Unconditionally Strong Stability Preserving Extensions of the TR-BDF2 Method

L. Bonaventura¹ · A. Della Rocca^{1,2}

Received: 25 October 2015 / Revised: 1 August 2016 / Accepted: 9 August 2016 / Published online: 17 August 2016

Abstract We analyze monotonicity, strong stability and positivity of the TR-BDF2 method, interpreting these properties in the framework of absolute monotonicity. The radius of absolute monotonicity is computed and it is shown that the parameter value which makes the method L-stable is also the value which maximizes the radius of monotonicity. In order to achieve unconditional monotonicity, hybrid variants of TR-BDF2 are proposed, that reduce the formal order of accuracy, while keeping the native L-stability property, which is useful for the application to stiff problems. Numerical experiments compare these different hybridization strategies to other methods used in stiff and mildly stiff problems. The results show that the proposed strategies provide a good compromise between accuracy and robustness at high CFL numbers, without suffering from the limitations of alternative approaches already available in literature.

Keywords TR-BDF2 \cdot Runge–Kutta \cdot Absolute monotonicity \cdot Positivity preservation \cdot TVD \cdot SSP

1 Introduction

The classical error analysis of numerical methods for ordinary differential equations (ODE) does not yield sufficient criteria for the conservation of special properties of the continuous

¹ Dipartimento di Matematica, Politecnico di Milano, MOX – Modelling and Scientific Computing, Via Bonardi 9, 20133 Milano, Italy

Partially supported by the INDAM - GNCS project 'Metodi numerici semi-impliciti e semi-Lagrangiani per sistemi iperbolici di leggi di bilancio'.

L. Bonaventura luca.bonaventura@polimi.it

A. Della Rocca alessandro.dellarocca@polimi.it; alessandro.dellarocca@tenova.com

² Tenova S.p.A., Global R&D, Via Albareto 31, 16153 Genova, Italy

This is a post-peer-review, pre-copyedit version of an article published in JOURNAL OF SCIENTIFIC COMPUTING[. The final authenticated version is available online at: http://dx.doi.org/10.1007% 2Fs10915-016-0267-9.

solutions during time integration. In many applications, for instance, the solution is required to remain non negative, or to take values in a certain range, or to preserve the total variation as a function of the space variable, when a time and space dependent partial differential equation is solved numerically. A number of different strategies have been proposed to address each of these issues separately, see, e.g. [5,32,42,43]. Especially in the ODE literature, many of these problems are empirically resolved by step size adaptation strategies that complement traditional ODE methods. However, in many applications, such as numerical weather prediction, environmental fluid dynamics or turbulent reactive flow simulations, the step size is usually kept fixed and/or relatively large, in order to minimize the number of expensive function evaluations required from the complex source terms involved. While dynamic time step adaptation [22] and multirate approaches [8] can overcome these problems, in this work we will study a robust but accurate fixed time step approach that can guarantee a good compromise between preservation of some relevant monotonicity properties, accuracy and efficiency. Splitting approaches [47] are commonly used to couple complex source terms to the discretized fluid dynamics in a relatively simple way, but we will restrict our attention to methods that do not resort to operator splitting, which may entail a loss of accuracy for advection-diffusion-reaction problems with space dependent source terms.

We will focus on the analysis of the monotonicity properties of the TR-BDF2 method, that was introduced in [1] and successively reformulated and analyzed in [28]. This second order accurate, A-stable and L-stable method has a number of interesting properties and it has been recently used with success in [48] as the key ingredient of an efficient semiimplicit, semi-Lagrangian discretization of fluid dynamics equations representative of many environmental models. It is therefore of interest to understand to which extent this method can also guarantee positivity and monotonicity for the equations of advected species. We will show that TR-BDF2 is conditionally monotone under a time step restriction that allows for time steps more than double with respect to those of explicit schemes. Our analysis relies on the results in [12,13,19,23,37,46], that allow to derive sufficient conditions for properties like positivity, monotonicity and total variation preservation in the unified framework of an extended monotonicity. We then propose two different modifications of the TR-BDF2 method, both based on a hybridization with the unconditionally monotone implicit Euler method. In this way, accuracy is sacrificed locally in space or time in order to preserve monotonicity, independently of the time step and stiffness of the problem. Other approaches, focusing specifically on the equations of chemical kinetics, have been proposed under more restrictive assumptions in [4,15,49]. The present approach represents an improvement over these results, since it does not require source term splitting, it is not limited to non stiff problems as [4], nor it requires a special form of the source term as [15] and differently than [49] it is second order when critical solution properties are not violated under the selected time step size.

Within the context of conservation laws, an alternative approach for enforcing monotonicity in implicit methods introduces the so called time-limiters [11] and obtains non-linear time integration schemes from single step methods. The TVD property is guaranteed by locally switching to the implicit Euler method on the basis of a parameter, whose cell value is assigned by checking the monotonicity condition on a quadratic temporal interpolant. A similar space-time limited method is analyzed in [10]. In both cases, the evolution is no longer space-time decoupled. Another strategy to relax step size monotonicity conditions when discretizing hyperbolic PDEs is introduced in [33], where a combination of upwind and downwind biased space semi-discretizations is analyzed and a family of downwind biased, second order, fully implicit Runge–Kutta methods with arbitrarily large SSP coefficients is derived. In contrast to the methods above, our approach is not limited to systems derived from the spatial discretization of conservation laws. Monotonicity is enforced here by combining an implicit high order single step method with a first order unconditionally monotone method. This coupling is analyzed in two possible variants. The second of these strategies consists of an automatically partitioned DIRK method and it is formally equivalent to an implicit-implicit ARK scheme. Thus it represents both an extension and an implicit variant of the SPERK schemes introduced in [36].

In Sect. 2, the theory of monotone and SSP methods is reviewed. The TR-BDF2 method is presented in detail in Sect. 3, where its absolute monotonicity property is analyzed. Two strategies to improve its monotonicity properties irrespective of time step size are described in Sect. 4, stemming from the classical results of absolute monotonicity. Other competitive time discretization approaches are introduced in Sect. 5 and interpreted under the SSP theory whenever possible. In Sect. 6 we present an empirical assessment of the properties of all these time integration methods in a variety of relevant benchmark problems. Conclusions, results and directions of further investigation are summarized in Sect. 7.

2 Review of Monotonicity and Strong Stability Results

We review the recent progresses in the field of strong stability preserving (SSP) methods introduced, among others, in [12,17,19,23] and [44]. In this work we consider an initial value problem for a system of ordinary differential equations (ODEs) of type

$$u'(t) = f(t, u(t)) \text{ and } t \in [0, T],$$
 (1a)

$$u(0) = u^0. \tag{1b}$$

We assume that $u^0 \in \mathbb{R}^m$, $f : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ such that the problem (1) has a unique solution. Moreover we assume also that $\|\cdot\| : \mathbb{R}^m \to \mathbb{R}$ is a convex functional

$$\|\lambda v + (1 - \lambda) w\| \le \lambda \|v\| + (1 - \lambda) \|w\|$$

for $0 \le \lambda \le 1$ and $v, w \in \mathbb{R}^m$. We shall deal with numerical methods for finding a numerical approximation u^n to the exact solution values u(n h), where *h* is a positive step size, assumed for simplicity to be constant and equal to $h = T/N_t$, with *T* the final time of integration and $n = 1, \ldots, N_t$.

Monotonicity of the total variation seminorm, positivity or range boundedness are all nonlinear properties that can be seen as specific realizations of some form of monotonicity. Even the discrete maximum principle and contractivity, a property relevant for the numerical stability of time integration methods, can be reintepreted as forms of monotonicity, see the review in "Appendix 1" . The recent theory of SSP integration methods [17] shows that all these concepts are strongly related. As a consequence, they can be guaranteed during the numerical integration using the same approach. In this section, we provide a synthetic presentation of this framework, that is based on the classical results in [37]. We will focus in particular on the properties of s-stages Runge–Kutta methods (RK)

$$g^{i} = u^{n} + h \sum_{j=1}^{s} a_{ij} f(t^{n} + c_{j}h, g^{j}) \quad (i = 1, \dots, s)$$
 (2a)

$$u^{n+1} = u^n + h \sum_{i=1}^{s} b_i f(t^n + c_i h, g^i)$$
(2b)

where a_{ij} , b_i and c_i are real parameters which characterize the method and g^i are the intermediate stages. It is usually assumed that $\sum_{j=1}^{s} a_{ij} = c_i$. The method is explicit if $a_{ij} = 0$ for $j \ge i$ and implicit otherwise. The parameters of the method are traditionally collected in compact form in the Butcher tableau as an $s \times s$ matrix $A = (a_{ij})$, a row vector $b = (b_1, \ldots, b_s)^T$ and a column vector $c = (c_1, \ldots, c_s)$. For homogeneous initial value problems, the coefficients c_i are not relevant, so that in this cases each RK method is completely identified by its coefficients (A, b).

Definition 1 *Monotonicity*. The RK method (2) is monotone with respect to the functional $\|\cdot\|$ if $\|u^n\| \le \|u^{n-1}\|$ for $n = 1, ..., N_t$ under the assumption that

$$\|u + hf(t, u)\| \le \|u\| \quad \text{for } 0 < h \le \tau_0.$$
(3)

In literature this is also commonly denoted as external monotonicity, in contrast to the internal monotonicity property

$$\|g^i\| \le \|u^{n-1}\|$$
 for $1 \le i \le s$. (4)

Assumption (3), commonly found in many references, see, e.g. [12,23-25,30,31,34,43] and [44], essentially amounts to define τ_0 as the maximum time step under which the explicit Euler method is monotone. In these studies, the critical step size for monotonicity is determined such that Definition 1 is verified for all

$$0 < h \le c \,\tau_0 \tag{5}$$

thus making the RK method *conditionally monotone*. If Definition 1 is verified for any step size h, then the method is called *unconditionally monotone*. In assessing monotonicity of different RK methods, the interest lies usually in determining the *maximal step size coefficient* c such that a time integration method is conditionally monotone.

Frequently, the convex functional $\|\cdot\|$ is intended either as the supremum norm $\|x\| = \|x\|_{\infty} = \sup_i |x_i|$ or as the total variation seminorm $\|x\| = \|x\|_{TV} = \sum_i |x_{i+1} - x_i|$, where x_i are the components of the vector x. We remind that $\|x\|_{TV}$ is a seminorm since it may vanish for $x \neq 0$ when $x_i = C$. Numerical methods statisfying Definition 1 under the total variation seminorm are called *total variation diminishing (TVD)*. They are especially important in the numerical solution of hyperbolic conservation laws, see e.g. [32, 39, 44].

For some classes of initial value problems, the properties of positivity and range boundedness play important roles in obtaining physically meaningful numerical solutions. Furthermore, due to the nonlinear form of some problems, the ability to maintain the native properties of the analytic solutions also in the numerical approximation is important in order to guarantee numerical stability of time integrations. In the following definitions, inequalities involving vectors should be interpreted compontentwise.

Definition 2 *Positivity.* The RK method (2) is positive if, whenever $u^0 \ge 0$, it guarantees that $u^{n+1} \ge 0$ under the assumption that

$$u + hf(t, u) \ge 0 \text{ for } 0 < h \le \tau_0.$$
 (6)

Definition 3 *Range boundedness.* The RK method (2) is range bounded in $[\chi, \psi]$ if, whenever $\chi \le u^0 \le \psi$, it guarantees that $\chi \le u^{n+1} \le \psi$ under the assumption that

$$\chi \le u + hf(t, u) \le \psi \quad \text{for } 0 < h \le \tau_0. \tag{7}$$

For linear methods of order $p \ge 2$, neither of these properties is guaranteed in general. When they are, they only hold if a time step limitation of the form (5) is respected. Even though these properties are formally different from the property of Definition 1, they can be derived from monotonicity after proper assumptions on the function f. In this respect, the generalization of monotonicity to arbitrary sublinear functionals $\|\cdot\|$ becomes relevant. Following the presentation in [29] and [46], it is useful to introduce two sublinear functionals, denoted respectively as the *floor* and *ceil* functional

$$\|u\|_{\lfloor\chi\rfloor} = -\min_{j}(\chi, u_{j}) \tag{8a}$$

$$\|u\|_{\lceil\psi\rceil} = \max_{j}(\psi, u_j).$$
(8b)

These functionals are not seminorms, since they both violate property $\|\lambda v\| = |\lambda| \|v\|$ for $\lambda = -1$. Using both functionals, the range boundedness Definition 3 naturally follows from the monotonicity Definition 1, while setting $\chi = 0$ in the floor functional the positivity Definition 2 is recovered. As a consequence, by introducing the floor and ceil functionals it possible to recast Definitions 2 and 3 in a form similar to Definition 1. Thus, positivity and range boundedness can just be interpreted as different forms of monotonicity. Similar considerations arise when guaranteeing the discrete maximum principle, see the definition in "Appendix 1".

