MOX-Report No. 56/2015

Monotonicity, positivity and strong stability of the TR-BDF2 method and of its SSP extensions

Bonaventura, L.; Della Rocca, A.

Monotonicity, positivity and strong stability of the TR-BDF2 method and of its SSP extensions

Luca Bonaventura[1], Alessandro Della Rocca[1],[2]

October 14, 2015

[1] MOX – Modelling and Scientific Computing,
Dipartimento di Matematica, Politecnico di Milano
Via Bonardi 9, 20133 Milano, Italy
luca.bonaventura@polimi.it
alessandro.dellarocca@polimi.it

[2] Tenova S.p.A.,
Global R&D,
Via Albareto 31, 16153 Genova, Italy
alessandro.dellarocca@tenova.com

Abstract

We analyze the one-step method TR-BDF2 from the point of view of monotonicity, strong stability and positivity. All these properties are strongly related and reviewed in the common 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. Two hybrid variants of TR-BDF2 are proposed, that reduce the formal order of accuracy and maximize the absolute monotonicity radius, while keeping the native L-stability useful in stiff problems. Numerical experiments compare these different hybridization strategies with other methods commonly used in the presence of stiff and mildly stiff source terms. The results show that both 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.

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 solutions during time integrations. In many applications, for instance, the solution is required to remain non negative or to take values in a certain range or again to preserve the total variation as a function of the space variable, in case 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], [28], [35], [36]. 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 [17] and multirate approaches [7] are able to 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 [40] 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 [24]. 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 [41] as the key ingredient of an efficient semi-implicit, 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 [9], [10], [15], [18], [31], [39], that allow to derive sufficient conditions for properties like positivity, monotonicity and total variation preservation in a unified framework of an extended concept of

3

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], [12], [43]. 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 [12] and differently than [43] it is second order when critical solution properties are not violated under the selected time step size.

In Section 2, the theory of monotone and SSP methods is reviewed. The TR-BDF2 method is presented in detail in Section 3 where its absolute monotonicity property is analyzed. Two strategies to improve its monotonicity properties irrespective of time step size are described in Section 4 stemming from the classical results of absolute monotonicity. Other competitive time discretization approaches are introduced in Section 5 and interpreted under the SSP theory whenever possible. In Section 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 investigations are summarized in Section 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 [9], [15], [14], [18] and [37]. In this work we consider an initial value problem for a system of ordinary differential equations (ODEs) of type

$$u'(t) = f(t, u(t)) \quad \text{and} \quad t \in [0, T], \tag{2.1a}$$
$$u(0) = u^0. \tag{2.1b}$$

We assume that $u^0 \in \mathbb{R}^m$, $f : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}^m$ such that the problem (2.1) has a unique solution. Moreover we assume also that $\|\cdot\| : \mathbb{R}^m \to \mathbb{R}$ is a convex functional

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

for $0 \leq \lambda \leq 1$ 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, discrete maximum principle, positivity or range boundedness are all nonlinear properties that can be seen as specific realizations of some form of monotonicity. Even contractivity, a property relevant for the numerical stability of time integration methods, can be reintepreted as a form of monotonicity. The recent theory of SSP integration methods [14] shows that all these concepts are strongly related. As a consequence, they can be guaranteed during the numerical integration using the same fundamental approach. In this section we provide a synthetic presentation of this framework, that is based on the classical results in [31]. We will focus in particular on the properties of s-stages Runge-Kutta methods (RK)

$$g^i = u^n + h \sum_{j=1}^{s} a_{ij}\, f(t^n + c_j h, g^j) \quad (i = 1, \ldots, s) \qquad (2.2a)$$

$$u^{n+1} = u^n + h \sum_{i=1}^{s} b_i\, f(t^n + c_i h, g^i) \qquad (2.2b)$$

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 $j \geq i$ 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 $b = (b_1, \ldots, b_s)^\intercal$ and a column vector $c = (c_1, \ldots, c_s)$. For homogeneous initial value problems, the coefficients c_i are not relevant, so that each RK method is completely identified by its coefficients (A, b).

Definition 2.1. *Monotonicity. The RK method (2.2) is monotone with respect to the functional $\|\cdot\|$ if $\|u^n\| \leq \|u^0\|$ under the assumption that*

$$\|u + hf(t, u)\| \leq \|u\| \quad for \quad 0 < h \leq \tau_0. \qquad (2.3)$$

Related definitions also commonly found in the literature are internal monotonicity

$$\|g^i\| \leq \|u^n\| \quad \text{for} \quad 1 \leq i \leq s \qquad (2.4)$$

and external monotonicity

$$\|u^{n+1}\| \leq \|u^n\|. \qquad (2.5)$$

Assumption (2.3), commonly found in many references, see e.g. [9], [18], [20], [21], [26], [27], [29], [36] and [37], 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 property 2.1 is verified for all

$$0 < h \leq c\,\tau_0 \tag{2.6}$$

thus making the RK method *conditionally monotone*. If property 2.1 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 2.1 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. [28], [33], [37].

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 strongly nonlinear form of such problems, the ability in maintaining such native properties also in the numerical solutions is important in order to guarantee numerical stability of time integrations.

Definition 2.2. *Positivity*. *The RK method* (2.2) *is positive if whenever $u^0 \geq 0$ it guarantees that $u^{n+1} \geq 0$ under the assumption that*

$$u + hf(t, u) \geq 0 \quad for \quad 0 < h \leq \tau_0. \tag{2.7}$$

Definition 2.3. *Range boundedness*. *The RK method* (2.2) *is range bounded in $[\chi, \psi]$ if whenever $\chi \leq u^0 \leq \psi$ it guarantees that $\chi \leq u^{n+1} \leq \psi$ under the assumption that*

$$\chi \leq u + hf(t, u) \leq \psi \quad for \quad 0 < h \leq \tau_0. \tag{2.8}$$

Both these properties are usually guaranteed if a time step limitation of the form (2.6) is respected. Even though these properties are formally different from the monotonicity property 2.1, they can be equally 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

6

presentation in [25] and [39], it is useful to introduce two sublinear functional, denoted respectively as *floor* and *ceil* functional

$$\|u\|_{\lfloor \chi \rfloor} = -\min_j(\chi, u_j) \tag{2.9a}$$

$$\|u\|_{\lceil \psi \rceil} = \max_j(\psi, u_j). \tag{2.9b}$$

These functionals are not seminorms, since they both violate property $\|\lambda v\| = |\lambda| \|v\|$ for $\lambda = -1$. Using both functionals the range boundedness property 2.3 naturally follows, while setting $\chi = 0$ in the floor functional the positivity property 2.2 is recovered. As a consequence, by introducing the floor and ceil functionals it possible to recast Definitions 2.2 and 2.3 in a form similar to Definition 2.1. Thus positivity and range boundedness can just be interpreted as different forms of monotonicity.

Alternatively, positivity and range boundedness can also be considered as two alternative realizations of the discrete maximum principle.

Definition 2.4. ***Discrete Maximum Principle***. *The RK method (2.2) follows the discrete maximum principle if it guarantees that*

$$\min_j u_j^0 \leq u_j^{n+1} \leq \max_j u_j^0$$

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

$$\min_{1 \leq q \leq m} u_q \leq u_p + \tau_0 f_p(t, u(t)) \leq \max_{1 \leq q \leq m} u_q, \quad (1 \leq p \leq m). \tag{2.10}$$

Similarly to the other properties, range boundedness may be verified under a step size restriction analogous to (2.6). Again following [25] and [39], we introduce two relevant sublinear functionals, denoted as *max* and *min* functional

$$\|u\|_> = \max_j u_j \tag{2.11a}$$

$$\|u\|_< = -\min_j u_j \tag{2.11b}$$

which allow us to write the assumption (2.10) in the form (2.3). By assuming the monotonicity property 2.1 under the functionals (2.11) the discrete maximum principle 2.4 directly follows.

Contractivity of numerical approximation methods has been extensively studied. Relevant conclusions on step size conditions for contractivity have been given in [38], while contractivity of RK for nonlinear problems was thoroughly examined in [31].

Definition 2.5. _Contractivity_. _The RK method_ (2.2) _is contractive if_ $\|\tilde{u}^{n+1} - u^{n+1}\| \le \|\tilde{u}^0 - u^0\|$ _under the assumption that_

$$\|\tilde{u} - u + h(f(t, \tilde{u}) - f(t, u))\| \le \|\tilde{u} - u\| \quad for \quad 0 < h \le \tau_0. \quad (2.12)$$

Usually, Definition 2.5 is verified under a step size restriction in the form of (2.6). For conditional contractivity, the _circle condition_ that was originally assumed in [31] is

$$\|f(t, \tilde{u}) - f(t, u) + \rho(\tilde{u} - u)\| \le \rho\|\tilde{u} - u\| \quad (2.13)$$

It was shown later in [20] that this condition can be considered as a special form of (2.3). See also [39] for additional considerations on these issues.

