respect to the PDE solution, but it was verified that the spatial errors are much smaller here than the temporal errors.

To see that the large errors for scheme CS2 in the maximum-norm are indeed caused by the interface points, the errors as function of x at the final time with $m = 800$ are displayed in Figure 1. The (relatively) large errors for CS2 at the interface points are clearly visible. For scheme TW2 there are no visible interface effects. The errors for SHV2 are almost the same as for TW2.

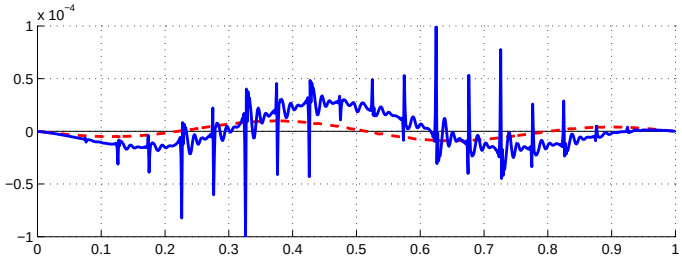

Figure 1: Errors versus $x_j \in (0,1)$ at final time $t_N = T$ for the schemes CS2 (thick solid line) and TW2 (thick dashed line), $m = 800$.

The CS2 scheme is not internally consistent at the interfaces, but we see in this test that it is still convergent. This is similar as with the OS1 scheme.

The linear advection test was repeated with an initial block-function with the aim of seeing the effect of the lack of mass-conservation for the TW2 and SHV2 schemes. In general, mass conservation is needed to guarantee a correct shock speed and shock

location. However, this test with a block function showed very little difference between the schemes.

4.1.2 Burgers' equation with stationary shock

In the above numerical test the lack of mass conservation for scheme TW2 only gave a very small effect. To make this effect more pronounced we consider the Burgers equation with a stationary shock at a grid interface. The equation is given by

$$u_t + \tfrac{1}{2}\big(u^2\big)_x = 0 \tag{4.6}$$

for $0 < t < T = 0.3$ and $-1 < x < 1$, with initial profile

$$u(x,0) = \begin{cases} 1 & \text{if } |x| < 0.3\,, \\ -1 & \text{otherwise}\,, \end{cases}$$

and boundary conditions $u(-1,t) = u(1,t) = -1$. This will lead to a rarefaction wave around $x = -0.3$ and a stationary shock at $x = 0.3$. In this experiment refinement is used at $\mathcal{D} = \cup_{k=1}^{10}[y_k, y_k + 0.1]$, $y_k = 0.2\,k - 1.1$. So the stationary shock is located at a grid interface.

The spatial discretization is given by the limited TVD scheme of Appendix A using a cell-centered non-uniform grid with mesh widths $\Delta x_j = \frac{1}{2}\Delta x$ if $x_j \in \mathcal{D}$, and $\Delta x_j = \Delta x$ otherwise. Also $\mathcal{I}_2 = \{j : x_j \in \mathcal{D}\}$ and $\mathcal{I}_1 = \mathcal{I} \setminus \mathcal{I}_2$, so that spatial and temporal refinements are taken at the same points.

Numerical solutions at the output time $t = T$ are shown in Figure 2 for $\Delta x = \frac{1}{80}$ and $\nu = \Delta t/\Delta x = 0.8$. The left picture shows the solution with $-1 < x < 1$ for the CS2 scheme. Differences between the schemes are not well visible on this scale. Therefore the right picture shows a zoom around $x = 0.3$ for the schemes TW2, CS2 and SHV2. One sees that with CS2 the shock location is correct; there is some smearing due to numerical diffusion in the spatial discretization, but it is more or less symmetric around $x = 0.3$. The solution of TW2 is leaning too much to the left, and for SHV2 too much to the right. This due to the lack of (local) conservation.

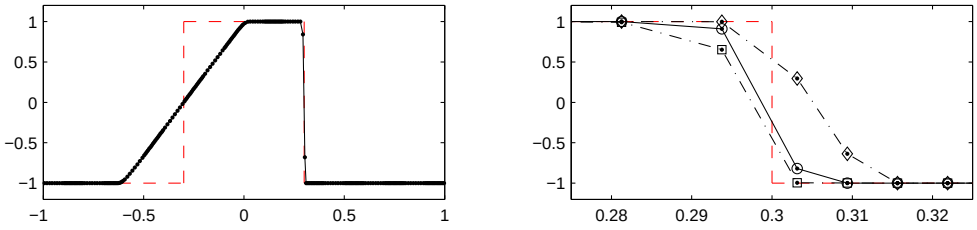

Figure 2: Numerical solutions at time $T = 0.3$ for $\Delta x = \frac{1}{80}$, $\nu = 0.8$. Left picture: initial profile (dashed), and semi-discrete solution for $-1 < x < 1$. Right picture: solutions around the stationary shock with the schemes TW2 (□ marks), CS2 (○ marks) and SHV2 (◇ marks), and with exact PDE solution (dashed line).

Let $M(v) = \sum_j \Delta x_j v_j$. (If the v_j were densities, this would be total mass; for Burgers' equation it is more natural to think of momenta.) Then $M(u(t_n)) - M(u_n)$ is a conservation defect. Figure 3 shows this defect at the final time $t_N = T$ for the three schemes on a fixed spatial mesh, $\Delta x = 1/160$, and with $\nu = \Delta t/\Delta x$ varying between 0 and 1.2. (We have taken $\nu = k/40$, $k = 1, 2, \ldots, 48$, with markers placed when ν is a multiple of 0.1.) In the same figure, middle plot, the increase of the total variation

$\|u_N\|_{\mathrm{TV}}$ is displayed. The total variation should be 4, as for the PDE solution, and this is the numerical value for the semi-discrete system (within machine precision). In this example it is conserved with larger Courant numbers for the scheme CS2 than for TW2 and SHV2. The right plot in the figure shows the increase of the maximum norm $\|u_N\|_\infty - 1$.

In these figures overflow values are not plotted. The schemes CS2 remained stable in this test up to $\nu = 1.2$, which is slightly larger than with the other two schemes. The instabilities did emerge at the stationary shock. Adding some initial perturbations results in instability for $\nu > 1$ with all three schemes.

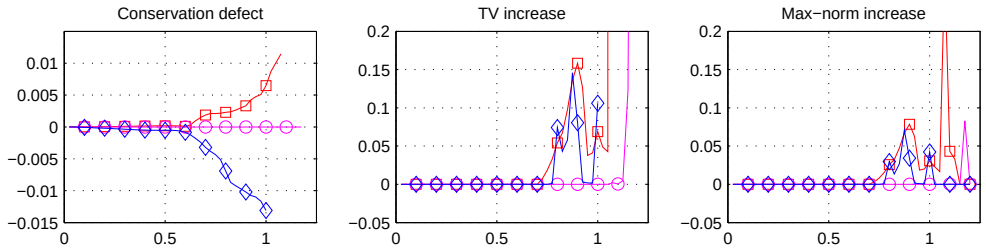

Figure 3: Conservation defects and increase of total variation and max-norm for $0 < \nu \leq 1.2$ with $\Delta x = \frac{1}{160}$, for the schemes TW2 (□ marks), CS2 (○ marks) and SHV2 (◇ marks).

Finally, in Figure 4 the logarithm (base 10) of the L_1-errors of the three schemes are given, again for $\Delta x = 1/160$ with varying ν. Both the errors with respect to the semi-discrete solution and the errors with respect to the PDE solution are plotted. It is seen that the ODE errors for CS2 are smaller than for the other two schemes for large Courant numbers. That is due to the fact that CS2 has a smaller error near the stationary shock. However, this scheme is more inaccurate than TW2 and SHV2 in the rarefaction wave, similar as in the previous test, and that reveals itself in the larger error for small Courant numbers. In the PDE errors the spatial errors will become dominant for small time steps, so there the best results are found for CS2 overall. From the PDE point of view, temporal errors less then 10^{-3} are not relevant on this spatial grid where we have a spatial error of $3.4 \cdot 10^{-3}$ approximately (PDE error for $\nu \to 0$).

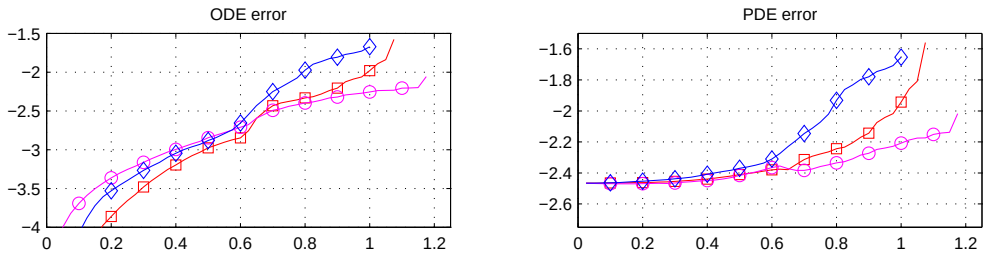

Figure 4: Logarithm ($\log_{10}$) of the L_1-errors, with respect to the exact semi-discrete solution (ODE error) and the exact PDE solution (PDE error), for $0 < \nu \leq 1.2$ with $\Delta x = \frac{1}{160}$. Results for the schemes TW2 (□ marks), CS2 (○ marks) and SHV2 (◇ marks).

4.1.3 Burgers' equation with moving shock

The last test is again Burgers' equation (4.6), but now with a moving shock. We take $0 < t < T = 0.6$, $-1 < x < 1$ with initial profile

$$u(x,0) = \begin{cases} 1 & \text{if } -0.6 < x < 0\,, \\ 0 & \text{otherwise.} \end{cases}$$

and boundary conditions $u(-1,t) = u(1,t) = 0$. This will lead to a rarefaction wave between $x = -0.6 + t$ and $x = 0$, together with a moving shock at $x = \frac{1}{2}t$. Further, we use the same set-up as in the previous test.

The solutions at time $T = 0.6$ are shown in Figure 5. The enlargement around the shock at $x = 0.3$ now shows very little difference between the three schemes. So the lack of mass conservation for the TW2 and SHV2 schemes does not have much impact for this test. This is similar as in the tests of [26] for the TW2 scheme.

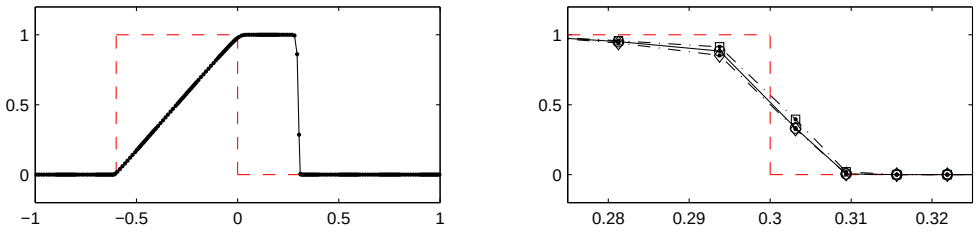

Figure 5: Numerical solutions at time $T = 0.6$ for $\Delta x = \frac{1}{80}$, $\nu = 0.8$. Left picture: initial profile (dashed), and semi-discrete solution for $-1 < x < 1$. Right picture: solutions around the moving shock with the schemes TW2 (□ marks), CS2 (○ marks) and SHV2 (◇ marks), and with exact PDE solution (dashed line).