All these properties appear as different forms of monotonicity under a proper choice of the convex functional $\|\cdot\|$. It is thus possible to extend the analytical results from the preservation of a specific property to that of other properties. It is thus advantageous to extend the contractivity results from [37], which again can be interpreted as a particular form of monotonicity, as explained in "Appendix 1". In our next discussion we will focus on *irreducible* RK methods, which are the only ones practically relevant. For a definition of irreducibility we refer to [21]. Following [37], we start from the absolute monotonicity of a RK method, analyzed in [37] to investigate necessry and sufficient conditions for contractivity. Define for $\xi \in \mathbb{R}$ the quantities

$$A(\xi) = A(I - \xi A)^{-1}, \qquad b^{\mathsf{T}}(\xi) = b^{\mathsf{T}}(I - \xi A)^{-1}, e(\xi) = (I - \xi A)^{-1}e, \qquad \varphi(\xi) = 1 + \xi b^{\mathsf{T}}(I - \xi)A^{-1}e.$$
(9)

Definition 4 Absolute monotonicity of RK. An irreducible *s*-stage RK with scheme (A, b) is absolutely monotone at $\xi \in \mathbb{R}$ if $A(\xi) \ge 0$, $b(\xi) \ge 0$, $e(\xi) \ge 0$ and $\varphi(\xi) \ge 0$ elementwise.

For the stability function φ , this entails that $\frac{d^k \varphi}{dz^k}(\xi) \ge 0$ for any $k \ge 0$, since the rationale behind Definition 4 lies in the Taylor expansion of some characteristic functions of a RK method, including the stability function. The quantities (9) form the coefficients of such expansions, see [26] and [37] for further details. These notations are useful to introduce the radius of absolute monotonicity.

Definition 5 *Radius of absolute monotonicity.* An *s*-stage RK with scheme (A, b) and $A \ge 0$ and $b \ge 0$ is characterized by its radius of absolute monotonicity, defined for all ξ in $-r \le \xi \le 0$ as

$$R(A, b) = \sup\{r | r \ge 0, A(\xi) \ge 0, \ b^{\mathsf{T}}(\xi) \ge 0, \ e(\xi) \ge 0, \ \varphi(\xi) \ge 0\}.$$
(10)

More compact definitions of the radius of absolute monotocity involve the use of matrices derived from the Butcher tableau, see, e.g., [35], but for the purpose of this work we found the use of the original definition more convenient.

In recent literature [17, 19], methods satisfying Definition 1 for a general convex functional are called *strong-stability preserving (SSP)*, in order to specify their ability to preserve any convex functional bound. Thus, they generalize classical TVD methods specifically developed for hyperbolic conservation laws.

Definition 6 Strong stability preserving (SSP). The RK method (2) is SSP with respect to the functional $\|\cdot\|$ if $\|u^n\| \le \|u^0\|$ under the assumption that

$$\|u + hf(t, u)\| \le \|u\| \quad \text{for } 0 < h \le \tau_0.$$
(11)

The SSP coefficient is the largest constant $c \ge 0$ such that this definition is verified for all $0 < h \le c\tau_0$.

The definition above closely follows Definition 1 for monotonicity. The SSP coefficient c turns out to be intimately related to the radius of absolute monotonicity.

Theorem 1 ([13], Theorem 3.4) For an irreducible RK, c = R(A, b).

Recent studies focused on the search of SSP-optimal RK methods having large SSP coefficients. While explicit SSP RKs are known since the seminal work in [44], a search for implicit SSP RKs started only recently in [14] and [35], where it was found that the SSP-optimal implicit RKs of order p = 2 and p = 3 are indeed SDIRK, while the optimal methods for p = 4 are DIRK. In the quest for improved monotonicity, the classes of *two-step RKs* [34] and *diagonally split RK* [2,3,41] have been investigated in the literature, with mixed success. Starting from the framework introduced above, in our analysis we will consider all these properties just as specific realizations of absolute monotonicity. As a consequence, in the next sections they will be briefly referred to as *monotonicity* of the time integration methods.

3 Strong Stability Preservation for the TR-BDF2 Method

We analyze here the monotonicity properties of the method originally introduced in [1] as TR-BDF2 (Trapezoidal Rule-Backward Differentiation Formula 2) method and successively reformulated in [28] as a DIRK. The same method was rediscovered in [6] and has been applied also in [16] to treat the implicit terms in an additive Runge–Kutta (ARK) scheme. A semi-Lagrangian reinterpretation of this method has recently been proposed in [48] for application to fluid mechanics problems in subcritical regimes.

We rely on the monotonicity and contractivity results reviewed in Sect. 2. In its original formulation, TR-BDF2 is defined as a one-step method resulting from the composition of the trapezoidal rule in the first substep, followed by BDF2 in the second substep. This combination is empirically justified under the rationale of combining the good accuracy of the trapezoidal rule with the stability and damping of fast modes guaranteed by BDF2. The TR-BDF2 method is

$$u^{n+\gamma} - \frac{\gamma}{2}hf^{n+\gamma} = u^n + \frac{\gamma}{2}hf^n$$
(12a)

$$u^{n+1} - \frac{(1-\gamma)}{(2-\gamma)} h f^{n+1} = \frac{1}{\gamma(2-\gamma)} u^{n+\gamma} - \frac{(1-\gamma)^2}{\gamma(2-\gamma)} u^n$$
(12b)

where $\gamma \in (0, 1)$ is a parameter whose value determines the stability and monotonicity properties of the method. By requiring that both stages have the same Jacobian, the value $\gamma = 2-\sqrt{2}$ was derived in [1]. As outlined in [28], the TR-BDF2 method can be rewritten as a DIRK method. (Indeed the TR-BDF2 method is embedded in a (2, 3) Runge–Kutta pair, as shown in [28], thus allowing an efficient estimation of the time discretization error, in case step size adaptation is required.) However, rather than closely following this reference, we will first derive the DIRK family associated to (12) before imposing the condition on the Jacobian. The Butcher tableau for this DIRK family is then



We remark that, by imposing the condition $a_{22} = a_{33}$, the value $\gamma = 2 - \sqrt{2}$ is readily obtained. The stability function associated to the above DIRK family is

$$\varphi(z) = \frac{\det(I - zA + zeb^{\mathsf{T}})}{\det(I - zA)} = \frac{[1 + (1 - \gamma)^2]z + 2(2 - \gamma)}{2(2 - \gamma)(1 - z\frac{\gamma}{2})(1 - z\frac{1 - \gamma}{2 - \gamma})}.$$
(13)

The rational polynomial (13) has a single zero in $z = \frac{2(\gamma-2)}{1+(1-\gamma)^2}$, which is always negative for all possible values of parameter γ , and two poles respectively in $z = \frac{2}{\gamma}$ and $z = \frac{2-\gamma}{1-\gamma}$ respectively, which are always positive. (While in [1] it is argued that there is a double pole at $z = \frac{2}{\gamma}$, this is true only for the numerical value of γ satisfying $a_{22} = a_{33}$.) Additionally, from results in "Appendix 1" it is immediate to find out that the stage order of TR-BDF2 is indeed $\tilde{p} = 2$, thus making the method resilient to order reduction in stiff problems (see, e.g., [23,41]).

We analyze the monotonicity properties of the DIRK family generalizing the TR-BDF2 by following [12, 13, 37].

Proposition 1 The radius of absolute monotonicity of the DIRK family at the base of TR-BDF2 is $R(A, b) = \frac{2(2-\gamma)}{1+(1-\gamma)^2}$ and it is maximized for $\gamma = 2 - \sqrt{2}$, which gives $R(A, b) \approx$ 2.414.

Proof Following Definition 5 and setting $\beta = 1 - \xi(1 - \gamma)/(2 - \gamma)$, we find

$$A(\xi) = \begin{bmatrix} 0 & 0 & 0 \\ \frac{\frac{\gamma}{2}}{1 - \frac{\gamma}{2}\xi} & \frac{\frac{\gamma}{2}}{1 - \frac{\gamma}{2}\xi} & 0 \\ \frac{1}{2(2 - \gamma)(1 - \frac{\gamma}{2}\xi)\beta} & \frac{1}{2(2 - \gamma)(1 - \frac{\gamma}{2}\xi)\beta} & \frac{1 - \gamma}{(2 - \gamma)\beta} \end{bmatrix}$$
$$b^{\mathsf{T}}(\xi) = \begin{bmatrix} \frac{\beta - \xi\gamma/2}{2(2 - \gamma)} \\ \frac{\beta - \xi\gamma/2}{2(2 - \gamma)} \\ \frac{(1 - \gamma)\beta}{2 - \gamma} \end{bmatrix} \quad e(\xi) = \begin{bmatrix} 1 \\ \frac{1 + \frac{\gamma}{2}\xi}{1 - \frac{\gamma}{2}\xi} \\ \frac{[1 + (1 - \gamma)^{2}]\xi + 2(2 - \gamma)}{2(2 - \gamma)(1 - \xi\frac{\gamma}{2})\beta} \end{bmatrix}$$
$$\varphi(\xi) = \frac{[1 + (1 - \gamma)^{2}]\xi + 2(2 - \gamma)}{2(2 - \gamma)(1 - \xi\frac{\gamma}{2})\beta}$$

from which by direct calculations we obtain the result.

From Theorem 1, we conclude that this is also the step size coefficient c for conditional monotonicity of the method under arbitrary seminorms and sublinear functionals in any nonlinear problem.

4 Two Unconditionally Monotone Variants of TR-BDF2

The classical results on contractivity order barriers reviewed in "Appendix 1", lead to the conclusion that there are no unconditionally monotone RK methods of order higher than one. For stiff initial value problems, this implies that even implicit higher order RK methods will always be subject to a CFL-like condition for monotonicity. Thus, the only RK method that does not need to comply with a time step restriction and that can be safely used without time step adaption is in practice the implicit Euler method. Due to such limitation, which is particularly relevant for problems with chemical kinetics, some solution methods found in literature rely exclusively on it, sacrificing accuracy for improving stability and consistency [50].

We propose here two hybridization strategies of TR-BDF2 with the implicit Euler method, that can be activated using a sensor detecting violations of relevant functional bounds. These hybrid schemes bring time integration back to a first order unconditionally monotone method whenever the sensor detects a violation of a selected functional bound during the current integration. Empirically, this local loss in accuracy should not degrade too much the solution accuracy if the time step is not too large. Our proposed methods apply this safe mode whenever it is expected from SSP theory that TR-BDF2 may produce non-monotone solutions.

The hybrid TR-BDF2 method is thus obtained by introducing a weighting parameter $\alpha \in [0, 1]$ in both stages of TR-BDF2

$$u^{n+\gamma} - \gamma h\left(1 - \frac{\alpha}{2}\right) f^{n+\gamma} = u^n + \gamma h \frac{\alpha}{2} f^n$$
(14a)

$$u^{n+1} - \frac{(1-\gamma)h}{\alpha(1-\gamma)+1} f^{n+1} = \frac{\alpha(\frac{1}{\gamma}-1)+1}{\alpha(1-\gamma)+1} u^{n+\gamma} - \frac{\alpha}{\alpha(1-\gamma)+1} \frac{(1-\gamma)^2}{\gamma} u^n.$$
 (14b)

For $\alpha = 1$, the first step is the trapezoidal rule and the second step is the BDF2 formula, thus reconstructing the original TR-BDF2 (12). For $\alpha = 0$, each of the two steps above is equivalent to an implicit Euler step, thus transforming the hybrid TR-BDF2 in succession of two substeps of the implicit Euler method (IE-IE) of length γh and $(1 - \gamma)h$, respectively, and making the method unconditionally monotone.