In the framework introduced above, all these properties appear as different forms of monotonicity under a proper choice of the convex functional $\|.\|$. It is thus possible to extend the analytical results from the preservation of a specific property to that of any other related property. Every convex functional will be retained under convex combinations of single RK methods preserving it. Even though in principle this approach may not lead to sharp bounds, experience shows that the time step limits obtained from monotonicity are representative for the preservation of the other properties as well [27]. The sharper results obtained in the context of inner product norms [19], for example, did not lead to time step limits significantly different for the practical use of conditionally monotone RK methods. It is to be noticed that RK methods that are non-monotone under arbitrary functionals may indeed be conditionally monotone under inner product norms, as in [19]. For linear problems these conditions relax even further, see e.g., [38], [42].

As a consequence, it appears useful to extend the contractivity results to the other properties. To this purpose, the absolute monotonicity introduced in [31] allows to investigate necessary and sufficient conditions for contractivity of RKs in dissipative problems under sublinear functionals, including the maximum norm. In [9] this approach was extended to study monotonicity of general RK under arbitrary seminorms. One of the relevant results is that the maximal step size coefficient for monotonicity is equal to the maximal step size for contractivity. Recently these conclusions were extended to the analysis of general linear methods under arbitrary convex functionals in [39].

For our next discussion we will focus on _irreducible_ RK methods, which are the only ones practically relevant. For a definition of irreducibility see [16]. Following [31], we introduce for real ξ the quantities

$$\begin{aligned} A(\xi) &= A(I - \xi A)^{-1}, & b^{\mathsf{T}}(\xi) &= b^{\mathsf{T}}(I - \xi A)^{-1}, \\ e(\xi) &= (I - \xi A)^{-1}e, & \varphi(\xi) &= 1 + \xi b^{\mathsf{T}}(I - \xi)A^{-1}e. \end{aligned} \quad (2.14)$$

8

Definition 2.6. ***Absolute monotonicity of RK***. *An irreducible s-stage RK* (A, b) *is absolutely monotone at* $\xi \in \mathbb{R}$ *if* $A \geq 0$, $b \geq 0$, $e \geq 0$ *and* $\varphi \geq 0$ *elementwise.*

For the stability function φ, this entails that $\frac{\mathrm{d}^k \varphi}{\mathrm{d}z^k}(\xi) \geq 0$ for any $k \geq 0$, since the rationale behind Definition 2.6 lies in the Taylor expansion of some characteristic functions of a RK method, including the stability function. The quantities (2.14) form the coefficients of such expansions, see [22] and [31] for further details. These notations are useful to introduce the radius of absolute monotonicity for a RK method.

Definition 2.7. ***Radius of absolute monotonicity***. *An s-stage RK with scheme* (A, b) *and* $A \geq 0$ *and* $b \geq 0$ *is characterized by its radius of absolute monotonicity defined for all* ξ *in* $-r \leq \xi \leq 0$ *as*

$$
\begin{aligned}
R(A, b) = \sup\{r : & r \geq 0, \\
& A(\xi) \geq 0, \ b^{\mathsf{T}}(\xi) \geq 0, \ e(\xi) \geq 0, \ \varphi(\xi) \geq 0\}.
\end{aligned} \tag{2.15}
$$

In [31], two useful results are derived, that simplify the practical estimation of the absolute monotonicity radius. We introduce the incidence matrix Inc(A) as the matrix containing 0 or 1 if the corresponding element in the matrix A is $a_{i,j} = 0$ or $a_{i,j} \neq 0$ respectively.

Theorem 2.1 ([31], Theorem 4.2). *For an irreducible RK* $R(A, b) > 0$ *iff* $A \geq 0$, $b > 0$ *and* $Inc(A^2) \leq Inc(A)$.

Lemma 2.1 ([31], Lemma 4.4). *For an irreducible RK* $R(A, b) \geq r$ *iff* $A \geq 0$ *and* (A, b) *is absolutely monotone at* $\xi = -r$.

More compact definitions of the radius of absolute monotocity involve the use of matrices derived from the Butcher tableau, see e.g., [30], but for the purpose of this work we found the use of the original definition more convenient. Other classical results for the practical identification of RK properties from the corresponding Butcher tableau stem from the *stage order*, that has practical relevance to avoid the order reduction phenomenon during stiff transients.

Theorem 2.2 ([31], Theorem 8.5). *Any RK with* $A \geq 0$ *has stage order* $\tilde{p} \leq 2$. *If* $\tilde{p} = 2$ *then* A *must have a zero row.*

Lemma 2.2 ([31], Lemma 8.6). *Any RK with* $b > 0$ *has stage order* $\tilde{p} \geq \lfloor \frac{p-1}{2} \rfloor$.

Theorem 2.3 ([31], Theorem 8.3). ***Unconditional contractivity***. *The order barrier for* $R(A, b) = \infty$ *in a RK is* $p \leq 1$. *Some first order unconditionally contractive RKs are the implicit Euler, the RADAU IA and the RADAU IIA methods.*

Theorem 2.4 ([31], Corollary 8.7). *Conditional contractivity.*
The order barriers for $R(A, b) \geq 0$ under the circle condition (2.13)
are $p \leq 4$ for explicit RK and $p \leq 6$ for implicit RK.

As a consequence, step size restrictions on RK of formal order $p >$
1 are inevitable to preserve contractivity of the numerical solution.
The analysis was recently extended in [39] to a much larger class of
methods, including IMEX methods (see also [22]). This implies that
the order barriers above are inevitable for a very large class of time
discretization methods.

In recent literature [14], [15], methods satisfying condition (2.5)
for a general convex functional are called *strong-stability preserving
(SSP)*, in order to specify their ability to preserve any convex func-
tional bound. Thus they generalize classical TVD methods specifically
developed for hyperbolic conservation laws.

Definition 2.8. *Strong stability preserving (SSP). The RK method*
(2.2) *is SSP with respect to the functional $\|\cdot\|$ if $\|u^n\| \leq \|u^0\|$ under
the assumption that*

$$\|u + hf(t, u)\| \leq \|u\| \quad for \quad 0 < h \leq \tau_0. \tag{2.16}$$

The SSP coefficient *is the largest constant $c \geq 0$ such that this defini-
tion is verified for all $0 < h \leq c\tau_0$.*

The definition above closely follows Definition 2.1 for monotonicity;
indeed, these definitions are equivalent. The SSP coefficient c turns
out to be strongly related to the radius of absolute monotonicity in-
troduced in Definition 2.7.

Theorem 2.5 ([10], Theorem 3.4). *For an irreducible RK, $c = R(A, b)$.*

Recent studies focused on the search of SSP-optimal RK methods hav-
ing large SSP coefficients. While explicit SSP RKs are known since
the seminal work in [37], a search for implicit SSP RKs started only
recently in [11] and [30], 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* [29] and *diagonally split RK*
[2], [3], [34] have been investigated in the literature, with mixed suc-
cess. Starting from the framework introduced above in our following
analysis we will consider all these properties just as specific realiza-
tions 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 TR-BDF2 method, originally introduced in [1] and successively studied in [24] as a DIRK. The same method was rediscovered in [6] and has been applied also in [13] to treat the implicit terms in an additive Runge-Kutta (ARK). A semi-implicit, semi-Lagrangian reinterpretation of this method has recently been proposed in [41].

We rely on the monotonicity and contractivity results reviewed in Section 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

$$u^{n+\gamma} - \frac{\gamma}{2}hf^{n+\gamma} = u^n + \frac{\gamma}{2}hf^n \tag{3.1a}$$

$$u^{n+1} - \frac{(1-\gamma)}{(2-\gamma)}hf^{n+1} = \frac{1}{\gamma(2-\gamma)}u^{n+\gamma} - \frac{(1-\gamma)^2}{\gamma(2-\gamma)}u^n \tag{3.1b}$$

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]. This value is also the only one for which the method (3.1) is L-stable. As outlined in [24], the TR-BDF2 method can be rewritten as a DIRK method. However, instead of closely following this reference, we will first derive the DIRK family associated to (3.1) before imposing the condition on the Jacobian. The Butcher tableau for this DIRK family is then

$$
\begin{array}{c|ccc}
0 & 0 & & \\
\gamma & \frac{\gamma}{2} & \frac{\gamma}{2} & \\
1 & \frac{1}{2(2-\gamma)} & \frac{1}{2(2-\gamma)} & \frac{1-\gamma}{2-\gamma} \\
\hline
& \frac{1}{2(2-\gamma)} & \frac{1}{2(2-\gamma)} & \frac{1-\gamma}{2-\gamma}
\end{array}
$$

We remark that, by imposing the condition $a_{22} = a_{33}$, the optimal value for γ able to guarantee L-stability is readily obtained. The stability function associated to the above DIRK family is

$$\varphi(\xi) = \frac{\det(I - \xi A + \xi e b^\intercal)}{\det(I - \xi A)} = \frac{[1 + (1-\gamma)^2]\xi + 2(2-\gamma)}{2(2-\gamma)(1 - \xi\frac{\gamma}{2})(1 - \xi\frac{1-\gamma}{2-\gamma})}. \tag{3.2}$$

11

The rational polynomial defining the stability function of the DIRK family has a single zero at $\xi = \frac{2(\gamma-2)}{1+(1-\gamma)^2}$, which is always negative for all possible values of parameter γ, and two poles at $\xi = \frac{2}{\gamma}$ and $\xi = \frac{2-\gamma}{1-\gamma}$ respectively, which are always positive. While in [1] it is argued that there is a double pole at $\xi = \frac{2}{\gamma}$, this is true only for the numerical value of γ satisfying $a_{22} = a_{33}$. As shown in [24], the TR-BDF2 method is embedded in a (2,3) Runge-Kutta pair, thus allowing an efficient estimation of the time discretization error, in case step size adaptation is required. Additionally, 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., [18], [34]).