The conservation defects and the increase of total variation and maximum-norm, with fixed mesh width $\Delta x = \frac{1}{160}$ and variable ν, are displayed in Figure 6. Here we see that all three schemes start to loose the TVD property when Courant numbers become larger than 0.8, approximately. The plot on the right of the overshoot values $\|u_N\|_\infty - 1$ looks similar, except that now the increase starts at Courant number one. The loss of the TVD property for $\nu \in [0.8, 1]$ is cause by oscillations at the shock, not in the rarefaction wave.

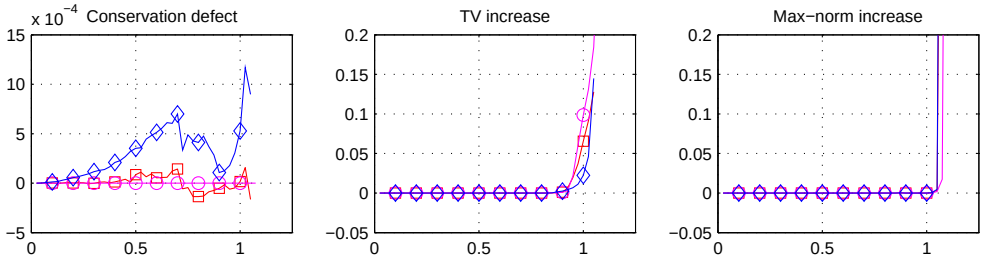

Figure 6: Conservation defects and increase of total variation and max-norm for $0 < \nu \leq 1.2$ with $\Delta x = \frac{1}{160}$, for the schemes TW2 (□ marks), CS2 (○ marks) and SHV2 (◇ marks).

We see that the conservation defect in this test is much smaller than in the previous test with a standing shock at a grid interface. Of course, both these tests are somewhat

academic, but for practical situations the present test with a moving shock seems more relevant. Monotonicity for the TW2 and SHV2 schemes holds with larger Courant numbers than in the previous test. This is caused by the fact that in the previous test there were two incoming fluxes at the standing shock, whereas now we have one incoming and one outgoing flux at each grid cell. In the standing shock test the conservation property of the CS2 scheme did suppress the tendency of increasing the total variation and maximum-norm.

In Figure 7 the temporal (ODE) errors and total (PDE) errors are plotted, again with fixed mesh width $\Delta x = \frac{1}{160}$ and variable ν. The ODE errors for the CS2 scheme are larger than for the other two schemes for small Courant numbers, but for the PDE errors this is not relevant here. In the plot of the PDE errors we see that here the SHV2 scheme gives somewhat larger errors than the TW2 and CS2 schemes. Detailed inspection of the solution plots revealed that this is due to a slight dissipation with SHV2 at the top and bottom of the rarefaction wave. We did notice, however, that these errors are quite sensitive to the precise set-up of the test. For example, with $T = 0.5$ and initial profile $u(0, x) = 1$ for $-T < x < 0$ and 0 otherwise, then the PDE errors of SHV2 were smaller than with the other two schemes for the larger Courant numbers.

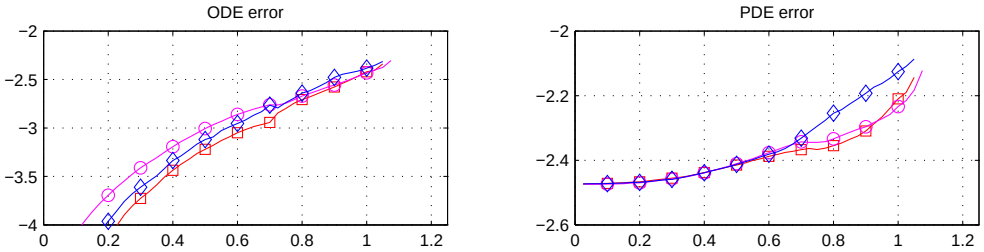

Figure 7: Logarithm ($\log_{10}$) of the L_1-errors, with respect to the exact semi-discrete solution (ODE error) and the exact PDE solution (PDE error), for $0 < \nu \leq 1.2$ with $\Delta x = \frac{1}{160}$. Results for the schemes TW2 ($\square$ marks), CS2 ($\circ$ marks) and SHV2 ($\diamond$ marks).

For theoretical purposes it is interesting to note that with the Burgers flux function $f(u) = \frac{1}{2}u^2$ we have $f'(u) \in [0, 1]$ in this test. Furthermore, the mesh width in space is $\Delta x_j = \Delta x/k$ for $j \in \mathcal{I}_k$, $k = 1, 2$, and $\mu = 1$ for the used spatial discretization. Therefore, as discussed in Example 2.2, the monotonicity assumption (2.13) will be satisfied with

$$\tau_0 = \tfrac{1}{2}\Delta x$$

for both the maximum-norm and for the total variation semi-norm. Note that with the first-order upwind discretization this would be $\tau_0 = \Delta x$.

5 Partitioned Runge-Kutta methods

5.1 General properties

In the multirate examples considered thus far, only one level of refinement was used to keep the notation simple. Generalizations will be formulated in this section in terms of partitioned Runge-Kutta methods; see also [3, 6]. This will enable us to present the schemes in a compact fashion. Since this paper is concerned with schemes for conservation laws, we will restrict ourselves to explicit methods.

For the ODE system in $\mathbb{R}^m$, arising from semi-discretization of a PDE with given initial value,

$$u'(t) = F(u(t)), \quad u(0) = u_0, \tag{5.1}$$

let $\mathcal{I} = \mathcal{I}_1 \cup \cdots \cup \mathcal{I}_r$ be an index partitioning with corresponding diagonal matrices $I = I_1 + \cdots + I_r$, where the entries of the I_k are zero or one, and I is the identity matrix. For a time step from t_n to $t_{n+1} = t_n + \Delta t$, an explicit partitioned Runge-Kutta method reads

$$\begin{aligned} v_{n,i} &= u_n + \Delta t \sum_{k=1}^{r} \sum_{j=1}^{i-1} a_{ij}^{(k)} I_k F(v_{n,j}), \qquad i = 1, \ldots, s, \\ u_{n+1} &= u_n + \Delta t \sum_{k=1}^{r} \sum_{j=1}^{s} b_j^{(k)} I_k F(v_{n,j}). \end{aligned} \tag{5.2}$$

The internal stage vectors $v_{n,i}$, $i = 1, \ldots, s$, give approximations at intermediate time levels. The multirate schemes of the previous sections all fit in this form with $r = 2$. With $r > 2$ more levels of temporal refinement are allowed.

5.1.1 Internal consistency and conservation

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

$$c_i^{(k)} = c_i^{(l)} \quad \text{for all } 1 \le k, l \le r \text{ and } 1 \le i \le s, \tag{5.3}$$

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

Apart from consistency, we will also regard global *conservation*, 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 (5.1). This will hold for arbitrary initial value u_0 provided that

$$h^T F(v) = 0 \qquad \text{for all } v \in \mathbb{R}^m. \tag{5.4}$$

For the partitioned Runge-Kutta scheme we have

$$\begin{aligned} h^T u_{n+1} &= h^T u_n + \Delta t \sum_{k=1}^{r} \sum_{j=1}^{s} b_j^{(k)} h^T I_k F(v_{n,j}) \\ &= h^T u_n + \Delta t \sum_{k \ne l} \sum_{j=1}^{s} \big(b_j^{(k)} - b_j^{(l)}\big) h^T I_k F(v_{n,j}), \end{aligned}$$

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

$$b_j^{(k)} = b_j^{(l)} \quad \text{for all } 1 \le k, l \le r \text{ and } 1 \le j \le s. \tag{5.5}$$

5.1.2 Order conditions for non-stiff problems

Below we shall use the order conditions for partitioned Runge-Kutta methods applied to non-stiff problems as found in [9, Thm. I.15.9] for $r = 2$. This classical order will be denoted by p. As we will see, it often does not correspond to the order of convergence for semi-discrete systems, and therefore p is often referred to as the *classical order*.

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

$$b_k^T e = 1 \qquad \text{for } k = 1,\ldots,r\,, \tag{5.6}$$

that is $\sum_{j=1}^s b_j^{(k)} = 1$ for all k. To have $p = 2$ the coefficients should satisfy

$$b_k^T A_l\, e = \tfrac{1}{2} \qquad \text{for } k,l = 1,\ldots,r\,. \tag{5.7}$$

The number of conditions quickly increase for higher orders; for $p = 3$ we get

$$b_k^T C_{l_1} A_{l_2} e = \tfrac{1}{3}\,, \quad b_k^T A_{l_1} A_{l_2} e = \tfrac{1}{6} \qquad \text{for } k,l_1,l_2 = 1,\ldots,r\,, \tag{5.8}$$

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

5.1.3 Formulation for non-autonomous systems

For non-autonomous systems

$$u'(t) = F(t,u(t))\,, \qquad u(0) = u_0\,, \tag{5.9}$$

we will use the partitioned method (5.2) with the stage function values $F(v_{n,j})$ replaced by $F(t_n + c_j\Delta t, v_{n,j})$. If (5.3) is valid, the abscissa are naturally taken as $c_i = c_i^{(k)}$, which is independent of k.

If (5.3) does not hold, then a proper choice of the abscissa is less obvious. For the OS1 and CS2 multirate schemes with $r = 2$ it is natural to take $c_i = c_i^{(2)}$. As generalization we will therefore use

$$c_i = c_i^{(r)}\,, \qquad i = 1,\ldots,s\,. \tag{5.10}$$

Note that if $h^T F(t,v) = 0$ for all $t \in \mathbb{R}$, $v \in \mathbb{R}^m$, then we still have the conservation property $h^T u_{n+1} = h^T u_n$ if the scheme satisfies (5.5).

The alternative of replacing $I_k F(v_{n,j})$ in (5.2) by $I_k F(t_n + c_j^{(k)}\Delta t, v_{n,j})$ will destroy this conservation property. If the non-autonomous form originates from a source term in the PDE, this loss of conservation may be of little concern, but for the advection equation $u_t + \big(a(x,t)u\big)_x = 0$ with time-dependent velocity it is still a very desirable property.