The hybrid TR-BDF2 method can be rewritten as a DIRK scheme as done for (12). By injecting the first step in the second one, the Butcher tableau of the hybrid TR-BDF2 method is found

$$\begin{array}{c|c} 0 & 0 \\ \gamma & \gamma \frac{\alpha}{2} & \gamma \left(1 - \frac{\alpha}{2}\right) \\ 1 & \frac{\alpha}{2} \frac{\alpha(1-\gamma)+\gamma}{\alpha(1-\gamma)+1} & \left(1 - \frac{\alpha}{2}\right) \frac{\alpha(1-\gamma)+\gamma}{\alpha(1-\gamma)+1} & \frac{1-\gamma}{\alpha(1-\gamma)+1} \\ & \frac{\alpha}{2} \frac{\alpha(1-\gamma)+\gamma}{\alpha(1-\gamma)+1} & \left(1 - \frac{\alpha}{2}\right) \frac{\alpha(1-\gamma)+\gamma}{\alpha(1-\gamma)+1} & \frac{1-\gamma}{\alpha(1-\gamma)+1} \end{array}$$

For the unconditionally monotone method $\alpha = 0$, equivalent to a double step of implicit Euler, the above reduces to

$$\begin{array}{cccc} 0 & 0 \\ \gamma & 0 & \gamma \\ 1 & 0 & \gamma & 1 - \gamma \end{array}$$

while for $\alpha = 1$ we recover the TR-BDF2 tableau. The stability function of the DIRK family associated to the hybrid TR-BDF2 method is thus

$$\varphi(z) = \frac{1 + \left[\frac{\alpha(1-\gamma)+\gamma}{\alpha(1-\gamma)+1} - \gamma\left(1-\frac{\alpha}{2}\right)\right]z}{1 - \left[\frac{1-\gamma}{\alpha(1-\gamma)+1} + \gamma\left(1-\frac{\alpha}{2}\right)\right]z + \gamma\left(1-\frac{\alpha}{2}\right)\frac{1-\gamma}{\alpha(1-\gamma)+1}z^2}$$
(15)

From this expression, for the case $\alpha = 1$ we recover the stability function (13) and for the unconditionally monotone case $\alpha = 0$ the stability function

$$\varphi(z) = 1/[1 - z + \gamma(1 - \gamma)z^2].$$
(16)

The DIRK family corresponding to the hybrid TR-BDF2 method (i.e., $\gamma = 2 - \sqrt{2}$) is thus entirely L-stable for every value of the parameter α , as visibile also in Figure 1. Additionally, for $\alpha = 1$ the value $\gamma = 2 - \sqrt{2}$ also maximizes the radius of absolute monotonicity, while



Fig. 1 Hybrid TR-BDF2 methods compared with implicit Euler: **a** stability functions along the negative real axis, **b** modulus of the stability functions along the imaginary axis, **c** absolute error functions along the negative real axis showing the asymptotic and non-asymptotic ranges, **d** absolute error functions along the positive imaginary axis

for $\gamma = 2 - \sqrt{2}$ starting from $\alpha = 1$ (TR-BDF2) the radius of absolute monotonicity is progressively increased by decreasing α , while the order is reduced to p = 1 for $\alpha \neq 1$, as evident from the absolute error $|\varphi(z) - e^z|$ in the asymptotic range ($Re \ z \to 0^-$, $Im \ z \to 0$) in Fig. 1. Formally, we have $\lim_{\alpha \to 0} R(A, b) = \infty$, thus recovering the unconditional monotonicity of the implicit Euler method.

The hybrid TR-BDF2 method can be exploited through different strategies, since by choosing values of the parameter $0 \le \alpha \le 1$ it is possible to produce a continuous blend of the two main schemes varying the radius of absolute monotonicity accordingly. In our work we adopt a simpler approach and we investigate two alternative modes for enforcing monotonicity under the selected step size.

4.1 TR-BDF2 Blended

Based on the analysis in the previous section, a first SSP extension of the TR-BDF2 method can be obtained by switching from the default $\alpha = 1$ (TR-BDF2) to the unconditionally monotone mode $\alpha = 0$ (IE-IE) in all time steps in which a suitable sensor detects a violation of a global functional bound on a provisional solution from TR-BDF2. After each detected violation by the TR-BDF2 solution, the time step integration is repeated in IE-IE mode. Clearly, in the case in which a violation is detected, a single IE step with the assigned time step value could also be carried out. We have chosen instead to maintain the two stage structure of the scheme. Numerical experiments not reported in this work have shown that the proposed procedure is slightly less dissipative when the time step size is kept fixed. We call this simple method TR-BDF2 blended, since it provides automatic adaption of the α value by enforcing unconditional monotonicity only during critical transients.

To this purpose we introduce the global sensor function

$$\sigma = s_g(u^{n+1}) = \begin{cases} 1 & \text{if } \|u^{n+1}\| > M \\ 0 & \text{otherwise.} \end{cases}$$
(17)

that is able to determine if the generic functional bound M on $\|\cdot\|$ is violated by the numerical solution u^{n+1} . For each time step, the TR-BDF2 blended algorithm can be described as follows:

- 1. Set $\alpha = 1$ and perform the current integration by (14) to find the tentative solution u^* .
- 2. Apply the sensor $\sigma = s_g(u^*)$.
 - If $\sigma = 0$, set $u^{n+1} = u^*$ and go to the next time step.
 - If $\sigma = 1$, set $\alpha = 0$, repeat the current integration by (14) to find the solution u^{n+1} and go to the next time step.

This basic method inherits all the relevant stability properties of its component methods.

Proposition 2 The TR-BDF2 blended strategy is L-stable.

Proof Since both component methods TR-BDF2 ($\alpha = 1$) and IE-IE ($\alpha = 0$) are L-stable, the resulting method using in each time step $\alpha = 0$ or $\alpha = 1$ will be L-stable too.

Proposition 3 The TR-BDF2 blended strategy is unconditionally monotone.

Proof Since the IE-IE method ($\alpha = 0$) is unconditionally monotone and it is activated whenever the TR-BDF2 ($\alpha = 1$) solution violates monotonicity, as expressed from the functional bound on *M*, the resulting method using in each time step $\alpha = 0$ or $\alpha = 1$ will be unconditionally monotone too. \Box

Proposition 4 The TR-BDF2 blended strategy preserves linear invariants of the solution.

Proof Both the IE-IE and TR-BDF2 methods are implicit Runge-Kutta methods and as such they conserve linear invariants (see, e.g. [20], Theorem 1.5). Since the TR-BDF2 blended method consists of a succession of IE-IE or TR-BDF2 steps, it preserves linear solution invariants as well. Consequently, it allows preservation of atomic mass of chemical species in chemical kinetics problems.

Clearly, the order of accuracy will be limited to p = 1 whenever the sensor function is activated in the current time step and similarly will occur for the stage order \tilde{p} .

4.2 TR-BDF2 Partitioned

Another extension of the TR-BDF2 method can be obtained by applying the switch from the default TR-BDF2 to the unconditionally monotone IE-IE mode only to those solution components which are likely to produce violations of monotonicity under the assigned time step size. In order to detect this possibility, we rely on the SSP theory results by applying Definition 1, together with the computed value of $R(A, b) \approx 2.414$ for TR-BDF2. We introduce the local sensor function

$$\sigma_i = s_l(u_i^{n+1}) = \begin{cases} 1 & \text{if } \|u_i^{n+1}\| > M_i \\ 0 & \text{otherwise.} \end{cases}$$
(18)

that detects any violation of the functional bound M_i for the solution component u_i^{n+1} . In particular, for each time step, the TR-BDF2 partitioned algorithm can be described as follows:

1. Perform a tentative step of the forward Euler method

$$u^* = u^n + h_{EE} f(t^n, u^n)$$

using a monotonicity-scaled time step $h_{EE} = h/R(A, b)$.

- 2. Apply the sensor $\sigma_i = s_l(u_i^*)$ (i = 1, ..., m) on the tentative solution, to construct the partitioning matrix $S = diag\{\sigma_i\}$.
- 3. Identifying with a_{ij} and b_i the coefficients corresponding to the tableau for $\alpha = 1$ and with \tilde{a}_{ij} and \tilde{b}_i the coefficients for $\alpha = 0$ in (14), construct the automatically partitioned RK method

$$g^{i} = u^{n} + h \sum_{j=1}^{s} [a_{ij}(I - S) + \tilde{a}_{ij}S] f(t^{n} + c_{j}h, g^{j})$$
(19a)

$$u^{n+1} = u^n + h \sum_{i=1}^{s} [b_i(I-S) + \tilde{b}_i S] f(t^n + c_i h, g^i)$$
(19b)

to find the solution u^{n+1} from stage values g^i (i = 1, ..., s).

Thus, we have effectively transformed the hybrid TR-BDF2 (14) into a partitioned Runge-Kutta method. Method (19) can also be interpreted as an additive Runge-Kutta method with Butcher tableau



Fig. 2 Stability regions (infinite portions of the plane): **a** stability boundaries for the hybrid TR-BDF2 and the implicit Euler methods, **b** stability region of the additive Runge–Kutta corresponding to the TR-BDF2 partitioned method

following [27]. Several approaches to the stability analysis of ARK methods have been proposed in the literature, see e.g. [7,40] and [26]. In [7], for instance, an ARK stability function was derived based on the scalar test problem $u'(t) = \lambda u(t) + i \mu u(t)$ with $u(0) = u^0$ and $\lambda, \mu \in \mathbb{R}$. Taking $\alpha = h\lambda$, $\beta = h\mu$ and setting $z = \alpha + i\beta$, the authors define

$$\varphi(z) = 1 + (i\beta b^{\mathsf{T}} + \alpha \tilde{b}^{\mathsf{T}})(I - i\beta A - \alpha \tilde{A})^{-1}e = \frac{P(z)}{Q(z)}.$$
(20)

Here, the polynomials at the numerator and denominator are

$$P(z) = \alpha i\beta \frac{\gamma (1-\gamma)^2}{2(2-\gamma)} + i\beta \frac{1+(1-\gamma)^2}{2(2-\gamma)}$$
$$Q(z) = -\beta^2 \frac{\gamma (1-\gamma)}{2(2-\gamma)} + \alpha^2 \gamma (1-\gamma) + \alpha i\beta \frac{\gamma (1-\gamma)(4-\gamma)}{2(2-\gamma)} - i\beta \frac{2-\gamma^2}{2(2-\gamma)} - \alpha + 1,$$

respectively. The ARK stability function (20) is represented in Fig. 2. In this case, the unconditionally monotone component α is computed by the IE-IE method, while the conditionally monotone component β is integrated using the original TR-BDF2. We call this second strategy TR-BDF2 partitioned, since it performs an equation-based partitioning, according to the definition in [36]. It introduces a small overhead in computational time, since it always performs an explicit tentative step, even in case of a successful integration from TR-BDF2.

When integrating conservation laws, a more suitable partitioning strategy is the flux-based partitioning proposed in [36]. In this approach, partitioning is performed on the numerical flux function defined on cell edges, rather than directly on the right hand side. In contrast to equation-based partitioning (19), it is conservative and as such it is able to capture the correct shock propagation speed, as shown for the spatially partitioned embedded Runge–Kutta (SPERK) methods introduced in [36]. Following this rationale, we introduce a variant of the above strategy which we call TR-BDF2 flux-partitioned. It is to be remarked that this approach is specific to ODE systems resulting from the space discretization of a conservation law, while the previous variants are applicable to completely general ODE problems. We start with the conservation law

$$u_t = -f(u)_x \tag{21}$$

for which a conservative spatial semi-discretization can be written as

$$u_i'(t) = -\frac{1}{\Delta x} \left(f_{i+\frac{1}{2}} - f_{i-\frac{1}{2}} \right)$$
(22)

where $f_{i\pm\frac{1}{2}}$ is the numerical flux at $x_{i\pm\frac{1}{2}}$ $(1 \le i \le m)$. By defining the vector of fluxes $\Phi(u)$ with components $\phi_i(u) = -\frac{1}{\Delta x} f_{i\pm\frac{1}{2}}(u)$ $(0 \le i \le m)$, the conservative method (22) can be recast as:

$$u_i'(t) = D\Phi,\tag{23}$$

in which the $m \times (m + 1)$ differencing matrix D is introduced, with -1 on the superdiagonal and 1 on the main diagonal.

The conservative semi-discrete form (23) can be partitioned now directly at the level of the flux vector $\Phi(u)$. For this purpose, we use a local sensor σ_i defined on each cell as in (18). The partitioning matrix *S* is here substituted by the $(m + 1) \times (m + 1)$ flux partitioning matrix $X = diag\{\zeta_{i+\frac{1}{2}}\}$ ($0 \le i \le m$) defined at cell edges as

$$\varsigma_{i+\frac{1}{2}} = \begin{cases} 1 & \text{if } \sigma_i = 1 \text{ or } \sigma_{i+1} = 1 \quad (0 \le i \le m) \\ 0 & \text{otherwise.} \end{cases}$$
(24)

Additionally, following similar steps taken in [36], in order to ensure that any local extremum requiring a monotone treatment does not fall outside the region where the IE-IE mode is applied, we widen it in downwind direction by applying the weights $\varsigma_{i+\frac{3}{2}}^* = 0.66$ and $\varsigma_{i+\frac{5}{2}}^* = 0.33$ whenever $\varsigma_{i+\frac{1}{2}} = 1$ and we modify the matrix *X* accordingly such that $\varsigma_{i+\frac{1}{2}} = \max\{\varsigma_{i+\frac{1}{2}}, \varsigma_{i+\frac{1}{2}}^*\}$.