We analyze the monotonicity properties of the DIRK family generalizing the TR-BDF2 by following [9], [10], [31]. Following the definition (2.15), we find

$$
A(\xi) = \begin{bmatrix} 0 & 0 & 0 \\ \frac{\frac{\gamma}{2}}{1-\frac{\gamma}{2}\xi} & \frac{\frac{\gamma}{2}}{1-\frac{\gamma}{2}\xi} & 0 \\ \frac{1}{2(2-\gamma)(1-\frac{\gamma}{2}\xi)\beta} & \frac{1}{2(2-\gamma)(1-\frac{\gamma}{2}\xi)\beta} & \frac{1-\gamma}{(2-\gamma)\beta} \end{bmatrix}
$$

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

$$
\varphi(\xi) = \frac{[1+(1-\gamma)^2]\xi+2(2-\gamma)}{2(2-\gamma)(1-\xi\frac{\gamma}{2})\beta}
$$

where we have set $\beta = 1 - \xi(1-\gamma)/(2-\gamma)$. The main conclusion is that the radius of absolute monotonicity of TR-BDF2 is $R(A,b) = \frac{2(2-\gamma)}{1+(1-\gamma)^2}$ and that it is maximized for $\gamma = 2 - \sqrt{2}$, which is exactly the value that makes the DIRK family L-stable. With this value of the parameter γ, the absolute monotonicity radius is $R(A,b) \approx 2.414$. From [9], we conclude that this is also the step size coefficient c for conditional monotonicity of the method under arbitrary seminorms and sublinear functionals for any nonlinear problem.

4 Two unconditionally monotone variants of TR-BDF2

The results reviewed in Section 2 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 be always 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 [44].

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 be too detrimental on solution accuracy if 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

$$u^{n+\gamma} - \gamma h(1 - \frac{\alpha}{2})f^{n+\gamma} = u^n + \gamma h \frac{\alpha}{2} f^n \tag{4.1a}$$

$$u^{n+1} - \frac{(1-\gamma)h}{\alpha(1-\gamma)+1} f^{n+1} =$$
$$= \frac{\alpha(\frac{1}{\gamma} - 1) + 1}{\alpha(1-\gamma)+1} u^{n+\gamma} - \frac{\alpha}{\alpha(1-\gamma)+1} \frac{(1-\gamma)^2}{\gamma} u^n. \tag{4.1b}$$

For $\alpha = 1$ the first step is the trapezoidal rule and the second step is the BDF2 formula, thus reconstructing the original TR-BDF2 (3.1). 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 (3.1). By injecting the first step in the second one, the Butcher tableau of the hybrid TR-BDF2 method is found

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

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

$$
\begin{array}{c|ccc}
0 & 0 & & \\
\gamma & 0 & \gamma & \\
1 & 0 & \gamma & 1-\gamma \\
\hline
 & 0 & \gamma & 1-\gamma
\end{array}
$$

The stability function of the DIRK family associated to the hybrid TR-BDF2 method is thus

$$
\varphi(\xi) = \frac{1 + \left[\frac{\alpha(1-\gamma)+\gamma}{\alpha(1-\gamma)+1} - \gamma\left(1 - \frac{\alpha}{2}\right)\right]\xi}{1 - \left[\frac{1-\gamma}{\alpha(1-\gamma)+1} + \gamma\left(1 - \frac{\alpha}{2}\right)\right]\xi + \gamma\left(1 - \frac{\alpha}{2}\right)\frac{1-\gamma}{\alpha(1-\gamma)+1}\xi^2} \quad (4.2)
$$

From this expression for the case $\alpha = 1$ we recover the stability function (3.2) and for the unconditionally monotone case $\alpha = 0$ the stability function $\varphi(\xi) = 1/[1 - \xi + \gamma(1-\gamma)\xi^2]$.

The DIRK family corresponding to the hybrid TR-BDF2 method is thus entirely L-stable for every value of the parameter α, as evident from Figure 1. Additonally for $\alpha = 1$ the value $\gamma = 2 - \sqrt{2}$ corresponding to TR-BDF2 maximizes the radius of absolute monotonicity, while 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 ($Re \to 0^-$, $Im \to 0$) in Figure 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 $0 \leq \alpha \leq 1$ 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

In the *hybrid in time mode* the strategy is to switch from the default $\alpha = 1$ (TR-BDF2) to the unconditionally monotone mode $\alpha = 0$ (IE-IE) whenever a suitable sensor detects a violation of some *global* functional bound on TR-BDF2 solution. After each detected violation by the TR-BDF2 solution, the time step integration is repeated in IE-IE mode. We call this simple method **TR-BDF2 blended**, since

14
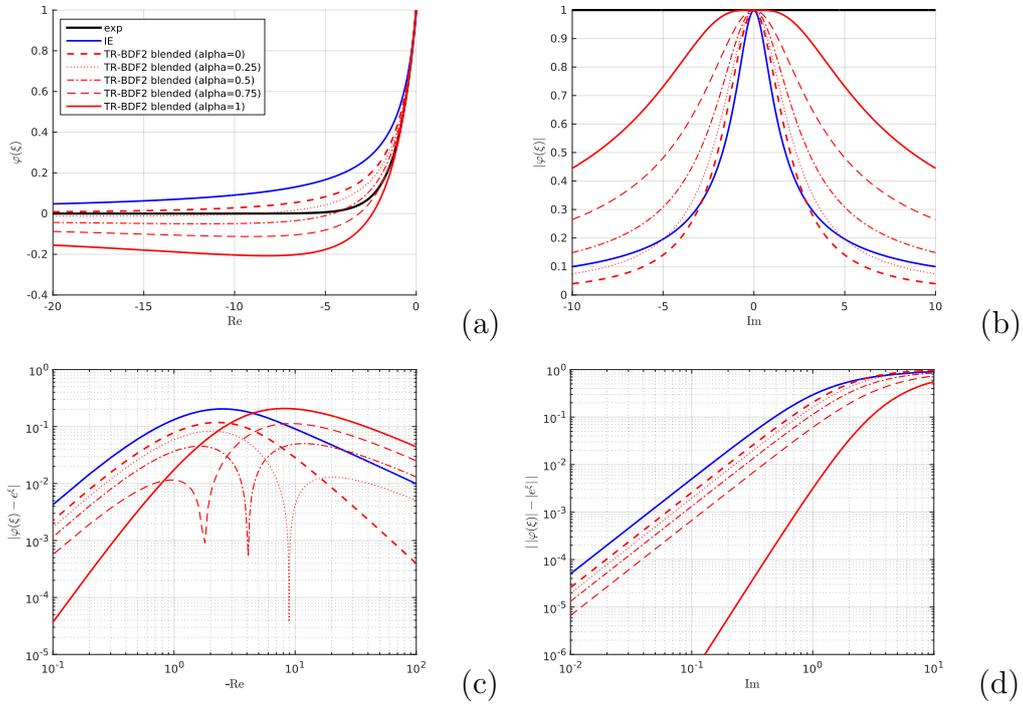
Figure 1: Hybrid TR-BDF2 in comparison with implicit Euler: a) stability functions along the negative real axis, b) modulus of the stability functions along the imaginary axis, c) absolute error functions along the negative real axis showing the asymptotic and non-asymptotic ranges, d) absolute error functions along the positive imaginary axis.
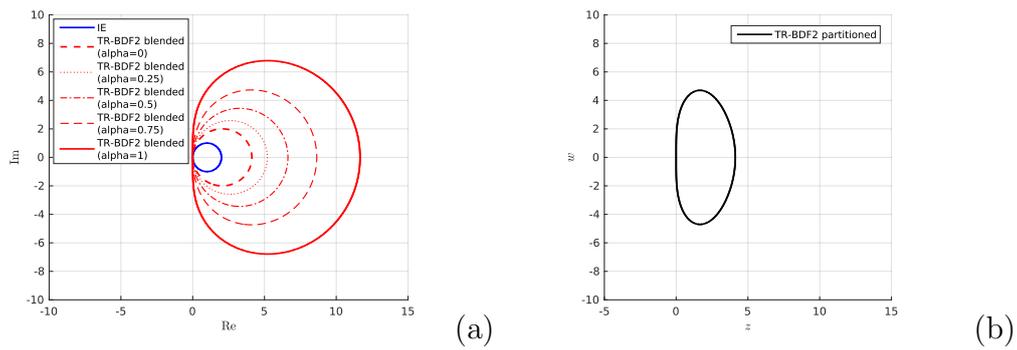
Figure 2: Stability regions (outside portions of the plane): a) stability boundaries for the hybrid TR-BDF2 and the implicit Euler methods, b) stability region of the additive Runge-Kutta corresponding to the TR-BDF2 partitioned method.

15

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

$$\sigma = s_g(u^{n+1}) = \begin{cases} 1 & \text{if } \|u^{n+1}\| \leq M \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

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 algorithm of TR-BDF2 blended with $\gamma = 2 - \sqrt{2}$ is:

1. Set $\alpha = 1$ and perform the current integration by (4.1) to find the tentative solution u^*.

2. Apply the sensor $\sigma = s_g(u^*)$.

   - If $\sigma = 1$, set $u^{n+1} = u^*$ and go to the next time step.
   - If $\sigma = 0$, set $\alpha = \sigma$, repeat the current integration by (4.1) to find the solution u^{n+1} and go to the next time step.