Example 5.1 The OS1 scheme (2.9) leads to the partitioned method (5.2) with $r = 2$ and coefficients given by

$$\begin{array}{c|c|cc|cc} a_{ij}^{(1)} & a_{ij}^{(2)} & & & & \\ \hline & & 0 & & 0 & \\ & & 0 & 0 & 1/2 & 0 \\ \hline b_j^{(1)} & b_j^{(2)} & 1/2 & 1/2 & 1/2 & 1/2 \end{array}$$

For non-autonomous systems $u'(t) = F(t,u(t))$ the scheme with (5.10) reads

$$\begin{cases} u_{n+\frac12} = u_n + \frac12\Delta t I_2 F(t_n,u_n)\,, \\ u_{n+1} = u_n + \frac12\Delta t F(t_n,u_n) + \frac12\Delta t F(t_{n+\frac12},u_{n+\frac12})\,. \end{cases}$$

The use of $I_k F(t_n + c_j^{(k)}\Delta t, v_{n,j})$ instead of $I_k F(t_n + c_j\Delta t, v_{n,j})$, $c_j = c_j^{(2)}$, would lead to the same formula for $u_{n+1/2}$ in the first stage, but then

$$u_{n+1} = u_n + \tfrac12\Delta t F(t_n,u_n) + \tfrac12\Delta t I_1 F(t_n,u_{n+\frac12}) + \tfrac12\Delta t I_2 F(t_{n+\frac12},u_{n+\frac12})\,,$$

which is no longer conservative. ◇

The above order conditions have been derived for autonomous systems, but with (5.10) they are also valid for non-autonomous systems. This follows from the fact that $u'(t) = F(t, u(t))$ can be written as an equivalent, augmented autonomous system $u'(t) = F(\vartheta(t), u(t))$, $\vartheta'(t) = 1$, with $\vartheta(0) = 0$, and application of the partitioned method to this augmented system gives the same result as to the original, non-autonomous system provided the additional equation $\vartheta'(t) = 1$ is included in the index set $\mathcal{I}_r$.

5.1.4 Conservation versus internal consistency

For the multirate schemes that have been considered in this paper, the conditions for internal consistency (5.3) and conservation (5.5) did not match. This incompatibility is valid for all 'genuine' multirate schemes that are based on one single method $\mathcal{M}_{\mathrm{RK}}$, that is, for schemes (5.2) that reduce to m_k applications (with step size $\Delta t/m_k$) of this base method $\mathcal{M}_{\mathrm{RK}}$ to cover $[t_n, t_{n+1}]$ in case that $\mathcal{I}_k = \mathcal{I}$ and the other $\mathcal{I}_l$ are empty.

Consider, as simple example, a quadrature problem $u'(t) = g(t) \in \mathbb{R}^m$, which is just a special case of (5.9). (In a PDE context, this can be viewed as a degenerate case of advection with a source term where the advective velocity happens to be zero.) Suppose (5.5) is valid, and let $\mathcal{J} = \{i \in \mathcal{I} : b_i \neq 0\}$. Then for the quadrature problem we simply get

$$u_{n+1} = u_n + \Delta t \sum_{i \in \mathcal{J}} b_i \, g(t_n + c_i \Delta t) \,,$$

which is independent of the partitioning. However, if this is the result of a base method $\mathcal{M}_{\mathrm{RK}}$ with $m_1 = 1$, $\mathcal{I}_1 = \mathcal{I}$, then the result for $m_2 = 2$, $\mathcal{I}_2 = \mathcal{I}$ should be

$$u_{n+1} = u_n + \tfrac{1}{2}\Delta t \sum_{i \in \mathcal{J}} b_i \Big(g\big(t_n + \tfrac{1}{2} c_i \Delta t\big) + g\big(t_n + \tfrac{1}{2}(1 + c_i)\Delta t\big) \Big) \,,$$

which is not the same for arbitrary source terms g.

Note that for general partitioned Runge-Kutta methods there is no conflict between (5.3) and (5.5). Given a scheme with the same $c_i^{(k)} = c_i^{(l)}$ (for all i, k, l), but different weights $b_i^{(k)} \neq b_i^{(l)}$ (for some i, k, l), we can add an extra stage with new weights b_i^* that are independent of k, to make it mass-conserving. Of course, this will increase the computational work per step, and for the TW1, TW2 and SHV2 schemes such a modification does not seem to lead to efficient schemes.

5.2 Monotonicity and convex Euler combinations

We are in particular interested in the case where the partitioned Runge-Kutta method (5.2) stands for a multirate scheme that takes m_k substeps of size $\Delta t/m_k$ on $\mathcal{I}_k$ to cover $[t_n, t_{n+1}]$, $k = 1, \ldots, r$, with $m_1 = 1 < m_2 < \cdots < m_r$. The corresponding monotonicity assumption is

$$\Big\| v + \sum_{k=1}^{r} \frac{\tau_k}{m_k} I_k F(v) \Big\| \;\leq\; \|v\| \qquad \text{for all } v \in \mathbb{R}^m \text{ and } \tau_k \leq \tau_0,\ k = 1, \ldots, r \,, \tag{5.11}$$

where $\|\cdot\|$ is a convex function or (semi-)norm. For theoretical purposes we will also consider

$$\big\| v + \frac{\tau_0}{m_k} I_k F(v) \big\| \;\leq\; \|v\| \qquad \text{for all } v \in \mathbb{R}^m \text{ and } k = 1, \ldots, r \,. \tag{5.12}$$

Of course, (5.11) implies (5.12). On the other hand, if (5.12) is valid, then the inequality in (5.11) will hold under the step size restriction $\tau_1 + \cdots + \tau_m \leq \tau_0$. If we are dealing with the maximum-norm, then (5.11) and (5.12) are equivalent.

In the following we denote for $l = 1, \ldots, r$,

$$\left\{ \begin{array}{rcll} \kappa_{ij}^{(l)} & = & m_l a_{ij}^{(l)}\,, & 1 \le i,j \le s\,, \\ \kappa_{s+1,j}^{(l)} & = & m_l b_j^{(l)}\,, & 1 \le j \le s\,, \\ \kappa_{i,s+1}^{(l)} & = & 0\,, & 1 \le i \le s+1\,. \end{array} \right. \tag{5.13}$$

These coefficients will be grouped in the $(s+1) \times (s+1)$ matrix $\mathcal{K}_l = [\kappa_{ij}^{(l)}]$. It is convenient to add $v_{n,s+1} = u_{n+1}$ to the internal vectors. Then (5.2) can be written as

$$v_{n,i} \;=\; u_n + \sum_{l=1}^{r} \sum_{j=1}^{i-1} \kappa_{ij}^{(l)} \frac{\Delta t}{m_l} I_l F(v_{n,j})\,, \qquad i = 1, \ldots, s+1\,. \tag{5.14}$$

Depending on the monotonicity assumption, we can consider various ways to represent this partitioned scheme in terms of convex Euler combinations. For this we will introduce new method coefficients $\alpha_{ij}^{(k)}$, $\beta_{ij}^{(k)}$ with corresponding lower triangular matrices $\mathcal{A}_k = [\alpha_{ij}^{(k)}]$ and $\mathcal{B}_k = [\beta_{ij}^{(k)}]$. Such convex Euler forms are also called Shu-Osher forms, after [24] where such representations were used originally to demonstrate the TVD property of certain Runge-Kutta methods.

Inequalities for matrices or vectors in this section are to be understood componentwise, that is, $P = [p_{ij}] \ge 0$ means that all p_{ij} are non-negative. Furthermore, if $P \in \mathbb{R}^{(s+1)\times q_1}$ and $Q \in \mathbb{R}^{(s+1)\times q_2}$, then $[P\ Q]$ stands for the matrix whose first q_1 columns equal those of P and the other columns equal those of Q. In this section we let $e = [1, 1, \ldots, 1]^T \in \mathbb{R}^{s+1}$, and we use the convention $\alpha/\beta = +\infty$ if $\alpha \ge 0$, $\beta = 0$.

5.2.1 Convex Euler form I: maximum-norm monotonicity.

A suitable form of (5.14) to obtain results on monotonicity in the maximum-norm is

$$v_{n,i} \;=\; \sum_{k=1}^{r} I_k \Big(\big(1 - \alpha_i^{(k)}\big) u_n + \sum_{j=1}^{i-1} \big(\alpha_{ij}^{(k)} v_{n,j} + \beta_{ij}^{(k)} \frac{\Delta t}{m_k} F(v_{n,j}) \big) \Big)\,, \tag{5.15}$$

where $\alpha_i^{(k)} = \sum_{j=1}^{i-1} \alpha_{ij}^{(k)}$ and $i = 1, \ldots, s+1$. To have correspondence between (5.14) and (5.15) the coefficients should satisfy

$$\mathcal{K}_k = \big(I - \mathcal{A}_k\big)^{-1} \mathcal{B}_k\,, \qquad k = 1, \ldots, r\,. \tag{5.16}$$

Further we want the coefficients to be such that

$$\alpha_i^{(k)} \le 1\,, \quad \alpha_{ij}^{(k)}, \beta_{ij}^{(k)} \ge 0 \qquad \text{for } 1 \le j < i \le s+1\,,\ 1 \le k \le r\,. \tag{5.17}$$

For such coefficients, let

$$C \;=\; \min_{i,j,k} \alpha_{ij}^{(k)} / \beta_{ij}^{(k)}\,. \tag{5.18}$$

If there are no coefficients such that (5.16) and (5.17) are satisfied, we set $C = 0$.

Theorem 5.2 *Consider (5.15) with (5.17) and let C be given by (5.18). Assume (5.11) is valid in the maximum-norm. Then $\|u_{n+1}\|_\infty \le \|u_n\|_\infty$ whenever $\Delta t \le C\tau_0$.*

Proof. The form (5.15) is equivalent to

$$I_k v_{n,i} \;=\; I_k \Big(\big(1 - \alpha_i^{(k)}\big) u_n + \sum_{j=1}^{i-1} \big(\alpha_{ij}^{(k)} v_{n,j} + \beta_{ij}^{(k)} \frac{\Delta t}{m_k} I_k F(v_{n,j}) \big) \Big)\,, \qquad k = 1, \ldots, r\,.$$

We have $v_{n,1} = u_n$. Suppose (induction assumption) that $\|v_{n,j}\|_\infty \le \|u_n\|_\infty$ for $j = 1, \ldots, i-1$. Since

$$\alpha_{ij}^{(k)} v_{n,j} + \beta_{ij}^{(k)} \frac{\Delta t}{m_k} I_k F(v_{n,j}) = \left(\alpha_{ij}^{(k)} - C\beta_{ij}^{(k)}\right) v_{n,j} + C\beta_{ij}^{(k)} \left(v_{n,j} + \frac{\Delta t}{C m_k} I_k F(v_{n,j})\right),$$

we then have

$$\|\alpha_{ij}^{(k)} v_{n,j} + \beta_{ij}^{(k)} \frac{\Delta t}{m_k} I_k F(v_{n,j})\|_\infty \le \alpha_{ij}^{(k)} \|v_{n,j}\|_\infty \le \alpha_{ij}^{(k)} \|u_n\|_\infty .$$

It follows that $\|I_k v_{n,i}\|_\infty \le \|u_n\|_\infty$ for $k = 1, \ldots, r$, and hence $\|v_{n,i}\|_\infty \le \|u_n\|_\infty$. Using induction with respect to $i = 1, \ldots, s+1$ the proof thus follows. □

It is obvious that we are in particular interested in the optimal value of C in (5.18) for a given method (5.14). To obtain a suitable expression for this optimal value, we can follow the construction of Ferracina & Spijker [7] and Higueras [10] for the individual Runge-Kutta methods given by the coefficients $\mathcal{K}_k$.