The overall algorithm is thus similar to the TR-BDF2 partitioned, where the automatically partitioned RK method (19) is now replaced by

$$g^{i} = u^{n} + h \sum_{j=1}^{s} [a_{ij}D(I - X) + \tilde{a}_{ij}DX] \Phi(g^{j})$$
(25a)

$$u^{n+1} = u^n + h \sum_{i=1}^{s} [b_i D(I - X) + \tilde{b}_i DX] \Phi(g^i)$$
(25b)

where g^i (i = 1, ..., s) are the stage values of the underlying ARK scheme applied to the functions $F_{I-X}(u) = D(I - X)\Phi(u)$ and $F_X(u) = DX\Phi(u)$. Differently from the SPERK schemes in [36], which are constructed under the constraint that $a_{ij} = \tilde{a}_{ij}$, both the partitioned and flux-partitioned TR-BDF2 methods are particular instances of a class of spatially partitioned RK methods generally allowing for $a_{ij} \neq \tilde{a}_{ij}$. Moreover, the partitioned methods analyzed here are implicit, while the SPERK schemes discussed in [36] are explicit methods. Again, the overall order of accuracy for both strategies will be limited to p = 1 whenever the sensor functions are activated in the current time step and similarly for the stage order \tilde{p} .

A more general stability result for the TR-BDF2 partitioned method can be proven considering the same model problem as in [40] and their definition of stability function (different from the one in [7] considered above!). Denote by $\varphi(w, z)$, $w, z \in \mathbb{C}$, the general ARK stability function defined [40]. The result can be summarized in the following proposition Proposition 5 The TR-BDF2 partitioned and flux-partitioned strategies satisfy

$$\lim_{z \to \infty} |\varphi(w, z)| = 0 , \quad \forall \ w \in \mathbb{C} \quad \lim_{w \to \infty} |\varphi(w, z)| = 0 , \quad \forall \ z \in \mathbb{C}$$
(26)

for $\gamma \neq 0$ *and* $\gamma \neq 1$ *.*

Proof We start by considering the general ARK stability function $\varphi(w, z)$ for $w, z \in \mathbb{C}$, see, e.g., [40], which can be computed as

$$\varphi(w,z) = \frac{\det(I - wA - z\tilde{A} + we \otimes b^{\mathsf{T}} + ze\tilde{\otimes}b^{\mathsf{T}})}{\det(I - wA - z\tilde{A})}.$$
(27)

By applying definition (27) we find

$$\begin{split} I - wA - z\tilde{A} + we \otimes b^{\mathsf{T}} + ze \otimes \tilde{b}^{\mathsf{T}} \\ &= \begin{bmatrix} 1 + w\frac{1}{2(2-\gamma)} & w\frac{1}{2(2-\gamma)} + z\gamma & w\frac{1-\gamma}{2-\gamma} + z(1-\gamma) \\ w\left[\frac{1}{2(2-\gamma)} - \frac{\gamma}{2}\right] \left\{ 1 + w\left[\frac{1}{2(2-\gamma)} - \frac{\gamma}{2}\right] \right\} w\frac{1-\gamma}{2-\gamma} + z(1-\gamma) \\ 0 & 0 & 1 \end{bmatrix} \\ I - wA - z\tilde{A} = \begin{bmatrix} 1 & 0 & 0 \\ -w\frac{\gamma}{2} & 1 - w\frac{\gamma}{2} - z\gamma & 0 \\ -\frac{w}{2(2-\gamma)} - \left\{\frac{w}{2(2-\gamma)} + z\gamma\right\} \left\{ 1 - w\frac{1-\gamma}{2-\gamma} - z(1-\gamma) \right\} \end{bmatrix}$$

from which the stability function of TR-BDF2 partitioned method follows

$$\varphi(w,z) = \frac{P(w,z)}{Q(w,z)}$$
(28)

where the numerator and denominator polynomials are respectively

$$P(w, z) = 1 + w \frac{2 - 2\gamma + \gamma^2}{2(2 - \gamma)} + wz \frac{2\gamma^2 - \gamma^3 - \gamma}{2(2 - \gamma)}$$
$$Q(w, z) = w^2 \frac{\gamma(1 - \gamma)}{2(2 - \gamma)} + z^2 \gamma(1 - \gamma) + wz \frac{\gamma(1 - \gamma)(4 - \gamma)}{2(2 - \gamma)} - w \frac{2 - \gamma^2}{2(2 - \gamma)} - z + 1.$$

As a consequence for $\gamma \neq 0$ and $\gamma \neq 1$ the stability function is always such that

$$\lim_{z \to \infty} |\varphi(w, z)| = 0, \quad \text{for any finite } w \in \mathbb{C}.$$
 (29)

Similarly we have $\lim_{w\to\infty} |\varphi(w, z)| = 0$ for any finite $z \in \mathbb{C}$.

Notice that this property can be interpreted as an extension of the classical L-stability concept to ARK methods.

Proposition 6 *The TR-BDF2 partitioned and flux-partitioned strategies are unconditionally monotone.*

Proof For ARK schemes the concept of radius of absolute monotonicity is substituted by the *region of absolute monotonicity* [26] defined as

$$R(A, \tilde{A}, b, \tilde{b}) = \sup\{(r_1, r_2) \mid r_1 \ge 0, r_2 \ge 0, \mathcal{A}(\xi_1, \xi_2) \ge 0, \ \tilde{\mathcal{A}}(\xi_1, \xi_2) \ge 0, \ \mathcal{E}(\xi_1, \xi_2) \ge 0, \text{for } (\xi_1, \xi_2) \in [-r_1, 0] \times [-r_2, 0]\}$$
(30)

where the quantities $\mathcal{A}(\xi_1, \xi_2) = (I + \xi_1 \mathbb{A} + \xi_2 \tilde{\mathbb{A}})^{-1}(A)$, $\tilde{\mathcal{A}}(\xi_1, \xi_2) = (I + \xi_1 \mathbb{A} + \xi_2 \tilde{\mathbb{A}})^{-1}(\tilde{A})$ and $\mathcal{E}(\xi_1, \xi_2) = (I + \xi_1 \mathbb{A} + \xi_2 \tilde{\mathbb{A}})^{-1}e$ have been introduced in the elementwise inequalities and the notation was simplified by

$$\mathbb{A} = \begin{pmatrix} A & 0 \\ b^{\mathsf{T}} & 0 \end{pmatrix}, \quad \tilde{\mathbb{A}} = \begin{pmatrix} \tilde{A} & 0 \\ \tilde{b}^{\mathsf{T}} & 0 \end{pmatrix}$$

Numerical monotonicity of an ARK method can be ensured under the step size restriction $0 < h \le \min\{r_1\tau_0, r_2\tilde{\tau}_0\}$, where r_1, r_2 are such that $(r_1, r_2) \in R(A, \tilde{A}, b, \tilde{b})$ and $\tau_0, \tilde{\tau}_0$ are the maximum time step sizes under which the explicit Euler method is monotone for the two right hand sides of the ARK, respectively. Since the TR-BDF2 partitioned is an ARK method composed by TR-BDF2 for \mathbb{A} and IE-IE for $\tilde{\mathbb{A}}$, we have that $r_1 \approx 2.414$ and $r_2 = \infty$.

Since this method automatically shifts components violating monotonicity of TR-BDF2 with the current step size to the right hand side Sf(t, u) (or $DX\Phi(u)$) integrated by IE-IE, the method is effectively unconditionally monotone, as long as the sensor function is correctly capturing monotonicity violations.

Proposition 7 The TR-BDF2 flux-partitioned strategy is conservative.

Proof Since the TR-BDF2 flux-partitioned is constructed by assigning weight values $\varsigma_{i+1/2}$ $(1 \le i \le m)$ on each cell edge, flux inconsistencies in the form of $\varsigma_{i+\frac{1}{2}} \ne \varsigma_{j-\frac{1}{2}}$ for j = i + 1 are a priori excluded from the construction of the right hand side. This does not hold for the TR-BDF2 partitioned strategy, see [36].

5 Potential Competitors of TR-BDF2

In order to compare the properties of TR-BDF2 to those of other similar methods, we introduce here some second order methods with analogous characteristics. We will make our assessment by comparing the performance of TR-BDF2 and its two variants from Sect. 4 against the following methods, as well as against other classic methods, such as implicit Euler ($R(A, b) = \infty$), the Crank-Nicolson (R(A, b!) = 2) and the implicit midpoint rule (R(A, b) = 2). Notice that the latter method has also been reinterpreted as a two stage second order SDIRK with R(A, b) = 4 in [35], so that its results are denoted by the label SDIRK 2(2) in the following.

Among the potential competitors of TR-BDF2 we consider the Rosenbrock method (ROS2) proposed in [9] and later applied in atmospheric chemistry problems, see, e.g. [42], [51]. The Butcher tableau of ROS2 is

with $\gamma = 1 + \frac{1}{\sqrt{2}}$. In addition to L-stability, from numerical experiments in [51] it was found that it also has interesting positivity properties. This was empirically justified by observing that the stability function is positive along the entire negative real axis. In [42] this observation was extended to the first two derivatives of the stability function. However, from the framework in Sect. 2 it turns out that R(A, b) = 0, since the third derivative of the stability function is negative along the negative portion of the real axis. In spite of this, the fact that up to the second derivative we have positivity for any negative real value can be interpreted as

a sort of *weak* absolute monotonicity. Even though small violations of arbitrary functionals cannot be a priori excluded at any step size, from numerical experiments in Sect. 6 we found an intrinsic resiliency against violations of the TVD property, even from initial conditions of limited regularity.

Another second order method for stiff chemical problems is the Modified Patankar Runge– Kutta (MPRK) introduced in [5] and later extended to third order accuracy in [15]. The MPRK method achieves unconditional positivity through proper weighting of production and destruction terms and it conserves the quantity $\sum_{j=1}^{m} u_j$, which represents the total mass of the system if species are expressed as mass concentrations. However, the mass of the atomic species is not conserved. Additionally, due to the specific form required to the right hand side, it can be used in a PDE framework only by introducing a source term splitting.

To complete the range of potential competing methods, we also consider the class of explicit SSP Runge–Kutta schemes. Since all the previously introduced methods are of second order of accuracy, we include the two-stages second order Shu–Osher scheme [44], whose SSP optimality was later demonstrated in [18], and we will refer to it as ERK 2(2) scheme. Starting from the original Shu–Osher form the standard Butcher tableau of ERK 2(2) can be derived

from which it is apparent that the method corresponds to the improved Euler (or second order Heun) method. Another relevant method for SSP time integrations is the three-stages third order Shu–Osher scheme [44], which it is also SSP-optimal [18]. The Butcher tableau of ERK 3(3) can be derived from the original Shu–Osher form and it is

Both these explicit methods cannot be A-stable and are characterized by R(A, b) = 1, similarly to the explicit Euler method.

6 Numerical Experiments

A number of numerical experiments have been carried out, in order to assess the performance of the TR-BDF2 method and its hybrid variants introduced in Sect. 4 against the other methods described in Sect. 5. The test cases include a reactive zero dimensional test problem, here adapted to the MPRK method, a one dimensional advection problem, an advection diffusion reaction problem for a mixture of chemical species, as well as two typical nonlinear conservation laws. For PDE tests, we have considered discontinous initial conditions, in order to show the emergence of critical issues for monotonicity. This provides the most stringent test, as we experienced from other computations, not reported here, using more regular initial conditions. In all test cases the nonlinear system associated to the implicit RK methods was solved using MATLAB fsolve function with a tolerance of tol = 10^{-10} , except when otherwise stated. All computations were performed on a single Intel[®] CoreTMi5-2540M (2.60 GHz) on a laptop with 4 GB RAM running Linux kernel 3.13.0-24-generic. In "Appendix 2" typical error magnitudes and workloads are reported from measurements during the numerical experiments.

6.1 A Chemical Model Problem: The Brusselator

As a first test we consider a typical nonlinear chemical kinetics problem. The same issues are shared by all chemistry modelling problems, which require positivity for each solution component. Consequently, we adopt both hybridization strategies from Sect. 4 with either global or slocal positivity sensors built from the floor norm (8) with $\chi = 0$. In the zero dimensional chemical model problem we measure the maximum error during time integrations from the l^{∞} -time absolute error norm for the *i*th species

$$\|e_i\|_{\infty} = \max_{n=1,\dots,N_t} |u_i^n - \tilde{u}_i^n|$$
(32)

where N_t is the number of time steps, u_i^n is the solution in the *n*th time step and \tilde{u}_i^n is the reference solution at the same time step, obtained with MATLAB ode15s using absolute and relative error tolerance levels of AbsTol=10⁻¹⁴ and RelTol=10⁻¹³, respectively. In our numerical results we will refer to the error on species i = 1, which is assumed to be representative of the problem.