This basic method is A-stable, L-stable and unconditionally monotone.

4.2 TR-BDF2 partitioned

In the *hybrid in solution space mode* the switch from the default TR-BDF2 to the unconditionally monotone IE-IE mode is applied only for those solution components which are likely to produce violations of a relevant property under the assigned time step size. In order to detect this behaviour we rely on SSP theory results by applying Definition 2.1 together with the computed value of $R(A, b) \approx 2.414$ for TR-BDF2. Since this strategy exploits locality, we introduce the *local* sensor function

$$\sigma_i = s_l(u_i^{n+1}) = \begin{cases} 1 & \text{if } \|u_i^{n+1}\| \leq M_i \\ 0 & \text{otherwise.} \end{cases} \tag{4.4}$$

that detects any componentwise violation of the functional bound M_i on $\|\cdot\|$. In particular for each time step the algorithm is:

1. Perform a tentative step of forward Euler

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

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

2. Apply the sensor $\sigma_i = s_l(u_i^*)$ $(i = 1, \ldots, m)$ on the tentative solution to construct the partitioning matrix $S = diag\{\sigma_i\}$.

16

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 (4.1), construct the automatically *partitioned* RK method

$$g^i = u^n + h \sum_{j=1}^{s} [\, a_{ij} S + \tilde{a}_{ij}(I - S) \,] f(t^n + c_j h, g^j) \quad \text{(4.5a)}$$

$$u^{n+1} = u^n + h \sum_{i=1}^{s} [\, b_i S + \tilde{b}_i(I - S) \,] f(t^n + c_i h, g^i) \qquad \text{(4.5b)}$$

to find the solution u^{n+1} from stage values g^i $(i = 1, \dots, s)$ with $\gamma = 2 - \sqrt{2}$.

Thus we have effectively transformed the hybrid TR-BDF2 (4.1) into a partitioned Runge-Kutta method, in which the solution space is automatically sorted into monotone and non-monotone components by performing a tentative forward Euler step on a suitably scaled time step. It may also be interpreted as an additive Runge-Kutta method (see, e.g., [23]) from which the usual ARK stability region $\varphi(z, w) = 1 + (iwb^{\intercal} + z\tilde{b}^{\intercal})(I - iwA - z\tilde{A})^{-1}e$ for the scalar test problem $u'(t) = \lambda u(t) + i\mu u(t)$ with $u(0) = u^0$ is represented in Figure 2. In this case the unconditionally monotone component z is computed by the IE-IE method while the conditionally monotone component w is integrated using the original TR-BDF2. We call this second strategy **TR-BDF2 partitioned**. 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. Clearly, 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 $\tilde{p}$, but both strategies preserve L-stability and do not spoil the linearity of the base methods, they also preserve any linear invariant and as such they allow atomic mass conservation (see, e.g., [35]).

5   Potential competitors of TR-BDF2

In order to compare the properties of TR-BDF2 to those of other similar methods in numerical test problems, we introduce here some second order methods with similar characteristics. We will make our assessment by comparing the performance of TR-BDF2 and its two hybrid variants from Section 4 against the following methods as well as against other classic methods, such as implicit Euler $(R(A, b) = \infty)$ and Crank-Nicolson $(R(A, b) = 2)$.

5.1 SSP-optimal SDIRK 2(2)

The 2-stage SSP-optimal Runge-Kutta method of second order, namely the second order RK with the largest radius of absolute monotonicity, was discovered in [11]. Later it was confirmed in [30] to be also the optimal 2-stage second order implicit RK through extensive numerical search. The Butcher tableau of DIRK 2(2) is

$$
\begin{array}{c|cc}
\frac{1}{4} & \frac{1}{4} & \\
\frac{3}{4} & \frac{1}{2} & \frac{1}{4} \\
\hline
 & \frac{1}{2} & \frac{1}{2}
\end{array}
$$

As apparent from the tableau, this method consist of two consecutive applications of the implicit midpoint rule, that is characterized by $R(A, b) = 2$, exactly as the Crank-Nicolson method. As a consequence the radius of absolute monotonicity of the SSP-optimal SDIRK 2(2) is $R(A, b) = 4$. Moreover the method inherits all the properties of the implicit midpoint rule, being A-stable and symplectic, but it is not L-stable as TR-BDF2.

5.2 ROS2

Among the potential competitors of TR-BDF2 we also include the Rosenbrock method proposed in [8] and later applied in atmospheric chemistry problems, see e.g. [35], [45]. The Butcher tableau is

$$
\begin{array}{c|cc|cc}
0 & 0 & & \gamma & \\
1 & 1 & 0 & -2\gamma & \gamma \\
\hline
 & \frac{1}{2} & \frac{1}{2} &
\end{array}
$$

with $\gamma = 1 + \frac{1}{\sqrt{2}}$. In addition to L-stability, from numerical experiments in [45] 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 [35] this observation was extended to the first two derivatives of the stability function. However, from the framework in Section 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 Section 6 we found an intrinsic resiliency against violations of the TVD property, even from initial conditions of limited regularity.

5.3 Modified Patankar Runge-Kutta

A suitable 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 [12]. 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.

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 Section 4 against the other methods described in Section 5. The range of the test cases covers 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 implicit RKs was solved using MATLAB's `fsolve` with a tolerance of `tol` $= 10^{-10}$, except when otherwise stated. All computations were performed on a single Intel® Core™i5-2540M (2.60 GHz) on a laptop with 4 GB RAM running Linux kernel `3.13.0-24-generic`. We dot not claim that the error-workload curves reported here are immediately relevant for the selection of step sizes or numerical methods, but we consider them as representative of the relative workload expected from the methods assessed.

6.1 0-D chemical model problem: the Brusselator

As a first tes,t we use 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 Section 4 with global and local positivity sensors built from the floor norm (2.9) 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

$$\|e_i\|_\infty = \max_{n=1,\ldots,N_t} |u_i^n - \tilde{u}_i^n| \tag{6.1}$$

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 in [32]:

$$u_1' = -k_1 u_1 \tag{6.2a}$$
$$u_2' = -k_2 u_2 u_5 \tag{6.2b}$$
$$u_3' = k_2 u_2 u_5 \tag{6.2c}$$
$$u_4' = k_4 u_5 \tag{6.2d}$$
$$u_5' = k_1 u_1 - k_2 u_2 u_5 + k_3 u_5^2 u_6 - k_4 u_5 \tag{6.2e}$$
$$u_6' = k_2 u_2 u_5 - k_3 u_5^2 u_6 \tag{6.2f}$$

since this form allows to write the right hand side in a form suitable to MPRK. In particular we follow the procedure in [12], which in this case can be applied, since the stoichiometric matrix has proper rank ([12], 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 Figure 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. MPRK shows similar accuracy at same step sizes, while it outperforms all the other methods in terms of workload, being it the only explicit method. 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.
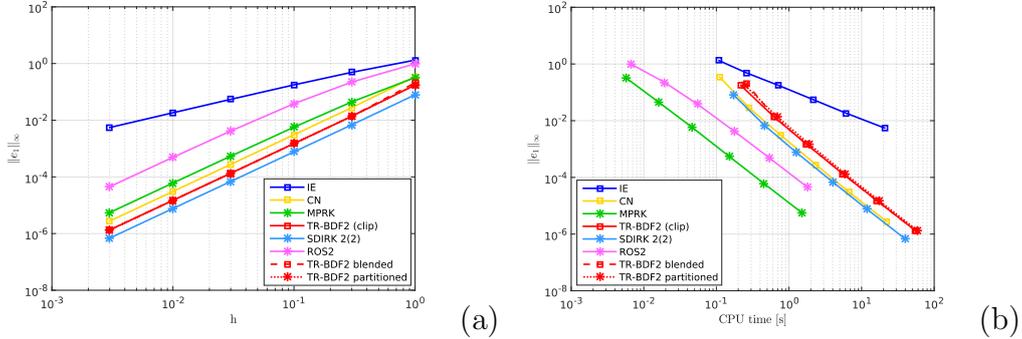
Figure 3: Original Brusselator problem: *a)* error-stepsize curves, *b)* error-workload curves.

6.2   1-D advection problem

As a first PDE case we consider the advection equation

$$u_t + vu_x = 0, \quad 0 \le x \le 1 \tag{6.3}$$

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 $x_i = i\Delta x$, $i = 1, \ldots, N_x$ 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

$$u(0, x) = \begin{cases} 1 & \text{if } |x - 0.5| < 0.25 \\ 0 & \text{otherwise.} \end{cases} \tag{6.4}$$

For this problem ,the explicit Euler method is stable under the well known condition $\text{Cou} = \frac{|a|\Delta t}{\Delta x} \le 1$. 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 (2.1), 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 $t = T$

$$\|e_i\|_{t=T}^\infty = \max_{j=1,\ldots,N_x} \left| u_{i,j}^T - \tilde{u}_{i,j}^T \right| \tag{6.5}$$

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 (6.3) and $\tilde{u}_{i,j}^T$ is the reference solution at the same point in space and time.