Theorem 5.3 *The optimal value for $C \ge 0$ in (5.18), under the constraints (5.16) and (5.17), equals the largest $\gamma \ge 0$ such that*

$$(I + \gamma\mathcal{K}_k)^{-1}[e\ \gamma\mathcal{K}_k] \ge 0\,, \qquad k = 1, \ldots, r\,. \tag{5.19}$$

Proof. Suppose $\gamma \ge 0$ is such that (5.19) holds. We take $\mathcal{B}_k = (I + \gamma\mathcal{K}_k)^{-1}\mathcal{K}_k$ and $\mathcal{A}_k = \gamma\mathcal{B}_k$. With this choice it is easily seen that (5.16) and (5.17) are valid and that (5.18) holds with $C = \gamma$.

On the other hand, suppose that we have (5.16), (5.17) and (5.18) with $C \ge 0$, and set $\gamma = C$. Then

$$\left(I + \gamma\mathcal{K}_k\right)^{-1}[\![e\ \gamma\mathcal{K}_k]\!] = \left(I - \mathcal{M}_k\right)^{-1}[(I - \mathcal{A}_k)e\ \ \gamma\mathcal{B}_k]\,,$$

where $\mathcal{M}_k = \mathcal{A}_k - \gamma\mathcal{B}_k$. From (5.18) we know that $\mathcal{M}_k \ge 0$, and since it is a strictly lower triangular matrix we also have

$$(I - \mathcal{M}_k)^{-1} = I + \mathcal{M}_k + \mathcal{M}_k^2 + \ldots + \mathcal{M}_k^s \ \ge\ 0\,.$$

It follows that (5.19) is valid. □

5.2.2 Convex Euler form II: monotonicity under (5.12)

If we assume (5.12) for a general (semi-)norm or convex function, then a suitable form for (5.14) is

$$v_{n,i} \ =\ \left(1 - \underline{\alpha}_i^{(0)}\right) u_n + \sum_{k=1}^{r} \sum_{j=1}^{i-1} \left(\underline{\alpha}_{ij}^{(k)} v_{n,j} + \underline{\beta}_{ij}^{(k)} \frac{\Delta t}{m_k} I_k F(v_{n,j})\right), \tag{5.20}$$

where $\underline{\alpha}_i^{(0)} = \sum_{j=1}^{i-1} \left(\underline{\alpha}_{ij}^{(1)} + \cdots + \underline{\alpha}_{ij}^{(r)}\right)$, $i = 1, \ldots, s+1$, and

$$\mathcal{K}_k = \left(I - \sum_{l=1}^{r} \underline{\mathcal{A}}_l\right)^{-1} \underline{\mathcal{B}}_k\,, \qquad k = 1, \ldots, r\,. \tag{5.21}$$

We want

$$\underline{\alpha}_i^{(0)} \le 1\,, \quad \underline{\alpha}_{ij}^{(k)}, \underline{\beta}_{ij}^{(k)} \ge 0 \qquad \text{for } 1 \le j < i \le s+1\,,\ 1 \le k \le r\,, \tag{5.22}$$

with an optimal

$$\underline{C} = \min_{i,j,k} \underline{\alpha}_{ij}^{(k)} / \underline{\beta}_{ij}^{(k)}\,. \tag{5.23}$$

Theorem 5.4 *Assume (5.12) is valid.*
(i) Consider (5.20) with (5.22) and let $\underline{C}$ *be given by (5.23). Then* $\|u_{n+1}\| \le \|u_n\|$ *whenever* $\Delta t \le \underline{C}\tau_0$.
(ii) The optimal $\underline{C} \ge 0$ *in (5.23), under the constraints (5.21) and (5.22), equals the largest* $\gamma \ge 0$ *such that*

$$\Big(I + \sum_{l=1}^{r} \gamma \mathcal{K}_l\Big)^{-1} [\![e\, \gamma \mathcal{K}_k]\!] \ge 0\,, \qquad k = 1, \ldots, r\,. \tag{5.24}$$

The proof of this result is similar to that of the Theorems 5.2 and 5.3. In fact, the result for $r = 2$ can be obtained directly from Higueras [11] and Spijker [25]. Further we note that the coefficient matrices $\underline{\mathcal{A}}_k$ and $\underline{\mathcal{B}}_k$ which lead to an optimal value $\underline{C}$ are in this case given by $\underline{\mathcal{B}}_k = (I + \sum_l \gamma \mathcal{K}_l)^{-1} \mathcal{K}_k$ and $\underline{\mathcal{A}}_k = \gamma \underline{\mathcal{B}}_k$.

5.2.3 Convex Euler form III: TVD property and monotonicity under (5.11)
Finally, if (5.11) is assumed for a general (semi-)norm or convex function, then we consider

$$v_{n,i} \;=\; \big(1 - \overline{\alpha}_i^{(0)}\big) u_n + \sum_{j=1}^{i-1} \Big(\overline{\alpha}_{ij}^{(0)} v_{n,j} + \sum_{k=1}^{r} \overline{\beta}_{ij}^{(k)} \frac{\Delta t}{m_k} I_k F(v_{n,j}) \Big)\,, \tag{5.25}$$

where $\overline{\alpha}_i^{(0)} = \sum_{j=1}^{i-1} \overline{\alpha}_{ij}^{(0)}$, $i = 1, \ldots, s+1$, and

$$\mathcal{K}_k = (I - \overline{\mathcal{A}}_0)^{-1} \overline{\mathcal{B}}_k\,, \qquad k = 1, \ldots, r\,. \tag{5.26}$$

Here we want

$$\overline{\alpha}_i^{(0)} \le 1\,, \quad \overline{\alpha}_{ij}^{(0)}, \overline{\beta}_{ij}^{(k)} \ge 0 \qquad \text{for } 1 \le j < i \le s+1\,,\ 1 \le k \le r\,. \tag{5.27}$$

such that

$$\overline{C} = \min_{i,j,k} \overline{\alpha}_{ij}^{(0)} / \overline{\beta}_{ij}^{(k)} \tag{5.28}$$

is optimal.

Theorem 5.5 *Consider (5.25) with (5.27) and let* $\overline{C}$ *be given by (5.28). Assume (5.11) is valid. Then* $\|u_{n+1}\| \le \|u_n\|$ *whenever* $\Delta t \le \overline{C}\tau_0$.

The proof is similar to that of Theorem 5.2. For this case there is no convenient representation of the optimal $\overline{C}$. An optimization code can be used to determine this optimal value. However, from the previous results we obtain useful upper and lower bounds for $\overline{C}$.

Theorem 5.6 *The optimal values* C, $\underline{C}$, $\overline{C}$ *in (5.18), (5.23) and (5.28) satisfy*

$$\frac{1}{r}\overline{C} \;\le\; \underline{C} \;\le\; \overline{C} \;\le\; C\,.$$

Consequently, if $\underline{C} = 0$ *then* $\overline{C} = 0$.

Proof. Given an optimal $\overline{C}$ with corresponding coefficient matrices $\overline{\mathcal{A}}_0$, $\overline{\mathcal{B}}_k$, we can take $\mathcal{A}_k = \overline{\mathcal{A}}_0$, $\mathcal{B}_k = \overline{\mathcal{B}}_k$. Then (5.16) and (5.17) hold and $\min_{i,j,k} \alpha_{ij}^{(k)} / \beta_{ij}^{(k)} \ge \overline{C}$. Consequently we have $C \ge \overline{C}$ for the optimal value C.

Likewise, for a given optimal $\underline{C}$ with corresponding $\underline{\mathcal{A}}_k$, $\underline{\mathcal{B}}_k$, we can choose $\overline{\mathcal{B}}_k = \underline{\mathcal{B}}_k$, $\overline{\mathcal{A}}_0 = \sum_{l=1}^{r} \underline{\mathcal{A}}_l$. Then (5.26) and (5.27) hold and we have $\min_{i,j,k} \overline{\alpha}_{ij}^{(0)} / \overline{\beta}_{ij}^{(k)} \ge \underline{C}$, showing that $\overline{C} \ge \underline{C}$.

On the other hand, for given optimal $\overline{C}$ with corresponding $\overline{\mathcal{A}}_0$, $\overline{\mathcal{B}}_k$, we can take $\underline{\mathcal{B}}_k = \overline{\mathcal{B}}_k$, $\underline{\mathcal{A}}_k = \frac{1}{r}\overline{\mathcal{A}}_0$. It follows that $\underline{C} \ge \frac{1}{r}\overline{C}$. □

5.2.4 Results for the multirate schemes with one level of refinement

The monotonicity results for the multirate schemes of the previous sections are presented in Table 2. The table gives the threshold values C, $\overline{C}$ and $\underline{C}$ for the various schemes. The results for the first-order schemes OS1 and TW1 can be derived analytically as in Section 3.1; we get $C = 1$, $\overline{C} = 2/3$, $\underline{C} = 1 - 1/\sqrt{3}$ for OS1, and $C = 1$, $\overline{C} = 2 - \sqrt{2}$, $\underline{C} = 1 - 1/\sqrt{3}$ for TW1. The threshold values C, $\overline{C}$ for the second-order schemes have been found numerically, using (5.19) and (5.24). For the TW2 and CS2 schemes we have $\underline{C} = 0$ and therefore also $\overline{C} = 0$. (The fact that $\underline{C} = 0$ for these two schemes can also be shown analytically, similar to [11], by considering (5.24) for small $\gamma > 0$.) The value of $\overline{C}$ for SHV2 was obtained with the MATLAB optimization code FMINIMAX. This does not provide a guarantee that the solution is a global optimum, and therefore this $\overline{C}$ is to be considered as a lower bound. The fact that we merely have $C = 1/2$ for the SHV2 scheme is due to the first stage. Finally we note that for the variant of that scheme with linear interpolation (4.5), instead of (4.4), it was found that $C = 1/2$, $\underline{C} = 0.304$, and the optimization code produced the same value $\overline{C} = 0.304$ for this variant.

Table 2: Threshold values for the multirate schemes with one level of refinement. The entry $\overline{C}$ for the scheme SHV2 is a lower bound.

	C	$\overline{C}$	$\underline{C}$
OS1	1	0.667	0.423
TW1	1	0.580	0.423
TW2	1	0	0
CS2	1	0	0
SHV2	0.5	0.284	0.284

As noted before, the result $C = 1$ for the OS1 and TW1 scheme was already given in [15, 18, 26] in terms of maximum principles. For the CS2 scheme the same result has been proved in [3].