We consider thus the original Brusselator system from [38]:

$$u_1' = -k_1 u_1$$
 (33a)

$$u_2' = -k_2 u_2 u_5 \tag{33b}$$

$$u'_3 = k_2 u_2 u_5$$
 (33c)

$$u_4' = k_4 u_5 \tag{33d}$$

$$u'_{5} = k_{1}u_{1} - k_{2}u_{2}u_{5} + k_{3}u_{5}^{2}u_{6} - k_{4}u_{5}$$
(33e)

$$u_6' = k_2 u_2 u_5 - k_3 u_5^2 u_6 \tag{33f}$$

since this form allows to write the right hand side in a form suitable to MPRK. In particular we follow the procedure in [15], which in this case can be applied, since the stoichiometric matrix has proper rank ([15], Assumption 2.1). We solve this problem for $0 \le t \le 10$ assuming $k_1 = k_2 = k_3 = k_4 = k_5 = k_6 = 1$ as in the reduced model and starting from the initial condition $u_1 = u_2 = 10, u_3 = u_4 = 0$ and $u_5 = u_6 = 0.1$.

The results in Fig. 3 show that Crank-Nicolson, SDIRK 2(2) and TR-BDF2 (clipped to avoid negative values) are almost equivalent in performance, with a slight advantage for the SSP-optimal SDIRK 2(2). MPRK shows similar accuracy at same step sizes, while it outperforms all the other methods in terms of workload, being the only explicit method, see Table 8 in "Appendix 2". ROS2 offers intermediate performance. Blended and partitioned TR-BDF2 are here equivalent to the clipped version, due to limited size of the integration interval T. Similar results, not shown here, were obtained for the simple geobiochemical problem from [5]. Even though the results from MPRK are promising, in the next tests we are forced to abandon it, since it would require a source splitting to the advection diffusion reaction problem that is out of our scope. We also consider the ERK 2(2) and ERK 3(3) schemes as alternative explicit methods.



Fig. 3 Error-stepsize curves for the original Brusselator problem

6.2 A Linear Advection Problem

As a first PDE case we consider the advection equation

$$u_t + vu_x = 0, \quad 0 \le x \le 1$$
 (34)

which we solve for $0 \le t \le 1$ and v = 1 using periodic boundary conditions. We discretize the interval [0, 1] by introducing N_x points $x_i = i \Delta x$, $i = 1, ..., N_x$ with $\Delta x = 0.01$ and we discretize the advective term using a first order upwind scheme to yield a contractive right hand side. We consider the non smooth initial condition

$$u(0, x) = \begin{cases} 1 & \text{if } |x - 0.5| < 0.25\\ 0 & \text{otherwise.} \end{cases}$$
(35)

For this problem ,the explicit Euler method is stable under the well known condition $C = \frac{|v| \Delta t}{\Delta x} \leq 1$. In our assessment, we disregard spatial discretization errors, since we compare the numerical solution obtained with any method to the exact solution of the ODE system (1), rather than to the exact solution of the original PDE. The exact solution is approximated by an accurate numerical solution obtained with the MATLAB ode45 solver, with absolute and relative error tolerances of AbsTol=10⁻¹⁴ and RelTol=10⁻¹³, respectively. During numerical tests we measured the l^{∞} -space absolute error norm at final time t = T

$$\|e_i\|_{t=T}^{\infty} = \max_{j=1,\dots,N_x} \left| u_{i,j}^T - \tilde{u}_{i,j}^T \right|$$
(36)

where $u_{i,j}^T$ represents the solution at point *j* and time t = T for the single species i = 1 in the one-dimensional advection problem (34) and $\tilde{u}_{i,j}^T$ is the reference solution at the same point in space and time. Furthermore, in order to assess any violation of the TVD property, we monitored also the *TV*-space l^{∞} -time seminorm for species *i*

$$\|TV_i\|_{\infty} = \max_{n=1,\dots,N_t} \sum_{j=1}^{N_x} \left| u_{i,j+1}^n - \tilde{u}_{i,j}^n \right|.$$
(37)



Fig. 4 Advection problem with non smooth initial condition: solution for $u_1(t=T)$ when using h = 0.100

Here, the TR-BDF2 blended from Sect. 4 relies on a global sensor with the floor functional with $\chi = 0$, while TR-BDF2 partitioned uses a local sensor using both floor and ceil functionals with $\chi = 0$ and $\psi = 1$. In Fig. 4 we report the solution at the final time step for the coarsest step size h = 0.1, corresponding to CFL = 10.00. At such high Courant number, the explicit methods ERK 2(2) and ERK 3(3) are unstable, as evident from Table 9 in "Appendix 2", so that we avoid to plot their solutions. On the other hand, when explcit methods are stable, the deviations in the solutions from the different methods are undistinguishable from the solution plots. As a consequence, we hardly find it relevant to plot the different solutions at small time step sizes. While implicit Euler and ROS2 are subject to strong smearing of the solution profile, SDIRK 2(2), TR-BDF2 and its variants show satisfactory behaviour, even though they are generating TVD violations, as from Table 1.

The discontinuous initial condition (35) generates TVD violations for all the conditionally monotone methods (i.e., not for implicit Euler and TR-BDF2 blended). The critical steps sizes at which violations occur closely follow the ratio between the radii of absolute monotonicity for the different methods, as from Sects. 3 and 5, see Table 1. Additionally, although ROS2 is never TVD, the violations of this property are always limited for any step size. Globally, however, the method does not show the typical behaviour of conditionally monotone methods, since it does not generate larger errors with increasing step sizes. This can be interpreted in light of the results discussed in Sect. 5.

TR-BDF2 blended from Sect. 4 never violates TVD property, while its clipped version shows the usual breakdown at large time steps. The possible explanation from Fig. 5 is that the critical step size for positivity is smaller than that for TVD property. As such TR-BDF2 blended starts reverting to the unconditionally monotone mode at relatively small step sizes, implicitly guaranteeing also the TVD property. On the other hand, TR-BDF2 clipped activates the clipping procedure for violations of positivity, but it is not able to manage TVD violations. This example illustrates well the advantage of using SSP theory to guarantee nonlinear solution properties. TR-BDF2 partitioned is always TVD, except for a small violation at the step size h = 0.06. We point out that the advection problem is solved here in non-conservative form using standard finite difference approximations. If the same problem is recast in conservative form, then the TR-BDF2 flux-partitioned strategy would guarantee a conservative integration. We postpone the assessment of the flux-partitioned strategy against the equation based partitioning until Sect. 6.4.

The accuracy results in Fig. 5 maintain the consistent ranking from the zero dimensional problem. ROS2 now features a non uniform convergence and it achieves higher accuracy for 0.01 < h < 0.04 by sacrificing monotonicity, as evident when comparing Fig. 5 and Table

h	CFL	ref	IE	CN	SDIRK 2(2)	ROS2
0.0025	0.25	2.000	2.000	2.000	2.000	2.008
0.0050	0.50	2.000	2.000	2.000	2.000	2.029
0.0100	1.00	2.000	2.000	2.000	2.000	2.076
0.0200	2.00	2.000	2.000	2.000	2.000	2.142
0.0241	2.41	2.000	2.000	2.375	2.000	2.147
0.0400	4.00	2.000	2.000	3.333	2.000	2.123
0.0600	6.00	2.000	2.000	4.062	2.768	2.073
0.1000	10.00	2.000	2.000	5.218	3.732	2.019
		TR-BDF2 (clipped)	TR-BDF2 blended	TR-BDF2 partitioned	ERK 2(2)	ERK 3(3)
0.0025		2.000	2.000	2.000	2.000	2.000
0.0050		2.000	2.000	2.000	2.000	2.000
0.0100		2.000	2.000	2.000	2.000	2.000
0.0200		2.000	2.000	2.000	∞	∞
0.0241		2.000	2.000	2.000	∞	∞
0.0400		2.278	2.000	2.000	∞	∞
0.0600		2.390	2.000	2.001	∞	∞
0.1000		2.477	2.000	2.000	∞	∞

Table 1 $||TV||_{\infty}$ for the advection problem with non smooth initial condition



Fig. 5 Error-stepsize curves for the advection problem with non smooth initial condition

1. ERK 2(2) reaches the formal accuracy only at very small stepsizes, resulting as the less accurate second order method with the exception of ROS2. ERK 3(3) is obviously more accurate, but its error is of similar magnitude to SDIRK 2(2) an TR-BDF2 at the coarsest time steps which still guarantees stability. As expected, both explicit methods generate less workload for sufficiently small step sizes with respect to their implicit competitors, see Table 9 in "Appendix 2", but they are unstable at the coarsest step sizes and thus require time

Table 2 Fraction of time steps of TR-BDF2 blended with activated sensor ($\sigma = 1$) for the advection problem with non smooth initial condition

h	0.0025	0.0050	0.0100	0.0200	0.0241	0.0400	0.0600	0.1000
$\sigma = 1$	0/400	0/200	0/100	0/50	0/42	5/25	4/17	2/10

step adaption strategies. TR-BDF2 blended shows a degradation in accuracy at larger step sizes, due to the intervention of the IE-IE mode triggered by the positivity monitor, activated especially at the three larger step sizes as reported in Table 2, also evidenced from the order reduction in Fig. 5. The TR-BDF2 partitioned is the only method able to obtain tighter accuracy levels similar to SDIRK 2(2), while additionally mantaining the TVD property with the exception of one step size. When repeating the same advection test with a smooth initial condition, the results obtained, not shown here, are similar to those in the chemical model problem and they do not exhibit the order reduction and TVD violations reported above.

6.3 An Advection Diffusion Reaction Problem

As a representative case for chemical transport of reacting species we consider the coupled advection diffusion reaction of three species given by

$$u_t + vu_x = Du_{xx} + f(u), \quad 0 \le x \le 1.$$
 (38)

with periodic boundary conditions. Here $u = [u_1, u_2, u_3]^T$ and $D = diag\{d_{ii}\}$ is the diffusivity matrix. The nonlinear source term f(u) is taken from the simple geobiochemical model in [5]

$$f(u)_1 = -\frac{u_1 u_2}{u_1 + 1} \tag{39a}$$

$$f(u)_2 = \frac{u_1 u_2}{u_1 + 1} - k u_2 \tag{39b}$$

$$f(u)_3 = ku_2 \tag{39c}$$

representing the interaction among three species identified as nutrients u_1 , phytoplankton u_2 and detritus u_3 . Since the total mass of the system must be conserved, there is an implicit linear invariant for this problem which is $\sum_i (u_{1,i} + u_{2,i} + u_{3,i})$ (or $e^{\top}u$). We solve (38) for $0 \le t \le 1$ with k = 0.3, an advection velocity of v = 0.1 and constant diffusivities for each species given by $d_{11} = 10^{-3}$, $d_{22} = 2 \times 10^{-3}$ and $d_{33} = 10^{-4}$. The grid and the discrete advection operator are the same of the advection problem, namely the interval [0, 1] is discretized by introducing N_x points $x_i = i \Delta x$, $i = 1, ..., N_x$ with $\Delta x = 0.01$. The diffusion term is approximated by central finite differencing and it is naturally contractive. The initial conditions for the three species are

$$u_1(0, x) = \begin{cases} 9.98 & \text{if } |x - 0.5| < 0.25\\ 0 & \text{otherwise,} \end{cases}$$
(40a)

$$u_2(0, x) = \begin{cases} 2 & \text{if } |x - 0.4| < 0.2\\ 0 & \text{otherwise,} \end{cases}$$
(40b)

$$u_3(0, x) = \begin{cases} 1 & \text{if } |x - 0.7| < 0.25\\ 0 & \text{otherwise.} \end{cases}$$
(40c)



Fig. 6 Advection diffusion reaction problem with non smooth initial condition: solution for $u_1(t = T)$ when using h = 0.100 (*top*), close-up in the region of positity violation from ROS2 (*bottom*)

In our assessment, we use a tolerance level $tol=10^{-8}$ for the nonlinear solver in implicit methods and the reference solution is obtained from MATLAB's ode15s with the same absolute and relative error tolerances from the advection test. The sensors for the hybrid TR-BDF2 variants are simply built from the floor functional with $\chi = 0$.

A sample solution for this problem is shown in Fig. 6, where a close-up shows the typical positivity violation on u_1 from ROS2. Again, we avoid plotting solutions from the explicit methods ERK 2(2) and ERK 3(3), under the same rationale as in the discussion on the advection problem. Violations of the TVD property are reported in Table 3 where it is evident that all the methods are TVD, with the exceptions of Crank-Nicolson for h = 0.100 and ROS2 that is never TVD nor positivity preserving.