21

Furthermore, in order to assess any violation of the TVD property, we monitored also the TV-space l^∞-time seminorm for species i

$$\|TV_i\|_\infty = \max_{n=1,...,N_t} \sum_{j=1}^{N_x} \left| u_{i,j+1}^n - \tilde{u}_{i,j}^n \right|. \qquad (6.6)$$

Here, the TR-BDF2 blended from Section 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 $\psi = 1$. The discontinuous initial condition (6.4) 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 Sections 3 and 5, see Figure 4 and Table 1. Additionally, ROS2 is never TVD, but the violations are always limited for any step size and globally it does not show the typical crisis of conditionally monotone methods. This can be interpreted in light of Section 5. TR-BDF2 blended from Section 4 never violates TVD property, while its clipped version shows the usual crisis at large time steps. The possible explanation from Figure 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 $h = 0.06$.
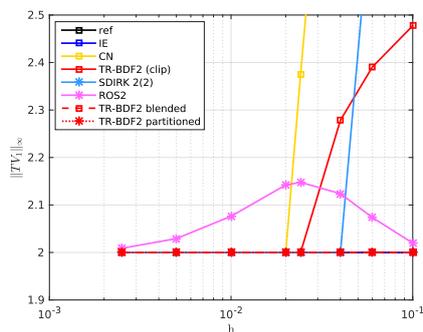
Figure 4: Advection problem with non smooth initial condition: TV-space l_∞-time seminorm for the single species.

22
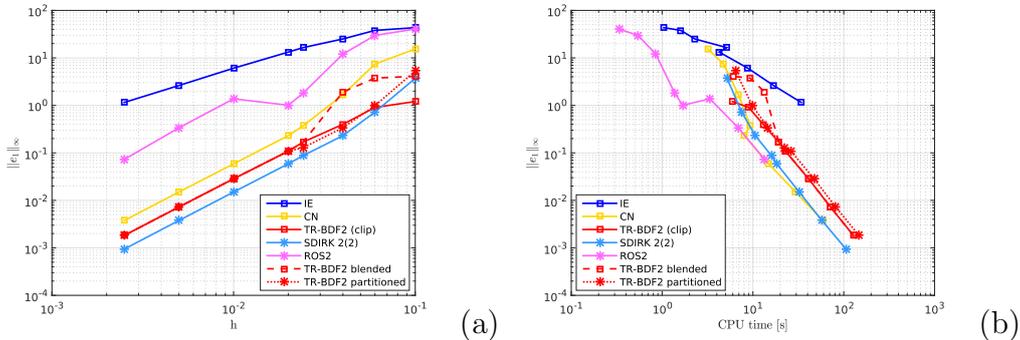
Figure 5: Advection problem with non smooth initial condition: *a)* error-stepsize curves, *b)* error-workload curves.

The accuracy results in Figure 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 Figures 4 and 5. 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, also evidenced from the order reduction. 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   1-D advection diffusion reaction problem

As a representative case for chemical transport of reacting species we introduce the advection diffusion reaction problem of three species

$$u_t + v u_x = D u_{xx} + f(u), \quad 0 \leq x \leq 1. \tag{6.7}$$

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]

$$f(u)_1 = -\frac{u_1 u_2}{u_1 + 1} \tag{6.8a}$$

$$f(u)_2 = \frac{u_1 u_2}{u_1 + 1} - k u_2 \tag{6.8b}$$

$$f(u)_3 = k u_2 \tag{6.8c}$$

23

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 $u_1 + u_2 + u_3$. We (6.7) for $0 \leq t \leq 1$ 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. The diffusion term is approximated by central finite differencing and it is naturally contractive. The initial conditions for the three species are

$$u_1(0, x) = \begin{cases} 9.98 & \text{if } |x - 0.5| < 0.25 \\ 0 & \text{otherwise.} \end{cases} \tag{6.9a}$$

$$u_2(0, x) = \begin{cases} 2 & \text{if } |x - 0.4| < 0.2 \\ 0 & \text{otherwise.} \end{cases} \tag{6.9b}$$

$$u_3(0, x) = \begin{cases} 1 & \text{if } |x - 0.7| < 0.25 \\ 0 & \text{otherwise.} \end{cases} \tag{6.9c}$$

In our assessment, we use a tolerance level $\texttt{tol} = 10^{-8}$ for the nonlinear solver in implicit methods and the reference solution is obtained from MATLAB's $\texttt{ode15s}$ 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 $\chi = 0$.
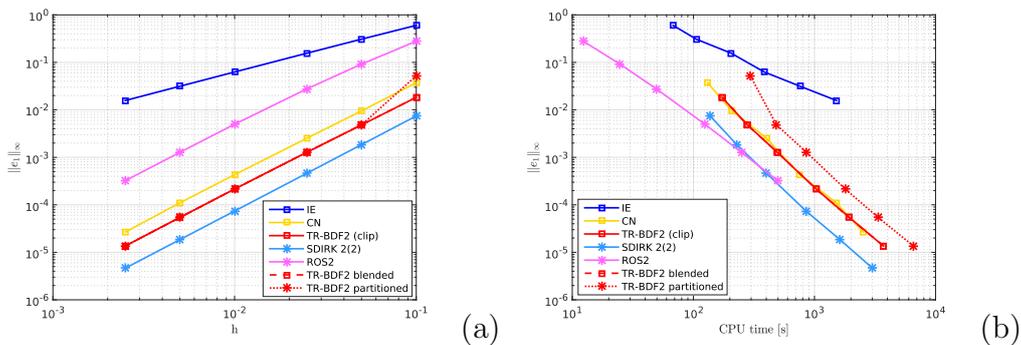
Figure 6: Advection diffusion reaction problem with non smooth initial condition: a) error-stepsize curves, b) error-workload curves.

The accuracy curves in Figure 6 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.

24

TR-BDF2 clipped and the blended variant are almost indistinguishable, while the partitioned version features slightly larger computational times due to the partitioning step.
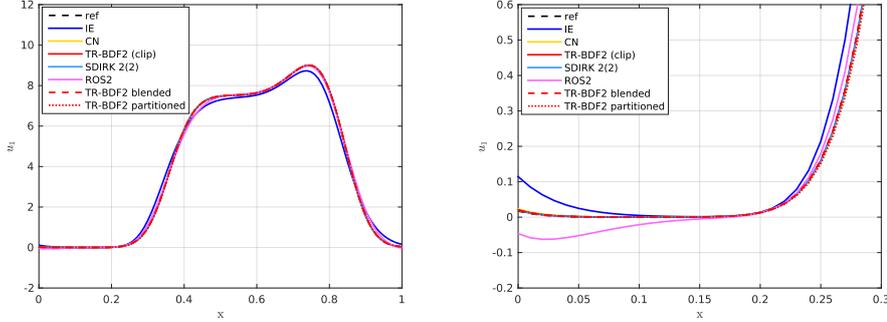
Figure 7: Advection diffusion reaction problem with non smooth initial condition: *a)* solution for $u_1(t = T)$ when using $h = 0.100$, *b)* close-up in the region of positity violation from ROS2.

A sample solution for this problem is shown in Figure 7, where a close-up shows the typical positivity violation on u_1 from ROS2. Violations of the TVD property are reported in Figure 8 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.
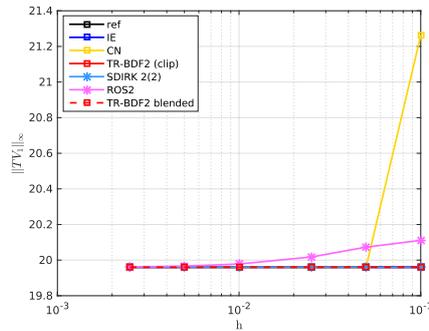
Figure 8: Advection diffusion reaction problem with non smooth initial condition: TV-space l_∞-time seminorm for the species u_1.

6.4   1-D conservation laws

We complete our assessment by considering two well known hyperbolic conservation laws, see e.g. [33] for a more detailed discussion. The

first is the Burgers equation

$$u_t = -f(u)_x = -\left(\frac{1}{2}u^2\right)_x, \quad 0 \le x \le 1. \tag{6.10}$$

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

$$u(0,x) = \frac{1}{2} + \frac{1}{4}sin(2\pi x). \tag{6.11}$$

The second is the Buckley-Leverett equation

$$u_t = -f(u)_x = -\left(\frac{u^2}{u^2 + \frac{1}{3}(1-u)^2}\right)_x, \quad 0 \le x \le 1. \tag{6.12}$$

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

$$u(0,x) = \begin{cases} \frac{1}{2} & \text{if } x \le 0.5 \\ 0 & \text{otherwise.} \end{cases} \tag{6.13}$$

Both equations are here solved with periodic boundary conditions and discretized by a high resolution finite volume method using flux limiters, see e.g. [28], [33]. More specifically, for the Burgers equation we adopt the *van Leer* limiter

$$\Psi(\theta) = \frac{\theta + |\theta|}{1 + |\theta|} \tag{6.14}$$

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

$$\Psi(\theta) = \max \left\{0 \; ; \; \min \left\{2 \; ; \; \frac{2}{3} + \frac{1}{3}\theta \; ; \; 2\theta\right\}\right\}. \tag{6.15}$$

Due to the flux limiters, strictly speaking the the Jacobian is not defined. Rather than approximating it by a finite difference discretization, we again exclude ROS2 from our assessment, since it is known a priori that it performs poorly on hyperbolic conservation laws.