Recall that the threshold values C are such that we will have monotonicity in the maximum-norm, as well as maximum principles, provided that $\Delta t \leq C\tau_0$. Likewise, for spatial discretization with limiting the TVD property will hold if $\Delta t \leq \overline{C}\tau_0$. All this under corresponding assumptions (2.13) for the semi-discrete system.

Comparison of these theoretical values with the experiments of Section 4.1 for Burgers' equation with the TW2, CS2 and SHV2 schemes does not show a clear correspondence. As was noted, in those experiments we had $\tau_0 = \frac{1}{2}\Delta x$ for both the maximum-norm and the total variation semi-norm. Therefore, with $\nu = \Delta t/\Delta x$, the TVD property is guaranteed by the above results for $\nu \leq \frac{1}{2}\overline{C}$ and the maximum principle for $\nu \leq \frac{1}{2}C$. For the Burgers' experiment with a moving shock it was noticed that for the schemes TW2, CS2 and SHV2 we had no overshoots for $\nu \leq 1$, whereas the TVD property was valid for $\nu \leq 0.8$ approximately. Therefore, for that test, the theoretical threshold values $\overline{C} = 0$ for the TW2 and CS2 schemes in Table 2 are much too pessimistic. The same seems to hold for the small value $C = \frac{1}{2}$ of the SHV2 scheme compared to the value $C = 1$ for TW2 and CS2. This may be caused by the fact that spatial discretizations with flux-limiting (or of WENO type) do add some local diffusion near very steep gradients, which may counteract an overshoot or increase of total variation of the time stepping scheme. However, for the discrepancy in the TVD results it is more likely that a more refined theory is needed. As noted before, it was

shown in [15] that the OS1 scheme is TVD for a class of limited discretizations under the same step size restriction as for the maximum principle, but that proof does not lend itself to generalization for the higher-order schemes.

Remark 5.7 Refined TVD results for the OS1 and TW1 scheme were also discussed in Section 3.1. It was shown that the TVD thresholds of both the OS1 and TW1 schemes become 1 for the system (3.8) arising from linear advection with first-order upwind discretization in space.

Experimentally, using various partitionings, including random partitionings, we observed that for this system the thresholds for monotonicity in the maximum-norm are 1 for the TW2 and CS2 schemes, and approximately 0.66 for the SHV2 scheme, whereas the thresholds for the TVD property are 0.5 for the TW2 and CS2 schemes, and 0.86 for the SHV2 scheme.

Furthermore, it should be noticed that having a bound $\|S\|_\infty \le 1$ for the amplification matrix S guarantees stability in the maximum norm for this linear problem, but this is not a necessary condition. The spectral radius of S was found to be bounded by 1 for Courant numbers $\nu_j = \Delta t/\Delta x_j \le k$ for $j \in \mathcal{I}_k$, $k = 1, 2$, for these three schemes, that is, including the SHV2 scheme. Note that having spectral radius bounded by 1 is of course necessary for stability, but it is not sufficient, not even in the L_2 norm because the amplification matrices S are not normal. ◇

5.3 Convergence for smooth problems

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

It is useful here to take also non-autonomous equations (5.9) into consideration. Then linear constant coefficient problems $u'(t) = Au(t) + g(t)$ with time dependent source terms are included. Such $g(t)$ may originate from a genuine source term in the PDE or from an inhomogeneous boundary condition.

To ensure stability, it will be assumed that

$$\big\|\tilde v - v + \frac{\tau_0}{m_k} I_k\big(F(t,\tilde v) - F(t,v)\big)\big\|_\infty \;\le\; \|\tilde v - v\|_\infty\,, \qquad k = 1,\dots,r\,, \tag{5.29}$$

for any two vectors $\tilde v, v \in \mathbb{R}^m$ and $t \in \mathbb{R}$. In applications to semi-discrete systems obtained from conservation laws this τ_0 will be proportional to the mesh widths used in the spatial discretization, and hence an upper bound $\Delta t \le C\tau_0$ on the step size will be a CFL restriction.

5.3.1 Perturbed schemes

Consider, along with (5.2) in non-autonomous form, the perturbed scheme

$$\begin{aligned}
\tilde v_{n,i} &= \tilde u_n + \Delta t \sum_{k=1}^{r}\sum_{j=1}^{i-1} a_{ij}^{(k)} I_k F(t_{n,j}, \tilde v_{n,j}) + \rho_{n,i}\,, \qquad i = 1,\dots,s\,,\\
\tilde u_{n+1} &= \tilde u_n + \Delta t \sum_{k=1}^{r}\sum_{j=1}^{s} b_{j}^{(k)} I_k F(t_{n,j}, \tilde v_{n,j}) + \sigma_n\,,
\end{aligned} \tag{5.30}$$

where $t_{n,j} = t_n + c_j\Delta t$ and the $\rho_{n,i}$, σ_n are perturbations. These perturbations will be used later on to obtain expressions for the discretization errors. In order to distinguish

the accuracy of the u_n from those of the internal stages we will mainly use the standard form (5.2) rather than (5.14).

As before, let the matrices $A_k = [a_{ij}^{(k)}] \in \mathbb{R}^{s\times s}$ and the vectors $b_k = [b_i^{(k)}] \in \mathbb{R}^s$ contain the coefficients of the scheme. Further, for the vector of abscissa $c = [c_i] \in \mathbb{R}^s$ we denote $c^j = [c_i^j]$ 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 $I = I_{m\times m}$. Likewise, $\boldsymbol{I}_k = I \otimes I_k$ with $s \times s$ identity matrix $I = I_{s\times s}$. To make the notation consistent, the $ms \times ms$ identity matrix is denoted by $\boldsymbol{I}$.

Let $\boldsymbol{Z}_n = \mathrm{diag}(Z_{n,i}) \in \mathbb{R}^{ms\times ms}$ with

$$Z_{n,i}(\tilde{v}_{n,i} - v_{n,i}) = \Delta t\big(F(t_{n,i}, \tilde{v}_{n,i}) - F(t_{n,i}, v_{n,i})\big)\,. \tag{5.31}$$

In view of (5.29) these $Z_{n,i} \in \mathbb{R}^{m\times m}$ can be taken such that[2]

$$\big\|I + \frac{1}{\gamma m_k} I_k Z_{n,i}\big\|_\infty \;\leq\; 1 \qquad \text{for } \Delta t \leq \gamma\tau_0\,,\ \gamma > 0\,,\ k = 1, \ldots, r\,. \tag{5.32}$$

To write the difference of (5.30) and (5.2) in a compact form, let also $\boldsymbol{\rho}_n = [\rho_{n,i}] \in \mathbb{R}^{sm}$ and $\boldsymbol{v}_n = [v_{n,i}]$, $\tilde{\boldsymbol{v}}_n = [\tilde{v}_{n,i}] \in \mathbb{R}^{sm}$. Then

$$\begin{aligned} \tilde{\boldsymbol{v}}_n - \boldsymbol{v}_n &= \boldsymbol{e}(\tilde{u}_n - u_n) + \sum_{k=1}^{r} \boldsymbol{A}_k \boldsymbol{I}_k \boldsymbol{Z}_n(\tilde{\boldsymbol{v}}_n - \boldsymbol{v}_n) + \boldsymbol{\rho}_n\,, \\ \tilde{u}_{n+1} - u_{n+1} &= \tilde{u}_n - u_n + \sum_{k=1}^{r} \boldsymbol{b}_k^T \boldsymbol{I}_k \boldsymbol{Z}_n(\tilde{\boldsymbol{v}}_n - \boldsymbol{v}_n) + \sigma_n\,. \end{aligned} \tag{5.33}$$

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

$$\tilde{u}_{n+1} - u_{n+1} = S_n(\tilde{u}_n - u_n) + \boldsymbol{r}_n^T \boldsymbol{\rho}_n + \sigma_n\,, \tag{5.34}$$

where

$$S_n = I + \boldsymbol{r}_n^T \boldsymbol{e}\,, \qquad \boldsymbol{r}_n^T = \Big(\sum_{k=1}^{r} \boldsymbol{b}_k^T \boldsymbol{I}_k \boldsymbol{Z}_n\Big)\Big(\boldsymbol{I} - \sum_{k=1}^{r} \boldsymbol{A}_k \boldsymbol{I}_k \boldsymbol{Z}_n\Big)^{-1}. \tag{5.35}$$

The following result provides stability for this recursion with a step size restriction $\Delta t \leq C\tau_0$, where C is the threshold for monotonicity in the maximum-norm. We can consider arbitrary matrices $\boldsymbol{Z}_n$ with blocks satisfying (5.32), so that these matrices are independent from the perturbations $\boldsymbol{\rho}_n$ and σ_n.

Lemma 5.8 *Consider (5.33). Assume (5.32) and $\Delta t \leq C\tau_0$. Then*

$$\|S_n\|_\infty \leq 1\,, \quad \|\boldsymbol{r}_n^T\|_\infty \leq 2s\,. \tag{5.36}$$

Proof. Denote $w_{n,i} = \tilde{v}_{n,i} - v_{n,i}$ and also $w_{n,s+1} = \tilde{u}_{n+1} - u_{n+1}$, $\rho_{n,s+1} = \sigma_n$. Then

$$w_{n,i} = \tilde{u}_n - u_n + \sum_{k=1}^{r}\sum_{j=1}^{i-1} \frac{1}{m_k} \kappa_{ij}^{(k)} I_k Z_{n,j} w_{n,j} + \rho_{n,i}\,, \qquad i = 1, \ldots, s+1\,.$$

[2] As noted before, if F is differentiable we can take the $Z_{n,i}$ as integrated Jacobian matrices, but also for non-differentiable F we can choose them to satisfy (5.31). This is similar to the fact that if $x, y \in \mathbb{R}^m$ with $\|y\|_\infty \leq \|x\|_\infty$, then there is an $V \in \mathbb{R}^{m\times m}$ such that $Vx = y$ and $\|V\|_\infty \leq 1$; for example, if $|x_k| = \|x\|_\infty$, the matrix with kth column $\frac{1}{x_k} y$ and the other columns zero.