The accuracy curves in Fig. 7 do not show the critical features of the advection test, due to the presence of the additional diffusive term that rapidly smooths out initial discontinuities. Again, RK methods maintain their relative ranking in terms of accuracy and workload. The

h	CFL	ref	IE	CN	SDIRK 2(2)	ROS2
0.0025	0.025	19.960	19.960	19.960	19.960	19.961
0.0050	0.050	19.960	19.960	19.960	19.960	19.966
0.0100	0.100	19.960	19.960	19.960	19.960	19.978
0.0250	0.200	19.960	19.960	19.960	19.960	20.017
0.0500	0.241	19.960	19.960	19.960	19.960	20.072
0.1000	1.000	19.960	19.960	21.261	19.960	20.110
		TR-BDF2 (clipped)	TR-BDF2 blended	TR-BDF2 partitioned	ERK 2(2)	ERK 3(3)
0.0025		19.960	19.960	19.960	19.960	19.960
0.0050		19.960	19.960	19.960	19.960	19.960
0.0100		19.960	19.960	19.960	19.960	19.960
0.0250		19.960	19.960	19.960	∞	19.961
0.0500		19.960	19.960	19.960	∞	∞
0.1000		19.960	19.960	19.960	∞	∞

Table 3 $||TV||_{\infty}$ for the advection diffusion reaction problem with non smooth initial condition



Fig. 7 Error-stepsize curves for the advection diffusion reaction problem with non smooth initial condition

TR-BDF2 clipped and blended variants are almost indistinguishable, while the partitioned version features slightly larger computational times due to the partitioning step.

6.4 Two Classic Conservation Laws

We complete our assessment by considering two well known hyperbolic conservation laws, see, e.g. [39] for a more detailed discussion. The first is the inviscid Burgers equation

$$u_t = -f(u)_x = -\left(\frac{1}{2}u^2\right)_x, \quad 0 \le x \le 1.$$
 (41)

which we solve for $0 \le t \le 1$ with the smooth initial condition

$$u(0,x) = \frac{1}{2} + \frac{1}{4}\sin(2\pi x).$$
(42)

The second is the Buckley-Leverett equation

$$u_t = -f(u)_x = -\left(\frac{u^2}{u^2 + \frac{1}{3}(1-u)^2}\right)_x, \quad 0 \le x \le 1.$$
(43)

which we solve for $0 \le t \le \frac{1}{8}$ with the discontinuous initial condition

$$u(0, x) = \begin{cases} \frac{1}{2} & \text{if } x \le 0.5\\ 0 & \text{otherwise.} \end{cases}$$
(44)

The interval [0, 1] is discretized by introducing N_x points $x_i = i \Delta x$, $i = 1, ..., N_x$ with $\Delta x = 0.01$. Both equations are here solved with periodic boundary conditions and discretized by a high resolution finite volume method using flux limiters, see, e.g. [32, 39]. Specifically, for the Burgers equation we adopt the van Leer limiter

$$\Psi(\theta) = \frac{\theta + |\theta|}{1 + |\theta|} \tag{45}$$

while for the Buckley-Leverett equation we select the Koren limiter

$$\Psi(\theta) = \max\left\{0; \min\left\{2; \frac{2}{3} + \frac{1}{3}\theta; 2\theta\right\}\right\}.$$
(46)

In this case, we have excluded ROS2 from our assessment, due to its poor performance in the previous test. Additionally, we introduce here the TR-BDF2 flux-partitioned method, specifically designed for conservation laws.

In the numerical tests, we measured the l^{∞} norm in space (36) as well as the *TV* seminorm in space (37). The reference numerical solution is obtained here by the MATLAB solver ode45, using again AbsTol=10⁻¹⁴ and RelTol=10⁻¹³, while the implicit stages of the RK methods are solved with a tolerance level of 10⁻¹⁰. While the TR-BDF2 blended exploits a global sensor for the TV seminorm, the TR-BDF2 partitioned relies on a local sensor for the floor and ceil functionals (8) with $\chi = 0.25$ and $\psi = 0.75$ for the Burgers equation and $\chi = 0$ and $\psi = 0.5$ for the Buckley–Leverett equation, respectively. Analogous values are employed for the TR-BDF2 flux-partitioned method. These choices follow from the initial conditions (42) and (44). Even though this local sensor is not properly a detector of TVD violations, we use it as an approximate TV sensor due to the known solution dynamics. This is not entirely correct, as we will see from the tests, but it comes from the difficulty of using a local test for a global property such as TVD.

In Fig. 8 we report the solution at the final time step for the coarsest step size h = 0.1, corresponding to about $C_{max} = 7.5$. While implicit Euler is able to maintain the TVD property, see Table 4, with a strong smoothing of the developing leading shock, all conditionally monotone methods develop visible oscillations downstream of this region. In particular, TR-BDF2 (here positive clipping is never activated) shows minor amplitudes with respect to SDIRK 2(2) and Crank-Nicolson. ERK 2(2) and ERK 3(3) are unstable at such high Courant numbers and thus their results are not shown. Their typical behaviour is to mantain stability and TVD property at sufficiently small step sizes, while they become unstable at $C_{max} \ge 1$. TR-BDF2 blended obtains a smoothed solution after several integrations in IE-IE mode. TR-BDF2 partitioned is qualitatively very close to the reference solution with the best



Fig. 8 Burgers equation with van Leer limiter and smooth initial condition: solution at final time t = 1 when using h = 0.1

Table 4 $||TV||_{\infty}$ for the Burgers equation with van Leer limiter and smooth initial condition

h	CFL _{max}	ref	EE	IE	CN	SDIRK 2(2)
0.0025	0.188	1.000	1.000	1.000	1.000	1.000
0.0050	0.375	1.000	1.000	1.000	1.000	1.000
0.0100	0.750	1.000	3.479	1.000	1.000	1.000
0.0200	1.500	1.000	∞	1.000	1.067	1.000
0.0400	3.000	1.000	∞	1.000	1.367	1.022
0.0600	4.500	1.000	∞	1.000	1.334	1.396
0.1000	7.500	1.000	∞	1.000	1.204	1.852
	TR-BDF2 (clipped)	TR-BDF2 blended	TR-BDF2 partitioned	TR-BDF2 flux-partitioned	ERK 2(2)	ERK 3(3)
0.0025	1.000	1.000	1.000	1.000	1.000	1.000
0.0050	1.000	1.000	1.000	1.000	1.000	1.000
0.0100	1.000	1.000	1.000	1.000	1.000	1.000
0.0200	1.000	1.000	1.000	1.000	∞	∞
0.0400	1.101	1.000	1.000	1.000	∞	∞
0.0600	1.186	1.000	1.000	1.000	∞	∞
0.1000	1.176	1.000	1.000	1.000	∞	∞



Fig. 9 Error-maximum Courant number curves for the Burgers equation with van Leer limiter and smooth initial condition

Table 5 Percentage of time steps of TR-BDF2 blended with activated sensor ($\sigma = 1$) for the Burgers problem with smooth initial condition

h	0.0025	0.0050	0.0100	0.0200	0.0400	0.0600	0.1000
$\sigma = 1$	0/400	0/200	0/100	5/50	11/25	11/17	9/10

approximation for the shock amplitude, but it features also a reduction in the propagation speed, as expected from its non conservative partitioning, see, e.g., [36]. On the other hand, the flux-partitioned TR-BDF2 obtains the correct shock propagation speed, even at $C_{max} = 7.5$. Significantly, all the three variants of TR-BDF2 remain TVD at any step size. The other methods show TVD violations with the usual critical step size progression, as evidenced in Table 4.

The accuracy results for the Burgers equation are reported in Fig. 9. Interesting behaviour arises for large step sizes, where the majority of the methods collapse about at the same accuracy of implicit Euler. ERK 2(2) and ERK 3(3) are here penalized in accuracy with respect to the previous test cases, with ERK 3(3) obtaining about the same accuracy as the second order implicit RK methods. Additionally, at $C_{max} = 1.5$ (h = 0.02) they both generate overflow and thus their accuracy curves abruptly interrupt at $C_{max} = 0.75$. TR-BDF2 blended realizes a smooth adaption from the monotone implicit Euler accuracy to the TR-BDF2 asymptotic curve, with the sensor being increasingly activated at larger step sizes, see Table 5. The error norm is penalized from the behaviour at the leading shock. TR-BDF2 flux-partitioned, instead, is able to obtain the most accurate solutions at large step sizes without violating TVD condition.

The Buckley–Leverett test provides a more stringest test due to its non-convex flux function. The solution at the final time step for the step size h = 0.025 is shown in Fig. 10, where ERK 2(2) and ERK 3(3) are omitted for the same reasons as in the previous tests. Again all conditionally monotone methods develop oscillations on the trailing shock, while only SDIRK 2(2) develops a stable rarefaction wave, Crank-Nicolson and TR-BDF2 (again with-



Fig. 10 Buckley–Leverett equation with Koren limiter and non smooth initial condition: solution at final time t = 0.125 when using h = 0.025

Table 6	$ TV _{\infty}$	for the	Buckley-	Leverett	equation	with	Koren	limiter	and	non	smooth	initial	condition
---------	-------------------	---------	----------	----------	----------	------	-------	---------	-----	-----	--------	---------	-----------

h	CFL_{max}	ref	EE	IE	CN	SDIRK 2(2)
0.0010	0.221	1.000	1.000	1.000	1.000	1.000
0.0025	0.551	1.000	1.000	1.000	1.000	1.000
0.0050	1.101	1.000	1.877	1.000	1.000	1.000
0.0075	1.654	1.000	6.747	1.000	1.270	1.000
0.0100	2.201	1.000	21.352	1.000	1.403	1.000
0.0150	3.301	1.000	21.107	1.000	8.097	1.282
0.0250	5.504	1.000	19.652	1.000	16.399	1.573
	TR-BDF2 (clipped)	TR-BDF2 blended	TR-BDF2 partitioned	TR-BDF2 flux-partitioned	ERK 2(2)	ERK 3(3)
0.0010	1.000	1.000	1.000	1.000	1.000	1.000
0.0025	1.000	1.000	1.000	1.000	1.000	1.000
0.0050	1.000	1.000	1.000	1.000	1.059	1.025
0.0075	1.045	1.000	1.000	1.000	13.766	2.474
0.0100	1.186	1.000	1.005	1.000	16.029	11.909
0.0150	1.322	1.000	1.031	1.030	26.285	30.532
0.0250	12.483	1.000	1.006	1.000	23.183	45.123



Table 7 Percentage of time steps of TR-BDF2 blended with activated sensor ($\sigma_i = 1$) for the Buckley–Leverett problem with non smooth initial condition

Fig. 11 Buckley–Leverett equation with Koren limiter and non smooth initial condition: solution at final time t = 0.125 when using h = 0.005

out clipping) develop spurious waves. Implicit Euler features a strongly smoothed behaviour and similarly TR-BDF2 blended, due to the frequent activation of the global sensor, see Table 7. TR-BDF2 partitioned remains free of oscillations due to the activation of the local sensor, also shown in the figure, that anyway allows to develop the shock and rarefaction waves correctly and without excessive smoothing. Anyway it shows a reduction in the shock speed, as in the Burgers test, due to is non-conservative partitioning and some limited vio-



Fig. 12 Error-maximum Courant number curves for the Buckley Leverett equation with Koren limiter and non smooth initial condition

lations of the TVD property at the largest step sizes, see Table 6. TR-BDF2 flux-partitioned is again able to predict the correct shock propagation speed, as well as the correct speed of the rarefaction wave. Additionally, in Fig. 11 we report the solution at the final time when using a sufficiently small time step (h = 0.005). In this case, implicit Euler features a strong smoothing, while Crank-Nicolson, SDIRK 2(2) and TR-BDF2 do not show significant differences. The TR-BDF2 variants are all equivalent, since the sensors are never activated for such a small step size. ERK 2(2) shows a slightly less accurate solution, while ERK 3(3) is sufficiently accurate but shows a TVD violation on the shock wave front.

The accuracy curves for the Buckley–Leverett equation in Fig. 12 confirm our previous findings. Conditionally monotone methods achieve worse results than implicit Euler at coarse step sizes, due to the impact of the relevant TVD violations. ERK 2(2) and ERK 3(3) achieve levels of accuracy which are similar to the implicit methods and while violating TVD condition at $C_{max} \ge 1$ they avoid here any overflow, probably as a side effect of the Koren limiter. TR-BDF2 blended offers a seamless compromise between accuracy at fine time steps and monotonicity at coarse time steps, while TR-BDF2 partitioned obtains qualitatively very good solutions, but it is penalized in the l^{∞} norm by the wrong prediction on the propagation speed of the leading shock. TR-BDF2 flux-partitioned method instead is able to obtain the most accurate solutions at coarse step sizes with respect to the other methods here analyzed due to its conservative partitioning.