In the numerical tests, we measured the l^∞ norm in space (6.5) as well as the *TV* seminorm in space (6.6). 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 `tol`$=10^{-10}$. 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 (2.9) 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. These choices follows from the initial conditions (6.11)

and (6.13). 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 as TVD.
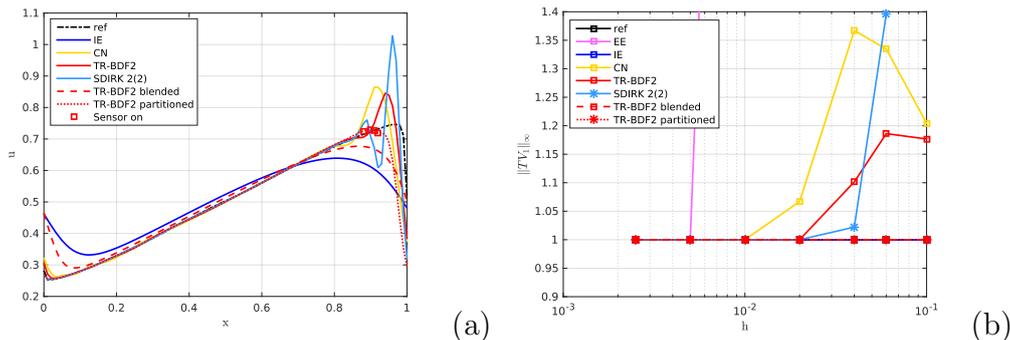
Figure 9: Burgers equation with van Leer limiter and smooth initial condition: a) solution at final time $t = 1$ when using $h = 0.1$, b) TV-space l_∞-time seminorm.

In Figure 9 we report the solution at the final time step for the coarsest step size $h = 0.1$, corresponding to about $Cou = 7.5$. While implicit Euler is able to maintain the TVD property 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. 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 approximation for the shock amplitude, but it features also a reduction in the propagation speed, probably due to the smoothing of the imaginary parts for the points integrated in IE-IE mode (see Figure 1). Significantly, both variants always remain TVD, even at $Cou = 7.5$. The other methods show TVD violations with the usual critical step size progression, see Figure 9. The accuracy results for the Burgers equation are reported in Figure 10. Interesting behaviour appears at coarse stepsizes where all the methods collapse about at the same accuracy of implicit Euler. TR-BDF2 blended realizes a smooth adaption from the monotone implicit Euler accuracy to the TR-BDF2 asymptotic curve. It preserves always monotonicity, while Crank-Nicolson, SDIRK 2(2) and TR-BDF2 clipped violate it, as evident from Figure 9. The error curves from TR-BDF2 partitioned follows tha same behaviour even

though the l^∞ error norm is penalized from the behaviour at the leading shock.
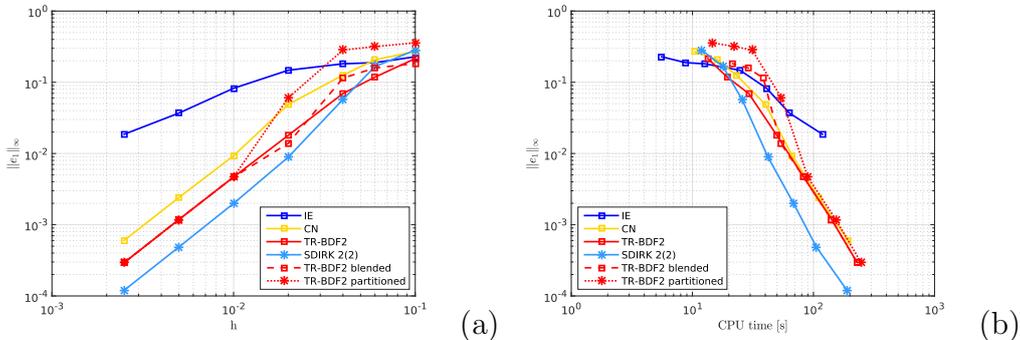
Figure 10: Burgers equation with van Leer limiter and smooth initial condition: a) error-stepsize curves, b) error-workload curves.

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 Figure 11. Again all conditionally monotone methods develop oscillation on the trailing shock, while only SDIRK 2 (2) develops a stable rarefaction wave, where Crank-Nicolson and TR-BDF2 (again without clipping) develop additional waves. Implicit Euler and TR-BDF2 blended feature a strongly smoothed behaviour, while TR-BDF2 partitioned remains free of oscillations due to the activation of the local sensor that allows to develop the shock and rarefaction waves. Anyway it shows a reduction in the shock speed, similarly as in the Burgers test, and some lmited violations of the TVD property at the largest step sizes, see Figure 11.

The accuracy curves for the Buckley-Leverett equation in Figure 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. 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 advection speed of the leading shock.

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

28
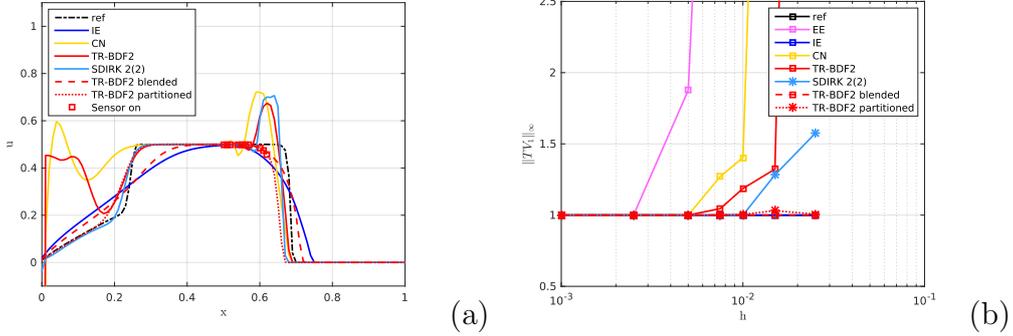
Figure 11: Buckley-Leverett equation with Koren limiter and non smooth initial condition: *a*) solution at final time $t = 0.125$ when using $h = 0.025$, *b*) TV-space l^{∞}-time seminorm.
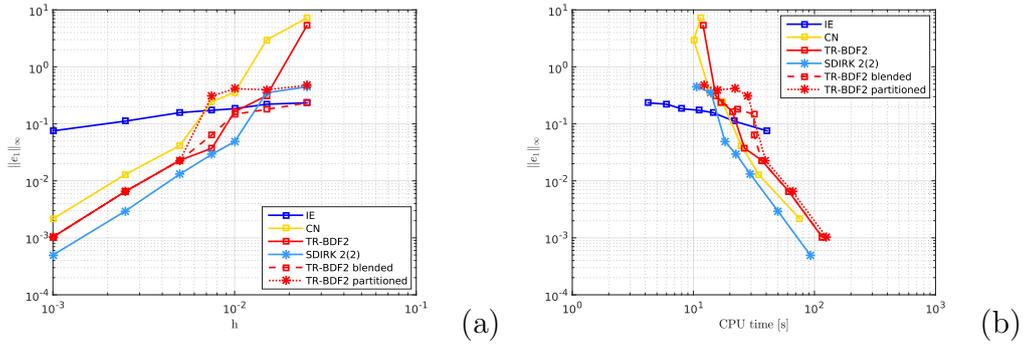
Figure 12: Buckley Leverett equation with Koren limiter and non-smooth initial condition. *Left*: Error-stepsize. *Right*: Error-workload.

to be of practical relevance for assessing contractivity, monotonicity, positivity and strong stability of RK methods. In particular we analyzed the monotonicity properties of the TR-BDF2 method, that was introduced in [1] and successively reformulated and analyzed in [24]. 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. Both strategies attempt to enforce monotonicity properties in constant time step integrations, as commonly found in meteorology, environmental fluid dynamics or turbulent reactive flow simulations. Both monotone strategies make use of sensor functions

able to detect local or global violations of suitable functional bounds, thus triggering a robust integration procedure when necessary to maintain monotonicity. In this way accuracy, is sacrificed locally 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 zero dimensional chemical rectors to advection diffusion reaction equations and nonlinear conservation laws. The results show that the time hybridization strategies are able to guarantee a seamless compromise between accuracy at fine step sizes and monotonicity at coarse step sizes, while the partitioned strategy obtains promising results penalized only by a reduced shock advection speed at high CFL. Further research may be useful to identify more appropriate sensors for triggering the partitioning method, as well as to extend the two strategies to SSP-optimal RK methods such as SDIRK 2(2).

8 Appendix

8.1 Relevant properties for chemical kinetics problems

Ordinary differential equations in the form (2.1) arise in the context of chemical kinetics when the reaction term is modeled by the *mass action law*. By ignoring the presence of additional terms in advection-diffusion-reaction PDEs, we briefly review here the reaction term.