Following the construction used in Theorem 5.3 with optimal coefficients $\beta_{ij}^{(k)} = \alpha_{ij}^{(k)}/\gamma$, $\gamma = C$, we obtain

$$I_k(w_{n,i} - \rho_{n,i}) = \big(1 - \alpha_i^{(k)}\big) I_k(\tilde{u}_n - u_n) + \sum_{j=1}^{i-1} \alpha_{ij}^{(k)} I_k\Big(w_{n,j} + \frac{1}{\gamma m_k} Z_{n,j} w_{n,j} - \rho_{n,j}\Big) .$$

This leads to

$$\|I_k w_{n,i}\|_\infty - \|\rho_{n,i}\|_\infty \;\le\; \big(1 - \alpha_i^{(k)}\big)\|\tilde{u}_n - u_n\|_\infty + \sum_{j=1}^{i-1} \alpha_{ij}^{(k)} \big(\|w_{n,j}\|_\infty + \|\rho_{n,j}\|_\infty\big) .$$

If we make the induction assumption

$$\|w_{n,j}\|_\infty \le \|\tilde{u}_n - u_n\|_\infty + L_j \max_{\iota \le j} \|\rho_{n,\iota}\|_\infty , \tag{5.37}$$

for $j = 1, \ldots, i-1$, with $L_j = 2j - 1$, then

$$\begin{aligned} \|I_k w_{n,i}\|_\infty \;&\le\; \|\tilde{u}_n - u_n\|_\infty + \sum_{j=1}^{i-1} \alpha_{ij}^{(k)} \big(L_j \max_{\iota \le j} \|\rho_{n,\iota}\|_\infty + \|\rho_{n,j}\|_\infty\big) + \|\rho_{n,i}\|_\infty \\ &\le\; \|\tilde{u}_n - u_n\|_\infty + (L_{i-1} + 1) \max_{j \le i-1} \|\rho_{n,j}\|_\infty + \|\rho_{n,i}\|_\infty . \end{aligned}$$

Hence (5.37) will also be satisfied for $j = i$, and the proof thus follows. □

Note that without the internal perturbations we obtain a result on contractivity in the maximum-norm:

$$\|\tilde{u}_{n+1} - u_{n+1}\|_\infty \;\le\; \|\tilde{u}_n - u_n\|_\infty \qquad \text{whenever } \Delta t \le C\tau_0 , \tag{5.38}$$

for any two parallel steps of the scheme (5.2), starting with $\tilde{u}_n$ and u_n, respectively. In the above proof, the arguments leading to monotonicity have been copied. A more elegant and direct way to deduce contractivity from monotonicity is found in [25, p. 1236], following a construction of [2] for inner-product norms.

5.3.2 Local and global discretization errors

Throughout this section we will denote by $\mathcal{O}(\Delta t^q)$ a term or vector that can be bounded in norm by $K\Delta t^q$, for $\Delta t > 0$ small enough, with K not depending on the mesh widths Δx_j in the spatial discretization. The norm in this section is the maximum-norm. Moreover it will be tacitly assumed that the exact solution is smooth, so that derivatives of $u(t)$ are $\mathcal{O}(1)$.

Let $e_n = u(t_n) - u_n$ be the global discretization error at time level t_n, $n \ge 0$. To obtain a recursion for these global errors we can employ the above perturbed scheme with $\tilde{u}_n = u(t_n)$ and $\tilde{v}_{n,i} = u(t_{n,i})$, $t_{n,i} = t_n + c_i\Delta t$, $i = 1, \ldots, s$. This choice for the $\tilde{v}_{n,i}$ defines the perturbations $\rho_{n,i}$ and σ_n. Assuming the exact solution u to be $l+1$ times differentiable, Taylor expansion directly leads to

$$\begin{aligned} \boldsymbol{\rho}_n &= \sum_{k=1}^{r} \sum_{j=1}^{l} \frac{\Delta t^j}{j!} \big(\boldsymbol{c}^j - j\boldsymbol{A}_k \boldsymbol{c}^{j-1}\big) I_k u^{(j)}(t_n) + \mathcal{O}(\Delta t^{l+1}) , \\ \sigma_n &= \sum_{k=1}^{r} \sum_{j=1}^{l} \frac{\Delta t^j}{j!} \big(I - j\boldsymbol{b}_k^T \boldsymbol{c}^{j-1}\big) I_k u^{(j)}(t_n) + \mathcal{O}(\Delta t^{l+1}) . \end{aligned} \tag{5.39}$$

It follows that the global errors $e_n = u(t_n) - u_n$ satisfy the recursion

$$e_{n+1} = S_n e_n + d_n\,, \qquad n \geq 0\,, \tag{5.40}$$

with local discretization errors d_n given by

$$d_n = \boldsymbol{r}_n^T \boldsymbol{\rho}_n + \sigma_n\,, \tag{5.41}$$

and with $S_n \in \mathbb{R}^{m\times m}$, $\boldsymbol{r}_n^T \in \mathbb{R}^{m\times ms}$ given by (5.35).

Note that from $\|S_n\|_\infty \leq 1$ it follows directly that consistency of order q (i.e., $\|d_n\|_\infty = \mathcal{O}(\Delta t^{q+1})$) implies convergence of order q (i.e., $\|e_n\|_\infty = \mathcal{O}(\Delta t^q)$), but we will see that the order of convergence can also be one larger than the order of consistency.

Let us first consider methods with classical order $p \geq 1$ that are not internally consistent, that is, $A_k e \neq A_l e$ for some k, l. Then the leading term in the local error is

$$d_n = \Delta t\, \boldsymbol{r}_n^T \sum_{k=1}^{r} (\boldsymbol{c} - \boldsymbol{A}_k \boldsymbol{e}) I_k u'(t_n) + \mathcal{O}(\Delta t^2)\,. \tag{5.42}$$

This gives an $\mathcal{O}(\Delta t)$ local error bound, which is of course quite poor. After all, d_n is the error that results after one step if $e_n = 0$. However, as we will see below, it can lead to convergence of order one.

Next assume the internal consistency condition (5.3) is satisfied, that is $A_k e = A_l e$ for $1 \leq k, l \leq r$. If $p = 1$ it follows directly that $\|d_n\|_\infty = \mathcal{O}(\Delta t^2)$. If $p \geq 2$ the leading term in the local discretization errors is given by

$$d_n = \Delta t^2 \boldsymbol{r}_n^T \sum_{k=1}^{r} \left(\tfrac{1}{2}\boldsymbol{c}^2 - \boldsymbol{A}_k \boldsymbol{c}\right) I_k u''(t_n) + \mathcal{O}(\Delta t^3)\,. \tag{5.43}$$

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.

For problems that are (mildly) stiff, such as semi-discrete systems from hyperbolic equations, the above derivation shows that *order reduction* is to be expected. This order reduction will appear primarily at interface points on the spatial grid, where the grid-functions $I_k u^{(j)}(t)$ have jumps. This is similar to the situation for standard Runge-Kutta methods, where order reduction appears at boundaries if the boundary values are time-dependent; see for instance the review with references in [14, Sect. II.2]. With the partitioned and multirate schemes, we are creating interfaces that act like (internal) boundaries with time-dependent boundary conditions.

Based on the local error behaviour, one would expect convergence with order one for the TW2 and SHV2 schemes, and lack of convergence for the scheme CS2. This is not what was seen in the numerical test in Section 4.1 for advection with a smooth solution. To obtain the correct (observed) order of convergence $q = 1, 2$, we need to study the propagation of the leading term in the local error. We already saw that the global error can be of the same order Δt^q as the local error if we have a suitable decomposition $d_n = (S_n - I)\xi_n + \eta_n$. In fact, we only need to study the principle term of the local error. It will be assumed that there exist vectors $\xi_n \in \mathbb{R}^m$, $n \geq 0$, such that

$$\left.\begin{aligned} &\left\| \left(\boldsymbol{r}_n^T \boldsymbol{e}\right)\xi_n \;-\; \Delta t^q \boldsymbol{r}_n^T \sum_{k=1}^{r} \tfrac{1}{q!}\left(\boldsymbol{c}^q - q\boldsymbol{A}_k \boldsymbol{c}^{q-1}\right) I_k u^{(q)}(t_n) \right\|_\infty = \mathcal{O}(\Delta t^{q+1})\,, \\ &\|\xi_n\|_\infty = \mathcal{O}(\Delta t^q)\,, \qquad \|\xi_{n+1} - \xi_n\|_\infty = \mathcal{O}(\Delta t^{q+1})\,. \end{aligned}\right\} \tag{5.44}$$

Then, following the proof of Theorem 3.3, we directly arrive at the following result.

Proposition 5.9 *Assume that (5.29) is valid, and let p be the (classical) order of the partitioned Runge-Kutta method.*
(i) If $p = 1$ and (5.44) holds with $q = 1$, then the method is convergent with order one in the maximum-norm.
(ii) Suppose that $p \geq 2$ and the method is internally consistent. Then, if (5.44) holds with $q = 2$, the method is convergent with order two in the maximum-norm.

The above result has been called a proposition, rather than a theorem, because it is far from clear how to verify the condition (5.44) in most situations of practical importance. In the next subsection we will consider this condition for a simple case: linear advection with first-order upwind spatial discretization. Of course, this is not the spatial discretization one would like to use with a high-order time stepping scheme, but it will give a heuristic explanation for the temporal orders observed in the accuracy experiment in Section 4.1.

Remark 5.10 The above expressions for the local errors are similar to those given in [13] for implicit-explicit Runge-Kutta methods, and in [19, 20] for a class of implicit additive Runge-Kutta methods with domain decomposition. Apart from the fact that these latter methods are implicit, because they are intended for parabolic problems, an interesting feature is that the matrices I_k are constructed from smooth grid functions, instead of the the step functions (zero-one entries) in this paper. This can have a positive influence on the accuracy of the schemes. ◇

5.3.3 Verification of condition (5.44) for linear advection

To study condition (5.44), let us consider linear problems with constant coefficients,

$$u'(t) = Au(t) + g(t)\,. \tag{5.45}$$

Denote $Z = \Delta t A$, $\boldsymbol{Z} = I \otimes Z$ with $I = I_{s\times s}$ the $s \times s$ identity matrix, and

$$\boldsymbol{r}(Z)^T = [r_1(Z), \ldots, r_s(Z)] = \Big(\sum_{k=1}^{r} \boldsymbol{b}_k^T \boldsymbol{I}_k \boldsymbol{Z}\Big)\Big(\boldsymbol{I} - \sum_{k=1}^{r} \boldsymbol{A}_k \boldsymbol{I}_k \boldsymbol{Z}\Big)^{-1}. \tag{5.46}$$

In this case we have $\boldsymbol{b}_k^T \boldsymbol{I}_k \boldsymbol{Z} = b_k^T \otimes I_k Z$ and $\boldsymbol{A}_k \boldsymbol{I}_k \boldsymbol{Z} = A_k \otimes I_k Z$. The matrices A_k are strictly lower triangular $s \times s$ matrices, and consequently a product of s such matrices vanishes. Writing the matrix inverse in (5.46) as a power series, it follows that

$$\boldsymbol{r}(Z)^T \boldsymbol{e} = \sum_{l=0}^{s-1} \sum_{k,j_1,\ldots,j_l=1}^{r} \big(b_k^T A_{j_1} \cdots A_{j_l} e\big)\, I_k Z\, I_{j_1} Z \cdots I_{j_l} Z\,. \tag{5.47}$$

In the same way it is seen that

$$\begin{aligned} &\boldsymbol{r}(Z)^T \sum_{i=1}^{r} \big(\boldsymbol{c}^q - q\boldsymbol{A}_i \boldsymbol{c}^{q-1}\big) I_i \\ = &\sum_{l=0}^{s-1} \sum_{k,j_1,\ldots,j_l,i=1}^{r} \big(b_k^T A_{j_1} \cdots A_{j_l} (c^q - q A_i c^{q-1})\big)\, I_k Z\, I_{j_1} Z \cdots I_{j_l} Z\, I_i\,, \end{aligned} \tag{5.48}$$

If there is a matrix $W \in \mathbb{R}^{m\times m}$ such that $\|W\|_\infty = \mathcal{O}(1)$ and

$$\big(\boldsymbol{r}(Z)^T \boldsymbol{e}\big) W = \boldsymbol{r}(Z)^T \sum_{i=1}^{r} \big(\boldsymbol{c}^q - q\boldsymbol{A}_i \boldsymbol{c}^{q-1}\big) I_i\,, \tag{5.49}$$

then we can take $\xi_n = \frac{1}{q!}\Delta t^q W u^{(q)}(t_n)$ in (5.44). Recall that $\|W\|_\infty = \mathcal{O}(1)$ means that W can be bounded uniformly in the mesh width and dimension m.