7 Conclusions

We have reviewed a general framework for the preservation of some relevant solution properties during numerical integrations with RK methods. The generality of the absolute monotonicity results proved to be of practical relevance for assessing monotonicity, positivity and strong stability of RK methods. In particular, we have analyzed the monotonicity properties of the TR-BDF2 method, that was introduced in [1] and successively reformulated and analyzed in [28]. We derived the characteristic SSP coefficient of the DIRK family associated to TR-BDF2, which expresses a CFL-like condition for monotonicity properties. We proposed two modifications, the first one based on a hybridization in time with the

implicit Euler method and the second one being an automatically partitioned RK method that tries to separate monotone and non-monotone solution components in each time step. A flux-partitioned version has been derived following [36], which allows a conservative partitioning and is thus able to capture the correct propagation speed of discontinuites, even at large Courant numbers. All these strategies attempt to enforce monotonicity properties in constant time step integrations, as commonly found in meteorology, environmental fluid dynamics, multiphase or turbulent reactive flow simulations. Both monotone strategies make use of suitably defined sensor functions, able to detect local or global violations of relevant functional bounds, thus triggering a robust integration procedure only when necessary. Thus accuracy is locally sacrificed, in order to preserve monotonicity independently of the time step and stiffness of the problem. Both strategies were assessed empirically against other RK methods on a series of benchmark problems, ranging from chemical kinetics to advection diffusion reaction equations and nonlinear conservation laws. While explicit methods are clearly superior for smaller time step sizes, the results show that the time hybridization strategy is able to guarantee a seamless compromise between accuracy at fine step sizes and monotonicity for large step sizes. Furthermore, the flux-partitioned strategy obtains relatively good accuracy even at high Courant numbers. Further research is required to identify more appropriate sensors for triggering the partitioning methods, as well as to extend the same strategies to other SSP RK methods.

Acknowledgements We would like to thank two anonymous reviewers for their constructive criticism, which has greatly helped to improve the original version of this paper. A.D.R. would like to thank Tenova S.p.A. for sponsoring his Executive Ph.D. at Politecnico di Milano and all the faculty members at the Department of Mathematics for their continuous support. L.B. acknowledges financial support from the INDAM - GNCS project 'Metodi numerici semi-impliciti e semi-Lagrangiani per sistemi iperbolici di leggi di bilancio'. Useful discussions with L. Formaggia and A. Scotti at MOX on the topics studied in this paper are kindly acknowledged.

Appendix 1: Additional Results on Monotonicity

Discrete maximum principle. The discrete maximum principle is closely related to the property of monotonicity.

Definition 7 *Discrete Maximum Principle*. The RK method (2) respects the discrete maximum principle if $\forall u \in \mathbb{R}^m$ it guarantees that

$$\min_{1 \le j \le m} u_j^0 \le u_i^{n+1} \le \max_{1 \le j \le m} u_j^0 \quad (1 \le i \le m)$$

under the assumption that for $0 < h \le \tau_0$ and $\forall u \in \mathbb{R}^m$ with components u_p

$$\min_{1 \le q \le m} u_q \le u_p + \tau_0 f_p(t, u(t)) \le \max_{1 \le q \le m} u_q, \quad (1 \le p \le m).$$
(47)

Similarly to the other properties, range boundedness may be verified under a step size restriction analogous to (5). Again following [29] and [46], we introduce two relevant sublinear functionals, denoted as *max* and *min* functional

$$\|u\|_{>} = \max_{j} u_{j} \tag{48a}$$

$$\|u\|_{<} = -\min_{j} u_{j} \tag{48b}$$

which allow us to write the assumption (47) in the form (3). By assuming the monotonicity property 1 under the functionals (48) the discrete maximum principle 7 directly follows.

Contractivity. Contractivity of numerical approximations has been extensively studied. Relevant conclusions on step size conditions for contractivity have been given in [45], while contractivity of RK for nonlinear problems was thoroughly examined in [37].

Definition 8 *Contractivity.* The RK method (2) is contractive if $\|\tilde{u}^n - u^n\| \le \|\tilde{u}^{n-1} - u^{n-1}\|$ under the assumption that

$$\|\tilde{u} - u + h(f(t, \tilde{u}) - f(t, u))\| \le \|\tilde{u} - u\| \text{ for } 0 < h \le \tau_0.$$
(49)

Usually, Definition 8 is verified under a step size restriction in the form of (5). For conditional contractivity, the *circle condition* that was originally assumed in [37] is

$$\|f(t,\tilde{u}) - f(t,u) + \rho(\tilde{u} - u)\| \le \rho \|\tilde{u} - u\|$$
(50)

It was shown later in [24] that this condition can be considered as a special form of (3). By introducing the auxiliary space $\mathbb{V} = \mathbb{R}^m \times \mathbb{R}^m$ and considering $||G_i|| = ||\tilde{g}^i - g^i||$ and $||U_i|| = ||\tilde{u}_i^n - u_i^n||$ the circle condition (49) can be reformulated as (3) in the space of perturbations. See also [46] for additional considerations on these issues.

Appendix 2: Tables of Errors and Workload in Numerical Experiments

Here we report the Tables 8, 9, 10, 11, and 12 of the errors and workload as measured during the numerical experiments to show the typical error magnitude and workload to be expected from the different methods. We dot not claim that the error-workload tables reported here are immediately relevant for the selection of step sizes or numerical methods, but we consider them as representative of the relative workload to be expected from the methods assessed.

h	IE		CN		MPRK		ROS2	
	$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _\infty$	t_{CPU}	$\ e_1\ _{\infty}$	t_{CPU}
0.0030	0.00551130	20.766	0.00000276	22.323	0.00000551	1.526	0.00004507	1.794
0.0100	0.01831771	6.142	0.00003066	6.650	0.00006086	0.450	0.00049018	0.533
0.0300	0.05449960	2.182	0.00027592	2.365	0.00053956	0.151	0.00415831	0.178
0.1000	0.17663848	0.713	0.00306899	0.769	0.00569421	0.046	0.03827381	0.054
0.3000	0.48933585	0.257	0.02772503	0.284	0.04422131	0.016	0.22248388	0.019
1.0000	1.32120559	0.108	0.34546108	0.110	0.32120559	0.005	0.98006844	0.006
	SDIRK 2(2)		TR-BDF2 (cl	ipped)	TR-BDF2 ble	ended	TR-BDF2 pa	rtitioned
	$\ e_1\ _{\infty}$	<i>t</i> _{CPU}	$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	<i>t</i> _{CPU}	$\ e_1\ _{\infty}$	<i>t</i> _{CPU}
0.0030	0.00000069	39.495	0.00000134	54.972	0.00000134	55.343	0.00000134	58.820
0.0100	0.00000766	11.764	0.00001489	16.365	0.00001489	16.324	0.00001489	17.521
0.0300	0.00006898	4.001	0.00013426	5.502	0.00013426	5.521	0.00013426	5.948
0.1000	0.00076662	1.256	0.00150218	1.731	0.00150218	1.742	0.00150218	1.895
0.3000	0.00687829	0.458	0.01376246	0.619	0.01376246	0.621	0.01376246	0.688
1.0000	0.07879441	0.174	0.17439178	0.217	0.20927048	0.259	0.17439178	0.248

Table 8 Error and workload in CPU time (s) for the Brusselator problem

h	IE		CN		ROS2		SDIRK 2(2)	
	$\ e_1\ _{\infty}$	t_{CPU}	$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	t_{CPU}	$\ e_1\ _{\infty}$	t_{CPU}
0.0025	0.02848154	33.87	0.00020777	58.55	0.00332937	13.44	0.00005193	105.48
0.0050	0.05262990	16.76	0.00083166	29.49	0.01242716	6.74	0.00020777	56.98
0.0100	0.09403148	8.77	0.00333283	14.69	0.03971792	3.36	0.00083166	31.98
0.0200	0.16387376	4.23	0.01332009	8.09	0.09985183	1.68	0.00333283	18.25
0.0241	0.18862172	5.11	0.01944843	9.29	0.12660793	1.39	0.00488201	16.07
0.0400	0.27194943	2.26	0.05011883	6.89	0.22309404	0.84	0.01332009	10.46
0.0600	0.34180965	1.58	0.09790173	4.69	0.30271373	0.53	0.02999527	7.42
0.1000	0.43671582	1.03	0.23795568	3.16	0.39800075	0.33	0.07244555	5.18
			TR-BDF2 (cl	ipped)	TR-BDF2 blended		TR-BDF2 partitioned	
			$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	t _{CPU}
0.0025			0.00010086	127.11	0.00010086	127.24	0.00010086	145.52
0.0050			0.00040372	71.08	0.00040372	70.70	0.00040372	81.14
0.0100			0.00161713	40.50	0.00161713	40.57	0.00161713	46.75
0.0200			0.00645944	22.77	0.00645944	22.79	0.00645944	26.19
0.0241			0.00945999	19.20	0.00945999	19.31	0.03744437	21.91
0.0400			0.03359402	12.78	0.05168372	13.19	0.08116569	14.42
0.0600			0.07366847	8.88	0.09263059	9.08	0.14552713	9.82
0.1000			0.15617064	5.94	0.14874170	6.02	0.25600533	6.39
					ERK 2(2)		ERK 3(3)	
					$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	t_{CPU}
0.0025					0.00041643	0.28	0.00000376	0.41
0.0050					0.00166677	0.15	0.00003014	0.20
0.0100					0.02649842	0.07	0.00024150	0.10
0.0200					∞	0.03	∞	0.05

 Table 9
 Error and workload in CPU time (s) for the advection problem with non smooth initial condition

h	IE		CN		ROS2		SDIRK 2(2)	
	$\ e_1\ _{\infty}$	t_{CPU}	$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	t_{CPU}	$\ e_1\ _{\infty}$	t_{CPU}
0.0025	0.01573010	1518.9	0.00002667	2530.8	0.00032172	498.5	0.00000463	3015.5
0.0050	0.03143022	760.3	0.00011020	1512.8	0.00127837	249.1	0.00001851	1622.0
0.0100	0.06273809	384.8	0.00043127	758.5	0.00498470	124.4	0.00007404	852.1
0.0250	0.15588305	205.0	0.00252878	401.6	0.02762922	49.5	0.00046285	401.1
0.0500	0.30832143	105.4	0.00950909	207.8	0.09207159	24.7	0.00185318	228.1
0.1000	0.60181856	67.9	0.03709329	131.1	0.27850471	12.3	0.00744174	137.5
			TR-BDF2 (cl	ipped)	TR-BDF2 ble	TR-BDF2 blended		rtitioned
			$\ e_1\ _{\infty}$	<i>t</i> _{CPU}	$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	<i>t</i> _{CPU}
0.0025			0.00001357	3708.9	0.00001357	3723.6	0.00001357	6563.7
0.0050			0.00005429	1938.8	0.00005429	1927.7	0.00005429	3357.7
0.0100			0.00021500	1033.9	0.00021500	1031.6	0.00021500	1794.4
0.0250			0.00128103	491.9	0.00128103	493.9	0.00128103	858.7
0.0500			0.00482422	279.1	0.00482422	278.3	0.00482422	485.8
0.1000			0.01833259	172.4	0.01833259	173.5	0.05210612	294.9
					ERK 2(2)		ERK 3(3)	
					$\ e_1\ _{\infty}$	<i>t</i> _{CPU}	$\ e_1\ _{\infty}$	<i>t</i> _{CPU}
0.0025					0.00003712	3.33	0.00000068	5.01
0.0050					0.00014888	1.65	0.00000536	2.48
0.0100					0.00059853	0.82	0.00003997	1.23
0.0250					∞	0.32	0.00085771	0.49
0.0500					∞	0.16	∞	0.24