Considering a chemical system having N_s species interacting in N_r chemical reactions

$$\sum_{q=1}^{N_s} l_{pq} u_q \xrightarrow{k_p} \sum_{q=1}^{N_S} r_{pq} u_q$$
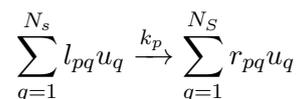
the evolution of the molar concentrations is described by the mass action law:

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t) = \dot{\omega} \quad \text{and} \quad t \geq t^0, \tag{8.1a}$$

$$u(t^0) = u^0. \tag{8.1b}$$

The source term is $\dot{\omega} = Q\omega(u)$ where Q is the matrix of stoichiometric coefficients

$$Q = R - L \in \mathbb{R}^{N_s \times N_r} \quad \text{with:} \quad R = [r_{pq}] \quad L = [l_{pq}]$$

30

and $\omega \in \mathbb{R}^{N_r}$ represents the vector of reaction rates that are usually expressed following the exponential form of the Arrhenius law

$$\omega_p(u) = k_p \prod_{q=1}^{N_s} (u_q)^{l_{pq}} \quad \text{for} \quad p = 1, \dots, N_r.$$

The relevant physical properties for the solution to (8.1) are:

- *Conservation of atomic mass*: the mass of single atomic species such as C, O and N forming the chemical species remains constant, since atomic species are conserved during chemical reactions. Algebraically this means that if $e \in \mathrm{Ker}(Q^\intercal)$ is a linear invariant of (8.1) and from $\mathrm{rank}(Q) = N_s - n$ we know that there are n linear invariants, then it is possible to collect the vectors defining the null space of $Q^\intercal$ in the columns of the matrix $A \in \mathbb{R}^{N_s \times n}$. As a consequence the solution of (8.1) must satisfy

$$A^\intercal u(t) = A^\intercal u^0 = const \quad \text{for} \quad t \geq t^0. \tag{8.2}$$

- *Positivity*: the concentrations are physical quantities that are bounded in the range of significant values $0 \leq u \leq 1$. By splitting the production and destruction terms in (8.1)

$$\frac{\mathrm{d}}{\mathrm{d}t} u(t) = P(u) - D(u)u \quad \text{and} \quad t \geq t^0, \tag{8.3a}$$

$$u(t^0) = u^0. \tag{8.3b}$$

where the production and destruction terms are

$$P(u) = R\,\omega(u) \quad \text{and} \quad D(u) = \mathrm{diag}\left[\frac{L(i, \cdot)\,\omega(u)}{u_p}\right].$$

This form ensures that $D_{ii}(u)$ are polynomials due to the functional form of the reaction rates. This implies that if all concentrations are nonnegative except $u_p = 0$ then

$$\frac{\mathrm{d}}{\mathrm{d}t} u_p(t) = P_p(u) \geq 0 \implies u(t) \geq 0 \quad \text{whenever} \quad u(0) \geq 0.$$

While conservation of linear invariants is automatic in general linear methods, positivity is more difficult to achieve. In practical applications positivity is usually enforced by a *clipping* step, in which all negative solution components are explicitly set to zero. Clipping alters the conservation of mass, since the error is introduced in a single direction only, namely adding mass to the system. While for short term computations under tight tolerances this is generally acceptable,

for long time integrations the mass added may be detrimental to solution accuracy. Moreover, as evident from Section 2, positivity is more than a constraint for physical significance of a numerical solution, since it is strongly related to nonlinear stability properties of numerical methods.

8.2 Tables of TV seminorm in numerical experiments

Here below we report the tables of the TV seminorm measured in numerical experiments to show that TV violations closely follow the expected behaviour from the absolute monotonicity radius.

h	Cou	ref	IE	CN	$TR\text{-}BDF2(clip)$
0.0025	0.25	2.00000000	2.00000000	2.00000000	2.00000000
0.0050	0.50	2.00000000	2.00000000	2.00000000	2.00000000
0.0100	1.00	2.00000000	2.00000000	2.00000000	2.00000000
0.0200	2.00	2.00000000	2.00000000	2.00000000	2.00000000
0.0241	2.41	2.00000000	2.00000000	2.37516991	2.00000000
0.0400	4.00	2.00000000	2.00000000	3.33333333	2.27858017
0.0600	6.00	2.00000000	2.00000000	4.06243821	2.39070772
0.1000	10.00	2.00000000	2.00000000	5.21857423	2.47739160
		$SDIRK2(2)$	$ROS2$	$TR\text{-}BDF2(blend)$	$TR\text{-}BDF2(part.)$
0.0025		2.00000000	2.00877086	2.00000000	2.00000000
0.0050		2.00000000	2.02925347	2.00000000	2.00000000
0.0100		2.00000000	2.07630970	2.00000000	2.00000000
0.0200		2.00000000	2.14215613	2.00000000	2.00000000
0.0241		2.00000000	2.14775690	2.00000000	2.00000000
0.0400		2.00000000	2.12378933	2.00000000	2.00000000
0.0600		2.76800000	2.07354078	2.00000000	2.00114309
0.1000		3.73260435	2.01991743	2.00000000	2.00000000

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

h	Cou	ref	IE	CN	$TR\text{-}BDF2(clip)$
0.0025	0.025	19.96000000	19.96000000	19.96000000	19.96000000
0.0050	0.050	19.96000000	19.96000000	19.96000000	19.96000000
0.0100	0.100	19.96000000	19.96000000	19.96000000	19.96000000
0.0250	0.200	19.96000000	19.96000000	19.96000000	19.96000000
0.0500	0.241	19.96000000	19.96000000	19.96000000	19.96000000
0.1000	1.000	19.96000000	19.96000000	21.26167041	19.96000000
		$SDIRK2(2)$	$ROS2$	$TR\text{-}BDF2(blend)$	$TR\text{-}BDF2(part.)$
0.0025		19.96000000	19.96178624	19.96000000	19.96000000
0.0050		19.96000000	19.96634023	19.96000000	19.96000000
0.0100		19.96000000	19.97806568	19.96000000	19.96000000
0.0250		19.96000000	20.01749991	19.96000000	19.96000000
0.0500		19.96000000	20.07270798	19.96000000	19.96000000
0.1000		19.96000000	20.11087432	19.96000000	19.96000000

Table 2: $\|TV\|_\infty$ for the advection diffusion reaction problem with non smooth initial condition.

h	Cou_{max}	ref	IE	CN	$TR\text{-}BDF2(clip)$
0.0025	0.188	1.00000000	1.00000000	1.00000000	1.00000000
0.0050	0.375	1.00000000	1.00000000	1.00000000	1.00000000
0.0100	0.750	1.00000000	1.00000000	1.00000000	1.00000000
0.0200	1.500	1.00000000	1.00000000	1.06702208	1.00000000
0.0400	3.000	1.00000000	1.00000000	1.36710392	1.10154954
0.0600	4.500	1.00000000	1.00000000	1.33460354	1.18618300
0.1000	7.500	1.00000000	1.00000000	1.20438152	1.17675243
		EE	$SDIRK2(2)$	$TR\text{-}BDF2(blend)$	$TR\text{-}BDF2(part.)$
0.0025		1.00000000	1.00000000	1.00000000	1.00000000
0.0050		1.00000000	1.00000000	1.00000000	1.00000000
0.0100		3.47978819	1.00000000	1.00000000	1.00000000
0.0200		∞	1.00000000	1.00000000	1.00000000
0.0400		∞	1.02205723	1.00000000	1.00000000
0.0600		∞	1.39639398	1.00000000	1.00000000
0.1000		∞	1.85272884	1.00000000	1.00000000

Table 3: $\|TV\|_\infty$ for the Burgers equation with van Leer limiter and smooth initial condition.

h	Cou_{max}	ref	IE	CN	$TR\text{-}BDF2(clip)$
0.0010	0.2206	1.00000000	1.00000000	1.00000000	1.00000000
0.0025	0.5514	1.00000000	1.00000000	1.00000000	1.00000000
0.0050	1.1007	1.00000000	1.00000001	1.00000000	1.00000000
0.0075	1.6543	1.00000000	1.00000000	1.27084161	1.04519501
0.0100	2.2007	1.00000000	1.00000000	1.40310073	1.18671375
0.0150	3.3010	1.00000000	1.00000000	8.09702136	1.32271889
0.0250	5.5037	1.00000000	1.00000000	16.39900045	12.48391471
		EE	$SDIRK2(2)$	$TR\text{-}BDF2(blend)$	$TR\text{-}BDF2(part.)$
0.0010		1.00000000	1.00000000	1.00000000	1.00000000
0.0025		1.00000000	1.00000000	1.00000000	1.00000000
0.0050		1.87710600	1.00000000	1.00000000	1.00000000
0.0075		6.74722895	1.00000000	1.00000000	1.00000000
0.0100		21.35210300	1.00000000	1.00000001	1.00495631
0.0150		21.10782320	1.28222058	1.00000000	1.03092403
0.0250		19.65284356	1.57372343	1.00000000	1.00663983

Table 4: $\|TV\|_\infty$ for the Buckley-Leverett equation with Koren limiter and non smooth initial condition.

Acknowledgements

References

[1] R.E. Bank, W.M. Coughran, W. Fichtner, E.H. Grosse, D.J. Rose, and R.K. Smith. Transient simulation of silicon devices and circuits. *IEEE Transactions on Electron Devices*, 32:1992–2007, 1985.

[2] A. Bellen, Z. Jackiewicz, and M. Zennaro. Contractivity of waveform relaxation Runge-Kutta iterations and related limit methods for dissipative systems in the maximum norm. *SIAM Journal of Numerical Analysis*, 31(2):499–523, 1994.