Consider as a simple example, the semi-discrete system (2.2) in $\mathbb{R}^m$ with $u_0(t) = 0$, corresponding to first-order upwind discretization of the advection equation 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 $\mathcal{I}_2 = \{j : \frac{1}{4}m < j \le \frac{3}{4}m\}$, 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 8 we have plotted the norm $\|W\|_\infty$ 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 SHV2 were similar to those of TW2. In this example, the matrix $\boldsymbol{r}(Z)^T\boldsymbol{e}$ is nonsingular, and it is well-conditioned for $\nu \le 1$. 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.

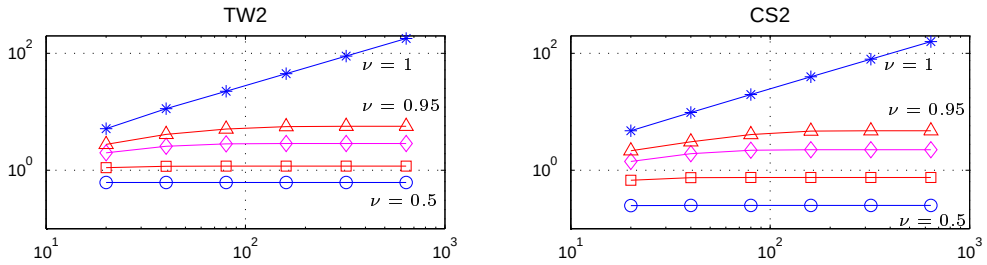

Figure 8: Norm $\|W\|_\infty$ versus $m = 20, 40, \ldots, 640$ for various values of $\nu = \Delta t/h$ with the schemes TW2 (left) and CS2 (right). Markers: $\circ$ for $\nu = 0.5$, $\square$ for $\nu = 0.75$, $\diamond$ for $\nu = 0.9$, $\triangle$ for $\nu = 0.95$ and $*$ for $\nu = 1$.

It is obvious that verification of condition (5.44) would be desirable for nonlinear problems and higher-order (nonlinear) spatial discretizations. Nevertheless, the combination of Proposition 5.9 and these experimental bounds for first-order advection discretization does provide a heuristic explanation for the numerical observations in Section 4.1 for the advection problem with smooth solution and WENO5 spatial discretization, where we saw convergence of the schemes TW2 and SHV2 with order two in the maximum-norm, and with order one for the CS2 scheme.

6 Final remarks

6.1 Partitioning based on fluxes

For conservation laws $u_t + f(u)_x = 0$, the semi-discrete system (2.8) will in general be of the form

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

Multirate methods can be based on these numerical fluxes $f_{j\pm 1/2}(u)$ rather than in terms of the components $F_j(u)$, and this is not well covered by the above formulations.

Suppose, as an example, that $\mathcal{I}_1 = \{j : j < i\}$ and $\mathcal{I}_2 = \{j : j \ge i\}$. Instead of $F = I_1F + I_2F$, we can consider the decomposition $F = F^1 + F^2$ with vector functions

F^1 and F^2 whose jth component is given by

$$\left.\begin{array}{lll} F_j^1(v) = \frac{1}{\Delta x_j}\big(f_{j-\frac{1}{2}}(v) - f_{j+\frac{1}{2}}(v)\big)\,, & F_j^2(v) = 0 & \text{for } j < i\,, \\ F_j^1(v) = \frac{1}{\Delta x_i} f_{i-\frac{1}{2}}(v)\,, & F_j^2(v) = \frac{-1}{\Delta x_i} f_{i+\frac{1}{2}}(v) & \text{for } j = i\,, \\ F_j^2(v) = \frac{1}{\Delta x_j}\big(f_{j-\frac{1}{2}}(v) - f_{j+\frac{1}{2}}(v)\big)\,, & F_j^1(v) = 0\,, & \text{for } j > i\,. \end{array}\right\} \tag{6.1}$$

We can consider any of the above schemes with $I_k F(v)$ replaced by $F^k(v)$. Since we are then dealing with fluxes, mass-conservation is guaranteed at any stage. However, there are two reasons why such schemes were not considered in this paper.

First, monotonicity assumptions such as (2.13) 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 (2.2). Writing this system as $u'(t) = Au(t)$, the above decomposition would correspond to $A = AI_1 + AI_2$, that is, $F^k = AI_k$, but it is easy to show that $\|I + \tau A I_k\|_\infty$ is larger than one for any $\tau > 0$.

Secondly, such a 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 (2.2), formula (2.10) with F^k replacing $I_k F$, $k = 1, 2$, leads to method (2.3) rather than (2.4). Using these F^1 and F^2 in (2.9) gives a completely inconsistent result.

6.2 Summary and conclusions

In this paper some multirate schemes based on the forward Euler method and the two-stage explicit trapezoidal rule have been analyzed. All these methods can be written as partitioned Runge-Kutta methods.

For the analysis of the monotonicity properties of the schemes we followed the TVD/SSP framework of [5, 24], assuming monotonicity of one forward Euler step with suitable local time steps. Different monotonicity thresholds were found for maximum-norm monotonicity and maximum principles on the one hand, and the TVD property on the other hand. However, these theoretical differences did not reveal themselves in the numerical tests. In practical situations, the threshold C found for maximum-norm monotonicity seems the most relevant.

Many multirate schemes are not internally consistent. This may lead to low accuracy at interface points. An analysis of the local discretization errors even suggests lack of convergence, but this is too pessimistic. Also for the other schemes, that are internally consistent, propagation of the leading local error terms has to be studied to understand the proper convergence behaviour.

Lack of mass conservation seems in many cases not a very serious defect because it only arises at interface points, so it will mainly be felt when a shock or very steep solution gradient passes such an interface. This conclusion is similar as in [26]. Of course, if mass conservation can be built in a scheme without affecting other essential properties, such as internal consistency and computational work per step, this is advisable. For the schemes considered in this paper lacking mass conservation we did not find such suitable modifications.

The use of a high-order Runge-Kutta methods as basis for a multirate scheme or a partitioned scheme will not directly lead to a high order of accuracy at interface points. The discretization errors have to be considered within the PDE context, leading to expressions for the local errors of the form (5.42) or (5.43). Regarding the semi-discrete as a fixed (non-stiff) ODE will in general lead to a too optimistic estimate of the rate of convergence.

A Appendix: a spatial discretization with TVD limiter on non-uniform grids

As an example of a discretization with limiting we will consider formulas on non-uniform grids that generalize the third-order upwind-biased scheme with the so-called Koren limiter on uniform grids.

A.1 Discretization and limiting

For a non-uniform grid with cells $\mathcal{C}_j = (x_j - \frac{1}{2}\Delta x_j, x_j + \frac{1}{2}\Delta x_j)$ and cell-average values u_j, the third-order upwind-biased spatial discretization can be derived by piecewise cubic reconstruction of the primitive grid-function $U_i = \sum_{j \le i} \Delta x_j u_j$ and differentiation.

On $\mathcal{C}_j$ we take $U(x)$ to be the cubic polynomial that passes through the points $(x_{j+k/2}, U_{j+k/2})$, $k = -3, -1, 1, 3$. Then the resulting values

$$u^R_{j-\frac{1}{2}} = U'(x_{j-\frac{1}{2}})\,, \qquad u^L_{j+\frac{1}{2}} = U'(x_{j+\frac{1}{2}})\,,$$

can be used as cell-boundary values in a numerical flux-function. In the following we only give the formulas for the left states $u^L_{j+1/2}$; those for $u^R_{j-1/2}$ are essentially the same, just the mirror image.

By some calculations (with Newton divided differences) it follows that

$$u^L_{j+\frac{1}{2}} = \gamma^L_{-1,j} u_{j-1} + \gamma^L_{0,j} u_j + \gamma^L_{1,j} u_{j+1}\,, \tag{A.1}$$

with coefficients $\gamma^L_{0,j} = 1 - \gamma^L_{-1,j} - \gamma^L_{1,j}$ and

$$\gamma^L_{-1,j} = \frac{-\Delta x_j \Delta x_{j+1}}{(\Delta x_{j-1} + \Delta x_j)(\Delta x_{j-1} + \Delta x_j + \Delta x_{j+1})}\,,$$
$$\gamma^L_{1,j} = \frac{(\Delta x_{j-1} + \Delta x_j)\Delta x_j}{(\Delta x_j + \Delta x_{j+1})(\Delta x_{j-1} + \Delta x_j + \Delta x_{j+1})}\,.$$

This provides the non-limited value.

To apply a limiter, we first write (A.1) in the form

$$u^L_{j+\frac{1}{2}} = u_j + \psi^*_j (u_{j+1} - u_j)\,, \qquad \psi^*_j = \frac{u^L_{j+\frac{1}{2}} - u_j}{u_{j+1} - u_j}\,. \tag{A.2}$$

Next we apply a limiter to this ψ^*_j,

$$\psi_j = \max\big(0\,,\, \min\big(1\,,\, \psi^*_j\,,\, \theta_j\big)\big)\,, \qquad \theta_j = \frac{u_j - u_{j-1}}{u_{j+1} - u_j}\,, \tag{A.3}$$

to obtain the limited value

$$u^L_{j+\frac{1}{2}} = u_j + \psi_j (u_{j+1} - u_j)\,. \tag{A.4}$$

This kind of limiting is often called 'target limitering' because the limited values are taken as close as possible to a target scheme (which is in our case the non-limited scheme) within the monotonicity constraints. It can be applied to any scheme producing non-limited values $u^L_{j+1/2}$. From (A.1), (A.2) it is seen that $\psi^*_j = \gamma^L_{1,j} - \gamma^L_{-1,j}\theta_j$, and therefore the limiter can also be written as

$$\psi_j = \max\big(0\,,\, \min\big(1\,,\, \gamma^L_{1,j} - \gamma^L_{-1,j}\theta_j\,,\, \theta_j\big)\big)\,. \tag{A.5}$$

To see that (A.4) will indeed introduce a spatial discretization with certain monotonicity properties, such as positivity and TVD, note that

$$u^L_{j-\frac{1}{2}} - u^L_{j+\frac{1}{2}} = \rho_j (u_{j-1} - u_j)\,, \qquad \rho_j = 1 - \psi_{j-1} + \psi_j \,/\, \theta_j\,.$$

In view of (A.3) we have $0 \le \psi_{j-1} \le 1$ and $0 \le \psi_j/\theta_j \le 1$, and therefore

$$0 \le \rho_j \le 2\,.$$

As explained in Example 2.2, this guarantees max-norm monotonicity and the TVD property for $u_t + f(u)_x = 0$ with $f'(u) \ge 0$ (for the relevant range of u values).