 Table 10
 Error and workload in CPU time (s) for the advection diffusion reaction problem with non smooth initial condition

h	EE		IE		CN		SDIRK 2(2)	
	$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	t_{CPU}	$\ e_1\ _{\infty}$	t_{CPU}	$\ e_1\ _{\infty}$	t_{CPU}
0.0025	0.01633914	0.263	0.01862666	116.52	0.00060337	189.94	0.00012032	184.09
0.0050	0.02957002	0.132	0.03690302	62.61	0.00241135	108.59	0.00048646	104.88
0.0100	0.34347101	0.065	0.08228971	40.70	0.00929184	65.98	0.00199032	67.89
0.0200	∞	0.033	0.14741042	24.76	0.04879019	39.99	0.00889496	41.77
0.0400			0.18165711	12.53	0.12592353	22.85	0.05726634	25.45
0.0600			0.18859609	8.61	0.20632052	15.82	0.16592343	17.61
0.1000			0.22793675	5.50	0.27086617	10.27	0.28283957	11.87
	TR-BDF2 (c	lipped)	TR-BDF2 bl	ended	TR-BDF2 pa	ırtitioned	TR-BDF2 flu	x-partitioned
	$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	t_{CPU}	$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	t _{CPU}
0.0025	0.00029418	223.82	0.00029418	225.19	0.00029418	248.28	0.00029418	288.41
0.0050	0.00118202	139.69	0.00118202	139.57	0.00118202	153.96	0.00118202	161.24
0.0100	0.00467327	82.03	0.00467327	81.99	0.00467327	90.51	0.00467327	90.63
0.0200	0.01802276	49.37	0.01389498	53.08	0.06063171	54.89	0.01276669	54.00
0.0400	0.06943195	28.94	0.11429848	38.13	0.28320460	30.92	0.06092638	32.68
0.0600	0.11801805	19.07	0.15883880	28.36	0.31682264	21.84	0.08505701	21.91
0.1000	0.21500270	13.50	0.18304840	21.10	0.35791166	14.68	0.11035705	16.07
					ERK 2(2)		ERK 3(3)	
					$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	<i>t</i> _{CPU}
0.0025					0.00119048	0.531	0.00004971	0.788
0.0050					0.00567482	0.261	0.00049326	0.391
0.0100					0.02144486	0.129	0.00716732	0.193
0.0200					∞	0.066	∞	0.097

 Table 11
 Error and workload in CPU time (s) for the Burgers equation with van Leer limiter and non smooth initial condition

h	EE		IE		CN		SDIRK 2(2)	
	$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	t_{CPU}	$\ e_1\ _{\infty}$	t_{CPU}	$\ e_1\ _{\infty}$	t _{CPU}
0.0010	0.10483401	0.088	0.07528013	39.20	0.00214926	72.83	0.00049777	90.01
0.0025	0.18946104	0.034	0.11166466	20.81	0.01288210	33.19	0.00294092	47.88
0.0050	0.49347629	0.019	0.15735687	14.04	0.04169737	24.01	0.01314232	28.53
0.0075	1.87246259	0.012	0.17298411	10.71	0.24450868	16.54	0.02926612	21.78
0.0100	5.34351807	0.010	0.18461063	7.65	0.35684065	13.34	0.04865308	17.79
0.0150	6.49759606	0.009	0.22144417	5.78	2.96767661	9.69	0.35139848	13.50
0.0250	7.53465548	0.006	0.23353878	4.06	7.17081275	11.09	0.44626267	10.28
	TR-BDF2 (c	lipped)	TR-BDF2 bl	ended	TR-BDF2 pa	ırtitioned	TR-BDF2 flu	x-partitioned
	$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	<i>t</i> _{CPU}	$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	<i>t</i> _{CPU}
0.0010	0.00103587	112.20	0.00103587	112.88	0.00103587	122.67	0.00103587	111.04
0.0025	0.00648607	58.99	0.00648607	58.92	0.00648607	64.38	0.00648607	48.92
0.0050	0.02290704	35.32	0.02290704	35.86	0.02290704	38.63	0.02290704	27.93
0.0075	0.03741508	25.37	0.06422935	30.56	0.30853798	27.68	0.03741508	22.28
0.0100	0.16334990	20.31	0.14744971	30.81	0.41362307	21.70	0.04941793	18.14
0.0150	0.31177240	14.69	0.17915385	22.53	0.39766696	15.42	0.07931282	15.71
0.0250	5.28844003	11.56	0.23345479	16.30	0.47365936	12.11	0.21773710	9.51
					ERK 2(2)		ERK 3(3)	
					$\ e_1\ _{\infty}$	t _{CPU}	$\ e_1\ _{\infty}$	t _{CPU}
0.0010					0.00415082	0.178	0.00031123	0.293
0.0025					0.02955096	0.069	0.00301782	0.097
0.0050					0.16754522	0.033	0.03301197	0.048
0.0075					0.55084230	0.022	0.22910285	0.032
0.0100					0.65771595	0.018	0.49697199	0.023
0.0150					5.72351502	0.016	2.62973611	0.017
0.0250					7.32290137	0.008	5.65912546	0.011

 Table 12
 Error and workload in CPU time (s) for the Buckley–Leverett equation with Koren limiter and non smooth initial condition

References

- Bank, R., Coughran, W., Fichtner, W., Grosse, E., Rose, D., Smith, R.: Transient simulation of silicon devices and circuits. IEEE Trans. Electron Devices 32, 1992–2007 (1985)
- Bellen, A., Jackiewicz, Z., Zennaro, M.: Contractivity of waveform relaxation Runge–Kutta iterations and related limit methods for dissipative systems in the maximum norm. SIAM J. Numer. Anal. 31(2), 499–523 (1994)
- Bellen, A., Torelli, L.: Unconditional contractivity in the maximum norm of diagonally split Runge–Kutta methods. SIAM J. Numer. Anal. 34(2), 528–543 (1997)
- Bertolazzi, E.: Positive and conservative schemes for mass action kinetics. Comput. & Math. Appl. 32(6), 29–43 (1996)
- Burchard, H., Deleersnijder, E., Meister, A.: A high-order conservative Patankar-type discretisation for stiff systems of production-destruction equations. Appl. Numer. Math. 47, 1–30 (2003)
- Butcher, J., Chen, D.: A new type of singly-implicit Runge–Kutta method. Appl. Numer. Math. 34, 179–188 (2000)
- Calvo, M., de Frutos, J., Novo, J.: Linearly implicit Runge–Kutta methods for advection-diffusion-reaction equations. Appl. Numer. Math. 37, 535–549 (2001)
- Constantinescu, E., Sandu, A.: Multirate timestepping methods for hyperbolic conservation laws. J. Sci. Comput. 33, 239–278 (2007)
- 9. Dekker, K., Verwer, J.: Stability of Runge–Kutta Methods for Stiff Nonlinear Differential Equations. Elsevier, North Holland (1984)
- Duraisamy, K., Baeder, J.: Implicit scheme for hyperbolic conservation laws using nonoscillatory reconstruction in space and time. SIAM J. Sci. Comput. 29(6), 2607–2620 (2007)
- Duraisamy, K., Baeder, J., Liu, J.G.: Concepts and application of time-limiters to high resolution schemes. J. Sci. Comput. 19, 139–162 (2003)
- Ferracina, L., Spijker, M.: Stepsize restrictions for the total-variation-diminishing property in general Runge–Kutta methods. SIAM J. Numer. Anal. 42(3), 1073–1093 (2004)
- Ferracina, L., Spijker, M.: An extension and analysis of the Shu–Osher representation of Runge–Kutta methods. Math. Comput. 74(249), 201–219 (2005)
- Ferracina, L., Spijker, M.: Strong stability of singly-diagonally-implicit Runge–Kutta methods. Appl. Numer. Math. 58, 1675–1686 (2008)
- Formaggia, L., Scotti, A.: Positivity and conservation properties of some integration schemes for mass action kinetics. SIAM J. Numer. Anal. 49(3), 1267–1288 (2011)
- Giraldo, F., Kelly, J., Constantinescu, E.: Implicit–explicit formulations of a three-dimensional nonhydrostatic unified model of the atmosphere (NUMA). SIAM J. Sci. Comput. 35(5), 1162–1194 (2013)
- Gottlieb, S., Ketcheson, D., Shu, C.W.: Strong Stability Preserving Runge–Kutta and Multistep Time Discretizations. World Scientific, Singapore (2011)
- Gottlieb, S., Shu, C.: Total variation diminishing Runge–Kutta schemes. Math. Comput. 67(221), 73–85 (1998)
- Gottlieb, S., Shu, C., Tadmor, E.: Strong stability-preserving high-order time discretization methods. SIAM Rev. 43(1), 89–112 (2001)
- Hairer, E., Lubich, C., Wanner, G. (eds.): Geometric Numerical Integration, 2nd edn. Springer, Heidelberg (2006)
- Hairer, E., Nørsett, S., Wanner, G.: Solving Ordinary Differential Equations I: Nonstiff Problems, 3 corr. edn. Springer, Heidelberg (2008)
- 22. Hairer, E.: Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems, 2 rev. edn. Springer, Heidelberg (2002)
- Higueras, I.: On strong stability preserving time discretization methods. J. Sci. Comput. 21, 193–223 (2004)
- Higueras, I.: Representations of Runge–Kutta methods and strong stability preserving methods. SIAM J. Numer. Anal. 43(3), 924–948 (2005)
- Higueras, I.: Strong stability for Runge–Kutta schemes on a class of nonlinear problems. J. Sci. Comput. 57(3), 518–535 (2005)
- Higueras, I.: Strong stability for additive Runge–Kutta methods. SIAM J. Numer. Anal. 44(4), 1735–1758 (2006)
- Higueras, I., Roldán, T.: Stage value predictors for additive and partitioned Runge–Kutta methods. Appl. Numer. Math. 56, 1–18 (2006)
- Hosea, M., Shampine, L.: Analysis and implementation of TR-BDF2. Appl. Numer. Math. 20, 21–37 (1996)

- Hundsdorfer, W., Mozartova, A., Spijker, M.: Special boundedness properties in numerical initial value problems. BIT 51(4), 909–936 (2011)
- Hundsdorfer, W., Ruuth, S., Spiteri, R.: Monotonicity-preserving linear multistep methods. SIAM J. Numer. Anal. 41(2), 605–623 (2003)
- Hundsdorfer, W., Spijker, M.: Boundedness and strong stability of Runge–Kutta methods. Math. Comput. 80(274), 863–886 (2011)
- 32. Hundsdorfer, W., Verwer, J.: Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations. Springer, Heidelberg (2003)
- Ketcheson, D.: Step sizes for strong stability preservation with downwind-biased operators. SIAM J. Numer. Anal. 49(4), 1649–1660 (2011)
- Ketcheson, D., Gottlieb, S., Macdonald, C.: Strong stability preserving two-step Runge–Kutta methods. SIAM J. Numer. Anal. 49(9), 2618–2639 (2011)
- Ketcheson, D., Macdonald, C., Gottlieb, S.: Optimal implicit strong stability preserving Runge–Kutta methods. Appl. Numer. Math. 59, 373–392 (2009)
- Ketcheson, D., Macdonald, C., Ruuth, S.: Spatially partitioned embedded Runge–Kutta methods. SIAM J. Numer. Anal. 51(5), 2887–2910 (2013)
- 37. Kraaijevanger, J.: Contractivity of Runge–Kutta methods. BIT **31**, 482–528 (1991)
- Lefever, R., Nicolis, G.: Chemical instabilities and sustained oscillations. J. Theor. Biol. 30, 267–284 (1971)
- LeVeque, R.: Finite Volume Methods for Hyperbolic Problems. Cambridge University Press, Cambridge (2002)
- Liu, H., Zou, J.: Some new additive Runge–Kutta methods and their applications. J. Comput. Appl. Math. 190, 74–98 (2006)
- Macdonald, C., Gottlieb, S., Ruuth, S.: A numerical study of diagonally split Runge–Kutta methods for PDEs with discontinuities. J. Sci. Comput. 35, 89–112 (2008)
- Sandu, A.: Positive numerical integration methods for chemical kinetic systems. J. Comput. Phys. 170, 589–602 (2001)
- Shu, C.: Total-variation-diminishing time discretization. SIAM J. Sci. Stat. Comput. 9(6), 1073–1084 (1988)
- Shu, C., Osher, S.: Efficient implementation of essentially non-oscillatory shock-capturing schemes. J. Comput. Phys. 77, 439–471 (1988)
- Spijker, M.: Contractivity in the numerical solution of initial value problems. Numerische Mathematik 42(3), 271–290 (1983)
- Spijker, M.: Stepsize conditions for general monotonicity in numerical initial value problems. SIAM J. Numer. Anal. 45(3), 1226–1245 (2007)
- Sportisse, B.: An analysis of operator splitting techniques in the stiff case. J. Comput. Phys. 161, 140–168 (2000)
- Tumolo, G., Bonaventura, L.: A semi-implicit, semi-Lagrangian discontinuous Galerkin framework for adaptive numerical weather prediction. Q. J. R. Meteorol. Soc. 141, 2582–2601 (2015)
- van Veldhuizen, S., Vuik, C., Klein, C.: A note on the numerical simulation of Kleijn's benchmark problem. Report 06–15, Delft University of Technology, Dept. of Applied Mathematical Analysis (2006)
- van Veldhuizen, S., Vuik, C., Klein, C.: Comparison of ODE methods for laminar reacting gas flow simulations. Numer. Methods Partial Differ. Equ. 24(3), 1037–1054 (2008)
- Verwer, J., Spee, E., Blom, J., Hundsdorfer, W.: A second-order Rosenbrock method applied to photochemical dispersion problems. SIAM J. Sci. Comput. 20(4), 1456–1480 (1999)