[3] A. Bellen and L. Torelli. Unconditional contractivity in the maximum norm of diagonally split Runge-Kutta methods. *SIAM Journal of Numerical Analysis*, 34(2):528–543, 1997.

[4] E. Bertolazzi. Positive and conservative schemes for mass action kinetics. *Computers & Mathematics with Applications*, 32(6):29–43, 1996.

[5] H. Burchard, E. Deleersnijder, and A. Meister. A high-order conservative Patankar-type discretisation for stiff systems of production-destruction equations. *Applied Numerical Mathematics*, 47:1–30, 2003.

[6] J. Butcher and D. Chen. A new type of singly-implicit Runge-Kutta method. *Applied Numerical Mathematics*, 34:179–188, 2000.

[7] E.M. Constantinescu and A. Sandu. Multirate timestepping methods for hyperbolic conservation laws. *Journal of Scientific Computing*, 33:239–278, 2007.

[8] K. Dekker and J.G. Verwer. *Stability of Runge-Kutta Methods for Stiff Nonlinear Differential Equations*. Elsevier-North Holland, Amsterdam, 1984.

[9] L. Ferracina and M.N. Spijker. Stepsize restrictions for the Total-Variation-Diminishing property in general Runge-Kutta methods. *SIAM Journal of Numerical Analysis*, 42(3):1073–1093, 2004.

[10] L. Ferracina and M.N. Spijker. An extension and analysis of the Shu-Osher representation of Runge-Kutta methods. *Mathematics of Computation*, 74(249):201–219, 2005.

[11] L. Ferracina and M.N. Spijker. Strong stability of singly-diagonally-implicit Runge-Kutta methods. *Applied Numerical Mathematics*, 58:1675–1686, 2008.

[12] L. Formaggia and A. Scotti. Positivity and conservation properties of some integration schemes for mass action kinetics. *SIAM Journal of Numerical Analysis*, 49(3):1267–1288, 2011.

[13] F.X. Giraldo, J.F. Kelly, and E.M. Constantinescu. Implicit-explicit formulations of a three-dimensional nonhydrostatic unified model of the atmosphere (NUMA). *SIAM Journal of Scientific Computing*, 35(5):1162–1194, 2013.

[14] S. Gottlieb, D. Ketcheson, and C.-W. Shu. *Strong stability preserving Runge-Kutta and multistep time discretizations*. World Scientific, Singapore, 2011.

[15] S. Gottlieb, C.W. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM Review*, 43(1):89–112, 2001.

[16] E. Hairer, S.P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*. Springer-Verlag, Berlin Heidelberg, 3rd corr. edition, 2008.

[17] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*. Springer-Verlag, Berlin Heidelberg, 2nd rev. edition, 2002.

[18] I. Higueras. On strong stability preserving time discretization methods. *Journal of Scientific Computing*, 21:193–223, 2004.

[19] I. Higueras. Monotonicity for Runge-Kutta methods: inner product norms. *Journal of Scientific Computing*, 24(1):97–117, 2005.

[20] I. Higueras. Representations of Runge-Kutta methods and strong stability preserving methods. *SIAM Journal of Numerical Analysis*, 43(3):924–948, 2005.

[21] I. Higueras. Strong stability for Runge-Kutta schemes on a class of nonlinear problems. *Journal of Scientific Computing*, 57(3):518–535, 2005.

[22] I. Higueras. Strong stability for additive Runge-Kutta methods. *SIAM Journal of Numerical Analysis*, 44(4):1735–1758, 2006.

[23] I. Higueras and T. Roldán. Stage value predictors for additive and partitioned Runge-Kutta methods. *Applied Numerical Mathematics*, 56:1–18, 2006.

[24] M.E. Hosea and L.F. Shampine. Analysis and implementation of TR-BDF2. *Applied Numerical Mathematics*, 20:21–37, 1996.

[25] W. Hundsdorfer, A. Mozartova, and M.N. Spijker. Special boundedness properties in numerical initial value problems. *BIT*, 51(4):909–936, 2011.

[26] W. Hundsdorfer, S.J. Ruuth, and R.J. Spiteri. Monotonicity-preserving linear multistep methods. *SIAM Journal of Numerical Analysis*, 41(2):605–623, 2003.

[27] W. Hundsdorfer and M.N. Spijker. Boundedness and strong stability of Runge-Kutta methods. *Mathematics of Computation*, 80(274):863–886, 2011.

[28] W. Hundsdorfer and J. Verwer. *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*. Springer-Verlag, Berlin Heidelberg, 2003.

[29] D.I. Ketcheson, S. Gottlieb, and C.B. Macdonald. Strong stability preserving two-step Runge-Kutta methods. *SIAM Journal of Numerical Analysis*, 49(9):2618–2639, 2011.

[30] D.I. Ketcheson, C.B. Macdonald, and S. Gottlieb. Optimal implicit strong stability preserving Runge-Kutta methods. *Applied Numerical Mathematics*, 59:373–392, 2009.

[31] J.F.B.M. Kraaijevanger. Contractivity of Runge-Kutta methods. *BIT*, 31:482–528, 1991.

[32] R. Lefever and G. Nicolis. Chemical instabilities and sustained oscillations. *Journal of Theoretical Biology*, 30:267–284, 1971.

[33] R.J. LeVeque. *Finite Volume Methods for Hyperbolic Problems*. Cambridge University Press, 2002.

[34] C.B. Macdonald, S. Gottlieb, and S.J. Ruuth. A numerical study of diagonally split Runge-Kutta methods for PDEs with discontinuities. *Journal of Scientific Computing*, 35:89–112, 2008.

[35] A. Sandu. Positive numerical integration methods for chemical kinetic systems. *Journal of Computational Physics*, 170:589–602, 2001.

[36] C.W. Shu. Total-Variation-Diminishing time discretization. *SIAM Journal on Scientific and Statistical Computing*, 9(6):10731084, 1988.

[37] C.W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77:439–471, 1988.

[38] M.N. Spijker. Contractivity in the numerical solution of initial value problems. *Numerische Mathematik*, 42(3):271–290, 1983.

[39] M.N. Spijker. Stepsize conditions for general monotonicity in numerical initial value problems. *SIAM Journal of Numerical Analysis*, 45(3):1226–1245, 2007.

[40] B. Sportisse. An analysis of operator splitting techniques in the stiff case. *Journal of Computational Physics*, 161:140–168, 2000.

[41] G. Tumolo and L. Bonaventura. A semi-implicit, semi-Lagrangian, DG framework for adaptive numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, DOI: 10.1002/qj.2544, 2015.

[42] J.A. van de Griend and J.F.B.M. Kraaijevanger. Absolute monotonicity of rational functions occurring in the numerical solution of initial value problems. *Numerische Mathematik*, 49(4):413–424, 1986.

[43] S. van Veldhuizen, C. Vuik, and C.R. Klein. A note on the numerical simulation of Kleijn's benchmark problem. Report 06-15, Delft University of Technology, Dept. of Applied Mathematical Analysis, 2006.

[44] S. van Veldhuizen, C. Vuik, and C.R. Klein. Comparison of ODE methods for laminar reacting gas flow simulations. *Numerical Methods for Partial Differential Equations*, 24(3):1037–1054, 2008.

[45] J.G. Verwer, E.J. Spee, J.G. Blom, and W. Hundsdorfer. A second-order Rosenbrock method applied to photochemical dispersion problems. *SIAM Journal of Scientific Computing*, 20(4):1456–1480, 1999.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

55/2015   Fumagalli, A.; Zonca, S.; Formaggia, L.
Advances in computation of local problems for a flow-based upscaling in fractured reservoirs

54/2015   Canuto, C.; Nochetto, R. H.; Stevenson, R.; Verani, M.
Adaptive Spectral Galerkin Methods with Dynamic Marking

53/2015   Menafoglio, A; Grujic, O.; Caers, J.
Universal kriging of functional data: trace-variography vs cross-variography? Application to forecasting in unconventional shales

49/2015   Ghiglietti, A.; Ieva, F.; Paganoni, A.M.
Statistical inference for stochastic processes: two sample hypothesis tests

50/2015   Grillo, A.; Guaily, A.; Giverso, C.; Federico, S.
Non-Linear Model for Compression Tests on Articular Cartilage

51/2015   Ballarin, F.; Faggiano, E.; Ippolito, S.; Manzoni, A.; Quarteroni, A.; Rozza, G.; Scrofani, R.
Fast simulations of patient-specific haemodynamics of coronary artery bypass grafts based on a POD–Galerkin method and a vascular shape parametrization

52/2015   Giverso, C.; Scianna, M.; Grillo, A.
Growing Avascular Tumours as Elasto-Plastic Bodies by the Theory of Evolving Natural Configurations

47/2015   Colombo, M. C.; Giverso, C.; Faggiano, E.; Boffano,C.; Acerbi, F.; Ciarletta, P.
Towards the personalized treatment of glioblastoma: integrating patient-specific clinical data in a continuous mechanical model

48/2015   Ambrosi, D.; Pettinati, V.; Ciarletta, P.
Active stress as a local regulator of global size in morphogenesis

46/2015   Giverso, C.; Verani, M.; Ciarletta P.
Emerging morphologies in round bacterial colonies: comparing volumetric versus chemotactic expansion