As mentioned already above, the formulas for the right states $u^R_{j-1/2}$ are essentially the same (reflexion around $x_{j-1/2}$), and these will be used if we have $f'(u) < 0$ for all (relevant) u values. With an arbitrary flux function $f(u)$ a suitable flux splitting is to be used, for example the simple Lax-Friedrich splitting given in [16, 23].

Remark A.1 The numerical fluxes $f_{j+1/2}(u) = f(u_{j+1/2})$ of the limited discretization are Lipschitz continuous,

$$|f_{j+1/2}(\tilde u) - f_{j+1/2}(u)| \;\le\; L\|\tilde u - u\|_\infty$$

for all $\tilde u = [\tilde u_j]$, $u = [u_j] \in \mathbb{R}^m$. This is not obvious from (A.3), (A.5), because the ratios θ_j will not satisfy a Lipschitz condition. However, if we denote $\sigma_j = u_{j+1} - u_j$, then by considering the different sign possibilities it is seen that

$$u^L_{j+\frac{1}{2}} = u_j + \mathrm{sign}(\sigma_j) \min\left(|\sigma_j|\,,\, \gamma^L_{1,j}|\sigma_j| - \gamma^L_{-1,j}|\sigma_{j-1}|\,,\, |\sigma_{j-1}|\right)$$

if $\mathrm{sign}(\sigma_j) = \mathrm{sign}(\sigma_{j-1})$, and $u^L_{j+1/2} = u_j$ otherwise. From this the Lipschitz condition can be deduced, with Lipschitz constant L determined by the actual grid. ◇

A.2 Accuracy test

Consider the advection equation $u_t + u_x = 0$, $0 < x, t < 1$, with spatial periodicity and initial value $u(x,0) = \sin^4(\pi x)$. The relative L_1-errors of the spatial discretization are given in Table 3 for various grids with m points, $m = 20, 40, 80, 160$. These results are to be compared with those in Appendix B of [1]. The random grids are chosen by first generating random numbers $\sigma_j \in [\frac{1}{2}, 1]$ and then setting $\Delta x_j = \sigma_j / \sum_{k=1}^m \sigma_k$. The grids indicated by 'Block1' and 'Block2' are cyclic repetitions of $(\Delta x_1, \Delta x_2, \Delta x_3, \Delta x_4) = (h, 2h, 3h, 4h)$ and $(\Delta x_1, \Delta x_2, \Delta x_3, \Delta x_4) = (h, 2h, 10h, 11h)$, respectively, with appropriate $h = 4/(10m)$, $h = 4/(14m)$, respectively.

The results compare favourably to those in [1], where it should be noted that the random grid used here has more variation in [1] and also the initial profile has been slightly changed to make it periodic.

We also note that the above limiter does not fit into the framework of slope limiting with linear reconstruction considered in [1]. There it is required that on each cell $\mathcal{C}_j$ we have an approximation $u(x) = u_j + (x - x_j)s_j$, with slope s_j that may be limited, and then

$$u^R_{j-\frac{1}{2}} = u_j - \tfrac{1}{2}\Delta x_j s_j\,, \qquad u^L_{j+\frac{1}{2}} = u_j + \tfrac{1}{2}\Delta x_j s_j\,.$$

To achieve this in the above algebraic framework one needs a certain 'symmetry' condition to ensure that u_j is the average of $u^R_{j-1/2}$ and $u^L_{j+1/2}$.

The spatial discretization used in [3] is of the same form as (A.5) but with different coefficients $\gamma_{k,j}$. In the above accuracy test this scheme gave less accurate results,

Table 3: Relative L_1-errors for scalar advection on non-uniform grids

	Uniform	Random	Block 1	Block 2
Non-lim., $m = 20$	$4.79 \cdot 10^{-2}$	$5.14 \cdot 10^{-2}$	$6.06 \cdot 10^{-2}$	$9.65 \cdot 10^{-2}$
Non-lim., $m = 40$	$6.82 \cdot 10^{-3}$	$7.49 \cdot 10^{-3}$	$9.13 \cdot 10^{-3}$	$1.58 \cdot 10^{-2}$
Non-lim., $m = 80$	$8.70 \cdot 10^{-4}$	$9.49 \cdot 10^{-4}$	$1.18 \cdot 10^{-3}$	$2.05 \cdot 10^{-3}$
Non-lim., $m = 160$	$1.09 \cdot 10^{-4}$	$1.19 \cdot 10^{-4}$	$1.49 \cdot 10^{-4}$	$2.60 \cdot 10^{-4}$
Limited, $m = 20$	$6.57 \cdot 10^{-2}$	$6.79 \cdot 10^{-2}$	$9.35 \cdot 10^{-2}$	$1.45 \cdot 10^{-1}$
Limited, $m = 40$	$1.36 \cdot 10^{-2}$	$1.49 \cdot 10^{-2}$	$2.02 \cdot 10^{-2}$	$3.32 \cdot 10^{-2}$
Limited, $m = 80$	$2.65 \cdot 10^{-3}$	$2.97 \cdot 10^{-3}$	$4.25 \cdot 10^{-3}$	$7.56 \cdot 10^{-3}$
Limited, $m = 160$	$4.97 \cdot 10^{-4}$	$5.73 \cdot 10^{-4}$	$8.11 \cdot 10^{-4}$	$1.58 \cdot 10^{-3}$

due to the fact that then the non-limited scheme is only of order two. The errors with limiter were then a factor three to four larger than in Table 3 on the fine grids, $m = 160$.

Finally we note that the limited schemes used in [26] are based on scaled ratios $\theta_j = \sigma_{j-1}/\sigma_j$ with $\sigma_k = (u_{k+1} - u_k)/\Delta x_k$. It is not too difficult to show that such schemes are not TVD or positivity preserving, but in tests they do perform quite well; there are overshoots, but these are very minor. Nevertheless, to remain within the theoretical framework outlined in Section 2.3, the discretization (A.5) seems preferable.

References

[1] M. Berger, M.J. Aftosmis, S.M. Murman, *Analysis of slope limiters on irregular grids*. AIAA Paper 2005-0490, 2005.

[2] K. Burrage, J.C. Butcher, *Nonlinear stability of a general class of differential equation methods*. BIT 20 (1980), 185–203.

[3] E.M. Constantinescu, A. Sandu, *Multirate timestepping methods for hyperbolic conservation laws*. Report TR-06-15 (913), Dept. Comp. Sc. Virginia Tech, 2006.

[4] C. Dawson, R. Kirby, *High resolution schemes for conservation laws with locally varying time steps*. SIAM J. Sci. Comput. 22 (2000), 2256–2281.

[5] S. Gottlieb, C.-W. Shu, E. Tadmor, *Strong stability preserving high-order time discretization methods*. SIAM Review 42 (2001), 89–112.

[6] M. Günther, A. Kværnø, P. Rentrop, *Multirate partitioned Runge-Kutta methods*. BIT 41 (2001), 504–514.

[7] L. Ferracina, M.N. Spijker, *An extension and analysis of the Shu-Osher representation of Runge-Kutta methods*. Math. Comp. 74 (2005), 201–219.

[8] A. Harten, *High resolution schemes for hyperbolic conservation laws*. J. Comput. Phys. 49 (1983), 357–393.

[9] E. Hairer, S.P. Nørsett, G. Wanner, *Solving Ordinary Differential Equations I – Nonstiff Problems*. Second edition, Springer Series Comput. Math. 8, Springer, 1993.

[10] I. Higueras, *Representations of Runge-Kutta methods and strong stability preserving methods.* SIAM J. Numer. Anal. 43 (2005), 924–948.

[11] I. Higueras, *Strong stability for additive Runge-Kutta methods.* SIAM J. Numer. Anal. 44 (2006), 1735–1758.

[12] R.A. Horn, C.R. Johnson, *Matrix Analysis.* Cambridge University Press, 1985.

[13] W. Hundsdorfer, S.J. Ruuth, *IMEX extensions of linear multistep methods with general monotonicity and boundedness properties.* CWI Report MAS-E0621, Amsterdam, 2006. To appear in J. Comput. Phys.

[14] W. Hundsdorfer, J.G. Verwer, *Numerical Solution of Advection-Diffusion-Reaction Equations.* Springer Series Comput. Math. 33, Springer, 2003.

[15] R. Kirby, *On the convergence of of high resolution methods with multiple time scales for hyperbolic conservation laws.* Math. Comp. 72 (2003), 1239–1250.

[16] R.J. LeVeque, *Finite Volume Methods for Hyperbolic Problems.* Cambridge Texts in Appl. Math., Cambridge Univ. Press, 2002.

[17] N.M. Maurits, H. van der Ven, A.E.P. Veldman, *Explicit multi-time stepping methods for convection dominated flow problems.* Comput. Meth. Appl. Mech. Engrg. 157 (1998), 133–150.

[18] S. Osher, R. Sanders, *Numerical approximations to nonlinear conservation laws with locally varying time and space grids.* Math. Comp. 41 (1983), 321–336.

[19] L. Portero, B. Bujanda, J.C. Jorge, *A combined fractional step domain decomposition method for the numerical integration of parabolic problems*, Lect. Notes in Comp. Sc. 3019 (2004), 1034–1041.

[20] L. Portero, *Fractional step Runge-Kutta methods for multidimensional evolutionary problems with time-dependent coefficients and boundary conditions.* Thesis, Univ. of Navarra, Pamplona, 2007.

[21] V. Savcenco, *Comparison of the asymptotic stability properties for two multirate strategies*, CWI Report MAS-R0705, Amsterdam, 2007.

[22] V. Savcenco, W. Hundsdorfer, J.G. Verwer, *A multirate time stepping strategy for stiff ordinary differential equations.* BIT 47 (2007), 137–155.

[23] C.-W. Shu, *High order ENO and WENO schemes for computational fluid dynamics.* In: *High-Order Methods for Computational Physics*, Eds. T.J. Barth, H. Deconinck, Lect. Notes Comp. Sc. Eng. 9, Springer, 1999, 439–582.

[24] C.-W. Shu, S. Osher, *Efficient implementation of essentially non-oscillatory shock-capturing schemes.* J. Comput. Phys. 77 (1988), 439–471.

[25] M.N. Spijker, *Stepsize restrictions for general monotonicity in numerical initial value problems.* SIAM J. Numer. Anal. 45 (2007), 1226–1245.

[26] 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.